Directed Evolution of Glycan-Binding Proteins

By

Elizabeth M. Ward

B.S. Chemistry
University of Utah, 2014

SUBMITTED TO THE MICROBIOLOGY GRADUATE PROGRAM IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN BIOLOGY
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

SEPTEMBER 2022

Signature of Author: _____
Microbiology Graduate Program
August 30, 2022

Certified by: _____
Barbara Imperiali
Class of 1922 Professor of Biology and Chemistry

Accepted by: _____
Jacquin C. Niles
Professor of Biological Engineering
Co-Director, Microbiology Graduate Program

Directed Evolution of Glycan-Binding Proteins

By

Elizabeth M. Ward

Submitted to the Microbiology Graduate Program on August 30, 2022 in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Biology

Abstract

Glycan-binding proteins (GBPs) are commonly used reagents for the study of glycans. They do not require specialized equipment or time-consuming experimental methods, making them widely used tools for basic research and clinical applications. Existing glycan recognition reagents,antibodies and lectins, are limited, and discovery or creation of reagents with novel specificities is time consuming and difficult.

This thesis details the generation of novel GBPs from a small, hyperthermostable DNA binding protein by directed evolution. A yeast surface display method for evolution of GBPs was developed and used to generate GBPs for the recognition of mammalian glycans sialic acid and the cancer-associated disaccharide Thomsen-Friedenreich (TF) antigen. Characterization of these proteins shows them to have specificities and affinities on par with currently available lectins. The proteins can be functionalized to create reagents that prove useful for glycoprotein blotting and cell staining applications.

Carbohydrate-protein interactions are often low affinity. Naturally occurring GBPs often oligomerize to make multivalent interaction with glycan ligands, increasing the avidity of the interaction. Fusion of the evolved GBPs to the coiled-coil trimerization domain of the lectin surfactant protein D (SP-D) leads to the formation of a trimeric GBP. These trimers are properly folded, stable, and have increased binding affinity compared to monomeric GBPs. Generation of trimeric Sso7d-based GBPs is a strategy for increasing the functional affinity of the evolved proteins, thereby making the proteins useful for a wider range of applications.

The overall goal is to create GBPs for glycans with no existing GBPs for their study. One area that can benefit from more GBP reagents is bacterial glycobiology. Many pathogens have glycans involved in virulence. One such organism is *Campylobacter jejuni.* The N-linked protein glycosylation pathway in *Campylobacter jejuni* is needed for pathogenicity of the organism, as loss of glycosylation decreases adhesion and invasion to intestinal epithelial cells and differentially modulates inflammatory responses in a gut-immune co-culture model. Future application of the developed GBP evolution platform toward bacterial glycans will have great impact on the field of bacterial glycobiology through powerful tools for studying the interactions of human pathogens, commensals and symbionts and their hosts together with novel diagnostic and analytical reagents.

Thesis supervisor: Barbara Imperiali
Title: Class of 1922 Professor of Biology and Chemistry

Acknowledgements

First, I would like to thank Professor Barbara Imperiali for her mentorship and support over the past six years. It has been a pleasure and privilege to be in your lab. I admire your scientific rigor and commitment to excellence. Thank you for allowing me to spread my wings and work on such an interesting project.

Second, I would like to thank the members of the Imperiali lab past and present, and Dr. Cristina Zamora in particular. You took on a fresh new graduate student who knew nothing and molded me into a scientist. I owe a huge part of my success to you and can never thank you enough for not only being a great mentor, but a wonderful friend. Leah, Greg, Alyssa, Hannah, Christine, Megan, Soumi, and Hugh, thank you for your endless help and support and being great labmates through a rough few years of COVID craziness. A huge shoutout to Megan is due for proofreading and editing nearly all the chapters of this thesis and being a great co-author on the review article comprising Chapter 1. And of course, a big thank you to past and present lab admins Elizabeth Fong and Meg Rheault. The lab would not function as smoothly as it does without your efforts.

Third, I would like to thank the Microbiology Graduate Program for giving me the opportunity to follow my dreams and come to MIT. Some days I still can't believe I made it here. The Micro Program would not function without the efforts of our wonderful admin Jacquie Carota. Thank you for keeping everyone organized and functioning and being one of the friendliest people I have ever met.

And finally, I would like to thank my family and friends. To my husband Jordan, thank you for supporting me through the good and the bad times these past few years in Boston. You have always been willing to listen to me practice my presentations, look at my figures, and give me pep talks when I thought I couldn't do it. You are the best and I love and appreciate you so much. To Mom and Dad, thank you for supporting me in everything I do. I know Dad has always wanted me to follow in his footsteps and become a doctor, and I hope being a PhD instead of an MD is close enough. To my twin Sarah, you have been a lifelong best friend and are the kindest, best person I know. And to all my other friends and family in Utah, Georgia, and Boston, thank you for your love and support.

Table of Contents

List of Figures

Chapter 1: Strategies and tactics for the development of selective glycan-binding proteins

This chapter has been adapted and reprinted with permission from:

Ward, E.M., Kizer, M.E., Imperiali, B. (2021). Strategies and Tactics for the Development of Selective Glycan-Binding Proteins, *ACS Chem. Biol. 16*, 1795-1813. doi.org/10.1021/acschembio.0c00880.

Contributions:

Introduction written by Elizabeth Ward and Megan Kizer
Applications of glycan-binding proteins written by Elizabeth Ward
Glycan-binding proteins written by Elizabeth Ward and Megan Kizer
Glycan protein interactions written by Elizabeth Ward
Strategies and tactics in protein engineering – methodology written by Megan Kizer
Current scope of lectin engineering written by Elizabeth Ward
Current scope of CBM engineering written by Elizabeth Ward
Current scope of antibody engineering written by Megan Kizer
Conclusions written by Elizabeth Ward and Megan Kizer

1.1 Abstract

Glycans influence all biological processes, disease states and pathogenic interactions. Glycan-binding proteins (GBPs) 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 to improve the properties of existing GBPs or to engineer novel specificities and potential applications. In this thesis, the de novo generation of GBPs from a non-GBP scaffold protein is explored using yeast surface display. This chapter reviews the current scope of proteins that bind carbohydrates and the engineering methods that have been applied to enhance affinity, selectivity, and specificity of these proteins.

1.2 Introduction

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.

Many research groups focus on diverse glycan-recognition methods, 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.[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 pre-processing. 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.

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. In this thesis, the de novo generation of GBPs from a non-GBP scaffold protein is explored using yeast surface display. This chapter serves to introduce the current state of protein binders that recognize carbohydrates and the engineering methods to develop superior GBPs. 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. 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.

1.3 Applications of glycan-binding proteins

GBPs are routinely utilized in many different ways (**Figure 1-1**). 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]

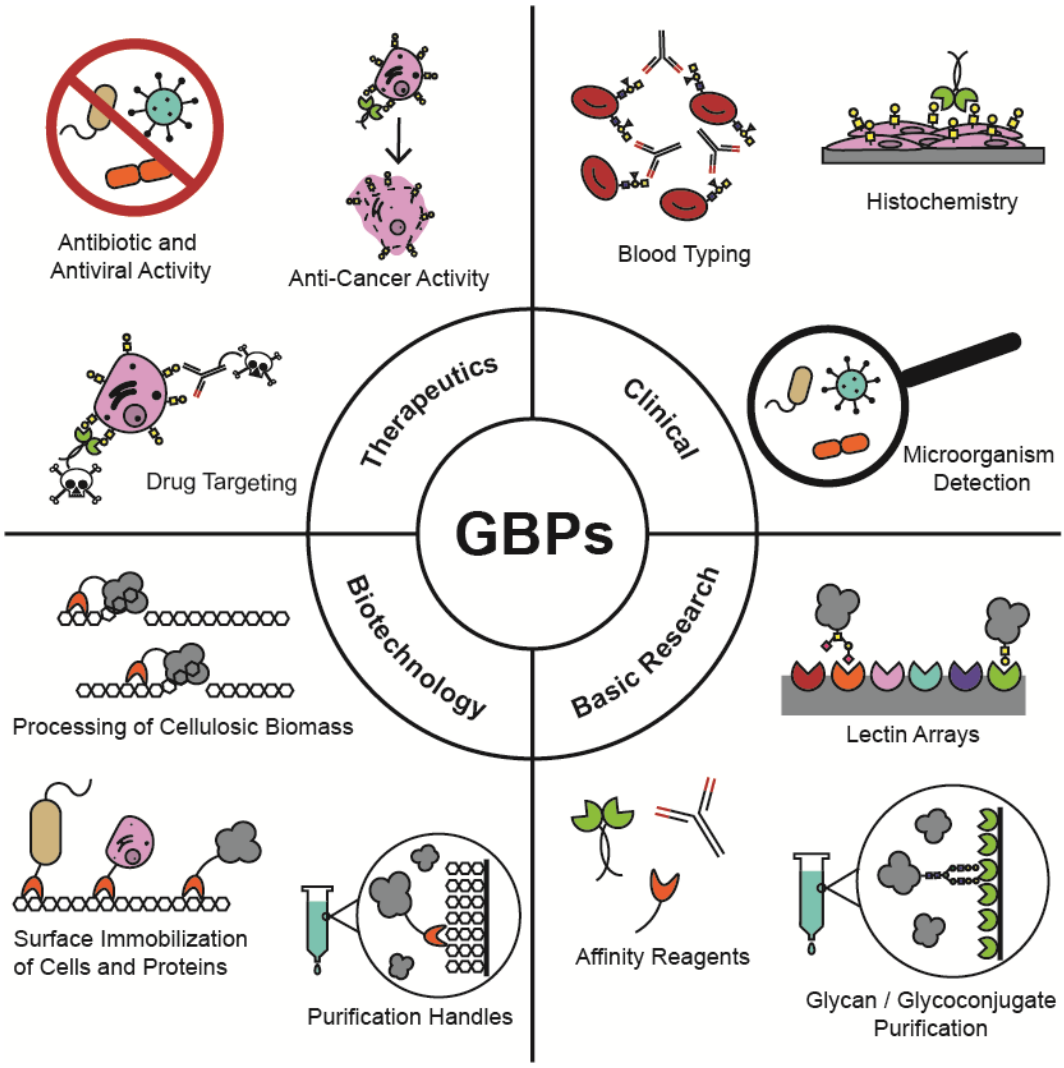


Figure 1-1. Applications of glycan-binding proteins (GBPs).
GBPs find many uses in therapeutic, clinical, biotechnological and basic research applications. Some of the prominent methods in which they have been utilized are depicted for each application area.

1.4 Glycan-binding proteins

There are a number of different types of proteins that recognize carbohydrates. The most represented GBPs in glycobiology can be categorized into three major groups: lectins, carbohydrate-binding modules (CBMs), and adaptive immune proteins (antibodies and variable lymphocyte receptors). Lectins describe a more general class of proteins that bind carbohydrates, while CBMs represent the carbohydrate-recognition domain of a larger processing enzyme, and

adaptive immune proteins include proteins with variable regions that evolve to bind glycan antigens through immune system function. Though differentiated by their classic descriptors, all three classes exhibit the commonality of binding glycans.

1.4.1 Lectins

Lectins are proteins that preferentially bind carbohydrates. 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.[25] 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.[25] 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.[26, 27] Lectins also suffer from broad specificity toward carbohydrate epitopes, making data interpretation difficult. Regardless, many lectins have been identified and their binding characteristics defined.[28]

1.4.2 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

enzymatic domain by either trafficking the enzyme to specific regions of the carbohydrate substrate, by increasing the concentration of the enzyme in the vicinity of the carbohydrate substrate, or by disrupting the polysaccharide structure to allow for easier enzyme access.[29-32] CBMs fall into three types.[33] 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.[34] They have also been grouped into seven families based on structure.[33] 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.[33]

1.4.3 Antibodies

Antibodies (Abs) are part of the mammalian immune system and form the foundation for the adaptive response to foreign antigens, including carbohydrate antigens. Although the mammalian immune system also requires lectin-type GBPs (selectins, galectins, and siglecs) to function, these proteins are not part of an adaptive system.[35, 36] Engineering efforts on these immune-related proteins are therefore discussed in Section 1.9. Ab proteins contain a constant (Fc) region and two variable (Fv) domains composed of heavy (H) and light (L) chains (**Figure 1-5A**). The Fv domains provide clonal diversity for specific antigen recognition; the Ab repertoire includes more than 10^{12} different variants in humans.[37] 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.[38, 39] Anti-glycan immune responses are less T-cell dependent, as polysaccharides do not always recruit T-cell assistance for B-cell maturation.[40, 41] 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 1-5A**). 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.[42] 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.[41] 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.

1.5 Glycan-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.[43] 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.[44] 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.[43] 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.[45] 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.[46] 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.[47] 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. GBP binding is highly dependent on the presentation mode of the glycan, as most GBPs preferentially bind densely-glycosylated glycoconjugates and cell surfaces over free monosaccharides and small oligosaccharides.[9] The gain in avidity from interaction of GBPs with multivalent carbohydrate ligands is known as the "cluster glycoside effect".[48] 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.[48] This effect does not require homogenous glycans either. The glycans presented on cell surfaces and even on single glycoproteins and glycolipids are diverse. There is evidence that these mixed glycans interact to form clustered saccharide patches, whereby heterogenous glycans are spatially organized in such a way that multiple may interact with a GBP.[49] As such, incorporating multivalency is frequently employed in GBP engineering.

1.6 Strategies and tactics in protein engineering – methodology

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.[50] 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 1-2A**).[51, 52] 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 1-2B**). 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,[53] bacterial,[54] or yeast[55]) or phage particle,[56] 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[57, 58] and mRNA display.[59] To effectively enrich 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 1-2C**).[60, 61]

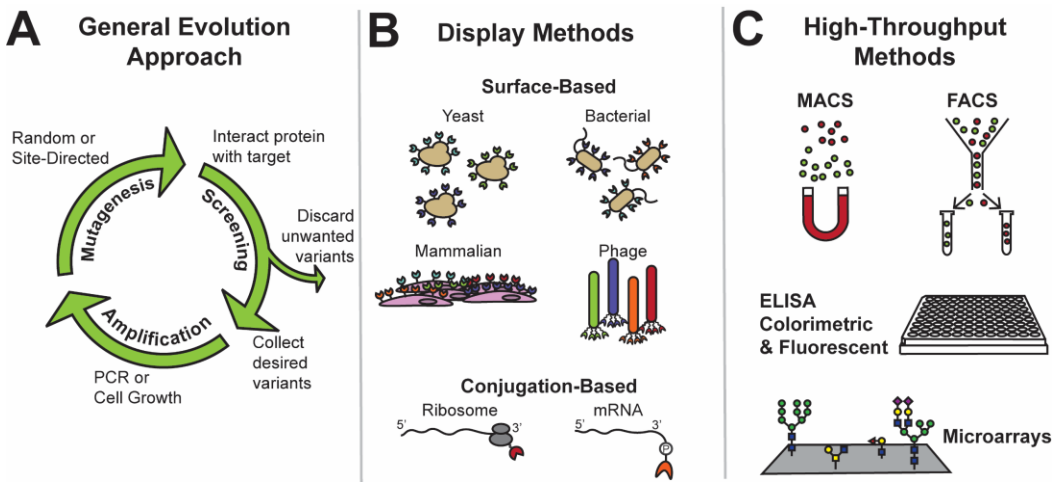


Figure 1-2. Evolutionary and screening methods.
A) The directed-evolution process includes an iteration of three main stages: mutagenesis, screening, and amplification. B) Many display methods have been developed for linking phenotype to genotype in directed evolution experiments. C) High throughput screening enhances the feasibility of directed evolution by efficient separation or detection of binders to intended targets.

Microarray technologies have revolutionized the way in which GBP screening is performed.[62-64] They have been employed in many studies involving lectins, viral protein interactions, immunological investigations, anti-glycan antibodies, and vaccine development.[60, 65-69] 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.[65, 66, 68, 69] The chemical combinations for immobilization are diverse and have been summarized in recent reviews.[65] 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 611 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 to generate a library with over 300 mono- to eicosaccharides.[70] 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 (microarrays consisting of biologically isolated and functionalized glycans) 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.[71, 72] 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.[67]

With various glycan arrays available, cross-comparisons of some of the more prominently utilized arrays have been employed to investigate glycan determinant diversity and validate array binding and consistency across different platforms.[70, 73-75] These studies reveal important considerations for GBP analysis using glycan arrays. While GBPs with known ligands were shown to recognize their reported substrates across various arrays, several factors led to discrepancies in weak binding events across platforms. The major influences include linkage type and glycan

density, while factors such as reducing end form (ring open or closed), presentation geometry (flat or spherical), slide type, and number of washes also play a role in generating inconsistencies across platforms.[74, 75] These differences in weak-binding events should therefore not be overlooked, as they may actually represent more physiologically relevant interactions. Improvements in chemical conjugation and the breadth of glycans sampled are therefore of great importance. One microarray development employs a chemoenzymatic approach to access N-linked glycosyl-asparagines that better mimic N-linked glycan epitopes.[76] Another conjugates glycans to polymers and peptides for clustered glycan presentation. Progress in multiplexing with Luminex® beads,[77] DNA-based arrays,[78, 79] and cell-based arrays have also been investigated.[80, 81] Still, the presentation of glycans on the solid array support is not entirely representative of their biological context. Therefore, GBP-glycan interactions observed on an array should be validated by an orthogonal, biologically-relevant method.

Furthermore, the current array landscape does not accurately represent the glycan complexity and diversity of both mammals and microbes. It has been estimated that there are 3,000 unique glycan species on glycoproteins and glycolipids in mammals,[82] meanwhile the bacterial glycome is even more diverse.[83] Yet many of these glycans are difficult to obtain in large quantities for biochemical assays, many biologically-isolated glycans are not sequence defined, and synthetic routes can be resource intensive and low-yielding. Efforts to overcome these limitations are in progress, and utilization of complementary arrays could provide useful information about cross-kingdom GBP-glycan interactions. For example, the MPS array has a diversity similar to that of the CFG array and the Imperial College London/Feizi array, is highly complementary.[70] Despite these limitations, the microarray remains an incredibly powerful tool for glycan-protein interaction analysis and GBP screening.

1.7 Current scope of lectin engineering

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®.[84] For other lectin engineering reviews, I direct the reader to Hu et al. and Hirabayashi and Arai[85, 86]

1.7.1 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.[87] 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 to create the Gln-Pro-Asp motif of C-type Gal-binding lectins.[88] 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 1-3A)**.[89] 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).[90] 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.[91] 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.[92] 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.[91] 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 1-3B)**.[93] 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.[94] 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.[95] 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 1-3B)**.[96, 97] Another mutant, Glu86Ala, lost binding affinity for all glycans tested except those containing a 3'-sulfo-Galβ1-4GlcNAc structure.[98]

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 1-3A**).[99] 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.[90] Phage display has also been used to pan a library of the α2-3Sia-specific plant lectin *Maackia amurensis* hemagglutinin (MAH) with human erythrocytes.[100] 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).[53] 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.

1.7.2 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.[101, 102] Sialic acid selections were carried out using the high throughput method known as ribosome display **(Figure 1-2B)**.[58, 103] Agarose beads modified with the glycoprotein Fetuin were used as the selection bait as this protein contains α2-6-sialylated tri-antennary 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 1-3C)**. 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.

1.7.3 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 (referred to as SRC for sia-recognition) 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.[104] 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 1-3C**).[105] 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.[106] 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).[107] 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 1-2C**).

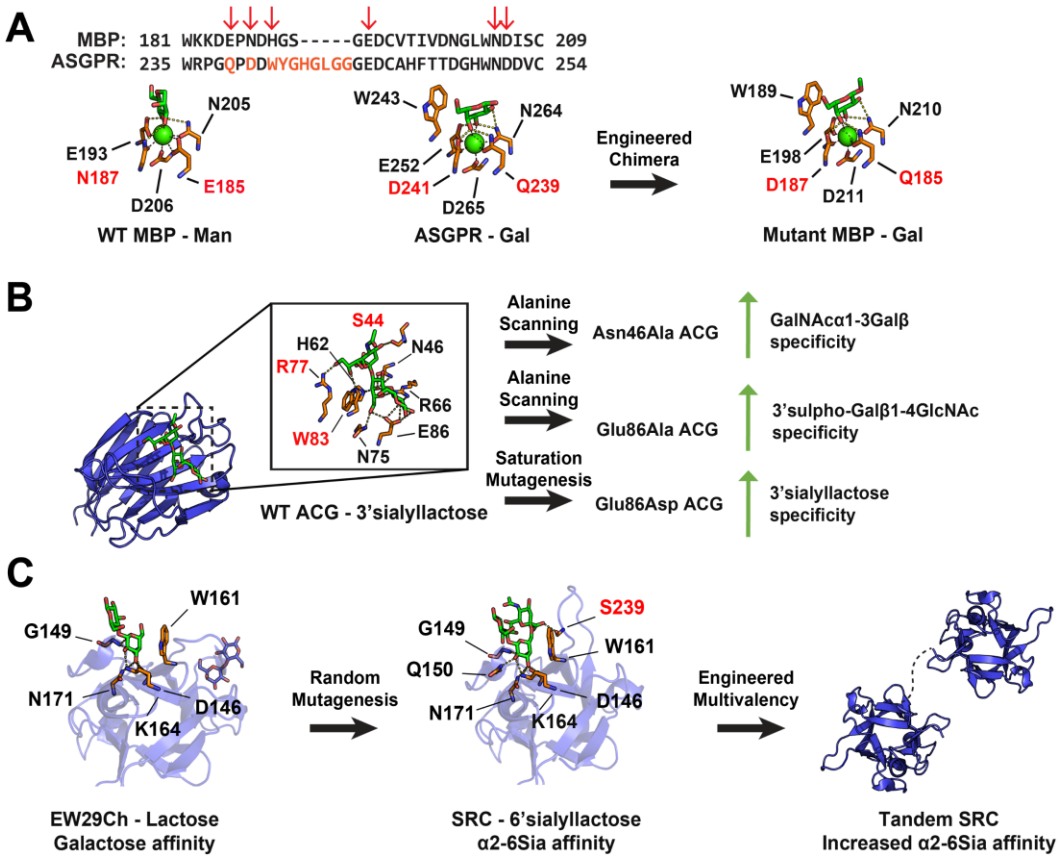


Figure 1-3. Examples of lectin engineering.
A) Sequence alignment of rat mannose binding protein (MBP) and rat asialoglycoprotein receptor (ASGPR) binding sites. Orange residues indicate mutations made in mutant MBP, and red arrows indicate residues shown in structures. Wild-type (WT) MBP structure with 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 sialic recognition 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.

1.8 Current scope of carbohydrate-binding module engineering

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.[108] 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.[33] 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.[109]

1.8.1 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.[110] 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 1-4A**). 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.[111] 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 1-4A**). 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.[112] 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.[113] 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 1-4B)**.[114] 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.[115] This phage library has also been screened for binders to xyloglucan, a plant cell wall polysaccharide which lacks appropriate reagents for its study.[116] 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 1-4C)**.[117]

1.8.2 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.[118] 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.[119] 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.[116] Error-prone PCR introduced random mutations

throughout the length of the gene to generate a new phage display library.[120] 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 1-4C)**. 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.

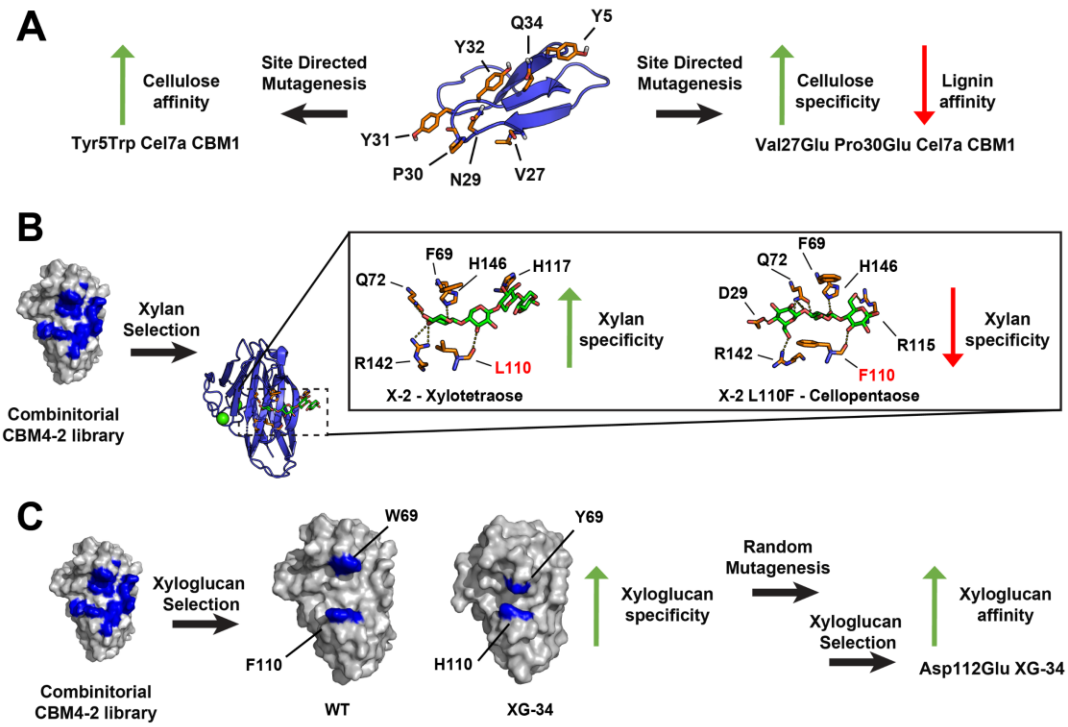


Figure 1-4. Examples of CBM engineering.
A) Cel7a CBM1 engineering. Site directed mutagenesis used to increase affinity for cellulose, and to decrease affinity for lignin. Structure shows residues involved in cellulose binding, and residues involved in lignin binding (PDB 1CBH). B) Wild-type CBM4-2 showing 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 XG-34 showing a 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.

1.9 Current scope of antibody engineering

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, 121-123] glycosaminoglycans,[124] human blood group oligosaccharides,[125] viral envelope proteins,[126-128] bacterial cell wall components,[129-131] 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.[41] Much of the current anti-glycan Ab landscape has been collated in the Database of Anti-Glycan Reagents (DAGR).[38] 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.

1.9.1 Fragmentation and display-based directed evolution

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 1-5B**). 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.[132] Further

fragmentation to the single-chain variable fragments (scFv) (**Figure 1-5B**) 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.[133] Due to their small size, they also present ideal scaffolds for display libraries. However, scFvs also exhibit prominent disadvantages including decreased stability, aggregation and misfolding when recombinantly produced in *E. coli*.[134, 135] Therefore, it is common to conjugate scFvs with a crystallizable fragment (Fc) to recapitulate some of the IgG structure (scFv-Fc) (**Figure 1-5B**) and regain stability.[136] 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.[137] 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.[138]

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.[139] The phage library was developed using extracted RNA from mouse spleen cells immunized with the Tn-positive Jurkat cells. 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 α-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 glycosaminoglycans (GAGs) chondroitin sulfate (CS), dermatan sulfate (DS), heparan sulfate (HS) and heparin epitopes.[140] 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 panning, 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.[141] 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.

More recently, Amon et al., utilized a yeast surface display approach to obtain scFVs binders to sialyl Lewis a (Neu5Acα2-3Galβ1-3(Fucα1-4)GlcNAcβ; SLe^a), a pancreatic and colorectal cancer-associated carbohydrate epitope.[142] Yeast expressed a randomized library of scFv scaffolds and were incubated with the SLe^a antigen on a polyacrylamide support (SLe^a-PAA). Clones that exhibited antigen binding were selected by FACS; three rounds of binding with successively decreased antigen concentration and increased gating stringency afforded five mutated scFv clones of interest. The scFv scaffolds of these isolated clones were then reconstituted as full-length IgG Abs for further characterization. One particular reconstituted full-length clone, RA9-23, was determined to have an apparent affinity of 12 nM by biolayer interferometry. This is a 3.5-fold enhancement of affinity compared to the native (non-randomized) scFv scaffold. RA9-23 also showed great specificity for SLe^a over some closely related glycans by ELISA, yet exhibited cross reactivity with $GcSLe^a$, 9-*O*-$AcSLe^a$, and 9-*O*-$GcSLe^a$ by sialoglycan microarray analysis. Incorporation of counter selections against these sialyated glycans would aid in imparting greater specificity to the RA9-23 Ab. Finally, this study showed the utility of the full-length RA9-23 Ab by showing its ability to bind SLe^a-expressing cancer cells (WiDr and Capan2) and increase cytotoxicity of the cell lines in vitro.

Together, these examples highlight the use of Ab scaffold fragments, namely scFvs, and display systems 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.

1.9.2 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.[143] 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.[144] A random codon-based mutagenesis strategy was applied to the anti-Lewis Y (Fucα1-2Galβ1-4(Fucα1-3)GlcNAc; Le^Y) Ab, BR96 (**Figure 1-5C**).[145] 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 tighter binding properties relative to the parent Ab, the specificities decreased (**Figure 1-5C**). 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.[125] 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*.[146] 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.[147] In 2G12, a phenomenon called domain swapping allows for an Ab conformation with four, proximal binding sites as opposed to two, distal ones (**Figure 1-5D**). 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 (**Figure 1-5D**). The Lai group developed a phage display approach for engineering glycan-binding Abs utilizing this domain-swapped chain arrangement.[148] 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 1-5D**). 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.[41] 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.

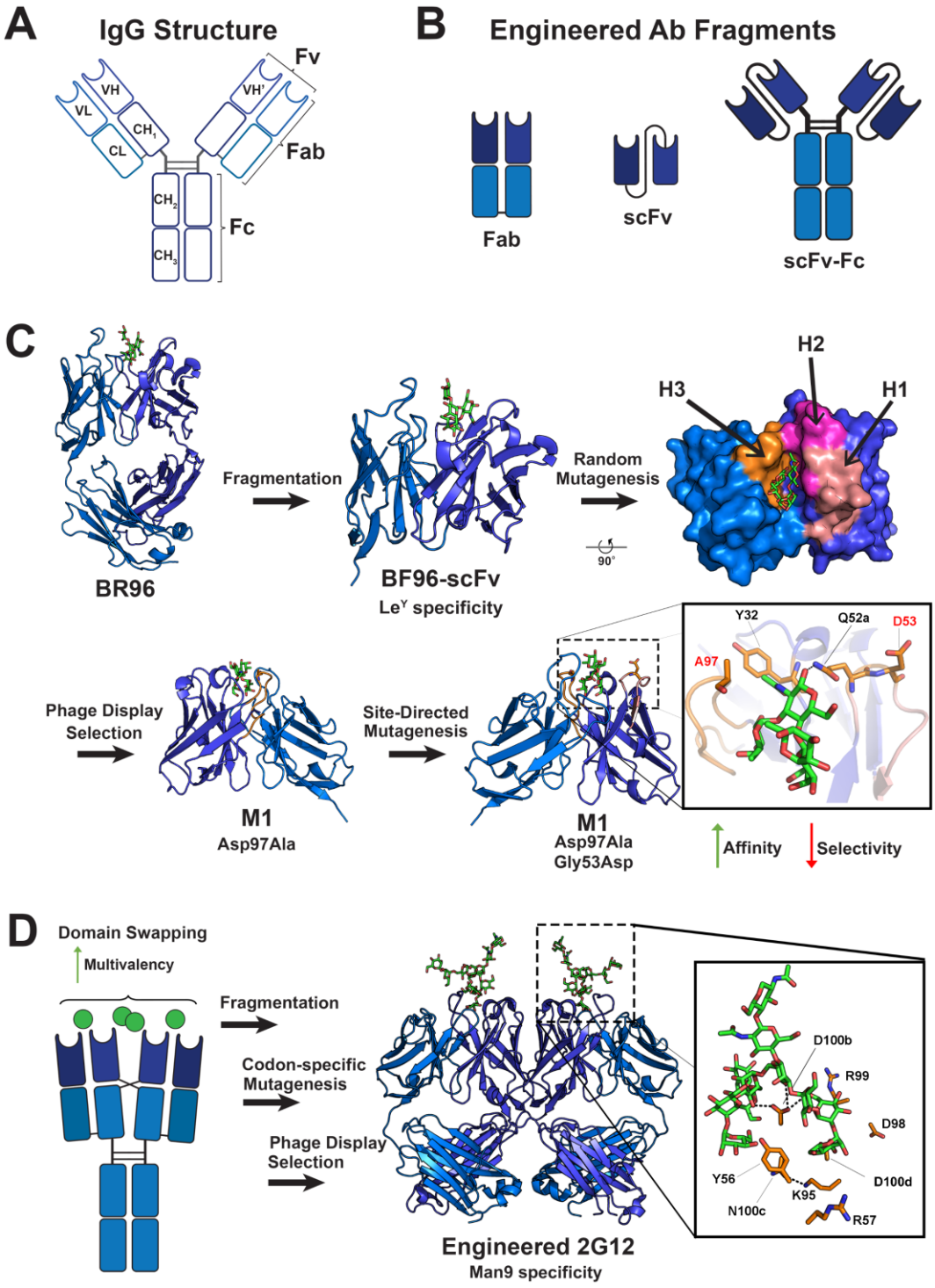


Figure 1-5. Antibody (Ab) structure and engineering.
A) Schematic of immunoglobulin G Ab denoting the heavy (H) and light (L) chains, crystallizable fragment (Fc), antigen binding fragment (Fab) and variable fragment (Fv). B) Engineered Ab fragments include the Fab, single-chain variable fragment (scFv), 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.

1.9.3 Lamprey variable lymphocyte receptors (VLRBs)

Lamprey, a jawless vertebrate, also exhibits an adaptive immune system, in this case *via* variable lymphocyte receptors (VLRs).[149] 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 1-6A**), providing a diversity of 10^{14} different proteins.[150] 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.[151]

Carbohydrates are highly immunogenic in lampreys;[152] 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.[151] 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.[153] 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,[149] 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 **Figure 1-6A**).[154] 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.[152] 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.

1.9.4 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 1-6C**). Tasumi et al. produced VLRBs in immunized lamprey and consequently developed a YSD vector for screening of the VLRB library.[150] 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).[155] 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.[156] 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.[156] 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 (**Figure 1-6A**). 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.

1.9.5 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.[157] YSD on the immunized library followed by MACS and FACS generated specific VLRBs. The isolated VLRBs were then conjugated to an Fc domain (**Figure 1-6B**) 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 6B**). 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 1-6C**).[158] 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.

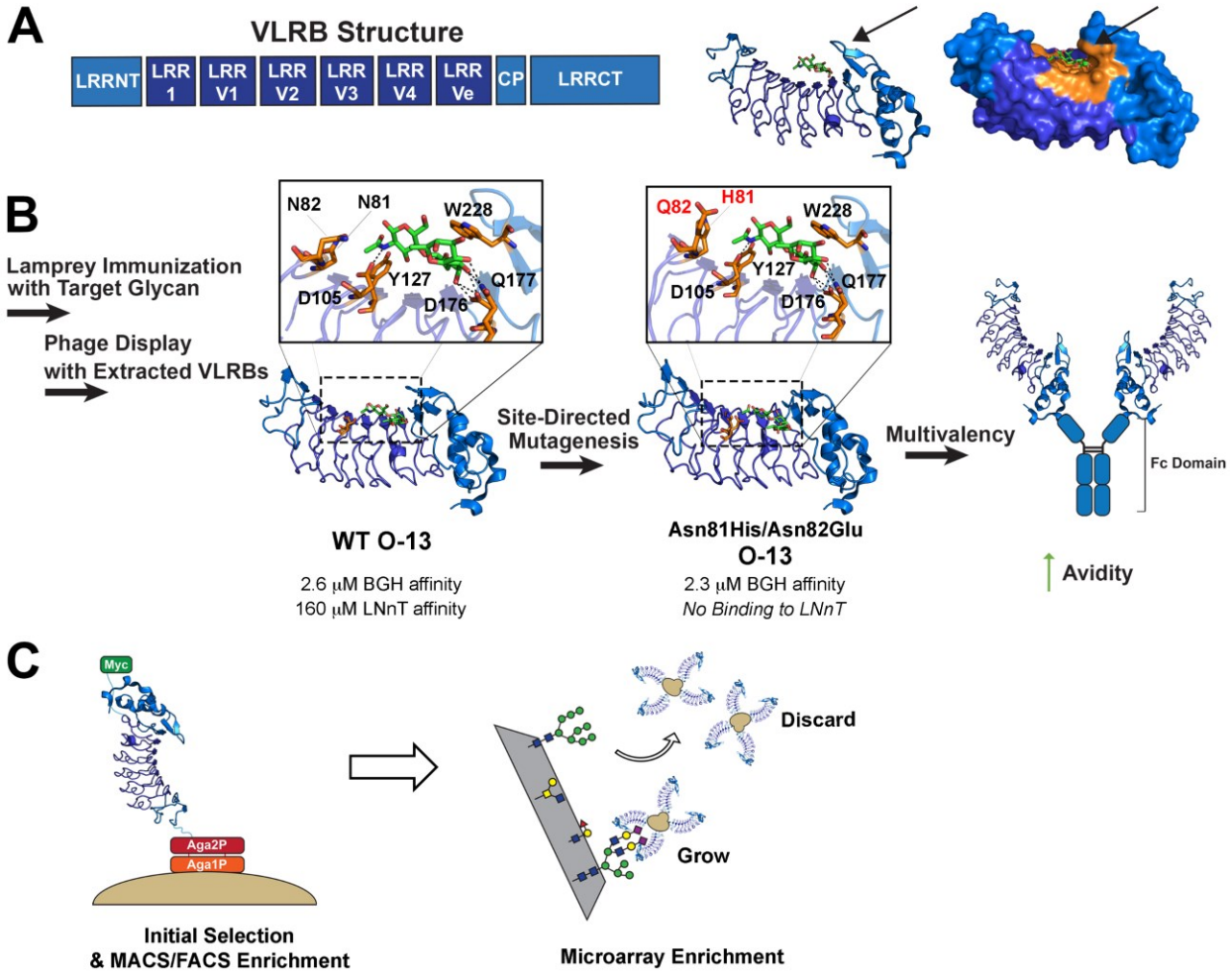


Figure 1-6. Variable lymphocyte receptor (VLRB) structure and engineering.
A) VLRBs are composed of leucine rich repeat (LRR) regions which fold to form a concave variable binding surface. A schematic of the protein sequence, highlighting the LRR motif, and a detailed structure of a monomeric protein 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 yeast surface display (YSD), affinity measurements and specificity determination. C) 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.

1.10 Conclusions

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 chapter 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 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,[159] although they have not yet been employed for anti-glycan development. These small Abs, also called “nanobodies,” are good scaffolds for phage[160]

and yeast display,[161] 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.[162]

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.[163, 164] 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.[40]

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".

1.11 Funding sources

This work was supported by National Institutes of Health (NIH) Interdepartmental Biotechnology Training Program Grant T32GM008334 (E.M.W.), NIH Postdoctoral Fellowship F32GM137477 (M.E.K.), and NIH Grant U01CA231078 (B.I.)

1.12 Acknowledgments

I would like to thank Dr. Megan Kizer who co-authored this review article with me, and Hannah Bernstein for careful proofreading of the manuscript.

1.13 References

1. Varki, A. (2017) Biological Roles of Glycans, *Glycobiology 27*, 3-49.

2. Toukach, P. V., and Egorova, K. S. (2015) Carbohydrate Structure Database Merged from Bacterial, Archaeal, Plant and Fungal Parts, *Nucleic Acids Res. 44*, D1229-D1236.

3. Imperiali, B. (2019) Bacterial Carbohydrate Diversity — a Brave New World, *Curr. Opin. Chem. Biol. 53*, 1-8.

4. Han, L., and Costello, C. E. (2013) Mass Spectrometry of Glycans, *Biochemistry (Mosc.) 78*, 710-720.

5. Leymarie, N., and Zaia, J. (2012) Effective Use of Mass Spectrometry for Glycan and Glycopeptide Structural Analysis, *Anal. Chem. 84*, 3040-3048.

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.

7. Sprinzl, M., Milovnikova, M., and Voertler, C. S. (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.

9. Taylor, M. E., Drickamer, K., Schnaar, R. L., Etzler, M. E., and Varki, A. (2015-2017) Discovery and Classification of Glycan-Binding Proteins, In *Essentials of Glycobiology* (Varki, A., Cummings, R. D., and Esko, J. D., 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.

11. Mazik, M., Cavga, H., and Jones, P. G. (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.

12. Wu, A. M., 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.

14. Brooks, S. A. (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, L. F. (2016) Blood Group Typing: From Classical Strategies to the Application of Synthetic Antibodies Generated by Molecular Imprinting, *Sensors 16*.

16. Slifkin, M., and Doyle, R. J. (1990) Lectins and Their Application to Clinical Microbiology, *Clin. Microbiol. Rev. 3*, 197-218.

17. Hashim, O. H., Jayapalan, J. J., and Lee, C.-S. (2017) Lectins: An Effective Tool for Screening of Potential Cancer Biomarkers, *PeerJ 5*, e3784-e3784.

18. Bies, C., Lehr, C.-M., and Woodley, J. F. (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.

20. Prendergast, J. M., Galvao da Silva, A. P., Eavarone, D. A., Ghaderi, D., Zhang, M., Brady, D., Wicks, J., DeSander, J., Behrens, J., and Rueda, B. R. (2017) Novel Anti-Sialyl-Tn Monoclonal Antibodies and Antibody-Drug Conjugates Demonstrate Tumor Specificity and Anti-Tumor Activity, *MAbs 9*, 615-627.

21. Coelho, L. C. B. B., Silva, P. M. d. S., Lima, V. L. d. M., Pontual, E. V., Paiva, P. M. G., Napoleão, T. H., and Correia, M. T. d. S. (2017) Lectins, Interconnecting Proteins with Biotechnological/Pharmacological and Therapeutic Applications, *Evid. Based Complement. Alternat. Med. 2017*, 1594074.

22. Bah, C. S. F., Fang, E. F., and Ng, T. B. (2013) Medicinal Applications of Plant Lectins, In *Antitumor Potential and Other Emerging Medicinal Properties of Natural Compounds* (Fang, E. F., and Ng, T. B., Eds.), pp 55-74, Springer Netherlands, Dordrecht.

23. Breitenbach Barroso Coelho, L. C., Marcelino dos Santos Silva, P., Felix de Oliveira, W., de Moura, M. C., Viana Pontual, E., Soares Gomes, F., Guedes Paiva, P. M., Napoleão, T. H., and dos Santos Correia, M. T. (2018) Lectins as Antimicrobial Agents, *J. Appl. Microbiol. 125*, 1238-1252.

24. Mitchell, C. A., Ramessar, K., and O'Keefe, B. R. (2017) Antiviral Lectins: Selective Inhibitors of Viral Entry, *Antiviral Res. 142*, 37-54.

25. 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.

26. Teschke, C. M., and King, J. (1992) Folding and Assembly of Oligomeric Proteins in Escherichia Coli, *Curr. Opin. Biotechnol. 3*, 468-473.

27. Sørensen, H. P., and Mortensen, K. K. (2005) Soluble Expression of Recombinant Proteins in the Cytoplasm of Escherichia Coli, *Microb. Cell Fact. 4*, 1.

28. 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.

29. Hervé, C., Rogowski, A., Blake, A. W., Marcus, S. E., Gilbert, H. J., and Knox, J. P. (2010) Carbohydrate-Binding Modules Promote the Enzymatic Deconstruction of Intact Plant Cell Walls by Targeting and Proximity Effects, *Proc. Natl. Acad. Sci. 107*, 15293-15298.

30. Din, N., Damude, H. G., Gilkes, N. R., Miller, R. C., Warren, R. A., and Kilburn, D. G. (1994) C1-Cx Revisited: Intramolecular Synergism in a Cellulase, *Proc. Natl. Acad. Sci. 91*, 11383-11387.

31. Southall, S. M., Simpson, P. J., Gilbert, H. J., Williamson, G., and Williamson, M. P. (1999) The Starch-Binding Domain from Glucoamylase Disrupts the Structure of Starch, *FEBS Lett. 447*, 58-60.

32. Gourlay, K., Arantes, V., and Saddler, J. N. (2012) Use of Substructure-Specific Carbohydrate Binding Modules to Track Changes in Cellulose Accessibility and Surface Morphology During the Amorphogenesis Step of Enzymatic Hydrolysis, *Biotechnol. Biofuels 5*, 51.

33. 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.

34. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2013) The Carbohydrate-Active Enzymes Database (Cazy) in 2013, *Nucleic Acids Res. 42*, D490-D495.

35. Lepenies, B., and Lang, R. (2019) Editorial: Lectins and Their Ligands in Shaping Immune Responses, *Front. Immunol. 10*.

36. Bochner, B. S., and Zimmermann, N. (2015) Role of Siglecs and Related Glycan-Binding Proteins in Immune Responses and Immunoregulation, *J. Allergy Clin. Immunol. 135*, 598-608.

37. Alberts, B., Johnson, A., Lewis, J., and al., e. (2002) The Generation of Antibody Diversity, In *Mol. Biol. Cell* 4th ed., Garland Science, New York, NY.

38. Sterner, E., Flanagan, N., and Gildersleeve, J. C. (2016) Perspectives on Anti-Glycan Antibodies Gleaned from Development of a Community Resource Database, *ACS Chem. Biol. 11*, 1773-1783.

39. Tommasone, S., Allabush, F., Tagger, Y. K., Norman, J., Köpf, M., Tucker, J. H. R., and Mendes, P. M. (2019) The Challenges of Glycan Recognition with Natural and Artificial Receptors, *Chem. Soc. Rev. 48*, 5488-5505.

40. Polonskaya, Z., Deng, S., Sarkar, A., Kain, L., Comellas-Aragones, M., McKay, C. S., Kaczanowska, K., Holt, M., McBride, R., Palomo, V., Self, K. M., Taylor, S., Irimia, A., Mehta, S. R., Dan, J. M., Brigger, M., Crotty, S., Schoenberger, S. P., Paulson, J. C., Wilson, I. A., Savage, P. B., Finn, M. G., and Teyton, L. (2017) T Cells Control the Generation of Nanomolar-Affinity Anti-Glycan Antibodies, *J. Clin. Invest. 127*, 1491-1504.

41. Haji-Ghassemi, O., Blackler, R. J., Martin Young, N., and Evans, S. V. (2015) Antibody Recognition of Carbohydrate Epitopes†, *Glycobiology 25*, 920-952.

42. Janeway, C. A. J., Travers, P., and Walport, M. (2001) Immunobiology: The Immune System in Health and Disease, 5th Edition ed., Garland Science, New York.

43. Hudson, K. L., Bartlett, G. J., Diehl, R. C., Agirre, J., Gallagher, T., Kiessling, L. L., and Woolfson, D. N. (2015) Carbohydrate–Aromatic Interactions in Proteins, *J. Am. Chem. Soc. 137*, 15152-15160.

44. Asensio, J. L., Ardá, A., Cañada, F. J., and Jiménez-Barbero, J. (2013) Carbohydrate–Aromatic Interactions, *Acc. Chem. Res. 46*, 946-954.

45. Drickamer, K. (1997) Making a Fitting Choice: Common Aspects of Sugar-Binding Sites in Plant and Animal Lectins, *Structure 5*, 465-468.

46. Taylor, M. E., and Drickamer, K. (2014) Convergent and Divergent Mechanisms of Sugar Recognition across Kingdoms, *Curr. Opin. Struct. Biol. 28*, 14-22.

47. Monsigny, M., Mayer, R., Roche, A.C. (1991) Sugar-Lectin Interactions: Sugar Cluster, Lectin Multivalency and Avidity, *Carbohydr. Lett. 4*, 35-52.

48. Lundquist, J. J., and Toone, E. J. (2002) The Cluster Glycoside Effect, *Chem. Rev. 102*, 555-578.

49. Cohen, M., and Varki, A. (2014) Chapter Three - Modulation of Glycan Recognition by Clustered Saccharide Patches, In *Int. Rev. Cell Mol. Biol.* (Jeon, K. W., Ed.), pp 75-125, Academic Press.

50. Marshall, S. A., Lazar, G. A., Chirino, A. J., and Desjarlais, J. R. (2003) Rational Design and Engineering of Therapeutic Proteins, *Drug Discov. Today 8*, 212-221.

51. Arnold, F. H. (1998) Design by Directed Evolution, *Acc. Chem. Res. 31*, 125-131.

52. Sen, S., Venkata Dasu, V., and Mandal, B. (2007) Developments in Directed Evolution for Improving Enzyme Functions, *Appl. Biochem. Biotechnol. 143*, 212-223.

53. 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.

54. Navaratna, T., Atangcho, L., Mahajan, M., Subramanian, V., Case, M., Min, A., Tresnak, D., and Thurber, G. M. (2020) Directed Evolution Using Stabilized Bacterial Peptide Display, *J. Am. Chem. Soc. 142*, 1882-1894.

55. Boder, E. T., and Wittrup, K. D. (2000) Yeast Surface Display for Directed Evolution of Protein Expression, Affinity, and Stability, *Methods Enzymol. 328*, 430-444.

56. 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.

57. Yan, X., and Xu, Z. (2006) Ribosome-Display Technology: Applications for Directed Evolution of Functional Proteins, *Drug Discov. Today 11*, 911-916.

58. Plückthun, A. (2012) Ribosome Display: A Perspective, In *Ribosome Display and Related Technologies: Methods and Protocols* (Douthwaite, J. A., and Jackson, R. H., Eds.), pp 3-28, Springer New York, New York, NY.

59. Xu, L., Aha, P., Gu, K., Kuimelis, R. G., Kurz, M., Lam, T., Lim, A. C., Liu, H., Lohse, P. A., Sun, L., Weng, S., Wagner, R. W., and Lipovsek, D. (2002) Directed Evolution of High-Affinity Antibody Mimics Using Mrna Display, *Chem. Biol. 9*, 933-942.

60. 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.*, 20191161.

61. 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.

62. Fukui, S., Feizi, T., Galustian, C., Lawson, A. M., and Chai, W. (2002) Oligosaccharide Microarrays for High-Throughput Detection and Specificity Assignments of Carbohydrate-Protein Interactions, *Nat. Biotechnol. 20*, 1011-1017.

63. Wang, D., Liu, S., Trummer, B. J., 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.

64. Blixt, O., Head, S., Mondala, T., Scanlan, C., Huflejt, M. E., Alvarez, R., Bryan, M. C., Fazio, F., Calarese, D., Stevens, J., Razi, N., Stevens, D. J., Skehel, J. J., van Die, I., Burton, D. R., Wilson, I. A., Cummings, R., Bovin, N., Wong, C.-H., and Paulson, J. C. (2004) Printed Covalent Glycan Array for Ligand Profiling of Diverse Glycan Binding Proteins, *Proc. Natl. Acad. Sci. 101*, 17033.

65. Gao, C., Wei, M., McKitrick, T. R., McQuillan, A. M., Heimburg-Molinaro, J., and Cummings, R. D. (2019) Glycan Microarrays as Chemical Tools for Identifying Glycan Recognition by Immune Proteins, *Front. Chem. 7*.

66. Heimburg-Molinaro, J., Song, X., Smith, D. F., and Cummings, R. D. (2011) Preparation and Analysis of Glycan Microarrays, *Curr. Protoc. Prot. Sci. 64*, 12.10.11-12.10.29.

67. Li, Z., and Feizi, T. (2018) The Neoglycolipid (Ngl) Technology-Based Microarrays and Future Prospects, *FEBS Lett. 592*, 3976-3991.

68. Rillahan, C. D., and Paulson, J. C. (2011) Glycan Microarrays for Decoding the Glycome, *Annu. Rev. Biochem. 80*, 797-823.

69. Smith, D. F., Song, X., and Cummings, R. D. (2010) Chapter Nineteen - Use of Glycan Microarrays to Explore Specificity of Glycan-Binding Proteins, In *Methods Enzymol.*, pp 417-444, Academic Press.

70. 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, H. S., Kandasamy, J., Leonori, D., Martin, C. E., Parameswarappa, S. G., Pasari, S., Schlegel, M. K., Tanaka, H., Xiao, G., Yang, Y., Pereira, C. L., Anish, C., and Seeberger, P. H. (2019) Microbe-Focused Glycan Array Screening Platform, *Proc. Natl. Acad. Sci. 116*, 1958.

71. Byrd-Leotis, L., Liu, R., Bradley, K. C., Lasanajak, Y., Cummings, S. F., Song, X., Heimburg-Molinaro, J., Galloway, S. E., Culhane, M. R., Smith, D. F., Steinhauer, D. A., and Cummings, R. D. (2014) Shotgun Glycomics of Pig Lung Identifies Natural Endogenous Receptors for Influenza Viruses, *Proc. Natl. Acad. Sci. 111*, E2241.

72. Song, X., Lasanajak, Y., Xia, B., Heimburg-Molinaro, J., Rhea, J. M., Ju, H., Zhao, C., Molinaro, R. J., Cummings, R. D., and Smith, D. F. (2011) Shotgun Glycomics: A Microarray Strategy for Functional Glycomics, *Nat. Methods 8*, 85-90.

73. Padler-Karavani, V., Song, X., Yu, H., Hurtado-Ziola, N., Huang, S., Muthana, S., Chokhawala, H. A., Cheng, J., Verhagen, A., Langereis, M. A., Kleene, R., Schachner, M., de Groot, R. J., Lasanajak, Y., Matsuda, H., Schwab, R., Chen, X., Smith, D. F., Cummings, R. D., and Varki, A. (2012) Cross-Comparison of Protein Recognition of Sialic Acid Diversity on Two Novel Sialoglycan Microarrays, *J. Biol. Chem. 287*, 22593-22608.

74. Wang, L., Cummings, R. D., Smith, D. F., Huflejt, M., Campbell, C. T., Gildersleeve, J. C., Gerlach, J. Q., Kilcoyne, M., Joshi, L., Serna, S., Reichardt, N.-C., Parera Pera, N., Pieters, R. J., Eng, W., and Mahal, L. K. (2014) Cross-Platform Comparison of Glycan Microarray Formats, *Glycobiology 24*, 507-517.

75. Temme, J. S., Campbell, C. T., and Gildersleeve, J. C. (2019) Factors Contributing to Variability of Glycan Microarray Binding Profiles, *Faraday Discuss. 219*, 90-111.

76. Gao, C., Hanes, M. S., Byrd-Leotis, L. A., Wei, M., Jia, N., Kardish, R. J., McKitrick, T. R., Steinhauer, D. A., and Cummings, R. D. (2019) Unique Binding Specificities of Proteins toward Isomeric Asparagine-Linked Glycans, *Cell Chem. Biol. 26*, 535-547.e534.

77. 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, R. D., Wang, P. G., and She, J.-X. (2018) Multiplex Glycan Bead Array for High Throughput and High Content Analyses of Glycan Binding Proteins, *Nat. Comm. 9*, 258.

78. 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.

79. Yan, M., Zhu, Y., Liu, X., Lasanajak, Y., Xiong, J., Lu, J., Lin, X., Ashline, D., Reinhold, V., Smith, D. F., and Song, X. (2019) Next-Generation Glycan Microarray Enabled by DNA-Coded Glycan Library and Next-Generation Sequencing Technology, *Anal. Chem. 91*, 9221-9228.

80. Chen, Y.-H., Narimatsu, Y., Clausen, T. M., Gomes, C., Karlsson, R., Steentoft, C., Spliid, C. B., Gustavsson, T., Salanti, A., Persson, A., Malmström, A., Willén, D., Ellervik, U., Bennett, E. P., Mao, Y., Clausen, H., and Yang, Z. (2018) The Gagome: A Cell-Based Library of Displayed Glycosaminoglycans, *Nat. Methods 15*, 881-888.

81. Briard, J. G., Jiang, H., Moremen, K. W., Macauley, M. S., and Wu, P. (2018) Cell-Based Glycan Arrays for Probing Glycan–Glycan Binding Protein Interactions, *Nat. Comm. 9*, 880.

82. Cummings, R. D. (2009) The Repertoire of Glycan Determinants in the Human Glycome, *Mol. Biosyst. 5*, 1087-1104.

83. Imperiali, B. (2019) Bacterial Carbohydrate Diversity - a Brave New World, *Curr. Opin. Chem. Biol. 53*, 1-8.

84. Woods, R. J., and Yang, L. (2018) Glycan-Specific Analytical Tools, *U.S. Patent 9926612 B2*.

85. Hu, D., Tateno, H., and Hirabayashi, J. (2015) Lectin Engineering, a Molecular Evolutionary Approach to Expanding the Lectin Utilities, *Molecules 20*.

86. Hirabayashi, J., and Arai, R. (2019) Lectin Engineering: The Possible and the Actual, *Interface Focus 9*, 20180068.

87. Cummings, R. D., and McEver, R. P. (2015) C-Type Lectins, In *Essentials of Glycobiology* (rd, Varki, A., Cummings, R. D., Esko, J. D., Stanley, P., Hart, G. W., Aebi, M., Darvill, A. G., Kinoshita, T., Packer, N. H., Prestegard, J. H., Schnaar, R. L., and Seeberger, P. H., Eds.), pp 435-452, Cold Spring Harbor (NY).

88. Drickamer, K. (1992) Engineering Galactose-Binding Activity into a C-Type Mannose-Binding Protein, *Nature 360*, 183-186.

89. Iobst, S. T., 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.

90. 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.

91. Adhikari, P., Bachhawat-Sikder, K., Thomas, C. J., Ravishankar, R., Jeyaprakash, A. A., 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.

92. 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.

93. 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.

94. Imamura, K., Takeuchi, H., Yabe, R., Tateno, H., and Hirabayashi, J. (2011) Engineering of the Glycan-Binding Specificity of Agrocybe Cylindracea Galectin Towards A(2,3)-Linked Sialic Acid by Saturation Mutagenesis, *J. Biochem. 150*, 545-552.

95. 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.

96. 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.

97. 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.

98. 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.

99. Yamamoto, K., Maruyama, I. N., and Osawa, T. (2000) Cyborg Lectins: Novel Leguminous Lectins with Unique Specificities1, *J. Biochem. 127*, 137-142.

100. 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.

101. 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.

102. 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.

103. 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.

