

Chemo-enzymatic Synthesis of *O*-Acetylated Sialosides and Their Unnatural Analogs to Probe Virus-Glycan Interactions

Chemo-enzymatische Synthese van *O*-Geacetyleerde Sialosides en Hun Onnatuurlijke Analogen om Virus-Glycaan Interacties te Bestuderen

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de
Universiteit Utrecht
op gezag van de
rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op

woensdag 13 april 2022 des middags te 12.15 uur

door

Zeshi Li

geboren op 21 december 1990
te Qingdao, China

Promotor:

Prof. dr. G.J.P.H. Boons

平静孤独

快乐幸福

在这条没有行人的路上

那钻石般的光芒

永远年轻

永远的热泪盈眶

节选自 万晓利《达摩流浪者》

Table of Contents

Key Abbreviations	1
Chapter 1 General Introduction.....	3
Part 1. Sialoglycobiology	4
Part 2. Glycochemistry and Chemical Biology of Sialic Acid	12
Part 3. Outline of This Thesis.....	23
Chapter 2 Synthetic O-Acetylated Sialosides Facilitate Functional Receptor Identification for Human Respiratory Viruses	33
Abstract	34
Introduction	35
Results and Discussion.....	37
Conclusion.....	45
Methods.....	46
References	49
Supplementary Information	53
Chapter 3 Synthetic O-Acetyl-N-glycolylneuraminic Acid Oligosaccharides Reveal Host-Associated Binding Patterns of Coronaviral Glycoproteins	93
Abstract	94
Main Text	95
References	104
Supplementary Information.....	106
Chapter 4 Synthetic O-Acetylated Sialosides and their Acetamido-deoxy Analogs as Probes for Coronaviral Hemagglutinin-Esterase Recognition	129
Abstract	130
Introduction	131

Results and Discussion	133
Conclusion.....	143
References	144
Supplementary Information	148
Chapter 5 Conjugation of a Toll-like Receptor Agonist to Glycans Preserves Neutralization Epitopes of HIV Native-like Envelope Trimer.....	197
Abstract	198
Introduction	199
Results	201
Discussion	205
References	206
Supplementary Information	209
Chapter 6 Discussion & Summary	221
General Discussion.....	222
Nederlandse Samaenvatting (Summary in Dutch).....	235
Acknowledgements (致谢).....	237
Curriculum Vitae (with a list of publications).....	241

Key Abbreviations

Ac	acetyl
ACN	acetonitrile
BCoV	bovine coronavirus
Bn	benzyl
bNAb	broadly neutralizing antibody
CASD1	CAS1 domain-containing protein 1
CMP	cytidine-5'-monophosphoryl
CoV	coronavirus
CRCoV	canine respiratory coronavirus
CsF	cesium fluoride
CuAAC	Copper(I)-catalyzed alkyne-azide cycloaddition
DCM	dichloromethane
DMF	N,N-dimethylformamide
DMSO	dimethylsulfoxide
ECoV	equine coronavirus
EDC	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide
Env	envelope spike protein of HIV-1
Et	ethyl
Gal	D-galactose
GalNAc	N-acetyl-D-galactosamine
GD3	disialoganglioside 3
GlcNAc	N-acetyl-D-glucosamine
GM3	monosialoganglioside 3
HE	hemagglutinin-esterase
HEF	hemagglutinin-esterase-fusion protein
HILIC	hydrophilic interaction liquid chromatography
HIV	human immunodeficiency virus
HOBt	1-hydroxybenzotriazole
HPLC-MS	high performance liquid chromatography-mass spectrometry
ICV	influenza C virus
IDV	influenza D virus
LacNAc	N-acetylglucosamine
Lc	6-aminocaproic acid linker
MD	molecular dynamics
Me	methyl
MHV	mouse hepatitis virus
Nap	2-naphthyl
Neu5Ac	N-acetylneuraminic acid

Neu5Gc	N-glycolylneuraminic acid
NMR	nuclear magnetic resonance
PBS	phosphate buffer saline
PEG	polyethylene glycol
RbCoV	rabbit coronavirus
S	spike protein of coronavirus
SEC	size-exclusion chromatography
Sia	sialic acid
SIAE	sialate-O-acyltransferase
ST8Sia1	alpha-N-acetylneuraminide alpha-2,8-sialyltransferase 1
STD	saturation transfer difference
TFA	trifluoroacetic acid
THF	tetrahydrofuran
TLR	toll-like receptor
TMSOTf	trimethylsilyl trifluoromethanesulfonate
ToV	torovirus

CHAPTER 1 | General Introduction

Chapter 1

PART 1. Sialoglycobiology

1.1 Diversity and complexity of sialic acids

Almost all cells are covered by a dense layer of carbohydrates, or glycans, linked to various carriers such as proteins and lipids. This layer is referred to as glycocalyx¹. The pool of glycans in the glycocalyx is structurally very diverse and complex, and play multi-faceted roles in numerous biological events. The composition of glycocalyx is regulated in a cell-/tissue-specific manner, and glycan structures can differ considerably across different tissues within the same organism. The glycocalyx is highly dynamic and cell state-dependent, and can rapidly change in response to outside stimuli. Because glycocalyx is located at the border between the plasma membrane and the extracellular environment, it often mediates crosstalk of a cell with “foreign” entities, such as microbes. For example, glycocalyx provides niches for commensal bacteria to colonize²; on the other hand, it can also be targeted by the smallest invaders, viruses, leading to diseases³.

This thesis focuses on the sialic acids which have the 3,5-dideoxy-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid backbone.⁴ Their deoxy and epimeric variants (for example, legionaminic acid and pseudaminic acid) found in a number of bacteria will not be discussed. Sialic acids (Sias) are arguably the most important component in the glycocalyx given its ubiquitous presence at the very frontline. Sias are a class of negatively charged, nine carbon keto sugar found across all three kingdoms of life. Sias normally occur at termini of complex glycans as a capping residue. Interestingly, Sias are structurally diversified by a remarkable array of natural modifications, which exceeds that of the common hexose building blocks. Although it is hitherto unclear why nature creates such a large collection of Sia variants, it is hypothesized that this may be due to evolutionary pressure posed by pathogens^{5, 6}. Essential endogenous functions of Sias critical to the host are preserved, whereas disadvantageous interactions with microbes can be blocked by, for example, *O*-acetylation^{7, 8}.

The “total complement of sialic acid types and linkages and their mode of presentation on a particular organelle, cell, tissue, organ or organism—as found at a particular time and under specific conditions” is given the term “sialome”⁹, which is a subset of glycome. The extraordinary complexity of the sialome comprises multiple layers. The first layer is the chemical structure of the Sia residue: *N*-acetylneuraminic acid (Neu5Ac) and *N*-glycolylneuraminic acid (Neu5Gc) are the most common forms of Sias; less abundant are KDN and the 5-de-*N*-acetylated version of neuraminic acid¹⁰. In particular, Neu5Ac and Neu5Gc can be further modified by acetyl esters at C4, C7, C8 and/or C9 providing up to 22 molecular variants (Fig. 1). Other modifications of hydroxyls are less common, including lactylation and phosphorylation at C-9, and methylation and sulfation at C-8.

The second layer is the glycosidic linkages between Sia and the underlying glycan. Carbohydrates are biosynthesized in a non-templated fashion and is catalyzed by various

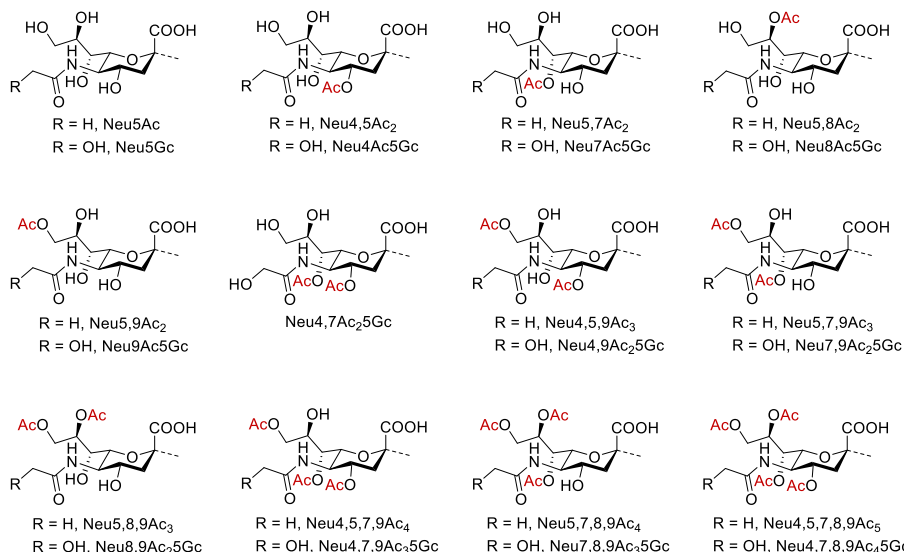


Figure 1 | Naturally occurring molecular variants of sialic acid, or sialoforms. Modification can occur at C4, 5, 7, 8 and/or 9 position.

glycosyltransferases specific for building one type of glycosidic linkage. In mammals, up to twenty glycosyltransferases capable of adding Sia residue to a specific hydroxyl of a glycosyl acceptor (sialyltransferases) have been identified¹¹. These enzymes use β -linked cytidine-5'-monophosphoryl-activated Sia (CMP-Sia) as glycosyl donor, which can be *O*-acetylated or hydroxylated (giving Neu5Gc) at this stage^{12, 13}. A Sia residue can be attached to a galactoside (Gal) *via* α 2,3- or 2,6-linkage, an *N*-acetyl galactosamine (GalNAc) *via* α 2,6 linkage, or to another Sia *via* α 2,8 linkage (Fig. 2).

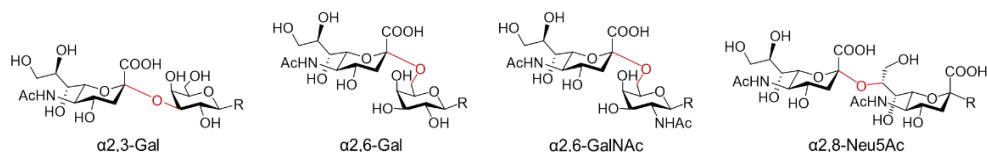


Figure 2 | Common glycosidic linkages between sialic acid and underlying sugars. Sialosidic bonds are shown in red. R denotes underlying carbohydrate residues.

In the third layer of complexity, Sia is placed in the context of glycotopes, which generally consist of several sugar units and normally occur at non-reducing termini (examples are provided in Fig. 3). The structure of a given glycotope is governed by an array of enzymes, including glycosyl transferases other than sialyltransferases, and sulfotransferases to modify underlying glycans. Glycotopes are key elements for molecular recognition and critically determine the fine specificities of Sia-binding proteins. For example, sialic acid-binding immunoglobulin-type lectins (Siglec's) are a family of inhibitory immunoreceptors, for

Chapter 1

which the presence of Sia residue is necessary for binding. However, the seventeen members of Siglecs in human differ in their preference for the underlying glycan structures which make up the complete glycotope¹⁴.

Sia can form very large glycotope consisting of over a hundred units. These special glycotopes are commonly referred to as polysialic acid (polySia), a homopolymer up to a hundred Neu5Ac residues linked in $\alpha 2,8$ glycosidic bond¹⁵. The biosynthesis thereof involves a priming enzyme, ST8Sia3¹⁶, which transfer an additional Neu5Ac residue to Neu5Ac $\alpha 2,3$ Gal moiety present in a number of glycoproteins; and later ST8Sia2 and 4, which polymerize Neu5Ac units on select glycoproteins. The occurrence of polySia is restricted to a subset of neural cells, dendritic cells and a number of tumor cells^{17, 18}. Interestingly, this particular structure is also biosynthesized by a number of pathogenic bacteria, including Serogroup B *Neisseria meningitidis* and *Escherichia Coli* K1, as a way to evade immune detection, because the glycans are viewed as ‘self’ by the host immune system¹⁹.

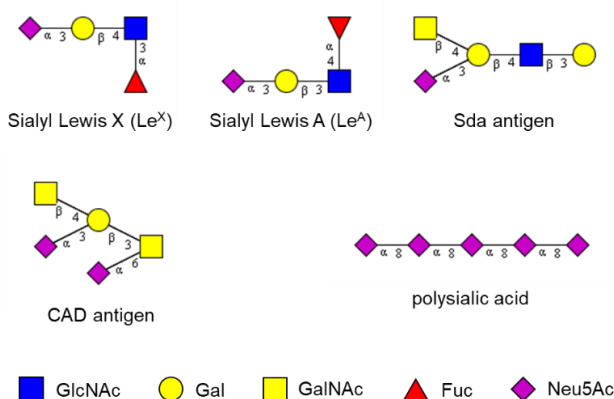


Figure 3 | Examples of sialylated glycotopes. Abbreviations: GlcNAc, *N*-acetyl glucosamine; Fuc, Fucose. Polysialic acid normally consists of more than a hundred Neu5Ac residues.

A given glycotope can occur as part of asparagine-linked glycans (*N*-glycans) or serine/threonine-linked glycans (*O*-glycans) on glycoproteins, as well as being part of glycolipids. Where the glycotopes are placed matters for their functions, because the localization affects greatly their presentation and local density. Mucins are heavily glycosylated proteins which bear a tandem repeating domain placed away from plasma membrane²⁰. This domain harbors densely *O*-linked glycans that comprises 80% of its molecular weight, adopting a brush-like morphology. This mucin-type hypervalent display of carbohydrates are reported to bear important roles in regulating membrane shape²¹, modulating colonic microbiota²² and effectively blocking microbial infection²³.

The final layer of complexity goes beyond glycans, and arises from the carrier molecules of the sialoglycan and how these sialoglycoconjugates are organized spatially in a cell. What are the *bona fide* functions of Sias? To provide a clear answer to this question, we have to restrict it to a specific glycoprotein or a glycolipid in a given recognition event. In other words, the functions of Sias are critically associated with the carrier molecules and the partner interacting with them. This is perhaps best exemplified by several immune-events. For example, the Sia-binding inhibitory receptor siglec-2 (CD22) does not non-specifically associate with any sialoglycoproteins; instead, it binds selectively the sialoglycans present

on a small panel of cell surface proteins, including siglec-2 on the same cell (forming homo-oligomers) and B cell receptor *in-trans*^{24, 25}. This spatial arrangement of siglec-2 and the binding partners plays important roles in regulating B cell activation. Another Sia-binding lectin, siglec-10, on tumor-associated macrophages specifically interacts with CD24, a highly sialylated GPI-anchored protein on tumor cells, which was shown to be an important mechanism for tumor evasion from phagocytic clearance²⁶.

1.2 Biological functions of *O*-acetylated sialic acid

O-acetylation is a common modification of sialic acids; the product of which are important yet understudied members of the sialome. Despite their wide occurrence, very limited knowledge regarding the biological significance of *O*-acetylated sialic acids has been established. In this section, what's known about these molecules is reviewed.

O-Acetylated sialic acid (*O*-Ac-Sia) is detected in many types of cells and is expressed in a tissue/organ-specific manner^{27, 28}. There are generally two ways to study the distribution of this class of glycotopes: *in-situ* lectin staining and GC/HPLC-MS analysis of cleaved, labeled sialic acids. The former harnesses plant- or microbe-derived *O*-Ac-Sia-binding proteins. The hemagglutinin-esterase (HE) proteins from nidoviruses and orthomyxoviruses with an inactive esterase domain (therefore term "virolectins") have been expressed as Fc fusion proteins for tissue array studies. Different *O*-acetylated forms of Sia can be stained by virtue of the inherent selectivities of the virolectins for *O*-acetylation patterns. For example, the HE from porcine torovirus has a selectivity for 9-*O*-acetylated Sias, whereas an additional 7-*O*-acetyl abolishes binding; while the one from bovine coronavirus greatly prefers 7,9-di-*O*-acetylated sialoform over a single acetyl ester at C-9. This method made it possible to discover that humans expresses 9- and 7,9-di-*O*-acetylated Sias across multiple tissues, including adrenal gland, colon, cerebellum, gray matter, pancreas, prostate and salivary gland.²⁸ The latter method for detecting *O*-Ac-Sias involves treating the sample (glycoproteins, cultured cells or tissue) with 2 M acetic acid at 80 °C to selectively cleave Sia residues from glycoconjugates and derivatizing the released Sia with 1,2-diamino-4,5-methylenedioxybenzene (DMB). The DMB-labeled Sias can then be subjected LC/GC-MS analysis and quantified with fluorescence^{13, 29}.

It remains unclear why and how an organism generates many forms of *O*-Ac-Sias. Only the biosynthesis of 9-*O*-acetylated sialosides has been studied in detail, whereas that of 4- or 7-*O*-acetyl containing structures is hitherto elusive. The addition of acetyl ester occurs in the Golgi and involves a sialate-*O*-acetyltransferase, namely CASD1, which employs acetyl-CoA as the acetyl donor and CMP-Sia as the acceptor substrate¹³. The 9-*O*-acetylated CMP-Sia can then be unitized by sialyltransferases to modify various types of glycoconjugates. The acetyltransferase activity was mapped to the *N*-terminus domain of CASD1, which is enzymatically active when expressed as a soluble protein. Interestingly, CASD1 also contains a multi-transmembrane C-terminus domain. Less is known about the function of this portion of the protein, but it has been suggested this domain may have regulatory function and is necessary for addition of an acetyl ester to C-7 of Sia, producing 7-*O*-mono and 7,9-di-*O*-acetylated sialosides (Unpublished data, Dr. Yifei Lang).

The overall level of *O*-Ac-Sia expression of a cell is dynamic and is determined by the interplay between CASD1 and a family of enzymes which hydrolyze 9-*O*-Ac of Sia, the sialate-9-*O*-acylesterases (SIAEs), which are localized in the lysosome or secreted extracellularly (Fig. 4). This interplay has been shown to play an important role in B cell immunity³⁰. It was found that the absence of SIAE significantly lowers the threshold for B cell receptor (BCR)-mediated activation. The molecular mechanism of this modulatory

Chapter 1

function has been elucidated. On B cell surface, BCR associates with Siglec-2 (CD22), an inhibitory Sia-binding protein that dampens the activating signals transduced from BCR clustering upon antigen engagement. Siglec-2-BCR association is dependent on unmodified Neu5Ac on the latter. 9-*O*-acetylation of BCR Sia abolishes Siglec-2 recognition of Neu5Ac, and thereby prevents Siglec-2 from associating with BCR. As a result, a B cell is more readily activated. The abnormally low level of SIAE resulting from missense single-nucleotide polymorphism in the SIAE gene therefore can lead to autoimmune diseases characterized by circulating auto-antibodies against self-proteins.

Another contributor to the dynamic regulation of sialate-*O*-acetylation is a family of neuraminidases in the lysosome and plasma membrane, which cleave Sia moiety from glycoconjugates. In human, there are four neuraminidases, namely NEU1, NEU2, NEU3 and NEU4 (Fig. 4)³¹. Although these enzymes commonly hydrolyze Sia residue from glycoconjugates, they exhibit differential hydrolytic activities relating to *O*-acetylation patterns on the Sia residue, sialosidic linkages and underlying glycoforms³². Thus, acetyl ester modification fine tunes the activity of these enzymes. Specifically, the presence of an acetyl ester at C-9 negatively affect the hydrolytic activities of NEU1-3 to varying degrees, with NEU2 activity almost completely abolished. Conversely, NEU4 activity is increased for 9-*O*-acetylated variants. The results also suggest that sialate-9-*O*-acetylation may have modulatory role in the metabolism of sialoglycans.

The best-studied glycoconjugate carrying *O*-Ac-Sia is 9-*O*-acetylated GD3 ganglioside, also known as CD60b. It has been reported as an oncofetal marker and found in many types of cancers such as pre-B cell acute lymphoblastic leukemia³³, melanoma³⁴, neuroblastoma and breast cancer³⁵. This molecule exhibits anti-apoptotic effect, thus promoting cancer cell survival and contributing to drug resistance³⁶. Remarkably, the removal of the sialate-9-*O*-acetyl *via* transient expression of viral sialate-9-*O*-acetyl esterases or knocking out CASD1 sensitizes cancer cells for cytotoxic agents, providing a potential target for cancer therapy³³. The ester hydrolysis of CD60b affords the non-modified GD3, or CD60a, a well-known pro-apoptotic messenger³⁷. CD60a largely resides on the cell surface and also in the Golgi. Upon initiation of programmed cell death, CD60a translocates to mitochondria, and through an unknown mechanism disrupt the transmembrane potential and increases permeability of mitochondria, leading to propagation of the apoptosis-related signals, such as cytochrome c release and activation of caspase-9³⁸.

Remarkably, the addition of an acetyl ester to C9 can simply neutralize the death-promoting effect by its non-acetylated counterpart. In other words, 9-*O*-acetylation abolishes the detrimental effect of GD3 on mitochondria. It was observed *in vitro* that 9-*O*-acetylation effectively prevent GD3 from oxidation into 7-aldehyde product, which is the postulated active species in disrupting mitochondria membrane³⁹. Thus, GD3 and its derivatives appear to be important regulators in Golgi-mitochondria crosstalk. However, what mediates the translocation event is still an open question. The observation also points to one or more hitherto elusive receptor(s) or interacting partner(s) in either one of the organelles.

1.3 Sialic acid as receptor for coronavirus

Coronaviruses (CoVs) are a group of enveloped, positive single-stranded RNA viruses. CoVs are divided into four genera: alpha, beta, gamma and delta. The betacoronavirus genus harbors most of the human-infecting CoVs and can be further categorized into several subgenera, including embeco-, hibeco-, merbeco- and sarbecoviruses. There are seven coronaviruses known to infect humans, among which three have caused deadly pneumonia outbreaks, namely SARS-CoV, MERS-CoV and SARS-CoV-2. While the former two

exhibited limited human-to-human transmission and have been geographically restricted, the latter has led to the devastating COVID-19 pandemic⁴⁰. The other four coronaviruses, OC43, 229E, HKU1 and NL63 have been circulating in the population worldwide, and are

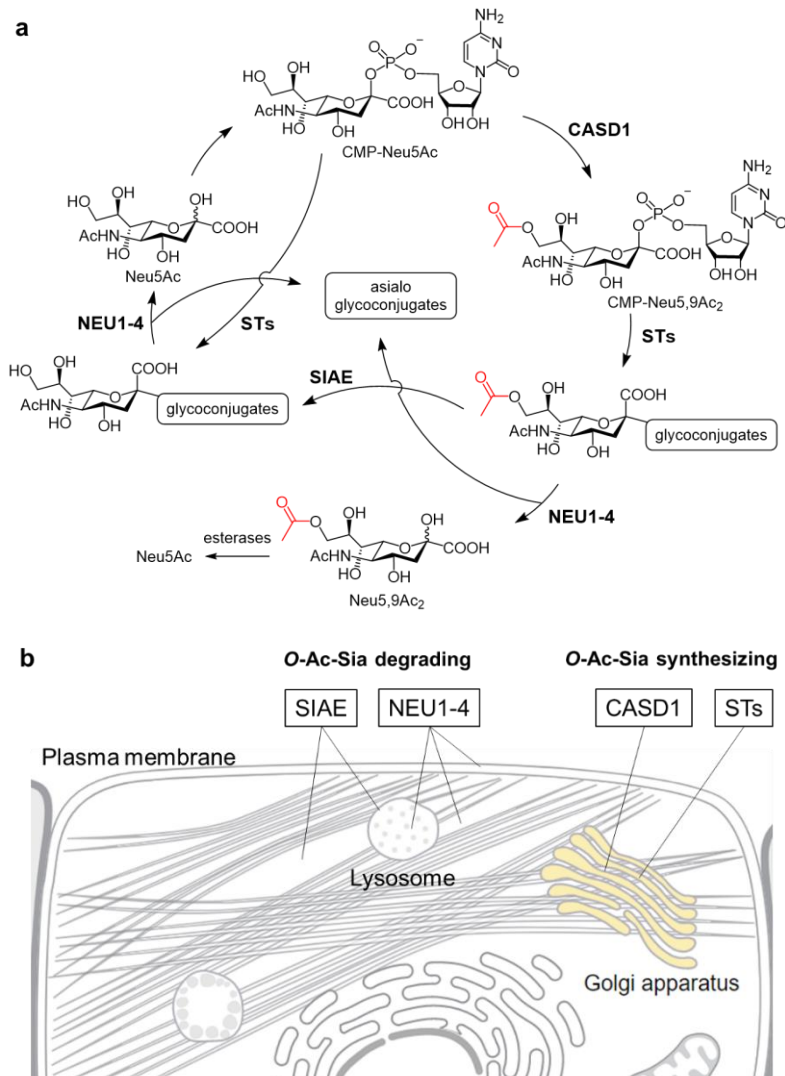


Figure 4 | Dynamic regulation of (de-)O-acetylation of sialoglycoconjugates. a, synthesis and degradation cycle of *O*-Ac-Sia. **b,** localization of enzymes related to *O*-Ac-Sia. CASD1 is multi-transmembrane protein in Golgi apparatus. ST denotes sialyltransferase and is also localized on the inner membrane of Golgi, facing the lumen. SIAEs have two isoforms, cytosolic and lysosomal. SIAE has also been found to be secreted extracellularly. Mammalian sialidases (NEU1-4) have four subtypes and are found in lysosome, plasma membrane and cytosol. Figure source, Uniprot (panel **b**).

Chapter 1

associated with generally mild symptoms in healthy adults, but can lead to serious complications in infants, the elderly and immunocompromised individuals⁴¹.

