UNIVERSITY OF CALIFORNIA SAN DIEGO

Glycomics platforms to investigate influenza binding interactions

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Chemistry

by

Taryn Marie Lucas

Committee in charge:

Professor Kamil Godula, Chair
Professor Michael D. Burkart
Professor Pascal Gagneux
Professor Michael Galperin
Professor Thomas Hermann

2021

The Dissertation of Taryn Marie Lucas is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2021

DEDICATION

I dedicate this dissertation to my family.
Thank you for all of your love and support
even as I was all the way across the country.

EPIGRAPH

The important thing is not to stop questioning.

Curiosity has its own reason for existing.

One cannot help but be in awe when he contemplates the

mysteries of eternity, of life, of the marvelous structure of reality.

It is enough if one tries merely to comprehend a little of this mystery every day.

Never lose a holy curiosity.

Albert Einstein

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SCHEMA

ACKNOWLEDGEMENTS

Receiving a PhD has been a tough but rewarding process. I started by moving across the country to a place where I knew no one. Luckily, I met some truly amazing people along the way. Here are just a few that I would like to acknowledge for their support and friendship. Sarah Zeng, my first roommate, was the first person I met in San Diego, and I am so thankful that we became friends. You eased the process of adjusting to life in San Diego.

During orientation, I met Ember who is now my best friend. She is hands down the strongest and most genuine person I have ever met. This year, especially, she has proven again and again how strong she truly is. Ember, you are my role model, and I admire you so much. Thank you for being the one to help me find my purse that I left behind at lunch. And thank you for making my life beautiful. I don't think that I would have made it through the program without your support. She was there through all the ups and downs, cheering me on along the way. She also introduced me to her family who became my San Diego family. I would attend holidays and birthday parties as another member of the "framily". Thank you, Nita, Michael, Alex, Lisa, Paige, Diora, Katie, and Alyson for being so inclusive and making the holidays that I couldn't go home for equally as special.

I would also like to thank my lab members, in particular Austen, Matt, Bryce, Chris, and Dan. We were some of the first members of Kamil's lab, and I think of you as my lab brothers. Thank you for all the scientific help and for just being really great friends.

Warren, I know that we really only became friends over the last year and a half or so, but I'm so glad you started inviting yourself on the weekly Scoops walks with Sarah. Thank you for listening to my practice talks and reading my application materials (even if

you made me cry), cooking me and Ember dinner, and changing your taste in beer. I would also like to Sven, Atzin, and Sarah Chang. Surviving the pandemic lockdown would have been so much more difficult if I didn't have such a great roommate and fantastic neighbors.

Lastly, I would like to thank my family for their love and support. Being so far away from you has been difficult these last six years, especially since I haven't been home since before the pandemic. But now, I will only be a three hour drive away so I can come visit with Freddie whenever I want.

Notes About the Chapters

Chapter two, in in full, is a reprint of the material as it appears Mucin-mimetic glycan arrays integrating machine learning for analyzing receptor pattern recognition by influenza A viruses in Chem Cell Press. Lucas, Taryn M.; Gupta, Chitrak; Altman, Meghan, O.; Sanchez, Emi; Nattichia, Matthew, R.; Gagneux, Pascal; Singharoy, Abhishek; Godula, Kamil., Elsevier, 2021. The dissertation author was the primary researcher and author of this material.

Chapter three, in part is currently being prepared for submission for publication. Lucas, Taryn M.; Brandenburg, Connor A.; Michalak, Austen L.; Godula, Kamil. The dissertation author was the primary researcher and author of this material.

Chapter four, in part contains coauthored/unpublished data. Lucas, Taryn M.; Robinson, Caressa M; Lewis, Nathan E.; Godula, Kamil. The dissertation author was a primary researcher and the primary author of this material.

VITA

Education

2015 Bachelor of Science, Chemistry and Biochemistry, West Chester University

2017 Master of Science, Chemistry, UC San Diego

2021 Doctor of Philosophy, Chemistry, UC San Diego

Publications

Lucas, T.M.; Gupta, C.; Altman, M.O.; Sanchez, E.; Naticchia, M.R.; Gagnuex, P.; Singharoy, A.; Godula, K. Mucin-mimetic glycan arrays integrating machine learning for analyzing receptor pattern recognition by influenza A viruses. *Chem Cell Press*. 2021 DOI:https://doi.org/10.1016/j.chempr.2021.09.015

de Rozières, C.M.; Pequeno, A.; Shahabi, S.; **Lucas, T.M.**; Godula, K.; Ghosh, G.; Joseph, S. PABP1 drives the selective translation of influenza A virus mRNA. *Submitted to J. Mol. Bio.*

Huang, M.L.; Tota, E.M.; **Lucas, T.M.**; Godula, K. Influencing early stages of neuromuscular junction formation through glycocalyx engineering. *ACS Chem Neurosci*. 2018, 3086-3093.

Naticchia, M.R.; Laubach, L.K.; Tota, E.M.; **Lucas, T.M.**; Huang, M.L.; Godula, K. Embryonic stem cell engineering with a glycomimetic FGF2/BMP4 co-receptor drives mesodermal differentiation in a three-dimensional culture. *ACS Chem Biol*. 2018, 2880-2887.

ABSTRACT OF THE DISSERTATION

Glycomics platforms to investigate influenza binding interactions

by

Taryn Marie Lucas

Doctor of Philosophy in Chemistry

University of California San Diego, 2021

Professor Kamil Godula, Chair

Glycans are one of the four major building blocks of life, along with nucleic acids, proteins, and lipids. These biomolecules are composed of monosaccharide sugar building blocks and are ubiquitously expressed by all animal, bacterial, and fungal cells. However, unlike RNA, DNA, and protein, the synthesis of this class of biomolecules is not directed by a template, making the study of glycans particularly cumbersome. Instead, each glycan's abundance changes in response to many stimuli, such as availability of

monosaccharide precursors, expression levels of enzymes required to elongate a glycan chain, and the amount of time a nascent glycan chain resides in the Golgi during synthesis and subsequent processing steps. There are approximately ten monosaccharides commonly found in human glycans that can be linked in a number of configurations to give hundreds, if not thousands, of unique structures within the glycome. Many of these glycans are attached to lipids or proteins as glycolipid or glycoprotein glycoconjugates that are presented above the cell surface in a forest of sugars known as the glycocalyx.

The biological relevance of glycans is becoming increasingly apparent. Not only are simple sugars, such as glucose, an important source of fuel for the cell, the glycans displayed at the cell surface have known implications in immunity, disease, cell signaling, and differentiation. Influenza A virus (IAV) is a well-studied pathogen that binds to sialic acid-containing glycans to gain entry into cells and initiate infection. Sialic acid frequently terminates *N*-and *O*-linked glycans on glycoproteins, as well as glycolipids that IAV can use as receptors. Because of the intrinsic heterogeneity of glycans, the exact molecular features defining the presentation of this viral receptor within the glycocalyx remain unknown. While there have been substantial advancements in methods to probe IAV binding in recent years, many of the investigations fail to recapitulate key aspects of the glycocalyx, such as the multivalency of glycans along a protein backbone and the heterogenous composition and arrangement of different sialylated glycoconjugates.

The following thesis chapters, describe, in detail, work I have done to improve current synthetic platforms for systematic examination IAV adhesion and release from its sialic acid receptors, as well as, establishing a method to study viral binding within the cellular glycocalyx.

1 Introduction

1.1 Motivation

Seasonal influenza A virus (IAV) infections pose an annual burden on the economic and healthcare systems across the world. Globally, influenza A is responsible for an estimated almost 400,000 respiratory deaths annually.[1] Influenza typically impacts the upper respiratory and can result in a myriad of symptoms including, fever, cough, sore throat, headache and body aches, weakness, and congestion.[2] Most infected individuals can make a full recovery at home, but a significant portion of the population, especially the elderly or those with comorbidities (e.g., heart disease or diabetes) develop pneumonia which requires hospitalization and can prove fatal.[2,3] Because current treatments lack long term efficacy as the virus evolves resistance, strategies for combatting the disease focus on prevention through vaccination,[4] which has its own limitations due to the guesswork involved in predicting the most prevalent viral strain in any given year.[5]

Influenza A virus belongs to the *Orthomyxoviridae* family and is comprised of eight segmented genes of negative sense single-stranded RNA[6] that encode ten structural and at least nine nonstructural/regulatory viral proteins.[7] Two of these glycoproteins, hemagglutinin (HA) and neuraminidase (NA), are expressed on the surface of the virus and are highly responsible for its pathogenicity. In fact, they are so important that they constitute the H and N used in the naming convention for the different influenza strains (H1N1 vs H3N2). HA is responsible for binding to sialic acid receptors on the host cell to begin the infection process, while NA cleaves the bond between sialic acid (Sia) and the following glycan in the chain (galactose, Gal) to allow the virus to move past decoy receptors found in mucus and detach from a newly infected cell.

A total of eighteen HA and eleven NA subtypes that exist in nature,[7] but only a few combinations circulate within human populations. They are H1N1, H2N2, and H3N2.[8] Avian species are asymptomatic reservoirs for IAV strains where the HA binds predominantly to Sia linked in an α2,3 configuration to Gal. Humans typically carry Sia linked by an α2,6 bond to Gal which results in a different three-dimensional structure of the receptor. As such, mutations in the HA protein are necessary for an avian virus to accommodate the different receptor structure and bind to human cells. It has been shown that as little as one mutation can effect a change in binding preference from an α2,3 to an α2,6 linked Sia.[9]

Unfortunately, IAV is constantly evolving to increase its infection potential and there exists an enormous natural reservoir in wild birds. The RNA polymerase used during viral replication is error prone and introduces mutations into the IAV genome as an evolutionary mechanism in what is known as antigenic drift. In addition to changing the binding preference of a particular virus, these mutations can also add a glycosylation site on the HA which can shield what was previously recognized by antibodies, leading to an immune evasive IAV strain. Antigenic shift is another way the virus can gain a foothold in infecting new hosts. This process can occur when two viral strains infect the same host. Because IAV has a segmented genome, recombination can occur to create a virus that expresses proteins from multiple viruses. Antigenic shift quickly produces a distinct new virus and is the major mechanism by which new pandemic strains are generated.[10]

The World Health Organization (WHO) is continually conducting surveillance on IAV to monitor strains capable of crossing from birds to humans. Current avian strains with zoonotic potential include H5N1,[11] H7N9,[12] H7N7,[13] and H9N2.[14] These viruses can

infect humans that have sustained contact with poultry but have yet to acquire the required mutations for respiratory transmission between humans. Along with tracking changes in the viral genome, the WHO and CDC measure differences in viral binding phenotype to prepare for the possibility of disease outbreak.

1.2 Glycan structure and symbol nomenclature

Much of this thesis contains references to glycans and their structures in relation to IAV binding and infection. Glycans may sometimes be drawn using symbol representation. All symbols used throughout this thesis will follow the accepted Symbol Nomenclature for Glycans (SNFG).[15] Below you can find the common symbols that will be used throughout this text (**Fig 1.1A**). In a glycan chain, the monosaccharide building

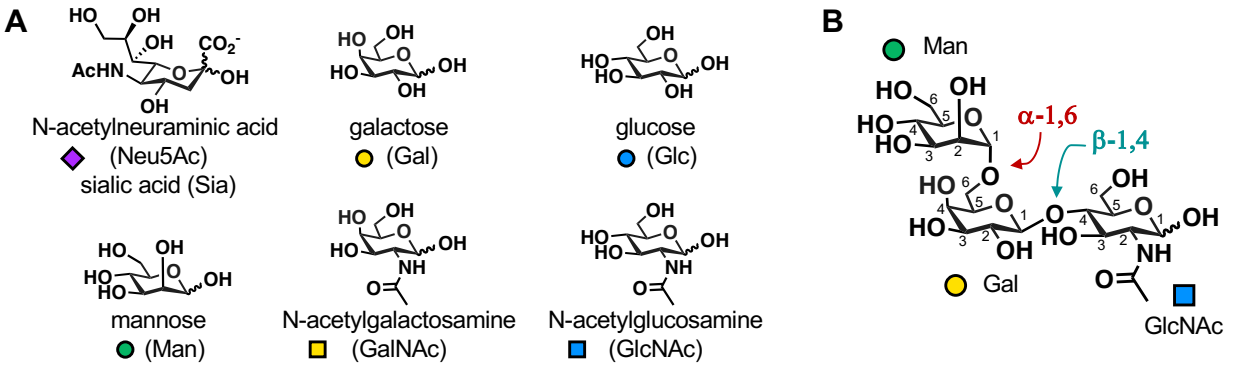


Figure 1.1 Glycan structures and symbols. A) N-acetylneuraminic acid (Neu5Ac, sialic acid, Sia) is denoted by a purple diamond, galactose (Gal) by a yellow circle, and glucose (Glc) by a blue circle. If monosaccharides contain an N-acetyl group at their 2-position, they are symbolized with squares (yellow for N-acetylgalactosamine, GalNac and blue for N-acetylglucosamine, GlcNAc). The squiggly line indicates that these glycans can exist as either the α or β anomer. B) To name the linkage that exists in a glycan, the carbons are numbered starting at the anomeric position and used to denote the hydroxyl groups used for forming bonds, as shown for Man (α-1,6) Gal and Gal (β-1,4) Glc.

blocks are attached through their various hydroxyl (-OH) groups in either an alpha (α) or beta (β) linkage. The type of linkage, along with the carbon number through which the bond is formed is used to name the glycan structure (**Fig 1.1B**).

The term glycan can refer to simple monosaccharides or complex polysaccharide structures containing many sugar residues stitched together (**Fig 1.2**). Because each monosaccharide has many hydroxyl groups through which it can be attached to another sugar moiety, the structural diversity of this class of biomolecules is very great, making them particularly difficult to synthesize and/or characterize.

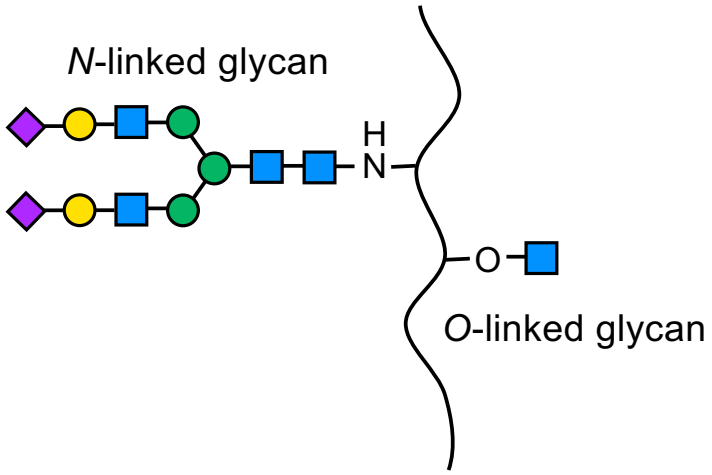


Figure 1.2 Structural diversity of glycans shown here attached in N- or O-linked linkage to a protein backbone, black line. The term glycans can refer to a simple monosaccharide, as in the case of *O*-GlcNAc, or longer chains of sugars linked together as shown for *N*-glycans that often display additional complexity due to various branch points.

1.3 Methods to investigate IAV binding

1.3.1 Hemagglutination and hemagglutination inhibition assays

In the early 1940s, virologist George Hirst discovered red blood cells (RBCs) agglutinate in the presence of allantoic fluid from eggs infected with influenza virus.[16] The reason why this occurs is because RBCs carry numerous copies of sialic acid on their cell surface that IAV can bind to and essentially crosslink the RBCs. During the late 1940s and into the next decade, scientists began looking for ways to inhibit this interaction by looking at electrolyte concentrations,[17] antisera,[18] urine,[19] carbohydrates,[20] and

mucoproteins.[21] Interestingly, when the virus-RBC aggregates were heated to 37 °C, the aggregation ceased, leading Hirst and others to hypothesize that there was a "receptor-destroying enzyme" which is known today as neuraminidase (a sialidase). Solution-based assays and colorimetric tests first characterized the released receptor as a carbohydrate-peptide conjugate containing an *N*-acetyl group,[23] however, it was later crystalized, and the structure was determined to be *N*-acetyl neuraminic acid,[24] commonly referred to as sialic acid.

In fact, treating RBCs with sialidase enzyme to remove sialic acid from the surface, abolishes IAV-induced agglutination, as shown by Paulson and Rogers[25] in their 1983 study where they also reintroduced sialic acid using sialyltransferase enzymes. They used two types of sialyltransferase enzymes: one to add the sialic acid in an α2,3-linkage to the underlying glycan and one to add it through an α2,6-linkage. By comparing the agglutination ability of viruses isolated from different species to these newly re-sialylated RBCs containing only one isomer, Paulson and Rogers introduced the commonly held belief that avian viruses prefer α2,3-linked sialic acid receptors and viruses able to infect humans predominantly recognize the α2,6 sialic acid linkage. Interestingly just this year, the Boons lab published a similar de-sialylation/re-sialylation approach in conjunction with enzymatic extension of polyLacNAc repeats on biantennary N-glycans to characterize recent H3N2 viruses that have lost the ability to agglutinate unmodified turkey and guinea pig RBCs without extended *N*-glycans[26] which was originally noted by the Paulson lab.[27]

A major reason why the HAs of these recent H3N2 strains lost the ability to agglutinate RBCs is due to the fact that HA has a relatively low affinity for sialic acid (~2-3 mM).[28] However, the cell surface is coated with millions of sialic acids and each virion

contains hundreds of copies of HA protein, so each low affinity interaction is enhanced through multivalency (**Fig 1.3**).

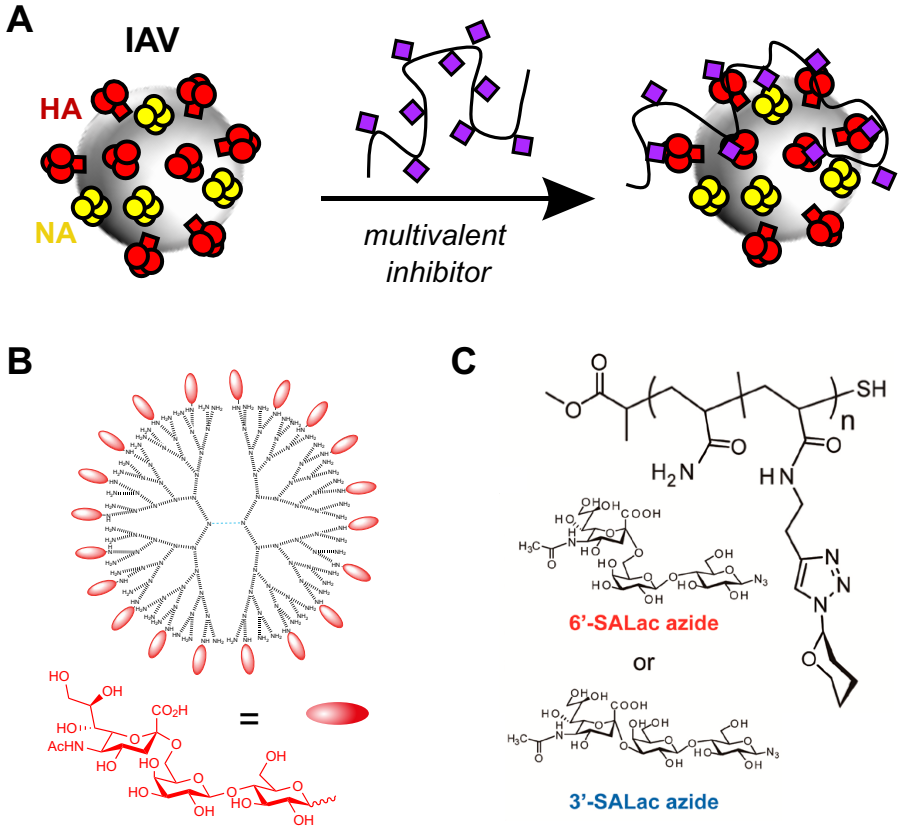


Figure 1.3 Multivalency in IAV binding interactions. A) Multivalent IAV inhibitors containing sialic acid enhance the relatively low affinity between a single HA subunit to sialic acid by binding to many HAs on the viral coat or to multiple binding sites within individual HA proteins. B) The dendrimer created by the Linhardt Lab is a high affinity inhibitor of IAV because the sialic acid moieties are spaced for optimal binding to all three binding sites within an HA trimer. C) The linear polymer synthesized by the Miura Lab inhibits IAV binding by interacting with multiple HAs on a virion.

Many labs have utilized multivalent constructs to inhibit IAV binding to the cell surface, and thus, prevent infection. The Whitesides group has done a lot of the initial work investigating polyacrylamide polymers containing sialic acid and the cooperativity of the multivalent systems.[29,30] Additionally, the Linhardt Lab synthesized a library of glycodendrimers with altered spacing between the terminal α2,6-sialyllactose and uncovered that the conjugate with glycans 3.1 nm apart showed the best inhibitory potential *in vitro* and *in vivo* because it most closely matched the spacing of the sialic acid

binging sites within the HA trimer.[31] Interestingly, work by the Haag group indicates that linear glycomaterials are better inhibitors than spherical dendrimers,[32] most likely through a steric shielding mechanism originally proposed by Whitesides.[29,30] While these inhibitors have provided insights into the binding modes of IAVs, unfortunately they have not been useful therapeutically because of the high concentrations required for successful inhibition.

1.3.2 Glycan microarrays

Due to the great successes that arose from DNA microarrays in the mid to late 1990s,[33] the glyco-community engineered their own versions containing synthetic and natural glycan structures immobilized onto glass slides, which they logically termed glycan microarrays.[34] In these arrays, the glycans are typically functionalized with a handle at their reducing end that can be used for immobilization onto a slide surface. The spots printed in the arrays are less than a few hundred microns in diameter and each slide can contain greater than five hundred spots printed in replicates, making this method well suited for high throughput screening of potential interactors with glycans.

Over the years, many labs generated their own array technology[35–38] but the one created by the Consortium for Functional Glycomics (CFG) has received the most fame. The CFG allows researchers to submit samples for testing on their array which contains over 600 glycan architectures.[39] Fluorescently labeled glycan binding proteins, or lectins, with binding preferences for distinct glycans are first used to ensure that the samples immobilized in a way that afforded recognition by the lectins.[40] The extent of binding can also be quantified through each spot's fluorescent intensity. Once the lectin specificity

was established, growth factors, serum, bacterial and viral proteins, and even whole microbes or eukaryotic cells have been tested on the array.

The CGF array has been used many times to test binding of HA proteins and whole virus. It was used to show that there were geographic differences in binding preferences of recombinantly expressed HAs that were circulating during the 1918 influenza pandemic.[41] It has also been used to investigate how viruses can cross species for 2009 H1N1 pandemic strain and the swine virus from which it was derived.[42] Even more recently, it has been employed to examine binding of an avian H5N1 strain capable of infecting humans. In this study, the researchers also mapped how HA mutations changed the binding behavior of the virus.[43,44] These studies, along with others, show agreement with the canonical human and avian receptor specificity, but there are caveats to drawing premature conclusions from array binding data. For example, when used to test the evolution of H3N2 binding preference since 1968, the initial conclusion was that some of the more recent human infective viruses preferred the avian-type receptor.[45] In actuality, the recent H3N2 viruses strongly bind the human-type receptor when it is presented in a bivalent manner on chains with extended LacNAc (disaccharide) repeats,[27] structures that were not included in the initial array screens. Furthermore, no correlation of infection potential was found when binding of human and avian influenza viruses were tested across multiple array formats and then compared to infection of live tissues.[46]

1.3.3 Glycomic screens and genetic engineering approaches

The chemical synthesis of glycans remains challenging due to the complexity of structures present within biological systems and the need for syntheses that maintain

sterio- and regio- chemical control. Additionally, isolating pure and well characterized glycans from natural sources is difficult due to each individual structure's low abundance. As such, the glycan arrays discussed above contain only a fraction of endogenous structures, and each presented in isolation.

To get a better indication of biologically relevant glycans in IAV infection, labs have begun to undertake glycomic analysis of tissues known to become infected with IAV. A comprehensive study that released the *N*- and *O*- linked glycans, as well as glycosphingolipids, from human lung samples for subsequent mass spectrometry analysis has expanded the knowledge of potential IAV receptors present in this tissue.[43] *N*-glycan analysis indicated a large variety of multi-antennary structures, with those containing polylactosamine extensions terminated with α2,3-linked sialic acid, and those without the extension carrying predominantly the α2,6-linked terminal modification. The *O*-glycan structures were mainly restricted to mono- and di-sialylated core 1 type structures containing both sialic acid linkages. Lastly, the glycosphingolipids expression showed sialylation on linearly elongated structures. Interestingly, tissue staining revealed that IAV could still bind lung when the sialic acid was removed through neuraminidase treatment, indicating the presence of a non-sialylated receptor for IAV adhesion. This additional receptor is most likely phosphorylated based on a shotgun glycomics analysis from the Cummings Lab where *N*-glycans were released from human lung, pooled into fractions based on HPLC separation, labeled with a bifunctional linker for immobilization onto a microarray, and then probed with IAV.[44]

The shotgun approach ideally requires access to multiple tissue samples for replicate analysis since glycan expression can vary greatly among individuals based on

age, disease status, and other factors. To circumvent this, the Clausen Lab has pioneered the creation of a cell-based glycan array through the differential knocking out and knocking in of specific genes within the glycosylation pathway.[45] This approach allows for the tuning of glycan expression based on class (i.e., *N*-linked, *O*-linked, and glycolipids), the amount of extension, and degree of sialylation and fucosylation. The glycans are also kept in their native presentation, whether that be attached to proteins or lipids. Testing of IAV binding to the various libraries has shown a preference for human IAVs to bind sialic acid on *N*-linked glycans and glycolipids, while viruses of avian origin can also bind to *O*-glycans.

This same lab recently employed their genetically engineered cell lines to focus attention on mucin glycoproteins.[46] This class of heavily *O*-glycosylated proteins can be attached to the cell surface or secreted as a main component of mucus. By differential knock out/knock in strategies for genes involved in mucin biosynthesis, they created a cell-based and secreted mucin library with tunable numbers of glycosylation domains that can be occupied by a variety of the mucin *O*-GalNAc architectures that are often heavily sialylated. Binding and competition studies revealed that IAV preferentially binds to mucins containing core2, and not core1, *O*-glycans.

1.4 Drawbacks to these approaches and project goals

While these approaches have provided useful information regarding IAV interactions, there are some limitations or drawbacks to each method. For instance, mammalian RBCs are not naturally infected with virus (given that these lack a nucelus),

so any of the hemagglutination inhibition assays have not been done in biologically relevant cell line. Additionally, the glycocalyx composition of RBCs51 is also not representative of structures in the lung or airways.[47] Microarrays also do not recapitulate the diverse set of glycans found in human glycome or their varied, mixed, and multivalent presentation, which is most likely why they are not good predictors of IAV infection.46 Lastly, glycomic analyses and genetic screening are both time consuming and require very specialized skillset.

The next chapter details the creation of a new microarray platform that contains some of the key structural parameters that are missing on conventional microarrays and may be responsible for their poor prediction of infection. It utilizes mucin-mimetics that can be synthesized in modular approach and tuned for valency to display IAV receptors. The fluorophore label included in the mimetic design allows for quantification of the immobilized receptors in the array. Like the CFG array, there is the potential for commercialization of this array. Work is ongoing to introduce a neuraminidase probe into an array of these mucin mimetics, as outlined in chapter three. Finally, chapter four describes current efforts to observe IAV binding in human alveolar cells and use mass spectrometry to identify protein receptors.

2 Mucin-mimetic glycan arrays integrating machine learning for analyzing receptor pattern recognition by influenza A viruses

2.1 Introduction

The periodic emergence of new respiratory viruses capable of spreading across the human population continues to exact a significant toll on human life and the global economy. The novel coronavirus, SARS-Cov2, which is responsible for the ongoing global COVID-19 pandemic,[53] provides a stark example of the risks of zoonotic virus adaptation to our society. Other animal pathogens, such as avian Influenza A viruses (IAVs), continuously pose a threat of crossing to human hosts and require close monitoring.[54] Many respiratory viruses, including IAVs, utilize specific glycan receptors on airway epithelial cells to initiate entry and replication.[55] Characterization of the glycan-binding phenotype of IAVs may provide an early indicator of increased infection potential.[56–58]

IAVs carry two types of glycoproteins in their viral coat with specificity for terminal sialic acid modifications on cell surface glycans – the receptor-binding hemagglutinins (HAs) and the receptor-destroying neuraminidases (NAs).28,59,60 The configuration of the sialic acid glycosidic bond linkage to the underlying glycans in glycoproteins and glycolipids plays a prominent role in defining IAV host specificity (**Fig 2.1A**). According to the prevailing paradigm[25,56] avian viruses preferentially recognize α2,3-linked sialic acids abundant in the gastrointestinal tract of birds, while human IAVs have affinity for α2,6-sialosides expressed on lung epithelial cells in our upper airways. A switch in HA

specificity from α2,3- to α2,6-linked sialic acids is associated with increased infection and transmission in humans.[9,54,61]

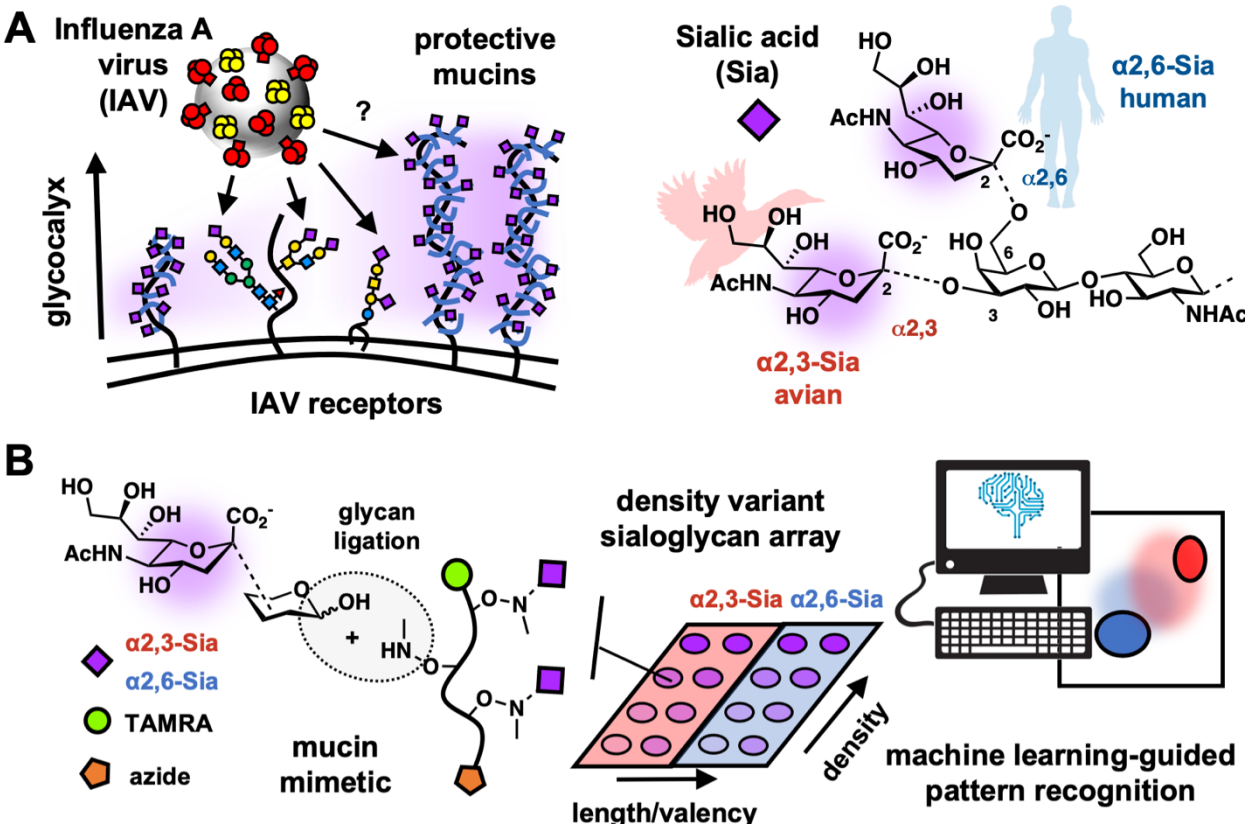


Figure 2.1 Machine learning-enabled glycomimetic array platform for assessing receptor pattern recognition by influenza A viruses (IAVs). A) IAVs begin their infection cycle by binding to sialylated host glycans, but these receptors are also present on mucins which have a proposed protective function. Avian and human IAVs show distinct preferences for the binding of α2,3- and α2,6-sialoglycan receptors. B) Glycopolymers, which mimic the architecture and composition of mucins, were used to build models of the mucinous glycocalyx on microarrays. A support vector machine (SVM) learning algorithm enabled analysis of viral binding response to changing receptor patterns in the synthetic glycocalyx arrays.

Glycomics screens[27,46] and cell-based studies using glycosylation mutants[49,62] have revealed that, in addition to a particular sialic acid linkage configuration, IAVs can also discriminate between distinct glycan classes and glycoconjugate types (i.e., *N*- and *O*-glycosylated proteins and glycolipids). Spatial combinations of these sialylated

glycoconjugates give rise to three-dimensional, hierarchically organized receptor patterns in the host cell glycocalyx that determine the specificity and avidity of IAV binding. Glycan arrays, which present ensembles of chemically defined glycans printed and immobilized on glass substrates, are routinely used to analyze viral HA-receptor specificity.[41–43,45,63] However, a recent cross-comparison between glycan composition of *ex vivo* human lung and bronchus tissues with glycan array binding data pointed to a limited ability of the arrays to predict infection events.[46] This indicates that the current platforms may not accurately recapitulate the receptor presentation in the glycocalyx environment as encountered by viruses at the mucosal epithelium.

The mucosal epithelial cell glycocalyx is dominated by membrane-tethered mucins (MUCs), which are large, heavily glycosylated proteins projecting tens to hundreds of nanometers above the cell surface (**Fig 2.1A**).[64,65] Mucins carry primarily, but not exclusively, *O*-glycans linked to tracks of serine and threonine residues within the core protein. As much as 80% of mucin mass derives from glycans, giving these glycoproteins an extended semi-flexible bottlebrush form.[66] The *O*-glycans in mucins are frequently terminated with sialic acids that can serve as IAV receptors;[67] however, epithelial mucins are believed to primarily provide protection against infection. Mucins can serve as decoys, which shed from the cell surface upon virus binding,**Error! Bookmark not defined.** or assemble into dense extended glycoprotein brushes that restrict virus access to apical membrane receptors and interfere with internalization.[68,69] Due to the prominence of these extended glycoproteins within the glycocalyx and their extensive modification with sialic acids, mucins are most likely the first, and likely non-productive, site of virus attachment in its infection cycle.[70] Interestingly, the IAV subtype, H1N1, was found to colocalize with some

(i.e., MUC1) but not other (i.e., MUC13 and MUC16) mucins on the surfaces of A549 lung epithelial cells,[71] revealing a preference of the virus for distinct mucin family members within the same cell and produced by a shared glycosylation machinery. The type of mucin and its presentation at the cell surface is likely to influence initial viral interactions at the epithelium and determine the course of infection. A more complete understanding of IAV interactions at the mucosal glycocalyx may, thus, provide a more accurate assessment of the potential of IAVs to infect human hosts.

Here, we report the development of an array platform, in which synthetic mucin-mimetics are used to model the mucosal epithelial cell glycocalyx, to evaluate receptor pattern recognition by IAVs (**Fig 2.1B**). We applied support vector machine (SVM) learning to identify and analyze effects of variations in glycan receptor type, mucin mimetic valency, nanoscale dimensions, and crowding in the glycocalyx models on shifts in the binding specificity of H1N1 and H3N2 IAV strains. We found that mucin-like polyvalent presentations of α2,3- and α2,6-sialoglycans and the surface crowding of the glycoconjugates differentially impacted adhesion of the viruses, consistent with the proposed protective functions of mucins in the airway epithelium. The mucin mimetic arrays also revealed an evolution of receptor pattern recognition by H1N1 produced in avian or mammalian cells, which could be characterized through machine learning.

2.2 Construction of glycopolymers for mucin-like glycan receptor presentation

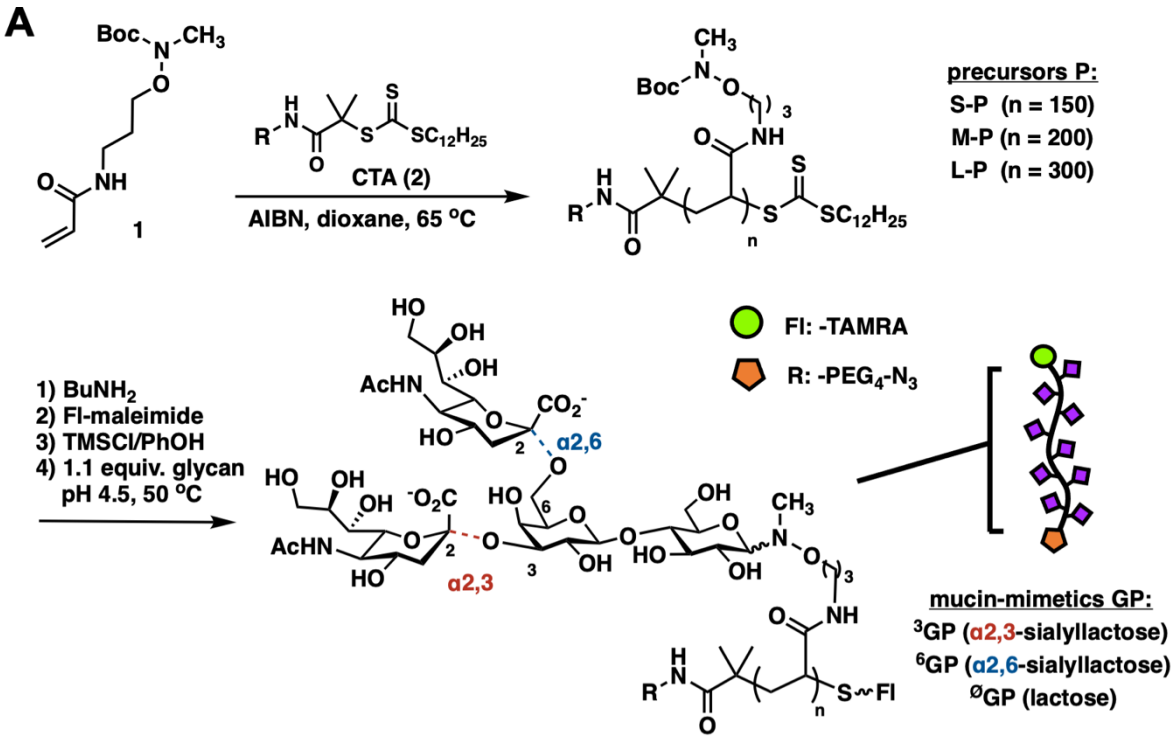
To model the mucinous glycocalyx environment in glycan arrays, we have devised a method for generating synthetic glycopolymers (**GPs**) that replicate key structural features of mucins (i.e., polyvalent glycans displayed along extended linear polypeptide

chains) while allowing for tuning of the mimetic size and glycosylation pattern (**Fig 2.2A**). Using the reversible addition-fragmentation chain transfer (RAFT) polymerization, we have generated a collection of mucin mimetics of increasing length glycosylated with α2,3- and α2,6-sialyllactose trisaccharides (**α2,3-SiaLac** and **α2,6-SiaLac**) as model avian and human IAV receptors, respectively. The polymers were terminated with an azide functionality and used either as soluble probes or covalently grafted on cyclooctyne-coated glass via the strain-promoted alkyne-azide cycloaddition (SPAAC) reaction to produce a mucin-like glycocalyx display.[72] A tetramethylrhodamine (TAMRA) fluorophore was appended to the opposing chain end to allow for characterization of mucin mimetic density on the arrays.

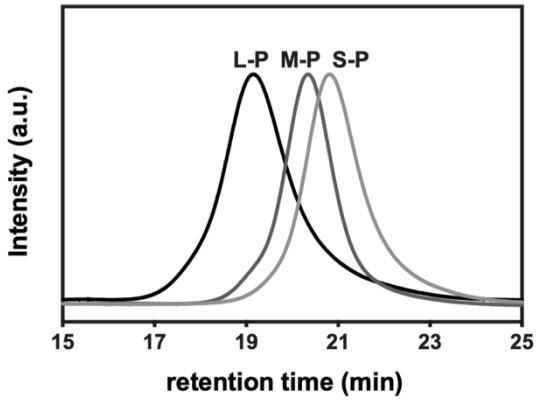
The mucin mimetic synthesis began with the polymerization of a Boc-protected *N*-methylaminooxypropyl acrylamide monomer (**1**) in the presence of a chain transfer agent (CTA, **2**) and a radical initiator (AIBN) to generate a set of azide-terminated short (**S**, DP ~ 150), medium (**M**, DP ~ 200) and long (**L**, DP ~ 300) polymeric precursors, **P** (**Fig 2.2A**). Size exclusion chromatography analysis (SEC, **Fig 2.2B** and **Table 2.2**) confirmed good control over the target molecular weight (Mw) and dispersity (Đ) of the polymers. Next, the trithiocarbonate end groups in polymers **P** were removed by aminolysis and the newly exposed thiol groups were capped with TAMRA-maleimide (**Fig 2.2A** and **Scheme 2.1**). The fluorophore labeling efficiency was determined for each polymer by UV-VIS spectrometry and ranged between 6-30% (**Tables 2.3**). Side-chain Boc-group deprotection in the presence of phenol and trimethylsilyl chloride (TMSCl)73 followed by conjugation of the released *N*-methylaminooxy groups with reducing glycans under acidic conditions completed the synthesis of the mucin mimetic glycopolymers **GP** (**Fig 2.2A**)

The mucin mimetic library was comprised of 27 short (**S**), medium (**M**) and long (**L**) sialylated glycopolymers, **^{3}GP** and **^{6}GP**, decorated with increasing amounts of the trisaccharides, **α2,3-SiaLac** and **α2,6-SiaLac**, respectively (**Fig 2.2C** and **Table 2.3**). In addition, we generated 11 control polymers lacking sialic acid modifications (**^{ø}GP**) displaying only the lactose disaccharide (**Lac**, **Fig 2.2C** and **Table 2.3**). The extent of glycosylation for all polymers was determined by ^{1}H NMR spectroscopy (**Data S1**) and varied according to glycan structure. Treatment with 1.1 equiv. of glycan per polymer sidechain was sufficient to achieve maximum polymer glycosylation of ~ 70% for **Lac** and ~ 45% for the negatively charged **α2,3-SiaLac** and **α2,6-SiaLac** (**Table 2.3**). The use of sub-stoichiometric amounts of glycans enabled tuning of glycan valency in the mucin mimetics (**Fig 2.2C** and **Fig 2.7**). The mucin mimetic lengths (*l*) were estimated to range from ~ 8 nm to 12 nm according to their DP using a method by Miura *et al*. for calculating theoretical end-to-end distances in sialylated glycopolymers (**Fig 2.2B** and **Equation 2.1**).[74]

In the airways, cell surface-associated mucins are organized into a dense, brush-like glycocalyx, which projects tens to hundreds of nanometers above the epithelial cell surface.[65] To gain insights into glycan receptor recognition by proteins and pathogens at the mucosal interface, we modeled the mucinous glycocalyx in arrays by printing mucin mimetic glycopolymers **GP** on cyclooctyne-functionalized glass (**Fig 2.3A**). In addition to varying the structure and glycosylation of the glycopolymer probes, we also modulated their surface crowding by increasing their concentration (c_{GP}) from 1 to 10 μM in the printing buffer (PBS supplemented with 0.05% Tween-20, pH = 7.4). The fluorescent TAMRA labels introduced synthetically into the probes were used to establish their



B



Pol.	*DP*	*Mw [kDa]*	*Đ*	*l [nm]*
S-P	150	39.2	1.28	8
M-P	200	54.0	1.17	9
L-P	300	98.1	1.36	12

C

glycan	*GP*	*valency [v]*
α2,3-SiaLac	S-^{3}GP$_{v}$	50, 85
	M-^{3}GP$_{v}$	50 – 110
	L-^{3}GP$_{v}$	110, 140
α2,6-SiaLac	S-^{6}GP$_{v}$	45, 65
	M-^{6}GP$_{v}$	50 – 105
	L-^{6}GP$_{v}$	105, 140
Lac	S-^{Ø}GP$_{v}$	75, 110
	M-^{Ø}GP$_{v}$	85 – 145
	L-^{Ø}GP$_{v}$	85, 170

Figure 2.2 Generation of mucin-mimetic probes. A) Fluorescently labeled azide-terminated short (S), medium (M) and long (L) mucin-mimetic glycopolymers GP were generated via RAFT polymerization. B) Size exclusion chromatography (SEC) analysis of the polymeric precursors P. C) The naming scheme for the GPs indicates the polymer backbone length (S-, M-, and L-), sialic acid linkage type (superscripts 3 and 6, and Ø designate α2,3-SiaLac, α2,6-SiaLac, and Lac, respectively), and glycan valency (final subscript). The oligomeric plant lectins, WGA and SNA, show distinct binding behavior in mucin-like receptor displays.

surface grafting efficiency (**Fig 2,9A**) and the overall glycan receptor density (**Fig 2.9C**) for each polymer condition. The printing conditions yielded spots of uniform morphology (**Fig 2.3A**) with linear increase in polymer density across the employed concentrations regardless of polymer size or glycosylation (**Fig 2.10** and **Fig 2.10**).

To confirm selective recognition of the mucin mimetics based on to the structure of their pendant glycans, the arrays were probed with Dylight649-labeled lectins wheat germ agglutin (WGA) and *Sambucus nigra* agglutinin (SNA) (**Fig 2.3A** and **Fig 2.11**). WGA, which primarily recognizes GlcNAc but has often been used to indicate the presence of α2,3-linked sialic acids, is specific for **α2,3-SiaLac** polymers on our arrays,[75,76] while SNA binds exclusively to the polymers containing the α2,6-linked isomer.[77] To obtain quantitative assessment of lectin binding to the mucin mimetics, we probed the arrays with increasing concentrations of the lectins to establish binding isotherms and extract apparent surface dissociation constants ($K_{D,surf}$) (**Fig 2.3B** and **Fig 2.12**). WGA binding to the medium sized **α2,3-SiaLac** polymers, **M-^{3}GP$_{50\text{-}110}$**, printed at low surface density (c_{GP} = 1μM) showed valency-dependent binding with autoinhibition at the highest valencies caused by glycan crowding on the polymer backbone. This behavior is frequently observed for lectin binding to glycopolymer probes in solution.[78] The low polymer printing concentration produced probe spacing on the array surface that allowed for the measurement of lectin binding responses to the underlying glycoconjugate architecture. Increasing the concentration of the polymers resulted in denser mucin mimetic arrays, attenuated WGA to **α2,3-SiaLac** valency of the individual probes, and increased overall avidity of the dimeric lectin toward the receptor display (**Fig 2.13** and **Table 2.4**). Our attempts to establish similar binding profiles for SNA were not successful

due to protein aggregation at concentrations needed to reach saturation binding (**Fig 2.14**).

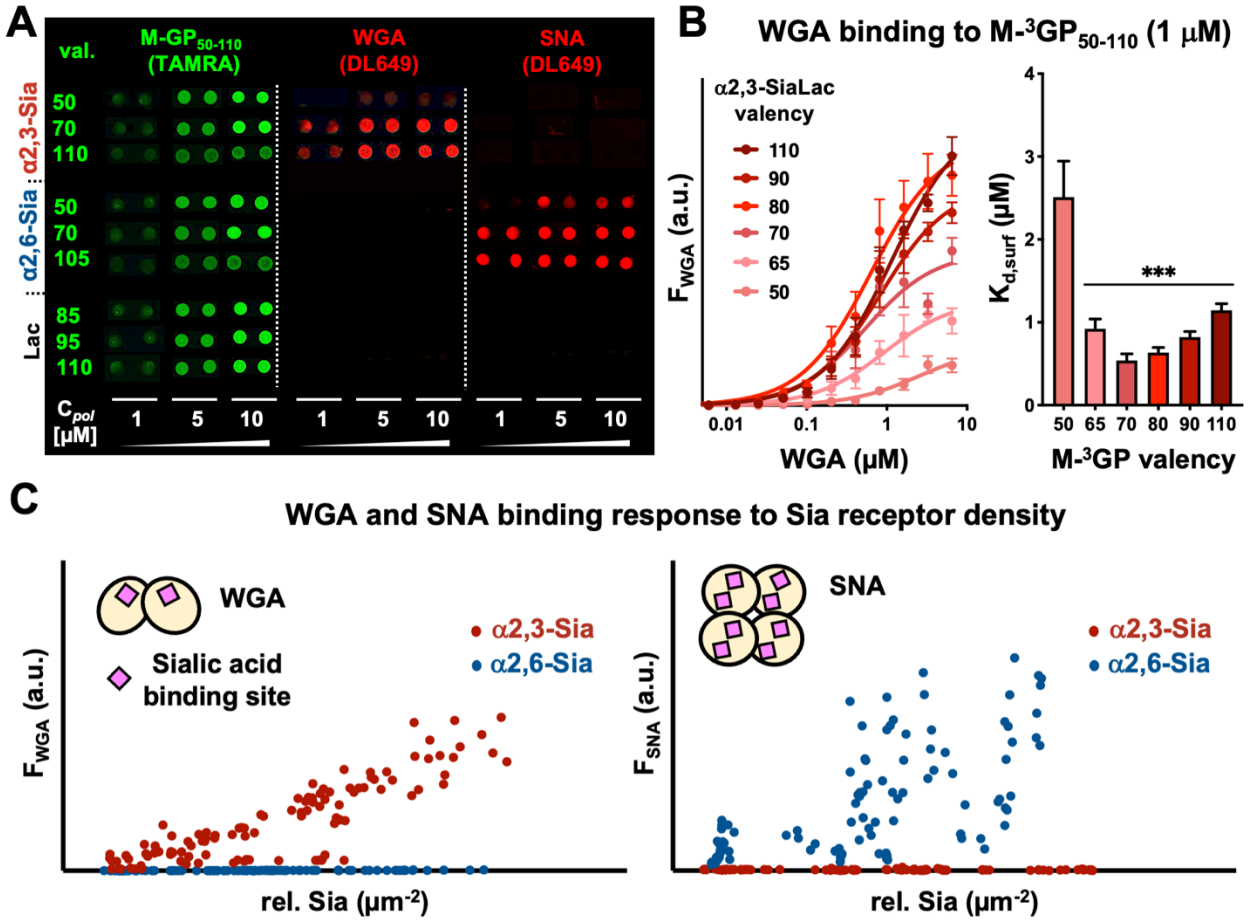


Figure 2.3 Construction and validation of mucin-mimetic arrays. A) Representative composite images of density-variant arrays of fluorescent mucin-mimetic glycopolymers GP (TAMRA, green) probed with Daylight649-labeled SNA and WGA lectins. Each condition is represented as a duplicate. Full array scans are provided in Fig S5. B) Binding isotherms and associated apparent surface dissociation constants (KD,surf) for binding of WGA to medium-sized mucin mimetics M-^{3}GP$_{50\text{-}110}$ with increasing α2,3-SiaLac valency printed at low surface density (cGP = 1 μM, ***p <0.0005 or greater). C) Binding responses of WGA and SNA to increasing glycan receptor density on the array. The dimeric WGA lectin binding is directly proportional to glycan density, whereas the tetrameric SNA lectin exhibits a more complex binding pattern. Insets represent graphical representation of lectin oligomeric state and orientation of sialic acid binding sites based on crystallographic data analysis (Figs S8 and S9).

Collecting thermodynamic binding data for each lectin-probe combination in the array can be time consuming and may not be possible for some lectins, as was the case for SNA. Simplified plots of lectin binding in response to changing relative sialoglycan

density in the arrayed glycopolymer spots provide a convenient way to discern different binding modes of the proteins. Using this analysis, we observed that the binding response of WGA to changing glycan density was generally linear, while SNA showed a less correlated binding pattern indicative of contributions from higher-order binding interactions, such as crosslinking of neighboring glycopolymers on the array (**Fig 2.3C**). Analysis of crystallographic data for WGA and SNA provide a structural basis for their differences in crosslinking capacity (**Fig 2.15** and **Fig 2.16**). WGA exists as a dimer with two sialic acid binding domains separated by 3.9 nm and positioned on the same face of the protein.[76] This arrangement reasonably favors WGA binding to glycans presented on the same mucin mimetic and may be responsible for the largely linear relationship between receptor density and lectin binding response. By contrast, SNA can exist as either a monomer, dimer, or tetramer.[79] Each monomer contains two glycan binding sites that are directed outward on opposite the edges of the protein.[80] The various oligomeric states and the orientation of the binding sites make SNA more likely to engage and crosslink multiple glycoconjugates on the surface, producing the more complex binding behavior observed on the mucin mimetic arrays.

2.3 H1N1 PR8 shows linkage-specific differences in binding to mucin-like sialoglycan presentations

Pathogens, which utilize oligomeric lectins and adhesins for binding to cell-surface glycan receptors, may be sensitive to the presentation of glycan receptors at the mucosal barrier.[81] We examined the binding of the H1N1 (A/Puerto Rico/8/1934 or PR8) virus to different presentations of sialoglycan receptors in our mucin mimetic platform (**Fig 2.4A**).

The PR8 strain is a well-characterized, laboratory-adapted human IAV strain, which has the ability to recognize both avian and human sialic acid receptor structures.[82,83] As such, it provides a useful model for assessing how receptor presentation may affect viral binding and selectivity.

H1N1, which was propagated in embryonated chicken eggs and henceforth labeled as H1N1 EGG, bound both receptor types in their soluble monovalent form, with ~ 4-fold preference for the **α2,3-SiaLac** isomer, as determined in red blood cell (RBC) agglutination inhibition assays ($K_{i,\alpha\mathbf{2,3}}$ = 13 μM vs $K_{i,\alpha\mathbf{2,6}}$ = 50 μM, **Fig 2.4B**, **Fig 2.17**, and **Table 2.5**). The array binding data mirrored this preference, while providing additional insights into the effects of receptor presentation on viral interactions (**Fig 2.4C-F**). H1N1 EGG virus binding to the medium size **α2,3-SiaLac** mucin mimetics **M-^{3}GP$_{50\text{-}110}$** immobilized at low surface densities (c_{GP} = 1 μM) indicated enhanced viral capture with increasing receptor valency, with a valency threshold for binding above 50 **α2,3-SiaLac** residues and a plateau at ~ 80 glycans per polymer (**Fig 2.4D** and **Fig 2.18**). Shortening the polymer length while maintaining a high receptor valency above 80 (**S-^{3}GP$_{85}$**) had no negative effect on viral capture (**Fig 2.4E** and **Fig 2.19**). We observed some decrease in binding to the longest mucin mimetic **L-^{3}GP$_{140}$** compared to **M-^{3}GP$_{105}$** despite its higher valency, presumably due to its increased chain conformational flexibility. RBC hemagglutination inhibition assays with soluble **α2,3-SiaLac** polymers **^{3}GP** confirmed the observed valency-dependent binding trend for H1N1 EGG ($K_{i,,\mathbf{3GP}}$ = 313 nM – 1.25 μM, **Fig 2.17**, and **Table 2.5**) and were consistent with prior reports using similar multivalent glycopolymers.[84] In contrast to the behavior of the arrayed mucin mimetics, increasing the polymer length resulted in a more effective inhibition of RBC agglutination by H1N1

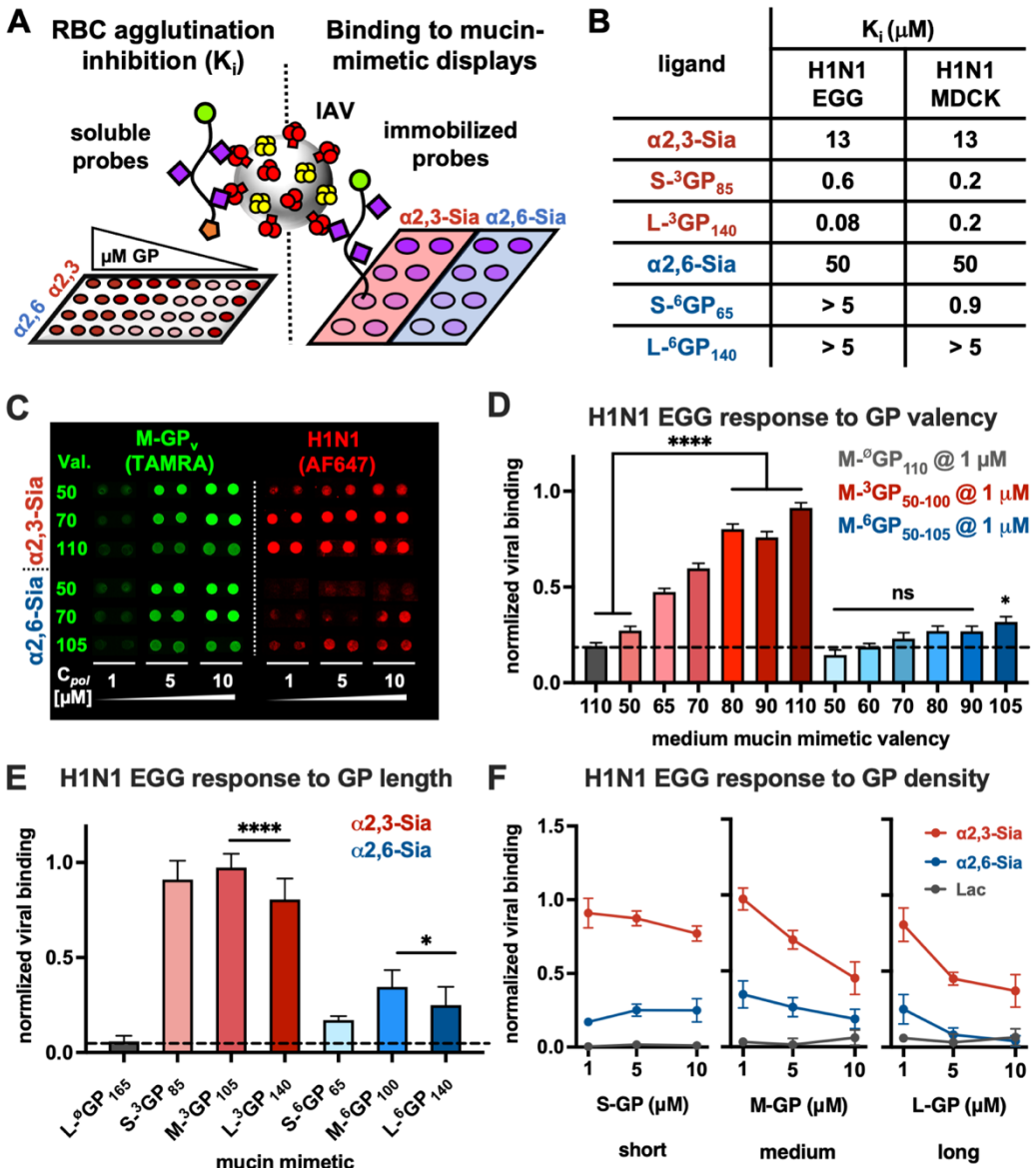


Figure 2.4 H1N1 EGG binding to mucin-mimetic displays of sialoglycan receptors. A) Red blood cell agglutination assays and array screens were used to probe the interactions of H1N1 produced in embryonated chicken eggs (H1N1 EGG) with soluble and surface bound mucin mimetics. B) Inhibitory activity, K_i, of soluble glycan receptors α2,3-SiaLac and α2,6-SiaLac and mucin mimetic glycopolymers GP in RBC agglutination assays expressed as the minimal ligand concentration needed to prevent cell aggregation. The experimental images are included in Figure S10. C) Representative composite images and bar graph representation of H1N1 EGG virus (red) binding to medium-sized mucin mimetics M-GP (green) printed at low surface density (c_{GP} = 1 μM) according to glycan receptor valency. Each array condition is represented as a duplicate and full array images are included in Figures S11, S12, and S14. Values and error bars represent averages and standard deviations of experiments from 9 different arrays. Significance is based on viral binding to the Lac polymer control M-^{Ø}GP$_{110}$ (black dashed line). E) H1N1 EGG binding to mucin mimetics of increasing length printed at low surface density (c_{GP} = 1 μM). Values and error bars represent averages and standard deviations of experiments from 6 different arrays. Significance was determined against Lac polymer control L-^{Ø}GP$_{165}$ (black dashed line). F) H1N1 EGG binding response to increasing crowding of mucin mimetics of all three lengths on the array surface. Values and error bars represent averages and standard deviations of experiments from 6 different arrays. (*p<0.05, **p<0.005, and ***p<0.0001)

EGG in solution ($K_{i,\mathbf{S\text{-}3GP}}$ = 625 nM vs $K_{i,\ \mathbf{L\text{-}3GP}}$ = 78 nM, **Fig 2.17**). It should be noted that the increase in inhibitory capacity of the glycopolymers compared to the monovalent receptor can be accounted for based on glycan valency and concentration alone, rather than avidity enhancements due to multivalency. In the case of the short polymer **S-^{3}GP$_{85}$**, when the total amount of glycan on the polymer is taken into account, the per glycan inhibitory activity ($K_{i,\mathbf{S\text{-}3GP}}$ x **α2,3-SiaLac** valency = 53 μM) was effectively reduced compared to the free **α2,3-SiaLac** ($K_{i,\alpha2,3\text{-SiaLac}}$ = 13 μM). Lactose glycopolymers, $^{\varnothing}$**GP**, lacking sialic acids served as negative controls in both assays (**Fig 2.4** and **Fig 2.17**).

Glycopolymers carrying the **α2,6-SiaLac** receptor (**M-^{6}GP$_{50\text{-}105}$**) showed only a limited ability to engage H1N1 and required glycan valency above 90 to reach binding above background (**Fig 2.4D** and **Fig 2.18**). Extending the length of the mucin mimetic, again, resulted in a decrease in viral capture (**Fig 2.4E** and **Fig 2.19**). All of the **α2,6-SiaLac** polymers failed to inhibiting RBC hemagglutination by the virus over the range of tested polymer concentrations ($K_{i,\mathbf{6GP}\ >}$ 5 μM or 325 μM with respect to **α2,6-SiaLac, Fig 2.17**). Considering that monovalent **α2,6-SiaLac** can prevent RBCs agglutination ($K_{i,\alpha\mathbf{2,6}\ =}$ 50 μM**, Fig 2.4B** and **2.17**), it appears that binding of H1N1 EGG to this glycan receptor is disfavored in the polyvalent glycopolymer presentation.

High levels of mucin expression on the surfaces of epithelial cells produces a dense glycoprotein brush, which has been proposed to restrict IAV access to membrane receptors necessary for infection.68,69 To examine the effects of polymer size and density on viral adhesion, we modelled glycocalyx crowding in our arrays by increasing the printing concentration of the mucin mimetics. We assayed H1N1 EGG binding to maximally glycosylated mucin mimetics of all three lengths arrayed at concentrations of

1, 5, and 10 μM (**Fig 2.4F** and **Fig 2.19**). The virus retained its overall preference for the **α2,3-SiaLac** probes across all surface densities; however, increased crowding of the polymers led to attenuated viral adhesion, which became more pronounced with increasing mucin mimetic length. Crowding of the **α2,6-SiaLac** glycopolymers both enhanced (**S-^{6}GP$_{65}$**) and inhibited (**L-^{6}GP$_{140}$**) viral adhesion depending on polymer length (**Fig 2.4F**). Our data show that, while the H1N1 EGG virus can utilize the less preferred **α2,6-SiaLac** receptors when presented in surface displays on short mucin mimetic scaffolds, increasing the length and density of the conjugates generally negatively impacted viral adhesion regardless of receptor type. Such negative impact of increasing receptor density was previously reported for the binding of nanoparticles bearing recombinant HA proteins to sialoglycans in supported lipid bilayers.[85] Thus, crowding of mucins in the glycocalyx may not only shield underlying glycan receptors from the virus,[68] but also limit viral adhesion to the heavily sialylated mucins themselves.