104. Hirabayashi, J., Dutta, S. K., 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.

105. Yabe, R., Itakura, Y., Nakamura-Tsuruta, S., Iwaki, J., Kuno, A., and Hirabayashi, J. (2009) Engineering a Versatile Tandem Repeat-Type A2-6sialic Acid-Binding Lectin, *Biochem. Biophys. Res. Commun. 384*, 204-209.

106. Mahajan, S., and Ramya, T. N. C. (2018) Nature-Inspired Engineering of an F-Type Lectin for Increased Binding Strength, *Glycobiology 28*, 933-948.

107. Hamorsky, K. T., Kouokam, J. C., Dent, M. W., Grooms, T. N., Husk, A. S., Hume, S. D., Rogers, K. A., Villinger, F., Morris, M. K., Hanson, C. V., and Matoba, N. (2019) Engineering of a Lectibody Targeting High-Mannose-Type Glycans of the Hiv Envelope, *Mol. Ther. 27*, 2038-2052.

108. Suits, M. D. L., Pluvinage, B., Law, A., Liu, Y., Palma, A. S., Chai, W., Feizi, T., and Boraston, A. B. (2014) Conformational Analysis of the Streptococcus Pneumoniae Hyaluronate Lyase and Characterization of Its Hyaluronan-Specific Carbohydrate-Binding Module, *J. Biol. Chem. 289*, 27264-27277.

109. 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. Bioinform. 85*, 1602-1617.

110. Linder, M., Lindeberg, G., Reinikainen, T., Teeri, T. T., 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.

111. Strobel, K. L., Pfeiffer, K. A., Blanch, H. W., and Clark, D. S. (2016) Engineering Cel7a Carbohydrate Binding Module and Linker for Reduced Lignin Inhibition, *Biotechnol. Bioeng. 113*, 1369-1374.

112. Cicortas Gunnarsson, L., Nordberg Karlsson, E., Albrekt, A. S., Andersson, M., Holst, O., and Ohlin, M. (2004) A Carbohydrate Binding Module as a Diversity-Carrying Scaffold, *Protein Eng. Des. Sel. 17*, 213-221.

113. Cicortas Gunnarsson, L., Montanier, C., Tunnicliffe, R. B., Williamson, M. P., Gilbert, H. J., Nordberg Karlsson, E., and Ohlin, M. (2007) Novel Xylan-Binding Properties of an Engineered Family 4 Carbohydrate-Binding Module, *Biochem. J. 406*, 209-214.

114. von Schantz, L., Håkansson, M., Logan, D. T., 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.

115. Gunnarsson, L. C., Dexlin, L., Karlsson, E. N., Holst, O., and Ohlin, M. (2006) Evolution of a Carbohydrate Binding Module into a Protein-Specific Binder, *Biomol. Eng. 23*, 111-117.

116. Gunnarsson, L. C., Zhou, Q., Montanier, C., Karlsson, E. N., Brumer, H., III, and Ohlin, M. (2006) Engineered Xyloglucan Specificity in a Carbohydrate-Binding Module, *Glycobiology 16*, 1171-1180.

117. Gullfot, F., Tan, T.-C., von Schantz, L., Karlsson, E. N., Ohlin, M., Brumer, H., and Divne, C. (2010) The Crystal Structure of Xg-34, an Evolved Xyloglucan-Specific Carbohydrate-Binding Module, *Proteins: Struct. Funct. Bioinform. 78*, 785-789.

118. Furtado, G. P., Lourenzoni, M. R., Fuzo, C. A., Fonseca-Maldonado, R., Guazzaroni, M.-E., Ribeiro, L. F., and Ward, R. J. (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.

119. Ribeiro, D. O., Viegas, A., Pires, V. M. R., Medeiros-Silva, J., Bule, P., Chai, W., Marcelo, F., Fontes, C. M. G. A., Cabrita, E. J., Palma, A. S., and Carvalho, A. L. (2020) Molecular Basis for the Preferential Recognition of B1,3-1,4-Glucans by the Family 11 Carbohydrate-Binding Module from Clostridium thermocellum, *FEBS J. 287*, 2723-2743.

120. von Schantz, L., Gullfot, F., Scheer, S., Filonova, L., Cicortas Gunnarsson, L., Flint, J. E., 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.

121. Jeffrey, P. D., Bajorath, J., Chang, C. Y., Yelton, D., Hellström, I., Hellström, K. E., and Sheriff, S. (1995) The X-Ray Structure of an Anti-Tumour Antibody in Complex with Antigen, *Nat. Struct. Biol. 2*, 466-471.

122. van Roon, A.-M. M., Pannu, N. S., de Vrind, J. P. M., van der Marel, G. A., van Boom, J. H., Hokke, C. H., Deelder, A. M., and Abrahams, J. P. (2004) Structure of an Anti-Lewis X Fab Fragment in Complex with Its Lewis X Antigen, *Structure 12*, 1227-1236.

123. Vazquez, A. M., Alfonso, M., Lanne, B., Karlsson, K. A., Carr, A., Barroso, O., Fernandez, L. E., Rengifo, E., Lanio, M. E., 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.

124. Conrad, G. W., Ager-Johnson, P., and Woo, M. L. (1982) Antibodies against the Predominant Glycosaminoglycan of the Mammalian Cornea, Keratan Sulfate-I, *J. Biol. Chem. 257*, 464-471.

125. Thomas, R., Patenaude, S. I., MacKenzie, C. R., To, R., Hirama, T., Young, N. M., and Evans, S. V. (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.

126. Mouquet, H., Scharf, L., Euler, Z., Liu, Y., Eden, C., Scheid, J. F., Halper-Stromberg, A., Gnanapragasam, P. N. P., Spencer, D. I. R., Seaman, M. S., Schuitemaker, H., Feizi, T., Nussenzweig, M. C., and Bjorkman, P. J. (2012) Complex-Type N-Glycan Recognition by Potent Broadly Neutralizing Hiv Antibodies, *Proc. Natl. Acad. Sci. 109*, E3268.

127. McLellan, J. S., 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, J. C., Chuang, G.-Y., Diwanji, D., Georgiev, I., Do Kwon, Y., Lee, D., Louder, M. K., Moquin, S., Schmidt, S. D., Yang, Z.-Y., Bonsignori, M., Crump, J. A., Kapiga, S. H., Sam, N. E., Haynes, B. F., Burton, D. R., Koff, W. C., Walker, L. M., Phogat, S., Wyatt, R., Orwenyo, J., Wang, L.-X., Arthos, J., Bewley, C. A., Mascola, J. R., Nabel, G. J., Schief, W. R., Ward, A. B., Wilson, I. A., and Kwong, P. D. (2011) Structure of Hiv-1 Gp120 V1/V2 Domain with Broadly Neutralizing Antibody Pg9, *Nature 480*, 336-343.

128. Calarese, D. A., Scanlan, C. N., Zwick, M. B., Deechongkit, S., Mimura, Y., Kunert, R., Zhu, P., Wormald, M. R., Stanfield, R. L., Roux, K. H., Kelly, J. W., Rudd, P. M., Dwek, R. A., Katinger, H., Burton, D. R., and Wilson, I. A. (2003) Antibody Domain Exchange Is an Immunological Solution to Carbohydrate Cluster Recognition, *Science 300*, 2065.

129. Vulliez-Le Normand, B., Saul, F. A., Phalipon, A., Bélot, F., Guerreiro, C., Mulard, L. A., and Bentley, G. A. (2008) Structures of Synthetic O-Antigen Fragments from Serotype 2a Shigella Flexneri in Complex with a Protective Monoclonal Antibody, *Proc. Natl. Acad. Sci. 105*, 9976-9981.

130. Brooks, C. L., Müller-Loennies, S., Borisova, S. N., Brade, L., Kosma, P., Hirama, T., Mackenzie, C. R., Brade, H., and Evans, S. V. (2010) Antibodies Raised against Chlamydial Lipopolysaccharide Antigens Reveal Convergence in Germline Gene Usage and Differential Epitope Recognition, *Biochemistry 49*, 570-581.

131. Murase, T., Zheng, R. B., Joe, M., Bai, Y., Marcus, S. L., Lowary, T. L., and Ng, K. K. (2009) Structural Insights into Antibody Recognition of Mycobacterial Polysaccharides, *J. Mol. Biol. 392*, 381-392.

132. Pejchal, R., Doores, K. J., Walker, L. M., Khayat, R., Huang, P. S., Wang, S. K., Stanfield, R. L., Julien, J. P., Ramos, A., Crispin, M., Depetris, R., Katpally, U., Marozsan, A., Cupo, A., Maloveste, S., Liu, Y., McBride, R., Ito, Y., Sanders, R. W., Ogohara, C., Paulson, J. C., Feizi, T., Scanlan, C. N., Wong, C. H., Moore, J. P., Olson, W. C., Ward, A. B., Poignard, P., Schief, W. R., Burton, D. R., and Wilson, I. A. (2011) A Potent and Broad Neutralizing Antibody Recognizes and Penetrates the Hiv Glycan Shield, *Science 334*, 1097-1103.

133. Monnier, P. P., Vigouroux, J. R., and Tassew, G. N. (2013) In Vivo Applications of Single Chain Fv (Variable Domain) (Scfv) Fragments, *Antibodies 2*.

134. Markiv, A., Anani, B., Durvasula, R. V., and Kang, A. S. (2011) Module Based Antibody Engineering: A Novel Synthetic Redantibody, *J. Immunol. Methods 364*, 40-49.

135. 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.

136. 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.

137. 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.

138. Blackler, R. J., Evans, D. W., Smith, D. F., Cummings, R. D., Brooks, C. L., Braulke, T., Liu, X., Evans, S. V., 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.

139. 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.

140. Smits, N. C., Lensen, J. F. M., Wijnhoven, T. J. M., ten Dam, G. B., Jenniskens, G. J., and van Kuppevelt, T. H. (2006) Phage Display-Derived Human Antibodies against Specific Glycosaminoglycan Epitopes, In *Methods Enzymol.*, pp 61-87, Academic Press.

141. van Kuppevelt, T. H., Dennissen, M. A., van Venrooij, W. J., Hoet, R. M., and Veerkamp, J. H. (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.

142. Amon, R., Rosenfeld, R., Perlmutter, S., Grant, O. C., Yehuda, S., Borenstein-Katz, A., Alcalay, R., Marshanski, T., Yu, H., Diskin, R., Woods, R. J., Chen, X., and Padler-Karavani, V. (2020) Directed Evolution of Therapeutic Antibodies Targeting Glycosylation in Cancer, *Cancers (Basel) 12*, 2824.

143. Bona, C. (1993) Molecular Characteristics of Anti-Polysaccharide Antibodies, *Springer Semin. Immunopathol. 15*, 103-118.

144. Brorson, K., Thompson, C., Wei, G., Krasnokutsky, M., and Stein, K. E. (1999) Mutational Analysis of Avidity and Fine Specificity of Anti-Levan Antibodies, *J. Immunol. 163*, 6694.

145. Yelton, D. E., Rosok, M. J., Cruz, G., Cosand, W. L., Bajorath, J., Hellström, I., Hellström, K. E., Huse, W. D., and Glaser, S. M. (1995) Affinity Maturation of the Br96 Anti-Carcinoma Antibody by Codon-Based Mutagenesis, *J. Immunol. 155*, 1994.

146. Deng, S. J., MacKenzie, C. R., Hirama, T., Brousseau, R., Lowary, T. L., Young, N. M., Bundle, D. R., and Narang, S. A. (1995) Basis for Selection of Improved Carbohydrate-Binding Single-Chain Antibodies from Synthetic Gene Libraries, *Proc. Natl. Acad. Sci. 92*, 4992-4996.

147. Wu, Y., West, A. P., Jr., Kim, H. J., Thornton, M. E., Ward, A. B., and Bjorkman, P. J. (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.

148. Lin, T.-Y., and Lai, J. R. (2017) Interrogation of Side Chain Biases for Oligomannose Recognition by Antibody 2g12 Via Structure-Guided Phage Display Libraries, *Biorg. Med. Chem. 25*, 5790-5798.

149. Altman, M. O., Bennink, J. R., Yewdell, J. W., and Herrin, B. R. (2015) Lamprey Vlrb Response to Influenza Virus Supports Universal Rules of Immunogenicity and Antigenicity, *eLife 4*, e07467.

150. Tasumi, S., Velikovsky, C. A., Xu, G., Gai, S. A., Wittrup, K. D., Flajnik, M. F., Mariuzza, R. A., and Pancer, Z. (2009) High-Affinity Lamprey Vlra and Vlrb Monoclonal Antibodies, *Proc. Natl. Acad. Sci. 106*, 12891.

151. Alder, M. N., Herrin, B. R., Sadlonova, A., Stockard, C. R., Grizzle, W. E., Gartland, L. A., Gartland, G. L., Boydston, J. A., Turnbough, C. L., and Cooper, M. D. (2008) Antibody Responses of Variable Lymphocyte Receptors in the Lamprey, *Nat. Immunol. 9*, 319-327.

152. Collins, B. C., Gunn, R. J., McKitrick, T. R., Cummings, R. D., Cooper, M. D., Herrin, B. R., and Wilson, I. A. (2017) Structural Insights into Vlr Fine Specificity for Blood Group Carbohydrates, *Structure 25*, 1667-1678.e1664.

153. McKitrick, T. R., Eris, D., Mondal, N., Aryal, R. P., McCurley, N., Heimburg-Molinaro, J., and Cummings, R. D. (2020) Antibodies from Lampreys as Smart Anti-Glycan Reagents (Sagrs): Perspectives on Their Specificity, Structure, and Glyco-Genomics, *Biochemistry*.

154. Han, B. W., Herrin, B. R., Cooper, M. D., and Wilson, I. A. (2008) Antigen Recognition by Variable Lymphocyte Receptors, *Science 321*, 1834.

155. Hong, X., Ma, M. Z., Gildersleeve, J. C., Chowdhury, S., Barchi, J. J., Mariuzza, R. A., Murphy, M. B., 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.

156. Luo, M., Velikovsky, C. A., Yang, X., Siddiqui, M. A., Hong, X., Barchi, J. J., Jr., Gildersleeve, J. C., Pancer, Z., and Mariuzza, R. A. (2013) Recognition of the Thomsen-Friedenreich Pancarcinoma Carbohydrate Antigen by a Lamprey Variable Lymphocyte Receptor, *J. Biol. Chem. 288*, 23597-23606.

157. Pollara, B., Litman, G. W., Finstad, J., Howell, J., and Good, R. A. (1970) The Evolution of the Immune Response, *J. Immunol. 105*, 738.

158. McKitrick, T. R., Goth, C. K., Rosenberg, C. S., Nakahara, H., Heimburg-Molinaro, J., McQuillan, A. M., Falco, R., Rivers, N. J., Herrin, B. R., Cooper, M. D., and Cummings, R. D. (2020) Development of Smart Anti-Glycan Reagents Using Immunized Lampreys, *Comm. Biol. 3*, 91.

159. Wilton, E. E., Opyr, M. P., Kailasam, S., Kothe, R. F., and Wieden, H.-J. (2018) Sdab-Db: The Single Domain Antibody Database, *ACS Synth. Biol. 7*, 2480-2484.

160. 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, *Antib. Ther. 3*, 1-9.

161. McMahon, C., Baier, A. S., Pascolutti, R., Wegrecki, M., Zheng, S., Ong, J. X., Erlandson, S. C., Hilger, D., Rasmussen, S. G. F., Ring, A. M., Manglik, A., and Kruse, A. C. (2018) Yeast Surface Display Platform for Rapid Discovery of Conformationally Selective Nanobodies, *Nat. Struct. Mol. Biol. 25*, 289-296.

162. Binz, H. K., Amstutz, P., Kohl, A., Stumpp, M. T., Briand, C., Forrer, P., Grütter, M. G., and Plückthun, A. (2004) High-Affinity Binders Selected from Designed Ankyrin Repeat Protein Libraries, *Nat. Biotechnol. 22*, 575-582.

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

164. Mettu, R., Chen, C.-Y., and Wu, C.-Y. (2020) Synthetic Carbohydrate-Based Vaccines: Challenges and Opportunities, *J. Biomed. Sci. 27*, 9.

Chapter 2: Directed evolution of archaeal protein Sso7d as a platform for development of novel glycan-binding proteins

Contributions:

Early sialic acid sort conditions with biotinylated ligands performed by Cristina Zamora
Alternating ligand selections performed by Elizabeth Ward and Cristina Zamora
Increased negative selections and competitive sorts with biotinylated TF ligands performed by Elizabeth Ward
Preparation and selections with magnetic beads performed by Cristina Zamora and Elizabeth Ward
Initial sialic acid sorts and affinity maturation with multivalent fluorescent sialic acid polymer performed by Cristina Zamora
Competitive selections using sialic acid and TF glycopolymers performed by Elizabeth Ward
Protein purifications and BLI analysis performed by Elizabeth Ward and Cristina Zamora

2.1 Abstract

Glycan-binding proteins (GBPs) are widely used for basic research and clinical applications. GBPs allow for the identification of glycan determinants without specialized equipment or time-consuming experimental methods. Existing GBPs, mainly antibodies and lectins, are limited, and discovery or creation of reagents with novel specificities is time consuming and difficult. Here I detail the generation of novel GBPs from a small, hyperthermostable DNA binding protein by directed evolution. A yeast surface display method for enrichment of GBPs was developed and used to generate sialic acid-binding proteins. Characterization of these proteins show them to have specificities and affinities on par with currently available lectins. The facile generation of GBPs that recognize specific saccharides of interest will have great impact on both biomedical and glycobiological research.

2.2 Introduction

All domains of life produce unique carbohydrates that can be secreted or appended to biomolecules. These carbohydrates, or glycans, are directly involved in diverse processes such as protein stability, cell adhesion, cell-cell communication, and host-cell interactions.[1] Cell surface glycans can also serve as important disease markers, as glycosylation is often aberrant in malignant tissues.[2] Despite the importance of glycans, there are significant difficulties in their study. This is due in part to the massive diversity of glycans, both at the monosaccharide level and at the glycosidic linkage level. Estimates of unique monosaccharides produced by bacteria are in the order of 800, much higher than nucleic acids and amino acids.[3, 4] Also unlike DNA or proteins, glycans are not linear polymers and can form glycosidic linkages at several hydroxyl groups in a sugar building block and can adopt an α- or β-anomeric configuration. This also allows for the formation of larger branched structures. Another major barrier to study of carbohydrates is that they are not encoded in the genome, so classical genetic techniques to manipulate and modify them cannot be employed. Many glycan structure analysis methods require release of the glycan from the glycoconjugate and separation from the resulting complex mixture, followed by methods such as mass spectrometry or NMR.[5] These methods have disadvantages such as large sample requirements, expensive equipment, and need for specialized personnel. Glycan-binding proteins (GBPs) are another tool for qualitative glycan structure analysis. GBPs are reagents that bind to specific glycan epitopes and allow for structure determination without the extensive processing or specialized equipment that other techniques require. These reagents can be used in the lab as affinity reagents for glycan/glycoconjugate purification and cell staining reagents, or used diagnostically for blood typing, microorganism detection, and histology.[6-9] Because of this, well-characterized GBPs are crucial for glycobiological research.

The predominant GBP reagents consist of fungal, plant, and animal derived lectins and monoclonal antibodies. These GBPs are important tools for glycobiology, but they have limitations. Lectins are non-enzymatic GBPs and are popular because they can recognize monosaccharide determinants and glycosidic linkages and are relatively cheap to isolate or purify. Commercially available lectins can recognize glycan determinants such as β-linked terminal Gal (RCA-I), α-linked fucose (AAL), terminal sialic acid (WGA), and various core O-glycans (PNA, Jacalin).[10] These proteins are very useful but the specificities of known lectins do not cover the large diversity of glycans found in nature. Antibodies against sugars can be directly elicited in animals by inoculation with free glycan (e.g. bacterial capsular polysaccharide), glycan-protein conjugates (native glycoproteins or glycans coupled to carrier proteins like bovine serum albumin or keyhole limpet hemocyanin), or whole cells (e.g. bacterial cells or tumor cells).[10] Anti-glycan antibodies have been developed for a number of glycan targets including O- and N-linked glycans, human Lewis and blood group antigens, glycolipids, glycosaminoglycans, and bacterial and plant cell wall components.[11] Antibody elicitation often produces unpredictable results due to the poor immunogenicity of carbohydrates and similarities to the host species' own glycome.[12] Antibodies are also expensive to produce, can have poor stability, and require large quantities of difficult-to-obtain glycan for production.

There have been numerous attempts to generate novel GBPs with higher affinities or different specificities by engineering of existing GBPs, as described in Chapter 1.[13, 14] These attempts have employed techniques such as mutation of a single amino acid to increase affinity to the native ligand as reported for residue Asn41 in peanut agglutinin, or more complicated directed-evolution approaches to change the specificity of the GBP to structurally distinct sugars, such as engineering of the Gal-binding earthworm lectin EW29 to a sialic acid-binding lectin.[15-18]

Recently, lamprey variable lymphocyte receptors (VLRBs) have shown promise as glycan recognition reagents.[19, 20] This scaffold, referred to as a "lambody", is a leucine-rich repeat protein that is part of the lamprey adaptive immune response. McKitrick and coworkers have shown that glycan-binding properties can be engineered by directed evolution of naïve libraries of VLRBs, or by elicitation in lampreys by injection of whole fixed cells, tissue homogenates, and human milk followed by rounds of *in vitro* selection.[19-22] Despite these advances, VLRBs can have properties nonideal for the creation of reagents, such as their large size and difficulties in purification in bacterial systems.

This chapter and the next will focus on the creation of GBPs for two glycans. The first is N-acetylneuraminic acid, known as Neu5Ac or sialic acid. Sialic acids are a class of alpha-keto acid sugars containing a nine-carbon sugar backbone, and Neu5Ac is the most common member of the group. Sialic acids are primarily located at the terminal non-reducing end of N- and O-glycans on glycoproteins and glycolipids, but are also common in bacterial polysaccharides and lipooligosaccharides.[23] Since they are terminal sugars of cell surface glycans, they interface with the environment including host immune system and pathogens. Because of this, there are many known lectins and adhesins that bind sialic acid. Sialic acid is an important molecule in human physiology playing diverse roles such as cell-cell recognition, signaling, tumor growth, bacterial and viral infections, and development.[24]

The second is a glycan that has been the subject of several GBP engineering efforts called the Thomsen-Friedenreich antigen, also known as the T or TF antigen.[15, 19, 25, 26] The TF antigen is the core 1 structure of O-linked mucin-type glycans, consisting of the disaccharide Galβ1-3GalNAcα linked to certain Ser or Thr residues. In healthy tissue, the core 1 structure is further modified by the addition of sialic acids, sulfates, or other glycans, but is often unsialylated in

cancerous and pre-cancerous tissue due to abnormal expression of glycotransferases and glycosidases in tumor cells. The epitope is a tumor associated carbohydrate antigen found in 90% of carcinomas, making it an important epitope for cancer diagnostics and immunotherapies.[27, 28]

In this chapter I present the development of a novel platform to generate GBPs, utilizing sialic acid and the TF antigen as model carbohydrate targets. The GBP development platform leverages yeast surface display-based directed evolution, utilizing a novel protein scaffold as our GBP library. The archaeal protein, Sso7d, is small, stable, and has features that show promise for GBP generation. As this protein is not natively known to bind carbohydrates, it is prudent to carefully develop our platform towards enriching Sso7d-based glycan-binding proteins and include appropriate controls to ensure secondary reagent or linker binding is not developed. As such, this chapter will rationalize the use of Sso7d as a GBP scaffold, and discuss the failures encountered and ultimate successes for platform development. Briefly, multivalent biotinylated versions of both sialic acid and TF antigen were used as the selection bait in the initial selection attempts. These biotinylated ligands proved unsuccessful due to the prevalence of biotin-binding variants being preferentially enriched over glycan binding variants in the population. A strategy using non-biotinylated ligands was then successfully developed and applied to the directed evolution of sialic acid binding variants. The result is a robust directed evolution platform that enables the generation, and subsequent engineering, of Sso7d-based GBPs.

2.3 Sso7d as a diversity generating scaffold

All directed evolution projects begin with selection of the appropriate scaffold protein. The term ‘scaffold’ refers to the protein framework that will be acted upon during the evolution process to obtain variants with certain desired characteristics. In general, the scaffold protein must have a defined tertiary structure that is amenable to mutations and insertions without loss of stability. This

is equivalent to the hypervariable complementarity-determining regions (CDRs) of antibodies that are capable of a high degree of sequence variability without compromising the overall folding and stability of the antibody structure. Although antibodies are excellent for creating diverse protein libraries, they suffer from many drawbacks that limit scalable recombinant production. The large, multidomain proteins contain disulfide bonds and glycosylation sites, rendering prokaryotic production problematic. Instead, mammalian expression systems must be implemented, increasing the overall production times and costs. The large size also makes many surface display systems challenging.

Alternate non-antibody scaffolds have shown great utility in the production of proteins for molecular recognition that are free from the drawbacks of antibodies. These include DARPins (ankyrin repeat proteins), knottins (highly stable cystine-rich miniproteins), affibodies (Z-domain of *S. aureus* protein A), and anticalins (derived from human lipocalins).[29, 30] The Sso7d DNA-binding protein of the hyperthermophilic archaea *Sulfolobus solfataricus* is a scaffold protein that has remarkable thermal stability with a melting temperature of nearly 100 ºC.[31] The protein is small at only 63 amino acids (**Figure 2-1**) and has no cysteine residues or glycosylation sites. It folds into an SH3-like fold with an incomplete β-barrel, containing five β- stands and one α-helix. The Sso7d scaffold has been used in many engineering efforts to bind proteins, small molecules (e.g. fluorescein), peptides, and even iron oxide.[32-34] Sso7d is highly stable under a wide pH range and when exposed to denaturing agents, even after the protein has undergone extensive mutagenesis.[32] The protein is also highly expressed in the cytoplasm of *E. coli*. Due to these properties, evolved Sso7d variants are superior affinity reagents to antibodies in terms of both cost of production and protein stability.[35, 36]

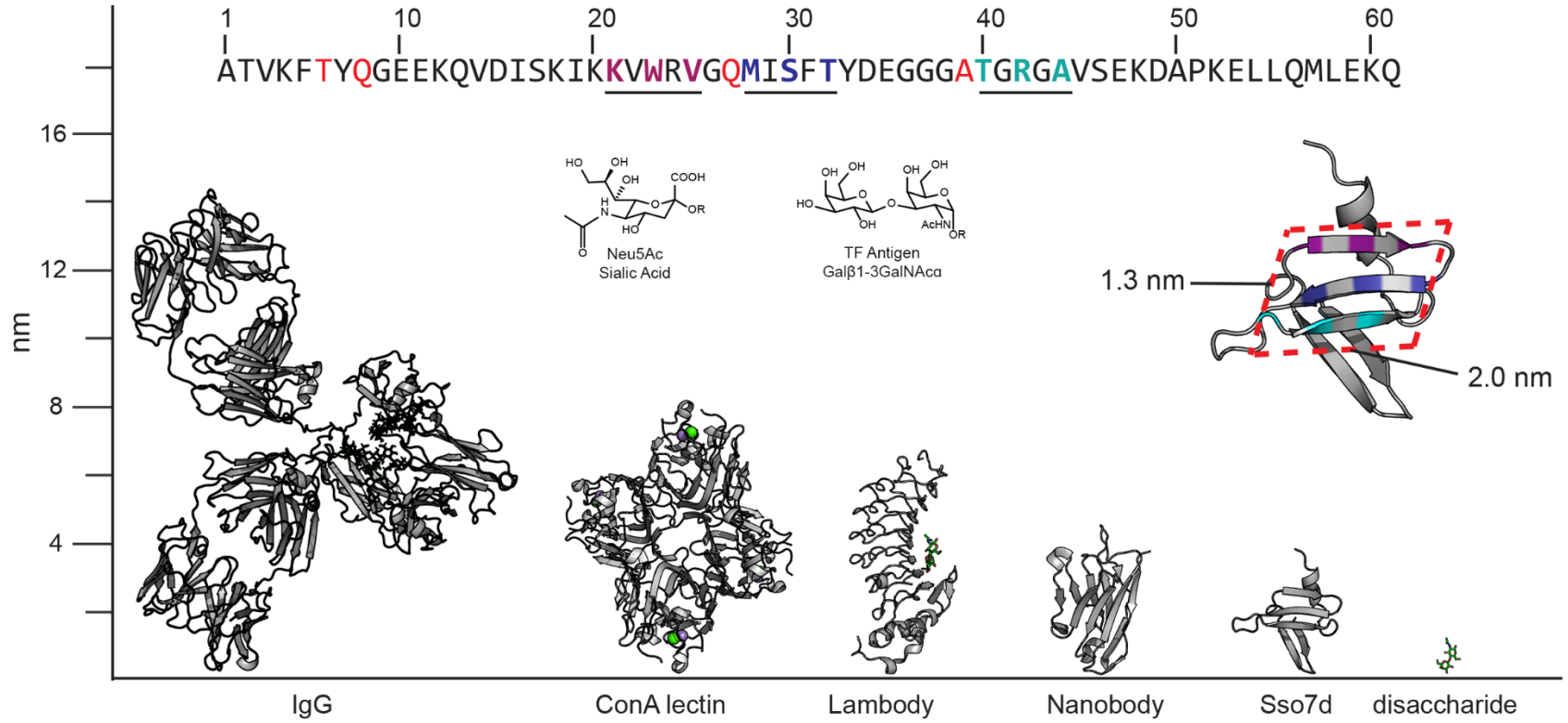


Figure 2-1. Approximate sizes of common glycan-binding proteins.
Phenobarbitol specific IgG1 antibody (PDB 1IGY), ConA lectin (PDB 1VAM), TF specific variable lymphocyte receptor of lamprey or "lambody" (PDB 5UFC), anti-SARS-CoV-2 spike nanobody mNb6 (PDB 7KKJ) and the DNA binding protein Sso7d (PDB 1SSO) compared to a disaccharide (TF antigen). Inset shows sequence of rcSso7d with residues mutated from Lys colored in red, and variable residues colored in maroon, dark blue, and light blue. The Sso7d structure shows variable residues color coded to those in the sequence as well as dimensions of the binding face. Sialic acid and TF antigen structures shown in inset.

Sso7d was chosen as a scaffold for GBP development because it meets the following criteria: stable to a range of temperatures, efficient expression in *E. coli*, amenable to conjugation with various tags for the creation of useful reagents, lack of disulfide bonds or glycosylation sites, and binding site dimensions and features comparable to those seen in other GBPs. The Sso7d scaffold contains a binding site with dimensions of approximately 1.3 nm by 2.0 (**Figure 2-1**) nm that can accommodate disaccharides and nine-carbon nonulosonic acids. Sso7d shares many features of native GBPs. First, the binding face of Sso7d is a relatively flat, solvent-exposed surface comprising three β-sheets. Shallow, solvent-exposed binding surfaces are common in lectins.[37] Second, the evolved Sso7d protein can be highly enriched in aromatic amino acids in its binding face without a subsequent loss in thermostability, expression level, and solubility.[38] Analysis of the amino acid residues involved in natural carbohydrate-protein interactions reveals a striking enrichment of aromatic amino acids.[39, 40] The polarized C-H bonds of carbohydrates are able to form CH-π interactions with the electron rich π-systems of aromatic amino acids.[41] Due to these properties the Sso7d was utilized as a scaffold for development of new GBPs.

2.4 Features of rcSso7d library

The Sso7d library used in this work was generated by Traxlmayr and coworkers in the Wittrup Lab at Massachusetts Institute of Technology.[38] Since Sso7d is a DNA-binding protein, it has a high degree of surface positive charge with the binding face surrounded by a ring of lysine residues. This positive charge can cause nonspecific interactions with negatively charged surfaces and potentially block interactions with positively charged targets. To decrease complication from these features, Traxlmayr and coworkers engineered a reduced-charge variant of Sso7d (rcSso7d) by making mutations K6T, K8Q, K27Q, K39A, and truncating the protein after Q61 to remove the terminal two lysine residues (see **Figure 2-1**). This reduced the net charge from +7 to +1 without

significantly changing the melting temperature of the protein (rcSso7d T_m of 95.5 ºC vs the WT Sso7d T_m of 98 ºC). This charge reduction significantly decreased nonspecific binding to mammalian HeLa cells compared to WT Sso7d which is important for downstream applications of the evolved rcSso7d variants.

Nine solvent-exposed residues in the three β-sheet binding face of rcSso7d, referred to as the paratope, were selected for randomization (**Figure 2-1**). Two libraries were created. The first containing amino acids found in protein-protein interactions using two data sets: an alanine scanning study by Bogan and Thorn that identified enriched amino acids in protein-protein interaction hot spots, and a study by Zemlin and coworkers that identified the frequency of amino acids in CDR-3 loops of human and mouse antibody heavy chains.[42, 43] This library, referred to as rcSso7d-11, contains an equal frequency (9.1%) of 11 different amino acids at all nine positions (amino acids Ala, Asp, Gly, His, Ile, Lys, Asn, Arg, Ser, Trp, Tyr). A second library, rcSso7d-18, was created with all amino acids except Pro and Cys at an equal frequency (5.6%). Trinucleotide synthesis was performed to control the frequency of amino acids in a precise manner. To allow for yeast surface display of the scaffold variants, the N-terminus of the rcSso7d variants were fused to the C-terminus of yeast protein Aga2P which forms disulfide bonds to the cell membrane anchored Aga1P (**Figure 2-2A**). An HA tag was fused to the N-terminus of the rcSso7d variant and a c-Myc tag fused to the C-terminus for confirmation of expressed, full length proteins. Both libraries were determined to have a high level of surface expression, indicating the majority of proteins are full length and properly folded.

Each library contained 1.5 x 10^9 members. To select glycan binding variants from the libraries, two types of sorts were performed. The first is magnetic-activated cell sorting (MACS) where ligands are adsorbed or coupled to the surface of superparamagnetic particles consisting of

iron oxide in a polymer shell, incubated with the yeast library, then used to separate binding cells from non-binding cells.[44] Positive selections have the glycan ligand immobilized to the magnetic bead surface and the selected yeast are grown to enrich the population in binding variants. Negative selections are utilized to remove yeast that bind to the bead itself or to the linker region of the glycan ligand, and the selected yeast are moved to waste to deplete the population in off-target binders. MACS is useful for quickly processing large libraries and is the first step in sorting the naïve Sso7d libraries. In all future experiments, the first selections with the naïve library are magnetic bead sorts of both library rcSso7d-11 and rcSso7d-18 sorted individually and then pooled once the theoretical diversity drops below 10^7 cells. After the magnetic bead sorts the library diversity was low enough to be processed with the second type of sorts, based on fluorescence-activated cell sorting (FACS) where fluorescent glycan ligands are incubated with the yeast library and a cell sorter is used to isolate fluorescent cells bound to the ligand. FACS is performed with dual labeling for both protein expression with an anti-HA antibody and ligand binding with the fluorescently labeled streptavidin. This allows not only for higher affinity variants to be isolated but for more stable variants to be isolated as expression efficiency is correlated to protein stability.[45] Many directed evolution projects utilize biotin on their ligands due to the ease of use and availability of streptavidin reagents, so initial selections proceeded with biotinylated ligands.

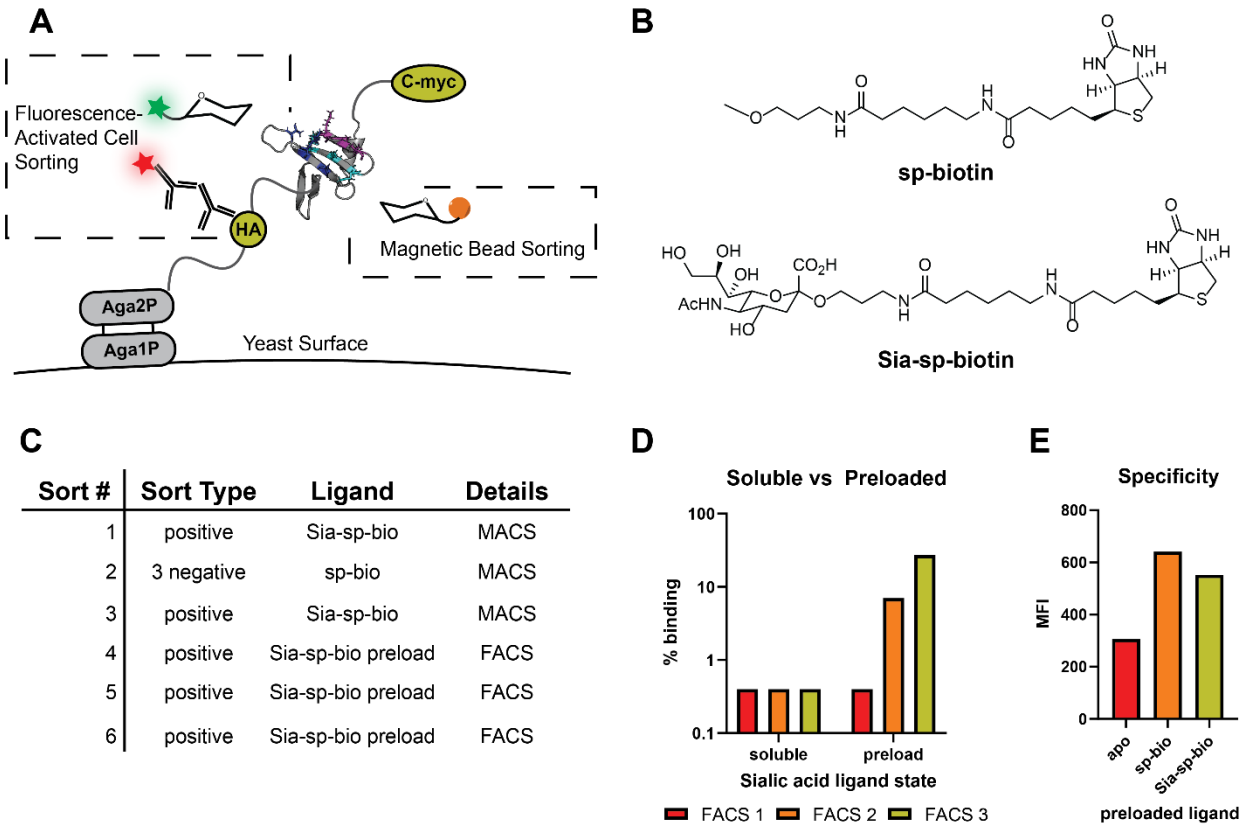
2.5 Library selections using biotinylated ligands

2.5.1 Early sort conditions

Early sort conditions were explored in collaboration with Dr. Cristina Zamora. Initially, sorts were carried out with a commercially available sialic acid glycosylated directly to a short hydrocarbon spacer (sp) with a terminal biotin, termed Sia-sp-biotin (**Figure 2-2B**). The control ligand containing just the sp linker and biotin (sp-biotin) was synthesized and characterized by

Teresa Naranjo Sanchez (**Figure 2-2B**). Biotin binder Dynabeads, 2.8 μm superparamagnetic beads coated in streptavidin, were incubated with Sia-sp-biotin or sp-biotin. Two positive selections with immobilized Sia-sp-biotin, and three negative selections with sp-biotin were performed on each library, bringing the diversity from 10^9 to 10^6 for each as determined by plate counts of isolated yeast. Libraries were then combined for further FACS selections.

Binding to the monovalent Sia-sp-biotin ligand was too weak to be detected at concentrations practical for use in solution. To improve this, pre-complexation of Sia-sp-biotin with the tetrameric secondary reagent streptavidin-AF647 was performed (**Figure 2-2D**). This creates a higher effective concentration when in proximity to the yeast surface and can be used to isolate weak binders. This preloading step is considered to create an intermediate concentration between the highest practical in-solution concentration of ligand one is willing to use and the effective concentration on magnetic beads. After three rounds of FACS, capturing all binding cells over background in each round, there was a significant enrichment in Sia-sp-biotin binding variants. However, these variants did not distinguish between Sia-sp-biotin and sp-biotin preloaded to streptavidin, while no binding above background was seen to apo streptavidin (**Figure 2-2E**).



Sort #	Sort Type	Ligand	Details
1	positive	Sia-sp-bio	MACS
2	3 negative	sp-bio	MACS
3	positive	Sia-sp-bio	MACS
4	positive	Sia-sp-bio preload	FACS
5	positive	Sia-sp-bio preload	FACS
6	positive	Sia-sp-bio preload	FACS

Figure 2-2. Early sort conditions
A) Yeast surface display system showing Sso7d fused to yeast protein Aga2P that makes disulfide bonds to yeast surface protein Aga1P with both N-terminal HA tag and C-terminal Myc tag. Yeast can be sorted using glycan conjugated magnetic beads or fluorescent glycan ligands. B) Structures of sp-biotin negative control and Sia-sp-biotin. C) Table of sorts performed using sp-biotin and Sia-sp-biotin. D) FACS rounds 1-3 showing the percentage of binding cells in each round using soluble ligand or ligand preloaded to fluorescent streptavidin. E) The specificity of the final population collected showing equivalent binding to the linker control and Sia-sp-bio (median fluorescence intensity, MFI).

2.5.2 Alternating ligand selection

The previous results indicate that binding to the linker arose over the course of the selections. To prevent the accumulation of linker-binding variants, a strategy was developed that alternated between two sialic acid-bearing ligands, each containing a different linkage type. Even if linker-binders were selected in one round, they would be lost in the subsequent round due to the lack of the specific epitope in the alternate ligand. This strategy was used previously to successfully isolate glycan-binding lambodies.[19] The first ligand is a biotinylated glycoprotein glycophorin A

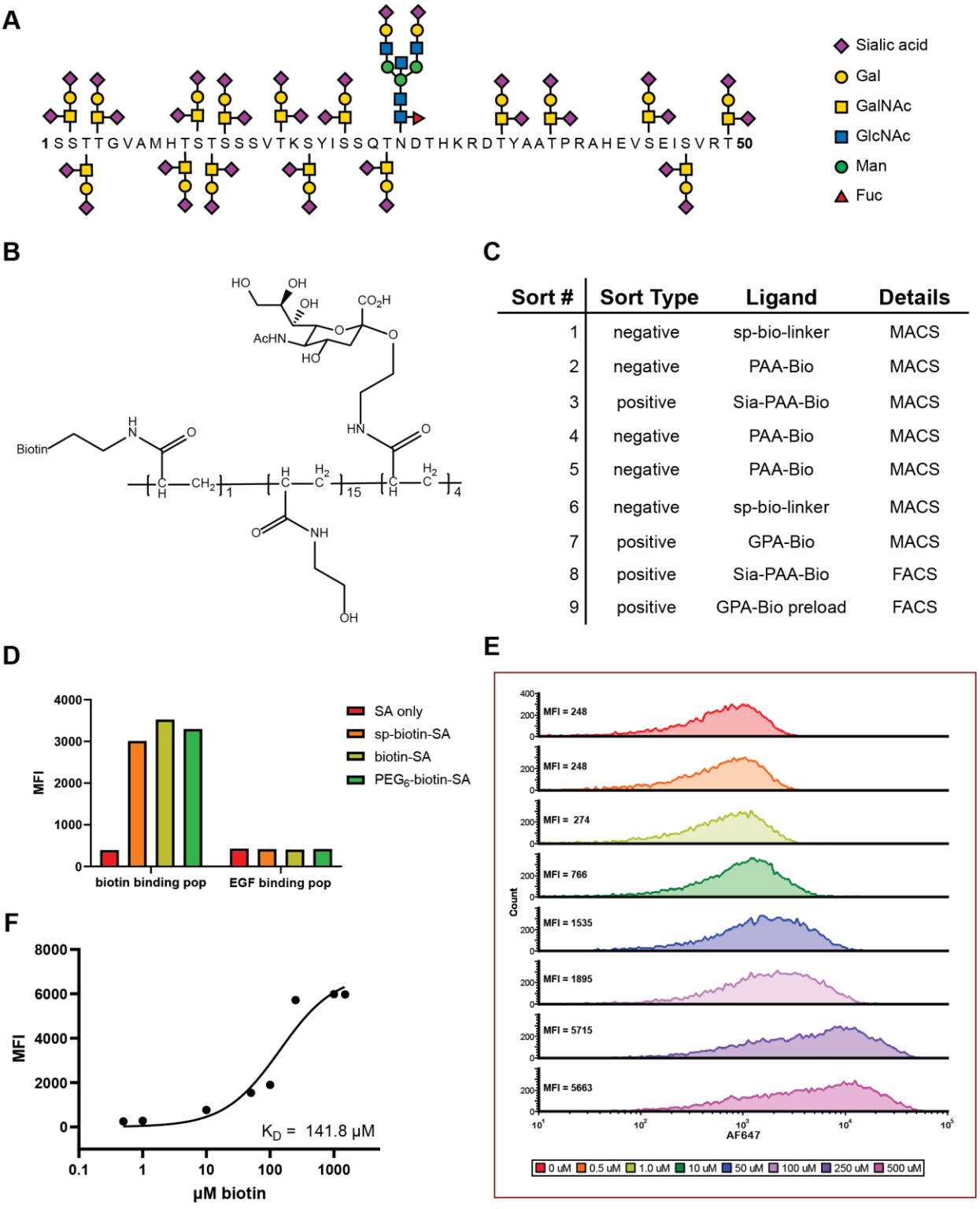
(GPA). GPA is a major component of red blood cell membranes and is a glycoprotein with a high degree of sialylated O-glycans (**Figure 2-3A**).[46] The second ligand is a multivalent, biotinylated polyacrylamide glycopolymer containing approximately twenty sialic acids and five biotins per polymer (Sia-PAA-Bio, **Figure 2-3B**).[47]

The library underwent five rounds of negative selection with the sp linker and the non-glycosylated biotinylated polyacrylamide control polymer (PAA-Bio) and two rounds of positive selection on magnetic beads with Sia-PAA-Bio and GPA-bio. Then two rounds of FACS were performed: the first with Sia-PAA-Bio collecting all binding cells above baseline, then with preloaded streptavidin-GPA collecting the top 0.1% of binding cells. The resulting population still showed binding to both streptavidin preloaded Sia-sp-biotin and sp-biotin despite the alternating strategy.

The only common feature between all tested ligands was the biotin used for immobilization onto scaffolds. Therefore, further flow cytometry analysis with streptavidin preloaded to biotin, sp-biotin or PEG_6-biotin were performed. The population shows equivalent binding to streptavidin preloaded with biotin, sp-biotin, and PEG_6-biotin, while apo streptavidin shows no binding signal (**Figure 2-3D**). This population was compared to a population that was previously evolved by Dr. Zamora to bind the cytokine epidermal growth factor (EGF) using biotinylated proteins and preloading of streptavidin. Compared to the population sorted with the biotinylated sialic acid ligands, the EGF-evolved population shows no binding to streptavidin preloaded with biotinylated ligands.

To determine if the binding is to biotin itself or the biotin-streptavidin complex, the population was then tested with soluble biotin by two-step labeling. At the highest concentration of 30 μM, there is a low level of binding in the population. A yeast surface titration was performed

by creating clonal yeast expressing a single variant isolated from this library, called 83H. Binding starts to develop around 50 μM and plateaus at 250 μM (**Figure 2-3E and F**). Fitting of the generated curve give a K_D of ~140 μM. This data shows that Sso7d variants binding to the small molecule biotin was enriched over the course of the selection process when biotinylated ligands were used.



Sort #	Sort Type	Ligand	Details
1	negative	sp-bio-linker	MACS
2	negative	PAA-Bio	MACS
3	positive	Sia-PAA-Bio	MACS
4	negative	PAA-Bio	MACS
5	negative	PAA-Bio	MACS
6	negative	sp-bio-linker	MACS
7	positive	GPA-Bio	MACS
8	positive	Sia-PAA-Bio	FACS
9	positive	GPA-Bio preload	FACS

Figure 2-3. Alternating ligand strategy.
A) Amino acids 1-50 of human GPA showing sites of O-linked and N-linked sialic acid-containing glycans. B) Structure of Sia-PAA-Biotin polymer. C) Table of sorts performed using the alternating ligand strategy. D) Final biotin-binding population and unrelated EGF-binding

population tested with streptavidin (SA, red), preloaded sp-biotin (sp-biotin-SA, orange), preloaded biotin (biotin-SA, yellow), and preloaded PEG_6-biotin (PEG_6-biotin-SA, green). E) Yeast displaying single biotin-binding variant 83H tested with two step labeling to biotin then streptavidin. F) Yeast displaying single biotin-binding variant 83H tested with two step labeling to biotin at 0, 0.5, 1.0, 10, 50, 100, 250, and 500 μM biotin fit to sigmoidal binding curve.

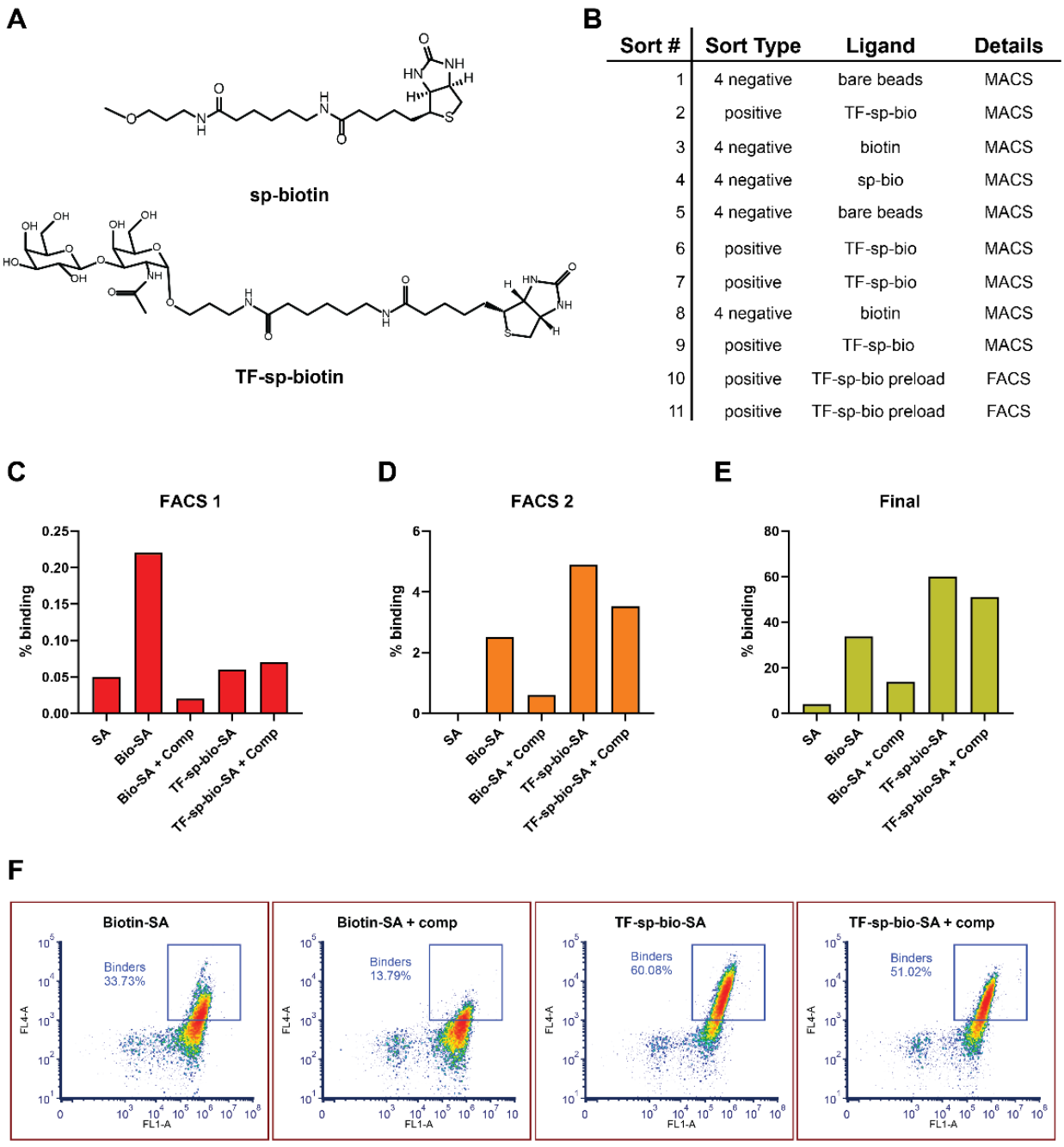
2.5.3 Increased negative selection and competition for removal of off-target binders

Because biotin-binding was enriched and not linker-binding as previously suspected, new strategies to remove biotin-binding variants early were tested. An increased number of negative selections during the initial magnetic bead sorts could deplete the population in these off-target binders and allow for the enrichment of sugar-binding variants. These experiments were performed with the TF antigen. The number of overall negative selections were increased four-fold from five total to twenty total against bare streptavidin coated beads, biotin bound beads, and sp-biotin bound beads. Four positive selections against TF-sp-biotin bound beads were also performed. Flow cytometry of the population after MACS shows no discrimination between biotin preloaded with streptavidin and TF-sp-bio preloaded with streptavidin. Increasing negative selections did not deplete the biotin-binding variants from the library, and they were still enriched in the population.

Increased negative selection alone was not enough to rid the library of biotin-binding variants. Another strategy explored was performing positive selections for TF-sp-biotin in the presence of excess soluble competitor. For MACS, excess soluble biotin-streptavidin was included during all four positive selections. This would ensure that yeast displaying a biotin-binding variant would have all sites occupied by the soluble biotin-streptavidin and would not be bound by the beads containing TF-sp-biotin. Twenty rounds of negative selection towards bare beads, biotin, and sp-biotin were interspersed between the positive selections (**Figure 2-4A**). For FACS, cells were incubated with TF-sp-biotin preloaded to a fluorescent streptavidin with excess competitor

(biotin preloaded to a non-fluorescent streptavidin). Any cells displaying biotin-binding variants in the population would therefore have all sites occupied by the non-fluorescent competitor, allowing for capture of cells binding only to the fluorescent TF-sp-biotin preloaded streptavidin.

All experiments carry a control with biotin preloaded fluorescent streptavidin with and without competition. This is a measure of how efficiently biotin binders are being removed from the population and allows for appropriate gates to be set up for FACS selection of TF-only binding cells. The percent of cells showing binding to the ligand provided is low for the first round of FACS (**Figure 2-4C**) but increases in round two (**Figure 2-4D**). Despite conservative gating to remove any biotin binding cells, each population still contained biotin binders. However, there is an increased percentage of TF binding in the second round of FACS and in the final population. Additionally, this TF binding was not totally depleted with added competitor. It is likely that a subset of the selected population can recognize TF, but it is mixed with a significant population of biotin-binding variants that are difficult to remove. Since biotin binding keeps persisting in the population despite many efforts put in place to attenuate it, it will be a consistent problem for all future GBP evolution projects. Many directed evolution projects utilize biotin on their ligands due to the ease of use and availability of streptavidin reagents but removing biotin from the system will be necessary to streamline GBP engineering.



Sort #	Sort Type	Ligand	Details
1	4 negative	bare beads	MACS
2	positive	TF-sp-bio	MACS
3	4 negative	biotin	MACS
4	4 negative	sp-bio	MACS
5	4 negative	bare beads	MACS
6	positive	TF-sp-bio	MACS
7	positive	TF-sp-bio	MACS
8	4 negative	biotin	MACS
9	positive	TF-sp-bio	MACS
10	positive	TF-sp-bio preload	FACS
11	positive	TF-sp-bio preload	FACS

Figure 2-4. Sorts with added competition.
A) Structure of sp-biotin control and TF-sp-biotin ligands. B) Table of sorts performed with added preloaded biotin-SA competitor. C-E) FACS sorts 1 and 2 and final resulting population binding signals with and without competition. F) Final population tested with preloaded biotin-SA with and without competition, and preloaded TF-sp-biotin-SA with and without competition.

2.6 Library selections without biotin-streptavidin

2.6.1 Tosyl-activated magnetic bead selections

Magnetic beads are available with a number of different chemistries for immobilization of ligands; we leveraged this to avoid the use of biotin-based immobilization. Tosyl-activated Dynabeads can be used for immobilization of ligands containing primary amines. Each bead contains 9-18 x 10^8 of *p*-toluenesulfonyl (tosyl) groups per bead compared to the Biotin Binder Dynabeads with 10^6 streptavidins per bead. Dr. Nathaniel Shocker synthesized sialic acid with a linker to a primary amine for immobilization on tosyl-modified beads (**Figure 2-5A**). Typically, tosyl-activated Dynabeads are used for protein and peptide conjugation where the level of protein can be assessed by UV-VIS at 280 nm. Glycan-NH_2 ligands cannot be assessed this way as they are not UV active, so confirmation of conjugation must be measured in a different way. Upon reaction with the amine-bearing ligand, the tosyl group is released into the solution and can be measured by UV/vis spectroscopy by observing a λ_{max} at 220 nm (**Figure 2-5B**). A standard curve of p-toluenesulfonic acid was made and used to assess tosylate release after incubation with the amine-bearing ligands (**Figure 2-5C**). After an 18-hour incubation with buffer only, linker-NH_2, or Sia-NH_2 the tosylate released into solution was measured. There is some background hydrolysis of the tosyl group in the buffer only sample of about 40 μM, but Sia-NH_2 was well over background at about 170 μM. The linker-NH_2 was less efficient than the Sia-NH_2 at only about 97 μM. Subtracting out the background hydrolysis leaves beads with approximately 130 μM Sia-NH_2 and negative control beads with approximately 60 μM linker-NH_2. Compared to the biotin binder Dynabeads, this is 10x and 5x over the number of biotinylated sugars immobilized to the same number of beads and should be sufficient for MACS. MACS was performed for a total of five positive selections interspersed with two to three negative selections with linker beads. Bare beads

were not used for negative sorts as they were quite hydrophobic and tended to non-specifically bind the yeast.