Coronaviruses encode a common trimeric surface projection, the spike protein (S), which is responsible for engaging cell surface receptors, supporting viral entry and membrane fusion⁴². S is proteolytically processed into S1 and S2 subunits. The receptor binding domain is normally located in the S1 C-terminus region, or S1^B, and interacts with protein receptors (Fig. 5). The N-terminus region of S1 (S1^A) has been shown to mainly interact with carbohydrates such as (*O*-acetylated) Sias. For example, besides engaging DPP4 via S1^B, MERS-CoV can weakly agglutinate to α 2,3-linked sialic acids via S1^A, which is not necessary but substantially enhance virion uptake^{43, 44}. Interestingly, this binding specificity correlates well with the tissue distribution of 2,3-linked sialosides in both human and the intermediate host, camel, as the lung epithelia from both species can be intensively stained with MAA-II, a sialic acid-binding lectin which prefers α 2,3-linkage. Other sialic acid-binding coronaviruses includes infectious bronchitis virus (IBV) which infects birds⁴⁵, and porcine deltacoronavirus (PDCoV)⁴⁶ and porcine epidemic diarrhea coronavirus (PEDV)⁴⁷. However, the functional consequences of this engagement still await elucidation.

Unlike other Coronaviridae members, the ones of the subgenus Embecovirus engage *O*-acetylated sialic acid *via* S1A for viral entry^{48, 49}. For example, OC43 and HKU1 entered human population independently and both have evolved to bind to 9-*O*-acetylated sialic acid via S1^A. One exception of embecoviruses in receptor recognition is MHV, which engages the glycoprotein Carcinoembryonic antigen-related cell adhesion molecule 1a (CEACAM-1a) *via* the same domain⁵⁰. Embecoviruses further differ from other coronaviruses by encoding another dimeric envelope glycoprotein, the hemagglutinin-esterase (HE). As the name suggests, HE is capable of binding to *O*-acetylated sialic acids via the lectin domain and hydrolyze sialate acetyl esters via the esterase domain²⁸. It serves as a receptor destroying enzyme and mediate dynamic interactions with multivalent (decoy) receptors⁵¹. Interestingly, homologous gene of HE has been found in viruses from a distinct taxon: influenza C and D viruses (Orthomyxoviridae)⁵². These viruses integrate receptor binding and destruction and membrane fusion functions in one single trimeric surface projection: the hemagglutinin-esterase-fusion protein (HEF).

Functioning as a two-component system that mediates dynamic interaction with receptors, S and HE critically determine host range and tissue tropism. However, the understanding of S and HE receptor binding has been limited to the global affinity for the terminus Sia monosaccharide residue present in a pool of poorly analyzed glycans, with most studies focusing only on Sia 9-*O*-acetylation. 9-*O*-Ac-Sia is bound by human⁴⁸, murine⁵³, bovine⁵⁴ and porcine viruses, which all form monophyletic clades and exhibit high degree of host specificities. The overlap in receptor usage by these viruses with distinct host specificities implies that important dimensions governing *O*-Ac-Sia recognition have been overlooked. These include 1) the contribution of other *O*-acetylated variants (also termed “sialoforms”), 2) the contribution of *N*-glycolyl modification, 3) the preference for distinct

sialoside topology or glycosidic linkages, and 4) quantitative differences in binding to the same receptor.

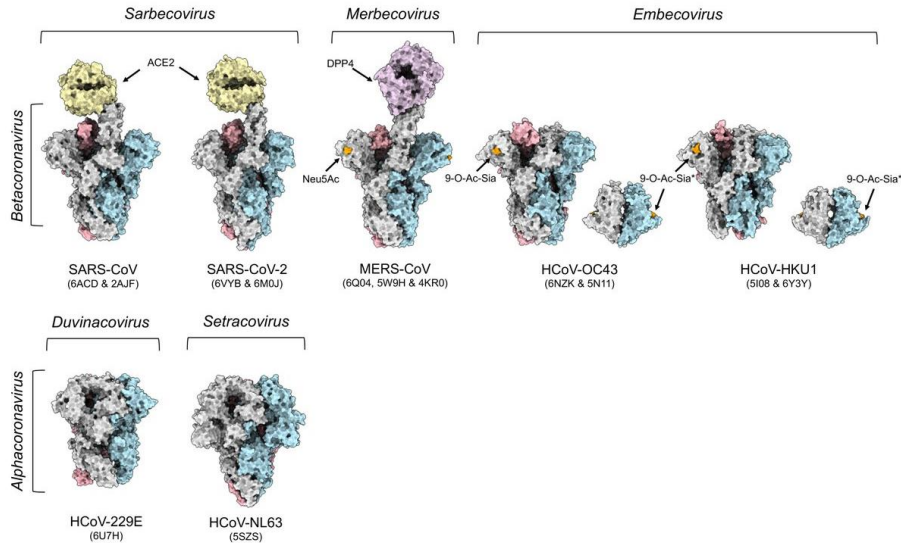


Figure 5 | Surface representation of spike proteins of human coronaviruses and the receptors.

The receptors ACE2, DPP4 and 9-*O*-Ac-Sia are shown in yellow, purple and orange, respectively. Protomers of spike protein are shown in grey, cyan and red. 9-*O*-Ac-Sias in cases of OC43 HE and HKU1 spike and HE were modeled in, and indicated with asterisk. Abbreviations: ACE2, angiotensin converting enzyme 2; DPP4, dipeptidyl peptidase 4; SARS-CoV, severe acute respiratory syndrome coronavirus; MERS-CoV, middle-east respiratory syndrome coronavirus. Figure source, @DanielHurdiss | Twitter.

PART 2. Glycochemistry and Chemical Biology of Sialic Acid

2.1 Chemical and chemoenzymatic synthesis of sialosides

Given the pivotal roles of Sias in mediating many biological processes, molecular understanding of sialoglycan recognition has been pursued by many laboratories. To facilitate this, synthetic glycochemistry provides homogenous, structurally well-defined glycans in high purity for probing the interaction with Sia-binding proteins. Critical advances in both chemical and (chemo-)enzymatic syntheses have granted access to many complex sialosides. The expanding collection of synthetic glycans are central to understanding the biological roles of sialoglycoconjugates. One key aspect of this chapter is to give an overview of the development of methods to synthesize sialosides, as well as the acetyl ester modified counterparts.

Constructing a glycosidic bond chemically generally involves coupling of a fully protected glycosyl donor which contains a potential leaving group at the anomeric center to a glycosyl acceptor exposing one free hydroxyl. The stereoselectivity is governed mostly by the nature of the adjacent protecting group. A 1,2-*trans* glycoside is readily formed when the C2-functionality is acyl-type which can participate in stabilizing the oxocarbenium ion (Fig. 6a). In the absence of such neighboring participation, both 1,2-*trans* and 1,2-*cis* glycosidations are possible (Fig. 6b). The former is normally a kinetically controlled product, whereas the latter is of lower free energy owing to the anomeric effect.

Building a glycosidic bond between sialic acid and other sugars is somewhat different. Unlike regular hexopyranoses which preferentially adopt a 4C_1 conformation, sialic acid has a 2C_5 ring, with a carboxylate functionality at anomeric center (C2), and an acetamido moiety and a bulky side chain at C5 and C6, respectively. The neighboring C3 is free of substituents and thus does not allow protection by participating functionality (Fig. 6c). As a result, the stereoselectivity of chemical sialylation does not subject to the empirical rule above. Moreover, while naturally occurring sialosides have an alpha-anomeric configuration, the unwanted beta isomer is favored thermodynamically. Thus, to construct one sialosidic bond, many factors such as the combination of protecting groups, nucleophilicity of acceptors, solvents and temperatures need to be examined to overrule the anomeric effect⁵⁵. Further problems stem from oxocarbenium ion destabilization by the electron-withdrawing C1-carboxylate, making 2,3-ene—or glycal—formation a competing side reaction.

The difficulties of glycosylation of sialic acid have been addressed by various approaches. Installation of an artificial C3-participating group represents the early strategies undertaken by Ito, Whitesides and Schmidt labs⁵⁶⁻⁵⁹, which are now referred to as indirect sialylation to achieve alpha-selectivity (Fig. 7a). In this approach, the equatorial C3-hydrogen is replaced with a chalcogen or a halogen, which is expected to form a three- or five-membered ring with anomeric C2 upon dissociation of leaving group, thus blocking the nucleophile attacking from β -side. However, the participating functionalities have to be removed with radical-mediated reduction post-sialylation to restore the natural structure. C1-

auxiliary was introduced as an alternative strategy (Fig. 7b), where the carboxylate is modified as a 2-(methylthio)ethyl⁶⁰, 2-(dimethylamino)-2-oxoethyl⁶¹ ester, or more recently with a picolinyl group⁶². Formation of a spiro[5.5]undecane-type species was postulated for these sialyl donors upon activation. It has been proposed that other C1-auxiliaries such as a thioester may work via a distinct mechanism when using nitrile solvent, in which the beta-

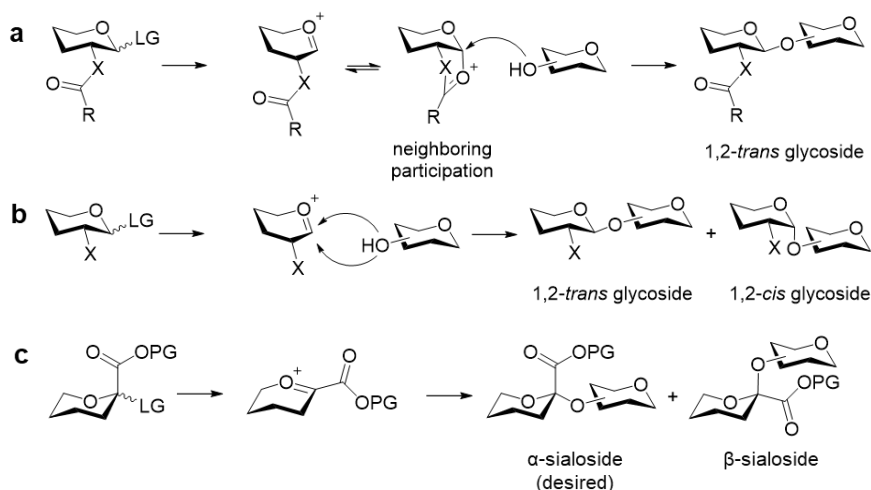


Figure 6 | General rules for building a glycosidic bond. **a**, 1,2-*trans* glycosidic bond is facilitated by neighboring participation. **b**, Without neighboring participation, a glycosylation reaction can give a mixture of 1,2-*trans* and 1,2-*cis* glycosides. Normally the former is the kinetically controlled and the latter is thermodynamically favored. **c**, α -Sialosidic bond cannot be constructed with neighboring participation, nor is thermodynamically favored.

sialyl nitrilium adduct is stabilized by the sulfur⁶³.

C5 modification was sparked by the finding by the Boons group that the attachment of an additional acetyl to the acetamido moiety⁶⁴ to form an imine not only reduces the unwanted nucleophilicity of the C5-acetamide but also significantly enhances the reactivity of the donor (Fig. 7c and d). Other electron-withdrawing amine protecting groups were thereafter investigated, which includes trifluoroacetyl (TFA)⁶⁵, azide⁶⁶, 2,2,2-trichloroethoxycarbonyl (Troc)⁶⁷, phthalyl⁶⁸ and isothiocyanate⁶⁹. 5*N*,4*O*-carbonyl protected sialyl donors were explored^{70, 71}, and its α -directing mechanism were studied in depth by Crich group⁷². The oxazolidinone *trans*-fused to the sialic acid pyran ring creates a strong dipole moment in the mean plane of the system, making it a powerful electron-withdrawing group favoring α -sialosidic bond formation.

The substituents on the glycerol side chain (C7-C9) which are seemingly remote from the anomeric center can also have a profound effect on the reactivity of the sialyl donor and the stereo-outcome (Fig. 7d and e)⁷³⁻⁷⁶. The absolute α -stereocontrol has been achieved only recently by Hiromune group using a macrocyclic sialyl donor, in which C1 and C5 is

Chapter 1

bridged with an alkyl chain of specific length (Fig. 7f)⁷⁷. The geminal dichloride present at the beta position of the C5-carbonyl not only boosts the reactivity of the donor, but allows facile removal of the alkoxy carbonyl with zinc, exposing the C5-amine for subsequent manipulation.

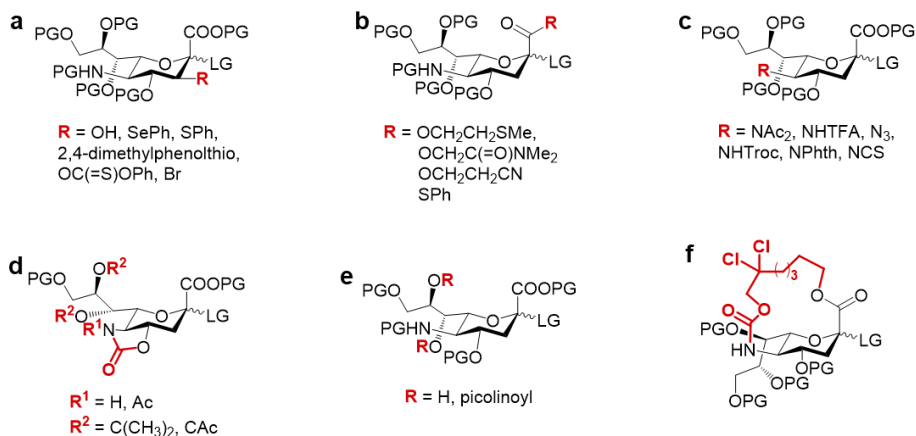


Figure 7 | Reported sialyl donors for constructing α -sialosides. Abbreviation: PG, protecting group; LG, leaving group; SePh, phenylselenyl; SPh, phenylthiol; Me, methyl; Ac, acetyl; TFA, trifluoroacetyl; Troc, 2,2,2-trichloroethoxycarbonyl; Phth, phthalyl.

In parallel with the advancement in chemical sialylation, biocatalytic conversion has been under active investigation and is recently becoming a mainstream strategy in sialyl oligosaccharide assembly, especially for those glycans that may be too complex to synthesize by conventional chemical approach, for example, the asymmetrically branched multi-antenna N-linked glycans and complex human milk oligosaccharides.

While early efforts focused on devising enzymatic cycles of sugar-nucleotide regeneration from readily available monosaccharide building blocks, pioneered by Whitesides and Wong groups^{78, 79}. More recent research mainly involved incorporating sialic acids into complex oligosaccharides, as more and more relevant enzymes from microbial sources became available. The scalability and flexibility of the conversion were significantly improved as Chen group developed a one-pot multi-enzyme (OPME) strategy, which led to the access to an expanded collection of diverse sialoglycans⁸⁰⁻⁸². In this strategy, sialosides are prepared starting from a six-carbon mannosamine precursor which undergoes condensation with pyruvic acid; is subsequently activated as sialic acid cytidine monophosphate; and is finally attached to glycosyl acceptors by bacterial sialyltransferases. By virtue of the promiscuity of these enzymes, a large variety of unnatural functionalities can be introduced to sialic acid residue, with C5 and C9 being the most investigated positions (reviewed in ⁸³). Recent reports from Cao group demonstrate that glycosyl acceptors presenting more than one potential sialylation site can be glycosylated at specific positions with excellent regiocontrol by exploiting simple chemical or chemoenzymatic manipulations

such as lactone formation⁸⁴ and galactose oxidase-catalyzed conversion combined with sodium cyanoborohydride-mediated reduction⁸⁵ (Fig. 8a).

Unlike the sialyltransferases of microbial origin, which exhibit broad substrate scope for glycosyl acceptors, many of those derived from mammals possess inherent selectivities for more intricate underlying structures. This interesting property has been exploited by Boons group to construct the asymmetrically branched *N*-linked glycans with unprecedented complexity⁸⁶. A critical step leading to desymetrization of the multi-antennary precursor prior to structural diversification involves treatment thereof with mammalian ST6Gal1, which selectively adds one sialic acid to the galactose residue in the bottom arm, thus allowing branch extension only from the top arm (Fig. 8b).

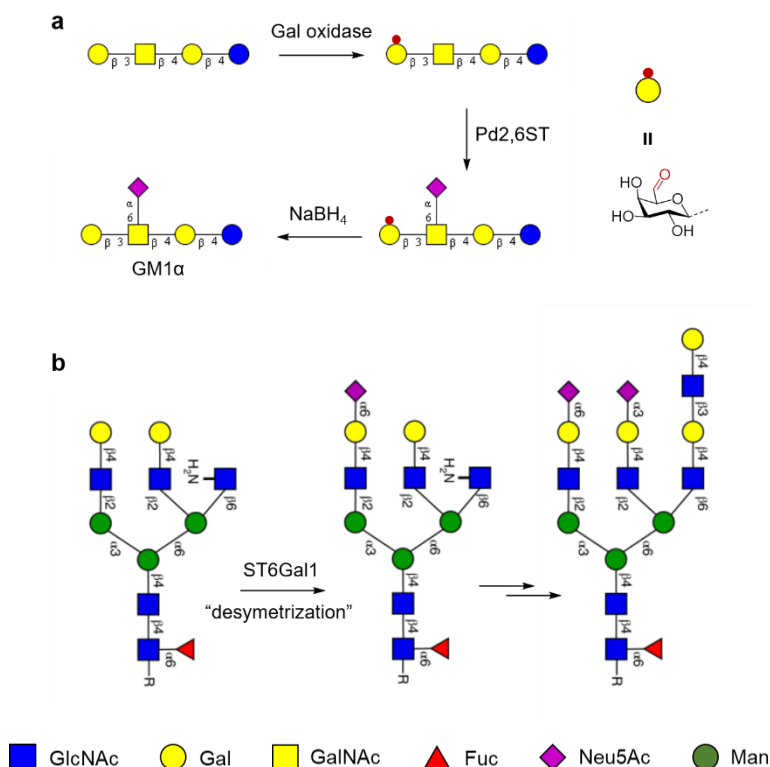


Figure 8 | Chemoenzymatic synthesis of complex glycans with regio-control. a, Regioselective sialylation of GalNAc residue using galactose oxidase to block terminal Gal. The oxidized 6-aldehyde form of Gal can be reversed using sodium borohydride to restore the natural glycan post-sialylation. **b,** Desymetrization of multi-branch *N*-glycan using inherent branch selectivity of ST6Gal-1. Treatment of bis-galactosylated precursor with ST6Gal1 affords the asymmetric *N*-glycan with Sia on the α 1,3 mannose arm.

O-acetylation of sialic acid in its own right poses another long-standing challenge in carbohydrate chemistry and biology. This is due to the inherent lability of sialate-*O*-acetyl

Chapter 1

esters. C9 acetyl ester is prone to non-enzymatic hydrolysis even under mildly basic condition (pH 8-9) and is readily hydrolyzed by esterases. C4 acetyl ester is also prone to rapid enzymatic degradation (Unpublished data, Dr. Yifei Lang). C7 and C8 acetyl esters, though resistant to esterases, can undergo migration to C9 position. The labile nature of acetyl esters has greatly complicated the isolation and analysis of these molecules from natural sources. Previous approaches for syntheses of O-acetylated sialic acid have been largely limited to monosaccharides and C9-monoacetylated sialyl oligosaccharides.

Early approaches for synthesizing O-acetylated sialic acid involves many protecting group manipulation steps to block non-acetylated hydroxyls for each acetylated form⁸⁷⁻⁹⁰. For monosaccharide synthesis, benzylidene, isopropylidene and tert-butyldimethylsilyl (TBS) groups are routinely used masking functionalities, the removal of which are compatible with acetyl esters (Fig. 9a). A complete monoacetylated series including C4, C7, C8 or C9 acetyl esters as well as several diacetylated forms can be accessed with these methods⁹¹. Based on the observation that the hydroxyls on sialic acid exhibit differential nucleophilicity, Park *et al.* reported a facile technique termed ReSET (regioselective silyl Neu5Ac starting from a C1-benzylated, per-*O*-TMSylated precursor is produced, which is then subjected to chromatographic separation to obtain each pure form. The ReSET approach grants access to a broader collection of naturally occurring and unnatural monosaccharides carrying multiple acetyl esters (Fig. 9b). The aforementioned OPME strategy provides readily C9 acetylated sialyl oligosaccharides in α 2,3, 2,6 and 2,8-linkages, while C4-acetylated counterpart can only be incorporated in α 2,3-linked sialosides⁹³. In these reactions, pH was carefully controlled to minimize acetyl hydrolysis. An alternative approach features the use of subtilisin-mediated transesterification in DMF with vinyl acetate as acetyl donor⁹⁴. This reaction can regioselectively add acetyl group to the 9-OH of terminal sialic acid residue of the glycolipid GD3 in moderate yield.

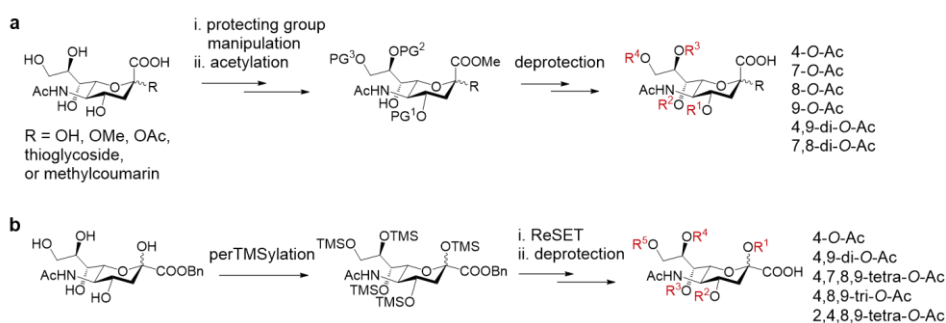


Figure 9 | Previous approaches to O-acetylated sialic acid monosaccharides. a, protecting group blocking-unblocking. **b**, Regioselective silyl ether exchange technology (ReSET). Abbreviations: TMS, trimethylsilyl; Bn, benzyl.

2.2 Modified sialic acid derivatives in chemical biology

Introducing non-native functionalities to glycans is a powerful approach to study these complex molecules. Sia has been a pursued target, given its many roles in glycobiology and considerable degree of toleration of “unnaturalness” in its biosynthesis. In this section, chemical biology approaches are reviewed to gain molecular understanding of Sia-protein interaction, selectively enhance desired functions of Sia, and incorporate designer functions *via* Sia into other biomolecules for downstream studies.

Glycosidic bond stabilization. One important aspect of unnatural modification is to stabilize covalent bonds that are chemically or enzymatically labile. Sialosides represent one such example: due to the presence of the C1-carboxylate, the glycosidic bond between sialic acid and the underlying sugar is more sensitive to acidic conditions as compared with other glycosidic linkages between regular hexoses¹⁰. Sialosides are substrate for many bacterial and viral sialidases, as well as for a family of lysosomal and membrane-associated neuraminidases in eukaryotes, potentially limiting the use of natural structures as probes for pull-down or on-surface binding studies. Although several potent neuraminidase inhibitors, such as oseltamivir, are available to minimize the degradation of sialosides, they are only selective for a subset of enzymes⁹⁵.

It is well known that monothioacetals are much more stable than the oxo-counterparts. Thus, substituting the extra-ring oxygen in a glycosidic bond with sulfur can make sialic acid remain glycosidically bound in the presence of a range of degrading enzymes and low pH (Fig. 10a). To synthesize *S*-linked sialosides, the highly nucleophilic thiol can be introduced to either the anomeric position of sialic acid or to the underlying sugar, followed by S_N2-type reaction with halogenated deoxy-sugar or sialyl halide, respectively⁹⁶⁻⁹⁹.

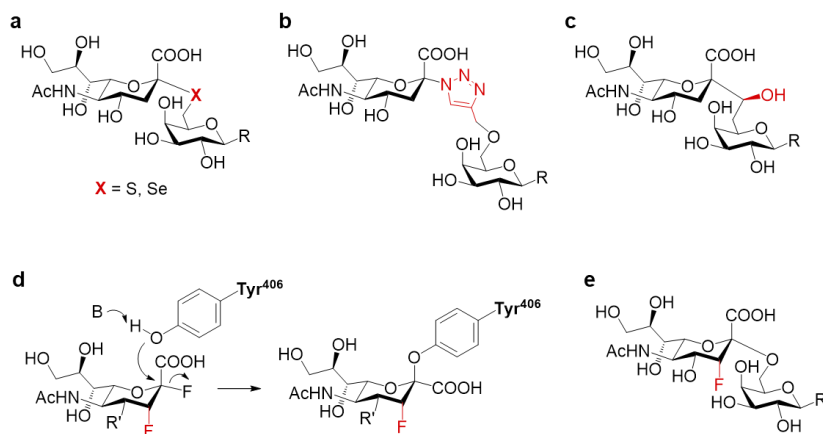


Figure 10 | Direct and indirect sialosidic bond stabilization. a-c, Replacement of glycosidic oxygen with other atom(s). The introduced non-native structures are highlighted in red. d, Working mode of mechanism-based neuraminidase inhibitor. B denotes the catalytic base residue. The covalent complex exhibits extended half-life of dissociation due the electron-withdrawing C3-fluorine. e, C3-fluorinated sialosides.

Chapter 1

Selenoglycosides were also reported using a similar synthetic strategy. Interestingly, the NMR-active ^{77}Se isotope provides a handle for investigating protein-carbohydrate interaction in solution¹⁰⁰. Alternative means to achieve stabilization is to prepare C- or triazole-linked glycosides¹⁰¹, in which additional atoms are introduced in between sialic acid and the penultimate residue (Fig. 10b and c). The stabilized sialosides have been utilized to capture influenza A virus particles¹⁰², and hold potential in virus detection and enrichment from biological samples.

