USING BIOLAYER INTERFEROMETRY TO QUANTIFY MONOMERIC SOLUTION AFFINITIES FOR CARBOHYDRATE-PROTEIN INTERACTIONS

by

YE JI

(Under the Direction of Robert J. Woods)

ABSTRACT

Many biological events involve carbohydrates binding to protein receptors (lectins, antibodies, enzymes, etc). The carbohydrates are often present as glycoproteins or glycopids that may be free in solution or anchored to membranes. The strengthes of such interactions have been studied and reported using various techniques in the past, for instance: surface plasmon resonance (SPR), enzyme-linked immunosorbent assay (ELISA), biolayer interferometry (BLI), isothermal titration calorimetry (ITC), microscale thermophoresis (MST), nuclear magnetic resonance spectroscopy (NMR), and more qualitatively by glycan array screening. However, among all these techniques the ones involving surface immobilization such as SPR, ELISA and BLI have high sensitivity and are require low amounts of reagents, but often are unable to generate monomeric affinity measurement due to the typically multimeric nature of the glycans or the receptors. Those techniques that do not require any immobilization such as NMR, ITC and MST can measure the monomeric affinity values, but are less sensitive and may require considerably more reagent, especially ITC and NMR. To understand the complicated relationships between monomeric affinity and multimeric avidity, a more convenient method for measuring affinity is urgently needed. This work introduced a Biolayer Interferometry

competition assay for the monomeric solution K_D determination. Firstly, this approach was tested on two well-studied cases: *Erythrina cristagalli* lectin (ECL) and Human influenza A/Hong Kong/1/1968 (X-31) H3N2 hemagglutinin with their ligands; resulting in good agreement with literature. Secondly, a survey of receptor conformational properties was presented. Results highly suggested that conformational entropy played a key role in defining specificity. Thirdly, our robust and accurate assay was applied to a current hot topic: pandemic influenza hemagglutinin-glycan interactions. This was the first time reporting solution K_D values for several important HAs. Fourthly, the inhibitory ability of several novel potential HA inhibitors where determined using the BLI competitive assay were presented. Finally, a study was completed using BLI to measure the direct binding of antibody binding: anti-blood group antibody specificity was demonstrated

INDEX WORDS: influenza hemagglutinin, glycan, carbohydrate, interaction, structure

USING BIOLAYER INTERFEROMETRY TO QUANTIFY MONOMERIC SOLUTION AFFINITIES FOR CARBOHYDRATE-PROTEIN INTERACTIONS

by

YE JI

B. S., Tianjin Medical University, 2006

M. S., Pittsburg State University, 2009

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2018

USING BIOLAYER INTERFEROMETRY TO QUANTIFY MONOMERIC SOLUTION AFFINITIES FOR CARBOHYDRATE-PROTEIN INTERACTIONS

by

YE JI

Major Professor: Robert J. Woods
Committee: James H. Prestegard
S. Mark Tompkins

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
May 2018

DEDICATION

I dedicate this work to my Mom, Guojuan Zhao.

She did not only give me the flesh but also the soul.

ACKNOWLEDGEMENTS

I would like to express acknowledgements of all the people who have helped me and have made with this work possible. I would like to acknowledge my academic advisor, Dr. Robert J. Woods, who prepared me as a successful scientist from University of Georgia. He did not only provide strong mentorship and financial research assistant during this program, but also strongly supported and cultivated me to think, solve and perform philosophical works those to improve humanity.

I would also like to thank my committee members, Dr. James H. Prestegard and Dr. S. Mark Tompkins for their helpful inside, guidance and collaboration on multiple projects over the years. I am especially thankful to all past and current Woods group members. I feel being cared and loved studying and living here. Especially, I want to thank Mark Baine and Dr. Oliver Grant's tremendous help from English grammar to writing to make this work become true.

I also want to express my grate to my family and friends for the encouragement during the process. Study aboard was very hard, in the past 8 years; I have lost two family members and three friends, fought against pre-cancer stage; and four newborn nieces and nephews whom I have not met yet at this point. Without any of these supports, I would not make it through. Lastly, the endless support from my Mom, Guojuan Zhao fulfilled with work with spirits.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

Bio-Layer Interferometry

Bio-Layer Interferometry (BLI) is a relatively new label-free high-throughput technique to study the interactions between an immobilized receptor and soluble analytes in real-time (1). This optical technique analyses the change in the interference pattern of white light reflected from a biosensor surface carrying a chemically immobilized ligand (such as a protein) that occurs as the analyte in solution (such as a carbohydrate) binds to the ligand (Figure 1.1).

Comparisons between BLI and traditional biophysical analysis instruments

The BLI approach shares some conceptual similarities with surface plasmon resonance (SPR), in the sense that the ligand is immobilized on a surface, both methods use similar immobilization chemistry, and the binding is detected in real time. However there are differences. Whereas SPR employs a microfluidic flow cell, BLI employs a multi-well plate (typically 96 or 384), with a row of eight sensors that are dipped into the wells. The association phase of the binding is thus dependent primarily on the concentration of the analyte in the wells and to the time the probe is exposed to the analyte. To measure analyte dissociation, the probes are automatically raised and transferred to wells containing buffer (Figure 1.2). The times for association and dissociation phases can be adjusted during the experiment, and more complex assays may be programmed. To assist in minimizing mass-transport effects, the analyte tray oscillates at up to 1000 rpm.

The performance of SPR and BLI has been reviewed recently (2-4) and while SPR may be slightly more sensitive (2), K_D values from BLI are generally within a factor of two (larger) than those from SPR (2,4). A significant advantage of BLI over SPR is the lack of any limit on measurement times for association or dissociation, which in SPR are limited by the volume of the microfluidics. On the other hand, very long exposures in the analyte wells can lead to solvent evaporation degrading the BLI measurements (3). Other benefits include the ability of BLI to measure interactions on relatively low-cost disposable probes that do not need to be regenerate (2). Overall the high throughput of the BLI platform may be its most significant benefit over SPR (3), and the popularity of the BLI method is steadily increasing (Figure 1.3). BLI provides a relatively high-throughput cost-effective method for quantifying the strengths of carbohydrate-protein interactions that bridges the gap between qualitative affinity data from high-throughput glycan array screening, and quantitative data from lower-throughput methods such as SPR, ELISA, ITC, MST and NMR. Additionally, like SPR, BLI provides the ability to determine kinetic binding rate constants (kon, koff), which are critical to a complete understanding of molecular recognition processes, such as between antibodies or lectins and carbohydrates (5-7).

BLI is recommended for characterizing interactions with affinities between 1 mM to 10 pM, and for analytes with molecular weights of at least 1.5 kDa, but preferably closer to 4 kDa (1). A convenient feature of the ForteBio implementation of BLI is the ability to alter the times allocated for the association or dissociation phases in real-time, in response to the observed data.

However, as is the case with all techniques for determining affinities that require immobilization of the receptor, multimeric analytes may form multivalent interactions. Such interactions boost the observed affinity through avidity, and hence do not result in monomeric

affinity values (10, 11). Additional issues, which may be particularly problematic in multivalent binding (12), include artificially slow dissociation rates, due to rapid analyte rebinding or "mass transport" (13). *Slow off rates ultimately lead to an overestimation of the affinity when K_D is derived from k_{off}/k_{on}.*

Protein-Carbohydrate Interactions

Protein-carbohydrate interactions mediate countless biological events, from normal development to cancer metastasis, and viral and bacterial infection (14). To overcome the notoriously low affinity (mM) of these interactions, biological glycan interactions are frequently inferred on the basis of their structure to take advantage of multimeric or multivalent binding (15). However, the structures of these multimeric complexes are generally unknown, while what is most often available instead are the three-dimensional (3D) structures of monomeric complexes, generated by X-ray diffraction. To develop robust structure-function relationships for glycans, a first step is therefore to be able to measure monomeric binding affinities that can be interpreted with regard to the available structural data. The ability to quantify monomeric binding affinities for glycans is essential for understanding the effects of protein mutations on biological function, as for example in defining the origin of species specificity of influenza infections (16). This information is equally important for the development of inhibitors that target glycan receptors (17,18).

The low affinity interactions associated with proteins and carbohydrates are challenging to measure by direct binding techniques such as NMR, MST, and ITC (19-22), often because of the limited availability of sufficient quantities of large biologically relevant glycans. A

competition assay overcomes this limitation by employing the minimum-binding determinant (or other small carbohydrate structure) as the inhibitor, while requiring only a minimal amount of the large glycan as the direct binding analyte.

Given that many carbohydrate-binding proteins are multimeric, as are most biologically relevant glycans, surface immobilization techniques are not ideal for quantifying carbohydrate-protein interactions. However, the extreme complexity of many glycans practically limits their availability, thus we sought to develop an approach to measuring monomeric solution K_D values for carbohydrate-protein interactions that nevertheless exploited the economies and convenience gained from the BLI approach. To do this, we created a competition-based protocol that reproduces K_D values from solution-phase equilibrium binding measurements (23).

BLI competitive assay Equation 1

Given that BLI is not designed to measure interactions weaker than mM, but that this is the range of many carbohydrate-protein affinities; the proposed BLI-based inhibition assay was developed to overcome this limitation. The solution K_D of the inhibitor (K_i) is obtained from **Equation 1** (23), and requires measurement of an IC_{50} for the analyte (oligosaccharide), and a K_D for binding to the immobilized receptor ($K_{D,\ surface}$).

$$K_i = IC_{50} / (1 + [protein]/K_{D,surface}) \qquad \textbf{Equation 1}$$

As illustrated in Figure 1.4, in the BLI based competition assay, there are two equilibria: that for the small oligosaccharide binding to the immobilized protein, and that for the mass-

amplified oligosaccharide (Fab-glycan) binding to the immobilized protein. When the oligosaccharides in both cases are exactly the same glycan, the K_I of small oligosaccharide is equivalent to its solution K_D (23), once any non-specific interactions with the Fab component are subtracted.

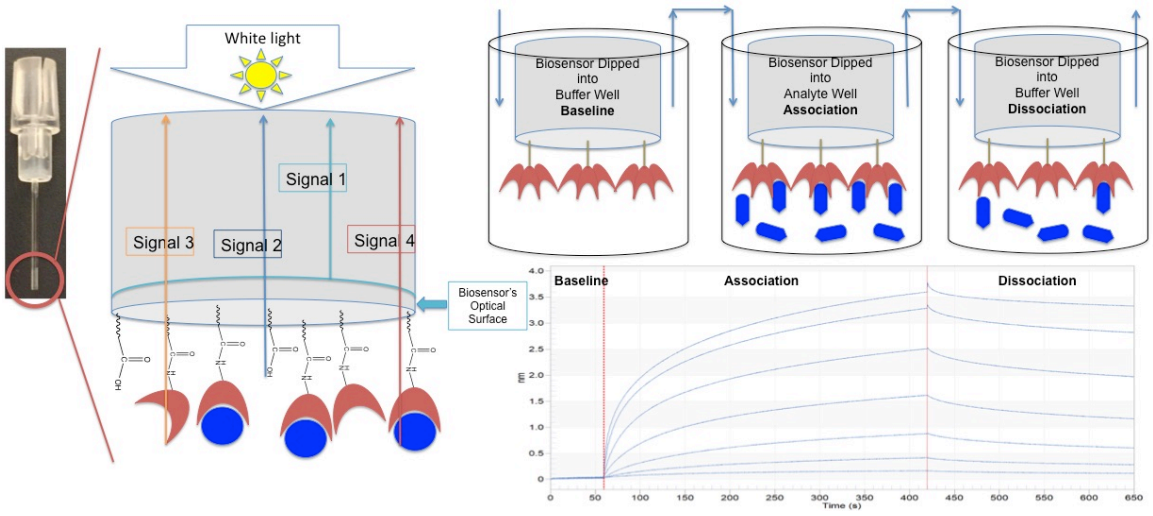


Figure 1.1 Mechanism of ForteBio Bio-Layer Interferometry.

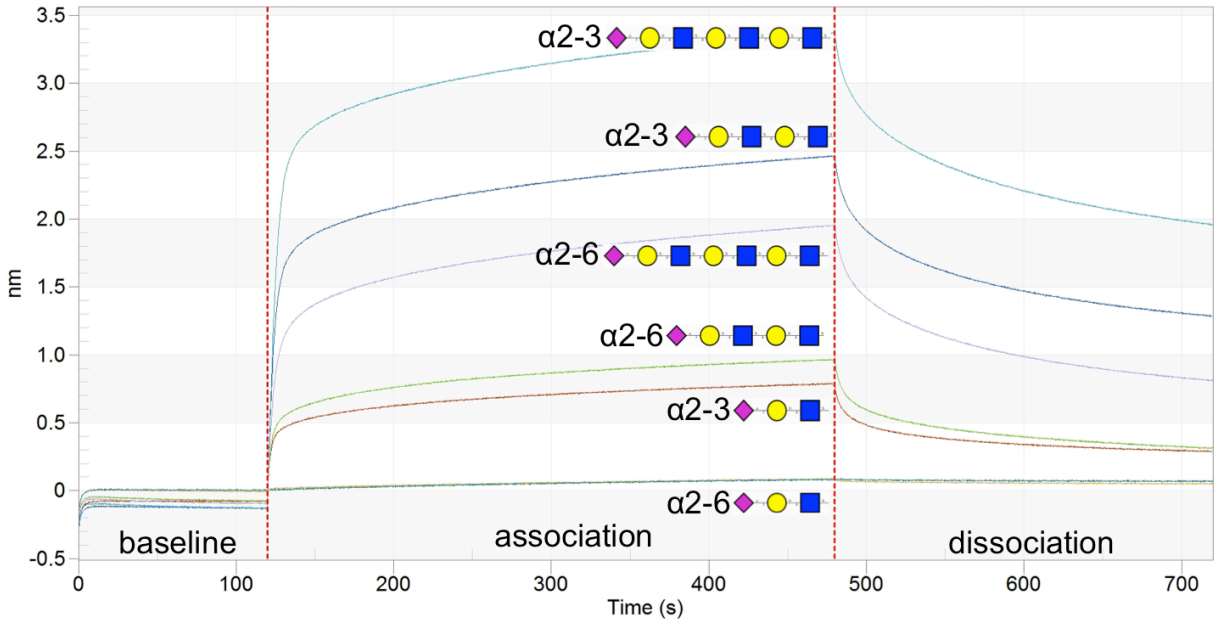


Figure 1. 2. An example in which BLI is used to screen the binding of six oligosaccharides to the lectin wheat germ agglutinin (WGA). Baseline (sensor soaking in buffer) 0~120s, association phase 120-480s, and dissociation phase 480-720s.

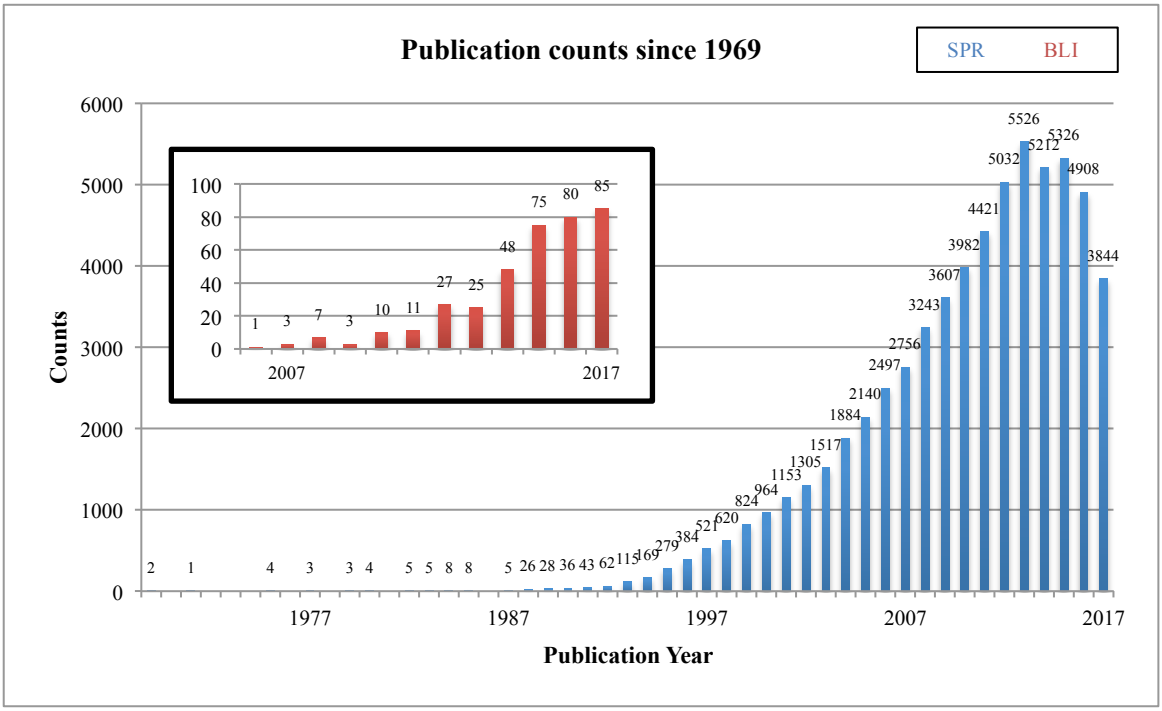


Figure 1. 3. Number of publications citing SPR and or BLI reported per year. Data obtained from SciFinder searched with the keyword "Surface Plasmon Resonance" or "BioLayer Interferometry". First publication of SPR (8) was reported in 1969. Then the first publication for BLI appeared over 35 years later in 2007 (9).

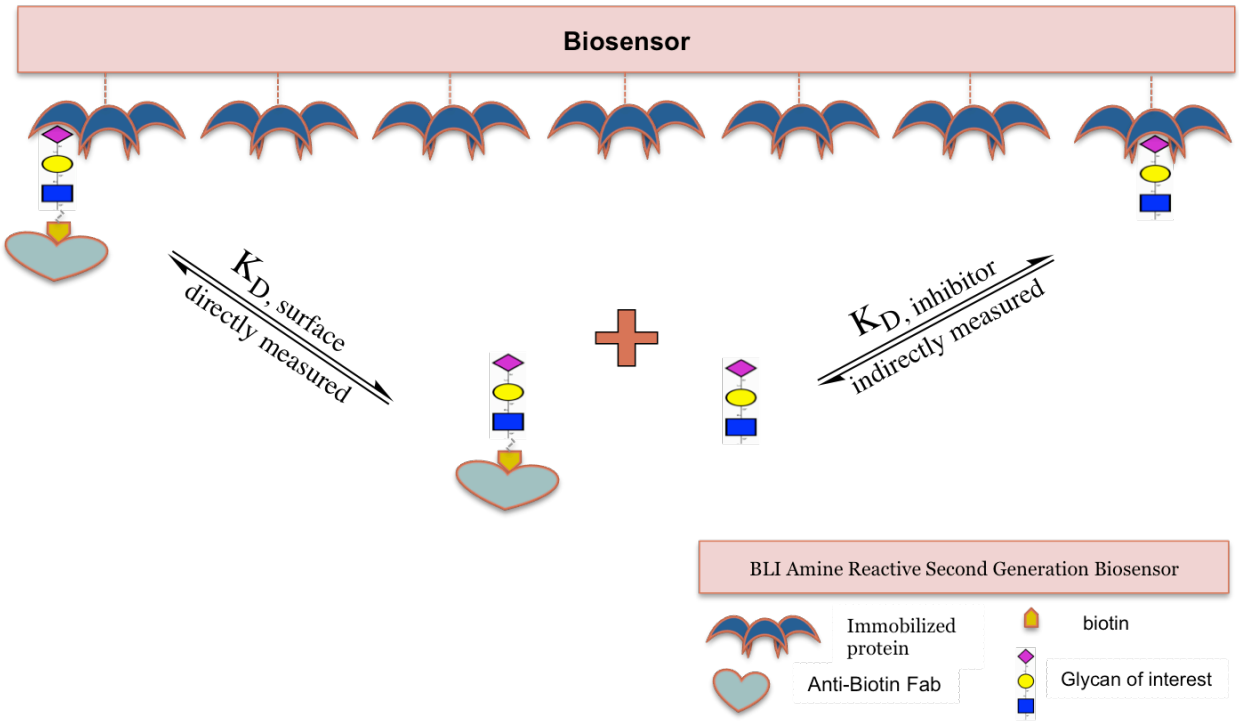


Figure 1. 4. Schematic representation of the BLI-based competition assay. The ability of a small carbohydrate (typically the minimal binding determinant) to inhibit (IC_{50}) the direct binding of the analyte of interest ($K_{D,\ surface}$) is determined in two sets of experiments. From these values the solution K_D ($K_{D,\ solution}$) of the minimal binding determinant may be determined.

References

1. Rich, R. L.; Myszka, D. G. (2007) Higher-throughput, label-free, real-time molecular interaction analysis. Anal Biochem 361, 1-6
2. Abdiche, Y., Malashock, D., Pinkerton, A., and Pons, J. (2008) Determining kinetics and affinities of protein interactions using a parallel real-time label-free biosensor, the Octet. *Anal Biochem* **377**, 209-217
3. Yang, D. L., Singh, A., Wu, H. L., and Kroe-Barrett, R. (2016) Comparison of biosensor platforms in the evaluation of high affinity antibody-antigen binding kinetics. *Analytical Biochemistry* **508**, 78-96
4. Abdiche, Y. N., Malashock, D. S., Pinkerton, A., and Pons, J. (2009) Exploring blocking assays using Octet, ProteOn, and Biacore biosensors. *Anal Biochem* **386**, 172-180
5. Lin, Y. P.; Xiong, X.; Wharton, S. A.; Martin, S. R.; Coombs, P. J.; Vachieri, S. G.; Christodoulou, E.; Walker, P. A.; Liu, J.; Skehel, J. J.et al. (2012) Evolution of the Receptor Binding Properties of the Influenza a(H3N2) Hemagglutinin. *Proc Natl Acad Sci***,** *USA* **109**, 21474-21479
6. Lian, W.; Wu, M.; Huang, N.; Gao, N.; Xiao, C.; Li, Z.; Zhang, Z.; Zheng, Y.; Peng, W.; Zhao, J. (2013) Anti-HIV-1 Activity and Structure–Activity-Relationship Study of a Fucosylated Glycosaminoglycan from an Echinoderm by Targeting the Conserved CD4 Induced Epitope. *BBA-Gen. Subjects,* **1830**, 4681-4691
7. Schieffelin, J.; Costin, J.; Nicholson, C.; Orgeron, N.; Fontaine, K.; Isern, S.; Michael, S.; Robinson, J. (2010) Neutralizing and Non-Neutralizing Monoclonal Antibodies against Dengue Virus E Protein Derived from a Naturally Infected Patient. *Virology J.* **7**, 1-11

8. Stanford, J. L., and Bennett, H. E. (1969) Enhancement of surface plasma resonance absorption in mirrors by overcoating with dielectrics. *Appl Opt* **8**, 2556-2557

9. Peng, L., Varma, M. M., Cho, W., Regnier, F. E., and Nolte, D. D. (2007) Adaptive interferometry of protein on a BioCD. *Appl Opt* **46**, 5384-5395

10. Mammen, M., Choi, S.-K., and Whitesides, G. M. (1998) Polyvalent Interactions in Biological Systems: Implications for Design and Use of Multivalent Ligands and Inhibitors. *Angewandte Chemie International Edition* **37**, 2754-2794

11. Gelinsky-Wersing, D., Wersing, W., and Pompe, W. (2017) Bivalent kinetic binding model to surface plasmon resonance studies of antigen-antibody displacement reactions. *Anal Biochem* **518**, 110-125

12. Liang, P.-H., Wang, S.-K., and Wong, C.-H. (2007) Quantitative Analysis of Carbohydrate−Protein Interactions Using Glycan Microarrays: Determination of Surface and Solution Dissociation Constants. *J. Am. Chem. Soc*. **129**, 11177–11184

13. Schuck, P., and Zhao, H. (2010) The role of mass transport limitation and surface heterogeneity in the biophysical characterization of macromolecular binding processes by SPR biosensing. *Methods Mol Biol* **627**, 15-54

14. Solís, D., Bovin, N. V., Davis, A. P., Jiménez-Barbero, J., Romero, A., Roy, R., Smetana Jr, K., and Gabius, H.-J. (2015) A guide into glycosciences: How chemistry, biochemistry and biology cooperate to crack the sugar code. *BBA-Gen. Subjects* **1850**, 168-235

15. Mammen, M., Choi, S.-K., and Whitesides, G. M. (1998) Polyvalent Interactions in Biological Systems: Implications for Design and Use of Multivalent Ligands and Inhibitors. *Angew. Chem. Int. Ed.* **37**, 2754-2794

16. Ji, Y., White, Y. J., Hadden, J. A., Grant, O. C., and Woods, R. J. (2017) New insights into influenza A specificity: an evolution of paradigms. *Curr Opin Struct Biol* **44**, 219-231

17. Ernst, B., and Magnani, J. L. (2009) From Carbohydrate Leads to Glycomimetic Drugs. *Nat. Rev. Drug Disc.* **8**, 661-677

18. Magnani, J. L., and Ernst, B. (2009) Glycomimetic drugs--a new source of therapeutic opportunities. *Discov. Med.* **8**, 247-252

19. Gupta, D., Cho, M., Cummings, R. D., and Brewer, C. F. (1996) Thermodynamics of Carbohydrate Binding to Galectin-1 from Chinese Hamster Ovary Cells and Two Mutants. A Comparison with Four Galactose-Specific Plant Lectins. *Biochemistry* **35**, 15236-15243

20. Notenboom, V., Boraston, A. B., Williams, S. J., Kilburn, D. G., and Rose, D. R. (2002) High-Resolution Crystal Structures of the Lectin-like Xylan Binding Domain from Streptomyces liVidans Xylanase 10A with Bound Substrates Reveal a Novel Mode of Xylan Binding. *Biochemisrty* **41**, 4246-4254

21. Sauter, N. K., Hanson, J. E., Glick, G. D., Brown, J. H., Crowther, R. L., Park, S. J., Skehel, J. J., and Wiley, D. C. (1992) Binding of Influenza Virus Hemagglutinin to Analogs of its Cell-surface Receptor, Sialic Acid: Analysis by Proton Nuclear Magnetic Resonance Spectroscopy and X-ray Crystallography. *Biochemistry* **31**, 9609-9621

22. Xiong, X., Coombs, P. J., Martin, S. R., Liu, J., Xiao, H., McCauley, J. W., Locher, K., Walker, P. A., Collins, P. J., Kawaoka, Y., Skehel, J. J., and Gamblin, S. J. (2013) Receptor binding by a ferret-transmissible H5 avian influenza virus. *Nature* **497**, 392-396

23. Cheng, Y., and Prusoff, W. (1973) Relationship between inhibition constant (K_i) and concentration of inhibitor which causes 50 percent inhibition (I_{50}) of an enzymatic reaction. *Biochem. Pharm.* **22**, 3099-3108

CHAPTER 2

NEW INSIGHTS INTO INFLUENZA A SPECIFICITY: AN EVOLUTION OF PARADIGMS

[1]Ye Ji, Yohanna J. B. White, Jodi A. Hadden, Oliver C. Grant, Robert J. Woods. *Current Opinion in Structural Biology* (2017) 44: 219-231. Reprinted here with permission of publisher.

Abstract

Understanding the molecular origin of influenza receptor specificity is complicated by the paucity of quantitative affinity measurements, and the qualitative and variable nature of glycan array data. Further obstacles arise from the varied impact of viral glycosylation and the relatively narrow spectrum of biologically-relevant receptors present on glycan arrays. A survey of receptor conformational properties is presented, leading to the conclusion that conformational entropy plays a key role in defining specificity, as does the newly-reported ability of biantennary receptors that terminate in Siaα2-6Gal sequences to form bidentate interactions to two binding sites in a hemagglutinin trimer. Bidentate binding provides a functional explanation for the observation that Siaα2-6 receptors adopt an open-umbrella topology when bound to hemagglutinins from human-infective viruses, and calls for a reassessment of virus avidity and tissue tropism.

Highlights

- Influenza specificity is influenced by differences in glycan conformational entropy
- Biantennary α2-6 sialosides can boost avidity by forming bidentate HA complexes
- All reported HA-oligosaccharide complexes can be grouped into one of four topologies
- Interpretations of specificity must consider the impact of changes in glycan substructure

Introduction

Wild birds are the primary natural reservoir for influenza A viruses [1], and the 1918 Spanish Flu pandemic that killed an estimated 50 million people [2] is believed to have originated from spontaneous mutations in an avian influenza virus that conferred human-to-human transmissibility [3, 4]. While zoonotic influenza can infect humans [5], close contact with infected animals is required [6]. Subsequent human-to-human transmission, leading to pandemics, requires that the virus undergo additional genetic alterations [5, 6]. As noted by Reperant et al. [5], in order for a zoonotic virus to become human-infective, it must overcome three sets of barriers: animal-to-human transmission, virus-cell interaction, and human-to-human transmission. Seasonal influenza epidemics arise from human-to-human transmission of circulating strains that have undergone sufficient mutation (antigenic drift) to circumvent established immunity within the population [7].

In contrast to the Spanish Flu, the Swine Flu pandemic of 2009 was relatively mild [8]. Nevertheless it raised concerns within the World Health Organization because of the rapidity with which it spread [9]; within 6 weeks of the first case, Swine Flu had spread to over 70 countries [10] and required the development of a new vaccine. Human adaptation is of particular concern in the case of highly pathogenic avian influenza (HPAI) subtypes, such as H5N1. Although infrequent, human infection by avian H5N1 has been reported in 16 countries, resulting in approximately 60% mortality [11]. Preparedness for pandemics therefore necessitates anticipation of the virulence of emerging strains, providing motivation for developing a deeper understanding of the basis for influenza specificity. Here, we reassess the

relationship between host glycan structure and influenza specificity in light of recent data that indicates critical roles for glycan sub-structure and dynamics.

Influenza A classification is based on the antigenic properties of the hemagglutinin (HA) and neuraminidase (NA) envelope proteins. Influenza HA is a homotrimeric glycoprotein whose protomers each comprise a globular head domain (HA1) and stalk region (HA2) [12]. Each HA1 domain contains a receptor binding site (RBS), through which the virus adheres to the host cell via binding to host glycans that contain sialic acid (Sia, neuraminic acid, Neu5Ac). There are currently 18 hemagglutinin subtypes, which are classified into two groups based on their antigenic properties: group 1 consists of H1-2, H5-6, H8-H9 H11-13, and H16; group 2 contains H3-4, H7, H10, H14, and H15. The most extensively studied HAs include H1, H3 and H5 [13, 14]. The NA protein mediates the cleaving of Sia from the host receptor glycan post cellular infection, enabling progeny virus to escape from the host cell surface [15]. Cryo-electron tomography indicates that there are approximately 300 HA proteins in the viral envelope [16], with the ratio of HA to NA varying between different strains from 4 to 6:1 [16, 17]. Compound factors affect the ability of a particular strain of influenza to infect humans, including the level of exposure, the replication rate in newly infected individuals, the glycan binding preferences of the viral surface HA, and the activity of the viral surface NA [15, 18-23]. Further, the enzyme activity of the NA must balance with the affinity of the HA [15, 22]. If the NA is too active, relative to the affinity of the HA, it will attenuate the ability of the virus to infect the host cell. Conversely, a relatively weak NA will impair shedding of the progeny virus.

In addition to receptor specificity, zoonotic infection is also sensitive to differences in the susceptibility of the HA to pH-mediated endosomal fusion [24], and differences in the efficiency

of translocation of the viral ribonucleoprotein complex to the host nucleus [25] (host adaptation). Moreover, ease of transmission and replication appears to be dependent on the distribution and composition of the receptors on host tissue. Viral attachment studies have shown that human influenza viruses adhere more strongly to human trachea and bronchi than avian viruses, and attach to different cell types [26]. Thus, the lack of a suitable receptor has been invoked as being responsible for the inefficient transmission [27] and replication of avian viruses in humans [28, 29]. Much work has been done to elucidate the molecular basis for the observed tissue tropism [28, 30-32].

Defining receptor specificity

The canonical view of the relationship between HA receptor specificity and species infectivity is that the HA in human-infective viruses prefers to bind to glycans present on the cell surface that terminate with the Siaα2-6Gal (α2-6) sequence; in avian-infective viruses, the HA prefers to bind to glycans that terminate in Siaα2-3Gal (α2-3). Some species, such as swine, can be co-infected by viruses that prefer either α2-3 or α2-6 structures, leading to the potential for genetic reassortment (antigenic shift) in swine that results in the introduction of α2-6 binding preference (enhanced human infectivity) into a zoonotic framework [5].

The discovery of the α2-6/α2-3 infectivity relationship originated not from quantitative biophysical studies, but from more qualitative, yet robust, hemagglutination assays [33]. Rogers and Paulson's [34] seminal work on enzymatically-modified red blood cells (RBCs) established that influenza receptor specificity depends, to an extent, on the species from which the virus was isolated. They reported that isolates of human subtype H3N2 agglutinated RBCs whose modified

surface glycans terminated in the α2-6 sequence, but that these isolates did not agglutinate RBCs with α2-3 glycans. Conversely, avian isolates preferentially agglutinated RBCs containing the α2-3 linkage. While hemagglutination by influenza is a general phenomenon not limited to chicken RBCs [35], not all virus strains agglutinate all types of RBCs equally well [35, 36]. Unmodified chicken RBCs contain a diversity of multiantennary glycans, roughly in an equal ratio of α2-3:α2-6 [36], but these represent only a limited subset of the glycans found on human epithelial tissue, which also include multiple lactosamine repeats in the antennae. The observation that the necessary human-type receptors are not present provides an explanation of the inability of certain human-adapted influenza strains to agglutinate chicken RBCs [36]. As noted by Ovsyannikova et al. [35], species selection of red blood cells (RBCs) is critical to determine antibody titers to influenza viruses reliably, however, further glycomics analyses are required to elucidate the origin of the differences in RBC agglutination behavior.

Affinity versus avidity

Monomeric binding affinities for HA-glycan interactions confirm the canonical view of HA specificity, but show remarkably modest differences between α2-3 and α2-6 receptors (Table 2.1). Avidity arising from interactions between multiple host glycans and multiple trimeric HAs on the viral surface has been invoked to explain the difference between the weak (mM) monomeric affinities for HA-glycan interactions and the sub-μM binding for whole virus [37-40]. Indeed, models of binding kinetics [38, 39] have shown that avidity can exponentially amplify the subtle differences in monomeric affinities, resulting in agreement with experimental virus binding kinetics.

In 2012, Lin et al. [44] reported that the avidity of H3N2 viruses for an α2-6 trisaccharide receptor decreased approximately four fold between 1968 and 2001, then progressively decreased a further 200 fold from 2001 to 2010, to such an extent that higher virus concentrations were required to observe any binding for the 2010 strains. This decrease in binding avidity was shown to be the result of mutations (antigenic drift) that weakened specific interactions between the RBS and the glycan receptors [44]. Recently, Peng et al. [45] screened the HAs from a number of H3N2 viruses against a custom glycan array that included multiantennary glycans of the type found in the human respiratory tract [46], and confirmed that binding to short, or linear, glycans had steadily decreased, consistent with the observations of Lin et al. [44]. However, strong binding to long biantennary sialoglycans was observed that was relatively insensitive to the effects of antigenic drift.

Recently, Peng et al. [45], and de Vries et al. [47] have raised the intriguing possibility that both branches in a biantennary glycan could bind simultaneously to two RBSs in an HA trimer, provided the branches were sufficiently long to reach two RBSs (Figure 2.1). Such bidentate binding would amplify the affinity of the glycan, potentially resulting in an apparent affinity of as much as the square of the monovalent K_D ($K_{D,mono}^2$) [40], although this would likely be reduced by entropic penalties. This binding enhancement would enable the HA to continue to retain affinity for certain biantennary glycans despite the overall negative impact of antigenic drift on receptor binding at a monovalent level. This hypothesis provides a basis for explaining the observation that, despite the general decrease in avidity displayed by H3N2 viruses [44, 48], they retain the ability to bind to biantennary glycans [45] and, thus, to infect and transmit in the human population [48].

Impact of HA glycosylation on specificity

Glycosylation of HA proteins varies both in location and composition depending on the strain of the virus [45, 50], as well as on the cell-type in which the virus was produced [51]. Over time, the number of glycosylation sites in circulating influenza strains has increased [50, 52], presumably shielding the protein surface from antibody recognition and assisting the virus in evading host immune surveillance [50, 53-56]. However, the more heavily glycosylated an HA1 domain, the more likely that its receptor binding ability will be impaired, either because the glycosylation directly blocks access to the RBS [57, 58], or because it forms a shield through which short receptor glycans may not be able to penetrate. Increased glycosylation, thus, potentially decreases affinity and virulence [59]. Three decades ago, it was observed that passaging of an avian infective H1N1 strain (A/WSN/1933) in mammalian (MDBK) cells led to the loss of glycosylation at N129 in the HA1 domain, leading to an increased affinity for host receptors, whereas passaging in chicken cells had no effect on glycosylation [60]. More recently, based on an analysis of 3D structures of HAs, Jayaraman et al. [57] predicted that, because of its proximity to the RBS, the loss of glycosylation at N91 in the HA from an H1N1 (A/South Carolina/1/18 and two variants, D225G and D190E/D225G) should affect receptor-binding properties. While loss of glycosylation at N91 was found to have no affect on the binding of the D190E/D225G (avian-like) variant to immobilized α2-3 oligosaccharides, it completely abrogated binding of the D225G variant to α2-3 and α2-6 oligosaccharides, and attenuated binding of wild-type HA to α2-6 oligosaccharides. The mechanism underlying the negative impact of loss of glycosylation on α2-6 binding was not identified.

In H5N1 strains, the N158 glycosylation site occupies a similar spatial position to that of

N129 in H1N1 strains, and appears to produce similar effects when glycosylated (attenuation of antigenicity, reduction of affinity for α2-6 receptors [61]). H5N1 viruses lacking glycosylation at N158 transmit efficiently by direct contact among guinea-pigs [62]. In 2015 Zhang et al. [63] examined the impact of glycosylation at three sites in the HA1 of an H5N1 virus (A/Mallard/Huadong/S/2005) and reached the conclusions that: 1) loss of glycosylation at N158 was a prerequisite for binding to α2,6-modified RBCs, and 2) viruses with a loss of glycosylation at N158 or N169 had higher lethality in mice. In 2010, Liao et al. [64] showed that deletion of glycosylation sites in an H5 derived from a consensus-based sequence [65] led to no major change in the glycan binding profiles for α2-3 oligosaccharides.

Yang et al. [54] noted in a study of H3N2 strains that the viruses had evolved to prefer longer linear glycans, and hypothesized that this preference was related to an increase in the number of glycosylation sites in the HA1. Alymova et al. [66] also recently examined H3N2 with varying glycosylation levels, and concluded that glycosylation of the HA1 could decrease binding affinity, without reducing virulence. They further introduced the hypothesis, based on the consistent binding of the HAs to linear α2-6 sialylated polylactosamine glycans, that physiologically relevant receptor binding had not changed over the past 40 years. However, their array did not include the large biantennary glycans used by Peng et al. [45], who concluded that H3N2 had evolved specificity for extended, branched α2-6 glycans.

While the current data regarding the impact of HA1 glycosylation show strain dependence, binding to α2-6 receptors generally appears to be markedly sensitive to variations in HA1 glycosylation. Further studies will be required to develop a clear understanding of the conditions under which HA1 glycosylation alters receptor binding and or virulence.

Relating HA structure to receptor specificity

Examination of pandemic HA sequences permits the identification of mutations in the RBS that appear to play a role in switching the virus specificity. A pair of mutations identified as E190D and G225D in H1N1 viruses has been shown to be critical for switching the binding preference from α2-3 to α2-6 glycans [3, 4, 56], and appears to have been responsible for the Spanish Flu pandemic [67]. Mutation at only one of these sites within an H1 typically leads to dual α2-3 and α2-6 receptor binding [3, 56, 68]. A different pair of mutations (Q226L and G228S) enabled the H2N2 and H3N2 pandemic viruses to gain specificity for α2-6 glycans [69]. However, these observations should not be considered to be specificity 'rules' – as part of a study to engineer α2-6 specificity into an H5N1 (A/Vietnam/1203/04), introduction of the E190D and G225D double mutations remarkably eliminated binding to all α2-3 and α2-6 glycans examined [70]. Additional host-adaptation is required in order to achieve this specificity switch in H5N1 viruses [71-74]. Very recently de Vries et al. [47] have shown that three mutations (V186K/G, K193T, and G228S) switch H7N9 influenza to human-type receptor specificity, with a binding profile practically identical to pandemic H1N1 A/California/04/2009.

The 3D structures of HA-oligosaccharide complexes are essential for understanding, and potentially predicting, the effect of mutations in HA on receptor specificity, and the structural features of influenza HA-glycan co-complexes have been well described [13, 15, 75, 76]. However, despite the large and growing number of co-crystal structures of HA-oligosaccharide complexes, rationalization of the observed specificity preferences in terms of 3D structural properties is far from straightforward [77]. As a case in point, consider the complexes of HAs from avian- (A/Wild Duck/JX/12416/2005) and human-infective (A/California/04/2009) H1N1

viruses (Figure 2.2). These HAs have been co-crystallized with both α2-3 and α2-6 glycans, and therefore provide an opportunity to illustrate the differences in hydrogen-bonding patterns in human- or avian-adapted HAs. It is clear from an examination of the hydrogen bond patterns between α2-3 and α2-6 oligosaccharides with the human-adapted HA (Figure 2.2, lower panels) that the α2-6 receptor makes several additional interactions (involving D190, D225 and K222) relative to the α2-3. These interactions are consistent with the observed α2-6 preference for human-adapted HAs. Why avian-adapted HAs generally bind more weakly, if at all, to α2-6 oligosaccharides is far less clear from these structures (Figure 2.2 upper panels). Indeed, as noted by Lin et al. [78], the mode of binding observed for the avian-adapted HA is not consistent with the view that the avian HA favors α2-3 receptors over α2-6.

The answer to these structural riddles must lie in properties that are not as readily apparent as hydrogen bond networks. These include van der Waals contacts, as well as contributions from conformational entropy, which may be significantly different for the two types of ligand.

Conformational entropy – the missing link

Given the relatively plastic nature of glycans, binding to a protein incurs an entropic penalty proportional to the degree of conformational constriction, and this has been proposed as an unfavorable contribution in HA-glycan binding [79]. Notably, computational analyses, based on molecular dynamics simulations of crystallographic HA co-complexes, fail to reproduce the observed binding specificities unless entropic contributions are explicitly included [80, 81]. The magnitude of the entropy penalty S for each rotamer state W that becomes constrained upon binding can be estimated from Boltzmann's expression ($ST = RT\ln W$) [82], or calculated from observed conformational populations [80, 81, 83]. In α2-3 linkages the ϕ-angle (C1'-C2'-O3-C3) typically populates two rotamers in solution (anti and –gauche with respect to C1') [84], but only one when bound to an HA, resulting in an estimated entropic penalty of approximately 0.4 kcal/mol (at 25 °C). The α2-6 linkage has an additional rotatable bond that leads to multiple conformations, giving rise to an estimated entropic penalty of at least 1.5 kcal/mol [80, 81, 83, 85]. Furthermore, in the case of α2-6 glycans, a curled or open-umbrella topology places more of the glycan substructure in contact with the HA surface than in the case of α2-3 glycans that adopt linear or cone-like topologies. These additional glycan-HA interactions can result in entropic penalties for α2-6 glycans that are larger than those for α2-3 glycans by as much as 5 kcal/mol [81].

Bidentate binding would also be expected to lead to a heightened entropic penalty, due to the overall restriction of motion for such large, flexible glycans, and in particular for the 1-6 linkage in the glycan core. Additionally, interactions between the amino acid side chains and the receptor in the RBS [82, 86] may be entropically disfavored. For example, for K222 to form its

hydrogen bond with the receptor, the long flexible side chain pays an entropic penalty of up to 2 kcal/mol (using $S = R\ln W$) [82]. The more constrained a flexible ligand is by enthalpically-favorable interactions, such as hydrogen bonds and van der Waals contacts, the higher the entropic penalty paid by the system [87], leading to the key concept of enthalpy-entropy compensation [88].

In order to prefer binding to α2-6 glycans over α2-3, the HA must evolve to form proportionally more or stronger interactions with the α2-6 receptor. Thus, although crystallography demonstrates that an avian-adapted HA can form as many (or more) interactions with an α2-6 glycan [78], the resultant entropically-disfavored stiffening of the α2-6 receptor results in a net free energy preference for the α2-3 glycan. For this reason, the number of receptor-HA interactions (Figure 2) is a poor metric for assessing subtle differences in affinity/specificity.

Relating glycan structure to specificity

Glycan array screening has been extensively applied to help define the specificity of influenza hemagglutinins. Overall, the data support the view that HAs from avian-adapted strains prefer α2-3 glycans, while human-infective strains generally prefer α2-6 [44-46, 48, 70, 89-97]. Nevertheless, glycan array screening has also brought to light many exceptions to the accepted view of specificity, and raised new and unanswered questions, particularly related to variations in response as a function of monosaccharide modifications (sulfation, acetylation, etc.) and glycan substructure [48, 97]. Common modifications to the Sia residue include acetylation of the glyceryl side chain (typically at the 9-position), or 5-N-glycolylation (Neu5Gc), which generally attenuate binding to HA from human-infective virus [98, 99]. Remarkably, in contrast to the effect of acetylation, a 9-*O*-lactoyl group appears to restore affinity (H1N1 and H3N2) to levels comparable to the non-derivatized sialoside [98]. Neu5Gc is not produced in humans [100], but can be abundant in non-human species; for example, Neu5Gc-containing glycans are the dominant moieties on epithelial cells from equine trachea [101]. Not surprisingly therefore, HAs from some (but not all) equine-infective influenza strains bind preferentially to glycans containing this modification [102], whereas HA from human-infective strains generally do not [98], explaining the equine/human zoonotic transmission barrier [101].

The sensitivity of binding to glycan substructure is an essential component when defining influenza specificity, but its assessment is complicated by the diversity of possible glycan structures, the influence of glycan substructure on the 3D structure of the sialylated terminus (Figure 2.3), and the differential impact of mutations in the RBS on interactions with glycan modifications [103]. It is impossible to separate the impact of modifications in the glycan from

the overall context of the glycan 3D shape, just as it is impossible to discuss the significance of mutations in the HA independently from the context of the particular subtype. For example, the Gal-2 residue may be linked to GlcNAc-3 at either the 3- or 4-positions (Galβ1-3GlcNAc or Galβ1-4GlcNAc). This chemically subtle difference is often left undefined in glycomic analyses [36], and yet has a dramatic impact on the orientation of the GlcNAc-3 residue relative to Gal-2, flipping the positions of the NAc and O6 moieties in the GlcNAc by approximately 180° degrees in the RBS (Figure 2.3). This difference in glycan substructure would be expected to have a noticeable influence on binding when the HA has evolved to prefer a receptor in which the GlcNAc is modified by sulfation at O6. For example, the HA from an equine H3N8 binds preferentially to 6-sulfated sialosides, but only when the Gal-2-GlcNAc-3 linkage is present in the β1-4 form [102]. For H5 subtypes, 6-O-sulfation of the receptor enhances binding [104] and was predicted to lead to the formation of a salt bridge between the sulfate moiety and K193 [105], which was recently confirmed by crystallography [106]. Similar favorable electrostatic interactions were observed between the same sulfated receptor and K158A in an avian H10 [43].

Another common modification of α2-3 sialosides is α-fucosylation at the 3- or 4-positions of GlcNAc-3. The site of fucosylation depends on the nature of the Gal-2-GlcNAc-3 linkage (β1-4 or β1-3), generating the well-known sialyl Le^X (SLe^x) and SLe^a motifs, respectively. Whether or not fucosylation attenuates affinity has been suggested to depend on the presence or absence of steric collisions with bulky side chains at positions 222 and/or 227 [102, 107-109]. Given the prevalence of SLe^x in mucins, they have been proposed as providing a barrier to infection [110].

In contrast to the 3D properties of the RBS, less attention has been given to a systematic analysis of the conformations of the receptors in the complexes, although it has frequently been observed that the α2-3 linkage adopts a "trans" orientation, resulting in a cone-like topology of the glycan relative to the HA surface [111]. The "cis" orientation of the α2-6 linkage [112, 113] has been further noted to lead such ligands to form a compact, curled, or folded conformation [114] that results in the receptor spanning a larger region of the HA surface, referred to as an open-umbrella topology [111]. The use of the "cis-" descriptor for the Siaα2-6Gal ϕ-angle has become widespread, however it is not useful when comparing the conformation of such linkages in HA complexes, as to date all such linkages adopt this conformation when co-complexed with HAs (Table 2.2). The conformation of the ψ angle in Siaα2-6Gal linkages does however vary, populating only two states, herein denoted "anti-ψ" or "eclipsed-ψ". Moreover, the terms "cis" and "trans" imply that the orientation of the bond is fixed, as in a double bond. As this is not the case for α2-6 or α2-3 linkages, we will refer to the so-called "cis" orientation as "gauche", and the "trans" as "anti". The receptor conformational properties extracted from well-resolved HA-oligosaccharide co-complexes are presented in Table 2.2.