The observed differential H1N1 EGG binding to the mucin-like receptor displays according to sialic acid linkage type supports the distinct functions of secreted and membrane bound mucins comprising the airway mucosal barrier.[68,86] Therein, secreted mucins produced by goblet cells and presenting primarily α2,3-sialic acid modifications serve as decoy receptors for viral capture and clearance. By contrast, the membrane-tethered mucins produced by epithelial cells display α2,6-linked sialic acid receptors and are thought to limit viral adhesion. The binding of H1N1 EGG to polyvalent **α2,3-SiaLac** mucin mimetics but not the **α2,6-SiaLac** analogs and the sensitivity of the virus to surface crowding of mucin mimetics carrying both receptor types would provide a rationale for the

synergistic but mechanistically distinct functions of secreted and surface-bound mucins in limiting viral infection.

2.4 Support vector machine (SVM) learning-enabled analysis of receptor pattern recognition by influenza A viruses

While the focused analysis of H1N1 EGG binding to some of the key features of glycan presentation in the mucin mimetic array (e.g., polymer size, valency, or surface density, **Fig 2.4**) was informative, the multidimensionality of receptor presentation on the array makes comprehensive assessment challenging and time consuming. For SNA and WGA, plots of lectin capture on the array according to glycan abundance, regardless of polymer structure or brush density, revealed qualitative differences in the lectin binding behavior (**Fig 2.3C**). Similar scatterplot representations can highlight major differences in receptor pattern recognition by different IAVs, as shown in **Figure 2.5A** for H1N1 EGG and the H3N2 A/Aichi/2/68 strain produced in Madin-Darby Canine Kidney (MDCK) cells, H3N2 MDCK.

Noting that molecular recognition is a multi-dimensional problem, we have applied a SVM learning approach (**Fig 2.5B**) to resolve the receptor patterns that are best recognized by the viruses. We leveraged the fact that the SVMs are known to predict accurate relationships from semi- or unstructured data.[87] First, we established a binding threshold for the virus based off the background signal from the control **Lac** polymers (**Fig 2.25**). Then, we randomly selected a portion (67 %) of the array binding data for each virus to train the SVM using a 5-dimensional parameter space of glycan type, glycan valency, glycan spacing on polymer, glycan density on the array, and polymer printing

concentration (**Fig 2.5B**). In combination, these parameters defined additional features of our mucin mimetic receptor displays, such as glycopolymer length (via glycan valency and glycan spacing on polymer) and glycopolymer density on the array surface (via glycan valency, glycan spacing on polymer, and glycan density on array). We used the remaining portion of the binding data set (33 %) to test the accuracy, recall, and precision of the model (**Fig 2.26** and **Fig 2.27**). This model predicted binding of H1N1 EGG to both receptor types, with majority of binding events occurring toward lower glycan valencies and polymer surface densities (**Fig 2.5C,** left), consistent with our manual analysis of the array data for this virus (**Fig 2.4**). The prediction plots in **Figure 2.5C** show the parameters of receptor presentation that most influenced viral binding (i.e., glycan receptor type, mucin mimetic valency and surface density). Additional parameters, such as polymer size and glycan spacing on the polymer, also contribute to viral recognition and can be analyzed. We confirmed that the performance of the model was similar when the training and testing data sets represented either separate array experiments or were selected randomly from combined data across multiple experiments (**Fig 2.28**).

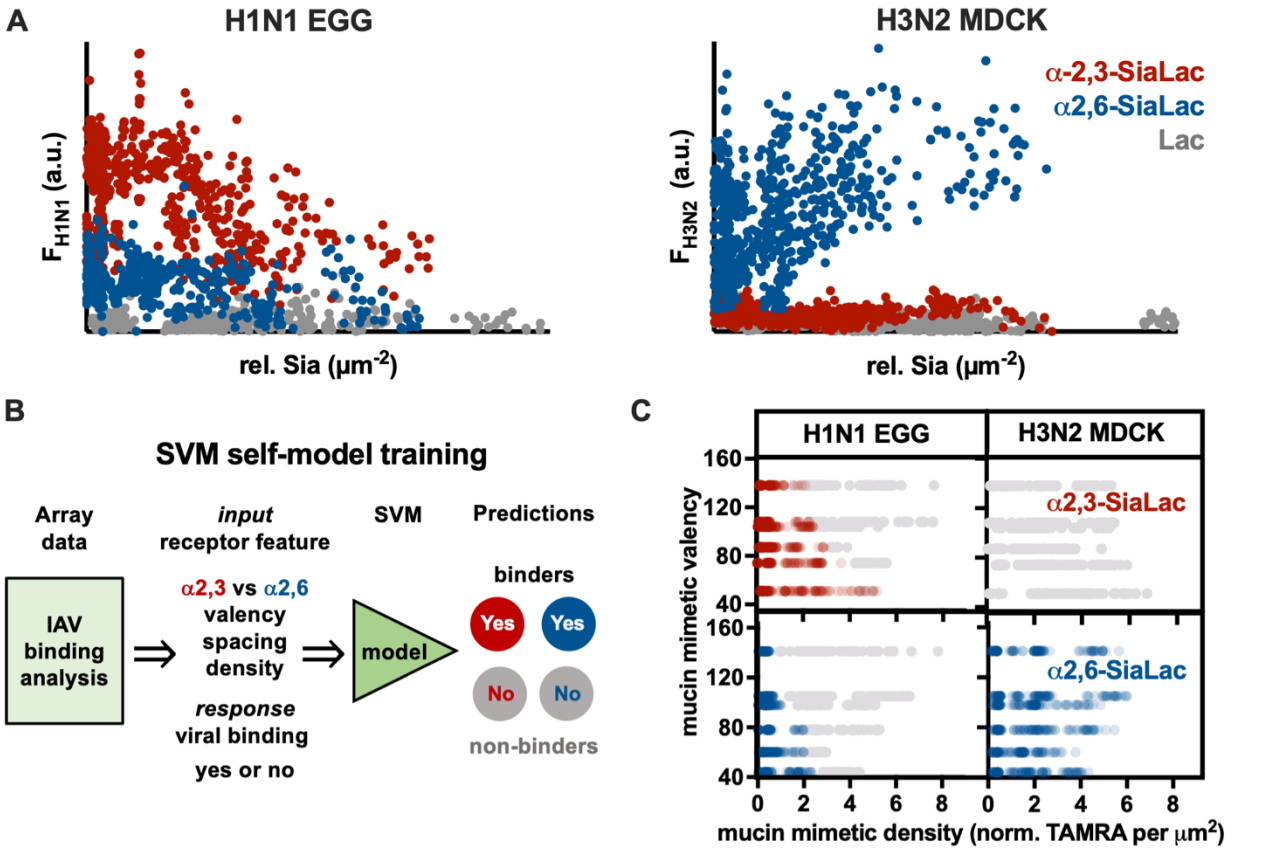


Figure 2.5 Analysis of receptor pattern recognition by H1N1 and H3N2 strains. A) Scatterplots of binding responses to changing glycan receptor density in mucin mimetic arrays for H1N1 EGG and H3N2 MDCK viruses. B) Workflow for creating SVM self-models using viral binding responses and receptor display parameters (i.e., glycan type, valency and spacing on mucin mimetic, mucin mimetic density and concentration in printing buffer) in mucin mimetic arrays. C) SVM analysis of H1N1 EGG and H3N2 MDCK binding in mucin mimetic arrays. In red and blue colors are predicted binding events for each virus to α2,3-SiaLac and α2,6-SiaLac, respectively. Non-binding events are shown in gray. Color intensity indicates the frequency of the predicted binding events according to the valency and surface density of the mucin mimetics. H1N1 EGG recognizes both α2,3- and α2,6-SiaLac with preference for low surface densities of mucin mimetics. H3N2 MDCK is specific for α2,6-SiaLac glycans and its binding does not diminish with increasing mucin mimetic density on the array.

The receptor binding preferences of other IAV strains on the mucin-mimetic array can be rapidly analyzed using the machine learning approach. The application of the SVM to the array binding for H3N2 MDCK correctly predicted the specificity of the virus for α2,6-sialoglycans,[25,88] and, newly, identified its better ability to utilize increasingly dense glycopolymer displays (**Fig 2.5C**). Manual analysis of the array binding data confirmed

these predictions (**Fig 2.20** and **Fig 2.21**), demonstrating the general applicability of the supervised models for analyzing receptor pattern recognition by IAVs.

2.5 H1N1 PR8 propagation in mammalian cells enhances interactions with α2,6-sialoglcyans in mucin-like displays

Having established the SVMs as an effective method to rapidly identify preferred receptor display parameters for viral binding in the mucin mimetic arrays, we set to explore how the glycan binding phenotype of the virus may change depending on the host in which it is propagated. Such information may enhance existing viral surveillance by either eliminating glycan-binding phenotype artifacts, which can be introduced during propagation of field-isolated viruses in the laboratory, or by establishing specific binding phenotype features associated with enhanced human transmission.

We performed a comparative solution and array binding analysis between H1N1 EGG and the same virus produced in MDCK cells (H1N1 MDCK). The MDCK cell line is a commonly used mammalian system for the propagation of IAVs in the laboratory. The binding of H1N1 MDCK to the soluble monovalent **α2,3-SiaLac** and **α2,6-SiaLac**, as measured via RBC hemagglutination inhibition, remained unchanged from that of H1N1 EGG (**Fig 2.4B**). However, H1N1 MDCK exhibited improved ability to engage mucin mimetics carrying the **α2,6-SiaLac** receptors both on arrays (**Fig 2.6A, Fig 2.22**, **and Fig 2.23**) and in their soluble form ($K_{i,\textbf{S-6GP65}}$ = 938 nM, **Fig 2.4B**). The increased ability to recognize α2,6-sialoglycans is expected for H1N1 propagated in MDCK cells, as the cells in the allantoic fluid of embryonated chicken eggs used for viral propagation display mainly α2,3-linked sialic acids and the surfaces of MDCK cells are populated by receptors

with both linkages.[89] During culture in MDCK cells, viruses can undergo selection for both receptor types, giving rise to virus population with an altered sialoglycan binding phenotype.[61,89] However, our hemagglutination inhibition assays indicate that the altered glycan-binding phenotype of H1N1 MDCK does not arise from changes in HA affinity for the individual monovalent receptors.

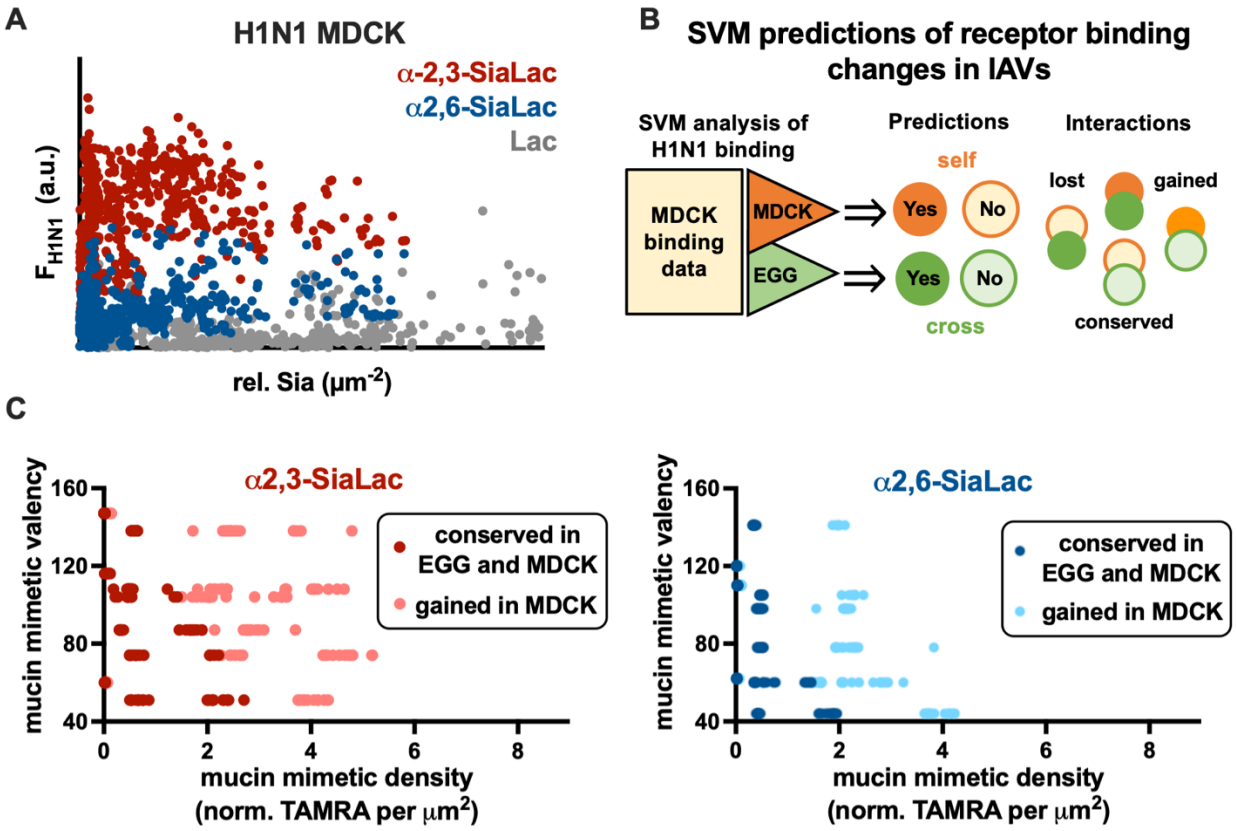


Figure 2.6 Analysis of changes in receptor pattern recognition by IAVs produced in avian and mammalian cells. A) Scatterplots of binding responses to changing glycan receptor density in mucin mimetic arrays for H1N1 MDCK. B) The H1N1 MDCK and H1N1 EGG SVM models can be applied to the MDCK array binding data to generated "self" and "cross"-predictions, respectively, that can be used to determine lost, conserved, and gained interactions. C) The SVM identified better utilization of increasingly crowded mucin mimetic displays by H1N1 upon transition from avian to mammalian cell culture.

Next, we adapted the SVM analysis to identify viral binding features for H1N1 that are conserved between viruses produced in embryonated eggs and those propagated in MDCK cells. We trained and validated a new SVM model for H1N1 MDCK as described above. The H1N1 EGG and H1N1 MDCK models were able to accurately produce

predictions of binding and non-binding events for each respective virus in the mucin mimetic array (**Fig 2.26C** and **Fig 2.27C**). These, we termed “self-predictions”. The EGG model was then applied to the array binding data for H1N1 MDCK to produce an EGG-to-MDCK “cross-prediction” that was compared with the H1N1 MDCK “self-prediction” results (**Fig. 2.6B** and **Fig 2.24**). The binding events correctly predicted by both models were termed “conserved in MDCK”. Those refer to receptor presentations that are recognized by H1N1 regardless of whether it was produced in avian or mammalian cells (**Fig 2.24**). The interactions correctly predicted as binding by the H1N1 MDCK model but were deemed non-binding by the H1N1 EGG model; i.e., interactions that were absent in the EGG-to-MDCK “cross-prediction”, were termed “gained in MDCK”. These refer to interactions with glycan receptor patterns in the array that were absent or in primarily nonbinding regions for the H1N1 virus produced in eggs but emerged when the virus was propagated in MDCK cells. Finally, the non-binders that were correctly predicted by the H1N1 MDCK model but were predicted to be binders by the H1N1 EGG model; i.e., interactions that were absent in H1N1 MDCK but were anticipated to occur based on the EGG-to-MDCK prediction, were termed “lost in MDCK”. These refer to glycan patterns recognized by the H1N1 virus produced in eggs but lost in mammalian cell culture.

As shown in **Figure 2.6C**, the conserved and gained interactions predicted by the SVM algorithm occur in distinct clusters with most of the interaction gain happening at higher polymer valencies or densities. The predictions are in agreement with our agglutination inhibition data and preliminary manual array analysis (**Fig 2.4** and **Fig 2.23**) pointing to improved ability of the H1N1 MDCK virus to engage mucin mimetic displays, both as individual polymers in solution as well as in ensembles on arrays. For **α2,6-**

SiaLac glycopolymers, sensitivity of the virus to receptor crowding continued to persist in H1N1 MDCK, as the majority of binding gains resulted either from higher valency polymers grafted at low densities or from surface crowding of low valency mimetics (**Fig 2.6C** and **Fig 2.29**). Likewise, the SVM analysis identified stronger interactions of H1N1 MDCK with high densities of **α2,3-SiaLac** receptors in the mucin-mimetic array (**Fig 2.6C**). The predictions did not identify loss of any interactions present in the H1N1 EGG upon transition into the mammalian culture system. To validate this method, we compared the results of the cross-prediction with experimental data from H1N1-EGG. We confirmed that the interactions identified as conserved and gained in H1N1 MDCK by the SVM also correlated with binding and non-binding events, respectively, observed for H1N1-EGG in the array. We observed a high level of agreement between the cross-predictions and experimental data in the low and high receptor valency-density regions on the array, where the conserved and gained interactions are well defined (**Fig 2.29**).

We note that an analysis of conservation, gain or loss of binding is made possible by application of the 5-dimensional SVM, wherein the algorithm learns the "rules-of-binding" from a set of interaction patterns (defined by glycan valency, density, polymer spacing and printing concentration data), which is then compared to another dataset to bring out the similarities and differences in the interactions. Arriving at these conclusions would be impossible using only chemical intuition or simple visualization of data because the interaction patterns are not apparent when the data is plotted with equal weights on all five dimensions. Using between 200 to 2000 iterations, the SVM finds linear combinations of the parameters in this 5-dimensional space, reweighting the original

dimensions, until most separated classes are created, and their underlying data features are segregated, to distinguish between binding and non-binding events.

It is unlikely that such combinations can be found manually, or using unsupervised clustering schemes, making SVM a natural choice to create an “intelligent” space that correlates binding rules to linear combinations of glycan valency and spacing on the mimetics and their density on the array surface. In contrast to popular simple clustering algorithms, such as K-means, which are limited by the number of dimensions, the strength of SVM is that it does not impose the ground-truth or prior knowledge of the binding patterns to learn the rules of binding.[90] Rather, SVM infers both the rule as well as the outcome directly from higher-dimensional data sets. Although we could have employed more sophisticated algorithms, such as neural networks, the use of a discrete binary classifier in this study (binder = 1 and non-binder = 0) made SVM a more convenient choice. The SVMs are not solved for local optima, thus, they handle well high dimensional data using one to two classifiers, as is the case here.

The improved range of binding interactions of H1N1 MDCK with $\alpha 2$,6-linked sialic acids is expected based on the higher expression of these glycans in the mammalian cells. Even in absence of *Muc* gene expression in MDCK cells,[91] the observed changes in receptor pattern recognition for both glycans do not stem from altered affinity of its HA toward the monovalent glycans in solution (**Fig 2.4B**). The HA protein carries several glycosylation sites near the sialic acid binding region and the addition of glycans can influence[27] binding**Error! Bookmark not defined.** and may give rise to the observed changes in receptor pattern recognition.

We performed PAGE analysis of the viral proteins before and after PNGase treatment to remove *N*-linked glycans. The glycosylated form of HA1 fragment from the egg-grown H1N1 virus had lower molecular weight than that of the MDCK virus **(Fig 2.30**). This difference in gel mobility was eliminated by PNGase treatment, which catalyzes the removal of *N*-linked glycans. This points to a more extensive glycosylation of the MDCK cell-derived virus HA proteins and is consistent with previously observed differences in the glycosylation, but not in the primary amino acid sequence, of HAs from isolated duck H1N1 viruses propagated in MDCK cells and egg cultures.[92] Whether the relationship between the changes in HA glycosylation and the altered receptor patter recognition of the virus observed in the current array study is causative or correlative is yet to be determined. Nonetheless, it presents one possible explanation for the altered binding behavior of the viruses. Such differences in glycosylation, may reflect the influence of species, cell type or combination of both on receptor pattern recognition by viruses and contribute to their emerging ability to cross between species. Similar analysis of glycan binding changes upon transition from egg culture to MDCK cells was not possible for the H3N2 virus used in this study, which, due to its restricted receptor specificity, does not propagate well in embryonated eggs without undergoing mutations in its HA sequence.[93] However, we anticipate that the array can help identify other viruses with glycan-binding phenotype shifts occurring in response to adaptation to a new host.

2.6 Conclusions

The glycocalyx exists as a complex assortment of membrane-tethered and secreted glycoconjugates that serve different roles in multicellular identity, function, and

pathogen invasion. In this work, we aimed to model the interactions of IAVs with the mucin glycoprotein components of the mucosal barrier by generating mucin-mimetic glycopolymers with tunable sizes and compositions displaying α2,3- and α2,6-linked sialyllactose glycans as prototypes for the avian and human receptors for IAVs. RBC hemagglutination inhibition assays with soluble forms of the probes revealed an enhancement in selectivity of H1N1 PR8 viruses produced in embryonated chicken eggs from ~ 4-fold preference for binding of **α2,3-SiaLac** in its soluble monomeric form to more than 60-fold for the polyvalent receptor displays. This differential arose from selectively disfavored binding of the virus to the **α2,6-SiaLac** glycans in the mucin-mimetics rather than increased avidity toward the **α2,3-SiaLac** polymers. Systematic evaluation of H1N1 EGG capture on arrays of immobilized mucin mimetics enabled by an SVM learning algorithm showed that the virus bound surface displays of both receptor types but was attenuated at high polymer densities. The receptor pattern recognition changed when the virus was propagated in MDCK cells toward improved utilization of **α2,6-SiaLac** mucin mimetics in both soluble and immobilized forms, and a lower sensitivity to surface crowding of the **α2,3-SiaLac** glycopolymers.

Our findings are consistent with the proposed protective functions of soluble mucins at the airway epithelium which present primarily α2,3-linked receptors and prevent infection through viral capture and clearance. Newly, our observations that the presentation of α2,6-linked sialoglycans on linear polyvalent scaffolds and the arrangement of such conjugates in increasingly dense surface ensembles disfavors the binding of H1N1 produced in avian cells also provides support for the role of membrane-associated mucins in limiting viral adhesion and infection at the epithelial cell surface. The

improved binding of H1N1 viruses produced in mammalian cells to the mucin-mimetic displays of α2,6-linked sialoglycans was not accompanied by changes in the affinity or selectivity of their HA proteins toward the individual glycan receptors. While the basis for the differences in receptor pattern recognition needs to be further investigated, our studies show that the mucin probes and their arrays may serve as useful tools to investigate viral interactions at the mucosal barrier and the evolution of their glycan receptor-binding phenotype.

The array and the SVM analysis can be extended to other IAV strains. The platform successfully predicted the α2,6-sialoglycan specificity of the H3N2 A/Aichi/2/68 virus and identified its better ability to utilize polyvalent and high-density mucin-like displays of the human receptor prototype, **α2,6-SiaLac**. The modularity of the mucin mimetic synthesis and the ease with which glycans can be incorporated into these materials[94] should be well suited for the introduction of more complex and biologically relevant glycans into the array, including mucin type *O*-glycans. Of particular interest for future array development are glycans with linear or branched core structures variously modified with extended *N*-acetyllactosamine disaccharide repeats and presenting fucosylation or sulfation motifs, as they continue to emerge from glycomics[27,46,48,70] and genetic screens[49,50] as relevant to IAV recognition and infection. Machine learning, as employed in this study, is well poised to support such increase in the dimensionality of parameter space on the array and can be expanded to non-binary classification for analysis. This may allow future inclusion of multiple continuous classifiers, such the strength and kinetics of a binding interaction between a virus and a receptor pattern, to construct more sophisticated neural networks that reveal higher-order relationships between pathogens and their hosts.

2.7 Materials and Methods

A detailed list of chemical and biological reagents including their sources and catalog numbers can be found as Table 2.1 in the Supporting Information.

2.7.1 Data and code availability

All SVM datasets and code generated for our machine-learning analysis of IAV binding to the mucin-mimetic arrays are available on Github and can be accessed through https://doi.org/10.5281/zenodo.5563187.

2.7.2 Glycopolymer synthesis and characterization

Fluorescently labeled glycopolymers end-functionalized with an azide (GP) used in this study and their polymeric precursors (**P**) were prepared using the RAFT polymerization method according to previously published procedures[95] and are summarized in Figure 2 and described in detail in SI.

2.7.3 Cell culture

Madin-Darby Canine Kidney (MDCK) cells*:* MDCK cells were cultured in Dulbeco's modified Eagle's medium supplemented with 10% FBS, 100 U/mL penicillin, and 100 U/ml streptomycin.

2.7.4 Viral culture

Influenza virus strain A/PR/8/34 (H1N1, ATCC VR-1469) was purchased from ATCC and propagated in MDCK cells that were transferred to DMEM medium supplemented with 0.2% BSA fraction V, 25mM HEPES buffer, 2 μg/ml TPCK-trypsin, and 1% penicillin/streptomycin ("DMEM-TPCK" media). The same strain was used for viral production in embryonated chicken eggs. Briefly, fertilized chicken eggs were obtained and stored at 37 ^{o}C. When the embryos were 10 days old (assessed by candeling), virus was injected through small holes made in the shell over the air sac. The holes were covered with parafilm and the eggs were placed back at 37 ^{o}C. After two days, the eggs were chilled to prepare for harvest. The eggshells were cut open above the airsac and the allantoic fluid was carefully collected into centrifuge tubes without rupturing the yolk. The tubes were centrifuged to pellet any debris and the supernatant containing virus was aliquoted into 1 mL cryovials and stored at -80 ^{o}C.

2.7.5 Viral titers

Turkey red blood cells were purchased from Lampire and a 1% solution was used to determine viral titers via the hemagglutination test. MDCK cells were used to determine the 50% tissue culture infective dose ($TCID_{50}$) using the Spearman-Karber method.

2.7.6 Hemagglutination inhibition

Glycopolymer solutions in PBS (25 μL, ranging from 20 μM to 20 nM by 2-fold dilutions) were added to a 96 well plate. The last well in each row was used as a PBS control and did not contain glycopolymer or virus. An equal volume (25 μL) of H1N1

diluted to HAU=4 (turkey RBCs) was added to each well and allowed to incubate at room temperature. After a ½ hr incubation, 50 µL of 1% turkey RBCs in PBS were added to all of the wells. The K_i value was read after a ½ hr as the lowest concentration of polymer to inhibit hemagglutination.

2.7.7 Array construction

Microarrays were printed on cyclooctyne-coated glass slides as previously described[Error! Bookmark not defined.] using a sciFLEXARRAYER S3 printer (Scienion) following passivation with a 1% BSA/0.1% Tween-20 solution in PBS for 1 hour. Polymer solutions were diluted in printing buffer (0.005% Tween-20 in PBS) to concentrations of 1, 5, and 10 µM polymer and printed in replicates of six at a humidity of 70%. Following an overnight reaction at 4°C, excess polymer was removed by vigorous washing in 0.1% Triton X/PBS solution. The slides were then imaged on an Axon GenePix4000B scanner (Molecular Devices) at the highest PMT possible without saturation of pixels. The GenePix software was used to calculate the relative polymer density, by dividing the fluorescent intensity at 532 nm by the labeling efficiency for each polymer (obtained through UV-Vis measurements) and the spot area (calculated from the spot diameter generated by the software). To obtain relative glycan density, the polymer density was multiplied by the glycan valency for each polymer (attained through NMR integration).

2.7.8 Array binding assays

Prior to array binding assays, subarrays were blocked with 3% BSA solution in PBS for 1 hr. *For lectin binding*, subarrays were washed three times with lectin binding

buffer (0.005% Tween-20 in PBS with 0.1 mM $CaCl_2$, $MnCl_2$, and $MgCl_2$). Dylight labeled SNA and WGA were diluted in the lectin binding buffer and incubated on the array for 1 hour in the dark. After washing with binding buffer, 0.1% Tween-20 solution in PBS, and rinsing with MilliQ water, the slides were imaged at the highest PMT possible without producing saturated pixels. *For H1N1 binding*, subarrays were washed with 1% BSA/PBST three times following passivation. H1N1 was diluted in 1% BSA/PBST and incubated on the array for 1 hr. The slide was washed two times with 1% BSA/PBST, and then fixed for 20 min with 4 % PFA in PBS. To visualize H1N1, binding a 1:500 dilution of anti-HA in 1% BSA/PBST was incubated on the array for 1 hour, followed by an hour incubation in the dark of a 1:500 dilution of anti-rabbit-AF647 antibody. The subarrays were washed two more times with 1% BSA/PBST, two times with the 0.1% Tween-20 solution in PBS, rinsed with MilliQ and imaged at the highest PMT possible without producing saturated pixels.

2.7.9 Machine learning workflow

The machine learning workflow is demonstrated in Figure S21. The H1N1 EGG, H3N2, and H1N1 MDCK data sets contained 8373, 1922, and 1273 data points respectively. Given the size of our data, the feature space (4-5 variables), and the category of problem (binary classification), we chose to use support vector machine (SVM) algorithm for this work. SVM is ideally suited for such binary classification tasks. Likewise, more sophisticated algorithms like random forest shows ideal performance with larger data sets and/or more complex features. Use of random forest in place of SVM did not show any improvements (data not shown). As a preprocessing step, all negative viral

fluorescent intensities (which resulted from background subtraction in the absence of viral binding) were adjusted to zero to indicate the lack of viral binding. The fluorescence intensities were then normalized over the entire data set in the range of [0,1]. This continuous data was converted to categorical ("binders"/"nonbinders") based on a cut-off fluorescence that was determined from the distribution of fluorescence intensities of lactose samples which served as negative control. Next, the features (glycan type, valency, and polymer density) were scaled to a range of [0,1] in order to avoid bias from higher values. The only categorical feature (glycan type) was transformed into numeric by mapping to a two-dimensional space were α2,3-SiaLac is represented by (1,0) and α2,6-SiaLac by (0,1). This data set was then split into a training and test set, where training set contained 67% of the data. Each experiment containing 6 fluorescence measurements was individually split between training and test sets. The SVM algorithm was used for learning from the training data, and predictions were validated against the test data. We confirmed that the performance of the model was similar when the data were split across all experiments for creating training and test sets (**Fig 2.28**). Training and testing were performed separately for H1N1 EGG, H3N2, and H1N1 MDCK data. Convergence was reached within 2000 iterations for H1N1 EGG data and 200 iterations for H3N2 and H1N1 MDCK data (**Fig 2.26**). Next, the algorithm trained on egg-virus data was used to predict results for MDCK-virus data, and the prediction results from this so-called "cross-model" was compared with the results obtained from the model trained on MDCK-virus (the "self-model"). Data was prepared in python using pandas, numpy, and scipy packages. SVM was performed using python's scikit-learn package.[96]

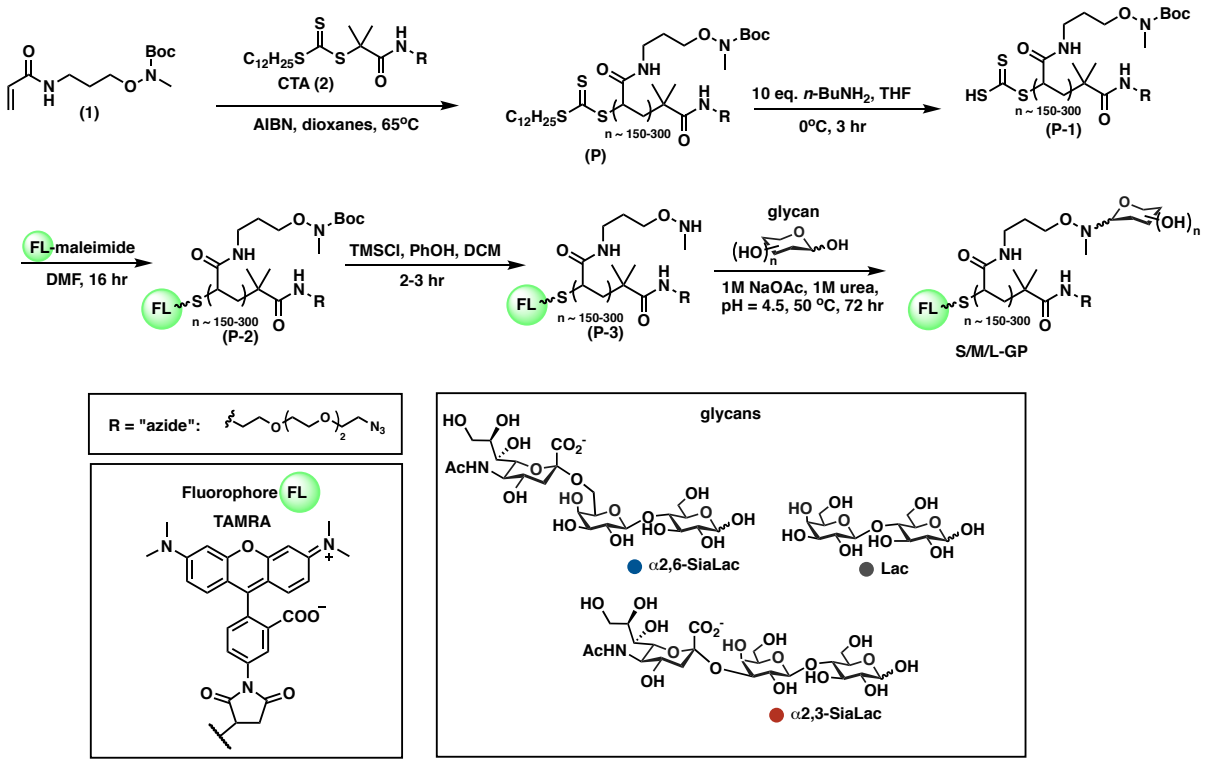
2.8 Supplemental information

2.8.1 Supplemental items

Table 2.1 Chemical and biological reagents

Reagent	Source	Cat. No.
3'-Sialyllactose sodium salt	Carbosynth	OS04397
6'-Sialyllactose sodium salt	Carbosynth	OS04398
Amicon Ultra-0.5 mL 10K Spin filters	Millipore	UFC5010
Dibenzocyclooctyne (DBCO) amine	BroadPharm	BP-22066
SuperChip Microarray Slides Epoxysilane	Thermo Fisher	EP10143261
TAMRA Maleimide	Thermo Fisher	T6027
BSA Fraction V	Spectrum	A3611
Dylight649 conjugated *Sambucus nigra* lectin	Ey Labs	DY649-6802-1
Dylight649 conjugated *Wheat germ agglutinin* lectin	Ey Labs	DY649-2101-1
Influenza virus A/PR/8/34 (H1N1)	ATCC	VR-1469
PBS (with Calcium and Magnesium)	Gibco	14040141
PCR tube	Thermo Fisher	AB0337
PD-10 Desalting Column	GE Life Sciences	17085101
Turkey Red Blood Cells	Lampire	7249408
Phosphate buffered saline	Gibco	10010023
A/California/06/09 H1N1 (polyclonal rabbit IgG α-HA antibody)	eEnzyme	IA-01SW-0100
goat α-rabbit-AF647	Thermo Fisher	A21244
Bolt 4-12% Bis-Tris Plus gel	Invitrogen	NW04120BOX
PNGase F	New England Biolabs	P0704L
Anti-rabbit IgG, HRP-linked Antibody	Cell Signaling Technologies	7074s
SeeBlue Plus2 Prestained standard ladder	Invitrogen	LC5925
BOLT LDS sample Buffer (4x stock)	Invitrogen	B0007
NuPAGE™ MES SDS Running Buffer (20X)	Invitrogen	NP0002
iBlot2 PVDF Mini Stacks	Invitrogen	IB24002
Luminata Classico HRP substrate	Millipore	WBLUC0500

*All other chemicals, if not otherwise indicated, were purchased from Sigma-Aldrich.



Scheme 2.1. Preparation of mucin mimetic glycopolymers GP

Table 2.2 Polymer backbone characteristics

Polymer ID	Mole ratio M(1)/CTA(2)	1 Monomer (mmol)	2 CTA (μmol)	AIBN (mmol)	Rxn time (hr)	% yield	Mw (kDa)	DP via SEC	*Đ*
S-P	200	2.57	12.9	2.56	8	90	39	150	1.28
M-P	300	1.94	6.47	1.28	10	87	54	200	1.17
L-P	600	3.59	6.03	2.44	12	91	98	300	1.36

Table 2.3 Structural characteristics of glycopolymers **GP**

Entry #	Polymer ID	DP	TAMRA labeling (%)	Glycan	Glycan equiv. added	Glycan ligation efficiency (%)	Glycan valency	Theoretical end-to-end length (nm)
1	**S-^{3}GP$_{50}$**	150	6.3	**α2,3 SiaLac**	0.49	34	50	8
2	**S-^{3}GP$_{60}$**	150	27.6	**α2,3 SiaLac**	1.10	40	60	8
3	**S-^{3}GP$_{85}$**	150	6.3	**α2,3 SiaLac**	1.09	58	85	8
4	**M-^{3}GP$_{50}$**	200	40.8	**α2,3 SiaLac**	0.09	27	50	9
5	**M-^{3}GP$_{65}$**	200	40.8	**α2,3 SiaLac**	0.25	34	65	9
6	**M-^{3}GP$_{70}$**	200	40.8	**α2,3 SiaLac**	0.35	37	70	9
7	**M-^{3}GP$_{75}$**	200	27.5	**α2,3 SiaLac**	0.53	39	75	9
8	**M-^{3}GP$_{80}$**	200	40.8	**α2,3 SiaLac**	0.49	43	80	9
9	**M-^{3}GP$_{90}$**	200	40.8	**α2,3 SiaLac**	0.71	49	90	9
10	**M-^{3}GP$_{110a}$**	200	40.8	**α2,3 SiaLac**	1.08	58	110	9
11	**M-^{3}GP$_{110b}$**	200	27.5	**α2,3 SiaLac**	1.18	54	110	9
12	**L-^{3}GP$_{110}$**	300	9.5	**α2,3 SiaLac**	0.68	36	110	12
13	**L-^{3}GP$_{140}$**	300	9.5	**α2,3 SiaLac**	1.14	46	140	12
14	**L-^{3}GP$_{145}$**	300	55.8	**α2,3 SiaLac**	1.09	49	145	12
15	**S-^{6}GP$_{45}$**	150	6.3	**α2,6 SiaLac**	0.47	29	45	8
16	**S-^{6}GP$_{60}$**	150	27.6	**α2,6 SiaLac**	1.09	41	60	8
17	**S-^{6}GP$_{65}$**	150	6.3	**α2,6 SiaLac**	1.10	43	65	8
18	**M-^{6}GP$_{50}$**	200	40.8	**α2,6 SiaLac**	0.11	28	50	9
19	**M-^{6}GP$_{60}$**	200	40.8	**α2,6 SiaLac**	0.26	32	60	9
20	**M-^{6}GP$_{70}$**	200	40.8	**α2,6 SiaLac**	0.35	37	70	9
21	**M-^{6}GP$_{80}$**	200	40.8	**α2,6 SiaLac**	0.51	41	80	9
22	**M-^{6}GP$_{90}$**	200	40.8	**α2,6 SiaLac**	0.69	46	90	9
23	**M-^{6}GP$_{100}$**	200	27.5	**α2,6 SiaLac**	1.13	51	100	9
24	**M-^{6}GP$_{105}$**	200	40.8	**α2,6 SiaLac**	1.07	55	105	9

Table 2.3 Structural characteristics of glycopolymers **GP**, continued

Entry #	Polymer ID	DP	TAMRA labeling (%)	Glycan	Glycan equiv. added	Glycan ligation efficiency (%)	Glycan valency	Theoretical end-to-end length (nm)
25	**L-^{6}GP$_{105}$**	300	9.5	**α2,6 SiaLac**	0.68	35	105	12
26	**L-^{6}GP$_{120}$**	300	37.4	**α2,6 SiaLac**	1.11	40	120	12
27	**L-^{6}GP$_{140}$**	300	9.5	**α2,6 SiaLac**	1.08	47	140	12
28	**S-^{Ø}GP$_{65}$**	150	27.6	**Lac**	0.51	44	65	8
29	**S-^{Ø}GP$_{75}$**	150	6.3	**Lac**	0.49	51	75	8
30	**S-^{Ø}GP$_{110}$**	150	6.3	**Lac**	1.08	74	110	8
31	**M-^{Ø}GP$_{85}$**	200	40.8	**Lac**	0.11	44	85	9
32	**M-^{Ø}GP$_{95a}$**	200	27.5	**Lac**	0.58	47	95	9
33	**M-^{Ø}GP$_{95b}$**	200	40.8	**Lac**	0.25	50	95	9
34	**M-^{Ø}GP$_{110}$**	200	27.5	**Lac**	0.37	57	110	9
35	**M-^{Ø}GP$_{130}$**	200	27.5	**Lac**	1.22	66	130	9
36	**L-^{Ø}GP$_{85}$**	300	9.5	**Lac**	0.22	28	85	12
37	**L-^{Ø}GP$_{155}$**	300	37.4	**Lac**	0.70	52	155	12
38	**L-^{Ø}GP$_{170}$**	300	9.5	**Lac**	0.66	56	170	12

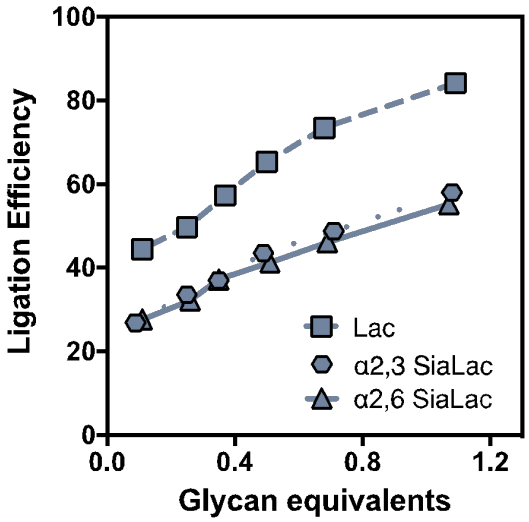


Figure 2.8 Control of glycopolymer valency. The number of glycans introduced into the polymers was controlled through stoichiometry of **SiaLac** or **Lac** with respect to the reactive side chains. The uncharged disaccharide **Lac** is incorporated more readily than the negatively charged **SiaLac** trisaccharides.

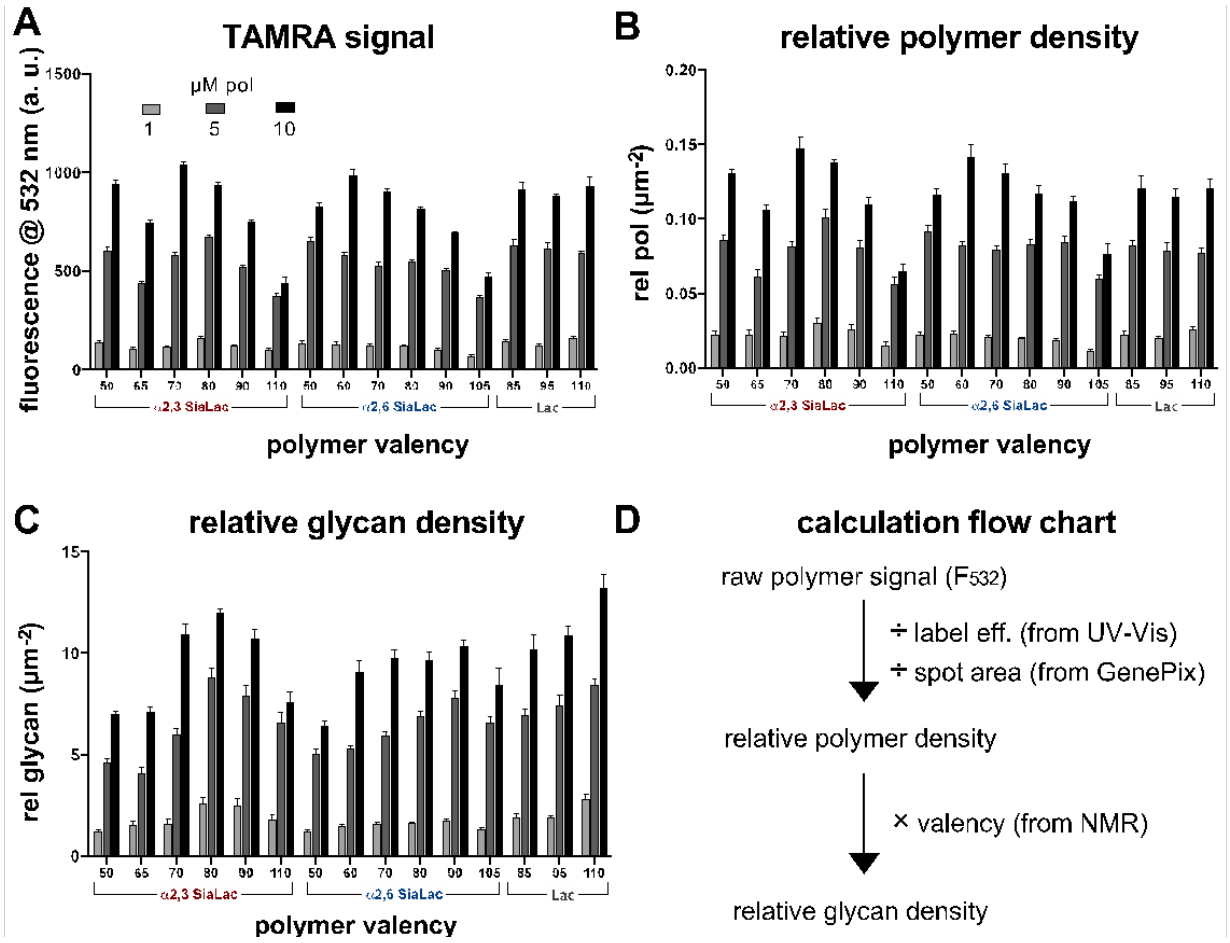


Figure 2.9 Conversion of TAMRA signal to relative polymer density and relative glycan density for the M-GP array. A) The TAMRA signal of the labeled polymers is obtained from analysis using GenePix software. B) The TAMRA signal can be converted to relative polymer density by dividing by the polymer labeling efficiency (41%) acquired through UV-Vis spectrometry measurements and dividing by the spot area which is calculated using the diameter from GenePix analysis. C) The relative glycan density can be determined by multiplying the relative polymer density by the valency of each polymer which is obtained through NMR integration. D) A flowchart showing the calculations.

.

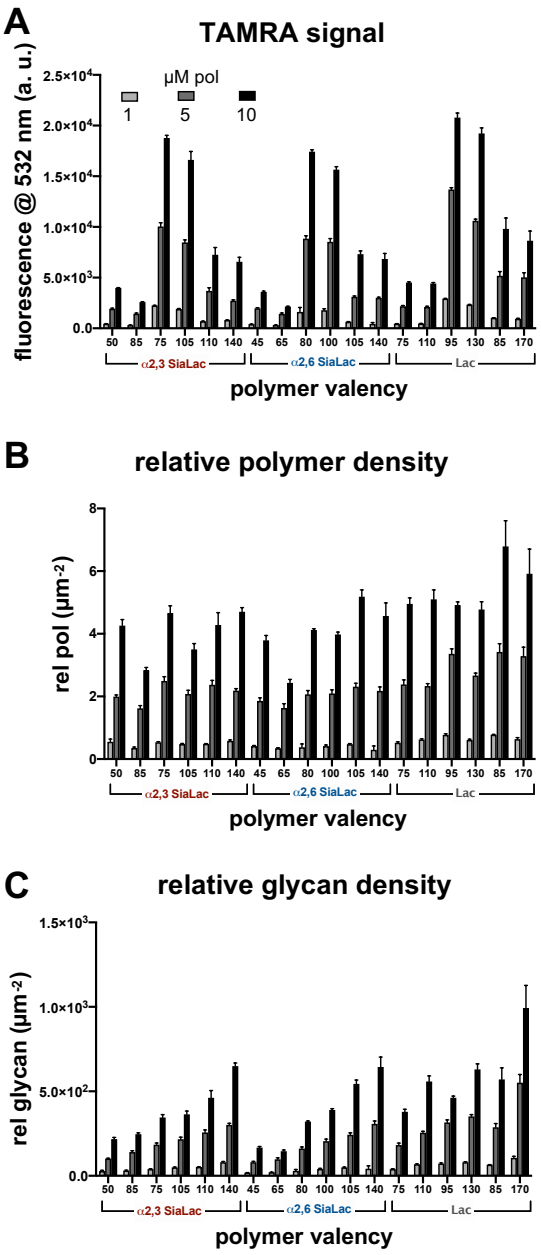


Figure 2.10 Polymer grafting and glycan density on arrays containing all polymer lengths. A) The TAMRA signal of the labeled polymers is obtained from analysis using GenePix software. B) The TAMRA signal can be converted to relative polymer density by dividing by the polymer labeling efficiency (S: 6%, M: 28%, L: 9%) acquired through UV-Vis spectrometry measurements and dividing by the spot area which is calculated using the diameter from GenePix analysis. C) The relative glycan density can be determined by multiplying the relative polymer density by the valency of each polymer which is obtained through NMR integration.

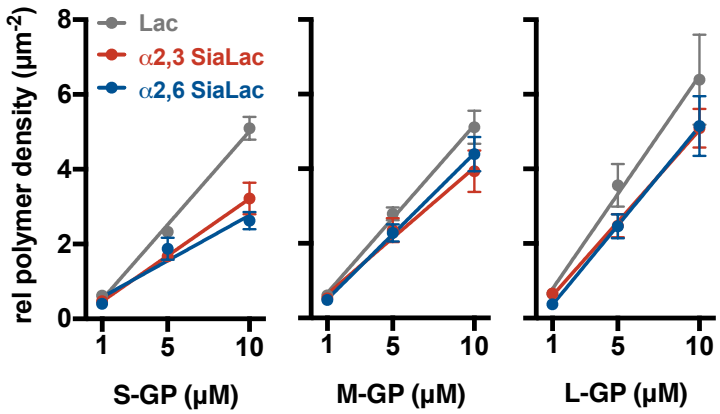


Figure 2.11 Relative array surface density for polymers S-, M-, and L- GPs with maximal glycosylation. Data points represent means ± standard deviation of 12 replicate spots and is fit using linear regression analysis (PRISM). All **GPs** remain in the linear range of grafting to the array surface. **Lac** polymers achieve the greatest density for all polymer sizes while the SiaLac polymers graft similarly under all conditions.

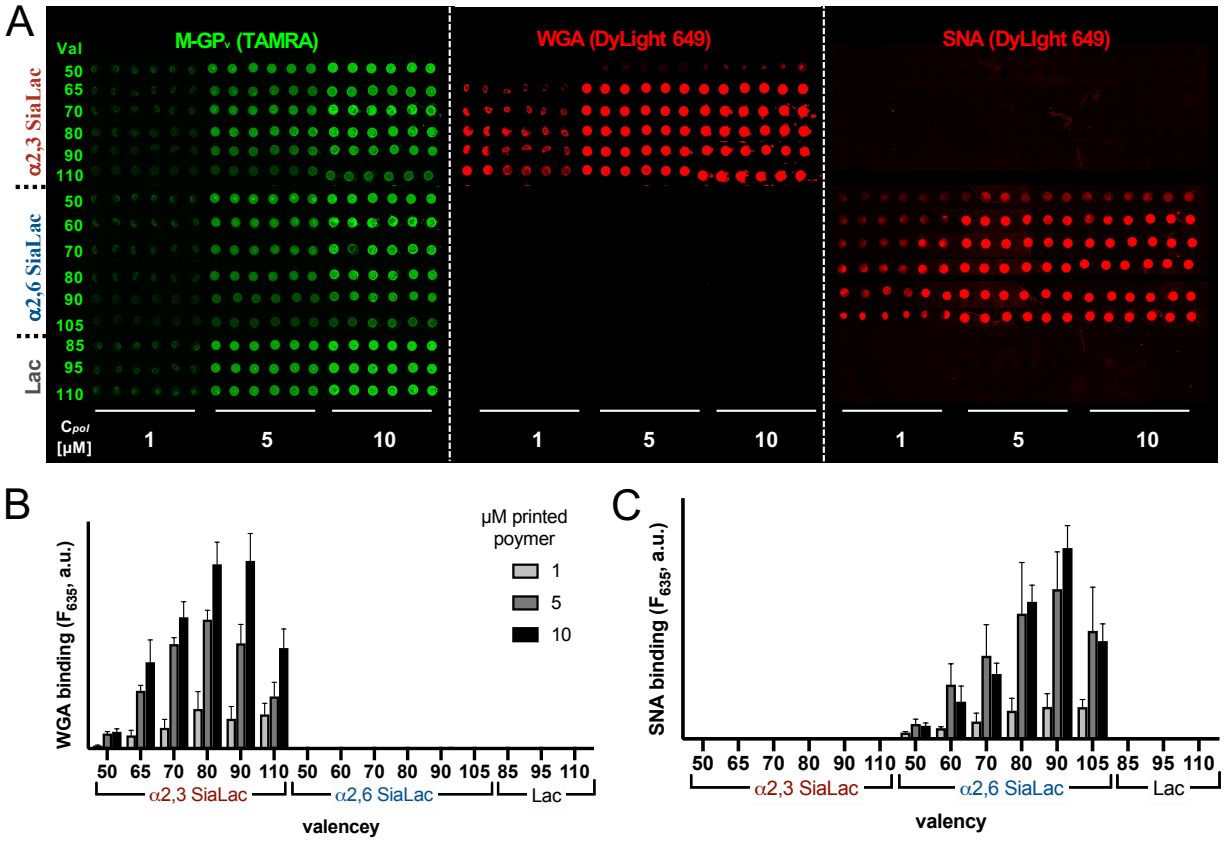


Figure 2.12 SNA and WGA binding in mucin mimetic arrays (extended data for Figure 2.3). A) Micrographs of the printed glycopolymers (TAMRA) and the bound lectins (DyLight649). WGA binds only to **M-^{3}GP$_{50\text{-}110}$** (middle panel), and DyLight649 labeled SNA binds only to **^{6}GP$_{50\text{-}105}$**. Bar graph representations for B) WGA binding at 500 nM and C) SNA binding at 500 nM. Error bars represent standard deviation from 6 replicate spots.

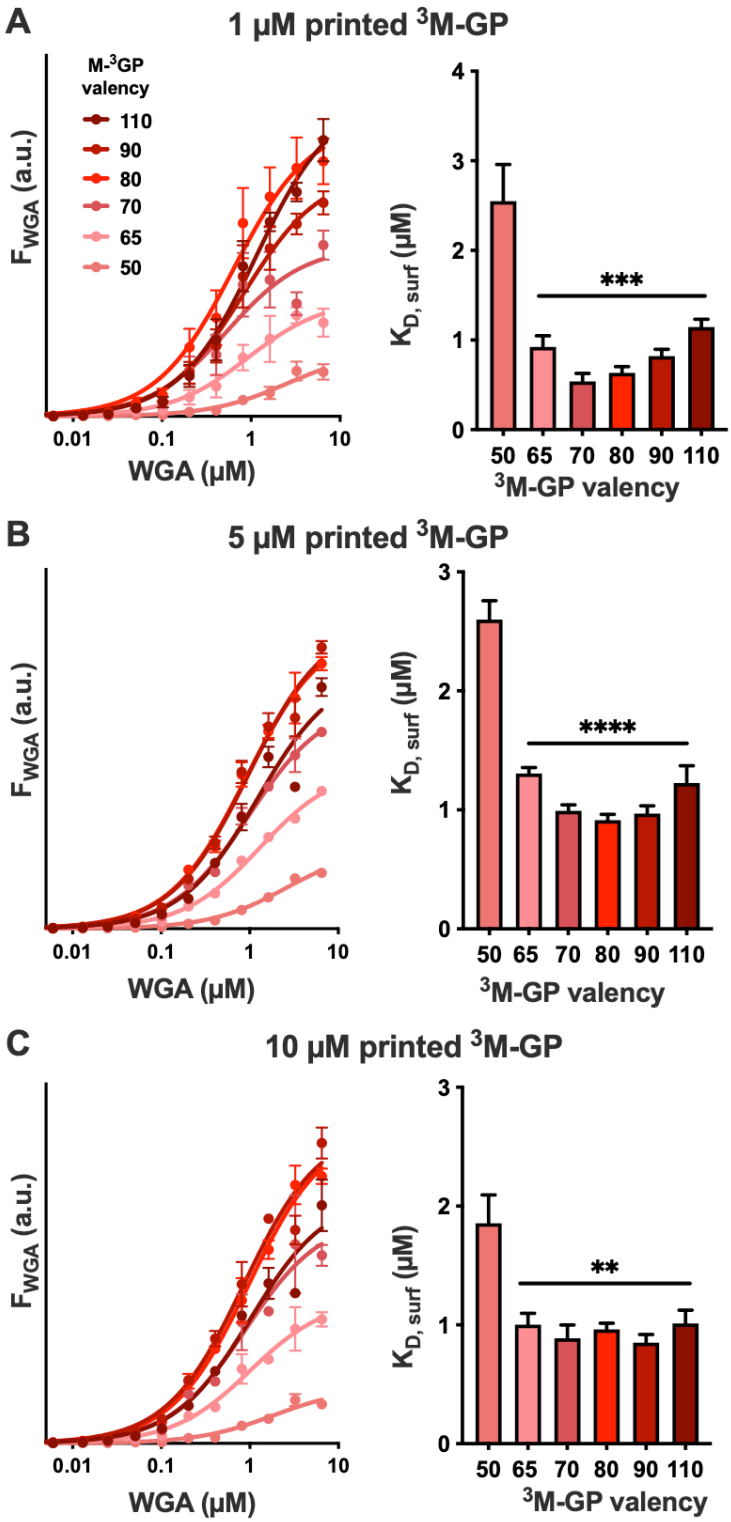


Figure 2.13 Determination of $K_{D,surf}$ for WGA binding to **M-^{3}GP$_{50\text{-}110}$** in arrays. Binding of WGA to the glycopolymers follows the same trends at all printing concentrations, however the valency dependence on $K_{D,\,surf}$ is most prominent at 1 μM printed polymer (A), decreases at 5 μM (B), and is essentially nonexistent for **^{3}M-GP$_{65\text{-}100}$** at the 10 μM printing concentration (C). (**p < 0.001, ***p < 0.0005, ****p < 0.0001)

Table 2.4 $K_{D,surf}$ values for the binding of WGA to **M-^{3}GP$_{50-110}$**

	μM printed polymer		
val	**1**	**5**	**10**
50	2.5 ± 0.4	26 ± 0.2	1.8 ± 0.3
65	0.9 ± 0.1	1.30 ± 0.06	0.98 ± 0.09
70	0.54 ± 0.08	1.01 ± 0.05	0.9 ± 0.1
80	0.63 ± 0.06	0.91 ± 0.05	0.95 ± 0.06
90	0.82 ± 0.07	0.95 ± 0.07	0.88 ± 0.08
110	1.15 ± 0.08	1.2 ± 0.2	1.1 ± 0.1

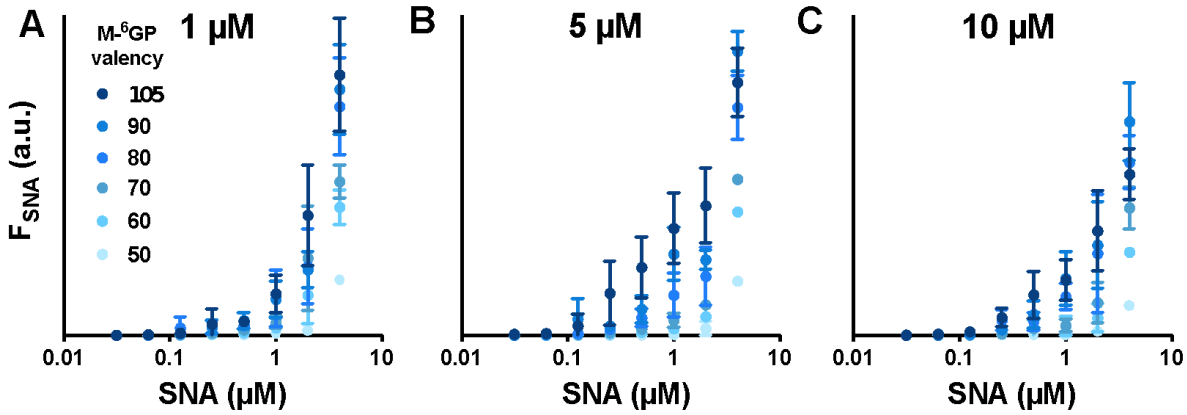


Figure 2.14 Determination of $K_{D,surf}$ for SNA binding to **M-^{6}GP$_{50-110}$** in arrays. (A-C) Fluorescence intensity of Dyight649-SNA bound to arrays printed at 1, 5 and 10 μM concentrations. Saturation binding could not be achieved due to SNA aggregation.

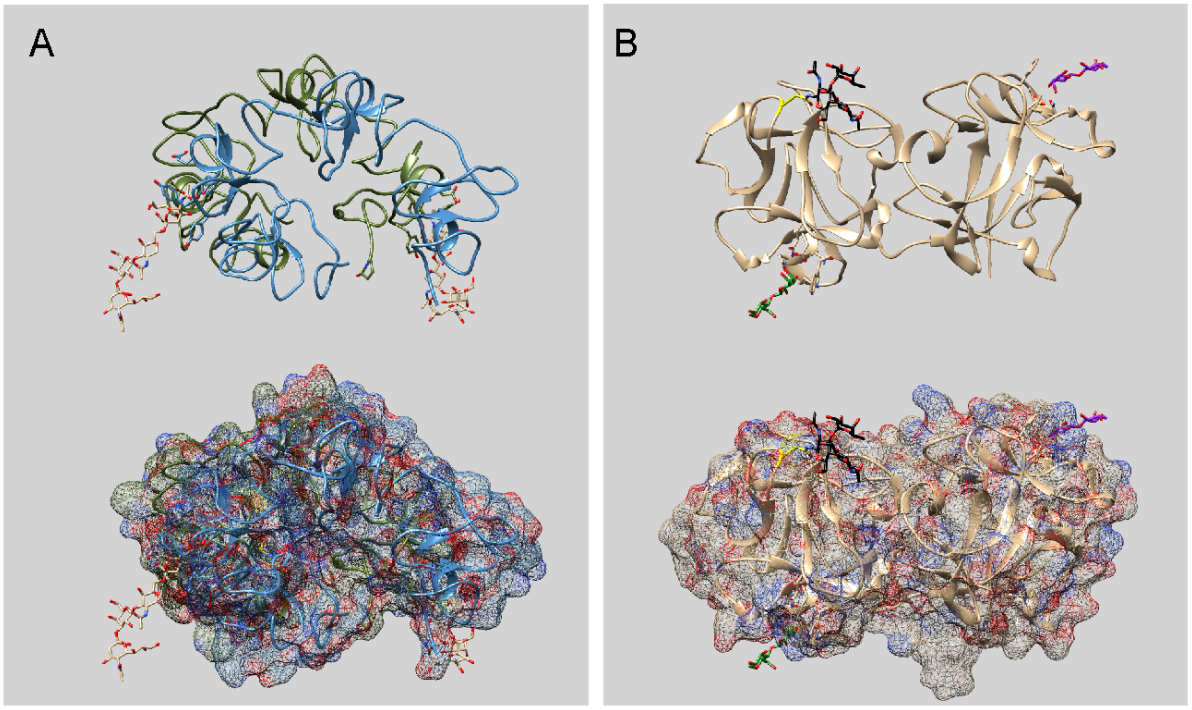


Figure 2.15 Crystal structures of dimeric WGA and monomeric SNA in ribbon and mesh representations. A) The WGA dimer is shown in complex with a sialyllated glycopeptide. This lectin contains two sialic acid binding sites on the same side of protein separated by 3.9 nm. B) Monomeric SNA contains two binding sites per subunit that are found on opposite ends of the protein. The glycan in the Iα binding site is colored in green and the one in the IIγ binding site is colored in purple. The black N-linked glycan included in the structure is important for the oligomerization state of the lectin.