Sialosidic bond can also be indirectly stabilized. The cleavage of Sia from the aglycon involves a transition state in which partial positive charge is developed at the anomeric center. Such transition state can be made more energy-demanding if positive charge is disfavored at the anomeric center, and as a result the sialosidic bond becomes more difficult to break. Fluorine atom is highly electronegative, and the attachment of this atom to C-3 position adjacent to the anomeric center poses a strong inductive effect which greatly increases the energy of the transition state. Thus, 3-fluoro-sialylated structures, especially those with an axially-oriented fluoride, is resistant to Sia hydrolase-mediated cleavage. This interesting property has been employed to develop a mechanism-based influenza A virus neuraminidase inhibitor with excellent potency and a broad inhibitory effect against multiple viral subtypes (Fig. 10d)¹⁰³. It was found that a covalent enzyme-substrate complex is formed upon the catalytic Tyr displacing the anomeric fluoride, and that dissociation of this complex was greatly retarded in the presence of an axial C-3 fluoride. In a different study from Wong group, glycosides containing axially 3-fluorinated Sia were prepared (Fig.10e). These sialosides were also resistant to sialidase treatment as compared with the natural counterpart. It appeared that the sialidase from *V. cholerae* does not recognize the fluorinated analog as substrate, as no competitive inhibition of the enzyme was observed¹⁰⁴.

Stabilization of sialate-acetyl modification. The labile nature of *O*-acetyl ester modifications of sialic acid discussed above is what hampers the research into this class of carbohydrates. Amide, which are much more stable than ester, is a well-established bioisosteres of the latter. It remains an open question whether the *O*-to-*N* substitution is well tolerated by the interacting protein, especially as such modification not only introduces an additional hydrogen atom but also alters the rigidity of the moiety, owing to the non-rotatable nature of amide bond. Furthermore, due to weaker electronegativity of nitrogen, electron density on the alpha-carbon increases, which is believed to be the cause of an altered distribution of rotamers. 9-*N*-acetylated sialic acid was shown to mimic 9-*O*-acetyl counterpart both computationally and experimentally. The unnatural sialoside analogs bind lectins equally well as the natural ones but is resistant to esterase degradation, providing a solution to the chemical lability of 9-*O*-acetylation¹⁰⁵. Whether *O*-to-*N* substitution applies to other positions remain unexplored.

Bioorthogonally tagged sialic acid. Bioorthogonal reactions are characterized with high degree of chemo- and regio- (site-) selectivities, which allows for the formation of

covalent bonds on a biomolecule, such as protein and DNA, at a designated residue without unwanted reactions on other ones¹⁰⁶. Because of its ubiquitous presence on glycoproteins, Sia has been extensively explored as a handle to introduce various biorthogonal tags, among which are ketone, azide, alkyne, cyclopropane or diazo at C5, C7 or C9 positions (reviewed in¹⁰⁷). These unnatural moieties appear to be well-tolerated by the biosynthetic pathway of sialoglycoconjugates.

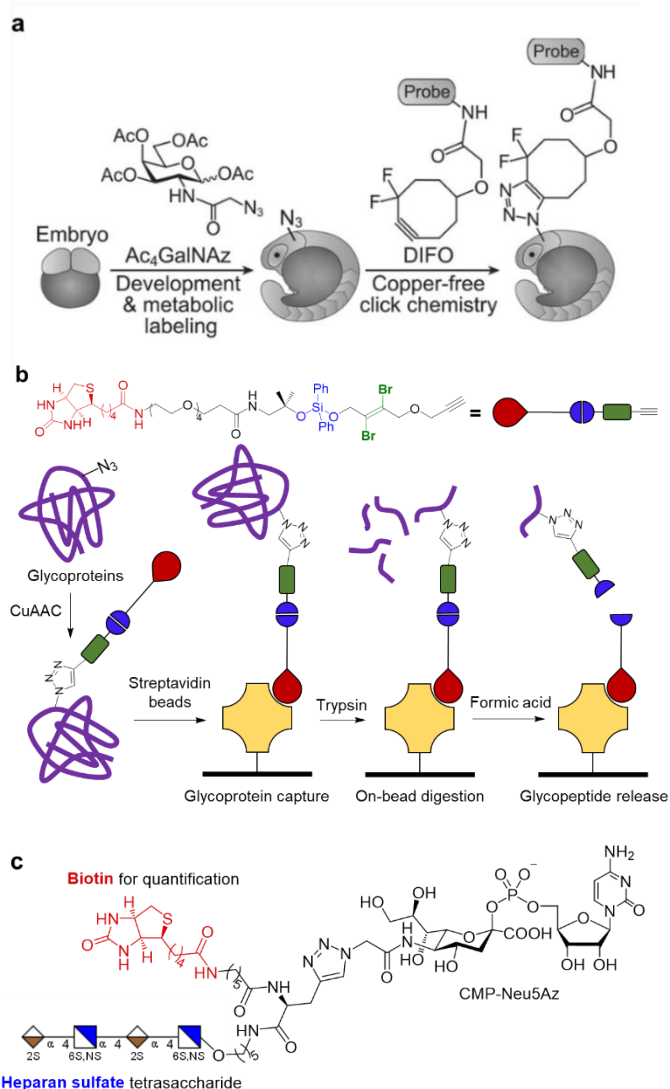


Figure 11 | Applications of bioorthogonally tagged sialic acid. a, Metabolic incorporation of azido sialic acid into complex glycans and *in vivo* labeling in zebra fish. **b**, Diagram of Isotope-targeted glycoproteomics (IsoTaG) workflow. Abbreviation: CuAAC, copper(I)-catalyzed alkyne-azide cycloaddition. **c**, CMP-Sia derivative for one-step selective exo-enzymatic labeling (SEEL).

Chapter 1

To the other half of a biorthogonal pair, a large variety of functionalities can be introduced. For example, fluorescent group conjugation via copper(I)-catalyzed or copper-free cycloaddition allows for tracking of sialoglycoconjugates not only on live cells but also *in vivo* at different developmental stages of zebra fish and mice (Fig. 11a)^{108, 109}. Alternatively, after metabolic incorporation of C-5 azido-Sia, or Neu5Az, into sialoglycoconjugates, a multifunctional linker can be clicked onto glycoproteins in cell lysate, which includes a biotin moiety for affinity enrichment; an acid labile silyl ether-based spacer in between biotin and Sia for glycopeptide release after on-bead digestion; and finally, a dibromo mass tag released together with the glycopeptides, enabling ion selection in mass spectrometry for peptide sequencing (Fig. 11b)¹¹⁰. A technology developed by Boons group termed selective exo-enzymatic labeling (SEEL) for cell surface glycoengineering involves a click pre-conjugation of the desired foreign functionality to C5 position of synthetic CMP-Neu5Az (Fig. 11c)^{111, 112}. Then in a one-step manner, endogenous Sia was firstly cleaved by *C. perfringens* sialidase to make glycosyl acceptors available to the modified CMP-Sia which was transferred together with the foreign functionality by ST6Gal1 onto glycoconjugates. Once attached, it cannot be cleaved by the sialidase, and therefore can retain on cell surface sufficiently long for functional studies. This method allows for gain of function experiments to study specific biological processes for which genetic manipulation remains difficult. For example, a heparan sulfate (HS) tetrasaccharide together with a biotin was clicked onto CMP-Neu5Az. The conjugate is well tolerated by ST6Gal1 and can be efficiently transferred to the *in situ* generated glycosyl acceptors. The introduction of the chimeric molecule to cell surface of a HS-deficient cell line restored formation of a ternary complex composed of fibroblast growth factor II (FGF-II), its cognate receptor, FGFR, and HS, which triggered cell proliferation.

Sialic acid derivatives as lead compounds in medicinal chemistry. Protein-carbohydrate interactions are generally weak, of which the dissociation constant ranges from micromolar to millimolar. The interaction is characterized by high solvent exposure, complex hydrogen bonding network between the hydroxyl and polar groups and limited hydrophobic interaction of sugar backbone C-H with non-polar residues in the binding pocket. Thus, strategies have been developed to enhance the affinity of ligands for protein receptors by introducing artificial hydrophobic moieties to carbohydrate skeleton to make additional contact with the amino acid residues in close proximity to the binding pocket. Knowles group discovered that by introducing a naphthyl and a dansyl group spaced with aliphatic chains to anomeric center and C4 of sialic acid, respectively (Fig. 12a)¹¹³, the affinity of the monosaccharide for hemagglutinin of influenza A virus was enhanced 10 times, thanks to hydrophobic interaction of the two functionalities with the surrounding hydrophobic residues. Zaccai et al. found that sialosides carrying aromatic groups via amide linkage at C9-position exhibited a dramatically increased binding to Siglecs-2 (Fig. 12b)¹¹⁴. It was later found out that this methodology can be extended to other members of the Siglec family, including Siglec-1, -3, -4, -5, -7, -8, -9 and -10 (reviewed in^{115, 116}). The functional group of choice can be screened in a high-throughput manner on-chip¹¹⁷ or *in cellulo*, demonstrated by Paulson

lab and Boltje lab¹¹⁸, respectively. Remarkably, high affinity and selective ligands for specific lectins make it possible to modulate specific biological pathways and provide lead compounds as potential therapeutic agents for disease intervention.

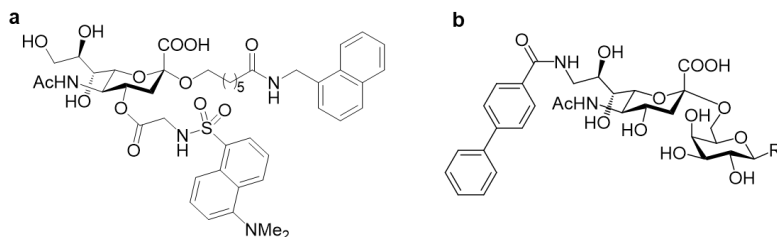


Figure 12 | Hydrophobic group incorporation into sialic acid. **a**, Potent monovalent blocker of influenza A virus hemagglutinin (HA). **b**, C9-Biphenyl derivative of sialoside as high affinity ligand for human Siglec-2.

Capturing carbohydrate-protein complexes. Affinity purification Sia-binding proteins for downstream analysis has been difficult owing to the low-affinity binding. To overcome this, photoaffinity tags such as diazirine¹¹⁹ and aryl azide²⁴ have been introduced to C5 or C9 positions of Sia, respectively. These photo-Sias are metabolically incorporated to cell surface glycans. Upon UV-irradiation, photoaffinity groups becomes highly reactive and non-selectively insert to C-H, N-H or O-H bond of the neighboring residues, thus forming a covalent bond between the sialylated glycans and the interacting partners, which can be subjected to proteomic identification.

Modulation of glycan biosynthesis. Non-natively modified sialic acid or its six-carbon precursor—*N*-acyl mannosamine—can be used in reversible biosynthetic modulation of specific type of glycoconjugates. An early work from Bertozzi group reported that by treating NT2 neurons with per-acetylated *N*-butanoyl mannosamine, which differs from the *N*-acetyl counterpart by two methylene groups, the polysialic acid chain on the carrier protein—NCAM—was dramatically truncated¹²⁰. Although *N*-butanoyl mannosamine readily enters the biosynthetic pathway of CMP-Sia to produce CMP-*N*-butanoyl neuraminic acid. The sialyl donor, however, is not efficiently transferred by ST8Sia2, the Sia polymerizing enzyme, and when the glycoconjugate is primed with *N*-butanoyl neuraminic acid, the enzymatic activity of ST8Sia2 on this glycosyl acceptor drops substantially, thus terminating polysialic acid synthesis. C3-fluorinated Sia has been reported to substantially and globally decrease sialylation of treated cells upon conversion to CMP-activated form and via feedback inhibition of an upstream biosynthetic enzyme in the sialic acid salvage pathway, UDP-GlcNAc 2-epimerase (GNE)^{121, 122}. These reagents represent useful tools for reversible modulation of glycan biosynthesis and is complementary to genetic manipulation which produces a permanent, irreversible effect on the cells.

Chapter 1

Augmenting immunogenicity of carbohydrate antigens. Glycans such as sialyl Tn and sialyl T antigens associated with cancer are generally weakly immunogenic, partly due to the toleration mechanism of host immune system towards these “self” molecules^{123, 124}. Likewise, several pathogenic bacteria express capsular polysaccharides which are structurally similar to host sialoglycans, and thereby camouflage themselves from host immune detection^{15, 19, 125}. Vaccines based on cancer-associated or bacteria-derived sialylated carbohydrate antigens generally elicit moderate to low level of glycan-specific humoral responses with limited potency, which may not be sufficient to offer protection.

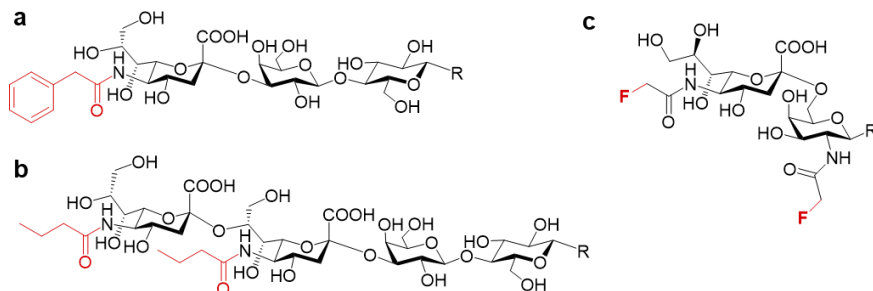


Figure 14 | Non-self or self-like moieties introduced to sialoside-based immunogens. **a** and **b**, Sialate-5-*N*-modified GM3 (**a**) and GD3 (**b**) carbohydrate antigens. Constructs containing these two sugars did not generate antibodies cross-reactive to their corresponding native structures. **c**, Fluorinated sialyl Tn-based immunogen that induced higher titer antibodies (compared with native sialyl Tn) cross-reactive to native structures. R denotes linker-carrier proteins (for **a** and **b**) or MUC1 peptide (**c**). “Non-self” or “self-like” moieties are highlighted in red.

Aiming to circumvent this issue, several laboratories undertook a strategy based on introducing “non-self” functionalities to C5 position of Sia to improve immunorecognition (Fig. 14). Because the glycans carrying these non-native moieties are broadly similar to the native counterparts, the antibodies raised against such designer antigens were expected to cross-react with the natural sugars. However, this approach has led to only limited success in the case of Sia-containing immunogens, with antibody responses directed towards non-self-antigens failing to cross-react with the native counterparts^{126, 127}; while in a few other studies, when fluorine atom was introduced to Sia-C5 and Gal-C2 acetamido groups of sialyl Tn portion to create “self-like” antigens, the immunogenicity was augmented significantly compared with the native sialyl Tn immunogen, and the antisera of the former was reactive to the native structures¹²⁸. These results suggest that there appears to be a narrow window for modified self-carbohydrates to be fairly immunogenic and in the meantime capable of inducing cross-reactive humoral responses to the natural structures. Future research into Sia-based immunogens may be guided by structural biology of available Sia-specific antibodies and incorporate structure-informed rational design.

PART 3. Outline of This Thesis

The transmission of viruses from animal reservoirs to humans poses great threats to public health. Preparedness for future zoonotic outbreaks requires a fundamental understanding how viruses of animal origin have adapted to binding to a cell surface component/receptor of the new host. In **Chapter 2**, we describe such specificities of human and animal viruses that engage with *O*-acetylated sialic acid, including betacoronaviruses, toroviruses, and influenza C and D viruses. Key to these studies was the development of a chemo-enzymatic methodology that can provide almost any sialate-acetylation pattern. The unprecedented collection of *O*-acetylated sialoglycans was printed as a microarray for receptor specificity determination. It showed host-specific patterns of receptor recognition and revealed that three distinct human respiratory viruses uniquely bind 9-*O*-acetylated α 2,8-linked disialoside. Immunofluorescence and cell entry studies support that such glycotope as part of a ganglioside is a functional receptor for human coronaviruses.

In **Chapter 3**, the coronaviral HE-mediated regioselective de-*O*-acetylation was extended to the preparation of a panel of *O*-acetylated *N*-glycolylneuraminic acid oligosaccharides. The resulted compound library was printed on streptavidin-coated glass slides to give a glycan microarray to investigate receptor binding specificities of viral envelope glycoproteins, including spike proteins and HEs, from animal and human coronaviruses. We have found that the binding pattern of these viral proteins to *N*-glycolylated sialosides differ considerable from that to *N*-acetylated counterparts; and that the ability of the spike proteins to tolerated *N*-glycolyl modification varies among viruses targeting different hosts, whereas the corresponding HEs showed a substantial decrease or loss of binding to *N*-glycolylated sialosides. These results suggest that coronaviruses have adjusted their receptor fine specificities to adapt to the sialoglycome of the corresponding host species.

The molecular basis by which viruses recognize specific *O*-acetylated sialosides is poorly understood, and it is unknown how viruses have evolved to recognize specific *O*-acetylated sialosides expressed by their host. In **Chapter 4**, we describe a chemoenzymatic approach that can readily provide sialoglycan analogs in which acetyl esters at C4 and/or C7 are replaced by stabilizing acetamide moieties. The analogs and their natural counterparts were used to examine ligand requirements of the lectin domain of coronaviral hemagglutinin-esterases (HEs). It revealed that HEs from viruses targeting different host species exhibit different requirements for 7-*O*-acetylation. It also showed that ester-to-amide perturbation results in decreased or loss of binding. STD-NMR and molecular modeling of the complexes of the HE of BCoV with the acetamido analogs and natural counterparts revealed that binding is governed by the complementarity between acetyl moieties of the sialosides and hydrophobic patches of the lectin. The precise spatial arrangement of these elements is important, and small perturbations such as *O*-to-*N* change results in substantial loss of binding. Molecular Dynamics simulations with HEs from coronaviruses infecting other

Chapter 1

species indicate that these viruses have adapted their HE specificities by the incorporation of hydrophobic or hydrophilic elements to modulates acetyl ester recognition.

Induction of broadly and durably protective response has been the ultimate goal of prophylactic vaccine development. The use of toll-like receptor (TLR) agonists has been increasingly pursued to augment such immune responses. Furthermore, attachment of these immunomodulating agents to immunogens offers several advantages over combined administration thereof as separate entities. However, there is a lack of general method to conjugate TLR agonists to complex immunogens, such as viral glycoproteins, without compromising important neutralization epitopes. In **Chapter 5** we describe a one-step enzymatic conjugation approach which allows for covalent attachment of small molecule immune activators to asparagine (*N*)-linked glycans of a native-like HIV envelope trimer immunogen. The drug-to-protein ratio can be readily quantified chromatographically. Binding studies with several broadly neutralizing antibodies (bNAbs) reveal that the glycan-targeted covalent modification of the glycoprotein immunogen minimally affects its neutralization epitopes.

References

1. Möckl L. The Emerging Role of the Mammalian Glycocalyx in Functional Membrane Organization and Immune System Regulation. *Front Cell Dev Biol* 2020, **8**: 253.
2. Sicard J-F, Le Bihan G, Vogeleeer P, Jacques M, Harel J. Interactions of Intestinal Bacteria with Components of the Intestinal Mucus. *Front Cell Infect Microbiol* 2017, **7**: 387.
3. Ströh LJ, Stehle T. Glycan Engagement by Viruses: Receptor Switches and Specificity. *Annu Rev Virol* 2014, **1**(1): 285-306.
4. Schauer R, Kamerling JP. Exploration of the sialic acid world. *Adv Carbohydr Chem Biochem* 2018, **75**: 1-213.
5. Varki A. Nothing in glycobiology makes sense, except in the light of evolution. *Cell* 2006, **126**(5): 841-845.
6. Corfield AP, Berry M. Glycan variation and evolution in the eukaryotes. *Trends Biochem Sci* 2015, **40**(7): 351-359.
7. Robinson LS, Lewis WG, Lewis AL. The sialate *O*-acetyltransferase EstA from gut Bacteroidetes species enables sialidase-mediated cross-species foraging of 9-*O*-acetylated sialoglycans. *J Biol Chem* 2017, **292**(28): 11861-11872.
8. Schauer R, Srinivasan GV, Wipfler D, Kniep B, Schwartz-Albiez R. *O*-Acetylated sialic acids and their role in immune defense. *Adv Exp Med Biol* 2011, **705**: 525-548.
9. Cohen M, Varki A. The Sialome—Far More Than the Sum of Its Parts. *OMICS: J Integrative Biol* 2010, **14**(4): 455-464.
10. Angata T, Varki A. Chemical Diversity in the Sialic Acids and Related α -Keto Acids: An Evolutionary Perspective. *Chem Rev* 2002, **102**(2): 439-470.
11. Moremen KW, Tiemeyer M, Nairn AV. Vertebrate protein glycosylation: diversity, synthesis and function. *Nat Rev Mol Cell Biol* 2012, **13**(7): 448-462.
12. Lee S, Roland S. The Biosynthesis of *N*-Glycolylneuraminic Acid Occurs by Hydroxylation of the CMP-Glycoside of *N*-Acetylneuraminic Acid. *Biol Chem* 1988, **369**(1): 477-486.
13. Baumann AM, Bakkers MJ, Buettner FF, Hartmann M, Grove M, Langereis MA, *et al.* 9-*O*-Acetylation of sialic acids is catalysed by CASD1 via a covalent acetyl-enzyme intermediate. *Nat Commun* 2015, **6**: 7673.
14. Macauley MS, Crocker PR, Paulson JC. Siglec-mediated regulation of immune cell function in disease. *Nat Rev Immunol* 2014, **14**(10): 653-666.
15. Mühlenhoff M, Rollenhagen M, Werneburg S, Gerardy-Schahn R, Hildebrandt H. Polysialic Acid: Versatile Modification of NCAM, SynCAM 1 and Neuropilin-2. *Neurochem Res* 2013, **38**(6): 1134-1143.
16. Volkens G, Worrall LJ, Kwan DH, Yu C-C, Baumann L, Lameignere E, *et al.* Structure of human ST8SiaIII sialyltransferase provides insight into cell-surface polysialylation. *Nat Struct Mol Biol* 2015, **22**(8): 627-635.
17. Kiermaier E, Moussion C, Veldkamp CT, Gerardy-Schahn R, de Vries I, Williams LG, *et al.* Polysialylation controls dendritic cell trafficking by regulating chemokine recognition. *Science* 2016, **351**(6269): 186.
18. Elkashef SM, Allison SJ, Sadiq M, Basheer HA, Ribeiro Morais G, Loadman PM, *et al.* Polysialic acid sustains cancer cell survival and migratory capacity in a hypoxic environment. *Sci Rep* 2016, **6**(1): 33026.
19. King MR, Steenbergen SM, Vimr ER. Going for baroque at the *Escherichia coli* K1 cell surface. *Trends Microbiol* 2007, **15**(5): 196-202.

Chapter 1

20. Martínez-Sáez N, Peregrina JM, Corzana F. Principles of mucin structure: implications for the rational design of cancer vaccines derived from MUC1-glycopeptides. *Chem Soc Rev* 2017, **46**(23): 7154-7175.
21. Shurer CR, Kuo JC-H, Roberts LM, Gandhi JG, Colville MJ, Enoki TA, *et al.* Physical Principles of Membrane Shape Regulation by the Glycocalyx. *Cell* 2019, **177**(7): 1757-1770.e1721.
22. Bergstrom K, Shan X, Casero D, Batushansky A, Lagishetty V, Jacobs JP, *et al.* Proximal colon–derived O-glycosylated mucus encapsulates and modulates the microbiota. *Science* 2020, **370**(6515): 467.
23. Linden SK, Sutton P, Karlsson NG, Korolik V, McGuckin MA. Mucins in the mucosal barrier to infection. *Mucosal Immunol* 2008, **1**(3): 183-197.
24. Han S, Collins BE, Bengtson P, Paulson JC. Homomultimeric complexes of CD22 in B cells revealed by protein-glycan cross-linking. *Nat Chem Biol* 2005, **1**(2): 93-97.
25. Ramya TNC, Weerapana E, Liao L, Zeng Y, Tateno H, Liao L, *et al.* In Situ trans Ligands of CD22 Identified by Glycan-Protein Photocross-linking-enabled Proteomics. *Mol Cell Proteomics* 2010, **9**(6): 1339.
26. Barkal AA, Brewer RE, Markovic M, Kowarsky M, Barkal SA, Zaro BW, *et al.* CD24 signalling through macrophage Siglec-10 is a target for cancer immunotherapy. *Nature* 2019, **572**(7769): 392-396.
27. Wasik BR, Barnard KN, Ossiboff RJ, Khedri Z, Feng KH, Yu H, *et al.* Distribution of O-Acetylated Sialic Acids among Target Host Tissues for Influenza Virus. *mSphere* 2017, **2**(5): e00379-00316.
28. Langereis MA, Bakkers MJ, Deng L, Padler-Karavani V, Vervoort SJ, Hulswit RJ, *et al.* Complexity and diversity of the mammalian sialome revealed by nidovirus virolectins. *Cell Rep* 2015, **11**(12): 1966-1978.
29. Barnard KN, Alford-Lawrence BK, Buchholz DW, Wasik BR, LaClair JR, Yu H, *et al.* Modified Sialic Acids on Mucus and Erythrocytes Inhibit Influenza A Virus Hemagglutinin and Neuraminidase Functions. *J Virol* 2020, **94**(9): e01567-01519.
30. Cariappa A, Takematsu H, Liu HY, Diaz S, Haider K, Boboila C, *et al.* B cell antigen receptor signal strength and peripheral B cell development are regulated by a 9-O-acetyl sialic acid esterase. *J Exp Med* 2009, **206**(1): 125-138.
31. Miyagi T, Yamaguchi K. Mammalian sialidases: Physiological and pathological roles in cellular functions. *Glycobiology* 2012, **22**(7): 880-896.
32. Hunter CD, Khanna N, Richards MR, Rezaei Darestani R, Zou C, Klassen JS, *et al.* Human Neuraminidase Isoenzymes Show Variable Activities for 9-O-Acetyl-sialoside Substrates. *ACS Chem Biol* 2018, **13**(4): 922-932.
33. Parameswaran R, Lim M, Arutyunyan A, Abdel-Azim H, Hurtz C, Lau K, *et al.* O-acetylated N-acetylneuraminic acid as a novel target for therapy in human pre-B acute lymphoblastic leukemia. *J Exp Med* 2013, **210**(4): 805-819.
34. Ravindranaths MH, Paulson JC, Irie RF. Human melanoma antigen O-acetylated ganglioside GD3 is recognized by Cancer antennarius lectin. *J Biol Chem* 1988, **263**(4): 2079-2086.
35. Cavdarli S, Delannoy P, Groux-Degroote S. O-acetylated Gangliosides as Targets for Cancer Immunotherapy. *Cells* 2020, **9**(3).
36. Malisan F, Franchi L, Tomassini B, Ventura N, Condò I, Rippon MR, *et al.* Acetylation Suppresses the Proapoptotic Activity of GD3 Ganglioside. *J Exp Med* 2002, **196**(12): 1535-1541.
37. De Maria R, Lenti L, Malisan F, Agostino F, Tomassini B, Zeuner A, *et al.* Requirement for GD3 Ganglioside in CD95- and Ceramide-Induced Apoptosis. *Science* 1997, **277**(5332): 1652.