An examination of the data in Table 2.2 indicates that α2-6 linkages adopt two conformations when bound to HAs, which can be defined by the value of the ψ (C2'-O6-C6-C5) angle. Two shapes are also adopted by bound α2-3 linkages, which vary in the ϕ (C1'C2'O3-C3) angle. The significance of these shapes, with respect to the presentation of the receptor in the RBS is illustrated in Figure 2.4. Figure 2.4 illustrates that the open-umbrella topology is associated with the "curled" anti-ψ conformation of an α2-6 linkage (panel A), while the cone-

like topology results from the "extended" anti-ϕ conformation of the α2-3 linkage (panel C). Presented in panels B and D are the alternative conformations of α2-6 (eclipsed-ψ) and α2-3 (–gauche-ϕ) linkages. The trisaccharides in the crystal structures presented in Figure 4 have each been extended to contain three lactosamine repeats to clearly illustrate the impact of the Sia-Gal linkage conformation on the orientation of the glycans. This analysis is consistent with the observations by Xu et al. [81] that the division of the glycan topologies into only cone-like or open-umbrella is insufficient to capture the diversity of glycan conformations in HA complexes.

The recent proposal by Peng et al. [45] that multiantennary α2-6 glycans can form bidentate interactions with trimeric HAs casts new light on the origin of glycan substructure differences. There are several constraints on the ability of a glycan to exhibit bidentate binding. One such constraint is the ability of the antennae to span the distance between two RBSs without steric blocking by HA surface residues, or by HA glycosylation. Another constraint arises from the topologies of the termini of individual glycan branches, which must facilitate orientations conducive to bidentate binding. As shown in Figure 4, *only α 2-6 receptors in a curled anti-ψ conformation satisfy this latter requirement; no known conformations of the α 2-3 receptors promote bidentate binding.* Although the α2-3 oligosaccharides in the –gauche ϕ-conformation (panel D) reach upward from the RBS rather than away (as in panels B and C), their spatial divergence from each other precludes their origination as branches of a single biantennary glycan. Biantennary binding requires that the bound oligosaccharides converge toward a common point in the glycan core (as in panel A). Ultimately, the inability of biantennary α2-3

receptors to form bidentate complexes arises from the linear shape of the α2-3 linkage, which controls the relative orientation of the Siaα2-3Gal disaccharide.

The observation that the α2-3 linkage precludes bidentate binding provides further insight into the functional significance of the cone like (α2-3) versus open-umbrella (α2-6) topologies [111]. The curled anti-ψ conformation of the α2-6 glycans promotes the formation of a bidentate complex, which may also be stabilized by glycan-protein interactions associated with the larger contact area of the open-umbrella topology. Thus, while both α2-3 and α2-6 glycans may in principle form multimeric interactions with an HA, only the α2-6 receptors appear to be able to form bidentate interactions. When glycan density is sufficiently high that two or more glycans can bind simultaneously to the same HA, bidentate binding may offer little enhancement to affinity [95]. However, the ability to form bidentate interactions provides a unique opportunity for the virus to achieve avidity-enhanced binding to α2-6 receptors on a single glycan. This unique capability explains why, despite the overall loss of avidity [44], human-adapted H3N2 viruses retain affinity for a subset of long biantennary α2-6 glycans [45]. Tissue tropism therefore needs to be interpreted not only in terms of composition and spatial distribution of the glycans, but also in the relative density of α2-3 and α2-6 glycans.

Conclusions

Although glycan array screening is a convenient method for examining specificity, developing structure-activity relationships solely on the basis of such data is perilous. Glycan array data should generally be treated qualitatively given that the data are sensitive to numerous factors, including glycan density, glycan linker chemistry, analyte concentration, and detection method [115]. While it is possible to determine surface binding constants ($K_{D,surf}$) using glycan arrays [64, 116], offering an important advantage by quantifying the binding properties of each of the glycans in an array, these protocols are not yet in widespread use. A further factor that significantly complicates the interpretation of array data is the extremely limited diversity of even the largest arrays [46]. This limitation has obvious consequences for ligand discovery [45], and for the elucidation of structure-specificity relationships. Although at present, data from glycan array screening need generally to be treated qualitatively, community-wide standards are being developed [117], which together with more quantitative approaches to data processing [64] and computational analysis [118, 119], will enhance the interpretability of such data. A powerful example of the generation and use of quantitative surface K_D values from glycan array screening was reported by Wong et al. [120]. They were able to dissect the energetic contributions made by each monosaccharide, including the sulfate moieties, in an array of sialosides binding to HAs, showing that the sulfate could enhance binding by nearly 100-fold. Further, by comparing the relative binding energies for each receptor, they were able to conclude that there is likely a competition between favorable binding interactions in the RBS, which the sulfate group maximizes and the fucose sterically blocks.

Crystallographic studies provide unique and crucial atomic-level insight into HA-receptor

interactions, but in the absence of entropic considerations, do not necessarily enable a clear rationalization of specificity. Such interpretations would greatly benefit from the generation of additional quantitative monomeric affinity measurements, as well as from modeling, which may guide the choice of targets for crystallography and array screening. Lastly, variations in HA glycosylation [45, 50, 51] can impact affinity and virulence [59], and should be considered in any analysis of specificity.

Based on agglutination data, glycan array screening, and (albeit limited) biophysical affinity measurements, avian-infective HAs have a clear preference for α2-3 glycans, consistent with the inability of these HAs to compensate for the entropic penalty associated with binding α2-6 glycans. The specificity of human-adapted HAs for α2-6 glycans is more complex, in part because the virus may retain residual affinity for α2-3-receptors, while evolving the ability to bind to α2-6-receptors. For preferential binding of α2-6 linked glycans, mutations must occur in the RBS that overcome the entropic penalty associated with binding to the more flexible α2-6 receptor, and/or which favor the formation of bidentate interactions with multiantennary glycans. The preference for bound α2-6 glycans to adopt an anti-ψ angle (required for bidentate binding) is seen in all well-resolved crystal structures of HAs from human transmissible viruses. This suggests that bidentate binding may be a general mechanism adopted by influenza A to boost affinity for α2-6 receptors, enabling human-to-human transmission.

This review has hopefully illustrated that, despite the challenges in reconciling all of the data relating to influenza A specificity, a molecular interpretation is emerging. The implications of glycan linkage α2-3 or α2-6 on specificity extend beyond the direct interactions between the

terminal Sia-Gal sequence and the HA to more macroscopic features, such as the ability to form multidentate complexes and the need to overcome the inherent entropic penalty associated with binding to α2-6 glycans. A complete understanding of specificity requires a continuous reevaluation of the paradigms with a view to integrating all available data into a holistic analysis.

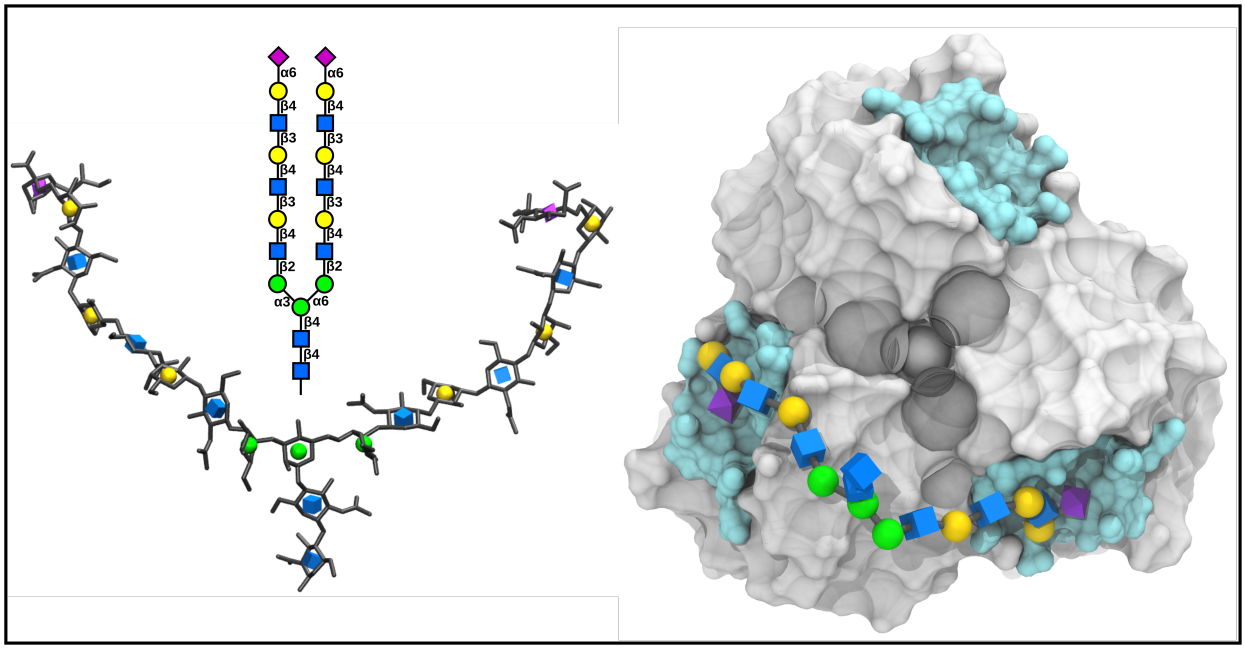


Figure 2. 1. Right: proposed [45] bidentate binding of a biantennary α2-6 glycan (left, 3D-SNFG representation [49]) to the HA (grey surface) from a pandemic H1N1 (A/California/04/2009), residues lining the RBS are shown in cyan. The glycan is shown in the conformation required for bidentate binding.

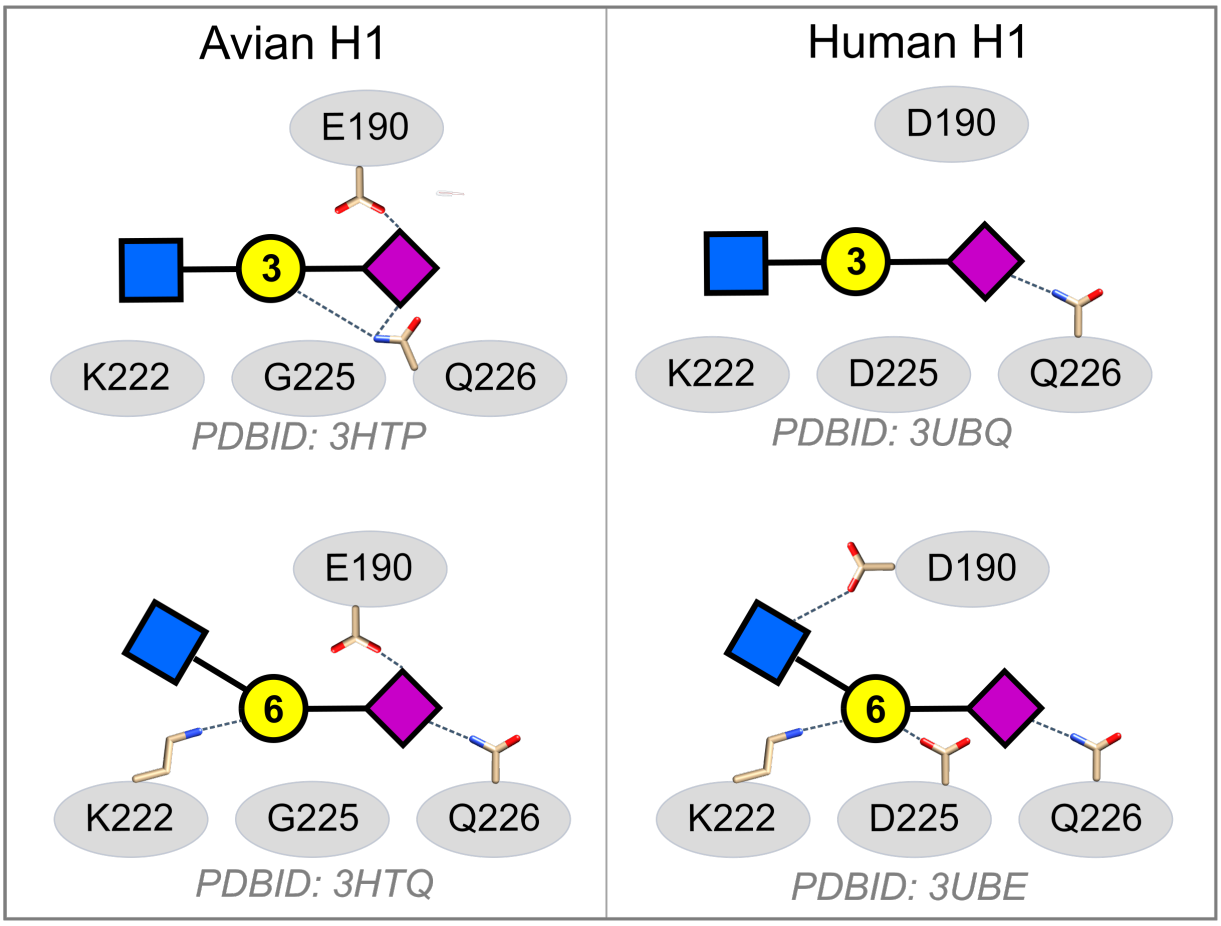


Figure 2.2. Schematic representations of the binding modes for representative avian- (left) and human-adapted (right) HAs from H1N1 viruses binding to α2-3 (upper) or α2-6 (lower) receptor analogs.

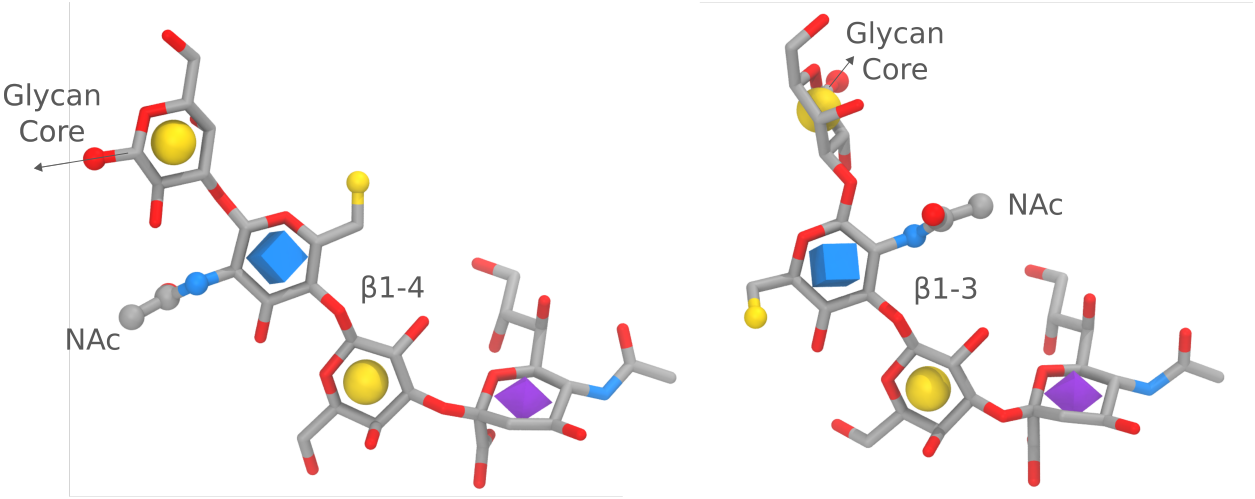


Figure 2. 3. HA receptor structures indicating the influence of the Gal-2 – GlcNAc3 linkage type (left: β1-4, right: β1-3) on conformation and presentation. The structures were retrieved from PDB IDs 4YYA and 4NRL, respectively, and aligned relative to the Sia residues. Note the reversal of the *N*-acetyl moieties relative to the Sia residues. The GlcNAc 6-position, which may be sulfated, is shown as a small yellow sphere.

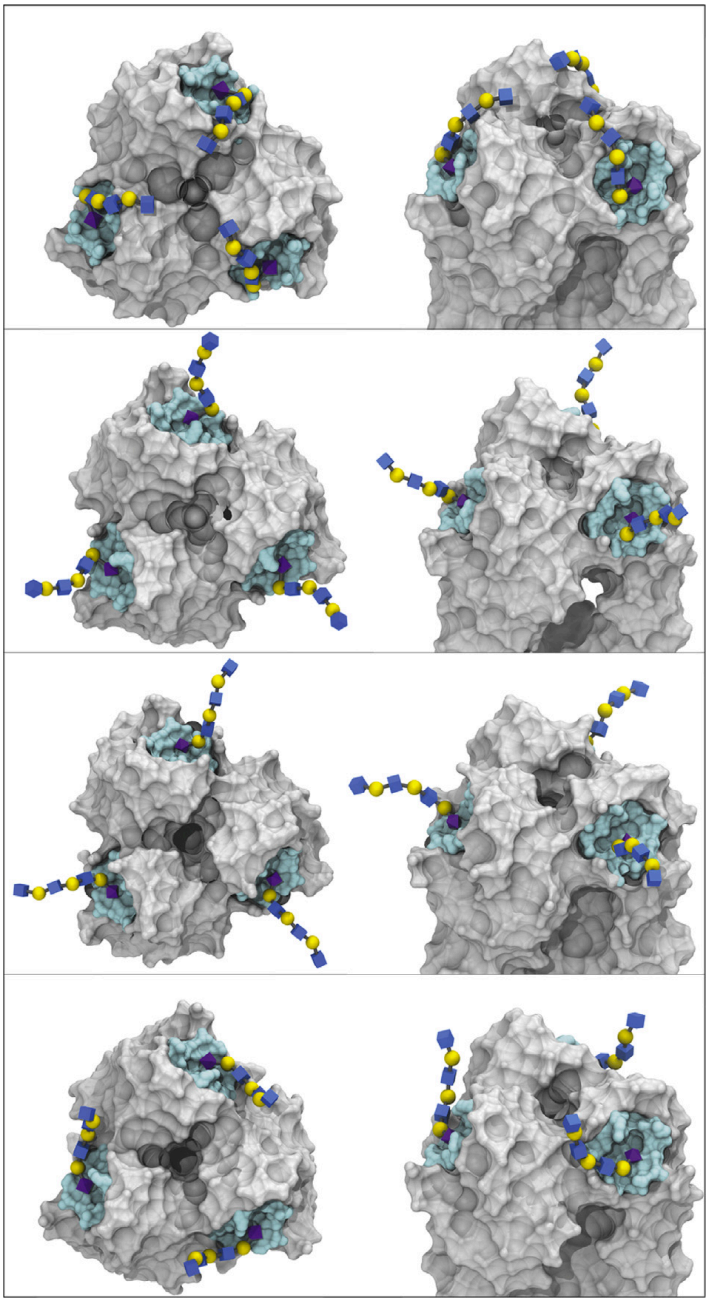


Figure 2. 4. Top (left) and side (right) views of the HA1 domains of four HA co-complexes that illustrate the four common ligand conformations seen in HA-oligosaccharide co-complexes.

Table 2. 1. Monomeric oligosaccharide – HA binding affinities

HA Viral Strain	Canonical Specificity	Ligand	K_D (mM)	ΔG (kcal/mol)[a]
H3N2				
A/Hong Kong/1/1968 [38, 41] (X-31)	α2-6	SiaαOMe	2.8 ± 0.3	-3.5
		3'SLN	3.1 ± 0.4	-3.4
		3'SL	3.6 ± 0.7, 3.2 ± 0.6	-3.3, -3.4
		LSTa	3.8 ± 0.8	-3.3
		6'SLN	2.0 ± 0.2	-3.7
		6'SL	1.7 ± 0.5, 2.1 ± 0.3	-3.8, -3.6
A/Memphis/102/72 [42]	α2-6	SiaαOMe	2.0 ± 1.1	-3.7
		LSTa	8.0	-2.9
		LSTc	1.2	-4.0
H3N2				
A/Hong Kong/1/1968 (X-31) L226Q [41]	α2-3	SiaαOMe	4.7 ± 0.5	-3.2
		3'SL	2.9 ± 0.3	-3.4
		6'SL	5.9 ± 0.7	-3.0
H5N1				
A/Vietnam/1194/04[38]	α2-3	3'SLN	1.1 ± 0.2	-4.0
		3'SL	0.7 ± 0.4	-4.3
		6'SLN	17 ± 3	-2.4
		6'SL	21 ± 6	-2.3
A/Vietnam/1194/04 (ferret transmissible) N158D/N224K/Q226L/T318I [38]	α2-6	3'SLN	32 ± 8	-2.0
		3'SL	43 ± 12	-1.9
		6'SLN	12 ± 2.5	-2.6
		6'SL	17 ± 5	-2.4
H10N8				
A/Jiangxi-Donghu/346/2013[43]	α2-6	3'SLN	1.8 ± 0.39	-3.7
		6'SLN	1.4 ± 0.32	-3.9

[a]At 25 °C.

Table 2.2: Ligand Conformation and HAs co-complexed to α2-6 and α2-3 oligosaccharides (resolution ≤ 2.5 Å)

PDB ID, Ligand	α2-6 ψ[a]	PDB ID, Ligand	α2-3 φ[b]	HA Virus Strain	Amino acids at 190,193,222, 225, 226, 227, 228
H1N1					
1RVT, LSTc	anti	1RV0, LSTa	---[c]	A/Swine/Iowa/15/1930	D,S,K,G,Q,A,G
1RVZ, LSTc	anti	1RVX, LSTa	anti	A/Puerto Rico/8/1934	E,D,K,D,Q,A,G
3UBE, LSTc	anti	3UBJ, LSTa	anti	A/California/04/2009	D,S,K,D,Q,E,G
3UBN, 6'SLN	anti	3UBQ, 3'SLN	anti	A/California/04/2009	D,S,K,D,Q,E,G
3HTQ[d], LSTc	anti	3HTP[d], LSTa	anti	A/Wild Duck/JX/12416/2005	E,T,K,G,Q,A,G
H2N2					
2WR1, LSTc	anti	2WR2, LSTa	anti	A/Chicken/New York/29878/91	E,T,K,G,Q,G,G
2WR4, LSTc	anti	2WR3, LSTa	anti	A/Duck/Ontario/1977	E,T,K,G,Q,G,G
2WR7, LSTc	anti	---		A/Singapore/1/57	E,T,K,G,L,G,S
H3N2					
2YP3, 6'SLN	anti	2YP5, 3'SLN	---[c]	A/Finland/486/2004	D,S,R,D,I,P,S
2YP4, LSTc	anti			A/Finland/486/2004	D,S,R,D,I,P,S
2YP8, 6'SLN	---[c]	2YP9, 3'SLN	---[c]	A/Hong Kong/4443/2005	D,F,R,N,I,P,S
2YPG[d], LSTc	anti			A/Aichi/2/1968-X31	E,S,W,G,L,S,S
H3N8					
		4WA2, 3'SLN	anti	A/harbor seal/Massachusetts/1/2011	E,N,L,G,Q,S,G
H5N1					
1JSO, LSTc	---[c]	1JSN, LSTa	anti	A/Duck/Singapore/3/1997	E,K,K,G,Q,S,G
4BGX, 6'SLN	eclipsed[b]	4BGY[d], 3'SLN	anti	A/Vietnam/1194/2004	E,K,K,G,Q,S,G
4BH0, 6'SLN	eclipsed	4BH1, 3'SLN	anti	A/Turkey/Turkey/1/2005	E,R,K,G,Q,S,G
4BH3, 6'SLN	anti	4BH4, 3'SLN	–gauche[b]	A/Vietnam/1203/2004 (N158D,N224Q,Q226L,T318I)	E,K,K,G,L,S,G
4KDO, LSTc	anti	4KDN, LSTa	---[c]	A/Vietnam/1203/2004 (N158D/N224K/Q226L/T318I)	E,K,K,G,L,S,G
4CQR, 6'SLN	eclipsed	4CQQ[d], 3'SLN	–gauche	A/Vietnam/1194/2004 (S227N,Q196R)	E,K,K,G,Q,R,G

4CQU, 6'SLN	eclipsed	5AJM, 3'SLN	–gauche	A/Vietnam/1194/2004 (N186K)	E,K,K,G,Q,S,G
		4CQY, LSTa	anti	A/turkey/Turkey/1/2005 (Δ133/I155T)	E,R,K,G,Q,S,G
4CQX, 6'SLN	eclipsed	4CQW, 3'SLN	anti	A/turkey/Turkey/1/2005 (Δ133/I155T)	E,R,K,G,Q,S,G
H6N1					
5BR6[a], LSTc	anti	5BR3[d], LSTa	–gauche	A/Taiwan/2/2013	V,N,A,G,Q,R,S
		4XKF, LSTa	–gauche	A/Taiwan/2/2013	V,N,A,G,Q,R,S
4XKG, 6'SLN	eclipsed	4XKE, 3'SLN	–gauche	A/Taiwan/2/2013	V,N,A,G,Q,R,S
H7N3					
4BSH, 6'SLN	eclipsed	4BSI, 3'SLN[d]	anti	A/Turkey/Italy/214845/2002	E,K,Q,G,Q,S,G
H7N9					
4BSB, LSTc	anti			A/Anhui/1/2013 (L20,T135)	E,K,Q,G,L,S,G
4BSC[d], 6'SLN	anti	4BSD, 3'SLN	–gauche	A/Anhui/1/2013 (L20,T135)	E,K,Q,G,L,S,G
4BSE[d], LSTc	anti			A/Anhui/1/2013 (V20,A135)	E,K,Q,G,L,S,G
4LKK[a], 6'SLNLN	anti	4LKJ[d], 3'SLNLN	anti	A/Anhui/1/2013 (L226Q)	E,K,Q,G,Q,S,G
		4N62, 3'SL(6S)N	+gauche[e]	A/Shanghai/2/2013 (L226)	E,K,Q,G,L,S,G
H9N2					
1JSI, LSTc	anti	1JSH, LSTa	–gauche	A/Swine/Hong Kong/9/1998	V,N,L,G,L,H,G
H10N2					
		4CYZ, LSTa	anti	A/mallard/Sweden/51/2002	E,D,Q,G,Q,S,G
H10N8					
4D00, 6'SLN	anti and eclipsed[f]			A/Jiangxi-Donghu/346/2013	E,D,Q,G,Q,S,G

[a]For α2-6 linkages, ψ (C2'-O6-C6-C5) adopts either an anti (188° ± 23) or eclipsed (113° ± 7) conformation. The remaining glycosidic angles adopt a single conformation characterized by average ϕ (C1'-C2'-O6-C6) = -55° ± 11 (-gauche, a.k.a. "cis") and average ω (O6-C6-C5-O5) = 63° ± 17, with the exceptions of 5BR6, where ϕ = 201°, and ω = -41°, 4LKK, where ω = 157°.

[b]For α2-3 linkages, ϕ (C1'-C2'-O3-C3) adopts either an anti (185° ± 10, a.k.a. "trans") or -gauche (-57° ± 8, a.k.a. "cis") conformation. The ψ (C2-O3-C3-C4) glycosidic angle adopt a single conformation average ψ = 100° ± 11. [c]Only the Sia residue is resolved. [d]Resolution > 2.5 Å. [e] Distorted Sia-1 and GlcNAc-3 rings, high B-Factors. [f]Eclipsed in chain E

Reference:

1. Morens, D.M. and J.K. Taubenberger, *Historical thoughts on influenza viral ecosystems, or behold a pale horse, dead dogs, failing fowl, and sick swine.* Influenza Other Respir Viruses, 2010. **4**: 327-37.
2. Johnson, N.P.A.S. and J. Mueller, *Updating the Accounts: Global Mortality of the 1918-1920 "Spanish" Influenza Pandemic* Bull. Hist. Med., 2002. **76**: 105-115.
3. Glaser, L., et al., *A Single Amino Acid Substitution in 1918 Influenza Virus Hemagglutinin Changes Receptor Binding Specificity.* J. Virol., 2005. **79**: 11533-11536.
4. Tumpey, T.M., et al., *A Two-Amino Acid Change in the Hemagglutinin of the 1918 Influenza Virus Abolishes Transmission.* Science, 2007. **315**: 655-659.
5. Reperant, L.A., T. Kuiken, and A.D. Osterhaus, *Adaptive pathways of zoonotic influenza viruses: from exposure to establishment in humans.* Vaccine, 2012. **30**: 4419-34.
6. Wong, S.S. and K.Y. Yuen, *Avian Influenza Virus Infections in Humans.* Chest, 2006. **129**: 156-168.
7. Smith, G.J., et al., *Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic.* Nature, 2009. **459**: 1122-5.
8. Al Hajjar, S. and K. McIntosh, *The first influenza pandemic of the 21st century.* Ann Saudi Med, 2010. **30**: 1-10.
9. Fineberg, H.V., *Pandemic preparedness and response--lessons from the H1N1 influenza of 2009.* N Engl J Med, 2014. **370**: 1335-42.

10. Swerdlow, D.L., L. Finelli, and C.B. Bridges, *2009 H1N1 influenza pandemic: field and epidemiologic investigations in the United States at the start of the first pandemic of the 21st century.* Clin Infect Dis, 2011. **52 Suppl 1**: S1-3.

11. Peiris, J.S., M.D. de Jong, and Y. Guan, *Avian influenza virus (H5N1): a threat to human health.* Clin Microbiol Rev, 2007. **20**: 243-67.

12. Skehel, J.J. and D.C. Wiley, *Receptor Binding and Membrane Fusion in Virus Entry: the Influenza Hemagglutinin.* Annu. Rev. Biochem., 2000. **69**: 531-569.

13. Sriwilaijaroen, N. and Y. Suzuki, *Molecular basis of the structure and function of H1 hemagglutinin of influenza virus.* Proc Jpn Acad Ser B Phys Biol Sci, 2012. **88**: 226-49. A particularly detailed and thorough review.

14. Medina, R.A. and A. Garcia-Sastre, *Influenza A viruses: new research developments.* Nat Rev Microbiol, 2011. **9**: 590-603.

15. Gamblin, S.J. and J.J. Skehel, *Influenza Hemagglutinin and Neuraminidase Membrane Glycoproteins.* J. Biol. Chem, 2010. **285**: 28403-28409.

16. Harris, A., et al., *Influenza virus pleiomorphy characterized by cryoelectron tomography.* Proc Natl Acad Sci U S A, 2006. **103**: 19123-7.

17. Subbarao, K. and T. Joseph, *Scientific barriers to developing vaccines against avian influenza viruses.* Nat Rev Immunol, 2007. **7**: 267-78.

18. Gustin, K.M., et al., *Comparison of the levels of infectious virus in respirable aerosols exhaled by ferrets infected with influenza viruses exhibiting diverse transmissibility phenotypes.* J Virol, 2013. **87**: 7864-73.

19. Thangavel, R.R. and N.M. Bouvier, *Animal models for influenza virus pathogenesis, transmission, and immunology.* J Immunol Methods, 2014. **410**: 60-79.

20. Heldt, F.S., T. Frensing, and U. Reichl, *Modeling the intracellular dynamics of influenza virus replication to understand the control of viral RNA synthesis.* J Virol, 2012. **86**: 7806-17.

21. Frensing, T., et al., *Influenza virus intracellular replication dynamics, release kinetics, and particle morphology during propagation in MDCK cells.* Appl Microbiol Biotechnol, 2016. **100**: 7181-92.

22. Wagner, R., M. Matrosovich, and H.D. Klenk, *Functional balance between haemagglutinin and neuraminidase in influenza virus infections.* Rev Med Virol, 2002. **12**: 159-66.

23. Zhang, H., *Tissue and host tropism of influenza viruses: importance of quantitative analysis.* Sci China C Life Sci, 2009. **52**): 1101-10.

24. Schrauwen, E.J. and R.A. Fouchier, *Host adaptation and transmission of influenza A viruses in mammals.* Emerg Microbes Infect, 2014. **3**: e9.

25. Gabriel, G., et al., *Differential use of importin-alpha isoforms governs cell tropism and host adaptation of influenza virus.* Nat Commun, 2011. **2**: 156.

26. van Riel, D., et al., *Human and avian influenza viruses target different cells in the lower respiratory tract of humans and other mammals.* Am J Pathol, 2007. **171**: 1215-23.

27. Zeng, H., et al., *Tropism and infectivity of influenza virus, including highly pathogenic avian H5N1 virus, in ferret tracheal differentiated primary epithelial cell cultures.* J Virol, 2013. **87**: 2597-607.

28. Ito, T., et al., *Molecular basis for the generation in pigs of influenza A viruses with pandemic potential.* J Virol, 1998. **72**: 7367-73.

29. Lakdawala, S.S., et al., *The soft palate is an important site of adaptation for transmissible influenza viruses.* Nature, 2015. **526**: 122-5.

30. Couceiro, J.N., J.C. Paulson, and L.G. Baum, *Influenza virus strains selectively recognize sialyloligosaccharides on human respiratory epithelium; the role of the host cell in selection of hemagglutinin receptor specificity.* Virus Res, 1993. **29**: 155-65.

31. Nicholls, J.M., et al., *Sialic acid receptor detection in the human respiratory tract: evidence for widespread distribution of potential binding sites for human and avian influenza viruses.* Respir Res, 2007. **8**: 73.

32. de Graaf, M. and R.A. Fouchier, *Role of receptor binding specificity in influenza A virus transmission and pathogenesis.* EMBO J, 2014. **33**: 823-41.

    A thorough review of influenza HA-glycan interactions and their impact on transmission and pathogensis from the perspective of tissue tropism.

33. Miller, G.L. and W.M. Stanley, *Quantitative Aspects of the Red Blood Cell Agglutination Test for Influenza Virus.* J Exp Med, 1944. **79**: 185-95.

34. Rogers, G.N. and J. Paulson, *Receptor Determinants of Human and Animal Influenza Virus Isolates: Differences in Receptor Specificity of the H3 Hemagglutinin Based on Species of Origin.* Virology, 1983. **127**: 361-373.

    The authors established HA binding preferences using RBCs with enzymatically modified glycan sialylation. This work forms the basis for the α2,3/α2,6 specificity paradigm.

35. Ovsyannikova, I.G., et al., *Turkey versus guinea pig red blood cells: hemagglutination differences alter hemagglutination inhibition responses against influenza A/H1N1.* Viral Immunol, 2014. **27**: 174-8.

36. Aich, U., et al., *Glycomics-based analysis of chicken red blood cells provides insight into the selectivity of the viral agglutination assay.* FEBS J, 2011. **278**: 1699-712.

37. Glick, G.D., et al., *Ligand recognition by influenza virus. The binding of bivalent sialosides.* J. Biol. Chem., 1991. **266**: 23660-9.

38. Xiong, X., et al., *Receptor binding by a ferret-transmissible H5 avian influenza virus.* Nature, 2013. **497**: 392-6.

Monomeric binding affinities determined for H3 and H5 HAs by microscale thermophoresis (MST), confirming that avidity is essential for explaining virus specificity.

39. Xu, H.F. and D.E. Shaw, *A Simple Model of Multivalent Adhesion and Its Application to Influenza Infection.* Biophysical Journal, 2016. **110**: 218-233.

40. Mammen, M., S.-K. Choi, and G.M. Whitesides, *Polyvalent Interactions in Biological Systems: Implications for Design and Use of Multivalent Ligands and Inhibitors.* Angew. Chem. Int. Ed., 1998. **37**: 2754-2794.

The paper provides a detailed assessment of the impact of multivalent interactions on binding kinetics.

41. Sauter, N.K., et al., *Hemagglutinins from Two Influenza Virus Variants Bind to Sialic Acid Derivatives with Millimolar Dissociation Constants: a 500-MHz Proton Nuclear Magnetic Resonance Study.* Biochemistry, 1989. **28**: 8388–8396.

Monomeric binding affinities were determined by NMR for H3-glycan interactions, showing that specificity arises from very small differences in binding free energies.

42. Pritchett, T.J., et al., *Recognition of monovalent sialosides by influenza virus H3 hemagglutinin.* Virology, 1987. **160**: 502-6.

The authors report IC_{50} data for small sialosides inhibiting virus-RBC interactions. The paper is a classic pre-glycan-array illustration of determining the α2,3/α2,6 specificity preferences of HAs.

43. Vachieri, S.G., et al., *Receptor binding by H10 influenza viruses.* Nature, 2014. **511**: 475-7.

44. Lin, Y.P., et al., *Evolution of the receptor binding properties of the influenza A(H3N2) hemagglutinin.* Proc Natl Acad Sci U S A, 2012. **109**: 21474-9.

The authors illustrate that H3 avidity loss has resulted from key mutations in the RBS.

45. Peng, W., et al., *Recent H3N2 Viruses Have Evolved Specificity for Extended, Branched Human-type Receptors, Conferring Potential for Increased Avidity.* Cell Host Microbe, 2017. **21**: 23-34.

The authors screen a glycan array that contains a larger diversity of glycans of the type present on human epithelial tissue. They propose that avidity loss from antigenic drift in HAs can be overcome by binding to long branched glycans that terminate in α2,6 linkages that can forming bidentate interactions with the HA.

46. Walther, T., et al., *Glycomic Analysis of Human Respiratory Tract Tissues and Correlation with Influenza Virus Infection.* PloS Pathog., 2013. **9**: e1003223.

An extensive glycomics analysis of glycan structure and distribution on human tissue, in which the authors note that none of the current glycan arrays contain all of the key glycans present in human airways, and thus glycan array data may not be predictive of replication.

47. de Vries, R.P., et al., *Three mutations switch H7N9 influenza to human-type receptor specificity.* PLoS Pathog, 2017. **13**: e1006390.
The authors predict and confirm specificity-switching mutations with potential pandemic significance.

48. Gulati, S., et al., *Human H3N2 Influenza Viruses Isolated from 1968 To 2012 Show Varying Preference for Receptor Substructures with No Apparent Consequences for Disease or Spread.* PLoS One, 2013. **8**: e66325.

49. Thieker, D.F., et al., *3D implementation of the symbol nomenclature for graphical representation of glycans.* Glycobiology, 2016. **26**: 786-7.

50. Sun, S., et al., *Glycosylation site alteration in the evolution of influenza A (H1N1) viruses.* PLoS One, 2011. **6**: e22844.

51. An, Y., et al., *Comparative Glycomics Analysis of Influenza Hemagglutinin (H5N1) Produced in Vaccine Relevant Cell Platforms.* J. Proteome Res., 2013. **12**: 3707-3720.

52. An, Y., et al., *Glycosylation Analysis of Engineered H3N2 Influenza A Virus Hemagglutinins with Sequentially Added Historically Relevant Glycosylation Sites.* J Proteome Res, 2015. **14**: 3957-69.

53. Sun, X., et al., *N-linked glycosylation of the hemagglutinin protein influences virulence and antigenicity of the 1918 pandemic and seasonal H1N1 influenza A viruses.* J Virol, 2013. **87**: 8756-66.

54. Yang, H., et al., *Structure and receptor binding preferences of recombinant human A(H3N2) virus hemagglutinins.* Virology, 2015. **477**: 18-31.

55. Zhang, Y., et al., *Glycosylation on hemagglutinin affects the virulence and pathogenicity of pandemic H1N1/2009 influenza A virus in mice.* PLoS One, 2013. **8**: e61397.

56. Stevens, J., et al., *Glycan Microarray Analysis of the Hemagglutinins from Modern and Pandemic Influenza Viruses Reveals Different Receptor Specificities.* J. Mol. Biol., 2006. **355**: 1143-1155.

57. Jayaraman, A., et al., *Glycosylation at Asn91 of H1N1 haemagglutinin affects binding to glycan receptors.* Biochem J, 2012. **444**: 429-35.

58. Mir-Shekari, S.Y., et al., *The glycosylation of the influenza A virus hemagglutinin by mammalian cells. A site-specific study.* J Biol Chem, 1997. **272**: 4027-36.

59. Tate, M.D., et al., *Playing hide and seek: how glycosylation of the influenza virus hemagglutinin can modulate the immune response to infection.* Viruses, 2014. **6**: 1294-316.

The authors illustrate the importance of HA glycosylation on viral fitness and antigenicity.

60. Deom, C.M., A.J. Caton, and I.T. Schulze, *Host cell-mediated selection of a mutant influenza A virus that has lost a complex oligosaccharide from the tip of the hemagglutinin.* Proc Natl Acad Sci U S A, 1986. **83**: 3771-5.

61. Wang, W., et al., *Glycosylation at 158N of the hemagglutinin protein and receptor binding specificity synergistically affect the antigenicity and immunogenicity of a live attenuated H5N1 A/Vietnam/1203/2004 vaccine virus in ferrets.* J Virol, 2010. **84**: 6570-7.

62. Gao, Y., et al., *Identification of amino acids in HA and PB2 critical for the transmission of H5N1 avian influenza viruses in a mammalian host.* PLoS Pathog, 2009. **5**: e1000709.

63. Zhang, X., et al., *Hemagglutinin glycosylation modulates the pathogenicity and antigenicity of the H5N1 avian influenza virus.* Vet Microbiol, 2015. **175**: 244-56.

64. Liao, H.-Y., et al., *Differential Receptor Binding Affinities of Influenza Hemagglutinins on Glycan Arrays.* J. Am. Chem. Soc., 2010. **132**: 14849-14856.
This paper introduces an analytical method for obtaining surface K_D values from glycan array data. Such approaches facilitate quantitative comparisons of ligand binding.

65. Chen, M.W., et al., *A consensus-hemagglutinin-based DNA vaccine that protects mice against divergent H5N1 influenza viruses.* Proc Natl Acad Sci U S A, 2008. **105**: 13538-43.

66. Alymova, I.V., et al., *Glycosylation changes in the globular head of H3N2 influenza hemagglutinin modulate receptor binding without affecting virus virulence.* Sci Rep, 2016. **6**: 36216.

67. Reid, A.H., et al., *1918 influenza pandemic caused by highly conserved viruses with two receptor-binding variants.* Emerg Infect Dis, 2003. **9**: 1249-53.

68. Zhang, W., et al., *Molecular Basis of the Receptor Binding Specificity Switch of the Hemagglutinins from both the 1918 and 2009 Pandemic Influenza A Viruses by a D225G Substitution.* J. Virol., 2013. **87**: 5949-5958.

69. Connor, R.J., et al., *Receptor Specificity in Human, Avian, and Equine H2 and H3 Influenza Virus Isolates.* Virology, 1994. **205**: 17-23.

70. Stevens, J., et al., *Structure and Receptor Specificity of the Hemagglutinin from an H5N1 Influenza Virus.* Science, 2006. **312**: 404-410.

71. Stevens, J., et al., *Recent avian H5N1 viruses exhibit increased propensity for acquiring human receptor specificity.* J Mol Biol, 2008. **381**: 1382-94.

72. Chen, L.-M., et al., *In Vitro Evolution of H5N1 Avian Influenza Virus Toward Human-Type Receptor Specificity.* Virology, 2012. **422**: 105-113.

73. Imai, M., et al., *Experimental adaptation of an influenza H5 HA confers respiratory droplet transmission to a reassortant H5 HA/H1N1 virus in ferrets.* Nature, 2012. **486**: 420-8.

74. Herfst, S., et al., *Airborne transmission of influenza A/H5N1 virus between ferrets.* Science, 2012. **336**: 1534-41.

75. Wilson, I.A., J.J. Skehel, and D.C. Wiley, *Structure of the Haemagglutinin Membrane Glycoprotein of Influenza Virus at 3 Å Resolution.* Nature, 1981. **289**: 366-373.

76. Russell, R.J., et al., *H1 and H7 influenza haemagglutinin structures extend a structural classification of haemagglutinin subtypes.* Virology, 2004. **325**: 287-96.

The authors illustrate the challenges associated with rationalizing HA specificity based only on HA-oligosaccharide 3D structures.

77. Raman, R., et al., *Glycan-receptor specificity as a useful tool for characterization and surveillance of influenza A virus.* Trends Microbiol, 2014. **22**: 632-41.

The authors illustrate the challenges associated with rationalizing HA specificity based only on HA-oligosaccharide 3D structures.

78. Lin, T., et al., *The hemagglutinin structure of an avian H1N1 influenza A virus.* Virology, 2009. **392**: 73-81.

The paper presents a rare structure of an avian H1, and the authors note the inability of the HA-oligosaccharide co-complexes to explain specificity.

79. vonderLieth, C.W. and T. Kozar, *Towards a better semiquantitative estimation of binding constants: Molecular dynamics exploration of the conformational behavior of isolated sialyllactose and sialyllactose complexed with influenza A hemagglutinin.* Theochem-Journal of Molecular Structure, 1996. **368**: 213-222.

80. Newhouse, E.I., et al., *Mechanism of Glycan Receptor Recognition and Specificity Switch for Avian, Swine, and Human Adapted Influenza Virus Hemagglutinins: A Molecular Dynamics Perspective.* J. Am. Chem. Soc., 2009. **131**: 17430-17422.

81. Xu, D., et al., *Distinct Glycan Topology for Avian and Human Sialopentasaccharide Receptor Analogues upon Binding Different Hemagglutinins: A Molecular Dynamics Perspective.* J. Mol. Biol., 2009. **387**: 465-491.

A good example of the strengths and limitations of computational simulations of HA-glycan complexes, which also indicated the importance of conformational entropy.

82. Doig, A.J. and M.J.E. Sternberg, *Side-Chain Conformational Entropy in Protein Folding.* Protein Sci., 1995. **4**: 2247-2251.

83. Kadirvelraj, R., et al., *Structure and Binding Analysis of Polyporus squamosus Lectin in Complex with the Neu5Acα2-6Galβ1-4GlcNAc Human-type Influenza Receptor.* Glycobiology, 2011. **21**: 973-984.

84. Veluraja, K. and V.S.R. Rao, *Theoretical studies on the conformation of monosialogangliosides and disialogangliosides.* Carbohydr. Polymers, 1983. **3**: 175-192.

85. Sawada, T., D.G. Fedorov, and K. Kitaura, *Role of the Key Mutation in the Selective Binding of Avian and Human Influenza Hemagglutinin to Sialosides Revealed by Quantum-Mechanical Calculations.* J. Am. Chem. Soc., 2010. **132**: 16862-16872.

86. Kadirvelraj, R., et al., *Involvement of Water in Carbohydrate-Protein Binding: Concanavalin A Revisited.* J. Am. Chem. Soc., 2008. **130**: 16933-16942.

87. Komath, S.S., R. Kenoth, and M.J. Swamy, *Thermodynamic analysis of saccharide binding to snake gourd (Trichosanthes anguina) seed lectin. Fluorescence and absorption spectroscopic studies.* Eur J Biochem, 2001. **268**: 111-9.

88. Sharp, K., *Entropy-Enthalpy Compensation: Fact or Artifact?* Protein Sci., 2001. **10**: 661-667.

89. Stevens, D., et al., *Glycan Microarray Technologies: Tools to Survey Host Specificity of Influenza Viruses* Nature, 2006. **4**: 857-864.

90. Blixt, O., et al., *Printed Covalent Glycan Array for Ligand Profiling of Diverse Glycan Binding Proteins.* Proc. Natl. Acad. Sci. USA, 2004. **101**: 17033-17038.

91. Childs, R.A., et al., *Receptor-binding Specificity of Pandemic Influenza A (H1N1) 2009 Virus Determined by Carbohydrate Microarray.* Nat. Biotechnol., 2009. **27**: 797-799.

92. Srinivasan, K., et al., *Quantitative Description of Glycan-Receptor Binding of Influenza A Virus H7 Hemagglutinin.* PLoS One, 2013. **8**: e49597.

93. McBride, R., J.C. Paulson, and R.P. de Vries, *A Miniaturized Glycan Microarray Assay for Assessing Avidity and Specificity of Influenza A Virus Hemagglutinins.* J. Vis. Exp., 2016. **111**: 1-11.

94. Huang, M.L., et al., *Determination of receptor specificities for whole influenza viruses using multivalent glycan arrays.* Chem Commun (Camb), 2015. **51**: 5326-9.

95. Nycholat, C.M., et al., *Recognition of sialylated poly-N-acetyllactosamine chains on N- and O-linked glycans by human and avian influenza A virus hemagglutinins.* Angew Chem Int Ed Engl, 2012. **51**: 4860-3.

96. Kumari, K., et al., *Receptor binding specificity of recent human H3N2 influenza viruses.* Virol J, 2007. **4**: 42.

97. Zhao, N., et al., *Association analyses of large-scale glycan microarray data reveal novel host-specific substructures in influenza A virus binding glycans.* Sci Rep, 2015. **5**: 15778.

98. Song, X., et al., *A sialylated glycan microarray reveals novel interactions of modified sialic acids with proteins and viruses.* J Biol Chem, 2011. **286**: 31610-22.
The paper demonstrates the critical impact of Sia modifications on affinity.

99. Bradley, K.C., et al., *Analysis of Influenza Virus Hemagglutinin Receptor Binding Mutants with Limited Receptor Recognition Properties and Conditional Replication Characteristics.* Journal of Virology, 2011. **85**: 12387-12398.

100. Varki, A., *Loss of N-glycolylneuraminic acid in humans: Mechanisms, consequences, and implications for hominid evolution.* Am J Phys Anthropol, 2001. **Suppl 33**: 54-69.

101. Suzuki, Y., et al., *Sialic Acid Species as a Determinant of the Host Range of Influenza A Viruses.* J. Virol., 2000. **74**: 11825–11831.

102. Gambaryan, A.S., et al., *Receptor-binding profiles of H7 subtype influenza viruses in different host species.* J Virol, 2012. **86**: 4370-9.

103. Stevens, J., et al., *Receptor specificity of influenza A H3N2 viruses isolated in mammalian cells and embryonated chicken eggs.* J Virol, 2010. **84**: 8287-99.

104. Gambaryan, A.S., et al., *H5N1 chicken influenza viruses display a high binding affinity for Neu5Acalpha2-3Galbeta1-4(6-HSO3)GlcNAc-containing receptors.* Virology, 2004. **326**: 310-6.

105. Gambaryan, A., et al., *Evolution of the receptor binding phenotype of influenza A (H5) viruses.* Virology, 2006. **344**: 432-8.

106. Xiong, X., et al., *Recognition of sulphated and fucosylated receptor sialosides by A/Vietnam/1194/2004 (H5N1) influenza virus.* Virus Res, 2013. **178**: 12-4.