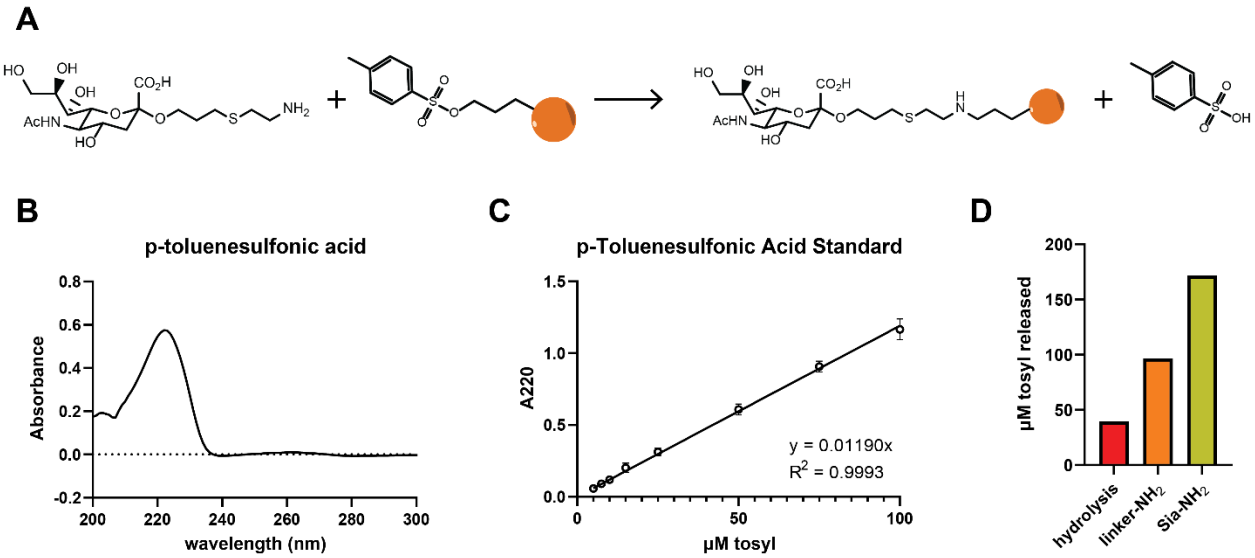


Figure 2-5. Conjugation of ligands to tosyl-activated magnetic Dynabeads.
A) Reaction of amine-modified sialic acid with tosyl activated Dynabeads. B) UV-VIS trace of 50 µM p-toluenesulfonic acid. C) Standard curve of p-toluenesulfonic acid from 5-100 µM. D) Concentration of tosyl released after conjugation with linker-NH_2, Sia-NH_2, or buffer only (hydrolysis) control.

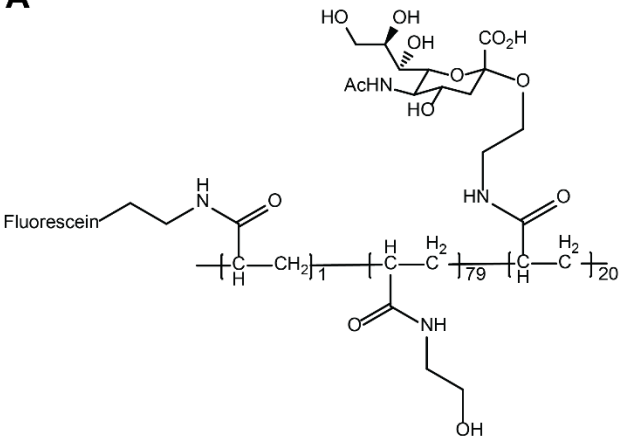
2.6.2 FACS with multivalent glycopolymers

A multivalent glycan ligand without biotin is required to enrich the library in glycan-binding Sso7d variants. Previous attempts used preloaded streptavidin with monovalent biotinylated sugars, a biotin-labeled densely glycosylated glycoprotein, or a biotinylated polyacrylamide (PAA) glycopolymer. A similar PAA glycopolymer is commercially available with a fluorophore instead of biotin, consisting of 20 mol% glycan and 1 mol% fluorescein, referred to as Sia-PAA-FITC (**Figure 2-6B**).

The population after five rounds of positive magnetic bead sorts was sorted by FACS with Sia-PAA-FITC. As expected, a low percentage of binders were present in the first round, but the percentage of binding cells increased on subsequent rounds of selection while non-glycosylated

control polymer PAA-FITC showed no enrichment in the population. With the introduction of the new reagents, tosyl-activated beads and PAA-FITC glycopolymers, enrichment of binders to the glycan target was successful.

A



B

Sort #	Sort Type	Ligand	Details
1	positive	Sia-NH_2	MACS
2	3 negative	linker-NH_2	MACS
3	positive	Sia-NH_2	MACS
4	2 negative	linker-NH_2	MACS
5	positive	Sia-NH_2	MACS
6	2 negative	linker-NH_2	MACS
7	positive	Sia-NH_2	MACS
8	2 negative	linker-NH_2	MACS
9	positive	Sia-NH_2	MACS

C

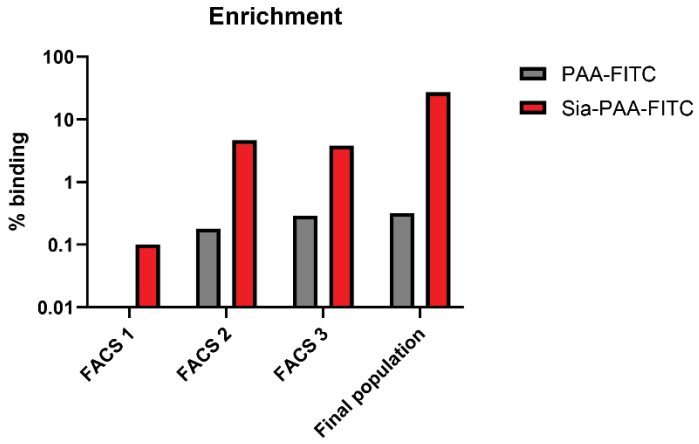


Figure 2-6. Enrichment of binding variants with Sia-PAA-FITC glycopolymer.
A) Structure of Sia-PAA-FITC. B) Tosyl bead sorts prior to sorts with Sia-PAA-FITC. C) Enrichment of binding variants for Sia-PAA-FITC compared to PAA-FITC over the course of three rounds of FACS.

2.6.3 Affinity maturation produces higher affinity binders

The isolated binders were low affinity and required a high concentration of glycopolymer for enrichment. To improve the binding affinity, the plasmids isolated from the final population underwent affinity maturation. First, the Sso7d gene underwent error prone PCR with addition of the nucleotides 8-oxo-2'-deoxyguanosine-5'-triphosphate (8-oxo-dGTP) and 2'-deoxy-p-nucleoside-5'-triphosphate (dPTP).[48] These two nucleotides act as mutagens, randomly incorporating into the gene during PCR and causing transition mutations to occur. The concentrations of nucleotide analogs used should give an average of one mutation per 500 base pairs. The newly randomized genes were then cloned back into the yeast display vector and transformed back into yeast to create a new variant library with a modest theoretical diversity of 4 x 10^6. The library was then sorted using a C-Myc antibody to remove any truncated variants from the population.

The new library underwent three rounds of FACS with four-fold less Sia-PAA-FITC to select variants with improved binding affinity. The median fluorescent intensity increased with each round of FACS, showing that binders were being enriched with each round (**Figure 2-7A**). The plasmids were isolated for sequencing, and a total of 182 variants produced 97 unique sequences. Pairwise distances were calculated based on the Blosum 62 substitution matrix, the results were hierarchically clustered, and a tree was built using the UPGMA method (**Figure 2-7B**). The sequences formed two large clusters with very distinct paratope sequences (**Figure 2-7C**). Forty sequences were selected that included multiple sequences from each major cluster, any repeated sequences, and the sequences with the largest distances from the other sequences. Clonal yeast cultures were made for each selected sequence and the binding of Sia-PAA-FITC was

analyzed using flow cytometry (**Figure 2-7D**). One sequence, 1.4.B, stood out above all other tested variants and was selected for further study.

Variant 1.4.B appeared twice out of the 182 plasmids sequenced. 1.4.B was compared to a variant selected from the library before affinity maturation called variant 0.8.E with a similar paratope sequence (**Figure 2-7F**). Both contain four aromatic residues in the paratope region: W28, W32, Y40, and Y44. This many aromatics could indicate multiple binding regions for carbohydrates in the binding site as is seen in some processive polysaccharide-processing enzymes, or could form a hydrophobic/aromatic "cage" as is seen in a disaccharide-binding lamprey variable lymphocyte receptor.[49, 50] They also have two basic residues in the paratope: R23 and K30. Outside of the paratope region, there are mutations of Q8R, D15G, K18R, G37G, and Q61R compared to the parental rcSso7d sequence. Strikingly 1.4.B has an additional basic residue in the paratope sequence, R18. Sialic acid is negatively charged at neutral pH, so gaining an additional basic residue in the binding face could explain the increased affinity of 1.4.B. This increased affinity is demonstrated by the yeast surface titration performed with 0.8.E and 1.4.B, where increasing concentrations of Sia-PAA-FITC was tested with each variant as a yeast surface fusion (**Figure 2-7D**). From this, functional affinity constants or K_{app} could be calculated as 2.5 ± 0.7 μM for 0.8.E and 1.5 ± 0.1 μM for 1.4.B. This is not a true K_D measurement due to the multivalent nature of the interaction.

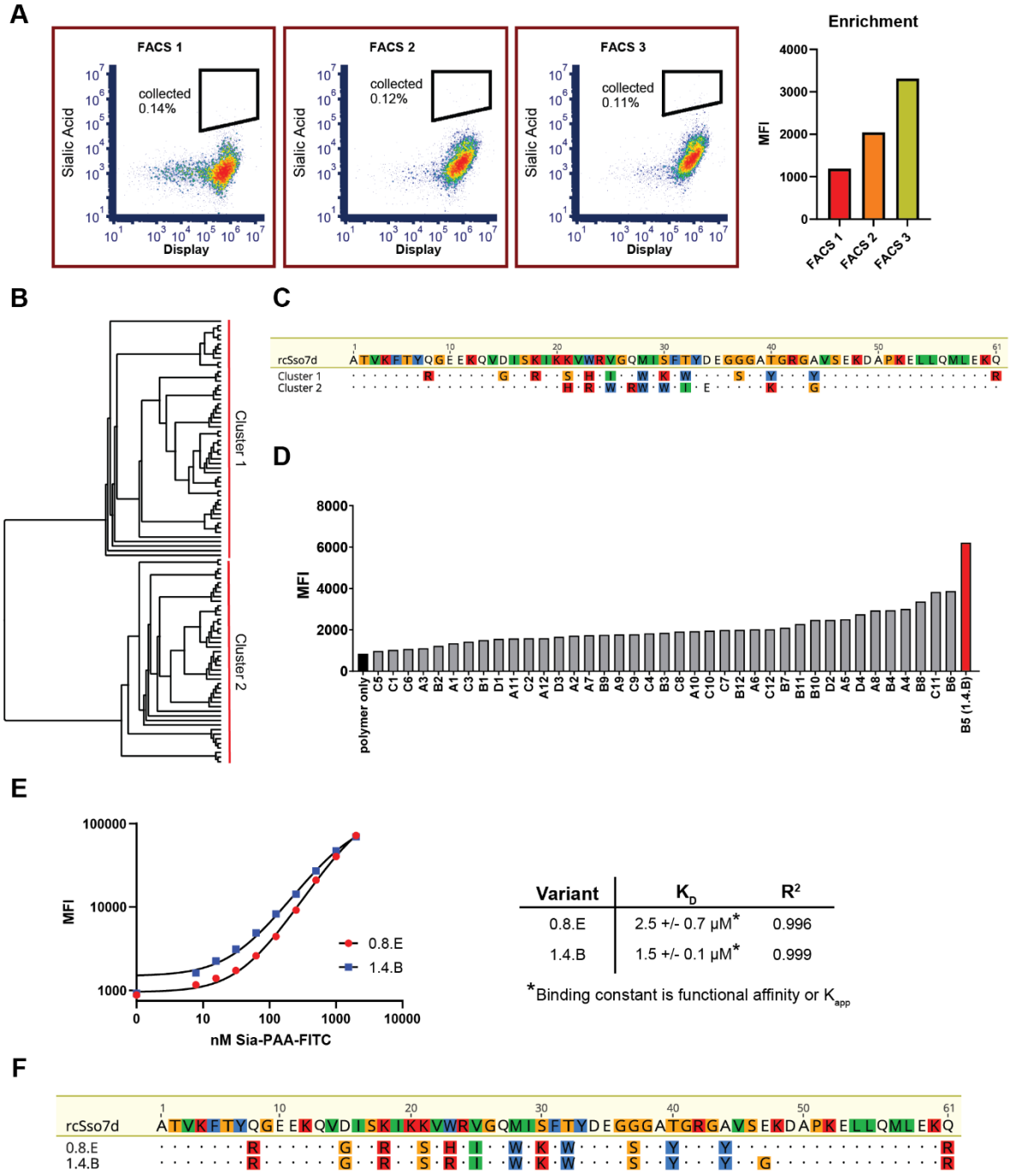


Figure 2-7. Affinity maturation improves binding affinity.
A) Dot plots and bar graph of median fluorescence intensity for each round of FACS showing enrichment in binding variants. Y-axis is sialic acid binding, and x-axis is expression of Sso7d on surface. Gates show cells collected during the sort. B) Tree of sequenced variants showing two distinct clusters. C) Sequences of the consensus sequence for cluster 1 and 2 from tree compared to rcSso7d parental sequence. D) Median fluorescence intensity of forty different variants as clonal yeast cultures. Polymer only in black and best binder 1.4.B in red. E) Yeast surface titration of

variant 0.8.E and affinity matured variant 1.4.B. fit with a sigmoidal binding function. Table shows K_{app} and R^2 of the fitting. F) Sequences of variants 0.8.E and 1.4.B compared to parent rcSso7d.

2.6.4 Competition can improve selectivity of isolated variants

The specificity of 1.4.B was tested with a panel of other glycopolymers that are highly similar or expected to be in proximity to sialic acid on glycoconjugates (**Figure 2-8A**). A striking result is that 1.4.B showed little affinity for N-glycolylneuraminic acid (Neu5Gc), a sugar that differs from sialic acid by the addition of one hydroxyl group. Variant 1.4.B was able to bind to other glycopolymers containing sialic acid such as the polysialic acid disaccharide Neu5Acα2-8Neu5Acα and Sia-TF trisaccharide. 1.4.B did show some off-target binding to several disaccharides containing no sialic acid. This includes chitobiose (GlcNAcβ1-4GlcNAcβ) which makes up the core of N-linked glycans, Le^c (Galβ1-3GlcNAβ) which is a Lewis antigen precursor, and TF (Galβ1-3GalNAcα).

A potential method to reduce this off-target binding to other glycans is through competition. This would involve the inclusion of a competitor glycopolymer during sorts with Sia-PAA-FITC. This was explored previously to remove biotin-binders using a non-fluorescent streptavidin-biotin competitor. However, inclusion of a glycopolymer with an orthogonal fluorophore would allow for a direct assessment of specificity and discrimination between yeast displaying variants that preferentially bind the competitor from those that preferentially bind sialic acid (**Figure 2-8C,D**). This strategy was used previously to evolve a high-specificity antibody for phosphorylated tau.[51] Variant 1.4.B underwent error prone PCR and a c-Myc sort to create a new library of full-length variants for competitive sorts. The competitor TF-PAA-Bio was introduced during sorts and visualized with streptavidin AF647 during FACS. The dot plot from the first round of FACS showed two groups of binding populations: one that preferred TF and one that bound both sialic acid and TF. After selecting the preferential sialic acid binders, the second round of

FACS showed that the preferential TF binding had been lost and most of the cells bound to both sialic acid and TF. The final population showed increased sialic acid signal and decreased TF binding signal (**Figure 2-8E**).

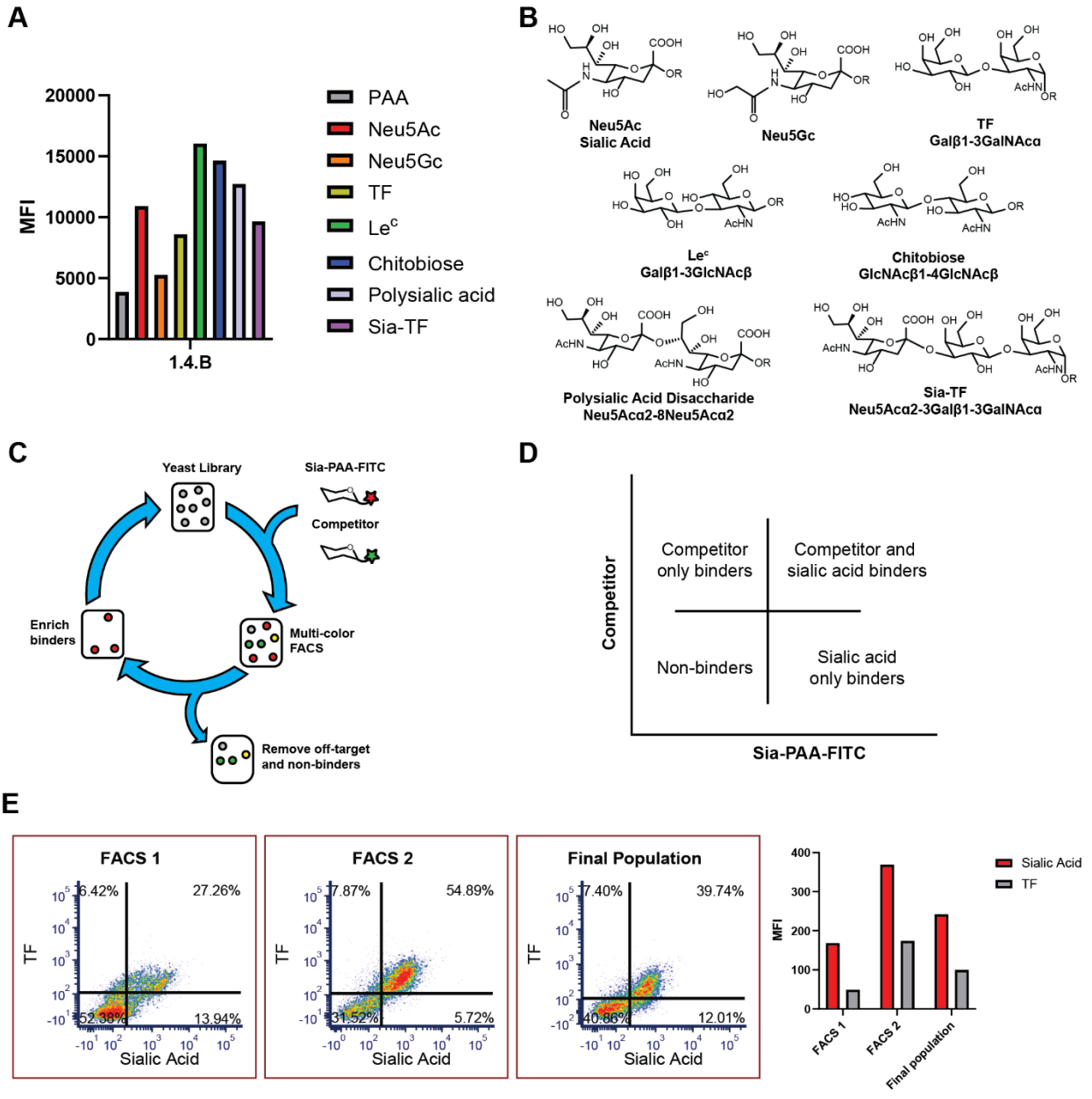
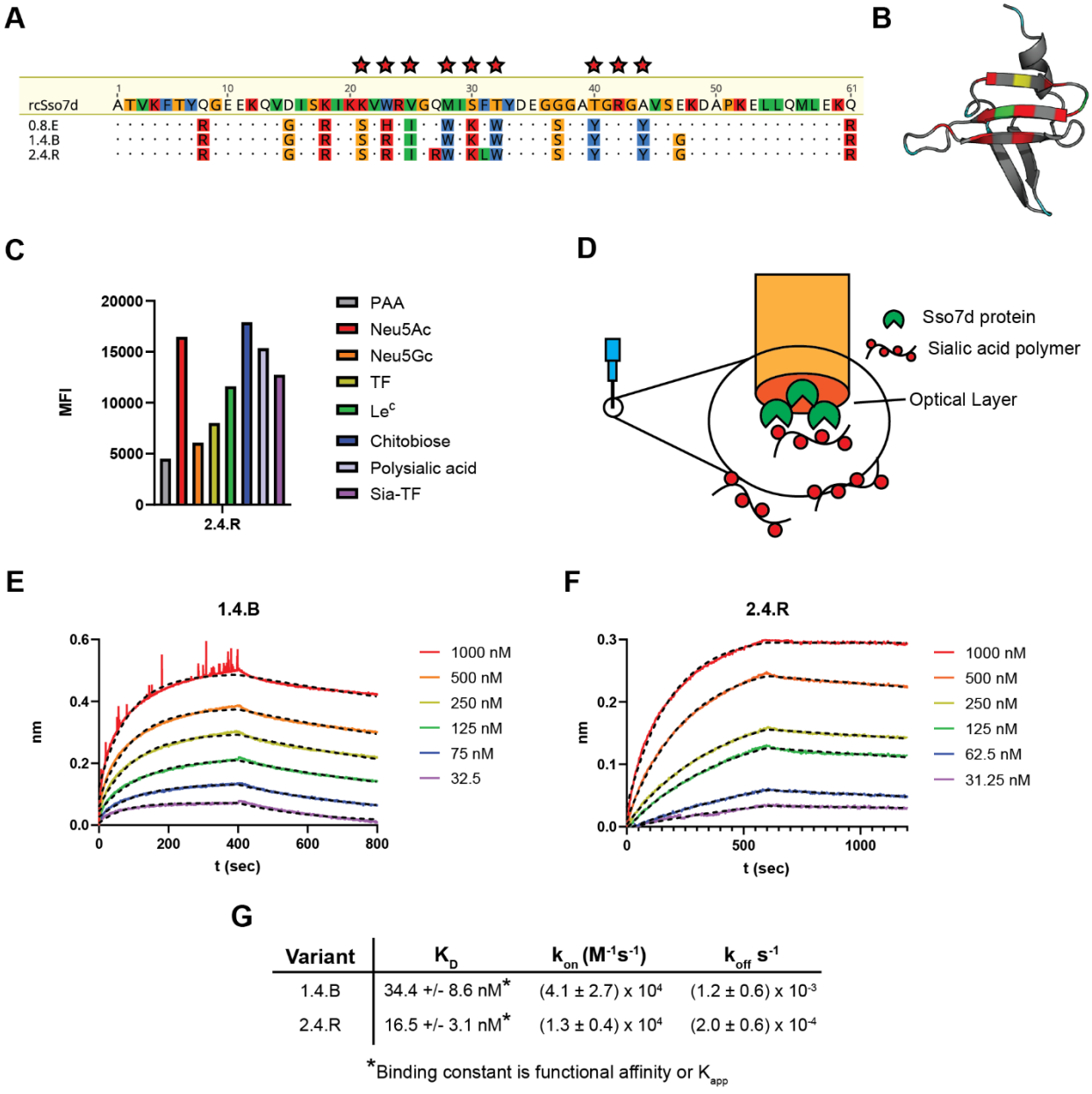


Figure 2-8. Competition sorts for increased specificity.
A) Specificity of variant 1.4.B with glycopolymer panel. B) Structures of glycans in glycopolymer specificity analysis. C) Competitive sort diagram showing sialic acid and competitor polymer with different fluorophores. Gating allows for removal off non-binders and off-target binders while catching specific sialic acid binders. D) Diagram of quadrants expected from sorting conditions.

Bottom left is non-binding cells, top left is binding to only competitor, bottom right is binding to only sialic acid, and top right is binding to both sialic acid and competitor. E) Dot plots and bar graph showing FACS round 1 and 2 and the final population. Bar graph shows median fluorescence intensity of y-axis (TF) and x-axis (sialic acid) from dot plots.

This population underwent sequencing and hierarchical clustering. Selected sequences from each cluster were analyzed and from this, variant 2.4.R was selected as it had the best specificity for sialic acid over TF (**Figure 2-9A,B**). Variant 2.4.R has the same paratope sequence as 1.4.B but gained the additional mutations Q27R and F31L. When 2.4.R was analyzed with the glycopolymer panel it showed that sialic acid binding had increased and off-target binding to Le^c, chitobiose, and TF decreased. Like 1.4.B, 2.4.R could still differentiate sialic acid from Neu5Gc and recognize Sia-TF and the polysialic acid disaccharide.

Both 1.4.B and 2.4.R were cloned into an expression vector containing a C-terminal hexahistidine tag and purified to assess the binding affinity of the soluble proteins. Biolayer interferometry (BLI) with Sia-PAA-FITC was used to measure the functional affinity, or K_{app} of binding. 1.4.B was measured to have a K_{app} of 34 ± 9 nM, and 2.4.R has a K_{app} of 16 ± 3 nM. While a true K_D measurement cannot be made this way due to the multivalency of the interaction, these K_{app} values are indicative of a strong GBP.



G

Variant	K_D	k_{on} ($M^{-1}s^{-1}$)	k_{off} s^{-1}
1.4.B	34.4 +/- 8.6 nM*	(4.1 ± 2.7) x 10^4	(1.2 ± 0.6) x 10^{-3}
2.4.R	16.5 +/- 3.1 nM*	(1.3 ± 0.4) x 10^4	(2.0 ± 0.6) x 10^{-4}

*Binding constant is functional affinity or K_{app}

Figure 2-9. Variant 2.4.R gained specificity and affinity.
A) Sequences of variants 0.8.E, 1.4.B, and 2.4.R. Red stars indicate the paratope randomized sequences. B) Structure of Sso7d showing paratope residues (red), mutations in 0.8.E outside the paratope residues (cyan), mutation found in paratope region of 1.4.B (yellow), and mutations found in 2.4.R (green). C) Specificity of 2.4.R on yeast tested with glycopolymer panel. D) Diagram of biolayer interferometry showing surface immobilized Sso7d variant on the fiber optic tip and soluble sialic acid polymer. E) BLI of 1.4.B with Sia-PAA-FITC from 32.5-1000 nM. F) BLI of 2.4.R with Sia-PAA-FITC from 31.25-1000 nM. F) Table showing K_D, k_{on}, and k_{off} values.

2.7 Conclusions

A method for production of custom GBPs from the DNA binding protein Sso7d was developed. Directed evolution procedures commonly use biotinylated ligands because of the stability, binding affinity, and availability of streptavidin and streptavidin conjugates. However, initial sorting conditions using biotinylated sugar ligands were unsuccessful for enrichment of GBPs as enrichment of biotin-binding Sso7ds from the library repeatedly occurred. Efforts to abate the biotin-binding were explored including increased negative selection and the presence of excess competitor during sorts. Though these were somewhat successful, it was not possible to fully separate the residual biotin-binders from the glycan-binders. This would be a persistent problem for every GBP evolution campaign with this library. Other directed evolution efforts using this rcSso7d library successfully utilized biotinylated ligands without enrichment of biotin-binding variants, demonstrating the additional challenges facing GBP engineering. The preferential binding to biotin over the desired carbohydrate ligand, compared to no observed biotin binding during a protein-protein directed evolution campaign, highlights the difficulty of obtaining carbohydrate binders *de novo*. Compared to proteins and peptides, the carbohydrate epitope is generally much smaller with fewer chemical functional groups that elicit strong binding interactions with a protein binder. As such, it is crucial to minimize other components of the directed evolution platform with strong chemical functionality that could lead to off target interactions.

The binding affinity of an isolated biotin-binding variant 83H was measured to be weak with a K_D of 140 μM, especially if compared to the extraordinary affinity of avidin to biotin. The avidin-biotin interaction is one of the strongest non-covalent interactions in nature with a K_D of $\sim 10^{-15}$, or one billion times stronger than the affinity of the Sso7d variant for biotin.[52] There is

interest in monomeric biotin-binding proteins as the tetrameric property of streptavidin can lead to issues for some applications, and progress has been made toward stable, high-affinity streptavidin monomers with low nM K_Ds.[53] For a biotin-binding Sso7d variant to be useful the affinity would need to be improved by further rounds of directed evolution.

Removal of biotin from the system was pursued to simplify the selection procedures towards a generalizable GBP evolution platform. Tosyl-activated magnetic beads replaced the streptavidin conjugated magnetic beads used previously and allowed a higher degree of ligand immobilization. This is advantageous for selection of the very rare and low affinity variants in the naïve library with sialic acid binding capabilities. The fluorescent glycopolymer Sia-PAA-FITC was used to successfully enrich the population in sialic acid binding variants without development of binding to the polyacrylamide backbone.

Two total rounds of affinity maturation and selections were performed and isolated variants 1.4.B and 2.4.R with superior affinities. These proteins showed a high degree of binding to sialic acid containing glycopolymers but exhibited some off-target binding to several disaccharides, most notably the core of N-linked glycans chitobiose. This is very similar to the commercially available wheat germ agglutinin (WGA), which is a well-known binder of β1-4-GlcNAc (such as chitobiose), but can also bind to sialic acid.[54] This off-target binding could be diminished by application of a competitive two-color sorting technique. The final variant 2.4.R had increased specificity toward sialic acid and a functional affinity of 16 nM, a small improvement from the 34 nM affinity of 1.4.B. The specificity and affinity of 2.4.R demonstrates that the GBPs evolved from the Sso7d scaffold can rival existing GBP reagents. The methods developed here can be applied to any glycan of interest for which multivalent fluorescent probes are available.

2.8 Materials and methods

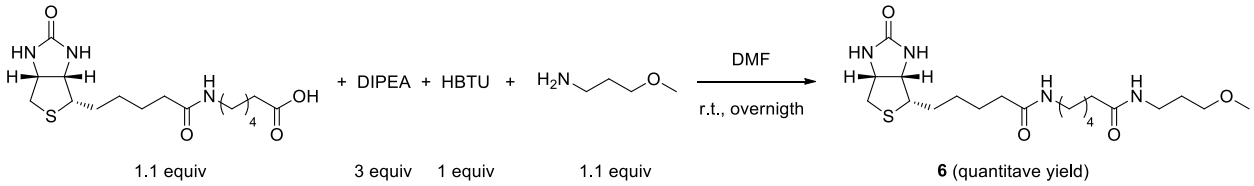
Materials

For flow cytometry and fluorescence-assisted cell sorting (FACS), detection reagents were as follows: chicken anti-HA (Exalpha Biologicals, Shirley, MA, Cat. No: AHA), chicken anti-cMyc (Exalpha Biologicals, Shirley, MA, Cat. No. ACMYC), goat anti-chicken AlexaFluor 647 or AlexaFluor 488 (ThermoFisher Scientific, Cat. No: A-21449 or A-11039), streptavidin AlexaFluor 647 (ThermoFisher Scientific, Cat. No: S21374). Monovalent Sia-sp-biotin and TF-sp-biotin were purchased from GlycoTech (Frederick, MD). Glycophorin A from blood group MN was purchased from Millipore-Sigma. Multivalent carbohydrate polymers were commercially purchased as PAA-FITC or PAA-biotin variations from GlycoTech (Frederick, MD.) All mammalian cell lines were purchased from ATCC and the appropriate growth media and supplements purchased from ATCC or MilliporeSigma, unless otherwise specified.

Yeast culture conditions

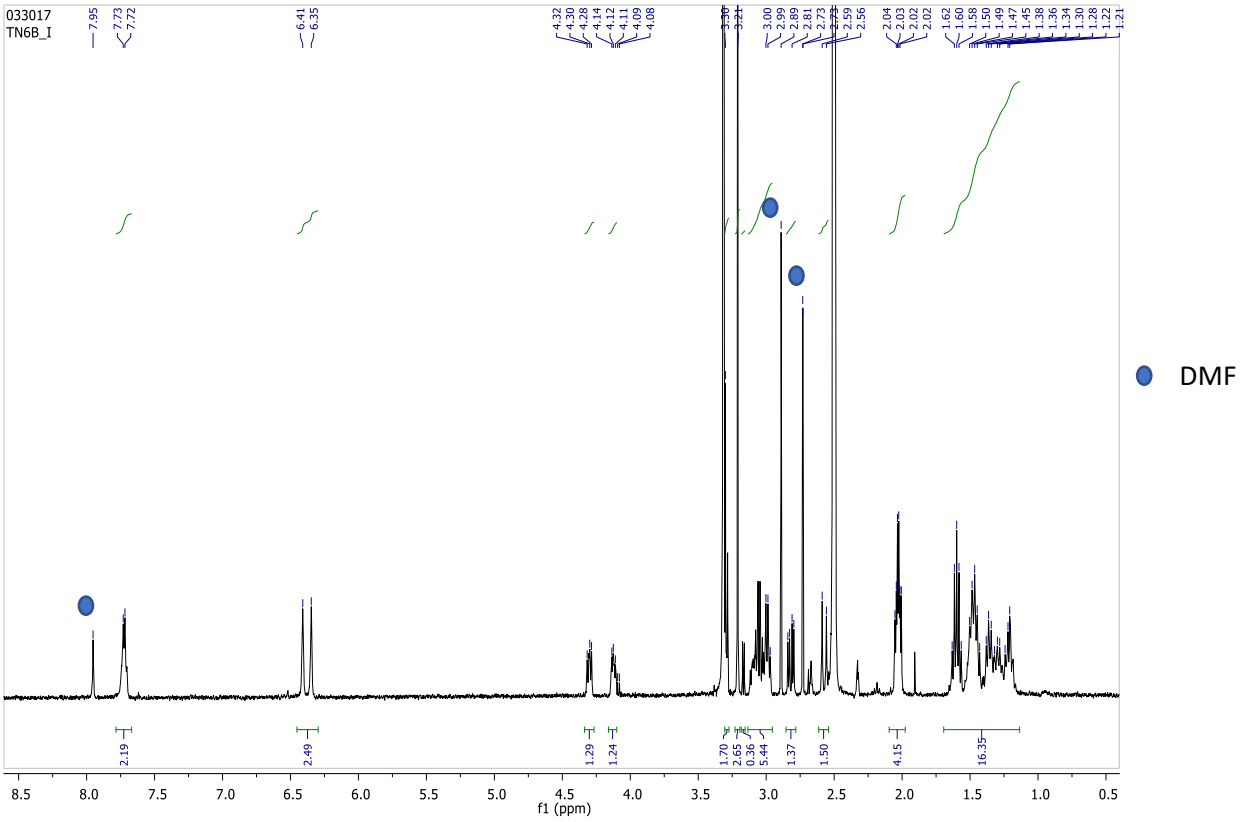
Use of yeast-surface display for directed evolution of glycan binders was carried out according to established protocols published by Wittrup and coworkers.[38, 55] All yeast work was done under aseptic conditions. Yeast populations containing yeast-display vector pCTCON2 were grown in SDCAA media containing 100 U/mL penicillin-streptomycin at 30 °C for routine culture and subculture, unless otherwise noted. Yeast-surface protein expression was induced by subculture and resuspension of log-phase yeast cultures in SGCAA media containing 100 U/mL pen-strep at 20 °C. All yeast washes and selections were carried out in PBSA (PBS at pH 7.4 with 0.1% bovine serum albumin, sterile filtered) at 4 °C.

Synthesis of sp-biotin

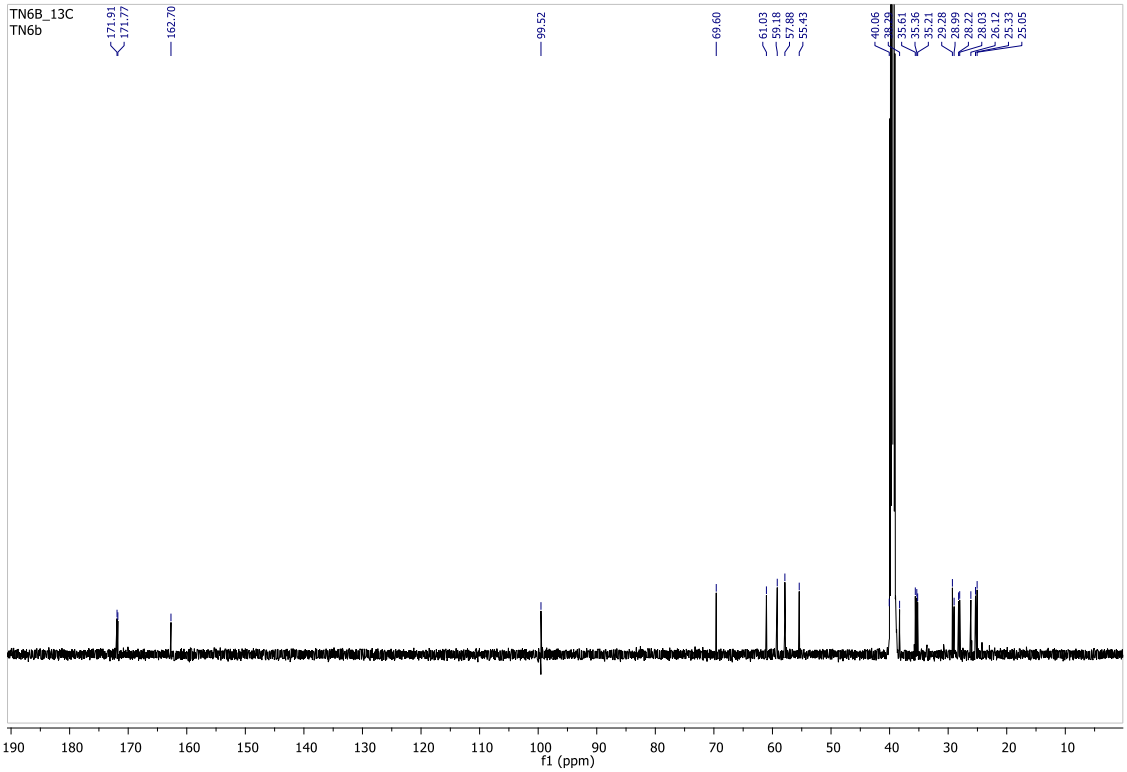


Synthetic route to sp-biotin

Experimental procedure: 40 mg of Biotin-LC were dissolved in DMF, DIPEA (58 µL) and HBTU (38mg) were added, after 20 minutes stirring at room temperature, 3-methoxypropylamine (14 µL) was added to the activated acid. The reaction mixture was stirred overnight. Once the reaction was completed (TLC), the solvent was removed under vacuum, giving a white solid, which was washed several times with methanol to obtain 48 mg of the white pure product (quantitative yield). ^{1}H NMR (DMSO d6, 400 MHz) δ: 7.73 (m, 2H), 6.41 (s, 1H), 6.35 (s, 1H), 4.30 (m, 1H), 4.12 (m, 1H), 3.29 (m, 2H), 3.21 (s, 3H), 3.09 (m, 1H), 3.05 (dd, J = 12.8, 7.0 Hz, 2H), 2.99 (dd, J = 12.8, 6.8 Hz, 2H), 2.82 (dd, J = 12.4, 5.1 Hz, 1H), 2.57 (d, J = 12.4 Hz, 1H), 2.03 (m, 3H), 1.60 (m, 3H), 1.47 (m, 4H), 1.28 (m, 7H). ^{13}C NMR (DMSO d6, 600 MHz) δ 171.9, 171.8, 162.7, 99.5, 69.6, 61.03, 59.2, 57.9, 55.4, 40.1, 38.3, 35.6, 35.4, 35.2, 29.3, 29.0, 28.2, 28.0, 26.1, 25.3, 25.1.

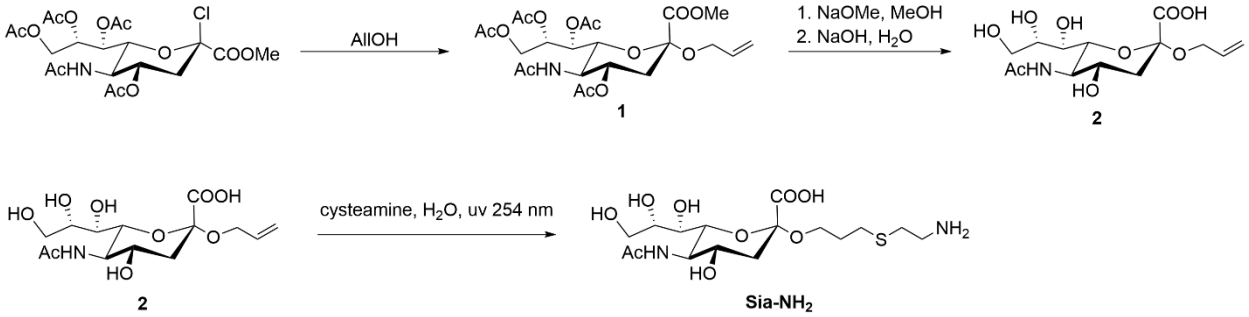


^{1}H NMR Spectrum (400 MHz, DMSO-d_6) of sp-biotin.



^{13}C NMR Spectrum (600 MHz, DMSO-*d*$_6$) of sp-biotin.

Synthesis of primary amine ligands

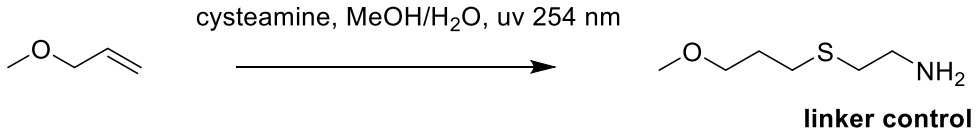


Synthetic route to Sia-NH$_2$

Synthesis of 1 and 2. Following a procedure from Roy and LaFerriere,[56] purchased chloride **1** (0.862 g, 1.69 mmol, CAS# 67670-69-3, Sigma-Aldrich) was dissolved in allyl alcohol (25 mL, 0.367 mol) with activated 4A molecular sieves. Silver salicylate (0.4 g) was added and the suspension was stirred at room temp in dark conditions for 2h. The mixture was filtered over Celite

and the precipitate washed with dichloromethane. The organic solution was washed 1x with $NaHCO_3$ solution, 1x with water, dried with magnesium sulfate, concentrated in vacuo, and the crude purified by flash chromatography (20:1 to 10:1 CH_2Cl_2/MeOH) to give allyl glycoside **1** (185 mg, 21%) matching physicochemical properties of the literature reference.[56] Purified **1** was then dissolved in a solution of 0.15 M NaOMe in anhydrous MeOH (2 mL) and mixture stirred for 2 h until spot-to-spot conversion observed by TLC. After, freshly washed Amberlite IR-120 H^+ resin was added to the reaction until pH 6 and the resin filtered away. The filtrate was concentrated twice by rotovap and the resulting residue containing **2** was carried to the next step with no further purification (119 mg, quant.) which matched the physicochemical properties of the literature reference.[56]

Synthesis of Sia-NH_2. Fully deprotected glycoside **3** (100 mg, 0.286 mmol) and cysteamine-HCl (65 mg, 0.572 mmol) were dissolved in deoxygenated H2O (3 mL) in a quartz cuvette and irradiated at 254 nm for 1 hr and the reaction monitored by TLC. Once the reaction appeared complete, the crude material was purified by reversed phase HPLC (Waters 1525 binary HPLC mounted with a 00G-4252-PO-AX (Luna Co.) preparative column, using a H2O:MeCN with 0.1% TFA gradient from 4-30% MeCN) to afford final product **Sia-NH_2** (52.6 mg, 43% yield.) The isolated product matched the physiochemical properties of the literature reference.[57]

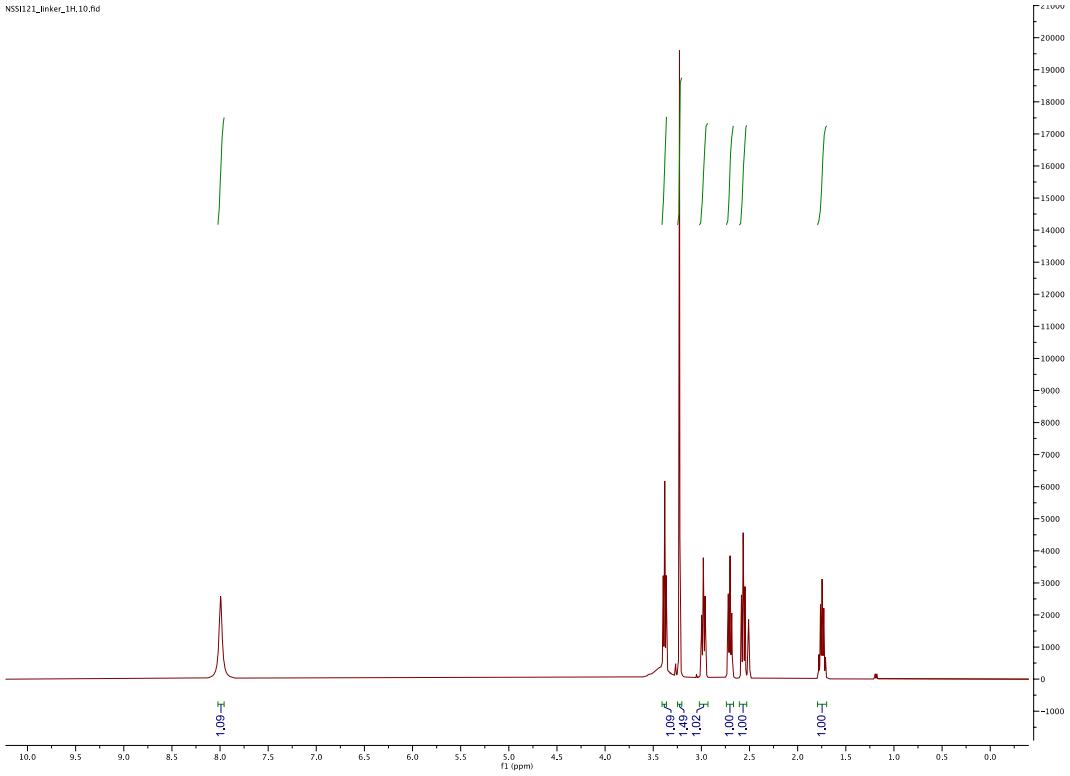


linker control

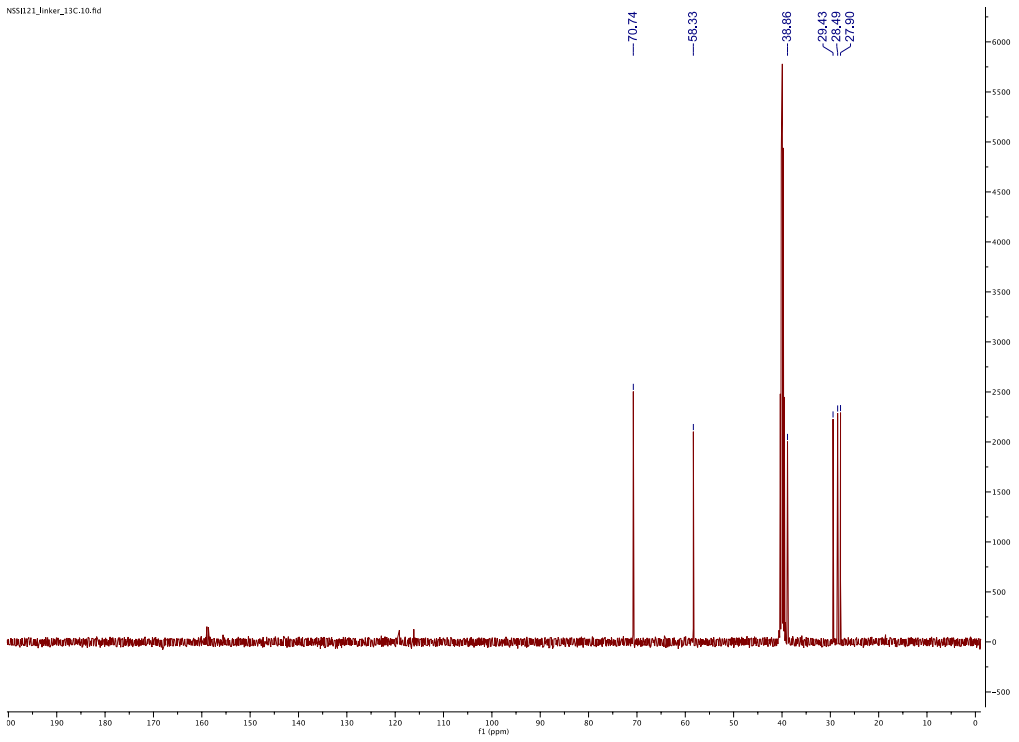
Synthetic route to control linker

Synthesis of 2-((3-methoxypropyl)thio)ethan-1-amine (linker-NH_2)

Allyl methyl ether (129 mg, 0.305 mmol, CAS # 627-40-7 Sigma-Aldrich) and cysteamine-HCl (42 mg, 0.370 mmol) were dissolved in deoxygenated H_2O (3 mL) in a quartz cuvette and irradiated at 254 nm for 1 hr. Once the reaction appeared complete by TLC, the crude material was purified by reversed phase HPLC (Waters 1525 binary HPLC mounted with a 00G-4252-PO-AX (Luna Co.) preparative column, using a H_2O/MeCN with 0.1% TFA gradient) to afford final product "linker-NH_2" (89.4 mg, 34% yield). ^{1}H NMR (400 MHz, DMSO-D_6) δ 7.99 (s, 2H); 3.35-3.41 (t, 2H, OCH_2); 3.23 (s, 3H, CH_3); 2.94-3.02 (t, 2H, CH_2); 2.66-2.74 (t, 2H, CH_2); 2.53-2.60 (t, 2H, CH_2); 1.70-1.79 (m, 2H, CH_2) ppm. ^{13}C-NMR (100 MHz, DMSO-D_6) δ 70.74; 58.33; 38.86; 29.43; 28.49; 27.90 ppm. LRMS (ESI+) m/z: [M+H]+ Calcd for $C_6H_{16}NOS^+$ 150.1; Found 150.1. HRMS (ESI+) m/z: [M+H]+ Calcd for $C_6H_{16}NOS^+$ 150.0947; Found 150.0948.



^{1}H NMR Spectrum (400 MHz, DMSO-*d*$_6$) of linker-NH_2.



^{13}C NMR Spectrum (100 MHz, DMSO-*d*$_6$) of linker-NH$_2$.

Magnetic bead preparations

Biotin Binder Dynabeads (ThermoFisher Scientific, Rockford, IL.) are first washed with PBSA buffer, incubated with the biotinylated ligand for 30 min, then washed twice. Concentration of streptavidin for volume of bead being treated is calculated using available information from manufacturer, and four-fold excess of biotinylated ligand is added to ensure all binding sites of streptavidin are occupied.

Tosylactivated Dynabeads (ThermoFisher Scientific, Rockford, IL) are washed in 0.1 M borate buffer pH 9.5. Ligand is added at 2 mM in 0.1 M borate buffer pH 9.5 + 1.2 M ammonium sulfate, then incubated for 18 hours at 37 ºC.[38] Coupling efficiency is measured by taking supernatant after overnight incubation and measuring A280, and concentration of coupled ligand is calculated from

a standard curve of toluene sulfonic acid. A control without ligand is carried along to measure background hydrolysis of the tosyl group from the beads, which is then subtracted to get the final concentration of ligand coupled.

Magnetic bead sorts

Binders were isolated by magnetic bead sorts from the reduced-charge Sso7d libraries rcSso7d-11 and rcSso7d-18 using either tosyl-activated Dynabeads or Biotin Binder Dynabeads in PBSA. Dynabeads loaded with carbohydrate ligand were utilized for positive selections, while bare beads or beads loaded with linker were utilized for negative selections. Positive and negative selections were iteratively performed until library diversities approached 10^4 variants. Selections use $4x10^5$ beads per 10^7 cells and were performed for one hour at 4 ºC.

FACS

Cells were washed in PBSA (PBS pH 7.4 with 0.1% bovine serum albumin) and then incubated with the sugar ligand and either chicken anti-HA or chicken anti-cMyc overnight with gentle agitation at 4 °C. Cells were then washed, followed by incubation with goat anti-chicken-AF 647 (if FITC polymers used) or goat anti-chicken-AF 488 (if streptavidin AF 647 used). Washed cells were sorted on a FACS Aria IIu (BD Biosciences, San Jose, CA) collecting the top 1% of binding cells and this process repeated for at least three rounds of further enrichment. For analysis without sorting, an Intellcyt iQue Screener flow cytometer (Essen Bioscience, Ann Arbor, MI) was used. For preloaded samples, streptavidin AF647 was incubated with four-fold excess biotinylated ligand for 30 min and then added in initial incubation. For two-step labeling, soluble biotinylated ligand was incubated with cells in first incubation, the streptavidin AF647 was added in the second

incubation. For competition experiments, streptavidin with no fluorophore label was incubated with four-fold excess biotin for 30 min before addition during first incubation.

Affinity maturation

To improve the affinities of enriched variants, plasmids from heterogeneous populations of variants or individual variants of interest were subjected to one or more rounds of affinity maturation. Error-prone PCR was initiated by amplifying Sso7d inserts from yeast display vectors in the presence of 2 μM nucleotide analogs (TriLink BioTechnologies, San Diego, CA) 8-oxo-2'-deoxyguanosine-5'-triphosphate (8-oxo-dGTP) and 2'-deoxy-p-nucleoside-5'-triphosphate (dPTP) as described previously.[55] PCR reactions were run using Taq polymerase for a minimum of 15 cycles, followed by purification and extraction on a 1% agarose gel. Mutated inserts were further amplified by PCR to provide sufficient material for electroporation. Separately, pCTCON2 plasmids bearing no Sso7d insert were digested with NheI, BamHI, and SalI, followed by ethanol precipitation with amplified mutagenized inserts. For electroporation, 25 mL freshly cultured EBY100 yeast at OD_{600} of 2 were made electrocompetent by incubation in 10 mM DTT, 100 mM LiOAc in YPD for 10 minutes at 37 C. Cells were washed twice with ice-cold H_2O, followed by resuspension in 200 μl H_2O containing 5 μg of insert and 1 μg of cut vector precipitated DNA. Yeast were electroporated 50 μl at a time in 2 mm cuvettes using a square-wave function (Gene Pulser Xcell, Bio-Rad, Hercules, CA.) Cells were immediately diluted in 1 mL YPD after electroporation for 30 min, followed by dilution and subculture in SDCAA. Samples of newly generated libraries were plated on SDCAA agar in order to determine maximum theoretical library diversity.

Recombinant expression and affinity purification of Sso7d constructs in *E. coli.*

Sso7d variants of interest were recombinantly expressed as LPETGG-His6 constructs for biophysical characterization. BL21-CodonPlus (DE3)-RIL *Escherichia coli* cells (Agilent Technologies, Santa Clara, CA) were transformed with pET24 plasmids containing Sso7d variants of interest. Single transformants were used to inoculate 8 mL overnight cultures of LB containing 30 μg/mL each of kanamycin and chloramphenicol. Grown overnight cultures were added to 1 L of TB containing 30 μg/mL each of kanamycin and chloramphenicol and grown at 37 °C until an OD_{600} of 0.8-1 reached. Cultures were brought to 16-18 °C and expression induced in the presence of 1 mM IPTG, followed by 16-18 h of growth. Cells were harvested by centrifugation at 3700 *x g* for 30 min and overexpressed proteins purified immediately or pelleted cells stored long term at -80 °C.

All protein purification steps were carried out at 4 °C. Pelleted cells were resuspended in Buffer A (50 mM HEPES pH 7.5, 150 mM NaCl, 20 mM imidazole) supplemented with 0.5 mg/mL lysozyme (Research Products International, Mount Prospect, IL), 1:1000 protease inhibitor cocktail (EMD Millipore, Burlington, MA), and 1 U/mL DNase I (NEB) and tumbled for 30 min. Homogenized cells were lysed by sonication at 50% amplitude for 3 cycles of 5 min (Vibra-Cell, Sonics, Newtown, CT), followed by clarification by ultracentrifugation at 35,000 *x g* for 60 min. Supernatants containing soluble Sso7d proteins were passed over Ni-NTA resin (HisPur, Thermo Scientific, Rockford, IL) pretreated with Buffer A. Resin was washed with Buffer A for 12 column volumes, followed by elution with high imidazole Buffer B (50 mM HEPES pH 7.5, 150 mM NaCl, 1 M imidazole) for 6-8 column volumes, with a majority of purified Sso7d eluting in the first 2 column volumes. Column fractions of interest were pooled and immediately desalted using

HiTrap desalting column (GE Healthcare, Chicago, IL). Proteins were stored at 4 °C for immediate use or flash-frozen and stored at - 80 °C.

Biophysical characterization of Sso7d-LPXTG-His6 constructs by bio-layer interferometry

Apparent binding affinities and binding specificities of Sso7d variantss of interest were determined by bio-layer inferometry (BLI). All BLI measurements were performed in PBSA supplemented with 0.05% Tween-20 at 22 °C on an Octet RED96 instrument (Pall ForteBio, Fremont, CA.) Sso7d-LPETGG-His_6 constructs were captured on Ni-NTA BLI tips to a final displacement of 2 nm, followed by a baseline reading for 60 s. Tips were placed into glycan polymer solutions of various concentrations to measure association, followed by buffer alone for dissociation measurements. If needed, buffer baseline measurements from Sso7d-loaded tips were used for background subtraction and all association and dissociation curves globally fitted to a 1:1 binding model to obtain apparent binding affinities.

2.9 Acknowledgements

I would like to thank Dr. Cristina Zamora for her collaboration and efforts on this project including her role in collecting and interpreting data, scientific chats, and overall support. I would like to thank Dr. Megan Kizer for her helpful comments and edits to this chapter. I would like to thank Dr. Nathaniel Schocker for the synthesis of the Sia-NH_2 compound for use with tosyl-activated magnetic beads and Teresa Naranjo Sanchez for synthesis of the sp-biotin control. Finally, a huge thanks to the Wittrup lab for sharing the Sso7d yeast surface display library, protocols, and equipment.

2.10 References

1. Varki, A. (2017) Biological Roles of Glycans, *Glycobiology 27*, 3-49.

2. Fuster, M. M., and Esko, J. D. (2005) The Sweet and Sour of Cancer: Glycans as Novel Therapeutic Targets, *Nat. Rev. Cancer 5*, 526-542.

3. Imperiali, B. (2019) Bacterial Carbohydrate Diversity - a Brave New World, *Curr. Opin. Chem. Biol. 53*, 1-8.

4. Herget, S., Toukach, P. V., Ranzinger, R., Hull, W. E., Knirel, Y. A., and von der Lieth, C.-W. (2008) Statistical Analysis of the Bacterial Carbohydrate Structure Data Base (Bcsdb): Characteristics and Diversity of Bacterial Carbohydrates in Comparison with Mammalian Glycans, *BMC Struct. Biol. 8*, 35.