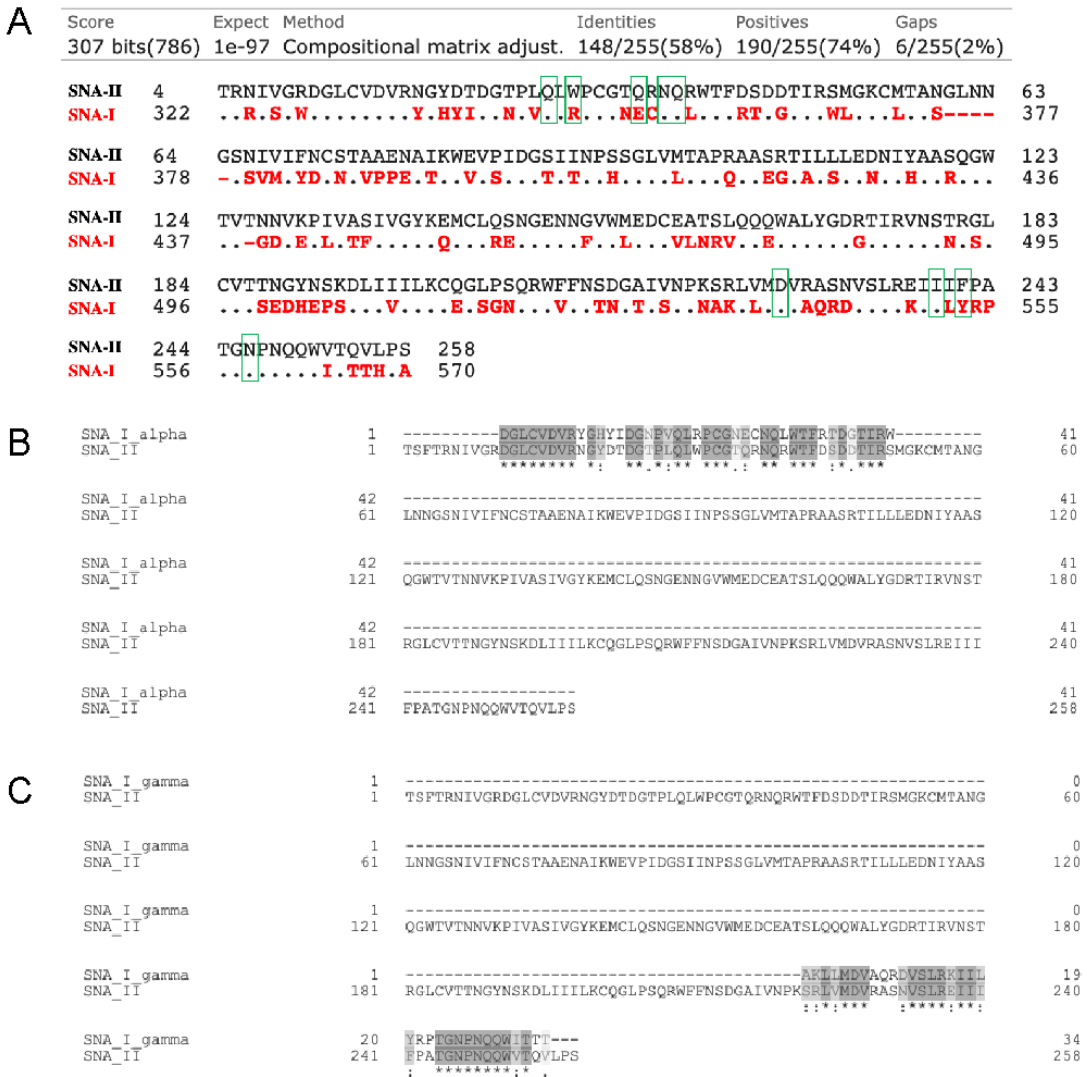


Figure 2.16 BLASTp sequence alignment of SNA-I and SNA-II. A) BLASTp alignment of the whole proteins for SNA-I and SNA-II. Amino acids that bind to the glycan receptors are outlined in green. Conserved residues are shown as dots in the SNA-I sequence. B) Local alignment of the Iα tandem repeat domain that contains the carbohydrate binding site to SNA-II. Alignment shows strong similarity to the reported Ia carbohydrate binding site of SNA-II. Dark gray boxes indicate conserved amino acids, and the light gray boxes denote amino acids of similar structure. C) Local alignment of the IIg tandem repeat domain carbohydrate binding site to SNA-II. Alignment shows strong homology to the reported IIg carbohydrate binding site of SNA-II.

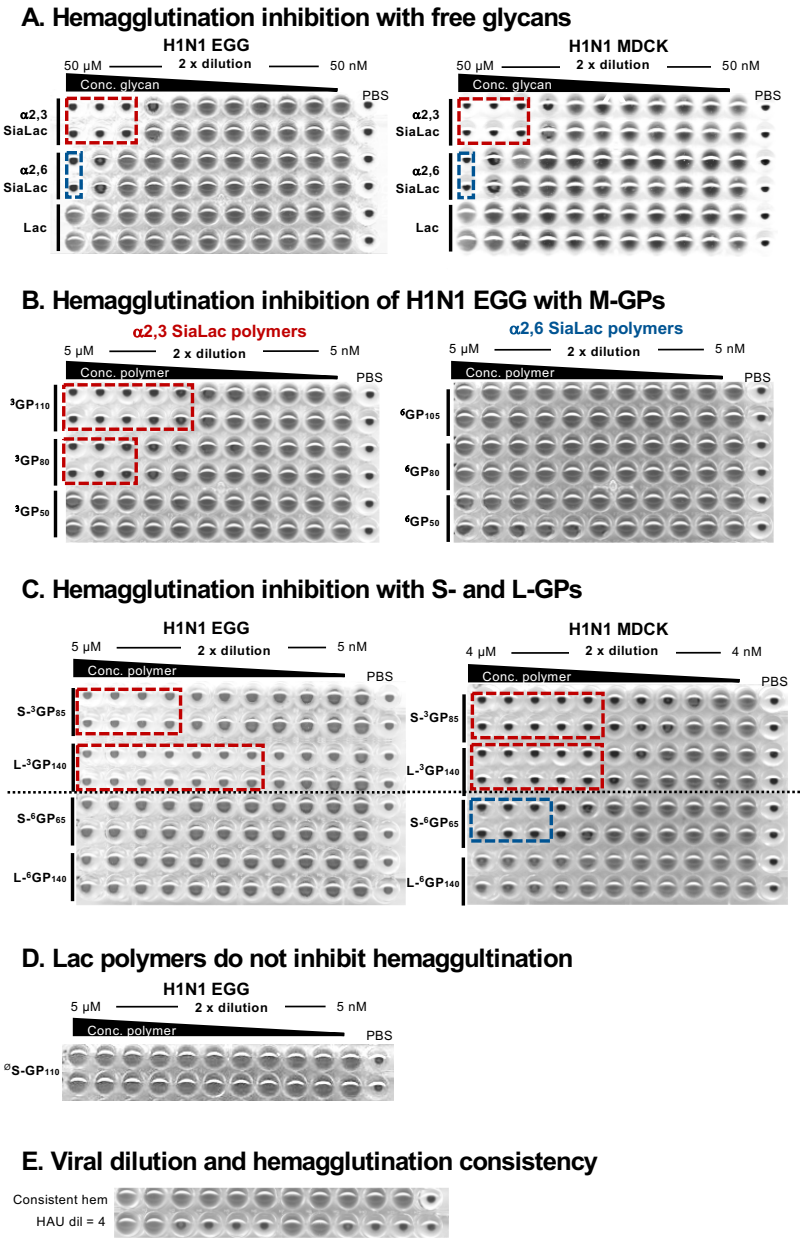


Figure 2.17 Hemagglutination inhibition assays. A) The free sialyllactose glycans inhibit H1N1 EGG and H1N1 MDCK to the same extent. B) The valency dependence **of M-^{3}GPs** can be seen for H1N1 EGG (K_i of **M-^{3}GP$_{110}$** = 313 nM, K_i of **M-^{3}GP$_{80}$** = 1.25 μM, and K_i of **M-^{3}GP$_{50}$** > 5 μM) while the **M-^{6}GPs** do not inhibit hemagglutination at any valency. C) Short and long **^{3}GPs** inhibit hemagglutination of the H1N1 EGG virus K_i of **S-^{3}GP$_{85}$** = 625 nM and K_i of **L-^{3}GP$_{140}$** = 78 nM. H1N1 MDCK virus is inhibited to equal extent by the short and long **^{3}GPs**, K_i = 234 nM and is also inhibited by **S-^{6}GP$_{65}$**, K_i = 938 nM. D) **Lac** polymers do not inhibit hemagglutination, shown for H1N1 EGG with **ØS-GP$_{110}$**. E) For all hemagglutination assays, a viral dilution with HAU = 4 was used and shown to be consistent across replicates.

Table 2.5 Hemagglutination inhibition constants. The HI constants are shown for each virus type and inhibitor used. Where inhibition was not seen, the value in the table was entered as the highest concentration used.

Virus	Inhibitor	K_i (μM)
H1N1 EGG	**α2,3-SiaLac**	13.0
H1N1 EGG	**α2,6-SiaLac**	50.0
H1N1 EGG	**Lac**	> 50.0
H1N1 MDCK	**α2,3-SiaLac**	13.0
H1N1 MDCK	**α2,6-SiaLac**	50.0
H1N1 MDCK	**Lac**	> 50.0
H1N1 EGG	**M-^{3}GP$_{110}$**	0.3
H1N1 EGG	**M-^{3}GP$_{80}$**	1.3
H1N1 EGG	**M-^{3}GP$_{50}$**	>5.0
H1N1 EGG	**M-^{6}GP$_{105}$**	>5.0
H1N1 EGG	**M-^{6}GP$_{80}$**	>5.0
H1N1 EGG	**M-^{6}GP$_{50}$**	>5.0
H1N1 EGG	**S-^{3}GP$_{85}$**	0.6
H1N1 EGG	**L-^{3}GP$_{140}$**	0.08
H1N1 EGG	**S-^{6}GP$_{65}$**	>5.0
H1N1 EGG	**L-^{6}GP$_{140}$**	>5.0
H1N1 MDCK	**S-^{3}GP$_{85}$**	0.2
H1N1 MDCK	**L-^{3}GP$_{140}$**	0.2
H1N1 MDCK	**S-^{6}GP$_{65}$**	0.9
H1N1 MDCK	**L-^{6}GP$_{140}$**	>4.0

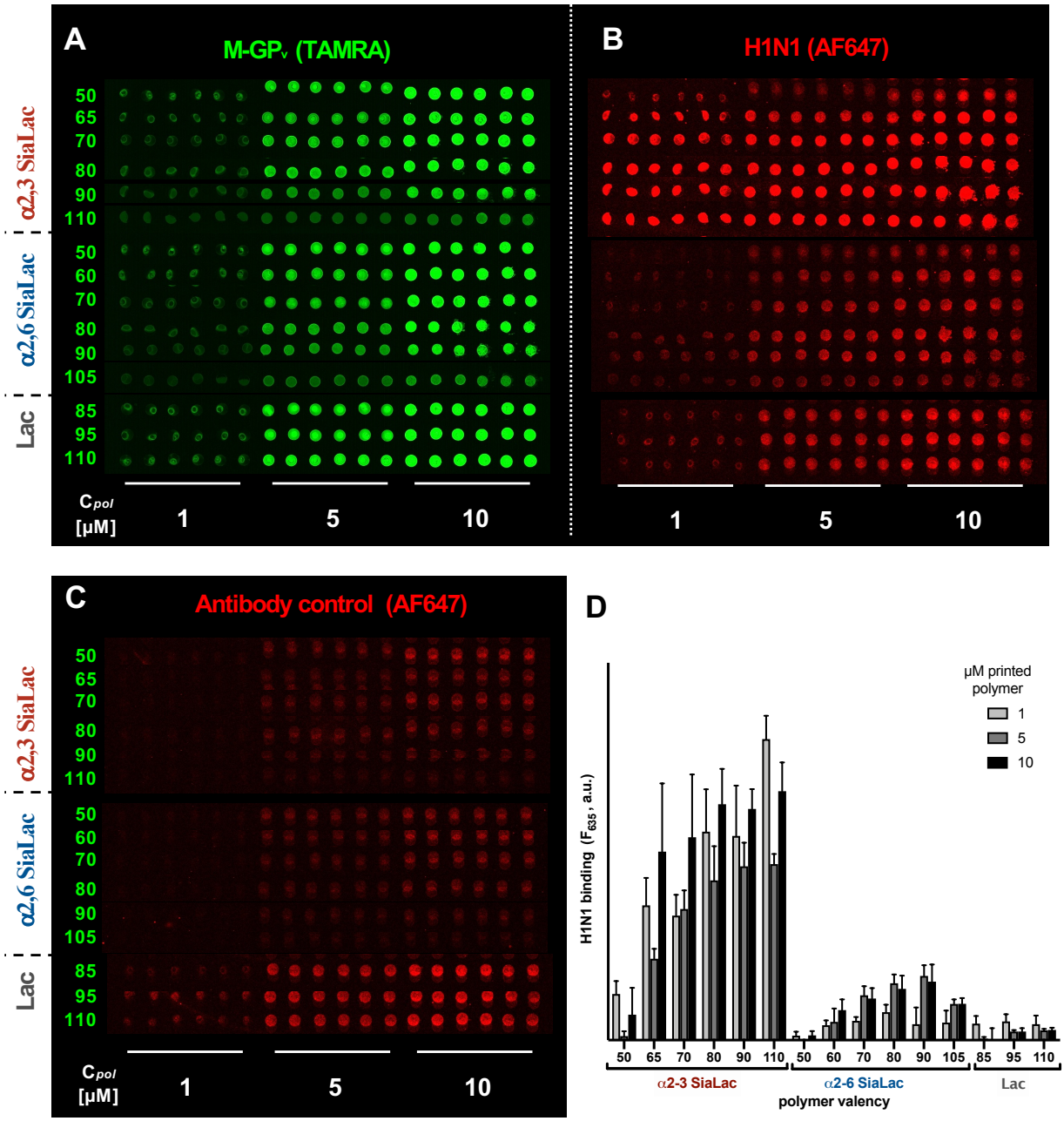


Figure 2.18 Full array fluorescence scan and bar graph representation of H1N1 EGG binding in mucin-mimetic arrays containing M-GPvs. A) Fluorescence scan (λem=532 nm) of density variant array comprising TAMRA-labeled medium (M) glycopolymers carrying α2,3-sialyllactose (^{3}GP), α2,6-sialyllactose (^{6}GP), or lactose ($^{\varnothing}$GP) printed at 1, 5 or 10 μM concentrations. B) Fluorescence scan (λem = 635 nm) of mucin mimetic array incubated with H1N1 and stained with AF647-labeled anti-HA antibody. C) Fluorescence scan (λem = 635 nm) of anti-HA antibody background binding. D) Bar graph representation of H1N1 binding intensity in mucin mimetic arrays after background signal (anti-HA only) subtraction. Values represent mean values for six replicate spots. Error bars represent standard deviation from 6 replicate spots.

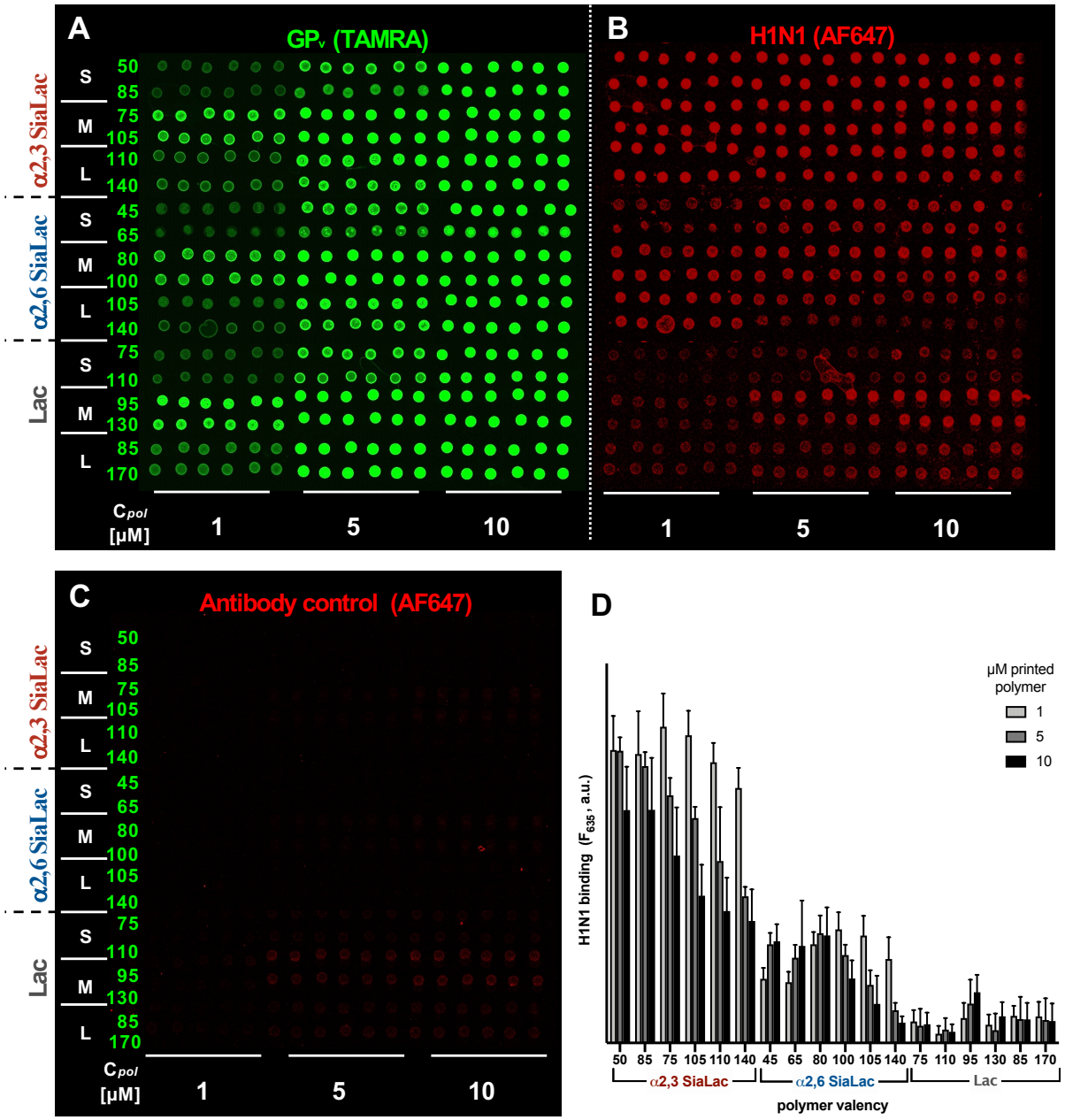


Figure 2.19 Full array fluorescence scan and bar graph representation of H1N1 EGG binding in mucin-mimetic arrays of all polymer lengths. A) Full microarray print containing S/M/L-GPs at half and full glycosylation printed in six replicates at 1, 5, and 10 μM. B) H1N1 from egg bound to the array and probed with anti-HA primary antibody and AF647 labeled secondary antibody. C) The antibody control contains both the HA primary antibody and labeled secondary and shows some background binding to the Lac polymers. D) The signal from the antibody control subarray is subtracted from the H1N1 array before plotting viral binding.

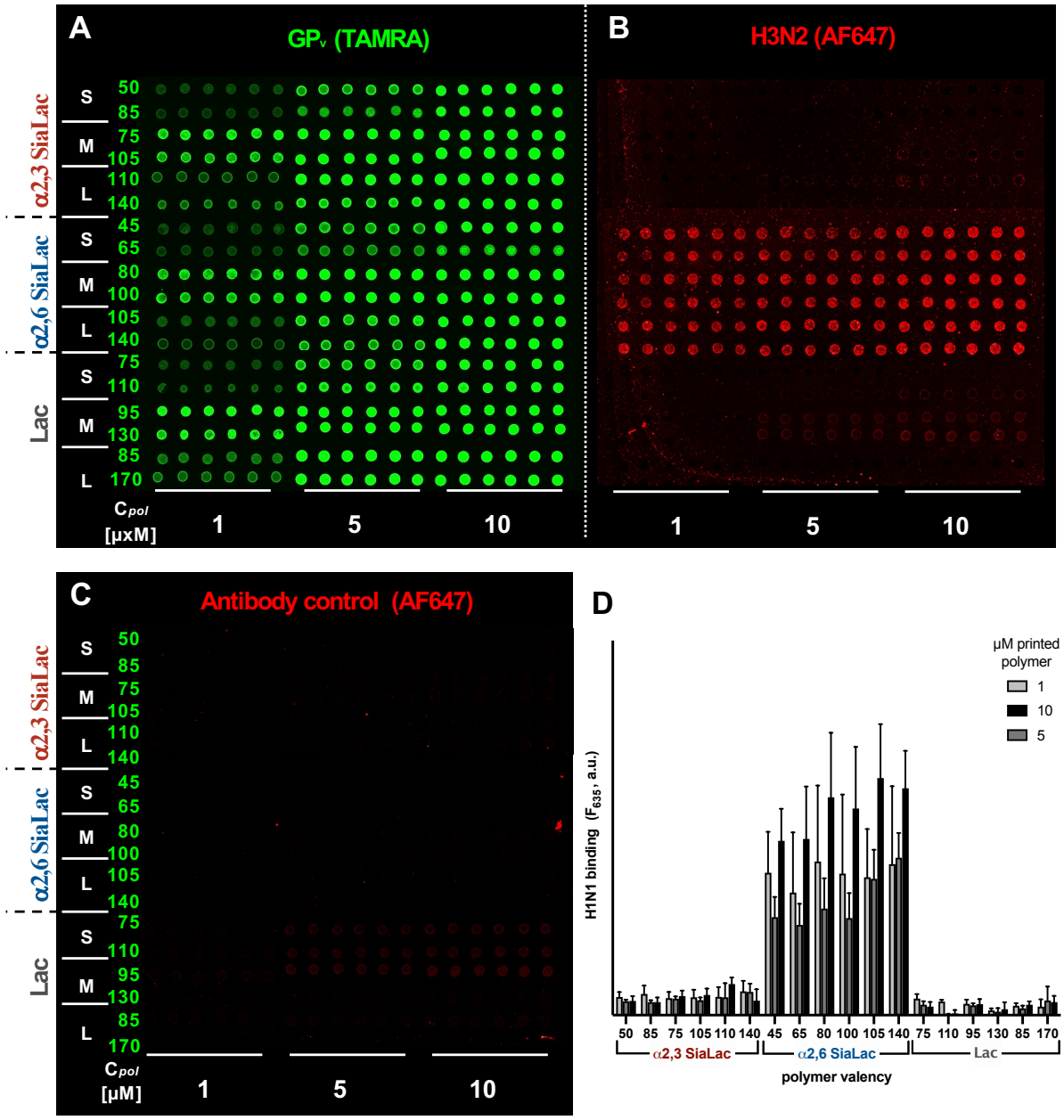


Figure 2.20 Full array fluorescence scan and bar graph representation of H3N2 MDCK binding in mucin-mimetic arrays of all polymer lengths. A) Full microarray print containing S/M/L-GPs at half and full glycosylation printed in six replicates at 1, 5, and 10 μM. B) H3N2 MDCK bound to the array and probed with anti-NP primary antibody and AF647 labeled secondary antibody. C) The antibody control contains both the NP primary antibody and labeled secondary and shows some background binding to the Lac polymers. D) The signal from the antibody control subarray is subtracted from the H3N2 array before plotting viral binding.

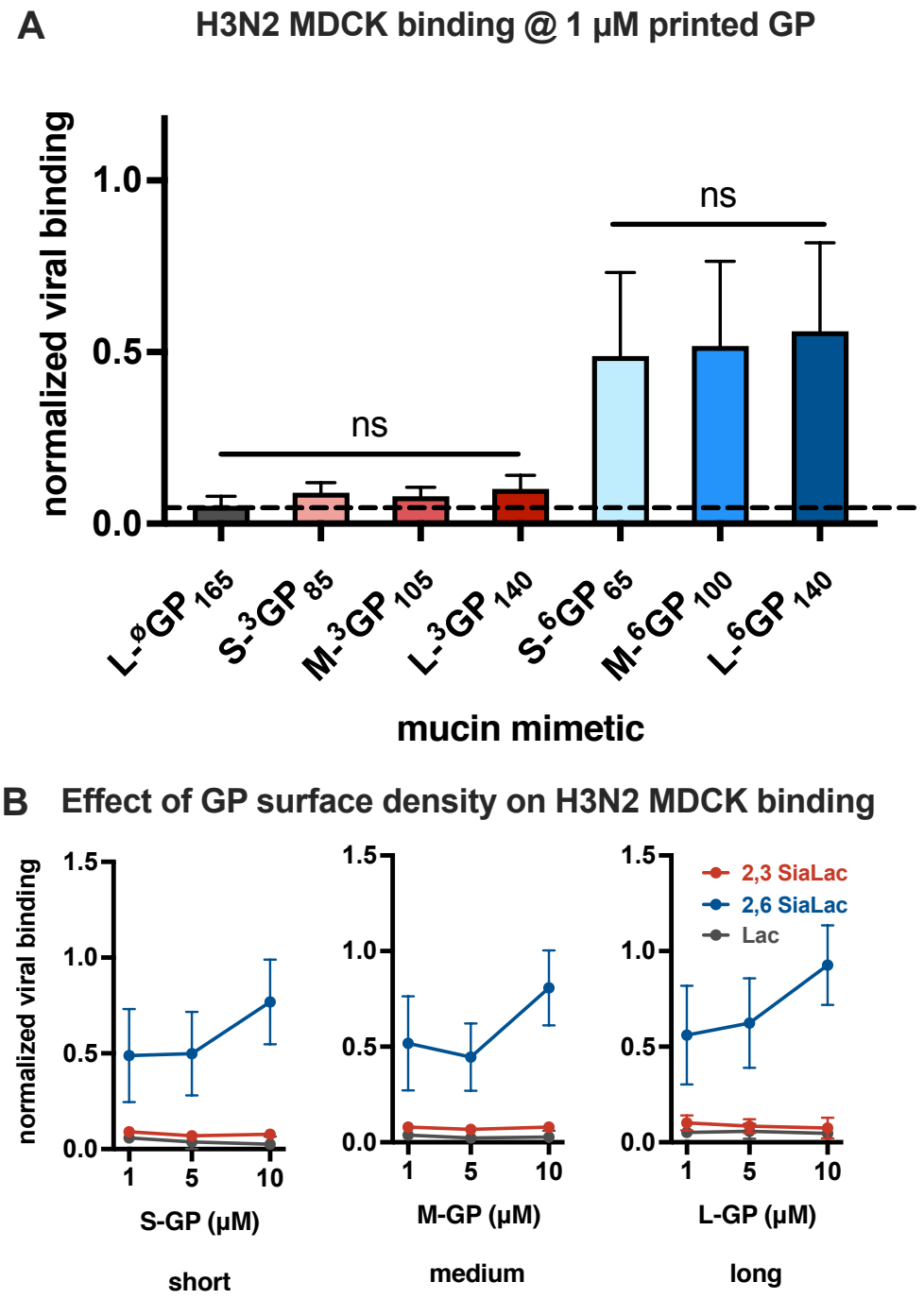


Figure 2.21 Analysis of H3N2 MDCK binding in mucin-mimetic arrays. A) Normalized binding of H3N2-MDCK to all glycopolymer lengths at 1 μM printing concentration. The dotted line signifies the Lac background. H3N2 MDCK is specific for **α2,6 SiaLac** and the extent of binding does not change with glycopolymer length. B) Binding trends for H3N2 MDCK indicates better ability of the virus to utilize increasingly dense displays of **α2,6 SiaLac**-containing **GPs** of all three lengths.

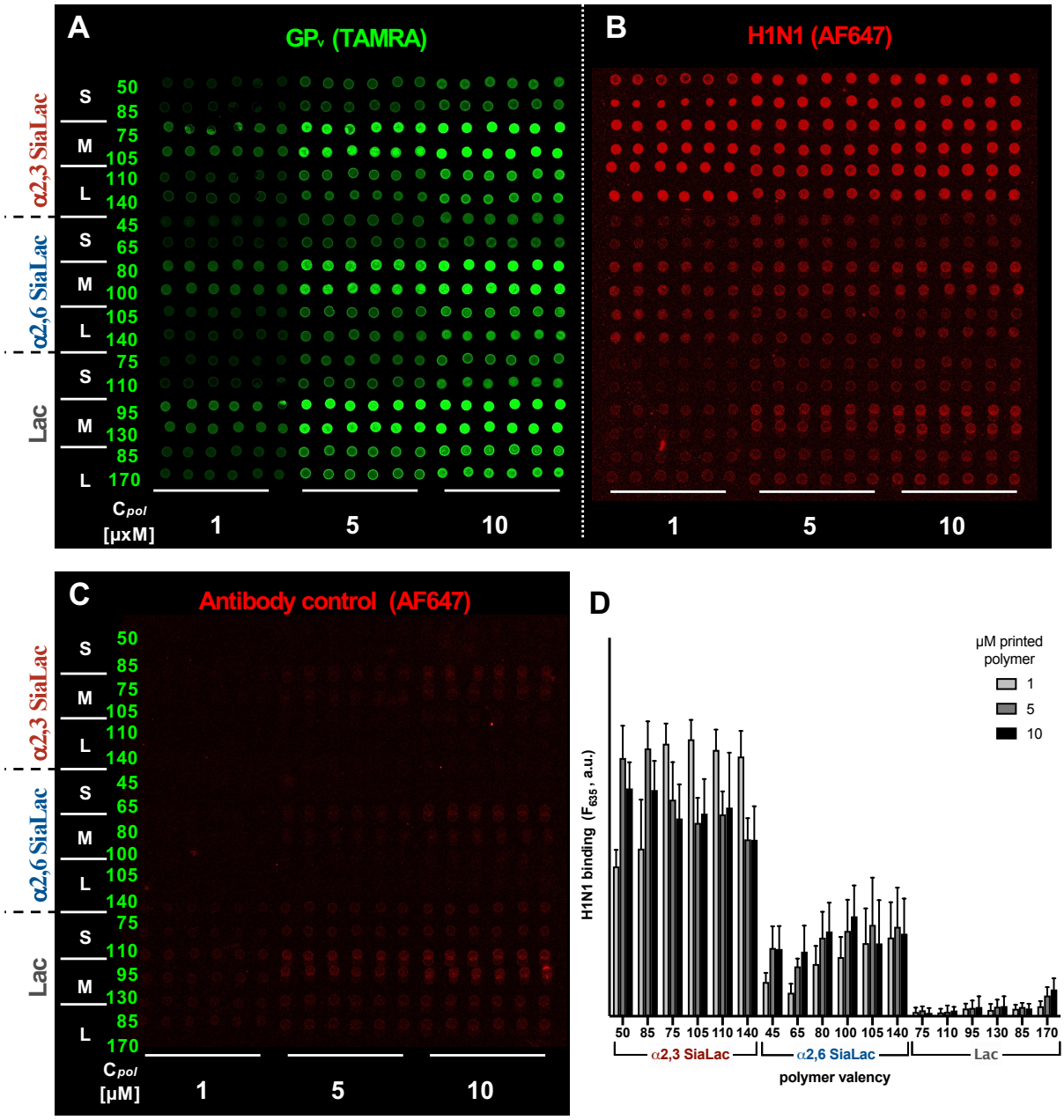


Figure 2.22 Full array fluorescence scan and bar graph representation of H1N1 MDCK binding in mucin-mimetic arrays of all polymer lengths. A) Full microarray print containing S/M/L-GPs at half and full glycosylation printed in six replicates at 1, 5, and 10 μM. B) H1N1 MDCK bound to the array and probed with anti-HA primary antibody and AF647 labeled secondary antibody. C) The antibody control contains both the HA primary antibody and labeled secondary and shows some background binding to the Lac polymers. D) The signal from the antibody control subarray is subtracted from the H1N1 array before plotting viral binding.

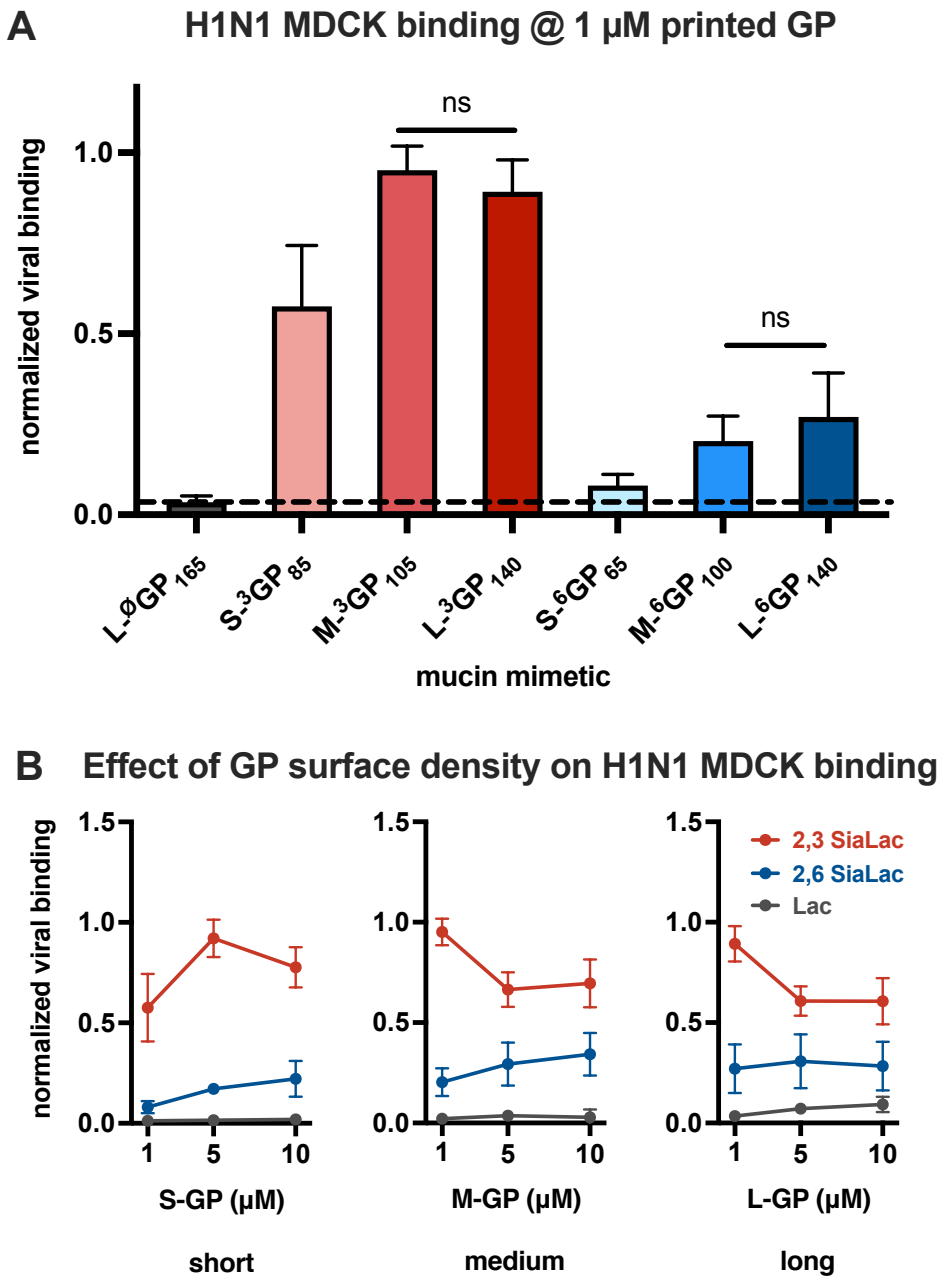


Figure 2.23 Analysis of H1N1 MDCK binding in mucin-mimetic arrays. A) Normalized binding of H1N1 MDCK to all glycopolymer lengths at 1 μM printing concentration. The dotted line signifies the **Lac** background. For H1N1 MCDK, there increasing polymer length does not negatively impact viral recognition. B) The effect of GP printing concentration on H1N1 MDCK binding shows that increasing surface density for the **α2,3 SiaLac** mimetics leads to a reduction in binding while no such relationship is observed for mimetics containing **α2,6 SiaLac**.

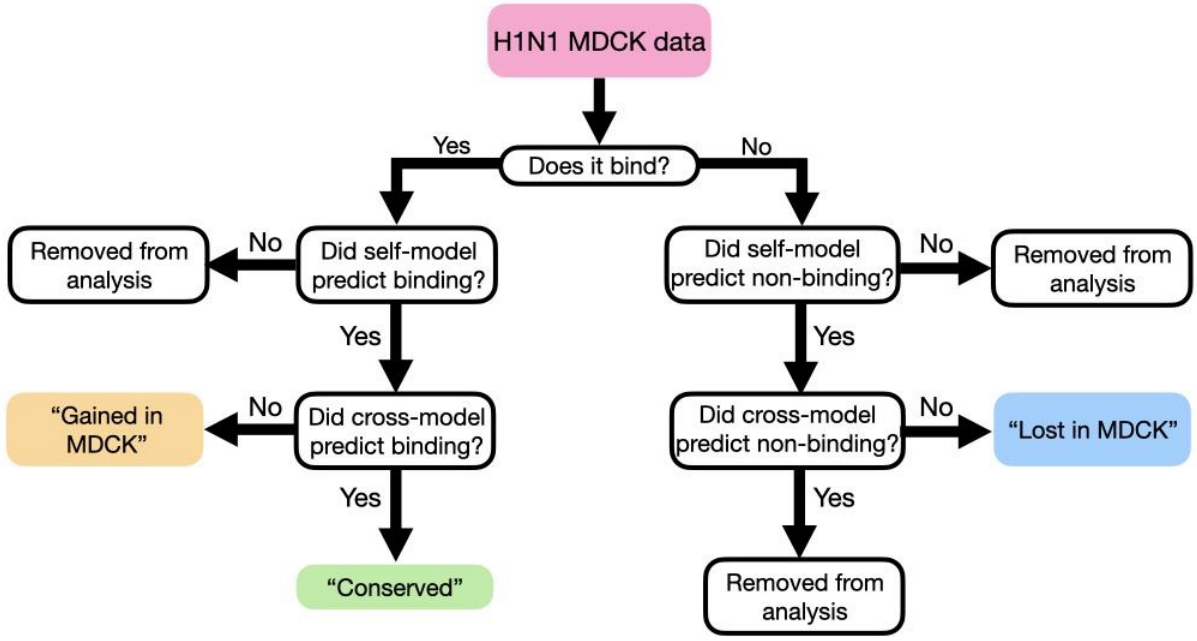


Figure 2.24 Comparison of interaction patterns. True binders (left branch) that were correctly predicted by both self- and cross-models were termed "Conserved". Those that were correctly predicted by the self-model but not by the cross-model (implying that these interactions did not exist in H1N1 EGG) were termed "Gained in MDCK". Those that were wrongly predicted by the self-model were not used in this analysis. True non-binders (right branch) that were correctly predicted by the self-model but not by the cross-model (implying that these interactions existed in H1N1 EGG but are absent in H1N1 MDCK) were termed "Lost in MDCK".

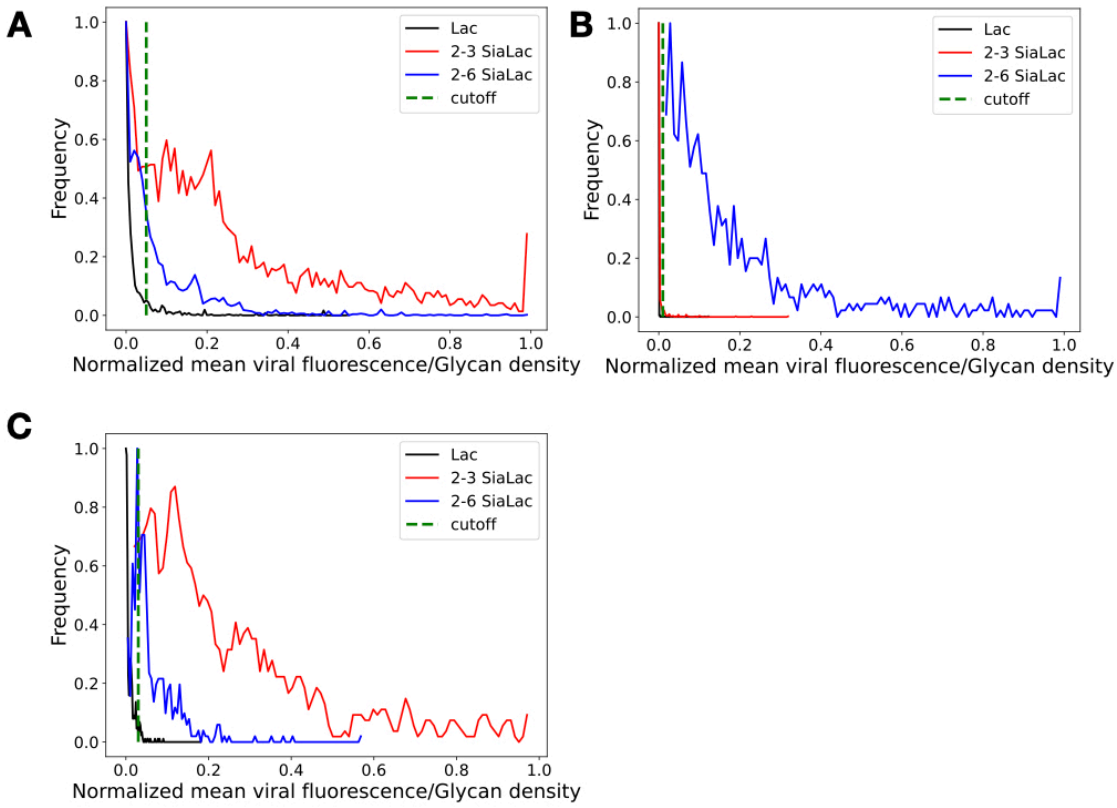


Figure 2.25 Establishing fluorescence threshold for binding. Distribution of fluorescence for lactose (black), α2,3-SiaLac (red), and α2,6-SiaLac (blue) are shown for A) H1N1 EGG, B) H3N2 MDCK, and C) H1N1 MDCK data sets. Cutoff value (broken green line) was determined to be 0.05 for H1N1 EGG, 0.15 for H3N2 EGG and 0.03 for H1N1 MDCK. The background fluorescence levels (Lac) for H1N1 and H3N2 viruses reflect the specificities of 1o antibodies used for their detection on the array.

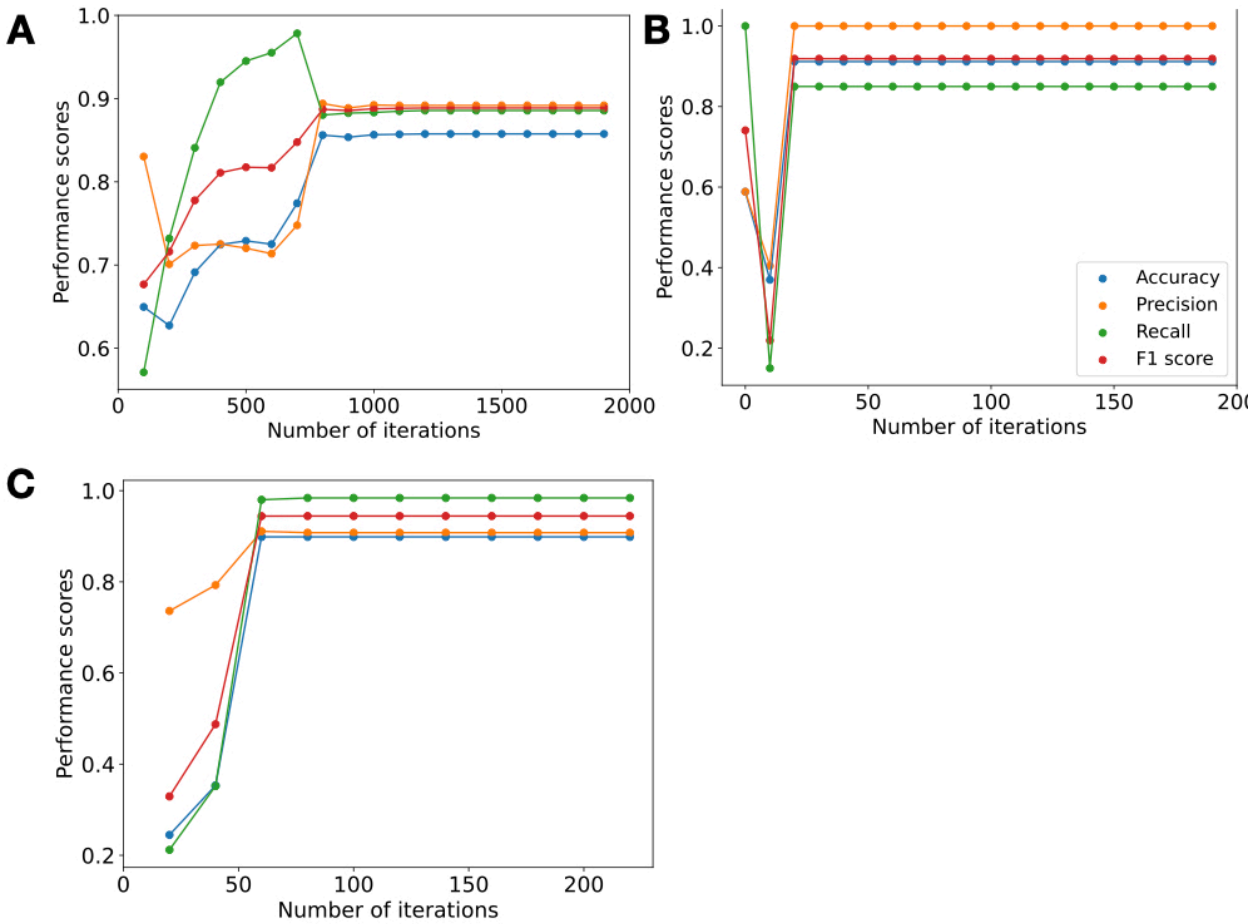


Figure 2.26 Convergence of SVM training. Prediction accuracy (blue), precision (yellow), Recall (green) and F-1 score (red) are shown as a function of number of iterations, for A) H1N1 EGG, B) H3N2 MDCK and C) H1N1 MDCK data sets. Training of the model is considered to have converged when more iterations show no further improvement of the model performance. Training with the H1N1 EGG data converged in 2000 iterations, while that of the other two data sets converged in 200 iterations. Hence, the model for the H1N1 EGG data was trained with 2000 iterations while that of the other two data sets was trained with 200 iterations.

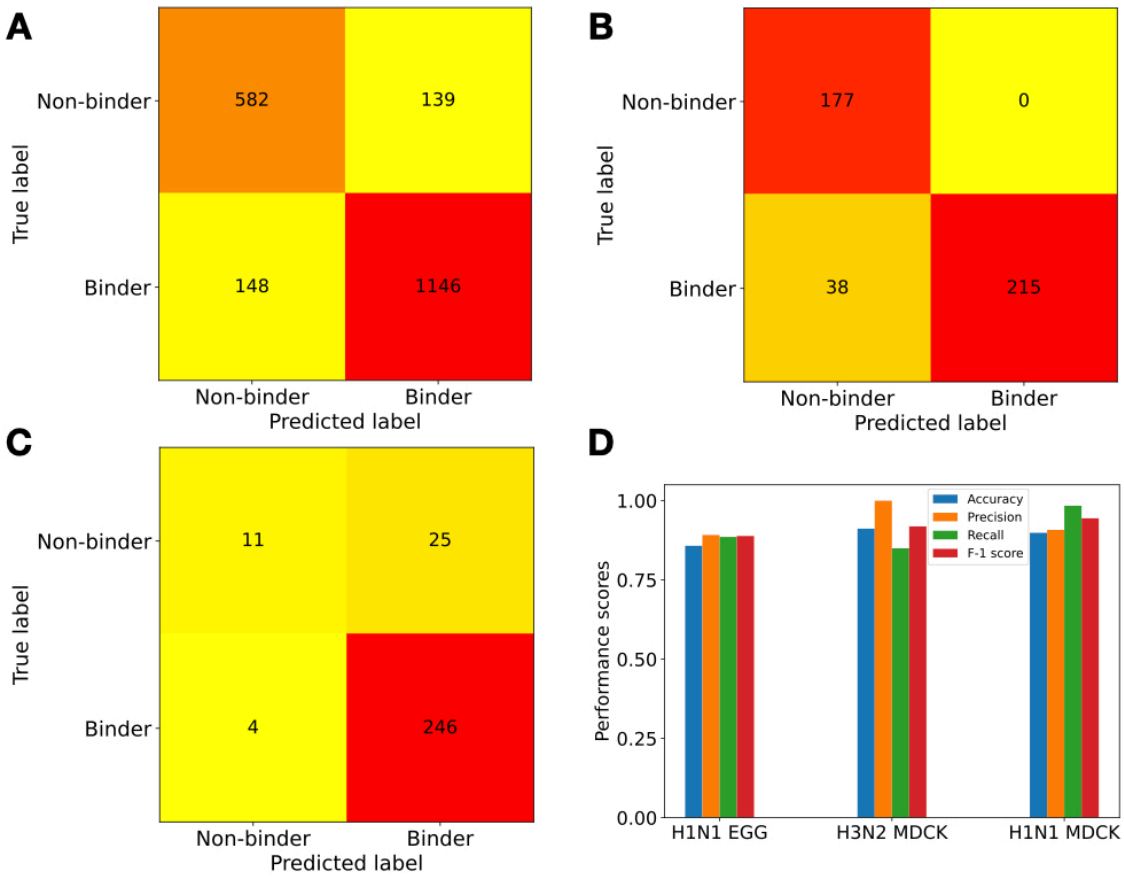


Figure 2.27 Validation of SVM testing. Confusion matrix for the prediction of testing data for A) H1N1 EGG, B) H3N2 MDCK and C) H1N1 MDCK data sets. The false positive rate of H1N1 MDCK is relatively high, presumably due to fewer non-binder labels (12% of total data set compared to 34% and 46% in H1N1 EGG and H3N2 MDCK respectively. (D) Prediction accuracy (blue), precision (yellow), recall (green) and F-1 score (red) are shown for all three data sets. For each of the trainings, accuracy and precisions were between 87% and 95%, while recall and F-1 scores were between 90% and 99%, validating the trained models.

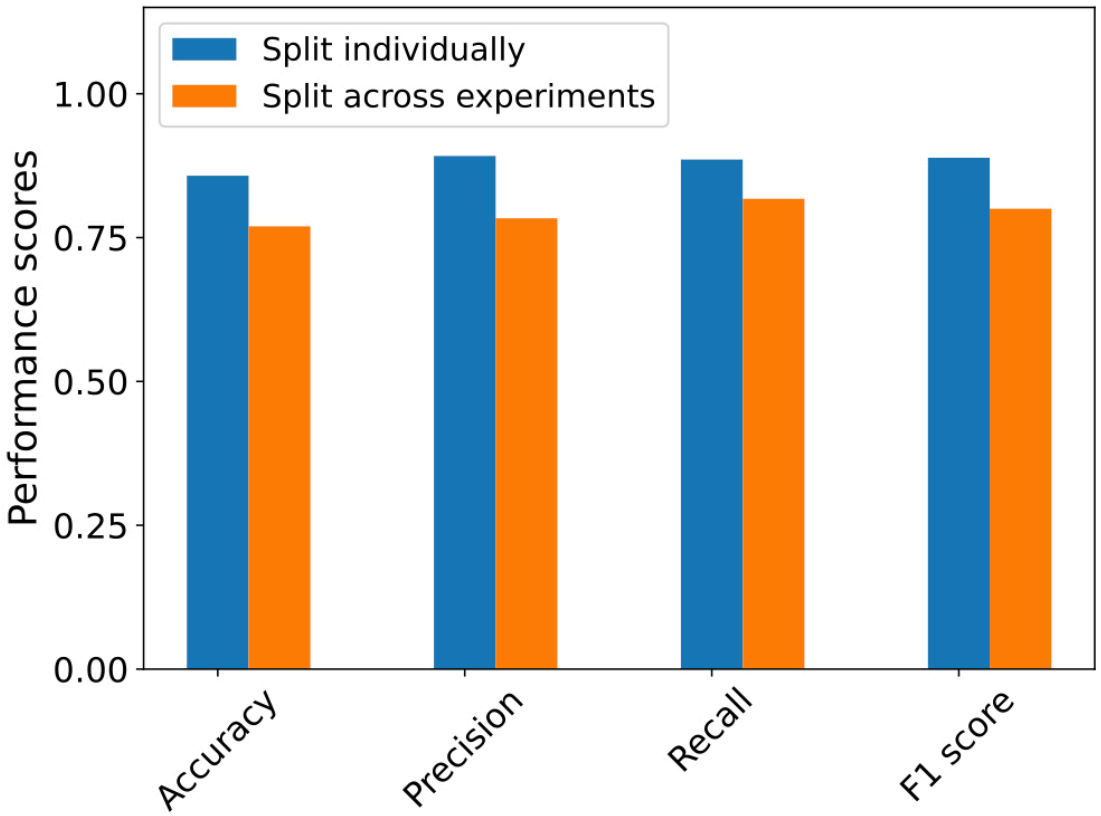


Figure 2.28 Impact of training/testing data set splitting method on SVM performance. Training and testing of H1N1 EGG was repeated, once by creating test/train splits where each experiment was individually represented in the two splits, while the other had no such criteria. Performance scores for these two predictions are shown in blue and yellow respectively. Splitting individually or across experiment yielded similar results for precision, accuracy, recall, and F1 scores. The model used for all further analyses was the one built by splitting the experiments individually.

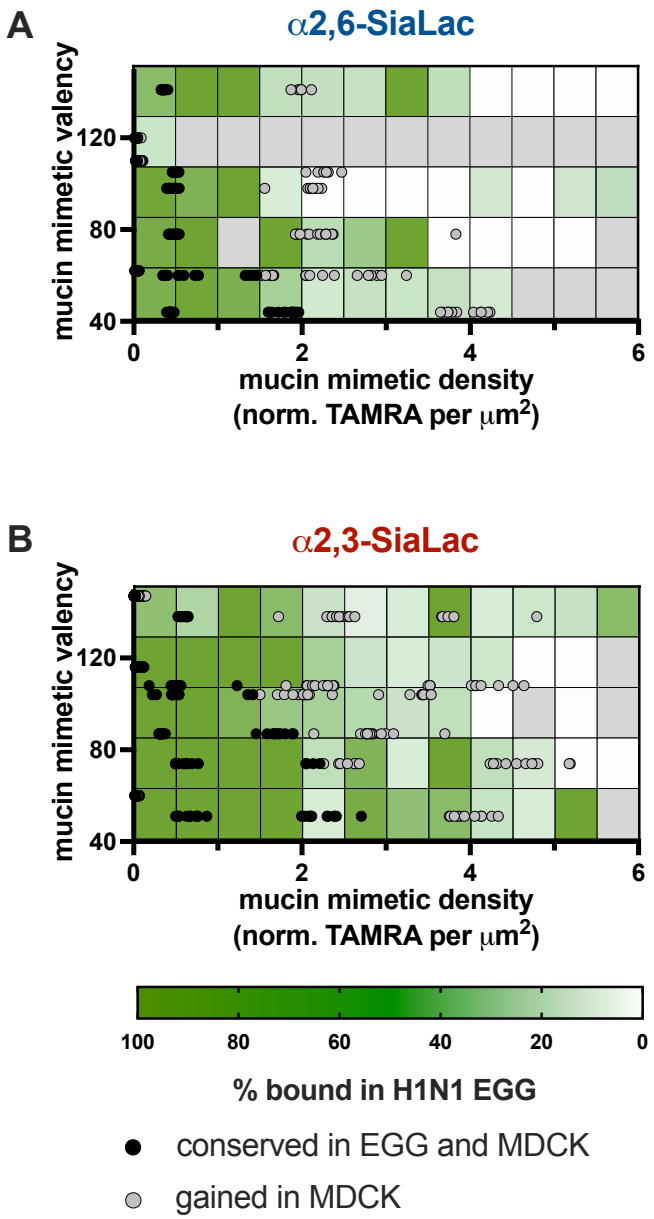


Figure 2.29 Validation of SVM cross-predictions for H1N1. The validation was performed for both **α2,6-SiaLac** A) and **α2,3-SiaLac** B) receptor types. The scatter plots represent SVM cross-predictions for conserved (black circles) and gained (gray circles) binding events when comparing H1N1 MDCK and H1N1 EGG (as shown in **Figure 6C**). The underlying heat maps show % of binding events for H1N1 EGG on the mucin mimetic array calculated by dividing the number of binding events by the sum of binding and non-binding events for each GP valency-density quadrant. Dark green squares represent a large % of binding events in the H1N1 EGG data, while white squares represent minimal binding events. The gray squares denote areas where no binding events were measured. Binding events predicted to be conserved in H1N1 MDCK overlap with the dark green regions of the heat map (high incidence of binding event for H1N1 EGG). Binding events predicted as gained in H1N1 MDCK overlap with areas where H1N1 EGG showed minimal binding (denoted by light green to white heat map cells).

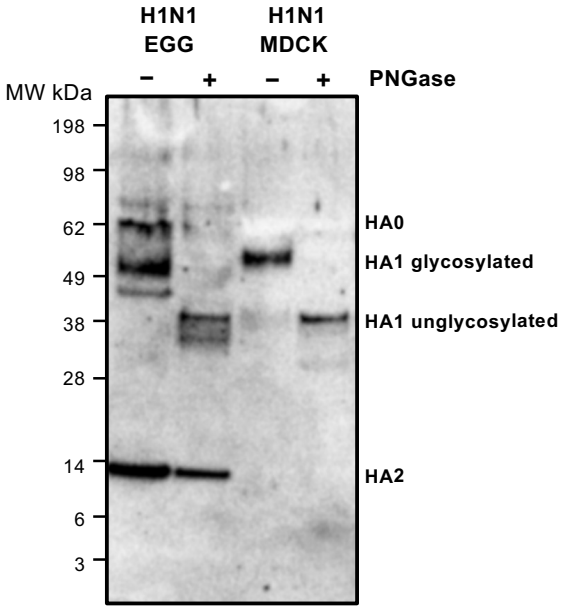


Figure 2.30 Western blot analysis of N-glycosylation of HA from H1N1. The HA of the H1N1 EGG is slightly smaller than the HA from the H1N1 MDCK. Removal of *N*-glycans through PNGase treatment yields HAs of the same molecular weight.

2.8.2 Supplemental procedures

Instrumentation

Proton nuclear magnetic resonance (^{1}H NMR) spectrum of polymer backbones and glycopolymers were obtained on either a 300 MHz (Bruker) or 500 MHz (Joel) NMR spectrometer, using deuterated solvents ($CDCl_3$ for the polyacrylaminde precursors, **P**, and sodium phosphate buffered saline (150 mM) in D_2O adjusted to pH 7.4 with DCl for the glycopolymers). The spectra were analyzed using MestReNova software and are reported in parts per million (ppm) on the δ scale relative to the residual solvent as an internal standard (for ^{1}H NMR: $CDCl_3$ = 7.26 ppm, D_2O = 4.79 ppm). Data are reported as follows: chemical shift, multiplicity (s = singlet, d = doublet, dd = doublet of doublets, t = triplet, q = quartet, br = broad, m = multiplet), and integration.

Size exclusion chromatography was performed on a Hitachi Chromaster system equipped with an RI detector and a 5 μm, mixed bed, 7.8 mm I.D. x 30 cm TSKgel column (Tosoh Bioscience). Polymer backbones **S/M/L-P** were analyzed in DMF (0.2% w/v LiBr, 70°C) using an isocratic method with a flow rate of 0.7 mL/min. Glycan conjugation reactions were performed in a Biorad MyCycler thermocycler (Hercules, CA).

Glycan microarrays were printed on cyclooctyne-coated glass slides as previously described[97] using a sciFLEXARRAYER S3 printer (Scienion) and imaged on an Axon GenePix4000B scanner (Molecular Devices) at the highest PMT possible without saturation of pixels.

Protein gels were run using the Invitrogen Mini Gel Tank (Cat no. A25977) set at constant 200 V for 22 minutes and transferred to PVDF membranes using the Invitrogen iBlot2 transfer system (Cat. no. IB21001) with P3 settings (20 V for 7 min). Following probing with antibodies, the blots were imaged using a BioRad GelDoc XRS+ imaging system.

Preparation and characterization of mucin mimetic glycopolymers GP

Preparation of polyacrylamide precursors (P)

A 10 mL flame-dried Schlenk flask equipped with a magnetic stir bar was charged with monomer **1** and CTA **2** according to ratios listed in Table 2.2. Dry dioxane was added, followed by an aliquot of AIBN stock solution (2 mg/mL, Table 2.2). The reaction volume was adjusted with dry dioxane to a final monomer concentration of 1.5 M. The yellow solution was thoroughly degassed through at least three freeze-pump-thaw cycles and backfilled with N_2. The flask was then heated to 65˚C and allowed to react for 8-12 hrs as

listed in Table 2.2. The reaction was stopped at ~ 60-70% conversion by precipitation into hexanes. The resulting yellow solid was dissolved in DCM and precipitated again into hexanes. This purification step was repeated twice. The final polymer was concentrated from $CHCl_3$ to remove residual hexanes and dried under vacuum to give polymers **P** as pale-yellow solids. To achieve different polymer lengths, the monomer to CTA ratio was adjusted based on the theoretical degree of polymerization (DP) and assuming 70% consumption of monomer (see Table 2.2). All **P** polymers were analyzed by SEC in DMF.

Preparation of chain end-deprotected polymer precursors (P-1)

The polymer backbones **P** were dissolved in anhydrous THF to give a 2 mM solution. Next, a solution of n-butylamine in THF (20 mM, 10 equiv. per trithiocarbonate end group) was added and stirred on ice for 2-3 hr. The product was precipitated in excess hexanes three times before concentrating in $CHCl_3$ and drying to give polymers as a white solid (yield: 85-95% for **P-1**).