107. Gambaryan, A.S., et al., *6-sulfo sialyl Lewis X is the common receptor determinant recognized by H5, H6, H7 and H9 influenza viruses of terrestrial poultry.* Virol J, 2008. **5**: 1-10.

108. Collins, P.J., et al., *Recent evolution of equine influenza and the origin of canine influenza.* Proc Natl Acad Sci U S A, 2014. **111**: 11175-80.

109. Hiono, T., et al., *Amino acid residues at positions 222 and 227 of the hemagglutinin together with the neuraminidase determine binding of H5 avian influenza viruses to sialyl Lewis X.* Arch Virol, 2016. **161**: 307-16.

110. Linden, S.K., et al., *Mucins in the mucosal barrier to infection.* Mucosal Immunol, 2008. **1**: 183-97.

111. Chandrasekaran, A., et al., *Glycan Topology Determines Human Adaptation of Avian H5N1 Virus Hemagglutinin.* Nat. Biotechnol., 2008. **26**: 107-113.

The paper introduces the cone-like and open-umbrella topology descriptions for α2-3 and α2-6 glycan-HA co-complexes.

112. Zhang, W., et al., *An airborne transmissible avian influenza H5 hemagglutinin seen at the atomic level.* Science, 2013. **340**: 1463-7.

113. Viswanathan, K., et al., *Glycans as receptors for influenza pathogenesis.* Glycoconj J, 2010. **27**: 561-70.

114. Xu, R., et al., *Structural Characterization of the Hemagglutinin Receptor Specificity from the 2009 H1N1 Influenza Pandemic.* J. Virol., 2012. **86**: 982-990.

115. Oyelaran, O. and J.C. Gildersleeve, *Glycan Arrays: Recent Advances and Future Challenges.* Curr. Opin. Chem. Biol., 2009. **13**: 406-413.

116. Raman, R., et al., *Glycan receptor specificity as a useful tool for characterization and surveillance of influenza A virus.* Trends Microbiol, 2014. **22**: 632-41.

117. Liu, Y., et al., *The minimum information required for a glycomics experiment (MIRAGE) project: improving the standards for reporting glycan microarray-based data.* Glycobiology, 2017. **27**: 280-284.

118. Grant, O.C., et al., *Combining 3D structure with glycan array data provides insight into the origin of glycan specificity.* Glycobiology, 2016. **26**: 772-783.

The paper illustrates an approach to interpreting glycan array data in terms of 3D binding motifs, providing a mechanism for rationalizing glycan-substructure effects.

119. Grant, O.C., et al., *Presentation, presentation, presentation! Molecular-level insight into linker effects on glycan array screening data.* Glycobiology, 2014. **24**: 17-25.

The paper illustrates the use of structure modeling to detect artifacts in glycan array data.

120. Wang, C.-C., et al., *Glycans on influenza hemagglutinin affect receptor binding and immune response.* P. Natl. Acad. Sci., 2009. **106**: 18137-18142.

The paper shows the insights gained from quantifying binding data from glycan array screening.

CHAPTER 3

QUANTIFYING WEAK GLYCAN-PROTEIN INTERACTIONS USING A BIOLAYER INTERFEROMETRY COMPETITION ASSAY: APPLICATIONS TO ECL LECTIN AND X-31 INFLUENZA HEMAGGLUTININ.

[1]Ye Ji and Robert J. Woods. Accepted by *Glycobiophysics*. Reprinted here with permission of publisher, [March 25th, 2018].

Abstract

This chapter introduces two formats for Biolayer Interferometry competition assays to determine the solution K_D of weak glycan-protein interactions. This approach overcomes the challenge of determining weak interactions, while minimizing the amounts of reagents required. Accurate solution K_D values aid in understanding the complex relationships between monomeric versus multimeric interactions, as well as affinity versus avidity. The assays have been applied to a well-studied lectin (*Erythrina cristagalli* lectin) and influenza hemagglutinin (X-31). The solution K_D values determined from this approach are in good agreement with previous reported literature values from isothermal titration calorimetry and NMR. Additionally, this approach appears robust and precise.

Keywords

Bio-Layer Interferometry, weak interaction, glycan-protein interaction, BLI-based inhibition assay, ECL, X-31

Abbreviations

Bio-Layer Interferometry (BLI), Consortium for Functional Glycomics (CFG), *erythrina cristagalli* lectin (ECL), enzyme-linked immunosorbent assay (ELISA), hemagglutinin (HA), isothermal titration calorimetry (ITC), microscale thermophoresis (MST), surface plasmon resonance (SPR), neuraminidase (NA), nuclear magnetic resonance spectroscopy (NMR), three-dimensional (3D).

Assay Format 1: Immobilized glycan, protein analyte.

Lectins such as ECL have long been used as carbohydrate detection reagents, and considerable data are available regarding their affinities. Thus ECL was chosen as a standard for development of the present protocol. With the protocol established, we then examined the biologically significant interaction of influenza hemagglutinin with glycans associated with human and avian infection.

Example: *Erythrina cristagalli* Lectin (ECL) – glycan interactions.

ECL is widely used as a reagent for the detection of terminal galactopyranose (Gal) residues in glycans (its canonical specificity is for Gal), yet it also binds to *N*-acetylgalactosamine (GalNAc) and fucosylated Galactose (L-Fucα1-2Gal). Although its function in the legume is unknown, understanding the complex specificity of lectins, such as ECL, is fundamental to the interpretation of specificity data, as for example in histology studies (1), and in the rational design of diagnostic and therapeutic agents (2) that target specific glycans. It has been reported that ECL binds to lactosamine (LacNAc) at 0.26mM (3) and 0.32mM (4) both from ITC measurement. Such weak binding affinities are on the edge of or below the detection limit of SPR and BLI. In this example, the biotinylated oligosaccharides will be immobilized on a streptavidin (SA) biosensor surface. And the ECL will be the analyte for direct binding as $K_{D,\ surface}$. As the second step in the assay, the ability of eight oligosaccharides to inhibit the ECL binding to the LacNAc biosensor surface was quantified in terms of their IC_{50} values (Table 3.1). Subsequently, solution K_i values of the inhibitors were derived from Equation 1 (Table 3.2). ECL, like many lectins is dimeric, so this assay format might be biased by multimeric (avidity)

interactions between the lectin and the immobilized oligosaccharides. However, the agreements (Table 3.2) between the reported monomeric solution K_D values from ITC measurements and ours suggests that the dimeric ECL this particular format does not form multiple simultaneous interactions with the surface. The surface K_D and representative sensor plots for direct binding are presented in Figure 3.1.

$$K_i = IC_{50} / (1 + [ECL]/K_{D,surface}) \quad [1]$$

Figure 3.1 shows that the surface K_D determined by fitting to a 1:1 binding model (0.92 μM ± 0.02 (stdev from triplicates)), which compared well to the value form the Scatchard analysis: 0.91 μM ± 0.02 (stdev from triplicates).

Assay Format 2: Immobilized protein, glycan analyte.

Many studies have applied direct ligand binding assays to measure the surface/apparent K_D of ligand-receptor interactions (5-7). However, effects from mass transfer, steric effects (8), and avidity (9) can lead to a considerable variation in the K_D values (3-4, 10-11). Additionally, when the ligands are low molecular weight, they may be below the confident detection limit of the assay. Nevertheless, the surface K_D values in many cases are useful for defining analyte specificities if not solution affinities (12). However, the understanding of monomeric molecular interactions, it would be extremely beneficial to compare to K_D values that do not have avidity effects (13).

It was fortuitous in the case of ECL that the protein did not appear to form multimeric interactions with the immobilized oligosaccharides. However, this would not be expected to be the case with a multimeric protein such as trimeric influenza hemagglutinin, which many have specifically evolved to prefer to bind in a multimeric mode to cell-surface oligosaccharides (12-15). To evade such multimeric interactions, we investigated the present assay by immobilizing the protein (HA), and detecting the binding to the relevant receptor oligosaccharides.

The canonical view of the relationship between HA receptor specificity and species infectivity is that the HA in human infective viruses prefers to bind to glycans present on the human cell surface that terminate with the Siaα2-6Gal sequence, whereas the HA in avian-infective viruses prefers to bind to glycans that terminate in Siaα2-3Gal (13). The discovery of the α2-6/α2-3 infectivity relationship originated not from quantitative biophysical studies, but from more qualitative, yet robust, hemagglutination assays (16). Turning from array screening, lectin staining, and agglutination assays to biophysical assays in order to quantify the

relationship between HA sequence and glycan specificity introduces additional challenges. What complicates the quest for a structural rationalization of HA specificity is that many HAs co-crystalize with sialosides independent of whether the sialic acid is in an α2-3 or α2-6Gal linkage. Additionally, solution binding data confirms that the interactions are weak at the monomeric level (mM) and that there is little measurable difference in affinity for either α2-3 or α2-6 sialosides (13).

The Case study: Human A/Hong Kong/1/1968 (X-31) H3N2 HA –glycan interactions.

X-31 is one of the first influenza HAs to have its solution binding K_D characterized (by NMR) (17), thus we chose it for the present study. The direct binding of α2-3 or α2-6 trisaccharides (3'or 6'-Sialyl-*N*-acetyllactoseamine, short for 3'/6'SLN) to X-31 resulted in signals that were too weak to interpret quantitatively, and the approximate equilibrium K_D values were not able to be determined (Figure 3.2 and 3.3).

In order to enhance the signal, the biotinylated glycans (molecular weight approximately 765 Da) were combined with an anti-biotin antibody Fab fragment (18), creating a monomeric neoglycoprotein with molecular weight approximately 50 kDa. The BLI sensorgram in Figure 8 shows that utilizing the neoglycoprotein dramatically amplified the maximum BLI signal (6.0 nm) relative to the oligosaccharide alone (0.12 nm).

An initial $K_{D,Surface}$ for 6'SLN-Fab binding to immobilized X-31 was measured as 1.15μM (see inset in Figure 3.4 for 1:1 association binding fit). When corrected for non-specific binding using an irrelevant neoglycoprotein (biotinylated mannose-Fab), the final $K_{D,Surface}$ for 6'SLN-Fab was 1.58μM.

In the second step of the assay, IC_{50} values are measured using soluble (non-biotinylated) oligosaccharides to inhibit the corresponding neoglycoprotein from binding to the immobilized X-31 (Figure 3.5).

The $K_{D,solution}$ values (Table 3.3) were then simply derived by employing the IC_{50} values and $K_{D,surface}$ values in Equation 1. Using the Man-Fab neoglycoconjugate as a correction for non-specific binding modestly altered the final $K_{D,solution}$ (for 6'SLN) from 2.17 mM (no Man-Fab subtraction) to 2.36 mM (after Man-Fab subtraction).

Thus accounting for the Fab component of the analyte altered the KD values by approximately 10%, but did not alter the relative specificity of 6'SLN versus 3'SLN for X-31 (Table 3.3)

Concluding remarks.

The competition assay formats appear to be highly suited to measuring weak (mM) solution K_D values for oligosaccharide-protein interactions. Not only does this approach provide accurate values, it minimized reagent use, which can be critical, especially for scarce reagents such as complex glycans.

For proteins that cannot form multimeric interactions, Assay Format 1 or 2 may be employed, however to minimize avidity effects from multivalent binding Assay Format 2 is likely to be preferable. In the case of ECL, it appears that although it is dimeric, if cannot form multimeric interactions with the immobilized oligosaccharides, possibly due to the spacing of the biotinylated oligosaccharides on the SA surface. The fact that the ECL ligands were all short (disaccharides) may also contribute to the inability of ECL to form multimeric interactions in Assay Format 1. Assay Format 1 may also be beneficial when the oligosaccharide is the limiting reagent.

When both the protein and the oligosaccharides are in scarce supply, Assay Format 2 offers the benefit of being able to recover the oligosaccharides used in the IC_{50} step using a molecular weight cutoff spin column to separate the oligosaccharide from the neoglycoprotien analyte. Oligosaccharide recovery is not as efficient in Assay Format 1 because some amount of the oligosaccharide will remain bound to the protein analyte.

Assay Format 1 Experimental Details

Experimental materials: Erythrina Cristagalli Lectin (ECL, Cat#: L-1140, Vector Lab, Burlingame, CA, USA), N-Acetyl-D-lactosamine (LacNAc, Cat# A7791, Sigma-Aldrich, St. Louis, MO, USA), D-Lactose monohydrate (Lac, Cat#: 61339, Sigma-Aldrich, St. Louis, MO, USA), 4-O-β-Galactopyranosyl-D-mannopyranose (epi-Lac, Cat#: G0886, Sigma-Aldrich, St. Louis, MO, USA), 2-Fucosyllactose (2'FucLac, Cat#:OF06739, Carbosynth Limited, Berkshire, UK), D-(+)-Cellobiose (Cellobiose, Cat# 22150, Sigma-Aldrich, St. Louis, MO, USA), and Maltose (Maltose, Cat# 0168-17, DIFCO&BD, Detroit, MI, USA) were purchased from their commercial resources. 2-Fucosyllactose- $OCH_2CH_2CH_2$-N_3 (2'FucLac-N3, CFG#: Tr120) and 2-Fucosyllactosamine-$OCH_2CH_2CH_2$-N_3 (2'FucLacNAc-N3, CFG#: Tr117) were requested from Consortium for Functional Glycomics (CFG). Biotinylated glycan Galβ1-4GlcNAcβ-$OCH_2CH_2CH_2NH$-biotin (LacNAc-biotin) was received as a gift from Dr. Nicolai Bovin. ECL was weighted and dissolved in the ECL buffer: 10 mM HEPES, 15 mM NaCl, 0.1 mM $CaCl_2$, and 0.1 mM $MnCl_2$ buffered at pH7.4, at 25°C.

Protein BLI direct binding assay ($K_{D,\ surface}$): Ligand LacNAc-biotin was loaded onto streptavidin biosensors (SA, Cat#: 18-5019, Pall ForteBio Corp., Menlo Park, CA, USA) at 1 μM for 1800s. Then the loaded LacNAc biosensors were dipped into 0.1μM EZ-link™ Hydrazide-Biocytin (biocytin, Cat#: 28020, Thermo Scientific, Rockford, IL, USA) for blocking the possible unoccupied biotin-SA binding sites for 1800s. The immobilization of ligand onto SA biosensors resulted in ~0.3nm as loading signal under this condition. ECL direct binding K_D (LacNAc biosensor surface K_D) was measured using a BioLayer Interferometer (BLI) Octet Red 96 system

(Pall ForteBio Corp., Menlo Park, CA, USA) and data acquired using ForteBio Data Acquisition 8.2 software (Pall ForteBio Corp., Menlo Park, CA, USA). The protein direct binding experiment was performed at 600s for association and 1800s for dissociation in ECL buffer. ECL was prepared in two-fold serial dilution in ECL buffer from 0~50 μM, in the replicates of three. Surface K_D ($K_{D,surface\ LacNAc\ biosensor}$) was then calculated by ForteBio Data Analysis 8.2 software (Pall ForteBio Corp., Menlo Park, CA, USA) and Microsoft Office Excel 2011 (Microsoft, USA). Surface K_D ($K_{D,surface\ LacNAc\ biosensor}$) was determined by 1:1 binding model from both steady state analysis and Scatchard plot (Figure 5) and resulted in 0.92 (STDEV: 0.02) μM of triplicates.

Protein BLI inhibition assay (IC_{50}): ECL protein was prepared at 2μM in ECL buffer in a large volume for protein inhibition assay. Eight compounds were tested in the inhibition assay including six compounds (inhibitors) and a negative control compound (non-ECL binder glycan). All the compounds were prepared in two-fold serial dilution in ECL buffer from 0,1.25, 2.5, 5, 10, 20, 40, and 80mM. 100μL of 2μM ECL, 20μL of prepared inhibitor/non-binder at its concentration, and 80μL of ECL buffer were mixed and incubated at room temperature for 1hour. ECL inhibition assay was performed on Octet Red 96 at baseline time 120s, association time 600s, and dissociation time 1800s at shaker speed 1000RPM at room temperature, in replicates of three. IC_{50} was calculated by using three-parameter dose-response inhibition model in GraphPad Prism 7 (GraphPad, La Jolla, CA, USA). The compounds LacNAc and LacNAc–sp-N_3 result in similar IC_{50} values, therefore, only the final values for LacNAc are reported. And

also this data shows that the azide group attached to the compound LacNAc–sp-N_3 does not affect binding.

Assay Format 2 Experimental Details

Experimental materials: Influenza A H3N2 (X-31) Hemagglutinin (X-31, Cat#: 40059-V08H-50, Sino Biological, China), 3'-Sialyl-*N*-acetyllactoseamine (3'SLN, Cat#: SLN302, Dextra, UK), 6'-Sialyl-*N*-acetyllactoseamine (6'SLN, Cat#: SLN306, Dextra, UK) were purchased from their commercial resources. Biotinylated glycan Siaα2-3Galβ1-4GlcNAcβ-$OCH_2CH_2CH_2NH$-biotin (3'SLN-biotin, CFG#B84), Siaα2-6Galβ1-4GlcNAcβ-$OCH_2CH_2CH_2NH$-biotin (6'SLN-biotin, CFG#B87), and azido glycan Siaα2-3Galβ1-4GlcNAcβ-$OCH_2CH_2CH_2$-N_3 (3'SLN-N_3, CFG#Tr33), and Siaα2-6Galβ1-4GlcNAcβ-$OCH_2CH_2CH_2$-N_3 (6'SLN-N_3, CFG#Tr36) were requested from Consortium for Functional Glycomics (CFG). Anti-biotin-Fab antibody was purchased from Rockland Inc. (Cat# 800-101-098). Fab-glycan was prepared as incubating Fab with biotinylated glycans at 1:1.1 mole ratio in dart at 4°C overnight. X-31 was weighted and dissolved in the HA buffer: PBS buffer at pH7.4, at 25°C. Fab-glycans were all prepared in HEPES buffer at pH7.4 at 25°C.

Immobilization of HA on streptavidin biosensors: Amine Reactive Second Generation biosensors (AR2G, Cat#: 18-5092, Pall ForteBio Corp., Menlo Park, CA, USA) were activated by dipped into a mixture of 20mM EDC and 10mM sulfo-NHS solution for 900s. Then H3N2 X-31 HA protein was coupled onto Amine Reactive Second Generation biosensors (AR2G, Cat#: 18-5092,

Pall ForteBio Corp., Menlo Park, CA, USA) at 20 μg/mL for 1800s. Then the loaded X-31 H3 HA biosensors were dipped into 1M ethanolamine (biocytin, Cat#: 28020, Thermo Scientific, Rockford, IL, USA) pH8.5 to quench the possible unreacted carboxylic group on the AR2G biosensor surface for 600s. The immobilization of HA onto AR2G biosensors resulted in ~3nm as loading signal under this condition.

Protein BLI direct binding assay ($K_{D,\ surface}$): Fab-glycan direct binding K_D (X-31 biosensor surface K_D) was measured using an BioLayer Interferometer (BLI) Octet Red 96 system (Pall ForteBio Corp., Menlo Park, CA, USA) and data acquired using ForteBio Data Acquisition 8.2 software (Pall ForteBio Corp., Menlo Park, CA, USA). The protein direct binding experiment was performed at 360s for association and 240s for dissociation in HEPES buffer pH7.4. Fab-glycan was prepared in two-fold serial dilution in HEPES buffer from 0~8μM, in the replicates of three. Surface K_D ($K_{D,\text{surface Fab-glycan biosensor}}$) was then calculated by ForteBio Data Analysis 8.2 software (Pall ForteBio Corp., Menlo Park, CA, USA) and Microsoft Office Excel 2011 (Microsoft, USA). Surface K_D ($K_{D,\text{surface Fab-glycan biosensor}}$) was determined by 1:1 binding model from both steady state analysis and scatchard plot and in triplicates.

Protein BLI inhibition assay (IC_{50}): Fab-glycan was prepared at 1μM in HEPES buffer in a large volume for protein inhibition assay. 3'SLN(Tr33), 6'SLN(Tr36), 3'SDLN(Te175), and 6'SDLN(Te176) were prepared in two-fold serial dilution in water from 0,1.25, 2.5, 5, 10, 20, 40, and 80mM. 100μL of 1μM Fab-glycan, 20μL of prepared azido-glycan at its concentration, and 80μL of HEPES buffer were mixed and incubated at room temperature for 1hour. X-31

inhibition assay was performed on Octet Red 96 at baseline time 120s, association time 360s, and dissociation time 240s at shaker speed 1000RPM at room temperature, in replicates of three. IC_{50} was calculated by using three-parameter dose-response inhibition model in GraphPad Prism 7 (GraphPad, La Jolla, CA, USA).

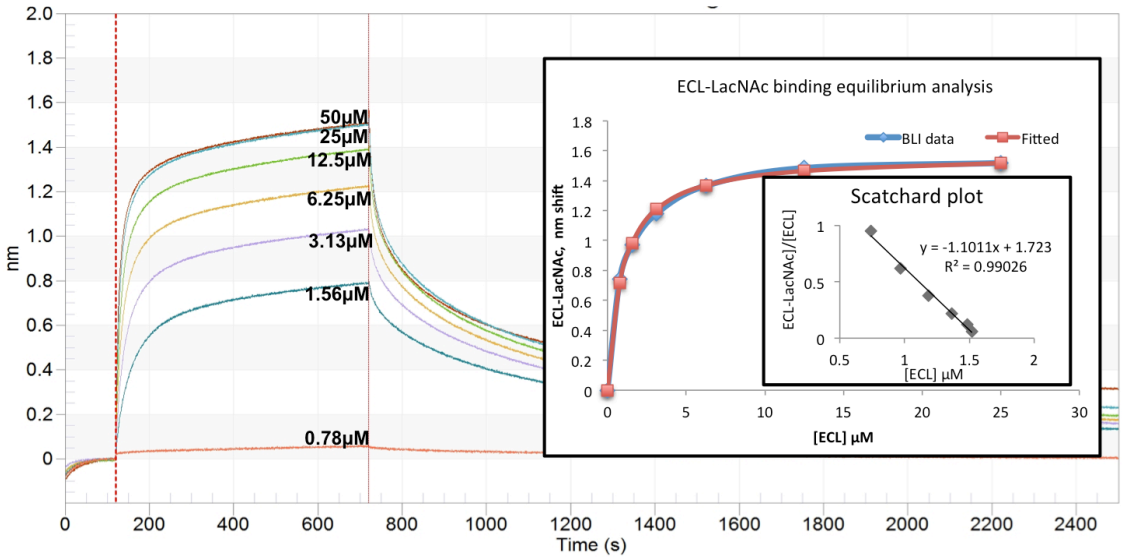


Figure 3.1. Left, BLI sensorgram of ECL direct binding to LacNAc on SA biosensors. Inset, the $K_{D,surface}$ resulting from equilibrium binding analysis and associated Scatchard plot.

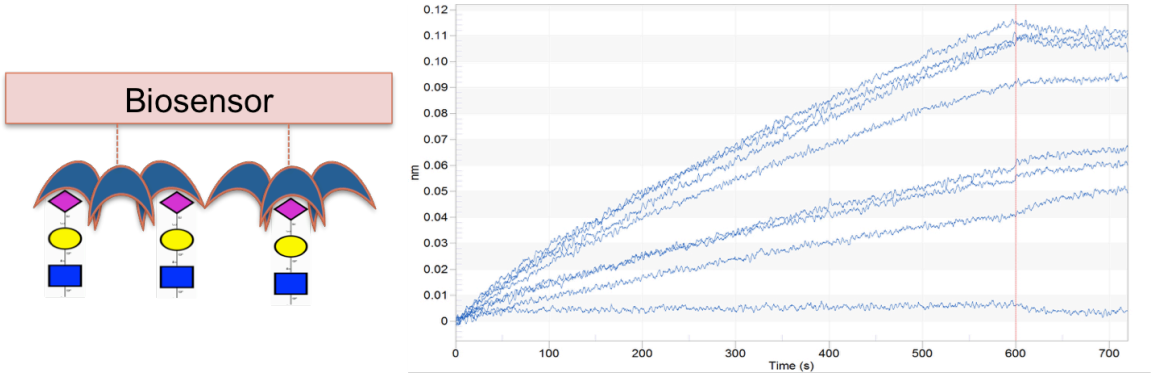


Figure 3.2. BLI sensorgrams for a serial dilution of the glycan (3'SLN-N_3) direct binding to X-31 H3 HA biosensors. The highest concentration of the glycan is 4mM

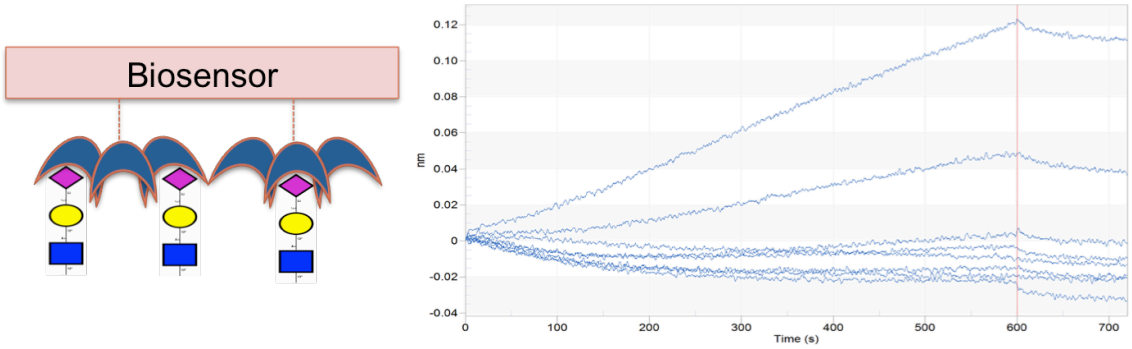


Figure 3.3. BLI sensorgrams of a serial dilution of the glycan (6'SLN-N_3) direct binding to X-31 H3 HA biosensors. The highest concentration of the glycan is 4mM.

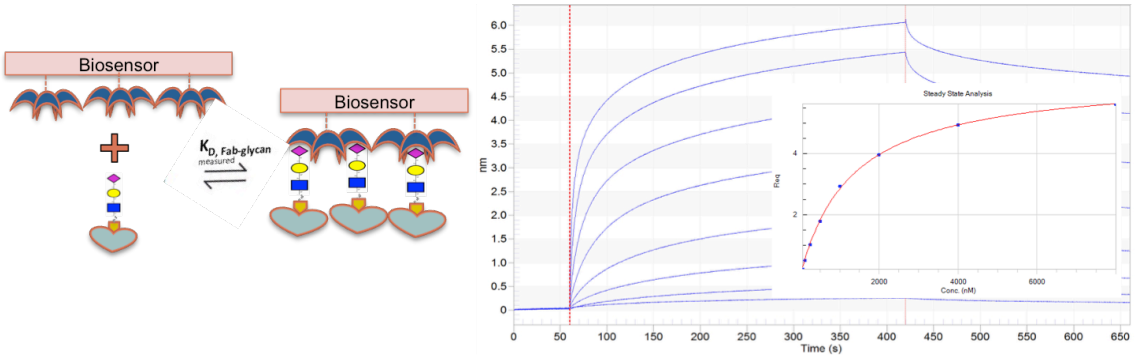


Figure 3.4. Scheme of Fab-glycan binding directly to HA immobilized biosensor surface.

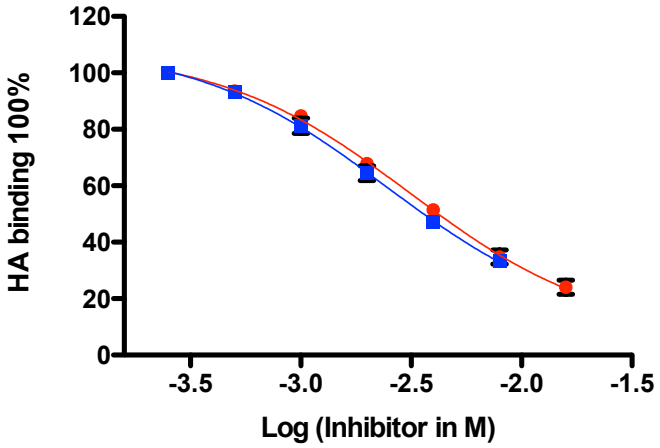


Figure 3.5. IC50 curve for 6'SLN inhibiting the binding of 6'SLN-Fab (red) and 3'SLN inhibiting the binding of 3'SLN-Fab (blue) to immobilized X-31.

Table 3.1. IC_{50} of all carbohydrate candidates, experiments reported in replicates of three.

Analyte	Average (mM)	STDEV (mM)
Lactose (Lac)	0.66	0.044
epi-Lactose	0.44	0.012
LacNAc	0.17	0.011
2'-FucosylLacNAc-sp[b]-N_3	0.07	0.014
2'-FucosylLactose	0.49	0.050
2'-FucosylLactose-N_3	0.47	0.027
Maltose	ND[a]	ND
Cellobiose	ND	ND

[a]ND not detected.

[b]sp = -CH2CH2-

Table 3.2. Solution K_D (mM) values for several analytes binding to ECL

Oligosaccharide	BLI	ITC (3)	ITC (4)
Lactose (Lac)	0.32	0.26	0.32
epi-Lactose	0.21	NR[a]	NR
LacNAc	0.08	0.09	NR
2'-FucosylLacNAc-sp[b]-N_3	0.03	NR	NR
2'-FucosylLactose-sp-N_3	0.22	NR	NR
2'-FucosylLactose	0.23	NR	0.31
Maltose	ND[c]	NR	NR
Cellobiose	ND	NR	NR

[a]NR not reported.

[b]sp = -CH2CH2-

[c]ND not detected.

Table 3.3. Solution K_D determined from BLI-based inhibition assay

HA Viral Strain	Canonical Specificity	Ligand	$K_{D,surface}$ (μM)[a]	IC_{50} (mM)	K_D (mM)	Reported K_D (mM)
A/Hong Kong/1/1968 (X-31)	α2-6	3'SLN-N_3	1.7	3.1 ± 0.47[b]	2.4 ± 0.12	NR[c]
		3'SLN	1.7	3.6 ± 0.13	2.7 ± 0.01	3.2 ± 0.6[(10)] 3.13 ± 0.4[(11)]
		6'SLN-N_3	1.2	3.1 ± 0.12	2.2 ± 0.03	NR
		6'SLN	1.2	3.8 ± 0.29	2.7 ± 0.07	2.1 ± 0.3[(10)] 2.03 ± 0.2[(11)]

[a]At 25°C.

[b]Errors are standard deviation of the mean

[c]NR not reported.

References

1. Tong, H. H., James, M., Grants, I., Liu, X., Shi, G., and DeMaria, T. F. (2001) Comparison of structural changes of cell surface carbohydrates in the eustachian tube epithelium of chinchillas infected with a Streptococcus pneumoniae neuraminidase-deficient mutant or its isogenic parent strain. *Microb Pathog* **31**, 309-317
2. Farhadi, S. A., and Hudalla, G. A. (2016) Engineering galectin-glycan interactions for immunotherapy and immunomodulation. *Exp Biol Med (Maywood)* **241**, 1074-1083
3. Gupta, D., Cho, M., Cummings, R. D., and Brewer, C. F. (1996) Thermodynamics of Carbohydrate Binding to Galectin-1 from Chinese Hamster Ovary Cells and Two Mutants. A Comparison with Four Galactose-Specific Plant Lectins. *Biochemistry* **35**, 15236-15243
4. Notenboom, V., Boraston, A. B., Williams, S. J., Kilburn, D. G., and Rose, D. R. (2002) High-Resolution Crystal Structures of the Lectin-like Xylan Binding Domain from Streptomyces liVidans Xylanase 10A with Bound Substrates Reveal a Novel Mode of Xylan Binding. *Biochemisrty* **41**, 4246-4254
5. Khurana, S., King, L. R., Manischewitz, J., Coyle, E. M., and Golding, H. (2014) Novel antibody-independent receptor-binding SPR-based assay for rapid measurement of influenza vaccine potency. *Vaccine* **32**, 2188-2197
6. Matsumoto, H., Shinzaki, S., Narisada, M., Kawamoto, S., Kuwamoto, K., Moriwaki, K., Kanke, F., Satomura, S., Kumada, T., and Miyoshi, E. (2010) Clinical application of a lectin-antibody ELISA to measure fucosylated haptoglobin in sera of patients with pancreatic cancer. *Clin. Chem. Lab. Med.* **48**, 505 - 512

7. Makeneni, S., Ji, Y., Watson, D. C., Young, N. M., and Woods, R. J. (2014) Predicting the origins of anti-blood group antibody specificity: A case study of the ABO A- and B-antigens. *Front. Immunol.* **5**, 1-9

8. Schuck, P., and Zhao, H. (2010) The role of mass transport limitation and surface heterogeneity in the biophysical characterization of macromolecular binding processes by SPR biosensing. *Methods Mol Biol* **627**, 15-54

9. Liao, H.-Y., Hsu, C.-H., Wang, S.-C., Liang, C.-H., Yen, H.-Y., Su, C.-Y., Chen, C.-H., Jan, J.-T., Ren, C.-T., Chen, C.-H., Cheng, T.-J. R., Wu, C.-Y., and Wong, C.-H. (2010) Differential Receptor Binding Affinities of Influenza Hemagglutinins on Glycan Arrays. *J. Am. Chem. Soc.* **132**, 14849-14856

10. Sauter, N. K., Hanson, J. E., Glick, G. D., Brown, J. H., Crowther, R. L., Park, S. J., Skehel, J. J., and Wiley, D. C. (1992) Binding of Influenza Virus Hemagglutinin to Analogs of its Cell-surface Receptor, Sialic Acid: Analysis by Proton Nuclear Magnetic Resonance Spectroscopy and X-ray Crystallography. *Biochemistry* **31**, 9609-9621

11. Xiong, X., Coombs, P. J., Martin, S. R., Liu, J., Xiao, H., McCauley, J. W., Locher, K., Walker, P. A., Collins, P. J., Kawaoka, Y., Skehel, J. J., and Gamblin, S. J. (2013) Receptor binding by a ferret-transmissible H5 avian influenza virus. *Nature* **497**, 392-396

12. de Vries, R. P., Peng, W., Grant, O. C., Thompson, A. J., Zhu, X., Bouwman, K. M., de la Pena, A. T. T., van Breemen, M. J., Ambepitiya Wickramasinghe, I. N., de Haan, C. A. M., Yu, W., McBride, R., Sanders, R. W., Woods, R. J., Verheije, M. H., Wilson, I. A., and Paulson, J. C. (2017) Three mutations switch H7N9 influenza to human-type receptor specificity. *PLoS Pathog* **13**, e1006390

13. Ji, Y., White, Y. J., Hadden, J. A., Grant, O. C., and Woods, R. J. (2017) New insights into influenza A specificity: an evolution of paradigms. *Curr Opin Struct Biol* **44**, 219-231

14. Peng, W., de Vries, R. P., Grant, O. C., Thompson, A. J., McBride, R., Tsogtbaatar, B., Lee, P. S., Razi, N., Wilson, I. A., Woods, R. J., and Paulson, J. C. (2017) Recent H3N2 Viruses Have Evolved Specificity for Extended, Branched Human-type Receptors, Conferring Potential for Increased Avidity. *Cell Host Microbe* **21**, 23-34

15. Khurana, S., Verma, S., Verma, N., Crevar, C. J., Carter, D. M., Manischewitz, J., King, L. R., Ross, T. M., and Golding, H. (2010) Properly Folded Bacterially Expressed H1N1 Hemagglutinin Globular Head and Ectodomain Vaccines Protect Ferrets against H1N1 Pandemic Influenza Virus. *PLoS One* **5**, e11548

16. Miller, G. L., and Stanley, W. M. (1944) Quantitative Aspects of the Red Blood Cell Agglutination Test for Influenza Virus. *J Exp Med* **79**, 185-195

17. Sauter, N. K., Bednarski, M. D., Wurzburg, B. A., Hanson, J. E., Whitesides, G. M., Skehel, J. J., and Wiley, D. C. (1989) Hemagglutinins from Two Influenza Virus Variants Bind to Sialic Acid Derivatives with Millimolar Dissociation Constants: a 500-MHz Proton Nuclear Magnetic Resonance Study. *Biochemistry* **28**, 8388–8396

18. Cornell, B. A., Braach-Maksvytis, V. L. B., King, L. G., Osman, P. D. J., Raguse, B., Wieczorek, L. and Pace, R. J. (1997) A biosensor that uses ion-channel switches. *Nat Lett* **387**, 580-583

CHAPTER 4

DEFINING THE SPECIFICITY OF "NON-SPECIFIC" CARBOHYDRATE-PROTEIN INTERACTIONS: QUANTIFYING FUNCTIONAL GROUP CONTRIBUTIONS

[1]Amika Sood, [1]Ye Ji, [1]Oksana O. Gerlits, , Nicolai V. Bovin, Leighton Coates, and Robert J. Woods

Submitted to *Journal of Chemical Information and Modeling* , March 5th, 201

Abstract

Protein-carbohydrate interactions are significant in a wide range of biological processes, disruption of which has been implicated in many different diseases. The capability of glycan-binding proteins (GBP) to specifically bind to the corresponding glycans allows GBPs to be utilized in glycan biomarker detection or conversely to serve as targets for therapeutic intervention. However, understanding the structural origins of GBP specificity has proven to be challenging due to their typically low binding affinities (mM) and their potential to display broad or complex specificities. Here we perform MD simulations and post-MD energy analyses with the Poisson Boltzman and Generalized Born solvent models (MM-PB/GBSA) of the *Erythrina cristagalli* lectin (ECL) with its known ligands, and with new co-crystal structures reported herein. While each MM-PB/GBSA parameterization resulted in different estimates of the desolvation free energy, general trends emerged that permit us to define GBP binding preferences in terms of ligand sub-structure specificity. Additionally, we have further decomposed the theoretical interaction energies into contributions made between chemically-relevant functional groups. Based on these contributions, the functional groups in each ligand can be assembled into a pharmacophore comprised of groups that are either critical for binding, or enhance binding, or are non-interacting. It is revealed that the pharmacophore for ECL consists of the galactopyranose (Gal) ring atoms along with C6, and the O3 and O4 hydroxyl groups. This approach provides a convenient method for identifying and quantifying the glycan pharmacophore and provides a novel method for interpreting glycan specificity that is independent of residue-level glycan nomenclature. A pharmacophore approach to defining

specificity is readily transferable to molecular design software, and therefore may be particularly useful in designing therapeutics (glycomimetics) that target GBPs.

KEYWORDS: lectin, carbohydrate-binding protein, crystal structure, carbohydrate structure, molecular dynamics, biolayer interferometry, GLYCAM, MM-GBSA, SMILES, InChI

ABBREVIATIONS

GBP, glycan-binding proteins;

ECL, *Erythrina cristagalli* lectin;

MD, Molecular Dynamics;

MM, Molecular Mechanics;

MM-PB/GBSA, Molecular Mechanics-Poisson–Boltzmann/Generalized Born surface area;

PDB, Protein Data Bank;

RMSD, Root Mean Square Deviation

Introduction

The recognition of glycans present on cell surfaces as glycoconjugates lies at the heart of a number of biological processes in animals, plants, and microorganisms (1). Non-covalent glycan-protein interactions are involved in cellular adhesion, innate immunity, bacterial and viral infection, as well as plant defense mechanisms and other processes (2-7). Glycan-binding proteins (GBPs) such as lectins, adhesins, toxins, antibodies, carbohydrate-binding modules, are often multimers that possess the ability to crosslink cells, which is essential for cell signaling (8) and the disruption of recognition can lead to conditions such as delay in muscle fiber development. The multimeric structure of most carbohydrate-binding proteins serves also to enhance the apparent affinity of the binding processes through avidity effects (9). The affinity of monomeric carbohydrate-protein interactions is typically weaker than μM, and yet the specificity appears to arise primarily from the structure of monomeric complexes (10).

Much of our understanding of carbohydrate recognition has come from crystallographic studies of plant lectins, because these proteins are often relatively stable, crystallize readily, and have a wide range of receptor specificities. More recently, glycan array screening has been widely applied to define specificity. However, the specificity of lectins (11) and anti-carbohydrate antibodies (12) can appear complex. Nevertheless, plant lectins have found widespread use as affinity reagents in the separation and characterization of oligosaccharides, and glycoconjugates (13), and are often employed in staining and histochemistry of cells and tissues (14-16). For example, the legume lectin from *Erythrina cristagalli* (ECL) is widely used as a reagent for the detection of terminal galactopyranose (Gal) residues in glycans (its canonical specificity is for Gal), yet it also binds to *N*-acetylgalactosamine (GalNAc) and fucosylated Gal

(Fucα1-2Gal). Although its function in the legume is unknown, understanding the complex specificity of lectins, such as ECL, is fundamental to the rational design of diagnostic and therapeutic agents that target specific glycans (17).

Numerous experimental methods have been used to quantify the affinity of GBP-carbohydrate interactions, including isothermal titration calorimetry (ITC), NMR spectroscopy, microscale thermophoresis (MST), biolayer interferometry (BLI), surface plasmon resonance (SPR), frontal affinity chromatography (FAC), and ELISA-based assays. Data from different experimental techniques can result in conflicting definitions of specificity, depending on the sensitivity of the method and on the presence or absence of avidity effects. This is particularly clear in the case of weak interactions, which may be observed by NMR (18) or MST (19), but not by glycan array screening (20-21). Given the widespread use of glycan array screening, it has become the *de facto* method for defining the specificity of GBPs, and yet often requires amplification of the signal through multimerization of the protein analyte (22). Although glycan array screening is a high throughput method capable of screening hundreds of glycans, it is often unable to detect weak monomeric interactions and does not provide structural insights into the origin of the observed specificity and cross-reactivity. While site-directed mutagenesis of the protein (23) or chemical modification of ligand (24) can be used to probe the mode of binding in the past, protein crystallography is by far the most widely used method to define the binding mode. However, crystallography often employs high ratios of ligand to a protein, and the ligand is typically only a small fragment of the intact glycan, leading to questions as to the biological relevance of the co-complex (25). Given the high flexibility of glycans, it is not surprising then that these complex macromolecules are resistant to crystallization, making it difficult to

determine the molecular structures for all but the simplest glycan fragments. Thus, experimental techniques alone can prove to be insufficient to understand the mechanism of low-affinity carbohydrate recognition. However, when these techniques are coupled with computational analyses, it can lead to an improved grasp of the underlying reasons behind the specificity of carbohydrate-protein interactions.

From a structural perspective, binding to the protein requires the carbohydrate to form interactions (hydrogen bonds, van der Waals contacts, hydrophobic contacts) that are specific in terms of geometry and charge complementarity. Discrimination between potential binders depends on differences in affinity, which depends on the strengths of individual interatomic (or inter-functional group) interactions. However, it is challenging to quantify these interactions experimentally, as any physical alteration to the protein (such as a point mutation) or to the ligand (such as a chemical modification) could perturb more than the local interaction, aside from the significant effort that may be required. Thus, an opportunity exists to exploit computational methods to estimate the energetic contributions made by individual interacting groups. There are a number of theoretical methods capable of estimating receptor-ligand affinities with varying levels of accuracy and computational cost (26), including thermodynamic integration (TI), free energy perturbation (FEP), and MM-PB/GBSA (molecular mechanics-Poisson–Boltzmann/Generalized Born surface area). While equilibrium methods such as TI and FEP are generally more accurate than end-point methods like MM-GBSA, achieving sufficient conformational sampling is only practical for TI/FEP calculations if the ligands differ only slightly in structure; calculating the binding energy difference between ligands that differ by one or more monosaccharide is currently impractical. In contrast, MM-PB/GBSA methods are less

size-limited, and by default are therefore the methods most widely applied for predicting the energetics of carbohydrate-protein complexes. Although the absolute interaction energies from MM-GBSA analyses typically overestimate the experimental binding free energies, the relative interaction energies can be useful in identifying structural features responsible for the observed experimental affinities (27).

Here we perform molecular dynamics (MD) simulations of complexes of ECL with six ligands: lactose (28-29) (Galβ1-4Glcβ, Lac, **1**), epi-lactose (Galβ1-4Manβ, Epilac, **2**), *N*-acetyllactosamine (Galβ1-4GlcNAcβ, LacNAc, **3**), *N,N*-diacetyllactosamine (GalNAcβ1-4GlcNAcβ , LacDiNAc, **4**), fucosylated lactose (Fucα1-2Galβ1-4Glcβ, FucLac, **5**) (28), and fucosylated *N*-acetyllactosamine (Fucα1-2Galβ1-4GlcNAcβ, FucLacNAc, blood group H trisaccharide, **6**) (Figure 4.1). The MM-GBSA method is then used to compute absolute affinities, as well as inter-residue and inter-group interaction energies. This approach enables us to identify key components of the ligand that are responsible for the observed experimental specificity and to quantify their relative contributions. In addition, we report a novel crystal structure of ECL in complex with *N*-acetyllactosamine and epi-lactose, and new experimental affinities for seven di- or trisaccharides. From a theoretical perspective, the results illustrate the current accuracy limitations of the computational methods.

The results from the present analysis provide an explanation for the observed specificity of ECL in terms of a sub-structure of ligand features, leading to the definition of a ligand pharmacophore that explains the inhibitory power of a range of reported monosaccharides (30). The ability to computationally detect glycan pharmacophores should advance both the

engineering of GBPs with modified ligand specificities (31) and conversely, the development of glycomimetic therapeutics (32-33).

Materials and Methods

Crystallization. A sample of ECL was dissolved in 100 mM NaCl, 20 mM HEPES pH 7.5, 0.1 mM $CaCl_2$ and 0.1 mM $MnCl_2$ to a concentration of ~ 7 mg/mL. About 1 hour prior to crystallization, the solution of ECL was combined with the aqueous solution, 0.25 mM, of the particular ligand at a molar ratio of 1:10 (ECL:ligand). Crystals were grown by the vapor diffusion at 20-22 ^{0}C using the sitting drop method. For ECL with *N*-acetyl-D-lactosamine complex screening with QIAGEN's the JCSG Core I Suite resulted in diffraction quality crystals of pyramidal shape from several conditions: #10, 12, 13, 20, 22, and 31. The best crystals were obtained from either 0.2 M calcium acetate hydrate or potassium sodium tartrate and 20 % PEG 3350, corresponding to conditions 20 and 22. The crystals grew from 1 µL sitting drop Intelli-Plates. Co-crystals of ECL with epi-lactose were obtained from 10 µL drops in microbridges using well solutions containing 0.2 M calcium acetate, 0.1 M HEPES pH 7.5, 14-16 % PEG 3350.

Data collection. For both complexes, X-ray crystallographic data were collected from frozen crystals at 100K. Prior to data collection crystals were placed in a cryoprotectant solution composed of 75% well solution and 25% glycerol and then flash cooled by immersion in liquid nitrogen. For ECL-*N*-acetyl-D-lactosamine complex diffraction data were collected using an ADSC Quantum 315r detector at the Advanced Photon Source (APS) on the ID19 beamline SBC-CAT to 1.9 Å resolution. For ECL-epi-lactose co-crystal crystallographic data were collected to 2.2 Å using a Rigaku HomeFlux system, equipped with a MicroMax-007 HF generator, Osmic VariMax optics, and an RAXIS-IV++ image-plate detector. X-ray diffraction

data were collected, integrated and scaled using HKL3000 software suite (34). The structure was solved by molecular replacement using CCP4 suite (35). The structure of the binary complex of ECL with lactose (PDB ID 1UZY) (29) was used as a starting model with all waters, ligands including the *N*-linked glycosylated saccharide and metal ions removed. Refinement was completed using the *phenix.refine* program in the *PHENIX* (36) suite and the resulting structure analyzed with molprobity (37). The structures were built and manipulated with program *Coot* (38), whereas the figures were generated using the *PyMol* molecular graphics software (v.1.5.0.3; Schrödinger LLC). A summary of the crystallographic data and refinement is given in Table 4.2.

BLI binding experiment: ECL (Cat#: L-1140, Vector Lab, Burlingame, CA, USA), **3** (Cat# A7791, Sigma-Aldrich, St. Louis, MO, USA), **1** (Cat#: 61339, Sigma-Aldrich, St. Louis, MO, USA), **2** (Cat#: G0886, Sigma-Aldrich, St. Louis, MO, USA), **5** (Cat#:OF06739, Carbosynth Limited, Berkshire, UK), **6** (provided by the Consortium for Functional Glycomics) and **7** (Cat# 22150, Sigma-Aldrich, St. Louis, MO, USA) were purchased from their commercial resources. Biotinylated glycan Galβ1-4GlcNAcβ-$OCH_2CH_2CH_2NH$-biotin (LacNAc-biotin) was received as a gift from Dr. Nicolai Bovin. ECL was weighted and dissolved in the ECL buffer: 10 mM HEPES, 15 mM NaCl, 0.1 mM $CaCl_2$, and 0.1 mM $MnCl_2$ buffered at pH7.4, at 25°C.

Protein BLI direct binding assay ($K_{D,surface}$): Ligand (LacNAc-biotin) was loaded onto streptavidin biosensors (SA, Cat#: 18-5019, Pall ForteBio Corp., Menlo Park, CA, USA) at 1 μM for 1800s. Then the loaded LacNAc biosensors were dipped into 0.1μM EZ-link™

Hydrazide-Biocytin (biocytin, Cat#: 28020, Thermo Scientific, Rockford, IL, USA) for blocking the possible unoccupied biotin-SA binding sites for 1800s. The immobilization of ligand onto SA biosensors resulted in ~0.3nm loading signal under this condition. ECL direct binding K_D (LacNAc biosensor surface KD) was measured using a BioLayer Interferometer (BLI) Octet Red 96 system (Pall ForteBio Corp., Menlo Park, CA, USA) and data acquired using ForteBio Data Acquisition 8.2 software (Pall ForteBio Corp., Menlo Park, CA, USA). The protein direct binding experiment was performed for 600s for association and 1800s for dissociation in ECL buffer. ECL was prepared in two-fold serial dilution in ECL buffer from 0~50 μM, in the replicates of three. The surface K_D ($K_{D,surface\ LacNAc\ biosensor}$) was calculated to be 0.92 μM with a standard deviation of 0.02 μM from triplicate measurements using the ForteBio Data Analysis 8.2 software (Pall ForteBio Corp., Menlo Park, CA, USA) assuming a 1:1 binding model.