5. Haslam, S. M., Freedberg, D. I., Mulloy, B., Dell, A., Stanley, P., and Prestegard, J. H. (2022) Structural Analysis of Glycans, *Essentials of Glycobiology [Internet]. 4th edition*.

6. Wu, A. M., Lisowska, E., Duk, M., and Yang, Z. (2008) Lectins as Tools in Glycoconjugate Research, *Glycoconj. J. 26*, 899.

7. Brooks, S. A. (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.

8. Slifkin, M., and Doyle, R. J. (1990) Lectins and Their Application to Clinical Microbiology, *Clin. Microbiol. Rev. 3*, 197-218.

9. Mujahid, A., and Dickert, L. F. (2016) Blood Group Typing: From Classical Strategies to the Application of Synthetic Antibodies Generated by Molecular Imprinting, *Sensors 16*.

10. Cummings, R. D., Etzler, M., Hahn, M. G., Darvill, A., Godula, K., Woods, R. J., and Mahal, L. K. (2022) Glycan-Recognizing Probes as Tools, *Essentials of Glycobiology [Internet]. 4th edition*.

11. Sterner, E., Flanagan, N., and Gildersleeve, J. C. (2016) Perspectives on Anti-Glycan Antibodies Gleaned from Development of a Community Resource Database, *ACS Chem. Biol. 11*, 1773-1783.

12. Kappler, K., and Hennet, T. (2020) Emergence and Significance of Carbohydrate-Specific Antibodies, *Genes Immun. 21*, 224-239.

13. Ward, E. M., Kizer, M. E., and Imperiali, B. (2021) Strategies and Tactics for the Development of Selective Glycan-Binding Proteins, *ACS Chem. Biol. 16*, 1795-1813.

14. Warkentin, R., and Kwan, D. H. (2021) Resources and Methods for Engineering "Designer" Glycan-Binding Proteins, *Molecules 26*, 380.

15. Adhikari, P., Bachhawat-Sikder, K., Thomas, C. J., Ravishankar, R., Jeyaprakash, A. A., 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.

16. 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.

17. Cicortas Gunnarsson, L., Nordberg Karlsson, E., Albrekt, A. S., Andersson, M., Holst, O., and Ohlin, M. (2004) A Carbohydrate Binding Module as a Diversity-Carrying Scaffold, *Protein Eng. Des. Sel. 17*, 213-221.

18. Lu, Z., Kamat, K., Johnson, B. P., Yin, C. C., Scholler, N., and Abbott, K. L. (2019) Generation of a Fully Human Scfv That Binds Tumor-Specific Glycoforms, *Sci. Rep. 9*, 5101.

19. Hong, X., Ma, M. Z., Gildersleeve, J. C., Chowdhury, S., Barchi, J. J., Mariuzza, R. A., Murphy, M. B., 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.

20. Tasumi, S., Velikovsky, C. A., Xu, G., Gai, S. A., Wittrup, K. D., Flajnik, M. F., Mariuzza, R. A., and Pancer, Z. (2009) High-Affinity Lamprey Vlra and Vlrb Monoclonal Antibodies, *Proc. Natl. Acad. Sci. 106*, 12891.

21. McKitrick, T. R., Goth, C. K., Rosenberg, C. S., Nakahara, H., Heimburg-Molinaro, J., McQuillan, A. M., Falco, R., Rivers, N. J., Herrin, B. R., Cooper, M. D., and Cummings, R. D. (2020) Development of Smart Anti-Glycan Reagents Using Immunized Lampreys, *Comm. Biol. 3*, 91.

22. McKitrick, T. R., Bernard, S. M., Noll, A. J., Collins, B. C., Goth, C. K., McQuillan, A. M., Heimburg-Molinaro, J., Herrin, B. R., Wilson, I. A., Cooper, M. D., and Cummings, R. D. (2021) Novel Lamprey Antibody Recognizes Terminal Sulfated Galactose Epitopes on Mammalian Glycoproteins, *Comm. Biol. 4*, 674.

23. Angata, T., and Varki, A. (2002) Chemical Diversity in the Sialic Acids and Related A-Keto Acids: An Evolutionary Perspective, *Chem. Rev. 102*, 439-470.

24. Ghosh, S. (2020) Sialic Acid and Biology of Life: An Introduction, *Sialic Acids and Sialoglycoconjugates in the Biology of Life, Health and Disease*, 1-61.

25. Ravn, P., Danielczyk, A., Bak Jensen, K., Kristensen, P., Astrup Christensen, P., Larsen, M., Karsten, U., and Goletz, S. (2004) Multivalent Scfv Display of Phagemid Repertoires for the Selection of Carbohydrate-Specific Antibodies and Its Application to the Thomsen–Friedenreich Antigen, *J. Mol. Biol. 343*, 985-996.

26. Sharma, V., Vijayan, M., and Surolia, A. (1996) Imparting Exquisite Specificity to Peanut Agglutinin for the Tumor-Associated Thomsen-Friedenreich Antigen by Redesign of Its Combining Site *J. Biol. Chem. 271*, 21209-21213.

27. Yu, L.-G. (2007) The Oncofetal Thomsen–Friedenreich Carbohydrate Antigen in Cancer Progression, *Glycoconj. J. 24*, 411-420.

28. Feng, D., Shaikh, A. S., and Wang, F. (2016) Recent Advance in Tumor-Associated Carbohydrate Antigens (Tacas)-Based Antitumor Vaccines, *ACS Chem. Biol. 11*, 850-863.

29. Könning, D., and Kolmar, H. (2018) Beyond Antibody Engineering: Directed Evolution of Alternative Binding Scaffolds and Enzymes Using Yeast Surface Display, *Microb. Cell Fact. 17*, 32.

30. Skerra, A. (2007) Alternative Non-Antibody Scaffolds for Molecular Recognition, *Curr. Opin. Biotechnol. 18*, 295-304.

31. Knapp, S., Karshikoff, A., Berndt, K. D., Christova, P., Atanasov, B., and Ladenstein, R. (1996) Thermal Unfolding of the DNA-Binding Protein Sso7d from the Hyperthermophilesulfolobus Solfataricus, *J. Mol. Biol. 264*, 1132-1144.

32. Gera, N., Hussain, M., Wright, R. C., and Rao, B. M. (2011) Highly Stable Binding Proteins Derived from the Hyperthermophilic Sso7d Scaffold, *J. Mol. Biol. 409*, 601-616.

33. Gera, N., Hill, A. B., White, D. P., Carbonell, R. G., and Rao, B. M. (2012) Design of Ph Sensitive Binding Proteins from the Hyperthermophilic Sso7d Scaffold, *PLoS One 7*, e48928.

34. Cruz-Teran, C. A., Bacon, K., McArthur, N., and Rao, B. M. (2018) An Engineered Sso7d Variant Enables Efficient Magnetization of Yeast Cells, *ACS Comb. Sci. 20*, 579-584.

35. Miller, E. A., Baniya, S., Osorio, D., Al Maalouf, Y. J., and Sikes, H. D. (2018) Paper-Based Diagnostics in the Antigen-Depletion Regime: High-Density Immobilization of Rcsso7d-Cellulose-Binding Domain Fusion Proteins for Efficient Target Capture, *Biosens. Bioelectron. 102*, 456-463.

36. Miller, E. A., Traxlmayr, M. W., Shen, J., and Sikes, H. D. (2016) Activity-Based Assessment of an Engineered Hyperthermophilic Protein as a Capture Agent in Paper-Based Diagnostic Tests, *Mol. Syst. Des. Eng. 1*, 377-381.

37. 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.

38. Traxlmayr, M. W., Kiefer, J. D., Srinivas, R. R., Lobner, E., Tisdale, A. W., Mehta, N. K., Yang, N. J., Tidor, B., and Wittrup, K. D. (2016) Strong Enrichment of Aromatic Residues in Binding Sites from a Charge-Neutralized Hyperthermostable Sso7d Scaffold Library, *J. Biol. Chem. 291*, 22496-22508.

39. Hudson, K. L., Bartlett, G. J., Diehl, R. C., Agirre, J., Gallagher, T., Kiessling, L. L., and Woolfson, D. N. (2015) Carbohydrate–Aromatic Interactions in Proteins, *J. Am. Chem. Soc. 137*, 15152-15160.

40. Kiessling, L. L., and Diehl, R. C. (2021) Ch−Π Interactions in Glycan Recognition, *ACS Chem. Biol. 16*, 1884-1893.

41. Asensio, J. L., Ardá, A., Cañada, F. J., and Jiménez-Barbero, J. (2013) Carbohydrate–Aromatic Interactions, *Acc. Chem. Res. 46*, 946-954.

42. Bogan, A. A., and Thorn, K. S. (1998) Anatomy of Hot Spots in Protein Interfaces11edited by J. Wells, *J. Mol. Biol. 280*, 1-9.

43. Zemlin, M., Klinger, M., Link, J., Zemlin, C., Bauer, K., Engler, J. A., Schroeder, H. W., and Kirkham, P. M. (2003) Expressed Murine and Human Cdr-H3 Intervals of Equal Length Exhibit Distinct Repertoires That Differ in Their Amino Acid Composition and Predicted Range of Structures, *J. Mol. Biol. 334*, 733-749.

44. Fonnum, G., Johansson, C., Molteberg, A., Mørup, S., and Aksnes, E. (2005) Characterisation of Dynabeads® by Magnetization Measurements and Mössbauer Spectroscopy, *J. Magn. Magn. Mater. 293*, 41-47.

45. Traxlmayr, M. W., and Obinger, C. (2012) Directed Evolution of Proteins for Increased Stability and Expression Using Yeast Display, *Arch. Biochem. Biophys. 526*, 174-180.

46. Fukuda, M., Lauffenburger, M., Sasaki, H., Rogers, M. E., and Dell, A. (1987) Structures of Novel Sialylated O-Linked Oligosaccharides Isolated from Human Erythrocyte Glycophorins, *J. Biol. Chem. 262*, 11952-11957.

47. Bovin, N. V. (1998) Polyacrylamide-Based Glycoconjugates as Tools in Glycobiology, *Glycoconj. J. 15*, 431-446.

48. Zaccolo, M., Williams, D. M., Brown, D. M., and Gherardi, E. (1996) An Approach to Random Mutagenesis of DNA Using Mixtures of Triphosphate Derivatives of Nucleoside Analogues, *J. Mol. Biol. 255*, 589-603.

49. Payne, C. M., Bomble, Y. J., Taylor, C. B., McCabe, C., Himmel, M. E., Crowley, M. F., and Beckham, G. T. (2011) Multiple Functions of Aromatic-Carbohydrate Interactions in a Processive Cellulase Examined with Molecular Simulation, *J. Biol. Chem. 286*, 41028-41035.

50. Luo, M., Velikovsky, C. A., Yang, X., Siddiqui, M. A., Hong, X., Barchi, J. J., Jr., Gildersleeve, J. C., Pancer, Z., and Mariuzza, R. A. (2013) Recognition of the Thomsen-Friedenreich Pancarcinoma Carbohydrate Antigen by a Lamprey Variable Lymphocyte Receptor, *J. Biol. Chem. 288*, 23597-23606.

51. Li, D., Wang, L., Maziuk, B. F., Yao, X., Wolozin, B., and Cho, Y. K. (2018) Directed Evolution of a Picomolar-Affinity, High-Specificity Antibody Targeting Phosphorylated Tau, *J. Biol. Chem. 293*, 12081-12094.

52. Green, N. M. (1975) Avidin, *Adv. Protein Chem. 29*, 85-133.

53. Lim, K. H., Huang, H., Pralle, A., and Park, S. (2013) Stable, High-Affinity Streptavidin Monomer for Protein Labeling and Monovalent Biotin Detection, *Biotechnol. Bioeng. 110*, 57-67.

54. Bojar, D., Meche, L., Meng, G., Eng, W., Smith, D. F., Cummings, R. D., and Mahal, L. K. (2022) A Useful Guide to Lectin Binding: Machine-Learning Directed Annotation of 57 Unique Lectin Specificities, *ACS Chem. Biol.*

55. Chao, G., Lau, W. L., Hackel, B. J., Sazinsky, S. L., Lippow, S. M., and Wittrup, K. D. (2006) Isolating and Engineering Human Antibodies Using Yeast Surface Display, *Nat. Protoc. 1*, 755-768.

56. Roy, R., and Laferrière, C. A. (1990) Synthesis of Protein Conjugates and Analogues of N-Acetylneuraminic Acid, *Can. J. Chem. 68*, 2045-2054.

57. Roy, R., and Laferrière, C. A. (1988) Synthesis of Antigenic Copolymers of N-Acetylneuraminic Acid Binding to Wheat Germ Agglutinin and Antibodies, *Carbohydr. Res. 177*, c1-c4.

Chapter 3: Engineered glycan-binding proteins for recognition of tumor associated carbohydrate antigen

3.1 Abstract

Glycan-binding proteins (GBPs) are commonly used reagents for the identification of glycan determinants. They do not require specialized equipment or time-consuming experimental methods, making them widely used tools for basic research and clinical applications. Existing glycan recognition reagents, mainly antibodies and lectins, are limited, and discovery or creation of reagents with novel specificities is time consuming and difficult. Here I apply the GBP generation platform developed in Chapter 2 towards a tumor-associated carbohydrate epitope. Yeast surface display of a variable library of rcSso7d proteins was screened to isolate variants with specificity toward the disaccharide Galβ1-3GalNAcα, or Thomsen-Friedenreich antigen, a disaccharide abundant on the surface of cancer cells. Characterization of the generated TF-binding proteins show them to have specificities and affinities on par with currently available lectins. The proteins can be functionalized with fluorophores or biotin to create reagents that prove useful for glycoprotein blotting and cell staining applications.

3.2 Introduction

Carbohydrates are ubiquitous biomolecules that are of vital importance in human health and disease. These carbohydrates, or glycans, can be appended to proteins and lipids or secreted as large extracellular polymers. The biological functions of glycans are broad, including energy metabolism, structural and stabilizing roles, and carriers of information.[1] Glycans often have protective, stabilizing, organizational, and barrier functions such as the cell walls of plants and bacteria, the exoskeleton of arthropods, or their contributions toward protein folding and stability. As information carriers, glycans play a role in cell adhesion, cell-cell communication, and host-pathogen interactions. Cell surface glycans can also serve as important disease markers, as glycosylation is often aberrant in malignant tissues.[2]

Glycan-binding proteins (GBPs) are widely used tools for qualitative glycan structure analysis. These proteins can distinguish between many glycan structures and do not require the sample handling and expertise that other glycan structure analysis methods, such as mass spectrometry and NMR, require. These proteins can be used in the lab as affinity reagents for glycan/glycoconjugate purification and cell staining reagents, or used diagnostically for blood typing, microorganism detection, and histology.[3-6] The predominant GBP reagents consist of fungal, plant, and animal derived lectins and monoclonal antibodies. Although these GBPs are important tools for glycobiology, they have major limitations. Lectins are non-enzymatic GBPs and are popular because they can recognize monosaccharide determinants and glycosidic linkages and are relatively cheap to isolate and purify. These proteins are very useful, but the specificities of known lectins do not cover the large diversity of glycans found distributed across biology. Unlike lectins, antibodies against sugars can be directly elicited in animals by inoculation with the free glycan or glycan-protein conjugates, however antibody elicitation often produces

unpredictable results due to the poor immunogenicity of carbohydrates and similarities to the host's own glycome.[7, 8] Antibodies are also expensive to produce, can have poor stability, and require large quantities of glycan, that can be difficult to obtain, for production.

Due to the drawbacks to antibody elicitation and lectin discovery, GBP engineering has become an increasingly popular field of study.[9, 10] These efforts are typically applied to existing lectins or adaptive immune proteins like antibodies, antibody fragments, and variable lymphocyte receptors. The methods used span from site-directed mutagenesis to large library directed evolution approaches that aim to improve the affinity for the native ligand or to change the specificity to structurally distinct sugars.[11-14]

In Chapter 2, the development of a GBP engineering platform based on the Sso7d DNA-binding protein of the hyperthermophilic archaea *Sulfolobus solfataricus* was presented. This protein was selected as it showed promise for GBP generation. Sso7d is a very small and stable protein that can be produced in the cytoplasm of *E. coli* and shares many features found in native GBPs. First, the binding face of the Sso7d protein is a flat, solvent-exposed surface comprising three β-sheets and shallow, solvent-exposed binding surfaces are common in lectins.[15] Second, the evolved Sso7d protein can be highly enriched in aromatic amino acids in its binding face without a subsequent loss in thermostability, expression level, and solubility.[16] This is important because the polarized C-H bonds of carbohydrates form CH-π interactions with aromatic amino acids and are a major contributor to carbohydrate binding.[17-19]. Using a yeast surface display library of reduced-charge Sso7d (rcSso7d) variants, a GBP for sialic acid-binding was successfully generated as described in Chapter 2 of this thesis.

This chapter will focus on the creation of GBPs for the Thomsen-Friedenreich antigen, also referred to as the T or TF antigen. The TF antigen is the Core 1 structure of O-linked mucin-type

glycans, consisting of the disaccharide Galβ1-3GalNAcα linked to certain Ser or Thr residues. In healthy tissue, the Core 1 structure is further modified by the addition of sialic acids, sulfates, or other glycans, but is often desialylated in cancerous and pre-cancerous tissue due to abnormal expression of glycosyl transferases and glycosidases in tumor cells. The epitope is a tumor associated carbohydrate antigen found in 90% of carcinomas, making it an attractive epitope for cancer diagnostics and immunotherapies.[20, 21] Because of this association to cancer, it has been the subject of several GBP engineering efforts and is an interesting target for Sso7d-based GBP engineering efforts.[11, 22-24]

In the following sections of this chapter, I will present the generation of an Sso7d-based GBP that recognizes the TF antigen by applying the yeast surface display-based directed evolution platform developed in Chapter 2, as well as modification and characterization for ultimate use as an affinity reagent. After two rounds of affinity maturation a TF-binding Sso7d variant was isolated and was demonstrated to have affinities and specificities on par with commercially available lectins. The GBP was developed into a reagent by site-directed installation of useful handles using sortase-mediated ligation and then applied to glycoprotein blotting and cancer cell labeling applications, demonstrating their utility as a reagent for the study of glycans in their native environments.

3.3 Enrichment of glycan-binding variants from naïve library

The rcSso7d library and selection procedures used in this study were described in detail in Chapter 2 of this thesis. Briefly, the library is a 1.4 x 10^9-member YSD library of a reduced charge variant of Sso7d (rcSso7d) (**Figure 3-1A,B**).[16] This unbiased library features nine variable amino acids in the β-sheet binding-face of the protein (paratope), randomly mutated to all amino acids except Pro and Cys. The rcSso7d variant is fused to the yeast surface protein Aga2P and contains

an N-terminal HA and C-terminal Myc tag to quantify proper expression of full-length proteins on the yeast surface. Yeast bearing binding variants can be isolated from the library manually using functionalized magnetic beads (magnetic-activated cell sorting or "MACS") or by fluorescence-activated cell sorting (FACS) with fluorescently labeled ligands (**Figure 3-1D,E**).

It is important that the selection system mimics nature to increase the opportunity to capture any low affinity variants with glycan-binding characteristics. Carbohydrate-protein interactions tend to be weak, with K_D values in the μM to mM range for a monovalent interaction.[25] In nature, GBPs often oligomerize to form multivalent interactions with glycans, increasing the strength of the interaction.[26, 27] This increased strength, or avidity, is presented as a functional K_D or apparent K_D (K_{app}) and can be orders of magnitude stronger than a monovalent interaction. In natural systems there also exists multivalency on the glycan side, either manifested as repeating units in one polysaccharide or as multiple proximal carbohydrates, for example the dense glycan-coating of a cell surface. A summary of reports of GBPs with multivalent saccharide ligands reveals that in all cases a multivalent ligand will show enhanced avidity, though the magnitude of this enhancement is highly variable.[28] The yeast surface display platform allows for multivalent display of the Sso7d because many copies of the protein are presented on the cell surface. Multivalent ligands were used for selections, consisting of either highly avid magnetic beads functionalized with the glycan of interest or commercially available polyacrylamide-based fluorescent glycopolymers (TF-PAA-FITC, **Figure 3-1E**).[29] These ligands allow for low affinity Sso7d-carbohydrate interactions to be captured, as monovalent biotinylated sugar ligands failed to capture glycan interactions and enrich the library in binding variants as described in detail in Chapter 2.

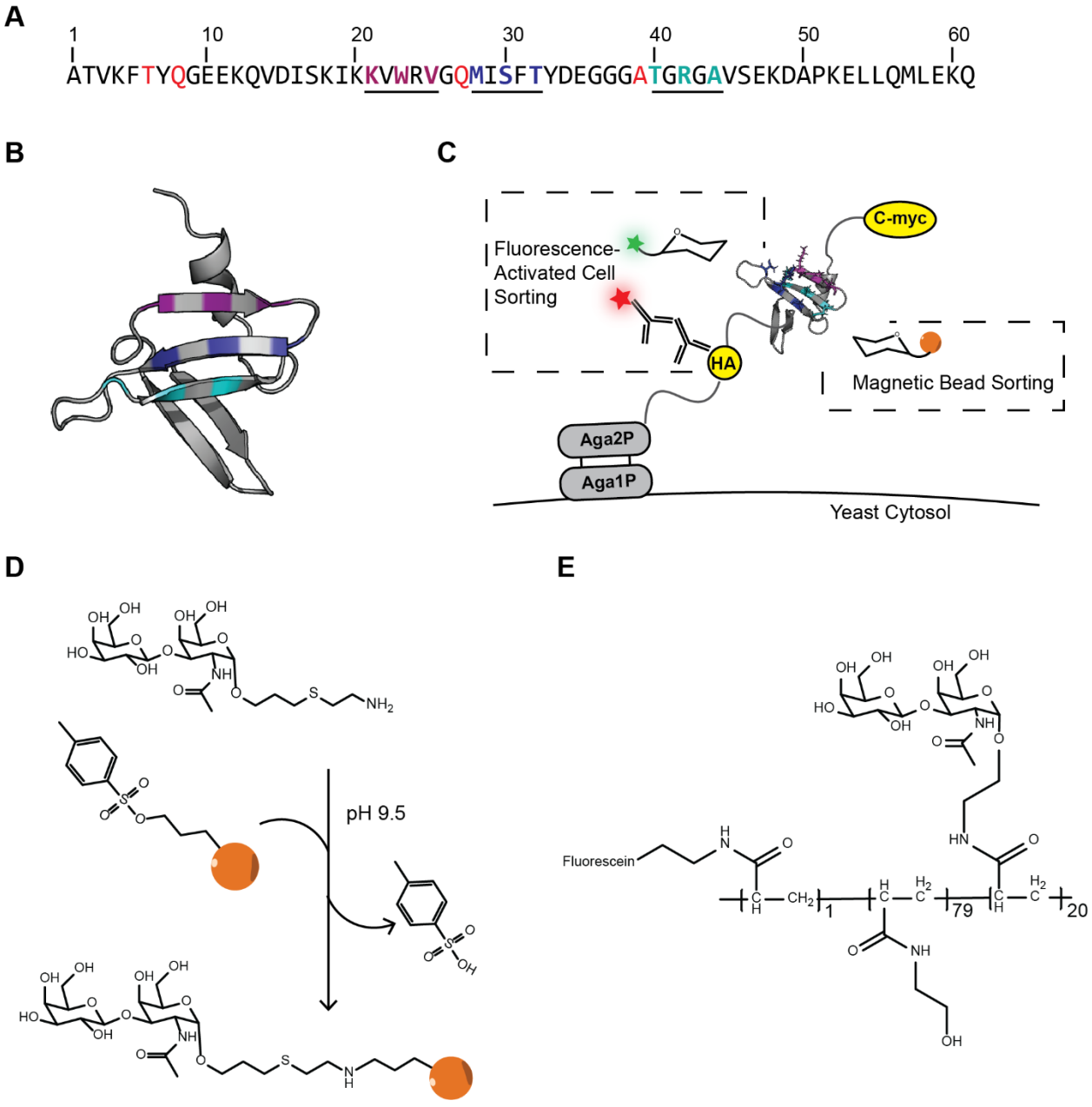


Figure 3-1. **Yeast surface display with rcSso7d-based library**.
A) Sequence of rcSso7d with residues mutated from Lys colored in red, and variable residues of the paratope colored in maroon, dark blue, and light blue. B) The Sso7d structure showing variable residues color coded to those in the sequence. C) Yeast surface display setup showing Sso7d fused to yeast protein Aga2P that makes disulfide bonds to yeast surface protein Aga1P with both N-terminal HA tag and C-terminal Myc tag. Yeast can be sorted using glycan conjugated magnetic beads or fluorescent glycan ligands. D) Tosyl-activated magnetic bead conjugated to amine-modified TF. E) Structure of TF-PAA-FITC polymer used for FACS sorts.

Selections with the large naïve library began with five rounds of enrichment using tosyl functionalized magnetic Dynabeads modified with a primary amine-bearing TF antigen as described for sialic acid in Chapter 2 (**Figure 3-1D**). To deplete the library in variants exhibiting nonspecific binding or linker-binding, three rounds of negative selections were interspersed with the positive selections using linker modified magnetic beads. The resulting population enriched in carbohydrate-binding variants was carried through three rounds of selection by FACS using the fluorescent TF-PAA-FITC polymer. After three rounds of FACS, the resulting enriched population exhibited a 14-fold increase in binding to the ligand as measured by the median fluorescence intensity (MFI) (**Figure 3-2A**) and the binding was glycan mediated as at no point did binding to secondary reagents or polyacrylamide polymer without carbohydrate (PAA-FITC) arise.

Plasmids were isolated from the final population and a subset were sequenced. Of 91 sequenced plasmids, eleven unique variants were represented. One variant, 0.8.F, repeated 28 times. Of the nine variable amino acids in the paratope region, 0.8.F had six aromatic amino acids: Trp25, Trp30, Tyr21, Tyr28, Tyr42, and Tyr44. (**Figure 3-2B**). Other variants present in the final population contained between two to six aromatic residues, primarily Trp and Tyr. This many aromatic amino acids were seen in the isolated sialic acid-binders from chapter 2 of this thesis, and could indicate multiple binding-sites or the presence of an aromatic "cage" around the sugar. The eleven unique variants were tested by making clonal yeast populations and all showed binding to TF-PAA-FITC over PAA-FITC though the magnitude of binding was highly variable (**Figure 3-2C**).

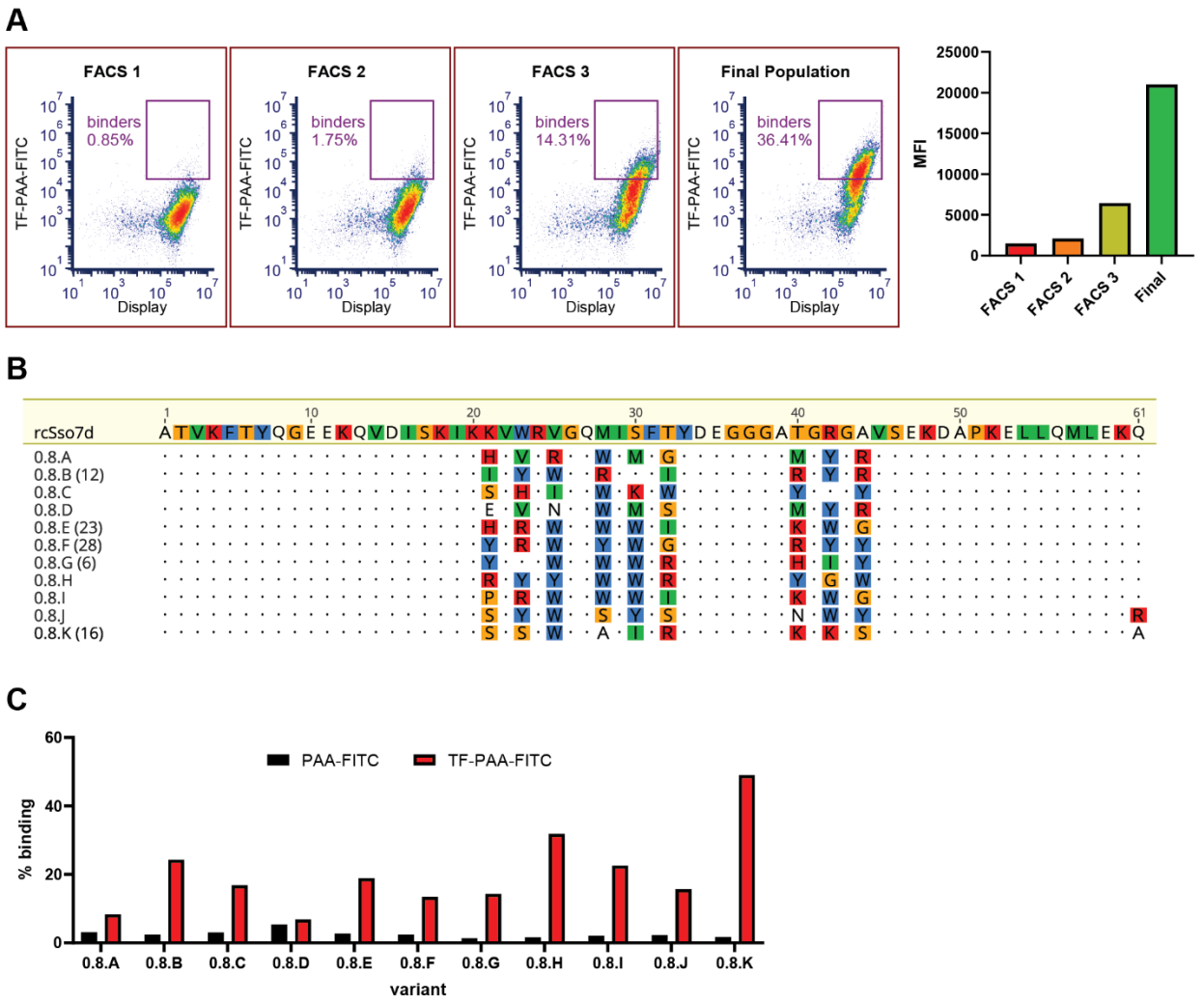


Figure 3-2. Enrichment of TF-binding variants from naïve library.
A) Dot plot and bar graph showing enrichment of cells binding to the TF-PAA-FITC polymer (y-axis) over the course of the three selections. B) Sequences of the eleven unique variants in the final population. Aromatic residues are colored in blue, basic residues in red, hydrophobic residues in green, and polar residues in yellow. C) % binding of PAA-FITC (black) and TF-PAA-FITC (red) to clonal yeast displaying the unique variants isolated after three selections.

3.4 Affinity maturation increases binding affinity and specificity of evolved glycan-binding proteins

To improve the binding affinity for the TF antigen, the selected variants were subjected to affinity maturation. Random mutations were introduced via error-prone PCR on all variants isolated from the final population giving rise to library 1.0. Three rounds of selection by FACS

were performed using 8-fold less TF polymer for increased stringency. As before, the new library was enriched in binding variants over the course of the selections. A subset of the population was sequenced, and hierarchical clustering was performed to curate representatives of variant families based on the paratope sequence. Of three clusters, the largest cluster contained variants that are descendants of 0.8.F, all containing two or more mutations outside of the binding face.

Several representatives from each cluster were chosen for further analysis by flow cytometry to assess affinity and specificity for TF using a panel of glycopolymers containing both highly similar and distinct mono-, di-, and trisaccharides (**Figure 3-3F,G,H**). An rcSso7d variant isolated in Chapter 2 that weakly binds biotin, called 83H, was tested with the glycopolymer panel to act as a negative control. The paratope region of this variant is distinct from the sugar-binding variants. It has two acidic Asp residues while the sugar-binding variants contain 1-3 basic Arg residues. 83H also has only two aromatic residues compared to the 5-6 in the sugar-binding variants, and importantly it shows no binding to any tested glycopolymer (**Figure 3-3A,B**).

TF-binding variant 1.3.D, a descendent of 0.8.F, was selected as it showed the highest specificity for TF over the other tested disaccharides (**Figure 3-3D**). 1.3.D has four additional mutations compared to 0.8.F: Q8R, K48T, K52E, and E53K all outside of the β-sheet binding face (**Figure 3-3A**). Variant 1.3.D showed specificity toward TF over the monosaccharide components Galβ and GalNAcα, differentiated between β vs α linkages of the Gal to GalNAcα (see Core 8 O-glycan Galα1-3GalNAcα), reversed directionality of the monosaccharide components (see Adi, GalNAcα-Galβ), 1,3 vs 1,4 linkages between highly similar monosaccharides (see LacNAc, Galβ1-4GlcNAcβ), and the inclusion of a large sulfate group (sulfo-TF). Variant 1.3.D also differentiated between a terminal and internal TF motif as seen with the H3 trisaccharide (Fucα1-2Galβ1-3GalNAcα).

There does exist some off-target binding to other disaccharides. Some are highly similar, like Le^c with a single stereochemical difference of GlcNAc vs GalNAc (see Le^c, Galβ1-3GlcNAcβ) and the Core 5 O-glycan (GalNAcα1-3GalNAcα) containing an additional N-acetyl and an α glycosidic linkage. Surprisingly, binding was also observed to more structurally distinct glycans like the negatively charged polysialic acid disaccharide (Neu5Acα2-8Neu5Acα), or chitobiose (GlcNAcβ1-4GlcNAcβ) which contains an additional N-acetyl group, two stereochemical differences, and a 1,4 linkage.

Compared to variant 0.8.F (**Figure 3-3C**), 1.3.D gained affinity for TF and the other binding glycans. We then hypothesized that further affinity maturation and selection could isolate a higher affinity variant, and subjected variant 1.3.D to an additional round of error prone PCR to generate library 2.0. At this point in the project, the TF-PAA-FITC glycopolymer reagent was out of stock and supply chain issues made it impossible to obtain additional polymer. To circumvent this problem, the glycopolymer with the highest degree of similarity that showed strong binding behavior, Le^c (Galβ1-3GlcNAcβ), was chosen for further selections to validate the affinity maturation hypothesis. Four rounds of selection with the Le^c glycopolymer were performed with reducing concentrations of polymer at each round. From this library variant 2.4.I was identified as having a high preference for Le^c over every other glycopolymer (**Figure 3-3E**). This variant has an additional three mutations compared to 1.3.D: W25R in the binding face, and D15A and D49A outside the binding face. The overall binding capacity, noted as y-median fluorescence intensity (MFI), for GalNAcα, Core 8, Core 5, H3 trisaccharide, and Le^c increased for 2.4.I while chitobiose, and polysialic acid decreased in MFI. Unfortunately, TF-PAA-FITC could not be tested but is expected to follow the trends of variant 0.8.F and 1.3.D and bind to variant 2.4.I to some degree.

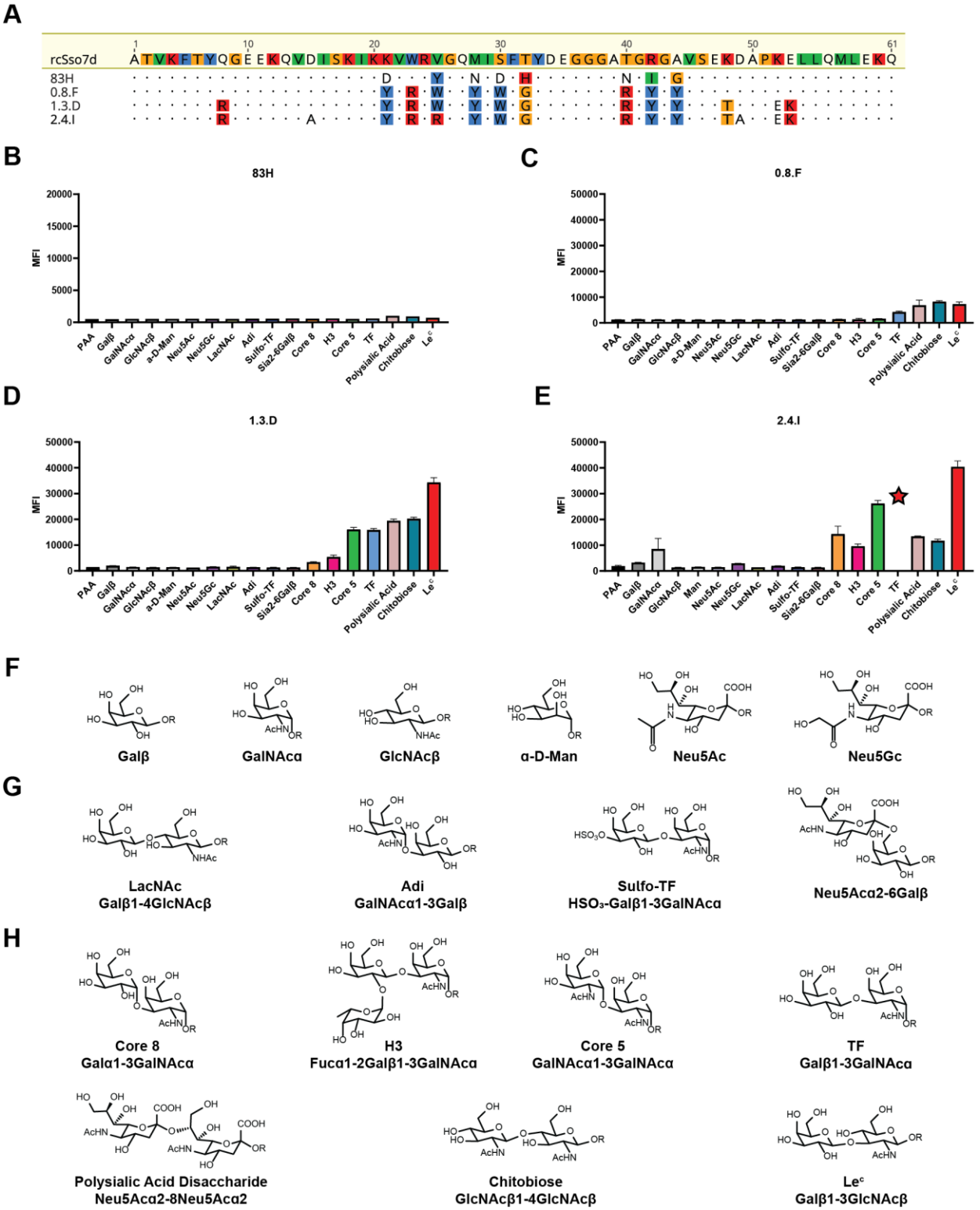


Figure 3-3. Variants isolated from affinity matured libraries exhibit selectivity to the target ligand.

A) Sequences of rcSso7d parent and selected non-sugar-binding and sugar-binding variants. B-E) Mono-, di-, and trisaccharide glycopolymers with clonal yeast expressing B) non-sugar-binding

variant 83H, C) 0.8.F, D) 1.3.D, and E) 2.4.I. Asterisk designates that TF could not be tested for this variant due to lack of material and supply chain issues. F) Structures of tested monosaccharides. G) Structures of non-binding disaccharides. H) Structures of binding di- and trisaccharides.

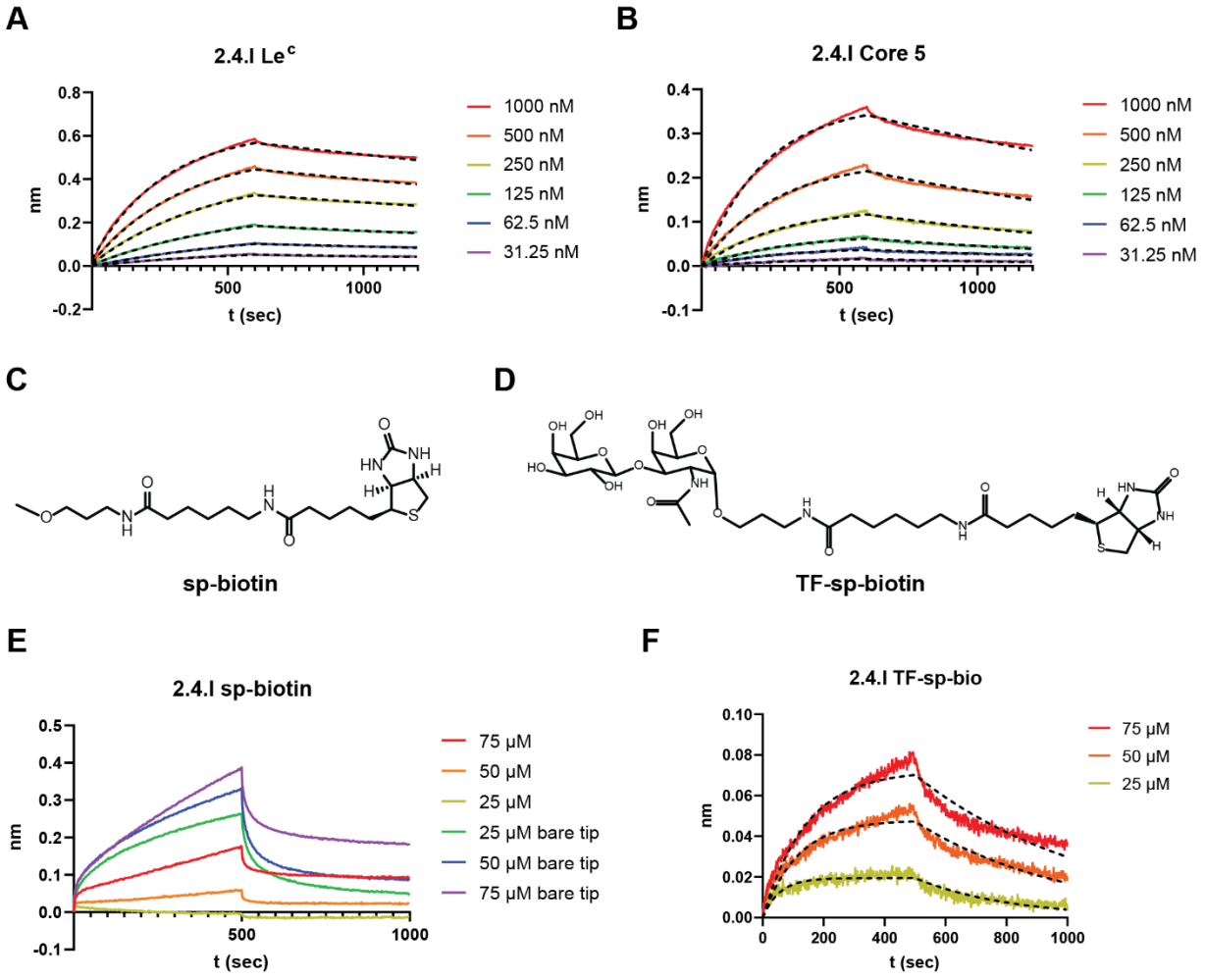
3.5 Isolated variants exhibit binding characteristics on par with commercial lectins

To characterize the binding affinities of variants removed from the yeast surface, the variant sequence was cloned into an expression vector containing a C-terminal hexahistidine tag and purified using immobilized metal affinity chromatography. Functional affinity constants were experimentally determined using bio-layer interferometry (BLI) with the fluorescent glycopolymers used in the flow-based selections and screenings. The resulting values are not true K_D's because the binding is multivalent. However, they are functionally relevant as they mimic the context of cell surface glycans being displayed in high density.

Soluble variant 2.4.I was immobilized on a Ni-NTA tip and dipped into a solution of Le^c or Core 5 glycopolymer to measure association rate constants (k_{on}) then moved to buffer to measure dissociation rate constants (k_{off}) (**Figure 3-4A-B**). The functional affinity measured for Le^c and Core 5 correlate with the YSD experiments, with Le^c having a K_{app} of 38.7 ± 1.4 nM and Core 5 having a higher K_{app} of 53.8 ± 2.9 nM for variant 2.4.I. The functional affinity of the TF glycopolymer could not be determined due to lack of sample but is expected to be in the range of the Core 5 and Le^c value based on the YSD experiments with variants 0.8.F and 1.3.D.

Although the functional affinity for TF could not be determined due to lack of polymeric reagent, the true K_D values for TF could be measured using a monovalent TF sugar. The disaccharide is linked to biotin by a short spacer unit (sp) which ensures the reducing end sugar is in the pyranose form (**Figure 3-4D**). The protein variant was immobilized and then dipped into a solution of the sugar (**Figure 3-4F**).[30] The monovalent K_D was measured at 16.4 μM ± 4.6 μM, a

value that correlates with those of monovalent lectin-sugar interactions.[25] When the control, sp-biotin, (**Figure 3-4C**) was tested in the same manner it showed a high degree of non-specific binding with the bare Ni-NTA BLI tips. This binding was reduced when 2.4.I was immobilized, indicating that sp-biotin did not bind specifically to 2.4.I (**Figure 3-4E**). This behavior was not seen with TF-sp-biotin, indicating a specific interaction with the protein and the sugar was taking place.



G

Ligand	K_D	k_{on} ($M^{-1}s^{-1}$)	k_{off} s^{-1}
Le^c-PAA-FITC	38.7 +/- 1.4 nM*	(8.5 ± 4.0) x 10^3	(2.8 ± 0.3) x 10^{-4}
Core 5-PAA-FITC	53.8 +/- 2.9 nM*	(2.0 ± 1.4) x 10^4	(7.1 ± 0.7) x 10^{-4}
TF-sp-bio	16.4 +/- 4.6 µM	(1.4 ± 0.4) x 10^2	(2.1 ± 0.4) x 10^{-3}

*Binding constant is functional affinity or K_{app}

Figure 3-4. Affinity measurements using biolayer interferometry.
A) BLI trace of immobilized 2.4.I with Le^c-PAA-FITC glycopolymer at a range from 31.25 to 1000 nM in solution. B) BLI trace of immobilized 2.4.I with Core 5-PAA-FITC at a range from 31.25-1000 nM in solution. C) Structure of sp-biotin. D) Structure of TF-sp-biotin. E) BLI trace of immobilized 2.4.I or bare tip with sp-biotin linker in solution at concentrations from 25-75 μM. F) BLI trace of immobilized 2.4.I with TF-sp-biotin monovalent sugar in solution at concentrations from 25-75 μM. G) Table of K_D, k_{on}, and k_{off} measured by BLI.

3.6 Protein labeling and applications

3.6.1 Sortase-mediated ligation installs handles for reagent-grade glycan binding proteins

To demonstrate the use of the evolved proteins as affinity reagents, they must be labeled with useful handles such as biotin and fluorophores. Sortase-mediated ligation (SML) is a powerful protein engineering method that allows for the site-specific incorporation of various biochemical and biophysical probes to the N or C-terminus of a protein. The SrtA enzyme of *S. aureus* catalyzes a transpeptidase reaction between a C-terminal LPXTG motif and an N-terminal oligoglycine sequence (**Figure 3-5A**).[31, 32] SML is advantageous because the soluble version of SrtA enzyme is robust and easy to produce in *E. coli*, has very high substrate specificity, and allows for incorporation of many different chemical handles. This method also allows for site specific labeling without the introduction of a reactive Cys residue into the protein or by labeling the primary amines of the many Lys residues contained in the evolved Sso7d proteins.

To label the C-terminus of the evolved GBPs with biotin or a FITC fluorophore, the variant was cloned into an expression vector containing a C-terminal LPETGG motif followed by a hexahistidine tag for immobilized metal affinity chromatography. Upon ligation the hexahistidine tag is excised allowing for purification of the final product away from the His-labeled SrtA enzyme. Oligoglycine-containing peptides appended to a biotin or FITC handle were ordered with the sequence GGGYK[K-biotin]T-amide or GGGYK[K-FITC]T-amide. SML was carried out by

incubating the variant 2.4.I with the nucleophilic peptide and SrtA with a hexahistidine tag in Ca^{2+} containing buffer. Labeled protein was separated from the excised His tag and SrtA by passing through Ni-NTA resin followed by desalting to remove excess peptide (**Figure 3-5B**). Protein labeling was confirmed by blotting with streptavidin or by fluorescence-based gel imaging (**Figure 3-5C**), validating SML as a robust means towards generating reagent-level Sso7d-based GBPs.

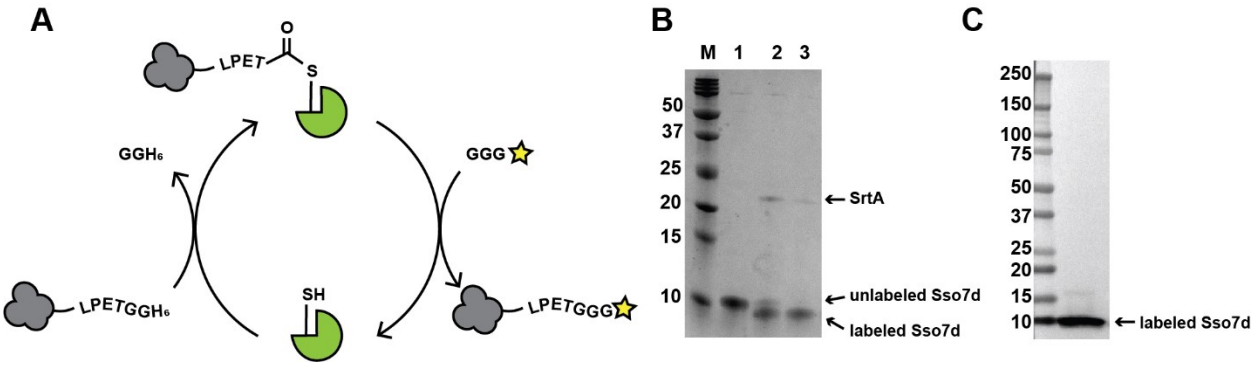


Figure 3-5. Modification of protein variants by sortase-mediated ligation.
A) Scheme of SML reaction. Sso7d variant shown in grey with C-teriminal sortase motif and His6 affinity tag reacting with GGG containing peptide, yellow star denotes handle such as biotin or fluorophore. B) Coomassie stained gel showing example reaction of variant 2.4.I with biotin peptide. 1: Unreacted 2.4.I 2: SML reaction mix after 30 minutes at room temperature 3: Purified labeled 2.4.I-biotin. C) Streptavidin Western blot showing presence of biotin labeled 2.4.I from reaction shown in B.

3.6.2 Engineered variants 2.4.I binds TF-containing glycoproteins

The biotin labeled TF-binding variant, 2.4.I-biotin, was then utilized in dot blotting application for detection of proteins known to carry the TF epitope and similar O-glycans. The selected glycoproteins tested were the mucin proteins MUC2, MUC5AC, and MUC5B, as well as glycophorin A, and fetuin.

Mucins are a family of large, heavily O-glycosylated proteins that are abundant in mucus. Up to 80% of their mass can made up of highly heterogeneous O-glycans.[33] MUC2, MUC5AC, and MUC5B are all secreted gel-forming mucins but their main tissue expression differs. MUC2

is mainly expressed in the intestine where it is a major component of colon mucus, while MUC5AC and MUC5B are both integral components of airway mucus, with MUC5B also present in salivary mucus and MUC5AC in gastric mucus.[34] Mucin O-glycans consist of Core 1 (Galβ1-3GalNAc, TF antigen), 2 (Galβ1-3(GlcNAcβ1-6)GalNAc), 3 (GlcNAcβ1-3GalNAc), and 4 (GlcNAcβ1-3(GlcNAcβ1-6)GalNAc), O-glycans (**Figure 3-6A**) that may be further modified with a sulfate group, or with additional monosaccharides such as sialic acid, and fucose. Glycophorin A (GPA) is a transmembrane sialoglycoprotein found on red blood cell membranes containing a high degree of sialylated TF antigen (**Figure 3-6B**). Removal of the sialic acids reveals the Core 1 TF antigen structure. Fetuin is a circulating glycoprotein containing sialylated N- and O-glycans. The N-glycans are typically di- or tri-branched with underlying lactosamine units (Galβ1-4GlcNAc), while the O-glycans are mono or di-sialylated TF antigen (**Figure 3-6C**). The carbohydrate composition of these glycoproteins is therefore well suited for post-enzymatic treatment detection with our engineered GBP.

Glycoprotein samples were treated with neuraminidase to remove α2-3,6,8 linked sialic acids, or with neuraminidase and O-glycosidase (Endo-α-N-Acetylgalactosaminidase) to remove Core 1 and Core 3 O-glycans. Treatment with neuraminidase will reveal more TF antigen by removing sialic acids from the many sialylated TF structures present on the glycoproteins and is expected to lead to increased binding signal. Removal of Core 1 and Core 3 O-glycans is expected to decrease binding signal as the glycoproteins will lose the epitope of interest to variant 2.4.I. After enzymatic treatment the glycoproteins were spotted on a nitrocellulose membrane and incubated with either 2.4.I-biotin or the non-sugar-binding Sso7d variant 83H-biotin. There is no observed binding to any glycoprotein sample with 83H-biotin or the streptavidin only control (**Figure 3-6D**). In contrast there is clear binding to MUC2 by 2.4.I with and without enzymatic

treatment, and a low level of binding to the four other glycoproteins. The dot blot was repeated with the best positive binders, MUC5AC and MUC2 (with and without neuraminidase and O-glycosidase treatment) and decreasing concentrations of 2.4.I. There is a clear preference for 2.4.I binding to MUC2 compared to MUC5AC, and as expected, the neuraminidase treated MUC2 sample exhibits the greatest binding signal (**Figure 3-6E**).

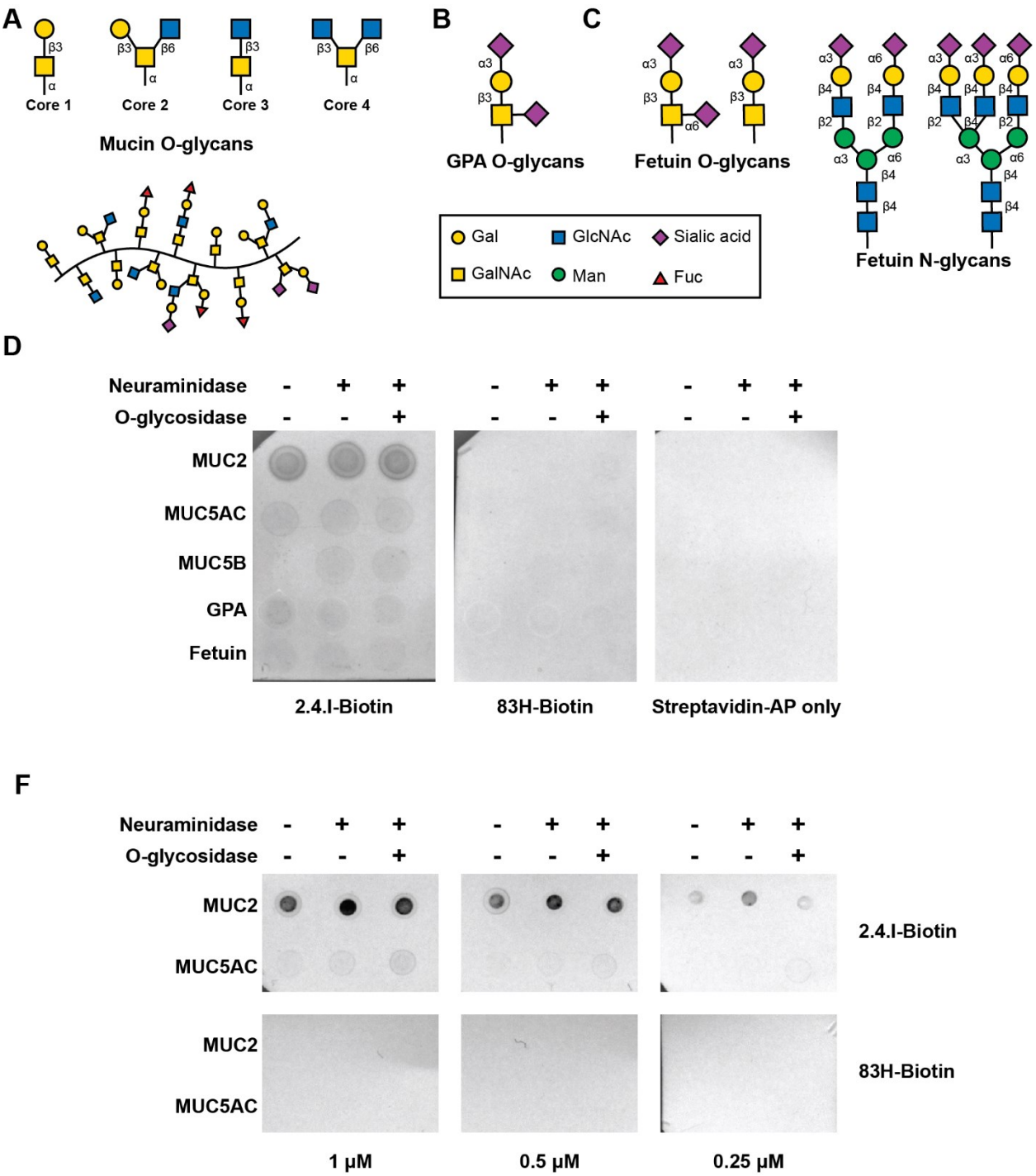


Figure 3-6. Variant 2.4.I binds TF antigen-containing glycoproteins by dot blot.
A) Structure of core O-glycans found in mucins with mucin cartoon shown below. B) GPA sialylated TF antigen. C) Fetuin glycans. Sialylated TF antigen O-glycans on left, examples of fetuin N-glycans on right. D) Dot blot of glycoproteins MUC2, MUC5AC, MUC5B, GPA, and Fetuin with sugar-binding variant 2.4.I, non-sugar-binding Sso7d variant 83H, and streptavidin-AP conjugate only. Samples are either untreated, treated with neuraminidase only, or treated with both neuraminidase and O-glycosidase. E) Dot blot of MUC5AC and MUC2 with 2.4.I at

concentrations 1, 0.5, and 0.25 μM to better show the impact of neuraminidase and O-glycosidase treatment.