38. García-Ruiz C, Colell A, Morales A, Calvo Ma, Enrich C, Fernández-Checa JC. Trafficking of Ganglioside GD3 to Mitochondria by Tumor Necrosis Factor- α . *J Biol Chem* 2002, **277**(39): 36443-36448.
39. Brenner C, Kniep B, Maillier E, Martel C, Franke C, Röber N, *et al.* GD3-7-aldehyde is an apoptosis inducer and interacts with adenine nucleotide translocase. *Biochem Biophys Res Commun* 2010, **391**(1): 248-253.
40. Coronaviridae Study Group of the International Committee on Taxonomy of V. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol* 2020, **5**(4): 536-544.
41. Su S, Wong G, Shi W, Liu J, Lai ACK, Zhou J, *et al.* Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses. *Trends Microbiol* 2016, **24**(6): 490-502.
42. Li F. Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annu Rev Virol* 2016, **3**(1): 237-261.
43. Raj VS, Mou H, Smits SL, Dekkers DHW, Müller MA, Dijkman R, *et al.* Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. *Nature* 2013, **495**(7440): 251-254.
44. Li W, Hulswit RJG, Widjaja I, Raj VS, McBride R, Peng W, *et al.* Identification of sialic acid-binding function for the Middle East respiratory syndrome coronavirus spike glycoprotein. *Proc Natl Acad Sci USA* 2017, **114**(40): E8508.
45. Wickramasinghe INA, de Vries RP, Gröne A, de Haan CAM, Verheije MH. Binding of Avian Coronavirus Spike Proteins to Host Factors Reflects Virus Tropism and Pathogenicity. *J Virol* 2011, **85**(17): 8903.
46. Zhang Y, Han L, Xia L, Yuan Y, Hu H. Assessment of hemagglutination activity of porcine deltacoronavirus. *J Vet Sci* 2020, **21**(1).
47. Liu C, Tang J, Ma Y, Liang X, Yang Y, Peng G, *et al.* Receptor Usage and Cell Entry of Porcine Epidemic Diarrhea Coronavirus. *J Virol* 2015, **89**(11): 6121.
48. Hulswit RJG, Lang YF, Bakkers MJG, Li WT, Li ZS, Schouten A, *et al.* Human coronaviruses OC43 and HKU1 bind to 9-O-acetylated sialic acids via a conserved receptor-binding site in spike protein domain A. *Proceedings of the National Academy of Sciences of the United States of America* 2019, **116**(7): 2681-2690.
49. Tortorici MA, Walls AC, Lang Y, Wang C, Li Z, Koerhuis D, *et al.* Structural basis for human coronavirus attachment to sialic acid receptors. *Nat Struct Mol Biol* 2019, **26**(6): 481-489.
50. Williams RK, Jiang GS, Holmes KV. Receptor for mouse hepatitis virus is a member of the carcinoembryonic antigen family of glycoproteins. *Proc Natl Acad Sci U S A* 1991, **88**(13): 5533-5536.
51. Lang Y, Li W, Li Z, Koerhuis D, van den Burg ACS, Rozemuller E, *et al.* Coronavirus hemagglutinin-esterase and spike proteins coevolve for functional balance and optimal virion avidity. *Proc Natl Acad Sci USA* 2020: 202006299.
52. Song H, Qi J, Khedri Z, Diaz S, Yu H, Chen X, *et al.* An Open Receptor-Binding Cavity of Hemagglutinin-Esterase-Fusion Glycoprotein from Newly-Identified Influenza D Virus: Basis for Its Broad Cell Tropism. *PLoS Pathog* 2016, **12**(1): e1005411.
53. Regl G, Kaser A, Iwersen M, Schmid H, Kohla G, Strobl B, *et al.* The Hemagglutinin-Esterase of Mouse Hepatitis Virus Strain S Is a Sialate-4-O-Acetylerase. *J Virol* 1999, **73**(6): 4721.
54. Schultze B, Gross HJ, Brossmer R, Herrler G. The S Protein of Bovine Coronavirus Is a Hemagglutinin Recognizing 9-O-Acetylated Sialic Acid as a Receptor Determinant. *J Virol* 1991, **65**(11): 6232-6237.

Chapter 1

55. Boons G-J, Demchenko AV. Recent Advances in *O*-Sialylation. *Chem Rev* 2000, **200**: 4539–4565.
56. Ito Y, Ogawa T. An efficient approach to stereoselective glycosylation of N-acetylneuraminic acid: Used of phenylselenenyl group as a stereocontrolling auxilliary. *Tetrahedron Lett* 1987, **28**(49): 6221-6224.
57. Ito Y, Ogawa T. Highly stereoselective glycosylation of N-acetylneuraminic acid aided by a phenylthio substituent as a stereocontrolling auxilliary. *Tetrahedron Lett* 1988, **29**(32): 3987-3990.
58. Martichonok V, Whitesides GM. A Practical Method for the Synthesis of Sialyl α -Glycosides. *J Am Chem Soc* 1996, **118**(35): 8187-8191.
59. Castro-Palomino JC, Tsvetkov YE, Schmidt RR. 8-O-Sialylation of Neuraminic Acid. *J Am Chem Soc* 1998, **120**(22): 5434-5440.
60. Takahashi T, Tsukamoto H, Yamada H. A new method for the formation of the α -glycoside bond of sialyl conjugates based on long-range participation. *Tetrahedron Lett* 1997, **38**(47): 8223-8226.
61. Haberman JM, Gin DY. A New C(1)-Auxiliary for Anomeric Stereocontrol in the Synthesis of α -Sialyl Glycosides. *Org Lett* 2001, **3**(11): 1665-1668.
62. Chen J, Hansen T, Zhang Q-J, Liu D-Y, Sun Y, Yan H, *et al.* 1-Picolinyl-5-azido Thiosialosides: Versatile Donors for the Stereoselective Construction of Sialyl Linkages. *Angew Chem Int Ed* 2019, **58**(47): 17000-17008.
63. Hanashima S, Akai S, Sato K-i. Thioester-assisted α -sialylation reaction. *Tetrahedron Lett* 2008, **49**(34): 5111-5114.
64. Demchenko AV, Boons G-J. A novel and versatile glycosyl donor for the preparation of glycosides of N-acetylneuraminic acid. *Tetrahedron Lett* 1998, **39**(19): 3065-3068.
65. Meo CD, Demchenko AV, Boons G-J. A Stereoselective Approach for the Synthesis of α -Sialosides. *The Journal of Organic Chemistry* 2001, **66**(16): 5490-5497.
66. Yu C-S, Niikura K, Lin C-C, Wong C-H. The Thioglycoside and Glycosyl Phosphite of 5-Azido Sialic Acid: Excellent Donors for the α -Glycosylation of Primary Hydroxy Groups. *Angew Chem Int Ed* 2001, **40**(15): 2900-2903.
67. Ando H, Koike Y, Ishida H, Kiso M. Extending the possibility of an N-Troc-protected sialic acid donor toward variant sialo-glycoside synthesis. *Tetrahedron Lett* 2003, **44**(36): 6883-6886.
68. Tanaka K, Goi T, Fukase K. Highly Efficient Sialylation towards α (2-3)- and α (2-6)-Neu5Ac-Gal Synthesis: Significant ‘Fixed Dipole Effect’ of N-Phthalyl Group on α -Selectivity. *Synlett* 2005, **2005**(19): 2958-2962.
69. Mandhapati AR, Rajender S, Shaw J, Crich D. The isothiocyanato moiety: an ideal protecting group for the stereoselective synthesis of sialic acid glycosides and subsequent diversification. *Angew Chem Int Ed Engl* 2015, **54**(4): 1275-1278.
70. Tanaka H, Nishiura Y, Takahashi T. Stereoselective synthesis of oligo- α -(2,8)-sialic acids. *J Am Chem Soc* 2006, **128**(22): 7124-7125.
71. Crich D, Li W. *O*-Sialylation with *N*-acetyl-5-*N*,4-*O*-carbonyl-protected thiosialoside donors in dichloromethane: facile and selective cleavage of the oxazolidinone ring. *J Org Chem* 2007, **72**(7): 2387-2391.
72. Kancharla PK, Navuluri C, Crich D. Dissecting the Influence of Oxazolidinones and Cyclic Carbonates in Sialic Acid Chemistry. *Angew Chem Int Ed* 2012, **51**(44): 11105-11109.
73. Tanaka H, Nishiura Y, Takahashi T. An Efficient Convergent Synthesis of GP1c Ganglioside Epitope. *Journal of American Chemical Society* 2008, **130**(51): 17244-17245.

74. Chuang HY, Ren CT, Chao CA, Wu CY, Shivatare SS, Cheng TJ, *et al.* Synthesis and vaccine evaluation of the tumor-associated carbohydrate antigen RM2 from prostate cancer. *J Am Chem Soc* 2013, **135**(30): 11140-11150.
75. Aoyagi T, Ohira S, Fuse S, Uzawa J, Yamaguchi Y, Tanaka H. The α -Glycosidation of Partially Unprotected N-Acetyl and N-Glycolyl Sialyl Donors in the Absence of a Nitrile Solvent Effect. *Chem Eur J* 2016, **22**(20): 6968-6973.
76. Wu Y-F, Tsai Y-F. Assistance of the C-7,8-Picoloyl Moiety for Directing the Glycosyl Acceptors into the α -Orientation for the Glycosylation of Sialyl Donors. *Org Lett* 2017, **19**(16): 4171-4174.
77. Komura N, Kato K, Udagawa T, Asano S, Tanaka H-N, Imamura A, *et al.* Constrained sialic acid donors enable selective synthesis of α -glycosides. *Science* 2019, **364**(6441): 677-680.
78. Simon ES, Bednarski MD, Whitesides GM. Synthesis of CMP-NeuAc from N-acetylglucosamine: generation of CTP from CMP using adenylate kinase. *J Am Chem Soc* 1988, **110**(21): 7159-7163.
79. Ichikawa Y, Shen GJ, Wong CH. Enzyme-catalyzed synthesis of sialyl oligosaccharide with in situ regeneration of CMP-sialic acid. *J Am Chem Soc* 1991, **113**(12): 4698-4700.
80. Yu H, Chokhawala H, Karpel R, Yu H, Wu B, Zhang J, *et al.* A Multifunctional Pasteurella multocida Sialyltransferase: A Powerful Tool for the Synthesis of Sialoside Libraries. *J Am Chem Soc* 2005, **127**: 17618-17619.
81. Yu H, Huang S, Chokhawala H, Sun M, Zheng H, Chen X. Highly efficient chemoenzymatic synthesis of naturally occurring and non-natural α -2,6-linked sialosides: a P. damsela α -2,6-sialyltransferase with extremely flexible donor-substrate specificity. *Angew Chem Int Ed Engl* 2006, **45**(24): 3938-3944.
82. Yu H, Cheng JS, Ding L, Khedri Z, Chen Y, Chin S, *et al.* Chemoenzymatic Synthesis of GD3 Oligosaccharides and Other Disialyl Glycans Containing Natural and Non-natural Sialic Acids. *J Am Chem Soc* 2009, **131**(51): 18467-18477.
83. Chen X, Varki A. Advances in the Biology and Chemistry of Sialic Acids. *ACS Chem Biol* 2010, **5**(2): 163-176.
84. Meng X, Yao W, Cheng J, Zhang X, Jin L, Yu H, *et al.* Regioselective Chemoenzymatic Synthesis of Ganglioside Disialyl Tetrasaccharide Epitopes. *J Am Chem Soc* 2014, **136**(14): 5205-5208.
85. Lu N, Ye J, Cheng J, Sasmal A, Liu C-C, Yao W, *et al.* Redox-Controlled Site-Specific α 2-6-Sialylation. *J Am Chem Soc* 2019, **141**(11): 4547-4552.
86. Liu L, Prudden AR, Capicciotti CJ, Bosman GP, Yang JY, Chapla DG, *et al.* Streamlining the chemoenzymatic synthesis of complex N-glycans by a stop and go strategy. *Nat Chem* 2019, **11**(2): 161-169.
87. Haverkamp J, Schauer R, Wember M, Kamerling JP, Vliegienthart JF. Synthesis of 9-O-acetyl- and 4,9-di-O-acetyl derivatives of the methyl ester of N-acetyl- β -d-neuraminic acid methylglycoside. Their use as models in periodate oxidation studies. *Hoppe-Seyler's Zeitschrift fur Physiologische Chemie*, vol. 356. Interactive Factory, 1975, pp 1575-1583.
88. Anazawa K, Furuhashi K, Ogura H. Synthesis of 7-O-Acetyl-N-Acetylneuraminic Acid-Derivative. *Chem Pharm Bull (Tokyo)* 1988, **36**(12): 4976-4979.
89. Furuhashi K, Ogura H. Studies on sialic acids .19. Syntheses of partially O-acetylated 4-methylcoumarin-7-yl 5-acetamido-3,5-dideoxy- α -D-glycero-D-galacto-2-nonulopyranosidonic acids. *Chem Pharm Bull (Tokyo)* 1989, **37**(8): 2037-2040.
90. Roth A, Faillard H. Synthesis of Fluorescent 7,8,9-Tri-O-acetyl-N-acetyl- and 4-O-Acetyl-N-acetylneuraminic Acid α -Thioketosides. *Liebigs Ann Chem* 1993, **1993**(5): 485-489.

Chapter 1

91. Clarke PA, Mistry N, Thomas GH. Synthesis of the complete series of mono acetates of N-acetyl-D-neuraminic acid. *Org Biomol Chem* 2012, **10**(3): 529-535.
92. Park SS, Gervay-Hague J. Synthesis of Partially O-Acetylated N-Acetylneuraminic Acid Using Regioselective Silyl Exchange Technology. *Org Lett* 2014, **16**(19): 5044-5047.
93. Yu H, Zeng J, Li Y, Thon V, Shi B, Chen X. Effective one-pot multienzyme (OPME) synthesis of monotreme milk oligosaccharides and other sialosides containing 4-O-acetyl sialic acid. *Org Biomol Chem* 2016.
94. Takayama S, Livingston PO, Wong C-H. Synthesis of the melanoma-associated ganglioside 9-O-acetyl GD3 through regioselective enzymatic acetylation of GD3 using subtilisin. *Tetrahedron Lett* 1996, **37**(52): 9271-9274.
95. Guo H, Rabouw H, Slomp A, Dai M, van der Vegt F, van Lent JWM, *et al.* Kinetic analysis of the influenza A virus HA/NA balance reveals contribution of NA to virus-receptor binding and NA-dependent rolling on receptor-containing surfaces. *PLoS Path* 2018, **14**(8): e1007233.
96. Liang C-F, Yan M-C, Chang T-C, Lin C-C. Synthesis of S-Linked $\alpha(2\rightarrow9)$ Octasialic Acid via Exclusive α S-Glycosidic Bond Formation. *J Am Chem Soc* 2009, **131**(9): 3138-3139.
97. Liang C-F, Kuan T-C, Chang T-C, Lin C-C. Stereoselective Synthesis of S-Linked $\alpha(2\rightarrow8)$ and $\alpha(2\rightarrow8)/\alpha(2\rightarrow9)$ Hexasialic Acids. *J Am Chem Soc* 2012, **134**(38): 16074-16079.
98. Huo C-X, Zheng X-J, Xiao A, Liu C-C, Sun S, Lv Z, *et al.* Synthetic and immunological studies of N-acyl modified S-linked STn derivatives as anticancer vaccine candidates. *Org Biomol Chem* 2015, **13**(12): 3677-3690.
99. Kuan T-C, Wu H-R, Adak AK, Li B-Y, Liang C-F, Hung J-T, *et al.* Synthesis of an S-Linked $\alpha(2\rightarrow8)$ GD3 Antigen and Evaluation of the Immunogenicity of Its Glycoconjugate. *Chem Eur J* 2017, **23**(28): 6876-6887.
100. Suzuki T, Hayashi C, Komura N, Tamai R, Uzawa J, Ogawa J, *et al.* Synthesis and Glycan-Protein Interaction Studies of Se-Sialosides by 77Se NMR. *Org Lett* 2019, **21**(16): 6393-6396.
101. Kuberan B, Sikkander SA, Tomiyama H, Linhardt RJ. Synthesis of a C-Glycoside Analogue of sTn: An HIV- and Tumor-Associated Antigen. *Angew Chem Int Ed* 2003, **42**(18): 2073-2075.
102. He Y, Yang Y, Iyer SS. Neuraminidase Resistant Sialosides for the Detection of Influenza Viruses. *Bioconjugate Chem* 2016, **27**(6): 1509-1517.
103. Kim J-H, Resende R, Wennekes T, Chen H-M, Bance N, Buchini S, *et al.* Mechanism-Based Covalent Neuraminidase Inhibitors with Broad-Spectrum Influenza Antiviral Activity. *Science* 2013, **340**(6128): 71.
104. Lo HJ, Krasnova L, Dey S, Cheng T, Liu H, Tsai TI, *et al.* Synthesis of Sialidase-Resistant Oligosaccharide and Antibody Glycoform Containing $\alpha(2,6)$ -Linked 3F(ax)-Neu5Ac. *J Am Chem Soc* 2019, **141**(16): 6484-6488.
105. Khedri Z, Xiao A, Yu H, Landig CS, Li WQ, Diaz S, *et al.* A chemical biology solution to problems with studying biologically important but unstable 9-O-acetyl sialic acids. *ACS Chem Biol* 2017, **12**(1): 214-224.
106. Devaraj NK. The Future of Bioorthogonal Chemistry. *ACS Central Science* 2018, **4**(8): 952-959.
107. Wratil PR, Horstkorte R, Reutter W. Metabolic Glycoengineering with N-Acyl Side Chain Modified Mannosamines. *Angew Chem Int Ed* 2016, **55**(33): 9482-9512.
108. Prescher JA, Dube DH, Bertozzi CR. Chemical remodelling of cell surfaces in living animals. *Nature* 2004, **430**(7002): 873-877.
109. Laughlin ST, Baskin JM, Amacher SL, Bertozzi CR. In Vivo Imaging of Membrane-Associated Glycans in Developing Zebrafish. *Science* 2008, **320**(5876): 664.

110. Woo CM, Iavarone AT, Spiciarich DR, Palaniappan KK, Bertozzi CR. Isotope-targeted glycoproteomics (IsoTaG): a mass-independent platform for intact N- and O-glycopeptide discovery and analysis. *Nat Methods* 2015, **12**(6): 561-567.
111. Sun T, Yu S-H, Zhao P, Meng L, Moremen KW, Wells L, *et al.* One-Step Selective Exoenzymatic Labeling (SEEL) Strategy for the Biotinylation and Identification of Glycoproteins of Living Cells. *J Am Chem Soc* 2016, **138**(36): 11575-11582.
112. Capicciotti CJ, Zong C, Sheikh MO, Sun T, Wells L, Boons G-J. Cell-Surface Glyco-Engineering by Exogenous Enzymatic Transfer Using a Bifunctional CMP-Neu5Ac Derivative. *J Am Chem Soc* 2017, **139**(38): 13342-13348.
113. Weinhold EG, Knowles JR. Design and evaluation of a tightly binding fluorescent ligand for influenza A hemagglutinin. *J Am Chem Soc* 1992, **114**(24): 9270-9275.
114. Zaccai NR, Maenaka K, Maenaka T, Crocker PR, Brossmer R, Kelm S, *et al.* Structure-Guided Design of Sialic Acid-Based Siglec Inhibitors and Crystallographic Analysis in Complex with Sialoadhesin. *Structure* 2003, **11**(5): 557-567.
115. Büll C, Heise T, Adema GJ, Boltje TJ. Sialic Acid Mimetics to Target the Sialic Acid–Siglec Axis. *Trends Biochem Sci* 2016, **41**(6): 519-531.
116. Movsisyan LD, Macauley MS. Structural advances of Siglecs: insight into synthetic glycan ligands for immunomodulation. *Org Biomol Chem* 2020, **18**(30): 5784-5797.
117. Rillahan CD, Schwartz E, McBride R, Fokin VV, Paulson JC. Click and pick: identification of sialoside analogues for siglec-based cell targeting. *Angew Chem Int Ed Engl* 2012, **51**(44): 11014-11018.
118. Büll C, Heise T, Beurskens DMH, Riemersma M, Ashikov A, Rutjes FPJT, *et al.* Sialic Acid Glycoengineering Using an Unnatural Sialic Acid for the Detection of Sialoglycan Biosynthesis Defects and On-Cell Synthesis of Siglec Ligands. *ACS Chem Biol* 2015, **10**(10): 2353-2363.
119. Tanaka Y, Kohler JJ. Photoactivatable Crosslinking Sugars for Capturing Glycoprotein Interactions. *J Am Chem Soc* 2008, **130**(11): 3278-3279.
120. Mahal LK, Charter NW, Angata K, Fukuda M, Koshland DE, Bertozzi CR. A Small-Molecule Modulator of Poly- α 2,8-Sialic Acid Expression on Cultured Neurons and Tumor Cells. *Science* 2001, **294**(5541): 380.
121. Rillahan CD, Antonopoulos A, Lefort CT, Sonon R, Azadi P, Ley K, *et al.* Global metabolic inhibitors of sialyl- and fucosyltransferases remodel the glycome. *Nat Chem Biol* 2012, **8**(7): 661-668.
122. Hinderlich S, Weidemann W, Yardeni T, Horstkorte R, Huizing M. UDP-GlcNAc 2-Epimerase/ManNAc Kinase (GNE): A Master Regulator of Sialic Acid Synthesis. *Top Curr Chem* 2015, **366**: 97-137.
123. Marcos NT, Pinho S, Grandela C, Cruz A, Samyn-Petit B, Harduin-Lepers A, *et al.* Role of the Human ST6GalNAc-I and ST6GalNAc-II in the Synthesis of the Cancer-Associated Sialyl-Tn Antigen. *Cancer Res* 2004, **64**(19): 7050.
124. Loureiro LR, Carrascal MA, Barbas A, Ramalho JS, Novo C, Delannoy P, *et al.* Challenges in Antibody Development against Tn and Sialyl-Tn Antigens. *Biomolecules* 2015, **5**(3).
125. Kahler CM, Martin LE, Shih GC, Rahman MM, Carlson RW, Stephens DS. The (α 2 \rightarrow 8)-Linked Polysialic Acid Capsule and Lipooligosaccharide Structure Both Contribute to the Ability of Serogroup B *Neisseria meningitidis* To Resist the Bactericidal Activity of Normal Human Serum. *Infect Immun* 1998, **66**(12): 5939.
126. Pan Y, Chefalo P, Nagy N, Harding C, Guo Z. Synthesis and Immunological Properties of N-Modified GM3 Antigens as Therapeutic Cancer Vaccines. *J Med Chem* 2005, **48**(3): 875-883.

Chapter 1

127. Zou W, Borrelli S, Gilbert M, Liu T, Pon RA, Jennings HJ. Bioengineering of Surface GD3 Ganglioside for Immunotargeting Human Melanoma Cells. *J Biol Chem* 2004, **279**(24): 25390-25399.
128. Yang F, Zheng XJ, Huo CX, Wang Y, Zhang Y, Ye XS. Enhancement of the immunogenicity of synthetic carbohydrate vaccines by chemical modifications of STn antigen. *ACS Chem Biol* 2011, **6**(3): 252-25.

CHAPTER 2 | Synthetic *O*-Acetylated Sialosides Facilitate Functional Receptor Identification for Human Respiratory Viruses

Zeshi Li^{1,6}, Yifei Lang^{2,6}, Lin Liu³, Mehman I. Bunyatov¹, Angelic Isaza Sarmiento², Raoul J. de Groot², Geert-Jan Boons^{1,3,4,5}

Affiliations

¹Department of Chemical Biology and Drug Discovery, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, 3584 CG Utrecht, The Netherlands.

²Virology Division, Department of Biomolecular Health Sciences, Faculty of Veterinary Medicine, Utrecht University, 3584 CL Utrecht, the Netherlands.

³Complex Carbohydrate Research Center, University of Georgia, 315 Riverbend Road, Athens, GA 30602, USA.

⁴Bijvoet Center for Biomolecular Research, Utrecht University, Utrecht, The Netherlands.