Protein BLI inhibition (IC_{50}) assay and K_D derivation: ECL protein was prepared at 2μM in ECL buffer. Eight compounds were tested in the inhibition assay including six inhibitors: **1**, **2**, **3**, **5**, **6** (present as the β-azido glycoside), and a non-ECL binder **7**. All the compounds were prepared in two-fold serial dilution in ECL buffer from 0,1.25, 2.5, 5, 10, 20, 40, and 80mM. 100μL of 2μM ECL, 20μL of prepared inhibitor/non-binder at its concentration, and 80μL of ECL buffer were mixed and incubated at room temperature for 1hour. ECL inhibition assay was performed on Octet Red 96 at baseline time 120s, association time 600s, and dissociation time 1800s at shaker speed 1000 RPM at room temperature, in replicates of three. IC_{50} was calculated by using three-parameter dose-response inhibition model in GraphPad Prism 7 (GraphPad, La Jolla, CA, USA). Solution K_D values for each inhibitor were calculated from the equation: $K_{D,solution} = IC_{50}/(1 +$

[ECL]/$K_{D,surface}$) (39). IC_{50} values and associated BLI sensorgrams are reported in Table 4.S1 and Figure 4.S1, respectively; inhibition curves are shown in Figure 4.S2.

Molecular Dynamics: Crystal structures of ECL in complex with **1**, **2**, **3** and **5**, along with the 3D models of **4** and **6** in complex with ECL were used for performing MD simulations. The GLYCAM-Web server (www.glycam.org) was used to generate 3D structures of **4** and **6**, which were then superimposed on **3** and **5** respectively to get the complex structures. All the waters of crystallization and ions were retained, while the *N*-glycan at N113 was removed from the crystal structures, retaining only N113. The missing hydrogen atoms were added to the protein and crystal waters using the Reduce tool, provided by AMBERTOOLS (58), which also sets the protonation state of HIS residues, and detects and corrects flipped amide or imidazole groups in the side chains of ASN, GLN, and HIS residues. The ionization states of the ionizable side chains (ASP, GLU, ARG, LYS) were set appropriately for a neutral pH, and kept in that state throughout the simulation. Hydrogen atoms in the ligand were assigned from the GLYCAM06 monosaccharide structure files using the tLeap module of AMBERTOOLS. These structures were then minimized *in vacuo* to get rid of any steric clashes by steepest descent (SD) minimization for 5000 steps followed by 20000 steps of conjugate gradient (CG) minimization. The net charge on the systems were neutralized by adding counter ions (6 Na^+ ions), followed by solvation in a truncated octahedral box with pre-equilibrated TIP3P water molecules, using the tLEAP module provided by AMBERTOOLS. Initially, the water molecules were allowed to relax around the solute, by performing SD minimization (5000 steps) followed by CG minimization (20000 steps), while the solute atoms were restrained (500 kcal/mol-$Å^2$). The final

stage of minimization was performed without any restraints using the same SD/CG steps involved in the previous stage. Each system was then heated from 5 K to 300 K over a span of 50 ps, under NVT conditions followed by a 1 ns equilibration under NPT conditions using the pmemd.cuda version of AMBER14 (40). The simulations were performed using the ECL monomer (extracted from the homodimer), therefore positional restraints (10 kcal/mol-$Å^2$) were applied to the Cα atoms in the protein backbone. The MD simulations were performed under the same conditions as equilibration for 100 ns.

Binding affinity and entropy calculations: These analyses were carried out on 30,000 snapshots extracted evenly from 30ns of MD simulation using a single trajectory method with the MMPBSA.py.MPI module of AMBER (41). The net binding energies (and entropies) were computed as the difference between those for the complex minus those for the protein and ligand. Quasi-harmonic (QH) entropies were calculated using the cpptraj module of AMBERTOOLS (42) and extrapolated to an infinite simulation period by fitting a linear regression curve to entropy as a function of inverse simulation period (43) (Figure 4.S4). Three different sets of snapshots were obtained by extracting every third frame from a 100ns simulation, starting from a different initial frame, generating three independent extrapolated entropies, which were then averaged to estimate the error range. Normal mode (NM) entropy calculations were performed using the MMPBSA.py.MPI module. As normal mode analysis is exceptionally computationally costly, it was performed using 100 snapshots from the simulation (44). Nevertheless, a trial calculation using 250 snapshots from a simulation of ECL in complex with **1** resulted in a net NM entropy value (-19.2 kcal/mol) comparable to that from 100

snapshots (-19.0 kcal/mol). Conformational entropies associated with changes in the glycosidic torsion angle distributions that occur upon binding were computed using the Karplus–Kushick approach.

Results

Specificity of ECL: ECL is a legume lectin with Galβ1-4GlcNAc as the preferred binding motif. A number of experimental studies have been performed to determine and compare the affinity of ECL for various monosaccharides and sugars (28,45). Binding studies performed here using BLI compare well with reported values obtained by ITC (28, 34), and show that lactose (Galβ1-4Glcβ**, 1**), epi-lactose (Galβ1-4Manβ, EpiLac, **2**), and fucosylated lactose (Fucα1-2Galβ1-4Glcβ, FucLac, **5**) are equivalent binders, while the introduction of an *N*-acetyl moiety into the Glc residue enhances affinity, as in *N*-acetyllactosamine (Galβ1-4GlcNAcβ, LacNAc, **3**) and 2′-Fucosyl-*N*-acetyllactosamine (Fucα1-2Galβ1-4GlcNAcβ, FucLacNAc, Blood group H trisaccharide, **6)** (Table 4.1). Neither cellobiose (Glcβ1-4Glcβ, **7**) nor maltose (Glcα1-4Glcβ, **8**) shows any measurable affinity for ECL. Interestingly, data from glycan array screening of ECL indicates that **1** and **5** are non-binders, while only **3**, **6**, and GalNAcβ1-4GlcNAcβ (LacDiNAc, **4**) are binders [46]. The false negative binding observed in the glycan array data for **1** and **5** may indicate the relative weakness of the binding of these ligands and suggests a need for caution when employing glycan array screening to define glycan-binding specificity for low-affinity ligands. While affinity measurements can indicate which regions of the ligand may be important for binding, a detailed rationalization can best be obtained from examination of the 3D structures of the complexes.

*Crystal structure of ECL in complex with **2** and **3**:* To study the structural effects of the ligand binding, the crystal structure of ECL bound to EpiLac and LacNAc was determined at 2.2 Å and

1.9 Å resolution respectively (Table 4.2). The electron density maps clearly demonstrate binding of **2** and **3** in the combining site (Figure 4.2B and 4.2C). The X-ray structures of native ECL, two in complex with Lac (PDB ID 1UZY, 1GZC) and one with FucLac (PDB ID 1GZ9) were determined previously (28-29). All ECL crystal structures indicate that there is only one binding site per monomer, which is characterized by a shallow groove. All the ligands occupy the same binding site with Gal and Glc residues residing in equivalent positions in each of the complexes (Figure 4.2). Assuming that all of the known ligands bind ECL in a similar fashion with Gal in the binding pocket, 3D models of **4** and **6** in complex with ECL were created. 3D structures for **4** and **6** ligands were retrieved from the GLYCAM-Web server (www.glycam.org), and models for their complexes with ECL were generated by superimposing the coordinates for the ring atoms on to those present in the complex with **3** and **5** respectively (Figure 4.2D and 4.2F).

The binding site for the Gal residue is formed by A88, D89, G107 and N133, which are highly conserved among related legume lectins (47) and participate in four important H-bond interactions with the sugar. In this hydrogen-bonding network, carboxylic oxygen atoms of D89 form two equivalently strong hydrogen bonds with O4 and O3 of Gal and are H-bond acceptors, whereas both the main chain NH of G107 and NH_2 group of N133 are H-bond donors in their weaker interactions with O3 of Gal. Relative to **1**, the fucosyl residue in **5**, and the *N*-acetyl group in **3**, **4** and **6** form additional hydrogen bonds and van der Waals contacts with the protein (Figure 4.3).

It is notable that despite the presence of presumably favorable interactions with the fucosyl residue, the affinity of **5** is not significantly different than **1**, while ECL possesses about 3-fold higher affinity and more favorable enthalpy for **3**, suggesting a need to examine the

interaction energies in detail. 3D structures alone can provide at best only a qualitative guide to the impact of any given intermolecular interaction on the affinity of the ligand. Computational simulations, employing accurate 3D structures, can permit structure-function relationships to be derived that include the critical contributions from molecular motion, solvation, and entropy.

Structural basis of ligand recognition: To examine and compare the stabilities and strengths of the interactions of each of the ligands with ECL, each complex was subjected to molecular dynamics (MD) simulation (100 ns) with backbone restraints (10 kcal/mol-Å^2) in the presence of explicit water, using the AMBER12SB/GLYCAM06j (48-49) force field. The ligand-protein complexes remained stable over the course of the simulations (average ligand displacement RMSD: **1** = 0.86 Å, **2** = 0.83 Å, **3** = 1.03 Å, **4** = 1.09 Å, **5** = 0.85 Å, **6** = 0.97 Å; average dihedral angle for glycosidic linkages from the MD simulation remain within the standard deviation of the averages from all the known PDB structures calculated using glytorsion at http://www.glycosciences.de/tools/glytorsion/ (Table S2)), which signified that the trajectories were equilibrated and appropriate for further analysis. Consistent with the crystal structures, each of the ligands formed stable hydrogen bonds between the O3/O4 hydroxyl groups of Gal and residues D89, N133, and A218 (Tables S3 and S4). In **5** and **6**, the Fuc-O2 group maintained its hydrogen bond with the side chain of N133. A hydrogen bond between the O3 group in the terminal reducing monosaccharide residue (Glc, Man, GlcNAc) in **1**-**6** was also observed, but found to be significantly more stable in the case of GlcNAc. Although a hydrogen bond is present between Gal-O3 and G107 in all the crystal structures, it was not highly occupied over

the course of the simulations. Similarly, the hydrogen bond between Fuc-O4 and Y108 in **5**, present in the crystal structure, only formed occasionally during the simulation.

Quantification of molecular contributions to affinity: The strengths of these interactions were quantified by performing MM-PB/GBSA energy analyses of the MD simulations. In addition to contributions from direct interactions (van der Waals and electrostatics), the MM-PB/GBSA energies also include estimates of desolvation free energy. Conformational entropies were estimated using three different methods. Firstly, we examined the quasi-harmonic (QH) approach, which derives entropy differences from a covariance analysis of the changes in atomic fluctuations that occur upon ligand binding (50). Secondly, an analysis of changes in the vibrational normal mode (NM) was performed (51), which estimates the entropic contributions for binding resulting from changes in the frequencies associated with bond stretching and angle bending. Lastly, the Karplus–Kushick QH approach [52] was employed to account for entropy differences arising from variations in the populations of conformational states of the glycosidic torsion angles in the bound versus free oligosaccharides (Table S5).

Five different Generalized Born (GBSA) desolvation free energy parametrizations (GB^{HCT}, igb=1 (53); $GB_1{}^{OBC}$, igb=2 (54); $GB_2{}^{OBC}$, igb=5 (54); GBn_1, igb=7 (55); GBn_2, igb=8 (56)), as well as a Poisson Boltzmann (PBSA) model using mbondi radii were employed to estimate binding affinities of all the six complexes. In agreement with the experimental data, and independent of the five GBSA desolvation models evaluated, **1** and **2** were always ranked the weakest binders and displayed essentially equivalent interaction energies, while **6** was correctly ranked as the highest affinity ligand. However, each of the models also ranked **5** amongst the

best binders, in disagreement with experiment. In contrast, the PBSA desolvation model, in agreement with the experiment, ranked **5** along with **1** and **2** amongst the weakest binders and correctly ranked **6** as the strongest binder. Overall none of the GBSA models could rank all the ligands accurately; on the other hand, the PBSA model could correctly rank every ligand (Figure 4.4A and Table 4.S6). As expected (57), incorporation of the entropic penalties always reduced the magnitude of the interaction energies. Surprisingly, the combination GBSA energies with QH entropy values did not lead to an improvement in the ranking of the relative affinities of the ligands (Figure 4.4B and Table 4.S7). In contrast, incorporation of NM entropies significantly improved the correlation between experimental and theoretical GBSA binding energies, and correctly ranked **3** and **6** amongst the best binders (Figure 4.4C and Table S9). The addition of ligand conformational entropic penalties further improved these correlations (Figure 4.4D and Tables S8 and S10) modestly. Remarkably, the addition of either QH or NM entropies to the PBSA binding energies significantly degraded their correlation with the experimental affinities.

The origin of the variations in absolute affinity arising from the desolvation model can be illustrated by an examination of contributions made by individual protein residues (Table 4.3). Each PB/GBSA model predicts similar (within approximately 2.2 kcal/mol) per-residue binding energies with **1** for interactions that do not involve hydrogen-bonds (A88, A222, F131, G217, P134, W135, Y106 and Y108). For hydrogen bond forming residues, this is not the case. For example, according to GB*n1* desolvation model N133 makes a negligible contribution to binding (-0.06 kcal/mol), despite the fact that this residue is involved in a stable hydrogen bond with the ligand. The most significant per-residue variation was seen in the predicted strength of the interaction with D89, which ranged from -7.7 to +8.5 kcal/mol. As a charged residue that makes

a stable hydrogen bond with the ligand, D89 would be expected to contribute significantly to binding, whereas the GB*n2* and PBSA desolvation methods both predicted its interaction to be unfavorable. Based on these observations, the GB*n1*, GB*n2*, and PBSA models were eliminated from further consideration, leaving GB^{HCT}, $GB_1{}^{OBC}$ and $GB_2{}^{OBC}$ for further analysis.

Quantification of per-residue contributions to affinity: Amino acids making significant interactions with the ligand were identified on the basis of their individual contributions to the total interaction energy, considering only the residues that contribute greater than 0.5 kcal/mol, which confirmed all of the expected interactions (Figure S5). In addition, stabilizing non-polar (van der Waals) interactions were observed between the Fuc residue and Y106, Y108, P134, and W135, which were confirmed by contact analyses of the crystal structure. Non-polar contacts were also observed in the presence of the GlcNAc residue, stabilizing its interaction with Q219. While the presence of the GalNAc residue introduced favorable van der Waals contacts with N133, it also introduced electrostatic repulsions, reducing the overall contribution of N133 to the binding. The significance of some of these residues (A88, Y106, F131, A218, D89, N133, and Q219, among others) has been confirmed experimentally by point mutations on a closely related protein called *Erythrina corallodendron* lectin (ECorL) (58).

On the basis of the current definition of glycan specificity, ECL is a Gal/GalNAc specific legume lectin (59). As expected, from the perspective of the ligand, the Gal/GalNAc residues were found to be the main contributors to binding, accounting for more than 65% of the interaction energy in all cases. According to $GB_1{}^{OBC}$ and $GB_2{}^{OBC}$ models, the Fuc residue in **5** and **6** contributed less than 6%, consistent with the observation that fucosylation impacts the

affinity only marginally (the GB^{HCT} model estimated the contribution from Fuc to be as high as 18.6%). The Glc and Man residues contributed less than 6.4%, while the presence of NAc group in the GlcNAc residue brings its contribution up to just over 8.6% in **3**, **4** and **6** (Figure 4.5). Despite their general utility, per-residue interaction energies include the contributions from all atoms in the interacting residues, and so do not provide direct measures of the strengths of specific interactions.

Quantification of per-functional group contributions to affinity: Using pairwise decomposition of the interaction energy, with per-atom and per-residue decomposition of the ligand and the protein respectively, the strength of all the hydrogen bonds was estimated and compared. Desolvation models $GB_1{}^{OBC}$ and $GB_2{}^{OBC}$ showed that Asp89 forms two favorable hydrogen bonds (contributing over -2.4 kcal/mol) with Gal/GalNAc (O3 and O4), whereas GB^{HCT} model was unable to capture the interaction accurately, by either underestimating its strength or by determining it to be unfavorable (between -1.2 to 0.2 kcal/mol), eliminating the GB^{HCT} model from further study. Only $GB_1{}^{OBC}$ and $GB_2{}^{OBC}$ presented comparable results for each of the individual interactions (Figure 4.6).

The assumption that the carbohydrate specificity of GBPs can be defined by the monosaccharide residues, fails to identify the underlying 3D structural features responsible for it. Not all exocyclic groups in a monosaccharide are equal participants in the interaction. Combining the per-atom decomposition values into contributions from individual functional groups (hydroxyl, NAc, etc) clearly revealed which of the functional groups in the ligand were most critical for binding. Six functional groups were created for the Gal/GalNAc residue (four

exocyclic hydroxyls, the NAc, and the ring structure including C6, which we refer to as the monosaccharide framework). The functional-group analysis showed that the main contribution to binding came from electrostatic interactions with the O3 and O4 hydroxyl groups (O3 over 25%, O4 over 20%) along with van der Waals contacts from the framework atoms of the Gal/GalNAc residue (over 18%) (Figure 4.7). The NAc moiety enhanced the interaction by contributing about 1 kcal/mol. It was also evident that some functional groups do not participate (such as the O6 and O2 hydroxyl groups of Gal/GalNAc residues). This approach provides an objective method to quantify features of the ligand that are critical/enhancing/unimportant for binding. Based on these observations it can be deduced that the conformation of the groups contributing most to the binding, defines the minimum 3D motif required for that protein-ligand interaction.

Carbohydrate 3D pharmacophore: The precise 3D spatial arrangement of functional groups in a ligand required for binding to a protein is often defined as a pharmacophore. As is evident from the present binding assays (Table 4.1), combined with the theoretical per-functional group contributions to binding (Figure 4.6), the 3D pattern that emerges as the pharmacophore required for binding to ECL is the spatial orientation of the O3 and O4 hydroxyl groups in the Gal/GalNAc residue along with the atoms forming the terminal ring structure (Figure S7). This implies that molecules that mimic the pharmacophore should be able to bind to ECL, provided that no unfavorable interactions are introduced. This observation is fully consistent with the present data, as well as with ECL inhibition data reported for a range of monosaccharides by Wu et.al. (30) (Figure 4.8). The pharmacophore analysis predicts that D-Gal, D-GalNAc, D-Fuc, L-Ara, L-Rha and D-Man should all be comparable inhibitors, whereas D-Glc, D-GlcNAc, D-Ara

and L-Fuc should be poor inhibitors due either to loss of hydrogen bond opportunities (D-Glc, D-GlcNAc, L-Fuc) or to unfavorable steric collisions (D-Ara, L-Fuc); as clashes were introduced or as the differences from the pharmacophore increased, inhibitory power decreased relative to D-Gal.

Discussion

Lectins typically display weak (mM) binding affinities and their specificities can appear complex, particularly when defined in terms of monosaccharide-based motifs (60). Here we introduce an alternative definition of binding motifs based on the observation that specificity arises from unique 3D arrangements of interacting groups in the ligand, which may be found in more than one monosaccharide. By understanding these spatial requirements, the specificity of a lectin may be defined by a subset of ligand atomic features. The ability to detect and computationally quantify these interactions and to use this information to define specificity has been illustrated here using the lectin ECL.

The affinity of ECL for a range of di- and trisaccharides has been quantified previously by ITC (28, 34) and here further using a competitive assay measured by BLI (61), which together provide reference values to evaluate the performance of MM-PB/GBSA calculations. Among the desolvation models used, in the absence of entropic corrections, PBSA performed the best at ranking the ligand affinities. However, ligand binding is also accompanied by significant configurational entropy penalties, and potentially also conformational entropy penalties arising from constraining the motion of the highly plastic oligosaccharides, which makes it essential to consider changes in entropy in the estimated binding affinities. The combination of QH entropies with MM-PB/GBSA data improved the correlation of the theoretical binding energies with the experimental data (from approximately 0.7 to 0.8 for GB_1^{OBC} or GB_2^{OBC}) but worsened the correlation when combined with PBSA data (dropping it from 0.95 to 0.71) (Table S7). A similar trend was observed with the combination of NM entropies with MM-GBSA data, which led to greater improvement in the correlation (from approximately 0.7 to 0.9 for GB_1^{OBC} or GB_2^{OBC}),

whereas with PBSA the correlation worsened from 0.95 to 0.48 (Table S9). While the QH entropies had not converged within 100 ns, it was possible to estimate their values at infinite sampling time by extrapolation (43). In contrast, the extreme computational expense of NM calculations reduced the number of snapshots that could practically be analyzed to 100 from each simulation. Thus, the NM values are less precise and probably less accurate than the QH values. Neither QH nor NM entropy estimates account for the decrease in the number of conformational states in the ligand that can be significant for oligosaccharides. Ligand conformational entropy penalties were therefore computed using the Karplus-Kushick (52) method based on the differences in glycosidic torsional states observed during MD simulations of the free and bound ligands. The addition of the relatively modest (< 1.5 kcal/mol) ligand conformational entropies on top of relatively large NM or QH entropies led to very slight improvements in the correlations with MM-GBSA models, but did not improve the poor correlations of the MM-PBSA data (Table S8 and S10).

The results from binding free energy analyses employing different desolvation models, along with entropy calculations indicated that improvements need to be made in the current desolvation models. It would likely be beneficial to re-calibrate the current GB/PB methods by including carbohydrate-protein interactions (62). Although no combination of solvation or entropy models led to an adequate agreement with the absolute experimental binding energies, much insight into the contributions from individual residues could be gained from a per-residue energy decomposition analysis. Most of GB/PB models identified the same set of key protein and ligand residues. However, a large variation in the contribution was seen for the negatively charged residue D89, which varied from -7.7 to +8.4 kcal/mol (Table 4.3). PBSA and GB*n2*

desolvation models predicted the contribution from D89 to be highly unfavorable, despite the fact that crystallographic data shows that it forms hydrogen bonds with the Gal residue of the ligand, and earlier mutational studies have shown D89 to be essential for binding. While GB*n1* model could successfully identify D89 as one of the residues favorable for binding, it appears to underestimate the contribution from another essential residue (N133) (58).

As is common practice, all the calculations were performed with a dielectric constant of unity ($\epsilon = 1$). It has however been suggested that using a higher dielectric value may be appropriate for systems where charge polarization is likely to be important (63). Given the extreme sensitivity of the contribution from D89 to the PB/GBSA model, a value of $\epsilon = 4$ was also examined. As expected, increasing ϵ proportionally decreased the interaction energies between polar groups (Table S11), while leaving non-polar interactions largely unaffected. However, the larger dielectric value did not correct the poor performance of the PBSA or GB*n2* models with D89. While several desolvation models showed good correlations with the experimental affinity data, when the per-residue interaction energies were examined, only the GB_1^{OBC} and GB_2^{OBC} models were consistently in agreement with expectations based on affinity data from point mutagenesis, and with the observed interactions in co-crystal structures.

By further decomposing the binding free energies on the per-group basis, it was possible to quantify the strengths of key interactions, such as hydrogen bonds (Tables S3 and S4). Such an analysis can be particularly useful in predicting or rationalizing the effects of protein mutations on ligand affinity. Conversely, this information can be crucial from the perspective of inhibitor design. A per-group energy analysis permits the identification of functional groups in each ligand that are responsible for the specificity of the interactions. The lack of participation of

the Glc-O2 group explains why epimerizing **1** at C2 (i.e. converting it to a mannose in **2**) resulted in equivalent binding affinities. Similarly, the O2 group of the Gal residue does not make a significant contribution to binding, thus its modification should also be tolerated, provided no new steric clashes arise. The ability to modify the ligand at the Gal-O2 position was confirmed by the binding of **4, 5** and **6**. Conversely, modification of groups with a high contribution (O3 and O4 groups of Gal residue) should significantly affect the binding. For example, replacing Gal residue with its O4 epimer i.e. Glc, resulting in cellobiose (Glcβ1-4Glcβ, **7**) should hamper its interaction with ECL, as demonstrated experimentally (Figure S1, S2, and S3).

Based on the range of strengths of their interactions, the functional groups could be characterized as critical, enhancing, or non-interacting. Critical groups are essential for achieving measurable affinity and define the pharmacophore. Enhancing groups improve the strength of the interaction relative to that of the pharmacophore, but are not required for binding, while non-interacting groups can be altered with no effect on binding, if doing so does not introduce unfavorable steric or electrostatic repulsions. The ability to rank the functional groups in terms of their importance to binding can be used to design novel ligands and can aid in explaining the specificity and affinity of different ligands for a protein.

By defining the glycan pharmacophore structurally, and separating it from residue-based nomenclature, it is possible to represent the pharmacophore in a number of alternative chemoinformatic formats. One such format is known as the Simplified Molecular Input Line Entry System (SMILES) (64), another is the IUPAC International Chemical Identifier (InChI) (65). SMILES and InChI strings are readily transferable between many software packages, facilitate the detection of similar features, and convertible back to 3D structures.

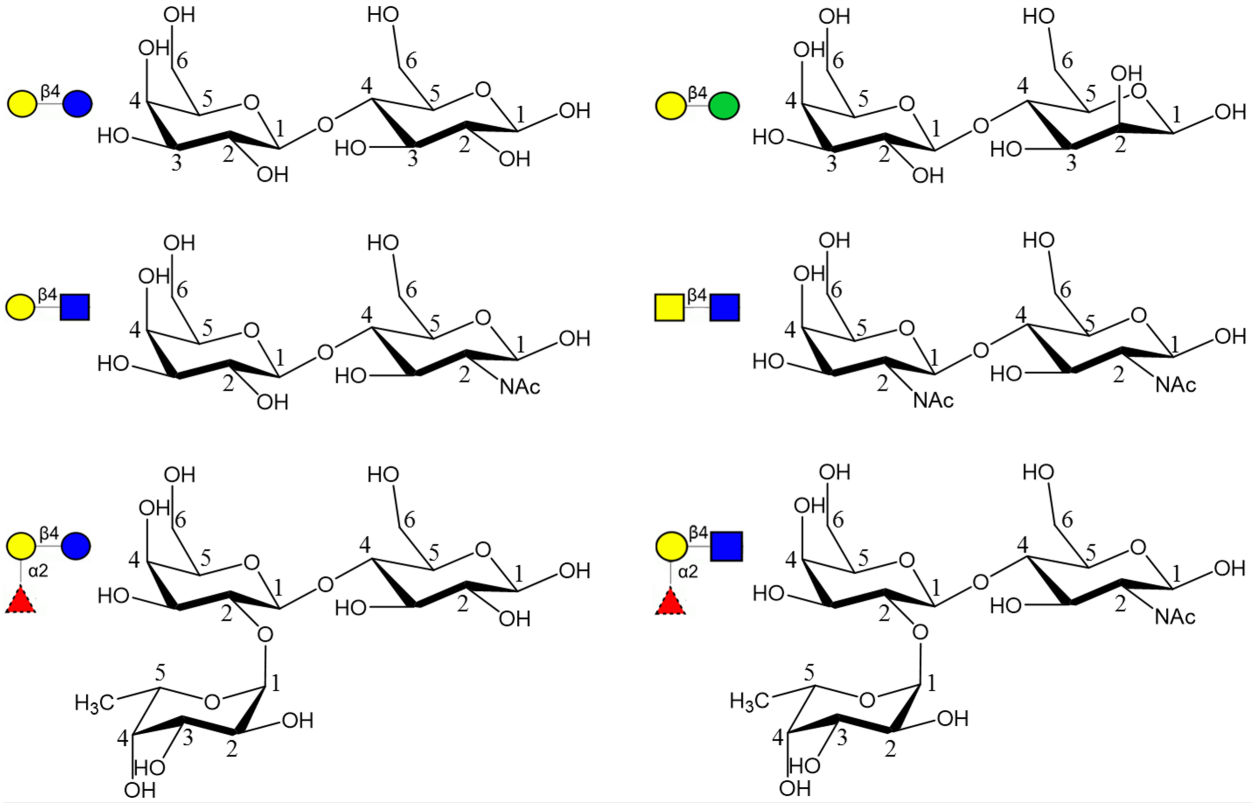


Figure 4.1. Six different ligands i.e. Lac (**1**, top left), Epilac (**2**, top right), LacNAc (**3**, middle left), LacDiNAc (**4**, middle right), Fuclac (**5**, bottom left) and FuclacNAc (**6**, bottom right) that interact with ECL. The monosaccharides are represented in SNFG notation [66] as Gal: yellow circle, Glc: blue circle, Man: green circle, GlcNAc: blue square, GalNAc: yellow square and Fuc: red triangle.

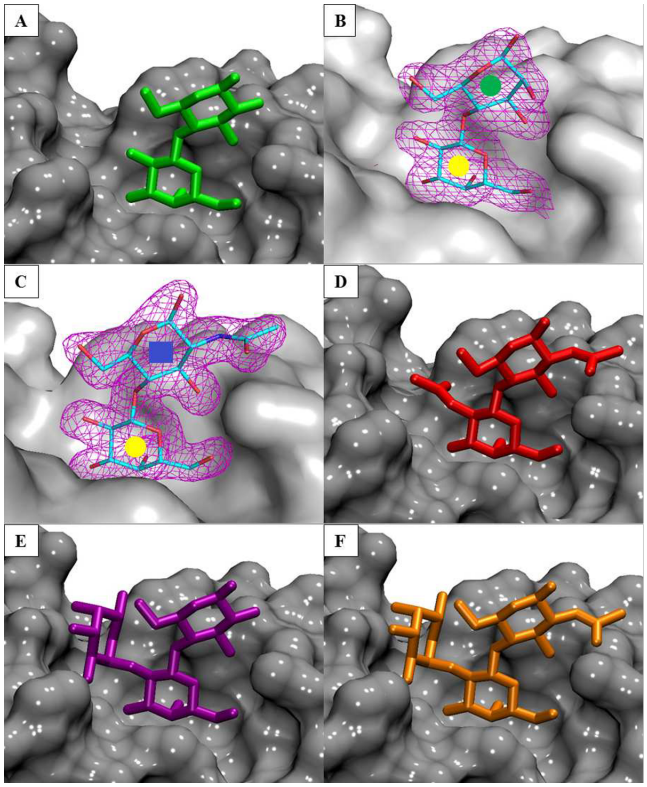


Figure 4.2. Co-crystal structures of ECL in complex with ligands **1**, **2**, **3**, and **5** presented in panels **A**, **B**, **C**, and **E**. Modelled structure of ECL in complex with ligands **4** and **6** presented in panels **D** and **F**. The protein is shown as a gray surface and the ligands are shown as sticks. Representative $2F_o$-F_c electron density maps (purple mesh at the 1.3σ level) are depicted ligands **2** (**B**) and **3** (**C**) colored by atom type, carbon is cyan, nitrogen is blue, oxygen is red.

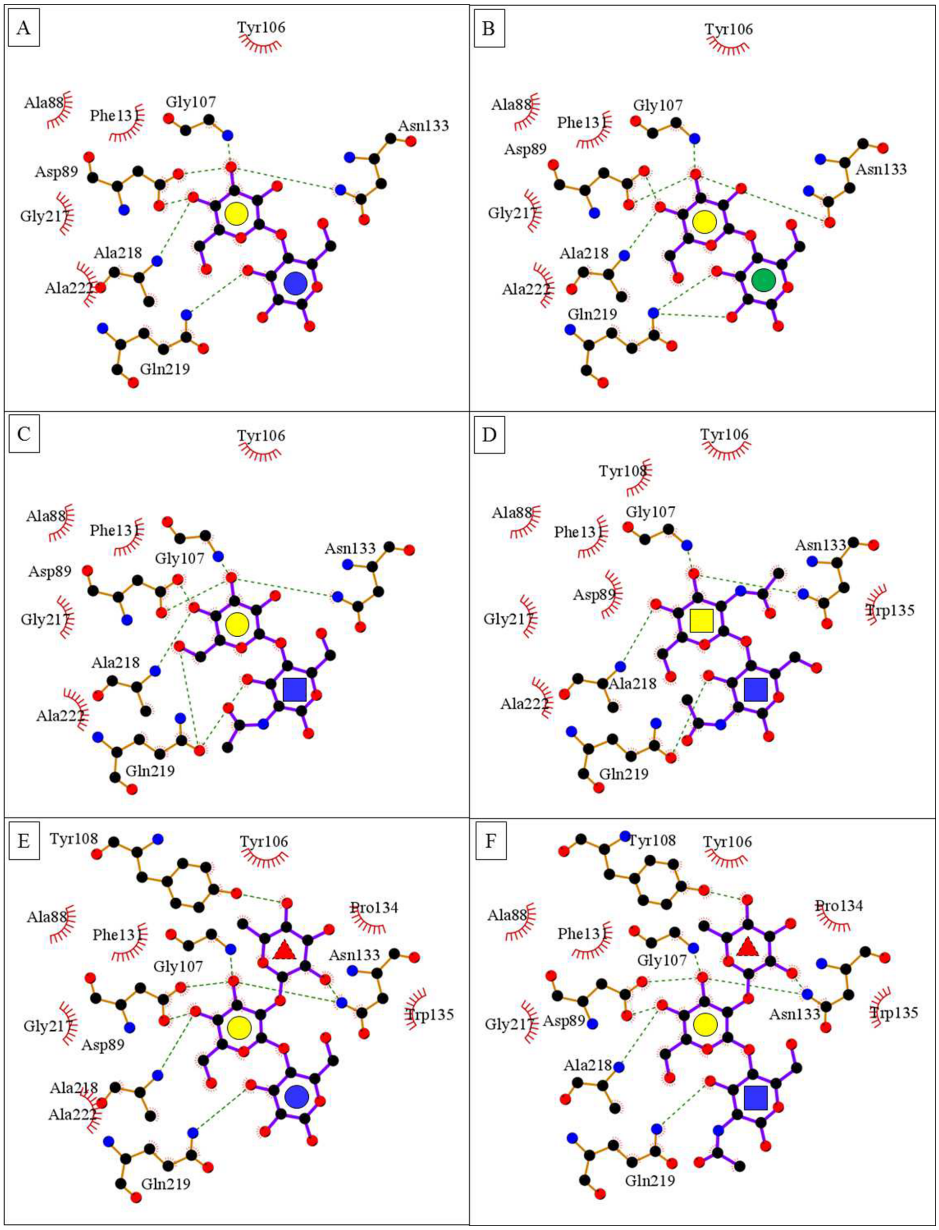


Figure 4.3. LigPlot (67) contacts between the amino acids in the binding pocket of ECL and ligands **1** to **6** presented from A to F. The red brackets show hydrophobic contacts, and green dotted lines show hydrogen bonds. The monosaccharides are represented as Gal: yellow circle, Glc: blue circle, Man: green circle, GlcNAc: blue square, GalNAc: yellow square and Fuc: red triangle.

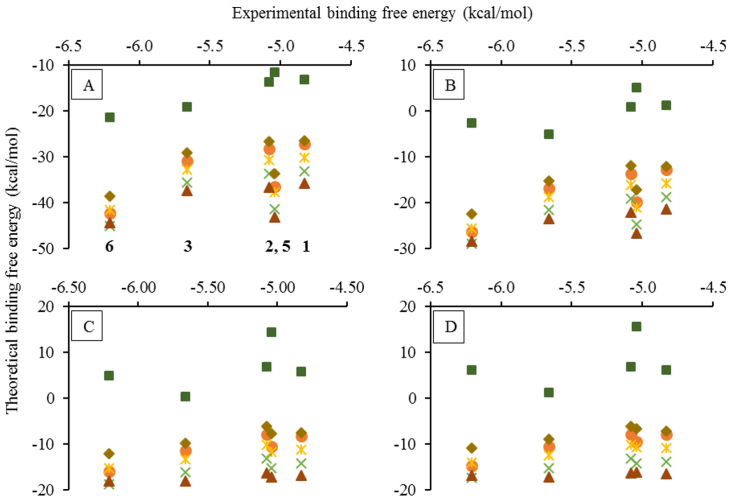


Figure 4.4. Comparison of theoretical (PBSA (green square), GB^{HCT} (orange circle), GB_1^{OBC} (yellow star), GB_2^{OBC} (green cross), GB*n1* (red triangle) and GB*n2* (brown rhombus)) and experimental binding free energies (Table 4.1) for five ligands (**1**, **2**, **3**, **5** and **6**). **A**. Binding free energies from MM-GB/PBSA calculation. **B.** Binding free energies from MM-GB/PBSA calculation employing quasi-harmonic entropies. **C.** Binding free energies from MM-GB/PBSA calculation employing normal mode entropies. **D**. Binding free energies from MM-GB/PBSA calculation employing normal mode and conformational entropies.

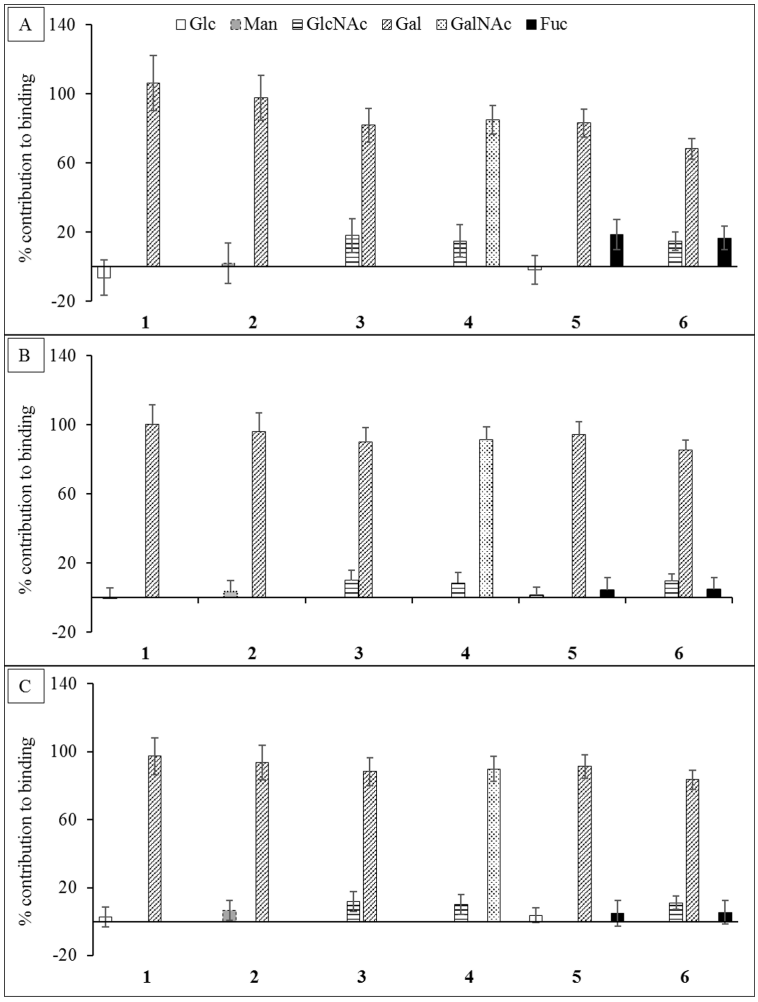


Figure 4.5. The percentage contribution to the total ΔG made by each monosaccharide in each ligand. The calculations were performed using three different desolvation models: **A**. GB^{HCT}, **B**. $GB_1{}^{OBC}$ and **C**. $GB_2{}^{OBC}$. In each ligand the Gal or GalNAc residue contributes the most to the total affinity.

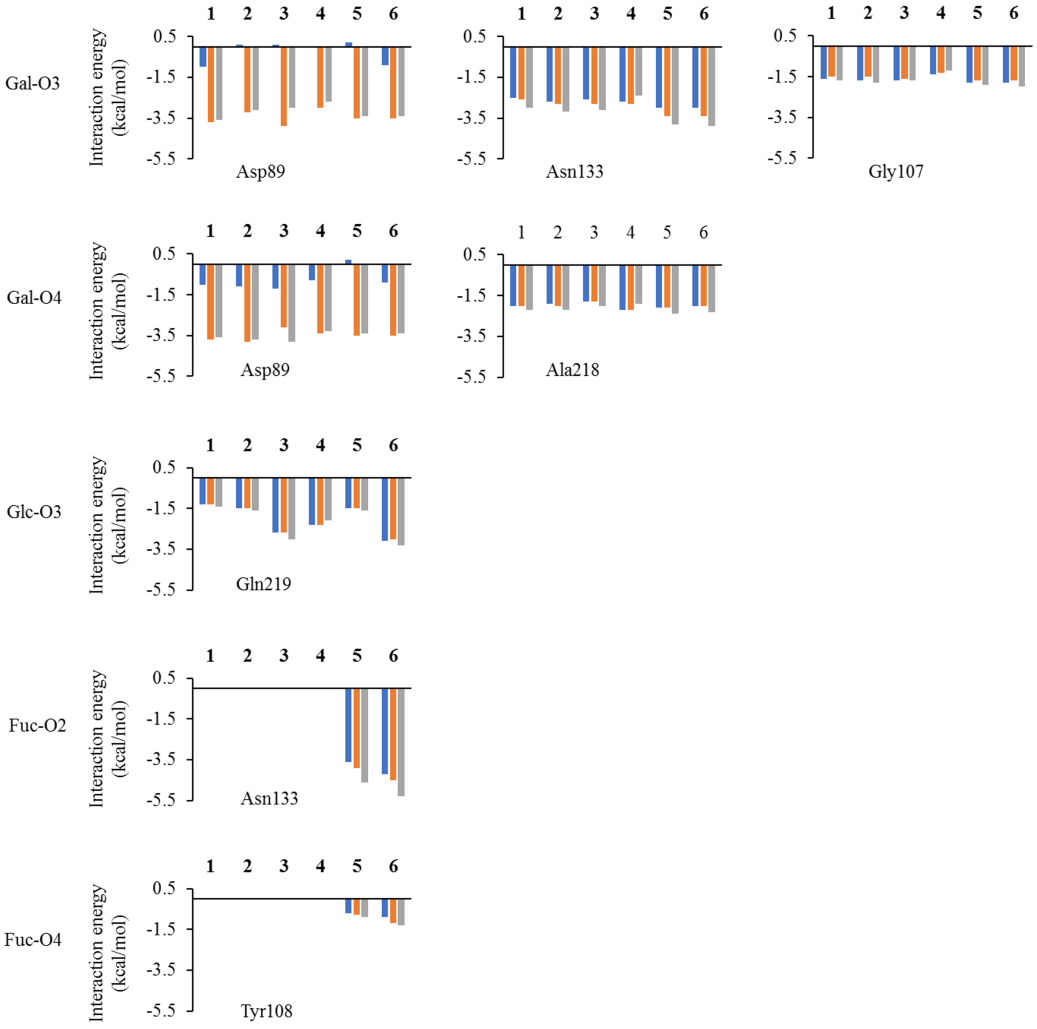


Figure 4.6. Interaction energies of per-hydroxyl group of the sugar interacting with different protein residues from the MD simulation of all the six ECL-ligand complexes. For ligand **4**, Gal-O3/O4 represents GalNAc-O3/O4. For ligand **2**, Glc-O3 represents Man-O3. For ligands **3** and **4**, Glc-O3 represents GlcNAc-O3. The blue bars indicate interaction energies calculated using GB^{HCT} (igb=1) desolvation parameters, while orange bars indicate calculations performed using GB_1^{OBC} (igb=2) and values represented by grey bars were calculated using GB_2^{OBC} (igb=5) parameters.

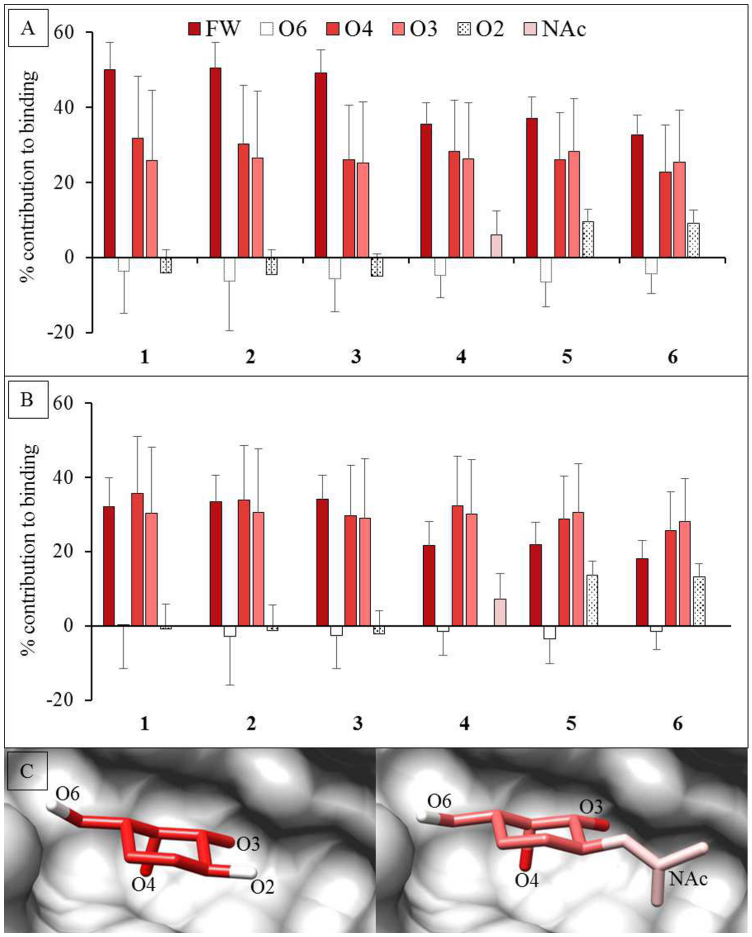


Figure 4. 7. The percentage contribution to the total ΔG made by specific functional groups in the Gal or GalNAc residues. The calculations were performed using two different desolvation models. **A**. GB_1 ^{OBC} and **B**. GB_2 ^{OBC}. The three most important contributors to binding are the ring framework (FW) atoms, and hydroxyl groups O3 and O4. **C.** Image of the D-Gal residue in Lac (**1**) (left) and D-GlcNAc in LacDiNAc (**4**) (right) interacting with ECL (grey surface). The per-functional group contribution to binding, where red to white indicates higher to lower contribution (using GB_2^{OBC} (igb=5) parameters).

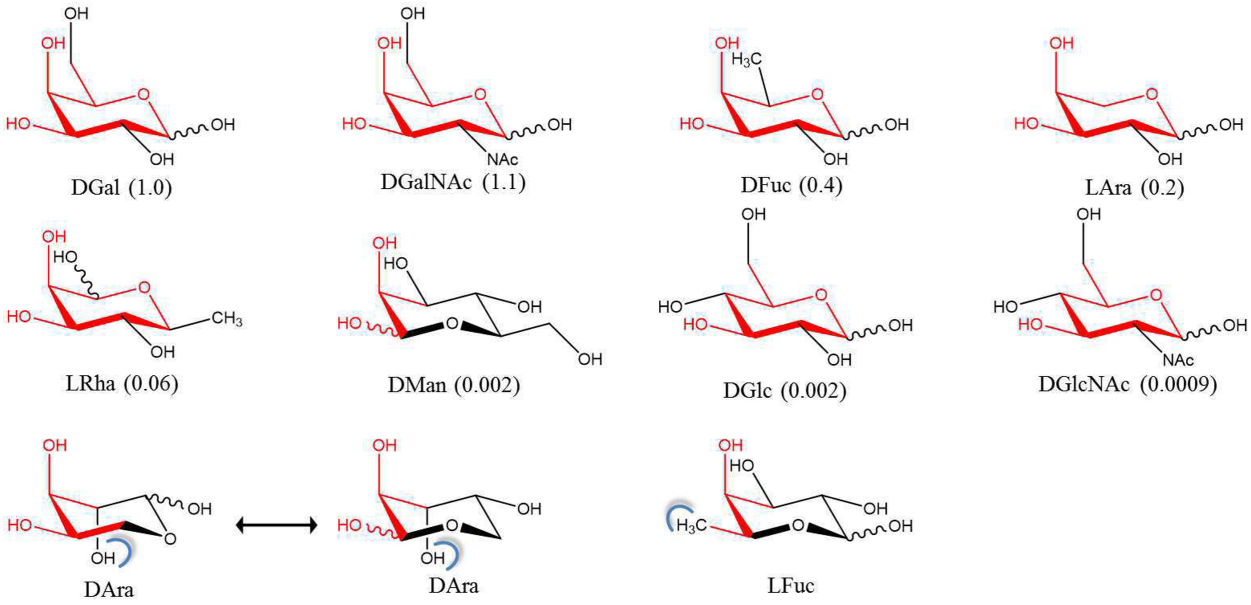


Figure 4.8. Relative ability of monosaccharides to inhibit the binding of ECL to the human asialo α1-acid glycoprotein, relative inhibitory potentials derived from IC_{50} values are shown in parentheses [30]. Pharmacophore positions are indicated in red and steric clashes (determined by Chimera [68]) are shown as blue brackets, based on alignment of the monosaccharides onto the Gal residue in the ECL co-crystal with lactose. The rings are oriented so as to most clearly show the similarity to D-Gal.

Table 4. 1. Binding parameters determined by BLI compared to reported values.