Preparation of TAMRA labeled polymer precursors (P-2)

This step was carried out immediately after end deprotection. A 2 mM solution of fluorophore-maleimide (1.5 eq., TAMRA per polymer end group) in dry DMF was used to dissolve end deprotected polymers (**P-1**) in a 4 mL vial equipped with a stir bar. The reactions were stirred in the dark overnight. The reaction mixtures were then diluted in ether before being precipitated in excess hexanes to remove unreacted fluorophore. The mixtures were centrifuged (1000 xg, 3 min) and the hexanes was decanted and replaced with fresh hexanes so the precipitation could be repeated. Finally, the products were

concentrated using $CHCl_3$ and dried under vacuum to give a dark pink solid (**P-2**). (Yields: 94-96% for **P-2**)

Preparation of side chain deprotected polymer precursors (P-3)

The TAMRA labeled polymers were charged into 4 mL vials and dissolved in 500 µL of 1 M TMS-Cl and 3 M phenol mixture freshly prepared in anhydrous DCM. The solutions were stirred for 2 hr in the dark. The product was precipitated into ether three times from DCM, redissolved in MilliQ water and purified on a PD10 column. The product was the lyophilized overnight to give polymers (**P-3**) as a fluffy pink solid. (Yields: 70-75%)

Preparation of mucin mimetic glycopolymers GP

The lyophilized deprotected polymers were dissolved in sugar ligation buffer (1M NaOAc, 1M urea, pH 4.5) to achieve a solution where the side chain concentration was kept constant at 200 mM. The solutions were then added to PCR tubes containing 0.1 to 1.1 equivalents of **α2,3-SiaLac**, **α2,6-SiaLac**, or **Lac** according to Table 2.3 and heated at 50˚C for 72 hr on a thermocycler. These reactions were purified using a Amicon Ultra Centrifugal Filter (pre-washed with Milli-Q water twice by spin dialysis, 10K MWCO, Millipore) and spin dialyzed (10000 xg, 15 min) five times using 500 µL of deuterated phosphate buffered saline solution (100 mM phosphate, 150 mM NaCl, pD 7.4). The glycan ligation efficiency was determined from the ^{1}H-NMR spectra of the glycopolymers by subtracting polymer backbone protons (total of 7 protons, found between 2.5-4.5 ppm) from the total integration in the region 2.5-4.5 ppm and dividing the difference by the

number of glycan protons which are also found between 2.5-4.5 ppm. Glycan valency was then calculated as the product of the degree of polymerization and ligation efficiency. To determine glycopolymer labeling efficiency, UV-Vis experiments were performed using a quartz cuvette (10 mm path length) in a Nanodrop2000c spectrophotometer (ThermoFisher). The TAMRA absorbance at a wavelength of 555 nm was measured for a known concentration (by weight) of glycopolymer. The fluorophore concentration of the solution was obtained using Beer's Law, and the fluorophore labeling efficiency was calculated by dividing the fluorophore concentration by the known polymer concentration. This value, which was determined for each polymer backbone, was then used to adjust the polymer concentration obtained in the UV-Vis measurements. All of the glycopolymer characteristics are displayed in Table 2.3.

Calculations for polymer end-to-end length in solution

The polymer length in solution was calculated as previously described[98] based on the theoretical distance between polymer chain termini (R), assuming a hard sphere model for atoms and imposing the excluded volume effect. The calculation is as follows:

$$Equation\ 2.1. \qquad R = \left(\frac{2}{3}\right)^{0.5} N^{0.6} C^{0.2} b^{0.4} \left(\frac{4\pi b^3}{3}\right)^{0.2}$$

where *N* is the number of C-C bonds between the side chain (DP x 2), *C* is the polymer characteristic ratio (8.5 for polyacrylamide),[99] and *b* is the C-C bond length (0.154 nm).

Mucin mimetic array preparation and analysis with SNA and WGA plant lectins

Construction and characterization of density-variant mucin mimetic arrays

All pre-print passivation solutions, printing buffers, and the system liquid were first filtered through 0.22 µm filters. For initial cyclooctyne functionalization, epoxysilane slides (Thermo Scientific SuperChip Microarray Slides) were incubated in a Coplin jar with rocking in a 1 mM solution of dibenzocyclooctyne (DBCO) amine (BroadPharm) in DIPEA/DMF overnight. Slides were then sonicated two times for 15 minutes in MeOH, rinsed with MilliQ, and spin-dried (550 rpm for 5 min). They were stored at 4˚C with desiccant until printing.

On the day of printing, the slides are passivated with a 1% BSA/0.1% Tween-20 solution in PBS for 1 hour with rocking, then washed with MilliQ three times for 15 min and spin-dried. A Scienion sciFLEXARRAYER S3 printer was used to print 6 replicates of polymers diluted in printing buffer (0.005% Tween-20 in PBS) at a humidity of 70%. After printing, the slides were re-humidified over an 80˚C water bath, snap dried on a glass plate set at 80˚C, and allowed to react at 4˚C overnight.

After outlining the subarrays with a glass cutter and snap drying again at 80˚C, the slides were immediately washed vigorously in a 0.1% TritonX/PBS solution for 2 minutes and rocked for another 15 minutes. The slides were next washed in PBS two times for 10 minutes, rinsed with MilliQ, and spin-dried. The slides can be stored at 4˚C with desiccant until further use.

Lectin binding in density variant mucin mimetic arrays

For lectin binding, the slides were washed for 15 minutes in a 0.1% Tween-20/PBS solution on a shaker and then spin-dried. They were then imaged at the highest PMT possible without saturation using an Axon GenePix 4000B microarray scanner (Molecular Devices). A gasket was added, and the subarrays were blocked with 3% BSA solution in PBS for 1 hour at RT, then washed three times with lectin binding buffer (LBB: 0.005% Tween-20 in PBS with 0.1 mM $CaCl_2$, $MnCl_2$, and $MgCl_2$). Dylight labeled SNA and WGA was diluted in the LBB and incubated on the array for 1 hour at room temperature with rocking and in the dark. The labeled lectin is removed, and the subarrays are washed two times with LBB, two times with the 0.1% Tween-20 solution, rinsed with MilliQ and spin-dried. The slides were imaged at the highest PMT possible without producing saturated pixels.

Array image collection and processing

GenePix Pro v7 software was used to image and analyze the microarrays. A block was constructed with the same dimensions as the printed array, aligned over the spots in the image, and analyzed. To determine the amount of polymer immobilized on the slide, the mean background subtracted 532 nm signal was used. Dividing this signal by the labeling efficiency and multiplying by the polymer valency provided the amount of relative glycan per spot (Fig S2). The spot area was calculated by using the spot diameter from the generated results page. Dividing the relative glycan amount by the spot area provided the relative glycan density per μm^2. The signal at 635 nm after background subtraction was used to analyze binding of Dylight649-SNA and WGA or H1N1 PR8 probed with

AF647-anti-HA Ab. Here, an AF647 anti-HA antibody only staining was used to determine background. An anti-NP primary antibody followed by an anti-rabbit-AF647 secondary antibody was used to visualize H3N2. Both primary and secondary antibodies were used to determine background signal for H3N2.

Crystal structures of WGA and SNA lectins

The lectin cartoons generated for Fig 2.3C were adapted from crystal structures of WGA complexed with a bivalent sialoglycan peptide[76] (PDB: 2CWG) and SNA-II complexed with lactose[80] (PDB: 3CA4). In the cartoons in Fig 2.3, ovals were used to approximate the backbone of the protein from the crystal structures and the binding sites were designated as gray diamonds.

WGA exists as a dimer, and while it is most commonly known for binding GlcNAc, it can also bind sialic acid, as shown in the crystal structure. SNA-I has not been crystalized, but it is a related lectin to SNA-II that exists can exist in multiple oligomeric states (monomer, dimer, or tetramer) and is specific for α2,6-linked sialic acid.[79]

We also performed a BLASTp analysis of the SNA-I binding site with that of SNA-II to determine the homology between the receptor binding domains.

Inhibition of H1N1 agglutination of RBCs by free glycans and soluble glycopolymers (GPs)

Hemagglutination inhibition (HI) assay procedure.

The viral stock solution was diluted to an HAU = 4 for HI experiments. This HAU was tested to ensure that it consistently hemagglutinated RBCs in the absence of soluble inhibitors. Glycopolymer solutions were diluted to the same starting concentration of 20

μM polymer (200 μM for the monovalent glycans) in the first lane of a 96 well plate, and a 2-fold dilution to a total volume of 25 μL was performed down the plate. The last well was used as a PBS only control. Then 25 μL of the viral dilution (HAU = 4) was added and incubated for a ½ hour at room temperature. After which time, 50 μL of a 1% turkey RBC solution was added to all wells. The assay was carried out in duplicate and read after ½ hour.

H1N1 and H3N2 binding to density variant mucin-mimetic arrays

H1N1 and H3N2 viruses were isolated and utilized neat from embryonated chicken eggs allantoic fluid or MDCK viral culture supernatant by incubating on the array for 1 hour at room temperature with rocking. The slide was washed two times with BSA/PBST. H1N1 was fixed to the array with a 20 minute incubation with 4% PFA. This was followed by an incubation of a 1:500 dilution of anti-HA in BSA/PBST on the array for 1 hour at room temperature with rocking. For H3N2, the virus was fixed on the array with a solution of 4% PFA for 10 minutes. Then the membrane was permeabilized in 70% ethanol for 10 min so that a 1:500 dilution of NP antibody could be incubated on the array for 1 hour at room temperature with rocking. Following primary antibody incubation, the slides were washed two more times with BSA/PBST and a 1:500 dilution of anti rabbit-AF647 antibody was incubated in the dark with shaking. The subarrays were washed two more times with BSA/PBST, two times with the 0.1% Tween-20 solution, rinsed with MilliQ, spin dried, and imaged at the highest PMT possible without producing saturated pixels.

Creation of SVM models for IAV binding

Establishing a fluorescent threshold for binding

Since the virus is known to be a non-binder to lactose, the lactose fluorescence was used as negative control. Based on the distribution of normalized fluorescence intensities of all 3 types of glycans (lactose, α2,3-SiaLac and α2,6-SiaLac), the cutoff was determined to be the least normalized fluorescence value for which lactose fluorescence was negligible.

Convergence of SVM training

Performance of binary classification, namely, accuracy, precision, recall, and F-1 score, were defined as:

$$accuracy = \frac{True\ positive + True\ negative}{All\ data}$$

(fraction of correctly classified data, calculated using the following equation)

$$precision = \frac{True\ positive}{True\ positive + False\ positive}$$

(fraction of total predicted positives that are true)

$$recall = \frac{True\ positivee}{True\ positive + False\ negative}$$

(fraction of all positives that are correctly predicted)

$$F1\ score = 2 * \frac{Precision * Recall}{Precision + Recall}$$

(balance between precision and recall)

Convergence of SVM training was defined when these quantities no longer changed with further iterations.

Validation of SVM training

SVM algorithm trained on the training data set (67% randomly selected data points from the data set) was used to predict the remaining, testing data set (33% of the data set). The results are shown in the form of a confusion matrix where the X-axis is the prediction output, and the Y-axis is the experimental outcome (**Fig 2.27**). Number of correct predictions are shown across the diagonal (correctly predicted non-binding shown on top left, correctly predicted binding shown on bottom right). False positives (predicted binding for points that showed non-binding in experiment) are shown top right of **Fig 2.27**. False negatives (predicted non-binding for points that showed binding in experiment) are shown on the bottom left of **Fig 2.27**.

Western blot analysis of PNGase treated H1N1

To prepare samples for PNGase treatment, 6 μL of IAV were combined with 1 μL of Glycoprotein Denaturing Buffer and 3 μL of MilliQ in PCR tubes. Samples were denatured at 100 ^{o}C for 10 min, then chilled on ice and centrifuged briefly (10 sec). The reaction volume of each sample was adjusted to 20 μL by adding GlycoBuffer 2 (2 μL), 10% NP-40 (2 μL), and of MilliQ (6 μL). Then 1 μL of PNGase (500 units) was added and

the tubes were mixed gently (1 μL of MilliQ was added to non-PNGase treated samples). The samples were incubated at 37 °C for 4 hr and placed at 4 °C until running the gel.

To prepare the samples for the gel, an equal volume (20 μL) of 2.5% loading dye with 0.4 M DTT was added and they were incubated at 95 °C for 10 min. A 4-12% Bis Tris gel was loaded with 15 μL of sample (8 μL of ladder was loaded to terminal wells). The gel was run at a constant 200 V for 22 min in 1x MES buffer, after which time it was rinsed with water and transferred to a PVDF membrane using the P3 settings (20 V for 7 min) on the iBlot2.

Following transfer, the membrane was block with a 5% blotto solution in PBST (0.1% Tween-20) for 1 hr at room temperature. Following rinses with PBST, the membrane was probed overnight at 4 °C with a 1:500 dilution of anti-HA antibody in PBST. The following day, it washed for 15 min (3 x 5 min) with PBST and incubated with a 1:2000 dilution of anti-rabbit IgG, HRP-linked antibody for 1 hr at room temperature. The membrane was washed again for 15 min (3 x 5 min) in PBST. The HRP substrate was added for 30 sec and the blot was imaged on a BioRad GelDoc XRS+ imager.

2.9 Acknowledgments

Chapter two, in in full, is a reprint of the material as it appears Mucin-mimetic glycan arrays integrating machine learning for analyzing receptor pattern recognition by influenza A viruses in Chem Cell Press. Lucas, Taryn M.; Gupta, Chitrak; Altman, Meghan, O.; Sanchez, Emi; Nattichia, Matthew, R.; Gagneux, Pascal; Singharoy, Abhishek; Godula, Kamil., Elsevier, 2021. The dissertation author was the primary researcher and author of this material.

3 Design of a self-reporting IAV probe

3.1 Introduction

The previous chapter focused on viral binding interactions that occur through HA. This chapter will concentrate on neuraminidase (NA), the protein that catalyzes enzymatic release of the virus. This protein exists as a tetramer made up of four identical subunits and extends from the viral membrane through an approx. 60 Angstrom long stalk segment.[100] The head of the protein, which sits atop the stalk, contains the catalytic site of the enzyme. There is one catalytic site for each of the four monomers that project outward, allowing NA to hydrolyze the glycosidic bond between a terminal sialic acid residue and the next glycan (typically galactose) on glycoproteins and glycolipids. While phylogenic mapping revealed two distinct groups of NAs among strains, all enzymes have extremely conserved active sites, including eight highly preserved residues (Arg118, Asp151, Arg152, Arg224, Glu276, Arg292, Arg371, and Tyr406) that interact directly with the sialic acid substrate. While the exact catalytic mechanism remains unclear, it is proposed that the charged residues form salt bridges to stabilize sialic acid in a conformation that allows Tyr406 to attack, resulting in eventual release of the sugar.[59]

Release of IAV from its receptor is essential to propagate infection.[101] As many virions are released from an infected cell at the same time, they coat themselves in host's cell membrane. For successful infection to have occurred, this same membrane needed to express the viral receptor on its surface. Without an active NA enzyme to cleave sialic acids, the progeny virions would form aggregates and be unable to propagate infection.

Neuraminidase is also useful at the start of infection when the virus is moving through the upper airways.[102] The epithelial cell surface is coated with both membrane-bound and secreted mucin glycoproteins which are a main component in mucus. The

mucin proteins in mucus are heavily glycosylated with decoy sialic acid receptors for IAV, and mucus can protect the underlying cells from infection by trapping the virus for its removal as mucus is sloughed away. NA activity makes it possible for IAV to release itself from the decoys in mucus so that it can reach its cell-surface receptor for internalization.

Because of the highly conserved nature of NA active sites and its essential function in initiating and propagating infection, this enzyme is a good candidate for inhibitor design as a therapeutic. In fact, NA inhibitors have been used in the past to prevent efficient viral release early a person's illness. Unfortunately, IAV has utilized its error prone PCR to introduce mutations close to the NA active site to limit the effectiveness of these inhibitors.[103] This same error prone reproduction can also cause mutations in HA. To obtain large enough quantities of virus to study, propagation is often necessary, and we have shown that even one passage in an alternate host can lead to changes in IAV binding preferences that can be assessed through machine learning,[104] making it difficult to investigate the true binding specificity of the original viral sample. To solve this problem, we envisioned the construction of a self-reporting IAV array composed of the mucin mimetic glycopolymers for measuring viral binding specificity and a fluorogenic neuraminidase activatable probe for detection. Because the NA enzyme can turn over many times, theoretically a single virus particle can be detected (Figure 3.1).

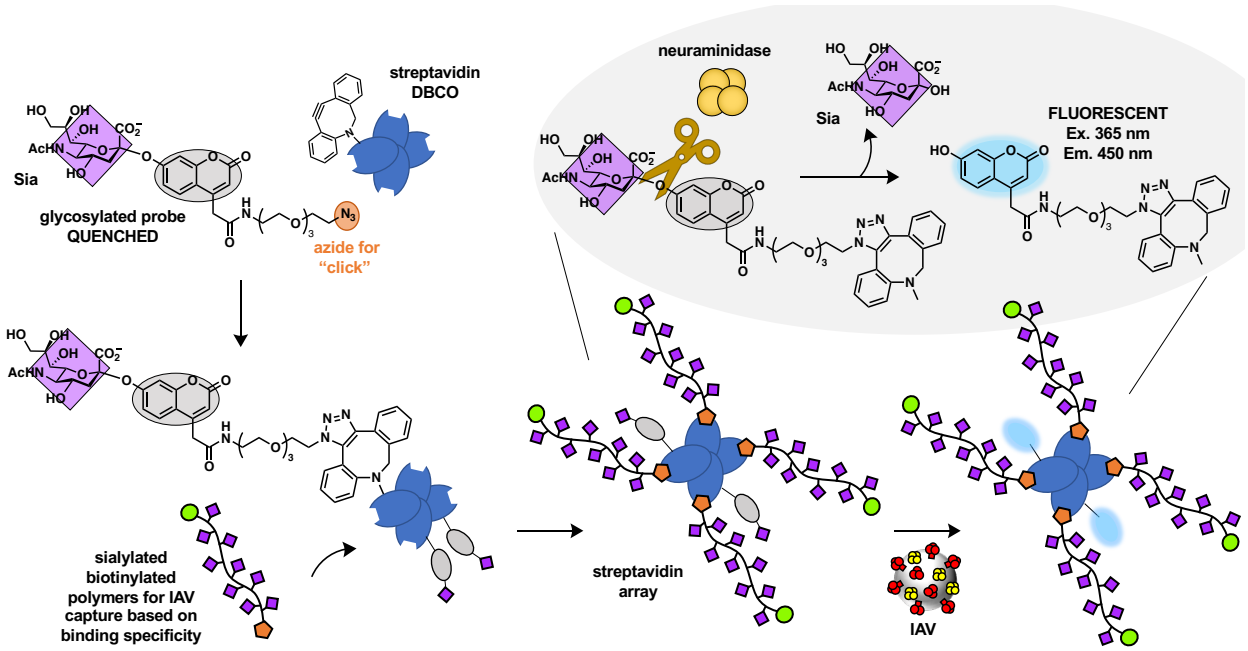


Figure 3.1 Design of a self-reporting IAV array. The design begins with the installation of a quenched probe onto dibenzocyclootyne (DBCO)-modified streptavidin. The DBCO moieties are added through NHS chemistry, leaving the biotin binding sites open for biotinylated glycopolymers containing sialic acid which can be used in affinity capture for IAVs based on their sialic acid binding specificity in a soluble streptavidin array. When the soluble array with bound IAV is heated, NA becomes active and releases sialic acid from the quenched probe to generate fluorescent signal.

3.2 Synthetic design and characterization of probe

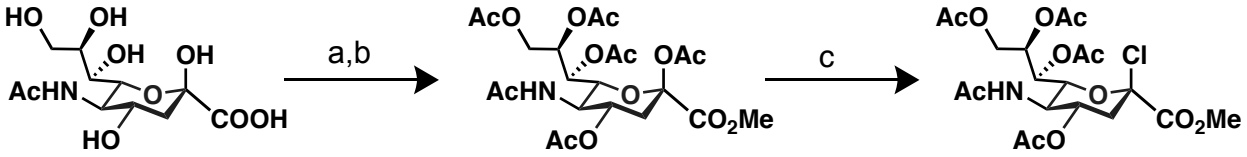
Previously made NA probes,[105] including the commonly used 4MU-NANA standard,[106] were used as inspiration for the probe. The synthetic design utilizes a coumarin-based probe quenched by the addition of sialic acid which is the substrate for NA. Once the sialic acid is cleaved, the coumarin becomes fluorescent due to the removal of the electron withdrawing sugar that allows the oxygen in the 7-hydroxy position to donate electrons to the conjugated π system (**Fig 3.1**). An azide handle was also added through a polyethylene glycol (PEG) linker for use in immobilization.

Synthesis of the probe occurred in three blocks that were then conjugated together to yield the probe containing a protected sialic acid. Following deprotection of the sugar, the final fluorogenic probe was generated. A brief description of the synthesis of each

block are below, starting with (1) functionalization of sialic acid and followed by (2) construction of the coumarin fluorophore and (3) synthesis of the azide-functionalized PEG linker. All NMR spectra for associated with the final probe and intermediates can be found in the Appendix for Chapter 3.

3.2.1 Protection and functionalization of sialic acid

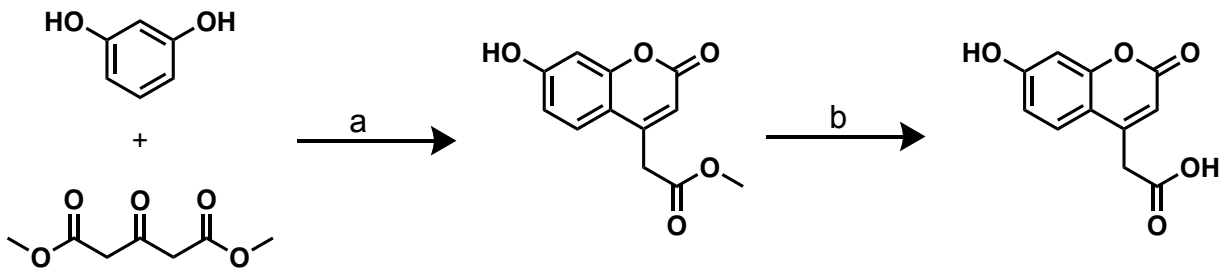
Protection of the many hydroxy groups on sialic acid with acetates was required at the start of the synthesis to mask their reactivity (**Scheme 3.1**). Additionally, the carboxylic acid was converted into a methyl ester.



Scheme 3.1. Protection and functionalization of sialic acid. a) MeOH, Amberlite-H^+. b) pyridine, Ac_2O. DMAP. c) AcCl, DCM, MeOH, -20 °C to RT

3.2.2 Construction of the coumarin fluorophore

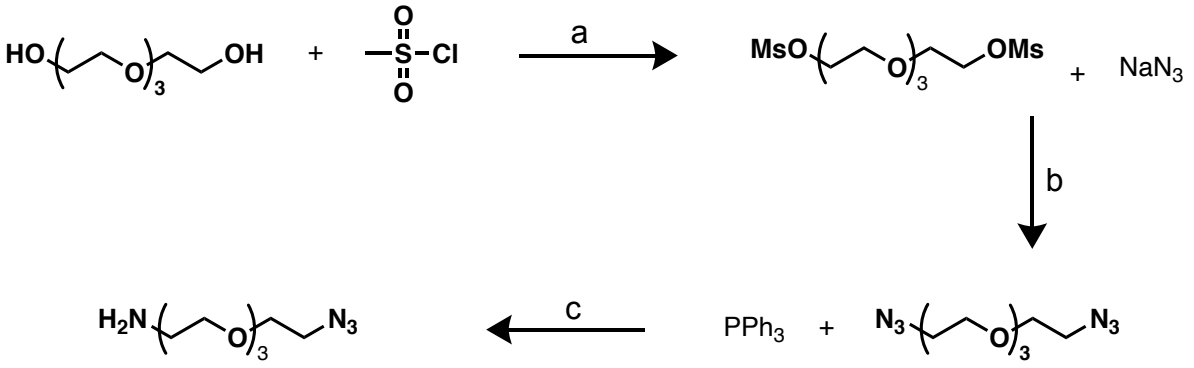
The formation of the coumarin fluorophore was adapted from previously published procedure.[107] The starting materials resorcinol and 1,3-acetonedicarboxylate were condensed in the presence of methanesulfonic acid to yield methyl(7-hydroxycoumarin-4-yl)acetate. Treatment with lithium hydroxide afforded the carboxylic acid that was used for attachment with the linker (**Scheme 3.2**).



Scheme 3.2. Construction of coumarin fluorophore. a) $MeSO_3H$, 50 °C, 4 hr. b) LiOH, THF, RT, 4 hr

3.2.3 Synthesis of the azide-functionalized PEG linker

The formation of the azido-PEG-amine linker began with the transformation of tetraethelyne glycol to the di-mesylate to allow for nucleophilic attack by sodium azide.[108] One of the azides was selectively reduced to an amine using triphenylphosphine[109] to give the linker in good yield (**Scheme 3.3**).

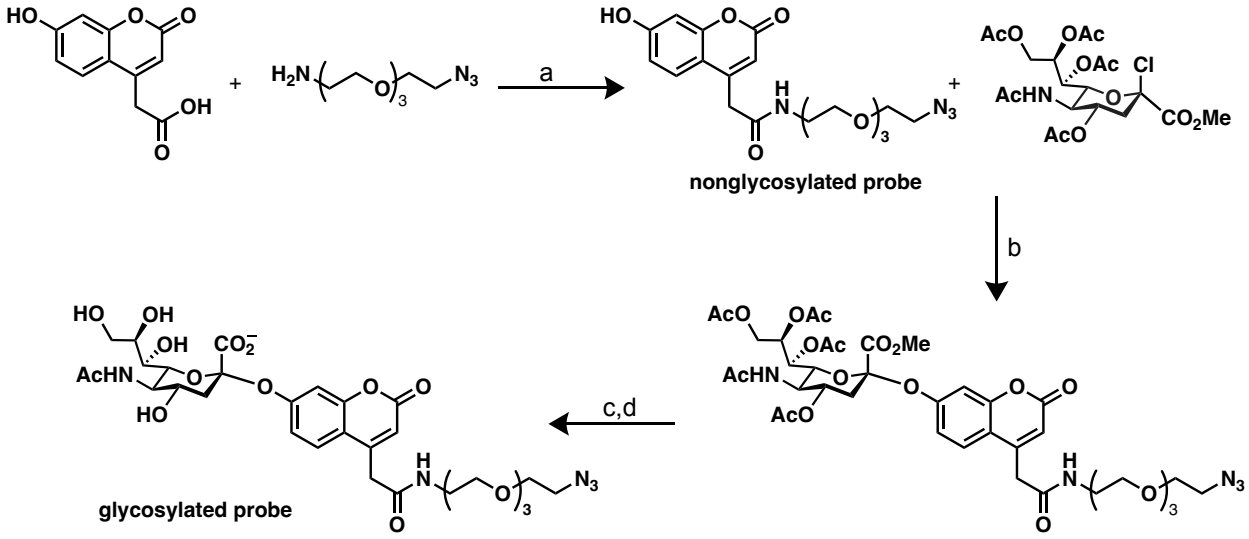


Scheme 3.3. Synthesis of the azido-PEG-amine linker. a) Et_3N, DCM, 0 °C to RT, overnight. b) DMF, 80 °C, overnight. c) 1M HCl, Et_2O, 0 °C to RT, overnight.

3.2.4 Putting the pieces together

To connect the coumarin and the linker, the coupling agent PyBOP was used.[110] This molecule is referred to as the nonglycosylated probe and is used as a positive control because it fluoresces at 450 nm when excited with a wavelength of light of 365 nm (**Fig 3.2**).

To create the fluorogenic probe, the protected sialic acid chloride dissolved in acetonitrile and mixed with the nonglycosylated probe in the presence of DIPEA.[105] The resulting sialic acid conjugate was deprotected using sodium carbonate and lithium hydroxide (**Scheme 3.4**).



Scheme 3.4. Synthesis of nonglycosylated and glycosylated probe. a) PyBOP, DIPEA, DMF, ACN, 3 hr. b) DIPEA, ACN, 5 hr, RT. c) Na_2CO_3, MeOH, 2 hr. d) LiOH, H_2O,2 hr

3.3 Neuraminidase successfully cleaves probe to turn on signal

Before moving forward with the construction of a self-reporting IAV array, the ability of neuraminidase for *Clostridium perfringens* to process the glycosylated probe (**Fig 3.2**). To do so, first the fluorescent profile of the nonglycosylated probe was determined so that the proper excitation and emission wavelengths could be chosen to detect processing. The coumarin-based molecule had an excitation maximum at 324 nm and an emission maximum at 365 nm (**Fig 3.2A**), which, as expected, is very close to that of 4-MU-NANA (Ex. of 365 nm and Em. of 450 nm). Fluorescent signal was generated in a concentration

dependent manner only in the presence of NA. Importantly, the probe was stable in the acidic (pH=4.5) buffer required for NA activity.

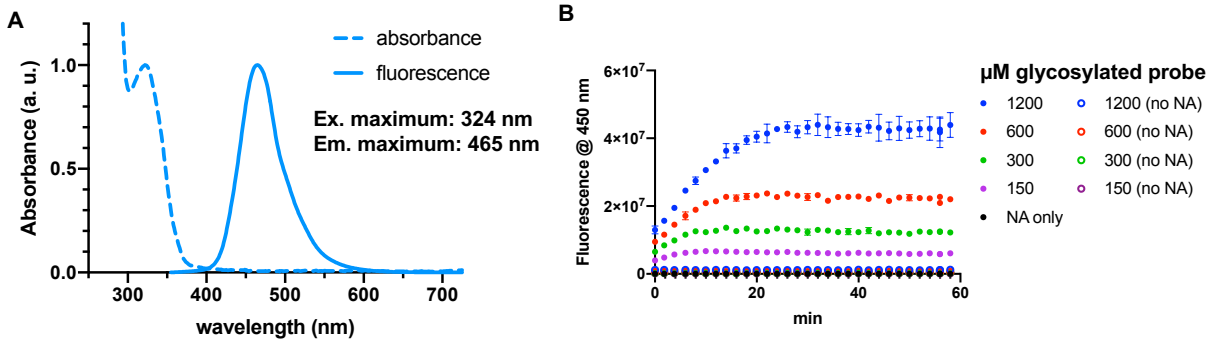
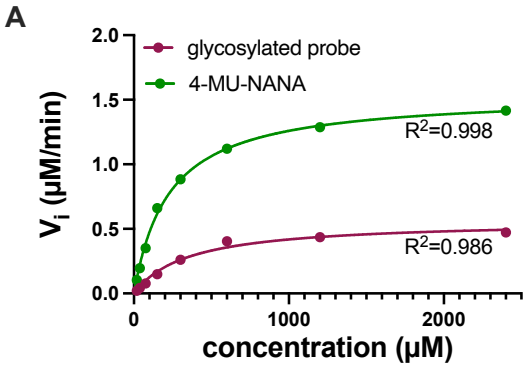


Figure 3.2 Spectral properties of fluorogenic NA probe. A) The excitation and emission wavelength for the nonglycosylated probe were empirically determined to be 324 and 465 nm, respectively. Because 4-MU-NANA was used as a control in many of the studies, its ex/em wavelength of 365 and 450 nm was routinely used. B) Incubating the glycosylated probe with NA form *C. perf* generated fluorescent signal at 450 nm in a concentration dependent manner. The probe was also stabile to background hydrolysis in the NaOAc buffer (pH=4.5) over the course of the experiment, as no signal was generated when NA was not added.

Next, the kinetic profile of the probe was determined and compared to 4-MU-NANA (**Fig 3.3**). The linear range of fluorescent turn on was established, and a standard curve using the nonglycosylated probe and/or 4-MU was generated to calculate the amount of product formation. The slope of these plots gives the rate of the enzyme cleavage. When plotted as a function of probe concentration, the curve can be fit and the Michaelis-Menten constant can be extracted (**Fig 3.3B**).



B

	glyosylated probe	4-MU-NANA
V_{max} (µM/min)	0.565	1.544
K_m (µM)	351.3	225.9
k_{cat} (min^{-1})	3.45	9.43
k_{cat}/K_m ($\mu M^{-1} * min^{-1}$)	0.009	0.042

Figure 3.3 A comparison of enzyme kinetics with 4-MU-NANA. A) Michaelis-Menten plots for the glycosylated probe and 4-MU-NANA. Each condition was run in duplicate, and the curves were fit using nonlinear regression with the Michaelis-Menten model. B) The best fit values for each curve.

Based on the derived enzyme kinetics, NA from *C. perf.* has a lower affinity for the glycosylated probe compared to 4-MU-NANA, as seen by the lower K_m for 4-MU-NANA, in addition to the larger k_{cat} and value for catalytic efficiency (k_{cat}/K_m).

3.4 Influenza A successfully cleaves probe to turn on signal

The probe was made with the intention of indicating the presence of IAV in a sample and so it is essential to show activity against viral NA. When tested against H1N1 EGG and H1N1 MDCK, the glycosylated probe once again generated fluorescent signal in a concentration dependent manner (**Fig 3.4**).

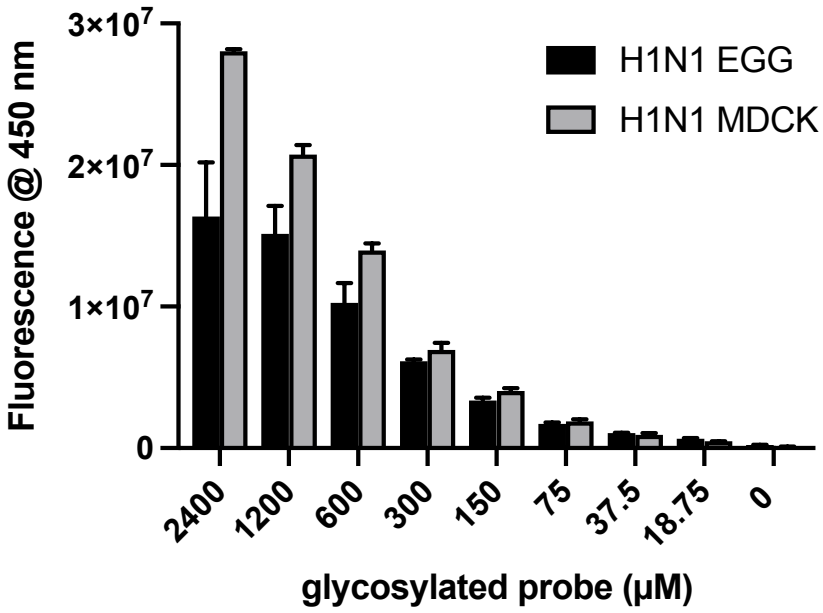
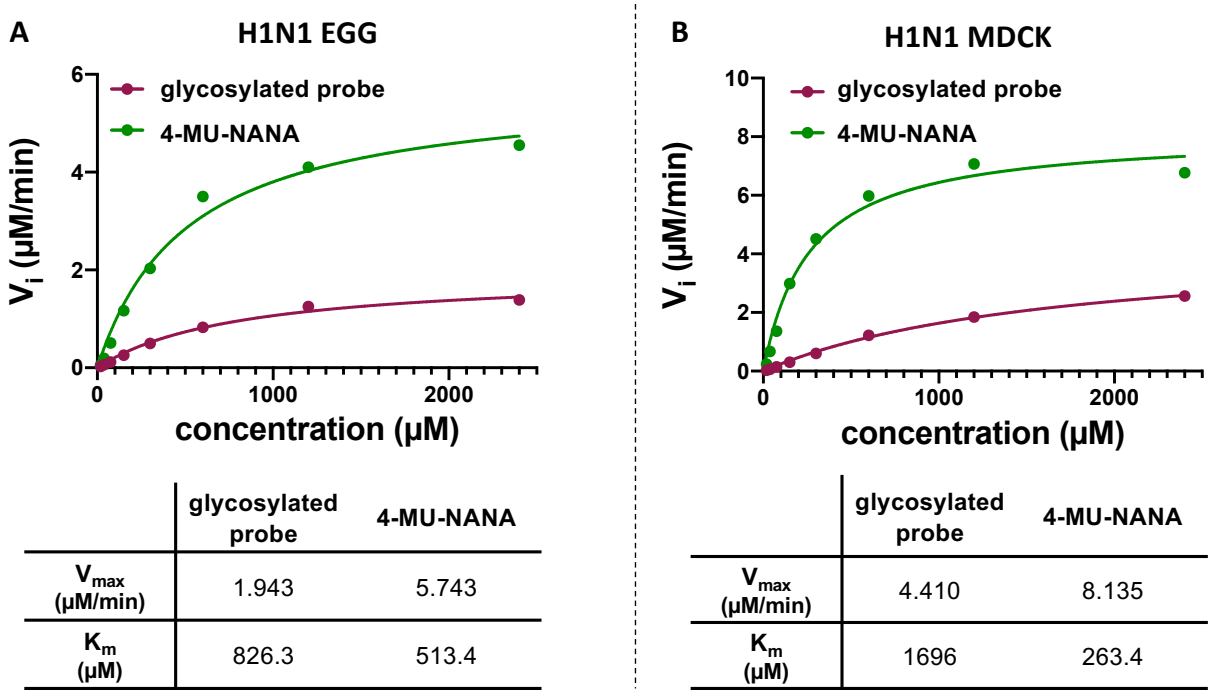


Figure 3.4 Viral cleavage of the glycosylated probe. H1N1 EGG and H1N1 MDCK were incubated at 37 °C with the indicated concentrations of the probe and the fluorescence was read after 1 hr.

Like NA from *C. perf.,* H1N1 EGG and H1N1 MDCK also displayed a lower affinity for the glycosylated probe than for 4-MU-NANA (**Fig 3.5**). The probed was designed for incorporation into a streptavidin array, where the mucin-mimetic glycopolymers would

capture the IAV based on receptor binding specificity. Heating to 37 °C would activate NA, allowing it to cleave the probe to generate the fluorescent signal that would report on the presence of the virus. IAV having a low affinity for the probe may be beneficial in limiting background from non-specific viral cleavage.



	glycosylated probe	4-MU-NANA
V_{max} (μM/min)	1.943	5.743
K_m (μM)	826.3	513.4

	glycosylated probe	4-MU-NANA
V_{max} (μM/min)	4.410	8.135
K_m (μM)	1696	263.4

Figure 3.5 A comparison of enzyme kinetics with 4-MU-NANA. A) Michaelis-Menten plots for the glycosylated probe and 4-MU-NANA. Each condition was run in duplicate, and the curves were fit using nonlinear regression with the Michaelis-Menten model. B) The best fit values for each curve.

3.5 Building the streptavidin array

The probe was clicked onto streptavidin that was functionalized with DBCO-PEG4-NHS. MALDI mass spectrometry analysis indicated that each monomer in the streptavidin tetramer was functionalized with at least one DBCO when 4 equivalents of the NHS coupling reagent were used (**Fig 3.6**).

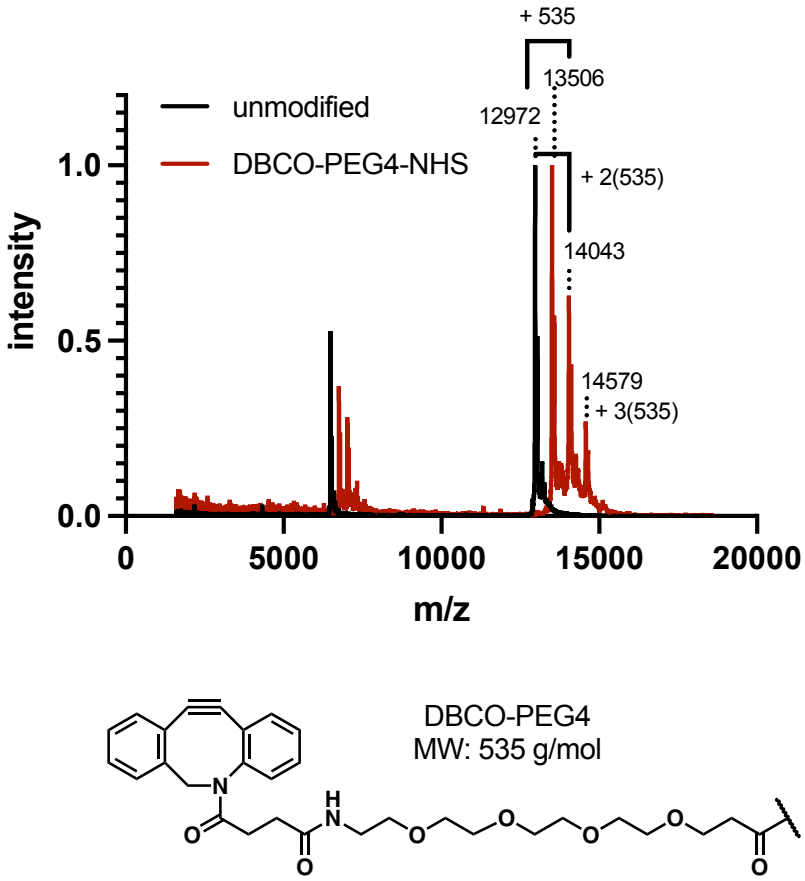


Figure 3.6 Modification of streptavidin with DBCO-PEG4-NHS. MALDI mass spectroscopy indicates that most of the streptavidin was modified with 1 or 2 molecules of DBCO based on the mass shift of the red spectra to peaks where the m/z increased by the weight of 1 or 2 of the DBCO-PEG4 fragment shown below the spectra.

The DBCO-modified streptavidin was then carried on to the next step where both the glycosylated and non-glycosylated probe were attached via click chemistry. The MALDI spectra for the streptavidin-probe conjugates can be seen in **Fig 3.7**. In MALDI mass spectroscopy, it is very common for sialic acids to be cleaved,[111] which is why the m/z peak for both the glycosylated and nonglycosylated probe have the same m/z which is increased by 420 units from the DBCO-modified streptavidin. Another peak with a mass increased by 956 was also seen in the streptavidin modified with the glycosylated probe

which corresponds to two probes (with the DBCO-PEG4 linker) conjugated to the streptavidin.

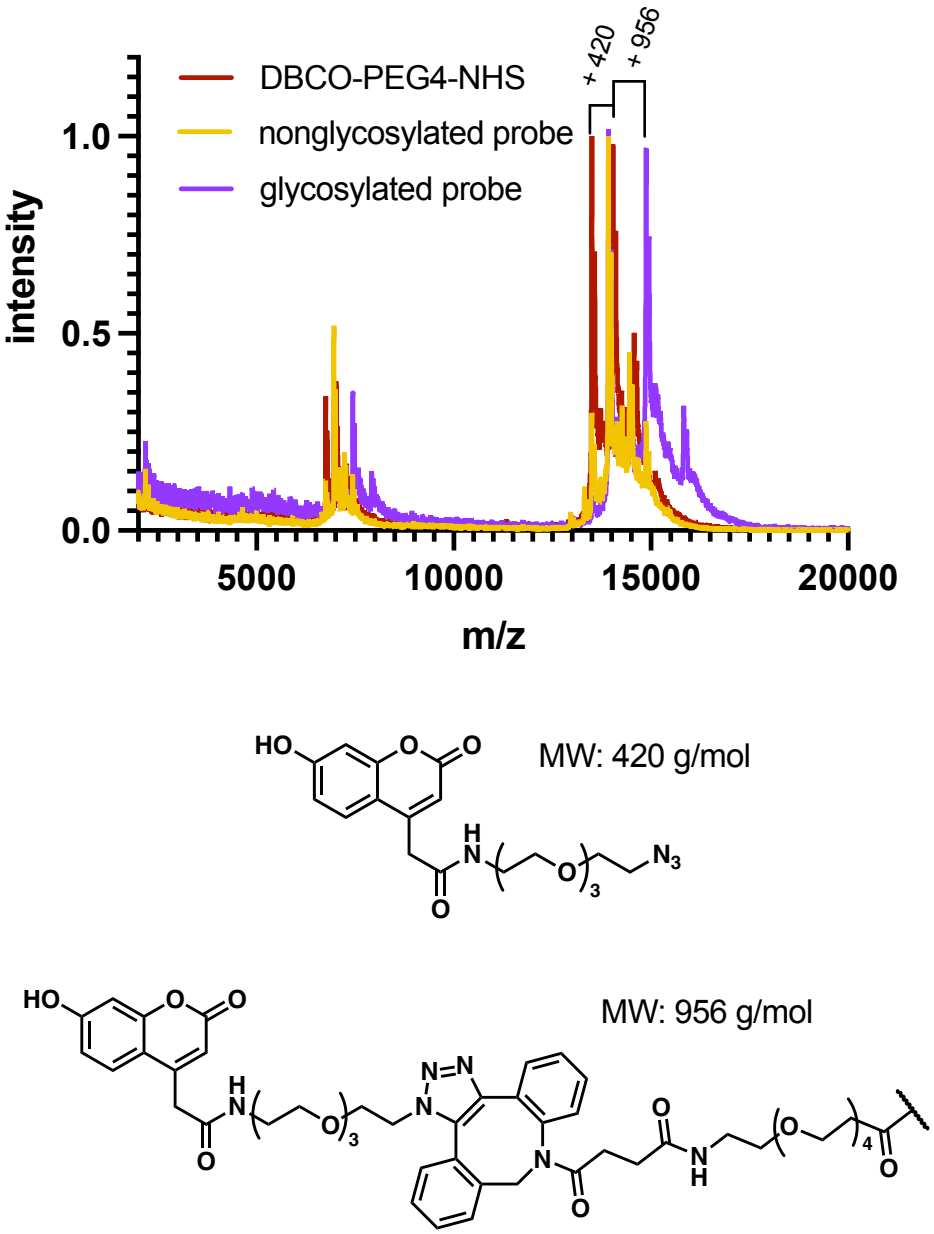


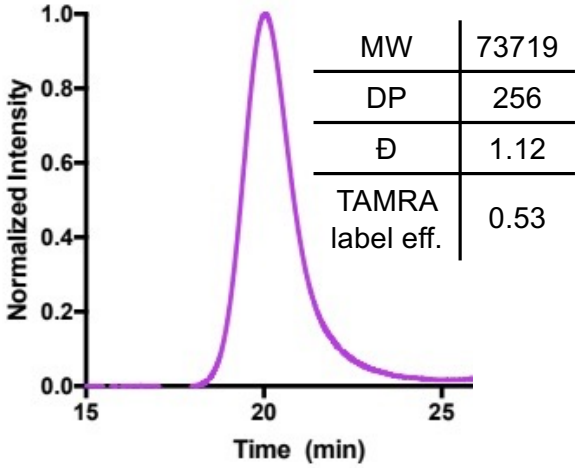
Figure 3.7 Modification of streptavidin with probes. MALDI mass spectroscopy indicates that 1 nonglycosylated probe was incorporated into the DBCO-modified streptavidin (yellow curve) and 1-2 glycosylated probes were incorporated (purple curve). The amount of incorporation is based on the molecular weights of the nonglycosylated probe and the additional m/z peak corresponding to the nonglycosylated probe conjugated to DBCO-PEG4 for the glycosylated probe. Unprotected salic acids are typically cleaved during MALDI ionization.

3.6 Conclusions and future outlooks

The results of the previous chapter illustrate that the mucin-mimetic glycopolymers can capture IAVs based on their binding specificity (H1N1 vs H3N2). Because the probes were conjugated through NHS chemistry to the streptavidin, glycopolymers containing a biotin handle can be incorporated into the open streptavidin binding sites to create a streptavidin array.

The polymers were made following the same synthetic procedures described in chapter two except that a biotin CTA was used instead of an azide CTA during the initial RAFT polymerization. Once again, sialyllactose glycans were employed as viral receptors and lactose was used as a control (**Fig 3.8**).

A



MW	73719
DP	256
Đ	1.12
TAMRA label eff.	0.53

B

Glycan	Ligation efficiency	valency
α2,3-SiaLac	0.36	91
α2,6-SiaLac	0.39	99
Lac	0.54	137

Figure 3.8 Biotinylated polymer characteristics. A) SEC trace of polymer backbone which has a degree of polymerization (DP) of 256 was labeled with TAMRA (53% labeling efficiency). B) The backbone was modified with both linkages of SiaLac and Lac.

The next step in the procedure is to test the conjugates for probe cleavage in solution. It is possible that the polymers will increase the effective concentration of IAV in

the vicinity of the probes based off its binding preference. So, by using small amounts of virus, NA enzymatic cleavage will only lead to fluorescent signal if IAV binds the glycans on the polymers. A second approach would include immobilization of the probe conjugates in a 96 well plate or on glass slides. Incubating the array with virus at 0 ^{o}C where NA is inactive, and HA still binds to its preferred receptor will provide a mechanism to wash away unbound virus. Then, plate can be heated to 37 ^{o}C to activate NA cleavage of the probe and measurement of its fluorescent.

Synthesis of a probe containing an alkyne handle will allow direct incorporation into a newer version of polymers[112] containing sialic acid glycans. These polymers can also be functionalized with biotin to afford their immobilization on streptavidin slides. The same protocol of binding at 0 ^{o}C, washing away unbound virus, and heating to 37 ^{o}C will result in fluorescent signal only in the polymers that initially captured IAV at 0 ^{o}C.

3.7 Materials and methods

3.7.1 Instrumentation

Proton nuclear magnetic resonance (^{1}H NMR) spectrum of polymer backbones and glycopolymers were obtained on either a 300 MHz (Bruker) or 500 MHz (Joel) NMR spectrometer, using deuterated solvents. The spectra were analyzed using MestReNova software and are reported in parts per million (ppm) on the δ scale relative to the residual solvent as an internal standard (for ^{1}H NMR: $CDCl_3$ = 7.26 ppm, D_2O = 4.79 ppm, MeOD = 4.87 ppm). Data are reported as follows: chemical shift, multiplicity (s = singlet, d = doublet, dd = doublet of doublets, t = triplet, q = quartet, br = broad, m = multiplet), and integration. NMR spectra are included in the Appendix to Chapter 3.

Mass spectra for select compounds were obtained on a Micromass Quattro Ultima Triple Quadrupole MS in either negative or positive ion mode using 50% MeOH in water as the mobile phase. Samples were dissolved to approximately 1 mg/mL in 50% MeOH/H_2O and were diluted 1:100 in the mobile phase before injecting into the instrument. Background from a MeOH injection was subtracted was the sample spectra for analysis.

A Bruker Autoflex Mx MALDI-TOFMS was used for MALDI analysis. Briefly, samples were dissolved to 1 µg/µL concentration in 0.1% TFA in MilliQ water. Then, 2 µL of sample was mixed with 4 µL of sinapinic acid matrix and 1.5 µL was plated in duplicate on the MALDI plate. Protein Standard I (used for proteins with a mass less than 20 kDa) was made and plated in a similar manner and used to calibrate the instrument prior to running the samples. The Flex Analysis program was used to analyze spectra which were then plotted in PRISM.

Fluorescent-based enzymatic assays were run in 96 well plates on a SpectraMax i3x plate reader from Molecular Devices. For the kinetic assays, the sample chamber was heated to 37 ^oC and each well was read every two minutes for an hour at an excitation wavelength of 365 nm and an emission wavelength of 450 nm. Endpoint fluorescent assays were also conducted after a one-hour incubation at 37 ^oC and read with the same excitation and emission wavelengths. All data was plotted in PRISM. The plate reader was also used to determine the presence of streptavidin protein through absorbance measurements at 280 nm.

3.7.2 Neuraminidase probe synthesis and characterization

Peracetylated methyl ester of sialic acid

Neuraminic acid (1.5 g) was suspended in MeOH, Amberlite-H^+ resin (4.6 g) was added, and the mixture was stirred overnight (18 hr). The resin was filtered off and the filtrate was concentrated and dried in vacuo. The crude beige solid was used directly in the next reaction where it was dissolved in 3.6 mL of pyridine and 4.4 mL of acetic anhydride. A catalytic amount of DMAP was added and the reaction was stirred overnight (18 hr). The following morning, the reaction was dried in vacuo and the residue was resuspended in 0.1 M HCl and DCM. To completely removed the remaining pyridine, the DCM layer was washed with saturated $CuSO_4$ and twice with brine. Toluene was then added to the DCM layer and evaporated. At the addition of hexanes, a white solid formed that was crystalized further in the freezer, filtered, and collected as the final product.

Protected chloride of sialic acid

The peracetylated methyl ester of sialic acid (0.7 g) was dissolved in DCM (28 mL) and cooled to -20 ^oC in a 1 L round bottom flask before acetyl chloride (6.5 mL) was added. A septum was secured with copper wire and MeOH (1.77 mL) was added dropwise. The reaction was allowed to stir at -20 ^oC for 1 hr before the salted ice water bath was removed so the reaction could stir at room temperature overnight (18 hr). The pressure was released from the reaction flask with a needle and the HCl gas was removed on the house vacuum to give the white sticky product.

Coumarin methyl ester

Methanesulfonic acid (4.5 mL) was added dropwise to a cooled (0 ^{o}C) mixture of resorcinol (500 mg) and 1,3-acetondicarboxylate (830 µL), causing the mixture to turn yellow. The mixture was the heated to 50 ^{o}C for 4 hours in the dark which dissolved all the resorcinol and generated a yellow viscous solution. The reaction was removed from heat and allowed to cool to room temperature before it was transferred with MeOH to a crushed ice/water mixture. An off-white precipitate formed immediately, and the mixture was allowed to stir before it was filtered and washed with cold water and hexanes to give a slightly yellow product.

Hydrolysis of coumarin methyl ester

The coumarin methyl ester was dissolved in THF and cooled in an ice bath. Then 2 equivalents of 2 M LiOH was added dropwise, causing the reaction to turn bright yellow. The flask was removed from ice and stirred at room temperature in the dark for 4 hours. The reaction was stopped by the addition of EtOAc and 1 M HCl. The organic layer was collected, and the aqueous layer was washed with additional aliquots of EtOAc. All organic layers were pooled together and dried over magnesium sulfate before filtration and drying gave an off-white solid.

Mesylating of tetraethylene glycol

Tetraethylene glycol was added to a flame dried shlenk flask, along with DCM and 2.5 equivalents of Et_3N. Then the reaction flask was cooled to 0 ^{o}C and 2.5 equivalents

of mesylchloride were added dropwise under N_2. The flask was the removed from the ice bath and allowed to stir at room temperature overnight (18 hr). Then more DCM was added, and the salts were filtered away before extracting the organic phase in water and $CHCl_3$. The organic layer was washed twice with brine and dried over magnesium sulfate before filtration and drying to isolate the product.

Making the di-azido-PEG

The mesylate was dissolved in DMF and 3.5 equivalents of sodium azide were added. The flask was stirred at 80 ^oC overnight (18 hr). The salts were filtered, and the reaction was concentrated to a yellow oil. The oil was dissolved in EtOAc and rinsed with water and brine before it was dried over magnesium sulfate, filtered, and the solvent was evaporated.

Selective reduction of one azide to an amine

1 M HCl was added to the di-azido-PEG and stirred on ice. 1.2 equivalents of triphenylphosphine was weighed into a separate vial and dissolved in Et_2O. This solution was added to the flask containing the di-azido-PEG and it was allowed to stir overnight from 0 ^oC to room temperature. The reaction was moved to a separatory funnel and the aqueous layer was washed with Et_2O. Then the aqueous layer was cooled to 0 ^oC and solid KOH was added. Once the KOH was dissolved, the product was extracted from the aqueous layer into DCM which was dried over magnesium sulfate, filtered, and the solvent was evaporated to give the azido-PEG-amine product as an oil.

Conjugating coumarin to the azido-PEG-amine (Nonglycosylated probe)

1.1 equivalents of PyBOP were added to a flask containing the coumarin carboxylic acid. 1.5 equivalents of the azido-PEG-amine was dissolved in a mixture of 10% DMF in ACN and transferred to the reaction flask. The flask was cooled on ice for the dropwise addition of 3 equivalents of DIPEA. Then it was removed from the ice bath and allowed to react in the dark at room temperature for 3 hr. The reaction was stopped by the addition of excess EtOAc and saturated NH_4Cl. The organic layer was washed with $NaHCO_3$ and brine. It was then dried using magnesium sulfate and the solvent was evaporated to give an orange oil. A column using a 1-10% mobile phase gradient of MeOH in EtOAc was used to purify the crude product. The fractions that had an Rf value of 0.32 in a 5% MeOH/EtOAc solvent system were collected to give the pure product as a white solid.

Attaching the sialic acid (protected sialic acid probe)

The protected sialic acid chloride (2.5 eq) was added to a round bottom flask containing the coumarin-PEG-azide and dissolved in ACN. DIPEA (1.2 eq) was added under positive N_2 pressure, causing the reaction to immediately turn orange. It was allowed to stir in the dark at room temperature for 5 hr. The reaction was quenched with addition of 3M HCl until orange color disappeared. Water was added and the aqueous layer was extracted with EtOAc, the organic layers were combined, washed with brine, dried over MgSO4, and the solvent was evaporated to give the crude product. It was purified on a column using a 1-10% mobile phase gradient of MeOH in EtOAc. The fractions that had an Rf value of 0.27 in a 5% MeOH/EtOAc solvent system were collected to give the pure product.

Deprotecting the sialic acid (glycosylate probe)

The protected sialic acid probe was dissolved in MeOH and Na_2CO_3 (3 eq) was added and allowed to stir in the dark for 2 hr. Then Amberlite-IR120 resin was added and allowed to stir for 40 min. After which time the resin was filtered off and the MeOH was evaporated. The crude residue was then dissolved in water and 2 equivalents of LiOH was added and allowed to stir at room temperature for 2 hr. Acetic acid was added to quench the reaction and the solvent was removed through lyophilization. To purify the crude product, a column was run using a 25 to 100% gradient of MeOH in EtOAc. The only UV active peak was collected, and the solvent was evaporated to give an off-white solid.

Creating DBCO-modified streptavidin

A spin filter was used to buffer exchange the storage buffer the streptavidin (New England Biolabs) came in to the reaction buffer (Na_3PO_4, at pH 8) used for NHS coupling. The resulting concentrated solution was diluted to 2.5 mg/mL and 4 equivalents per streptavidin tetramer of DBCO-PEG4-NHS (Avanti) dissolved in DMSO was added and allowed to react overnight at room temperature. The following morning, a PD10 desalting column was run using MilliQ water as the eluent. The fractions containing protein were determined through absorbance readings at 280 nm. They were pooled and lyophilized to give a white solid. The extent of modification was determined through MALDI analysis.

Clicking on the probes

The azide on both the glycosylated and nonglycosylated probe was clicked to the DBCO-modified streptavidin. The streptavidin conjugated was suspended to approximately 200 μM in PBS and split into two aliquots. The glycosylated probe in 1.2 equivalent excess was added to one aliquot and the other aliquot was reacted with 1.2 equivalents of the nonglycosylated probe. Both reactions were allowed to proceed overnight and the following day a PD10 desalting column was run using MilliQ water as the eluent. The fractions containing protein were determined through absorbance readings at 280 nm. Each reaction was pooled and lyophilized to give a white solid. The extent of modification was determined through MALDI analysis. The initial analysis of the glycosylated probe showed little conjugation, so the protein was redissolved in PBS and the reaction and purification was repeated using two equivalents of the glycosylated probe. The new analysis indicated 1-2 probes per streptavidin monomer.

3.7.3 Biotin polymer synthesis

The polymers were synthesized as described in Chapter 2 except a CTA with biotin was used instead of one with an azide. Briefly, Boc-protected *N*-methylaminooxypropyl acrylamide monomer was polymerized via RAFT polymerization using AIBN as the radical initiator and a chain transfer agent containing biotin to produce the polymeric precursor. Next, the trithiocarbonate end group was removed by aminolysis and the newly exposed thiol group was capped with TAMRA-maleimide. The fluorophore labeling efficiency was determined to be 53 % by UV-VIS spectrometry. The Boc protecting groups were deprotected by a mixture of phenol and trimethylsilyl chloride (TMSCl),[73] allowing the

conjugation of reducing glycans under acidic conditions to the newly exposed *N*-methylaminooxy group to complete the synthesis of the biotinylated glycopolymers.

3.7.4 Creating the streptavidin array

Biotin polymers containing α2,6-SiaLac, α2,3-SiaLac, and Lac (10 eq, 2.5 eq per biotin binding site on streptavidin) were mixed overnight at 4 ^{o}C with 3-5 μM of streptavidin-probe conjugates in PBS supplemented with additional calcium and magnesium. The samples were then run on a 1% agarose gel to determine the extent of polymer incorporation.

3.8 Acknowledgements

I would like to thank Austen Michalak for his help in attaching biotin polymers I made to create the final streptavidin array. He also ran the gel to characterize how the polymers incorporated into the array. I would also like to thank Connor Brandenburg. He recently joined the lab and will be continuing the NA probe project once I leave. He is currently scaling up synthesis of the probe and devising ways to incorporate an alkyne handle for incorporation into a newer version of mucin mimetic polymers.

Chapter three, in part is currently being prepared for submission for publication. Lucas, Taryn M.; Brandenburg, Connor A.; Michalak, Austen L.; Godula, Kamil. The dissertation author was the primary researcher and author of this material.

4 Uncovering the identity of IAV receptors

4.1 Introduction

While the systems described in the previous two chapters are useful for studying IAV binding preferences to receptor types and identifying if virus is present in a sample, they rely on the use of mimetic glycoproteins that approximate the receptor presentation within the glycocalyx. By employing proximity-based labelling strategies, one could investigate viral binding on the cell surface to gain insight into receptors IAV uses during the infection process. There are several proximity labelling approaches that have been developed.[112–116] Each utilizes a peroxidase enzyme either fused with a protein of interest or fused to a secondary antibody to the protein of interest to promote a reaction of biotin (or a biotin derivative) with nearby proteins. This results in the interactors with the protein of interest becoming labelled with a biotin that can be used for purification and a unique mass signature for identification by mass spectrometry.

The Lewis Lab at UCSD has used the BioID method to investigate transient protein-protein interactions that occur in the secretory pathway to gain insights into the therapeutically relevant issue of limited secretion of recombinantly expressed protein drug candidates.[117] This method requires the expression of a fusion protein containing the protein of interest (the bait) and BirA biotin ligase from *E.*coli with a R118G mutation in its active site to allow for biotinylation of any protein, even those without the normally required biotin acceptor peptide (BAP).[118] Biotinylation by antibody recognition (BAR) utilizes an HRP-conjugated secondary antibody catalyze the formation of a free radical biotin derivative that quickly reacts with nearby proteins through tyrosine residues.[115] Because it circumvents the need to generate bait-BirA fusion constructs, the Lewis Lab

has also started experimenting with this approach which is very well suited to tag the glycoproteins IAV employs on its path toward infection of a host.

4.2 Biotinylation by antibody recognition (BAR) to find IAV binding partners

The overall goal of this project is to create a model of IAV binding interactions that occur during infection. It is known that IAV engages sialic acid, but many proteins contain this terminal glycan, so how exactly does IAV discriminate between sialylated proteins when it is moving through the glycocalyx? IAV can bind through its HA protein and then use its enzymatic NA activity to move across fetuin-coated glass,[119] and it is entirely plausible that it uses this same machinery to reach its receptor for internalization. Our current hypothesis is that IAV utilizes mucin glycoproteins as a first point of contact since they are heavily sialylated and extend microns from the epithelial cell surface. We believe that these interactions are non-productive for internalization, but that IAV uses the dense sites of sialylation as stepping stones to the underlying glycoproteins and glycolipids that are responsible for internalization.[120] To test this hypothesis we initiated a collaboration with the Lewis Lab at UCSD to carry out BAR experiments (**Fig 4.1**) in human alveolar cells. We chose to use A549 cells because they are biologically relevant for IAV infection, and a recent paper has shown that IAV colocalizes with some mucins expressed by these cells, but not others.[71] By conducting the IAV binding to these cells at 0 ^{o}C, where NA is inactive, and then following the BAR protocol with a primary antibody against HA, we should get a snapshot of all the proteins IAV interacts with on the cell surface. Examining the structures and known glycosylation profiles of the protein hits will be useful in generating a model of IAV interactions.