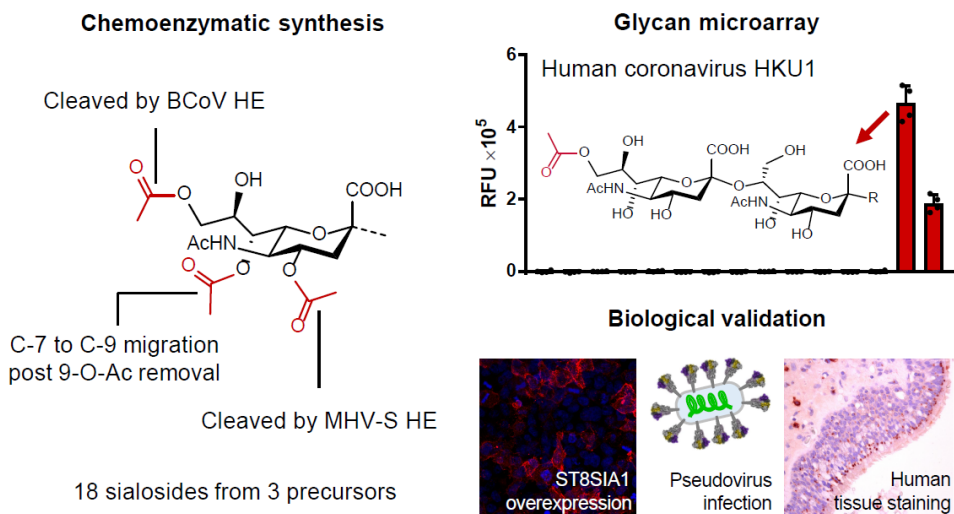
⁵Chemistry Department, University of Georgia, Athens, GA 30602, USA.

⁶These authors contributed equally: Zeshi Li, Yifei Lang.

Chapter 2

Abstract

The transmission of viruses from animal reservoirs to humans poses great threats to public health. Preparedness for future zoonotic outbreaks requires a fundamental understanding how viruses of animal origin have adapted to binding to a cell surface component/receptor of the new host. Here, we describe such specificities of human and animal viruses that engage with *O*-acetylated sialic acid, including betacoronaviruses, toroviruses, and influenza C and D viruses. Key to these studies was the development of a chemo-enzymatic methodology that can provide almost any sialate-acetylation pattern. The unprecedented collection of *O*-acetylated sialoglycans was printed as a microarray for receptor specificity determination. It showed host-specific patterns of receptor recognition and revealed that three distinct human respiratory viruses uniquely bind 9-*O*-acetylated α 2,8-linked disialoside. Immunofluorescence and cell entry studies support that such glycotope as part of a ganglioside is a functional receptor for human coronaviruses.



Introduction

Eukaryotic cells are covered by a dense layer of glycans. These carbohydrates are structurally very diverse and expressed in tissue- and cell-specific manners. During evolution, glycan biosynthesis diverged and as a result, structures of glycans differ considerably even between closely related species. These differences are in part driven by pathogenic pressure to hamper cross species transmission^{1,2}.

Numerous viruses initiate infection by binding to cell surface glycans of the host. The selectivities of viral receptor binding proteins for specific glycan structures critically determine host range, targeted tissues and cells and pathogenesis³. Prototypic are influenza A viruses (IAVs), which have an obligatory dependence for host glycans modified by a terminal sialic acid to infect cells⁴. The hemagglutinin (HA) of avian IAVs preferentially binds sialic acid α 2,3-linked to galactose (Gal), whereas human viruses recognize the α 2,6-linked isomer. The IAVs circulating in humans are of avian origin and acquired an ability to use α 2,6-linked sialosides through mutations in the receptor binding pocket of HA. This adaptation occurred because human upper airway tissue expresses high levels of α 2,6-sialosides whereas duck intestinal mucosa is rich in α 2,3-sialosides.

For many other vertebrate viruses that employ sialoglycans as receptors, it has been difficult to determine receptor binding specificities^{5,6}. This information is critical to establish to what extent differences in sialoglycan repertoire between vertebrate species hamper cross species transmission and how human viruses of zoonotic origin have overcome these barriers to establish infections in humans. Insight into receptor binding specificities is also critical for the development of surveillance, prevention and intervention strategies^{7,8}.

Here, we describe receptor binding specificities of a group of viruses that engage with *O*-acetylated variants of sialic acid, including several betacoronaviruses and toroviruses (subfamily *Orthocoronavirinae* and family *Tobaniviridae*, respectively), and influenza C and D viruses (family *Orthomyxoviridae*), and demonstrate host-specific patterns of receptor recognition relating to both the *O*-acetylation pattern and sialic acid glycosidic linkage type. Our data revealed that viruses adapted to humans underwent convergent evolution to become highly selective for *O*-acetylated forms of α 2,8-linked di-sialylated structures commonly found on glycolipids. Key to these discoveries was the implementation of a chemo-enzymatic methodology that provided a previously inaccessible collection of *O*-acetylated sialoglycans.

N-acetylneuraminic acid (Neu5Ac) and *N*-glycolylneuraminic acid (Neu5Gc) are the most common forms of sialic acid in vertebrates⁹. They occur in a variety of different glycosidic forms and typical ones are α 2,3-linked to Gal, α 2,6-linked to Gal and *N*-acetyl-galactosamine (GalNAc) or *via* an α 2,8-linkage to another sialic acid (Fig. 1a). Further structural diversity arises from the substitution of the C-4, C-7, C-8 and/or C-9 hydroxyls

Chapter 2

by acetyl esters (Fig. 1b). Despite their presence in many different tissues^{10,11}, *O*-acetylated sialic acids are Cinderella molecules¹². There is limited knowledge of their biosynthesis, precluding analysis through genetic manipulation. This class of glycans exhibit inherent chemical labilities and are prone to *O*-acetyl ester migration and hydrolysis, which complicates isolation and characterization. A lack of methods to prepare panels of *O*-acetylated sialosides is another major stumbling block, and reported synthetic approaches are largely limited to monosaccharide derivatives and *C*-9-modified sialosides (see Supplementary Fig. S1a-d for an overview of previously reported synthetic approaches)¹³⁻²³. The synthetic inaccessibility of a complete series of *O*-acetylated sialosides stems from the stringent requirements to select protecting groups for a sialyl donor and acceptor to achieve high-yielding and stereoselective sialylations²⁴. The selected protecting groups of the donor and acceptor also need to be compatible with the installation of different patterns of acetyl esters that have a propensity to migrate or hydrolyze even under mildly basic or acidic conditions resulting in a yet unmet synthetic challenge (Supplementary Fig. S1e).

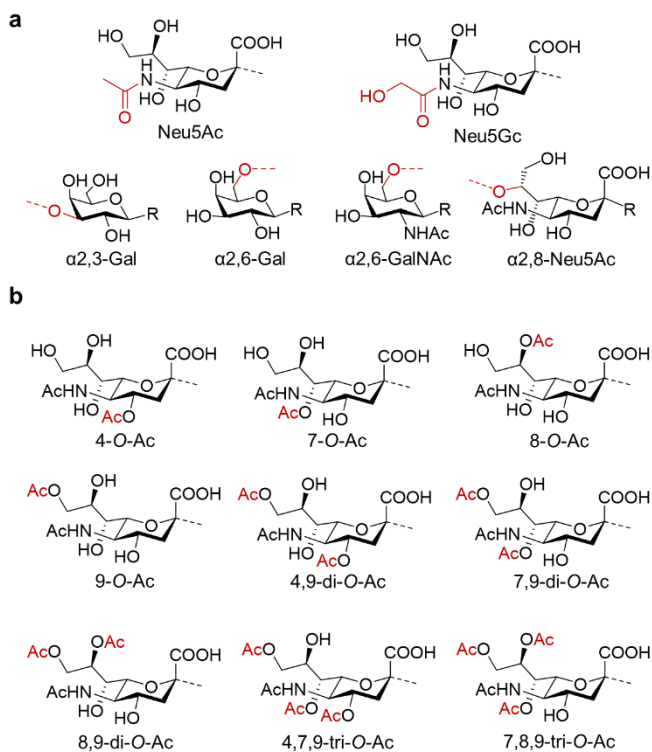
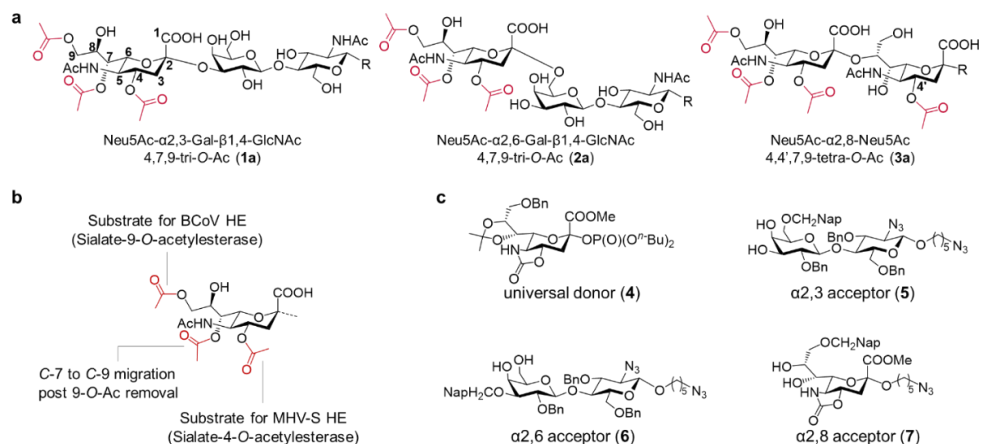


Figure 1 | Structural diversity of sialosides in vertebrates. **a**, Common backbone structures and glycosidic linkages. **b**, Naturally occurring partially *O*-acetylated variants of Neu5Ac⁹. Ac denotes acetyl.

Results and discussion

A general approach to diverse *O*-acetylated sialyl oligosaccharides. To develop a general method for the preparation of *O*-acetylated sialosides, we drew inspiration from the sialate-*O*-acetyltransferase activities of viral hemagglutinin-esterases (HE). These enzymes, which are expressed by various nido- and orthomyxoviruses, cleave acetyl esters of sialic acids with exquisite regioselectivity to facilitate dynamic virion attachment to host cells. The bovine coronavirus (BCoV) HE is a 9-*O*-acetyltransferase with a preference for 7,9-di-*O*-acetyl-sialosides¹⁰, while mouse hepatitis virus strain S (MHV-S) produces HEs specific for sialate-4-*O*-acetyl esters²⁵. We envisaged that diversification of α 2,3-, 2,6- and 2,8-linked sialosides **1a**, **2a** and **3a**, respectively carrying acetyl esters at C-4, C-7 and C-9 (Fig. 2a), by treatment with these enzymes combined with controlled acetyl ester migration from C-7 to C-9²⁶ (Fig. 2b), can provide almost any sialate-acetylation pattern. The attraction of this approach is that from a common oligosaccharide precursor as many as seven sialate-*O*-acetylation patterns can be installed thereby alleviating the need to selected for each target proper protecting groups and establish appropriate glycosylation conditions, which is almost an impossible task to achieve.

The three common precursors **1a**, **2a** and **3a** were prepared from universal donor **4** and glycosyl acceptors **5**, **6** and **7**, respectively (Fig. 2c). The 7,8-isopropylidene of **4** combined with the presence of a 4,5-carbamate ensured that glycosylations proceeded in high yield with exquisite α -anomeric selectivity²⁷. The latter was facilitated by using glycosyl acceptors **5**, **6** and **7** which have highly reactive 3,4-, 4,6- and 7,8-diols that in a regioselective manner can be sialylated to provide α 2,3-, α 2,6- and α 2,8-linked products, respectively. Oxidative acetal ring closure of the 2-naphthylmethyl ether of the glycosylation products protected the remaining alcohol allowing selective manipulation of the hydroxyls of the sialic acid moiety to install the required acetyl esters only at C-4, C-7 and C-9 (see Supplementary Fig. S1f for the analysis of general strategy).



Chapter 2

Figure 2 / A general strategy for synthesizing *O*-acetylated sialosides. a, Chemically synthesized 4,7,9-tri-*O*-acetylated common precursors for HE-mediated regioselective de-*O*-acetylation. The biotinylated anomeric linker is omitted in the structures and is denoted as R. **b,** Methods for chemoenzymatic remodeling of the common precursors. Bovine coronavirus (BCoV) and mouse hepatitis strain S (MHV-S) HEs regio-specifically cleave 9- and 4-*O*-acetyl groups, respectively. **c,** The glycosyl donor and acceptors for chemical assemblies of **1-3a**. See Supplementary Fig. S2 preparation of glycosyl acceptors. Abbreviations: Bn, benzyl; Bu, butyl; Me, methyl; Nap, 2-naphthylmethyl.

Thus, a trimethylsilyl triflate (TMSOTf) mediated glycosylation of **4** with **5** gave a trisaccharide that was immediately subjected to wet trifluoroacetic acid (TFA) to remove the isopropylidene acetal to provide **8** in a yield of 79% over two steps as only the α -anomer (Fig. 3a). 2,3-Dichloro-5,6-dicyano-*p*-benzoquinone (DDQ)-mediated oxidative acetal ring closure of **8** in anhydrous dichloromethane (DCM) gave **9**, which was converted into **10** by a one-pot procedure involving hydrolysis of the methyl ester and 4,5-carbamate of the sialic acid moiety followed by selective acetylation of the resulting amine by reaction with acetyl chloride. Next, the azido moieties of **10** were reduced by Staudinger reaction to give diamine **11**, which was converted into **12** by selective amine manipulations. The free amine and the C-4, C-7 and C-8 hydroxyls of **12** were acetylated using acetic anhydride in pyridine to give **13**. Hydrogenation of the latter compound resulted in the removal of the benzyl ethers and reduction of the azido moiety into an amine. The target trisaccharide **1a** was obtained in high overall yield by controlled migration of the acetyl ester at C-8 to C-9 using CsF-acetic acid buffer in DMSO followed by biotinylation of the amine of the anomeric spacer.

Exposure of **1a** to MHV-S, or BCoV HE in ammonium formate buffer provided the 4,7- and 7,9-di-*O*-acetylated derivatives **1b** and **1d**, respectively (Fig. 3b). Although bovines do not express C-4 acetylated sialosides, BCoV HE readily converted **1a** into **1b**. Exposure of the product of MHV-S (**1d**) with BCoV HE gave a 7-*O*-acetylated derivative **1f**. Treatment of **1b** and **1f** with aqueous ammonium bicarbonate (100 mM) resulted in an efficient migration of the C-7 acetyl ester to C-9 providing **1c** and **1g**, respectively. Finally, subjecting **1c** to BCoV gave the 4-*O*-acetyl derivative **1e**. Compound **2a** (see Supplementary Fig. S3 for chemical synthesis) could be diversified into the mono- and di-*O*-acetylated derivatives **2b-e** by using MHV-S and BCoV HE combined with controlled *O*-acetyl ester migrations.

The preparation of 2,8-linked di-sialosides is particularly challenging due to the low reactivity of the C-8 hydroxyl of a sialic acid acceptor²⁸. Coupling of **4** with **7**²⁹ using TMSOTf as activator gave the corresponding 2,8-linked di-sialoside in high yields which could easily be converted into **3a** (Fig. 3c) and then diversified by MHV-S and BCoV HE treatment and controlled acetyl ester migration to provide **3d, f** and **g**. We found that the 4-*O*-acetyl ester of the terminal and the internal sialoside of **3a** can readily be cleaved by MHV-S HE thereby providing access to biologically highly relevant glycans⁹.

The results demonstrate that regardless of glycosidic linkage type and *O*-acetylation pattern, MHV-S and BCoV HEs efficiently remove *C*-4 and *C*-9 acetyl esters, respectively. Furthermore, we found that the viral esterases are highly enzymatically active, can be stored for prolonged period of times and tolerate a wide range of pH values, making them attractive tools for chemoenzymatic synthesis. All compounds were fully characterized by 2D NMR experiments, which confirmed homogeneity and a proper pattern of *O*-acetylation. The latter was evident from a large downfield shift of α -proton peaks (Supplementary Table S2), which in HMBC experiments showed correlation with carbonyl carbons (Supplementary Fig. S6-S8).

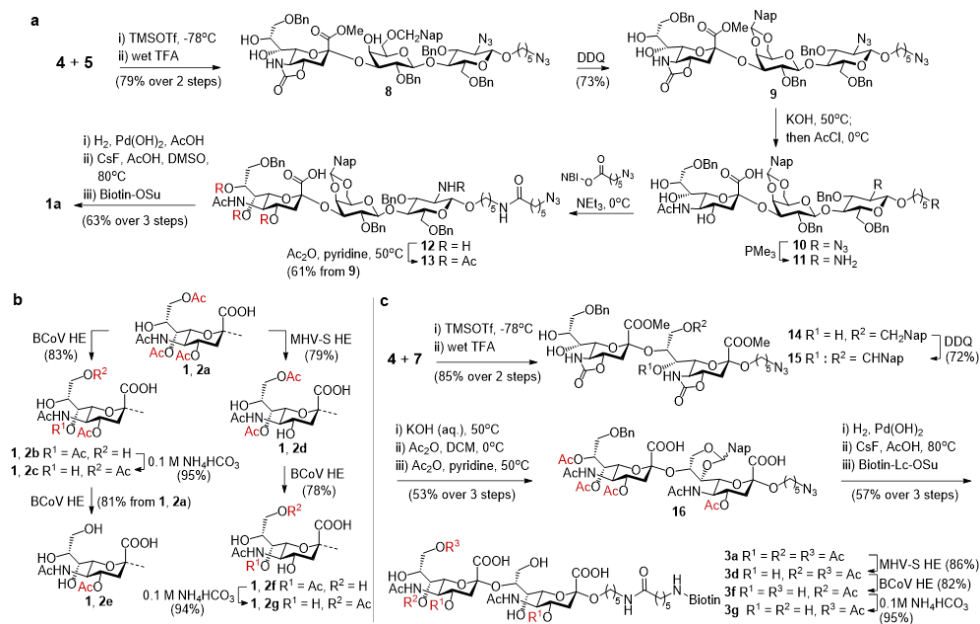


Figure 3 | Chemoenzymatic syntheses of *O*-acetylated sialosides. a, Chemical synthesis of **1a**. See Supplementary Fig. S4 for details on regioselective *C*-8-to-*C*-9 acetyl ester migration performed on deprotected oligosaccharides in DMSO. Abbreviations: Lc, 6-aminocaproic acid linker; NBI, 5-norbornene-2,3-dicarboxylic acid imido; Su, succinimido. **b**, Modification of **1a** and **2a** using HE-catalyzed *O*-acetyl removal and base-mediated *O*-acetyl migration. Enzymatic reactions went to full conversion. Yields were determined post-HPLC purification. 7-*O*-acetyl remained stable (for sialoforms **b**, **d**, and **f**) during the course of reaction and purification. See Supplementary Figs. S5-S8 and Table S2 for structural confirmation. **c**, Chemical synthesis and enzymatic modification of **3a** using HEs. MHV-S HE removed 4-*O*-acetyl groups of both non-reducing and reducing end Neu5Ac residues.

Viral receptor analysis using glycan microarray technology. The comprehensive library of *O*-acetylated sialosides allowed systematic examination of receptor specificities of

Chapter 2

an extensive range of viruses known to use this class of glycans for host cell attachment and entry (see Supplementary Table S1 for specifications). Among these were three established human viruses, namely coronaviruses OC43 and HKU1 (genus *Betacoronavirus*, subgenus *Embecovirus*) and influenza C virus (ICV). Infections with these viruses can cause mild upper respiratory illnesses in healthy adults, but can lead to serious complications in newborns, the elderly and immunocompromised individuals³⁰. In addition, receptor specificities of a wide range of related animal viruses were investigated including bovine coronavirus (BCoV), which is the presumed ancestor of OC43, and influenza D virus (IDV), which is related to ICV. We also examined receptor specificities of toroviruses from swine and cattle that are of veterinary importance and can cause respiratory and gastrointestinal illnesses. The CoVs attach to *O*-acetylated sialosides *via* their spike protein (S) for virion uptake. They also express a homodimeric HE, composed of a sialic acid binding and esterase domain, which is important for destruction of decoy receptors and virion release. Toroviruses encode a gene homologous to HE, whereas ICV and IDV express a multifunctional hemagglutinin-esterase-fusion protein (HEF), in which the hemagglutinin domain is responsible for receptor binding, the esterase region for receptor-destruction and fusion domain for merging viral envelope with host membrane.

The receptor-binding (S1^A) domains of the spike proteins, and enzymatically inactive forms of the ectodomains of HEs and HEFs were expressed in HEK293T cells as Fc fusion proteins. The *O*-acetylated sialosides were printed on streptavidin coated glass slides to give a microarray that was exposed to different concentrations of the Fc-fusion proteins and binding was detected by an anti-Fc antibody tagged with AlexaFluor-647 (full structures of compounds on the microarray are shown in Supplementary Fig. S9).

Spike proteins of the human CoVs OC43 and HKU1 showed a strong preference for α 2,8-linked disialic acid modified by 9-*O*-acetyl ester (Fig. 4a). HKU1, which entered the human population centuries earlier than OC43³¹, exhibited the highest selectivity for this glycoform. Both virolectins also bound disialic acid having an additional *C*-7 acetyl ester, but with much lower responsiveness. The HEF of ICV/JHB, a human orthomyxovirus, targets the same structure but does not tolerate an acetyl ester at *C*-7. The binding selectivities of the spike proteins from animal CoVs differ from those of humans. In the case of BCoV, the presumable ancestor of OC43, the spike is rather promiscuous and binds several 9-mono- and 7,9-di-*O*-acetylated sialosides preferentially linked in an α 2,3-fashion. Equine, canine and rabbit CoVs also encode spikes that recognize *C*-9 acetylated sialosides but again with preferences for specific glycosidic forms distinctive from that of the human viruses. Similarly, IDV/OK, an orthomyxovirus of swine and distantly related to the human pathogen ICV³², prefers α 2,3- and 2,6-sialosides.

Surprisingly, the HEs of the animal CoVs display ligand fine specificities different from that of the corresponding spike protein (Fig. 4b). For example, the spikes of ECoV, RbCoV and CRCoV bind preferentially 2,3-linked sialosides modified by a 9-*O*-mono-acetyl ester whereas their HEs recognize di- or tri-*O*-acetylated structures. The mouse CoV (MHV-S)

spike and HE represent an extreme case of engaging with different structures, and bind a proteinaceous receptor (CEACAM-1a)³³ and 4-*O*-acetylated sialosides, respectively. The lectin domain of the HEs has probably evolved to recognize *O*-acetylated glycans rich in the mucus layer of the host. It contributes to virion attachment but also regulates the sialate-*O*-acetyl-esterase activity of HE to destroy decoy receptors to facilitate virion migration through mucus and release of viral progeny^{34,35}. The spike protein has as a main function to bind to a cell surface receptor to promote virion uptake followed by membrane fusion and ultimately releasing the nucleocapsid into the cytoplasm³⁶. Due to promiscuity, a spike may also have affinity for decoy receptors in the mucus, and for BCoV and OC43 it has been shown that HE and spike function as a two-component system for dynamic receptor interactions³⁵.

The binding selectivities of the HEs for specific *O*-acetylation forms correlate well with the sialic acid repertoire of the mucus of the respective hosts. Cattle produce ample quantities of 7,9-di-*O*-acetylated sialosides³⁷, which is preferred by BCoV and BToV. Equine, murine and rabbits harbor 4-*O*-acetylated sialic acids across multiple tissues³⁸, which is in agreement with the specificities of the HEs from ECoV, MHV-S and RbCoV, respectively. Loss of HE lectin function in the human coronaviruses OC43 and HKU1 resulted in a substantial reduction in sialate-*O*-acetyl-esterase activity towards multivalently displayed 9-*O*-acetylated sialosides³⁹, possibly as an adaptation to low receptor densities in the human respiratory mucus⁴⁰.

Identification of functional receptor for human coronaviruses. While reversible virion attachment to sialosides is key during transit through the mucus, a more stable association at the epithelial cell surface is required for entry. The array studies showed that the spikes of the human viruses OC43 and HKU1 recognize 9-*O*-acetylated α 2,8-linked disialosides. These observations prompted us to examine whether spike-mediated binding to such a glycotope can facilitate infection.

Humans express six α 2,8-sialyltransferases (ST8Sia1-6). ST8Sia2, 3 and 4 are involved in the biosynthesis of α 2,8-linked linear homopolymers of sialic acid, which are mainly found in the brain where they play a role in the developmental plasticity of the nervous system^{41,42}. These sialosides have also been detected in secondary lymphatic tissues where they are involved in the recruitment of immune cell⁴³, and on cancer cells where they play a role in metastasis⁴⁴. Polysialic acids of mammals are not known to be modified by *O*-acetyl esters, and have not been observed on epithelial cells, and thus it is unlikely they function as receptor for the respiratory viruses. ST8Sia1 (GD3 synthase) and ST8Sia5 (GT3 synthase) and ST8Sia6 transfer a single sialic acid residue to NeuAc α 2-3Gal moieties, thereby forming α 2,8-linked di-sialosides. The first two enzymes modify gangliosides whereas the latter one acts on *O*-linked glycans⁴⁵⁻⁴⁷. A recent study demonstrated that the human respiratory tract expresses α 2,8-linked disialylated glycolipids, and in particular the ganglioside GD3 (NeuAc α 2,8NeuAc α 2,3Gal β 1,4Glc β 1-Cer) was abundantly present⁴⁸. This ganglioside is one of a few glycoconjugates in humans known to be modified by a C-9 acetyl ester (9-Ac-GD3).