Ligand	K_D (mM)	ΔG (kcal/mol)	Reference ΔG [28, 45a]	Reference ΔG [28, 45b]
1	0.32 (0.02)[c]	-4.83 (0.04)	-4.9 (0.2)	-4.8 (<0.1)
2	0.21 (0.01)	-5.08 (0.02)		
3	0.08 (0.01)	-5.66 (0.04)	-5.5 (0.1)	
5	0.22 (0.01)	-5.04 (0.06)		-4.8 (<0.1)
6	0.032 (0.01)	-6.21 (0.14)		
D-Gal				-4.0 (<0.1)

[a]Experiments performed by isothermal titration calorimetry (ITC) at 27°C.

[b]Experiments performed by ITC at 25°C.

[c]Standard deviations shown in parentheses.

Table 4. 2. X-ray crystallographic data-collection and refinement statistics[a].

	ECL-**2**	ECL-**3**
Beamline/Facility	Rigaku HighFlux HomeLab/ORNL	SBC-CAT 19ID/APS
Space group	$P6_5$	$P6_5$
Cell dimensions:		
a, b, c (Å)	134.95, 134.95, 81.79	134.67, 134.67, 81.21
α, β, γ (°)	90, 90, 120	90, 90, 120
Resolution (Å)	40.00-2.20 (2.28-2.20)	44.08-1.90 (1.93-1.90)
R_{merge} (%)[b]	8.50 (49.60)	6.80 (46.10)
$I / \sigma I$	13.1 (2.1)	38.4 (4.4)
No. reflections measured	42764 (4263)	65144 (3227)
Completeness (%)	98.8 (98.8)	99.2 (98.1)
Redundancy	3.3 (3.1)	6.7 (5.8)
*R*work / *R*free (%)	18.14 / 20.42	22.17 / 26.36
No. atoms (non-H)	4142	4274
Water Molecules	296	394
R.m.s.d. bonds (Å)	0.003	0.007
R.m.s.d. bond angles (°)	0.684	1.188
PDB ID	6AQ5	6AQ6

[a]Data in parentheses is for the highest resolution shell.

[b]R_{merge} = å|I-<I>|/å<I>.

Table 4. 3. The impact of desolvation free energy on per-residue MM-PB/GBSA values[a].

	GB^{HCT} (igb = 1)	$GB_1{}^{OBC}$ (igb = 2)	$GB_2{}^{OBC}$ (igb = 5)	GBn_1 (igb = 7)	GBn_2 (igb = 8)	PBSA
Residues forming hydrogen bonds with the ligand						
A218	-3.30	-2.63	-2.45	-1.34	-2.97	-2.98
D89	-1.54	-5.25	-6.77	-7.72	**3.93**[b]	**8.45**
G107	-1.30	-0.75	-0.71	-0.28	-1.32	-1.73
N133	-1.81	-0.95	-0.91	**-0.06**	-0.55	-2.15
Q219	-2.92	-2.23	-2.25	-1.57	-2.26	-2.04
Residues involved in other interactions with the ligand						
A88	-0.88	-0.64	-0.54	-0.49	-1.13	-1.41
A222	-0.42	-0.45	-0.46	-0.66	-0.41	-0.32
F131	-2.07	-2.36	-2.53	-2.52	-2.22	-0.54
G217	-1.32	-0.32	-0.17	0.67	-0.49	-1.49
P134	-0.15	-0.19	-0.21	-0.13	-0.17	-0.14
W135	-0.03	-0.13	-0.24	-0.44	-0.09	-0.37
Y106	-2.51	-1.70	-1.68	-1.78	-2.22	-2.86
Y108	0.02	-0.08	-0.12	-0.19	-0.13	-0.17

[a]Energies in kcal/mol.

[b]Numbers in bold represent residues with structurally inconsistent values.

References

1. Wang, B.; Boons, G.-J., *Carbohydrate Recognition: Biological Problems, Methods, and Applications*. Wiley: 2011; p 448 pages.
2. Lasky, L. A., Selectins: interpreters of cell-specific carbohydrate information during inflammation. *Science* **1992**, 258, 964-969.
3. Perillo, N. L.; Pace, K. E.; Seilhamer, J. J.; Baum, L. G., Apoptosis of T cells mediated by galectin-1. *Nature* **1995**, 378, 736-739.
4. Haltiwanger, R. S.; Lowe, J. B., Role of Glycosylation in Development. *Annu. Rev. Biochem.* **2004**, 73, 491-537.
5. Brown, G. D.; Gordon, S., Immune recognition: A new receptor for [beta]-glucans. *Nature* **2001**, 413, 36-37.
6. Cobb, B. A.; Kasper, D. L., Coming of age: carbohydrates and immunity. *Eur. J. Immunol.* **2005**, 35, 352-356.
7. Nakahara, S.; Raz, A., Biological Modulation by Lectins and Their Ligands in Tumor Progression and Metastasis. *Anticancer Agents Med. Chem.* **2008**, 8, 22-36.
8. Elola, M.; Wolfenstein-Todel, C.; Troncoso, M.; Vasta, G.; Rabinovich, G., Galectins: matricellular glycan-binding proteins linking cell adhesion, migration, and survival. *Cell. Mol. Life Sci.* **2007**, 64, 1679-1700.
9. Monsigny, M.; Mayer, R.; Roche, A.-C., Sugar-lectin interactions: sugar clusters, lectin multivalency and avidity. *Carbohydrate letters* **2000**, 4, 35-52.

10. Grant, O. C.; Smith, H. M.; Firsova, D.; Fadda, E.; Woods, R. J., Presentation, presentation, presentation! Molecular-level insight into linker effects on glycan array screening data. *Glycobiology* **2014**, 24, 17-25.

11. Manimala, J. C.; Roach, T. A.; Li, Z.; Gildersleeve, J. C., High-Throughput Carbohydrate Microarray Analysis of 24 Lectins. *Angew. Chem. Int. Ed.* **2006**, 45, 3607-3610.

12. Manimala, J. C.; Roach, T. A.; Li, Z.; Gildersleeve, J. C., High-throughput carbohydrate microarray profiling of 27 antibodies demonstrates widespread specificity problems. *Glycobiology* **2007**, 17, 17C-23C.

13. Cummings, R. D. Use of lectins in analysis of glycoconjugates. In *Methods Enzymol.*; 1994; Vol. 230, pp 66-86.

14. Ching, C. K.; Black, R.; Helliwell, T.; Savage, A.; Barr, H.; Rhodes, J. M., Use of lectin histochemistry in pancreatic cancer. *J. Clin. Pathol.* **1988**, 41, 324-328.

15. Ambepitiya Wickramasinghe, I. N.; de Vries, R. P.; Weerts, E. A. W. S.; van Beurden, S. J.; Peng, W.; McBride, R.; Ducatez, M.; Guy, J.; Brown, P.; Eterradossi, N.; Gröne, A.; Paulson, J. C.; Verheije, M. H., Novel Receptor Specificity of Avian Gammacoronaviruses That Cause Enteritis. *J. Virol.* **2015**, 89, 8783-8792.

16. Ambrosi, M.; Cameron, N. R.; Davis, B. G., Lectins: tools for the molecular understanding of the glycocode. *Org. Biomol. Chem.* **2005**, 3, 1593-1608.

17. Tessier, M. B.; Grant, O. C.; Heimburg-Molinaro, J.; Smith, D.; Jadey, S.; Gulick, A. M.; Glushka, J.; Deutscher, S. L.; Rittenhouse-Olson, K.; Woods, R. J., Computational Screening of the Human TF-Glycome Provides a Structural Definition for the Specificity of Anti-Tumor Antibody JAA-F11. *PLoS One* **2013**, 8, e54874.

18. Sauter, N. K.; Bednarski, M. D.; Wurzburg, B. A.; Hanson, J. E.; Whitesides, G. M.; Skehel, J. J.; Wiley, D. C., Hemagglutinins from two influenza virus variants bind to sialic acid derivatives with millimolar dissociation constants: a 500-MHz proton nuclear magnetic resonance study. *Biochemistry* **1989**, 28, 8388-8396.

19. Xiong, X.; Coombs, P. J.; Martin, S. R.; Liu, J.; Xiao, H.; McCauley, J. W.; Locher, K.; Walker, P. A.; Collins, P. J.; Kawaoka, Y., Receptor binding by a ferret-transmissible H5 avian influenza virus. *Nature* **2013**, 497, 392-396.

20. Yang, H.; Carney, P. J.; Chang, J. C.; Villanueva, J. M.; Stevens, J., Structural analysis of the hemagglutinin from the recent 2013 H7N9 influenza virus. *J. Virol.* **2013**, 87, 12433-12446.

21. Xu, R.; de Vries, R. P.; Zhu, X.; Nycholat, C. M.; McBride, R.; Yu, W.; Paulson, J. C.; Wilson, I. A., Preferential recognition of avian-like receptors in human influenza A H7N9 viruses. *Science* **2013**, 342, 1230-1235.

22. Cao, Z.; Partyka, K.; McDonald, M.; Brouhard, E.; Hincapie, M.; Brand, R. E.; Hancock, W. S.; Haab, B. B., Modulation of glycan detection on specific glycoproteins by lectin multimerization. *Anal. Chem.* **2013**, 85, 1689-1698.

23. Vyas, N. K., Atomic features of protein-carbohydrate interactions. *Curr. Opin. Struct. Biol.* **1991**, 1, 732-740.

24. Lee, Y. C.; Lee, R. T., Carbohydrate-protein interactions: basis of glycobiology. *Acc. Chem. Res.* **1995**, 28, 321-327.

25. Xu, R.; McBride, R.; Nycholat, C. M.; Paulson, J. C.; Wilson, I. A., Structural characterization of the hemagglutinin receptor specificity from the 2009 H1N1 influenza pandemic. *J. Virol.* **2012**, 86, 982-990.

26. Hadden, J. A.; Tessier, M. B.; Fadda, E.; Woods, R. J. Calculating Binding Free Energies for Protein–Carbohydrate Complexes. In *Glycoinformatics*; 2015, pp 431-465.

27. Sun, H.; Li, Y.; Tian, S.; Xu, L.; Hou, T., Assessing the performance of MM/PBSA and MM/GBSA methods. 4. Accuracies of MM/PBSA and MM/GBSA methodologies evaluated by various simulation protocols using PDBbind data set. *Phys. Chem. Chem. Phys.* **2014**, 16, 16719-16729.

28. Svensson, C.; Teneberg, S.; Nilsson, C. L.; Kjellberg, A.; Schwarz, F. P.; Sharon, N.; Krengel, U., High-resolution Crystal Structures of Erythrina cristagalli Lectin in Complex with Lactose and 2′-α-l-Fucosyllactose and Correlation with Thermodynamic Binding Data. *J. Mol. Biol.* **2002**, 321, 69-83.

29. Turton, K.; Natesh, R.; Thiyagarajan, N.; Chaddock, J. A.; Acharya, K. R., Crystal structures of Erythrina cristagalli lectin with bound N-linked oligosaccharide and lactose. *Glycobiology* **2004**, 14, 923-929.

30. Wu, A. M.; Wu, J. H.; Tsai, M.-S.; Yang, Z.; Sharon, N.; Herp, A., Differential affinities of Erythrina cristagalli lectin (ECL) toward monosaccharides and polyvalent mammalian structural units. *Glycoconj. J.* **2007**, 24, 591-604.

31. Imamura, K.; Takeuchi, H.; Yabe, R.; Tateno, H.; Hirabayashi, J., Engineering of the glycan-binding specificity of Agrocybe cylindracea galectin towards α (2, 3)-linked sialic acid by saturation mutagenesis. *J. Biochem.* **2011**, 150, 545-552.

32. Ernst, B.; Magnani, J. L., From Carbohydrate Leads to Glycomimetic Drugs. *Nat. Rev. Drug Disc.* **2009**, 8, 661-677.

33. Magnani, J. L.; Ernst, B., Glycomimetic drugs--a new source of therapeutic opportunities. *Discov. Med.* **2009**, 8, 247-52.

34. Minor, W.; Cymborowski, M.; Otwinowski, Z.; Chruszcz, M., HKL-3000: the integration of data reduction and structure solution–from diffraction images to an initial model in minutes. *Acta Crystallogr. D Biol. Crystallogr.* **2006**, 62, 859-866.

35. Collaborative Computational Project, N., The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **1994**, 50, 760-763.

36. Adams, P. D.; Afonine, P. V.; Bunkóczi, G.; Chen, V. B.; Davis, I. W.; Echols, N.; Headd, J. J.; Hung, L.-W.; Kapral, G. J.; Grosse-Kunstleve, R. W., PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, 66, 213-221.

37. Davis, I. W.; Leaver-Fay, A.; Chen, V. B.; Block, J. N.; Kapral, G. J.; Wang, X.; Murray, L. W.; Arendall, W. B.; Snoeyink, J.; Richardson, J. S., MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* **2007**, 35, W375-W383.

38. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, 66, 486-501.

39. Cheng, Y.-C.; Prusoff, W. H., Relationship between the inhibition constant (KI) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. *Biochem. Pharmacol.* **1973**, 22, 3099-3108.

40. Case D.A., B. V., Berryman J.T., Betz R.M., Cai Q., Cerutti D.S., Cheatham T.E., III, Darden T.A., Duke R.E., Gohlke H., Goetz A.W., Gusarov S., Homeyer N., Janowski P., Kaus J., Kolossváry I., Kovalenko A., Lee T.S., LeGrand S., Luchko T., Luo R., Madej B., Merz K.M., Paesani F., Roe D.R., Roitberg A., Sagui C., Salomon-Ferrer R., Seabra G., Simmerling C.L., Smith W., Swails J., Walker R.C., Wang J., Wolf R.M., Wu X., Kollman P.A., AMBER14. *University of California; San Francisco: 2014* **2014**.

41. Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E., MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* **2012**, 8, 3314-3321.

42. Roe, D. R.; Cheatham, T. E., PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, 9, 3084-3095.

43. Schlitter, J., Estimation of absolute and relative entropies of macromolecules using the covariance matrix. *Chem. Phys. Lett.* **1993**, 215, 617-621.

44. Xue, W.; Yang, Y.; Wang, X.; Liu, H.; Yao, X., Computational study on the inhibitor binding mode and allosteric regulation mechanism in hepatitis C virus NS3/4A protein. *PLoS One* **2014**, 9, e87077.

45. Gupta, D.; Cho, M.; Cummings, R. D.; Brewer, C. F., Thermodynamics of Carbohydrate Binding to Galectin-1 from Chinese Hamster Ovary Cells and Two Mutants. A Comparison with Four Galactose-Specific Plant Lectins. *Biochemistry* **1996**, 35, 15236-15243.

46. Grant, O. C.; Tessier, M. B.; Meche, L.; Mahal, L. K.; Foley, B. L.; Woods, R. J., Combining 3D structure with glycan array data provides insight into the origin of glycan specificity. *Glycobiology* **2016**, 26, 772-783.

47. Sharon, N.; Lis, H., How Proteins Bind Carbohydrates: Lessons from Legume Lectins. *J. Agric. Food Chem.* **2002**, 50, 6586-6591.

48. Case, D.; Darden, T.; Cheatham III, T.; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Walker, R.; Zhang, W.; Merz, K., AMBER12. *University of California: San Francisco* **2012**.

49. Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González‐Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J., GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem.* **2008**, 29, 622-655.

50. Levy, R. M.; Karplus, M.; Kushick, J.; Perahia, D., Evaluation of the configurational entropy for proteins: application to molecular dynamics simulations of an α-helix. *Macromolecules* **1984**, 17, 1370-1374.

51. Case, D. A., Normal mode analysis of protein dynamics. *Curr. Opin. Struct. Biol.* **1994**, 4, 285-290.

52. Karplus, M.; Kushick, J. N., Method for estimating the configurational entropy of macromolecules. *Macromolecules* **1981**, 14, 325-332.

53. Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G., Pairwise solute descreening of solute charges from a dielectric medium. *Chem. Phys. Lett.* **1995**, 246, 122-129.

54. Onufriev, A.; Bashford, D.; Case, D. A., Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Structure, Function, and Bioinformatics* **2004**, 55, 383-394.

55. Mongan, J.; Simmerling, C.; McCammon, J. A.; Case, D. A.; Onufriev, A., Generalized Born model with a simple, robust molecular volume correction. *J. Chem. Theory Comput.* **2007**, 3, 156-169.

56. Nguyen, H.; Roe, D. R.; Simmerling, C., Improved Generalized Born Solvent Model Parameters for Protein Simulations. *J. Chem. Theory Comput.* **2013**, 9, 2020-2034.

57. Gohlke, H.; Case, D. A., Converging free energy estimates: MM‐PB (GB) SA studies on the protein–protein complex Ras–Raf. *J. Comput. Chem.* **2004**, 25, 238-250.

58. Adar, R.; Sharon, N., Mutational Studies of the Amino Acid Residues in the Combining Site of Erythrina corallodendron Lectin. *Eur. J. Biochem.* **1996**, 239, 668-674.

59. Iglesias, J. L.; Lis, H.; Sharon, N., Purification and Properties of a d-Galactose/N-Acetyl-d- galactosamine-Specific Lectin from Erythrina cristagalli. *Eur. J. Biochem.* **1982**, 123, 247-252.

60. Kletter, D.; Cao, Z.; Bern, M.; Haab, B., Determining lectin specificity from glycan array data using motif segregation and GlycoSearch software. *Curr. Protoc. Chem. Biol.* **2013**, 5, 157-169.

61. Ji, Y.; Woods, R. J. Quantifying weak glycan-protein interactions using a Biolayer lnterferometry competition assay: Applications to lectin and influenza hemagglutinin. In *Glycobiophysics*, Kato, K.; Yamaguchi, Y., Eds.; Springer: in press.

62. Eid, S.; Saleh, N.; Zalewski, A.; Vedani, A., Exploring the free-energy landscape of carbohydrate–protein complexes: development and validation of scoring functions considering the binding-site topology. *J. Comput. Aided Mol. Des.* **2014**, 28, 1191-1204.

63. Hou, T.; Wang, J.; Li, Y.; Wang, W., Assessing the performance of the MM/PBSA and MM/GBSA methods: I. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J. Chem. Inf. Model.* **2011**, 51, 69-82.

64. Anderson, E.; Veith, G. D.; Weininger, D., *SMILES, a line notation and computerized interpreter for chemical structures*. US Environmental Protection Agency, Environmental Research Laboratory: 1987.

65. Heller, S.; McNaught, A.; Stein, S.; Tchekhovskoi, D.; Pletnev, I., InChI-the worldwide chemical structure identifier standard. *J. Cheminform.* **2013**, 5, 1.

66. Varki, A.; Cummings, R. D.; Aebi, M.; Packer, N. H.; Seeberger, P. H.; Esko, J. D.; Stanley, P.; Hart, G.; Darvill, A.; Kinoshita, T., Symbol nomenclature for graphical representations of glycans. *Glycobiology* **2015**, 25, 1323-1324.

67. Wallace, A. C.; Laskowski, R. A.; Thornton, J. M., LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Engineering, Design and Selection* **1995**, 8, 127-134.

68. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., UCSF Chimera—A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, 25, 1605-1612.

CHAPTER 5

Quantifying monomeric binding affinities (K_D, K_i) of influenza hemagglutinin using BioLayer Interferometry

[1]Ye Ji, Oliver Grant, Jen Hendel, and Robert Woods.

To be submitted to *Journal of Biological Chemistry*.

Abstract:

Defining the specificity of influenza A has become the topic of intense study for decades, in part because all human infective strain originate in avian species, and thus there is the potential for highly pathogenic avian strains to evolve into human pandemics. While many crystal structures have been reported for influenza hemagglutinin bound to host glycans, very little research has been done to quantify the strength of these monomeric interactions. This is mainly because of the weak affinities of these interactions, which are often difficult to determine accurately. The rationalization of specificity has depended largely on the interpretation of contacts made by the glycan receptors in the hemagglutinin co-crystal structures, and on the qualitative affinity data from techniques such as glycan array screening. Glycan array screening is a rapid method to determine binding preferences, but it rarely provides quantitative affinity values, and even in those cases the contribution from monomeric binding (as seen in the crystal structures) can't be separated from multimeric effects arising from the array format. Thus quantitative structure/specificity relationships have been slow to emerge. From a theoretical perspective, computational simulations can predict monomeric binding affinities, however the lack of experimental values for comparison has hindered the development of these approaches. This study reports the development of an inhibition assay that permits monomeric affinity values to be accurately determined for human and avian-infective hemagglutinin (HA) with their host cell receptor glycans using BioLayer Interferometry (BLI). The measured solution K_D values agree with the reported values determined by NMR titration experiments and provide a basis for interpreting the interactions observed in hemaglutinin co-crystal complexes.

Introduction:

The World Health Organization reported the annual cost of influenza epidemics to the global economy at $71-167 billion (USD) [1]. In the United States alone, the direct costs of influenza have been estimated at $2.2 billion, and indirect costs at $8.8 billion [2]. In seasonal influenza epidemics, between 5 and 15% of the population are affected, with hospitalization and deaths mainly occurring in high-risk groups (elderly, chronically ill) [3]. Although difficult to assess, these annual epidemics are thought to result in between 3 and 5 million cases of severe illness, and between 250,000 and 500,000 deaths globally each year [4]. In influenza A there are currently 18 hemagglutinin (HA) and 11 neuraminidase (NA) subtypes, classified in accordance with the antigenic properties of HA and NA, leading to the familiar nomenclature: H1N1, H5N1, etc [5]. Viral influenza infection starts with hemagglutinin adhesion to its host cell surface [6]. Hemagglutinin binds to host cell surface glycans terminating with sialic acid (Sia, Neu5Ac) and linked to galactose (Gal) [7]. Neuraminidase on the other hand, is responsible for cleaving the Sia from the host receptor glycan, post infection, enabling the progeny virus to escape from the host cell surface. Factors that affect the ability of a particular strain of influenza to infect humans include level of exposure [8], glycan binding preferences of the viral surface HA [9], and activity of the viral surface NA [10]. If NA is too active, relative to the affinity of the HA, it will attenuate the ability of the virus to infect the host cell. Conversely, if the NA is too weak, it will impair shedding of the progeny virus, thus a balance between the activities of the NA and HA proteins are essential [10-11]. Virulence is further affected by such factors as the replication rate [12] and host immune competence [13]. This study will focus on the viral surface HA-host cell glycan interaction.

It has been reported that the HA in human infective viruses prefers to bind to glycans present on the human cell surface that terminate with the Siaα2-6Gal sequence, whereas the HA in avian-infective viruses prefers to bind to glycans that terminate in Siaα2-3Gal [9]. Some species, such as swine, can be co-infected by viruses that prefer Siaα2-3 or Siaα2-6Gal, leading to the potential for genetic reassortment (antigenic shift) to result in the introduction of Siaα2-6 binding preference (enhanced human infectivity) into a zoonotic framework [14].

Measuring monomeric glycan-protein interactions presents multiple challenges and technical difficulties. The majority of reported affinity values of glycan-protein interactions contain some component of avidity, due to the multimeric nature of hemagglutinin, which is a trimer, or the assay formats which may involve multimerization of the HA or glycan. Moreover, monomeric glycan-protein interactions are difficult to measure because they are weak (mM); current reported available techniques are solution Nuclear Magnetic Resonance (NMR) [15], Isothermal Titration Calorimetry (ITC) [16], and Microscale Thermophoresis (MST) [8].

As shown in Figure 5.1 [17], BLI is a similar technique to SPR in that it requires surface immobilization of one of the binding partners on a biosensor surface. Direct binding of small monomeric oligosaccharides to the immobilized HA is too weak to detect by BLI [17], and if the glycan is immobilized, it creates a multimeric binding interaction when a trimeric HA is employed as analyte. To boost the signal for glycan binding, while still maintaining the glycan as a monomeric analyte, a biotinylated form of the glycan was employed that could be combined with an anti-biotin Fab antibody fragment. This is comparable to the use of streptavidin to increase the analyte signal [18], but has the advantage of not multimerizing the analyte. The resulting Fab-glycans were used in all assays.

Two binding events are measured in this assay 1) the direct binding of Fab-labeled glycan to the immobilized HA, leading to the surface binding constant ($K_{D,surface}$), and 2) the inhibition of the direct binding by oligosaccharides that contain the minimal binding determinant typically observed in crystallographic co-complexes, leading to the IC_{50}. The inhibition constant (K_i) for the oligosaccharide is derived using Equation 1. When the oligosaccharide has the same structure as the glycan in the Fab-glycan analyte, the measured K_i equals the solution K_D of the glycan.

The solution K_D of the inhibitor (K_i) is obtained from **Equation 1**, and requires measurement of an IC_{50} for the analyte (oligosaccharide), and a surface K_D ($K_{D,surface}$) for the glycan.

$$K_i = IC_{50} / (1 + [protein]/K_{D,surface}) \qquad \textbf{Equation 1}$$

Here, we applied our previously reported BLI competitive assay [17] to quantify the monomeric affinity of influenza HA to 3' and 6'sialylated glycans (sialosides), with HAs from avian (A/Vietnam/1194/2004, a.k.a. VN1194) and human-infective (A/California/04/2009, a.k.a. CA04) viruses. We then compared the monomeric K_D or K_i values for linear glycans versus biantennary glycans to see whether or not the data supported a proposed bidentate mode of binding [17], in which both branches of a glycan simultaneously bind to a single HA trimer. Lastly, we quantified the inhibitory properties of six synthetic putative HA inhibitors, synthesized in the Woods group.

Results and Discussion:

Quantifying oligosaccharide-HA affinity

The affinities for six sialylated glycans (3/6'SLN, 3/6'SLN-N_3, and 3/6'SDLN-N_3), which are most commonly employed as ligands in HA co-complexes, were determined using the BLI competitive assay in order to quantify their solution K_D values (Table 5.1). Surprisingly, the avian VN1194 HA didn't show any affinity preference for the presumed cognate 3'sialylated glycans over the corresponding 6'glycans (Table 5.1). This observation is in contrast to the specificity patterns generally seen in glycan array screening [20]. As noted above, glycan array screening inevitably introduces avidity effects, which have been cited as being essential for amplifying the specificity differences among HAs [21].

An examination of the crystal contacts made in the complexes of VN1194 HA with 3'SLN (PDBID: 4BGY) and 6'SLN (PDBID: 4BGX) shows that many hydrogen bonds are the same in each complex, however there are differences (Figure 5.2). Overall there is one net additional interaction present in the complex with 3'SLN (GLN 222 – Gal O3/O4). An important value of the monomeric interaction energies determined herein is that they provide a basis for the discussion of the importance and strengths of interactions such as those in Figure 5.2.

The present K_D values for VN1194 also contrast with those reported by Xiong et al. [8], who reported solution K_D values from MST experiments indicating that the HA from VN1194 prefers that 3'SLN oligosaccharide over the 6' by a factor of 17 fold. This Micro Scaled Thermophresis (MST) data is the only existing data for this avian H5 HA-glycan monomeric affinity, and suggests a far weaker affinity for 6'SLN (-2.40 kcal/mol) than observed to date for

any other HA [17]. It is important to note that the MST technique has a dynamic range of 10 nM to mM [8], and thus may not be sufficiently accurate for this application.

As regards the HA from human-infective influenza (CA04), as expected the 6'SLN glycan is significantly preferred over the 3'SLN (Table 5.1). This observation is supported by crystallographic data that shows that the 6'SLN forms three additional hydrogen bonds compared 3'SLN (Figure 5.3). This is the first report of the solution K_D values of these glycans to the CA04 HA, hence there is no other literature with which to compare.

In terms of binding energies, the differences among the HAs and relevant oligosaccharides in Table 1 are small, the strongest being -3.83 kcal/mol to weakest of only -3.15 kcal/mol. While each of the small 3' or 6' oligosaccharides bind equally tightly to the HA from VN1194, the HA from CA04 displays a slightly weaker affinity for 3'sialosides and a slightly enhanced affinity for 6'sialosides. However the differences are remarkably subtle (less than 1 kcal/mol) and call into question the significance of the affinities derived for short non-biologically relevant glycans.

Quantifying glycan-HA affinity

To address the role of glycan substructure, we then determined the inhibitory ability of 3'SLN and 6'SLN against a range of large biologically relevant glycans that included branched structures of the type found on eukaryotic cell surfaces [21]. The glycan inhibition data (Figure 5.4) show remarkable differences depending on whether the HA has evolved to prefer 3' (VN1194) or 6'sialosides (CA04). In the case of the HA from VN1194, lengthening the glycan leads to a decrease in binding affinity of as much as approximately 1 kcal/mol as the glycans begin to approach biologically-relevant lengths. This decrease appears to plateau with chain

lengths beyond two LacNAc (TLN) repeats (Table 5.1). Furthermore, there is no appreciable difference in the affinity for biantennary glycans over linear structures of the same branch length. These data strongly support the view that the 3'sialosides do not form bidentate interactions, but rather, each arm functions independently.

In contrast to the case of VN1194 binding to avian flu receptors, the affinity of the human infective CA04 HA shows a marked preference for longer glycans, plateauing again with chain lengths beyond two LacNAc (LN) repeats. However in the case of 6'sialosides binding to CA04, a small (0.25 kcal/mol) but statistically significant boost in the affinity is observed for biantennary glycans compared to linear structures of the same branch length (Table 5.2). This is consistent with the hypothesis that biantennary 6'sialosides of sufficient length may form bidentate binding [17]. However, the boost from this interaction is less than would be expected from a simplistic additive model of affinity. That is, if the linear 6'STLN has an approximate affinity of -4.12 (from $\Delta G_{binding} + \Delta\Delta G = -3.99 + -0.13$), then we would expect bidentate binding to result in an affinity of twice the linear value, or, -8.24 kcal/mol. The fact that almost 8 kcal/mol is lost upon bidentate binding, is explicable by conformational entropic penalties, as would be expected for immobilizing such a large flexible glycan (Figure 5.5) [21].

Quantifying HA-inhibitor affinity

Lastly, the competition assay was employed to quantify the inhibitory ability of various synthetic glycomimetics of 3'SG (Table 5.3).). IC_{50} values were reported and categorized into two groups. Group one was categorized when compounds inhibit human H1 protein from A/California/04/2009 binding to 6'SLN-Fab. Similarly, avian H5 protein from A/Vietnam/1194/2004 binding to 3'SLN-Fab was inhibited in Group two. Among all six

synthesized mimetics, compound FB127 showed the best inhibition. Inhibitor 127 was able to inhibit both H1 and H5 proteins from binding the cognate ligand 3'SG. Furthermore, inhibition curves of compound FB127 (FB127) against 3'SLN-Fab binding to four HA subtypes (H1 A/California/04/2009, H3 A/Brisbane/10/2007, H5 A/Vietnam/1194/2004, and H7 A/Anhui/1/2013) were illustrated in Figure 5.6. Compound FB127 presented good inhibition in all four HA subtypes comparing to the natural glycan 3'SG. This data strongly suggested that compound FB127 was a good candidate for influenza hemagglutinin inhibition in major subtypes.

FB127 is the first small molecule glycomimetic reported that has higher affinity than the endogenous glycan 3'SG.

Conclusion:

We have demonstrated the precision and accuracy of a BLI-based competition assay and used it to generate solution K_D values for monomeric glycan-HA interactions. Avian HA did not show binding preference to short 3'sialosides over 6'sialosides. However, human HA preferred 6'sialosides 2-3 fold in solution K_D values or 0.4 to 0.6 kcal/mol in binding free energy. It is remarkable that such small energy differences are sufficient to translate into species infectivity preferences.

We observed that the length of the glycan have a significant affect on the affinity. Longer 3' sialosides bind weaker to Avian HA than shorter 3'sialosides. Therefore short (biologically irrelevant) glycans should be considered to be a relatively limited model for interpreting biological recognition. It was also noted that longer 6'sialosides bind tighter to human HA than shorter sialosides. Again this emphasized the danger of employing short (biologically irrelevant) glycans as a model for biological recognition. Long 3' biantennary sialosides showed no difference in affinity for avian HA, relative to long linear glycans, consistent with the prediction [21] that they can't form bidentate interactions. Whereas, long 6' biantennary sialosides showed a small *but statistically significant* boost in affinity for human HA, relative to long linear glycans, consistent with the formation of bidentate binding [21].

Using the relationship $\Delta G_{L2} = \Delta G_{L1} + \Delta\Delta G_{L1,2}$, the longest linear 3'sialoside (3'STetraLN) displayed an approximate K_D of -4.02 + 1.00 = -3.02 kcal/mol (Table 5.2), while the longest linear 6'sialoside (6'STetraLN) displayed an approximate K_D of -3.99 + -0.18 = -4.17 kcal/mol (Table 5.2). Thus the human HA prefers long glycans by just over 1 kcal/mol more than the avian HA. When biantennary glycans are compared, the human HA prefers long biantennary glycans by -4.43 kcal/mol (6'STetraLN-Bi) compared to the corresponding avian

value of -2.81 kcal/mol. These values suggest that more multimeric interactions involving multiple 3'sialosides would be required to achieve the same level of affinity/infectivity as an HA evolved to recognize human 6'sialosides.

Methodology:

Experimental materials: Influenza A Hemagglutinin H1N1 A/California/04/2009 (SinoBiological Inc. Cat#11055-V08H) and H5N1 A/Vietnam/1194/2004 HA (SinoBiological Inc. Cat#11062-V08H1), 3'-Sialylgalactose sodium salt (3'SG, Prod#: OS09486, Carbosynth, UK), 6'-Sialylgalactose sodium salt (6'SG, Prod#: OS15416, Carbosynth, UK), 3'-Sialyl-*N*-acetyllactoseamine (3'SLN, Cat#: SLN302, Dextra, UK), 6'-Sialyl-*N*-acetyllactoseamine (6'SLN, Cat#: SLN306, Dextra, UK) were purchased from their commercial resources. All azido and biotinylatd glycans were received from two providers (Table 5. 4). A Hemagglutinin H3N2 A/Brisbane/10/2007 (BEI resources, Cat#NR-19238, USA) and H7N9 A/Anhui/1/2013 (BEI resources, Cat#NR-45118, USA) were given by Dr. Mark Tompkins from Infectious Disease in University of Georgia, Athens, GA. Anti-biotin-Fab antibody was purchased from Rockland Inc. (Cat# 800-101-098). Fab-glycan was prepared as incubating Fab with biotinylated glycans at 1:1.1 mole ratio in dart at 4°C overnight. HA proteins were weighted and dissolved in the HA buffer: PBS buffer at pH7.4, at 25°C. Fab-glycans were all prepared in HEPES buffer at pH7.4 at 25°C.

Additionally, with the availability of BLI-based competition assay, IC_{50} values were also determined and reported in this work for 6 inhibitors (Figure 5.7) previously synthesized by group members.

Immobilization of HA on streptavidin biosensors: Amine Reactive Second Generation biosensors (AR2G, Cat#: 18-5092, Pall ForteBio Corp., Menlo Park, CA, USA) were activated after being dipped into a mixture of 20mM EDC and 10mM sulfo-NHS solution for 900s. Then HA proteins

were coupled onto Amine Reactive Second Generation biosensors (AR2G, Cat#: 18-5092, Pall ForteBio Corp., Menlo Park, CA, USA) at 20 μg/mL for 1800s. Then the loaded HA biosensors were dipped into 1M ethanolamine (Cat#: 110167-100mL, Sigma-Aldrich, St. Louis, MO, USA) pH8.5 to quench the possible unreacted carboxylic group on the AR2G biosensor surface for 600s. The immobilization of HA onto AR2G biosensors resulted in ~3nm as loading signal under this condition.

Protein BLI direct binding assay ($K_{D,\ surface}$): Linear Fab-glycan direct binding K_D (HA biosensor surface K_D) was measured using an BioLayer Interferometer (BLI) Octet Red 96 system (Pall ForteBio Corp., Menlo Park, CA, USA) and data acquired using ForteBio Data Acquisition 8.2 software (Pall ForteBio Corp., Menlo Park, CA, USA). The direct binding experiment was performed at 360s for association and 240s for dissociation in HEPES buffer pH7.4. Fab-glycan was prepared in two-fold serial dilution in HEPES buffer from 0~8μM, in the replicates of three. Surface K_D ($K_{D,\text{surface Fab-glycan biosensor}}$) was then calculated by ForteBio Data Analysis 8.2 software (Pall ForteBio Corp., Menlo Park, CA, USA) and Microsoft Office Excel 2011 (Microsoft, USA). Surface K_D ($K_{D,\text{surface Fab-glycan biosensor}}$) was determined by 1:1 binding model from both steady state analysis and scatchard plot and in triplicates.

Protein BLI inhibition assay (IC_{50}): Fab-glycan was prepared at 1μM in HEPES buffer in a large volume for protein inhibition assay. 3'SLN-N_3(Tr33), 6'SLN-N_3 (Tr36), 3'SDLN-N_3 (Te175), 6'SDLN-N_3 (Te176), 3'SLN, and 6'SLN were prepared in two-fold serial dilution in water from 0,1.25, 2.5, 5, 10, 20, 40, and 80mM. 100μL of 1μM Fab-glycan, 20μL of prepared azido-glycan or free glycan at its concentration, and 80μL of HEPES buffer were mixed and incubated at room

temperature for 1 hour. The inhibition assay was performed on Octet Red 96 at baseline time 120s, association time 360s, and dissociation time 240s at shaker speed 1000RPM at room temperature, in replicates of three. IC_{50} was calculated using three-parameter dose-response inhibition model in GraphPad Prism 7 (GraphPad, La Jolla, CA, USA).

Fab-glycan IC_{50} value measurement: Fab-labeled both linear and biantennary glycans (SLN, S-diLN, S-triLN, S-tetraLN) were prepared at 1μM in HEPES buffer in a large volume for protein inhibition assay. 3'SLN-N_3(Tr33) and 6'SLN-N_3 (Tr36) were prepared in two-fold serial dilution in water from 0,1.25, 2.5, 5, 10, 20, 40, and 80mM. 100μL of 1μM Fab-glycan (6'SLN-Fab was used on CA04-immobilized biosensors, and 3'SLN-Fab was used when inhibition measures on VN1194-immobilized biosensors), 20μL of prepared glycan at its concentration, and 80μL of HEPES buffer were mixed and incubated at room temperature for 1 hour. The inhibition assay was performed on Octet Red 96 at baseline time 120s, association time 360s, and dissociation time 240s at shaker speed 1000RPM at room temperature, in replicates of three. IC_{50} was calculated using three-parameter dose-response inhibition model in GraphPad Prism 7 (GraphPad, La Jolla, CA, USA).

Inhibitor IC_{50} value measurement: Fab-glycan was prepared at 1μM in HEPES buffer in a large volume for protein inhibition assay. 3'SLN-N_3(Tr33), 6'SLN-N_3 (Tr36), 3'SDLN-N_3 (Te175), 6'SDLN-N_3 (Te176), 3'SG, 6'SG and six inhibitors (FB127, 143, 122, 142, 145, and 146) were prepared in two-fold serial dilution in water from 0,1.25, 2.5, 5, 10, 20, 40, and 80mM. 100μL of 1μM Fab-glycan (6'SLN-Fab was used on CA04-immobilized biosensors, and 3'SLN-Fab was used when inhibition measures on VN1194-immobilized biosensors), 20μL of prepared glycan at

its concentration, and 80µL of HEPES buffer were mixed and incubated at room temperature for 1 hour. The inhibition assay was performed on Octet Red 96 at baseline time 120s, association time 360s, and dissociation time 240s at shaker speed 1000RPM at room temperature, in replicates of three. IC_{50} was calculated using three-parameter dose-response inhibition model in GraphPad Prism 7 (GraphPad, La Jolla, CA, USA).

Equation correction: in Equation 1, the concentration of the Fab-glycan was considered as free concentration before the inhibiting glycan was introduced into the equilibrium. This equation may introduce errors in generating inhibition equilibrium constant when the initial concentration is not the same as the total concentration at certain situations [19]. Hence, Munson and Rodbard [19] made a further correction to the Equation 1 as shown in Equation 2. In their study, y_0 was the initial bound to free ratio for the Fab-glycan before the inhibitor was introduced into the equilibrium. When y_0 is very small ($y_0<0.1$), such a correction can be ignored. Theoretically, there is no more than 1.6 femto mole of HA molecules immobilized on the biosensor surface. Hence, the bound to the surface concentration should be 8nM=1.6E-12/200E-6) as in 200µL assay volume. With the known free Fab-glycan concentration in solution 0.5µM, the y_0 values equates to 0.016=(8nM)/(0.5µM). This theoretical number is smaller than 0.1 and should not contribute to errors of Ki. However, a further glycan density assay would be needed to explore this concern.

$$K_i = \frac{IC_{50}}{1 + \frac{L_T(y_0 + 2)}{2 \times K_d(y_0 + 1)} + y_0} - K_d \left(\frac{y_0}{y_0 + 2} \right)$$

Equation 2

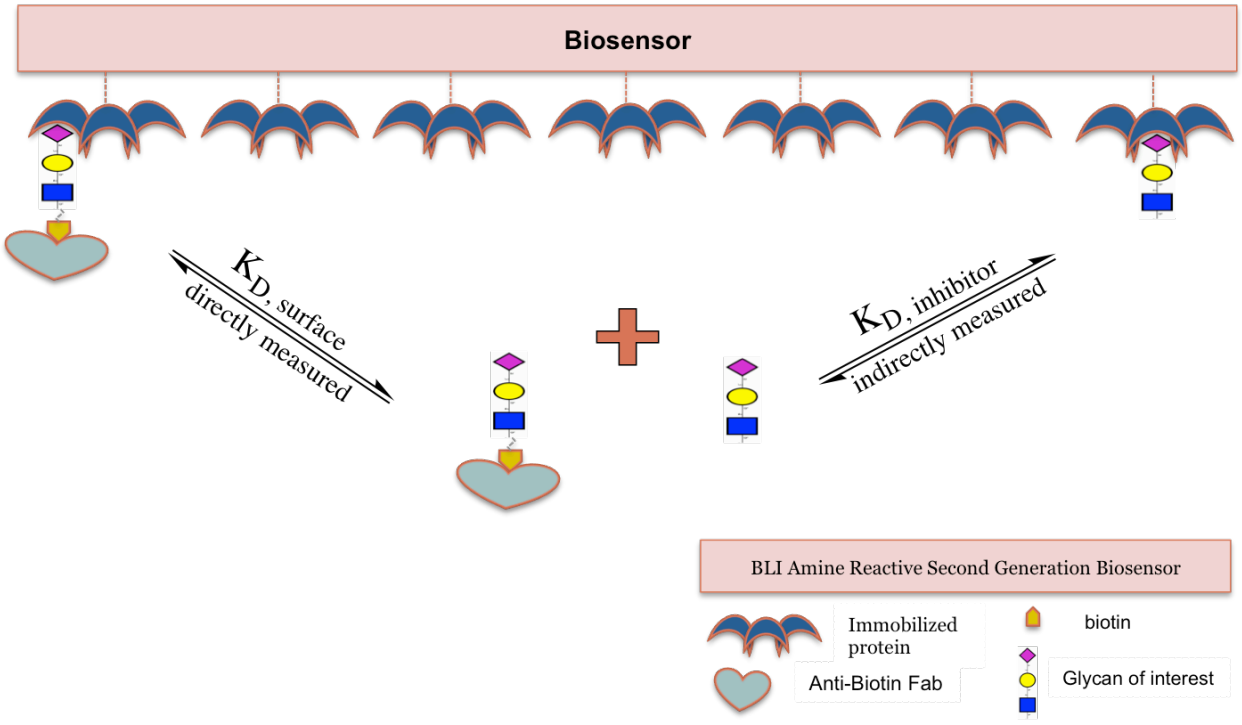


Figure 5.1. Schematic representation of the BLI-based competition assay. The ability of a small carbohydrate (typically, the minimal binding determinant) to inhibit (IC_{50}) the direct binding of the analyte of interest ($K_{D,\ surface}$) is determined in two sets of experiments. From these values, the solution K_D ($K_{D,\ inhibitor}$) of the minimal binding determinant may be concluded.

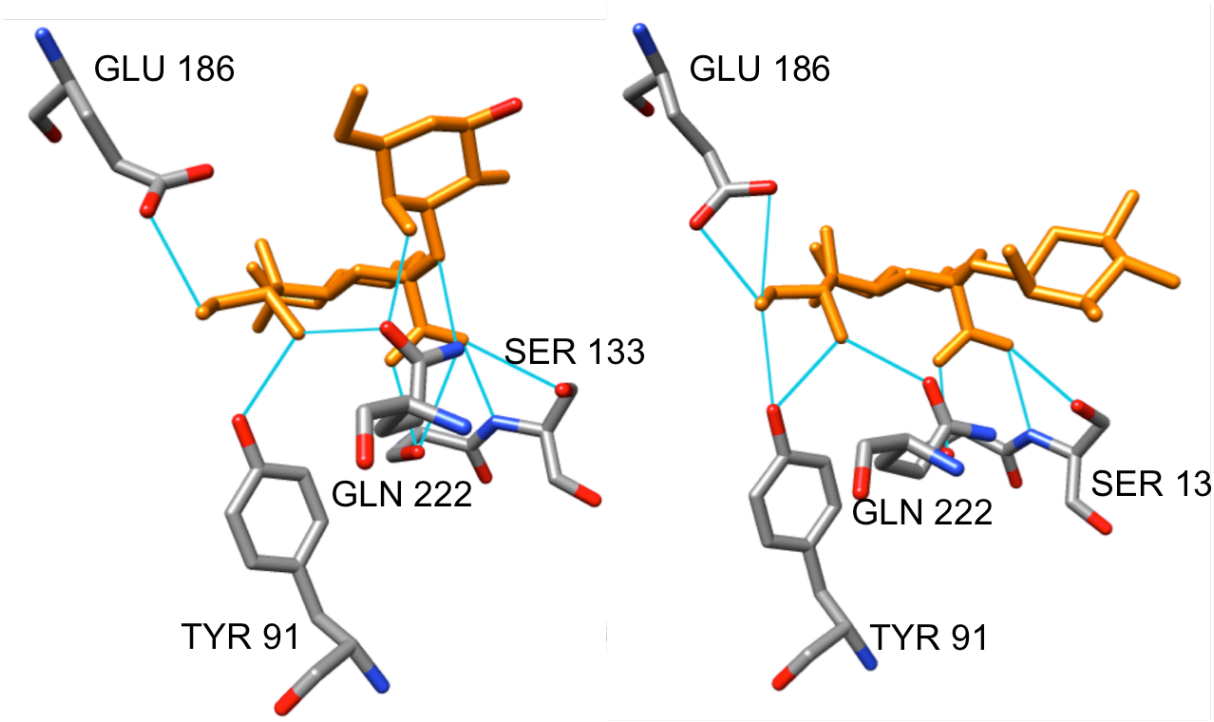


Figure 5. 2. Hydrogen bond contacts identified using Chimera [22] in the complexes of 3'SLN (left) and 6'SLN (right) with the HA from avian-infective VN1194.

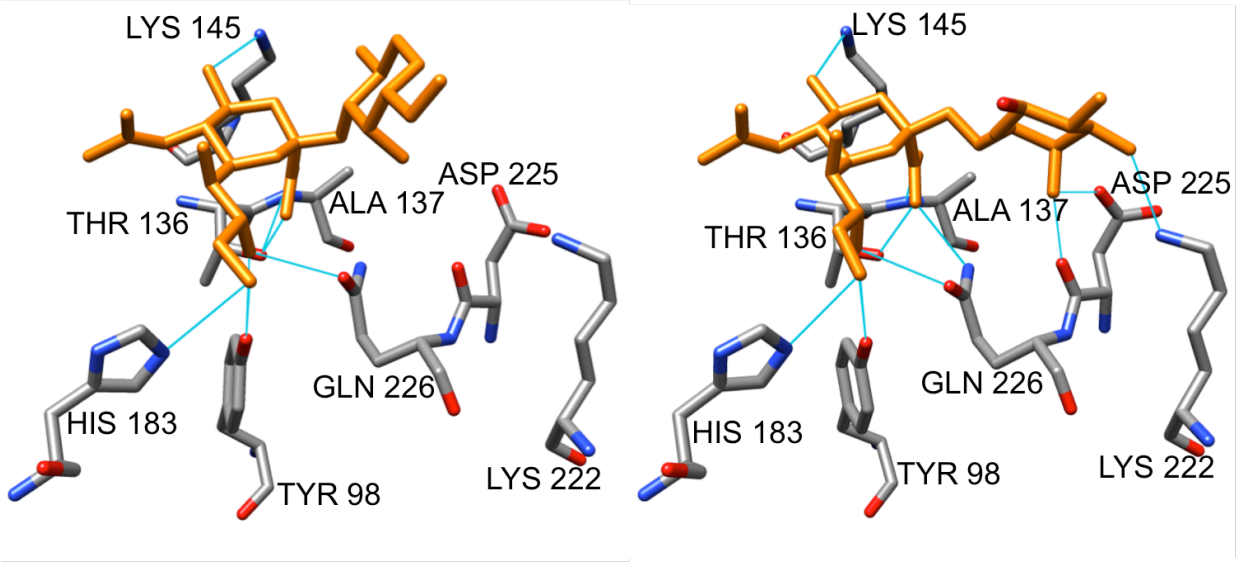


Figure 5.3. Hydrogen bond contacts identified using Chimera [22] in the complexes of 3'SLN (left) and 6'SLN (right) with the HA from human-infective CA04.