3.6.3 Mammalian cell staining

The FITC labeled variant, 2.4.I-FITC, was utilized *in vitro* to stain cancerous mammalian cells containing a surface-presented TF epitope. MCF7 cells, a human breast cancer epithelial cell line, were used for this purpose as they express high levels of TF antigen on the surface.[35] MCF7 cells were treated with neuraminidase and O-glycosidase to remove sialic acid, and Core 1 and Core 3 O-glycans respectively. However, treatment with O-glycosidase did not decrease binding as expected. In fact, this enzymatic treatment increased binding of 2.4.I-FITC to the cell (**Figure 3-7A**). It has already been demonstrated that 2.4.I can bind to other epitopes than the TF antigen, therefore it is possible that removal of Core 1 and 3 glycans may reveal other epitopes for 2.4.I to bind, leading to the observed increased signal. This binding increase could also be explained by a reduction in steric hinderance by removal of some of the many cell-surface glycans. To remove more surface glycans that may be responsible for the observed phenomena, cells were treated with a protein deglycosylation cocktail that contains PNGase F (N-glycan), α2-3,6,8,9 neuraminidase A (sialic acid), O-glycosidase (Core 1 and 3 O-glycan), β1-4 galactosidase S (galactose), β-N-acetylhexosamindase (GalNAc and GlcN). This cocktail can remove N-glycans and more complex O-glycans. However, this treatment did not significantly increase or decrease binding signal (**Figure 3-7B**). In all cases, 2.4.I-FITC stained cells better than the non-sugar-binding variant 83H, which showed a minor amount of staining above background.

Treatment with glycosidases typically requires denatured proteins or long incubation times, neither of which are possible when staining whole cells. MCF7 cells instead were harvested, lysed, and denatured, before extended enzymatic treatment. The treated and mock treated samples were

fractionated by SDS-PAGE and proteins were transferred to a nitrocellulose membrane and blotted with 2.4.I-biotin or 83H-biotin. Even with extended enzymatic treatment there was no difference between the mock treated and enzyme treated samples (**Figure 3-7C**). However, there was no binding to any protein with variant 83H or secondary reagents only.

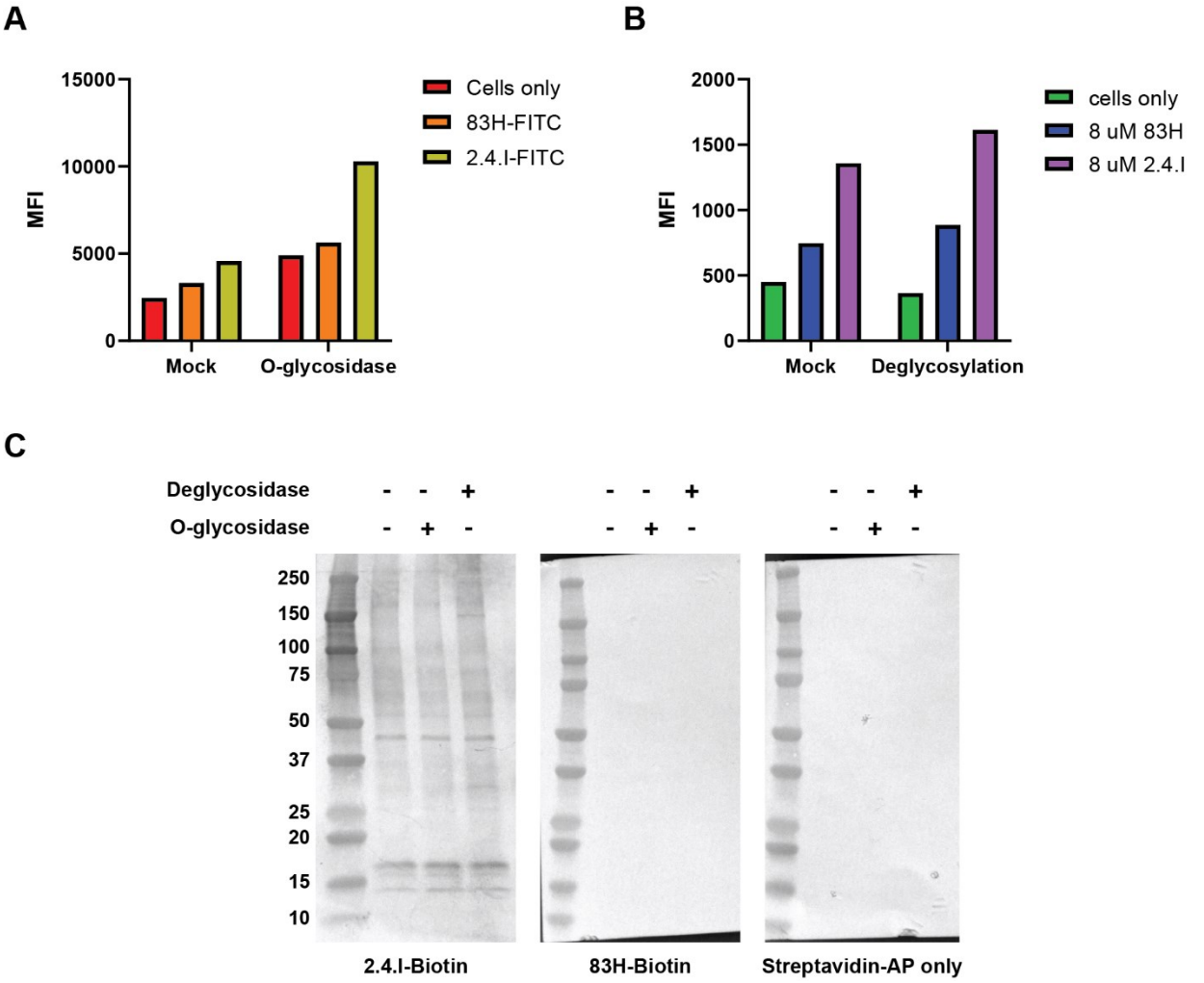


Figure 3-7. Mammalian cell staining with engineered GBP 2.4.I-FITC.
A) Median fluorescence intensity (MFI) of untreated and O-glycosidase treated MCF7 cells unstained, stained with non-sugar-binding variant 83H-FITC, or 2.4.I-FITC. B) Mock and deglycosylation treatment of MCF7 cells unstained, stained with non-sugar-binding variant 83H-FITC, or 2.4.I-FITC. C) Western blot using 2.4.I-bio of untreated and treated MCF7 cell lysates.

3.7 Computational analysis provides insight into 2.4.I-mediated disaccharide binding

Computational methods were applied to provide insight into the protein determinants responsible for binding to disaccharide ligands utilized during this GBP evolution campaign. The HADDOCK webserver was used to dock the binding disaccharides onto a model of variant 2.4.I.[36] The protein model was built using ColabFold, a structure prediction program using Alphafold[37] and MMseqs2[38] without side chain relaxation. The paratope region residues were assigned as the active residues and TF, Le^c, and Core 5 disaccharide were fit using the protein-glycan setting of HADDOCK. Disaccharide structures were modeled using the CSDB/SNFG structure editor, an online glycan builder for 3D structure visualization.[39] For all disaccharides, between 191-197 total structures were grouped into 4-6 clusters. The top scoring (lowest energy) HADDOCK model for each disaccharide is shown in **Figure 3-8**. TF is predicted to make four polar contacts with residues Arg25 and Trp30 and protons H4 and H6 of the α-GalNAc in TF are in proximity to residue Tyr44 and capable of forming CH-π interactions. The Core 5 disaccharide is predicted to make three polar contacts with Arg25 and Trp30 and like TF, the H4 and H6 protons of the α-GalNAc are in proximity with Tyr44 for CH-π interactions. Le^c is predicted to make one polar contact with Trp30 and the axial protons H1, H3, and H5 of the β-GlcNAc are all in proximity to Tyr44 for CH-π interactions. From this model, the carbohydrates are making both polar contacts and CH-π interactions with surrounding residues as is commonly seen in protein-carbohydrate interactions.[40] Based on these models, residues Try44 and Arg25 are particularly important for binding. Arg25 is a mutation gained during affinity maturation of 1.3.D and could, in part, explain the increased binding affinity gained by 2.4.I.

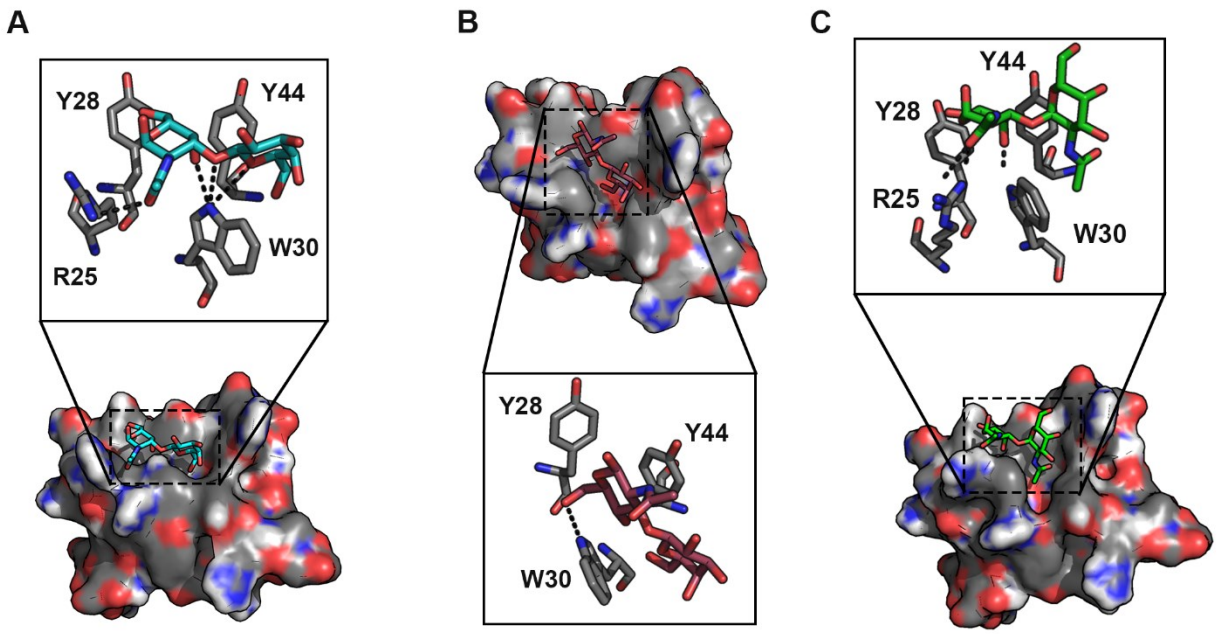


Figure 3-8. Docking of disaccharides to 2.4.I model.
A-C) Surface structure of an Alphfold model of variant 2.4.I docked to A) TF, B) Le^c, C) Core 5 using HADDOCK webserver. Residues making contacts to the disaccharide ligand are shown enlarged in boxes with dashed lines indicating polar contacts between the disaccharide and protein.

3.8 Conclusions

In this chapter, the process of evolving and characterizing a GBP for the tumor associated carbohydrate antigen TF was described. Using yeast surface display and multivalent glycan ligands, a small and stable DNA-binding protein was evolved into a carbohydrate-binding protein. The evolved protein shows affinities and specificities on par with commercially available lectins and was demonstrated to be functional as an affinity-binding reagent for Western blot analysis and cell staining applications. This method allows for the generation of GBPs for a specific glycan of interest and does not suffer from the drawbacks of carbohydrate-specific antibody elicitation.

Evolved GBPs bind TF but show a broader selectivity to other glycans with different monosaccharide components and linkages such as Le^c, chitobiose, and the Core 5 O-glycan

disaccharide. This is not unlike commercially available lectins and is an important piece of understanding the capabilities and limitations of a commercial reagent. Analysis of the selectivities of several TF-binding lectins show binding motifs like the evolved 2.4.I variant generated in this chapter.[41] One is *Artocarpus integrifolia* agglutinin (AIA or Jacalin). This lectin predominantly binds Core 1 (TF) and Core 3 (GlcNAcβ1-3GalNAcα) O-glycans, but additionally can bind substituents at the C3 position of the GalNAc residue such as GalNAc, GlcNAc, Gal, or longer oligosaccharides with an α or β linkage. Modeling of glycans bound to 2.4.I explains some of the observed binding behavior. The Le^c disaccharide shows the greatest binding to 2.4.I based on yeast surface data and BLI measurements. This glycan was able to adopt a planar configuration and fit into the groove made by Trp30 and Tyr44, allowing for CH-π interactions between Tyr44 and the axial hydrogens H1, H3, and H5 of the β-GlcNAc. This was not possible with the α-GalNAc residues present in TF and the Core 5 disaccharide because it does not contain the three axial hydrogens present in β-GlcNAc. The presence of β-GlcNAc alone is not enough to explain the binding, as other β-GlcNAc containing glycan LacNAc (Galβ1-4GlcNAcβ) did not bind 2.4.I. More structural data is required to gain a better insight into the binding determinants of 2.4.I.

The measured K_D of the evolved GBP 2.4.I toward monovalent TF is high at 1.5 μM. Although this is in line with commercially available lectins that have K_D values of 1-10 μM for complex glycans, it is much stronger than the mM K_D values that these lectins exhibit for monosaccharides.[8, 42] This affinity was achieved after only two rounds of affinity maturation, which introduced mutation Arg25 in variant 2.4.I that is anticipated to make polar contacts with the bound glycan. Other mutations were introduced outside of the binding face that lead to an increase in affinity from variant 0.8.F to 1.3.D. The role these mutations play in binding will require further structural information and highlight the power of directed evolution by introduction

of mutations that are distal to the binding surface but impact binding in a way that could not be predict. We anticipate that additional rounds of affinity maturation could be performed to achieve further optimization. Many YSD campaigns have involved greater than two rounds of affinity maturation to achieve sub-nanomolar binding affinities, but whether this type of affinity is possible for a carbohydrate-binding protein is unknown.[43, 44] There has been work towards increasing the pace of affinity maturation, such as 'autonomous hypermutation yeast surface display' (AHEAD), a new strategy that pairs an orthogonal error-prone DNA replication system with yeast surface display.[45] This mimics somatic hypermutation of the scaffold protein without performing the mutagenesis steps typically required by display technologies. Future endeavors with this method could more quickly sample the sequence space of the Sso7d for development of even higher affinity GBPs. Engineered multivalency is another tactic that can be deployed with the Sso7d-based GBPs. This has been applied to GBP engineering before to create multimers with superior binding affinities to monomeric GBPs, although this strategy could impact the commercial scalable production that makes the Sso7d unique.[46, 47]

Even without further affinity engineering the evolved 2.4.I variant proved useful as a reagent for glycoprotein blotting and cell staining. The 2.4.I GBP can bind to the glycoproteins MUC2, MUC5AC, MUC5B, GPA, and fetuin, all of which contain the TF epitope that has been evolved for, usually as the sialylated form. However, there is a clear preference for MUC2. Analysis of human MUC2 shows over 100 unique glycan structures with a high degree of sialylation, sulfation, and fucosylation.[48] Most MUC2 glycans are Core 3 structures of GlcNAcβ1-3GalNAc which was not tested for binding with the glycopolymer panel and could be a recognized epitope for 2.4.I. Alternatively, gastric mucins (including MUC5AC) are primarily Core 2 structures and MUC5B is primarily Core 1 and 2.[49, 50] This difference in core O-glycan structures

could explain the difference in binding abilities, but further testing into the specificity of variant 2.4.I is needed.

The work outlined herein therefore demonstrates how the platform developed in Chapter 2, which was successfully applied to the negatively charged nine-carbon sialic acid, can be applied to a neutral disaccharide, the cancer-associated epitope TF antigen, for the generation of selective, tight binding Sso7d-based GBPs. This underscores versatility; the platform can be applied to other glycans of interest, especially those with no current reagents for their detection. For the platform to be successful, multivalent presentation of the glycan of interest is critical. Multivalent polyacrylamide-based glycopolymers are commercially available for many mammalian glycans like those used in this report, but in-house synthesis of multivalent ligands might be necessary for unique or unusual glycans. This platform, coupled with glycan isolation and multivalent synthetic procedures, enables engineering of GBPs that can specifically recognize a vast array of user-defined glycans of interest. This will have great impact on glycobiological research, allowing for the study of important glycans with no existing affinity tools.

3.9 Materials and methods

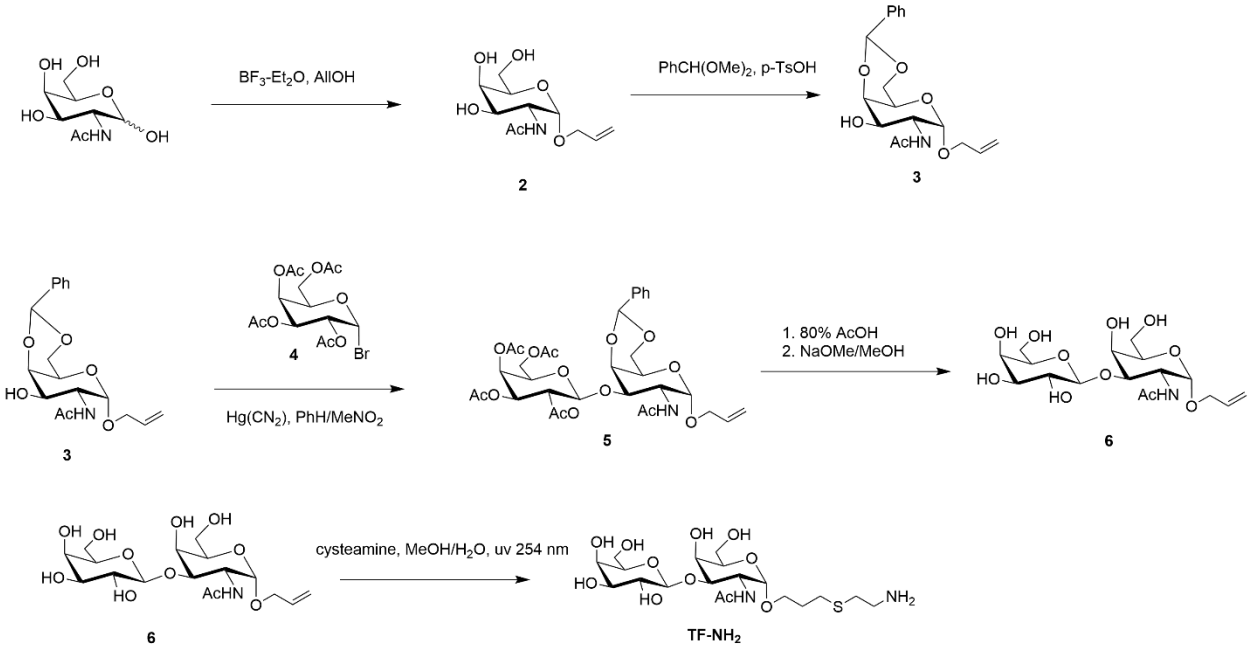
Materials

For flow cytometry and fluorescence-assisted cell sorting (FACS), detection reagents were as follows: chicken anti-HA (Exalpha Biologicals, Shirley, MA, Cat. No: AHA), chicken anti-cMyc (Exalpha Biologicals, Shirley, MA, Cat. No. ACMYC), goat anti-chicken AlexaFluor 647 (ThermoFisher Scientific, Cat. No: A-21449.) Multivalent carbohydrate polymers were commercially purchased as PAA-FITC, PAA-biotin, or PAA variations from GlycoTech (Frederick, MD.) All mammalian cell lines were purchased from ATCC and the appropriate growth media and supplements purchased from ATCC or MilliporeSigma, unless otherwise specified. All restriction enzymes, Taq DNA polymerase and glycosidases were purchased from New England Biolabs (Ipswich, MA) unless otherwise noted.

Yeast culture conditions

Use of yeast-surface display for directed evolution of glycan binders was carried out according to established protocols published by Wittrup and coworkers [51]. All yeast work was done under aseptic conditions using proper techniques. Yeast populations bearing yeast-display vector pCTCON2 were grown in SDCAA media containing 100 U/mL penicillin-streptomycin at 30 °C for routine culture and subculture, unless otherwise noted. Yeast-surface protein expression was induced by subculture and resuspension of log-phase yeast cultures in SGCAA media containing 100 U/mL pen-strep at 20 °C. All yeast washes and selections were carried out in PBSA (PBS at pH 7.4 with 0.1% bovine serum albumin, sterile filtered) at 4 °C.

Synthesis of TF-NH_2 and control linker



Scheme of synthetic route to glycoconjugate TF-NH_2

General Information. LCMS analyses were conducted on a nominal mass Agilent 6125B mass spectrometer with an electrospray (ESI) source attached to an Agilent 1260 Infinity LC. High-resolution mass spectral analyses were obtained on a Q-TOF high-resolution Agilent 6545 mass spectrometer coupled to an Agilent Infinity 1260 LC system with a Jet Stream ESI source.

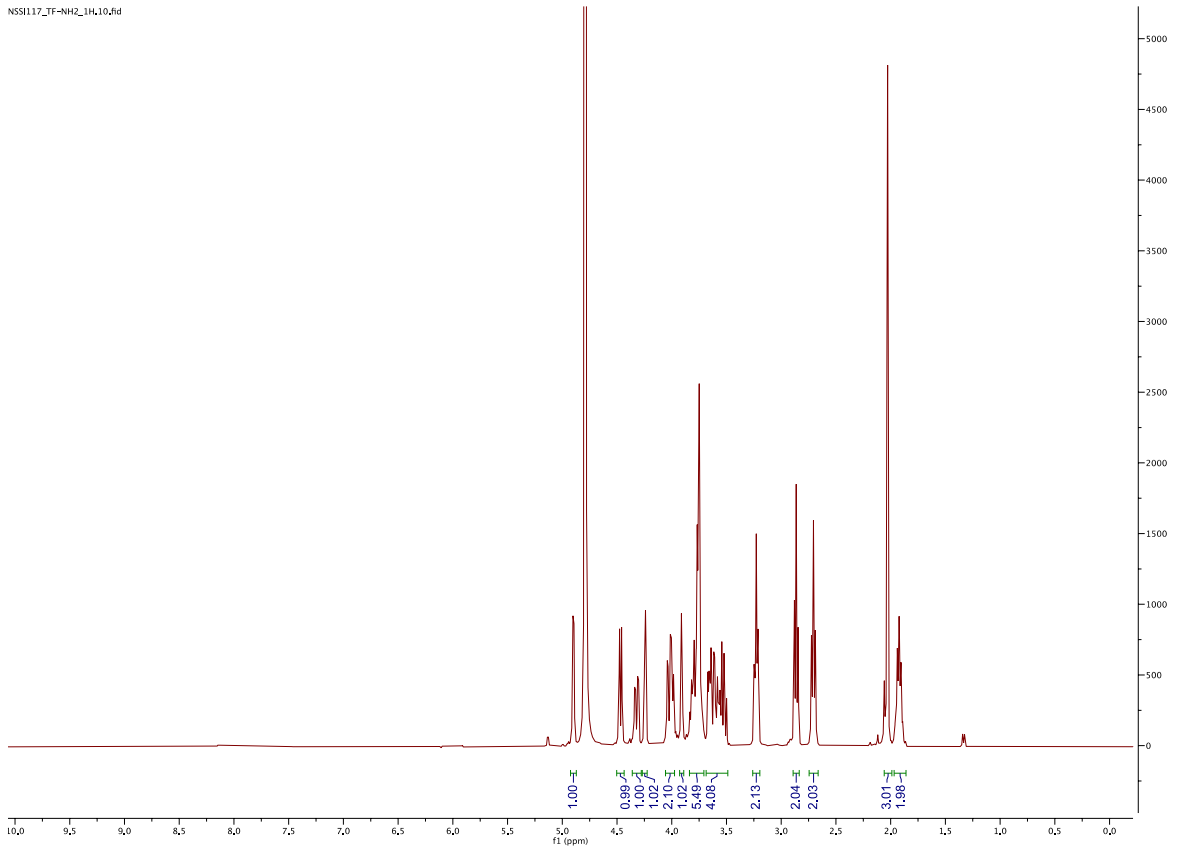
Synthesis of O-allyl 2. Following a procedure by Feng et. al.[52], to a solution of *N*-acetylgalactosamine (GalNAc) (1 g, 4.52 mmol) in allyl alcohol (18 mL) was added BF_3-Et_2O dropwise (642 µL, 4.52 mmol). The reaction mixture was stirred vigorously at 70 °C for 2 h, followed by cooling to ambient temperature and neutralization with Et_3N (1 equiv). The mixture was concentrated *in vacuo* and the crude purified by flash chromatography (12:1 to 7:1 CH_2Cl_2/MeOH) to afford **2** (849.1 mg, 72% yield), which matched the physicochemical properties

of the literature reference.

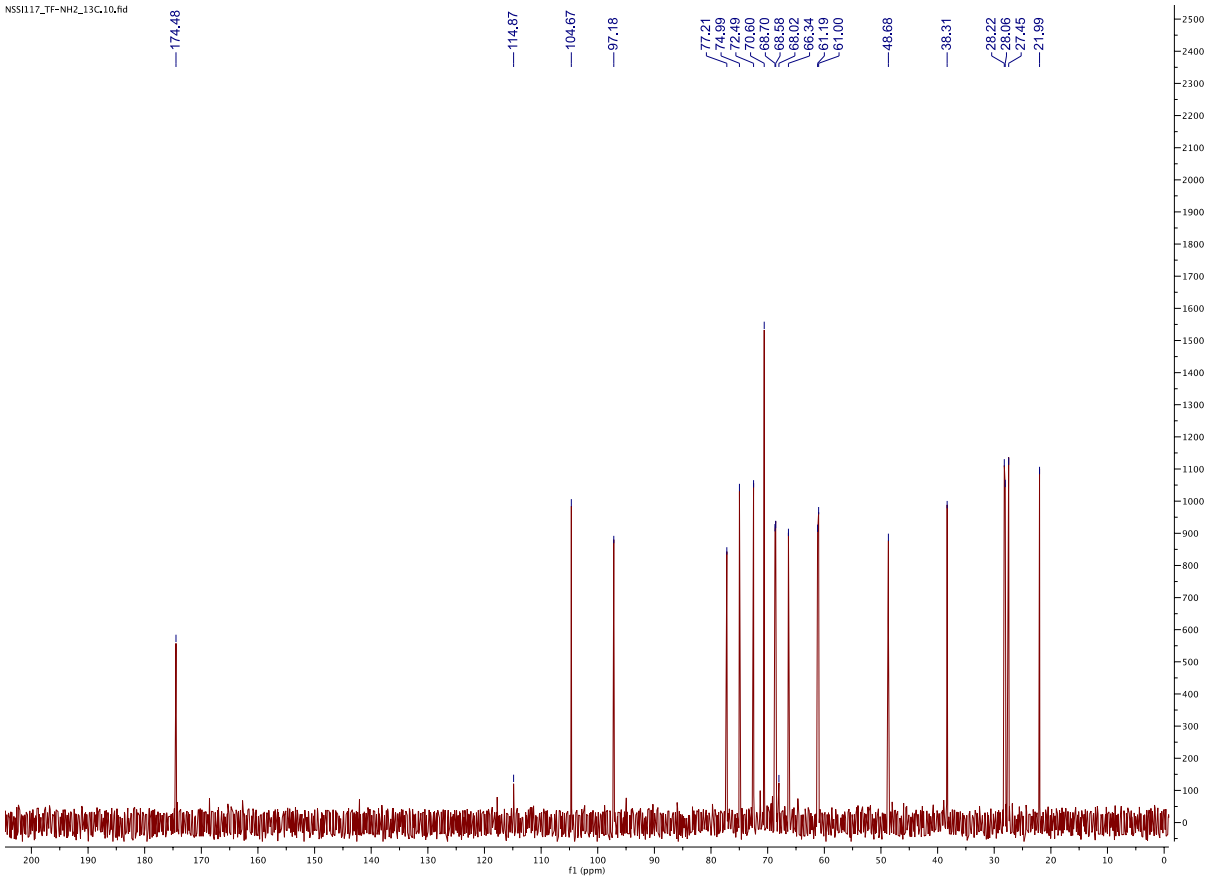
Synthesis of 3, 5, and 6. Following a procedure from Baek and Roy[53], *O*-allyl sugar **2** (773 mg, 2.957 mmol) and anhydrous tosylic acid (57 mg, 0.3 mmol) were added to a flame-dried round-bottom flask and the flask was charged with N_2. To this, anhydrous DMF (11 mL) was added under nitrogen and $PhCH(OMe)_2$ added dropwise with stirring. The reaction was heated to 60 °C with nitrogen bubbling through. After 1 h, the reaction was concentrated *in vacuo* and the crude compound was purified by flash chromatography twice (15:1 CH_2Cl_2/MeOH) to afford benzylidene **3** (456 mg, 44% yield) which matched the physicochemical properties of the literature reference.[53] Acceptor **3** (438 mg, 1.25 mmol) and purchased glycosyl halide donor **4** (773 mg, 1.88 mmol, CAS # 3068-32-4, Sigma-Aldrich) were dissolved in anhydrous 1:1 nitromethane/benzene (10 mL) in a flame-dried flask. The reaction was charged with promoter $Hg(CN)_2$ (475 mg, 1.880 mmol) and the reaction proceeded for 3 h. The crude reaction mixture was filtered through Celite, concentrated and purified by flash chromatography (20:1 to 10:1 CH_2Cl_2/MeOH) to afford disaccharide **5** (263.4 mg, 31% yield) which matched the physicochemical properties of the literature reference.[53] Disaccharide **5** (456 mg, 0.671 mmol) was brought up in 80% AcOH/H_2O and held at 60 °C with stirring for 2 h. Once deprotection of 1,2 diol was confirmed by TLC, the reaction was concentrated *in vacuo* with addition of toluene and the crude purified by flash chromatography (20:1 CH_2Cl_2/MeOH). The desired product was isolated in 52% yield (204 mg) with identical NMR spectra to those previously reported. Purified deprotected disaccharide (0.204 mg, 0.345 mmol) was dissolved in a solution of 0.15 M NaOMe in anhydrous MeOH (2 mL) and the mixture stirred for 2 h until spot-to-spot conversion observed by TLC. After, freshly washed Amberlite IR-120 H^+ resin was added to the reaction until pH 6 was obtained. The resin was

subsequently filtered away and the filtrate was concentrated twice by rotary evaporation resulting in a residue containing **6** which was carried to the next step without further purification (146 mg, quant.). The isolated product matched the physicochemical properties of the literature reference.[53]

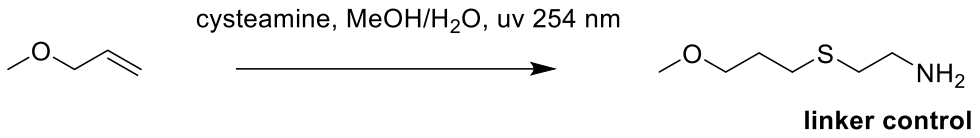
Synthesis of TF-NH$_2$. Fully deprotected glycoside **6** (129 mg, 0.305 mmol) and cysteamine-HCl (42 mg, 0.370 mmol) were dissolved in deoxygenated H_2O (3 mL) in a quartz cuvette and irradiated at 254 nm for 1 hr. Once the reaction appeared complete by TLC, the crude material was purified by reversed phase HPLC (Waters 1525 binary HPLC mounted with a 00G-4252-PO-AX (Luna Co.) preparative column, using a H_2O/MeCN with 0.1% TFA gradient from 4-30% MeCN) to afford final product **TF-NH$_2$** in 41% yield (63.6 mg). The isolated product matched the physicochemical properties of the literature reference.[54] ^{1}H NMR (400 MHz, D_2O) δ 4.89-4.91 (d, 1H, J = 3.8 Hz, GalNac-H1); 4.45-4.49 (d, 1H, J = 7.7 Hz, Gal-H1); 4.29-4.35 (dd, 1H, $J_{1,2}$ = 3.7 Hz, $J_{2,3}$ = 11.1 Hz, GalNac-H2); 4.24 (broad d, 1H, GalNAc-H4); 3.97-4.06 (m, 2H, GalNac-H3, GalNAc-H5); 3.89-3.93 (broad d, 1H, $J_{3,4}$ = 3.3 Hz, Gal-H4); 3.70-3.85 (m, 5H); 3.49-3.69 (m, 4H); 3.19-3.26 (t, 2H, J = 6.7 Hz, CH_2); 2.83-2.89 (t, 2H, J = 6.7 Hz, CH_2); 2.66-2.74 (dd, 2H, J = 7.1 Hz, CH_2); 2.03 (s, 3H, Ac); 1.88-1.97 (m, 2H, J = 6.8 Hz, CH_2) ppm. ^{13}C-NMR (100 MHz, D_2O) δ 174.48; 104.67; 97.18; 77.21; 74.99; 72.49; 70.60; 68.70; 68.58; 66.34; 61.19; 61.00; 48.68; 38.31; 28.22; 28.06; 27.45; 21.99 ppm. LRMS (ESI+) m/z: [M+H]+ Calcd for $C_{19}H_{37}N_2O_{11}S^+$ 501.2; Found 501.1. HRMS (ESI+) m/z: [M+Na]+ Calcd for $C_{19}H_{36}N_2O_{11}SNa$ 523.1938; Found 523.1936.



^{1}H NMR Spectrum (400 MHz, D_2O) of TF-NH_2.



^{13}C NMR Spectrum (100 MHz, D_2O) of TF-NH_2.



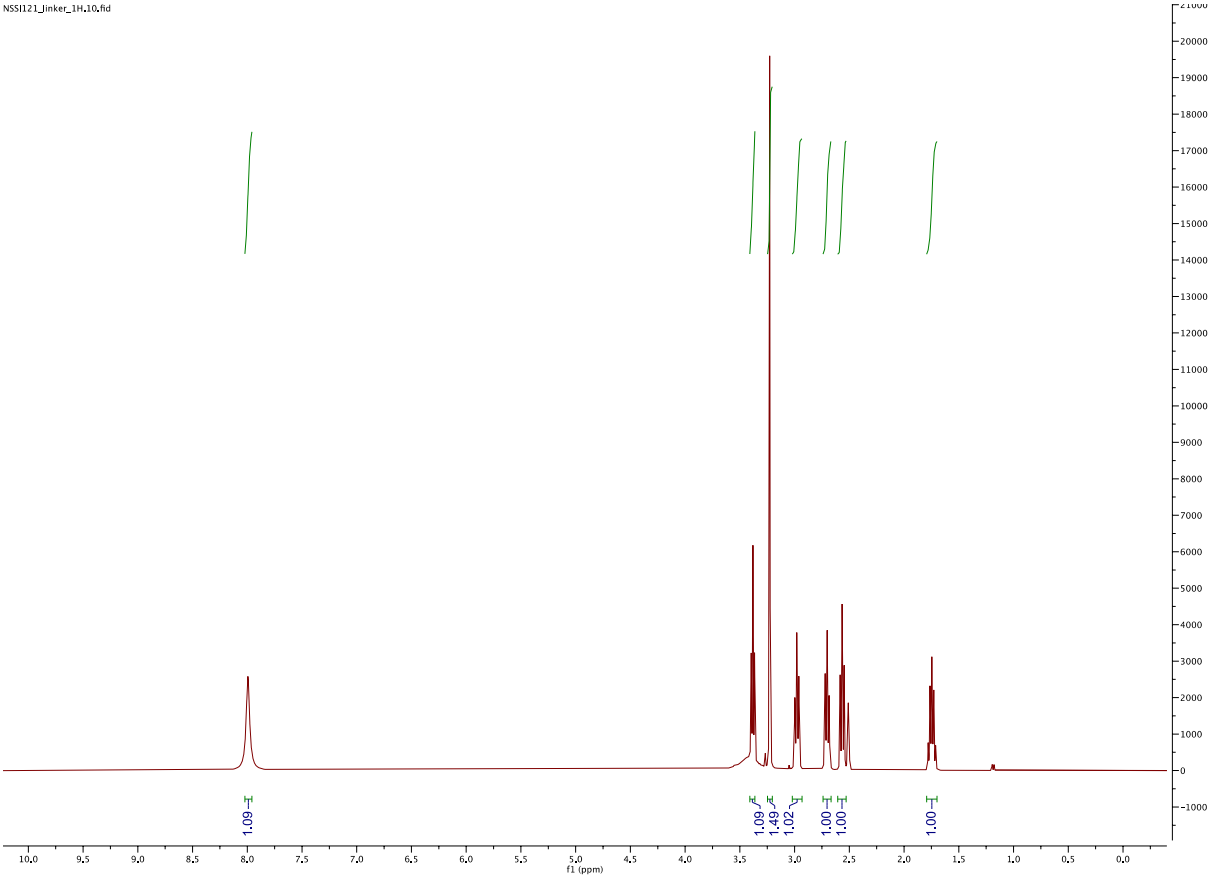
linker control

Synthetic route to control linker

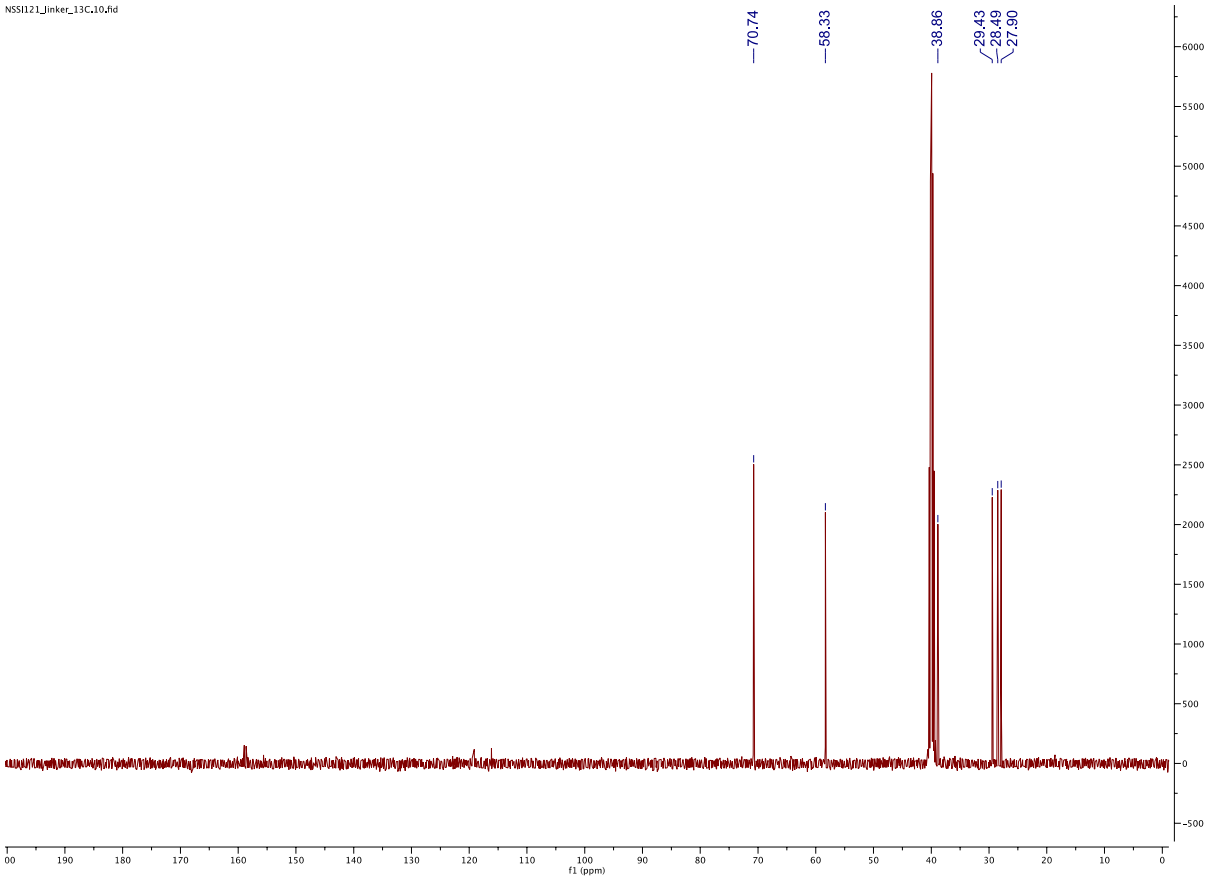
Synthesis of 2-((3-methoxypropyl)thio)ethan-1-amine linker-NH_2

Allyl methyl ether (129 mg, 0.305 mmol, CAS # 627-40-7 Sigma-Aldrich) and cysteamine-HCl (42 mg, 0.370 mmol) were dissolved in deoxygenated H_2O (3 mL) in a quartz cuvette and irradiated at 254 nm for 1 hr. Once the reaction appeared complete by TLC, the crude material was purified by reversed phase HPLC (Waters 1525 binary HPLC mounted with a 00G-4252-PO-AX (Luna Co.) preparative column, using a H_2O/MeCN with 0.1% TFA gradient) to afford final

product "linker-NH_2" (89.4 mg, 34% yield). ^{1}H NMR (400 MHz, DMSO-D_6) δ 7.99 (s, 2H); 3.35-3.41 (t, 2H, OCH_2); 3.23 (s, 3H, CH_3); 2.94-3.02 (t, 2H, CH_2); 2.66-2.74 (t, 2H, CH_2); 2.53-2.60 (t, 2H, CH_2); 1.70-1.79 (m, 2H, CH_2) ppm. ^{13}C-NMR (100 MHz, DMSO-D_6) δ 70.74; 58.33; 38.86; 29.43; 28.49; 27.90 ppm. LRMS (ESI+) m/z: [M+H]+ Calcd for $C_6H_{16}NOS^+$ 150.1; Found 150.1. HRMS (ESI+) m/z: [M+H]+ Calcd for $C_6H_{16}NOS^+$ 150.0947; Found 150.0948.



^{1}H NMR Spectrum (400 MHz, DMSO-*d*$_6$) of linker-NH_2.



^{13}C NMR Spectrum (100 MHz, DMSO-d_6) of linker-NH_2.

Magnetic bead selections

Tosyl-activated Dynabeads (ThermoFisher Scientific, Rockford, IL) are washed in 0.1 M borate buffer pH 9.5. Ligand is added at 2 mM in 0.1 M borate buffer pH 9.5 + 1.2 M ammonium sulfate, then incubated for 18 hours at 37 ºC.[16] Coupling efficiency is measured by taking supernatant after overnight incubation and measuring the A280 absorption, and the concentration of coupled ligand is calculated from a standard curve of toluene sulfonic acid. A control without ligand is carried along to measure background hydrolysis of the tosyl group from the beads, which is then subtracted to get the final concentration of ligand coupled.

Binders were isolated by magnetic bead sorts from the reduced-charge Sso7d libraries rcSso7d-11 and rcSso7d-18 using tosyl-activated Dynabeads in PBSA. Dynabeads loaded with carbohydrate ligand were utilized for positive selections, while beads loaded with linker were utilized for negative selections. Positive and negative selections were iteratively performed until library diversities approached 10^4 variants. Selections use $4x10^5$ beads per 10^7 cells.

Selection of TF-binding Sso7d variants

Sso7d variants capable of binding carbohydrates were isolated using yeast-surface display [51]. Briefly, cells washed in PBSA (PBS pH 7.4 with 0.1% bovine serum albumin) were incubated with chicken anti-HA or chicken anti-cMyc and TF-PAA-FITC, followed by incubation with goat anti-chicken-AF 647. Washed cells were sorted on a FACS Aria IIu (BD Biosciences, San Jose, CA) and this process repeated for at least three rounds of further enrichment.

Sequencing of individual variants from enriched heterogeneous populations

Yeast-display vector pCTCON2 from enriched yeast populations were isolated using Zymoprep Yeast Plasmid Miniprep II (Zymo Research, Irvine, CA) following manufacturer's protocols. XL1-Blue Competent Cells (Agilent Technologies, Santa Clara, CA.) were transformed with isolated plasmids and transformants plated on agar with selection antibiotics. Individual colonies were re-streaked on selective media in 96-well plates, plasmids extracted, and sequencing performed. Sequences were subjected to multiple sequence alignment, followed by hierarchal cluster analysis performed with Geneious Prime 2021.2.

Analytical flow cytometry of heterogeneous or clonal yeast populations

Populations comprising multiple variants or clonal populations were characterized for binding specificity by flow cytometry on an Intellcyt iQue Screener flow cytometer (Essen Bioscience, Ann Arbor, MI.) Sso7d expression was induced on yeast surfaces and washed cells were tested for binding to glycan-PAA-FITC or PAA-FITC at various concentrations in the presence of chicken anti-HA in PBSA. After washing, cells were stained with goat anti-chicken AF647 and analyzed. To characterize binding of individual Sso7d variants, competent EBY100 were transformed with selected plasmids using the Frozen-EZ Transformation II kit (Zymo Research, Irvine, CA.) to produce clonal population of yeast. Variants expressed on yeast surfaces were treated with ligand and analyzed on iQue as above. Flow cytometry data were processed using FCS Express (De Novo Software,Glendale, CA), and binding analyzed using GraphPad Prism (GraphPad Software,San Diego, CA.)

Affinity maturation

To improve the affinities of enriched variants, plasmids from heterogeneous populations of variants or individual variants of interest were subjected to one or more rounds of affinity maturation. Error-prone PCR was initiated by amplifying Sso7d inserts from yeast display vectors in the presence of 2 μM nucleotide analogs (TriLink BioTechnologies, San Diego, CA) 8-oxo-2'-deoxyguanosine-5'-triphosphate (8-oxo-dGTP) and 2'-deoxy-p-nucleoside-5'-triphosphate (dPTP) as described previously.[51] PCR reactions were run using Taq polymerase for a minimum of 15 cycles, followed by purification and extraction on a 1% agarose gel. Mutated inserts were further amplified by PCR to provide sufficient material for electroporation. Separately, pCTCON2 plasmids bearing no Sso7d insert were digested with NheI, BamHI, and SalI, followed by ethanol precipitation with amplified mutagenized inserts. For electroporation, 25 mL freshly cultured

EBY100 yeast at OD_{600} of 2 were made electrocompetent by incubation in 10 mM DTT, 100 mM LiOAc in YPD for 10 minutes at 37 C. Cells were washed twice with ice-cold H_2O, followed by resuspension in 200 µl H_2O containing 5 µg of insert and 1 µg of cut vector precipitated DNA. Yeast were electroporated 50 µl at a time in 2 mm cuvettes using a square-wave function (Gene Pulser Xcell, Bio-Rad, Hercules, CA.) Cells were immediately diluted in 1 mL YPD after electroporation for 30 min, followed by dilution and subculture in SDCAA. Samples of newly-generated libraries were plated on SDCAA agar in order to determine maximum theoretical library diversity.

Cloning of Sso7d variants into expression vector

For soluble recombinant expression in *E. coli*, Sso7d inserts were cloned into pET-24a expression vector (EMD Millipore, Burlington, MA.), which encodes a C-terminal His_6 affinity tag. To produce constructs with a C-terminal SML site, followed by His_6, sequences for Sso7d variants were PCR amplified from the yeast display vector pCTCON2 using a reverse primer encoding the chosen sortase motif LPETGG. Forward and reverse primers contain overhangs for the vector, and Gibson assembly was used to assemble the final product (NEB, Ipswich, MA).

Recombinant expression and affinity purification of Sso7d constructs in *E. coli.*

Sso7d variants of interest were recombinantly expressed as LPETGG-His6 constructs for biophysical characterization, sortase-mediated ligation and mammalian cell staining. BL21-CodonPlus (DE3)-RIL *Escherichia coli* cells (Agilent Technologies, Santa Clara, CA) were transformed with pET24 plasmids containing Sso7d variants of interest. Single transformants were used to inoculate 8 mL overnight cultures of LB containing 30 µg/mL each of kanamycin

and chloramphenicol. Grown overnight cultures were added to 1 L of TB containing 30 μg/mL each of kanamycin and chloramphenicol and grown at 37 °C until an OD_{600} of 0.8-1 reached. Cultures were brought to 16-18 °C and expression induced in the presence of 1 mM IPTG, followed by 16-18 h of growth. Cells were harvested by centrifugation at 3,700 *x g* for 30 min and overexpressed proteins purified immediately or pelleted cells stored long term at -80 °C.

All protein purification steps were carried out at 4 °C. Pelleted cells were resuspended in Buffer A (50 mM HEPES pH 7.5, 500 mM NaCl, 20 mM imidazole) supplemented with 0.5 mg/mL lysozyme (Research Products International, Mount Prospect, IL), 1:1000 protease inhibitor cocktail (EMD Millipore, Burlington, MA), and 1 U/mL DNase I (NEB) and tumbled for 30 min at 4 °C. Homogenized cells were lysed by sonication at 50% amplitude for 3 cycles of 5 min (Vibra-Cell, Sonics, Newtown, CT), followed by clarification by ultracentrifugation at 35,000 *x g* for 60 min. Supernatents containing soluble Sso7d proteins were passed over Ni-NTA resin (HisPur, Thermo Scientific, Rockford, IL) pretreated with Buffer A. Resin was washed with Buffer A for 12 column volumes, followed by elution with high imidazole Buffer B (50 mM HEPES pH 7.5, 500 mM NaCl, 500 mM imidazole) for 6-8 column volumes, with a majority of purified Sso7d eluting in the first 2 column volumes. Column fractions of interest were pooled and immediately desalted using HiTrap desalting column (GE Healthcare, Chicago, IL). Proteins were stored at 4 °C for immediate use or flash-frozen and stored at -80 °C.

Biophysical characterization of Sso7d-LPETG-His6 constructs by bio-layer interferometry

Apparent binding affinities of Sso7d variant 2.4.I to Le^c and core 5 were determined by bio-layer inferometry. All BLI measurements were performed in PBSA supplemented with 0.05% Tween-20 at 22 °C on an Octet RED96 instrument (Pall ForteBio, Fremont, CA.) Sso7d-LPETGG-His_6

constructs were captured on Ni-NTA BLI tips to a final displacement of 1.2 nm, followed by a baseline reading for 60 s. Tips were placed into glycan polymer solutions of various concentrations (1 μM to 31.25 nM) for 600 s to measure association, followed by 600 s in buffer alone for dissociation measurements. If needed, buffer baseline measurements from Sso7d-loaded tips were used for background subtraction and all association and dissociation curves fitted to a 1:1 binding model of one phase exponential association and exponential dissociation using the "associate then dissociate" function of GraphPad Prism (GraphPad Software,San Diego, CA).

The 1:1 binding affinity to TF was measured in the same way using Ni-NTA-tip immobilized 2.4.I dipped into a solution of TF-sp-biotin between 25-75 μM. Bare Ni-NTA tips were dipped into 25-75 μM TF-sp-biotin for baseline subtraction as reported by Wartchow and coworkers.[30] Binding curves were fitted to a 1:1 binding model of one phase exponential association and exponential dissociation using the "associate then dissociate" function of GraphPad Prism (GraphPad Software,San Diego, CA) to obtain the K_D.

Sortase mediated ligation

Sortase-mediated ligation was used to label the evolved GBP 2.4.I with FITC or biotin for use with glycoprotein blotting and cell labeling. Peptides of the sequence GGGYK[K-5-FAM]T-amide and GGGYK[K-PEG2-biotin]T-amide were prepared (21st Century Biochemicals, Inc., Marlboro, MA). His-tagged sortase enzyme (SrtA) (40 μg) was immobilized on 200 μL Ni-NTA resin in sortase buffer (150 mM HEPES pH 7.0, 150 mM NaCl, 10 mM $CaCl_2$). Then, 25 μM of sortase motif bearing protein and 250 μM biotin or FITC peptide were incubated with the resin at room temperature for 30 min. The resin was applied to a small column and the flow through collected,

followed by several washes with sortase buffer. The flow through and washes were pooled and desalted using HiTrap desalting columns to remove excess peptide.

Dot blot of glycoproteins

Human MUC2, MUC5AC, and MUC5B were received from the Ribbeck Lab at MIT. Bovine fetuin was purchased from New England Biolabs and human glycophorin A from Millipore Sigma. A 10-µg aliquot of protein was treated with α-2-3,6,8 neuraminidase or α-2-3,6,8 neuraminidase and O-glycosidase following manufacturers protocols. Proteins were dialyzed overnight before use in dot blot applications. Glycoprotein (1-2 µl) was spotted on a 0.2 µm nitrocellulose membrane and allowed to dry completely before being blocked for 1 hr at RT in PBS + 3% BSA + 0.1% Tween-20. Membranes were incubated in biotinylated 2.4.I or non-sugar-binding Sso7d variant overnight in blocking buffer at 4 ºC. Membrane was washed and incubated with streptavidin-alkaline phosphatase conjugate (Promega, Madison, WI) for 1 hour in PBS before developing with NBT/BCIP one step reagent (ThermoFisher, Rockford, IL).

Flow cytometry staining of mammalian cell line with FITC-labeled proteins

The binding of evolved and functionalized rcSso7d-FITC proteins to mammalian cell-surface glycans was quantified by flow cytometry. For staining of TF antigen, adherent human mammary epithelial cells (MCF7) were maintained in EMEM (ATCC, Manassas, VA) supplemented with 10% v/v FBS and 1% v/v penicillin-streptomycin and grown at 37 °C in a humidified incubator at 5% CO2 to a confluency of 80% before use. Adherent cells were washed once with DPBS and trypsinized with TrypLE Express. Cells were then centrifuged at 300 x g for 3 min, followed by resuspension in ice-cold PBSA at a concentration of 1 x 10^6 cells/mL. From this, 1 x 10^5 cells were

dispensed into individual microcentrifuge tubes and washed with PBS twice, then treated with α-2-3,6,8 neuraminidase and O-glycosidase or Protein Deglycosylation Mix II (NEB) for 1 hour at 37 ºC. Cells were washed twice to remove excess enzyme and resuspended in 50-100 µL PBSA containing 8 µM rcSso7d-FITC. Samples were incubated at 4 °C for 2 hr shaking at 300 rpm. Samples were then washed twice with ice-cold PBSA, resuspended in 150 µL and analyzed on an iQue Flow Cytometer. Flow cytometry data were processed using FCS Express (De Novo Software,Glendale, CA), and binding analyzed using GraphPad Prism (GraphPad Software,San Diego, CA.)

Western blot analysis of MCF7 lysates

1 x 10^5 MCF7 cells were incubated in lysis buffer (50 mM Tris pH 7.5, 150 mM NaCl, 1% Triton X-100) on ice for 1 hr. A 30-µg aliquot of protein was denatured at 100 ºC for 10 minutes and then treated with α-2-3,6,8 neuraminidase and O-glycosidase or Protein Deglycosylation Mix II (NEB) for four hours at 37 ºC. A mock treated sample without enzymes was carried along as a control. Ten µg of each sample was separated by SDS-PAGE and transferred to nitrocellulose then blotted with 2 µM biotinylated protein overnight at 4 ºC and developed with NBT/BCIP one step reagent (ThermoFischer, Rockford, IL).

3.10 Acknowledgements

I would like to thank Dr. Megan Kizer for her work with MCF7 cell cultivation, chapter edits, and experimental planning. I would like to thank Dr. Cristina Zamora for her guidance and mentorship with yeast surface display techniques and assistance with experimental planning. I would also like to thank Dr. Nathaniel Schocker for synthesis of TF-NH_2 for use with magnetic

bead sorts. I would also like to thank the Ribbeck lab for providing mucin glycoproteins MUC2, MUC5AC, and MUC5B. Finally, a huge thanks to the Wittrup lab for sharing the Sso7d yeast surface display library, protocols, and equipment.

3.11 References

1. Gagneux, P., Hennet, T., and Varki, A. (2022) Biological Functions of Glycans, *Essentials of Glycobiology [Internet]. 4th edition*.

2. Fuster, M. M., and Esko, J. D. (2005) The Sweet and Sour of Cancer: Glycans as Novel Therapeutic Targets, *Nat. Rev. Cancer 5*, 526-542.

3. Wu, A. M., Lisowska, E., Duk, M., and Yang, Z. (2008) Lectins as Tools in Glycoconjugate Research, *Glycoconj. J. 26*, 899.

4. Brooks, S. A. (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.

5. Slifkin, M., and Doyle, R. J. (1990) Lectins and Their Application to Clinical Microbiology, *Clin. Microbiol. Rev. 3*, 197-218.

6. Mujahid, A., and Dickert, L. F. (2016) Blood Group Typing: From Classical Strategies to the Application of Synthetic Antibodies Generated by Molecular Imprinting, *Sensors 16*.

7. Kappler, K., and Hennet, T. (2020) Emergence and Significance of Carbohydrate-Specific Antibodies, *Genes Immun. 21*, 224-239.

8. Cummings, R. D., Etzler, M., Hahn, M. G., Darvill, A., Godula, K., Woods, R. J., and Mahal, L. K. (2022) Glycan-Recognizing Probes as Tools, *Essentials of Glycobiology [Internet]. 4th edition*.

9. Ward, E. M., Kizer, M. E., and Imperiali, B. (2021) Strategies and Tactics for the Development of Selective Glycan-Binding Proteins, *ACS Chem. Biol. 16*, 1795-1813.

10. Warkentin, R., and Kwan, D. H. (2021) Resources and Methods for Engineering "Designer" Glycan-Binding Proteins, *Molecules 26*, 380.

11. Adhikari, P., Bachhawat-Sikder, K., Thomas, C. J., Ravishankar, R., Jeyaprakash, A. A., 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.

12. 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.

13. Cicortas Gunnarsson, L., Nordberg Karlsson, E., Albrekt, A. S., Andersson, M., Holst, O., and Ohlin, M. (2004) A Carbohydrate Binding Module as a Diversity-Carrying Scaffold, *Protein Eng. Des. Sel. 17*, 213-221.

14. Lu, Z., Kamat, K., Johnson, B. P., Yin, C. C., Scholler, N., and Abbott, K. L. (2019) Generation of a Fully Human Scfv That Binds Tumor-Specific Glycoforms, *Sci. Rep. 9*, 5101.

15. 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.

16. Traxlmayr, M. W., Kiefer, J. D., Srinivas, R. R., Lobner, E., Tisdale, A. W., Mehta, N. K., Yang, N. J., Tidor, B., and Wittrup, K. D. (2016) Strong Enrichment of Aromatic Residues in Binding Sites from a Charge-Neutralized Hyperthermostable Sso7d Scaffold Library, *J. Biol. Chem. 291*, 22496-22508.

17. Hudson, K. L., Bartlett, G. J., Diehl, R. C., Agirre, J., Gallagher, T., Kiessling, L. L., and Woolfson, D. N. (2015) Carbohydrate–Aromatic Interactions in Proteins, *J. Am. Chem. Soc. 137*, 15152-15160.