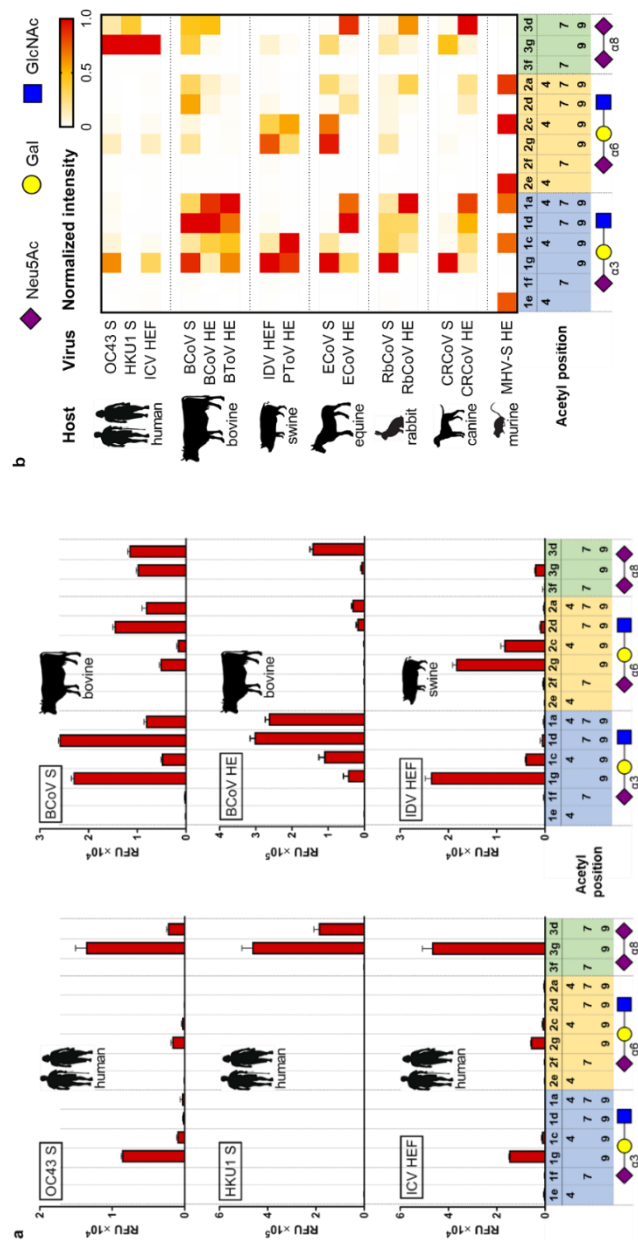


Figure 4 | Glycan microarray studies reveal binding patterns associated with host specificities. **a**, Column graph of viral receptor selectivities. **b**, Heatmap presentations of viral receptor selectivities. In the heat map, the signal intensities are normalized with the highest value in each protein defined as 1.0 and are shown with a dual-color gradient. Host of the virus is indicated with silhouettes. Concentrations of Fc-tagged proteins used: BCoV S1^A 0.3 μg/mL, OC43 S1^A 0.3 μg/mL, HKU1 S1^A 30 μg/mL, ICV HEF 0.3 μg/mL, IDV HEF 0.3 μg/mL, PTov HE 3 μg/mL, BCoV HE 3 μg/mL, BCoV HE 3 μg/mL, RbCoV HE 3 μg/mL, MHV-S HE 3 μg/mL, ECoV S1^A 3 μg/mL, RbCoV S1^A 3 μg/mL, CRCov S1^A 10 μg/mL. See Supplementary Fig. S9 for chemical structures of the compounds. Additional results with different concentrations of proteins are presented in Supplementary Fig. S10. Representative surface dissociation constants ($K_{D,surf}$) were obtained for the binding of HKU1 S1^A with compounds **3g** (9-*O*-Ac, 60 nM) and **3d** (7-9-di-*O*-Ac, 85 nM). See supplementary Fig. S11 for binding curves. Sialoform **b** (4,7-di-*O*-Ac Neu5Ac) was not included in the library because it has not been documented to be naturally occurring.

At the plasma membrane, gangliosides cluster in microdomains, where they can engage in multivalent interactions with receptor binding proteins resulting in high avidity binding, receptor clustering and endocytosis⁴⁹. Indeed, they are known to promote cellular uptake of bacterial toxins⁵⁰ and viruses including ICV⁵¹. In the case of CoVs, the attachment of virions to gangliosides *via* the 20-nm spike proteins is expected to keep the 8-nm HE proteins at some distance from the cell surface, thereby protecting these receptors from HE-mediated destruction preventing virion elution³⁹. Thus, we asked whether gangliosides, such as 9-Ac-GD3, may function as a receptor for OC43 and HKU1.

The biosynthesis of 9-Ac-GD3 involves the concerted action of *N*-acetylneuraminase 9-*O*-acetyltransferase (CASD1), which catalyzes the addition of an acetyl ester to *C*-9 of cytidine-monophosphate-linked sialic acid to give CMP-Neu5,9Ac₂, and ST8Sia1 which can transfer this sugar nucleotide to GM3⁵². HEK293T cells naturally express CASD1, and by transfecting these cells with ST8Sia1, produce 9-Ac-GD3. Thus, we examined by immunofluorescence the binding of OC43 and HKU1 S1^A-Fc fusion proteins to cells transfected by ST8Sia1 (Fig. 5a). OC43 S1^A-Fc bound to mock-transfected cells, in accordance with the susceptibility of HEK293T-cells to OC43-1967/USA infection³⁹, but binding was strongly increased upon ST8Sia1-overexpression. HKU1 can only be propagated in epithelial cell cultures and select cell lines⁵³, but not in HEK293T cells. As anticipated, HKU1 S1^A-Fc did not detectably bind to mock-transfected HEK293T cells. However, it strongly bound to ST8Sia1-transfected cells demonstrating the latter enzyme is involved in the biosynthesis of the receptor for this virus. Substitution of the critical S1^A receptor binding site (RBS) residue Trp90 by Ala abolished binding⁵⁴, showing an intact RBS is required for binding. Treatment of transfected cells with BCoV HE strongly reduced binding confirming a dependence of *O*-acetylation of sialic acid for binding. Finally, a strong reduction in binding was also observed when transfected cells were pretreated with an inhibitor of glycolipid biosynthesis (Supplementary Fig. S12)⁵⁵, supporting that S1^A-Fc binds to 9-*O*-acetylated disialogangliosides.

To determine whether 9-*O*-acetylated disialogangliosides support cell entry, infection assays were performed with vesicular stomatitis virus (VSV) particles pseudotyped with CoV spikes. Mock-transfected HEK293T cells were susceptible to infection with VSV pseudotyped with OC43 spike, but the infection increased significantly upon ST8Sia1 over-expression (Fig. 5b). HEK293T cells were not detectably infected by particles pseudotyped with HKU1 spike but became susceptible upon overexpression of ST8Sia1. VSV particles, pseudotyped with RBS-defective spike proteins, were non-infectious, while infection by particles bearing wildtype spikes was prohibited in CASD1-knockout HEK293T cells transfected with ST8Sia1. The microarray and cell-based studies indicate that the spikes of OC43 and HKU1 can bind to α 2,8-disialosyl containing gangliosides, such as 9-Ac-GD3, *via* their S1^A in a 9-*O*-acetyl-dependent fashion, thereby facilitating cell entry. Future studies will focus on extending the chemo-enzymatic methodology to the preparation of intact ganglioside glycans for direct binding studies to investigate the importance of the underlying glycan for recognition by the human coronaviruses.

44

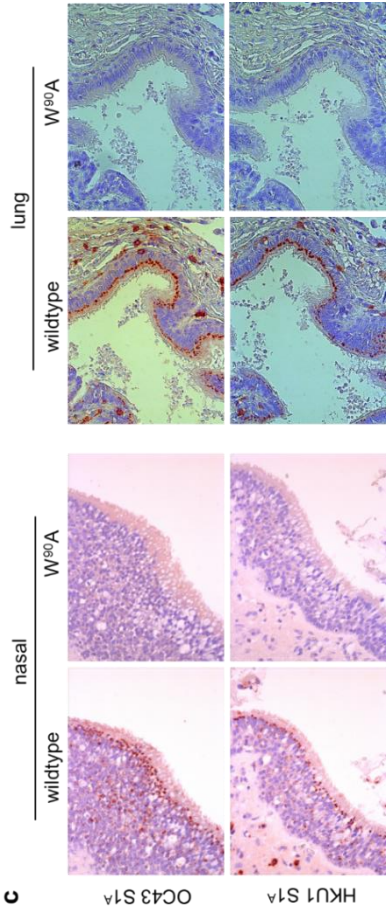
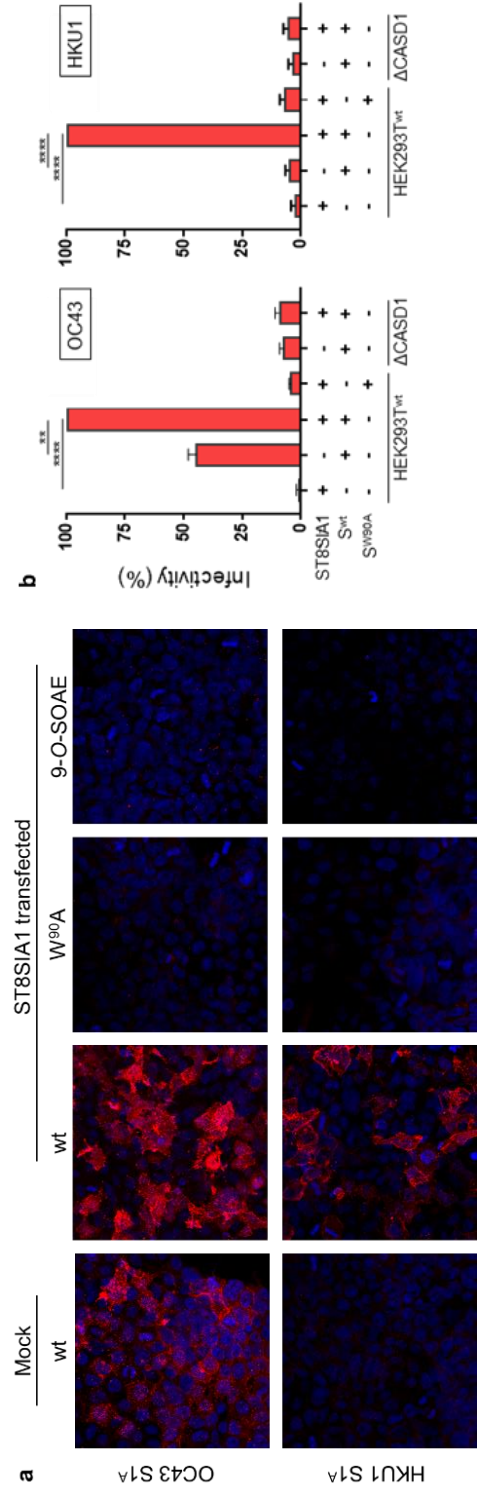


Figure 5 | 9-*O*-acetylated disialogangliosides serve as receptor for human coronavirus OC43 and HKU1. **a**, Immunofluorescent staining of mock- or ST8Sia1-transfected HEK293T cells using wildtype (denoted wt) or W90A-mutant S1^A-mouse-Fc from OC43 or HKU1. Cells pre-treated with the sialate-9-*O*-acetyl esterase (denoted 9-*O*-SOAE), BCoV HE, were used as receptor-depleted controls. W stands for amino acid tryptophan (Trp) and A for alanine (Ala). See Supplementary Fig. S14 for cell treatment with glycolipid biosynthesis inhibitor. **b**, Infectivity assays using OC43 and HKU1 spike-pseudotyped virions. Wildtype or CASD1-knockout (denoted Δ CASD1) HEK293T cells were mock transfected or with ST8Sia1 before infection. VSV particles encoding a luciferase gene were pseudotyped with wildtype or W90A-mutant spike from OC43 or HKU1. Infectivity is quantified with relative light units (RLUs) by measuring signal produced from reaction of luciferase with the substrate. RLUs are normalized by setting the highest value in OC43 or HKU1 as 100%. **c**, Immunohistochemical staining of nasal (left) and lung (right) epithelia using wildtype or W90A-mutated OC43 or HKU1 S1^A fused with mouse Fc.

Finally, we examined the distribution of *O*-acetylated disialoside receptors in the human respiratory tract using OC43 and HKU1 S1A fused with mouse Fc (Fig. 5c). The two wild-type S1^A proteins stained ciliated pseudostratified columnar epithelial cells of human nasal cavity and lung tissue. No binding was observed for a receptor-binding-site-deficient S1^A, indicating the staining is dependent on the interaction with sialate-9-*O*-acetyl. The results demonstrate that 9-*O*-acetylated disialoside glycotopes are expressed in human upper and lower respiratory tracts, with the former being the major site for viral replication in natural infections.

Conclusions

CoVs emerging from zoonotic introduction pose threats to public health as poignantly illustrated by the current SARS-CoV-2 pandemic⁵⁶. Four CoVs of zoonotic origin have become endemic in the human population, which include the alphacoronaviruses NL63 and 229E and the betacoronavirus OC43 and HKU1, causing up to 30% of mild respiratory tract infections but can also lead to severe disease and fatality in young children, the elderly and immunocompromised individuals. Future zoonotic transmissions of CoVs are likely given there is large animal reservoir and their ability to cross species barriers. Preparedness for future outbreaks will require a fundamental understanding how these viruses adapt to replication in humans, which includes insight in human-associated patterns of receptor preference. Here, we describe a platform to examine receptor requirements of viruses engaging with *O*-acetylated sialosides including Embecoviruses. Our findings highlight that the human CoVs OC43 and HKU1, and human orthomyxovirus ICV, preferentially bind to 9-*O*-acetylated α 2,8-linked sialosides. Immunofluorescence and cell entry studies indicate that such glycotope as part of glycolipids is employed for viral entry. Preferential usage of a common receptor type by three distinct human respiratory viruses is indicative of convergent evolution by adapting to the sialoglycome of the human respiratory tract. This knowledge will facilitate predicting future cross-species transmissions and will be important for the

Chapter 2

implementation of prevention and intervention strategies. In addition to host-virus interactions, *O*-acetylated sialosides have also been implicated in innate and adaptive immunity³⁷, cancer⁵⁷, and bacterial foraging and pathogenesis^{58,59}. The glycan array described here will find utility to examine at a molecular level how *O*-acetylated sialosides modulate physiological and disease processes.

Methods

General procedure for sialylation and isopropylidene removal. A suspension of the diol acceptors **5**, **6** or **7** and the sialyl donor **4** (1.2 eq. to **5** and **6**, 2.0 eq. to **7**) and 4 Å molecular sieves in anhydrous DCM (50 mM based on acceptors) was placed under an atmosphere of argon and then stirred at room temperature for 1 h. The mixture was cooled (-78°C) and TMSOTf (1.2 eq. to **5** and **6**, 2.0 eq. to **7**) was added dropwise. The reaction mixture was stirred at -78°C for 5-10 min after which it was then quenched with triethylamine. The mixture was filtered through Celite, diluted with DCM and washed with sat. NaHCO₃. The organic phase was dried (Na₂SO₄) and concentrated *in vacuo*. The residue was dissolved in DCM to which a few drops of water was added. TFA (10 - 15% v/v) was added, and the biphasic reaction mixture was stirred vigorously at room temperature for 2 h. The mixture was directly poured into cold sat. NaHCO₃. The organic phase was collected when no more CO₂ was generated, dried (Na₂SO₄), filtered and the filtrate concentrated *in vacuo*. The residue was purified by silica gel column (toluene: ethyl acetate 3:1-2:1 for **4+5** and **4+6**, toluene: acetonitrile 4:1-3:1 for **4+7**).

General procedure for hydrogenation, regioselective acetyl migration and biotin conjugation. In a double-necked flask, the starting material (**13**, **16** or **S15**) was dissolved in THF/water containing acetic acid (5 eq. to starting materials). The flask was filled with hydrogen followed by the addition of palladium hydroxide (Degussa type). Upon completion of the reaction, the catalyst was filtered off over Celite and the filtrate concentrated *in vacuo* and further freeze-dried. The residue was dissolved in DMSO-*d*₆ in presence of 6 eq. acetic acid-*d*₄ and 3 eq. cesium fluoride. The mixture was heated (80 °C) for 14-20 h, after which a proton NMR was recorded which showed a shift of Sia *H*-8 from ~5 ppm to ~4 ppm confirming completion of the reaction. The mixture was lyophilized and the residue was directly dissolved in DMSO to which *N*-hydroxysuccinimide activated biotin or Lc-biotin was added. Upon completion of the reaction as indicated by ESI-MS, the solution was freeze-dried and the resulted residue was passed over a Biogel p-2 column and then further purified with RP-HPLC (C18, acetonitrile/water 5% - 40%, 50 min, for **1a** and **2a**) or HILIC HPLC (acetonitrile/water 95% - 65%, 60 min, for **3a**).

HE-mediated de-*O*-acetylation. The precursor **1a**, **2a** or **3a** was dissolved in 50 mM ammonium formate buffer (pH 6.4) to achieve 5 mM concentration. BCoV and/or MHV-S HEs was added to achieve a concentration of 10 µg/mL. The progress of the reaction was monitored by ESI-MS (negative mode). Upon completion, the reaction mixture was

lyophilized and purified with RP-HPLC (C18, acetonitrile/water 5% - 40%, 50 min, for the **1** and **2** series) or HILIC-HPLC (acetonitrile/water 95% - 65%, 60 min, for the **3** series).

Glycan microarray printing and screening. The biotinylated compounds were printed on streptavidin coated glass slides (SuperStreptavidin Microarray Substrate Slides, ArrayIt Inc) using a Scienion sciFLEXARRAYER S3 non-contact microarray equipped with a Scienion PDC80 nozzle (Scienion Inc). Compounds were dissolved in MilliQ water at the concentration of 1 mM to prepare a stock solution. This solution was placed at -80 °C for long-term storage. A printing solution of biotinylated sialosides at a concentration of 100 μ M was freshly made each time from by diluting the stock solution in the printing buffer (10 mM PBS buffer, pH 7). The compounds were printed in replicates of 6 with spot volume ~ 400 pL, at 20 °C and 50% humidity. The slides were stored at 4 °C, and blocked with TSM binding buffer (20 mM Tris·HCl, pH 7, 150 mM NaCl, 2 mM CaCl_2 and 2 mM MgCl_2 , 0.05% Tween-20, 1% BSA) for 1 h at 4 °C prior to use. The human-Fc tagged proteins were mixed with goat anti-human IgG antibody (Alexa Fluor 647 conjugated, 109-605-008, Jackson ImmunoResearch) in a 1:1 ratio in TSM binding buffer and incubated at 4°C for 1 h. The slides were incubated with the pre-mixed binding solution at 4 °C for 1 h. Post-incubation, the slides were washed with **1**) TSM washing buffer (20 mM Tris·HCl, pH 7, 150 mM NaCl, 2 mM CaCl_2 and 2 mM MgCl_2 , 0.05% Tween-20), **2**) TSM buffer (20 mM Tris·HCl, pH 7, 150 mM NaCl, 2 mM CaCl_2 and 2 mM MgCl_2) and **3**) water, and were then spun dry. The slides were scanned using a GenePix 4000B microarray scanner (Molecular Devices) at the appropriate excitation wavelength. The image was analyzed using GenePix Pro 7 software (version 7.2.29.2, Molecular Devices). This procedure complies with MIRAGE Glycan Array Guideline v 1.0⁶⁰. See Supplementary Information for method of obtaining surface dissociation constants ($K_{D,\text{surf}}$).

Immunofluorescence assay of cultured cells. HEK-293T cells were grown on glass coverslips in 24-well format and either mock transfected or transfected with pcDNA-ST8Sia1 plasmid using fugene according to manufacturer's instructions. At 36 h post transfection, cell monolayers were fixed with paraformaldehyde (PFA; 3.7% in PBS, 30 min), incubated with blocking buffer (PBS, 0.05% Tween-20, 2% BSA) for 30 min. For soluble HE, cells were treated with BCoV-LUN HE (20 μ g/mL in blocking buffer) for 2 h at 37°C. Cells were then incubated with S1^A-mFc lectins (50 μ g/mL) for 45 min, and then for 30 min with goat-anti-mouse IgG-Alexa 594 (A32742, Thermo Scientific; 1:100), and with DAPI (1:1000). All incubations were in blocking buffer. Washing was performed with PBS containing 0.05% Tween-20. The cells were then embedded in FluorSave (Calbiochem) and examined by confocal fluorescence microscopy (Nikon STORM). Glycan biosynthesis inhibitor N-butyl deoxygalactonojirimycin were added to culture supernatants of mock transfected or pcDNA-ST8SIA1 transfected HEK 293T cells at 8 h post transfection. Incubation was continued for 72 h. Cells were then fixed and stained with S1^A-mFc lectins as above.

Virolectin staining of human tissues. Human tissue staining was performed using procedures adapted from previous reports¹⁰. Briefly, paraffin-embedded human tissue

Chapter 2

sections were dewaxed in xylene and rehydrated. Endogenous peroxidases were inactivated with peroxide (0.3% in methanol). Tissue sections were then successively incubated with S1^A-mFc virolectins (50 µg/mL) and biotinylated goat- α -mouse IgG antibodies (SAB4600004, Sigma; 1:250; lectins and antibodies all diluted in blocking buffer), with avidin-biotin HRPO complex (ABC-PO staining kit, Thermo Scientific) and with 3,3'-diaminobenzidine (DAB) (Sigma). Blocking, incubations and washing was performed as for IFA. Tissue sections were counterstained with Mayer's Hematoxylin, embedded in Eukitt® mounting medium (Fluka) and examined by standard light microscopy.

Vesicular stomatitis virus (VSV)-pseudotyping and infectivity assay. The production of OC43/HKU1 S-pseudotyped G-protein-deficient VSV particles and infectivity assays were as described⁶¹. In brief, C-terminal truncated and flag-tagged OC43 and HKU1 S (wildtype and W90A mutant) were cloned in pCAGGS expression vector. A plasmid mixture of pCAGGS-OC43/HKU1-flag and pCD5-BCoV HE-Fc (molar ratios 8:1) were complexed with PEI and transfected to HEK 293T cells at 70% confluency. 48 h post transfection, cells were transduced with VSV-G pseudotyped VSV Δ G/Fluc with MOI of 1. Cell-free supernatants were then harvested at 24 h after transduction, filtered through 0.45µm membranes, and virus particles were further purified and concentrated by sucrose cushion ultracentrifugation at 29,000 rpm for 2.5 h. Pelleted virions were resuspended in PBS and stored at -80°C before usage. For the infection assay, HEK-293T or HEK-293T^{ΔCASD1} cells were seeded in 96-well cluster format, and were transfected or mock transfected with pcDNA-ST8SIA1 plasmid using fugene. At 36 h post transfection, infection was performed with equal amounts S-pseudotyped VSVs as calculated from VSV-N content. 18 h later, infected cells were lysed using passive lysis buffer (Promega). Firefly luciferase expression was measured via a homemade firefly luciferase assay system as described. Infection experiments were conducted in independent triplicates, each of which with three technical replicates.

Acknowledgments We thank Dr. Debby van Riel (Erasmus MC) for providing paraffin-embedded tissue sections; Dr. Richard Wubbolts (Utrecht University) for offering guidance for confocal fluorescence microscopy; Dr. Herman Egberink and Ms. Shan Zhao (Utrecht University) for providing plasmids of ECoV spike protein. This work was supported by TOP-PUNT Grant 718.015.003 of the Netherlands Organization for Scientific Research (G.-J.B.); ECHO Grant 711.011.006 of the Council for Chemical Sciences of the Netherlands Organization for Scientific Research (R.J.d.G.); and China Scholarship Council 2014-03250042 (Y.L.).

Author contributions G.-J.B., R.J.d.G., Z.L. and Y.L. conceived the study. G.-J.B. and R.J.d.G. supervised the study. Z.L. performed chemical and enzymatic synthesis. M.I.B. assisted chemical synthesis. Y.L. produced enzymes. Z.L. performed analysis of synthetic compounds. L.L. performed glycan microarray screening. L.L. and Z.L. processed microarray data. Y.L. produced viral glycoproteins used in microarray analysis. G.-J.B., Z.L., and L.L. analyzed glycan microarray data. Y.L. and A.I.S. performed immunofluorescence

assay, histochemical tissue staining and infectivity assay. G.-J.B., R.J.d.G., Z.L. and Y.L. wrote the paper.