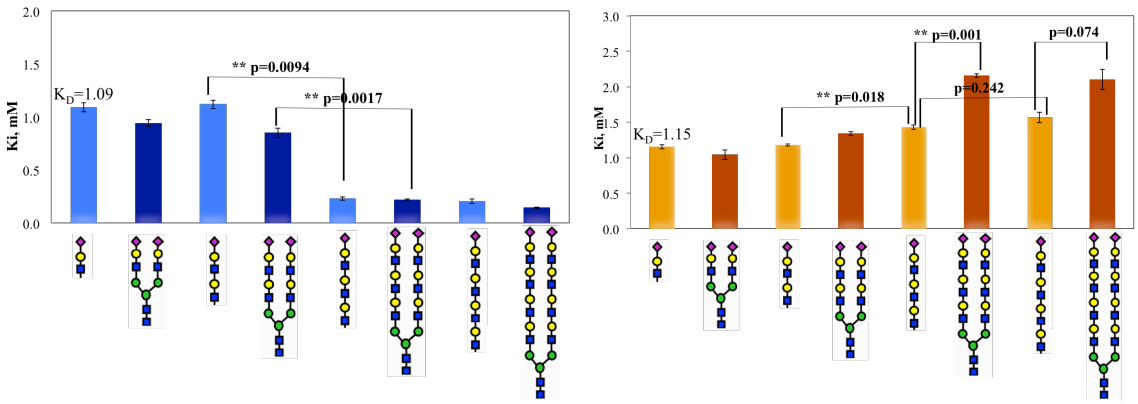


Figure 5.4. Left: K_i values for 3'SLN-N_3 inhibiting Fab-glycans from binding to HA from VN1194; Fab-labeled linear glycans (light blue) and biantennary glycans (dark blue). Right: K_i values for 6'SLN-N_3 inhibiting Fab-glycans from binding to HA from CA04; Fab-labeled linear glycans (light orange) and biantennary glycans (dark orange). Glycans are shown in 2D-SNFG representation [24]). Standard deviation of the mean presented in the error bars. Experiments were done in replicates of three. P values are calculated from one tailed student t test.

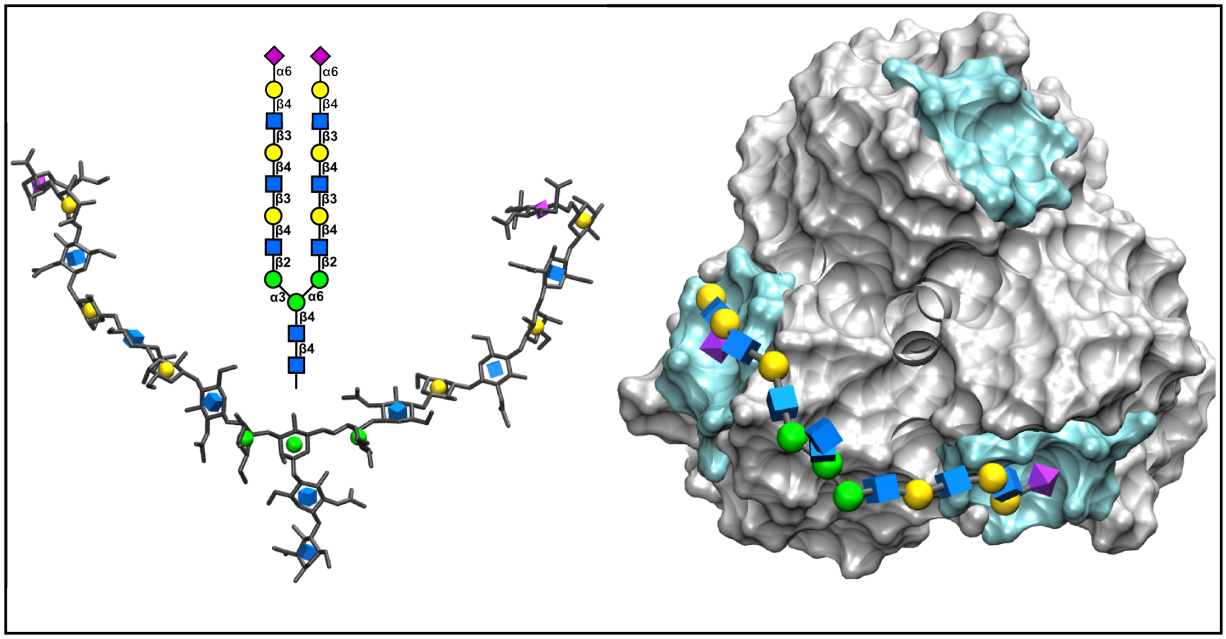


Figure 5.5. Proposed [21] bidentate binding of a biantennary α2-6 glycan (3D-SNFG representation [24]) to the HA (grey surface) from a pandemic H1N1 (CA04).

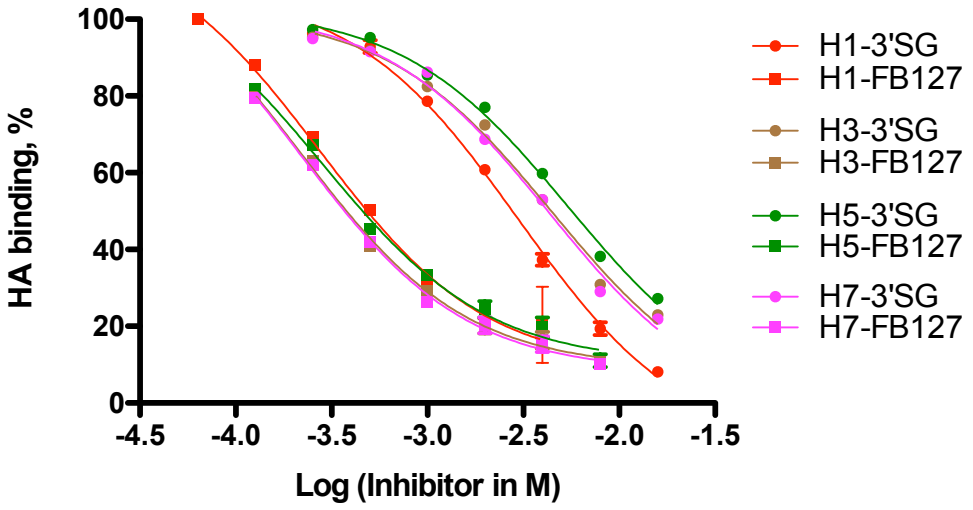


Figure 5.6. Inhibitor FB127 inhibits 3'SLN-Fab binding to all H1, H3, H5, and H7 HA compared to the natural glycan 3'SG.

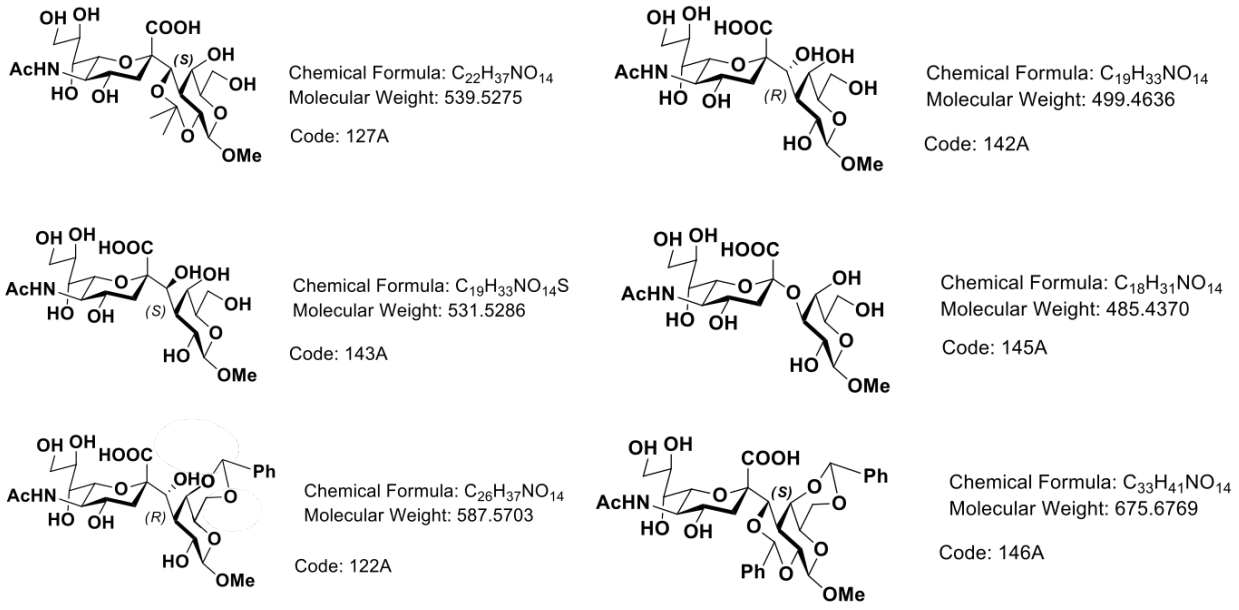


Figure 5.7. Structures of six synthesized HA inhibitors.

Table 5.1. Solution K_D values (mM) and derived binding free energies (ΔG, kcal/mol) determined from BLI competitive assay.

Viral Strain	Canonical Specificity	Inhibitor	Fab-Glycan	K_D	ΔG	Reported K_D (mM)
VN1194	α2-3	3'SLN-N_3	3'SLN-sp-Fab	2.6 ± 0.07[a]	-3.57	
		3'SLN	3'SLN-sp-Fab	2.4 ± 0.01[b]	-3.62	1.1 ± 0.2[8]
		3'SDLN-N_3	3'SDLN-sp-Fab	2.1 ± 0.02[c]	-3.70	
		6'SLN-N_3	6'SLN-sp-Fab	2.5 ± 0.01[a]	-3.59	
		6'SLN	6'SLN-sp-Fab	2.1 ± 0.01[b]	-3.70	17 ± 3[8]
		6'SDLN-N_3	6'SDLN-sp-Fab	2.3 ± 0.05[c]	-3.64	
Viral Strain	**Canonical Specificity**	**Ligand**	**Ligand**	**K_D**	**ΔG**	**Reported**
CA04	α2-6	3'SLN-N_3	3'SLN-sp-Fab	3.2 ± 0.02[d]	-3.45	
		3'SLN	3'SLN-sp-Fab	4.8 ± 0.13[e]	-3.20	
		3'SDLN-N_3	3'SDLN-sp-Fab	4.4 ± 0.00[f]	-3.26	
		6'SLN-N_3	6'SLN-sp-Fab	1.6 ± 0.04[d]	-3.86	
		6'SLN	6'SLN-sp-Fab	1.6 ± 0.03[e]	-3.86	
		6'SDLN-N_3	6'SDLN-sp-Fab	1.5 ± 0.01[f]	-3.90	

[a]Not statistically different ($p = 0.9876$). [b]Not statistically different ($p = 0.1365$). [c]Not statistically different ($p = 0.0733$). [d]Statistically highly different ($p < 0.0001$). [e]Statistically highly different ($p = 0.0002$). [f]Statistically highly different ($p < 0.0001$). p values calculated from two tailed t test. All experiments were done in replicates of three. Errors were standard deviation of the mean.

Table 5.2. Relative Binding Free Energies (kcal/mol)

3'Glycan – VN1194	ΔΔG[a]	6'Glycan – CA04	ΔΔG
3'SLN[b]	0.00	6'SLN[c]	0.00
3'SLN-Bi	0.09	6'SLN-Bi	0.06
3'SDLN	-0.02	6'SDLN	-0.02
3'SDLN-Bi	0.15	6'SDLN-Bi	-0.09
3'STLN	0.92	6'STLN	-0.13
3'STLN-Bi	0.94	6'STLN-Bi	-0.37
3'STetraLN	1.00	6'STetraLN	-0.18
3'STetraLN-Bi	1.21	6'STetraLN-Bi	-0.27

The K_D and K_i data in Figures 8 and 9 were converted into relative binding Free Energy using $\Delta\Delta G = RT\ln(K_D/K_i)$. [b]$\Delta G_{binding}$ = -4.02 kcal/mol. [c]$\Delta G_{binding}$ = -3.99 kcal/mol

Table 5.3. IC_{50} values of vary compounds to inhibit two binding groups.

Compound	Group One IC_{50} (mM)	Group Two IC_{50} (mM)
6'SG	4.51 ± 0.082	10.47 ± 0.198
3'SG	12.59 ± 0.024	4.51 ± 0.004
145	3.05 ± 0.031	4.81 ± 0.059
122 R-Bz	2.08 ± 0.010	2.25 ± 0.131
127 S-iP	0.28 ± 0.028	0.40 ± 0.008
142 R-OH	0.63 ± 0.010	1.03 ± 0.015
146 S-Bz 4,6Bz	0.98 ± 0.016	ND[b]
143 S-OH	1.00 ± 0.017	3.63 ± 0.031

[a] sp = -CH2CH2-

[b]ND not detected.

All experiments were done in replicates of three. Errors were standard deviation of the mean.

Table 5.4: Glycans received from courtesy providers.

Glycan structures	Abbreviation	Provider
Siaα2-3Galβ1-4GlcNAcβ-sp[a]-NH-LCLC[b]-biotin	3'SLN-biotin (CFG#B84)	Consortium for Functional Glycomics (CFG)
Siaα2-6Galβ1-4GlcNAcβ-sp-NH-LCLC-biotin	6'SLN-biotin (CFG#B87)	
Siaα2-3Galβ1-4GlcNAcβ-sp-N_3	3'SLN-N_3 (CFG#Tr33)	
Siaα2-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ-sp-N_3	3'SDLN-N_3 (CFG#Te175)	
Siaα2-6Galβ1-4GlcNAcβ-sp-N_3	6'SLN-N_3 (CFG#Tr36)	
Siaα2-6Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ-sp-N_3	6'SDLN-N_3 (CFG#Te176)	
Siaα2-3/6Galβ1-4GlcNAcβ-spNH-LCLC-biotin	3/6'SLN-biotin	Dr. James Paulson from Scripps, CA
Siaα2-3/6Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ-spNH-LCLC-biotin	3/6'SDLN-biotin	
Siaα2-6Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ-spNH-LCLC-biotin	3/6'STLN-biotin	
Siaα2-6Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ-spNH-LCLC-biotin	3/6'STetraLN-biotin	
Siaα2-3/6Galβ1-4GlcNAcβ-Bi-Asn-LCLC-biotin	3/6'SLN-Bi-biotin	
Siaα2-3/6Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ-Bi-Asn-LCLC-biotin	3/6'SDLN-Bi-biotin	
Siaα2-6Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ-Bi-Asn-LCLC-biotin	3/6'STLN-Bi-biotin	
Siaα2-6Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAcβ-Bi-Asn-LCLC-biotin	3/6'STetraLN-Bi-biotin	

[a]sp: -CH_2CH_2-

[b]LCLC: -CO-$(CH_2)_5$-NHCO-$(CH_2)_5$-NH-

References:

1. Meltzer, M. I., Cox, N. J., and Fukuda, K. (1995) The economic impact of pandemic influenza in the United States: Priorities for intervention. *Emerg Infect Dis*. **5**, 659–671
2. US Congress. Office of Technology Assessment: Cost-effectivness of Influenza Vaccination. GPO, Washington, DC; 1981.
3. https://www.cdc.gov/media/releases/2017/p1213-flu-death-estimate.html
4. Lemartre, M, and Carrat, F. (2010) Comparative age distribution of influenza morbidity and mortality during seasonal influenza epidemics and the 2009 H1N1 pandemic. *Emerg Infect Dis*. **10**, 162–167
5. Schrauwen, E. J., and Fouchier, R. A. (2014) Host adaptation and transmission of influenza A viruses in mammals. Emerg Microbes Infect 3, e9
6. Reuter, J. D., Myc, A., Hayes, M. M., Gan, Z., Roy, R., Qin, D., Yin, R., Piehler, L. T., Esfand, R., Tomalia, D. A., and Baker, J. R. (1999) Inhibition of Viral Adhesion and Infection by Sialic-Acid-Conjugated Dendritic Polymers. Bioconjugate Chem. 10, 271-278
7. Rogers, G. N., and Paulson, J. (1983) Receptor Determinants of Human and Animal Influenza Virus Isolates: Differences in Receptor Specificity of the H3 Hemagglutinin Based on Species of Origin. Virology 127, 361-373
8. Xiong, X., Coombs, P. J., Martin, S. R., Liu, J., Xiao, H., McCauley, J. W., Locher, K., Walker, P. A., Collins, P. J., Kawaoka, Y., Skehel, J. J., and Gamblin, S. J. (2013) Receptor binding by a ferret-transmissible H5 avian influenza virus. Nature 497, 392-396
9. de Graaf, M., and Fouchier, R. A. (2014) Role of receptor binding specificity in influenza A virus transmission and pathogenesis. EMBO J 33, 823-841

10. Wagner, R., Matrosovich, M., and Klenk, H. D. (2002) Functional balance between haemagglutinin and neuraminidase in influenza virus infections. Rev Med Virol 12, 159-166 1.

11. Gamblin, S. J., and Skehel, J. J. (2010) Influenza Hemagglutinin and Neuraminidase Membrane Glycoproteins. J. Biol. Chem 285, 28403-28409

12. Hirst, G.K., *The Agglutination of Red Cells by Allantoic Fluid of Chick Embryos Infected with Influenza Virus.* Science, 1941. **94**(2427): p. 22-3.

13. Thangavel, R. R., and Bouvier, N. M. (2014) Animal models for influenza virus pathogenesis, transmission, and immunology. J Immunol Methods 410, 60-79

14. Newhouse, E. I., Xu, D., Markwick, P. R. L., Amaro, R. E., Pao, H. C., Wu, K. J., Alam, M., McCammon, J. A., and Li, W. W. (2009) Mechanism of Glycan Receptor Recognition and Specificity Switch for Avian, Swine, and Human Adapted Influenza Virus Hemagglutinins: A Molecular Dynamics Perspective. J. Am. Chem. Soc. 131, 17430-17422

15. Sauter, N. K., Bednarski, M. D., Wurzburg, B. A., Hanson, J. E., Whitesides, G. M., Skehel, J. J., and Wiley, D. C. (1989) Hemagglutinins from Two Influenza Virus Variants Bind to Sialic Acid Derivatives with Millimolar Dissociation Constants: a 500-MHz Proton Nuclear Magnetic Resonance Study. Biochemistry 28, 8388–8396

16. Bewley, C. A., Cai, M., Ray, S., Ghirlando, R., Yamaguchi, M., and Muramoto, K. (2004) New Carbohydrate Specificity and HIV-1 Fusion Blocking Activity of the Cyanobacterial Protein MVL: NMR, ITC and Sedimentation Equilibrium Studies. J. Mol. Biol. 339, 901-914

17. Ji, Y and Woods, R. (2018) Quantifying weak glycan-protein interactions using a Biolayer Interferometry competition assay: Applications to ECL lectin and X-31 influenza hemagglutinin. Glycobiophysics March 25, 2018 In press.

18. Fialova, D., Krejcova, L., Janu, L., Blazkova, I., Krystofova, O., Hynek, D., Kopel, P., Drbohlavova, J., Konecna, M., Vaculovicova, M., Kynicky, J., Hubalek, J., Babula, P., Kizek, R., and Adam, V. (2013) Flow Injection Electrochemical Analysis of Complexes of Influenza Proteins with CdS, PbS and CuS Quantum Dots. International Journal of Electrochemical Science 8, 10805-10817

19. Munson, P. J., and Rodbard, D. (1988) An exact correction to the "Cheng-Prusoff" correction. *J. Recptor Res*. 8, 533-546

20. Liao, H.-Y., Hsu, C.-H., Wang, S.-C., Liang, C.-H., Yen, H.-Y., Su, C.-Y., Chen, C.-H., Jan, J.-T., Ren, C.-T., Chen, C.-H., Cheng, T.-J. R., Wu, C.-Y., and Wong, C.-H. (2010) Differential Receptor Binding Affinities of Influenza Hemagglutinins on Glycan Arrays. J. Am. Chem. Soc. 132, 14849-14856

21. Peng, W., de Vries, R. P., Grant, O. C., Thompson, A. J., McBride, R., Tsogtbaatar, B., Lee, P. S., Razi, N., Wilson, I. A., Woods, R. J., and Paulson, J. C. (2017) Recent H3N2 Viruses Have Evolved Specificity for Extended, Branched Human-type Receptors, Conferring Potential for Increased Avidity. Cell Host Microbe 21, 23-34

22. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004) UCSF Chimera - A Visualization System for Exploratory Research and Analysis. J. Comp. Chem. 25, 1605-1612

23. Thieker, D. F., Hadden, J. A., Schulten, K., and Woods, R. J. (2016) 3D implementation of the symbol nomenclature for graphical representation of glycans. Glycobiology 26, 786-787 23234

CHAPTER 6

PREDICTING THE ORIGINS OF ANTI-BLOOD GROUP ANTIBODY SPECIFICITY: A CASE STUDY OF THE ABO A- AND B-ANTIGENS

Ye Ji. Partial work from *Frontier in Immunology* 2014, 5:1-9. Reprint here with permission of publisher

Abstract

The ABO blood group system is the most important blood type system in human transfusion medicine. Here, we explore the specificity of antibody recognition towards ABO blood group antigens using computational modeling and biolayer interferometry. Automated docking and molecular dynamics (MD) simulations were used to explore the origin of the specificity of an anti-blood group A antibody variable fragment (Fv AC1001). The analysis predicts a number of Fv-antigen interactions that contribute to affinity, including a hydrogen bond between a His^{L49} and the carbonyl moiety of the GalNAc in antigen A. This interaction was consistent with the dependence of affinity on pH, as measured experimentally; at lower pH there is an increase in binding affinity. Binding energy calculations provide unique insight into the origin of interaction energies at a per-residue level in both the scFv and the trisaccharide antigen. The calculations indicate that while the antibody can accommodate both blood group A and B antigens in its combining site, the A antigen is preferred by approximately 4 kcal/mol, consistent with the lack of binding observed for the B antigen.

Keywords: Molecular docking, MD simulations, Blood group antigens, Antibody specificity, GLYCAM, AMBER

Introduction

Since its discovery in 1900 (1), the ABO blood group system has played a crucial role in defining human blood and tissue compatibility. The blood type of an individual indicates the presence or absence of relevant antigens and antibodies. The three blood types share a core oligosaccharide antigen (H), and based on the glycosyl transferases inherited, different antigens are synthesized (2-4); type A transferase adds a terminal non-reducing *N*-acetylgalactosamine (GalNAc) residue; type B transferase adds galactose (Gal), whereas individuals with blood group O retain the unmodified H antigen. During the first years of life, the immune system forms antibodies upon exposure to non-self antigens from various exogenous factors. Thus an A-type individual will have circulating antibodies specific for the B-antigen, and vice-versa. The high degree of specificity is notable given that the only difference between the structures of the A- and B-antigens is the replacement of an acetamido moiety (in A) with a hydroxyl group (in B). Because of the presence of circulating antibodies, a mismatched blood transfusion or organ transplant can lead to hyperacute immune response and death (5, 6). Additionally, under certain circumstances, incompatibilities in blood groups between mother and child can trigger the mother's immune system to produce antibodies against the fetus, causing hemolytic disease (7).

Alterations in the structures of the ABO antigens often occur during carcinogenesis and therefore they have also been considered tumor markers (8, 9). Recently, strong correlations have been established between the presence of particular ABO and Lewis antigens, and susceptibility to infectious diseases, such as *Helicobacter pylori*, norovirus, and cholera (10), wherein the blood group antigens can be exploited as receptors for bacterial and viral adhesion. Conversely,

it has been suggested that endogenous anti-blood group antibodies can recognize blood-group-like carbohydrate antigens on pathogen surfaces, conferring protection against infection (11).

Despite their clinical importance, relatively little is known about the structural basis for these highly specific antibodies – antigen interactions. Although X-ray crystallography has been used to characterize antibody-carbohydrate complexes, the generally enhanced flexibility and conformational heterogeneity of oligosaccharides, detracts from the ability to generate co-crystals (12). Additionally, anti-carbohydrate antibodies bind to their antigens with an affinity that is 3-5 orders of magnitude lower than typical antibodies that bind to protein or peptide antigens. Difficulties in generating 3D structures for carbohydrate-antibody complexes has led to the increasing use of theoretical structure prediction methods (13, 14), which, while convenient, are prone to predicting false positives due to inaccuracies in pose scoring functions(15) and to the omission of carbohydrate conformational preferences(16) .

In this study, we examined the structural origin of the antigenicity (the specificity and affinity) of a monoclonal antibody raised against blood group A (BGA) antigen, for which an apo structure of the single-chain variable fragment (scFv AC1001) has been reported (17). The specificity data from screening two independent glycan arrays (Consortium for Functional Glycomics (v4.0, request ID: 1808) and from the group of Jeff Gildersleeve) confirmed that the scFv displayed no detectable binding to any B-antigens and only bound to BGA-containing glycans. To provide a structural interpretation for the specificity of AC1001 for BGA over blood groups H (BGH) and B (BGB), we generated a 3D model of the immune complex using molecular docking and refined it by molecular dynamics (MD) simulation. Despite its limitations, molecular docking, with or without additional experimental constraints, such as from

NMR data, is often the only approach that may be employed to generate the structure of a ligand-protein complex, in the absence of direct crystallographic data. To enhance the success rate a recent carbohydrate conformational energy function (16) was employed with AutoDock VINA (18), which quantifies the conformational preferences of oligosaccharides based on their glycosidic torsion angles. MD simulations (50 ns) were subsequently performed to ensure that the docked complexes were stable under physically realistic conditions, and in that event, the MD data were employed in binding free energy calculations. A particular advantage of MD-based energy calculations is that they provide statistically converged values that may be partitioned into contributions from individual residues in the protein and ligand(19).

Biolayer Interferometry (BLI)

Affinity measurements were performed on a biolayer interferometer (Octet Red96, ForteBio). Data were processed using the Data Acquisition and Analysis 8.0 software (ForteBio), and kinetic binding constants determined from a 1:1 binding model using the OriginPro software (OriginLab). The scFv was immobilized on an amine reactive second-generation (AR2G) biosensor (Lot No. 1311212, ForteBio). The BGA trisaccharide was analysed as the conjugate to bovine serum albumin (BSA-BGA) and was dissolved in an analysis buffer containing 10 mM HEPES, 150 mM NaCl , 3.4 mM EDTA , 0.005% Tween 20 at a range of pH values (5, 5.5, 6, 6.5, and 7). A BSA-Le^X trisaccharide conjugate (Prod. No. NGP0302, V-Labs, Inc.) and BSA (Prod. No. 23209, Pierce Thermo Scientific, Rockford, IL, USA) were used as negative controls. Details of the BLI conditions are provided in supplemental info (Supplemental Methods).

Results and Discussion

Involvement of His^{L49} in binding affinity

All Histidines in the scFv were protonated by default for modeling with a hydrogen atom at the δ nitrogen position. During the MD simulation of the BGA-scFv complex, the χ2 angle of His^{L49} flips (-73° to 115°) enabling a hydrogen bond to form with the carbonyl moiety of the NAc group in the GalNAc residue in BGA, which would be expected to be significant for enhancing the stability of the BGA-scFv complex. In the BGB complex, the same His^{L49} forms an interaction with the non-terminal Gal residue. The interaction with $HisL^{49}$ suggests that there might also be a pH dependence on binding; at lower pH all Histidines would be positively charged, potentially enhancing the strength of the His^{L49} - BGA hydrogen bond, leading to higher binding affinity. This prediction was confirmed by BLI measurements, which showed a marked decrease in the apparent K_D as the pH dropped below the pK_A of histidine (Figure 6.1). It should be noted that this protonation would not be localized to His^{L49} nevertheless, no enhanced non-specific binding was observed at low pH for either BSA or BGA-Le^x (Figures 7.2-6.4), supporting a role for a direct interaction between His^{L49} and the BGA antigen.

Conclusions

In this study, 3D models of the BGA and BGB trisaccharides in complex with scFv AC1001 were generated that provided a detailed atomic level rationalization of the interactions and dynamics responsible for antigen specificity. Quantification of the binding affinities identified key residues in the binding site that are predicted to contribute to specific and non-specific interactions with each antigen and led to the confirmed prediction of enhanced binding at lower pH. The spontaneous dissociation of antigen B from the scFv-BGB complexes (in three different simulations) indicated that MD simulations confirm the known preference of this antibody for the A antigen, and support a role for MD simulations in overcoming limitations associated with ligand docking. The present study illustrates that integration of multiple experimental (affinity measurements, glycan array screening, and crystallography) and theoretical (ligand docking, MD simulation, and energy decomposition) methods provides a powerful platform for predicting the origin of antibody-carbohydrate specificity

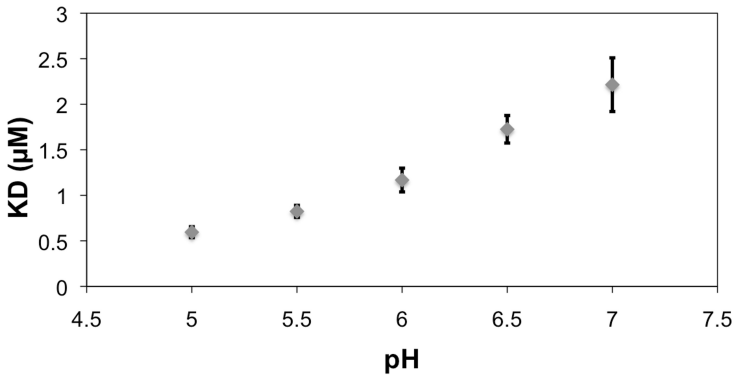


Figure 6.1. The reference (BSA)-subtracted pH dependence of the apparent K_D for the interaction between scFv AC1001 and the BSA-BGA conjugate. Error bars are derived from replicates of five measurements. Note, the pKa of Histidine is 6.04. (20)

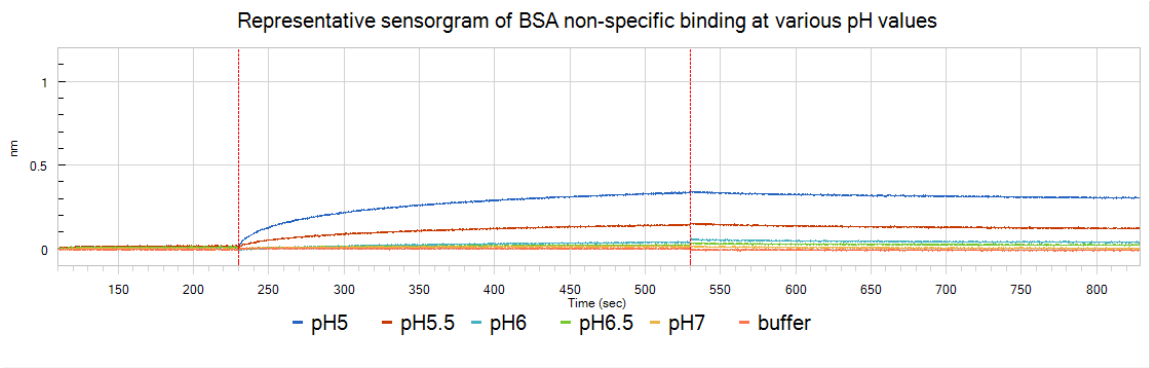


Figure 6.2. Binding assay of BSA at various pH values. scFv-immobilized biosensor was dipped into 1 μM BSA at pH5 (blue), pH5.5 (red), pH6 (cyan), pH6.5 (green), pH7 (yellow), and buffer (orange). BSA showed no binding to scFv-immobilized biosensor at pH 6.5, and 7, but a relative small non-specific binding to scFv-immobilized biosensor at acidic pH 5, 5.5, and 6. Analysis buffer (reference in orange) did not display any binding at all.

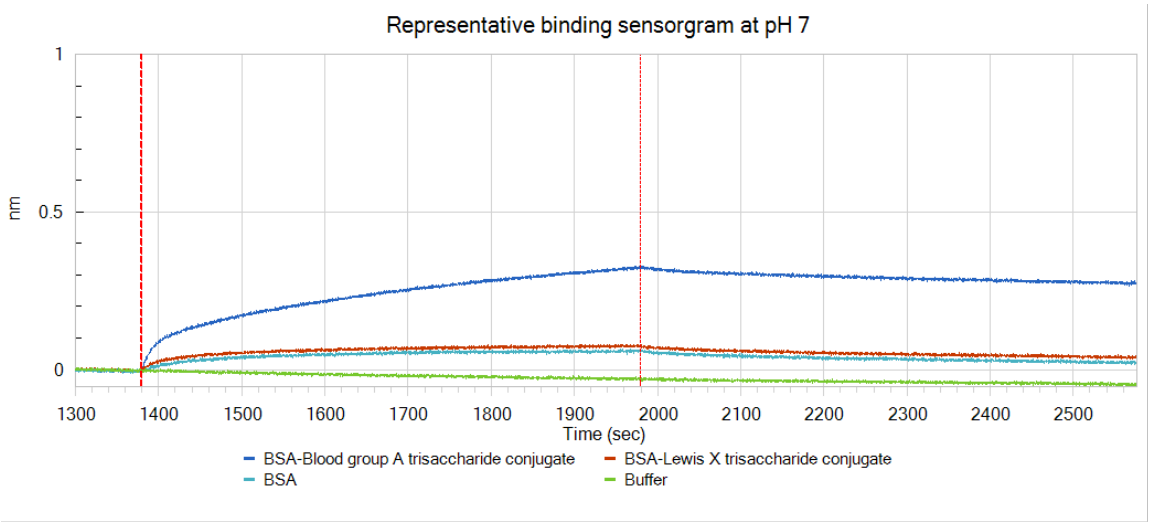


Figure 6.3. Representative binding sensorgram for BGA-conjugate (blue), Le^x-conjugate (red), BSA (cyan) and buffer (green) at pH 7. BSA-Le^x and BSA showed a similar signal.

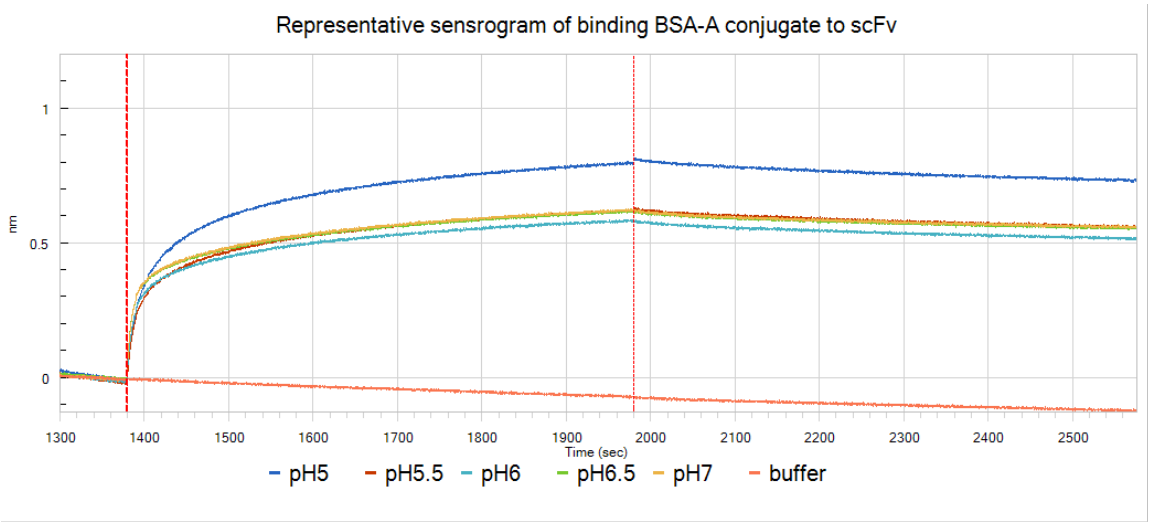


Figure 6.4. BLI binding assay of scFv to BSA-blood group A trisaccharide conjugate at pH5 (blue), pH5.5 (red), pH6 (cyan), pH6.5 (green), pH7 (yellow), and buffer (orange).

Reference:

1. Landsteiner K. Zur Kenntnis der antifermentativen, lytischen und agglutinierenden Wirkungen des Blutserums und der Lymphe. *Zentralblatt Bakteriologie* (1900) **27**(10):357-62.
2. Yamamoto F. Review: ABO blood group system--ABH oligosaccharide antigens, anti-A and anti-B, A and B glycosyltransferases, and ABO genes. *Immunohematology / American Red Cross* (2004) **20**(1):3-22. Epub 2004/09/18. PubMed PMID: 15373665.
3. Chester MA, Olsson ML. The ABO blood group gene: A locus of considerable genetic diversity. *Transfusion medicine reviews* (2001) **15**(3):177-200.
4. Marionneau S, Cailleau-Thomas A, Rocher J, Le Moullac-Vaidye B, Ruvoën N, Clément M, et al. ABH and Lewis histo-blood group antigens, a model for the meaning of oligosaccharide diversity in the face of a changing world. *Biochimie* (2001) **83**(7):565-73. doi: http://dx.doi.org/10.1016/S0300-9084(01)01321-9.
5. Williamson LM, Lowe S, Love EM, Cohen H, Soldan K, McClelland DBL, et al. Serious hazards of transfusion (SHOT) initiative: analysis of the first two annual reports. *BMJ* (1999) **319**(7201):16-9. doi: 10.1136/bmj.319.7201.16.
6. Sazama K. Reports of 355 transfusion-associated deaths: 1976 through 1985. *Transfusion* (1990) **30**(7):583-90. doi: 10.1046/j.1537-2995.1990.30790385515.x.
7. Ozolek JA, Watchko JF, Mimouni F. Prevalence and lack of clinical significance of blood group incompatibility in mothers with blood type A or B. *The Journal of pediatrics* (1994) **125**(1):87-91.
8. Le Pendu J, Marionneau S, Cailleau-Thomas A, Rocher J, Le Moullac-Vaidye B, Clement M. ABH and Lewis histo-blood group antigens in cancer. *APMIS : acta*

pathologica, microbiologica, et immunologica Scandinavica (2001) **109**(1):9-31. Epub 2001/04/12. PubMed PMID: 11297197.

9. Dabelsteen E, Gao S. ABO Blood-group antigens in oral cancer. *Journal of Dental Research* (2005) **84**(1):21-8. doi: 10.1177/154405910508400103.

10. Anstee DJ. The relationship between blood groups and disease. *Blood* (2010) **115**(23):4635-43. doi: 10.1182/blood-2010-01-261859.

11. Ceravolo IP, Sanchez BAM, Sousa TN, Guerra BM, Soares IS, Braga EM, et al. Naturally acquired inhibitory antibodies to Plasmodium vivax Duffy binding protein are short-lived and allele-specific following a single malaria infection. *Clinical & Experimental Immunology* (2009) **156**(3):502-10. doi: 10.1111/j.1365-2249.2009.03931.x.

12. DeMarco ML, Woods RJ. Structural glycobiology: A game of snakes and ladders. *Glycobiology* (2008) **18**(6):426-40. doi: 10.1093/glycob/cwn026.

13. Woods R, Yongye A. Computational techniques applied to defining carbohydrate antigenicity. In: Kosma P, Müller-Loennies S, editors. *Anticarbohydrate Antibodies*. Springer Vienna (2012). p. 361-83.

14. Paula S, Monson N, Ball WJ, Jr. Molecular modeling of cardiac glycoside binding by the human sequence monoclonal antibody 1B3. *Proteins* (2005) **60**(3):382-91. Epub 2005/06/23. doi: 10.1002/prot.20484. PubMed PMID: 15971203.

15. Woods RJ, Tessier MB. Computational glycoscience: characterizing the spatial and temporal properties of glycans and glycan-protein complexes. *Current opinion in structural biology* (2010) **20**(5):575-83. Epub 2010/08/17. doi:

10.1016/j.sbi.2010.07.005. PubMed PMID: 20708922; PubMed Central PMCID: PMC3936461.

16. Nivedha AK, Makeneni S, Foley BL, Tessier MB, Woods RJ. Importance of ligand conformational energies in carbohydrate docking: Sorting the wheat from the chaff. *Journal of Computational Chemistry* (2014) **35**(7):526-39. doi: 10.1002/jcc.23517.

17. Thomas R, Patenaude SI, MacKenzie CR, To R, Hirama T, Young NM, et al. Structure of an anti-blood group A Fv and improvement of its binding affinity without loss of specificity. *The Journal of biological chemistry* (2002) **277**(3):2059-64. Epub 2001/10/27. doi: 10.1074/jbc.M104364200. PubMed PMID: 11679577.

18. Trott O, Olson AJ. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry* (2010) **31**(2):455-61. doi: 10.1002/jcc.21334.

19. Hadden JAT, M. B.; Fadda, E.; Woods, R. J. Calculating binding free energies for protein-carbohydrate complexes. *Methods in Molecular Biology: Glycoinformatics*. NJ(In press).

20. Wood EJ. Data for biochemical research (third edition) by R M C Dawson, D C Elliott, W H Elliott and K M Jones. *Biochemical Education* (1987) **15**(2):97-. doi: 10.1016/0307-4412(87)90110-5.

CHAPTER 7

CONCLUSION

Carbohydrates are involved with multiple biological events; many of them are mediated by carbohydrate-protein interactions, such as viral infection, for instance. As mentioned at the beginning of this study, a variety of techniques and methods have been applied to this type of interaction. Despite the availability of these various implementations, many have failed due to low experimental sensitivity to generate accurate monomeric affinity measurements. This work introduced a competition assay using Biolayer Interferometry (BLI) to establish the monomeric solution K_D determination for weak carbohydrate-protein complexes. Chapter One explained this method in detail showing its good reproducibility when compared to other techniques such as NMR, ITC, and MST. Chapter Two not only surveyed conformational properties of HA-glycan complexes, but also summarized the rationale of this work on Influenza HA glycan interactions. Both HA from human infective and avian infective virus could bind to short, small glycans such as di-, tri-, and even penta-saccharides without displaying the canonical binding preferences. Chapter Three reported the monomeric affinity values of such interactions using the BLI competition assay. ECL and X-31 HA glycan interactions were tested separately in this work with two assay formats: where glycans were immobilized and where proteins were immobilized. Both formats generated highly accurate and reproducible solution K_D values, compared to NMR and ITC reports. Chapter Four detailed how computational analysis of ECL-glycan binding agreed with monomeric affinity resulted from the BLI competition assay.

Solution K_D values of multiple common glycans binding to human H1N1 A/California/04/2009 HA and avian H5N1 A/Vietnam/1194/2004 HA were both determined and discussed in Chapter Five. The monomeric affinity values of both 6' and 3'sialylated tri- and penta-saccharides were very close in the case of avian H5 HA, but human H1 HA showed 2-3 fold preferences to 6' sialylated glycans. This result was consistent with the binding energy calculation and computational studies mentioned in Chapter Two. The explanation for the lowly boos in affinity from presumed bidentate binding was explained as arising from an entropy penalty. The entropy penalty in the binding event could be observed through solution K_D value measurement of complex branched glycans according to this conclusion. As noted in Chapter Five, this conclusion had been observed in SPR experiments whereas human H1 HA binds tighter to the long-branched glycan structures. This observation should be examined and compared using the BLI competitive assay. With the advantage of this assay, such entropy penalties could be determined via experiment.

Other than monomeric affinity values, BLI competitive assay also provided a reliable and reproducible method to measure the IC_{50} values for six previously designed influenza inhibitors by our group. Among the six inhibitors, as well as natural glycans, the compound FB127 exhibited the 20 times better in inhibition to all H1, H3, H5, and H7 HA compared to its natural ligand in Chapter Five. Chapter Six also displayed the accurate measurement of anti-blood group antibody specificity via direct binding format.

In the future, the assay format will be reversed so that Fab-glycan will be immobilized on the biosensor surface at varied densities to test the hypothesis that longer biantennary 6'sialosides are capable of bidentate binding. In this assay format, they would continue to bind the HA after equivalent 3'sialosides lost measureable affinity. The entropy penalty of glycan-

HA binding should be computed from simulations and used to help explain the large apparent entropy penalty paid by the bidentate binding. This could be used to explain the bidentate binding of human biantennary glycans to human HA which exhibits a binding boost with entropy penalty with both calculated and determined values. Better understanding of glycan binding preferences would help create a better understanding of the specificity preferences of influenza, and should help advance therapeutic strategies that target these interactions.

APPENDIX A: THE SUPPLEMENTARY MATERIAL FOR CHAPTER 4

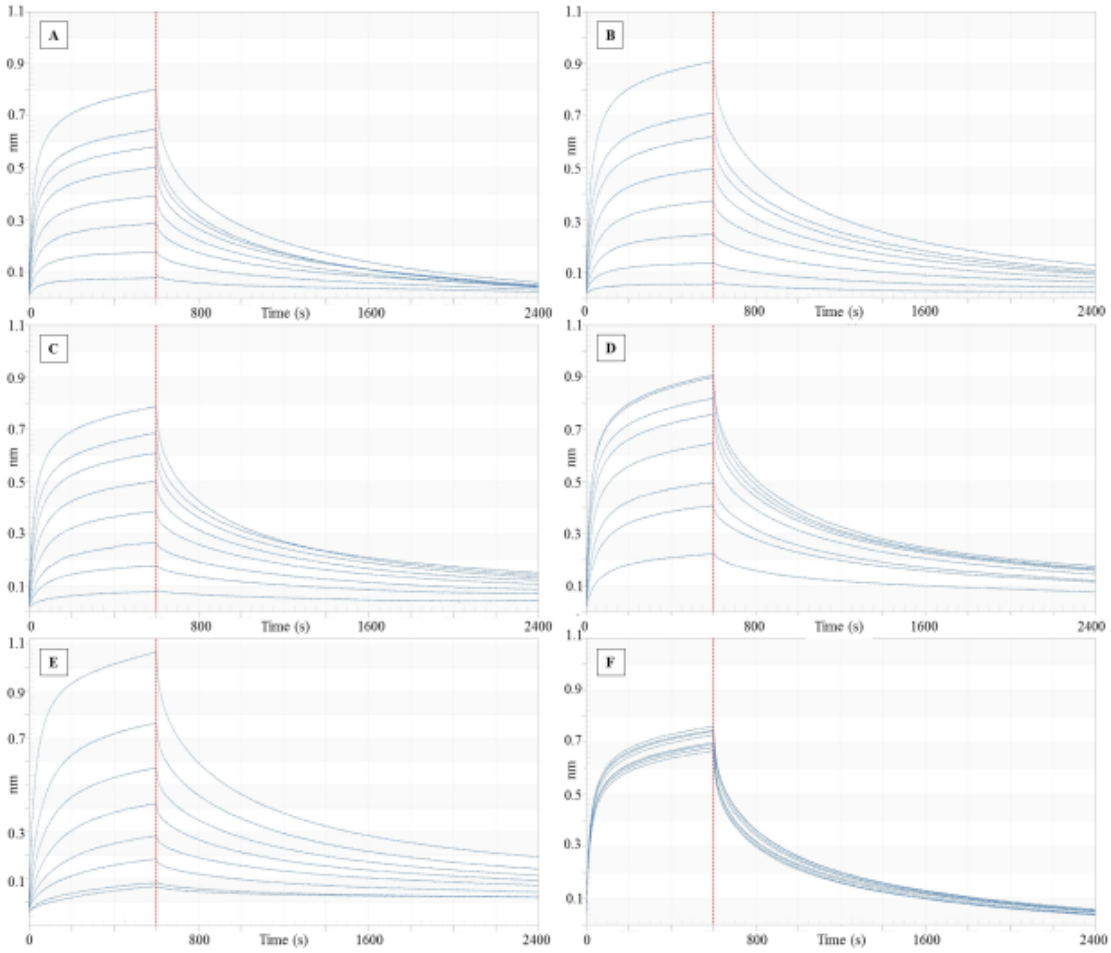


Figure S1. Inhibition of binding of ECL to **3** (immobilized on the BLI biosensor) for six ligands. A. Effect of **1** (concentration 0μM, 62.5μM, 0.125mM, 0.25mM, 0.5mM, 1mM, 2mM, 4mM) on binding. B. Effect of **2** (0μM, 62.5μM, 0.125mM, 0.25mM, 0.5mM, 1mM, 2mM, 4mM) on binding. C. Effect of **3** (0μM, 15.6μM, 31.3μM, 62.5μM, 0.125mM, 0.25mM, 0.5mM, 1mM) on binding. D. Effect of **5** (0μM, 15.6μM, 31.3μM, 62.5μM, 0.125mM, 0.25mM, 0.5mM, 1mM, 2mM) on binding. E. Effect of **6** (0μM, 15.6μM, 31.3μM, 62.5μM, 0.125mM, 0.25mM, 0.5mM, 1mM, 2mM) on binding. F. Effect of **7** (0μM, 0.125mM, 0.25mM, 0.5mM, 1mM, 2mM, 4mM, 8mM) on binding, indicating that **7** did not exhibit any inhibition or dose-dependence.

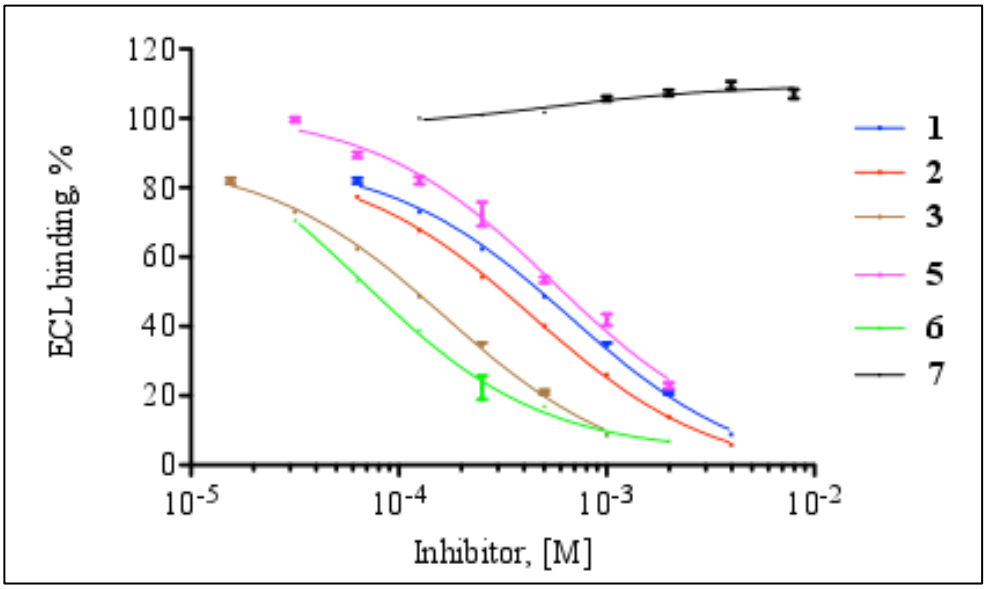


Figure S2. Inhibition curves used to compute IC_{50} values for **1**-**7** inhibiting the binding of ECL to immobilized **3**.