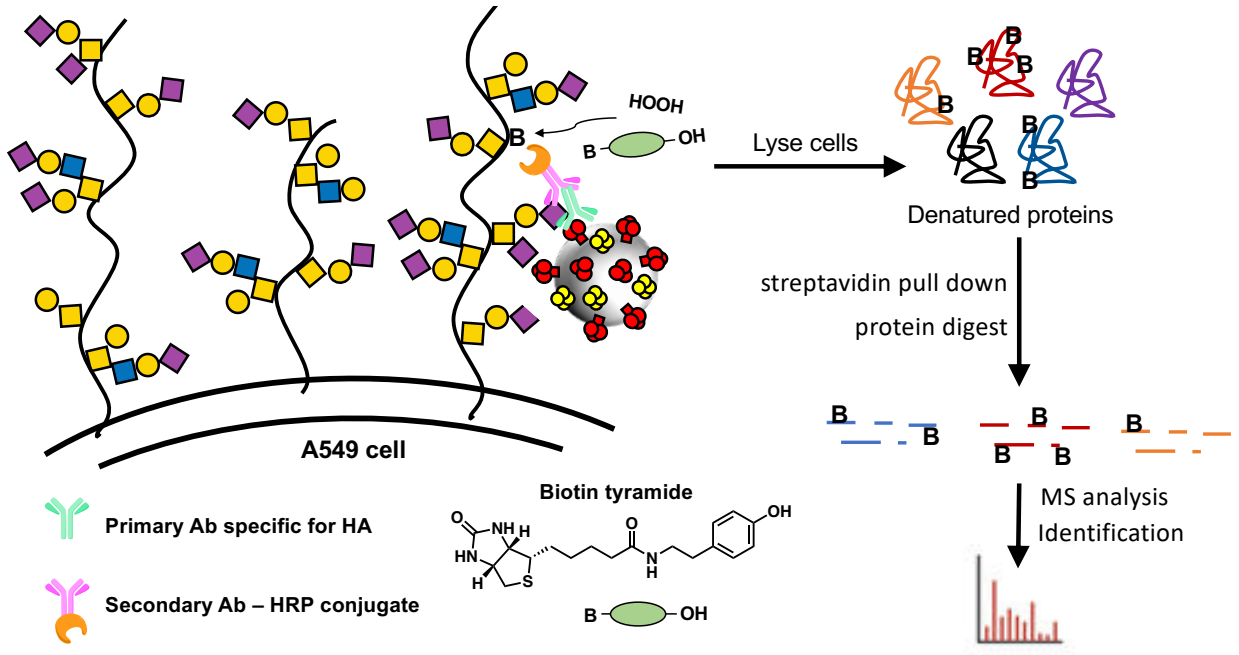


Figure 4.1 BAR approach for identifying IAV receptors. A549 cells are incubated with IAV and fixed with PFA. Then a primary antibody for HA is added, followed by a secondary antibody-HRP conjugate. The HRP catalyzes the generation of free radicals on biotin tyramide after the addition of hydrogen peroxide which leads to biotinylation of proteins near the site of the radical formation on HA. Once the cells are lysed, the biotin handle can be used to pull down labelled proteins that are then digested for MS analysis and identification.

4.2.1 Validating primary antibodies

A major consideration when conducting BAR experiments is the specificity of the primary antibody because that is ultimately what localizes the formation of free radicals and subsequent biotin labelling. If the primary antibody lacks specificity for the protein of interest, then background will be high. Three different antibodies were tested (data not shown) before one was found that was did not display cross reactivity to components of the cell lysate (**Fig 4.2**).

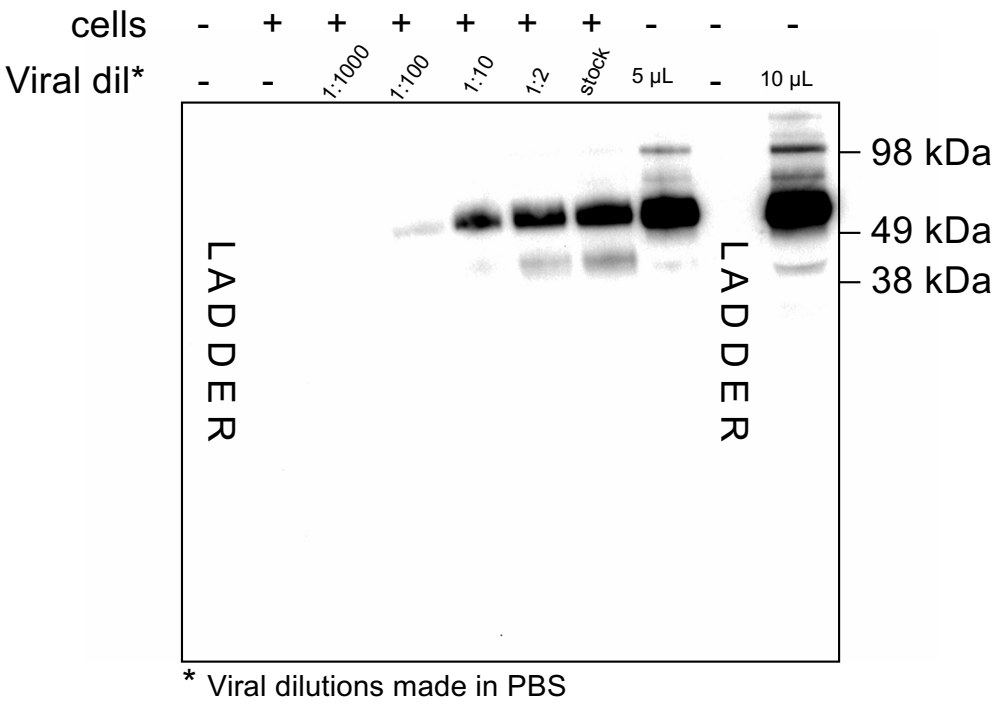


Figure 4.2 Testing HA antibody specificity by western blot. A549 cells were incubated with IAV stock or dilutions of IAV in PBS for 1 hr. Two lanes contain only virus at 5 and 10 µL. A blot of the cell lysate that was probed with a 1:1000 dilution of H28 E23 antibody against HA is specific for conditions containing virus. Appreciable signal is seen up to the 1:10 viral dilution and no signal is seen in the cell only control.

Antibody specificity was also confirmed through microscopy after subjecting the cells to the BAR protocol (**Fig 4.3**). These results also indicate that the H28 E23 antibody is specific and can be used for biotin labelling experiments.

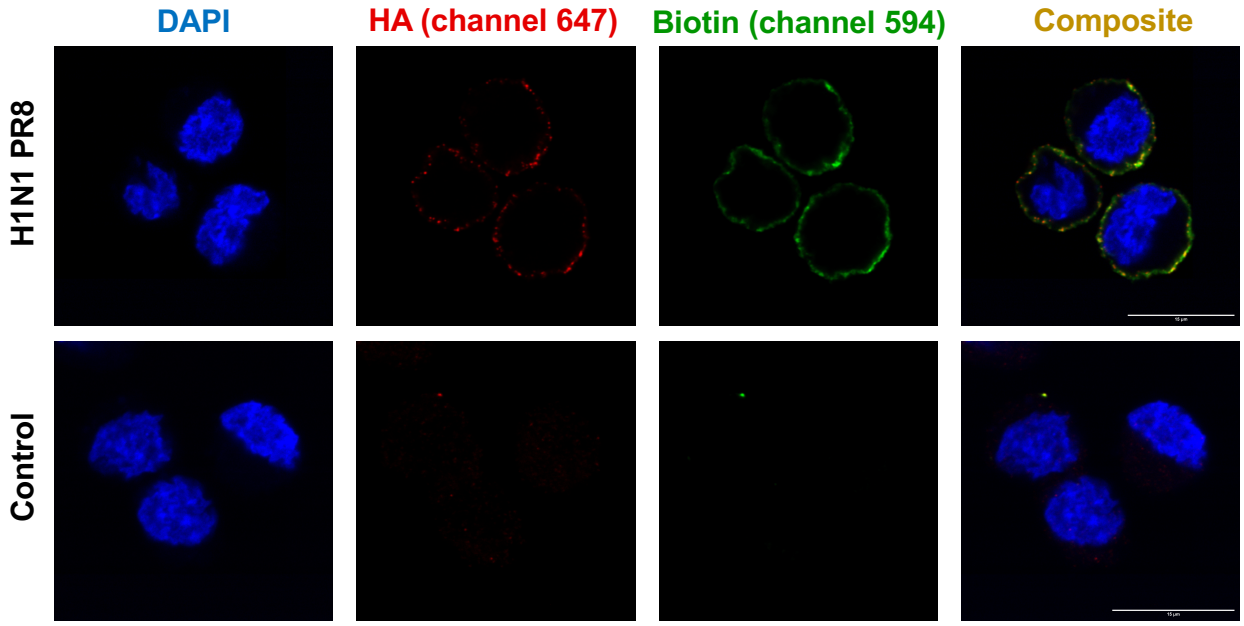


Figure 4.3 Testing HA antibody specificity by microscopy. A549 cells were incubated with IAV stock or infection media not containing virus for 1 hr. They were then fixed, and the BAR protocol was carried out. HA staining can be seen in the AF647 channel, and the biotin labelling can be seen in the AF594 channel because it was visualized using streptavidin-AF594. In both cases, signal is seen when IAV was present and there is no signal in the control population.

Additionally, western blots of the cell lysate only show streptavidin signal in the conditions incubated with virus (**Fig 4.4**). The smear of stain for the IAV condition is expected since there are many sialylated glycoproteins on the cell surface that IAV can interact with. The few minor bands that appear in the control lane are most likely endogenously biotinylated proteins.

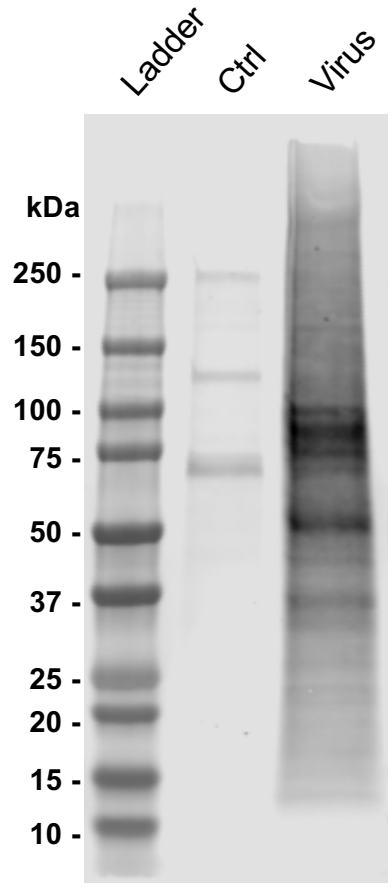


Figure 4.4 Western blot of A549 cell lysate following BAR protocol. A549 cells were incubated with IAV stock or infection media not containing virus for 1 hr. They were then fixed, and the BAR protocol was carried out. Then the cells were lysed, and the lysate was run on a gel, transferred to a blot, and stained with streptavidin. A smear of signal is seen for the IAV condition because many sialylated proteins can interact with the virus.

4.2.2 BAR experiment with MS identification

Having validated the antibody being used, the BAR experiment was scaled up to allow triplicate measurements of both the IAV and control conditions. All experiments were performed at the same time and validated through microscopy and western blots (data not shown). Each replicate was sent to the mass spectrometry facility at Sanford Consortium for protein digestion and MS analysis. The resulting volcano plot of IAV interactors can be seen in **Figure 4.5**. The cells that were incubated with virus exhibit a much greater level of biotinylation. Many of the enriched proteins in the control samples

are mitochondrial in nature (e.g., FDXR, MRPL12, HSPE1) where endogenously biotinylated proteins are found.[121] The cells incubated with virus contain an abundance of plasma membrane proteins (e.g., ITGA2, EGFR, PVR, CDH2) which was expected since the IAV binding was done at 0 °C where internalization would not occur.

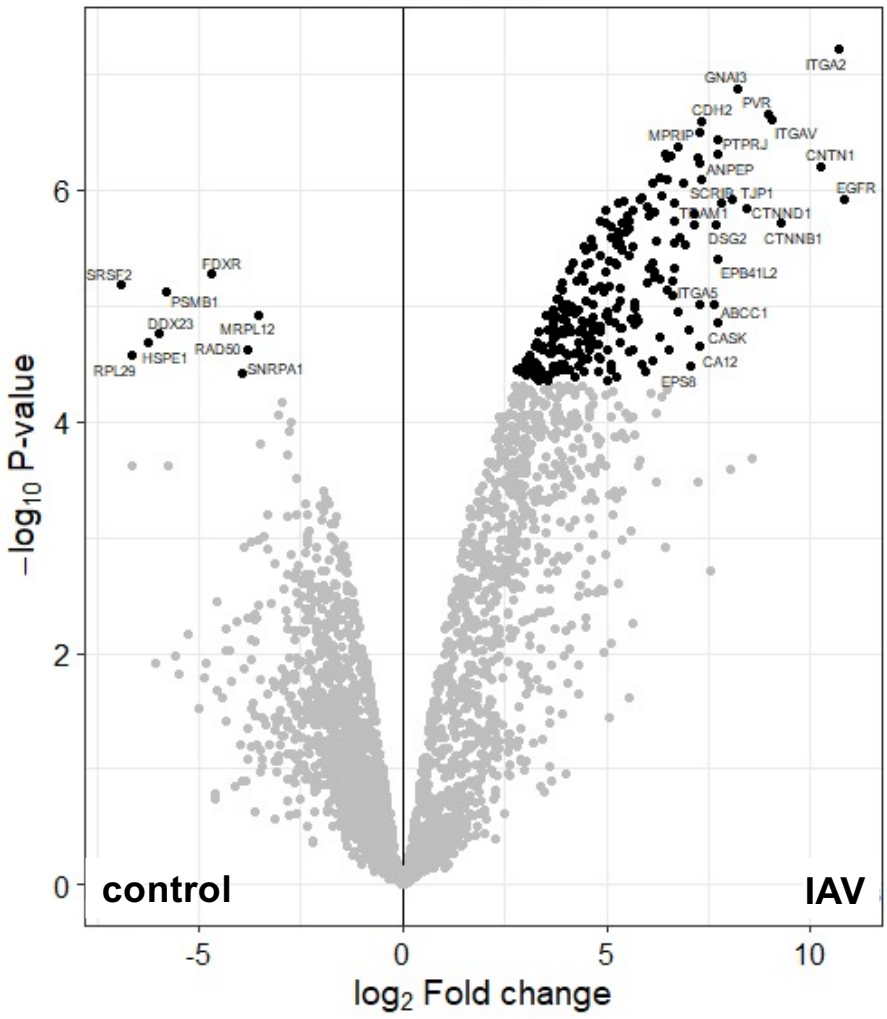


Figure 4.5 Volcano plot showing biotinylated proteins identified through MS analysis. The IAV samples contained a high enrichment for membrane bound proteins, while the control cells exhibited much less biotin labelling in general that is mostly restrained to mitochondrial proteins.

4.3 Conclusions and future outlooks

We have shown that the BAR method can be used to identify protein interactors with whole IAV particles. We are currently evaluating the enriched proteins in this large dataset for their glycosylation status and known levels of sialylation. We were interested in using this cell line because of the prior work showing that IAV selectively binds some mucins expressed by these cells, but not others.[71] Interestingly, the mucin shown to colocalize with IAV (MUC1) did not appear in our MS analysis. This could be due to various reasons, including a limited number of accessible mucin tyrosine residues that are preferentially labelled by the biotin tyramide reagent, incomplete digestion of mucins which are notoriously resistant to proteolysis which would restrict their ability to fly on the mass spec, or A549 cells have a low abundance of MUC1 so they were obscured in the pool of all other enriched proteins. Searching Uniprot protein sequences, identified five tyrosine residues in the extracellular domain that are all relatively close to the transmembrane domain. There is also no defined sequon for *O*-glycosylation so even though some of these tyrosine residues are close to serine, threonine, and proline, they may not actually be near a glycosite, making them too distant from an HA binding site. Mucin antibodies can be used to reaffirm that the cells express mucins to appreciable extents, and perhaps a mucin enrichment using StcE beads[122] would lead to their identification even if their expression is low.

4.4 Materials and methods

4.4.1 Instrumentation

SDS PAGE gels for determining antibody specificity were run using a Mini Gel Tank (Life Technologies, cat# A25977) using 4-12% Bis-Tris Plus gels (Invitrogen, cat#

NW04120BOX) in 1X MES buffer. Loading dye (Novex BOLT LDS sample Buffer 4X stock, cat# B0007) was diluted in 1M dithiothreitol (DTT) in H_2O to a 2.5X loading dye and 0.4M DTT. Equal volumes of sample and loading dye were incubated at 95 ^oC for 10 min before 15 µL were loaded on the gel. 8 µL of protein ladder (Invitrogen SeeBlue Plus2 Prestained standard ladder, cat# LC5925) were loaded in terminal lanes. The gel running conditions were 200 V for 22 min. Gels were transferred to PVDF membranes (Invitrogen iBlot2 PVDF Mini or Regular Stacks, Cat# IB24002 or IB24001) using Invitrogen iBlot2 transfer system (cat# IB21001) with P3 settings of 20 V for 7 min). Blots were imaged using a BioRad GelDoc XRS+ imaging device.

The BAR cell lysate samples (15 µg) were run on a Biorad system 4-12% TGX gels in Tris-glycine-SDS running buffer. Samples were mixed with Biorad loading dye and β–mercaptoethanol and loaded onto the gel, along with the precision plus dual color ladder. Blot transfer was done using Biorad TurboBlot system. These blots were read on a Licor IR instrument. A subsample of the cell pellet from BAR experiment were imaged following fluorescent secondary incubations using Leica SP8 with Lightning Deconvolution.

Mass spectrometry was run at the Sanford Burnham Prebys Proteomics Core where affinity purified samples were analyzed by LC-MS/MS using a Proxeon EASY-nanoLC system (ThermoFisher) coupled to a Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific). Peptides were separated using an analytical C18 Aurora column (75µm x 250 mm, 1.6 µm particles; IonOpticks) at a flow rate of 300 nL/min (60 ^oC) using a 75-min gradient: 1% to 5% B in 1 min, 6% to 23% B in 45 min, 23% to 34% B in 28 min, and 34% to 48% B in 1 min (A= formic acid 0.1%; B=80% ACN: 0.1% formic

acid). The mass spectrometer was operated in positive data-dependent acquisition mode. MS1 spectra were measured in the Orbitrap in a mass-to-charge (m/z) of 375 – 1500 with a resolution of 60,000 at m/z 200. Automatic gain control target was set to 4 x 10^5 with a maximum injection time of 50 ms. The instrument was set to run in top speed mode with 2-second cycles for the survey and the MS/MS scans. After a survey scan, the most abundant precursors (with charge state between +2 and +7) were isolated in the quadrupole with an isolation window of 0.7 m/z and fragmented with HCD at 30% normalized collision energy. Fragmented precursors were detected in the ion trap as rapid scan mode with automatic gain control target set to 1 x 10^4 and a maximum injection time set at 35 ms. The dynamic exclusion was set to 20 seconds with a 10 ppm mass tolerance around the precursor.

All mass spectra from were analyzed with MaxQuant software version 1.6.11.0. MS/MS spectra were searched against the Homo sapiens Uniprot protein sequence database (downloaded in January 2020) and GPM cRAP sequences (commonly known protein contaminants). Precursor mass tolerance was set to 20ppm and 4.5ppm for the first search where initial mass recalibration was completed and for the main search, respectively. Product ions were searched with a mass tolerance 0.5 Da. The maximum precursor ion charge state used for searching was 7. Carbamidomethylation of cysteine was searched as a fixed modification, while oxidation of methionine and acetylation of protein N-terminal were searched as variable modifications. Enzyme was set to trypsin in a specific mode and a maximum of two missed cleavages was allowed for searching. The target-decoy-based false discovery rate (FDR) filter for spectrum and protein identification was set to 1%.

4.4.2 *Cell and viral culture*

A549 cells were cultured in Dulbecco's Modified Eagle Medium (DMEM) with 10% Fetal Bovine Serum in an incubator set a 37 ºC with 5% CO_2.

Influenza virus strain A/PR/8/34 (H1N1, ATCC VR-1469) was purchased from ATCC and propagated in MDCK cells that were transferred to DMEM medium supplemented with 0.2% BSA fraction V, 25mM HEPES buffer, 2 µg/ml TPCK-trypsin, and 1% penicillin/streptomycin ("DMEM-TPCK" media). After 2 days of incubation, the cell supernatant was collected and characterized for viral titer. Viral titers were characterized through hemagglutination assays. Turkey red blood cells were purchased from Lampire and a 1% solution was used to determine viral titers via the hemagglutination test. MDCK cells were used to determine the 50% tissue culture infective dose ($TCID_{50}$) using the Spearman-Karber method. Viral aliquots were stored at -80 ^{o}C until use.

4.4.3 *Viral binding to A549 cells and BAR protocol*

A549 cells were incubated with IAV stock or infection media not containing virus for 1 hr at 0 ^{o}C. They were then fixed with 4 % PFA for 10 minutes at room temperature. The cells were incubated with 0.25% H_2O_2, blocked with 5% goat serum, and incubated overnight at 4 ^{o}C with a 1:800 dilution of E28 E23 (HA antibody from mouse). The samples were the incubated with a 1:1000 dilution of secondary goat-anti-mouse HRP conjugate and the BAR reaction was performed with TSA-biotin kit and a 3 min reaction time.

For preliminary analysis before MS analysis, an aliquot of the cell pellet was taken for immunofluorescent microscopy and the remainder was lysed with 2% SDS in PBST for western blot studies. For microscopy, a 1:500 dilution of goat-anti-mouse-AF647 was used to visualize the HA protein bound to the cells and a 1:1000 dilution of streptavidin-AF594 was used to detect the biotin labelling. As mentioned above, samples were imaged using Leica SP8 with Lightning Deconvolution. Biotin labeling in the western blot analyses was detected using a 1:3000 dilution of streptavidin-AF630 for imaging on a Licor IR.

4.5 Acknowledgements

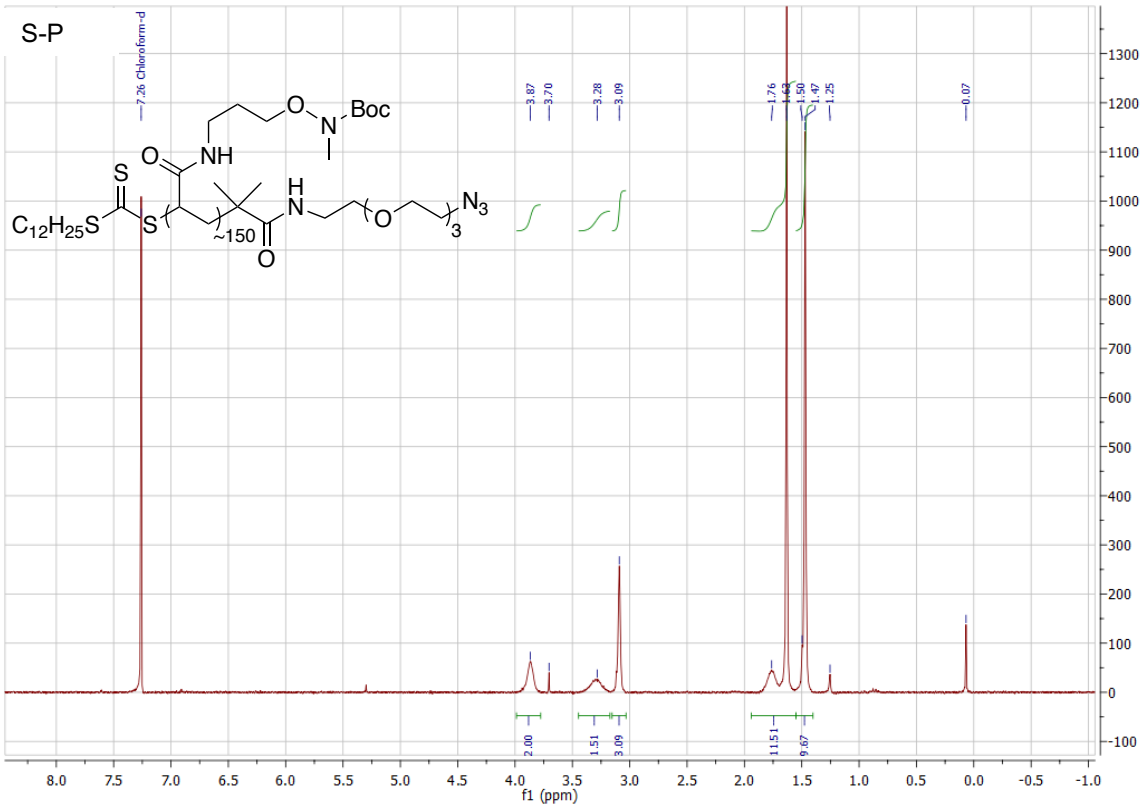
I would like to thank Caressa Robinson who carried out the BAR protocol on the A549 cells and submitted them for MS analysis. I would also like to thank Meghan O. Altman who gifted me with the highly specific HA antibody that was used for the BAR experiments. Additionally, I would like to thank the mass spec core at Sanford Consortium for isolating and running all the mass spec samples.

Chapter four, in part contains coauthored/unpublished data. Lucas, Taryn M.; Robinson, Caressa M; Lewis, Nathan E.; Godula, Kamil. The dissertation author was a primary researcher and the primary author of this material.

Appendix for Chapter 2: NMR

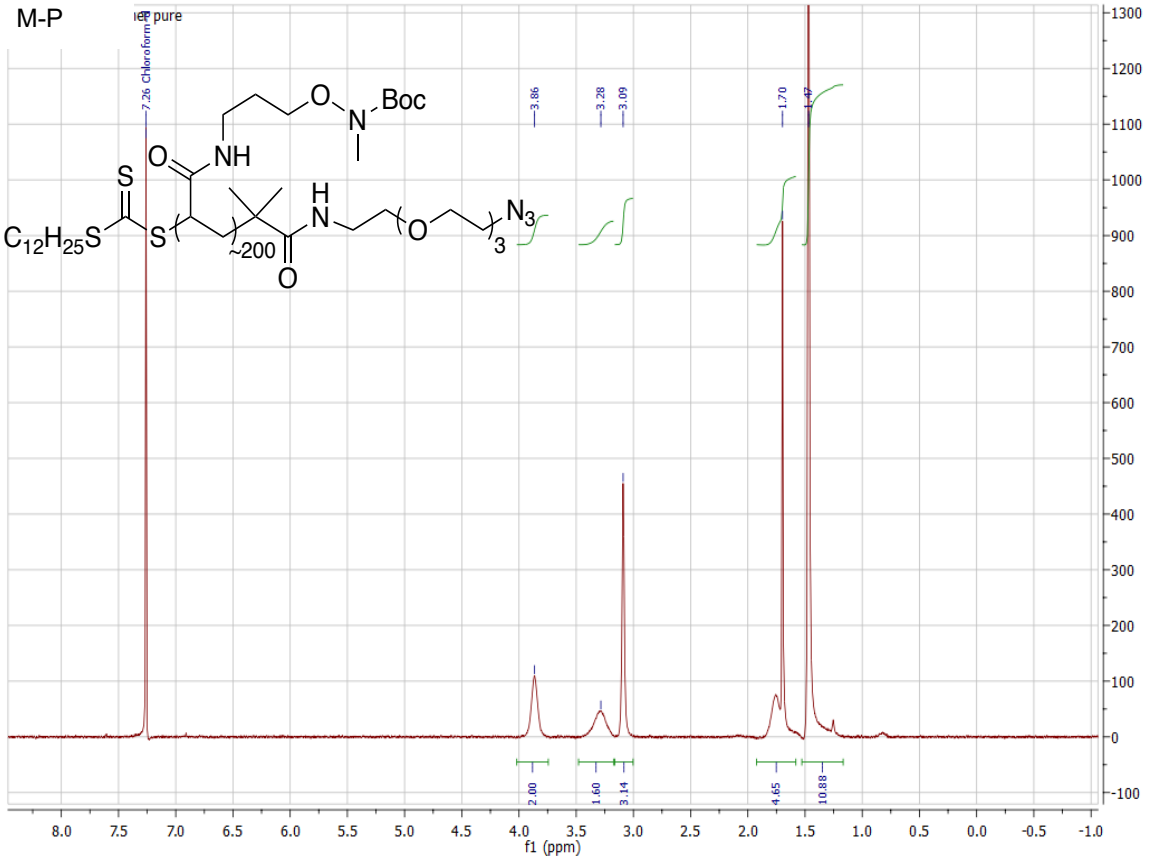
^{1}H NMR of S-P backbone

$CDCl_3$, 300 MHz, δ (ppm): 3.90-3.65 (bs. 2H), 3.45-3.00 (bm, 5H), 1.80-1.40 (bm, 21H)



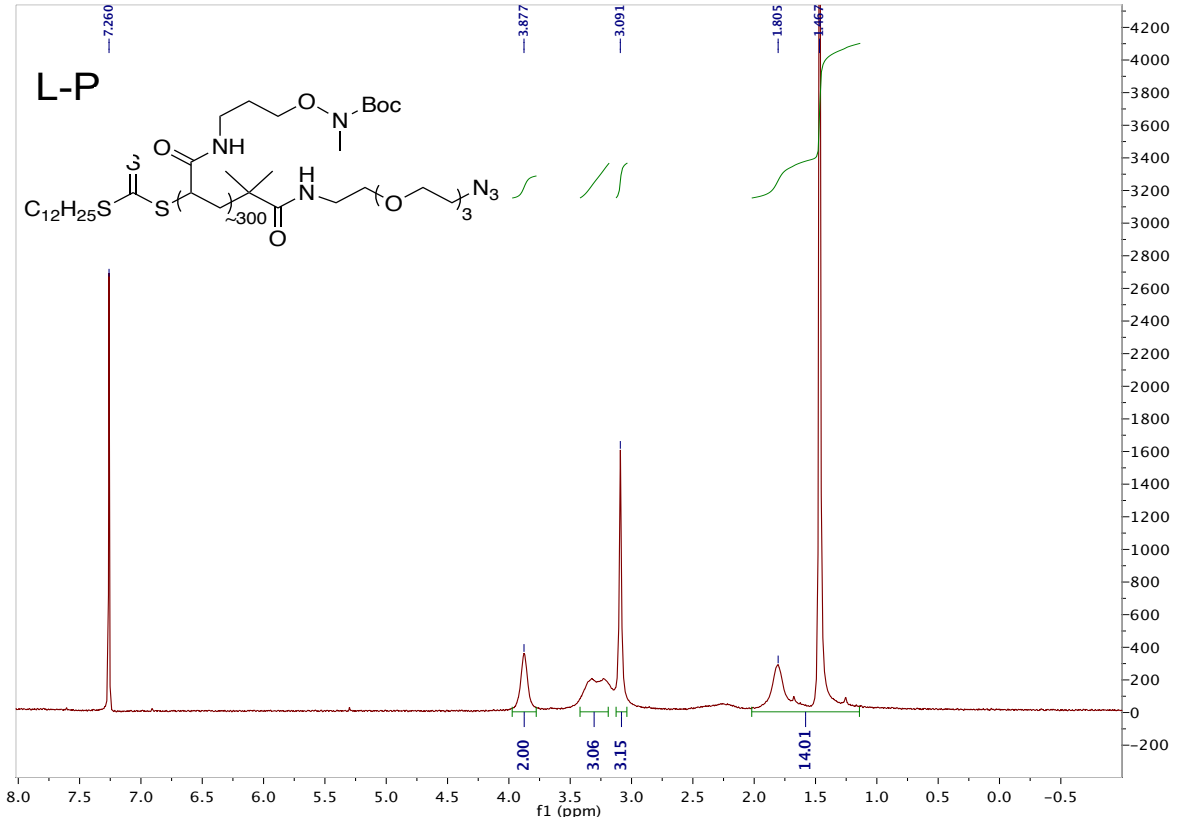
^{1}H NMR of M-P backbone

$CDCl_3$, 300 MHz δ (ppm): 3.90-3.75 (bs. 2H), 3.50-3.00 (bm, 6H), 1.80-1.25 (bm, 15H).



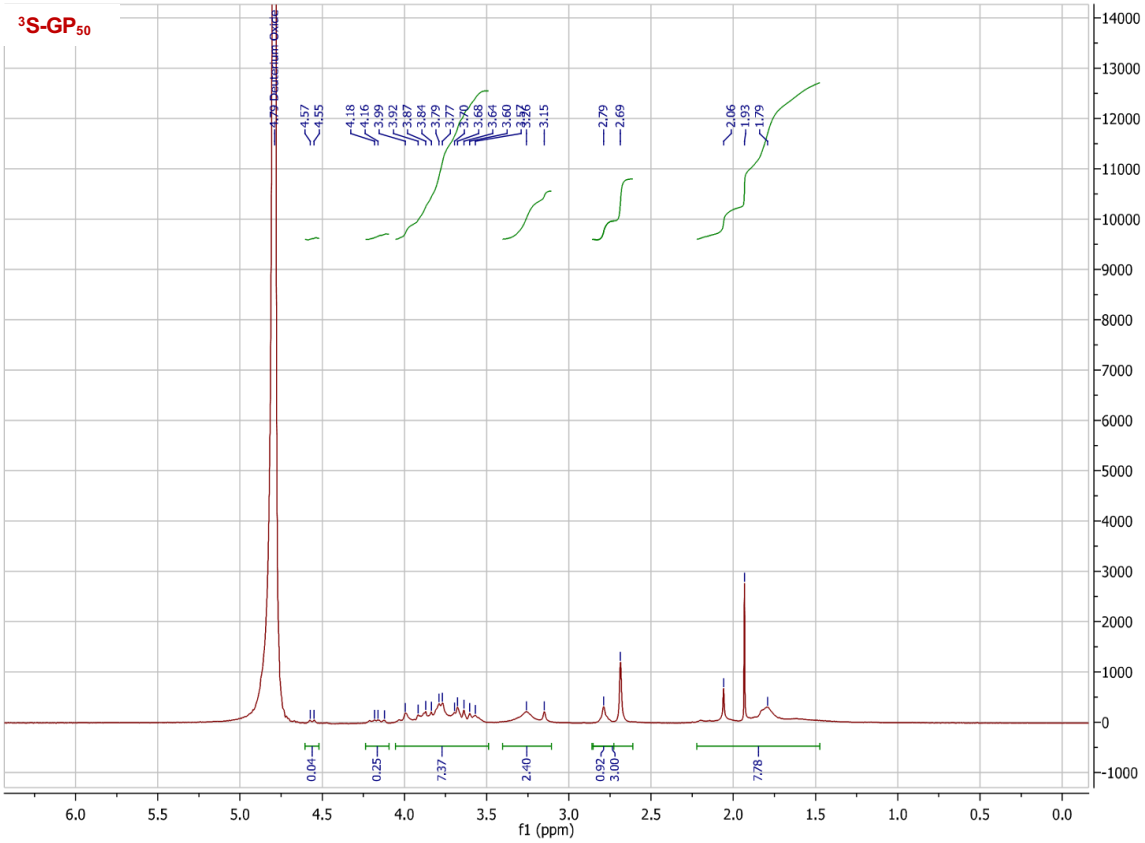
^{1}H NMR of L-P backbone

$CDCl_3$, 300 MHz δ (ppm): 3.90-3.75 (bs. 2H), 3.50-3.00 (bm, 6H), 1.90-1.05 (bm, 14H).



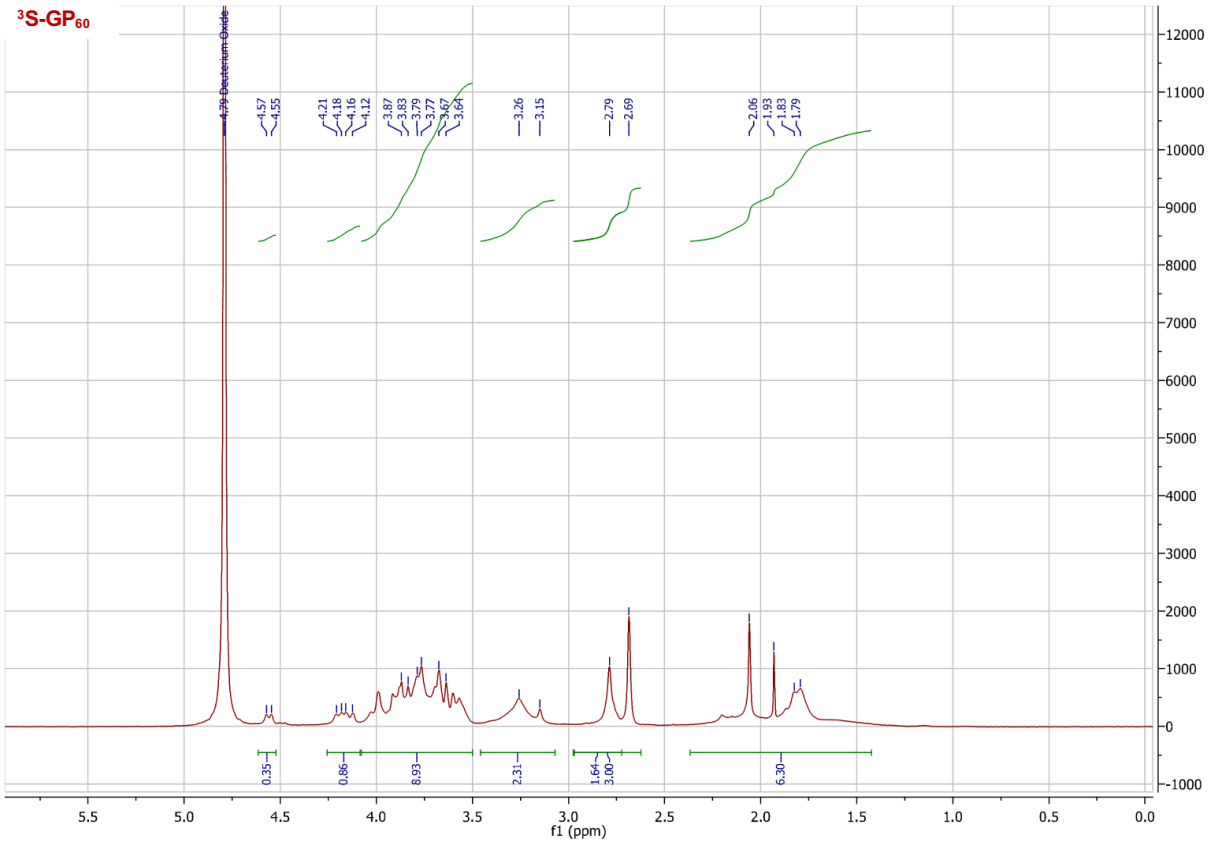
1H NMR of ^{3}S-GP$_{50}$

D_2O, 300 MHz δ (ppm): 4.60-4.50 (bd, 0.04H), 4.25-4.10 (bm, 0.25H), 4.05-3.45 (bm, 7.37H), 3.40-3.10 (bm, 2.40H), 2.85-2.60 (bm, 3H), 2.20-1.45 (bm, 7.78H).



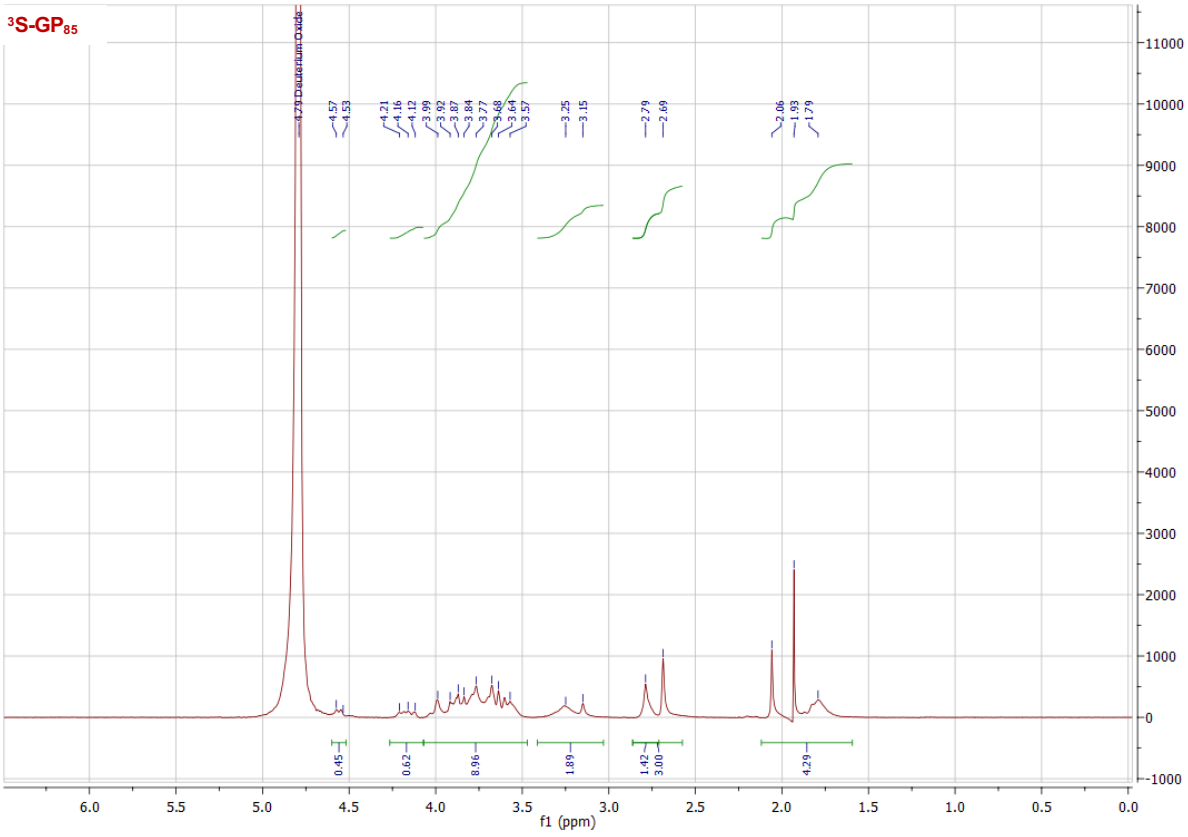
^{1}H NMR of ^{3}S-GP$_{60}$

D_2O, 300 MHz δ (ppm): 4.60-4.50 (bd, 0.35H), 4.25-4.10 (bm, 0.86H), 4.10-3.50 (bm, 8.93H), 3.45-3.10 (bm, 2.31H), 2.95-2.60 (bm, 3H), 2.35-1.40 (bm, 6.30H).



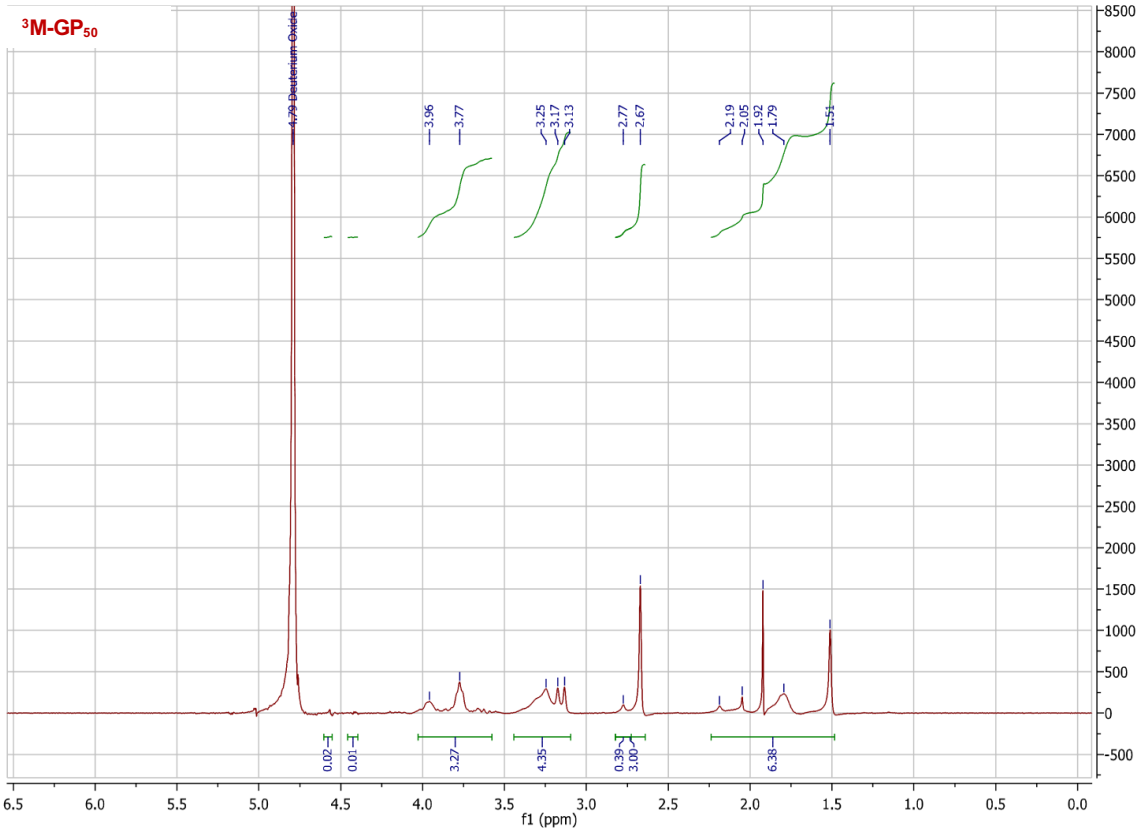
^{1}H NMR of ^{3}S-GP$_{85}$

D_2O, 300 MHz δ (ppm): 4.60-4.50 (bd, 0.45H), 4.25-4.10 (bm, 0.62H), 4.10-3.45 (bm, 8.96H), 3.40-3.10 (bm, 1.89H), 2.80-2.60 (bm, 3H), 2.15-1.55 (bm, 4.29H).



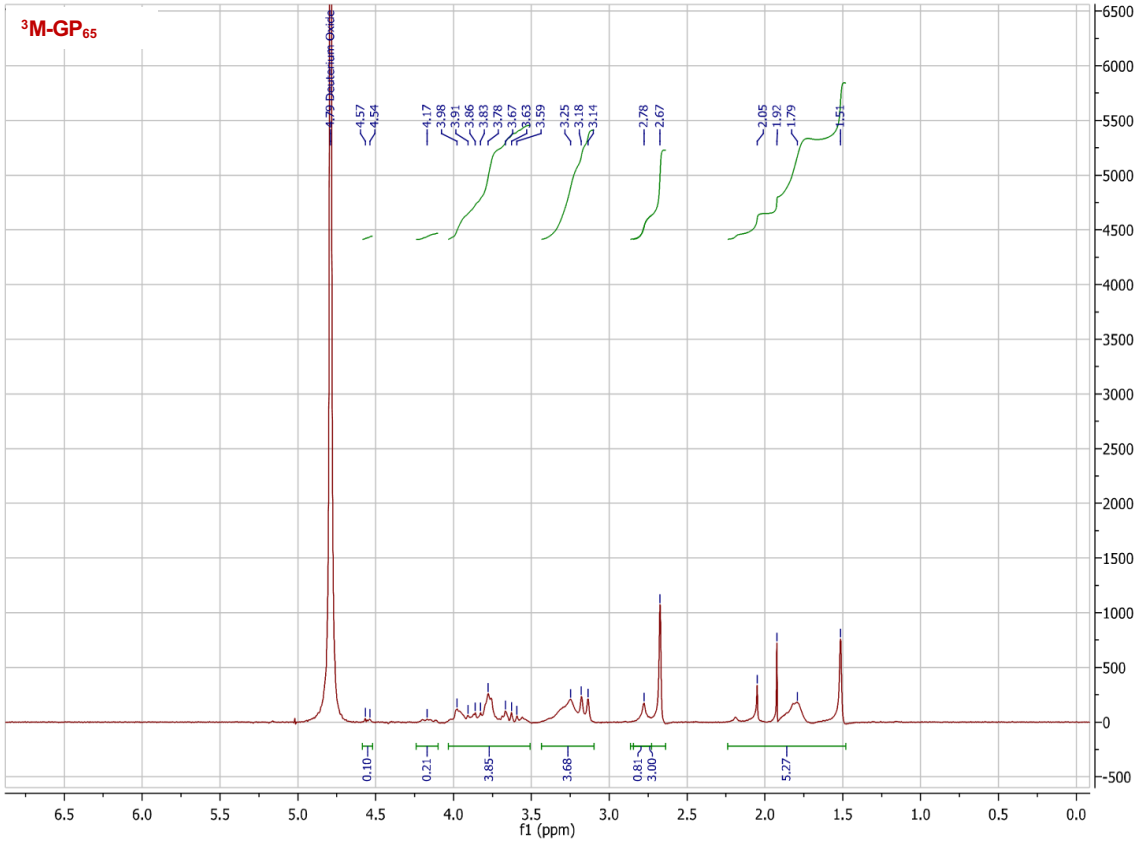
^{1}H NMR of ^{3}M-GP$_{50}$

D_2O, 300 MHz δ (ppm): 4.60-4.55 (bd, 0.02H), 4.45-4.40 (bm 0.01), 4.05-3.55 (bm, 3.27H), 3.40-3.10 (bm, 4.35H), 2.80-2.60 (bm, 3H), 2.25-1.50 (bm, 6.38H).



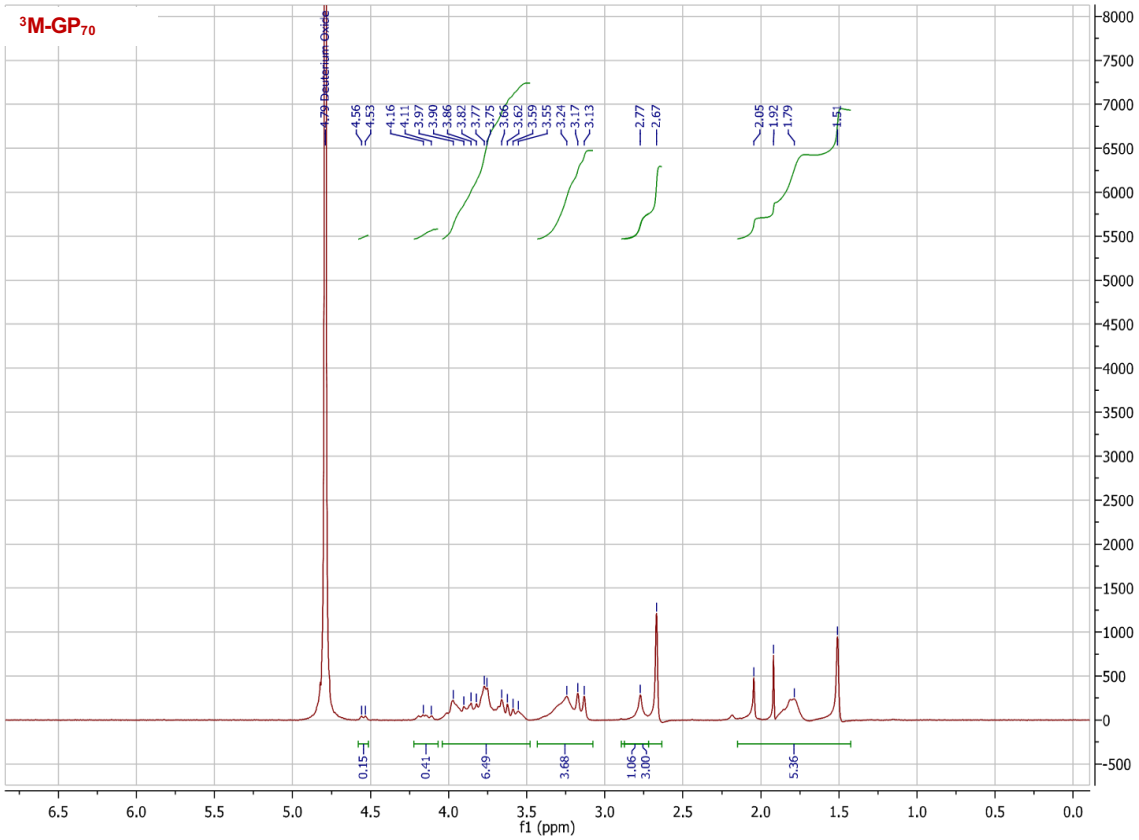
^{1}H NMR of ^{3}M-GP$_{65}$

D_2O, 300 MHz δ (ppm): 4.55-4.50 (bd, 0.10H), 4.25-4.10 (bm, 0.21H), 4.05-3.50 (bm, 3.85H), 3.40-3.10 (bm, 3.68H), 2.85-2.60 (bm, 3H), 2.25-1.50 (bm, 5.27H).



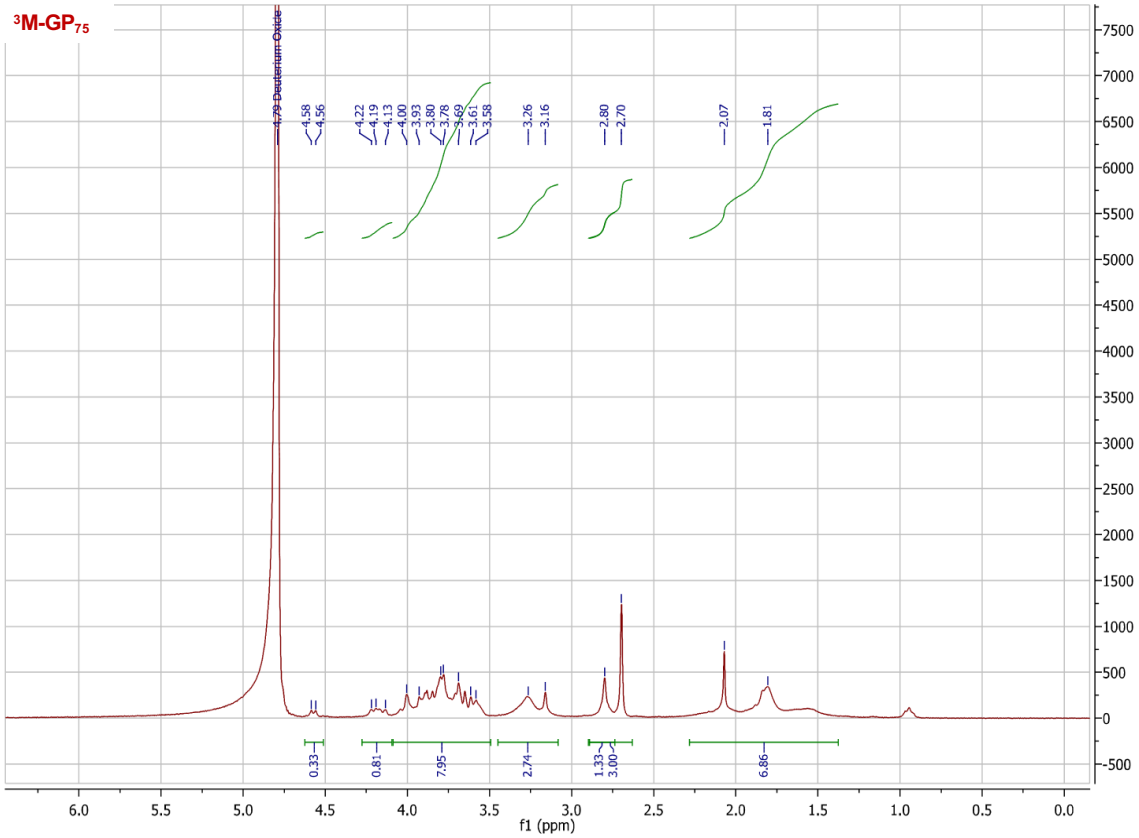
^{1}H NMR of ^{3}M-GP$_{70}$

D_2O, 300 MHz δ (ppm): 4.55-4.50 (bd, 0.15H), 4.25-4.10 (bm, 0.41H), 4.05-3.50 (bm, 6.49H), 3.45-3.10 (bm, 3.68H), 2.85-2.60 (bm, 3H), 2.20-1.40 (bm, 5.36H).



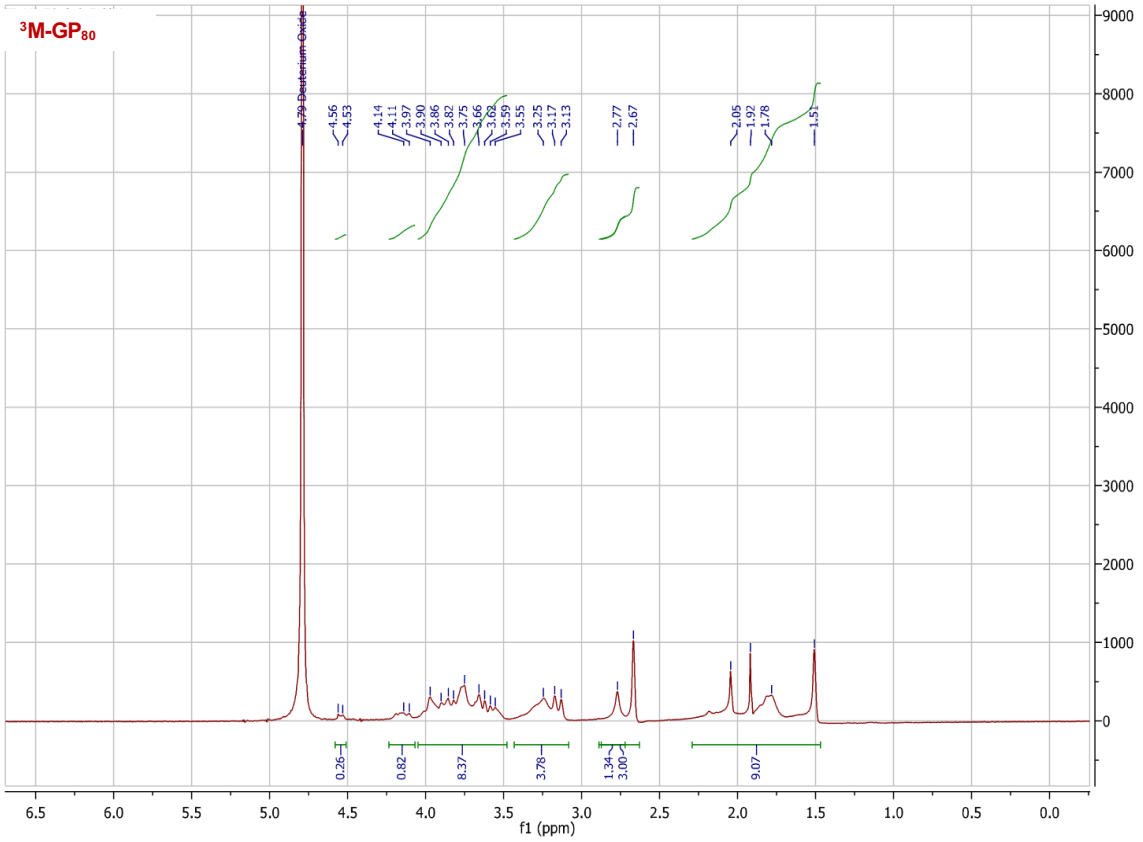
^{1}H NMR of ^{3}M-GP$_{75}$

D_2O, 300 MHz δ (ppm): 4.55-4.50 (bd, 0.33H), 4.25-4.10 (bm, 0.81H), 4.10-3.50 (bm, 7.95H), 3.45-3.10 (bm, 2.74H), 2.85-2.60 (bm, 3H), 2.25-1.40 (bm, 6.86H).



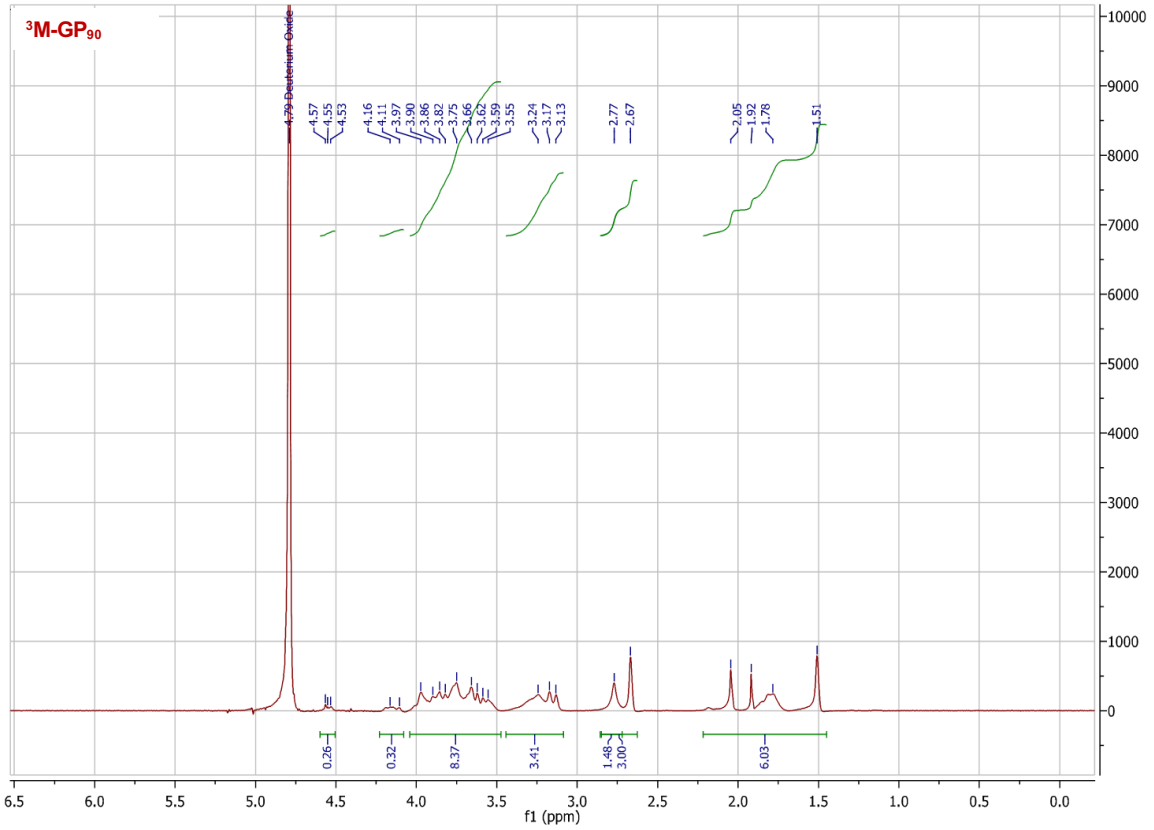
^{1}H NMR of ^{3}M-GP$_{80}$

D_2O, 300 MHz δ (ppm): 4.55-4.50 (bd, 0.26H), 4.25-4.10 (bm, 0.82H), 4.10-3.50 (bm, 8.37H), 3.45-3.10 (bm, 3.78H), 2.85-2.60 (bm, 3H), 2.25-1.45 (bm, 9.07H).