References

1. Varki, A. Nothing in glycobiology makes sense, except in the light of evolution. *Cell* **126**, 841-845 (2006).
2. Le Pendu, J., Nyström, K. & Ruvoën-Clouet, N. Host-pathogen co-evolution and glycan interactions. *Curr Opin Virol* **7**, 88-94 (2014).
3. Thompson, A. J., De Vries, R. P. & Paulson, J. C. Virus recognition of glycan receptors. *Curr Opin Virol* **34**, 117-129 (2019).
4. Long, J. S., Mistry, B., Haslam, S. M. & Barclay, W. S. Host and viral determinants of influenza A virus species specificity. *Nat Rev Microbiol* **17**, 67-81 (2018).
5. Stencel-Baerenwald, J. E., Reiss, K., Reiter, D. M., Stehle, T. & Dermody, T. S. The sweet spot: Defining virus-sialic acid interactions. *Nat Rev Microbiol* **12**, 739-749 (2014).
6. Wasik, B. R., Barnard, K. N. & Parrish, C. R. Effects of sialic acid modifications on virus binding and infection. *Trends Microbiol* **24**, 991-1001 (2016).
7. Morse, S. S. et al. Prediction and prevention of the next pandemic zoonosis. *Lancet* **380**, 1956-1965 (2012).
8. Pepin, K., et al. Identifying genetic markers of adaptation for surveillance of viral host jumps. *Nat Rev Microbiol* **8**, 802-813 (2010).
9. Schauer, R. & Kamerling, J. P. Exploration of the sialic acid world. *Adv Carbohydr Chem Biochem* **75**, 1-213 (2018).
10. Langereis, M. A. et al. Complexity and diversity of the mammalian sialome revealed by nidovirus virolectins. *Cell Rep* **11**, 1966-1978 (2015).
11. Wasik, B. R. et al. Distribution of *O*-acetylated sialic acids among target host tissues for influenza virus. *mSphere* **2**, e00379-00316 (2017).
12. Chen, X. & Varki, A. Advances in the biology and chemistry of sialic acids. *ACS Chem Biol* **5**, 163-176 (2010).
13. Haverkamp, J., Schauer, R., Wember, M., Kamerling, J. P. & Vliegenthart, J. F. in *Hoppe-Seyler's Zeitschrift für Physiologische Chemie* Vol. 356 1575-1583 (Interactive Factory, 1975).
14. Anazawa, K., Furuhashi, K. & Ogura, H. Synthesis of 7-*O*-acetyl-*N*-acetylneuraminic acid-derivative *Chem Pharm Bull* **36**, 4976-4979 (1988).
15. Furuhashi, K. & Ogura, H. Studies on sialic acids. Syntheses of partially *O*-acetylated 4-methylcoumarin-7-yl 5-acetamido-3,5-dideoxy- α -D-glycero-D-galacto-2-nonulopyranosidonic acids *Chem Pharm Bull* **37**, 2037-2040 (1989).
16. Roth, A. & Faillard, H. Synthesis of fluorescent 7,8,9-tri-*O*-acetyl-*N*-acetyl- and 4-*O*-acetyl-*N*-acetylneuraminic acid α -thioketosides *Liebigs Ann Chem* **1993**, 485-489 (1993).
17. Clarke, P. A., Mistry, N. & Thomas, G. H. Synthesis of the complete series of mono acetates of *N*-acetyl-D-neuraminic acid *Org Biomol Chem* **10**, 529-535 (2012).
18. Park, S. S. & Gervay-Hague, J. Synthesis of partially *O*-acetylated *N*-acetylneuraminic acid using regioselective silyl exchange technology *Org Lett* **16**, 5044-5047 (2014).
19. Yu, H. et al. Highly efficient chemoenzymatic synthesis of naturally occurring and non-natural α -2,6-linked sialosides: A P. Damsela α -2,6-sialyltransferase with extremely flexible donor-substrate specificity *Angew Chem Int Ed Engl* **45**, 3938-3944 (2006).

Chapter 2

20. Sugiarto, G. et al. A sialyltransferase mutant with decreased donor hydrolysis and reduced sialidase activities for directly sialylating Lewis X. *ACS Chem Biol* **7**, 1232-1240 (2012).
21. Yu, H. et al. Chemoenzymatic synthesis of GD3 oligosaccharides and other disialyl glycans containing natural and non-natural sialic acids *J Am Chem Soc* **131**, 18467-18477 (2009).
22. Yu, H. et al. Effective one-pot multienzyme (OPME) synthesis of monotreme milk oligosaccharides and other sialosides containing 4-*O*-acetyl sialic acid *Org Biomol Chem* **14**, 8586-97 (2016).
23. Takayama, S., Livingston, P. O. & Wong, C.-H. Synthesis of the melanoma-associated ganglioside 9-*O*-acetyl GD3 through regioselective enzymatic acetylation of GD3 using subtilisin *Tet Lett* **37**, 9271-9274 (1996).
24. Guberman, M & Seeberger, P. H. Automated glycan assembly: A perspective. *J Am Chem Soc* **141**, 5581-5592 (2019).
25. Regl, G. et al. The hemagglutinin-esterase of mouse hepatitis virus strain s is a sialate-4-*O*-acetyl esterase. *J Virol* **73**, 4721-4727 (1999).
26. Kamerling, J. P. et al. Migration of *O*-acetyl groups in *N,O*-acetylneuraminic acids. *Eur J Biochem* **162**, 601-607 (1987).
27. Chuang, H. Y. et al. Synthesis and vaccine evaluation of the tumor-associated carbohydrate antigen RM2 from prostate cancer. *J Am Chem Soc* **135**, 11140-11150 (2013).
28. Komura, N. et al. Constrained sialic acid donors enable selective synthesis of α -glycosides. *Science* **364**, 677-680 (2019).
29. Tanaka, H., Nishiura, Y. & Takahashi, T. An efficient convergent synthesis of GP1c ganglioside epitope. *J Am Chem Soc* **130**, 17244-17245 (2008).
30. Su, S. et al. Epidemiology, genetic recombination, and pathogenesis of coronaviruses. *Trends Microbiol* **24**, 490-502 (2016).
31. Lau, S. K. et al. Discovery of a novel coronavirus, China Rattus coronavirus HKU24, from Norway rats supports the murine origin of betacoronavirus 1 and has implications for the ancestor of betacoronavirus lineage A. *J Virol* **89**, 3076-3092 (2015).
32. Hause, B. M. et al. Isolation of a novel swine influenza virus from Oklahoma in 2011 which is distantly related to human influenza C viruses. *PLoS Pathog* **9**, e1003176 (2013).
33. Williams, R. K., Jiang, G. S. & Holmes, K. V. Receptor for mouse hepatitis virus is a member of the carcinoembryonic antigen family of glycoproteins. *Proc Natl Acad Sci U.S.A* **88**, 5533-5536 (1991).
34. Linden, S. K., Sutton, P., Karlsson, N. G., Korolik, V. & McGuckin, M. A. Mucins in the mucosal barrier to infection. *Mucosal Immunol* **1**, 183-197 (2008).
35. Lang, Y. et al. Coronavirus hemagglutinin-esterase and spike proteins coevolve for functional balance and optimal virion avidity. Preprint at <https://www.biorxiv.org/content/10.1101/2020.04.03.003699v1> (2020).
36. Li, F. Structure, function, and evolution of coronavirus spike proteins. *Annu Rev Virol* **3**, 237-261 (2016).
37. Schauer, R., Srinivasan, G. V., Wipfler, D., Knipe, B. & Schwartz-Albiez, R. *O*-acetylated sialic acids and their role in immune defense. *Adv Exp Med Biol* **705**, 525-548 (2011).
38. Park, S. S. Post-glycosylation modification of sialic acid and its role in virus pathogenesis. *Vaccines (Basel)* **7**, 171-189 (2019).
39. Bakkers, M. J. et al. Betacoronavirus adaptation to humans involved progressive loss of hemagglutinin-esterase lectin activity. *Cell Host Microbe* **21**, 356-366 (2017).

40. Barnard, K. N. et al. Modified sialic acids on mucus and erythrocytes inhibit influenza A virus hemagglutinin and neuraminidase functions. *J Virol* **94**, e01567-01519 (2020).
41. Mühlenhoff, M., Rollenhagen, M., Werneburg, S., Gerardy-Schahn, R. & Hildebrandt, H. Polysialic acid: Versatile modification of NCAM, SynCAM-1 and neuropilin-2. *Neurochem Res* **38**, 1134-1143, (2013).
42. Volkers, G. et al. Structure of human ST8SiaIII sialyltransferase provides insight into cell-surface polysialylation. *Nature Struct Mol Biol* **22**, 627-635, (2015).
43. Kiermaier, E. et al. Polysialylation controls dendritic cell trafficking by regulating chemokine recognition. *Science* **351**, 186, (2016).
44. Elkashef, S. M. et al. Polysialic acid sustains cancer cell survival and migratory capacity in a hypoxic environment. *Sci Rep* **6**, 33026, (2016).
45. Bobowski, M., Harduin-Lepers, A. & Delannoy, P. ST8 Alpha-N-Acetyl-Neuraminide Alpha-2,8-Sialyltransferase 1 (ST8SIA1). in *Handbook of Glycosyltransferases and Related Genes* (eds Naoyuki Taniguchi et al.) 767-780 (Springer Japan, 2014).
46. Tsuji, S. & Takashima, S. ST8 Alpha-N-Acetyl-Neuraminide Alpha-2,8-Sialyltransferase 5 (ST8SIA5). in *Handbook of Glycosyltransferases and Related Genes* (eds Naoyuki Taniguchi et al.) 813-821 (Springer Japan, 2014).
47. Takashima, S. & Tsuji, S. ST8 Alpha-N-Acetyl-Neuraminide Alpha-2,8-Sialyltransferase 6 (ST8SIA6). in *Handbook of Glycosyltransferases and Related Genes* (eds Naoyuki Taniguchi et al.) 823-832 (Springer Japan, 2014).
48. Jia, N. et al. The human lung glycome reveals novel glycan ligands for influenza A virus. *Sci Rep* **10**, 5320, (2020).
49. Regina Todeschini, A. & Hakomori, S. Functional role of glycosphingolipids and gangliosides in control of cell adhesion, motility, and growth, through glycosynaptic microdomains. *Biochim Biophys Acta* **1780**, 421-433 (2008).
50. Ewers, H. & Helenius, A. Lipid-mediated endocytosis. *Cold Spring Harbor Perspect Biol* **3**, a004721 (2011).
51. Herrler, G. & Klenk, H.-D. The surface receptor is a major determinant of the cell tropism of influenza C virus. *Virology* **159**, 102-108 (1987).
52. Baumann, A. M. et al. 9-*O*-acetylation of sialic acids is catalysed by CASD1 via a covalent acetyl-enzyme intermediate. *Nat Commun* **6**, 7673 (2015).
53. Pyrc, K. et al. Culturing the unculturable: Human coronavirus HKU1 infects, replicates, and produces progeny virions in human ciliated airway epithelial cell cultures. *J Virol* **84**, 11255-11263 (2010).
54. Hulswit, R. J. G. et al. Human coronaviruses OC43 and HKU1 bind to 9-*O*-acetylated sialic acids via a conserved receptor-binding site in spike protein domain. *A Proc Natl Acad Sci U.S.A* **116**, 2681-2690 (2019).
55. Andersson, U., Butters, T. D., Dwek, R. A. & Platt, F. M. *N*-butyldeoxygalactonojirimycin: A more selective inhibitor of glycosphingolipid biosynthesis than *N*-butyldeoxynojirimycin, in vitro and in vivo. *Biochem Pharmacol* **59**, 821-829 (2000).
56. Coronaviridae Study Group of the International Committee on Taxonomy of Viruses The species severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nat Microbiol* **5**, 536-544 (2020).
57. Parameswaran, R. et al. *O*-acetylated *N*-acetylneuraminic acid as a novel target for therapy in human pre-B acute lymphoblastic leukemia. *J Exp Med* **210**, 805-819 (2013).

Chapter 2

58. Robinson, L. S., Lewis, W. G. & Lewis, A. L. The sialate *O*-acetyltransferase EstA from gut *Bacteroidetes* species enables sialidase-mediated cross-species foraging of 9-*O*-acetylated sialoglycans. *J Biol Chem* **292**, 11861-11872 (2017).
59. Nguyen, T. et al. The role of 9-*O*-acetylated glycan receptor moieties in the typhoid toxin binding and intoxication. *Plos Pathog* **16**, e1008336 (2020).
60. Liu, Y. et al. The minimum information required for a glycomics experiment (MIRAGE) project: Improving the standards for reporting glycan microarray-based data *Glycobiology* **27**, 280-284 (2017).
61. Tortorici, M. A. et al. Structural basis for human coronavirus attachment to sialic acid receptors. *Nat Struct Mol Biol* **26**, 481-489 (2019).

Supplementary Information for

Synthetic *O*-Acetylated Sialosides Facilitate Functional Receptor Identification for Human Respiratory Viruses

Part I. Supplementary Figures and Tables

Supplementary Table S1 | Specifications of viral glycoproteins. Proteins studied in this work are in bold. Abbreviations: RBP, receptor binding protein; S, spike protein; HE, hemagglutinin-esterase; HEF, hemagglutinin-esterase-fusion protein; 9-*O*-Ac Sia, 9-*O*-acetylated sialic acid; CEACAM, carcinoembryonic antigen-related cell adhesion molecule. All proteins used in the microarray studies were fused with human Fc (see Experimental Section 1.3 in this file for recombinant protein expression protocol), and were detected with AlexaFluor647-conjugated goat anti human Fc antibody (109-605-008, Jackson ImmunoResearch).

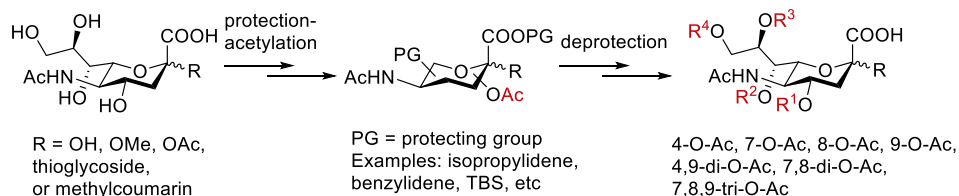
Virus	Strain/variant	Abbreviation	RBP	Receptor ^{Ref.1-4}
Embecovirus				
Human coronavirus OC43	USA/1967	OC43	S	9-O-Ac Sia
			HE	no binding
Human coronavirus HKU1	Hong Kong/2004	HKU1	S	9-O-Ac Sia
			HE	no binding
Bovine coronavirus	LUN	BCoV	S	9-O-Ac Sia
			HE	9- and 7,9-di-O-Ac Sia
Equine coronavirus	NC99	ECoV	S	?
			HE	9- and 7,9-di-O-Ac Sia
Canine respiratory coronavirus	240/05	CRCoV	S	?
			HE	9-O-Ac Sia
Rabbit coronavirus	HKU14	RbCoV	S	?
			HE	9-O-Ac Sia
Mouse hepatitis virus	S	MHV-S	S	CEACAM-1a
			HE	4-O-Ac Sia
Torovirus				
Bovine torovirus	Breda	BToV	S	?
			HE	9- and 7,9-di-O-Ac Sia

Chapter 2

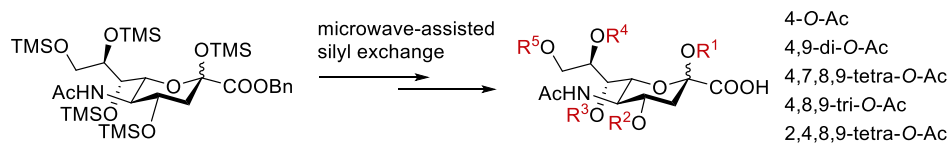
Porcine torovirus	P4	PToV	S HE	? 9-O-Ac Sia
Orthomyxovirus				
Influenza C virus	Johannesburg/66	ICV/JHB	HEF	9-O-Ac Sia
Influenza D virus	OK/1334/2011	IDV/OK	HEF	9-O-Ac Sia

Supplementary Figure S1 | Synthetic strategies for *O*-acetylated sialic acid. **a-d**, Previous approaches. Panels **a** and **b** describe the synthesis of monosaccharides. Panels **c** and **d** describe *O*-acetylated sialic acid-containing oligosaccharide synthesis. Only 9- or 4-monoacetylated sialic acid have been synthesized as oligosaccharides. **e**, Chemical liabilities of sialate-*O*-acetyl groups potentially hampering synthesis and analysis. **f**, The general strategy for assembly of sialoside common precursor for enzymatic diversification (demonstrated with **1a**). Key steps are shown in the text boxes. Abbreviation: Ac, acetyl; Bn, benzyl; LG, leaving group; Me, methyl; Nap, naphthyl; PG, protecting group; R, reducing terminus substitution; TBS, *tert*-butyldimethylsilyl; R¹-R⁵, potential sites of acetylation; TMS, trimethylsilyl.

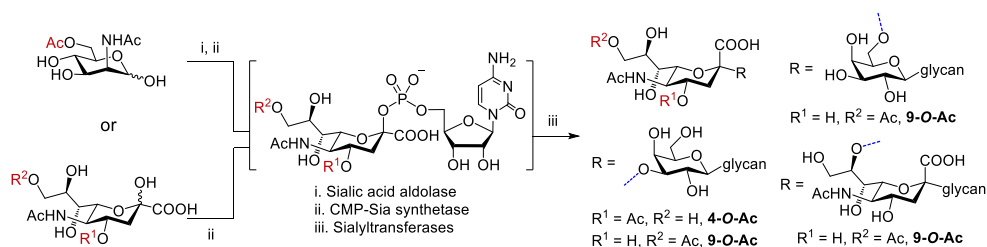
a. conventional protecting group manipulation (Ref. 13-17 in the main text)



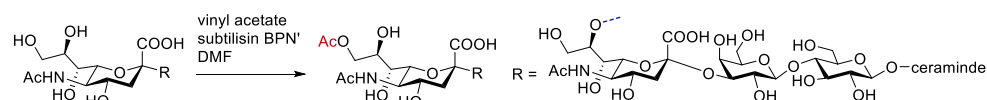
b. Regioselective Silyl Exchange Technology (ReSET, Ref. 18 in the main text)



c. One-pot multi-enzyme synthesis (OPME, Ref. 19-22 in the main text)

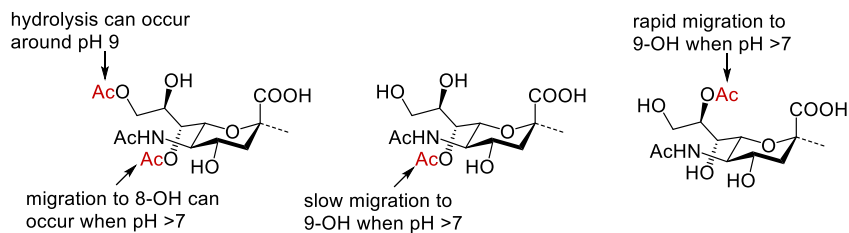


d. Enzymatic derivatization (Ref. 23 in the main text)

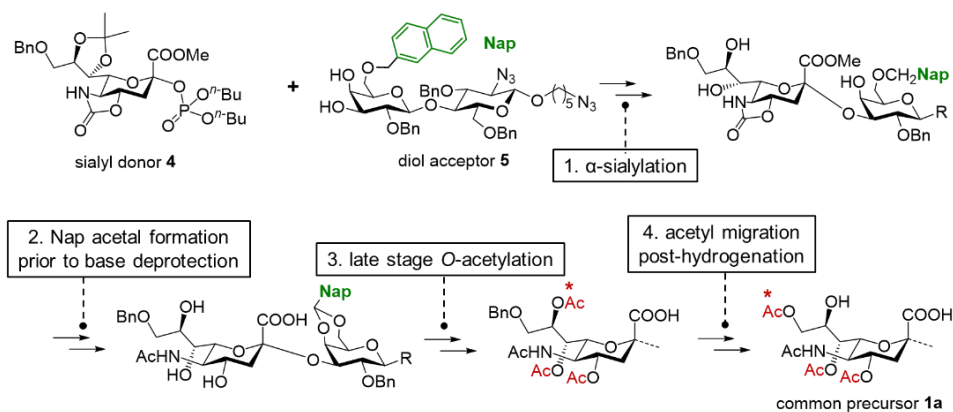


Chapter 2

e. Inherent liabilities of sialic acid O-acetyl esters

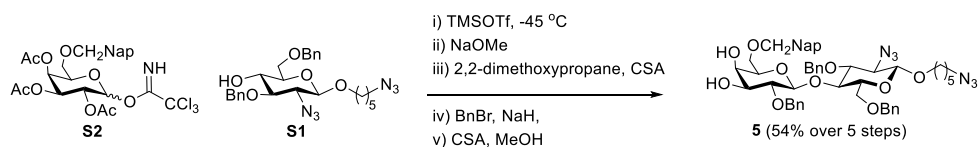


f. Strategy of this work

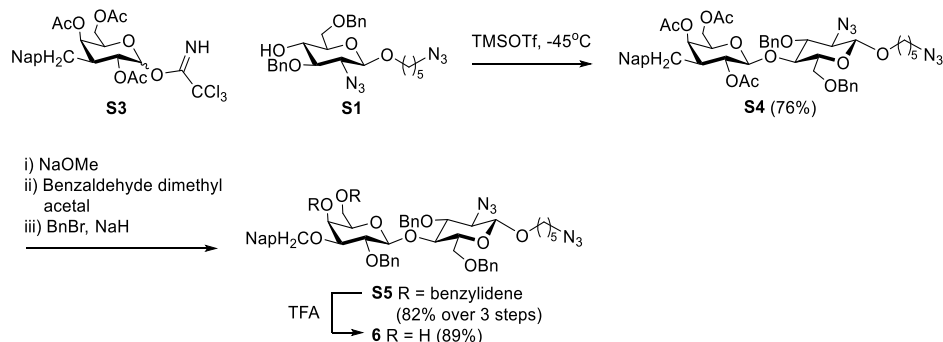


Supplementary Figure S2 | Chemical synthesis of glycosyl acceptors. a, α 2,3 acceptor 5. b, α 2,6 acceptor 6. c, α 2,8 acceptor 7. Abbreviations: TMSOTf, trimethylsilyl trifluoromethylsulfonate; NaOMe, sodium methoxide; BnBr, benzyl bromide; NaH, sodium hydride; Ac, acetyl; Bn, benzyl; NapCH₂, 2-naphthylmethyl; Me, methyl.

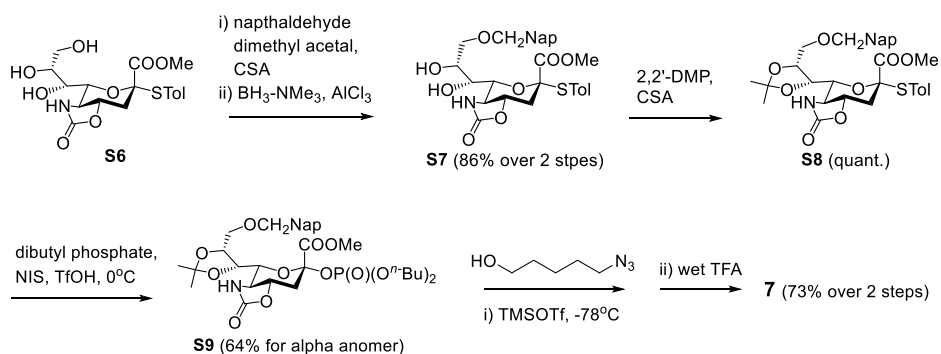
a



b

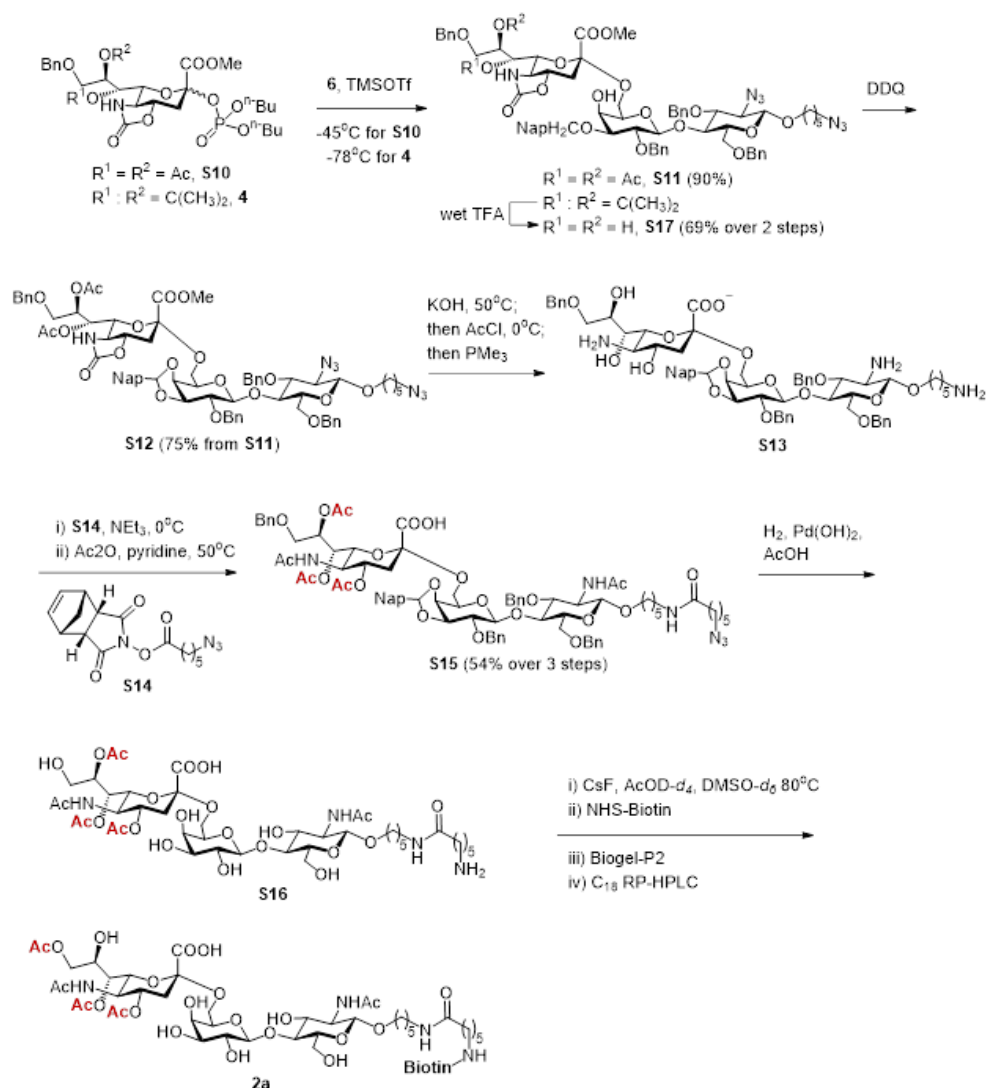


c

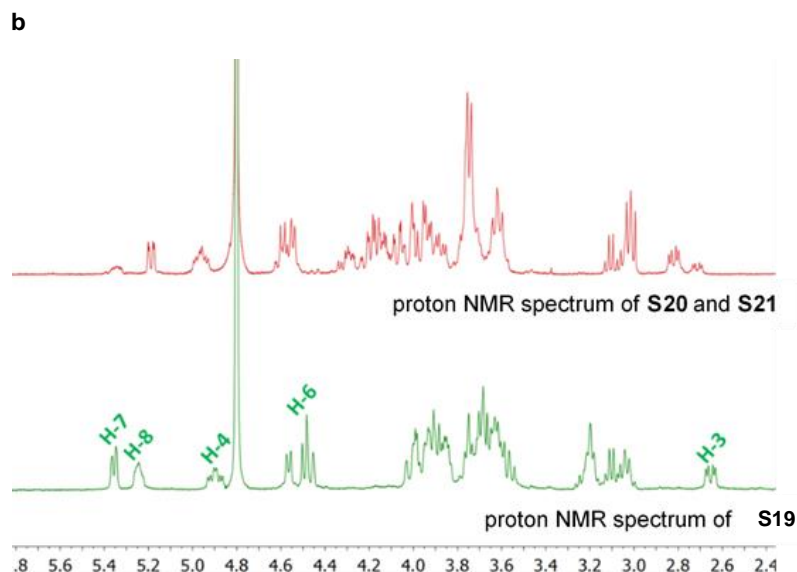
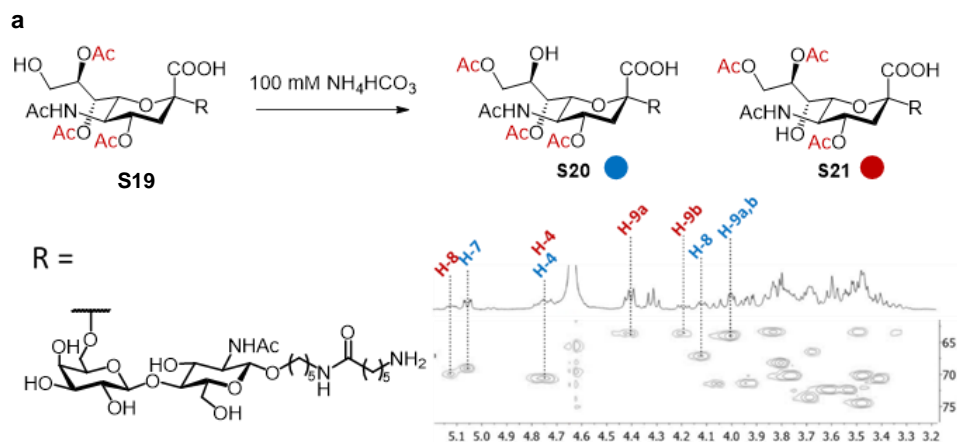


Chapter 2

Supplementary Figure S3 | Chemical synthesis of 2a. The conversions after sialylation was proceeded using **S11**. Abbreviations: DDQ, 2,3-Dichloro-5,6-dicyano-p-benzoquinone; KOH, potassium hydroxide; AcCl, acetyl chloride; PMe_3 , trimethylphosphine; NEt_3 , triethylamine; Ac_2O , acetic anhydride; $\text{Pd}(\text{OH})_2$, palladium(II) hydroxide; AcOH, acetic acid; CsF, cesium fluoride; NHS, N-hydroxysuccinimide; RP-HPLC, reverse-phase high performance liquid chromatography.