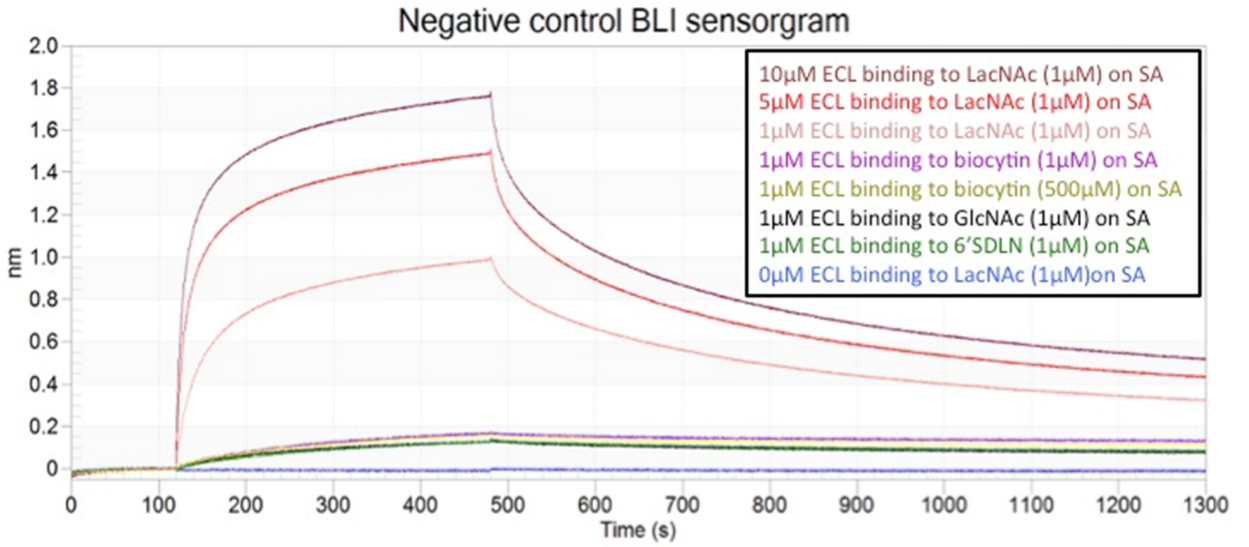


Figure S3. BLI control sensorgrams for ECL binding to positive (LacNAc) and negative controls (GlcNAc, 6'SDLN, and biocytin) surfaces.

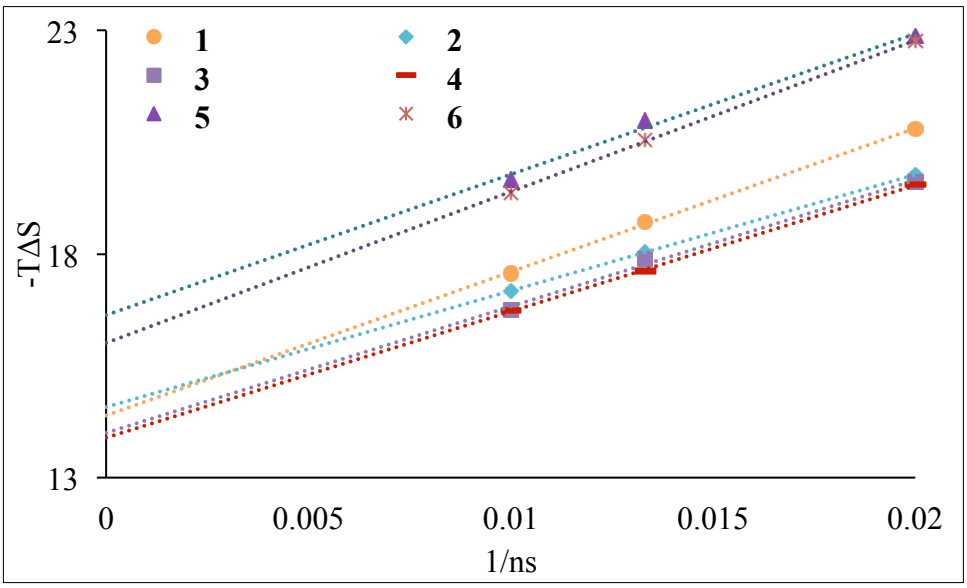


Figure S4. Extrapolation of quasi-harmonic entropy to infinite time for all the ligands. The coefficient of determination (R^2) in all the cases is greater than 0.99.

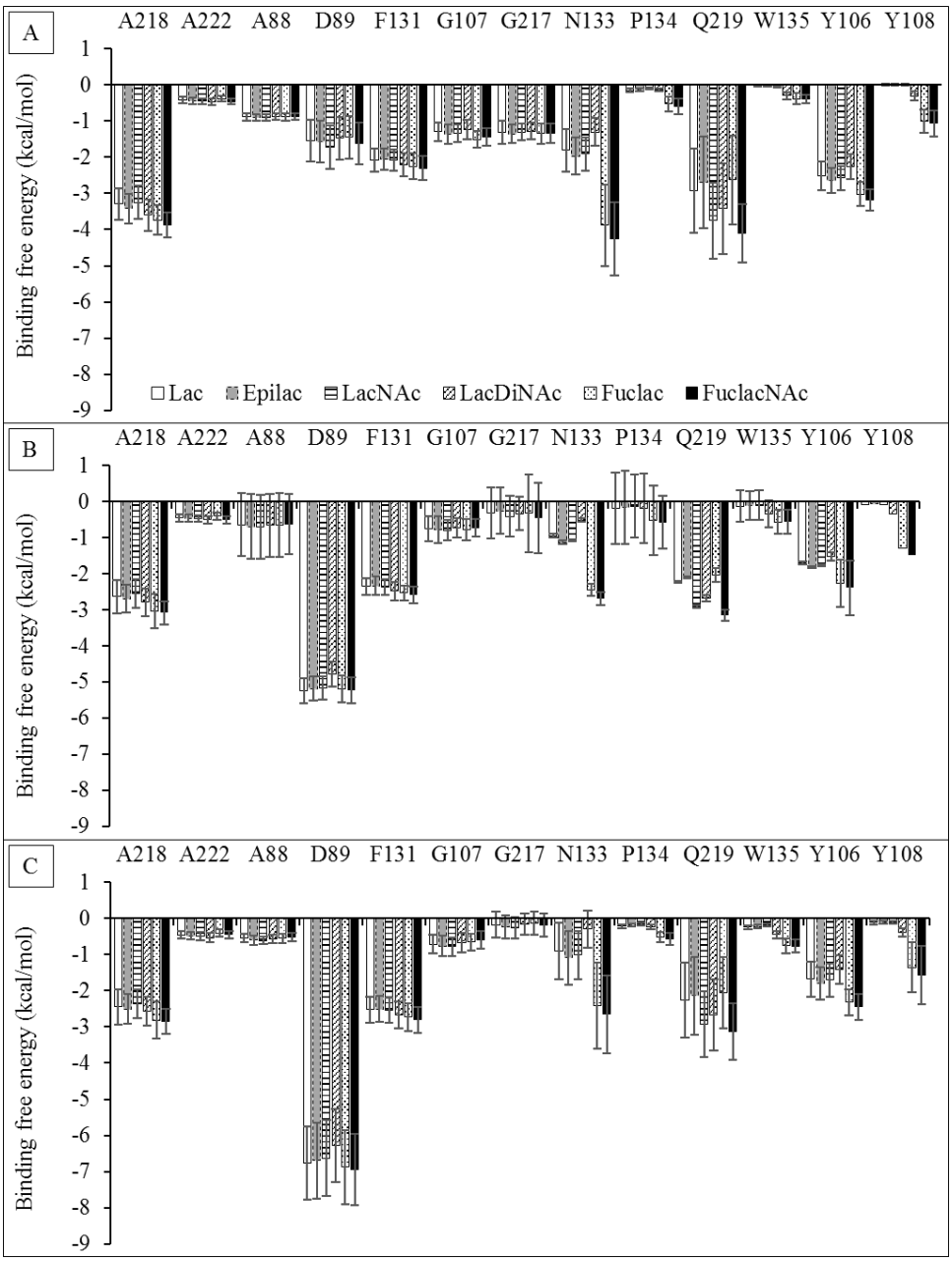


Figure S5. The binding free energy contributions of amino acids making significant interactions with the ligand. The calculations were performed using three different desolvation models. **A**. GB^{HCT}, **B**. $GB_1{}^{OBC}$ and **C**. $GB_2{}^{OBC}$.

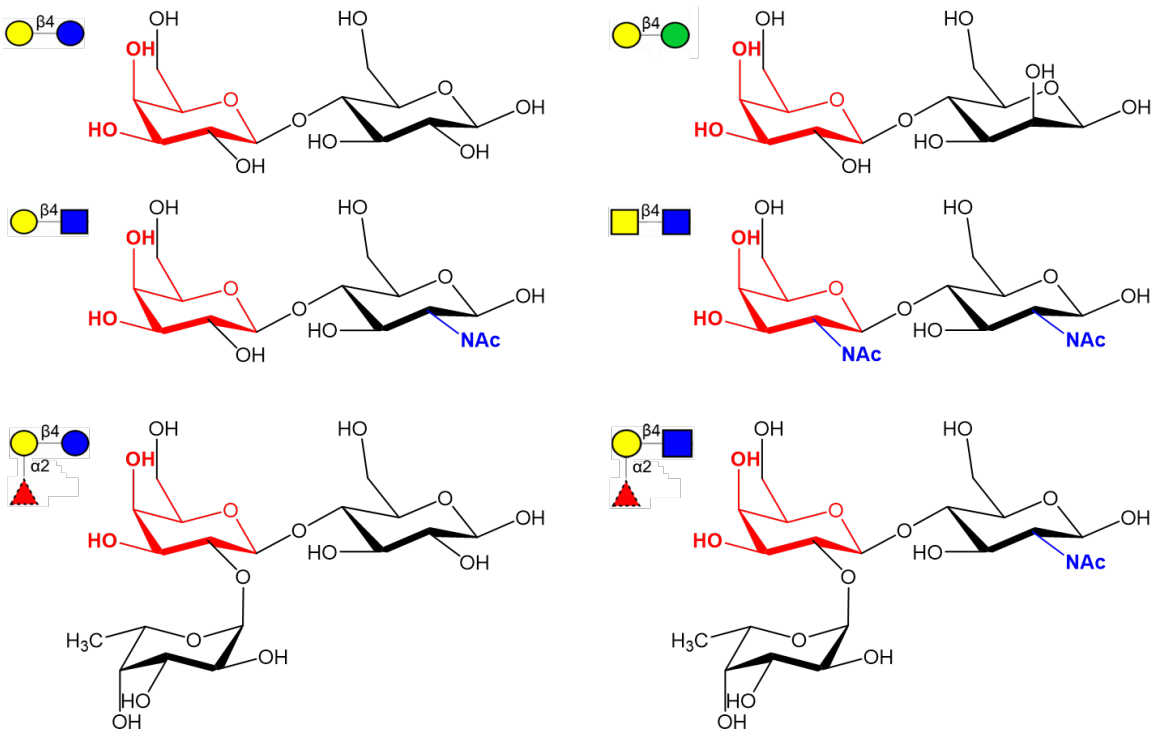


Figure S6. The pharmacophore (red) required for binding to ECL is defined by the spatial orientation of the O3 and O4 hydroxyl groups in the Gal/GalNAc residue along with the atoms forming the ring structure; indicated for six different ligands: Lac (top left), Epilac (top right), LacNAc (middle left), LacDiNAc (middle right), Fuclac (bottom left) and FuclacNAc (bottom right). The NAc groups (blue) enhance the affinity, but are not required for binding. The monosaccharides are represented in SNFG notation as Gal: yellow circle, Glc: blue circle, Man: green circle, GlcNAc: blue square, GalNAc: yellow square and Fuc: red triangle.

Table S1. IC_{50} values for oligosaccharides.

Ligand	IC_{50}[a]
1	0.66 (0.04)[b]
2	0.44 (0.01)
3	0.17 (0.01)
5	0.49 (0.05)
6	0.07 (0.01)
7	No Binding

[a]In mM.

[b]Error are

Table S2. Average dihedral angles for the glycosidic linkages from the PDB structures and MD simulations of the six complexes (**1** to **6**) [a.]

		Galβ1-4Glcβ		Fucα1-2Galβ	
	Ligand	ϕ	ψ	ϕ	ψ
MD	**1**	-80.5 (22)	-134.0 (19)		
	2	-78.1 (21)	-133.2 (19)		
	3	-78.4 (15)	-131.1 (12)		
	4	-75.1 (7)	-134.0 (10)		
	5	-75.6 (7)	-135.9 (20)	-62.4 (7)	-86.5 (8)
	6	-74.9 (6)	-133.5 (8)	-63.4 (7)	-87.4 (8)
Exptl.	Average[b]	-73.7 (23)	-113.4 (41)	-78.5 (21)	-104.9 (17)

[a]Data in parentheses is the standard deviation.

[b]Average of dihedral angles for the glycosidic linkages from all the structures in the PDB database containing that linkage obtained using glytorsion.

Table S3. Experimental and theoretical hydrogen bonds and pairwise hydrogen bond interaction energies.

Ligand	Protein residue	Ligand residue	Distance[a]		Occupancy (%)	Interaction energy[b]		
			Exptl.	MD		GB^{HCT} (igb = 1)	GB_1^{OBC} (igb = 2)	GB_2^{OBC} (igb = 5)
1	D89-Oδ2	Gal-O4	2.6	2.6	100	-1.0	-3.7	-3.6
2		Gal-O3	2.6	2.8	98	**0.1**[c]	-3.2	-3.1
3		Gal-O3	2.6	2.6	99	**0.1**	-3.9	-3.0
5		Gal-O4	2.6	2.6	99	**0.2**	-3.5	-3.4
1	D89-Oδ1	Gal-O3	2.7	2.7	98	**0.0**	-3.3	-3.3
2		Gal-O4	2.7	2.6	100	-1.1	-3.8	-3.7
3		Gal-O4	2.6	2.7	100	-1.2	-3.1	-3.8
5		Gal-O3	2.7	2.7	100	-0.8	-2.8	-2.6
1	N133-Nδ2	Gal-O3	2.9	2.9	81	-2.5	-2.6	-3.0
2		Gal-O3	4.0	3.0	90	-2.7	-2.8	-3.2
3		Gal-O3	3.1	3.0	94	-2.6	-2.8	-3.1
5		Gal-O3	2.9	3.0	96	-3.0	-3.4	-3.8
1	A218-N	Gal-O4	3.1	3.0	99	-2.0	-2.0	-2.2
2		Gal-O4	3.2	3.1	99	-1.9	-2.0	-2.2
3		Gal-O4	3.1	3.2 (0.2)	97	-1.8	-1.8	-2.0
5		Gal-O4	3.1	3.0	99	-2.1	-2.1	-2.4
1	G107-N	Gal-O3	3.0	3.0	28	-1.6	-1.5	-1.7
2		Gal-O3	3.0	3.0	29	-1.7	-1.5	-1.8
3		Gal-O3	2.9	3.0	39	-1.7	-1.6	-1.7
5		Gal-O3	3.0	3.0	41	-1.8	-1.7	-1.9
1	D219-Nε2	Glc-O3	3.1	4.0 (1.1)	21	-1.3	-1.3	-1.4
2		ManO3	3.0	3.9 (0.9)	33	-1.5	-1.5	-1.6
3		GlcNAc-O3	2.9	3.4 (0.8)	71	-2.7	-2.7	-3.0
5		Glc-O3	3.1	4.0 (1.0)	34	-1.5	-1.5	-1.6
5	N133-Nδ2	Fuc-O2	2.7	3.1 (0.3)	64	-3.6	-3.9	-4.6
5	Y108-OH	Fuc-O4	3.0	4.4 (0.8)	14	-0.7	-0.8	-0.9

[a]In Å, with standard deviations greater than 0.1 shown in parentheses.

[b]In kcal/mol, standard error of mean is less than 0.01 in all cases.

[c]Numbers in bold represent residues with structurally inconsistent values.

Table S4. Theoretical hydrogen bonds and pairwise hydrogen bond interaction energies in modeled complexes.

Ligand	Protein residue	Ligand residue	Distance[a] MD	Occupancy (%)	Interaction energy[b] GB^{HCT} (igb = 1)	$GB_1{}^{OBC}$ (igb = 2)	$GB_2{}^{OBC}$ (igb = 5)
4	D89-Oδ2	GalNAc-O3	2.6	99	**0.0**[c]	-3.0	-2.7
6		Gal-O4	2.6	100	-0.9	-3.5	-3.4
4	D89-Oδ1	GalNAc-O4	2.7	100	-0.8	-3.4	-3.3
6		Gal-O3	2.7	100	**0.2**	-2.7	-2.4
4	N133-Nδ2	GalNAc-O3	3.0	97	-2.7	-2.8	-2.4
6		Gal-O3	3.0	97	-3.0	-3.4	-3.9
4	A218-N	GalNAc-O4	3.0	100	-2.2	-2.2	-1.9
6		Gal-O4	3.1	100	-2.0	-2.0	-2.3
4	G107-N	GalNAc-O3	3.1 (0.2)	13	-1.4	-1.3	-1.2
6		Gal-O3	3.0	43	-1.8	-1.7	-2.0
4	D219-Nε2	GlcNAc-O3	3.9 (1.4)	60	-2.3	-2.3	-2.1
6		GlcNAc-O3	3.2 (0.6)	84	-3.1	-3.0	-3.3
6	N133-Nδ2	Fuc-O2	3.0 (0.2)	77	-4.2	-4.5	-5.3
6	Y108-OH	Fuc-O4	4.0 (0.8)	28	-0.9	-1.2	-1.3

[a]In Å, with standard deviations greater than 0.1 shown in parentheses.

[b]In kcal/mol, standard error of mean is less than 0.01 in all cases.

[c]Numbers in bold represent residues with structurally inconsistent values.

Table S5. Entropy contributions (-TΔS) in kcal/mol at 300 K[a].

Ligand	VRT Quasiharmonic[b] (QH)	VRT Normal Mode[c] (NM)	Ligand Conformational Karplus–Kushick[d]
1	14.38 (0.01)	19.0 (0.87)	0.34
2	14.60 (0.02)	20.5 (1.00)	0.03
3	13.98 (0.02)	19.4 (0.95)	0.89
4	13.88 (0.01)	22.3 (0.98)	0.50
5	16.64 (0.01)	26.0 (1.04)	1.08
6	16.02 (0.01)	26.4 (0.96)	1.32

[a]Data in parentheses is the standard error of mean.

[b]Employing100,000 frames interpolated to infinite sampling.

[c]Employing 100 frames.

[d]Employing 100,000 frames.

Table S6. Experimental (BLI) and theoretical (not including entropy corrections) binding free energies[a].

Ligand	Exptl.	GB^{HCT} (igb = 1)	GB_1^{OBC} (igb = 2)	GB_2^{OBC} (igb = 5)	GB*n1* (igb = 7)	GB*n2* (igb = 8)	PBSA
Dielectric Constant (ϵ) = 1							
1	-4.8	-27.2	-30.2	-33.1	-35.7	-26.4	-13.1
2	-5.1	-28.3	-30.7	-33.6	-36.7	-26.5	-13.6
3	-5.7	-30.7	-32.7	-35.5	-37.3	-29.1	-19.0
4		-31.5	-32.2	-34.5	-36.7	-28.1	-14.5
5	-5.0	-36.5	-37.7	-41.3	-43.2	-33.7	-11.5
6	-6.2	-42.3	-41.6	-45.0	-44.4	-38.4	-21.4
	r[b]	0.74	0.7	0.68	0.57	0.74	0.95
ϵ = 4							
1	-4.8	-25.0	-25.7	-26.4	-27.0	-24.8	-13.4
2	-5.1	-25.9	-26.6	-27.3	-28.1	-25.7	-14.0
3	-5.7	-28.0	-28.5	-29.1	-29.6	-27.6	-19.3
4		-30.7	-30.9	-31.4	-31.9	-29.9	-15.3
5	-5.0	-35.8	-36.1	-36.9	-37.4	-35.1	-12.3
6	-6.2	-38.9	-38.7	-39.5	-39.4	-37.9	-21.9
	r	0.63	0.62	0.61	0.59	0.62	0.96

[a]In kcal/mol. All experimental errors less than 0.1 kcal/mol and all theoretical standard error of mean values less than 0.1 kcal/mol. [b]Pearson correlation coefficient.

Table S7. Experimental (BLI) and theoretical (employing QH entropies) binding free energies[a]

Ligand	Exptl.	GB^{HCT} (igb = 1)	$GB_1{}^{OBC}$ (igb = 2)	$GB_2{}^{OBC}$ (igb = 5)	GBn_1 (igb = 7)	GBn_2 (igb = 8)	PBSA
1	-4.8	-12.8	-15.8	-18.7	-21.4	-12.0	1.3
2	-5.1	-13.7	-16.1	-19.0	-22.1	-11.9	1.0
3	-5.7	-16.8	-18.7	-21.5	-23.4	-15.1	-5.0
4		-17.6	-18.4	-20.6	-22.8	-14.2	-0.6
5	-5.0	-19.8	-21.1	-24.6	-26.5	-17.0	5.1
6	-6.2	-26.3	-25.5	-28.9	-28.3	-22.4	-5.4
	r[b]	0.82	0.80	0.77	0.69	0.83	0.71

[a]In kcal/mol. All experimental errors less than 0.1 kcal/mol and all theoretical standard error of mean values less than 0.1 kcal/mol.

[b]Pearson correlation coefficient

Table S8. Experimental (BLI) and theoretical (employing QH and conformational entropies) binding free energies[a]

Ligand	Exptl.	GB^{HCT} (igb = 1)	$GB_1{}^{OBC}$ (igb = 2)	$GB_2{}^{OBC}$ (igb = 5)	GBn_1 (igb = 7)	GBn_2 (igb = 8)	PBSA
1	-4.8	-12.5	-15.5	-18.4	-21.0	-11.7	1.6
2	-5.1	-13.6	-16.0	-19.0	-22.0	-11.9	1.0
3	-5.7	-15.9	-17.8	-20.6	-22.5	-14.2	-4.1
4		-17.1	-17.9	-20.1	-22.3	-13.7	-0.1
5	-5.0	-18.7	-20.0	-23.6	-25.4	-16.0	6.2
6	-6.2	-24.9	-24.2	-27.6	-27.0	-21.1	-1.2
	r[b]	0.82	0.81	0.77	0.67	0.84	0.63

[a]In kcal/mol. All experimental errors less than 0.1 kcal/mol and all theoretical standard error of mean values less than 0.1 kcal/mol.

[b]Pearson correlation coefficient

Table S9. Experimental (BLI) and theoretical (employing NM entropies) binding free energies[a]

Ligand	Exptl.	GB^{HCT} (igb = 1)	$GB_1{}^{OBC}$ (igb = 2)	$GB_2{}^{OBC}$ (igb = 5)	GBn_1 (igb = 7)	GBn_2 (igb = 8)	PBSA
1	-4.8	-8.2(0.9)	-11.2(0.9)	-14.1(0.9)	-16.7(0.9)	-7.4(0.9)	5.9(0.9)
2	-5.1	-7.8(1.0)	-10.1(1.0)	-13.1(1.0)	-16.2(1.0)	-6.1(1.0)	6.9(1.0)
3	-5.7	-11.3(1.0)	-13.3(1.0)	-16.1(1.0)	-17.9(1.0)	-9.7(1.0)	0.4(1.0)
4		-11.0(1.0)	-11.8(1.0)	-14.0(1.0)	-16.2(1.0)	-7.6(1.0)	7.8(1.0)
5	-5.0	-10.4(1.0)	-11.7(1.0)	-15.3(1.0)	-17.1(1.0)	-7.7(1.0)	14.5(1.0)
6	-6.2	-15.8(1.2)	-15.1(1.2)	-18.5(1.2)	-17.9(1.2)	-12.0(1.2)	4.9(1.2)
	r[b]	0.93	0.93	0.91	0.83	0.93	0.48

[a]In cal/mol. All experimental errors less than 0.1 kcal/mol. Theoretical standard error of mean shown in parentheses.

[b]Pearson correlation coefficient.

Table S10. Experimental (BLI) and theoretical (employing NM and conformational entropies) binding free energies[a]

Ligand	Exptl.	GB^{HCT} (1gb = 1)	$GB_1{}^{OBC}$ (1gb = 2)	$GB_2{}^{OBC}$ (igb = 5)	GBn_1 (igb = 7)	GBn_2 (igb = 8)	PBSA
1	-4.8	-7.8(0.9)	-10.8(0.9)	-13.7(0.9)	-16.4(0.9)	-7.1(0.9)	6.3(0.9)
2	-5.1	-7.7(1.0)	-10.2(1.0)	-13.1(1.0)	-16.2(1.0)	-6.0(1.0)	6.9(1.0)
3	-5.7	-10.4(1.0)	-12.4(1.0)	-15.2(1.0)	-17.1(1.0)	-8.8(1.0)	1.3(1.0)
4		-10.5(1.0)	-11.3(1.0)	-13.5(1.0)	-15.7(1.0)	-7.1(1.0)	8.3(1.0)
5	-5.0	-9.4(1.0)	-10.6(1.0)	-14.2(1.0)	-16.1(1.0)	-6.6(1.0)	15.6(1.0)
6	-6.2	-14.6(1.2)	-13.9(1.2)	-17.3(1.2)	-16.7(1.2)	-10.8(1.2)	6.3(1.2)
	r[b]	0.95	0.95	0.94	0.68	0.94	0.40

[a]In kcal/mol. All experimental errors less than 0.1 kcal/mol. Theoretical standard error of mean shown in parentheses.

[b]Pearson correlation coefficient.

Table S11. The impact of polar desolvation free energy on per-residue MM-PB/GBSA energies with dielectric constant of 4[a].

	GB^{HCT} (igb = 1)	$GB_1{}^{OBC}$ (igb = 2)	$GB_2{}^{OBC}$ (igb = 5)	GBn_1 (igb = 7)	GBn_2 (igb = 8)	PBSA
Residues forming hydrogen bonds with the ligand						
A218	-2.50	-2.35	-2.30	-2.03	-2.43	-2.43
D89	**0.25**[b]	-0.64	-1.01	-1.24	**1.57**	**0.20**
G107	-0.51	-0.38	-0.37	-0.26	-0.51	-0.51
N133	-0.75	-0.54	-0.53	-0.33	-0.45	-0.70
Q219	-2.18	-2.01	-2.02	-1.85	-2.02	-2.12
Residues involved in other interactions with the ligand						
A88	-0.78	-0.72	-0.70	-0.68	-0.84	-0.79
A222	-0.49	-0.50	-0.50	-0.55	-0.49	-0.49
F131	-2.14	-2.21	-2.25	-2.25	-2.18	-2.14
G217	-0.56	-0.32	-0.28	-0.08	-0.36	-0.53
P134	-0.16	-0.17	-0.18	-0.16	-0.17	-0.16
W135	-0.11	-0.14	-0.16	-0.21	-0.13	-0.12
Y106	-1.97	-1.77	-1.77	-1.79	-1.90	-1.95
Y108	-0.13	-0.15	-0.16	-0.17	-0.16	-0.13

[a]In kcal/mol, standard error of mean is less than 0.01 in all cases.

[b]Numbers in bold represent residues with structurally inconsistent values.

APPENDIX B: THE SUPPLEMENTARY MATERIAL FOR CHAPTER 5

Table 1. Surface KD values of 6'SLN-Fab to immobilized H1, H3, and H5 HA from Octet BLI analysis software in which only produced two scientific position, replicates of three. Errors were standard deviation of the mean.

		$K_{D, surface}$ μM		AVG	STDEV	ERR
VN1194	2.00	2.10	2.10	2.07	0.06	0.02
CA04	0.77	0.75	0.83	0.78	0.04	0.01
X-31	0.96	1.20	1.30	1.15	0.17	0.06

Table 2. Surface KD values of 3'SLN-Fab to immobilized H1, H3, and H5 HA from Octet BLI analysis software in which only produced two scientific position, replicates of three. Errors were standard deviation of the mean.

		$K_{D, surface}$ μM		AVG	STDEV	ERR
VN1194	0.93	0.97	1.00	0.97	0.04	0.01
CA04	1.50	1.70	1.90	1.70	0.20	0.07
X-31	1.30	1.70	2.00	1.67	0.35	0.12

Table 3. IC_{50}, $K_{D, surface}$, and $K_{D, solution}$ values for 6'SLN purchased from Dextra to three HAs: H1, H3, and H5, replicates of three. Errors were standard deviation of the mean.

HA	IC_{50}, mM			$K_{D, surface}$	$K_{D, solution}$ mM		
strains	AVG	STDEV	ERR	μM	AVG	STDEV	ERR
VN1194	2.669	0.048	0.016	2.07	2.149	0.038	0.013
CA04	2.582	0.169	0.056	0.78	1.576	0.103	0.034
X-31	3.842	0.295	0.098	1.15	2.678	0.206	0.069

Table 4. IC_{50}, $K_{D, surface}$, and $K_{D, solution}$ values for 3'SLN purchased from Dextra to three HAs: H1, H3, and H5, replicates of three. Errors were standard deviation of the mean.

HA	IC_{50}, mM			$K_{D, surface}$	$K_{D, solution}$ mM		
strains	AVG	STDEV	ERR	μM	AVG	STDEV	ERR
VN1194	3.711	0.416	0.139	0.97	2.446	0.274	0.091
CA04	6.287	0.519	0.173	1.70	4.858	0.401	0.134
X-31	3.554	0.013	0.004	1.67	2.735	0.010	0.003

Table 5. IC_{50}, $K_{D, surface}$, and $K_{D, solution}$ values for 6'SLN-N3 requested from CFG to several HAs, replicates of three. Errors were standard deviation of the mean.

	IC_{50}, mM			$K_{D, surface}$	$K_{D, solution}$ mM		
HA strains	AVG	STDEV	ERR	μM	AVG	STDEV	ERR
Cal04 D190E/D225G /I219A/E227A	1.601	0.099	0.033	1.01	1.071	0.066	0.022
X-31	3.114	0.115	0.038	1.15	2.172	0.080	0.027
Victoria/2011	2.293	0.053	0.018	1.40	1.690	0.039	0.013
VN1194	3.166	0.033	0.011	1.77	2.468	0.026	0.009
CA04	2.193	0.022	0.007	1.20	1.548	0.016	0.005

Table 6. IC_{50}, $K_{D, surface}$, and $K_{D, solution}$ values for 3'SLN-N3 requested from CFG to several HAs, replicates of three. Errors were standard deviation of the mean.

	IC_{50}, mM			$K_{D, surface}$	$K_{D, solution}$ mM		
HA strains	AVG	STDEV	ERR	μM	AVG	STDEV	ERR
Cal04							
D190E/D225G/I219A/E227A	1.126	0.228	0.076	1.50	0.844	0.171	0.057
X-31	3.057	0.473	0.158	1.67	2.352	0.364	0.121
Victoria/2011	1.910	0.255	0.085	2.07	1.538	0.206	0.069
VN1194	2.843	0.425	0.142	1.00	1.895	0.283	0.094
CA04	2.970	0.531	0.177	1.60	2.263	0.405	0.135

Table 7. IC_{50}, $K_{D, surface}$, and $K_{D, solution}$ values for 6'SDLN-N3 requested from CFG to several HAs, replicates of three. Errors were standard deviation of the mean.

	IC_{50}, mM			$K_{D, surface}$	$K_{D, solution}$ mM		
HA strains	AVG	STDEV	ERR	μM	AVG	STDEV	ERR
VN1194	3.581	0.218	0.073	0.870	2.274	0.138	0.046
CA04	2.474	0.148	0.049	0.780	1.508	0.111	0.037

Table 8. IC_{50}, $K_{D, surface}$, and $K_{D, solution}$ values for 3'SDLN-N3 requested from CFG to several HAs, replicates of three. Errors were standard deviation of the mean.

	IC_{50}, mM			$K_{D, surface}$	$K_{D, solution}$ mM		
HA strains	AVG	STDEV	ERR	μM	AVG	STDEV	ERR
VN1194	3.198	0.086	0.029	1.000	2.132	0.070	0.023
CA04	7.571	0.005	0.002	0.700	4.416	0.003	0.001

Table 9. Surface direct binding affinity $K_{D, surface}$, replicates of three. Errors were standard deviation of the mean.

Fab-glycan	CA04 H1 HA		VN1194 H5 HA	
$K_{D, surface}$ μM	Monomeric	Trimeric	Monomeric	Trimeric
6'SLN-Fab	1.2 ± 0.094	1.2 ± 0.12	2.4 ± 0.21	2.5 ± 0.23
3'SLN-Fab	3.4 ± 0.46	2.1 ± 0.26	0.86 ± 0.11	0.96 ± 0.14
6'SDLN-Fab	0.85 ± 0.11	0.78 ± 0.095	0.9 ± 0.12	0.87 ± 0.092
3'SDLN-Fab	1.2 ± 0.22	0.70 ± 0.087	0.64 ± 0.075	0.66 ± 0.079

Table 10. IC_{50} values of list of glycans inhibit 6'SLN-Fab binding to immobilized CA04 H1 HA, replicates of three. Errors were standard deviation of the mean.

	Monomeric IC_{50}		Trimeric IC_{50}	
Compound	AVG (mM)	STDEV	AVG (mM)	STDEV
6'SDLN-N3	0.286	0.021	0.148	0.001
3'SDLN-N3	0.481	0.013	0.393	0.017
6'SLN-N3	1.491	0.120	2.107	0.176
3'SLN-N3	1.132	0.081	1.404	0.132
6'SG	3.931	0.386	4.505	0.245
3'SG	6.792	0.280	12.590	0.071
145	2.240	0.060	3.054	0.092
122 R-Bz	2.022	0.040	2.080	0.031
127 S-iP	0.228	0.046	0.281	0.083
142 R-OH	0.620	0.044	0.630	0.029
146 S-Bz 4,6Bz	1.382	0.065	0.983	0.049
143 S-OH	1.054	0.072	1.003	0.052

Table 11. IC_{50} values of list of glycans inhibit 3'SLN-Fab binding to immobilized VN1194 H5 HA, replicates of three. Errors were standard deviation of the mean.

	Monomeric IC_{50}		Trimeric IC_{50}	
Compound	AVG (mM)	STDEV	AVG (mM)	STDEV
6'SDLN-N3	1.316	0.267	1.409	0.107
3'SDLN-N3	1.294	0.250	1.040	0.376
6'SLN-N3	1.865	0.056	1.711	0.041
3'SLN-N3	3.900	0.291	3.781	0.344
6'SG	9.163	0.489	10.470	0.594
3'SG	4.461	0.036	4.511	0.012
145	4.717	0.186	4.814	0.177
122 R-Bz	1.468	0.232	2.253	0.393
127 S-iP	0.235	0.020	0.401	0.024
142 R-OH	1.203	0.113	1.032	0.044
146 S-Bz 4,6Bz	ND[a]	ND	ND	ND
143 S-OH	4.923	0.285	3.625	0.094

[a] ND. Note Detected.

Table 12. K_i values of list of glycans binding to immobilized CA04 H1 HA, replicates of three. Errors were standard deviation of the mean.

	Monomeric K_i		Trimeric K_i	
Compound	AVG (mM)	STDEV	AVG (mM)	STDEV
6'SDLN-N3	0.202	0.015	0.105	0.000
3'SDLN-N3	0.340	0.009	0.277	0.012
6'SLN-N3	1.053	0.084	1.487	0.124
3'SLN-N3	0.799	0.057	0.991	0.093
6'SG	2.775	0.273	3.180	0.173
3'SG	4.794	0.198	8.887	0.050
145	1.581	0.042	2.156	0.065
122 R-Bz	1.427	0.028	1.468	0.022
127 S-iP	0.161	0.033	0.199	0.058
142 R-OH	0.437	0.031	0.445	0.020
146 S-Bz 4,6Bz	0.976	0.046	0.694	0.034
143 S-OH	0.744	0.051	0.708	0.037

Table 13. K_i values of list of glycans binding to immobilized VN1194 H5 HA, replicates of three. Errors were standard deviation of the mean.

	Monomeric K_i		Trimeric K_i	
Compound	AVG (mM)	STDEV	AVG (mM)	STDEV
6'SDLN-N3	0.832	0.169	0.926	0.070
3'SDLN-N3	0.818	0.158	0.684	0.247
6'SLN-N3	1.180	0.035	1.125	0.027
3'SLN-N3	2.466	0.184	2.486	0.226
6'SG	5.794	0.309	6.884	0.391
3'SG	2.821	0.022	2.966	0.008
145	2.983	0.118	3.166	0.116
122 R-Bz	0.928	0.147	1.481	0.259
127 S-iP	0.149	0.012	0.264	0.016
142 R-OH	0.761	0.071	0.678	0.029
146 S-Bz 4,6Bz	ND[a]	ND	ND	ND
143 S-OH	3.113	0.180	2.384	0.062

[a] ND. Note Detected.

Table 14. IC50 values of natural glycan 145 and inhibitor FB127 to H1, H3, H5, and H7 HA, replicates of three.

	H1N1		H5N1		H3N2		H7N9	
$IC5_0$, mM	A/California/04/ 2009		A/Vietnam/1194/ 2004		A/Brisbane/10/ 2007		A/Anhui/1/2013	
	AVG	STDEV	AVG	STDEV	AVG	STDEV	AVG	STDEV
145	3.054	0.092	5.412	0.108	4.451	0.153	4.278	0.143
127 S-iP	0.281	0.083	0.283	0.012	0.216	0.008	0.214	0.007

APPENDIX C: THE COMPLETE PUBLICATION FOR CHAPTER 6

Spandana Makeneni, Ye Ji, David C. Watson, N. Martin Young, and Robert J. Woods

Frontier in Immunology 2014, 5:1-9. Reprint here with permission of publisher.

Author's words:

The BLI experimental data and results were equally important as the computational study in this publication. This work was collaboration between three institutes. The BLI evaluation of the scFv binding to BSA-conjugates was critical in order to support the computational study results. Without either BLI experimental evaluation or the computational analysis, this work would not be complete and published. Computational analysis examined the binding site, H-bonds, and rotational bonds contributed to the binding of scFv to blood group carbohydrates. The BLI work in this study applied a biophysical method to determine the binding as well as the pH values that affected such interaction in real-time. When experimental results agreed with theoretical prediction that turned the study into the publication.

Both major professor and committee members strongly suggested displaying the whole publication in the chapter instead of only a part of this work. However graduate school does not allow any second author paper to be appearing in the dissertation, this work now is attached as appendix C.

Abstract

The ABO blood group system is the most important blood type system in human transfusion medicine. Here, we explore the specificity of antibody recognition towards ABO blood group antigens using computational modeling and biolayer interferometry. Automated docking and molecular dynamics (MD) simulations were used to explore the origin of the specificity of an anti-blood group A antibody variable fragment (Fv AC1001). The analysis predicts a number of Fv-antigen interactions that contribute to affinity, including a hydrogen bond between a His^{L49} and the carbonyl moiety of the GalNAc in antigen A. This interaction was consistent with the dependence of affinity on pH, as measured experimentally; at lower pH there is an increase in binding affinity. Binding energy calculations provide unique insight into the origin of interaction energies at a per-residue level in both the scFv and the trisaccharide antigen. The calculations indicate that while the antibody can accommodate both blood group A and B antigens in its combining site, the A antigen is preferred by approximately 4 kcal/mol, consistent with the lack of binding observed for the B antigen.

Keywords: Molecular docking, MD simulations, Blood group antigens, Antibody specificity, GLYCAM, AMBER

Introduction

Since its discovery in 1900 (1), the ABO blood group system has played a crucial role in defining human blood and tissue compatibility. The blood type of an individual indicates the presence or absence of relevant antigens and antibodies. The three blood types share a core oligosaccharide antigen (H), and based on the glycosyl transferases inherited, different antigens are synthesized (2-4); type A transferase adds a terminal non-reducing *N*-acetylgalactosamine (GalNAc) residue; type B transferase adds galactose (Gal), whereas individuals with blood group O retain the unmodified H antigen. During the first years of life, the immune system forms antibodies upon exposure to non-self antigens from various exogenous factors. Thus an A-type individual will have circulating antibodies specific for the B-antigen, and vice-versa. The high degree of specificity is notable given that the only difference between the structures of the A- and B-antigens is the replacement of an acetamido moiety (in A) with a hydroxyl group (in B). Because of the presence of circulating antibodies, a mismatched blood transfusion or organ transplant can lead to hyperacute immune response and death (5, 6). Additionally, under certain circumstances, incompatibilities in blood groups between mother and child can trigger the mother's immune system to produce antibodies against the fetus, causing hemolytic disease (7).

Alterations in the structures of the ABO antigens often occur during carcinogenesis and therefore they have also been considered tumor markers (8, 9). Recently, strong correlations have been established between the presence of particular ABO and Lewis antigens, and susceptibility to infectious diseases, such as *Helicobacter pylori*, norovirus, and cholera (10), wherein the blood group antigens can be exploited as receptors for bacterial and viral adhesion. Conversely, it has

been suggested that endogenous anti-blood group antibodies can recognize blood-group-like carbohydrate antigens on pathogen surfaces, conferring protection against infection (11).

Despite their clinical importance, relatively little is known about the structural basis for these highly specific antibodies – antigen interactions. Although X-ray crystallography has been used to characterize antibody-carbohydrate complexes, the generally enhanced flexibility and conformational heterogeneity of oligosaccharides, detracts from the ability to generate co-crystals (12). Additionally, anti-carbohydrate antibodies bind to their antigens with an affinity that is 3-5 orders of magnitude lower than typical antibodies that bind to protein or peptide antigens. Difficulties in generating 3D structures for carbohydrate-antibody complexes has led to the increasing use of theoretical structure prediction methods (13, 14), which, while convenient, are prone to predicting false positives due to inaccuracies in pose scoring functions(15) and to the omission of carbohydrate conformational preferences(16) .

In this study, we examined the structural origin of the antigenicity (the specificity and affinity) of a monoclonal antibody raised against blood group A (BGA) antigen, for which an apo structure of the single-chain variable fragment (scFv AC1001) has been reported (17). The specificity data from screening two independent glycan arrays (Consortium for Functional Glycomics (v4.0, request ID: 1808) and from the group of Jeff Gildersleeve) confirmed that the scFv displayed no detectable binding to any B-antigens and only bound to BGA-containing glycans. To provide a structural interpretation for the specificity of AC1001 for BGA over blood groups H (BGH) and B (BGB), we generated a 3D model of the immune complex using molecular docking and refined it by molecular dynamics (MD) simulation. Despite its limitations, molecular docking, with or without additional experimental constraints, such as from NMR data, is often the only

approach that may be employed to generate the structure of a ligand-protein complex, in the absence of direct crystallographic data. To enhance the success rate a recent carbohydrate conformational energy function (16) was employed with AutoDock VINA (18), which quantifies the conformational preferences of oligosaccharides based on their glycosidic torsion angles. MD simulations (50 ns) were subsequently performed to ensure that the docked complexes were stable under physically realistic conditions, and in that event, the MD data were employed in binding free energy calculations. A particular advantage of MD-based energy calculations is that they provide statistically converged values that may be partitioned into contributions from individual residues in the protein and ligand(19).

Methods

Cloning, Expression and Purification of scFv: An scFv gene containing a short linker (RADAA) and the Leu 103H Val mutation (17), with a His_6 tag was assembled by PCR and cloned into the phagemid pSK4. The construct was maintained in *Escherichia coli* TG1 cells. Cells from positive clones, as judged by DNA sequence analysis, were grown in minimal media, induced, and subjected to periplasmic extraction. The scFv dimer was purified from the extract by Ni^{2+} immobilised metal affinity chromatography, by elution with an imidazole gradient.

Biolayer Interferometry (BLI): Affinity measurements were performed on a biolayer interferometer (Octet Red96, ForteBio). Data were processed using the Data Acquisition and Analysis 8.0 software (ForteBio), and kinetic binding constants determined from a 1:1 binding model using the OriginPro software (OriginLab). The scFv was immobilized on an amine

reactive second-generation (AR2G) biosensor (Lot No. 1311212, ForteBio). The BGA trisaccharide was analysed as the conjugate to bovine serum albumin (BSA-BGA) and was dissolved in an analysis buffer containing 10 mM HEPES, 150 mM NaCl , 3.4 mM EDTA , 0.005% Tween 20 at a range of pH values (5, 5.5, 6, 6.5, and 7). A BSA-Le^X trisaccharide conjugate (Prod. No. NGP0302, V-Labs, Inc.) and BSA (Prod. No. 23209, Pierce Thermo Scientific, Rockford, IL, USA) were used as negative controls. Details of the BLI conditions are provided in supplemental info (Supplemental Methods).

Automated Docking: Docking was performed using AutoDock VINA (18) with twenty docked poses generated for each experiment. The protein and the ligand files were prepared using Autodock tools (ADT) (20) with Gassteiger (21) partial atomic charges assigned to both the protein and ligand residues. The crystal structure of the scFv (PDB ID: 1JV5) was employed, together with a 3D structure of BGA obtained from the GLYCAM-Web server (www.glycam.org). Crystal waters were removed prior to docking and hydrogen atoms were added to the protein using ADT, whereas hydrogen atoms in the ligand were assigned from the GLYCAM residue templates. The glycosidic ϕ and φ torsion angles were allowed to be flexible during docking, as were all the hydroxyl groups. The protein was maintained rigid. The docking grid box (dimensions: 26.25 x 26.25 x 37.5 Å) was centered relative to the complimentarity determining regions (CDRs) of the antibody as described previously(16). For the mutational-docking approach, TrpH100 was mutated to Ala by deleting the side chain atoms of the Trp residue in the crystal structure, followed by processing with the tleap module in AMBER (22). AlaH100 was reverted back to Trp by restoring the crystal coordinates of the side chain of TrpH100. The docked poses from the mutational approach were filtered based on the clashes

with the reverted Trp. Poses in which the clashes could not be eliminated by implicit energy minimization (details in the MD simulations section) of were rejected. Ligand conformations of all the docked poses from both the flexible and mutational docking approaches were scored using the recently reported Carbohydrate Intrinsic (CHI) energy scoring function (16). Any conformations with total CHI-energies greater than 5 kcal/mol were rejected. The BGB complex was generated directly from that generated for BGA by simple replacement of the NAc group by an OH group.

MD Simulations: All the MD simulations were performed with the GPU implementation of the pmed code, pmed.cud_SPDP(23), from AMBER12(22). The calculations employed the ff99SSB(24) parameters for the protein and the GLYCAM06h(25) parameters for the carbohydrate. For the BGA, BGB-scFv complex simulations, an implicit solvent energy minimization (5000 steps of steepest descent followed by 5000 steps of conjugate gradient) was performed to optimize the side chain positions of the reverted Trp residue. During this minimization the backbone atoms of the framework regions were restrained with a 5 kcal/mol-$Å^2$ while the CDR regions and the ligand were allowed to be flexible. The systems were then solvated in a cubic water box (120Å per side, with a TIP3P water(26)). Each system was energy minimized using explicit solvent (10000 steps of steepest descent, 10000 steps of conjugate gradient). During this energy minimization, the protein residues were restrained with a force constant of 100 kcal/mol-$Å^2$ allowing only the solvent and ligand to relax. This minimization was followed by heating from 5 to 300 K over the course of 50 ps at constant volume. Production MD simulations were performed for 50ns at constant pressure (NPT ensemble) with the temperature held constant at 300 K using a Langevin thermostat. During the heating and the

production MD, the backbone atoms of the protein were restrained with a force constant of 5 kcal/mol-Å^2, with the protein side chains and ligand atoms allowed to be flexible. The backbone atoms were restrained in order to ensure that the protein fold remained stable during the course of the simulation. For the BGA trisaccharide MD simulation, the system was solvated in a cubic water box (120Å per side, with a TIP3P water) and energy minimized using explicit solvent (5000 steps of steepest descent, 5000 steps of conjugate gradient). This was followed by heating from 5 to 300 K for a period of 50ps at constant volume. Production MD simulations were performed for 50ns at constant pressure (NPT). During the minimization, heating and production MD simulations, there were no restraints placed on the trisaccharide. For both BGA, BGB-scFv complexes and BGA trisaccharide simulations, all covalent bonds involving hydrogen atoms were constrained using the SHAKE(27) algorithm, allowing a time step of 2 fs. A non-bonded cut-off of 8 Å was used and long-range electrostatics were employed using the particle mesh Ewald (PME) method(28). Snapshots were collected at 1 ps intervals for subsequent analysis.