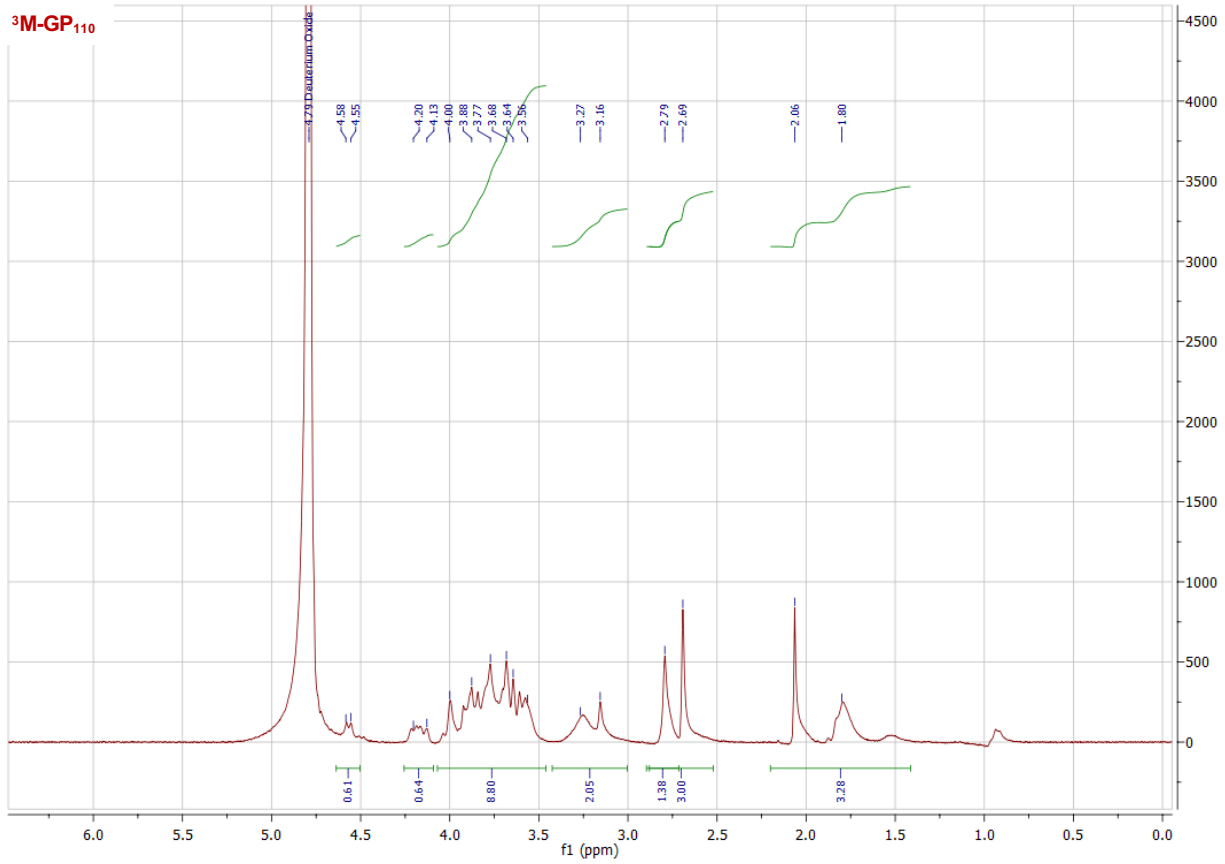
^{1}H NMR of ^{3}M-GP$_{90}$

D_2O, 300 MHz δ (ppm): 4.55-4.50 (bd, 0.26H), 4.20-4.10 (bm, 0.32H), 4.05-3.50 (bm, 8.37H), 3.45-3.10 (bm, 3.41H), 2.80-2.60 (bm, 3H), 2.20-1.45 (bm, 6.03H).



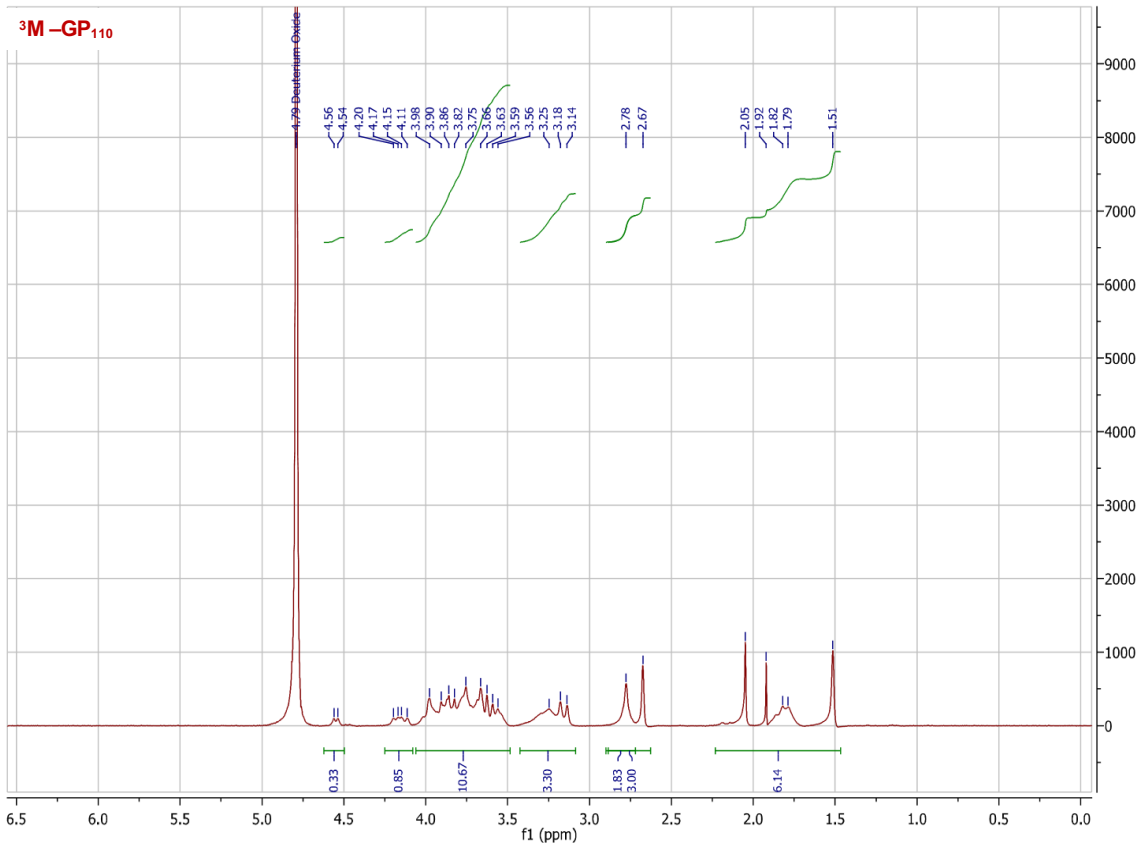
^{1}H NMR of ^{3}M-GP$_{110}$

D_2O, 300 MHz δ (ppm): 4.55-4.50 (bd, 0.61H), 4.25-4.10 (bm, 0.64H), 4.10-3.45 (bm, 8.80H), 3.40-3.00 (bm, 2.05H), 2.85-2.50 (bm, 3H), 2.20-1.40 (bm, 3.28H).



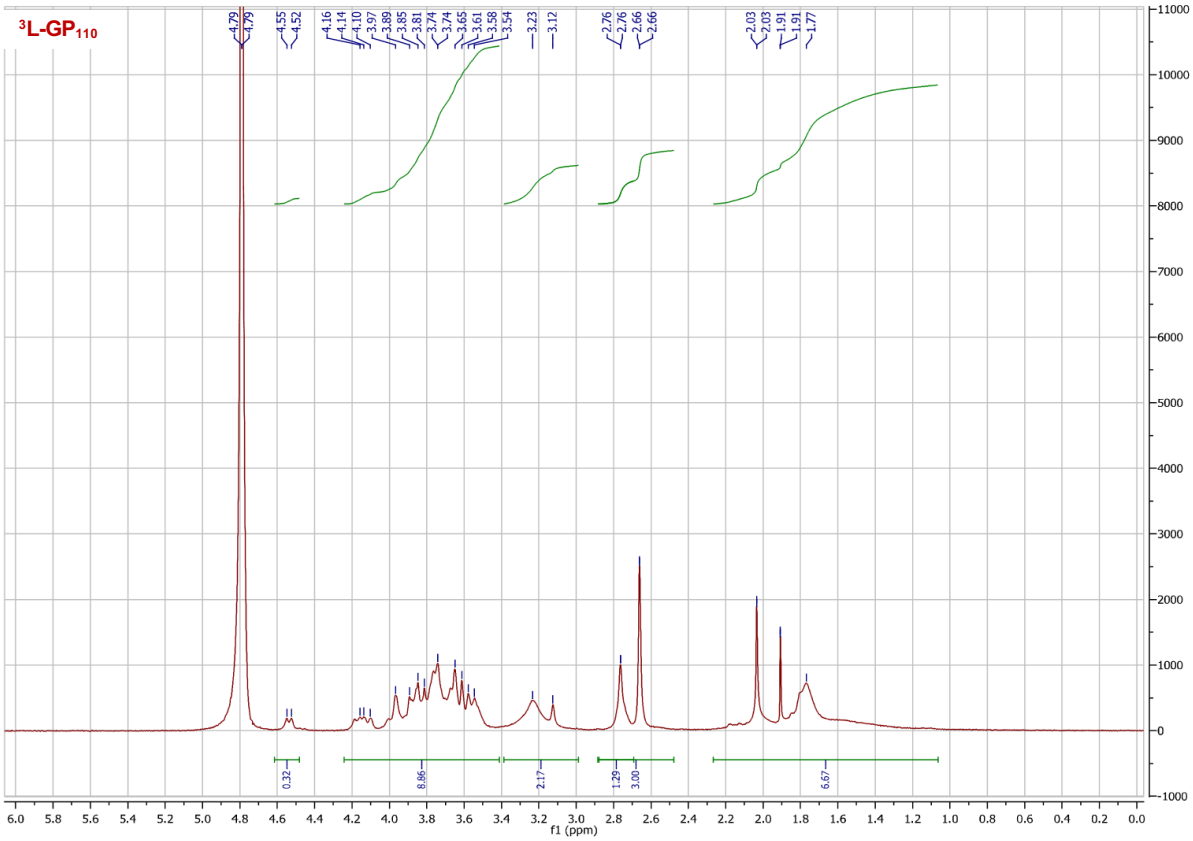
^{1}H NMR of ^{3}M-GP$_{110}$ (2)

D_2O, 300 MHz δ (ppm): 4.60-4.50 (bd, 0.33H), 4.20-4.10 (bm, 0.85H), 4.05-3.50 (bm, 10.67H), 3.45-3.10 (bm, 3.30H), 2.85-2.60 (bm, 3H), 2.20-1.45 (bm, 6.14H).



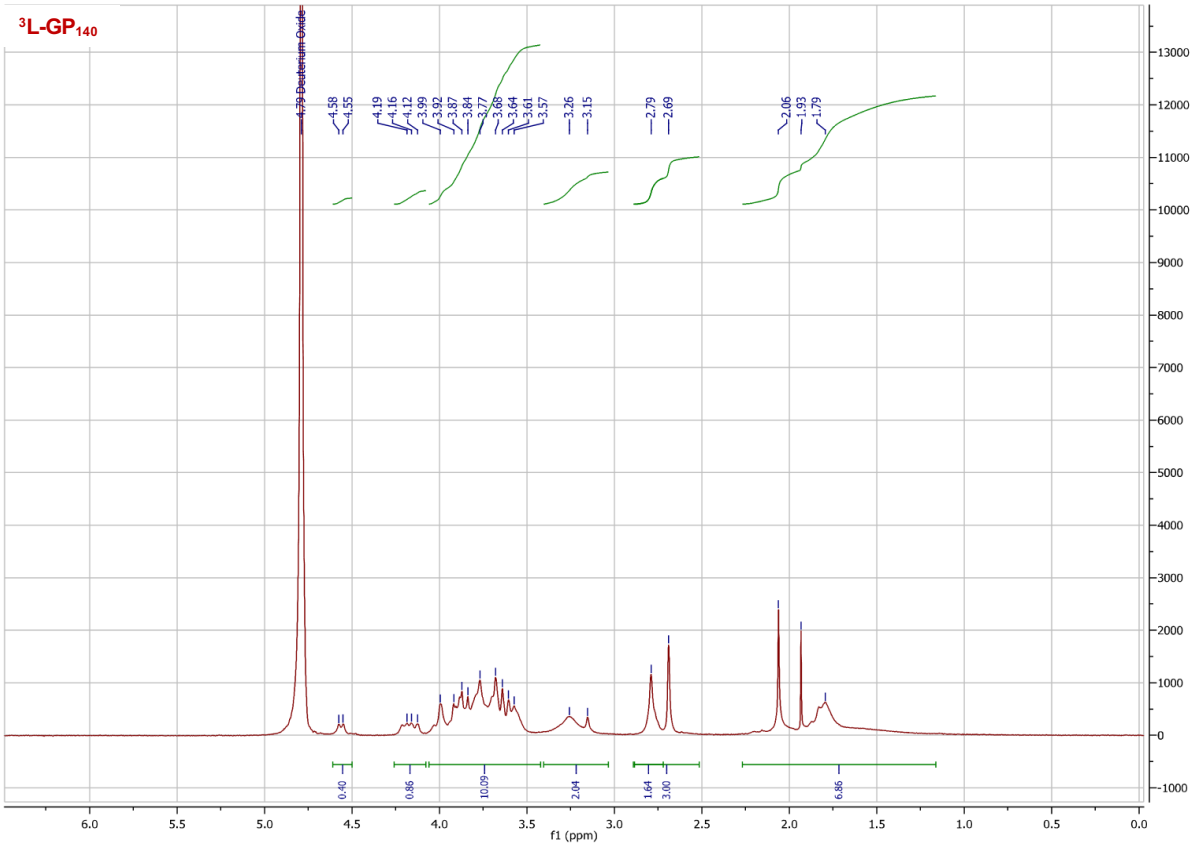
^{1}H NMR of ^{3}L-GP$_{110}$

D_2O, 300 MHz δ (ppm): 4.60-4.50 (bd, 0.32H), 4.25-3.40 (bm, 8.86H), 3.40-3.00 (bm, 2.17H), 2.85-2.50 (bm, 3H), 2.25-1.05 (bm, 6.67H).



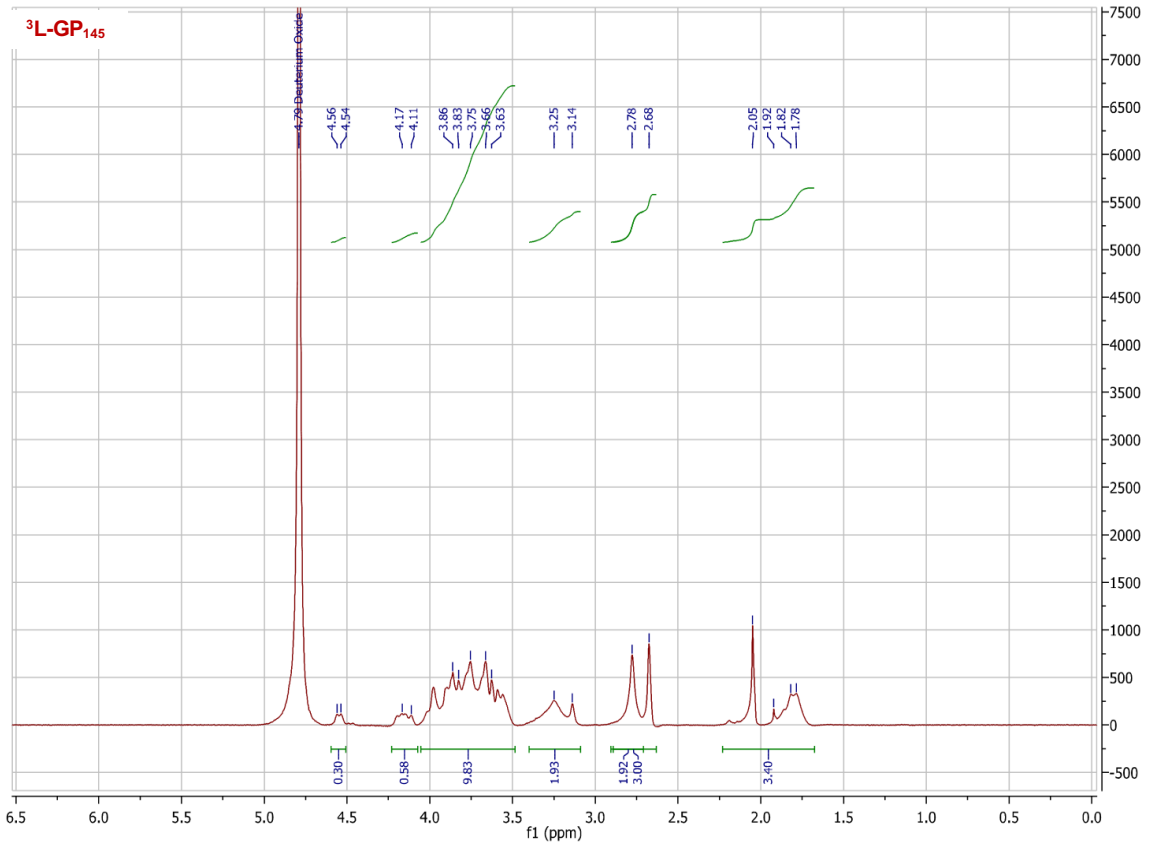
^{1}H NMR of ^{3}L-GP$_{140}$

D_2O, 300 MHz δ (ppm): 4.55-4.50 (bd, 0.40H), 4.25-4.10 (bm, 0.86H), 4.10-3.45 (bm, 10.09H), 3.45-3.05 (bm, 2.04H), 2.85-2.50 (bm, 3H), 2.25-1.15 (bm, 6.86H).



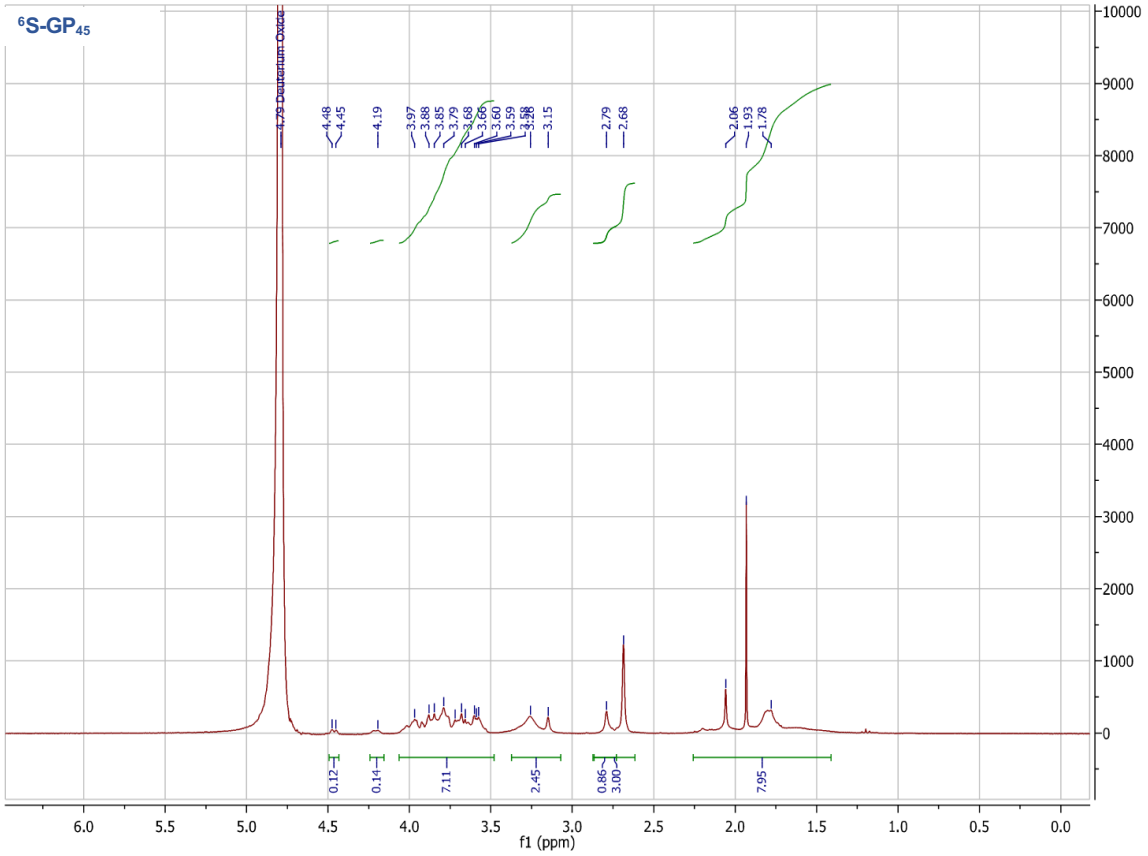
^{1}H NMR of ^{3}L-GP$_{145}$

D_2O, 300 MHz δ (ppm): 4.55-4.50 (bd, 0.30H), 4.25-4.10 (bm, 0.58H), 4.10-3.50 (bm, 9.83H), 3.40-3.10 (bm, 1.93H), 2.85-2.60 (bm, 3H), 2.25-1.60 (bm, 3.40H).



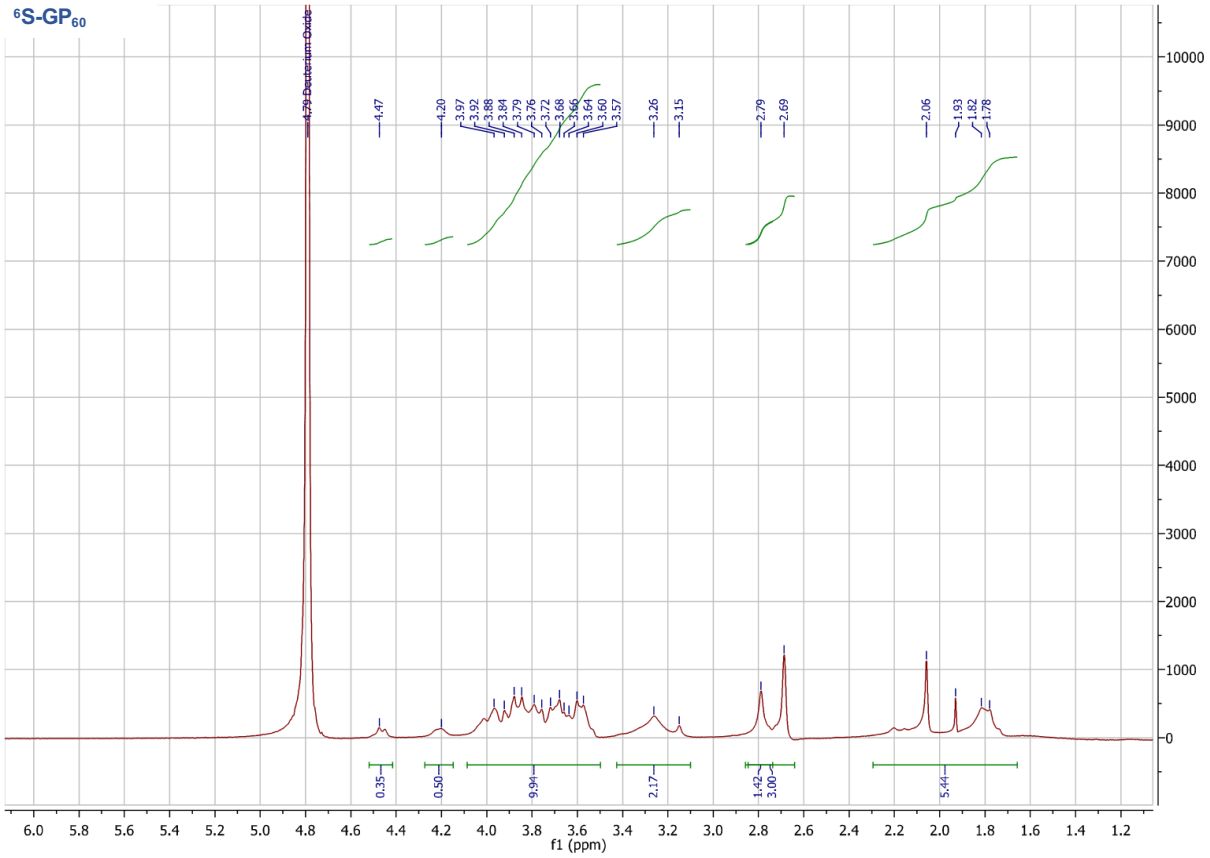
^{1}H NMR of ^{6}S-GP$_{45}$

D_2O, 300 MHz δ (ppm): 4.50-4.45 (bd, 0.12H), 4.25-4.15 (bm, 0.14H), 4.05-3.50 (bm, 7.11H), 3.40-3.10 (bm, 2.45H), 2.85-2.60 (bm, 3H), 2.25-1.40 (bm, 7.95H).



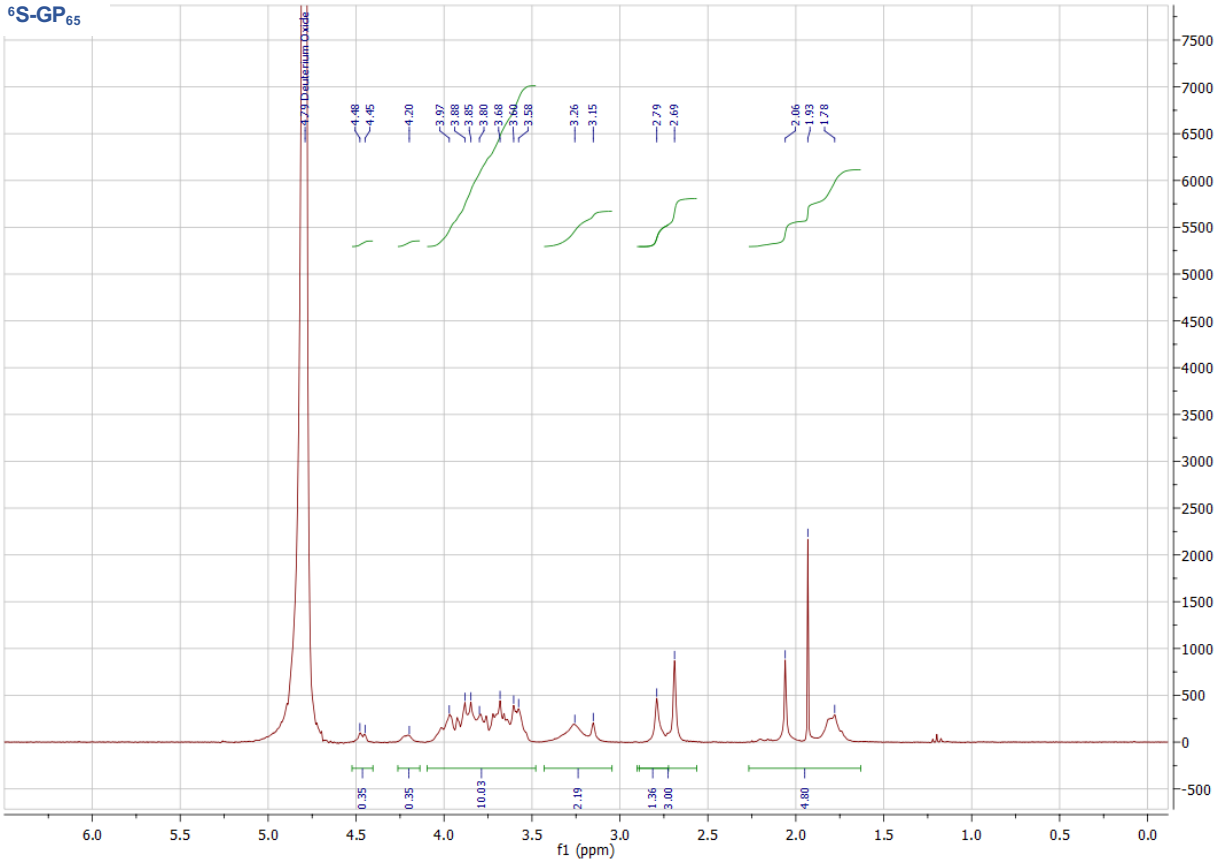
^{1}H NMR of ^{6}S-GP$_{60}$

D_2O, 300 MHz δ (ppm): 4.50-4.40 (bd, 0.35H), 4.25-4.15 (bm, 0.50H), 4.10-3.50 (bm, 9.94H), 3.40-3.10 (bm, 2.17H), 2.85-2.65 (bm, 3H), 2.30-1.65 (bm, 5.44H).



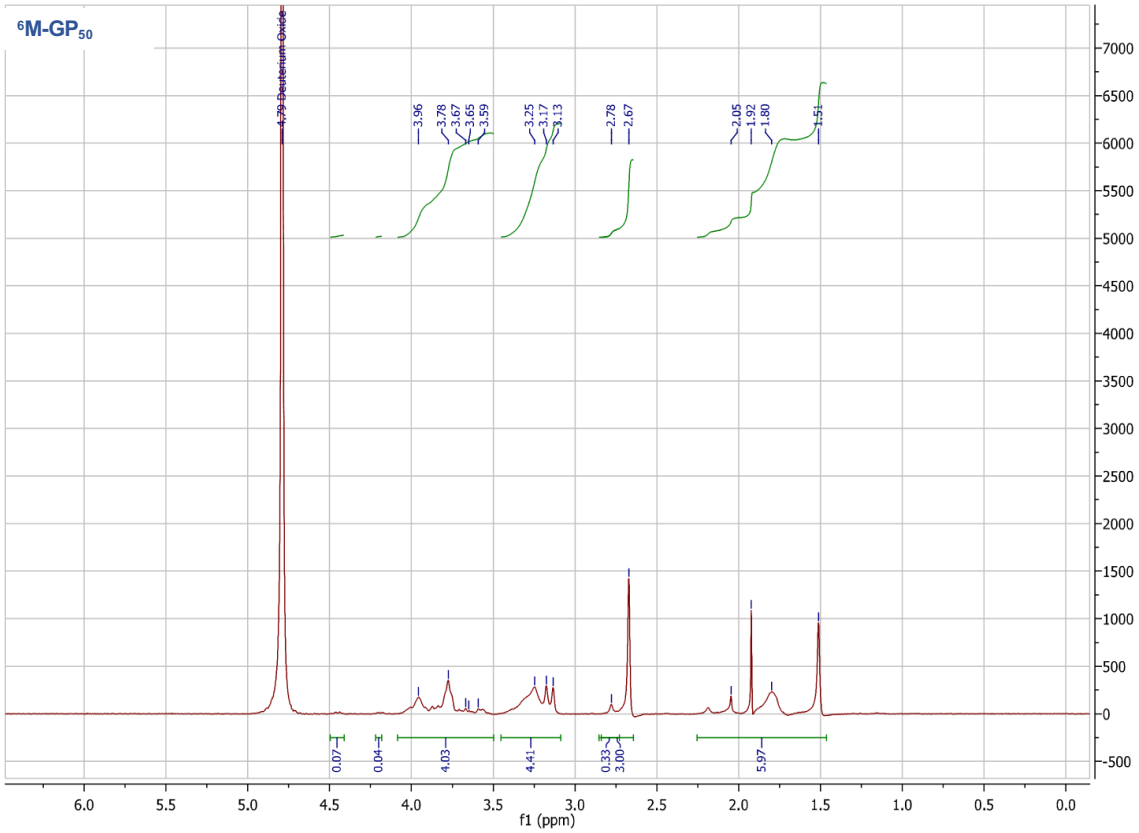
^{1}H NMR of ^{6}S-GP$_{65}$

D_2O, 300 MHz δ (ppm): 4.50-4.40 (bd, 0.35H), 4.25-4.15 (bm, 0.35H), 4.10-3.50 (bm, 10.03H), 3.45-3.05 (bm, 2.19H), 2.85-2.60 (bm, 3H), 2.25-1.60 (bm, 4.80H).



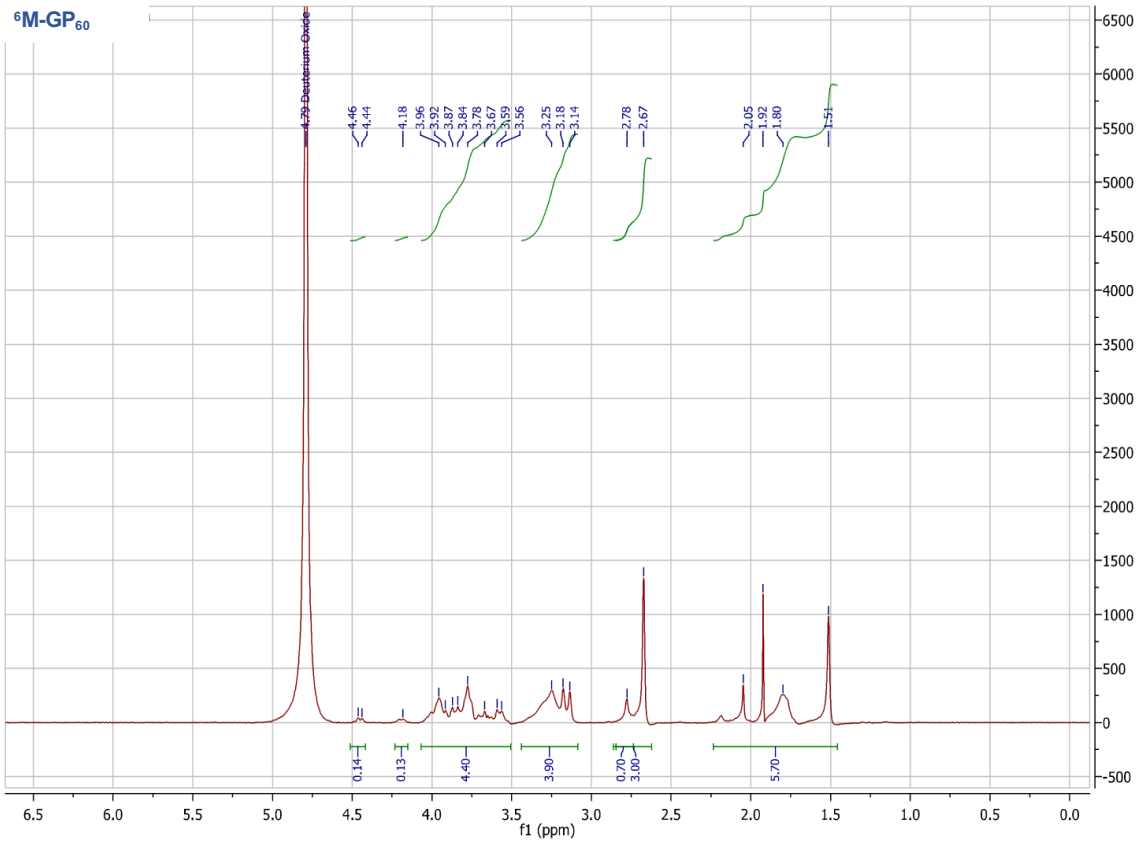
^{1}H NMR of ^{6}M-GP$_{50}$

D_2O, 300 MHz δ (ppm): 4.50-4.45 (bd, 0.07H), 4.20-4.15 (bm, 0.04H), 4.10-3.50 (bm, 4.03H), 3.45-3.10 (bm, 4.41H), 2.85-2.60 (bm, 3H), 2.25-1.45 (bm, 5.97H).



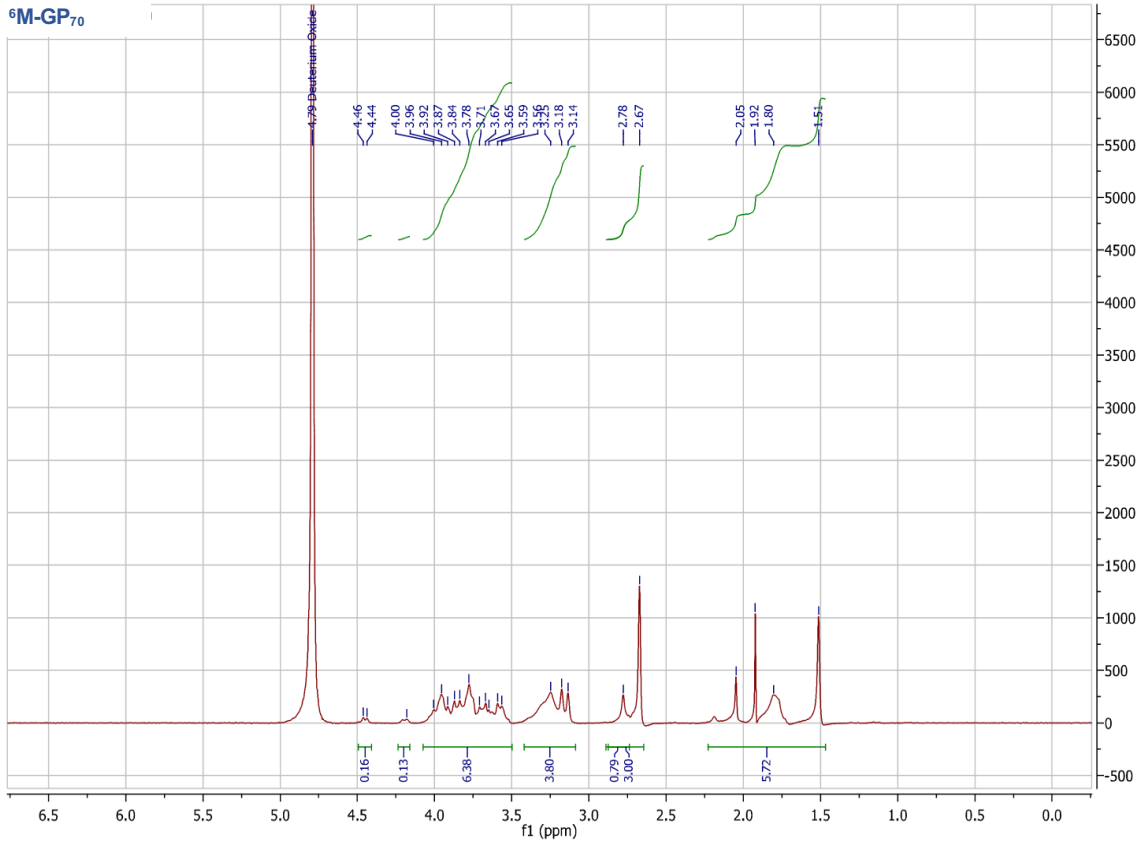
^{1}H NMR of ^{6}M-GP$_{60}$

D_2O, 300 MHz δ (ppm): 4.50-4.45 (bd, 0.14H), 4.25-4.20 (bm, 0.13H), 4.15-3.50 (bm, 4.40H), 3.45-3.10 (bm, 3.90H), 2.85-2.60 (bm, 3H), 2.25-1.45 (bm, 5.70H).



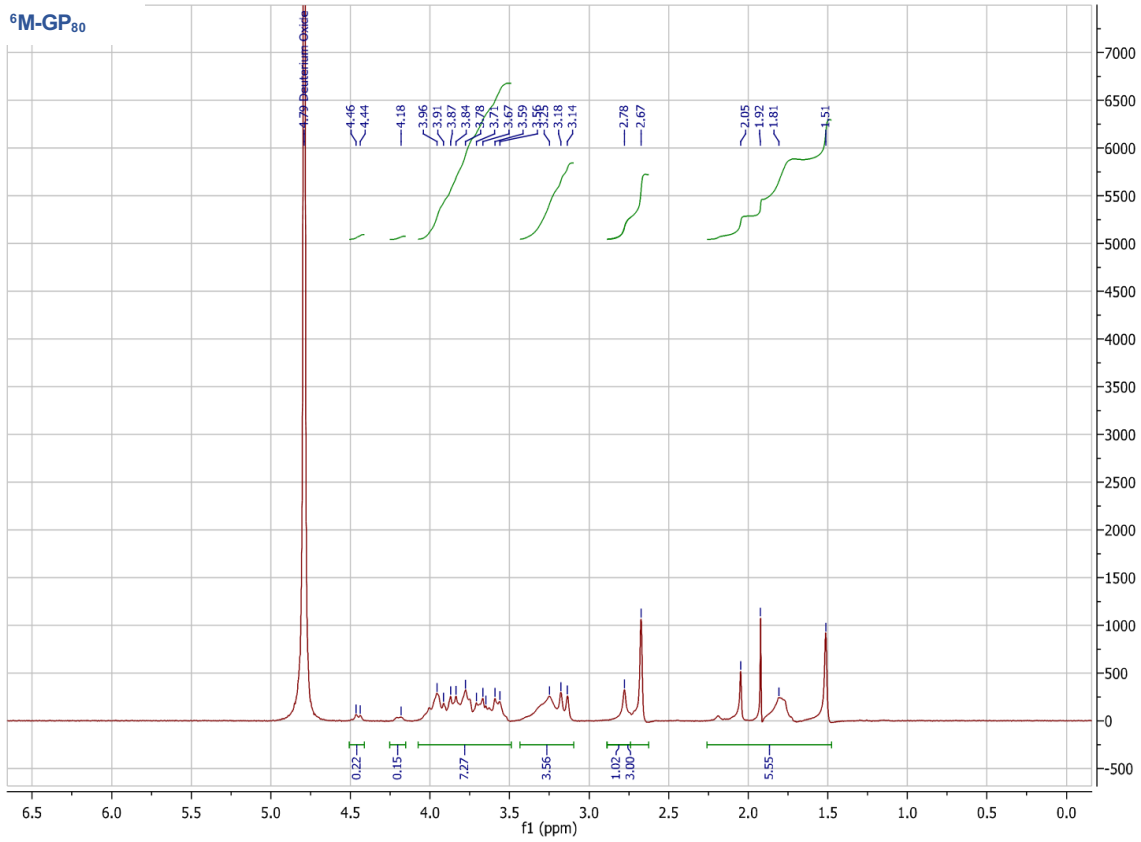
^{1}H NMR of ^{6}M-GP$_{70}$

D_2O, 300 MHz δ (ppm): 4.50-4.45 (bd, 0.16H), 4.25-4.20 (bm, 0.13H), 4.10-3.50 (bm, 6.38H), 3.40-3.10 (bm, 3.80H), 2.85-2.60 (bm, 3H), 2.25-1.50 (bm, 5.72H).



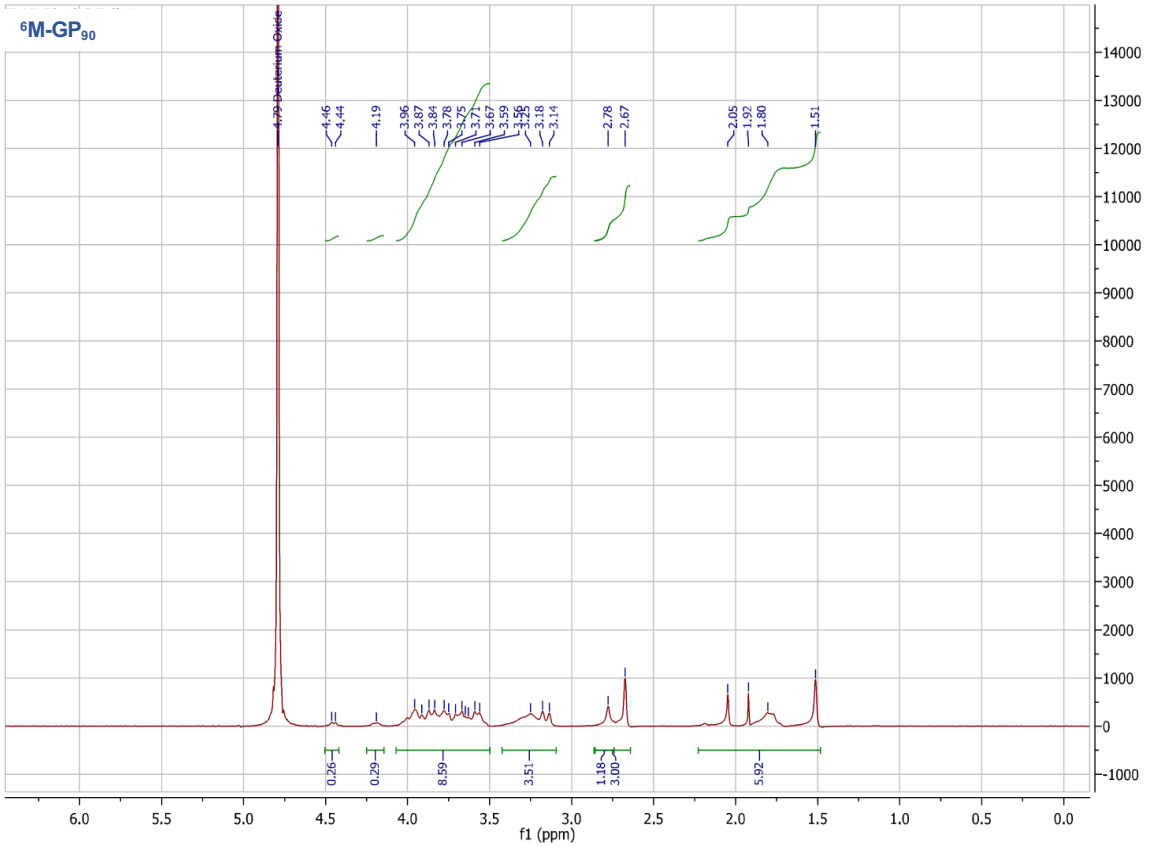
^{1}H NMR of ^{6}M-GP$_{80}$

D_2O, 300 MHz δ (ppm): 4.50-4.45 (bd, 0.22H), 4.25-4.15 (bm, 0.15H), 4.05-3.50 (bm, 7.27H), 3.45-3.10 (bm, 3.56H), 2.85-2.60 (bm, 3H), 2.25-1.50 (bm, 5.55H).



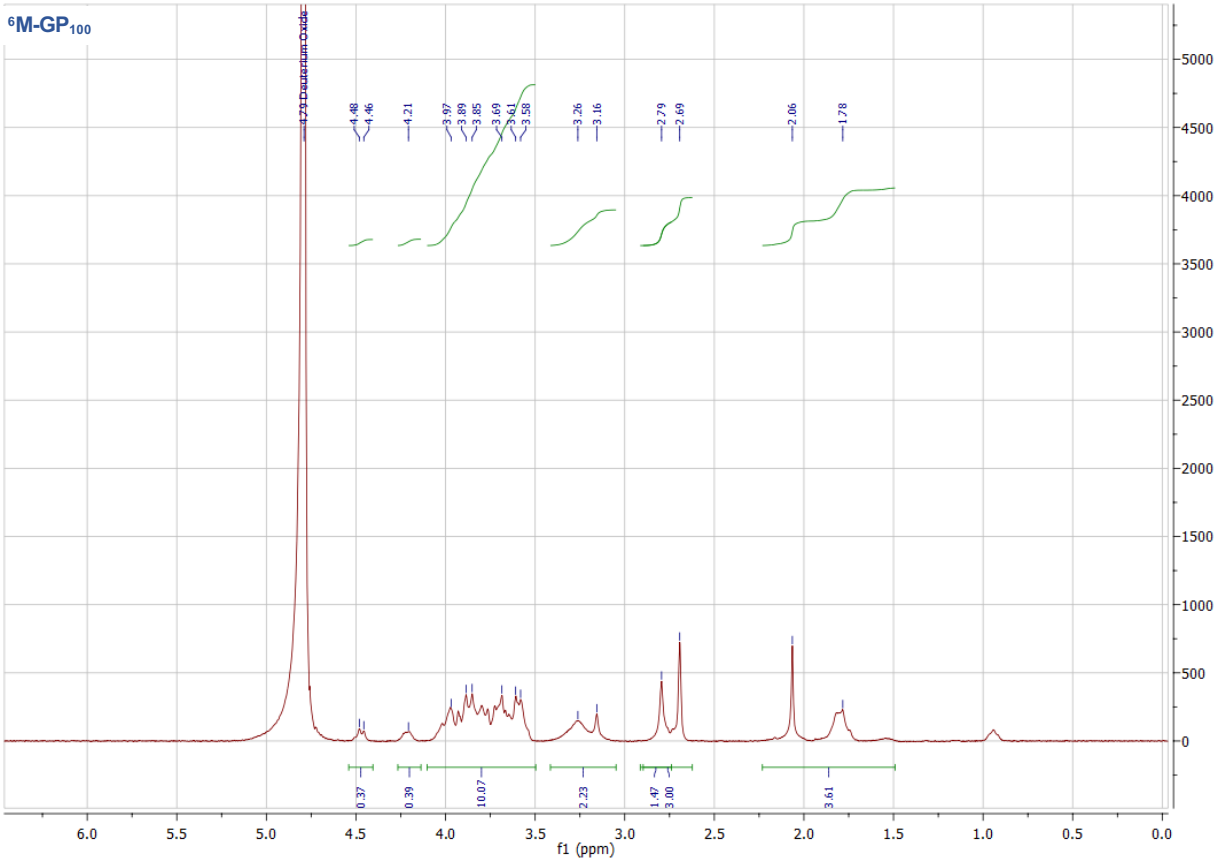
^{1}H NMR of ^{6}M-GP$_{90}$

D_2O, 300 MHz δ (ppm): 4.50-4.45 (bd, 0.26H), 4.25-4.10 (bm, 0.29H), 4.05-3.50 (bm, 8.59H), 3.40-3.10 (bm, 3.51H), 2.85-2.60 (bm, 3H), 2.25-1.50 (bm, 5.92H).



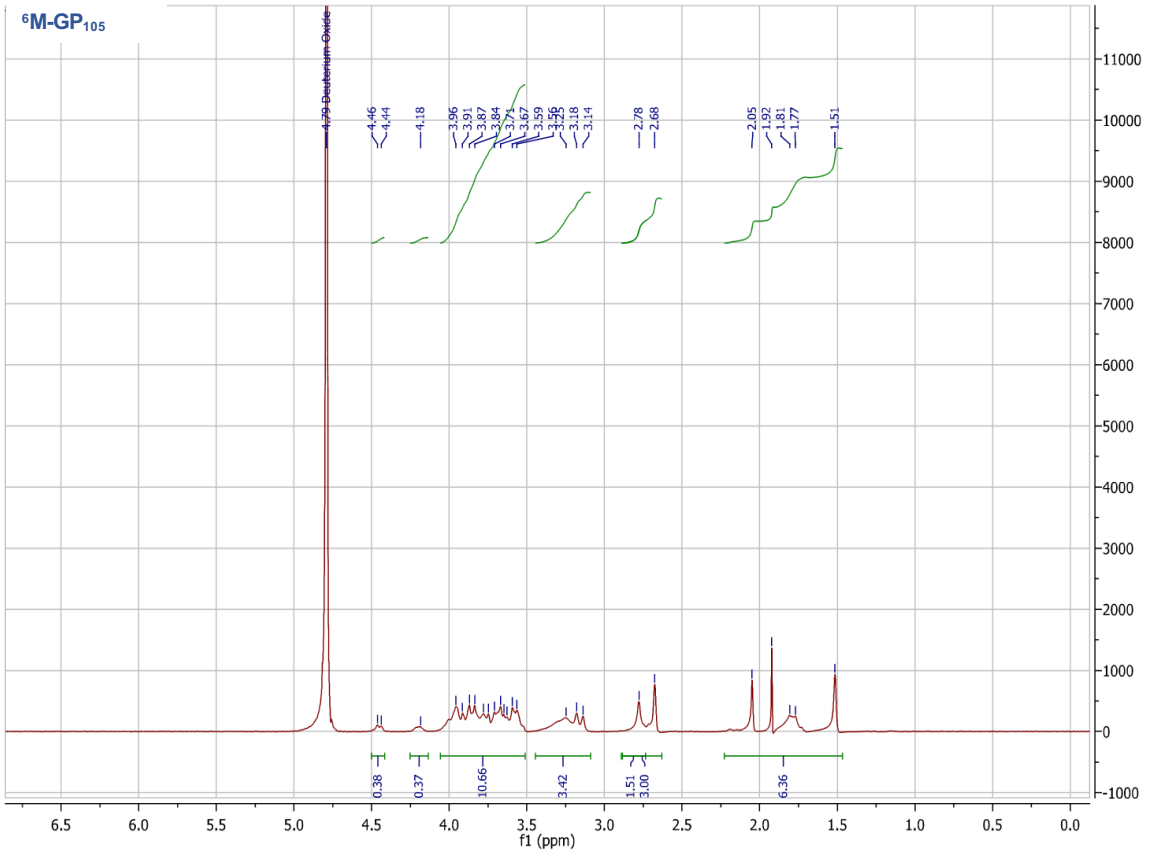
^{1}H NMR of ^{6}M-GP$_{100}$

D_2O, 300 MHz δ (ppm): 4.55-4.45 (bd, 0.37H), 4.25-4.15 (bm, 0.39H), 4.10-3.50 (bm, 10.07H), 3.40-3.05 (bm, 2.23H), 2.85-2.60 (bm, 3H), 2.25-1.50 (bm, 3.61H).



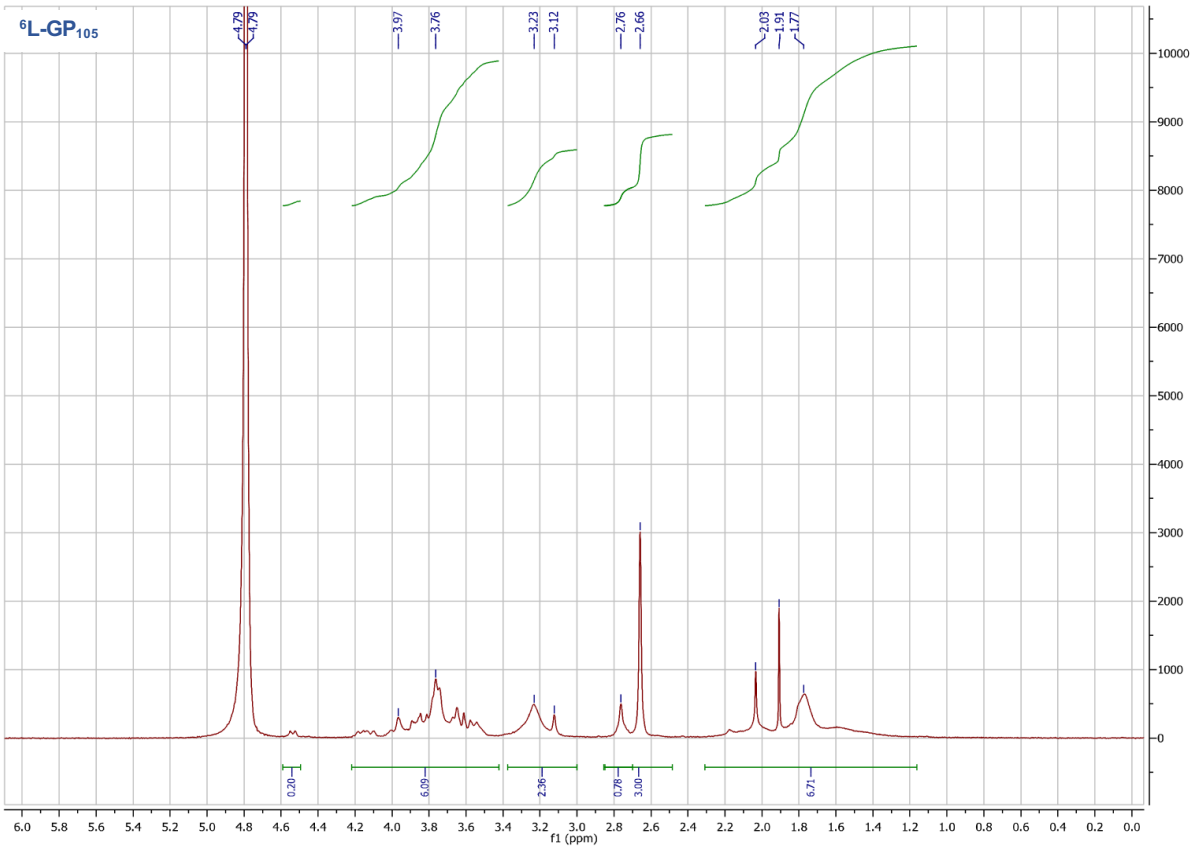
^{1}H NMR of ^{6}M-GP$_{105}$

D_2O, 300 MHz δ (ppm): 4.50-4.45 (bd, 0.38H), 4.25-4.10 (bm, 0.0.37H), 4.05-3.50 (bm, 10.66H), 3.45-3.10 (bm, 3.42H), 2.85-2.60 (bm, 3H), 2.25-1.50 (bm, 6.36H).



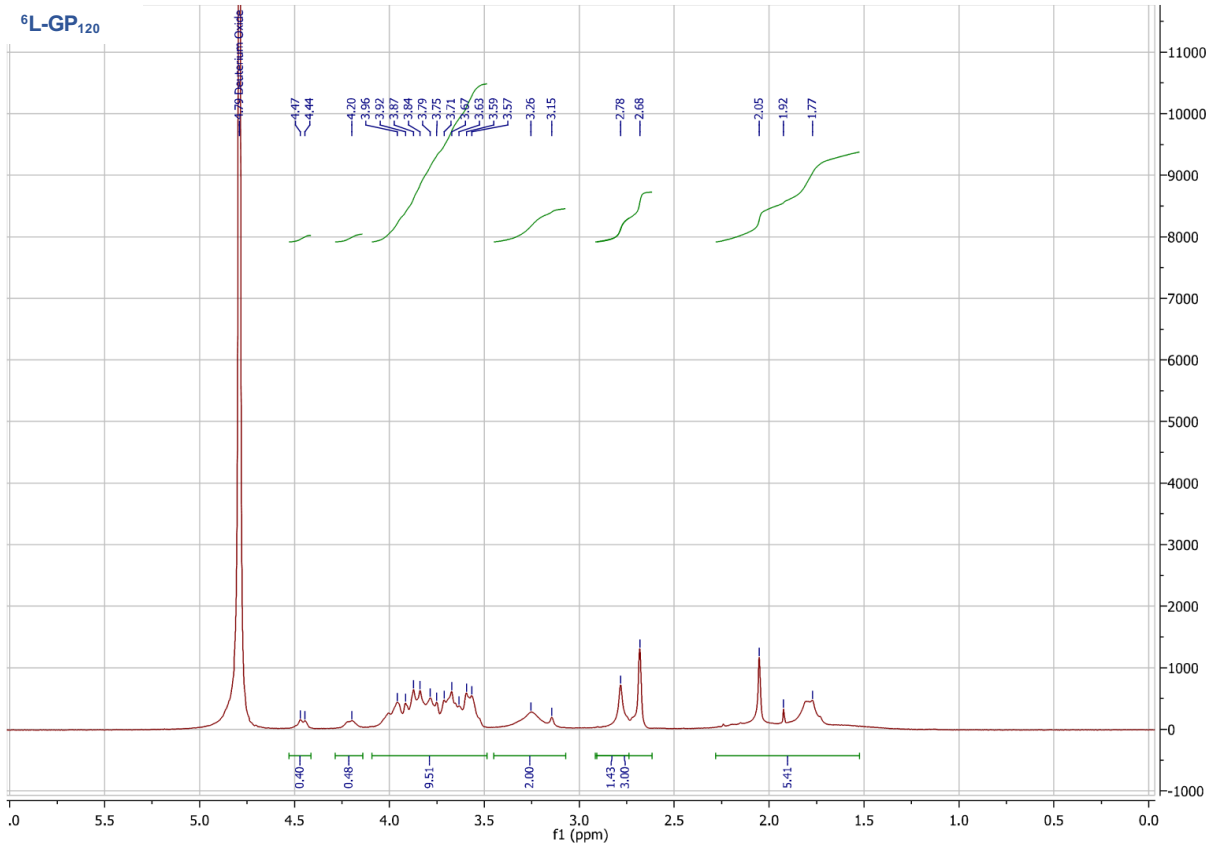
^{1}H NMR of ^{6}L-GP$_{105}$

D_2O, 300 MHz δ (ppm): 4.60-4.45 (bd, 0.0.20H), 4.20-3.40 (bm, 6.09H), 3.40-3.00 (bm, 2.36H), 2.85-2.50 (bm, 3H), 2.30-1.15 (bm, 6.71H).



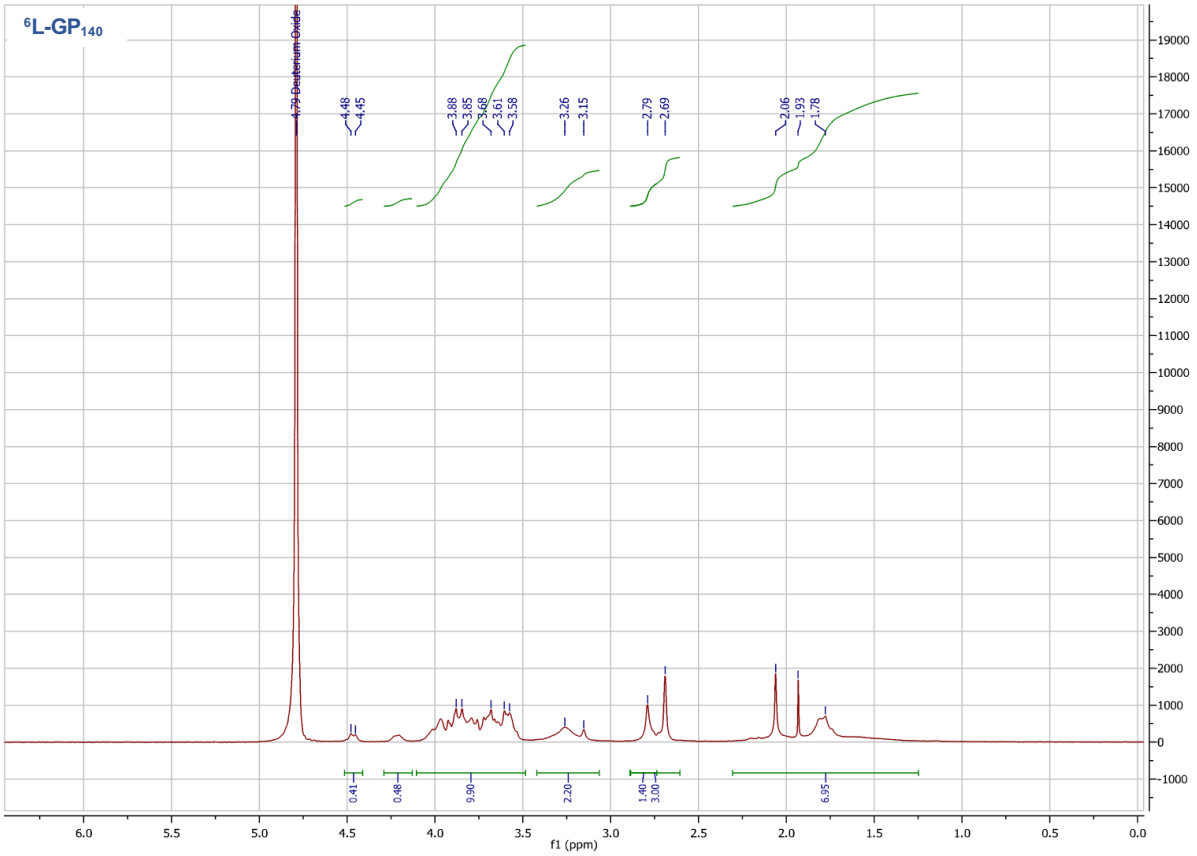
^{1}H NMR of ^{6}L-GP$_{120}$

D_2O, 300 MHz δ (ppm): 4.55-4.45 (bd, 0.0.40H), 4.30-4.10 (bm, 0.48H), 4.05-3.50 (bm, 9.51H), 3.45-3.10 (bm, 2.00H), 2.85-2.60 (bm, 3H), 2.25-1.50 (bm, 5.41H).