18. Kiessling, L. L., and Diehl, R. C. (2021) Ch−Π Interactions in Glycan Recognition, *ACS Chem. Biol. 16*, 1884-1893.

19. Asensio, J. L., Ardá, A., Cañada, F. J., and Jiménez-Barbero, J. (2013) Carbohydrate–Aromatic Interactions, *Acc. Chem. Res. 46*, 946-954.

20. Yu, L.-G. (2007) The Oncofetal Thomsen–Friedenreich Carbohydrate Antigen in Cancer Progression, *Glycoconj. J. 24*, 411-420.

21. Feng, D., Shaikh, A. S., and Wang, F. (2016) Recent Advance in Tumor-Associated Carbohydrate Antigens (Tacas)-Based Antitumor Vaccines, *ACS Chem. Biol. 11*, 850-863.

22. Hong, X., Ma, M. Z., Gildersleeve, J. C., Chowdhury, S., Barchi, J. J., Mariuzza, R. A., Murphy, M. B., 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.

23. Ravn, P., Danielczyk, A., Bak Jensen, K., Kristensen, P., Astrup Christensen, P., Larsen, M., Karsten, U., and Goletz, S. (2004) Multivalent Scfv Display of Phagemid Repertoires for the Selection of Carbohydrate-Specific Antibodies and Its Application to the Thomsen–Friedenreich Antigen, *J. Mol. Biol. 343*, 985-996.

24. Sharma, V., Vijayan, M., and Surolia, A. (1996) Imparting Exquisite Specificity to Peanut Agglutinin for the Tumor-Associated Thomsen-Friedenreich Antigen by Redesign of Its Combining Site *J. Biol. Chem. 271*, 21209-21213.

25. Lee, R. T., and Lee, Y. C. (2000) Affinity Enhancement by Multivalent Lectin–Carbohydrate Interaction, *Glycoconj. J. 17*, 543-551.

26. Monsigny, M., Mayer, R., Roche, A.C. (1991) Sugar-Lectin Interactions: Sugar Cluster, Lectin Multivalency and Avidity, *Carbohydr. Lett. 4*, 35-52.

27. Kiessling, L. L., Young, T., Gruber, T. D., and Mortell, K. H. (2008) Multivalency in Protein–Carbohydrate Recognition, In *Glycoscience: Chemistry and Chemical Biology* (Fraser-Reid, B. O., Tatsuta, K., and Thiem, J., Eds.), pp 2483-2523, Springer Berlin Heidelberg, Berlin, Heidelberg.

28. Lundquist, J. J., and Toone, E. J. (2002) The Cluster Glycoside Effect, *Chem. Rev. 102*, 555-578.

29. Ryckaert, S., Callewaert, N., Jacobs, P. P., Dewaele, S., Dewerte, I., and Contreras, R. (2007) Fishing for Lectins from Diverse Sequence Libraries by Yeast Surface Display – an Exploratory Study, *Glycobiology 18*, 137-144.

30. Wartchow, C. A., Podlaski, F., Li, S., Rowan, K., Zhang, X., Mark, D., and Huang, K.-S. (2011) Biosensor-Based Small Molecule Fragment Screening with Biolayer Interferometry, *J. Comput. Aided Mol. Des. 25*, 669.

31. Mao, H., Hart, S. A., Schink, A., and Pollok, B. A. (2004) Sortase-Mediated Protein Ligation: A New Method for Protein Engineering, *J. Am. Chem. Soc. 126*, 2670-2671.

32. Antos, J. M., Ingram, J., Fang, T., Pishesha, N., Truttmann, M. C., and Ploegh, H. L. (2017) Site-Specific Protein Labeling Via Sortase-Mediated Transpeptidation, *Curr. Protoc. Prot. Sci. 89*, 15.13.11-15.13.19.

33. Dolan, B., and Hansson, G. C. (2021) Mucins, In *Reference Module in Life Sciences*, Elsevier.

34. Johansson, M. E., Phillipson, M., Petersson, J., Velcich, A., Holm, L., and Hansson, G. C. (2008) The Inner of the Two Muc2 Mucin-Dependent Mucus Layers in Colon Is Devoid of Bacteria, *Proc. Natl. Acad. Sci. 105*, 15064-15069.

35. Geiger, P., Mayer, B., Wiest, I., Schulze, S., Jeschke, U., and Weissenbacher, T. (2016) Binding of Galectin-1 to Breast Cancer Cells Mcf7 Induces Apoptosis and Inhibition of Proliferation in Vitro in a 2d- and 3d- Cell Culture Model, *BMC Cancer 16*, 870.

36. van Zundert, G. C. P., Rodrigues, J. P. G. L. M., Trellet, M., Schmitz, C., Kastritis, P. L., Karaca, E., Melquiond, A. S. J., van Dijk, M., de Vries, S. J., and Bonvin, A. M. J. J. (2016) The Haddock2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes, *J. Mol. Biol. 428*, 720-725.

37. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., and Hassabis, D. (2021) Highly Accurate Protein Structure Prediction with Alphafold, *Nature 596*, 583-589.

38. Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022) Colabfold: Making Protein Folding Accessible to All, *Nat. Methods 19*, 679-682.

39. Bochkov, A. Y., and Toukach, P. V. (2021) Csdb/Snfg Structure Editor: An Online Glycan Builder with 2d and 3d Structure Visualization, *J. Chem. Inf. Model. 61*, 4940-4948.

40. Weis, W. I., and Drickamer, K. (1996) Structural Basis of Lectin-Carbohydrate Recognition, *Annu. Rev. Biochem. 65*, 441-473.

41. Bojar, D., Meche, L., Meng, G., Eng, W., Smith, D. F., Cummings, R. D., and Mahal, L. K. (2022) A Useful Guide to Lectin Binding: Machine-Learning Directed Annotation of 57 Unique Lectin Specificities, *ACS Chem. Biol.*

42. Hirabayashi, J., Tateno, H., Shikanai, T., Aoki-Kinoshita, K. F., and Narimatsu, H. (2015) The Lectin Frontier Database (Lfdb), and Data Generation Based on Frontal Affinity Chromatography, *Molecules 20*.

43. Boder Eric, T., Midelfort Katarina, S., and Wittrup, K. D. (2000) Directed Evolution of Antibody Fragments with Monovalent Femtomolar Antigen-Binding Affinity, *Proc. Natl. Acad. Sci. 97*, 10701-10705.

44. Hackel, B. J., Kapila, A., and Dane Wittrup, K. (2008) Picomolar Affinity Fibronectin Domains Engineered Utilizing Loop Length Diversity, Recursive Mutagenesis, and Loop Shuffling, *J. Mol. Biol. 381*, 1238-1252.

45. Wellner, A., McMahon, C., Gilman, M. S. A., Clements, J. R., Clark, S., Nguyen, K. M., Ho, M. H., Hu, V. J., Shin, J.-E., Feldman, J., Hauser, B. M., Caradonna, T. M., Wingler, L. M., Schmidt, A. G., Marks, D. S., Abraham, J., Kruse, A. C., and Liu, C. C. (2021) Rapid Generation of Potent Antibodies by Autonomous Hypermutation in Yeast, *Nat. Chem. Biol. 17*, 1057-1064.

46. Yabe, R., Itakura, Y., Nakamura-Tsuruta, S., Iwaki, J., Kuno, A., and Hirabayashi, J. (2009) Engineering a Versatile Tandem Repeat-Type A2-6sialic Acid-Binding Lectin, *Biochem. Biophys. Res. Commun. 384*, 204-209.

47. Hamorsky, K. T., Kouokam, J. C., Dent, M. W., Grooms, T. N., Husk, A. S., Hume, S. D., Rogers, K. A., Villinger, F., Morris, M. K., Hanson, C. V., and Matoba, N. (2019) Engineering of a Lectibody Targeting High-Mannose-Type Glycans of the Hiv Envelope, *Mol. Ther. 27*, 2038-2052.

48. Holmén Larsson, J. M., Karlsson, H., Sjövall, H., and Hansson, G. C. (2009) A Complex, but Uniform O-Glycosylation of the Human Muc2 Mucin from Colonic Biopsies Analyzed by Nanolc/Msn, *Glycobiology 19*, 756-766.

49. Jin, C., Kenny, D. T., Skoog, E. C., Padra, M., Adamczyk, B., Vitizeva, V., Thorell, A., Venkatakrishnan, V., Lindén, S. K., and Karlsson, N. G. (2017) Structural Diversity of Human Gastric Mucin Glycans *Mol. Cell. Proteomics 16*, 743-758.

50. Thomsson, K. A., Prakobphol, A., Leffler, H., Reddy, M. S., Levine, M. J., Fisher, S. J., and Hansson, G. C. (2002) The Salivary Mucin Mg1 (Muc5b) Carries a Repertoire of Unique Oligosaccharides That Is Large and Diverse, *Glycobiology 12*, 1-14.

51. Chao, G., Lau, W. L., Hackel, B. J., Sazinsky, S. L., Lippow, S. M., and Wittrup, K. D. (2006) Isolating and Engineering Human Antibodies Using Yeast Surface Display, *Nat. Protoc. 1*, 755-768.

52. Feng, F., Okuyama, K., Niikura, K., Ohta, T., Sadamoto, R., Monde, K., Noguchi, T., and Nishimura, S.-I. (2004) Chemo-Enzymatic Synthesis of Fluorinated 2-N-Acetamidosugar Nucleotides Using Udp-Glcnac Pyrophosphorylase, *Org. Biomol. Chem. 2*, 1617-1623.

53. Baek, M. G., and Roy, R. (2002) Synthesis and Protein Binding Properties of T-Antigen Containing Glycopamam Dendrimers, *Bioorg. Med. Chem. 10*, 11-17.

54. Roy, R., Baek, M.-G., and Rittenhouse-Olson, K. (2001) Synthesis of N,N‘-Bis(Acrylamido)Acetic Acid-Based T-Antigen Glycodendrimers and Their Mouse Monoclonal Igg Antibody Binding Properties, *J. Am. Chem. Soc. 123*, 1809-1816.

Chapter 4: Design and characterization of trimeric glycan-binding proteins

4.1 Abstract

Glycan-binding proteins (GBPs) make low affinity interactions with their carbohydrate ligands and often oligomerize in their native context to form high-avidity interactions which boosts apparent affinity. The GBPs evolved in Chapter 2 and 3 have modest functional affinities in the low to mid nM range when tested with multivalent carbohydrate antigens, but the proteins are monomeric and can suffer from poor affinity in certain contexts. Fusion of the evolved GBPs to the coiled-coil trimerization domain of the lectin surfactant protein D (SP-D) leads to the formation of a trimeric GBP. These trimers are properly folded, stable, and have increased binding affinity compared to monomeric GBPs. Generation of trimeric Sso7d-based GBPs is a strategy for increasing the functional affinity of the evolved proteins, thereby making the proteins useful for a wider range of applications.

4.2 Introduction

Glycan-binding proteins (GBPs) are biologically important molecules that perform a myriad of functions, including trafficking of glycoproteins, adhesion of cells to other cells, and immune responses to invading pathogens.[1] GBPs are also useful tools for qualitative glycan structure analysis and are often applied as affinity reagents in the lab or clinic for glycan/glycoconjugate purification, blood typing, and histology.[2, 3]

Carbohydrate-protein interactions are typically low affinity, with K_D values in the μM to mM range for a monovalent interaction.[4] To compensate for this, many naturally occurring GBPs form higher-order oligomers to interact with multivalent glycan ligands, manifested as repeating units in one polysaccharide or as multiple proximal carbohydrates such as the densely glycosylated cell surface. This allows for the formation of interactions that increase the functional affinity, or avidity of the interaction, which can be orders of magnitude stronger than a monovalent interaction.[5, 6] Lectins exist in many oligomerization states from monomers (certain galectins, selectins, siglecs), dimers (certain galectins), trimers (collectins), tetramers (ConA), to pentamers (pentraxin).[5] Certain lectins can further multimerize to create even higher order structures, such as bouquets of 3-5 trimers, cruciform structures made up of four trimers, and "fuzzy balls" made up of six or more trimers.[7, 8] In some instances, a lectin will not oligomerize but contain multiple carbohydrate recognition domains (CRDs) in one polypeptide such as the mannose receptor with a total of nine CRDs, or β-propeller lectins that adopt a donut shape made up of multiple carbohydrate-binding repeats.[9, 10]

It is well understood that oligomeric lectins gain affinity toward their targets compared to their respective monomeric form. Therefore, many glycan-binding protein engineering strategies have attempted to develop oligomers from monovalent lectins.[11] One strategy is the introduction

of tandem repeats. This strategy has been applied to an engineered sialic acid-binding lectin that increased the affinity tenfold compared to the monomeric form, and to a fucose-binding lectin to improve the affinity twelve-fold.[12, 13] Another strategy is fusion of the lectin to an oligomerization domain. One fusion partner used in previous GBP engineering efforts is the fraction crystallizable (Fc) domain of immunoglobulins to form a dimeric protein, called a "lectibody", that increased the binding affinity ten-fold for its target glycan.[14] GBPs developed using lamprey variable lymphocyte receptors, or "lambodies", were fused to either an Fc domain or to a leucine zipper dimerization domain to produce dimeric proteins with increased affinity for their target.[15] Other groups have developed *de novo* designed four-helix bundle proteins that can assemble into species from dimers to dodecamers, and when fused to the *Agrocybe cylindracea* galectin improved the affinity 10-100 fold for 3'-sialyllactose depending on the oligomeric state.[16]

One lectin that naturally trimerizes is surfactant protein D (SP-D). SP-D is part of the innate immune system and is highly expressed in pulmonary epithelia to defend against inhaled microorganisms.[17] SP-D consists of a short N-terminal domain, a collagen-like triple-helical domain containing 59 GXY repeats, a short α-helical coiled-coil neck, and a C-terminal CRD (**Figure 4-1A**). SP-D trimer formation is mediated through the coiled-coil neck domain (**Figure 4-1B**).[18] Coiled-coils are one of the principal subunit oligomerization motifs of proteins consisting of two to five amphipathic α-helices that twist around each other like a rope. The coiled-coil of SP-D is a 35-residue sequence made up of a heptad repeat with nonpolar residues in the first and fourth positions that form the hydrophobic core of the helical packing (**Figure 4-1C**).[19] The SP-D neck domain is a promising oligomerization domain for GBP engineering as it has been shown to induce the formation of stable trimers *in vitro* when fused to the heterologous protein thrioredoxin, a monovalent non-lectin protein.[20]

In the previous chapters a platform for creation of GBPs from the small, stable DNA-binding protein Sso7d was described. Yeast-surface display-based directed evolution was performed and variants were isolated that can recognize sialic acid and the TF antigen, important components of mammalian glycans. The evolved GBPs have nM functional affinities in multivalent systems with true K_D values measured to be low μM for one TF-binding variant. The evolved GBPs have been demonstrated to be functional as reagents for cell staining and blotting, but improved binding affinities would be beneficial for future applications. One evolved TF-binding variant, 1.3.D, had affinities too low to be useful and was chosen for further affinity engineering by fusion to a trimerization domain.

The following chapter details the design, purification, and analysis of Sso7d variant 1.3.D fused to the coiled-coil neck domain of human SP-D. The fusion proteins are readily produced in the cytoplasm of *E. coli* and the C-terminal fusion construct was shown to form stable trimers in solution with no negative impact to the protein stability. Addition of the coiled-coil domain increased the functional affinity of the protein, showing binding to a multivalent polymer that was not seen with the monomer 1.3.D. Fusion of evolved Sso7d-based GBPs to an oligomerization domain is a strategy to improve the functionality of the protein when the affinity of the monovalent protein is too low to be useful.

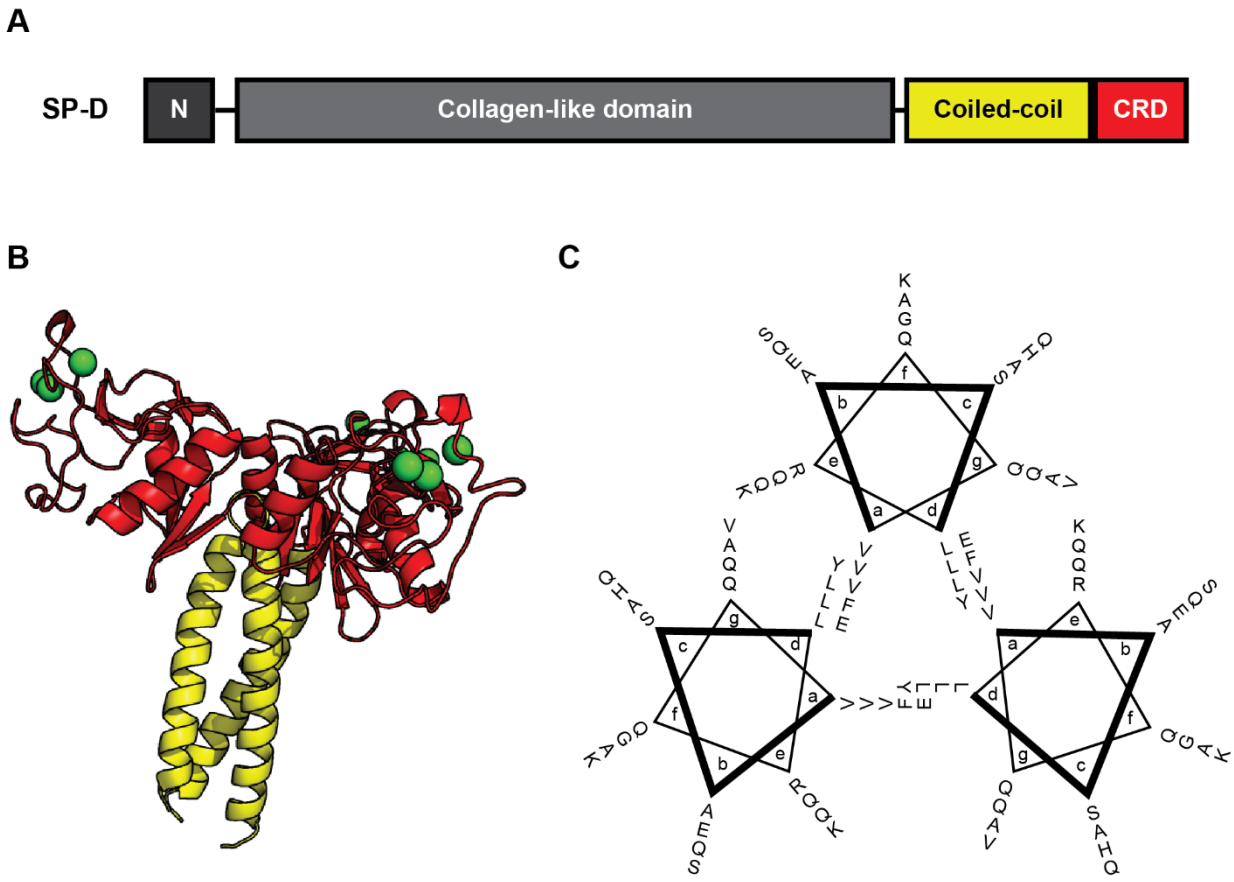


Figure 4-1. Coiled-coil neck domain of SP-D forms stable trimers.
A) The organization of human SP-D showing the N-terminal domain, the collagen-like domain, the coiled-coil neck domain, and the C-terminal carbohydrate recognition domain (CRD). B) Crystal structure of human SP-D neck and CRD (PDB 1B08). C) Helical wheel diagram of the coiled-coil domain of SP-D showing the hydrophobic core.

4.3 Design and purification of SP-D fusion constructs

The neck region (abbreviated as N) of human SP-D consists of the 32 residues from Val224 through Pro255. Sso7d variant 1.3.D was selected for creation of the trimeric fusion proteins. Both the N- and C-terminus of Sso7d is solvent accessible and fusion to either would not occlude the binding site. Fusions to both the N-terminus (N1.3.D) and C-terminus (1.3.DN) were designed to test which functioned better (**Figure 4-2A**). A flexible $(GGGS)_2$ linker was placed between the neck region and 1.3.D to allow free movement of the 1.3.D module. The C-terminus of each construct contains a sortase recognition motif (LPETGG) and a hexahistidine tag allowing for

respective labeling and purification of the protein. The fusion constructs were designed as gene blocks with codons optimized for *E. coli* expression and cloned into a bacterial vector. Full length SP-D typically requires eukaryotic expression systems due to issues with producing the collagen-like domains in *E. coli*, but the neck-CRD fragment has been successfully produced in *E. coli* previously.[21]

Protein overexpression was induced by addition of IPTG to log phase cells and was successful, with both constructs expressing in high yield. Protein was purified from the cleared lysate using Ni-NTA resin (**Figure 4-2B,D**). It has been observed that sugar-binding Sso7d variants have some affinity to agarose-based resins, which is not surprising as agarose is a polysaccharide of β-D-galactose and 3,6-anhydro-L-galactose. Like other sugar-binding Sso7d variants, the trimeric proteins had affinity to the agarose-based resins required for Ni-NTA purification and required high concentrations of salt and imidazole for elution. The proteins were rapidly desalted to remove the excess salt and imidazole, yielding reasonably pure protein in milligram quantities for further analysis (**Figure 4-2C,E**). The proteins run at ~13 kDa which aligns well with the expected molecular weights of 13.0 kDa for N1.3.D and 13.5 kDa for 1.3.DN.

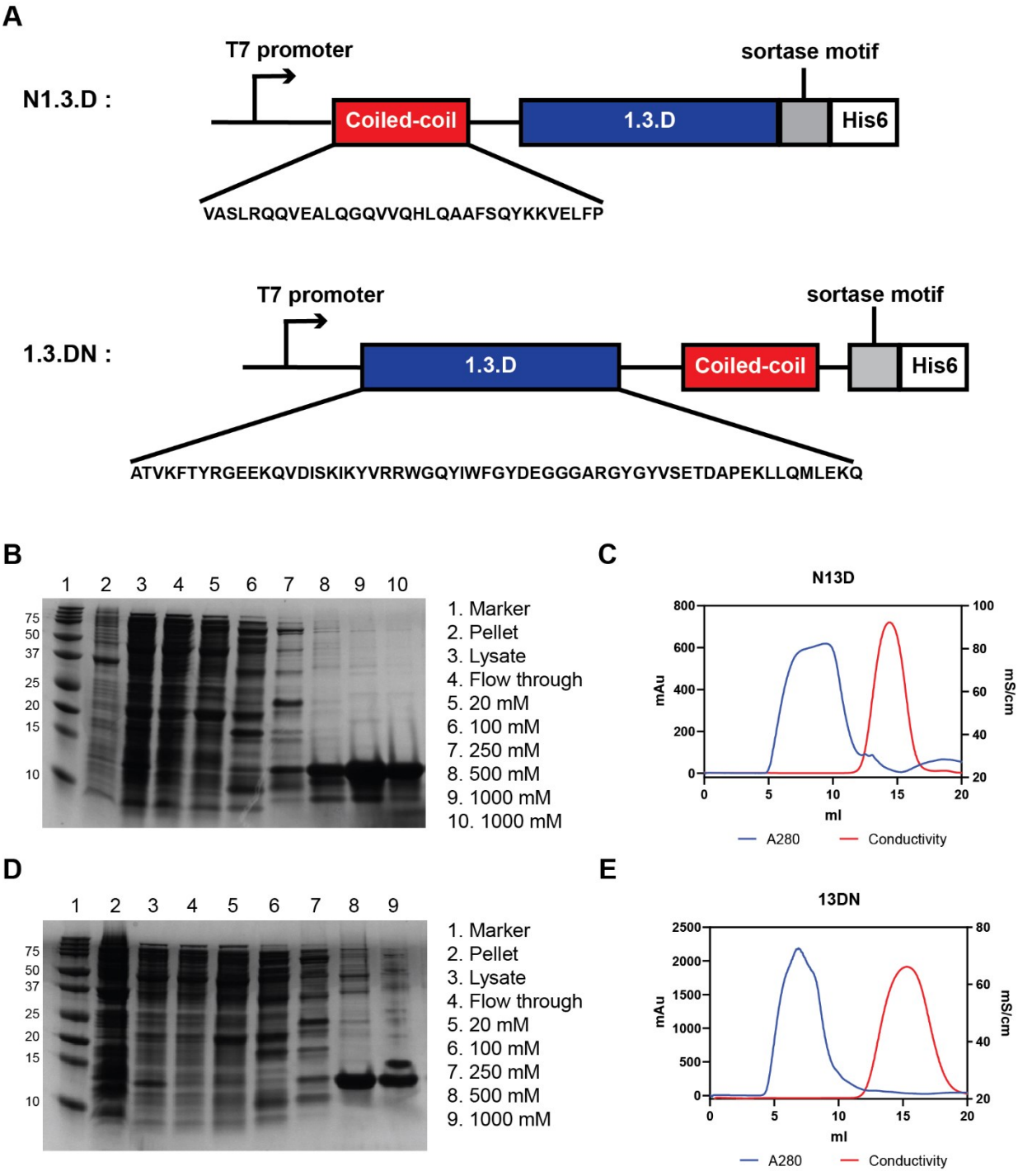


Figure 4-2. Design and purification of SP-D neck 1.3.D fusions.
A) Schematic representations of coiled-coil fusion proteins N1.3.D and 1.3.DN. B) SDS-PAGE gel stained with Coomassie blue showing steps in the purification of N1.3.D. Steps are labeled with concentrations of imidazole used in wash and elution steps. C) Desalting of purified N1.3.D, blue shows A280 trace, red is conductivity trace. D) SDS-PAGE gel showing steps in purification of 1.3.DN. E) Desalting of purified 1.3.DN, blue is A280, red is conductivity.

4.4 Fusion to coiled-coil domain produces folded and stable proteins

4.4.1 Application of far-UV circular dichroism spectroscopy to determine protein secondary structure

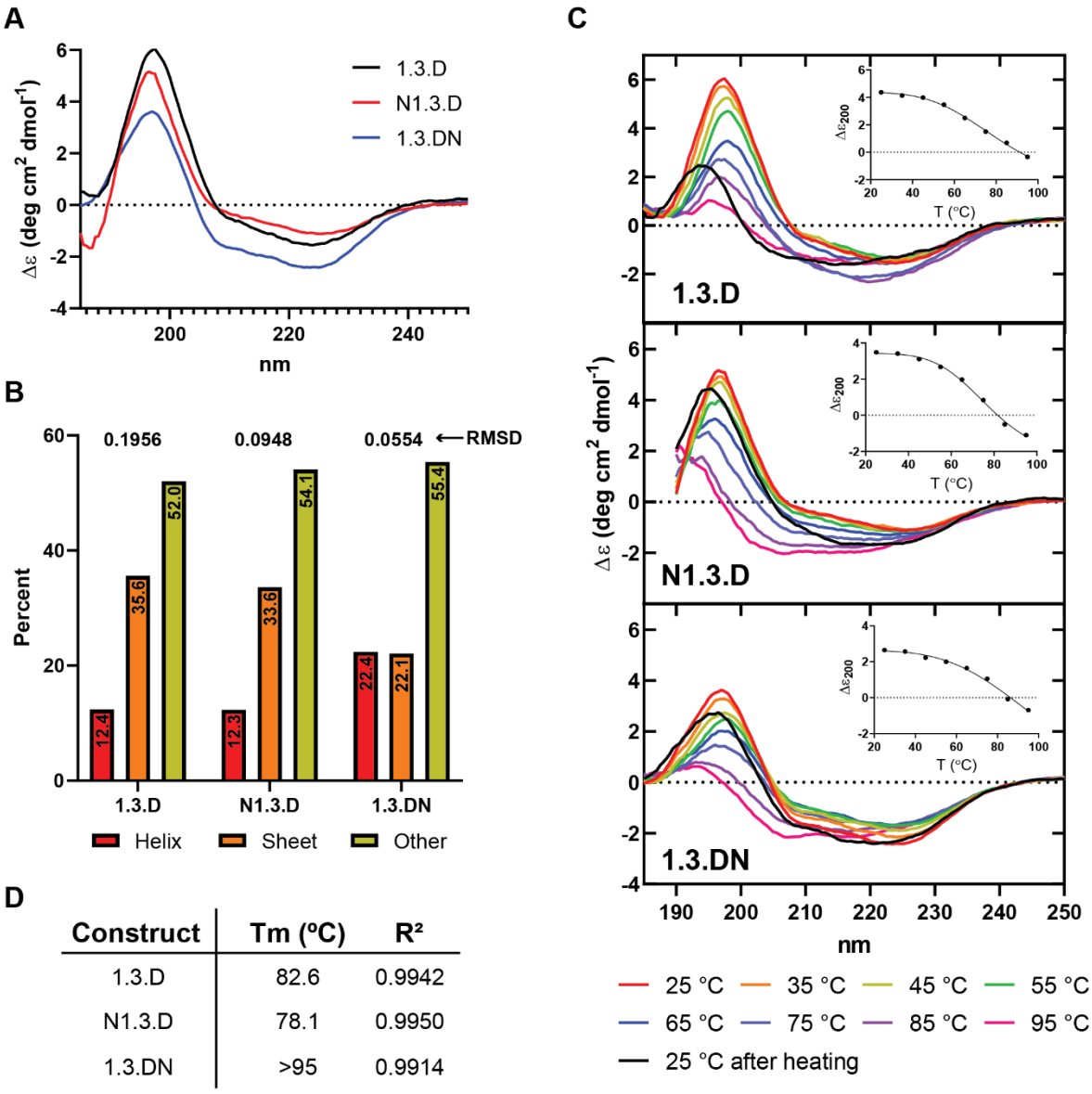
Circular dichroism (CD) spectroscopy was performed to analyze the secondary structure of 1.3.D, N1.3.D, and 1.3.DN (**Figure 4-3A**). Far-UV CD of 1.3.D showed a shallow negative peak around 222 nm, a cross over on the x-axis at 207 nm, and a large positive peak at 196 nm. Far-UV CD of WT Sso7d has been performed and showed a negative peak at 218 nm.[22] However for Sso7d variants evolved to bind other targets (e.g. fluorescein, cIgY, lysozyme, mIgG, streptavidin, and β-catenin peptide) the far-UV CD showed a shift of the negative peak toward 222 nm, like that of 1.3.D.[23] Addition of the coiled-coil helical domain would increase the expected percent helicity of the protein. 1.3.DN showed strong negative peaks at 222 nm and 208 nm and crossed over the x-axis at 203 nm, all indicative of α-helices.[24] The $\theta_{222}/\theta_{208}$ ratio can be used to gauge the α-helicity of the protein: a $\theta_{222}/\theta_{208} \geq 1$ is indicative of coiled-coils while a $\theta_{222}/\theta_{208} \leq 0.86$ is indicative of isolated helices.[25] 1.3.DN had a $\theta_{222}/\theta_{208}$ ratio of 1.85, suggesting the protein forms a coiled-coil in solution. N1.3.D did not have as pronounced negative peaks at 222 and 208 nm, indicating less α-helical character than 1.3.DN. It also crossed over the x-axis at 207 nm as seen for 1.3.D.

The BeStSel (Beta Structure Selection) web server was used to determine the secondary structure based on the acquired spectra (**Figure 4-3B**).[26, 27] This web server was selected because it has improved structure prediction accuracy for β-sheet rich proteins like Sso7d, and can estimate α-helix content more accurately than other methods. Fitting of 1.3.D had a root-mean-square deviation (RMSD) of 0.1595 and was determined to have 12.4% α-helix and 35.6% β-sheet. This is in agreement with prior CD analysis of Sso7d and with the NMR structure of Sso7d.[22, 28] 1.3.DN

fitting had an RMSD of 0.0554 and has increased helicity as expected at 22.4% α-helix and 22.3% β-sheet. N1.3.D fitting had an RMSD of 0.0948 and was determined to have a nearly identical secondary structure as 1.3.D with 12.3% α-helix and 33.6% β-sheet. The BeStSel web server was also able to correctly identify the β-sheets as being anti-parallel. All together, these data indicate that all the proteins are folded, but that 1.3.DN contains more α-helical character and is consistent with the presence of a coiled-coil domain.

4.4.2 Fusion proteins maintain high thermal stability

The thermal stability of the proteins was assessed by CD. A solution of each protein was measured from 185 to 250 nm at ten-degree intervals from 25 ºC to 95 ºC (**Figure 4-3C**). As temperature increases, the secondary structure will transition to an unfolded state. The delta molar ellipticity (Δε) at 200 nm changes significantly upon unfolding for all three proteins, so this was plotted against temperature and fit with a sigmoidal binding function to determine the melting temperature (Tm) of the protein (**Figure 4-3C,D**). The monomeric 1.3.D had a Tm of 82 ºC, which is quite high and close to the Tm of rcSso7d of 95 ºC. The N1.3.D construct has a slightly reduced, yet still high Tm of 78 ºC. However, for 1.3.DN there was a stabilizing effect that increased the Tm to >95 ºC. Protein stabilization after coiled-coil fusion was seen previously for the thioredoxin-SP-D neck fusion construct which increased the Tm by 13 ºC.[20] After thermal unfolding the solutions were cooled back to 25 ºC and scanned to assess protein refolding. 1.3.DN but not 1.3.D or N1.3.D showed partial refolding as shown by the black vs red traces in **Figure 4-3C**. For 1.3.DN, the negative peaks at 222 and 208 nm were identical before and after heating, but the positive peak at 197 nm was slightly diminished.



Construct	Tm (ºC)	R^2
1.3.D	82.6	0.9942
N1.3.D	78.1	0.9950
1.3.DN	>95	0.9914

Figure 4-3. Circular dichroism spectroscopy.
A) Comparison of far-UV CD spectra of 1.3.D (black), N1.3.D (red), 1.3.DN (blue) from 185-250 nm. 5 μM protein in 2 mm cuvette, 10 mM sodium phosphate pH 6.7. B) Percent secondary structure as determined using BeStSel web server, showing α-helix (red), β-sheet (orange), and other structure (yellow). The RMSD for BeStSel fitting of each construct is shown above the columns. C) Far-UV CD spectra of 1.3.D, N1.3.D, and 1.3.DN from 25 ºC to 95 ºC. Black trace is protein cooled to 25 ºC post heating. Inset shows Δε at 200 nm at each temperature fit with sigmoidal binding curve. D) Table showing calculated Tm for each construct after fitting of $\Delta\varepsilon_{200}$ to a sigmoidal binding curve. R^2 of fitting is shown.

4.5 Fusion of 1.3.D to coiled-coil domain induces trimerization

4.5.1 Crosslinking

Protein crosslinking was used to assess the oligomeric state of the fusion proteins. The coiled-coil neck contains two consecutive lysine residues toward the C-terminal end of the domain. According to the crystal structure of SP-D neck and CRD, the lysine residues are solution accessible and 13-15.4 Å apart (**Figure 4-4A**).[29] The crosslinker BSOCOES (bis[2-(succinimidyloxycarbonyloxy)ethyl]sulfone) is a homobifunctional *N*-hydroxysuccimide ester that forms covalent bonds with the primary amine groups of lysine residues (**Figure 4-4B**). It has a spacer arm of 13 Å that can bridge the lysine residues of the coiled-coil. Previous crosslinking of the coiled-coil domain of SP-D was performed successfully with a crosslinker with an 11.4 Å spacer, indicating the lysine residues should crosslink with a 13 Å spacer arm.[20, 29]

Crosslinking was performed at a pH of 8.0 with increasing concentrations of BSOCOES, with the lowest concentration at ten-fold molar excess over the protein. The monomer 1.3.D shows no crosslinking at any concentration as expected (**Figure 4-4AC,D**). N1.3.D shows presence of a dimer at all concentrations tested, even without crosslinker present, and shows a faint trimer signal in the highest concentration sample (**Figure 4-4C**). 1.3.DN shows concentration-dependent formation of dimer starting at 0.1 mM and trimer as low as 0.5 mM (**Figure 4-4D**). From this data it is clear that 1.3.DN forms a trimer under the conditions used for crosslinking (e.g. 10 μM protein in PBS). The N-terminal fusion N1.3.D does not crosslink as efficiently, but the presence of a dimer even under the denaturing conditions of SDS-PAGE in the 0 mM crosslinker sample suggests that the protein forms a highly stable dimer under crosslinking conditions.

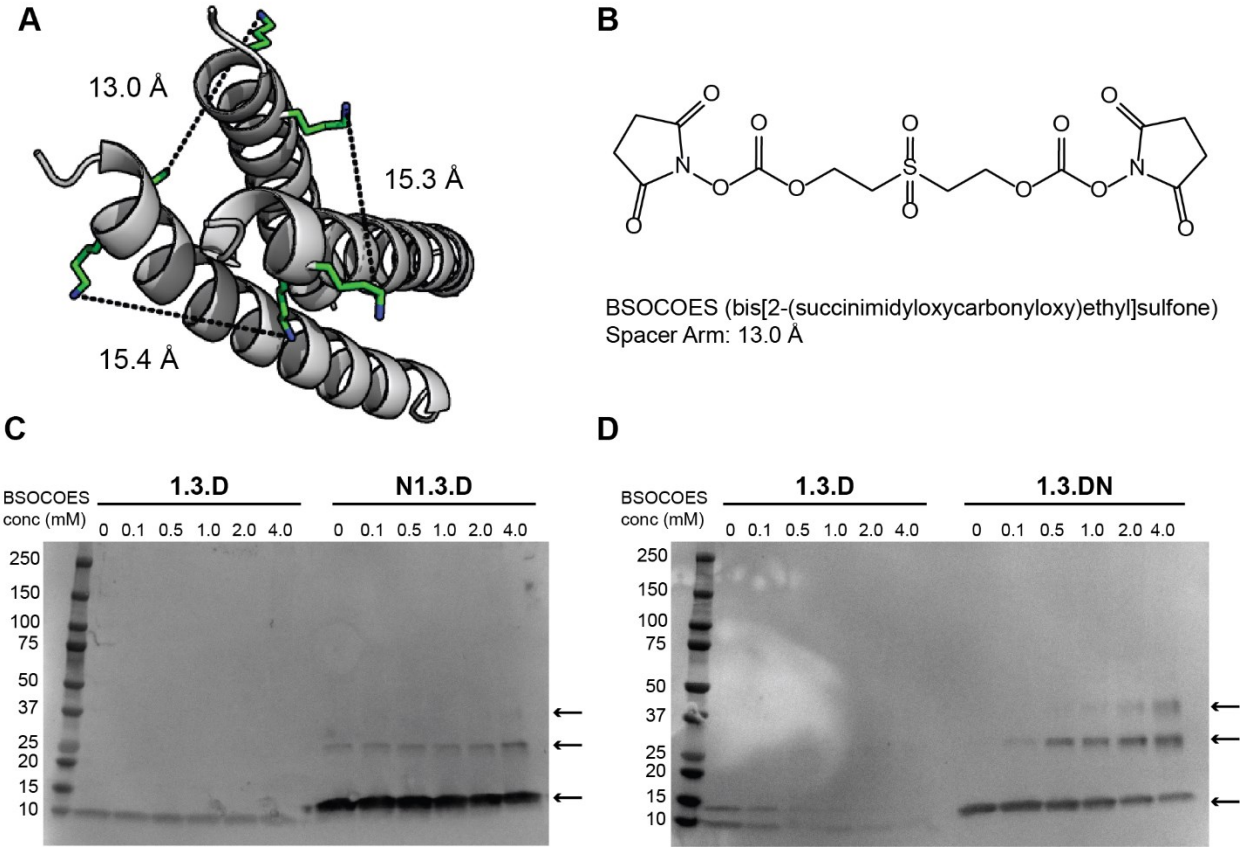


Figure 4-4. Crosslinking of coiled-coil fusions.
A) Coiled-coil region of human SP-D with lysine residues shown in green and distances between lysine residues labeled, PDB 1B08. B) Structure of heterobifunctional crosslinker BSOCOES. C-D) Crosslinking of 1.3.D, N1.3.D, and 1.3.DN with 0, 0.1, 0.5, 1.0, 2.0, and 4.0 mM of BSOCOES lysine-lysine crosslinker. Arrows indicate monomer, dimer, and trimer species.

4.5.2 Matrix-Assisted Laser Desorption/Ionization (MALDI) Mass Spectroscopy

Samples of 1.3.D, N1.3.D and 1.3.DN were taken for analysis with MALDI (**Figure 4-5B**). 1.3.D showed a single peak at the expected molecular weight of 8.9 kDa. N1.3.D only had signal for the monomer at the expected molecular weight of 13.0 kDa. However, the 1.3.DN sample shows a large monomer peak at 13.2 kDa, a smaller dimer peak at 26.5 kDa, and a smaller trimer peak at 39.7 kDa. This correlates with the calculated monomer, dimer, and trimer masses of 13.5, 27.0, and 40.5 kDa respectively. A sample of 1.3.DN crosslinked with 1 mM BSOCOES (**Figure 4-5A**) analyzed by MALDI exhibits three peaks at similar molecular weights to native 1.3.DN.

The crosslinked dimer and trimer peaks appear at roughly the same positions as the non-crosslinked peaks, though the crosslinked peaks are quite broad and cannot be assigned definitive molecular weights. This observed peak broadening is a result of multiple lysine residues in the protein. There are a total of seven lysine residues in the fusion proteins: two in the coiled-coil neck and five in 1.3.D. BSOCOES may singly label these lysine residues in the protein, leading to additional mass on the protein which will broaden the MALDI peaks.

4.5.3 Size exclusion chromatography

To assess the oligomeric state in solution, size exclusion chromatography (SEC) was performed. The first attempt used a HiLoad 16/600 Superdex 200 column (**Figure 4-5C**). The bed volume of the column is 120 ml, and 1.3.DN did not elute until 120 ml, giving a calculated size of about 1 kDa based on a calibration curve of known standards. Superdex is a carbohydrate polymer composed of an agarose-dextran composite matrix and sugar-binding Sso7d variants have shown affinity to agarose-based resins during purifications before. This affinity slowed the elution of the protein on the column preventing an accurate size to be calculated using this method.

To get a more accurate sizing of the proteins, an ENrich SEC 650 column was used. The bead chemistry of this column is hydrophilic polymethacrylate and not carbohydrate based, so protein is less likely to interact with the resin and slow the elution. The monomer 1.3.D and 1.3.DN were run on the column (**Figure 4-5D**). 1.3.D eluted at 16.6 ml and is 6.5 kDa based on a calibration curve of known standards. This is close to the expected molecular weight of 8.5 kDa. 1.3.DN eluted at 15.9 ml as a broad peak with a trailing shoulder and had a calculated molecular weight of 11.0 kDa, close to the 13.5 kDa monomer weight. A 1.3.DN sample crosslinked with 1 mM BSOCOES was also ran on the ENrich column. This sample had two peaks. The first was a

broad peak that eluted at 14.2 ml and had a calculated molecular weight of 42.6 kDa, which is close to the expected trimer weight of 40.5 kDa. The second peak was larger and eluted at 17.9 ml and had a calculated molecular weight of 2.3 kDa. This is likely excess crosslinker that was not removed during the buffer exchange that was performed with a centrifugal ultrafiltration device, as the N-hydroxysuccinimide ester has a λ_{max} of 260 nm and would be detected with the A280 used to monitor protein elutions. Based on the aforementioned CD, MALDI and SEC data, the fusion protein 1.3.DN forms a trimer in solution.

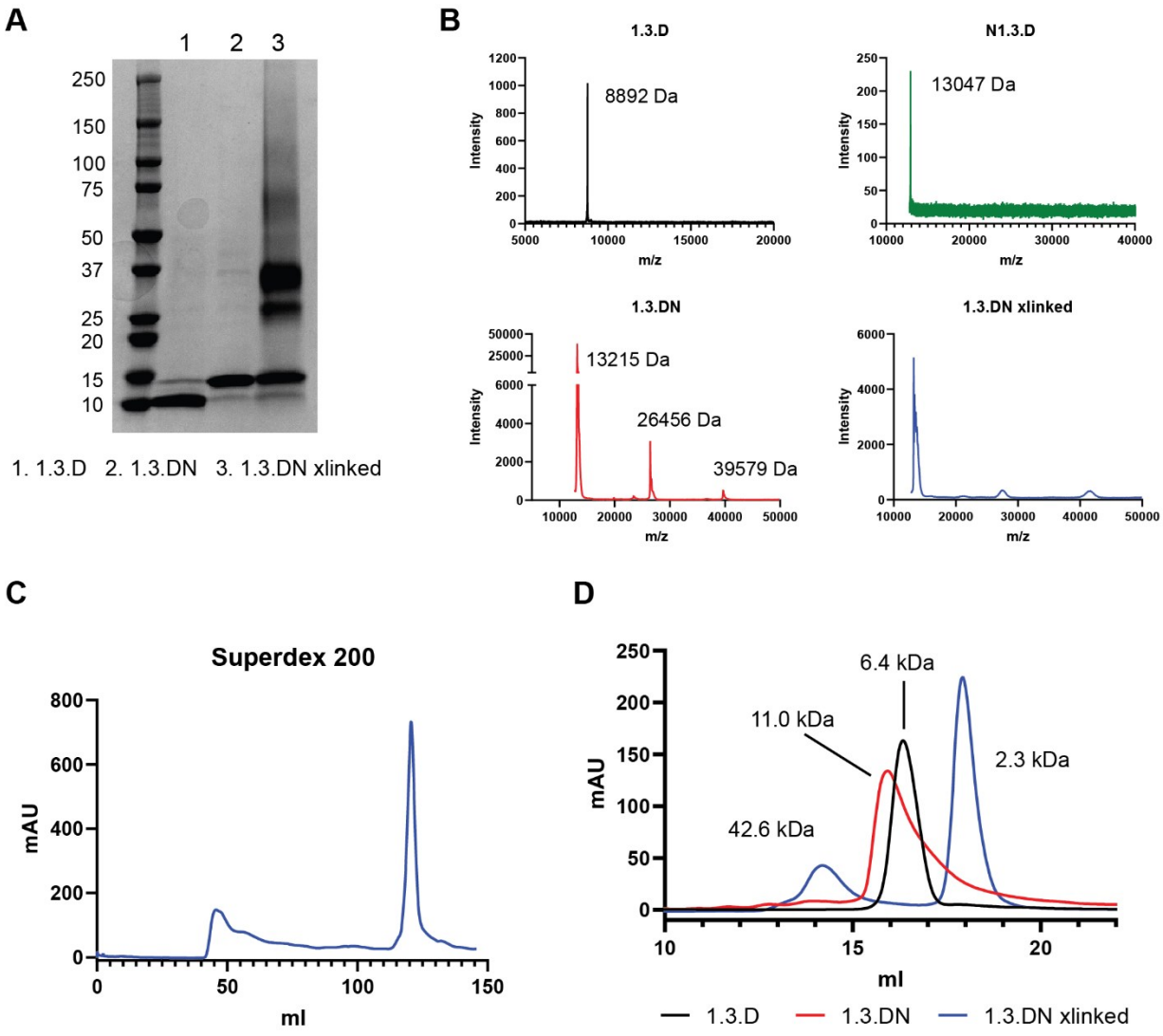
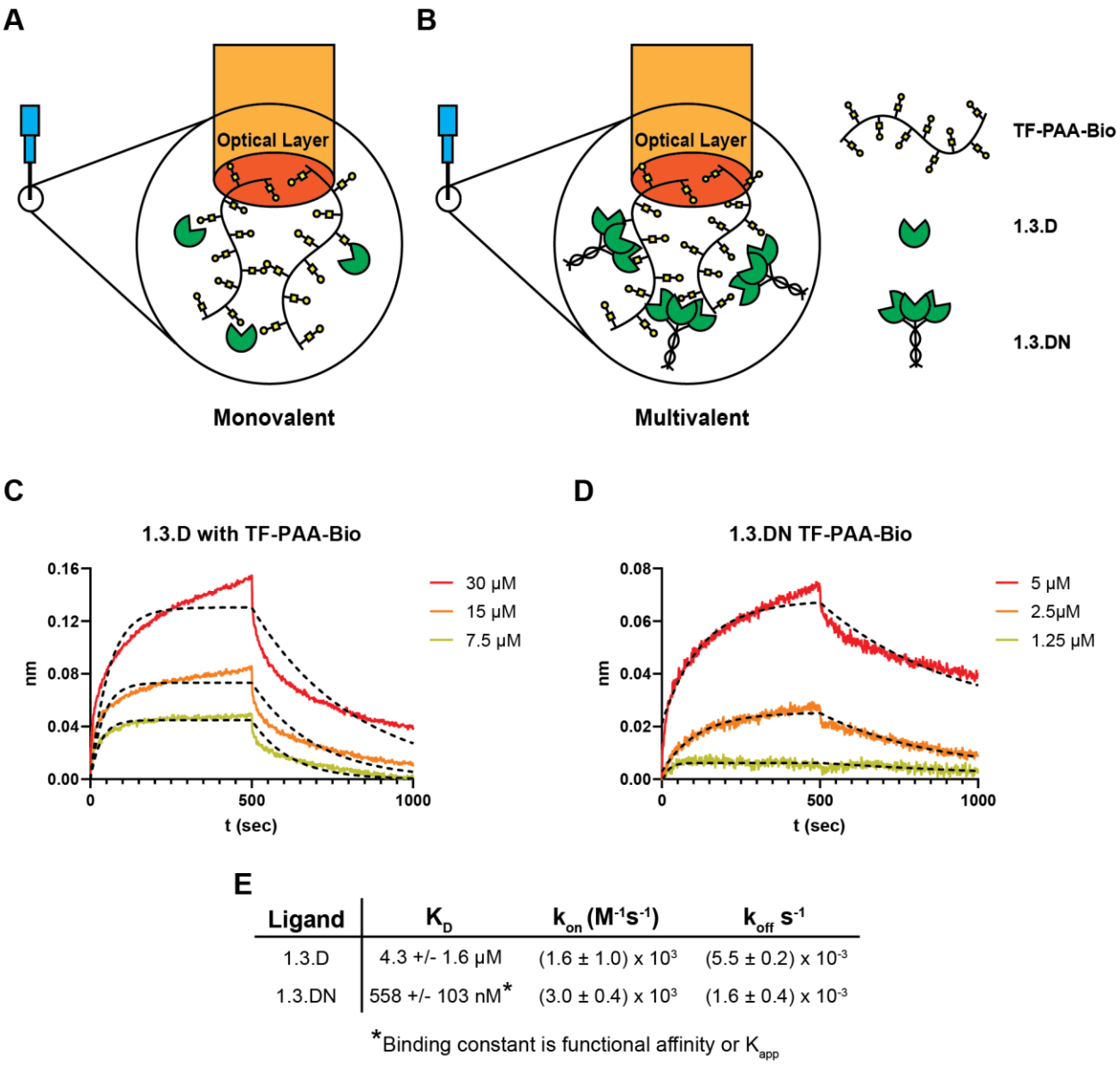


Figure 4-5. Fusion to coiled-coil domain induces trimerization.
A) SDS-PAGE gel stained with Coomassie blue showing 1: 1.3.D, 2: 1.3.DN, 3: crosslinked 1.3.DN used for figures B and D. B) MALDI traces of 1.3.D, N1.3.D, non-crosslinked 1.3.DN,

and crosslinked 1.3.DN. C) Superdex 200 trace of 1.3.DN. D) SEC using ENrich 650 column of 1.3.D (black), 1.3.DN (red), and crosslinked 1.3.DN (blue).

4.6 Trimeric GBP has improved binding affinity

Biolayer interferometry (BLI) was used to determine if 1.3.DN was still functional after fusion to the coiled-coil domain and if the oligomerization of the protein produced higher affinity interactions. To do this, the protein must be soluble in solution with the glycan immobilized on the BLI tip (**Figure 4-6A,B**). Immobilization of protein on the tip would produce multivalent interactions for both 1.3.D and 1.3.DN when dipped into a glycopolymer solution, and the influence of trimerization would not be able to be determined. TF-PAA-Bio was immobilized onto a streptavidin-coated BLI tip and dipped into a solution of 1.3.D or non-crosslinked 1.3.DN. Monomeric variant 1.3.D had a measured K_D of 4.3 ± 1.6 μM (**Figure 4-6C**). The trimeric 1.3.DN has higher affinity for the immobilized TF-PAA-Bio as expected (**Figure 4-6D**). The interaction is multivalent so a true K_D cannot be measured, but the functional affinity or K_{app} for 1.3.DN is 558 nM ± 103 nM.



E

Ligand	K_D	k_{on} ($M^{-1}s^{-1}$)	k_{off} s^{-1}
1.3.D	4.3 +/- 1.6 µM	(1.6 ± 1.0) x 10^3	(5.5 ± 0.2) x 10^{-3}
1.3.DN	558 +/- 103 nM*	(3.0 ± 0.4) x 10^3	(1.6 ± 0.4) x 10^{-3}

*Binding constant is functional affinity or K_{app}

Figure 4-6. 1.3.DN makes multivalent interactions with TF-PAA-Bio.
A) Diagram of biolayer interferometry experiment with immobilized TF-PAA-Bio making monovalent interactions with 1.3.D. B) Diagram of biolayer interferometry experiment with immobilized TF-PAA-Bio making multivalent interactions with 1.3.DN. C) BLI trace of immobilized TF-PAA-Bio with 1.3.DN at 30, 15, and 7.5 µM. D) BLI trace of immobilized TF-PAA-Bio with 1.3.DN at 5, 2.5, and 1.25 µM.

4.7 Conclusions

Protein oligomerization has several advantages, including multivalency, higher binding strength and increased stabilization.[30] Oligomerization is very important for GBPs, allowing the

usually low affinity proteins to gain avidity and improve their binding capabilities. Fusion of monomeric lectins to an oligomerization domain has shown binding affinity improvements before and are a worthwhile strategy for improvement of the evolved Sso7d-based GBPs developed in Chapters 2 and 3. Fusion of the coiled-coil neck domain of the lectin SP-D to the variant 1.3.D was explored for generation of an oligomeric Sso7d-based GBP. Coiled-coil fusion induced trimerization of 1.3.D without compromising stability, with the monomer 1.3.D having a calculated Tm of about 80 ºC and the trimeric 1.3.DN having a calculated Tm approaching 100 ºC. Trimerization of 1.3.D improved the binding affinity compared to the monomeric protein as expected.

The location of the coiled-coil fusion was shown to be important. The C-terminal fusion 1.3.DN was superior to the N-terminal fusion N1.3.D. The N-terminal fusion of the coiled-coil domain resulted in protein with secondary structure similar to the monomeric 1.3.D and did not increase in α-helicity as would be expected of a properly folded coiled-coil. This was also demonstrated by the lack of trimer present after chemical crosslinking of the protein even at high concentration of crosslinker. The SP-D lectin is organized with the neck domain N-terminal to the CRD like the N1.3.D construct, but for SP-D the coiled-coil is preceded by a cystine-rich N-terminal domain and a collagen-like domain. These domains are not necessary for production of functional trimers. Others have purified a recombinant SP-D neck-CRD only construct from *E. coli* and found that the proteins formed functional, stable trimers, though it was initially purified with an N-terminal fusion to maltose binding protein.[21] Although the C-terminal fusion of the SP-D coiled-coil is not the natural orientation of the neck in the WT SP-D, this orientation successfully produced stable trimers when fused to the small monomeric protein thioredoxin.[20]

Trimerization of the low affinity variant 1.3.D increased the functional affinity of the protein to around 550 nM while the monomeric 1.3.D affinity was about an order of magnitude weaker at about 4 μM. In Chapter 3, variant 1.3.D was further evolved into a higher affinity variant, 2.4.I, that had a K_D of 16 μM when tested with a soluble monovalent TF sugar, which is in the range of 1.3.D as measured here. However, variant 2.4.I could be used for Western blotting and cell staining applications which were unsuccessful with variant 1.3.D. The TF-PAA-Bio immobilized on the tip does increase the local concentration of TF and could lead to a higher measured affinity than that of soluble monovalent sugar. Future Sso7d-based GBP engineering efforts may yield variants with affinities in the range of variant 1.3.D, and this trimerization strategy is a way to increase the functionality of the proteins in such cases.

4.8 Materials and methods

Cloning of SP-D neck constructs

For cloning of the N-terminal fusion N1.3.D, a gene block (IDT, Coralville, IA) was ordered consisting of residues Val224 through Pro255 of human lectin surfactant protein D, a $(GGGS)_2$ sequence, Sso7d variant 1.3.D, and a C-terminal sortase motif LPETGG. For the C-terminal fusion 1.3.DN, a gene block was ordered consisting of Sso7d variant 1.3.D, a $(GGGS)_2$ sequence, residues Val224 through Pro225 of SP-D, followed by a GGGS sequence and sortase motif of LPETGG. Both gene blocks have overlaps for pET-24a expression vector (EMD Millipore, Burlington, MA.), which encodes a C-terminal His_6 affinity tag. The vector was linearized using PCR and Gibson assembly was used to assemble the final product (NEB, Ipswich, MA).

Purification of SP-D neck 1.3.D constructs

Fusion proteins were recombinantly expressed as LPETGG-His_6 constructs. BL21-CodonPlus (DE3)-RIL *Escherichia coli* cells (Agilent Technologies, Santa Clara, CA) were transformed with pET24 plasmids containing the fusion protein. Single transformants were used to inoculate 8 mL overnight cultures of LB containing 30 µg/mL each of kanamycin and chloramphenicol. Grown overnight cultures were added to 1 L of TB containing 30 µg/mL each of kanamycin and chloramphenicol and grown at 37 °C until an OD_{600} of 0.8-1 reached. Cultures were brought to 16-18 °C and expression induced in the presence of 1 mM IPTG, followed by 16-18 h of growth. Cells were harvested by centrifugation at 3700 *x g* for 30 min and overexpressed proteins purified immediately or pelleted cells stored long term at -80 °C.

All protein purification steps were carried out at 4 °C. Pelleted cells were resuspended in Buffer A (50 mM Na-phosphate pH 6.7, 500 mM NaCl, 20 mM imidazole) supplemented with 0.5

mg/mL lysozyme (Research Products International, Mount Prospect, IL), 1:1000 protease inhibitor cocktail (EMD Millipore, Burlington, MA), and 1 U/mL DNase I (NEB) and tumbled for 30 min. Homogenized cells were lysed by sonication at 50% amplitude for 3 cycles of 5 min (Vibra-Cell, Sonics, Newtown, CT), followed by clarification by ultracentrifugation at 35,000 *x g* for 60 min. Supernatants containing soluble Sso7d proteins were passed over Ni-NTA resin (HisPur, Thermo Scientific, Rockford, IL) pretreated with Buffer A. Resin was washed with Buffer A for 3 column volumes, followed by washes at 50, 100, 250, and 500 mM imidazole containing buffer before elution with high imidazole Buffer B (50 mM Na-phosphate pH 6.7, 500 mM NaCl, 1 M imidazole) for 6-8 column volumes. Column fractions of interest were pooled and immediately desalted using HiTrap desalting column (GE Healthcare, Chicago, IL). Proteins were stored at 4 °C for immediate use or flash-frozen and stored at - 80 °C.