Analysis

The stability of the complexes was assessed by monitoring the root-mean-squared-displacement (RMSD) of the ligand position, the glycosidic torsion angles, the ring conformations, and the protein-ligand hydrogen bonds. All these values except for the ring conformation analysis were generated using the ptraj module of AMBERTOOLS 12(29). Ligand displacement RMSD values were calculated for the ring atoms, relative to the first time step of the simulation. Hydrogen bond interactions between the protein and the ligand were measured with distance and angle cut-off values of 3.5 Å and 120° respectively. The ring conformations of each individual residue in

the ligand during the course of simulation were analyzed using the recently reported BFMP method (Makeneni et al., submitted). Binding free energies were calculated with the MMGBSA(30, 31) module in AMBERTOOLS12. All the water molecules were removed prior to the MM-GBSA calculation, and desolvation free energies approximated using the generalized born implicit solvation model (igb = 2)(32).

Results and Discussion

Docking Analysis

In preliminary experiments, docking to the rigid scFv structure yielded complexes that failed to remain stable during subsequent 10 ns MD simulations (Supplemental Table 1). The spontaneous dissociation of the complex during MD simulation suggested that the docking had failed to detect the correct, high affinity, pose(33). Upon inspection of the MD data, it was observed that light chain residue His49 (His^{L49}) forms a stacking interaction with heavy chain residue Trp100 (Trp^{H100}), which occupies a large volume of the presumed binding site, potentially preventing deeper penetration of the ligand (Figure 6.1).

As Trp residues can also form stacking interactions with the apolar face of monosaccharides in antibody complexes(34), we hypothesized that the trisaccharide ligand might compete for formation of such an interaction with Trp^{H100}. For example, the Galactose (Gal) residue in a *Salmonella* trisaccharide antigen stacks against Trp^{L93} in the complex with Fab Se155-4(34). In addition, in the same complex, Trp^{H33} stacks against the C-6 position in the 6-deoxy sugar Abequose. The BGA antigen contains GalNAc and a 6-deoxy monosaccharide

(Fucose, Fuc), thus a revised docking experiment was sought that would permit the formation of such interactions with the aromatic residues in the binding pocket. Thus, two alternative docking experiments were designed: in the first, the side chain torsion angles of Trp^{H100} were allowed to be flexible during docking (termed flexible residue docking); while in the second, Trp^{H100} was mutated to Ala prior to docking, and then reverted back to Trp after docking (mutational residue docking). The docked poses were filtered based on three criteria. Firstly, poses in which the GalNAc was not located within the binding pocket were eliminated (Figure 2c). This criterion was adopted based on the results from two array screenings, which indicated that the antibody interacts exclusively with the BGA antigens (Appendix B Supplementary Table 2 and 3) and because the only structural difference between BGA and BGB is the presence of the NAc moiety in the former. Therefore it was hypothesized that the ability of the antibody to discriminate between these two antigens would be dependent on interactions with this residue. Secondly, in the case of the mutational approach, poses were rejected if the Ala-Trp mutation led to irreconcilable steric clashes with the antigen (Figure 2b). All the docked poses obtained from each of these approaches were then scored using a carbohydrate intrinsic scoring function. After applying these criteria, both docking approaches identified essentially equivalent antigen poses (0.48Å RMSD between ligand positions) (Figure 2a), in which the C6 atom of the GalNAc forms a CH/φ stacking interaction with the Trp^{H96}. This complex was selected for further analysis by MD simulation.

Structural Stability of the immune complexes

BGA

The final docked model of the blood group antigen A bound to the antibody remained stable during the course of a 50ns simulation based on the RMSD of the ring atoms of the ligand, which remained between 2 – 4 Å over the course of the simulation (Figure 5). An analysis of the ring conformational preferences showed that all three residues in the trisaccharide remained in the $^4C_1$ chair conformations. The ϕ- and φ-glycosidic torsion angles for the GalNAcα(1,3)Gal (ϕ_1, φ_1) and Fucα(1,2)Gal (ϕ_2, φ_2) linkages were monitored throughout both the simulations (BGA-scFv complex and BGA trisaccharide in solution), and the average values were found to be in agreement with the values observed for the same trisaccharide in the complex with *Dolichos biflorus* lectin as well as the conformations of the trisaccharide in solution(35) (Table 1). The stacking interactions between the GalNAc and Trp^{H96} interactions were characterized by the angle (φ) between the normals to the ring planes, and the distance (R) between their centroids (36). For an ideal stacking conformation, φ should be around 180° or 0°, and for CH/φ, it should be around 90°. The average θ value was close to the latter at 108° (with a standard deviation of 9°) at a distance of 6.5 Å.

During the course of the MD simulation, the side chain of His^{L49} was observed to flip from its initial orientation (χ_2 = <-73°>) to one (<115°>) in which it could form a hydrogen bond with the *N*-acetyl group of the GalNAc residue (Figure 3, Table 2, Figure 4). This interaction remained stable for the remainder of the 50 ns simulation. This side-chain flip may represent an

example of induced fit during ligand binding, however, at the resolution of the present x-ray data (2.2 Å), it is not possible to reliably discriminate between Histidine χ_2 rotamers.(37)

BGB

To probe the specificity of the antibody for antigen B, the scFv was screened experimentally against an array of neoglycoconjugates including ABO and related blood group antigens. The screening confirmed the exclusive specificity of the antibody for BGA-related antigens (Appendix B Supplemental Tables 2 and 3). Computational carbohydrate grafting(39) of the relevant glycans from the array onto the bound BGA trisaccharide in the scFv complex confirmed that all of the BGA- and BGB-related glycans could be accommodated in the binding pocket (Appendix B Supplemental Table 3). Therefore, the lack of binding of the BGB-glycans does not appear to be due to steric collisions, but rather to the loss of affinity arising from the absence of the NAc group in the BGA congeners. MD simulation of the BGB-scFv complex was employed to examine the effect of the loss of the NAc moiety on the stability and affinity of the structural difference in the antigens on the stability and affinity of the putative immune complex. Despite the fact that the MD simulations of the two complexes (BGA and BGB) were started with the antigens aligned in identical binding modes, the BGB antigen dissociated from the antibody after a relatively short simulation period of 10 ns. In order to eliminate the possibility that this instability arose due to artifacts from the conversion of the BGA to BGB antigen, two additional simulations were performed with independent initial atomic velocities. In both cases, the ligand appeared to dissociate from the antibody after approximately 10 ns (Figure

5). To enable comparison with the BGA complex, only the data from the initial stable 10 ns period of the BGB complex were chosen for analysis.

In antigen-scFv complexes, the Gal or GalNAc residues are flanked by residues Tyr^{L50}, Asn^{L34}, His^{L49} on one side of the antigen (Group 1) and residues Trp^{H100}, Trp^{L96} (Group 2) on the other; the Fuc interacts with Gly^{L91} and Asn^{L92} (Group 3) (Figure 6). In contrast to the case of the BGA antigen, in the BGB-scFv simulation His^{L49} does not form a stabilizing interaction with the terminal Gal residue. Additionally, the Gal and Fucl residues display enhanced flexibility owing to the loss of stabilizing interactions with residues from Groups 2 and 3.

Involvement of His^{L49} in binding affinity

All Histidines in the scFv were protonated by default for modeling with a hydrogen atom at the δ nitrogen position. During the MD simulation of the BGA-scFv complex, the χ2 angle of His^{L49} flips (-73° to 115°) enabling a hydrogen bond to form with the carbonyl moiety of the NAc group in the GalNAc residue in BGA, which would be expected to be significant for enhancing the stability of the BGA-scFv complex. In the BGB complex, the same His^{L49} forms an interaction with the non-terminal Gal residue. The interaction with $HisL^{49}$ suggests that there might also be a pH dependence on binding; at lower pH all Histidines would be positively charged, potentially enhancing the strength of the His^{L49} - BGA hydrogen bond, leading to higher binding affinity. This prediction was confirmed by BLI measurements, which showed a marked decrease in the apparent K_D as the pH dropped below the pK_A of histidine (Figure 6.7). It should be noted that this protonation would not be localized to His^{L49} nevertheless, no enhanced non-specific binding was observed at low pH for either BSA or BGA-Le^x (Appendix B

Supplemental Figures 1, 2 and 3), supporting a role for a direct interaction between His^{L49} and the BGA antigen.

Binding Energy Analysis

A per-residue decomposition of the interaction energies in the immune complexes indicated that, in the case of BGA, the GalNAc residue contributed 25% (-8.1 kcal/mol) towards the binding energy, compared to a reduced (-4.3 kcal/mol) contribution from the corresponding Gal residue in BGB (Table 3). This loss of approximately 4 kcal/mol of interaction energy is the predominant difference between the two antigens, and would be enough to reduce the affinity by nearly 800 fold, consistent with the lack of apparent binding of the BGB analogs in the glycan array screening. In addition, this analysis identified the residues that contributed significantly towards antigen binding.

In the BGA-scFv complex, residues from CDR L3 make the maximum contributions to binding (Gly^{L91} + Trp^{L96} + Asn^{L92} +Thr^{L93} = -7.2 kcal/mol) followed by H3 (Asn^{H98} + Trp^{H100} + Leu^{H99} = -5.5 kcal/mol), L1 (Tyr^{L32} + Asn^{L34} = -4.5 kcal/mol) and L2 (Tyr^{L50} = -1.02 kcal/mol). In contrast, in the case of BGB, the same residues from L3 contribute less than a total of 1 kcal/mol to the interaction energies. The most significant single residues are Tyr^{L32}, Gly^{L91}, Trp^{H100} and Trp^{L96}, which each contributes more than 2 kcal/mol and together account for approximately 50% of the total affinity. Residues Gly^{L91} and Asn^{L92} that form hydrogen bonds with the Fuc residue together contribute -4.0 kcal/mol to the binding of BGA, but fail to make any stable interactions in the BGB simulation and therefore contribute negligibly to the affinity. It is these interactions that provide the predominant contributions to the preferential binding of

the BGA antigen. While in the BGB complex His^{L49} does not form any stable hydrogen bonds with the terminal Gal, it is able to form new, albeit transient, interactions with the non-terminal Gal for approximately 30% of the stable simulation period. Therefore, while the per-residue contribution values indicate that His^{L49} makes a contribution greater than -1.5 kcal/mol in both cases, the interactions it forms in BGA are more stable when compared to the interactions in BGB.

Conclusions

In this study, 3D models of the BGA and BGB trisaccharides in complex with scFv AC1001 were generated that provided a detailed atomic level rationalization of the interactions and dynamics responsible for antigen specificity. Quantification of the binding affinities identified key residues in the binding site that are predicted to contribute to specific and non-specific interactions with each antigen and led to the confirmed prediction of enhanced binding at lower pH. The spontaneous dissociation of antigen B from the scFv-BGB complexes (in three different simulations) indicated that MD simulations confirm the known preference of this antibody for the A antigen, and support a role for MD simulations in overcoming limitations associated with ligand docking. The present study illustrates that integration of multiple experimental (affinity measurements, glycan array screening, and crystallography) and theoretical (ligand docking, MD simulation, and energy decomposition) methods provides a powerful platform for predicting the origin of antibody-carbohydrate specificity.

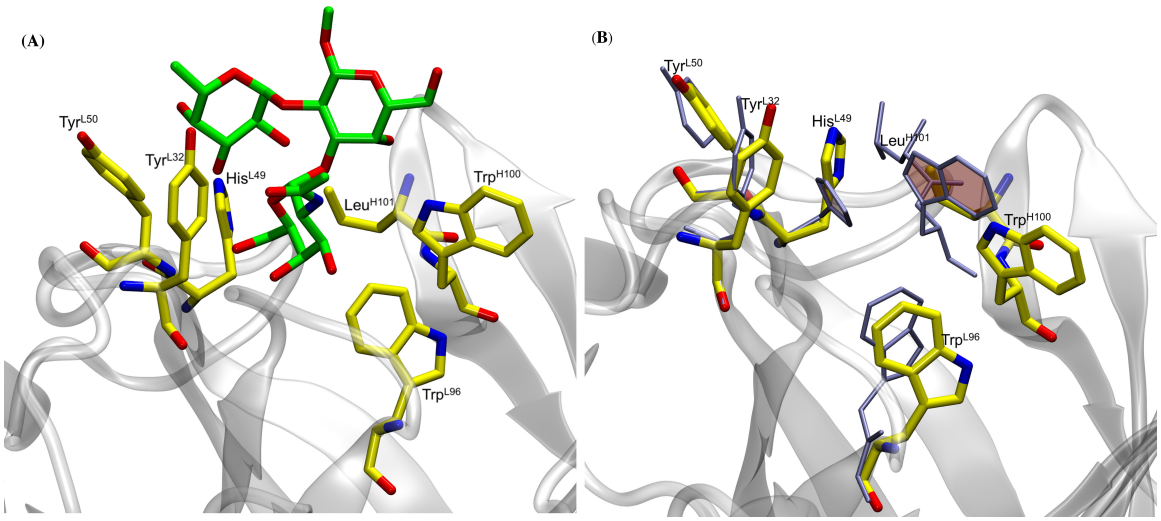


Figure 1. (A) Docked antigen A (green) from preliminary docking experiments with residues lining the binding pocket (shown in yellow). The antibody is shown in grey. (B) Residues lining the binding pocket before (yellow) and after (ice blue) the 50 ns MD simulation. Residues His^{L49} and Trp^{H100} (shaded rings) form stacking interactions during the course of the simulation thereby causing the ligand to become unstable.

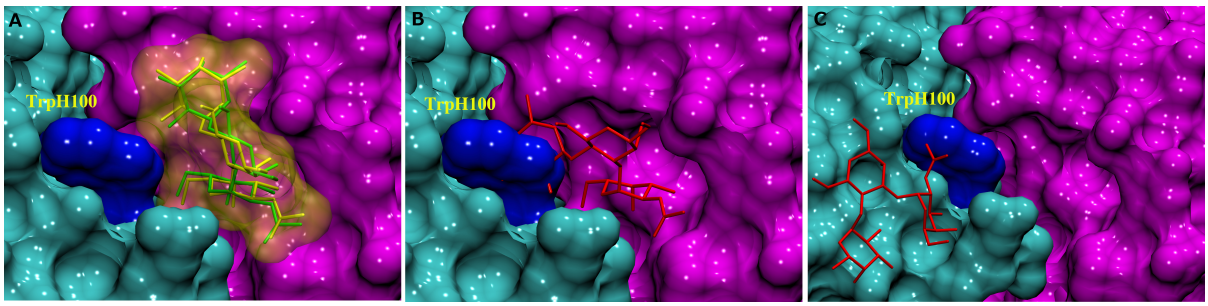


Figure 2. Docked complexes of BGA (stick structure) in the scFv binding site (heavy and light chains shown as solvent accessible surfaces in cyan and pink, respectively, the Trp^{H100} surface is shown in dark blue). (A) The stick structures in green and yellow represent the best-docked poses from the Trp^{H100}-mutagenesis and the flexible residue docking approaches, respectively. (B) An example of a docked pose (red) that was eliminated on the basis of clashes ensuing from the $Ala^{H100}Trp$ mutation. (C) An example of a docked pose (red) that was eliminated on the basis of the orientation of the ligand in the binding pocket.

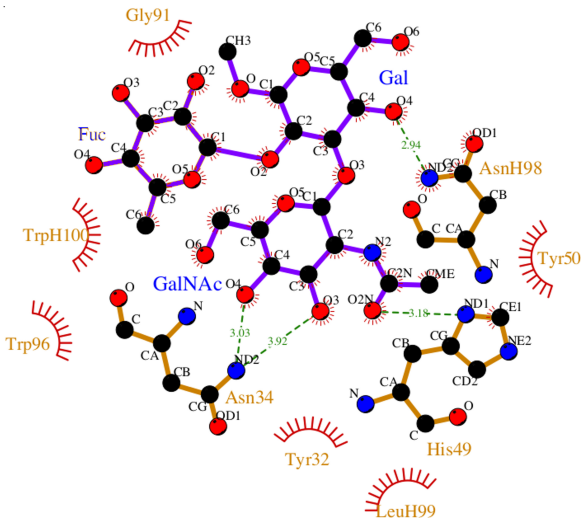


Figure 3. Non-bonded interactions between the BGA and Fab AC1001 (prepared using LigPlot(38)). The structure represents a single frame of the MD simulation that is closest to the average RMSD of the structure during the simulation.

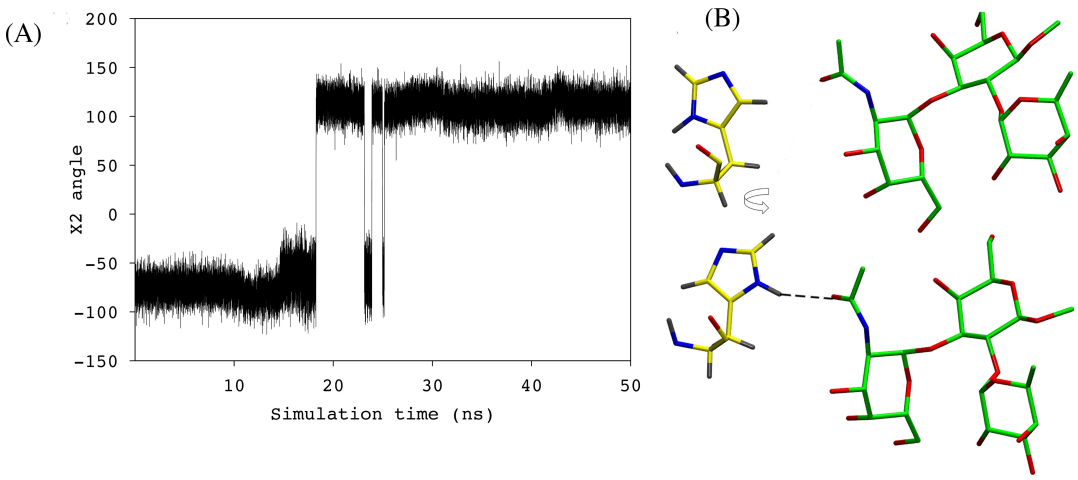


Figure 4. (A) χ_2 angle of the His^{L49} during the course of the simulation. (B) His^{L49} (shown in yellow) during the first 18 ns of the simulation (top) and the remainder of the simulation (bottom)

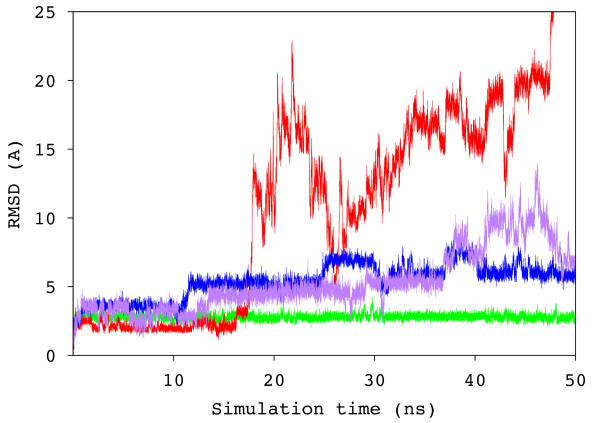


Figure 5. Time series of the RMSD values for the ring atoms of the BGA (green) and BGB (from three independent simulations, blue, purple, and red) antigens, relative to the starting conformation of the complex.

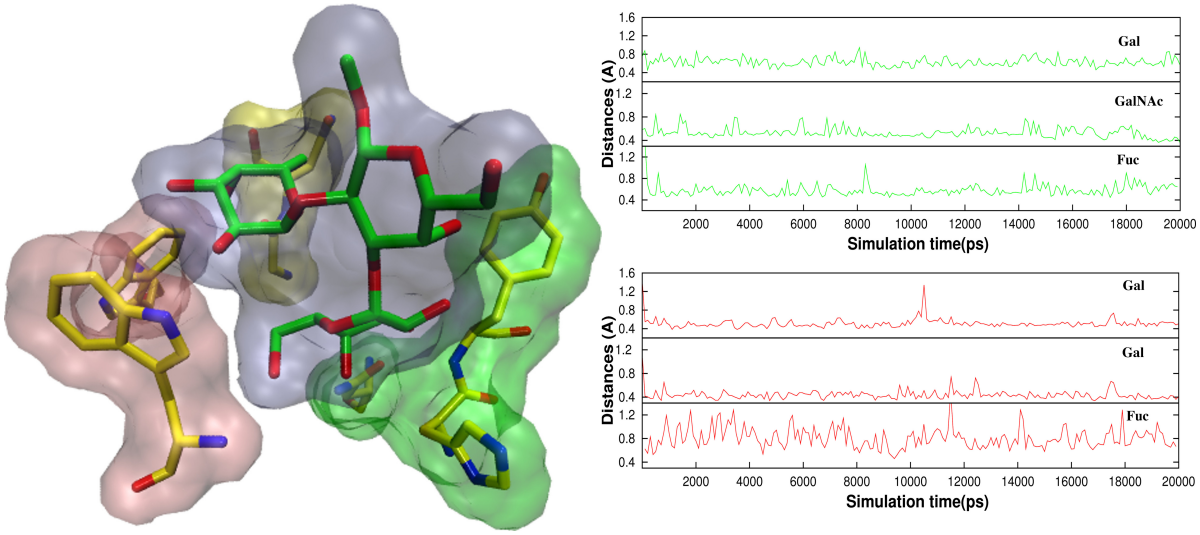


Figure 6. (Left) The antigens are flanked by Tyr^{L50}, Asn^{L34}, His^{L49} (Group 1, green surface), and Trp^{H100}, Trp^{L96} (Group 2, pink surface). Fuc interacts with Gly^{L91} and Asn^{L92} (Group3, yellow surface). (Right) Atomic fluctuations of residues Gal, GalNAc/Gal (BGA/BGB) and Fuc as a function of time.

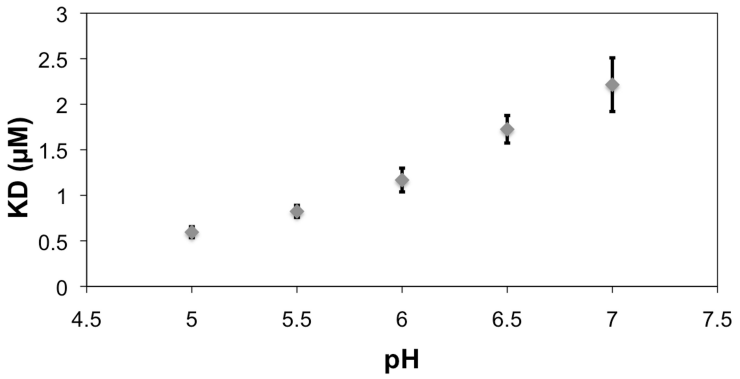


Figure 7. The reference (BSA)-subtracted pH dependence of the apparent K_D for the interaction between scFv AC1001 and the BSA-BGA conjugate. Error bars are derived from replicates of five measurements. Note, the pKa of Histidine is 6.04.(40)

Table 1: Comparison of glycosidic torsion angles between experimentally observed values and average values obtained from the MD simulations.

	(ϕ_1, φ_1)[a]		(ϕ_2, φ_2)[b]	
	Experimental	Theoretical	Experimental	Theoretical
BGA trisaccharide	$62^o<\phi_1<82^o$, $61^o<\varphi_1<74^o$	-68 ± 14^o, 51 ± 25^o	$-77^o< \phi_2<-67^o$, $-109^o< \varphi_2<-86^o$	-69 ± 11^o, -101 ± 26^o
BGA-scFv complex	68^o, 77^o	82 ± 11^o, 68 ± 7^o	-68^o, -90^o	-69 ± 8^o, -113 ± 10^o

[a]glycosidic torsion angles for the GalNAcα(1,3)Gal (ϕ_1, φ_1)

[b]torsion angles for Fucα(1,2)Gal (ϕ_2, φ_2)

Table 2: Hydrogen bonds between BGA and the scFv during the MD simulation.

Donor		Acceptor		MD Period: 0-18ns		MD Period: 18-50ns	
				Distance[a,b]	Occupancy[c]	Distance	Occupancy
GalNAc	O3	Asn^{L34}	Hδ1	3.1 (0.18)[c]	67	>3.5	---
	O4	Asn^{L34}	Hδ1	3.1(0.22)	32	3.0 (0.17)	77
	O2N	His^{L49}	Hδ	>3.5	---	2.9 (0.16)	91
Gal	O4	GalNAc	H2N	3.2 (0.17)	65	3.2 (0.17)	31
	O4	Asn^{H98}	Hδ1	3.1 (0.18)	45	3.1 (0.17)	41

[a]In Å.

[b]Standard deviations in parentheses.

[c]Percentage.

Table 3: Key[a] per-residue contributions towards the energy for the BGA and BGB-scFv complexes.

Residue	vdW	Electrostatic	Polar Desolvation	Non-Polar Desolvation	Total
Antibody					
Tyr^{L32}	-4.2[b]	-1.4	2.1	-0.5	-4.0
	-3.0[b]	-1.7	2.0	-0.5	-3.2
Gly^{L91}	-0.8	-4.8	3.2	-0.2	-2.5
	---	---	---	---	---
Trp^{H100}	-2.5	-1.2	1.7	-0.3	-2.2
	-2.8	-1.0	1.2	-0.4	-2.9
Trp^{L96}	-1.4	-1.1	0.7	-0.2	-2.0
	-0.7	-0.3	0.2	-0.1	-0.9
Asn^{L34}	-0.8	-1.6	0.9	-0.1	-1.6
	-0.6	-3.7	1.5	-0.1	-2.9
Asn^{H98}	-2.2	-2.9	4.0	-0.5	-1.5
	-2.8	-1.6	3.8	-0.5	-1.1
Asn^{L92}	-1	-2.9	2.6	-0.2	-1.5
	---	---	---	---	---
Tyr^{L50}	-1.6	-0.9	1.1	-0.1	-1.5
	-1.7	-1.0	1.4	-0.1	-1.5
Leu^{H99}	-1.4	-1	1	-0.1	-1.4
	-1.4	-0.3	0.8	-0.1	-1
His^{L49}	-0.6	-3.3	2.5	-0.1	-1.4
	-1.3	-2.8	2.5	-0.2	-1.8
Thr^{L93}	-0.5	-0.6	0.6	0.0	-0.5
	---	---	---	---	---
Subtotal	-17.1	-20.5	20.5	-2.2	-19.3
	-11.3	-12.4	13.4	-2	-15.3
Antigen					
Gal	-3.1	-0.4	1.2	-0.4	-2.7
	-6.2	-3.7	5.2	-0.9	-5.6
GalNAc	-13.1	-12.5	19.6	-2.1	-8.1
Gal	-10.1	-8.2	15.8	-1.7	-4.3
Fuc	-4.0	-9.7	12.2	-0.8	-2.3
	-2.5	-1.2	4.5	-0.4	0.4
Ligand Total	-20.2	-22.6	32.9	-3.2	-13.2

	-18.9	-13.2	25.4	-1.4	-9.6

[a]Key residues defined as those that contribute greater than 0.5 kcal/mol to the total interaction energy for either the BGA or BGB in the complexes. Only the initial stable 10 ns period of the BGB simulation was employed, whereas the entire 50 ns trajectory for BGA was analyzed.

[b]Upper row, values for BGA, lower, BGB.

References

1. Landsteiner K. Zur Kenntnis der antifermentativen, lytischen und agglutinierenden Wirkungen des Blutserums und der Lymphe. *Zentralblatt Bakteriologie* (1900) 27(10):357-62.
2. Yamamoto F. Review: ABO blood group system--ABH oligosaccharide antigens, anti-A and anti-B, A and B glycosyltransferases, and ABO genes. *Immunohematology / American Red Cross* (2004) 20(1):3-22. Epub 2004/09/18. PubMed PMID: 15373665.
3. Chester MA, Olsson ML. The ABO blood group gene: A locus of considerable genetic diversity. *Transfusion medicine reviews* (2001) 15(3):177-200.
4. Marionneau S, Cailleau-Thomas A, Rocher J, Le Moullac-Vaidye B, Ruvoën N, Clément M, et al. ABH and Lewis histo-blood group antigens, a model for the meaning of oligosaccharide diversity in the face of a changing world. *Biochimie* (2001) 83(7):565-73. doi: http://dx.doi.org/10.1016/S0300-9084(01)01321-9.
5. Williamson LM, Lowe S, Love EM, Cohen H, Soldan K, McClelland DBL, et al. Serious hazards of transfusion (SHOT) initiative: analysis of the first two annual reports. *BMJ* (1999) 319(7201):16-9. doi: 10.1136/bmj.319.7201.16.
6. Sazama K. Reports of 355 transfusion-associated deaths: 1976 through 1985. *Transfusion* (1990) 30(7):583-90. doi: 10.1046/j.1537-2995.1990.30790385515.x.
7. Ozolek JA, Watchko JF, Mimouni F. Prevalence and lack of clinical significance of blood group incompatibility in mothers with blood type A or B. *The Journal of pediatrics* (1994) 125(1):87-91.

8. Le Pendu J, Marionneau S, Cailleau-Thomas A, Rocher J, Le Moullac-Vaidye B, Clement M. ABH and Lewis histo-blood group antigens in cancer. *APMIS : acta pathologica, microbiologica, et immunologica Scandinavica* (2001) 109(1):9-31. Epub 2001/04/12. PubMed PMID: 11297197.

9. Dabelsteen E, Gao S. ABO Blood-group antigens in oral cancer. *Journal of Dental Research* (2005) 84(1):21-8. doi: 10.1177/154405910508400103.

10. Anstee DJ. The relationship between blood groups and disease. *Blood* (2010) 115(23):4635-43. doi: 10.1182/blood-2010-01-261859.

11. Ceravolo IP, Sanchez BAM, Sousa TN, Guerra BM, Soares IS, Braga EM, et al. Naturally acquired inhibitory antibodies to Plasmodium vivax Duffy binding protein are short-lived and allele-specific following a single malaria infection. *Clinical & Experimental Immunology* (2009) 156(3):502-10. doi: 10.1111/j.1365-2249.2009.03931.x.

12. DeMarco ML, Woods RJ. Structural glycobiology: A game of snakes and ladders. *Glycobiology* (2008) 18(6):426-40. doi: 10.1093/glycob/cwn026.

13. Woods R, Yongye A. Computational techniques applied to defining carbohydrate antigenicity. In: Kosma P, Müller-Loennies S, editors. *Anticarbohydrate Antibodies*. Springer Vienna (2012). p. 361-83.

14. Paula S, Monson N, Ball WJ, Jr. Molecular modeling of cardiac glycoside binding by the human sequence monoclonal antibody 1B3. *Proteins* (2005) 60(3):382-91. Epub 2005/06/23. doi: 10.1002/prot.20484. PubMed PMID: 15971203.

15. Woods RJ, Tessier MB. Computational glycoscience: characterizing the spatial and temporal properties of glycans and glycan-protein complexes. *Current opinion in*

structural biology (2010) 20(5):575-83. Epub 2010/08/17. doi: 10.1016/j.sbi.2010.07.005. PubMed PMID: 20708922; PubMed Central PMCID: PMC3936461.

16. Nivedha AK, Makeneni S, Foley BL, Tessier MB, Woods RJ. Importance of ligand conformational energies in carbohydrate docking: Sorting the wheat from the chaff. *Journal of Computational Chemistry* (2014) 35(7):526-39. doi: 10.1002/jcc.23517.

17. Thomas R, Patenaude SI, MacKenzie CR, To R, Hirama T, Young NM, et al. Structure of an anti-blood group A Fv and improvement of its binding affinity without loss of specificity. *The Journal of biological chemistry* (2002) 277(3):2059-64. Epub 2001/10/27. doi: 10.1074/jbc.M104364200. PubMed PMID: 11679577.

18. Trott O, Olson AJ. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry* (2010) 31(2):455-61. doi: 10.1002/jcc.21334.

19. Hadden JAT, M. B.; Fadda, E.; Woods, R. J. Calculating binding free energies for protein-carbohydrate complexes. *Methods in Molecular Biology: Glycoinformatics*. NJ(In press).

20. Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry* (2009) 30(16):2785-91. doi: 10.1002/jcc.21256.

21. Gasteiger J, Marsili M. Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges. *Tetrahedron* (1980) 36(22):3219-28. doi: http://dx.doi.org/10.1016/0040-4020(80)80168-2.

22. Case DA, Darden TA, Cheatham TE, Simmerling CL, Wang J, Duke RE, et al. *AMBER 12*. University of California, San Francisco (2012).

23. Götz AW, Williamson MJ, Xu D, Poole D, Le Grand S, Walker RC. Routine microsecond molecular dynamics simulations with AMBER on GPUs. 1. Generalized Born. *Journal of Chemical Theory and Computation* (2012) 8(5):1542-55. doi: 10.1021/ct200909j.

24. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins* (2006) 65(3):712-25. Epub 2006/09/19. doi: 10.1002/prot.21123. PubMed PMID: 16981200.

25. Kirschner KN, Yongye AB, Tschampel SM, González-Outeiriño J, Daniels CR, Foley BL, et al. GLYCAM06: A generalizable biomolecular force field. Carbohydrates. *Journal of Computational Chemistry* (2008) 29(4):622-55. doi: 10.1002/jcc.20820.

26. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* (1983) 79(2):926-35. doi: doi:http://dx.doi.org/10.1063/1.445869.

27. Ryckaert J-P, Ciccotti G, Berendsen HJC. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics* (1977) 23(3):327-41. doi: http://dx.doi.org/10.1016/0021-9991(77)90098-5.

28. Darden T, York D, Pedersen L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* (1993) 98(12):10089-92. doi: doi:http://dx.doi.org/10.1063/1.464397.

29. Roe DR, Cheatham TE. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *Journal of Chemical Theory and Computation* (2013) 9(7):3084-95. doi: 10.1021/ct400341p.

30. Kollman PA, Massova I, Reyes C, Kuhn B, Huo S, Chong L, et al. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Accounts of Chemical Research* (2000) 33(12):889-97. doi: 10.1021/ar000033j.

31. Srinivasan J, Miller J, Kollman PA, Case DA. Continuum solvent studies of the stability of RNA hairpin loops and helices. *Journal of Biomolecular Structure and Dynamics* (1998) 16(3):671-82. doi: 10.1080/07391102.1998.10508279.

32. Onufriev A, Bashford D, Case DA. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins: Structure, Function, and Bioinformatics* (2004) 55(2):383-94. doi: 10.1002/prot.20033.

33. Proctor Elizabeth A, Yin S, Tropsha A, Dokholyan Nikolay V. Discrete molecular dynamics distinguishes nativelike binding poses from decoys in difficult targets. *Biophysical Journal* (2012) 102(1):144-51. doi: http://dx.doi.org/10.1016/j.bpj.2011.11.4008.

34. Cygler M, Rose DR, Bundle DR. Recognition of a cell-surface oligosaccharide of pathogenic Salmonella by an antibody Fab fragment. *Science* (1991) 253(5018):442-5. doi: 10.2307/2878892.

35. Casset F, Peters T, Etzler M, Korchagina E, Nifant'ev N, Pérez S, et al. Conformational analysis of blood group A trisaccharide in solution and in the binding site of Dolichos biflorus lectin using transient and transferred nuclear overhauser enhancement (NOE) and

Rotating-Frame NOE Experiments. *European Journal of Biochemistry* (1996) 239(3):710-9. doi: 10.1111/j.1432-1033.1996.0710u.x.

36. Ford MG, Weimar T, Kohli T, Woods RJ. Molecular dynamics simulations of galectin-1-oligosaccharide complexes reveal the molecular basis for ligand diversity. *Proteins* (2003) 53(2):229-40. Epub 2003/10/01. doi: 10.1002/prot.10428. PubMed PMID: 14517974.

37. Glusker JPL, Mitchell; Rossi, Miriam. crystal structure analysis for chemists and biologists *Journal of Chemical Education* (1995) 72(3):A73. doi: 10.1021/ed072pA73.9.

38. Wallace AC, Laskowski RA, Thornton JM. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Engineering* (1995) 8(2):127-34. doi: 10.1093/protein/8.2.127.

39. Tessier MB, Grant OC, Heimburg-Molinaro J, Smith D, Jadey S, Gulick AM, et al. Computational screening of the human TF-glycome provides a structural definition for the specificity of anti-tumor antibody JAA-F11. *PLoS ONE* (2013) 8(1):e54874. doi: 10.1371/journal.pone.0054874.

40. Wood EJ. Data for biochemical research (third edition) by R M C Dawson, D C Elliott, W H Elliott and K M Jones. *Biochemical Education* (1987) 15(2):97-. doi: 10.1016/0307-4412(87)90110-5.

APPENDIX D: SUPPLEMENTARY DATA OF CHAPTER 6

Details of BLI experiments

Measurements were set as 60s (equilibration) - 300s (activation) – 600s (immobilization) – 300s (quenching) – 120s (baseline) – 600s (association) – 600s (dissociation) at 25°C. For details about operation of BLI please see Octet BLI technical note 26 from ForteBio webpage. All BSA and BSA conjugates were prepared in analysis buffer at 1 μM concentration. scFv was loaded onto AR2G biosensor at 1 μM in water.

Supplementary Table 1. Results from preliminary docking experiments

Rank	Residue[a]	CHI Energy[b]
1	GalNAc	6.9
2	GalNAc	4.4
3	Gal	4.5
4	Fuc	9.4
5	Gal	3.4
6	Gal	5.6
7	Fuc	9.1
8	Gal	9.9
9	Fuc	1.8
10	Gal	9.8
11	GalNAc	4.5
12	Fuc	2.2
13	-	3.3
14	-	9.0
15	Gal	2.1
16	Gal	6.0
17	Gal	4.0
18	GalNAc	3.9
19	GalNAc	5.0
20	Gal	9.4

[a]Indicates the residue located in the V-shaped binding pocket

[b]Energies of the conformation of the docked pose calculated using a CHI energy scoring function[1]. All values are in kcal/mol.Twenty docked poses were generated using Autodock VINA. Each of the docked poses were scored using a CHI Energy scoring function. Docked poses with energy higher than 5kcal/mol were eliminated. Of the remaining docked structures, poses in which GalNAc was not within the V-shaped deep binding pocket were eliminated. The remaining docked poses (Rank 2,11,18 and 19) were subjected to MD simulations. None of these complexes remained stable.

Supplementary Table 2. Binders from the glycan array screening of scFv AC1001 against the Consortium of Functional Glycomics (CFG) printed glycan array (v4.0). Complete data can be accessed on the CFG website (request ID: 1808)

Glycan Sequence	Experimental RFU[a]
GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	53921
GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp0	51949
GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb-Sp0	47339
GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp8	46920
GalNAca1-3(Fuca1-2)Galb1-4Glcb-Sp0	41543
GalNAca1-3(Fuca1-2)Galb-Sp8	37136
GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-3Galb1-4GlcNAcb1-3Galb1-4GlcNAcb-Sp0	36786
GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-3(GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	35917
GalNAca1-3(Fuca1-2)Galb1-3GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp0	35914
GalNAca1-3(Fuca1-2)Galb1-4(Fuca1-3)GlcNAcb-Sp0	34747
GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-3(GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-2Mana1-6)Manb1-4GlcNAcb1-4GlcNAcb-Sp20	34620
GalNAca1-3(Fuca1-2)Galb-Sp18	33170
GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-3GalNAca-Sp14	30552
GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-3GalNAc-Sp14	29972
Fuca1-2Galb1-3GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb-Sp0	4748

[a]Relative Fluorescence Units from the array screening

Supplementary Table 3. Comparison of theoretical (computational carbohydrate grafting) and experimental data for glycan array screening of scFv AC1001 against a carbohydrate array.

Glycan Sequence	CCG[a]	Experimental RFU[c]				
	Score[b]	0.2[d]	2	20	100	200
Cy3-BSA (20mg/mL + BSA, 125mg/mL total)	0	79997	85412	111073	57627	93252
GalNAca1-3[Fuca1-2]Galb1-4GlcNAcb-Sp-BSA	0	4007	15442	16935	19376	31929
GalNAca1-3(Fuca1-2)Galb1-3GalNAcb1-3Gala1-4Galb1-BSA	0	4552	17206	22662	22679	22362
GalNAca1-3(Fuca1-2)Galb1-3Galb1-linker-BSA	0	3454	14997	18395	19402	22027
GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-linker-BSA	0	4634	15633	17443	19354	21289
GalNAca1-3[Fuca1-2]Galb1-3GlcNAcb-Sp-BSA	0	4441	14639	17987	18327	20932
GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-linker-BSA	0	5758	15841	19124	20324	20864
GalNAca1-3(Fuca1-2)Galb1-4Glcb1-linker-BSA	0	3927	15086	18092	19159	20785
GalNAca1-3(Fuca1-2)Galb1-3GalNAcb1-linker-BSA	0	4888	15188	17374	20457	20739
GalNAca1-3(Fuca1-2)Galb1-3GalNAcα1-linker-BSA	0	3302	12027	15352	19104	20220
GalNAca1-3[Fuca1-2]Galb1-4[Fuca1-3]GlcNAcb-Sp-BSA	0	5711	15196	18043	18027	18947
GalNAca1-3(Fuca1-2)Galb- -BSA	0	985	10291	15891	18071	17572
use 25ug/mL + 100ug/mL BSA	0	15926	11208	13943	8524	16557
GalNAca1-3[Fuca1-2]Galb1-4GlcNAcb-Sp-BSA	0	1306	9807	14607	13612	16062
GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-3Galb1-4(Glc)-APD-HSA	0	1662	9483	12090	14604	15524
GalNAca1-3[Fuca1-2]Galb1-3GlcNAcb-Sp-BSA	0	2006	10983	20920	16997	15481
GalNAca1-3(Fuca1-2)Galb1-3GalNAcb1-3Gala1-4Galb1-BSA	0	722	7078	9496	15315	14869
GalNAca1-3[Fuca1-2]Galb1-4[Fuca1-3]GlcNAcb-Sp-BSA	0	4152	15498	23244	24923	14708
GalNAca1-3[Fuca1-2]Galb1-4GlcNAcb-Sp-BSA	0	978	8354	12292	13545	14688
GalNAca1-3(Fuca1-2)Galb1-3(Fuca1-4)GlcNAcb1-3Galb1-	0	4957	11959	14146	15916	13625

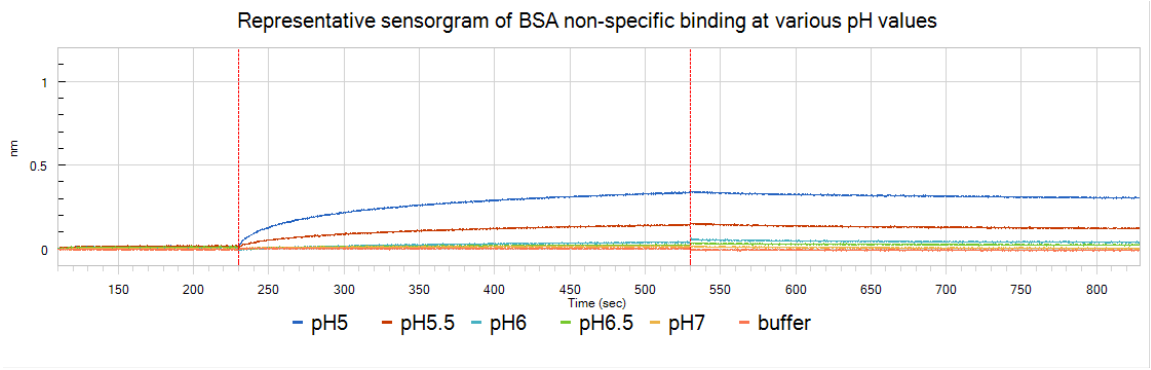
BSA						
GalNAca1-3(Fuca1-2)Galb1-3GlcNAcb1-linker-BSA	0	779	7088	11132	12175	12878
GalNAca1-3(Fuca1-2)Galb1-3GalNAcb1-linker-BSA	0	587	6762	11120	13016	12831
GalNAca1-3(Fuca1-2)Galb1-4GlcNAcb1-linker-BSA	0	343	4564	9932	11241	11679
GalNAca1-3(Fuca1-2)Galb1-3Galb1-linker-BSA	0	247	4306	8506	11066	11549

[a]Computational Carbohydrate Grafting[2]

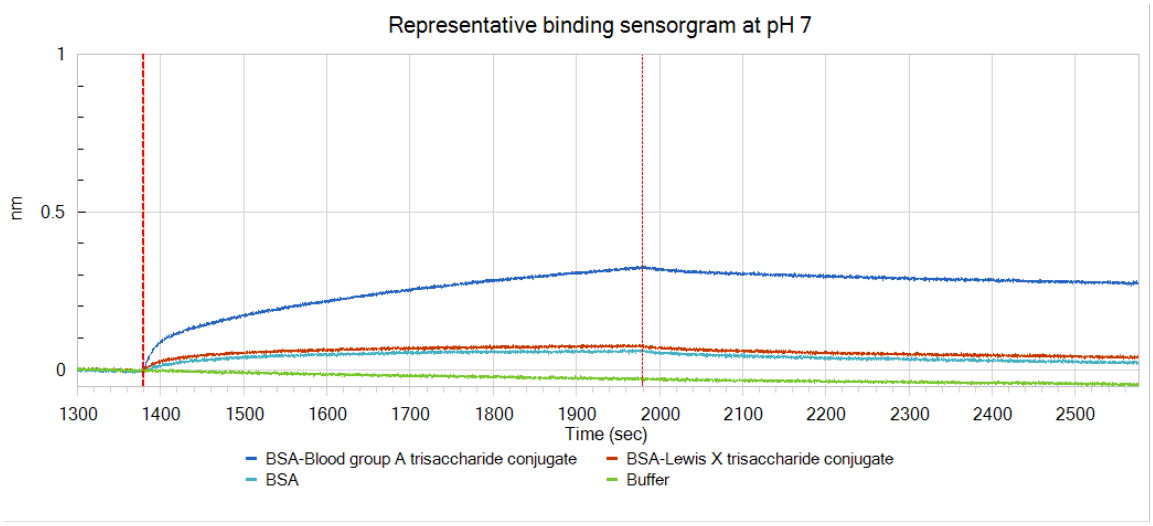
[b]Relative van der Waals overlap[2]

[c]Relative Fluorescence Units from the array screening

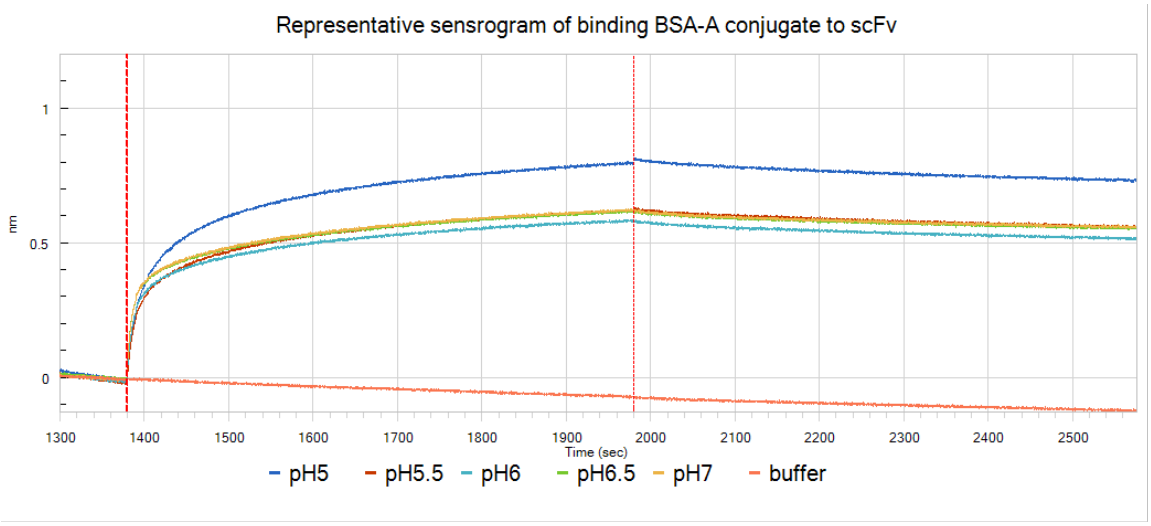
[d]Concentration in μg/ml



Supplementary Figure 1. Binding assay of BSA at various pH values. scFv-immobilized biosensor was dipped into 1 μM BSA at pH5 (blue), pH5.5 (red), pH6 (cyan), pH6.5 (green), pH7 (yellow), and buffer (orange). BSA showed no binding to scFv-immobilized biosensor at pH 6.5, and 7, but a relative small non-specific binding to scFv-immobilized biosensor at acidic pH 5, 5.5, and 6. Analysis buffer (reference in orange) did not display any binding at all.



Supplementary Figure 2. Representative binding sensorgram for BGA-conjugate (blue), Le^x-conjugate (red), BSA (cyan) and buffer (green) at pH 7. BSA-Le^x and BSA showed a similar signal.



Supplementary Figure 3. BLI binding assay of scFv to BSA-blood group A trisaccharide conjugate at pH5 (blue), pH5.5 (red), pH6 (cyan), pH6.5 (green), pH7 (yellow), and buffer (orange).

Supplementary References:

1. Nivedha, A. K.; Makeneni, S.; Foley, B. L.; Tessier, M. B.; Woods, R. J., Importance of ligand conformational energies in carbohydrate docking: Sorting the wheat from the chaff. *Journal of Computational Chemistry* 2014, *35* (7), 526-539.

2. Tessier, M. B.; Grant, O. C.; Heimburg-Molinaro, J.; Smith, D.; Jadey, S.; Gulick, A. M.; Glushka, J.; Deutscher, S. L.; Rittenhouse-Olson, K.; Woods, R. J., Computational Screening of the Human TF-Glycome Provides a Structural Definition for the Specificity of Anti-Tumor Antibody JAA-F11. *PLoS ONE* 2013, *8* (1), e54874.