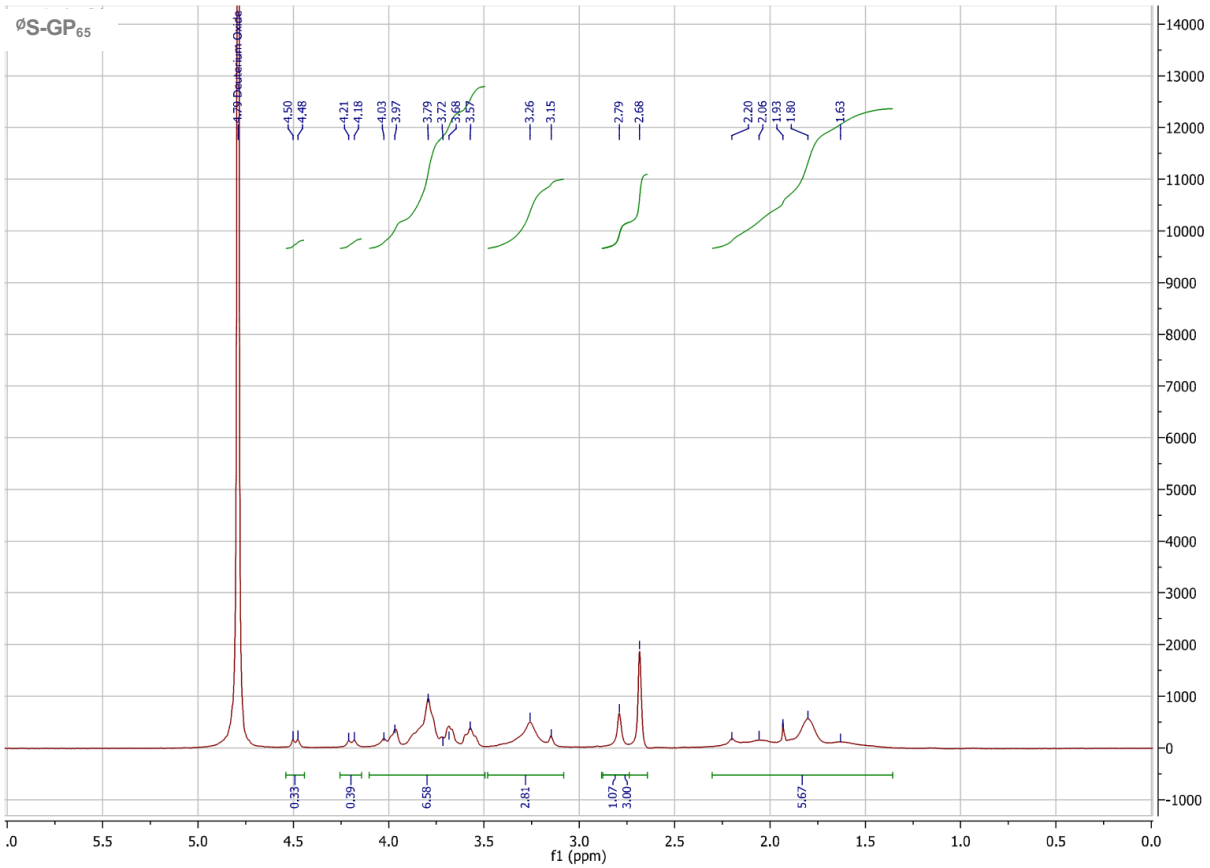
^{1}H NMR of ^{6}L-GP$_{140}$

D_2O, 300 MHz δ (ppm): 4.50-4.45 (bd, 0.41H), 4.25-4.10 (bm, 0.48H), 4.10-3.50 (bm, 9.90H), 3.40-3.10 (bm, 2.20H), 2.85-2.60 (bm, 3H), 2.30-1.25 (bm, 6.95H).



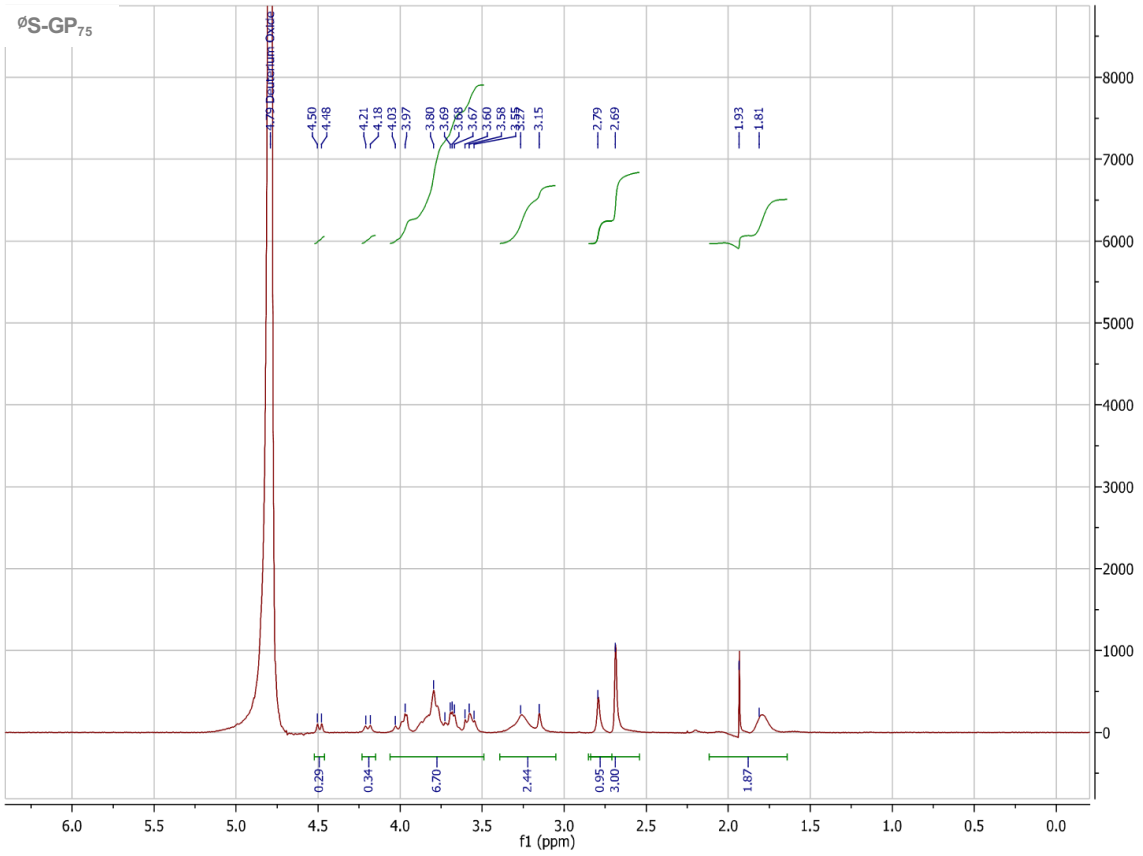
^{1}H NMR of ^{Ø}S-GP$_{65}$

D_2O, 300 MHz δ (ppm): 4.55-4.45 (bd, 0.33H), 4.25-4.15 (bm, 0.39H), 4.10-3.50 (bm, 6.58H), 3.50-3.10 (bm, 2.81H), 2.85-2.60 (bm, 3H), 2.30-1.40 (bm, 5.67H).



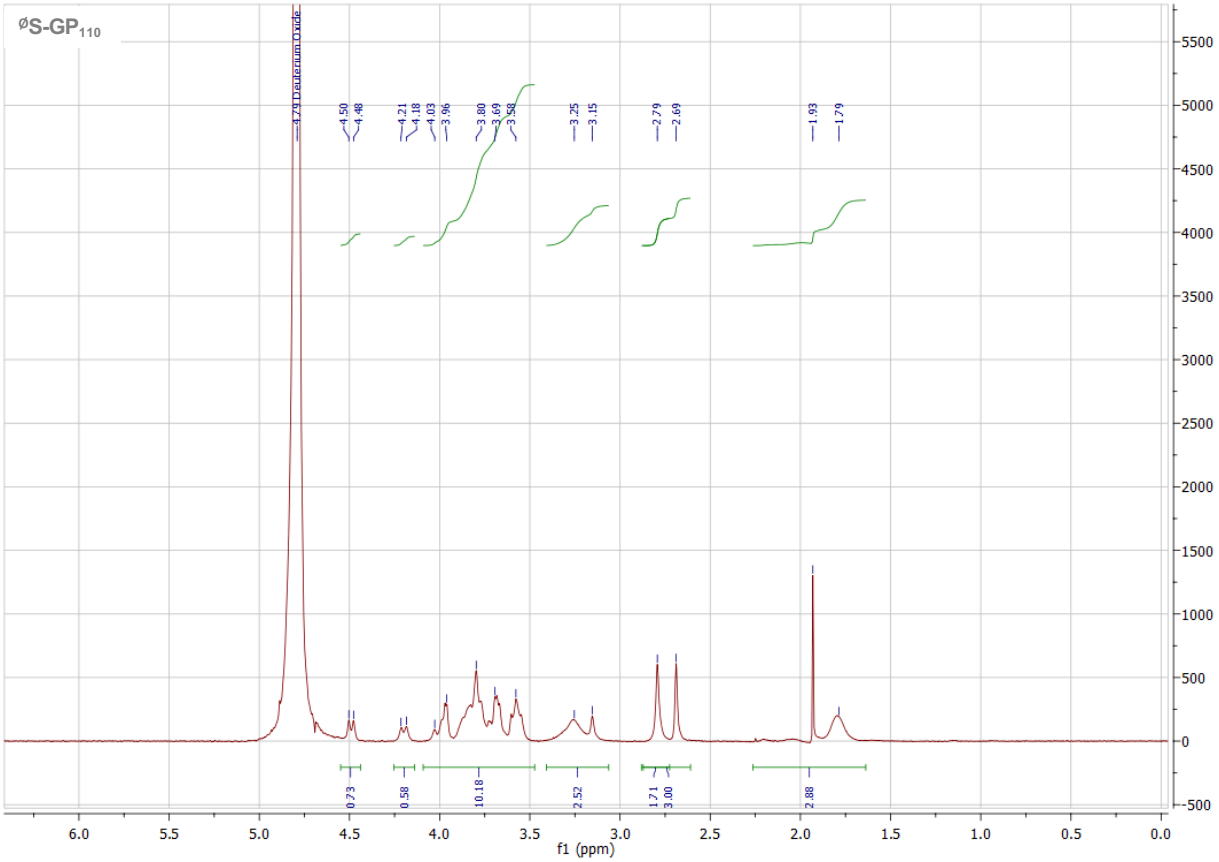
^{1}H NMR of ^{Ø}S-GP$_{75}$

D_2O, 300 MHz δ (ppm): 4.56-4.53 (bd, 0.29H), 4.25-4.20 (bm, 0.34H), 4.05-3.50 (bm, 6.70H), 3.40-3.05 (bm, 2.44H), 2.85-2.60 (bm, 3H), 2.10-1.60 (bm, 1.87H).



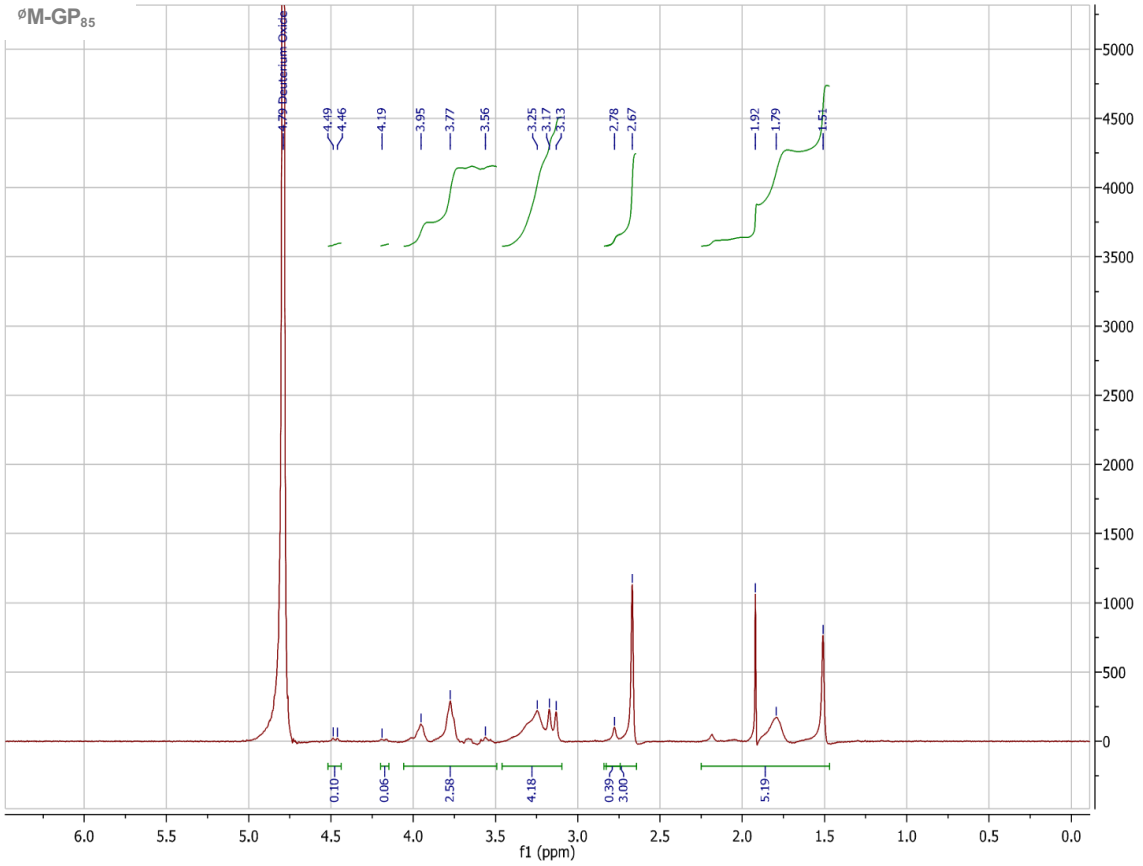
^{1}H NMR of ^{Ø}S-GP$_{110}$

D_2O, 300 MHz δ (ppm): 4.55-4.45 (bd, 0.73H), 4.25-4.15 (bm, 0.58H), 4.05-3.50 (bm, 10.18H), 3.40-3.10 (bm, 2.52H), 2.85-2.60 (bm, 3H), 2.25-1.60 (bm, 2.88H).



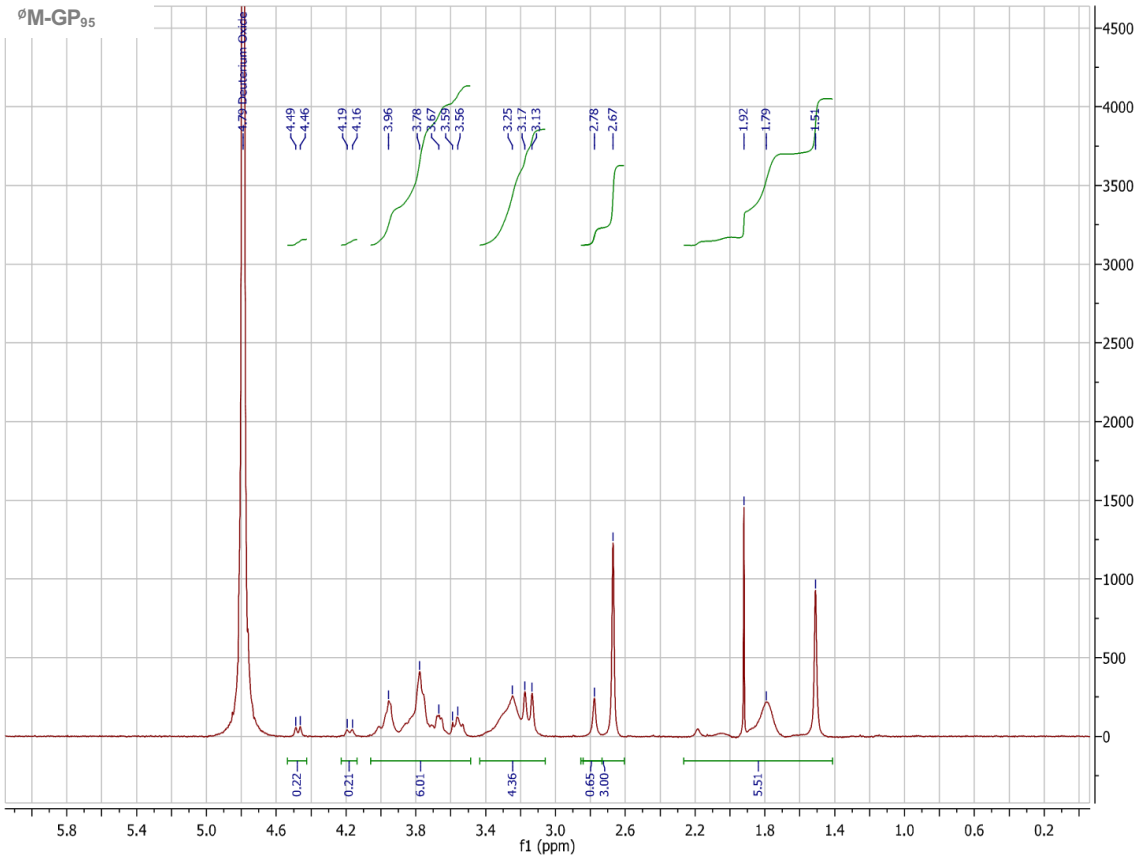
^{1}H NMR of $^{\emptyset}$M-GP$_{85}$

D_2O, 300 MHz δ (ppm): 4.50-4.45 (bd, 0.10H), 4.20-4.15 (bm, 0.06H), 4.05-3.50 (bm, 2.58H), 3.45-3.10 (bm, 4.18H), 2.85-2.60 (bm, 3H), 2.25-1.45 (bm, 5.19H).



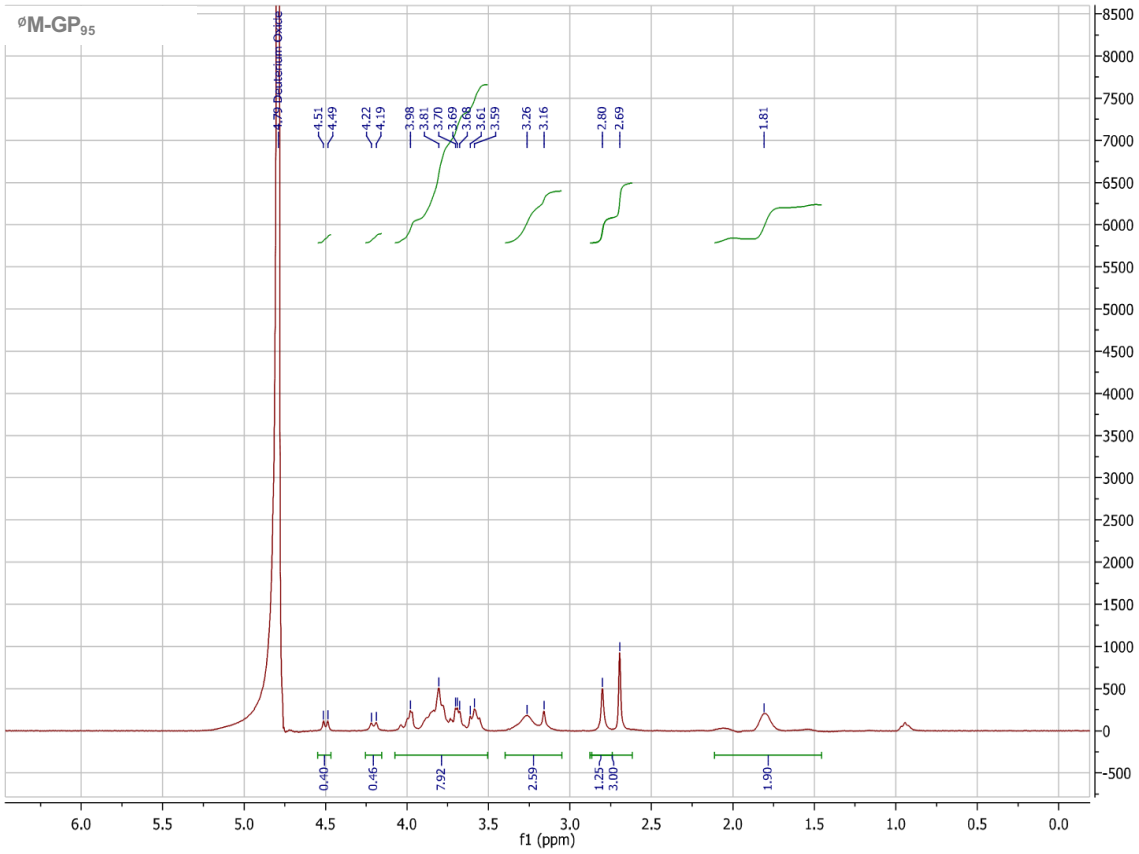
^{1}H NMR of ^{Ø}M-GP$_{95}$

D_2O, 300 MHz δ (ppm): 4.55-4.45 (bd, 0.22H), 4.25-4.10 (bm, 0.21H), 4.15-3.50 (bm, 6.01H), 3.45-3.05 (bm, 4.36H), 2.85-2.60 (bm, 3H), 2.25-1.40 (bm, 5.51H).



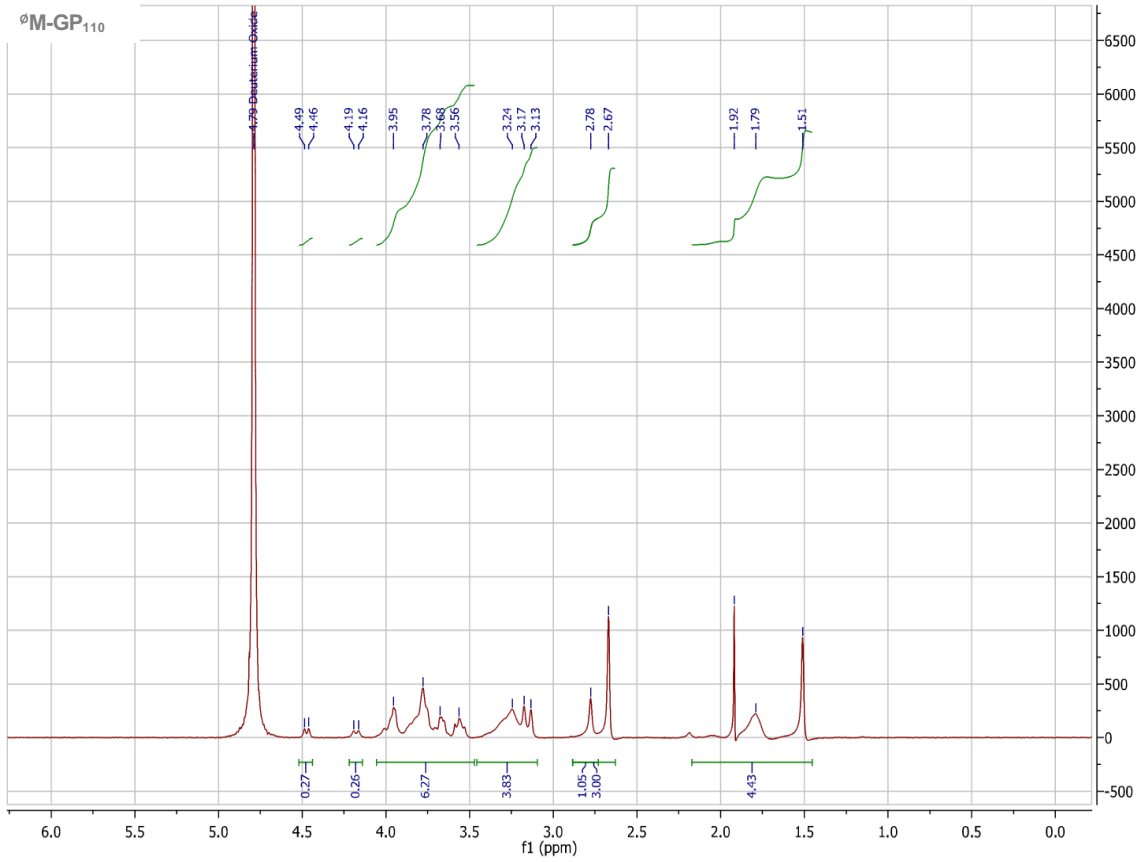
^{1}H NMR of $^{\emptyset}$M-GP$_{95}$ (2)

D_2O, 300 MHz δ (ppm): 4.53-4.48 (bd, 0.40H), 4.25-4.15 (bm, 0.46H), 4.05-3.50 (bm, 7.92H), 3.40-3.10 (bm, 2.59H), 2.85-2.60 (bm, 3H), 2.10-1.45 (bm, 1.90H).



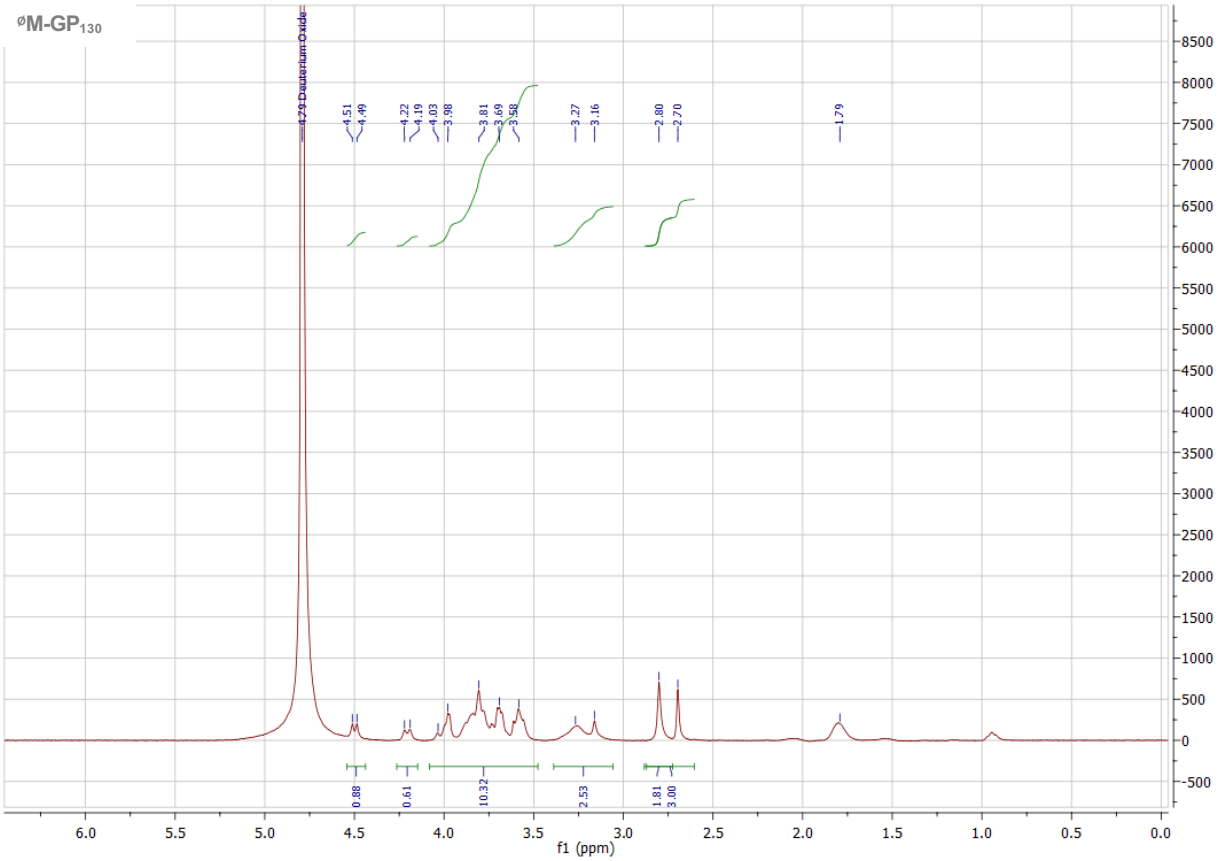
^{1}H NMR of ^{Ø}M-GP$_{110}$

D_2O, 300 MHz δ (ppm): 4.51-4.47 (bd, 0.27H), 4.20-4.15 (bm, 0.26H), 4.05-3.45 (bm, 6.27H), 3.45-3.10 (bm, 3.83H), 2.85-2.60 (bm, 3H), 2.20-1.45 (bm, 4.43H).



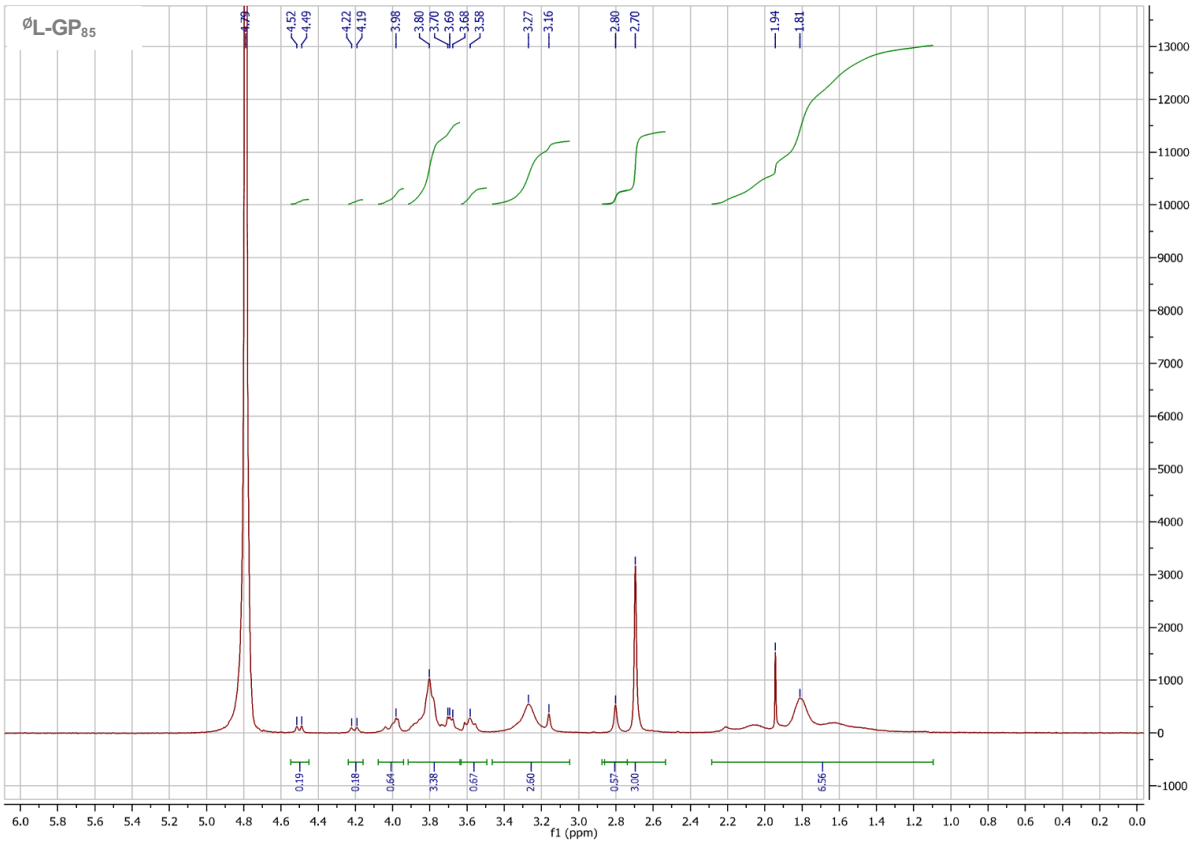
^{1}H NMR of ^{Ø}M-GP$_{130}$

D_2O, 300 MHz δ (ppm): 4.55-4.45 (bd, 0.88H), 4.25-4.15 (bm, 0.61H), 4.10-3.50 (bm, 10.32H), 3.40-3.05 (bm, 2.53H), 2.85-2.60 (bm, 3H).



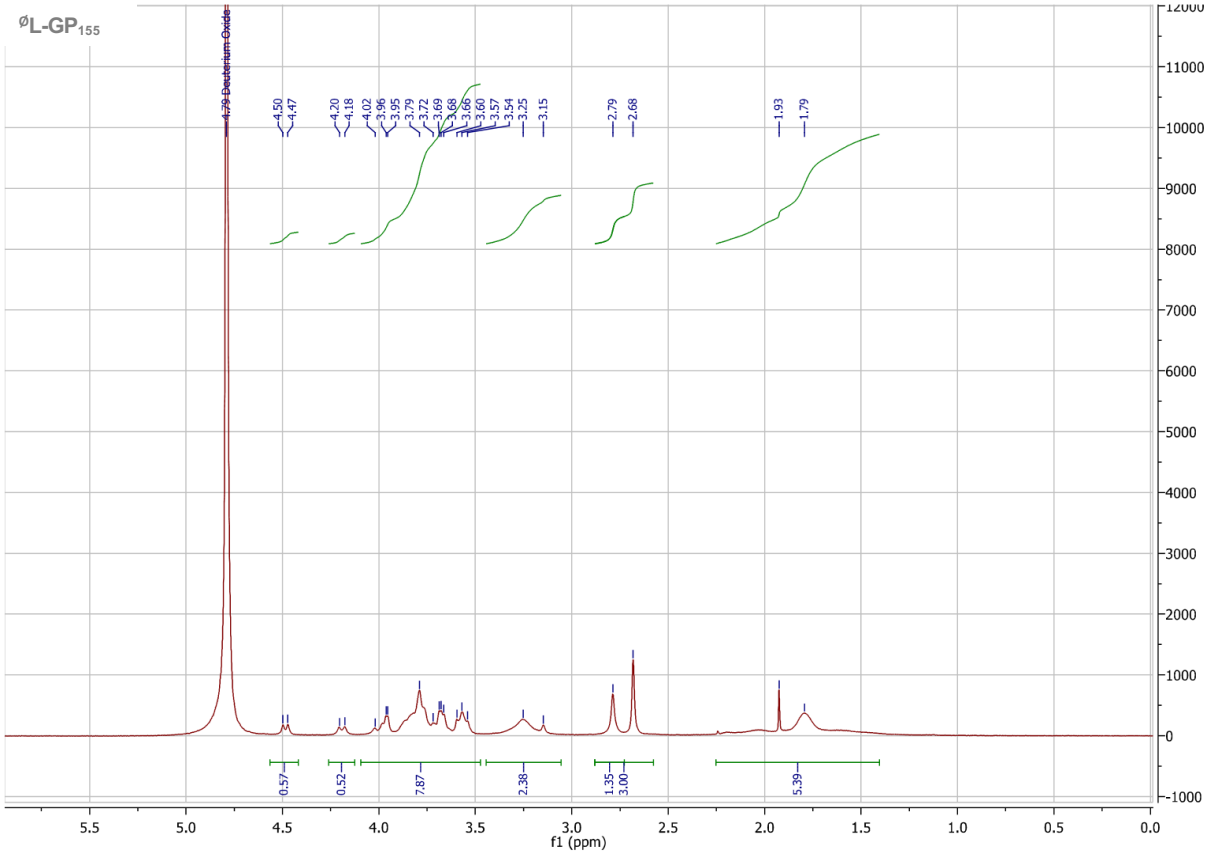
^{1}H NMR of ^{Ø}L-GP$_{85}$

D_2O, 300 MHz δ (ppm): 4.55-4.45 (bd, 0.19H), 4.25-4.15 (bm, 0.18H), 4.10-3.95 (bm, 0.64H), 3.90-3.65 (bm, 3.38), 3.65-3.50 (bm, 0.67), 3.45-3.05 (bm, 2.60H), 2.85-2.55 (bm, 3H), 2.25-1.10 (bm, 6.56H).



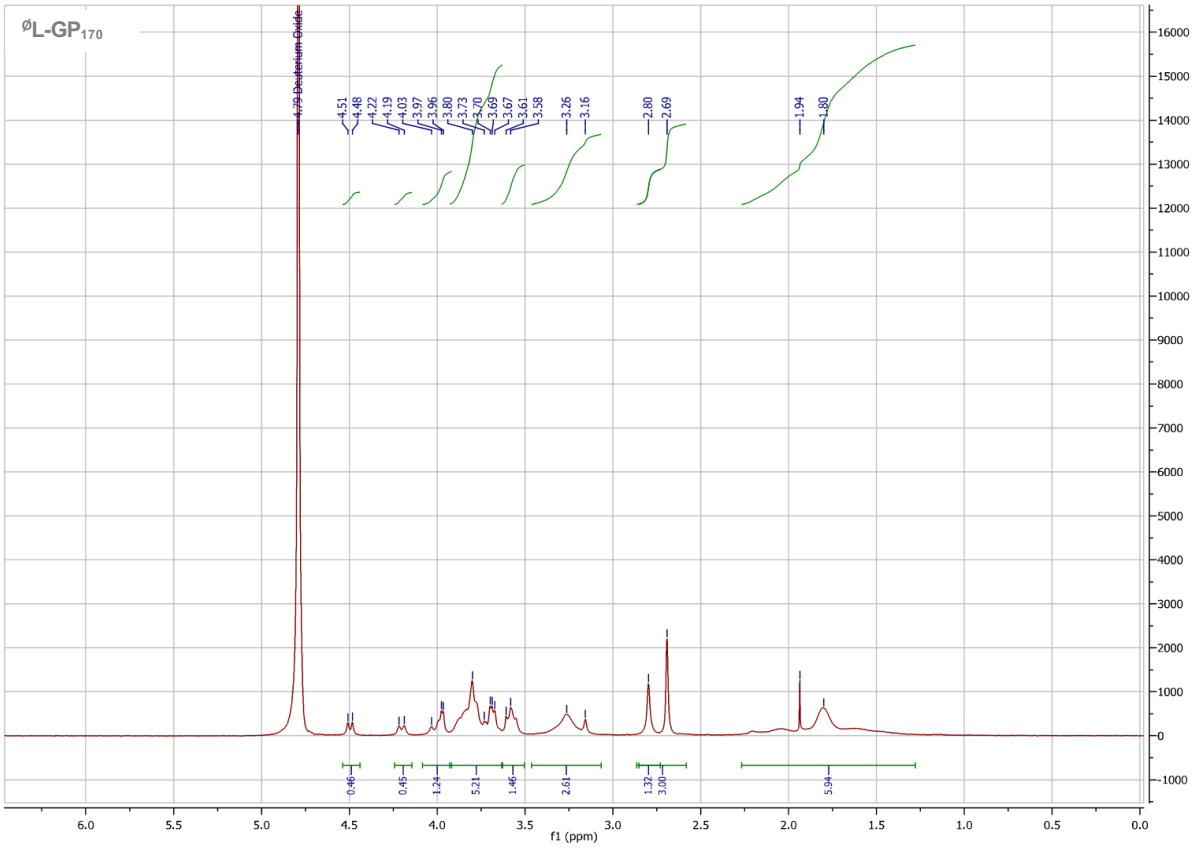
^{1}H NMR of ^{Ø}L-GP$_{155}$

D_2O, 300 MHz δ (ppm): 4.55-4.45 (bd, 0.57H), 4.25-4.15 (bm, 0.52H), 4.10-3.50 (bm, 7.87H), 3.45-3.05 (bm, 2.38H), 2.85-2.60 (bm, 3H), 2.25-1.40 (bm, 5.39H).



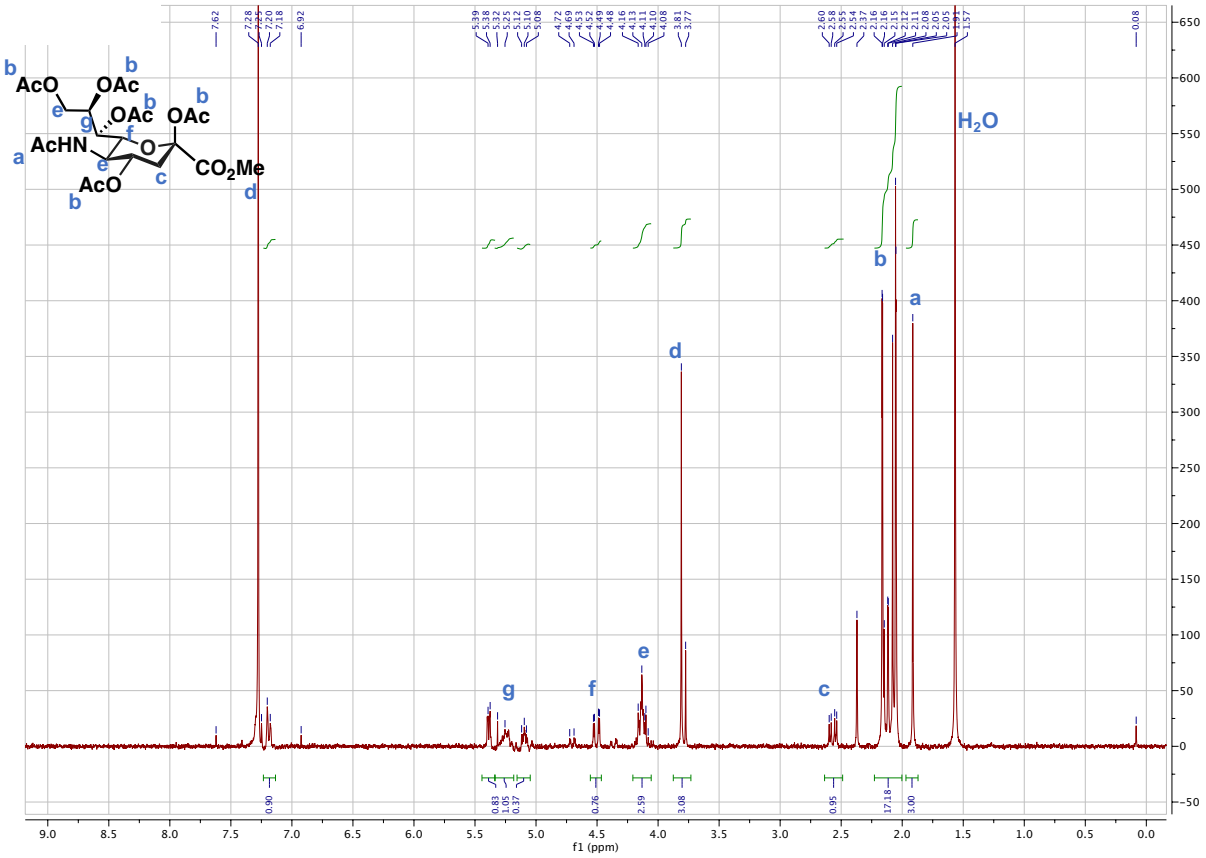
^{1}H NMR of ^{Ø}L-GP$_{170}$

D_2O, 300 MHz δ (ppm): 4.55-4.45 (bd, 0.46H), 4.25-4.15 (bm, 0.45H), 4.10-3.90 (bm, 1.24H), 3.90-3.60 (bm, 5.21H), 3.60-3.50 (bm, 1.46H), 3.05-3.45 (bm, 2.61H), 2.90-2.55 (bm 3H), 2.25-1.25 (bm, 5.94).

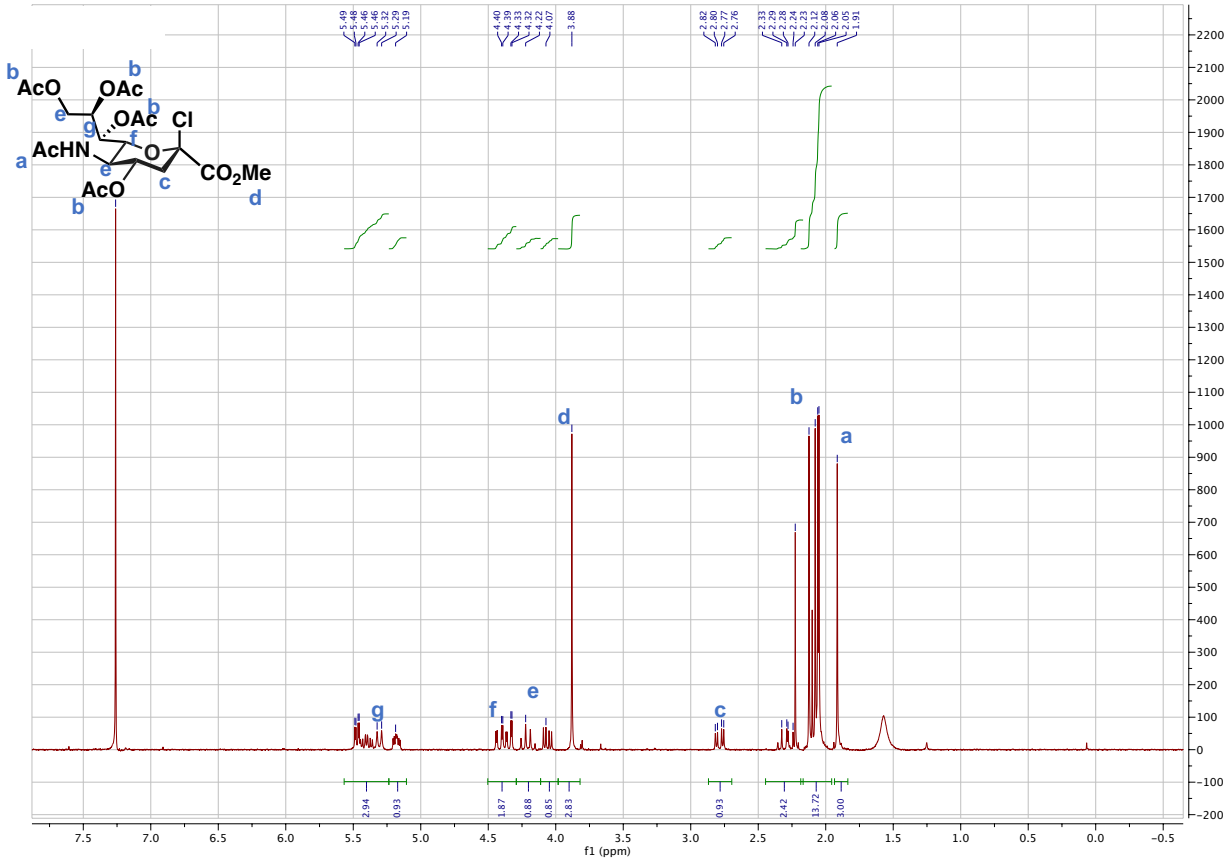


Appendix for Chapter 3: NMR and mass spectra

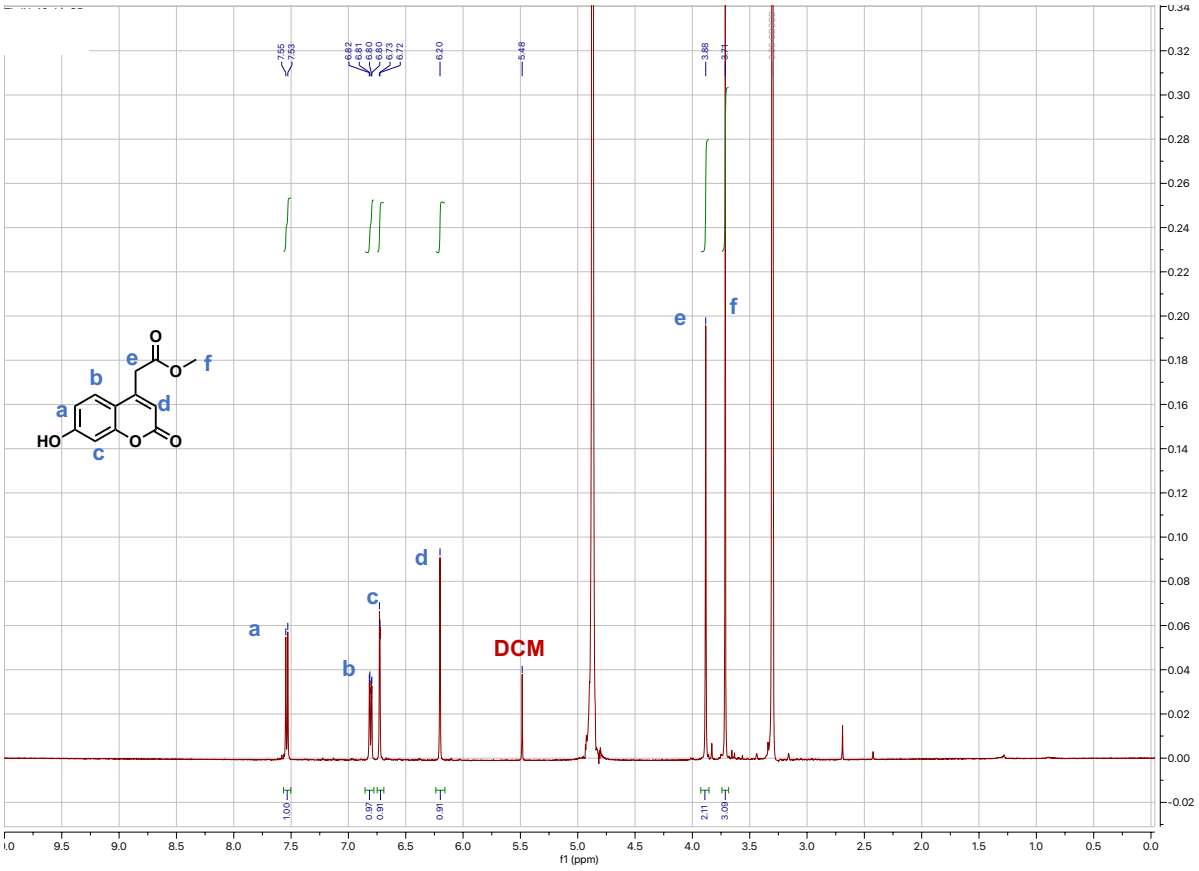
^{1}H NMR of peracetylated methyl ester of sialic acid in $CDCl_3$



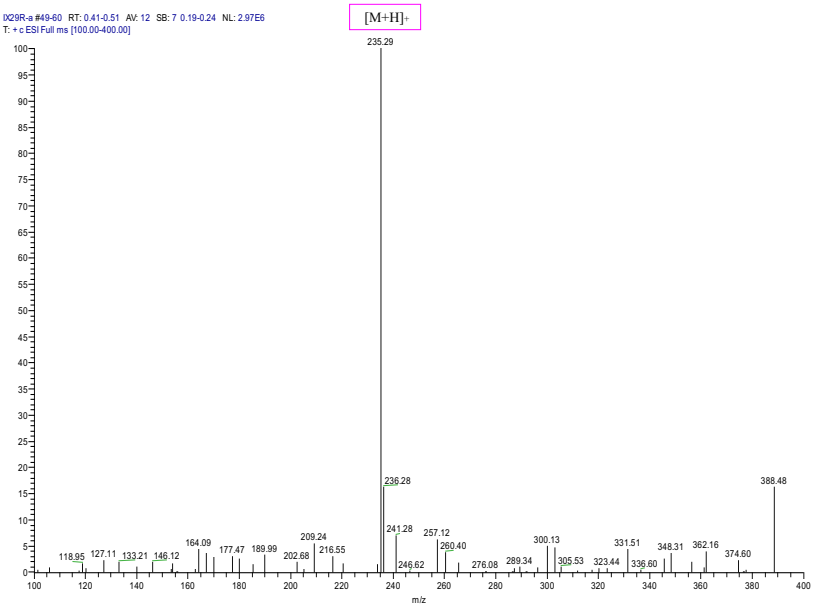
^{1}H NMR of peracetylated methyl ester chloride of sialic acid in $CDCl_3$



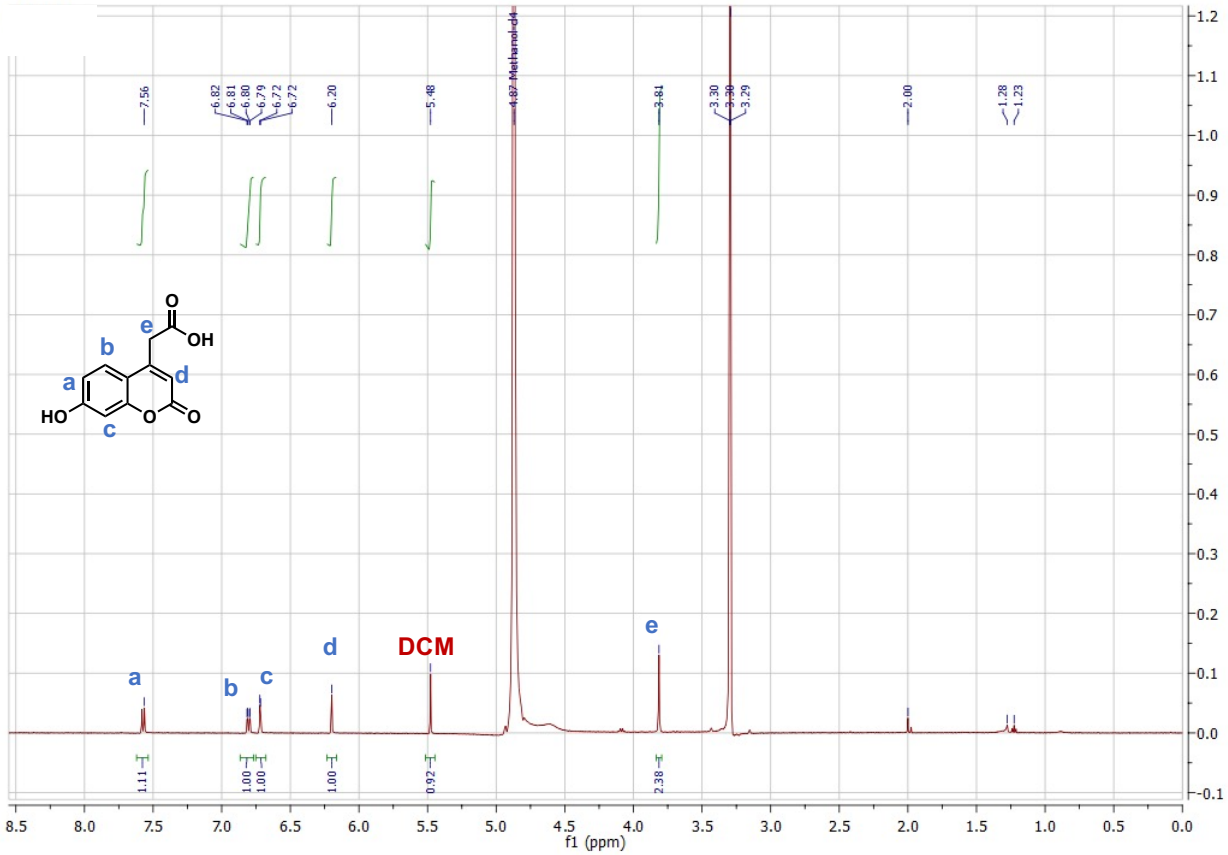
^{1}H NMR of coumarin methyl ester in MeOD



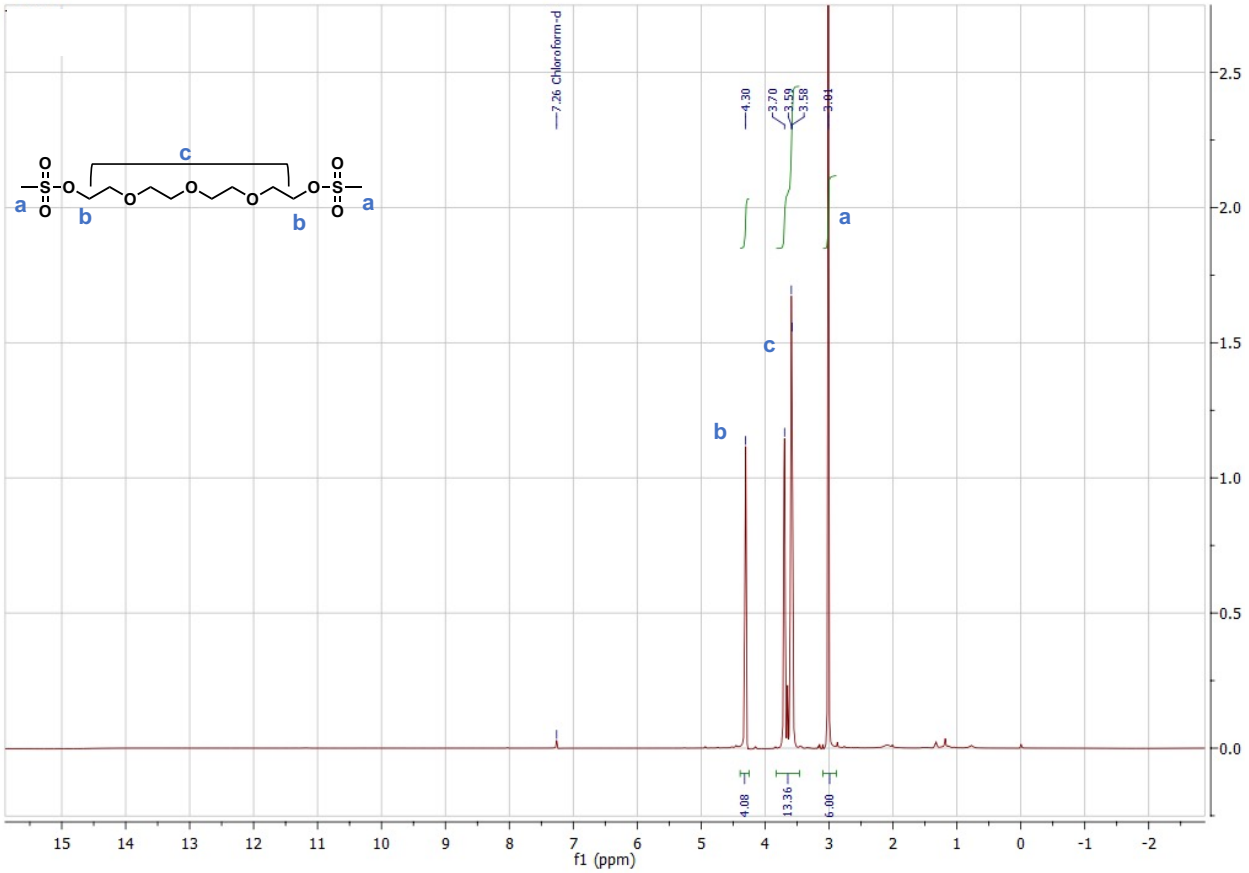
Mass spectra of coumarin methyl ester (pos ion mode)



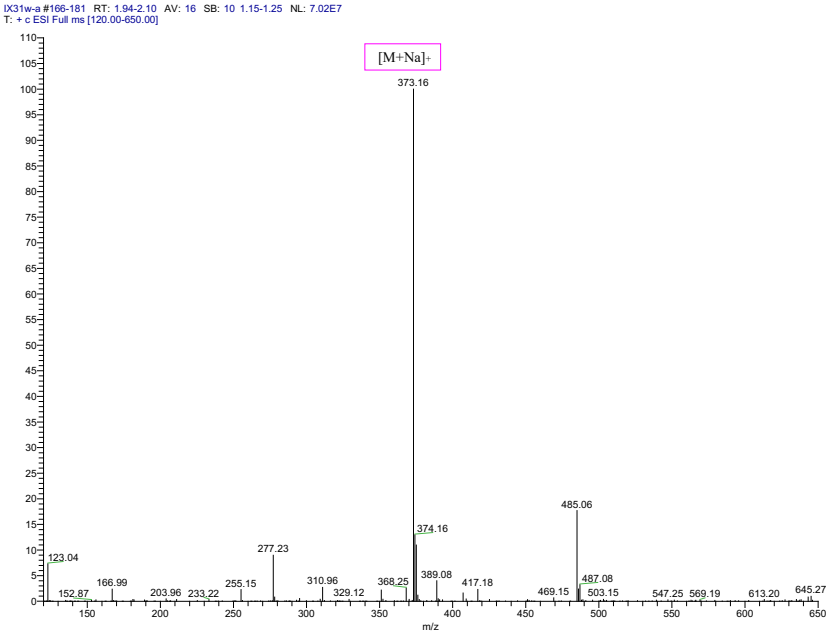
^{1}H NMR of hydrolyzed coumarin (free carboxylic acid) in MeOD



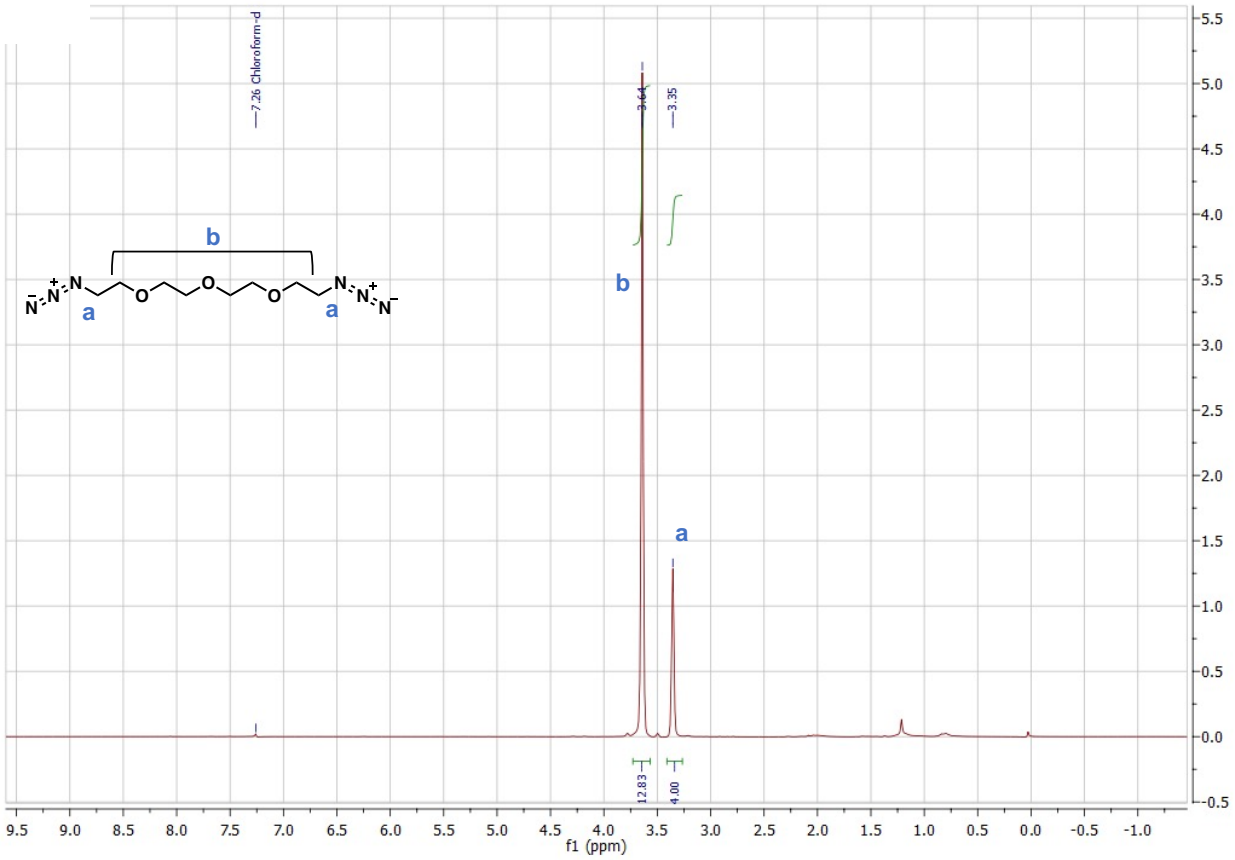
^{1}H NMR of mesylated tetraethylene glycol in $CDCl_3$