Circular dichroism

CD spectra were obtained using a Jasco J-1500 (Jasco Inc, Easton, MD). Scans were performed with 5 μM protein in 10 mM Na-phosphate pH 6.7 using 2 mm cuvettes. For thermal melts, proteins were heated at 2 °C/min from 25-95 °C. Every 10 °C spectra were obtained from 185-250 nm, after being held at temperature for 10 minutes. Data is the average of three scans. Proteins were allowed to return to 25 °C before a final scan to determine if proteins refold after melting. The delta epsilon (Δε) at 200 nm was plotted at each temperature and data was fit using sigmoidal fit function of Graphpad prism for Tm determination.

Crosslinking

Protein was diluted to 10 μM in 50 mM Na-phosphate pH 8.0, 150 mM NaCl. BSOCOES crosslinking reagent (Thermo Scientific, Rockford, IL) dissolved in DMSO was added at concentrations from 0.1-4 mM. Crosslinking occurred for 30 minutes at room temperature. 6X Laemmli buffer was added to samples and gel electrophoresis and Western blotting was performed. Blots were probed with mouse anti-His (Thermo Fisher, Rockford, IL) and goat anti-mouse-AP conjugate. Blots were developed with NBT/BCIP one step reagent (Thermo Fisher, Rockford, IL).

MALDI

A high-resolution Bruker Autoflex LRF Speed mass spectrometer (Billerica, MA) was used for MALDI analysis of the proteins. The matrix, sinapinic acid (3,5-Dimethoxy-4-hydroxycinnamic acid, SA; ProteoChem, Hurricane, UT), was prepared at 10 mg/mL in 50% acetonitrile, 50% water, 0.1% TFA. Proteins at concentrations between 13-100 μM were then mixed 1:1 (v/v) with the matrix, and 1 μL was spotted onto the MALDI plate and dried fully before analysis.

Size exclusion chromatography

HiLoad 16/600 Superdex 200 prep grade column (Cytiva, Marlborough, MA) was equilibrated with 50 mM Na-phosphate pH 6.7, 150 mM NaCl buffer. Approximately 1 mg of 1.3.DN was injected, and molecular weight of eluted fractions was determined by calibration curve of gel filtration standards (#1511901, Bio-Rad, Hercules, CA).

ENrich SEC 650 column (Bio-Rad, Hercules CA) was equilibrated with 50 mM Na-phosphate pH 6.7, 150 mM NaCl buffer. Injections of approximately 0.6 mg 1.3.D, 1 mg 1.3.DN, and 0.7 mg

1.3.DN crosslinked were ran for one column volume and molecular weight of eluted fractions was determined by calibration curve of gel filtration standards (#1511901, Bio-Rad, Hercules, CA).

Biolayer interferometry

Apparent binding affinities of 1.3.DN was determined by bio-layer inferometry. All BLI measurements were performed in PBS supplemented with 1% BSA and 0.05% Tween-20 at 30 °C on an Octet RED96 instrument (Pall ForteBio, Fremont, CA). TF-PAA-Bio was captured on streptavidin BLI tips to a final displacement of about 1 nm, followed by a baseline reading for 60 sec. Tips were placed into protein solutions of various concentrations (10 μM to 2.5 μM) for 500 sec to measure association, followed by 500 sec in buffer alone for dissociation measurements. If needed, buffer baseline measurements from TF-PAA-Bio-loaded tips were used for background subtraction and all association and dissociation curves globally fitted to a 1:1 binding model to obtain apparent binding affinities.

4.9 Acknowledgments

I would like to thank Dr. Leah Seebald for her assistance with the MALDI experiments, Dr. Megan Kizer for her assistance in chapter edits, and Dr. Gregory Dodge for his help with SEC experiments.

4.10 References

1. Taylor, M., Drickamer, K., Imberty, A., van Kooyk, Y., Schnaar, R., Etzler, M., and Varki, A. (2022) Discovery and Classification of Glycan-Binding Proteins, *Essentials of Glycobiology, 4th edition*.

2. Wu, A. M., Lisowska, E., Duk, M., and Yang, Z. (2008) Lectins as Tools in Glycoconjugate Research, *Glycoconj. J. 26*, 899.

3. Brooks, S. A. (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.

4. Lee, R. T., and Lee, Y. C. (2000) Affinity Enhancement by Multivalent Lectin–Carbohydrate Interaction, *Glycoconj. J. 17*, 543-551.

5. Monsigny, M., Mayer, R., Roche, A.C. (1991) Sugar-Lectin Interactions: Sugar Cluster, Lectin Multivalency and Avidity, *Carbohydr. Lett. 4*, 35-52.

6. Kiessling, L. L., Young, T., Gruber, T. D., and Mortell, K. H. (2008) Multivalency in Protein–Carbohydrate Recognition, In *Glycoscience: Chemistry and Chemical Biology* (Fraser-Reid, B. O., Tatsuta, K., and Thiem, J., Eds.), pp 2483-2523, Springer Berlin Heidelberg, Berlin, Heidelberg.

7. Weis, W. I., Taylor, M. E., and Drickamer, K. (1998) The C-Type Lectin Superfamily in the Immune System, *Immunol. Rev. 163*, 19-34.

8. Crouch, E., Persson, A., Chang, D., and Heuser, J. (1994) Molecular Structure of Pulmonary Surfactant Protein D (Sp-D), *J. Biol. Chem. 269*, 17311-17319.

9. Boskovic, J., Arnold, J. N., Stilion, R., Gordon, S., Sim, R. B., Rivera-Calzada, A., Wienke, D., Isacke, C. M., Martinez-Pomares, L., and Llorca, O. (2006) Structural Model for the Mannose Receptor Family Uncovered by Electron Microscopy of Endo180 and the Mannose Receptor, *J. Biol. Chem. 281*, 8780-8787.

10. Bonnardel, F., Kumar, A., Wimmerova, M., Lahmann, M., Perez, S., Varrot, A., Lisacek, F., and Imberty, A. (2019) Architecture and Evolution of Blade Assembly in B-Propeller Lectins, *Structure 27*, 764-775.e763.

11. Ward, E. M., Kizer, M. E., and Imperiali, B. (2021) Strategies and Tactics for the Development of Selective Glycan-Binding Proteins, *ACS Chem. Biol. 16*, 1795-1813.

12. Yabe, R., Itakura, Y., Nakamura-Tsuruta, S., Iwaki, J., Kuno, A., and Hirabayashi, J. (2009) Engineering a Versatile Tandem Repeat-Type A2-6sialic Acid-Binding Lectin, *Biochem. Biophys. Res. Commun. 384*, 204-209.

13. Mahajan, S., and Ramya, T. N. C. (2018) Nature-Inspired Engineering of an F-Type Lectin for Increased Binding Strength, *Glycobiology 28*, 933-948.

14. Hamorsky, K. T., Kouokam, J. C., Dent, M. W., Grooms, T. N., Husk, A. S., Hume, S. D., Rogers, K. A., Villinger, F., Morris, M. K., Hanson, C. V., and Matoba, N. (2019)

Engineering of a Lectibody Targeting High-Mannose-Type Glycans of the Hiv Envelope, *Mol. Ther. 27*, 2038-2052.

15. Hong, X., Ma, M. Z., Gildersleeve, J. C., Chowdhury, S., Barchi, J. J., Mariuzza, R. A., Murphy, M. B., 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.

16. Irumagawa, S., Hiemori, K., Saito, S., Tateno, H., and Arai, R. (2022) Self-Assembling Lectin Nano-Block Oligomers Enhance Binding Avidity to Glycans, *Int. J. Mol. Sci. 23*.

17. Reid, K. B. M. (1998) Functional Roles of the Lung Surfactant Proteins Sp-a and Sp-D in Innate Immunity, *Immunobiology 199*, 200-207.

18. Hoppe, H.-J., Barlow, P. N., and Reid, K. B. M. (1994) A Parallel Three Stranded A-Helical Bundle at the Nucleation Site of Collagen Triple-Helix Formation, *FEBS Lett. 344*, 191-195.

19. Burkhard, P., Stetefeld, J., and Strelkov, S. V. (2001) Coiled Coils: A Highly Versatile Protein Folding Motif, *Trends Cell Biol. 11*, 82-88.

20. Li, P., Zhou, Y. J., Zhou, Y. Y., Qian, D. C., Li, O., Min, H., and Wu, C. X. (2009) The Coiled-Coil Neck Domain of Human Pulmonary Surfactant Protein D Drives Trimerization and Stabilization of Thioredoxin, a Heterologous Non-Collagenous Protein, *Protein Peptide Lett. 16*, 306-311.

21. Kishore, U., Wang, J. Y., Hoppe, H. J., and Reid, K. B. (1996) The Alpha-Helical Neck Region of Human Lung Surfactant Protein D Is Essential for the Binding of the Carbohydrate Recognition Domains to Lipopolysaccharides and Phospholipids, *Biochem. J. 318 (Pt 2)*, 505-511.

22. Shehi, E., Granata, V., Del Vecchio, P., Barone, G., Fusi, P., Tortora, P., and Graziano, G. (2003) Thermal Stability and DNA Binding Activity of a Variant Form of the Sso7d Protein from the Archeon Sulfolobus Solfataricus Truncated at Leucine 54, *Biochemistry 42*, 8362-8368.

23. Gera, N., Hussain, M., Wright, R. C., and Rao, B. M. (2011) Highly Stable Binding Proteins Derived from the Hyperthermophilic Sso7d Scaffold, *J. Mol. Biol. 409*, 601-616.

24. Greenfield, N. J., and Fasman, G. D. (1969) Computed Circular Dichroism Spectra for the Evaluation of Protein Conformation, *Biochemistry 8*, 4108-4116.

25. Zhou, N. E., Kay, C. M., and Hodges, R. S. (1992) Synthetic Model Proteins. Positional Effects of Interchain Hydrophobic Interactions on Stability of Two-Stranded Alpha-Helical Coiled-Coils, *J. Biol. Chem. 267*, 2664-2670.

26. Micsonai, A., Wien, F., Bulyáki, É., Kun, J., Moussong, É., Lee, Y. H., Goto, Y., Réfrégiers, M., and Kardos, J. (2018) Bestsel: A Web Server for Accurate Protein Secondary Structure Prediction and Fold Recognition from the Circular Dichroism Spectra, *Nucleic Acids Res. 46*, W315-w322.

27. Micsonai, A., Moussong, É., Wien, F., Boros, E., Vadászi, H., Murvai, N., Lee, Y.-H., Molnár, T., Réfrégiers, M., Goto, Y., Tantos, Á., and Kardos, J. (2022) Bestsel: Webserver for Secondary Structure and Fold Prediction for Protein Cd Spectroscopy, *Nucleic Acids Res. 50*, W90-W98.

28. Baumann, H., Knapp, S., Lundbäck, T., Ladenstein, R., and Härd, T. (1994) Solution Structure and DNA-Binding Properties of a Thermostable Protein from the Archaeon Sulfolobus Solfataricus, *Nat. Struct. Biol. 1*, 808-819.

29. Håkansson, K., Lim, N. K., Hoppe, H.-J., and Reid, K. B. M. (1999) Crystal Structure of the Trimeric A-Helical Coiled-Coil and the Three Lectin Domains of Human Lung Surfactant Protein D, *Structure 7*, 255-264.

30. Engel, J., and Kammerer, R. A. (2000) What Are Oligomerization Domains Good For?, *Matrix Biol. 19*, 283-288.

Chapter 5: Application of a gut-immune co-culture system for the study of N-glycan-dependent host-pathogen interactions of *Campylobacter jejuni*

This chapter has been adapted and reprinted with permission from:

Zamora, C.Y., **Ward, E.M.**, Kester, J.C., Chen, W.L.K., Velazquez, J.G., Griffith, L.G., Imperiali, B. (2020). Application of a Gut-Immune Co-Culture System for the Study of N-Glycan-Dependent Host-Pathogen Interactions of *Campylobacter jejuni*, *Glycobiol. 30*, 374-381. doi.org/10.1093/glycob/cwz105.

Contributions:

Tissue culture and preparation of GICs performed by Jason Velazquez and Wen Li Kelly Chen
Campylobacter jejuni culturing and GIC incubations performed by Cristina Zamora and Elizabeth Ward
Outer membrane vesicle preparation, application on GICs, protease profiling, and cadherin digestion performed by Elizabeth Ward
Mucin analysis, Lucifer yellow permeability assays, TEER measurements, and inflammatory marker immunoassays performed by Jason Velazquez and Jemila Kester
Manuscript prepared by Cristina Zamora and Elizabeth Ward
Research designed by Cristina Zamora, Elizabeth Ward, Wen Li Kelly Chen, Jemila Kester, Linda Griffith, and Barbara Imperiali

5.1 Abstract

An *in vitro* gut-immune co-culture model with apical and basal accessibility, designed to more closely resemble a human intestinal microenvironment, was employed to study the role of the N-linked protein glycosylation (Pgl) pathway in *Campylobacter jejuni* pathogenicity. The gut-immune co-culture (GIC) was developed to model important aspects of the human small intestine by the inclusion of mucin producing goblet cells, human enterocytes, and dendritic cells, bringing together a mucus-containing epithelial monolayer with elements of the innate immune system. The utility of the system was demonstrated by characterizing host-pathogen interactions facilitated by N-linked glycosylation, such as host epithelial barrier functions, bacterial invasion and immunogenicity. Changes in human intestinal barrier functions in the presence of 11168 *C. jejuni* (wildtype) strains were quantified using GICs. The glycosylation-impaired strain 11168 Δ*pglE* was 100-fold less capable of adhering to and invading this intestinal model in cell infectivity assays. Quantification of inflammatory signaling revealed that 11168Δ*pglE* differentially modulated inflammatory responses in different intestinal microenvironments, suppressive in some but activating in others. Virulence-associated outer membrane vesicles produced by wildtype and 11168Δ*pglE C. jejuni* were shown to have differential composition and function, with both leading to immune system activation when provided to the gut-immune co-culture model. This analysis of aspects of *C. jejuni* infectivity in the presence and absence of its N-linked glycome is enabled by application of the gut-immune model and we anticipate that this system will be applicable to further studies of *C. jejuni* and other enteropathogens of interest.

5.2 Introduction

The Gram-negative pathogen *Campylobacter jejuni* is a leading cause of gastroenteritis and diarrheal disease.[1] Infections can be severe, especially in children, the elderly and immunocompromised individuals, with infection leading to fatality in 1 in 3,000 cases by some estimates.[2] It has been suggested that adhesion to and invasion of host epithelial cells is critical for disease development.[3] *C. jejuni* has been shown to transcellularly invade intestinal epithelial cells, becoming encased in vacuoles that protect them from lysosomal degradation and immune detection.[4, 5] Characterization of the roles of virulence determinants in adhesion and invasion is crucial for identification of pathways, enzymes, and molecules to target for therapeutic intervention. This is of particular importance in the era of antimicrobial resistance, where additional insight into the contributors to microbial pathogenicity would allow for investigation of anti-virulence agents that may be less likely to elicit such resistance. Many virulence factors have been associated with host cell invasion, including flagellar motility and chemotaxis, outer membrane vesicles, adhesins and proteases.[6-8]

One such determinant of microbial pathogenicity is protein glycosylation.[9] N-linked glycosylation in *C. jejuni* has been shown to impact several cellular processes, including proteome stability, protein quality control, stress response, nutrient uptake, chemotaxis, cell morphology, and virulence.[10, 11] The correlation between N-linked protein glycosylation and virulence has been demonstrated in avian models of infection and *in vitro* mammalian systems. The N-linked protein glycosylation (*pgl*) operon and N-glycan of *C. jejuni* are illustrated in **Figure 5-1**. Cecal analysis of chick intestines has shown significantly-decreased *C. jejuni* colonization upon individual knockouts of *pglB, pglD, pglE, pglH* and *pglK* genes.[12-14] Additionally, studies of 81-176 *C. jejuni* in murine models have revealed similar decreases in colonization by *pglB* and *pglE* knockouts.[14]

In vitro studies of *C. jejuni* virulence have been carried out in unpolarized monolayers of cell lines HeLa and T84 [15], among others.[16-18] However, these systems do not feature tight junctions, brush borders or other important components of intestinal epithelia.[19] Additionally, studies of infectivity on polarized monolayers of human epithelial colorectal adenocarcinoma (Caco-2) [20] do not account for the intestinal mucosal layer and its barrier functions, which in some strains of *C. jejuni* has been shown to play a role in bacterial motility [21] and assist in binding and invasion of host cells.[22] Bacterial interactions with mucosal components of the gut have been shown to be an important part of virulence behaviors, in some cases influencing whether *C. jejuni* acts as a commensal as seen in chickens and other animals, or as a pathogen as seen in humans.[23, 24] Although these *in vitro* studies on components of human intestinal physiology have provided valuable insight, a more comprehensive understanding of the mechanism and determinants of *C. jejuni* pathogenicity could be potentially achieved with model systems incorporating several aspects and more closely resembling human gut physiology.

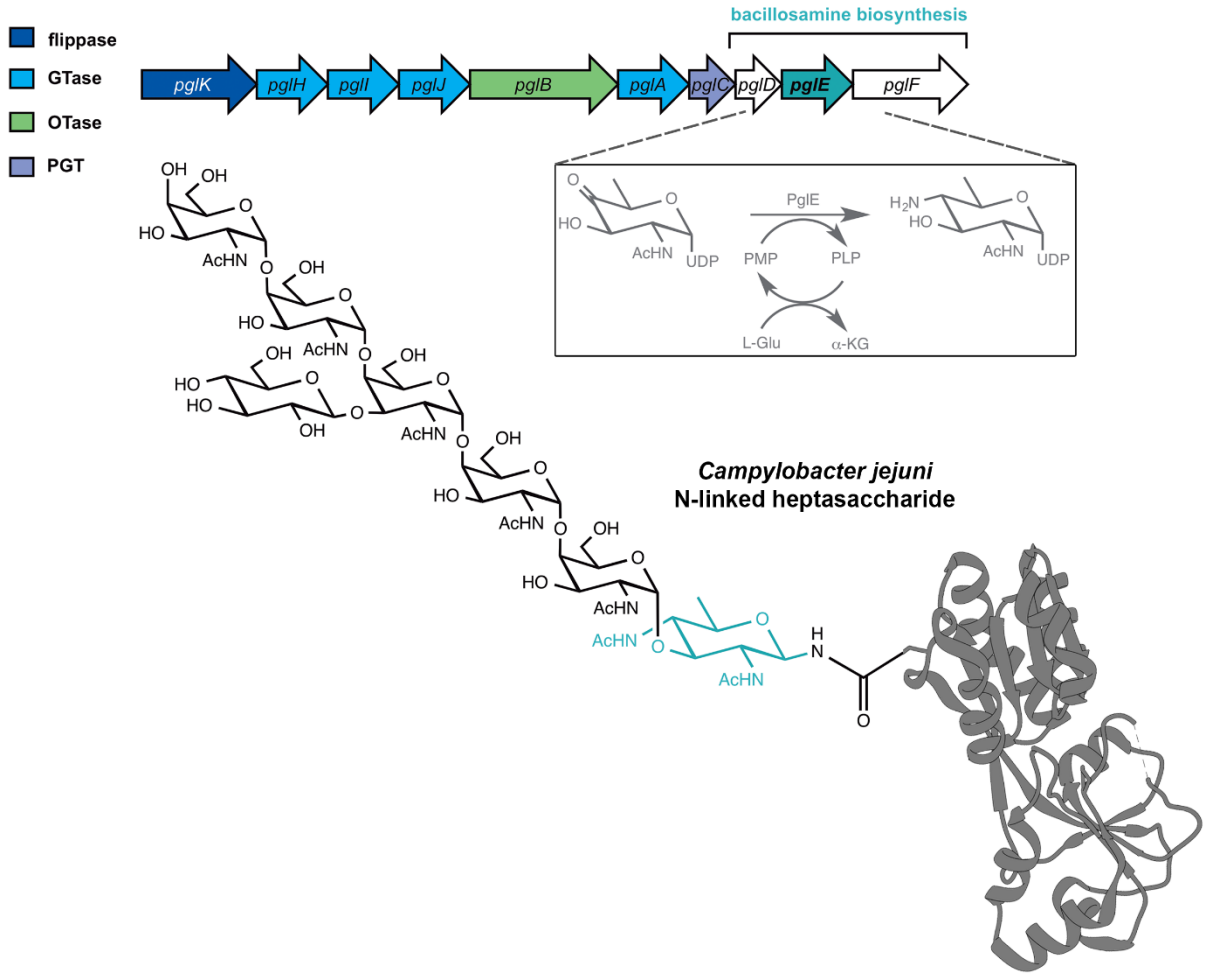


Figure 5-1. Illustration of the *pgl* locus in *C. jejuni* and the resulting N-linked heptasaccharide
Ten enzymes act in concerted fashion to produce N-linked glycoproteins in C. jejuni. Knockout of transaminase PglE results in near-complete loss of function in this biosynthetic pathway. Shown as a cartoon is putative adhesin PEB3 from C. jejuni (PDB: 2HXW), N-glycosylated with the C. jejuni heptasaccharide.

To address this need, herein we present the adaptation and application of a gut-immune co-culture model (GIC, **Figure 5-2**), engineered to mirror important aspects of native human small intestinal tissue including barrier functions and pharmacological properties, for the characterization of *C. jejuni* N-linked glycans in host-pathogen interactions.[25-28] Each GIC is composed of a permeable Transwell insert with an epithelial monolayer of absorptive human

enterocytes and human mucin producing goblet cells on the apical face, and dendritic cells on the basolateral face. Thus, each GIC brings together a mucus-bearing gut with essential elements of the innate immune system.

This gut-immune co-culture model was used to investigate the impact of *C. jejuni* infection on different aspects of human intestinal function, such as barrier integrity, mucus production and innate immune system activation, and the importance of N-glycosylation in these host-pathogen interactions. We anticipate that the application of the gut-immune model to the characterization of this recalcitrant microaerophilic organism will advance studies of the pathogen in a relevant, quantitative multidimensional platform.

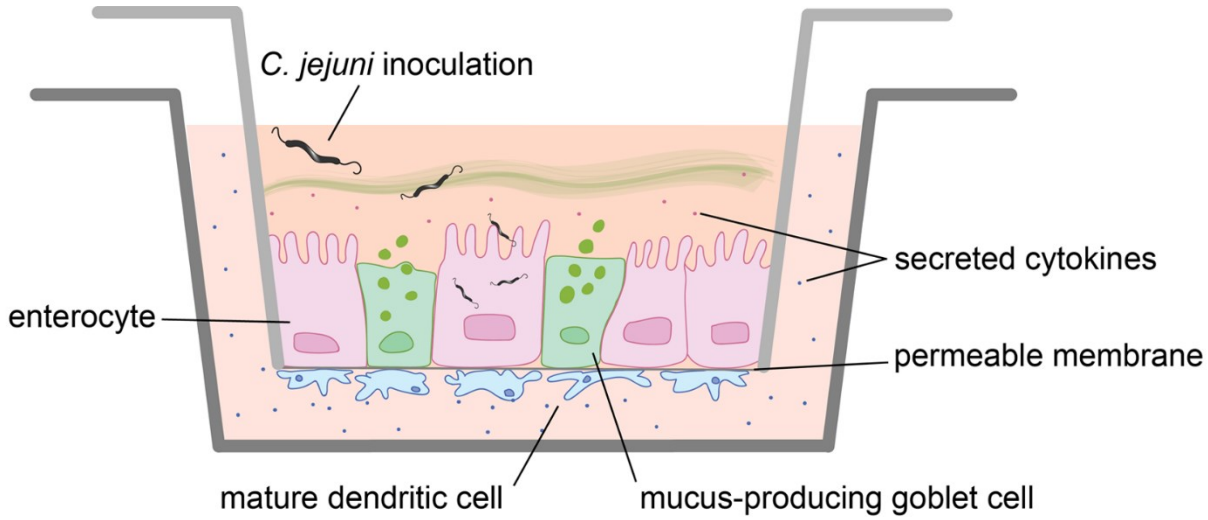


Figure 5-2. Schematic illustration of the gut-immune co-culture model system.
Gut-immune co-cultures (GICs), comprised of enterocytes, goblet cells and mature dendritic cells colocalized to Transwell inserts, act as a useful model of important intestinal barrier and immunoregulatory functions. GICs produce a wide variety of readouts, such as measures of the glyco-interactome between bacterial pathogen and gut, quantification of secreted mucins, glycan profiling before and after bacterial infection, gut epithelium viability, barrier function measurements, bacterial adhesion and invasion, mRNA profiling and transcriptome changes, gut-immune inflammatory crosstalk and several others.

5.3 Culturing of *C. jejuni* on gut-immune co-cultures

To demonstrate the utility of the gut-immune model for glycobiological studies of bacterial pathogenicity, the impact of *C. jejuni* N-glycosylation was characterized and quantified using several analytical readouts of the gut-immune model, listed in **Figure 5-2**. Each GIC contains an epithelial monolayer comprising a 9:1 ratio of absorptive human enterocytes (C2BBe1) and human mucin-producing goblet cells (HT29-MTX) seeded and grown on a permeable Transwell insert. The C2BBe1 cell line was chosen for its ability to form a brush border morphologically comparable to that of human colon, with heterogeneous microvillar presentation. Dendritic cells, derived by *in vitro* differentiation of primary human monocytes, were present on the basolateral face of the Transwell insert at approximately a 1:10 ratio to epithelial cells in the mature monolayer. In these studies, the apical face of GICs was challenged with wildtype NCTC 11168 *C. jejuni* or a 11168Δ*pglE* strain, which exhibits near-complete abolition of N-linked protein glycosylation without a significant growth defect (**Figure 5-3**).[29]

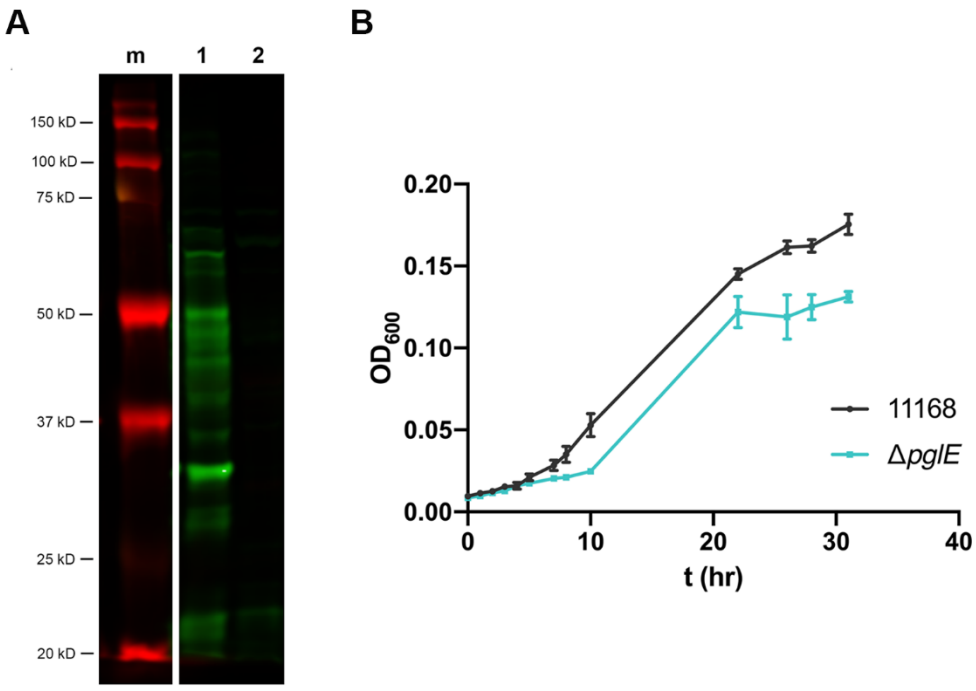


Figure 5-3. Knockout of *pglE* in 11168 *C. jejuni* results in significantly depleted *N*-glycosylated proteomes.
A) Equal amounts of whole cell lysates of wildtype and Δ*pglE C. jejuni* were analyzed by immunoblotting, probing with rabbit anti-N-glycan antibodies. *Lanes*: m: molecular weight markers, 1: 11168 *C. jejuni*, 2: 11168 Δ*pglE*. B) 11168 *C. jejuni* and 11168Δ*pglE* was inoculated into Mueller-Hinton media in a 96 well plate and incubated at 37°C in microaerophilic conditions. OD_{600} was measured periodically and un-inoculated Mueller-Hinton broth was used for background subtraction. Points show the mean of six replicates with error bars showing one standard deviation.

It was initially unclear whether this fastidious microaerophilic organism, typically cultured to low optical densities of 0.2-0.3 in Mueller Hinton broth under 85% N_2, 5% O_2 and 10% CO_2, would be viable or culturable during and after prolonged studies on GICs. Conversely, it was unknown if the nutrient rich GIC growth media would cause bacterial overgrowth. It was determined that the typical mammalian tissue culture conditions used to grow and maintain GICs (37 °C and 5% CO_2, see Generalized Procedure E in SI) were sufficient to sustain *C. jejuni* viability on GICs throughout these experiments. Several multiplicities of infection (MOI) were screened in order to determine the minimum number of bacteria needed to generate an immune response from

the apical epithelium within 24-48 h, with MOI between 10-50 resulting in strong immune responses without risk of bacterial overgrowth. It is worth noting that adherence and infection of GICs by *C. jejuni* at MOI of 10 is measurable at soon as 2 h, similar to studies on other monolayer models.[30] When applied to GICs, all wildtype and mutant *C. jejuni* strains tested remained viable and entered strong log-phase growth in the nutrient-rich GIC media, but only in the presence of a GIC epithelium – mammalian growth media and incubator conditions alone were not sufficient to maintain bacterial viability (**Figure 5-4**). These results were suggestive of the oxygen gradient and anoxic conditions known to be formed at the face of the intestinal epithelium.[31] Despite growth of *C. jejuni* on the apical face of GICs, no significant epithelial or dendritic cell distress or cytotoxicity was ever observed. Ultimately, these multiplicities of infection, culture conditions and experimental time frames were able to support bacterial and intestinal cell survival and allow for measurements to be taken of both bacterial and mammalian processes.

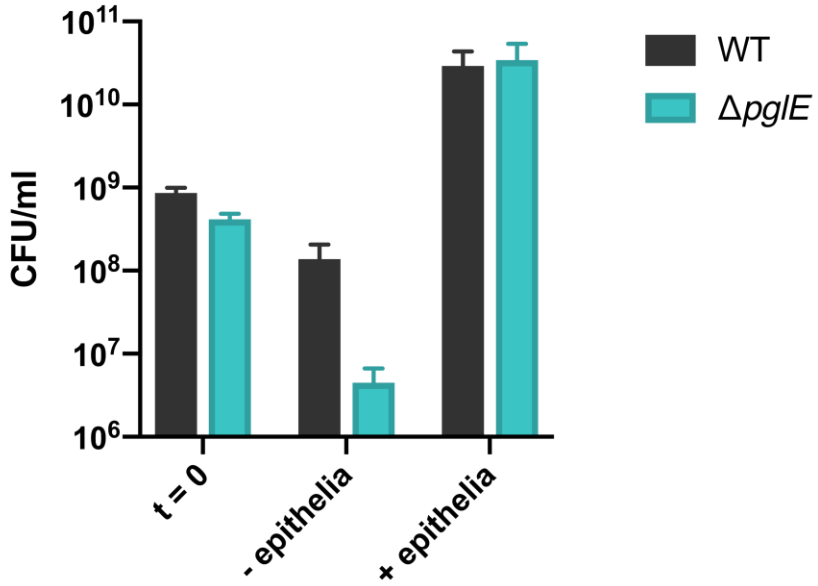


Figure 5-4. *C. jejuni* strains can grow in GIC culture conditions only in the presence of a GIC epithelium.
11168 *C. jejuni* and 11168Δ*pglE* in apical growth media was applied to transwells containing a C2BBe1 epithelial monolayer or an empty well at MOI=10 and incubated under GIC culture

conditions for 24 h. Apical media was sampled for initial and final CFU determination. Bars show mean of three independent replicates, with error bars showing one standard deviation.

5.4 Epithelial barrier integrity upon *C. jejuni* infection

Using these conditions, we characterized the impact of *C. jejuni* infection on aspects of GIC epithelial barrier integrity and probed the role of N-glycosylation. GICs were challenged with wildtype NCTC 11168 *C. jejuni* or 11168Δ*pglE* and intestinal secreted mucin content was examined. As shown in **Figure 5-5A**, the presence of both wildtype and Δ*pglE C. jejuni* resulted in a significant reduction of secreted mucins associated with the monolayer as compared with uninfected GICs. Next, changes in the barrier integrity in response to infection were quantified by Lucifer yellow permeability, measuring apical-to-basal transport of dye to indicate paracellular permeability in response to infection.[32] In contrast to wildtype, which showed modest changes in barrier function, infection with Δ*pglE* resulted in significantly increased monolayer permeability (**Figure 5-5B**).

5.5 Influence of glycosylation on adherence and invasion

Next, as pathogen/host association and invasion are primary hallmarks of *C. jejuni* infection *in vivo*, we quantified these behaviors in the GIC. In this analysis, a nearly 100-fold decrease in bacterial association with the epithelial monolayer between wildtype and Δ*pglE* was observed (**Figure 5-5C**). Monolayer infection was also quantified after gentamycin treatment of the apical compartment, killing all bacteria except those within cells, and a similar difference between wildtype and Δ*pglE* infectivity was observed. Additionally, a statistically significant, but modest, 2.4-fold decrease in adhesion and invasion of wildtype *C. jejuni* was observed (**Figure 5-5D**) following treatment of monolayers with 2 mM free reducing heptasaccharide from *C. jejuni,* shown in **Figure 5-1**. Competition between soluble *C. jejuni* N-glycan, also known as free

oligosaccharide or FOS, and wildtype *C. jejuni* suggests a molecular role for the N-glycan in the adhesion and invasion of host intestinal cells.[33]

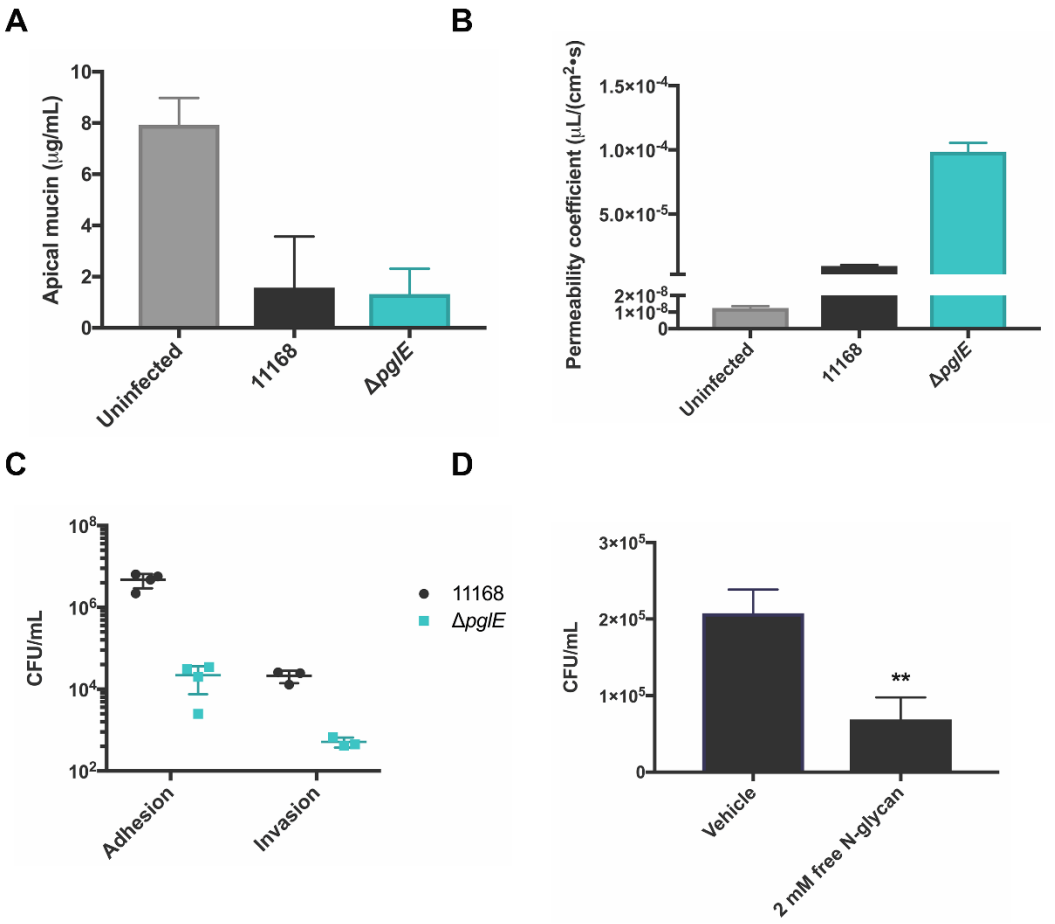


Figure 5-5. Effects of loss of N-glycosylation in 11168 *C. jejuni* on barrier functions and on epithelial adhesion and invasion.
A) Infection of GICs with C. jejuni wildtype and ΔpglE results in significantly-reduced levels of soluble mucins in transwells. GICs were inoculated with 11168 or 11168ΔpglE C. jejuni, incubated for 24 h and soluble mucins quantified by Alcian blue colorimetric assay. Data shown are the average of three independent experiments, with error bars denoting one standard deviation. B) Glycosylation-impaired C. jeuni destabilize host tight junctions to a significantly higher degree than wildtype. GICs were inoculated with 11168 or 11168ΔpglE C. jejuni for 48 h and monolayer integrity characterized by Lucifer yellow permeability assay. Data shown are the average of 4 independent replicates, with error bars denoting one standard deviation. C) Loss of N-glycosylation in 11168 C. jejuni attenuates adhesion to and invasion of gut epithelial monolayer. GICs were inoculated with 11168 or 11168ΔpglE C. jejuni in DMEM growth media and incubated at 37 °C for 24 h, followed by quantification of bacterial adhesion or invasion by colony plate counting. Data shown are independent biological replicates and representative of four independent data sets, with error bars denoting one standard deviation. Longer horizontal bar gives the mean

of all measurements. D) Bacterial infectivity appears directly mediated by C. jejuni N-linked heptasaccharide. GICs were pre-incubated with fresh media (vehicle) or 2 mM free N-glycan, followed by 2 h incubation with 11168 C. jejuni. Data shown are the average of 3 independent replicates, in technical duplicate, with error bars denoting one standard deviation. ** indicates $P<0.001$ (Welch's unpaired t-test).

5.6 Inflammatory response upon *C. jejuni* infection

A major advantage of the gut-immune model for studying *C. jejuni* pathogenicity is its utility in quantifying both intestinal epithelium inflammation and the resulting inflammatory response of the basal innate immune component. The extent of immunogenicity of *C. jejuni* infection was measured by apical infection of the GICs with wildtype or Δ*pglE C. jejuni* and sampling the basolateral media. Total cytokine and chemokine release into the basolateral compartment by both the intestinal epithelium and dendritic cells on the basal surface was measured. TNF-a, IL-6, CCL-2, CCL-3 and several other cytokines and chemokines were found to be increased in response to treatment with *C. jejuni*, with infection by Δ*pglE C. jejuni* resulting in slightly increased inflammation across many cytokines that were measured (**Figure 5-6**). Together, these experiments show the utility of the GICs for studying the roles of protein glycosylation in the complex immune signaling resulting from the multiple cell types that compose human intestinal epithelia.

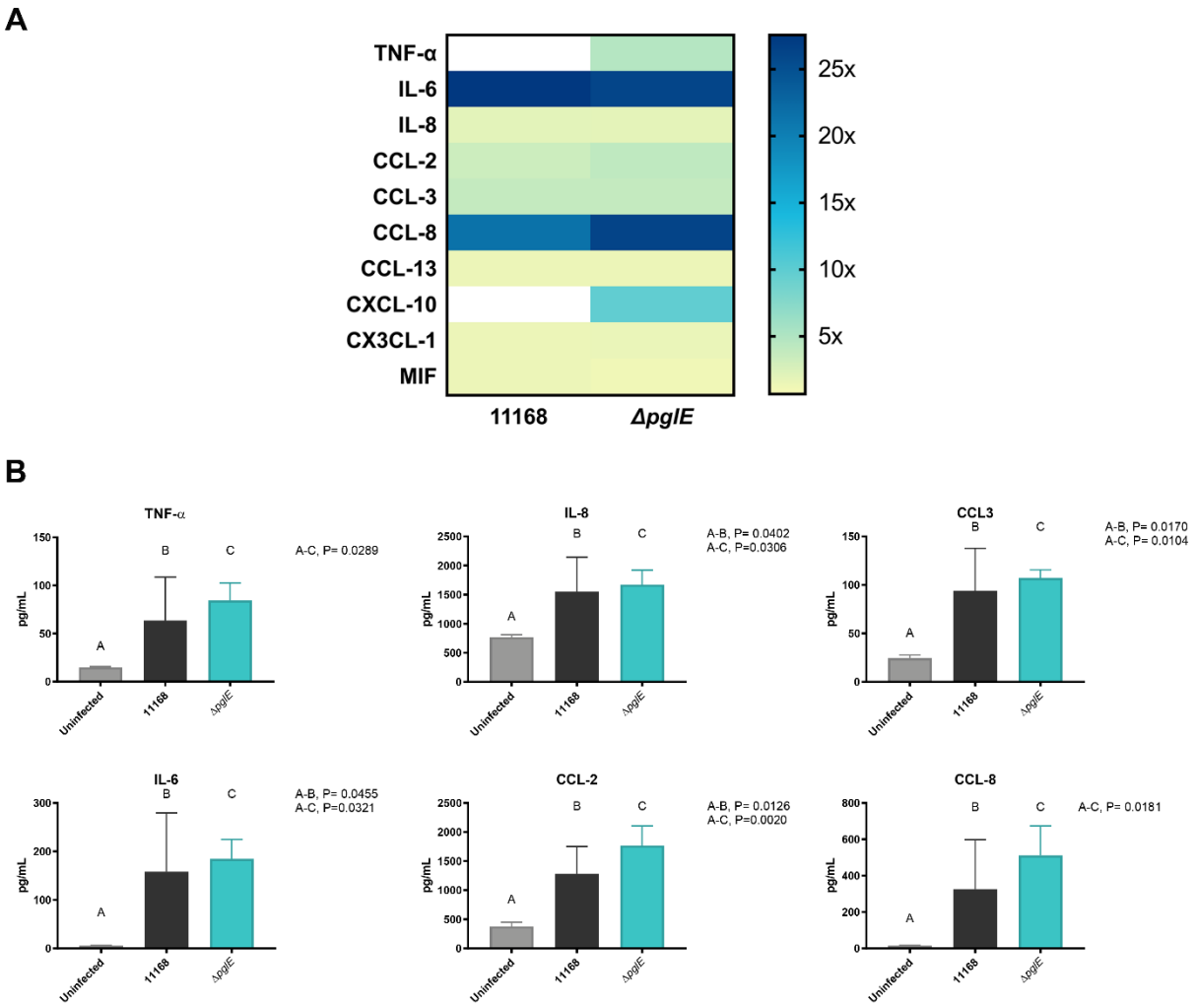


Figure 5-6. Loss of N-linked glycosylation in 11168 *C. jeuni* results in immune response on GICs.

A) 11168 C. jejuni on apical face of gut-immune co-cultures elicit an immune response with dendritic cells present on the basolateral face. Growth media from GICs inoculated with 11168 C. jejuni, 11168ΔpglE or vehicle alone as a control were taken at 24 h post-inoculation from basolateral compartments. Inflammatory markers were quantified by 40-plex immunoassay. The average of four independent replicates, normalized to cytokine or chemokine concentrations in control GICs, are shown graphed as a heat map showing fold-over vehicle alone. Cytokine and chemokine concentrations in all three data sets were compared to each other via 1-way ANOVA with no correction for multiple comparisons and those with an absolute $P<0.05$ shown. B) Some of the cytokines and chemokines shown in A).

5.7 Outer membrane vesicle isolation and protease profiling

Finally, to show the versatility of the GICs in glycobiological studies of secreted bacterial virulence factors, we characterized the effect of outer membrane vesicles (OMVs). OMVs are 25-

250 nm diameter vesicles released by Gram-negative bacteria into the extracellular milieu. These bacterial "cargo drops" deliver glycoproteins, effector proteins, DNA, endotoxins and proteases directly to the cytosol of host cells, initiating and facilitating pathogenic processes prior to direct cell-cell contact.[7, 34-36] OMVs from 11168 [37], hypermotile variant 11168H [7] and 81-176 *C. jejuni* [38] have been shown to contain cytolethal distending toxin, and are hypothesized to be the primary mechanism of toxin delivery to the host. To study the role of the *pgl* locus on the behavior of this important *C. jejuni* virulence factor, OMVs were isolated from cultures of 11168 and 11168Δ*pglE* *C. jejuni* and analyzed for glycoprotein content by immunoblotting. A significant loss of N-linked glycoproteins was observed in OMVs isolated from Δ*pglE* when compared to wildtype (**Figure 5-7A**). Analysis of wildtype and Δ*pglE* OMV content by activity-based serine protease profiling with the biotinylated fluorophosphonate FP-biotin (**Figure 5-7B**) detected an enrichment of proteases in Δ*pglE* OMV over those of wildtype 11168 *C. jejuni*.[39] Proteases in OMVs from both strains were capable of cleaving human tight junction protein E-cadherin *in vitro*, with Δ*pglE* OMV resulting in more cleavage products (**Figure 5-7C**).

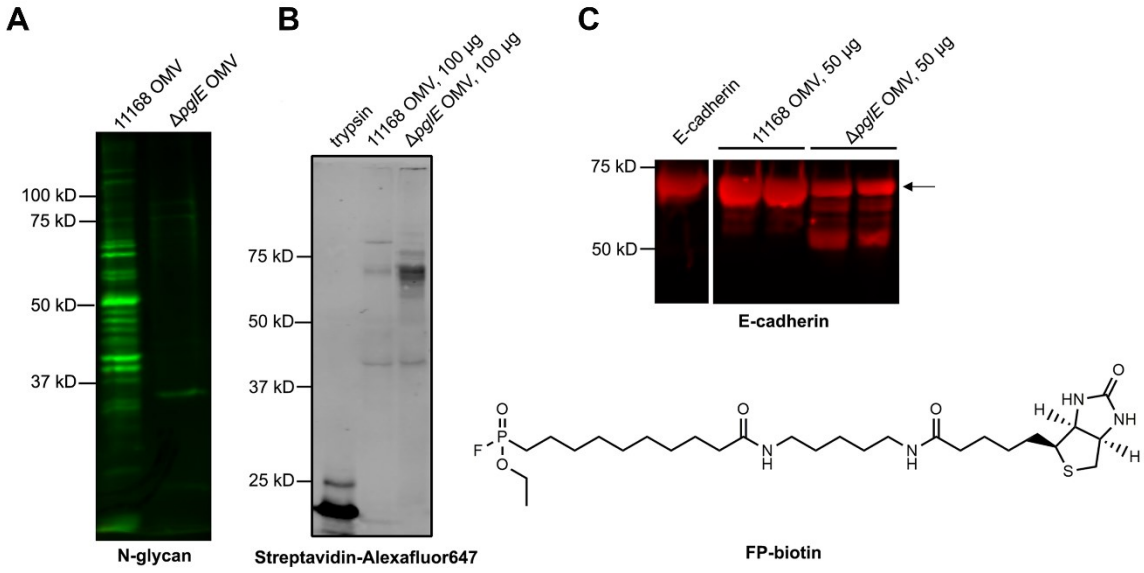


Figure 5-7. Loss of N-linked glycosylation in 11168 *C. jeuni* results changes in virulence factor function.

A) Loss of Pgl pathway function in 11168 *C. jejuni* results in the production of glycoprotein-depleted outer membrane vesicles. Proteins from a 10 μg aliquot of OMV isolated from 11168 *C. jejuni* and 11168Δ*pglE* were resolved by SDS-PAGE and analyzed by immunoblotting with a rabbit anti-N-glycan antibody. B) Activity-Based Protein Profiling (ABPP) reveals OMV from 11168Δ*pglE C. jejuni* are enriched in serine proteases. Biotinylated proteases within 100 μg aliquots of OMVs isolated from 11168 *C. jejuni* and 11168Δ*pglE* were resolved by SDS-PAGE, followed by imaging with Streptavidin-Alexafluor-647. Trypsin was used as a positive control for ABPP. C) OMV from 11168Δ*pglE C. jejuni* cleave human tight junction proteins more readily compared to wildtype OMV. Recombinant His-tagged human E-cadherin was incubated for 16 h at RT with 50 μg OMV isolated from 11168 *C. jejuni* and 11168Δ*pglE*, followed by resolution by immunoblotting and probing with mouse anti-His antibody. Full length E-cadherin is indicated with an arrow.

5.8 Influence of outer membrane vesicles on gut-immune co-culture inflammatory response

A useful feature of GICs is the ability to vary composition of the apical epithelial monolayer according to investigator needs, while maintaining dendritic cells in the basolateral compartment. To study the functions of glycosylated and non-glycosylated *C. jejuni* OMV, barrier function and inflammation were measured on GICs lacking goblet cells and dendritic cells to ensure effective vesicle-epithelial cell contact. C2bbe1 monolayers were treated with 100 μg doses of OMVs from both strains and showed no change in barrier integrity by TEER (**Figure 5-8A**). However, IL-8, CXCL1, and several other immune markers were secreted by these GICs in response to OMV treatment (**Figure 5-8B**). Importantly, these results illustrate the utility of GICs in quantitatively characterizing several aspects of virulence factor function such as those exhibited by *C. jejuni* OMV.

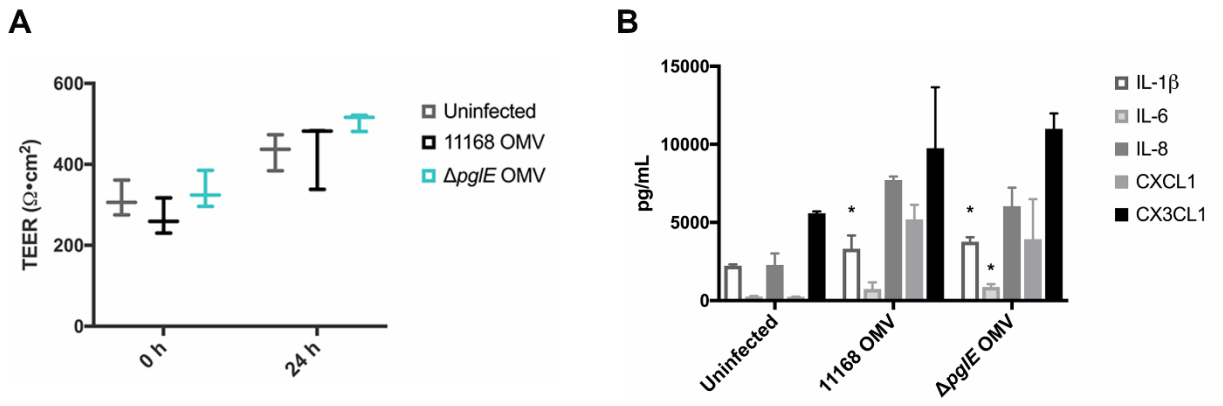


Figure 5-8. OMV from 11168 *C. jejuni* provoke an immune response from gut epithelia without changes in barrier integrity.
A) Barrier integrity of each C2BBe1-only GIC was quantified by transepithelial electrical resistance (TEER) measurements at the start of the experiment. GICs were treated with vehicle, 100 µg OMV isolated from 11168 *C. jejuni,* or 100 µg 11168Δ*pglE* OMV and incubated at 37 °C for 24 h. After incubation, TEER values were again measured for all monolayers and the apical media was sampled for immunoassay analysis. Data shown are the average of three independent replicates, with error bars denoting one standard deviation. The mean is shown by a longer horizontal bar. B) GICs were treated with vehicle alone, 100 µg OMV isolated from 11168 *C. jejuni,* or 100 µg 11168Δ*pglE* OMV, followed by incubation for 24 h. Apical media from all GICs was sampled and cytokines and chemokines present quantified by multiplexed immunoassay. Data shown are the average of three independent replicates, in technical duplicate, with error bars denoting one standard deviation. Mean marker concentrations in 11168 OMV and 11168Δ*pglE* OMV samples were compared to uninfected GICs and each other via 1-way ANOVA with no correction for multiple comparisons. * indicates statistically significant differences over background with absolute P-values ≤0.05.

5.9 Conclusions

Use of the gut-immune model allows for convenient and simultaneous measurement of physiological and immunological responses to infection by *C. jejuni*, whose protein glycosylation pathway has been shown to be important in pathogenicity. We anticipate that this model system will be a valuable alternative for studying interactions of *C. jejuni* and other enteropathogens with host cells. These studies also show that GICs can be interfaced with microaerophilic pathogens such as *C. jejuni* for periods of time amenable to immunological study and under conventional culture conditions in normoxia and may be used to study other facets of this cell-cell interaction.

Furthermore, the presented conditions should be readily adaptable to the study of other "difficult-to-culture" Gram-negative pathogens (e.g. *Helicobacter pylori*).

The analyses on GICs reflected known direct relationships between *C. jejuni* N-glycosylation and several pathogenic functions, in addition to shedding light on new aspects of its biology. Soluble mucin content on GICs was largely depleted during incubation with *C. jejuni* - a finding not previously quantified. Mucin degradation is a common tactic employed by enteric pathogens to breach the mucus layer, and increased gene expression of putative mucin degrading enzymes has been observed in *C. jejuni* in the presence of MUC2.[40, 41] The reduction of monolayer integrity was also directly correlated with loss of bacterial N-glycosylation and could be explained by a concomitant increase in proteases within Δ*pglE* OMVs capable of cleaving tight-junction proteins, shown in **Figure 5-7C**. Importantly, the degree of *C. jejuni* adhesion and invasion in the GICs was in direct correlation with global N-linked protein glycosylation (**Figure 5-5C**) and in agreement with what has been observed in other systems.[14, 42, 43]

Although direct release of cytokines by non-polarized and polarized epithelial monolayers and dendritic cells has been reported separately in response to wildtype *C. jejuni* exposure [30, 44, 45], this is the first characterization of the impact of the Pgl pathway on cytokine response in the intestinal epithelium in conjunction with basolateral innate immune cells, with many cytokines and chemokines that were measured not previously reported in studies of *C. jejuni* infection. Here, it appears that basolateral inflammatory cytokine and chemokine secretion was modestly increased when the GICs were infected with the Δ*pglE* strain as compared to wildtype (**Figure 5-6A**). Although cell death resulting from overstimulation could explain the decreased inflammation of GICs infected with wildtype, we did not observe any cell death, distress or monolayer disruption as a result of infection with *C. jejuni* strains over 2 – 48 hours. The cytokine profiles measured as

a result of insult with *C. jejuni* were likely dominated by factors produced by dendritic cells with some contribution from the epithelial cells; however, the exact contribution from each cell type is difficult to ascertain given the highly complex nature of the multicellular crosstalk. The concentrations of cytokines and chemokines measured in basolateral media lead us to believe that direct dendritic cell contact with *C. jejuni* cells is not a significant component of the inflammatory response, as exploratory experiments have shown inflammation in this case to be up to an order of magnitude higher (data not shown.)

Outer membrane vesicles (OMV) are produced by *C. jejuni* strains and considered a significant virulence factor for this organism. As an increase in OMV production during infection has been observed in other bacterial pathogens [46], it was surmised that they were present during our experiments on GICs. As such, we isolated OMVs from wildtype and Δ*pglE C. jejuni* in order to examine their role in infection and inflammation of the gut model. Activity-based protease profiling revealed a curious enrichment of proteases in glycosylation-impaired OMVs from Δ*pglE* capable of cleaving tight junction proteins. Analyses of the proteome of glycosylation-deficient *C. jejuni* have shown an enrichment in protein quality control machineries such as chaperones and proteases to mitigate protein misfolding and aggregration as a consequence of *pgl* loss of function, with one such example being serine protease HtrA.[10, 11] HtrA is known to degrade tight-junction proteins and has been identified as one of several serine proteases in OMVs of *C. jejuni* capable of rendering OMVs proteolytically active against occludin and E-cadherin.[37, 47] This suggests a global increase in cellular proteases in Δ*pglE C. jejuni* results in production of protease enriched OMVs. However, despite containing increased proteolytic power, the glycosylation state of isolated wildtype and Δ*pglE* OMV did not change the inherent immunogenicity on C2bbe1-only

GICs, suggesting that the exact role and mechanisms of OMVs in infection requires further characterization, perhaps using other readouts provided by the gut-immune model.

In conclusion, the tripartite design of the gut-immune model, containing a recapitulated human intestinal epithelium and innate immune component, allowed for examination of the various roles played by the gut monolayer, the mucus, and dendritic cells and highlighted the importance of proper N-glycan presentation in various aspects of *C. jejuni* infectivity. This system also provided an opportunity to study the physiological and inflammatory interplay between all of the host components when exposed to this pathogen and its virulence factors – such insights would have otherwise been unobservable in other culture models. We anticipate this system will be advantageously applied to the study of other enteropathogens or aspects of pathogenic interactions.