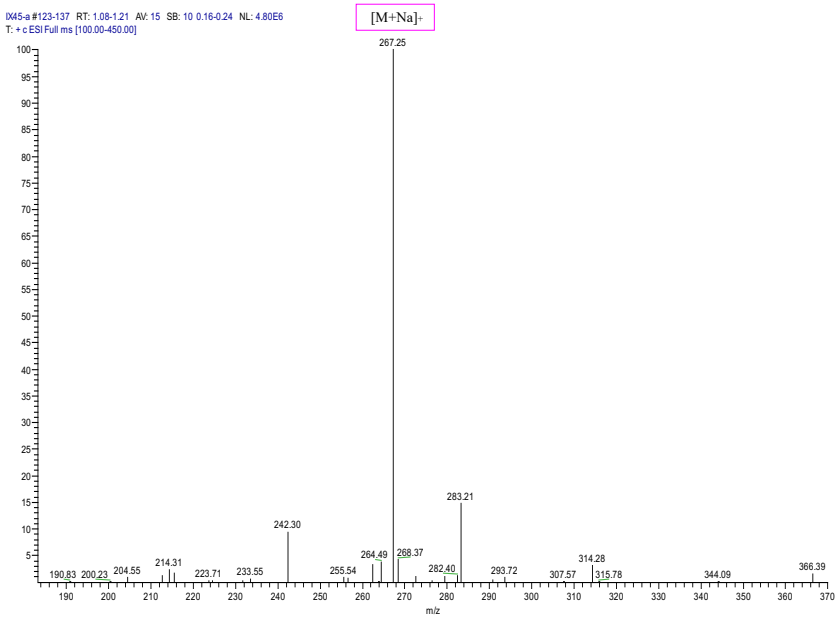
Mass spectra of mesylated tetraethylene glycol (pos ion mode)



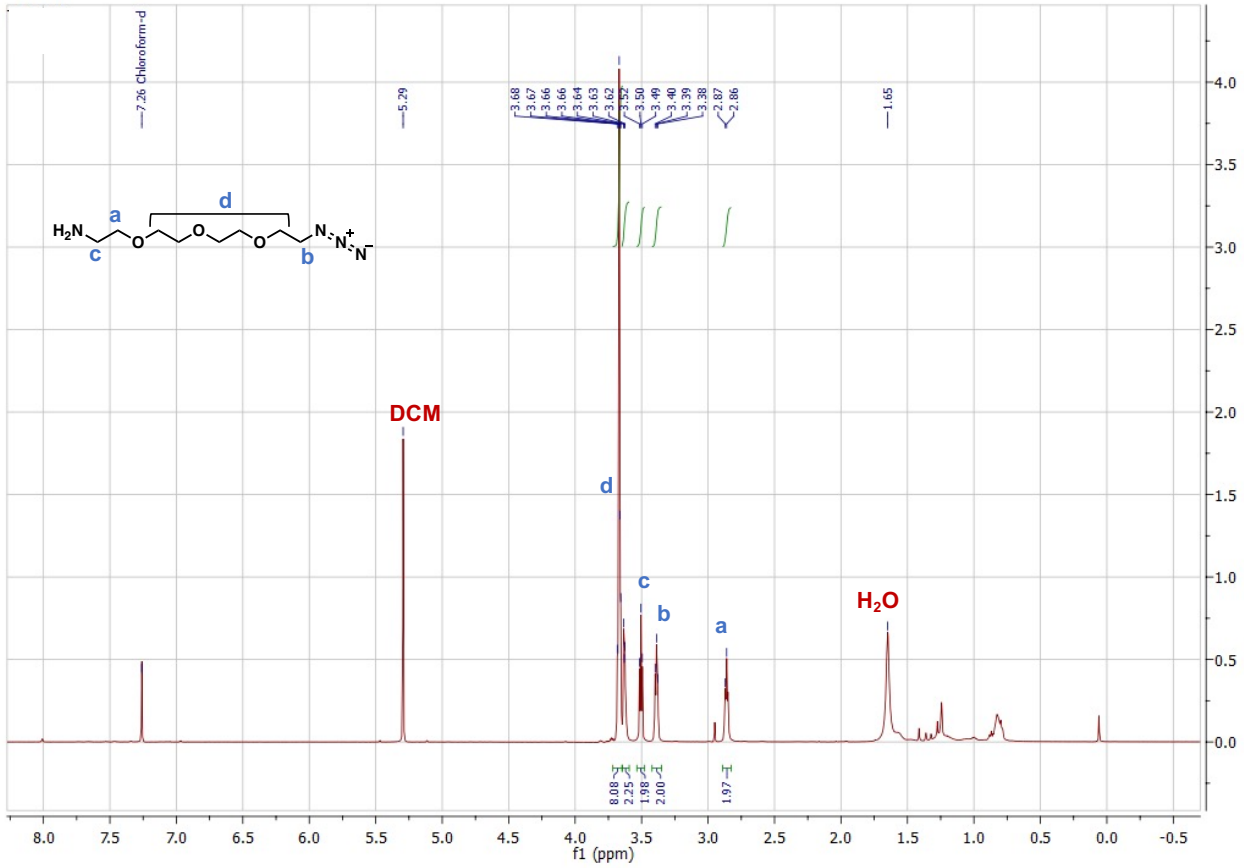
^{1}H NMR of di-azide version of tetraethylene glycol in $CDCl_3$



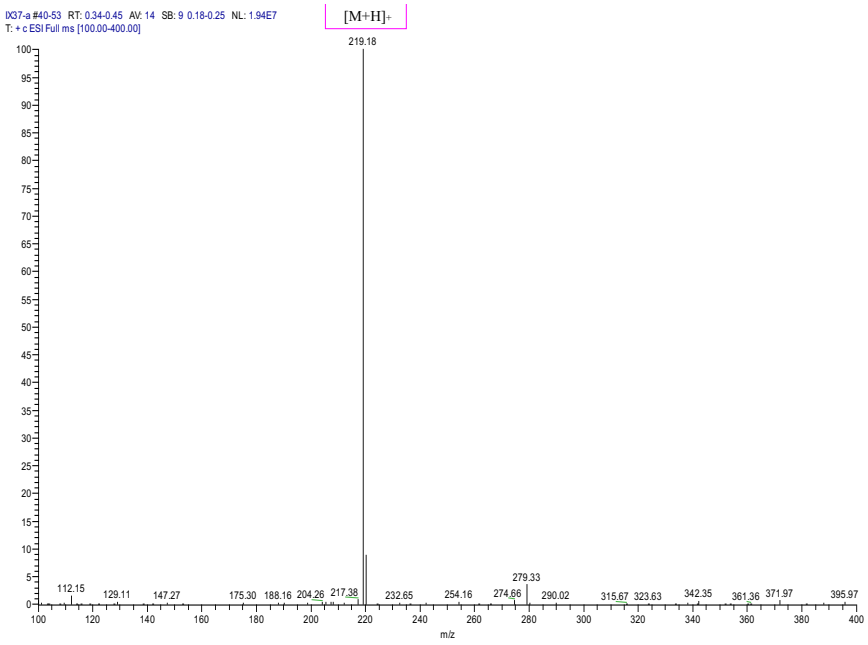
Mass spectra of di-azide version of tetraethylene glycol (pos ion mode)



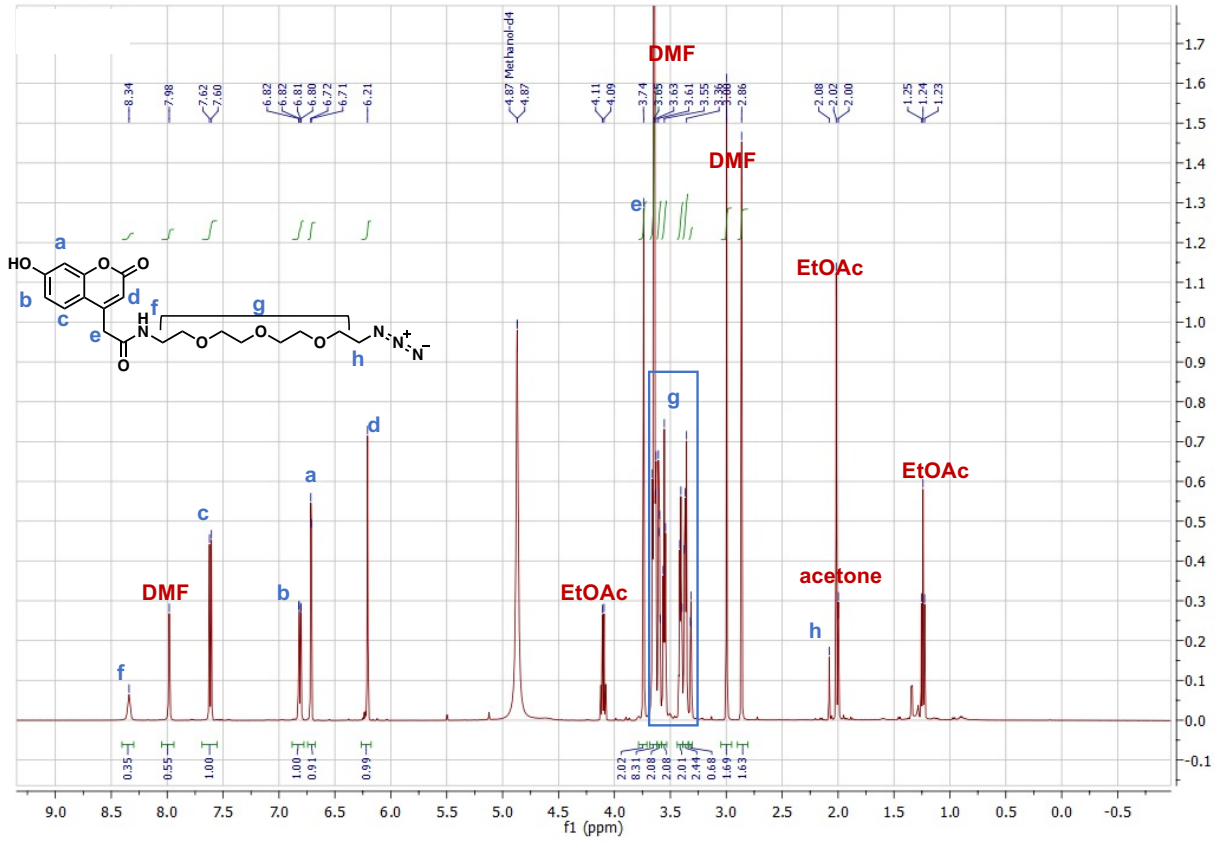
^{1}H NMR of azido-PEG-amine linker in $CDCl_3$



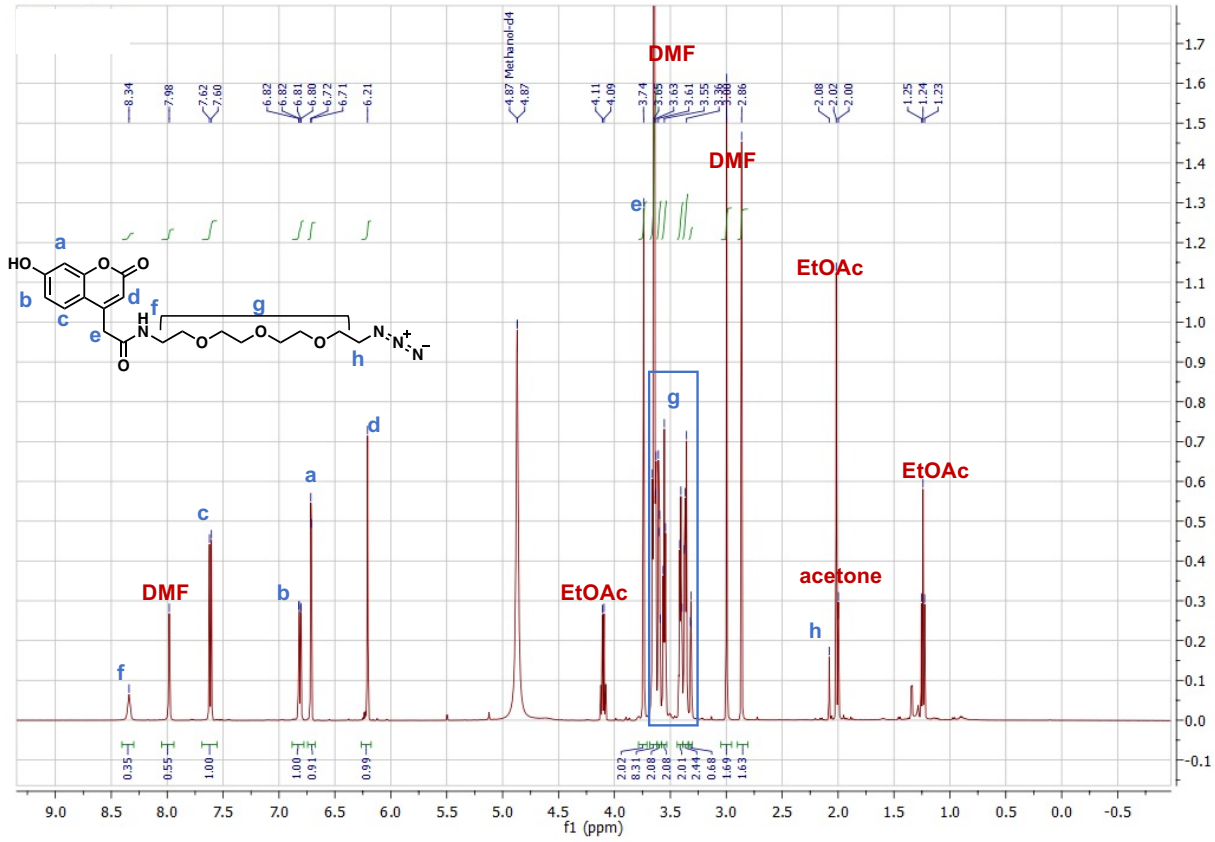
Mass spectra of azido-PEG-amine linker (pos ion mode)



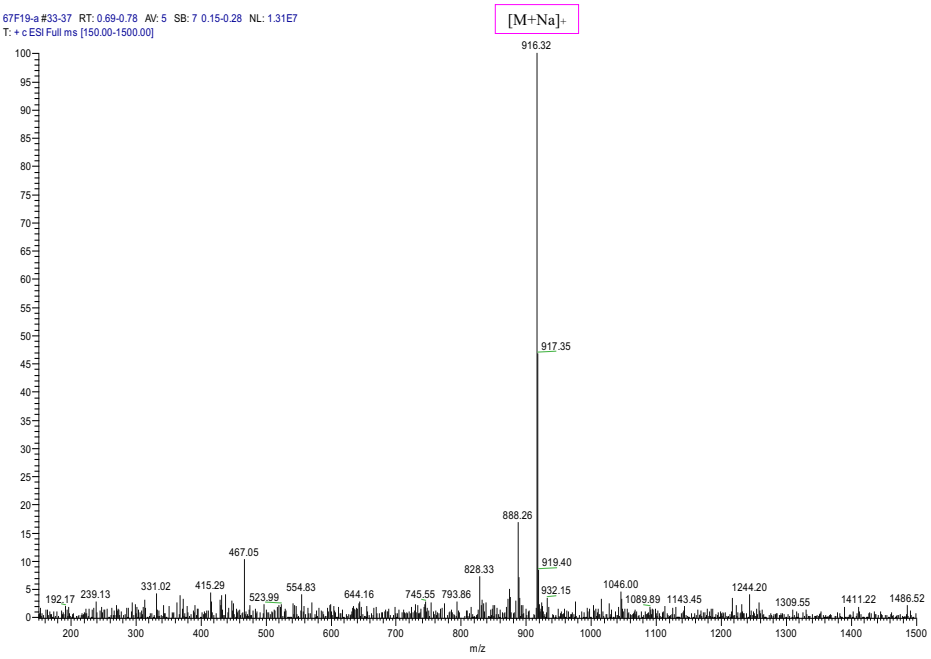
^{1}H NMR of nonglycosylated probe in MeOD



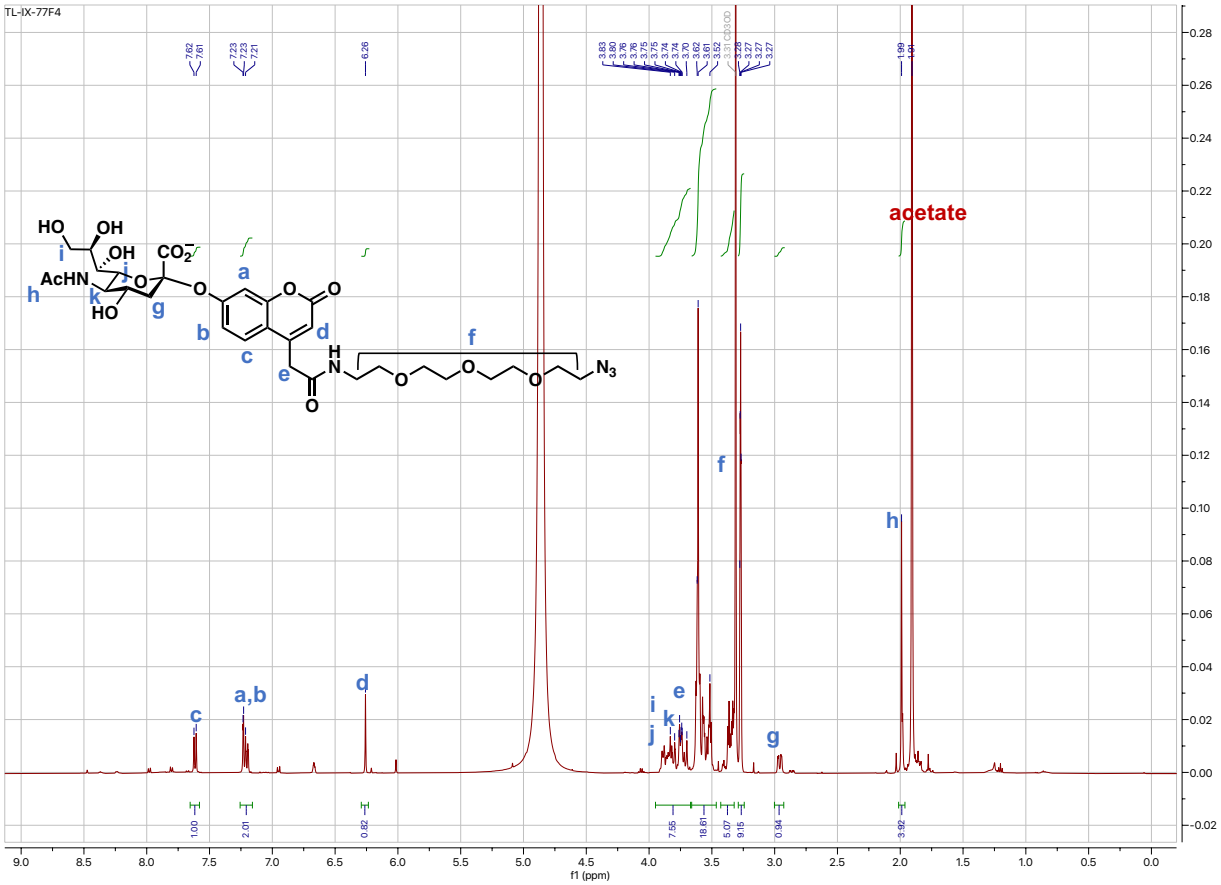
^{1}H NMR of protected glycosylated probe in MeOD



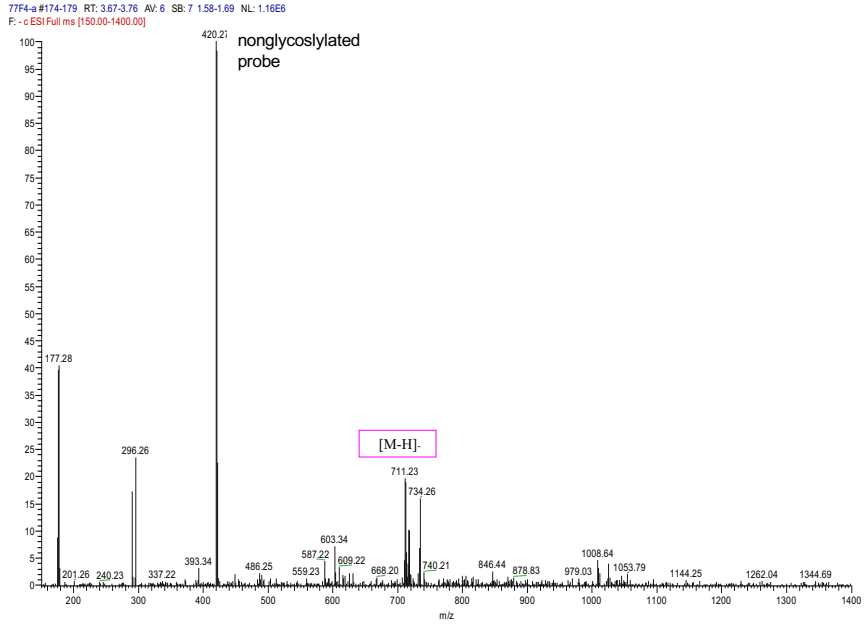
Mass spectra of protected glycosylated probe (pos ion mode)



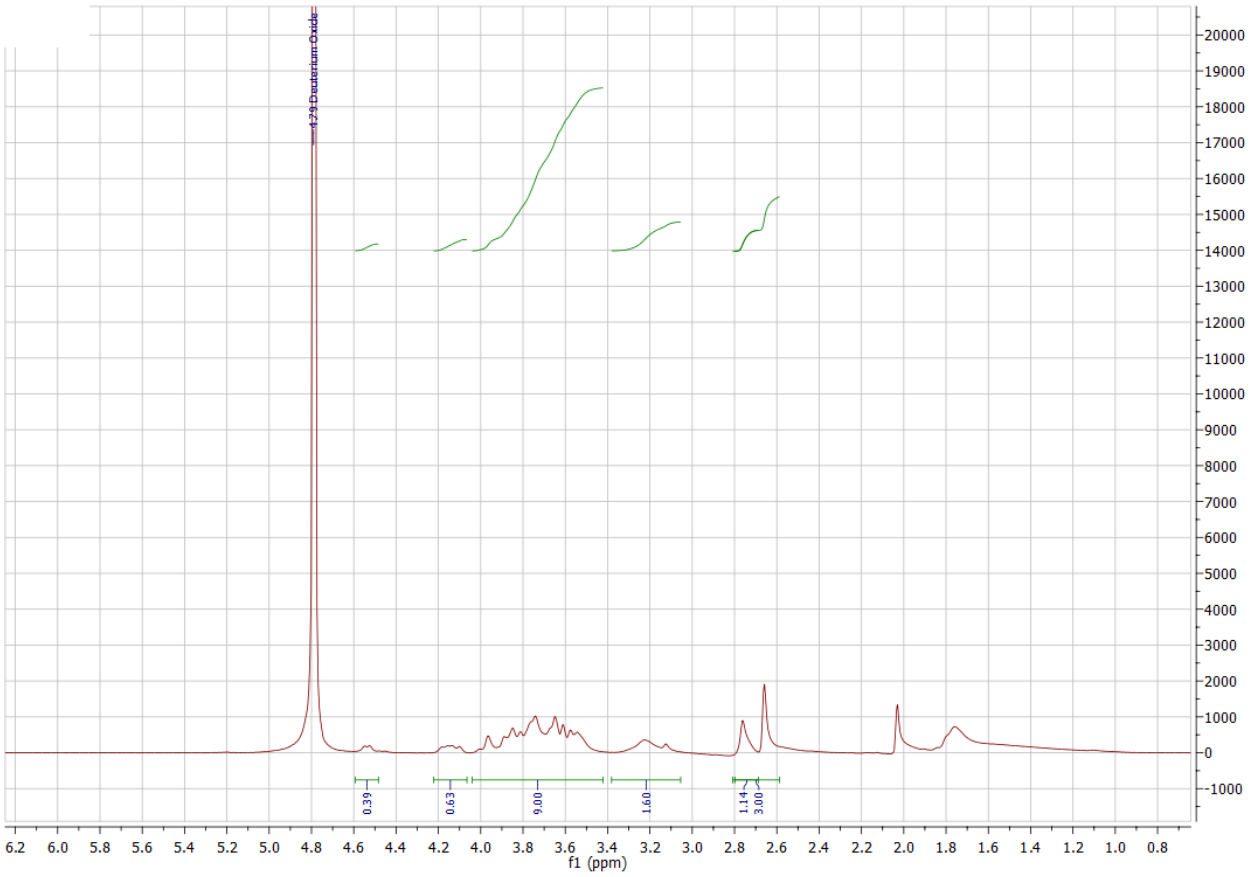
^{1}H NMR of glycosylated probe in MeOD



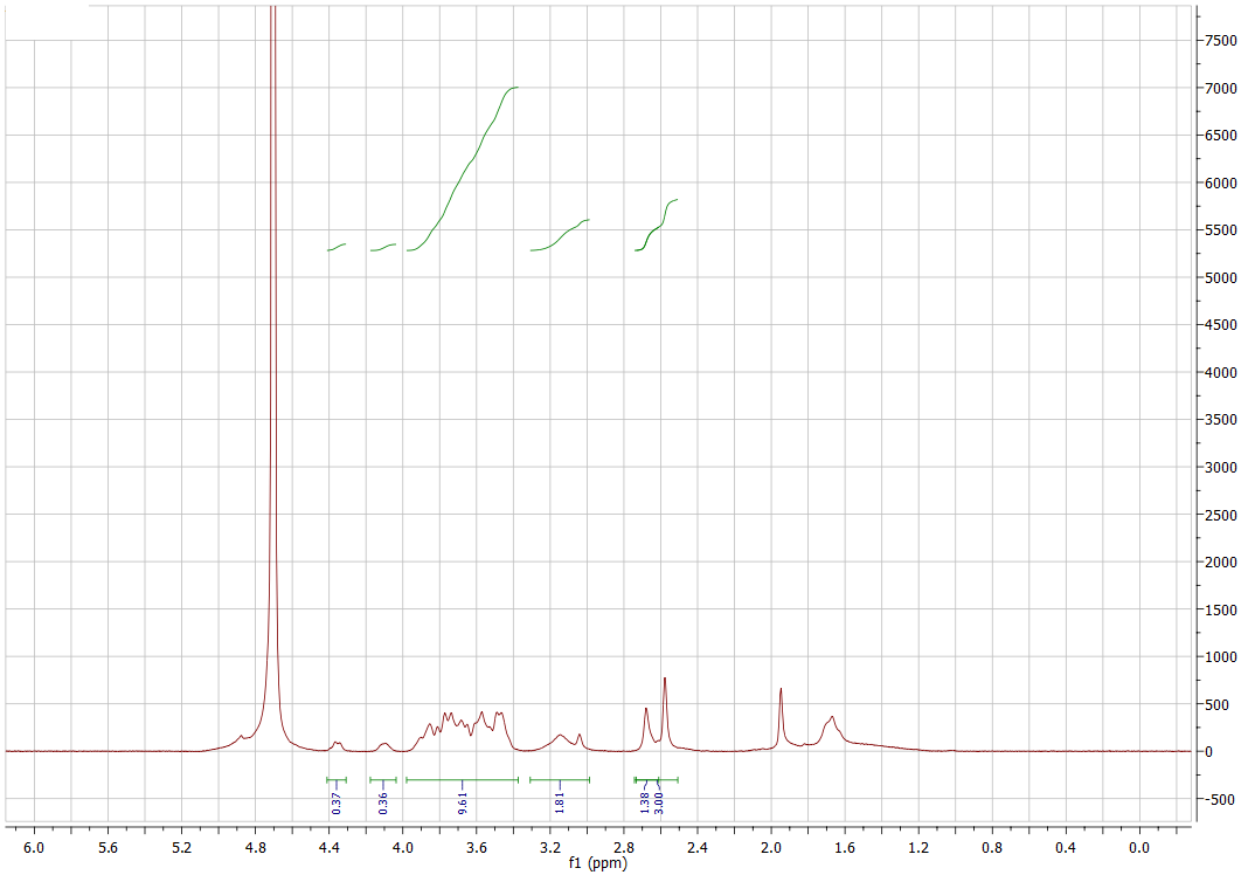
Mass spectra of glycosylated probe (in neg ion mode)



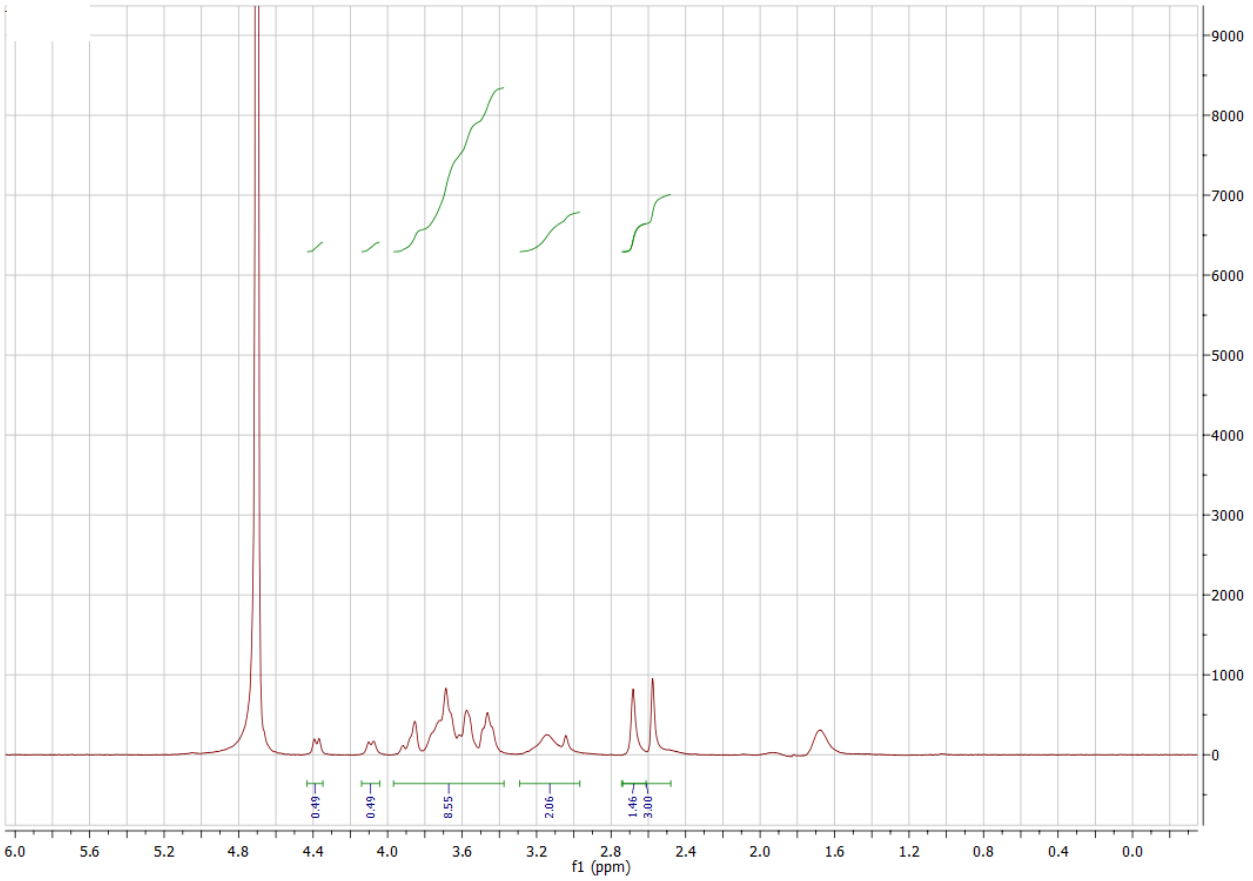
^{1}H NMR of biotin polymer containing α2,3 SiaLac



^{1}H NMR of biotin polymer containing α2,6 SiaLac



^{1}H NMR of biotin polymer containing Lac



References

1. Paget, J., Spreeuwenberg, P., Charu, V., Taylor, R. J., Iuliano, A. D., Bresee, J., Simonsen, L. & Viboud, C. Global mortality associated with seasonal influenza epidemics: New burden estimates and predictors from the GLaMOR Project. *Journal of Global Health* **9**, (2019).

2. van den Dool, C., Hak, E., Wallinga, J., van Loon, A. M., Lammers, J. W. J. & Bonten, M. J. M. Symptoms of Influenza Virus Infection in Hospitalized Patients. *Infection Control & Hospital Epidemiology* **29**, 314–319 (2008).

3. Putri, W. C. W. S., Muscatello, D. J., Stockwell, M. S. & Newall, A. T. Economic burden of seasonal influenza in the United States. *Vaccine* **36**, 3960–3966 (2018).

4. Beigel, J. H. Influenza. *Critical Care Medicine* **36** 2660–2666 (2008).

5. Kamali, A. & Holodniy, M. Influenza treatment and prophylaxis with neuraminidase inhibitors: A review. *Infection and Drug Resistance* **6** 187–198 (2013).

6. Bouvier, N. M. & Palese, P. The biology of influenza viruses. *Vaccine* vol 26 (2008).

7. Kosik, I. & Yewdell, J. W. Influenza hemagglutinin and neuraminidase: Yin–yang proteins coevolving to thwart immunity. *Viruses* **11** (2019).

8. Slepushkin, V. A., Staber, P. D., Wang, G., McCray, P. B. & Davidson, B. L. Infection of human airway epithelia with H1N1. H2N2, and H3N2 influenza A virus strains. *Molecular Therapy* **3**, 395–402 (2001).

9. Rogers, G. N., Paulson, J. C., Daniels, R. S., Skehel, J. J., Wilson, I. A., & Wiley, D.C. Single amino acid substitutions in influenza haemagglutinin change receptor binding specificity. *Nature* **304** 76-78 (1983).

10. Kim, H., Webster, R. G. & Webby, R. J. Influenza Virus: Dealing with a Drifting and Shifting Pathogen. *Viral Immunology* **31**, 174–183 (2018).

11. Stevens, J., Blixt, O., Chen, L. M., Donis, R. O., Paulson, J. C. & Wilson, I. A. Recent Avian H5N1 Viruses Exhibit Increased Propensity for Acquiring Human Receptor Specificity. *Journal of Molecular Biology* **381**, 1382–1394 (2008).

12. Shi, J., Deng, G., Ma, S., Zeng, X., Yin, X., Li, M., Zhang, B., Cui, P., Chen, Y., Yang, H., Wan, X., Liu, L., Chen, P., Jiang, Y., Guan, Y., Liu, J., Gu, W., Han, S., Song, Y., Liang, L., Qu, Z., Hou, Y., Wang, X., Bao, H., Tian, G., Li, Y., Jiang, L., Li, C., & Chen, H. Rapid evolution of H7N9 highly pathogenic viruses that emerged in China in 2017. *Cell Host and Microbe* **24**, 558-568.e7 (2018).

13. Puzelli, S., Rossini, G., Facchini, M., Vaccari, G., di Trani, L., di Martino, A., Gaibani, P., Vocale, C., Cattoli, G., Bennett, M., McCauley, J. W., Rezza, G., Moro, M. L., Rangoni, R., Finarelli, A. C., Landini, M. P., Castrucci, M. R. & Donatelli, I. Human infection with highly pathogenic a(H7N7) avian influenza virus, Italy, 2013. *Emerging Infectious Diseases* **20**, 1745–1749 (2014).

14. Sun, Y. & Liu, J. H9N2 influenza virus in China: a cause of concern. *Protein and Cell* **6**, 18–25 (2014).

15. Neelamegham, S., Aoki-Kinoshita, K., Bolton, E., Frank, M., Lisacek, F., Lütteke, T., O'Boyle, N., Packer, N. H., Stanley, P., Toukach, P., Varki, A. & Woods, R. J. Updates to the symbol nomenclature for glycans guidelines. *Glycobiology* **29**, 620–624 (2019).

16. Hirst, G. K. Absorption of influenza hemagglutinins and virus by red blood cells. *Journal of Experimental Medicine* **76** 195-209 (1942).

17. Davenport, F. M., Horsfall Jr., F. L. The associative reactions of pneumonia virus of mice (PVM) and influenza viruses: the effects of pH and electrolytes upon virus-host cell combinations. *Journal of Experimental Medicine* **88** 621-644 (1948).

18. Hilleman, M. R., Buescher, E. L. & Smadel, J. E. Preparation of dried antigen and antiserum for the agglutination-inhibition test for virus influenza. *Public Health Reports* **66** 1195-1203 (1951).

19. Tamm, I. & Horsfall Jr., F. L. Characterization and separation of an inhibitor of viral hemagglutination present in urine. *Experimental Biology and Medicine* **74** 108-114 (1950).

20. Green, R. H. & Woolley, D. W. Inhibition by certain polysaccharides of hemagglutination and of multiplication of influenza virus. *Journal of Experimental Medicine* **86** 55-64 (1947).

21. Stulberg, C. S., Schapira, R., Robinson, B. R., Basinski, D. H. & Freund, H. A. Inhibition of influenza virus hemagglutination by purified plasma mucoproteins *Experimental Biology and Medicine* **76** 704-706 (1951).

22. Gottschalk, A. & Lind, P. E. Product of Interaction between Influenza Virus Enzyme and Ovomucin. *Nature* **164**, (1949).

23. Klenk, V. E., Faillard, H. & Lempfrid, H. Über die enzymatische Wirkung von Influeiizavirus. Croonian Lecture*, Proc. Roy. Soc. [London], Ser. B* **301** (1955).

24. Rogers, G. N. & Paulson, J. C. 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 (1983).

25. Broszeit, F., van Beek, R. J., Unione, L., Bestebroer, T. M., Chapla, D., Yang, J. Y., Moremen, K. W., Herfst, S., Fouchier, R. A. M., de Vries, R. P. & Boons, G. J. Glycan remodeled erythrocytes facilitate antigenic characterization of recent A/H3N2 influenza viruses. *Nature Communications* **12**, (2021).

26. 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. & Paulson, J. C. Recent H3N2 viruses have evolved specificity for extended, branched human-type receptors, conferring potential for increased avidity. *Cell Host and Microbe* **21**, 23–34 (2017).

27. Skehel, J. J. & Wiley, D. C. Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. *Annual Review of Biochemistry* **63** 531-569 (2000).

28. Mammen, M., Dahmann, G. & Whitesides, G. M. Effective inhibitors of hemagglutination by influenza virus synthesized from polymers having active ester groups. insight into mechanism of inhibition. *Journal of Medicinal Chemistry* **38**, 4179–4190 (1995).

29. Sigal, G. B., Mammen, M., Dahmann, G. & Whitesides, G. M. Polyacrylamides bearing pendant r-sialoside groups strongly inhibit agglutination of erythrocytes by influenza virus: the strong inhibition reflects enhanced binding through cooperative polyvalent interactions. *Journal of American Chemical Society* **118** 3789-3800 (1996).

30. Kwon, S. J., Na, D. H., Kwak, J. H., Douaisi, M., Zhang, F., Park, E. J., Park, J. H., Youn, H., Song, C. S., Kane, R. S., Dordick, J. S., Lee, K. B. & Linhardt, R. J. Nanostructured glycan architecture is important in the inhibition of influenza A virus infection. *Nature Nanotechnology* **12**, 48–54 (2017).

31. Bhatia, S., Lauster, D., Bardua, M., Ludwig, K., Angioletti-Uberti, S., Popp, N., Hoffmann, U., Paulus, F., Budt, M., Stadtmüller, M., Wolff, T., Hamann, A., Böttcher, C., Herrmann, A. & Haag, R. Linear polysialoside outperforms dendritic analogs for inhibition of influenza virus infection in vitro and in vivo. *Biomaterials* **138**, 22–34 (2017).

32. Gresham, D., Dunham, M. J. & Botstein, D. Comparing whole genomes using DNA microarrays. *Nature Reviews Genetics* **9,** 291–302 (2008).

33. Blixt, O., Head, S., Mondala, T., Scanlan, C., Huflejt, M. E., Alvarez, R., Bryan, M. C., Fazio, F., Calarese, D., Stevens, J., Razi, N., Stevens, D. J., Skehel, J. J., van Die, I., Burton, D. R., Wilson, I. a, Cummings, R., Bovin, N., Wong, C.-H., & Paulson, J. C. Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 17033–17038 (2004).

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

35. Palma, A. S., Feizi, T., Childs, R. A., Chai, W. & Liu, Y. The neoglycolipid (NGL)-based oligosaccharide microarray system poised to decipher the meta-glycome. *Current Opinion in Chemical Biology* **18,** 87–94 (2014).

36. 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.-Jen. R., Wu, C.-Y. & Wong, C.-H. Differential receptor binding affinities of influenza hemagglutinins on glycans arrays. *J. Am. Chem. Soc.* **132**, 14849–14856 (2010).

37. Song, X., Yu, H., Chen, X., Lasanajak, Y., Tappert, M. M., Air, G. M., Tiwari, V. K., Cao, H., Chokhawala, H. A., Zheng, H., Cummings, R. D. & Smith, D. F. A sialylated glycan microarray reveals novel interactions of modified sialic acids with proteins and viruses. *Journal of Biological Chemistry* **286**, 31610–31622 (2011).

38. Song, X., Heimburg-Molinaro, J., Cummings, R. D. & Smith, D. F. Chemistry of natural glycan microarrays. *Current Opinion in Chemical Biology* **18**, 70–77 (2014).

39. Alvarez, R. A. & Blixt, O. Identification of Ligand Specificities for Glycan-Binding Proteins Using Glycan Arrays. *Methods in Enzymology* vol. 415 292–310 (2006).

40. Stevens, J., Blixt, O., Glaser, L., Taubenberger, J. K., Palese, P., Paulson, J. C. & Wilson, I. A. Glycan microarray analysis of the hemagglutinins from modern and pandemic influenza viruses reveals different receptor specificities. *Journal of Molecular Biology* **355**, 1143–1155 (2006).

41. Bradley, K. C., Jones, C. A., Tompkins, S. M., Tripp, R. A., Russell, R. J., Gramer, M. R., Heimburg-Molinaro, J., Smith, D. F., Cummings, R. D. & Steinhauer, D. A. Comparison of the receptor binding properties of contemporary swine isolates and early human pandemic H1N1 isolates (Novel 2009 H1N1). *Virology* **413**, 169–182 (2011).

42. Stevens, J., Blixt, O., Tumpey, T. M., Taubenberger, J. K., Paulson, J. C. & Wilson, I. A. Structure and Receptor Specificity of the Hemagglutinin from an H5N1 Influenza Virus. *Science* **312**, 404-410 (2006).

43. Stevens, J., Blixt, O., Chen, L.-M., Donis, R. O., Paulson, J. C. & Wilson, I. A. Recent avian h5n1 viruses exhibit increased propensity for acquiring human receptor specificity. *Journal of Molecular Biology* **381**, 1382-1394 (2008).

44. Gulati, S., Smith, D. F., Cummings, R. D., Couch, R. B., Griesemer, S. B., st. George, K., Webster, R. G. & Air, G. M. 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* **8**, (2013).

45. Walther, T., Karamanska, R., Chan, R. W. Y., Chan, M. C. W., Jia, N., Air, G., Hopton, C., Wong, M. P., Dell, A., Malik Peiris, J. S., Haslam, S. M. & Nicholls, J. M. Glycomic analysis of human respiratory tract tissues and correlation with influenza virus infection. *PLoS Pathogens* **9**, (2013).

46. Jia, N., Byrd-Leotis, L., Matsumoto, Y., Gao, C., Wein, A. N., Lobby, J. L., Kohlmeier, J. E., Steinhauer, D. A. & Cummings, R. D. The human lung glycome reveals novel glycan ligands for influenza A virus. *Scientific Reports* **10**, (2020).

47. Byrd-Leotis, L., Jia, N., Dutta, S., Trost, J. F., Gao, C., Cummings, S. F., Braulke, T., Müller-Loennies, S., Heimburg-Molinaro, J., Steinhauer, D. A. & Cummings, R. D. Influenza binds phosphorylated glycans from human lung. *Sci. Adv* **5** (2019).

48. Narimatsu, Y., Joshi, H. J., Nason, R., van Coillie, J., Karlsson, R., Sun, L., Ye, Z., Chen, Y. H., Schjoldager, K. T., Steentoft, C., Furukawa, S., Bensing, B. A., Sullam, P. M., Thompson, A. J., Paulson, J. C., Büll, C., Adema, G. J., Mandel, U., Hansen, L., Bennet, E. P., Varki, A., Vakhrushev, S. Y., Yang, Z., & Clausen, H. An atlas of human glycosylation pathways enables display of the human glycome by gene engineered cells. *Molecular Cell* **75**, 394-407.e5 (2019).

49. Nason, R., Büll, C., Konstantinidi, A., Sun, L., Ye, Z., Halim, A., Du, W., Sørensen, D. M., Durbesson, F., Furukawa, S., Mandel, U., Joshi, H. J., Dworkin, L. A., Hansen, L., David, L., Iverson, T. M., Bensing, B. A., Sullam, P. M., Varki, A., de Vries, E., de Haan, C. A. M., Vincentelli, R., Henrissat, B., Vakhrushev, S. Y., Clausen, H., & Narimatsu, Y. Display of the human mucinome with defined O-glycans by gene engineered cells. *Nature Communications* **12**, (2021).

50. Hochmuth, R. M., Evans, C. A., Wiles, H. C. & McCown, J. T. Mechanical measurement of red cell membrane thickness. *Science* **220**, 101-102 (1983).

51. Uchimido, R., Schmidt, E. P. & Shapiro, N. I. The glycocalyx: A novel diagnostic and therapeutic target in sepsis. *Critical Care* **23** 292-2296 (2019).

52. Keni, R., Alexander, A., Nayak, P. G., Mudgal, J. & Nandakumar, K. COVID-19: Emergence, spread, possible treatments, and global burden. *Frontiers in Public Health* **8**, (2020).

53. Long, J. S., Mistry, B., Haslam, S. M. & Barclay, W. S. Host and viral determinants of influenza A virus species specificity. *Nature Reviews Microbiology* **17**, 67–81 (2019).

54. Lipsitch, M., Barclay, W., Raman, R., Russell, C. J., Belser, J. A., Cobey, S., Kasson, P. M., Lloyd-smith, J. O., Maurer-stroh, S., Riley, S., Beauchemin, C. A.,

Bedford, T., Friedrich, T. C., Handel, A., Herfst, S., Murcia, P. R., Roche, B., Wilke, C. O. & Russell, C. A. Viral factors in influenza pandemic risk assessment. *eLife* **5** (2006).

55. Suzuki, Y. Sialobiology of influenza- molecular mechanism of host range variation of influenza viruses. *Biological and Phrnological Bulletin* **28,** 399-408 (2005).

56. Raman, R., Tharakaraman, K., Shriver, Z., Jayaraman, A., Sasisekharan, V. & Sasisekharan, R. Glycan receptor specificity as a useful tool for characterization and surveillance of influenza A virus. *Trends in Microbiology* **22**, 632–641 (2014).

57. Air, G. M. Influenza virus-glycan interactions. *Current Opinion in Virology* **7,** 128–133 (2014).

58. McAuley, J. L., Gilbertson, B. P., Trifkovic, S., Brown, L. E. & McKimm-Breschkin, J. L. Influenza virus neuraminidase structure and functions. *Frontiers in Microbiology* 1**0,** (2019).

59. Gaymard, A., le Briand, N., Frobert, E., Lina, B. & Escuret, V. Functional balance between neuraminidase and haemagglutinin in influenza viruses. *Clinical Microbiology and Infection* **22,** 975–983 (2016).

60. Nelson, J., Couceiro, S. S., Paulson, J. C. & Baum, L. G. Influenza virus strains selectively recognize sialyloligosaccharides on human respiratory epithelium; the role of the host cell in selection of hemagglutinin receptor specificity. *Virus Research* **29**, 156-165 (1993).

61. Chu, V. C. & Whittaker, G. R. Influenza virus entry and infection require host cell N-linked glycoprotein. *Proceedings of the National Academy of Sciences* **101,** 18153-18158 (2004).

62. 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. & Wong, C. H. Differential receptor binding affinities of influenza hemagglutinins on glycan arrays. *Journal of the American Chemical Society* **132**, 14849–14856 (2010).

63. Möckl, L. The emerging role of the mammalian glycocalyx in functional membrane organization and immune system regulation. *Frontiers in Cell and Developmental Biology* **8**, (2020).

64. Linden, S. K., Sutton, P., Karlsson, N. G., Korolik, V. & McGuckin, M. A. Mucins in the mucosal barrier to infection. *Mucosal Immunology* **1**, 183–197 (2008).

65. Hang, H. C. & Bertozzi, C. R. The chemistry and biology of mucin-type O-linked glycosylation. *Bioorganic and Medicinal Chemistry* **13,** 5021–5034 (2005).

66. Mayr, J., Lau, K., Lai, J. C. C., Gagarinov, I. A., Shi, Y., McAtamney, S., Chan, R. W. Y., Nicholls, J., von Itzstein, M. & Haselhorst, T. Unravelling the role of o-glycans in influenza A virus infection. *Scientific Reports* **8**, (2018).

67. Delaveris, C. S., Webster, E. R., Banik, S. M., Boxer, S. G. & Bertozzi, C. R. Membrane-tethered mucin-like polypeptides sterically inhibit binding and slow fusion kinetics of influenza A virus. *Proceedings of the National Academy of Sciences* **117**, 12643–12650 (2020).

68. Button, B., Cai, L.-H., Ehre, C., Kesimer, M., Hill, D. B., Sheehan, J. K., Boucher, R. C. & Rubinstein, Michael. A periciliary brush promotes the lung health by separating the mucus layer from airway epithelia. *Science* **337**, 937-941 (2012).

69. Byrd-Leotis, L., Liu, R., Bradley, K. C., Lasanajak, Y., Cummings, S. F., Song, X., Heimburg-Molinaro, J., Galloway, S. E., Culhane, M. R., Smith, D. F., Steinhauer, D. A. & Cummings, R. D. Shotgun glycomics of pig lung identifies natural endogenous receptors for influenza viruses. *Proceedings of the National Academy of Sciences of the United States of America* **111**, E2241–E2250 (2014).

70. McAuley, J. L., Corcilius, L., Tan, H. X., Payne, R. J., McGuckin, M. A. & Brown, L. E. The cell surface mucin MUC1 limits the severity of influenza A virus infection. *Mucosal Immunology* **10**, 1581–1593 (2017).

71. J. Agard, N., A. Prescher, J. & R. Bertozzi, C. A Strain-Promoted [3 + 2] Azide−Alkyne Cycloaddition for Covalent Modification of Biomolecules in Living Systems. *Journal of the American Chemical Society* **126**, 15046–15047 (2004).

72. Kaiser, E., Picart, F., Kubiak, T., Tam, J. P. & Merrifield, R. B. Selective deprotection of the N.alpha.-tert-butyloxycarbonyl group in solid phase peptide synthesis with chlorotrimethylsilane in phenol. *The Journal of Organic Chemistry* **58**, 5167–5175 (1993).

73. Nagao, M., Fujiwara, Y., Matsubara, T., Hoshino, Y., Sato, T. & Miura, Y. Design of glycopolymers carrying sialyl oligosaccharides for controlling the interaction with the influenza virus. *Biomacromolecules* **18**, 4385–4392 (2017).

74. Armstrong, G. D., Howard, L. A. & Peppler, M. S. Use of glycosyltransferases to restore pertussis toxin receptor activity to asialoagalactofetuin. *The Journal of biological chemistry* **263**, 8677–8684 (1988).

75. Wright, C. S. & Jaeger, J. Crystallographic refinement and structure analysis of the complex of wheat germ agglutinin with a bivalent sialoglycopeptide from glycophorin A. *Journal of Molecular Biology* **232**, 620–638 (1993).

76. Shibuya, N., Goldstein, I. J., Broekaert, W. F., Nsimba-Lubaki, M., Peeters, B. & Peumans, W. J. The elderberry (Sambucus nigra L.) bark lectin recognizes the

Neu5Ac (alpha 2-6)Gal/GalNAc sequence. *The Journal of Biological Chemistry* **262**, 1596–1601 (1987).

77. Lees, W. J., Spaltenstein, A., Kingery-Wood, J. E. & Whitesides, G. M. Polyacrylamides Bearing Pendant α-Sialoside Groups strongly inhibit agglutination of erythrocytes by influenza A virus: multivalency and steric stabilization of particulate biological systems. *Journal of Medicinal Chemistry* **37**, 3419–3433 (1994).

78. Damme, E. J. M. van, Peumans, W. J., Barre, A. & Rougé, P. Plant Lectins: A composite of several distinct families of structurally and evolutionary related proteins with diverse biological roles. *Critical Reviews in Plant Sciences* **17**, 575–692 (1998).

79. Maveyraud, L., Niwa, H., Guillet, V., Svergun, D. I., Konarev, P. v., Palmer, R. A., Peumans, W. J., Rougé, P., van Damme, E. J. M., Reynolds, C. D. & Mourey, L. Structural basis for sugar recognition, including the Tn carcinoma antigen, by the lectin SNA-II from *Sambucus nigra*. *Proteins: Structure, Function, and Bioinformatics* **75**, 89–103 (2009).

80. Raman, R., Tharakaraman, K., Sasisekharan, V. & Sasisekharan, R. Glycan–protein interactions in viral pathogenesis. *Current Opinion in Structural Biology* **40**, 153–162 (2016).

81. Meng, B., Marriott, A. C. & Dimmock, N. J. The receptor preference of influenza viruses. *Influenza and other Respiratory Viruses* **4**, 147–153 (2010).

82. Koerner, I., Matrosovich, M. N., Haller, O., Staeheli, P. & Kochs, G. Altered receptor specificity and fusion activity of the haemagglutinin contribute to high virulence of a mouse-adapted influenza A virus. *Journal of General Virology* **93**, 970–979 (2012).

83. Nagao, M., Fujiwara, Y., Matsubara, T., Hoshino, Y., Sato, T. & Miura, Y. Design of glycopolymers carrying sialyl oligosaccharides for controlling the interaction with the influenza virus. *Biomacromolecules* **18**, 4385–4392 (2017).

84. di Iorio, D., Verheijden, M. L., van der Vries, E., Jonkheijm, P. & Huskens, J. Weak multivalent binding of influenza hemagglutinin nanoparticles at a sialoglycan-functionalized supported lipid bilayer. *ACS Nano* **13**, 3413–3423 (2019).

85. Zanin, M., Baviskar, P., Webster, R. & Webby, R. The interaction between respiratory pathogens and mucus. *Cell Host and Microbe* 1**9**, 159–168 (2016).

86. Uddin, S., Khan, A., Hossain, M. E. & Moni, M. A. Comparing different supervised machine learning algorithms for disease prediction. *BMC Medical Informatics and Decision Making* **19**, (2019).

87. Suzuki, Y., Nagao, Y., Kato, H., Suzuki, T., Matsumoto, M., Murayama, J.-I. & Suzuki, Y. The hemagglutinins of the human influenza viruses A and B recognize different receptor microdomains. *Biochimica et Biophysica Acta* **903**, 417-424 (1987).

88. Ito, T., Suzuki, Y., Takada, A., Kawamoto, A., Otsuki, K., Masuda, H., Yamada, M., Suzuki, T., Kida, H. & Kawaoka, Y. Differences in sialic acid-galactose linkages in the chicken egg amnion and allantois influence human influenza virus receptor specificity and variant selection. *Journal of virology* **71**, 3357–62 (1997).

89. Yao, Y., Liu, Y., Yu, Y., Xu, H., Lv, W., Li, Z. & Chen, X. K-SVM: An effective SVM algorithm based on K-means clustering. *Journal of Computers (Finland)* **8**, 2632–2639 (2013).

90. Hudson, M. J., Stamp, G. W., Chaudhary, K. S., Hewitt, R., Stubbs, A. P., Abel, P. D. & Lalani, E. N. Human MUC1 mucin: a potent glandular morphogen. *The Journal of pathology* **194**, 373–383 (2001).

91. Inkster, M. D., Hinshaw, V. S. & Schulze, I. T. The hemagglutinins of duck and human H1 influenza viruses differ in sequence conservation and in glycosylation. *Journal of virology* **67**, 7436–7443 (1993).

92. Li, J., Liu, S., Gao, Y., Tian, S., Yang, Y. & Ma, N. Comparison of N-linked glycosylation on hemagglutinins derived from chicken embryos and MDCK cells: a case of the production of a trivalent seasonal influenza vaccine. *Applied Microbiology and Biotechnology* **105,** 3559–3572 (2021).

93. Cohen, M., Fisher, C. J., Huang, M. L., Lindsay, L. A. L., Plancarte, M., Boyce, W. M., Godula, K. & Gagneux, P. Capture and characterization of influenza A virus from primary samples using glycan bead arrays. *Virology* **493**, 128–135 (2016).

94. Huang, M. L., Smith, R. A. A., Trieger, G. W. & Godula, K. Glycocalyx remodeling with proteoglycan mimetics promotes neural specification in embryonic stem cells. *Journal of the American Chemical Society* **136**, 10565–10568 (2014).

95. Pedregosa F, F., Michel, V., Grisel Oliviergrisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Vanderplas, J., Cournapeau, D., Pedregosa, F., Varoquaux, G., Gramfort, A., Thirion, B., Grisel, O., Dubourg, V., Passos, A., Brucher, M., Perrot., & Duchesnay, É. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research* **12**, 2825-2830 (2011).

96. Cohen, M., Senaati, H. P., Fisher, C. J., Huang, M. L., Gagneux, P. & Godula, K. Synthetic mucus nanobarriers for identification of glycan-dependent primary influenza a infection inhibitors. *ACS Central Science* **2**, 710–714 (2016).

97. Nagao, M., Fujiwara, Y., Matsubara, T., Hoshino, Y., Sato, T. & Miura, Y. Design of Glycopolymers Carrying Sialyl Oligosaccharides for Controlling the Interaction with the Influenza Virus. *Biomacromolecules* **18**, 4385–4392 (2017).

98. Bohdaneck$, M., Petrus, V. & Sedldtek, B. Estimation of the characteristic ratio of polyacrylamide in water and in a mixed theta-solvent. *Makromol. Chem* **184,** (1983).
99. Gamblin, S. J. & Skehel, J. J. Influenza hemagglutinin and neuraminidase membrane glycoproteins. *Journal of Biological Chemistry* **285**, 28403–28409 (2010).

100. Air, G. M. & Laver, W. G. The neuraminidase of influenza virus. *PROTEINS: Structure, Function, and Genetics* **6** (1989).

101. Cohen, M., Zhang, X. Q., Senaati, H. P., Chen, H. W., Varki, N. M., Schooley, R. T. & Gagneux, P. Influenza A penetrates host mucus by cleaving sialic acids with neuraminidase. *Virology Journal* **10**, (2013).

102. Hurt, A. C., Holien, J. K., Parker, M. W. & Barr, I. G. Oseltamivir resistance and the H274Y neuraminidase mutation in seasonal, pandemic and highly pathogenic influenza viruses. *Drugs* **69,** 2523–2531 (2009).

103. Lucas, T. M., Gupta, C., Altman, M. O., Sanchez, E., Naticchia, M. R., Gagneux, P., Singharoy, A. & Godula, K. Mucin-mimetic glycan arrays integrating machine learning for analyzing receptor pattern recognition by influenza A viruses. *Chem Cell Press* (2021).

104. Gao, Z., Thompson, A. J., Paulson, J. C. & Withers, S. G. Proximity Ligation-Based Fluorogenic Imaging Agents for Neuraminidases. *Angewandte Chemie* **130**, 13726–13729 (2018).

105. Eichelberger, M. C., Hassantoufighi, A., Wu, M. & Li, M. Neuraminidase activity provides a practical read-out for a high throughput influenza antiviral screening assay. *Virology Journal* **5**, (2008).

106. Lizzul-Jurse, A., Bailly, L., Hubert-Roux, M., Afonso, C., Renard, P. Y. & Sabot, C. Readily functionalizable phosphonium-tagged fluorescent coumarins for enhanced detection of conjugates by mass spectrometry. *Organic and Biomolecular Chemistry* **14**, 7777–7791 (2016).

107. Kimura, Y., Miyabara, Y., Terashima, T. & Sawamoto, M. Polyacrylamide pseudo crown ethers via hydrogen bond-assisted cyclopolymerization. *Journal of Polymer Science, Part A: Polymer Chemistry* **54**, 3294–3302 (2016).

108. Okoth, R. & Basu, A. End-labeled amino terminated monotelechelic glycopolymers generated by ROMP and Cu(I)-catalyzed azide-alkyne cycloaddition. *Beilstein Journal of Organic Chemistry* **9**, 608–612 (2013).

109. Spinnler, K., von Krüchten, L., Konieczny, A., Schindler, L., Bernhardt, G. & Keller, M. An Alkyne-functionalized Arginine for Solid-Phase Synthesis Enabling "bioorthogonal" Peptide Conjugation. *ACS Medicinal Chemistry Letters* **11**, 334–339 (2020).

110. Harvey, D. J. Analysis of carbohydrates and glycoconjugates by matrix-assisted laser desorption/ionization mass spectrometry: An update for 2011–2012. *Mass Spectrometry Reviews* **36**, 255–422 (2017).

111. Honigfort, D. J., Altman, M. O., Gagneux, P. & Godula, K. Glycocalyx crowding with mucin mimetics strengthens binding of soluble and virus-associated lectins to host cell glycan receptors. *Proceedings of the National Academy of Sciences* **118**, e2107896118 (2021).

112. Roux, K. J., Kim, D. I., Burke, B. & May, D. G. BioID: A Screen for Protein-Protein Interactions. *Current Protocols in Protein Science* **91**, 19.23.1-19.23.15 (2018).
113. Lam, S. S., Martell, J. D., Kamer, K. J., Deerinck, T. J., Ellisman, M. H., Mootha, V. K. & Ting, A. Y. Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nature Methods* **12**, 51–54 (2014).

114. Jiang, S., Kotani, N., Ohnishi, T., Miyagawa-Yamguchi, A., Tsuda, M., Yamashita, R., Ishiura, Y. & Honke, K. A proteomics approach to the cell-surface interactome using the enzyme-mediated activation of radical sources reaction. *Proteomics* **12**, 54–62 (2012).

115. Bar, D. Z., Atkatsh, K., Tavarez, U., Erdos, M. R., Gruenbaum, Y. & Collins, F. S. Biotinylation by antibody recognition - a method for proximity labeling. *Nature Methods* **15**, 127–133 (2018).

116. Rees, J. S., Li, X. W., Perrett, S., Lilley, K. S. & Jackson, A. P. Selective proteomic proximity labeling assay using tyramide (SPPLAT): A quantitative method for the proteomic analysis of localized membrane-bound protein clusters. *Current Protocols in Protein Science* **88**, (2017).

117. Samoudi, M., Kuo, C. C., Robinson, C. M., Shams-Ud-Doha, K., Schinn, S. M., Kol, S., Weiss, L., Petersen Bjorn, S., Voldborg, B. G., Rosa Campos, A. & Lewis, N. E. In situ detection of protein interactions for recombinant therapeutic enzymes. *Biotechnology and Bioengineering* **118**, 890–904 (2021).

118. Chen, C. L. & Perrimon, N. Proximity-dependent labeling methods for proteomic profiling in living cells. *Wiley Interdisciplinary Reviews: Developmental Biology* **6**, (2017).

119. Sakai, T., Nishimura, S. I., Naito, T. & Saito, M. Influenza A virus hemagglutinin and neuraminidase act as novel motile machinery. *Scientific Reports* **7**, (2017).

120. Jia, N., Byrd-Leotis, L., Matsumoto, Y., Gao, C., Wein, A. N., Lobby, J. L., Kohlmeier, J. E., Steinhauer, D. A. & Cummings, R. D. The human lung glycome reveals novel glycan ligands for influenza A virus. *Scientific Reports* **10**, (2020).

121. Hollinshead, M., Sanderson, J. & Vaux Sir William, D. J. Anti-biotin antibodies offer superior organelle-specific labeling of mitochondria over avidin or streptavidin. *The Journal of Histochemistry & Cytochemistry* **45**, 1053-1057 (1997).

122. Malaker, S. A., Riley, N. M., Shon, D. J., Pedram, K., Krishnan, V., Dorigo, O. & Bertozzi, C. R. Revealing the human mucinome. *bioRxiv* (2021).