5.10 Materials and methods

General information

All common chemicals, gases and reagents were purchased from Sigma-Aldrich, AirGas, Fisher Scientific and VWR unless otherwise noted. Following is a list of the sources of other key reagents and expendable materials used in these studies: BBL Mueller-Hinton broth (BD and Co., Cat. No. 211443); Trimethoprim (Chem Impex, Cat. No. 01634); kanamycin sulfate (Teknova, Cat. No. K2151); Oxoid AGS plastic pouches (Invitro Diagnostics, Cat. No., AG0020C); sealing clips for plastic pouches (Invitro Diagnostics, Cat. No. AN0005C); culture tubes for bacterial cultures (17 x 100 mm) (VWR, Cat. No. 60818-689); Pierce BCA assay kit (Thermo Fisher Scientific, Cat. No. 23227); mouse anti-6x-His epitope tag monoclonal antibody (Thermo Fisher Scientific, Cat. No. MA1-21315); LI-COR IRDye 800CW goat anti-rabbit IgG (LI-COR, Cat. No. 926-32211); Licor IRDye 680LT goat anti-mouse IgG (LI-COR, Cat. No. 680 926-68020); LI-COR IRDye 800CW goat anti-mouse IgG (LI-COR, Cat. No. 926-32210); streptavidin-AlexaFluor 647 (Thermo Fisher Scientific, Cat. No. S32357); 75 cm^2 tissue culture flasks (Corning, Cat. No. 353136); trypsin-EDTA (Gibco, Cat. No. 252000-56); 12 mm Transwell inserts with 0.4 μm polyester membranes (Corning, Cat. No. 3460); trypsin (Corning, Cat. No. 25-051-CI); recombinant E-cadherin (R & D Systems, Cat. No. 8505-EC-050).

Tissue culture media compositions

Gut Cell Line Media: DMEM (Gibco, Cat. No. 11965), 10% fetal bovine serum (FBS, Atlanta Biologicals, Cat. No. S11150), 1x non-essential amino acids (NEAA, Gibco, Cat. No. 11140-148), 1x GlutaMAX (Gibco, Cat. No. 35050-061), 1% penicillin-streptomycin (P/S, Gibco, Cat. No. 15140-148.)

Gut Insert Seeding Media: Advanced DMEM (Gibco, Cat. No. 12491), 10% FBS, 1x GlutaMAX, 1% P/S.

Serum-Free Insert Medium: Advanced DMEM, 1x insulin-transferrin-selenium supplement (ITS, Roche, Cat. No. 11074547001), 1x GlutaMAX, 1% P/S. Apical Media: DMEM, ITS, 1x GlutaMAX, 1x NEAA and 1% P/S. Basal Media: Advanced DMEM, 1x ITS, 1x GlutaMAX, 0.85-1.25 mg/mL bovine serum albumin (BSA, Sigma-Aldrich, Cat. No. A9576) and 1% P/S.

Generalized Procedure A: Culture of *C. jejuni* strains

Samples from glycerol stocks of *C. jejuni* (NCTC 11168 or 11168Δ*pglE*) were streaked onto MH agar plates containing selection antibiotics (10 μg/mL trimethoprim for 11168 wildtype, 10 μg/mL trimethoprim and 20 μg/mL kanamycin sulfate for 11168Δ*pglE*). Plates were placed inside a plastic pouch (along with an atmospheric CO_2 indicator) and purged with a microaerophilic gas mixture (85% N_2, 10% CO_2, 5% O_2) multiple times until the correct concentration of CO_2 was reached. Sealed pouches were then incubated at 37 °C for 24 h, after which colonies were taken up in 3 mL of PBS broth. Cells taken from these master stocks of *C. jejuni* were utilized in the subsequent experiments. In our hands, an OD_{600} of 0.015 corresponded to 1 x 10^7 cfu/mL.

Generalized Procedure B: Maintenance and passaging of intestinal epithelial cell lines C2BBe1 and HT29-MTX

C2BBe1 (ATCC, passage 48-58) and HT29-MTX (Sigma, passage 20-30) were maintained with Gut Cell Line Media in separate 75 cm^2 flasks at 37 °C at 5% CO_2 and 95% humidity. Cells at 80-90% confluence were passaged by washing with PBS without calcium and magnesium (PBS-/-) and lifting with 0.25% trypsin-EDTA at 37°C. After 10 minutes, cells were manually

dissociated, quenched with media, and centrifuged at 300 *x g* for 5 min in 15 mL conical tubes. After, cells were re-suspended and quantified by Trypan Blue exclusion for continued passaging or seeding into Transwell membrane inserts according to *Generalized Procedures C* and *D*. Passages of C2BBe1 and HT29-MTX were seeded with 6 x10^5 and 2 x10^6 cells, respectively, every 6-7 days and were fed with Gut Cell Line Medium every 2-3 days.

Generalized Procedure C: Preparation and maintenance of polarized epithelial monolayers

Polarized epithelial monolayers were prepared as previously described [25]. Prior to seeding, Transwell membrane inserts were coated with 50 mg/mL rat tail collagen I (Corning, Cat No. 354236) in PBS-/- for 2 h at rt. Intestinal epithelial cells were collected according *Generalized Procedure B*, ensuring cells had been passaged twice post-thawing prior to use. For C2BBe1-only monolayers, after rinsing away collagen coating solution with Advanced DMEM, 1.12 x 10^5 C2BBe1 cells (10^5 cells/cm^2) were seeded in each insert in 500 µL of Gut Seeding Media to form C2BBe1 monolayers. For C2BBe1/HT29-MTX monolayers, cells seeded on inserts were a 9:1 mixture of the C2BBe1 and HT29-MTX cell lines. All inserts were maintained at 37 °C at 5% CO_2 and 95% humidity and fed every 2-3 days with 500 µL and 1.5 mL Gut Insert Seeding media in apical and basolateral chambers, respectively. From Day 7 on, inserts were fed every 2-3 days with Serum-Free Insert Medium. Inserts were cultured in this manner for 21-27 days before use in any experiments.

Generalized Procedure D: Immune cell isolation, differentiation and seeding

Peripheral blood mononuclear cells (PBMCs) were processed from Leukopak (STEMCELL Technologies, Cat. No. 70500) using the recommended protocol for cell isolation

without density centrifugation. Isolated cells were aliquoted and stored in liquid nitrogen. To derive dendritic cells, monocytes were isolated from thawed PBMCs using the EasySep Human Monocyte Enrichment Kit without CD16 Depletion (STEMCELL Technologies, Cat. No. 19058). The purified monocytes were seeded into a 24-well tissue culture treated plate at ~1.2 million cells/well in Advanced RPMI medium (Gibco, Cat. No. 12633-012) supplemented with 1X GlutaMax, 1% P/S, 50 ng/mL GM-CSF (Biolegend, Cat. No. 572903), 35 ng/mL IL-4 (Biolegend Cat. No. 574004) and 10 nM retinoic acid. After 7 days of differentiation (at day 19-20 of the epithelial culture) at 37 °C at 5% CO_2 and 95% humidity, mature dendritic cells were lifted using PBS-/- and Accutase (Gibco, Cat. No. A11105-01) and seeded onto the basolateral side of inverted gut Transwell inserts at 10^5/insert, allowing 2 h for attachment.

Preparation of retinoic acid stock solution

Retinoic acid was dissolved in 200 proof EtOH and concentration was determined by measuring absorbance at 350 nm (ε_{max} (EtOH) = 44,300 $M^{-1}cm^{-1}$). The solution was diluted to 50μM in PBS-/- with 1% BSA. Frozen stocks were kept at -20°C for up to 3 months.

Generalized Procedure E: Preparation of gut-immune co-cultures (GICs)

To begin, intestinal epithelia were prepared as described in *Generalized Procedure C*. At Day 13, monocytes were isolated and differentiated as in *Generalized Procedure D*. On Day 20 post-monolayer seeding (7 days post-monocyte isolation), dendritic cells were recovered and seeded onto the basal membrane of the gut cultures. From this point, gut-immune co-cultures were maintained with 500μL Apical Media and 1.5 mL Basal Media in their respective chambers. GICs were maintained at 37 °C at 5% CO_2 and 95% humidity and were utilized in experiments 21 days

post-monolayer seeding. GICs were rinsed with Serum-Free Maintenance Media without antibiotics prior to inoculation with *C. jejuni.*

Generalized Procedure F: Treatment of GICs with strains and mutants of *C. jejuni*

The apical compartment of GICs containing gut epithelial monolayers or co-cultures were inoculated with *C. jejuni* strains and mutants ahead of experiments characterizing bacterial adhesion and internalization under various conditions and ahead of immunoassays and other measures of epithelial health. To begin, *C. jejuni* stocks were prepared according to General Procedure A and aliquots sufficient to give an OD = 1.0 in 1 mL were taken into microcentrifuge tubes. Cells were centrifuged at 16,000 *x g* for two minutes to give a pellet (pink in color,) the supernatant removed and the pellet resuspended in 1 mL DPBS. Apical growth media without P/S was warmed to 37 °C and inoculated with washed *C. jejuni* to give a solution with initial OD of 0.03 (multiplicity of infection: 10) or 0.15 (multiplicity of infection: 50). GICs to be treated were provided fresh basal growth media right before innoculation. Growth media was aspirated from the apical face of each GIC to be treated and replaced with 500 µL of media containing *C. jejuni* strains.

Measurement of bacterial adhesion adhesion to GIC epithelia

GICs inoculated with wildtype *C. jejuni* NCTC 11168 or 11168Δ*pglE* according to General Procedure F were incubated at 37 °C for 24 h, after which inserts were visually inspected to confirm lack of epithelial cell rounding or distress. Apical compartments were washed four times with 500 µL DPBS to remove non-adherent bacteria. GICs were given 500 µL DPBS with 0.1% Triton and placed on an orbital shaker shaking at 300 rpm for 20 min to lyse epithelial monolayers.

Lysates containing bacteria from monolayer surfaces and cell interiors were serially diluted in MH broth and 50 μL plated on MH agar containing appropriate selection antibiotics for colony counting.

Measurement of invasion by C. jejuni into epithelial monolayers by gentamycin treatment

GICs inoculated with wildtype *C. jejuni* NCTC 11168, or 11168Δ*pglE* according to General Procedure C were incubated at 37 °C for 24 h, after which inserts were visually inspected and monolayer washed 3 times with 500 μL DPBS to remove non-adherent bacteria. GICs were then treated with 500 μL apical media containing 210 μM gentamycin sulfate and incubated at 37 °C for 45 min. GICs were washed four times with 500 μL DPBS to remove bacteria and residual antibiotic, followed by treatment with 500 μL DPBS with 0.1% Triton. GICs were placed on an orbital shaker shaking at 300 rpm for 20 min to lyse and lysates were serially diluted in MH broth and plated on MH agar containing appropriate selection antibiotics for colony counting.

Secreted mucin quantification by Alcian blue colorimetric assay

Secreted mucins in apical media collected from experiments were quantified *via* a colorimetric assay adapted from Hall et al [48]. Briefly, collected samples stored in low-binding microcentrifuge tubes were spun at 6,000 *x g* for 5 min to pellet bacteria and/or cell debris prior to analysis. Supernatants were analyzed immediately or flash frozen with liquid nitrogen for storage at -80 °C. Samples were mixed with Alcian Blue solution (Richard Allan Scientific Co., San Diego, CA) in a ratio of 3:1 and incubated for 2 h. After, samples in 96-well plates were centrifuged at 1,640 *x g* for 30 min at rt. Supernatants were removed by inversion and the resulting pellet resuspended twice with a 40% ethanol/60% 0.1M sodium acetate buffer with 25 mM $MgCl_2$, pH 5.8, centrifuging 10 min as above after each resuspension. Washed pellets were fully dissolved in

10% SDS solution and absorption was measured at 620 nm on a plate reader (Spectramax m3/m2e). Mucin from bovine submaxillary glands (Sigma, Cat. No. M3895) served as a standard and was prepared and analyzed identically in parallel with experimental samples.

Fluorometric quantification of paracellular permeability in epithelial monolayers

Paracellular permeability of epithelial monolayers was quantified utilizing Lucifer Yellow reagent according to manufacturer's instructions. The epithelial monolayers of GICs were washed with transport buffer (HBSS with $CaCl_2$, $MgCl_2$ supplemented with 10 mM HEPES, pH 7.4) and 500 µl of 100 mM Lucifer Yellow in transport buffer was added to the apical compartment of each insert. Transport buffer (1.5 mL) was added to the basal compartment and inserts were incubated at 37 °C in 5% CO_2 for 1-2 h with shaking. Following incubation, inserts were removed and 150 µl of each sample transferred to a 96-well plate and fluorescence measured (λ_{ex}= 485 nm, λ_{em}= 530 nm.) Standard curves of Lucifer Yellow alone in transport buffer were generated to determine fluorophore concentrations in experimental samples. Apparent permeability coefficients were calculated according to manufacturer's instructions.

Quantification of cytokine/chemokine release by multiplexed immunoassay

Cytokine/chemokine amounts in apical or basal media samples collected from experiments were measured using Bio-Plex Pro Human Chemokine Panel, 40-plex (Bio-Rad Laboratories, Inc., Cat. No. 171AK99MR2) according to manufacturer's instructions. Collected samples stored in low-binding microcentrifuge tubes were spun at 6,000 *x g* for 5 min to pellet bacteria and/or cell debris prior to analysis. Supernatants were analyzed immediately or flash frozen with liquid nitrogen for storage at -80 °C. BSA was added to all samples (final concentration 5 mg/mL) in order to minimize non-specific binding to beads. Multiple dilutions of each sample were analyzed

to ensure to ensure measurements were within the linear dynamic range of the assay. Assays were performed on a Bio-Plex 3D Suspension Array System and data collected using xPONENT for FLEXMAP 3D software, version 4.2 (Luminex Corporation, Austin, TX). Concentrations of analytes were determined utilizing a standard curve generated by fitting a 5-parameter logistic regression of mean fluorescence on known concentrations of each analyte with Bio-Plex Manager software.

Isolation of outer membrane vesicles from NCTC 11168 and 11168 Δ*pglE*

NCTC 11168 and NCTC 11168Δ*pglE* from 24 h plates were used to inoculate 1 L of Mueller Hinton media. Cultures were grown overnight under microaerophilic conditions at 37 °C with shaking to mid log phase. OMV isolation was adapted from Valguarnera et al [49]. Briefly, bacteria were pelleted for 30 min at 3,600 *x g* and the supernatant filtered through 0.22 μm membrane. The filtrate was ultracentrifuged for 2 h at 100,000 *x g*. The vesicles were resuspended in PBS and ultracentrifuged an additional 2 h at 100,000 *x g*. Fresh PBS was added and the vesicles were ultracentrifuged overnight at 200,000 *x g*. The supernatant was removed and the vesicles were brought up in 2 ml in PBS and the protein content was assessed by BCA assay.

Characterization of *C. jejuni* OMV by SDS-PAGE and immunoblotting

For analysis of total protein content, 100 μg of purified OMV was mixed with 20 μl 2x Laemmli buffer and boiled for 10 min. The samples were separated by 10% SDS-PAGE and stained with Instant Blue overnight. For N-glycan analysis, 20 μg of the OMV was separated by 10% SDS PAGE, transferred to nitrocellulose membrane by wet transfer, and blocked with 5% BSA in TBS for 1 h. The membrane was blotted with rabbit anti-N-glycan antibody (1:10,000

dilution) overnight at 4 °C, followed by secondary anti-rabbit antibody incubation 1:10,000 dilution for 2 h. Blots were visualized by LI-COR Odyssey scanner.

Serine protease profiling of *C. jejuni* OMV with FP-biotin

Active serine protease profiling was performed using 100 μg purified OMV as previously described [50]. Samples were incubated with 2 μM **FP-biotin** (10-(fluoroethoxyphosphinyl)-*N*-(biotinamidopentyl)decanamide) in a final volume of 100 μl in PBS pH 7.3 for 30 minutes at room temperature. Four μg of trypsin was included as a positive control for serine protease activity. The samples were evaporated to dryness and reconstituted in 20 μl 1x Laemmli buffer and boiled for 10 min. The samples were separated by SDS PAGE (10% polyacrylamide gels) and transferred to nitrocellulose membrane by wet transfer method. The blot was blocked overnight in 10% wt/vol dry milk in TBS then probed with streptavidin-Alexafluor 647 at a 1:1,000 dilution for one h and visualized with LI-COR Odyssey scanner.

E-cadherin cleavage assay with *C. jejuni* OMV

The cleavage of recombinant E-cadherin was determined as described previously.[47] Briefly, 10 μg of purified OMV was incubated with 1 μg of recombinant human E-cadherin containing a C-terminal 6xHis tag in PBS at 37 °C for 16 h. The reactions were mixed 1:1 with 2 x Laemmli buffer and boiled for 10 minutes. The samples were separated by SDS PAGE and transferred to nitrocellulose membrane by wet transfer. The membrane was blocked with 5% BSA in TBS for one h then incubated with mouse anti-His antibody overnight at 4 °C. The blot was incubated with anti-mouse secondary antibody at a 1:10,000 dilution for one h then imaged with LI-COR Odyssey scanner.

Barrier function characterization via transepithelial electrical resistance (TEER)

TEER measurement across epithelial monolayers in GIC transwells was carried out using EndOhm-12 chambers and an EVOM2 meter (World Precision Instruments, Sarasota, FL). Transwells and EndOhm chamber were maintained at 37 °C during all measurements to minimize variability. TEER of C2BBe1-only GICs in transwells (Day 21) was measured at the start of the experiment. Next, the apical media of the transwells was aspirated and replaced with 500 μL apical media alone or 500 μL containing 10 or 100 μg of wildtype or Δ*pglE* OMV. Transwells were incubated at 37 °C for 24 h, after which apical media was sampled and TEER measured a second time. Sampled media was flash frozen in liquid nitrogen for immunoassay analysis.

Statistical Analysis

All experiments were performed at least in triplicate, with additional replicates, technical replicates and independent biological replicates noted in each figure caption. Statistical analyses of data sets were performed using GraphPad Prism software. Statistical significance ($P < 0.05$) and P values were calculated utilizing Student's t-tests or 1-way ANOVA as appropriate.

5.11 Funding sources

National Institutes of Health (R01-GM097241 to B.I., R01EB021908 to L.G.G.); Defense Advanced Research Projects Agency (DARPA) (W911NF-12-2-0039 to L.G.G); National Institutes of Health Interdepartmental Biotechnology Training Program (T32-GM008334 to E.M.W.).

5.12 Acknowledgements

We thank Prof. Christine Szymanski for her valued advice and productive scientific discussions. The authors also thank the Szymanski laboratory at the Complex Carbohydrate Research Center (Athens, Georgia) and Alberta Glycomics Centre (Edmonton, Alberta, Canada) for the generous gifts of the 11168Δ*pglE C. jejuni* strain, free oligosaccharide (FOS) and anti-*N*-glycan antibody. The authors also thank Prof. Ben Cravatt and his laboratory at Scripps Research Institute (La Jolla, CA) for providing a sample of FP-biotin, and Dr. Enrique Valguarnera and Prof. Mario Feldman (Washington University School of Medicine, St. Louis, MO) for technical advice on OMV preparation.

5.13 References

1. Kaakoush, N. O., Castano-Rodriguez, N., Mitchell, H. M., and Man, S. M. (2015) Global Epidemiology of Campylobacter Infection, *Clin. Microbiol. Rev. 28*, 687-720.

2. Ternhag, A., Torner, A., Svensson, A., Giesecke, J., and Ekdahl, K. (2005) Mortality Following Campylobacter Infection: A Registry-Based Linkage Study, *BMC Infect. Dis. 5*, 70.

3. Young, K. T., Davis, L. M., and Dirita, V. J. (2007) Campylobacter Jejuni: Molecular Biology and Pathogenesis, *Nat. Rev. Microbiol. 5*, 665-679.

4. Dorrell, N., Mangan, J. A., Laing, K. G., Hinds, J., Linton, D., Al-Ghusein, H., Barrell, B. G., Parkhill, J., Stoker, N. G., Karlyshev, A. V., Butcher, P. D., and Wren, B. W. (2001) Whole Genome Comparison of Campylobacter Jejuni Human Isolates Using a Low-Cost Microarray Reveals Extensive Genetic Diversity, *Genome Res. 11*, 1706-1715.

5. Watson, R. O., and Galan, J. E. (2008) Campylobacter Jejuni Survives within Epithelial Cells by Avoiding Delivery to Lysosomes, *PLoS Pathog. 4*, e14.

6. Dasti, J. I., Tareen, A. M., Lugert, R., Zautner, A. E., and Groß, U. (2010) Campylobacter Jejuni: A Brief Overview on Pathogenicity-Associated Factors and Disease-Mediating Mechanisms, *Int. J. Med. Microbiol. 300*, 205-211.

7. Elmi, A., Watson, E., Sandu, P., Gundogdu, O., Mills, D. C., Inglis, N. F., Manson, E., Imrie, L., Bajaj-Elliott, M., Wren, B. W., Smith, D. G. E., and Dorrell, N. (2012) Campylobacter Jejuni Outer Membrane Vesicles Play an Important Role in Bacterial Interactions with Human Intestinal Epithelial Cells, *Infect. Immun. 80*, 4089-4098.

8. Backert, S., Bernegger, S., Skórko-Glonek, J., and Wessler, S. (2018) Extracellular Htra Serine Proteases: An Emerging New Strategy in Bacterial Pathogenesis, *Cell. Microbiol. 20*, e12845.

9. Lu, Q., Li, S., and Shao, F. (2015) Sweet Talk: Protein Glycosylation in Bacterial Interaction with the Host, *Trends Microbiol. 23*, 630-641.

10. Cain, J. A., Dale, A. L., Niewold, P., Klare, W. P., Man, L., White, M. Y., Scott, N. E., and Cordwell, S. J. (2019) Proteomics Reveals Multiple Phenotypes Associated with N-Linked Glycosylation in Campylobacter Jejuni, *Mol. Cell. Proteomics 18*, 715-734.

11. Abouelhadid, S., North, S. J., Hitchen, P., Vohra, P., Chintoan-Uta, C., Stevens, M., Dell, A., Cuccui, J., and Wren, B. W. (2019) Quantitative Analyses Reveal Novel Roles for N-Glycosylation in a Major Enteric Bacterial Pathogen, *mBio 10*, e00297-00219.

12. Jones, M. A., Marston, K. L., Woodall, C. A., Maskell, D. J., Linton, D., Karlyshev, A. V., Dorrell, N., Wren, B. W., and Barrow, P. A. (2004) Adaptation of Campylobacter Jejuni Nctc11168 to High-Level Colonization of the Avian Gastrointestinal Tract, *Infect. Immun. 72*, 3769-3776.

13. Kelly, J., Jarrell, H., Millar, L., Tessier, L., Fiori, L. M., Lau, P. C., Allan, B., and Szymanski, C. M. (2006) Biosynthesis of the N-Linked Glycan in Campylobacter Jejuni and Addition onto Protein through Block Transfer, *J. Bacteriol. 188*, 2427-2434.

14. Szymanski, C. M., Burr, D. H., and Guerry, P. (2002) Campylobacter Protein Glycosylation Affects Host Cell Interactions, *Infect. Immun. 70*, 2242-2244.

15. MacCallum, A. J., Harris, D., Haddock, G., and Everest, P. H. (2006) Campylobacter Jejuni-Infected Human Epithelial Cell Lines Vary in Their Ability to Secrete Interleukin-8 Compared to in Vitro-Infected Primary Human Intestinal Tissue, *Microbiology 152*, 3661-3665.

16. Friis, L. M., Pin, C., Pearson, B. M., and Wells, J. M. (2005) In Vitro Cell Culture Methods for Investigating Campylobacter Invasion Mechanisms, *J. Microbiol. Methods 61*, 145-160.

17. Bahrami, B., Macfarlane, S., and Macfarlane, G. T. (2011) Induction of Cytokine Formation by Human Intestinal Bacteria in Gut Epithelial Cell Lines, *J. Appl. Microbiol. 110*, 353-363.

18. Guerry, P., Szymanski, C. M., Prendergast, M. M., Hickey, T. E., Ewing, C. P., Pattarini, D. L., and Moran, A. P. (2002) Phase Variation of Campylobacter Jejuni 81-176 Lipooligosaccharide Affects Ganglioside Mimicry and Invasiveness in Vitro, *Infect. Immun. 70*, 787-793.

19. Backert, S., Boehm, M., Wessler, S., and Tegtmeyer, N. (2013) Transmigration Route of Campylobacter Jejuni across Polarized Intestinal Epithelial Cells: Paracellular, Transcellular or Both?, *Cell Commun. Signal. 11*, 72.

20. Hu, L., Tall, B. D., Curtis, S. K., and Kopecko, D. J. (2008) Enhanced Microscopic Definition of Campylobacter Jejuni 81-176 Adherence to, Invasion of, Translocation across, and Exocytosis from Polarized Human Intestinal Caco-2 Cells, *Infect. Immun. 76*, 5294-5304.

21. Ferrero, R. L., and Lee, A. (1988) Motility of Campylobacter Jejuni in a Viscous Environment: Comparison with Conventional Rod-Shaped Bacteria, *J. Gen. Microbiol. 134*, 53-59.

22. Szymanski, C. M., King, M., Haardt, M., and Armstrong, G. D. (1995) Campylobacter Jejuni Motility and Invasion of Caco-2 Cells, *Infect. Immun. 63*, 4295-4300.

23. Alemka, A., Corcionivoschi, N., and Bourke, B. (2012) Defense and Adaptation: The Complex Inter-Relationship between Campylobacter Jejuni and Mucus, *Front. Cell. Infect. Microbiol. 2*, 15.

24. Alemka, A., Clyne, M., Shanahan, F., Tompkins, T., Corcionivoschi, N., and Bourke, B. (2010) Probiotic Colonization of the Adherent Mucus Layer of Ht29mtxe12 Cells Attenuates Campylobacter Jejuni Virulence Properties, *Infect. Immun. 78*, 2812-2822.

25. Chen, W. L. K., Edington, C., Suter, E., Yu, J., Velazquez, J. J., Velazquez, J. G., Shockley, M., Large, E. M., Venkataramanan, R., Hughes, D. J., Stokes, C. L., Trumper, D. L., Carrier, R. L., Cirit, M., Griffith, L. G., and Lauffenburger, D. A. (2017) Integrated Gut/Liver Microphysiological Systems Elucidates Inflammatory Inter-Tissue Crosstalk, *Biotechnol. Bioeng. 114*, 2648-2659.

26. Edington, C. D., Chen, W. L. K., Geishecker, E., Kassis, T., Soenksen, L. R., Bhushan, B. M., Freake, D., Kirschner, J., Maass, C., Tsamandouras, N., Valdez, J., Cook, C. D., Parent, T., Snyder, S., Yu, J., Suter, E., Shockley, M., Velazquez, J., Velazquez, J. J., Stockdale,

L., Papps, J. P., Lee, I., Vann, N., Gamboa, M., LaBarge, M. E., Zhong, Z., Wang, X., Boyer, L. A., Lauffenburger, D. A., Carrier, R. L., Communal, C., Tannenbaum, S. R., Stokes, C. L., Hughes, D. J., Rohatgi, G., Trumper, D. L., Cirit, M., and Griffith, L. G. (2018) Interconnected Microphysiological Systems for Quantitative Biology and Pharmacology Studies, *Sci. Rep. 8*, 4530.

27. Louwen, R., Nieuwenhuis, E. E., van Marrewijk, L., Horst-Kreft, D., de Ruiter, L., Heikema, A. P., van Wamel, W. J., Wagenaar, J. A., Endtz, H. P., Samsom, J., van Baarlen, P., Akhmanova, A., and van Belkum, A. (2012) Campylobacter Jejuni Translocation across Intestinal Epithelial Cells Is Facilitated by Ganglioside-Like Lipooligosaccharide Structures, *Infect. Immun. 80*, 3307-3318.

28. Tsamandouras, N., Chen, W. L. K., Edington, C. D., Stokes, C. L., Griffith, L. G., and Cirit, M. (2017) Integrated Gut and Liver Microphysiological Systems for Quantitative in Vitro Pharmacokinetic Studies, *AAPS J 19*, 1499-1512.

29. Linton, D., Allan, E., Karlyshev, A. V., Cronshaw, A. D., and Wren, B. W. (2002) Identification of N-Acetylgalactosamine-Containing Glycoproteins Peb3 and Cgpa in Campylobacter Jejuni, *Mol. Microbiol. 43*, 497-508.

30. Hu, L., Bray, M. D., Osorio, M., and Kopecko, D. J. (2006) Campylobacter Jejuni; Induces Maturation and Cytokine Production in Human Dendritic Cells, *Infect. Immun. 74*, 2697.

31. Espey, M. G. (2013) Role of Oxygen Gradients in Shaping Redox Relationships between the Human Intestine and Its Microbiota, *Free Radic. Biol. Med. 55*, 130-140.

32. DeAngelis, I., and Turco, L. (2011) Caco-2 Cells as a Model for Intestinal Absorption, *Curr. Protoc. Toxicol. 47*, 20.26.21-20.26.15.

33. Dwivedi, R., Nothaft, H., Reiz, B., Whittal, R. M., and Szymanski, C. M. (2013) Generation of Free Oligosaccharides from Bacterial Protein N-Linked Glycosylation Systems, *Biopolymers 99*, 772-783.

34. Jan, A. T. (2017) Outer Membrane Vesicles (Omvs) of Gram-Negative Bacteria: A Perspective Update, *Front. Microbiol. 8*, 1053.

35. Schwechheimer, C., and Kuehn, M. J. (2015) Outer-Membrane Vesicles from Gram-Negative Bacteria: Biogenesis and Functions, *Nat. Rev. Microbiol. 13*, 605-619.

36. O'Donoghue, E. J., and Krachler, A. M. (2016) Mechanisms of Outer Membrane Vesicle Entry into Host Cells, *Cell Microbiol 18*, 1508-1517.

37. Jang, K.-S., Sweredoski, M. J., Graham, R. L. J., Hess, S., and Clemons, W. M. (2014) Comprehensive Proteomic Profiling of Outer Membrane Vesicles from Campylobacter Jejuni, *J. Proteomics 98*, 90-98.

38. Lindmark, B., Rompikuntal, P. K., Vaitkevicius, K., Song, T., Mizunoe, Y., Uhlin, B. E., Guerry, P., and Wai, S. N. (2009) Outer Membrane Vesicle-Mediated Release of Cytolethal Distending Toxin (Cdt) from Campylobacter Jejuni, *BMC Microbiol. 9*, 220-220.

39. Kidd, D., Liu, Y., and Cravatt, B. F. (2001) Profiling Serine Hydrolase Activities in Complex Proteomes, *Biochemistry 40*, 4005-4015.

40. Sicard, J.-F., Le Bihan, G., Vogeleer, P., Jacques, M., and Harel, J. (2017) Interactions of Intestinal Bacteria with Components of the Intestinal Mucus, *Front. Cell. Infect. Microbiol. 7*, 387-387.

41. Tu, Q. V., McGuckin, M. A., and Mendz, G. L. (2008) Campylobacter Jejuni Response to Human Mucin Muc2: Modulation of Colonization and Pathogenicity Determinants, *J. Med. Microbiol. 57*, 795-802.

42. Karlyshev, A. V., Everest, P., Linton, D., Cawthraw, S., Newell, D. G., and Wren, B. W. (2004) The Campylobacter Jejuni General Glycosylation System Is Important for Attachment to Human Epithelial Cells and in the Colonization of Chicks, *Microbiology 150*, 1957-1964.

43. Hendrixson, D. R., and DiRita, V. J. (2004) Identification of Campylobacter Jejuni Genes Involved in Commensal Colonization of the Chick Gastrointestinal Tract, *Mol. Microbiol. 52*, 471-484.

44. Bakhiet, M., Al-Salloom, F. S., Qareiballa, A., Bindayna, K., Farid, I., and Botta, G. A. (2004) Induction of A and B Chemokines by Intestinal Epithelial Cells Stimulated with Campylobacter Jejuni, *J. Infect. 48*, 236-244.

45. Hu, L., Bray, M. D., Geng, Y., and Kopecko, D. J. (2012) Campylobacter Jejuni-Mediated Induction of Cc and Cxc Chemokines and Chemokine Receptors in Human Dendritic Cells, *Infect. Immun. 80*, 2929-2939.

46. Bauwens, A., Kunsmann, L., Marejkova, M., Zhang, W., Karch, H., Bielaszewska, M., and Mellmann, A. (2017) Intrahost Milieu Modulates Production of Outer Membrane Vesicles, Vesicle-Associated Shiga Toxin 2a and Cytotoxicity in Escherichia Coli O157:H7 and O104:H4, *Environ. Microbiol. Rep. 9*, 626-634.

47. Elmi, A., Nasher, F., Jagatia, H., Gundogdu, O., Bajaj-Elliott, M., Wren, B., and Dorrell, N. (2016) Campylobacter Jejuni Outer Membrane Vesicle-Associated Proteolytic Activity Promotes Bacterial Invasion by Mediating Cleavage of Intestinal Epithelial Cell E-Cadherin and Occludin, *Cell. Microbiol. 18*, 561-572.

48. Hall, R. L., Miller, R. J., Peatfield, A. C., Richardson, P. S., Williams, I., and Lampert, I. (1980) A Colorimetric Assay for Mucous Glycoproteins Using Alcian Blue *Biochem. Soc. Trans. 8*, 72.

49. Valguarnera, E., and Feldman, M. F. (2017) Glycoengineered Outer Membrane Vesicles as a Platform for Vaccine Development, *Methods Enzymol. 597*, 285-310.

50. Liu, Y., Patricelli, M. P., and Cravatt, B. F. (1999) Activity-Based Protein Profiling: The Serine Hydrolases, *Proc. Natl. Acad. Sci. 96*, 14694-14699.

Chapter 6: Future directions – bacterial glycan analysis probes

6.1 Future direction

This thesis described the development of a glycan-binding protein (GBP) generation platform and its application to create GBPs that recognize two mammalian glycans, the TF antigen and sialic acid. That the platform was able to develop GBPs to both a neutral disaccharide and a negatively charged nine-carbon monosaccharide demonstrates the versatility of the method. The overall goal of the platform is to create GBPs for glycans with few or no existing GBPs for their study. One area that can greatly benefit from more GBP reagents is bacterial glycobiology. The bacterial kingdom is estimated to produce upwards of 800 individual monosaccharides, while only 36 building blocks are needed to construct 75% of mammalian oligosaccharides in the Glycosciences.de databank.[1, 2] Bacterial glycans are important from a human health perspective. Chapter 5 of this thesis described the influence of the N-linked heptasaccharide on *Campylobacter jejuni* virulence when applied to a gut-immune co-culture model. This is just one example of the many pathogenic bacteria with glycans involved in virulence. Reagents for identification and characterization of bacterial glycans and glycoconjugates are needed. Existing GBPs to recognize the huge diversity of bacterial glycan structures are unavailable. For this reason, application of the GBP development platform described in this thesis toward bacterial glycans is an exciting future direction of the project.

6.2 Bacterial glycans

Bacterial glycans are often secreted or displayed on the cell surface as exopolysaccharides, capsular polysaccharides, lipooligosaccharides, glycolipids, wall teichoic acids, peptidoglycan, and glycoproteins. These glycans play protective and structural roles, but many are also important for virulence or colonization of host organisms. Protein O- and N-linked glycosylation is important for virulence in many pathogens, with adhesion factors, flagella, and pilin proteins often bearing

glycans.[3] Lipid-linked glycans such as the lipopolysaccharide/lipooligosaccharide (LPS/LOS) make up a significant portion of the outer membrane of Gram-negative bacteria and are important for structural integrity of the organism, but can also aid the bacterium in immune system evasion or lead to the development of septic shock in humans.[4] Bacteria can also secrete polysaccharides that form a capsule around the bacterium and protect it from the host immune system, or produce a matrix allowing bacteria to develop antibiotic resistant biofilms.[5]

Three interesting bacterial glycans for future application of the GBP engineering method are the N-linked heptasaccharide of *C. jejuni*, the bacterial nonulosonic acid pseudaminic acid (Pse), and the lipoarabinomannan (LAM) of *Mycobacterium tuberculosis* (**Figure 6-1**).

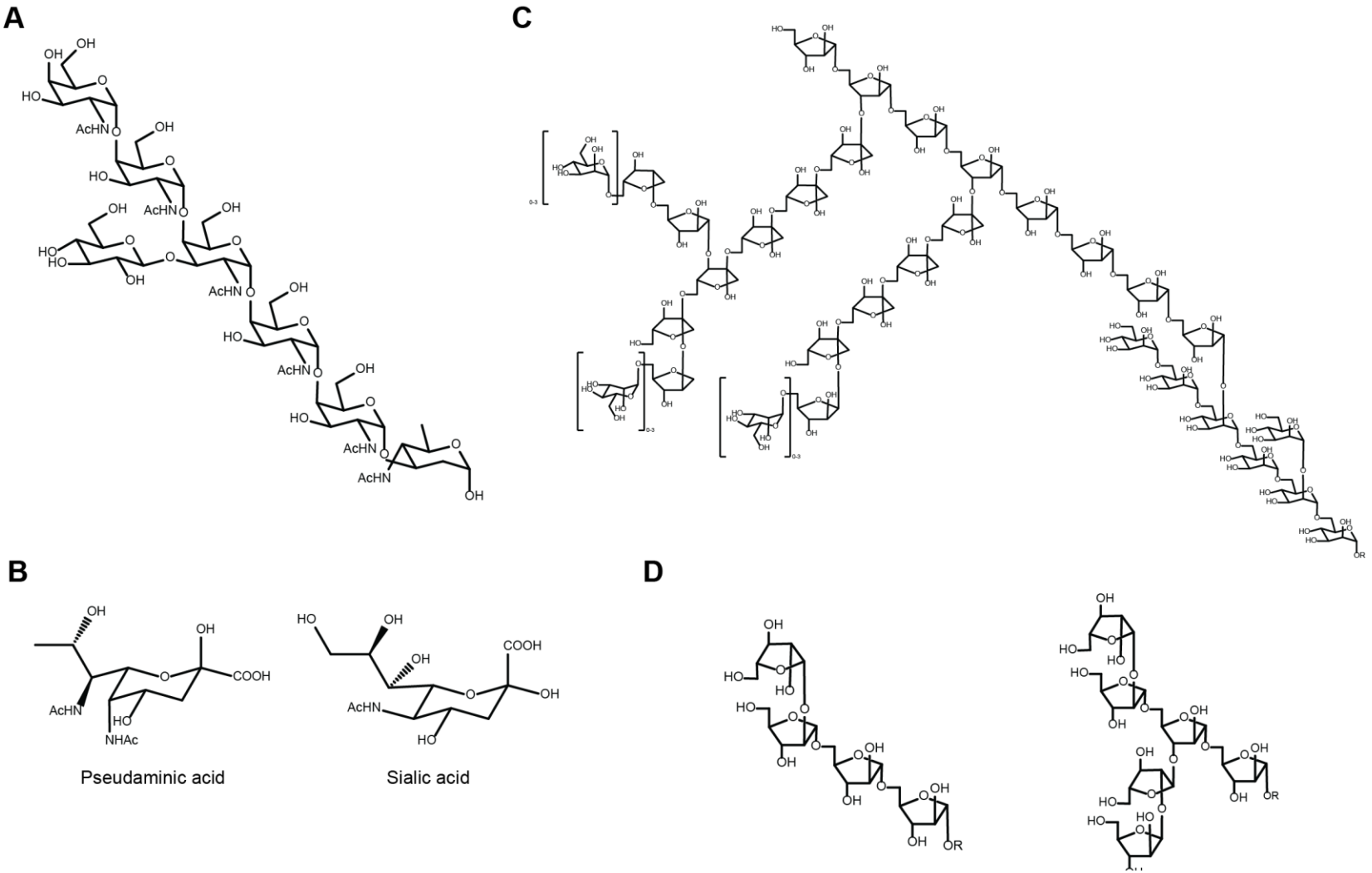


Figure 6-1. Potential bacterial glycans for glycan-binding protein generation.
A) N-linked heptasaccharide of *C. jejuni* consisting of GalNAc-α1,4-GalNAc-α1,4-[Glcβ GalNAc-α1,3-Bac. B) Pseudaminic acid structure compared to sialic acid structure. C) Full *M. tuberculosis*. D) LAM terminal branch fragments Ara*f*4 and Ara*f*6.

6.2.1 *Campylobacter jejuni* heptasaccharide

C. jejuni is a Gram-negative pathogen and is a leading cause of gastroenteritis and diarrheal disease.[6, 7] *C. jejuni* contains a well-characterized system for N-linked protein glycosylation. The N-glycan is a heptasaccharide with the structure GalNAc-α1,4-GalNAc-α1,4-[Glcβ1,3]-GalNAc-α1,4-GalNAc-α1,4-GalNAc-α1,3-Bac (**Figure 6-1A**), where Bac is the bacterial sugar bacillosamine (2,4-diacetamido-2,4,6-trideoxyglucose).[8] This N-glycan modifies more than 60 periplasmic and membrane proteins, and has been shown to impact proteome stability, protein quality control, stress response, nutrient uptake, chemotaxis, cell morphology, and virulence in *C. jejuni*.[9, 10] Chapter 5 of this thesis demonstrated the role of the N-glycan in *C. jejuni* virulence as loss of glycosylation decreased adhesion and invasion of bacteria 100-fold in a human gut-immune co-culture model. *C. jejuni* N-glycan antibodies have been developed, but a smaller GBP capable of selectively binding "signature" sugar epitopes the N-glycan is desirable making this a worthwhile target for Sso7d-based GBPs.[11]

6.2.2 Pseudaminic acid (Pse)

Pseudaminic acid (Pse, 5,7-diacetamido-3,5,7,9-tetradeoxy-L-glycero-α-L-manno-nonulosonic acid) is a member of the 9-carbon α-keto acid sugars that includes sialic acid (Neu5Ac) (**Figure 6-1B**). Pse is a bacterial sugar that was initially discovered in the LPS of pathogens *Pseudomonas aeruginosa* and *Shigella boydii*.[12] In *H. pylori* and *C. jejuni*, Pse modifies Ser and Thr residues of the flagellin proteins FlaA and FlaB.[13-15] Pse is important for virulence in these organisms as O-glycosylation of these flagellar proteins are necessary for motility, and motility is needed to initiate infection.[16, 17] Two of the "ESKAPE" pathogens for which multidrug resistance is becoming widespread, *Acinetobacter baumannii* and *Enterobacter* species, contain Pse in their exopolysaccharide.[18, 19] The presence of Pse on so many pathogenic bacteria has made

it an attractive target for therapeutics and diagnostics, and some progress has been made towards the production of Pse-binding antibodies.[20] Development of a Pse-binding Sso7d variant would be useful for the field and would complement the sialic acid-binding Sso7d variant developed in Chapter 2 of this thesis.

6.2.3 Lipoarabinomannan of *Mycobacterium tuberculosis*

Mycobacteria are a unique bacterial genus that have a thick, waxy cell wall that makes them particularly tolerant to antibiotic treatment. The cell envelope contains many different glycan species including lipoarabinomannan (LAM). LAM is a glycolipid and virulence factor of *M. tuberculosis* consisting of a linear mannan backbone made up of α1,6-mannose, D-arabinan core made up of α1,5-D-arabinofuranose (Ara*f*), and branching arabinan α1,3-linked to the arabinan core (**Figure 6-1C**). The non-reducing ends of these branches are often linear Ara*f*4 or branched Araf6 (**Figure 6-1D**), and are often capped with 1,3-linked mannose residues.[21] LAM is considered a virulence factor primarily for its role in modulating phagocyte function, and is the focus of many TB diagnostic efforts as LAM is released from metabolically active or degrading cells during infection and excreted in soluble form in urine.[21] Urine-based TB diagnostics are highly desirable because sample collection is easy and LAM-binding assays are fast and inexpensive compared to current diagnostic tests like radiography, sputum smears, and culturing of *M. tuberculosis*. Development of LAM antibodies for TB diagnostics is an active field of research, but the sensitivity of the antibody approach has been an issue.[22] Notably, in the context of the current thesis achievements, Sso7d paper-based diagnostics have been demonstrated to be superior to antibody paper-based diagnostics in terms of cost of production and protein stability.[23, 24] Recent studies have also found that fusion of an Sso7d variant to a cellulose-binding domain enabled high-

density immobilization to cellulose that improves target capture, and could overcome sensitivity issues with urine-based LAM diagnostics.[23]

6.3 Bacterial GBP generation

Bacterial GBP generation will have the following workflow: selection of bacterial glycan of interest, design of multivalent glycan ligand, GBP evolution by yeast surface display, and conjugation of tags for application of the evolved GBP (**Figure 6-2**). Creation of multivalent glycan ligands is required to utilize the GBP generation platform developed in this thesis. The glycan of interest must be obtained in pure form. This can be done by isolation of the bacterial glycan of interest from the organism, or *in vitro* assembly synthetically, enzymatically, or a combination of the two. For the *C. jejuni* heptasaccharide, there are protocols for isolation of free oligosaccharide that exists in the periplasm of *C. jejuni*, and protocols for *in vitro* assembly of the heptasaccharide chemoenzymatically or by chemical synthesis.[25-27] Pseudaminic acid can also be produced chemoenzymatically or by total synthesis.[28, 29] LAM is a large molecule so total chemical synthesis is not possible. However, various LAM fragments consisting of terminal branches, arabinan core, and core mannan have been synthesized for use on a TB glycan array.[30] These fragments would be superior to full length LAM for GBP development as the epitope is more defined. Finally, the glycan will need to be modified and attached to a support such as a peptide for multivalent display. Methods for installation of an azido-linker and click reactions to an alkyne-modified peptide are currently being developed in the Imperiali lab toward this end.[31]

Bacterial glycans have significant implications in human health and disease, and development of GBPs to study these important biomolecules will have great impact on the field of bacterial glycobiology through powerful tools for studying the interactions of human pathogens, commensals and symbionts and their hosts together with novel diagnostic and analytical reagents.

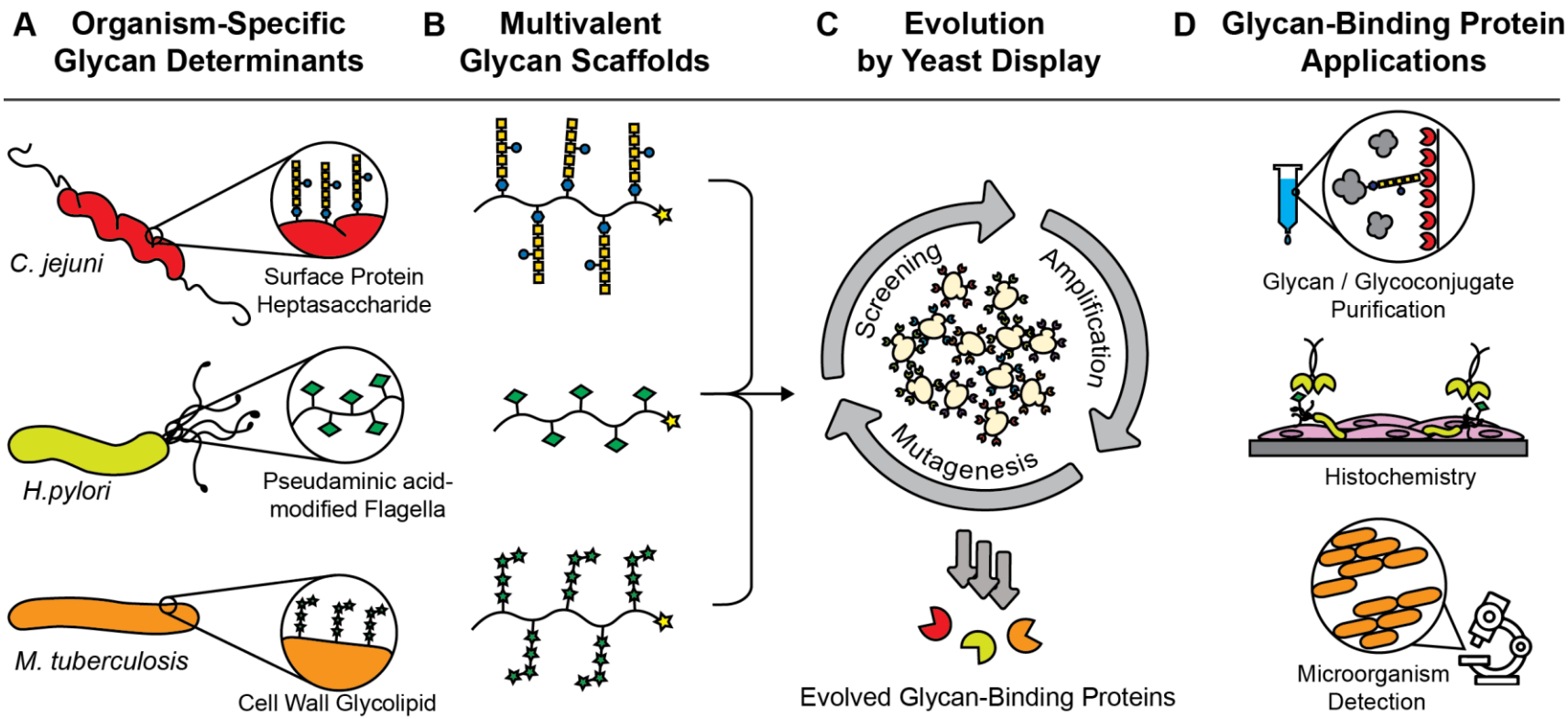


Figure 6-2. Bacterial glycan-binding protein generation workflow.
A) Bacterial glycan determinant is selected. B) Multivalent glycan ligands are synthesiz surface display. D) GBPs are applied for study and detection of bacterial glycan of inter

6.5 References

1. Imperiali, B. (2019) Bacterial Carbohydrate Diversity — a Brave New World, *Curr. Opin. Chem. Biol. 53*, 1-8.

2. Werz, D. B., Ranzinger, R., Herget, S., Adibekian, A., von der Lieth, C.-W., and Seeberger, P. H. (2007) Exploring the Structural Diversity of Mammalian Carbohydrates ("Glycospace") by Statistical Databank Analysis, *ACS Chem. Biol. 2*, 685-691.

3. Nothaft, H., and Szymanski, C. M. (2010) Protein Glycosylation in Bacteria: Sweeter Than Ever, *Nat. Rev. Microbiol. 8*, 765.

4. Matsuura, M. (2013) Structural Modifications of Bacterial Lipopolysaccharide That Facilitate Gram-Negative Bacteria Evasion of Host Innate Immunity, *Front. Immunol. 4*, 109-109.

5. Nwodo, U. U., Green, E., and Okoh, A. I. (2012) Bacterial Exopolysaccharides: Functionality and Prospects, *Int. J. Mol. Sci. 13*, 14002-14015.

6. Kaakoush, N. O., Castano-Rodriguez, N., Mitchell, H. M., and Man, S. M. (2015) Global Epidemiology of Campylobacter Infection, *Clin. Microbiol. Rev. 28*, 687-720.

7. Lu, Q., Li, S., and Shao, F. (2015) Sweet Talk: Protein Glycosylation in Bacterial Interaction with the Host, *Trends Microbiol. 23*, 630-641.

8. Young, N. M., Brisson, J.-R., Kelly, J., Watson, D. C., Tessier, L., Lanthier, P. H., Jarrell, H. C., Cadotte, N., St. Michael, F., Aberg, E., and Szymanski, C. M. (2002) Structure of the *N*-Linked Glycan Present on Multiple Glycoproteins in the Gram-Negative Bacterium, *Campylobacter Jejuni*, *J. Biol. Chem. 277*, 42530-42539.

9. Cain, J. A., Dale, A. L., Niewold, P., Klare, W. P., Man, L., White, M. Y., Scott, N. E., and Cordwell, S. J. (2019) Proteomics Reveals Multiple Phenotypes Associated with N-Linked Glycosylation in Campylobacter Jejuni, *Mol. Cell. Proteomics 18*, 715-734.

10. Abouelhadid, S., North, S. J., Hitchen, P., Vohra, P., Chintoan-Uta, C., Stevens, M., Dell, A., Cuccui, J., and Wren, B. W. (2019) Quantitative Analyses Reveal Novel Roles for N-Glycosylation in a Major Enteric Bacterial Pathogen, *mBio 10*, e00297-00219.

11. Nothaft, H., Scott, N. E., Vinogradov, E., Liu, X., Hu, R., Beadle, B., Fodor, C., Miller, W. G., Li, J., Cordwell, S. J., and Szymanski, C. M. (2012) Diversity in the Protein *N*-Glycosylation Pathways within the *Campylobacter* Genus *Mol. Cell. Proteomics 11*, 1203-1219.

12. Knirel, Y. A., Vinogradov, E. V., L'Vov, V. L., Kocharova, N. A., Shashkov, A. S., Dmitriev, B. A., and Kochetkov, N. K. (1984) Sialic Acids of a New Type from the Lipopolysaccharides of Pseudomonas Aeruginosa and Shigella Boydii, *Carbohydr. Res. 133*, C5-C8.

13. Schirm, M., Schoenhofen, I. C., Logan, S. M., Waldron, K. C., and Thibault, P. (2005) Identification of Unusual Bacterial Glycosylation by Tandem Mass Spectrometry Analyses of Intact Proteins, *Anal. Chem. 77*, 7774-7782.

14. Schirm, M., Soo, E. C., Aubry, A. J., Austin, J., Thibault, P., and Logan, S. M. (2003) Structural, Genetic and Functional Characterization of the Flagellin Glycosylation Process in Helicobacter Pylori, *Mol. Microbiol. 48*, 1579-1592.

15. Hitchen, P., Brzostek, J., Panico, M., Butler, J. A., Morris, H. R., Dell, A., and Linton, D. (2010) Modification of the Campylobacter Jejuni Flagellin Glycan by the Product of the Cj1295 Homopolymeric-Tract-Containing Gene, *Microbiology 156*, 1953-1962.

16. Josenhans, C., Vossebein, L., Friedrich, S., and Suerbaum, S. (2002) The Neua/Flmd Gene Cluster of Helicobacter Pylori Is Involved in Flagellar Biosynthesis and Flagellin Glycosylation, *FEMS Microbiol. Lett. 210*, 165-172.

17. Ottemann, K. M., and Lowenthal, A. C. (2002) Helicobacter Pylori Uses Motility for Initial Colonization and to Attain Robust Infection, *Infect. Immun. 70*, 1984-1990.

18. Perepelov, A. V., Wang, M., Filatov, A. V., Guo, X., Shashkov, A. S., Wang, L., and Knirel, Y. A. (2015) Structure and Genetics of the O-Antigen of Enterobacter Cloacae G3054 Containing Di-N-Acetylpseudaminic Acid, *Carbohydr. Res. 407*, 59-62.

19. Filatov, A. V., Perepelov, A. V., Shashkov, A. S., Burygin, G. L., Gogoleva, N. E., Khlopko, Y. A., and Grinev, V. S. (2021) Structure and Genetics of the O-Antigen of Enterobacter Cloacae K7 Containing Di-N-Acetylpseudaminic Acid, *Carbohydr. Res. 508*, 108392.

20. Lee, I. M., Tu, I. F., Yang, F.-L., and Wu, S.-H. (2020) Bacteriophage Tail-Spike Proteins Enable Detection of Pseudaminic-Acid-Coated Pathogenic Bacteria and Guide the Development of Antiglycan Antibodies with Cross-Species Antibacterial Activity, *J. Am. Chem. Soc. 142*, 19446-19450.

21. Turner, J., and Torrelles, J. B. (2018) Mannose-Capped Lipoarabinomannan in Mycobacterium Tuberculosis Pathogenesis, *Pathogens and Disease 76*, fty026.

22. Flores, J., Cancino, J. C., and Chavez-Galan, L. (2021) Lipoarabinomannan as a Point-of-Care Assay for Diagnosis of Tuberculosis: How Far Are We to Use It?, *Front. Microbiol. 12*, 638047.

23. Miller, E. A., Baniya, S., Osorio, D., Al Maalouf, Y. J., and Sikes, H. D. (2018) Paper-Based Diagnostics in the Antigen-Depletion Regime: High-Density Immobilization of Rcsso7d-Cellulose-Binding Domain Fusion Proteins for Efficient Target Capture, *Biosens. Bioelectron. 102*, 456-463.

24. Miller, E. A., Traxlmayr, M. W., Shen, J., and Sikes, H. D. (2016) Activity-Based Assessment of an Engineered Hyperthermophilic Protein as a Capture Agent in Paper-Based Diagnostic Tests, *Mol. Syst. Des. Eng. 1*, 377-381.

25. Dwivedi, R., Nothaft, H., Reiz, B., Whittal, R. M., and Szymanski, C. M. (2013) Generation of Free Oligosaccharides from Bacterial Protein N-Linked Glycosylation Systems, *Biopolymers 99*, 772-783.

26. Glover, K. J., Weerapana, E., and Imperiali, B. (2005) In Vitro Assembly of the Undecaprenylpyrophosphate-Linked Heptasaccharide for Prokaryotic N-Linked Glycosylation, *Proc. Natl. Acad. Sci. 102*, 14255-14259.

27. Amin, M. N., Ishiwata, A., and Ito, Y. (2007) Synthesis of N-Linked Glycan Derived from Gram-Negative Bacterium, Campylobacter Jejuni, *Tetrahedron 63*, 8181-8198.

28. Zamora, C. Y., Schocker, N. S., Chang, M. M., and Imperiali, B. (2017) Chapter Five - Chemoenzymatic Synthesis and Applications of Prokaryote-Specific Udp-Sugars, In *Methods Enzymol.* (Imperiali, B., Ed.), pp 145-186, Academic Press.

29. Williams, J. T., Corcilius, L., Kiefel, M. J., and Payne, R. J. (2016) Total Synthesis of Native 5,7-Diacetylpseudaminic Acid from N-Acetylneuraminic Acid, *J. Org. Chem. 81*, 2607-2611.

30. Zheng, R. B., Jégouzo, S. A. F., Joe, M., Bai, Y., Tran, H. A., Shen, K., Saupe, J., Xia, L., Ahmed, M. F., Liu, Y. H., Patil, P. S., Tripathi, A., Hung, S. C., Taylor, M. E., Lowary, T. L., and Drickamer, K. (2017) Insights into Interactions of Mycobacteria with the Host Innate Immune System from a Novel Array of Synthetic Mycobacterial Glycans, *ACS Chem. Biol. 12*, 2990-3002.

31. Munneke, S., Prevost, J. R. C., Painter, G. F., Stocker, B. L., and Timmer, M. S. M. (2015) The Rapid and Facile Synthesis of Oxyamine Linkers for the Preparation of Hydrolytically Stable Glycoconjugates, *Org. Lett. 17*, 624-627.