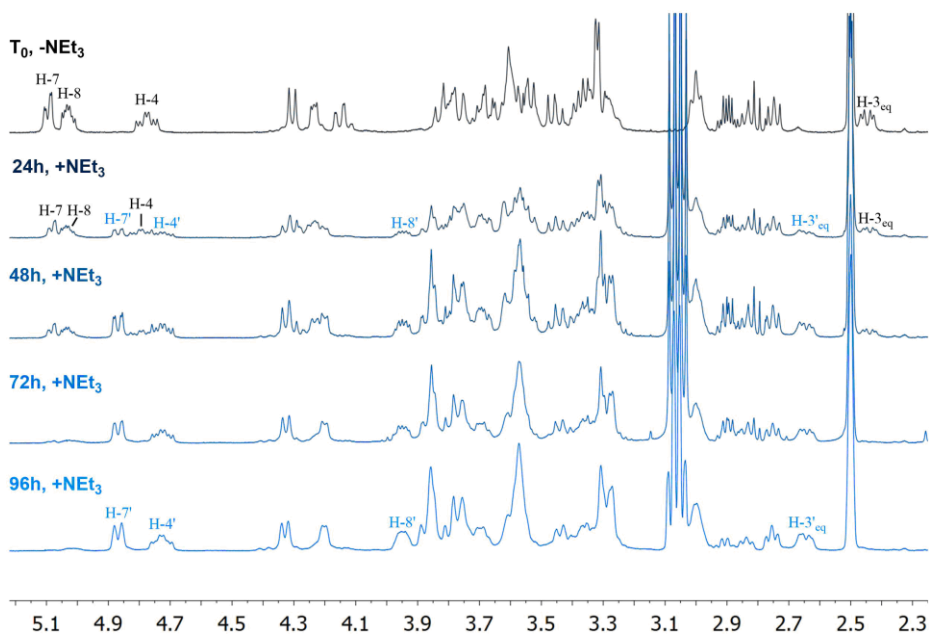


Supplementary Figure S4 | NMR analysis of controlled *O*-acetyl migration. **a**, 4,7,9-tri-*O*-Ac (**S20**) and 4,8,9-tri-*O*-Ac (**S21**) sialoforms resulted from base-mediated *O*-Ac migration. **b**, Comparison of ^1H NMR spectrum of starting material **S19** with that of **S20** and **S21** mixture. **c**, NMR monitoring of acid-mediated *O*-Ac migration. The reaction was originally carried out in 20% AcOD- d_4 in DMSO- d_6 in the presence of ~ 2 equivalents (to **S19**) triethylamine. The reaction conditions were then optimized and performed in one-pot fashion with biotinylation (described in *general procedure C*). **d**, Comparison of 4,7,9-tri-*O*-Ac (**S20**) ^1H NMR spectrum with that of 4,7,8-tri-*O*-Ac (**S19**).

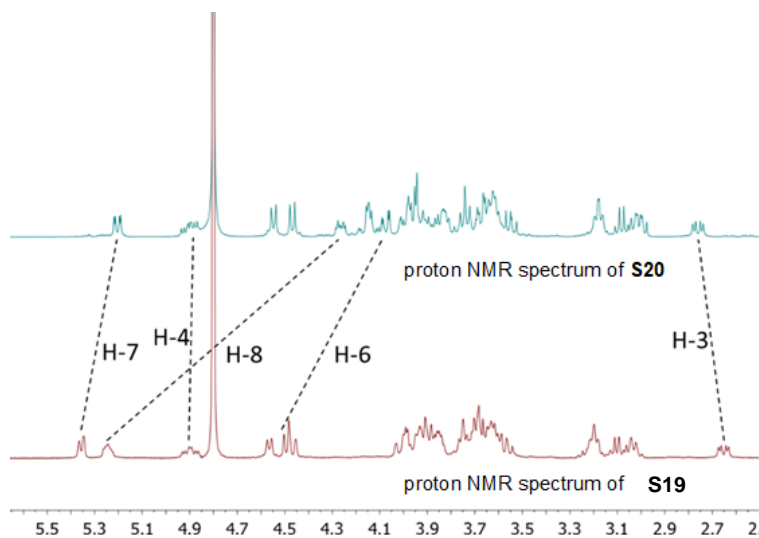


Chapter 2

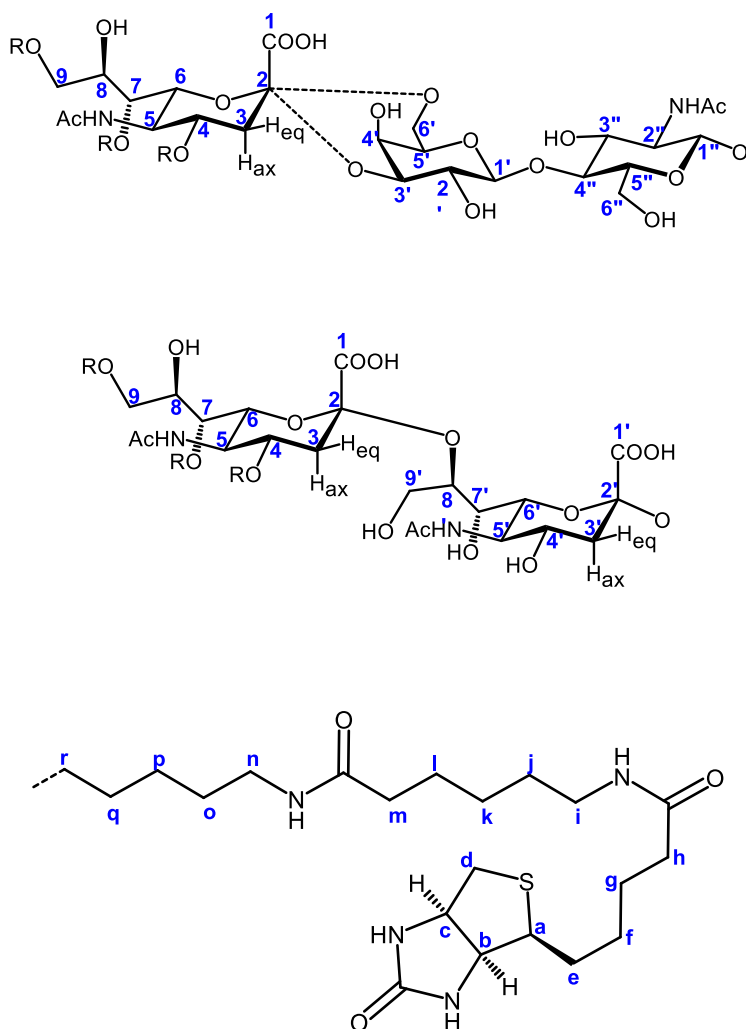
c



d



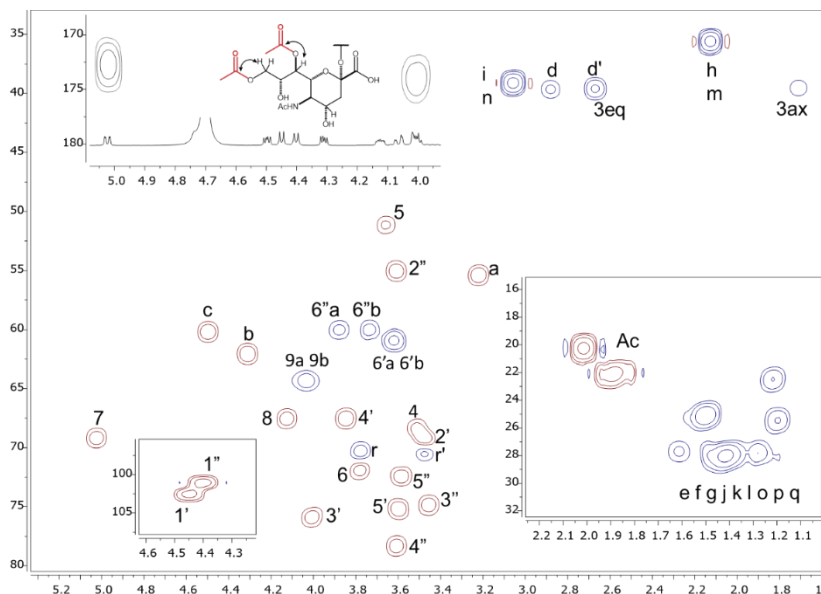
Supplementary Figure S5 | Proton numbering for NMR assignment. α 2,3/6-linked sialyl LacNAc (top), α 2,8-linked disialic acid (middle) and linker-biotin (bottom).



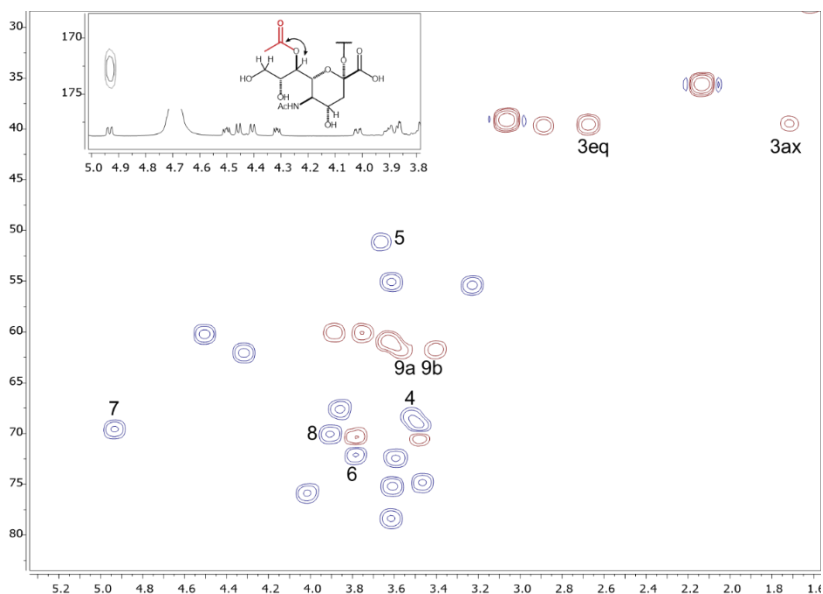
Chapter 2

Supplementary Figure S6 | HSQC and HMBC analyses for 1d (a) and 1f (b). Preservation of 7-*O*-Ac is demonstrated by chemical shifts H-7 of sialic acid in HSQC spectra and correlation of H-7 with carbonyl carbon of acetyl group in HMBC.

a. 7,9-di-*O*-Ac (1d)

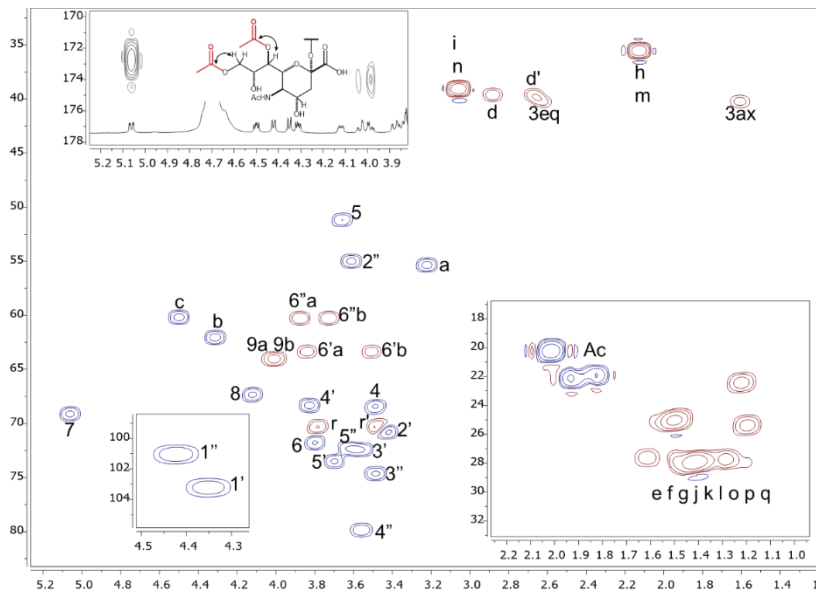
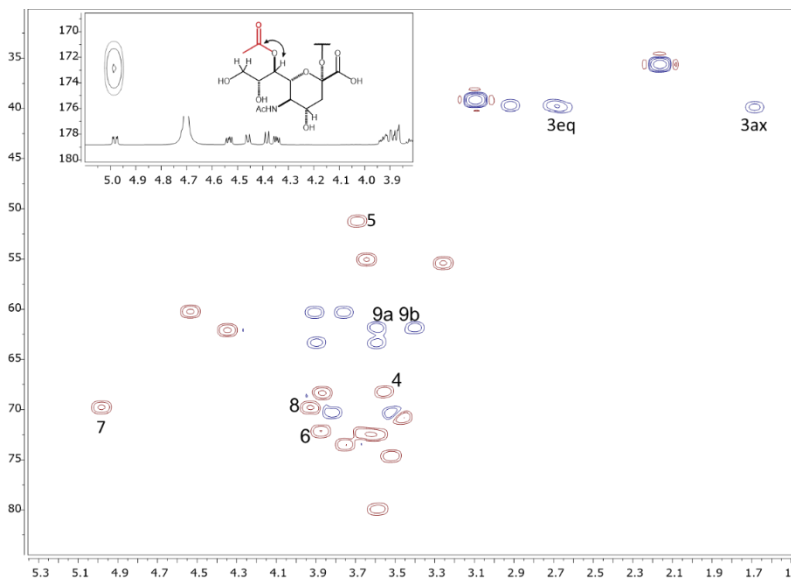


b. 7-*O*-Ac (1f)



Supplementary Figure S7 | HSQC and HMBC analyses for 2d (a) and 2f (b). Preservation of 7-*O*-Ac is demonstrated by H-7 chemical shift of sialic acid in HSQC spectra and correlation of H-7 with carbonyl carbon of acetyl group in HMBC.

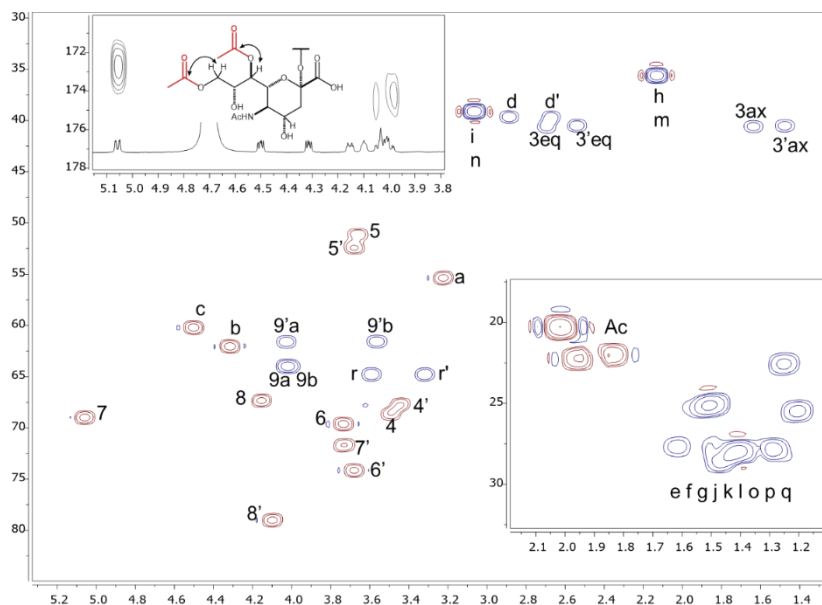
a. 7,9-di-*O*-Ac (2d)

**b. 7-*O*-Ac (2f)**

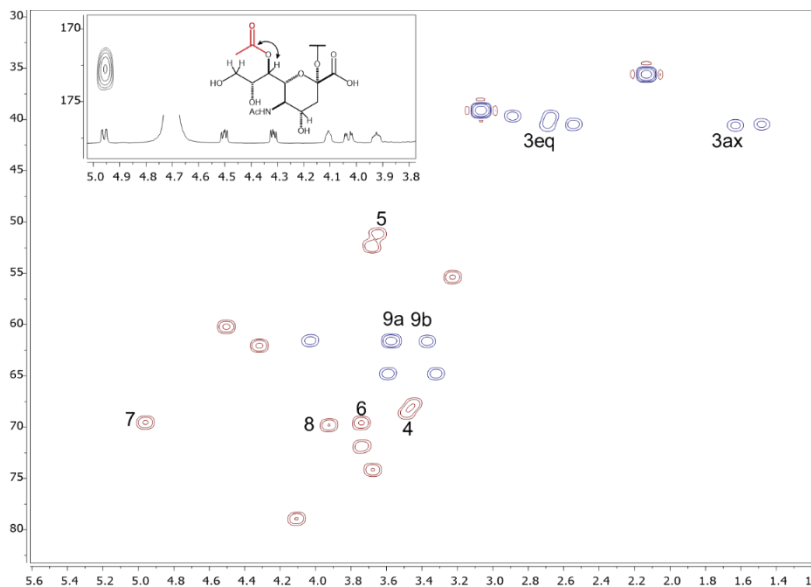
Chapter 2

Supplementary Figure S8 | HSQC and HMBC analyses for 3d (a) and 3f (b). Preservation of 7-*O*-Ac is demonstrated by chemical shift of H-7 of sialic acid in HSQC spectra and correlation of H-7 with carbonyl carbon of acetyl group in HMBC.

a. 7,9-di-*O*-Ac (3d)



b. 7-*O*-Ac (3f)



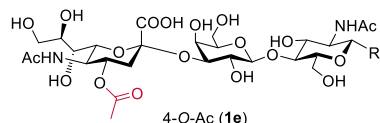
Supplementary Table S2 | NMR assignments. Comparison of ^1H and ^{13}C chemical shifts (ppm) of Sia residue among **1a, c-g, 2a, c-g** and **3a, d, f** and **g**. Proton chemical shifts of *O*-acetylated positions are shown in red. Water peak is referenced at 4.800 ppm.

	H3	C3	H4	C4	H5	C5	H6	C6	H7	C7	H8	C8	H9	C9
1a	$\frac{2.79}{2.00}$	36.29	4.98	70.39	3.94	48.91	4.09	71.57	5.17	69.04	4.28	67.22	4.22-4.15	64.38
1c	$\frac{2.76}{1.91}$	36.71	4.94	71.04	4.08	49.22	3.81	71.22	3.67	67.91	4.10	69.21	4.40-4.18	65.59
1d	$\frac{2.77}{1.81}$	39.53	3.61	68.33	3.76	51.13	3.88	71.92	5.12	69.17	4.22	67.54	4.18-4.09	64.33
1e	$\frac{2.76}{1.91}$	36.64	4.95	71.13	4.08	49.21	3.79	72.40	3.61	67.88	3.90	71.73	3.86-3.64	62.49
1f	$\frac{2.77}{1.81}$	39.46	3.61	68.29	3.76	51.12	3.88	72.10	5.03	69.57	4.00	70.06	3.67-3.50	61.72
1g	$\frac{2.75}{1.79}$	39.69	3.66	68.20	3.85	51.59	3.68	72.56	3.64	68.11	4.09	69.30	4.40-4.18	65.67
2a	$\frac{2.76}{1.90}$	39.96	4.95	70.38	3.96	48.97	4.12	71.58	5.20	69.18	4.28	67.03	4.20-4.13	64.28
2c	$\frac{2.66}{1.82}$	37.04	4.89	70.79	4.05	49.38	3.89	71.88	3.62	68.18	4.11	69.00	4.40-4.20	65.56
2d	$\frac{2.73}{1.71}$	40.22	3.59	68.43	3.76	51.15	3.90	71.76	5.16	69.11	4.22	67.33	4.14-4.07	64.02
2e	$\frac{2.66}{1.83}$	37.06	4.89	70.85	4.04	49.45	3.86	72.10	3.58	68.26	3.90	71.67	3.87-3.63	62.60
2f	$\frac{2.75}{1.78}$	39.85	3.65	68.22	3.79	51.22	3.97	72.14	5.08	69.77	4.02	69.79	3.69-3.49	61.84
2g	$\frac{2.69}{1.74}$	40.04	3.62	68.40	3.84	51.81	3.84	73.73	3.68	68.18	4.12	69.06	4.43-4.23	65.76
3a	$\frac{2.84}{1.90}$	37.49	4.96	70.05	3.93	48.92	3.86	68.34	5.17	68.78	4.23	67.10	4.17-4.10	64.03
3d	$\frac{2.79}{1.74}$	40.58	3.59	68.49	3.76	51.08	3.83	69.62	5.16	68.99	4.25	67.34	4.15-4.08	63.99
3f	$\frac{2.79}{1.73}$	40.58	3.58	68.63	3.74	51.12	3.84	69.58	5.06	69.54	4.02	69.80	3.67-3.47	61.62
3g	$\frac{2.75}{1.72}$	40.45	3.66	68.17	3.82	51.68	3.83	69.55	3.63	68.17	4.10	69.17	4.37-4.20	65.45

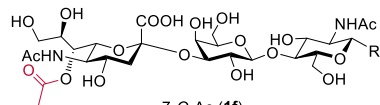
Chapter 2

Supplementary Figure S9 | List of compounds printed as a glycan microarray.

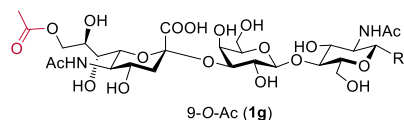
α 2,3-linked sialyl *N*-acetylglucosamine



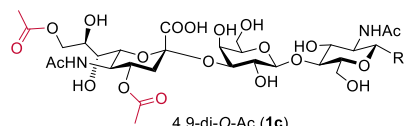
4-O-Ac (1e)



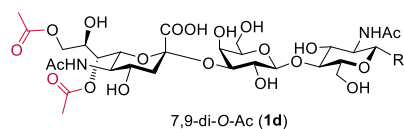
7-O-Ac (1f)



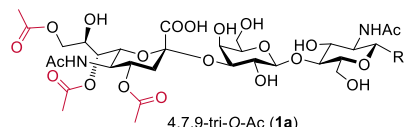
9-O-Ac (1g)



4,9-di-O-Ac (1c)

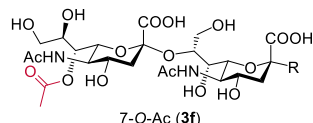


7,9-di-O-Ac (1d)

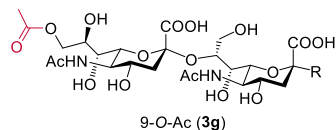


4,7,9-tri-O-Ac (1a)

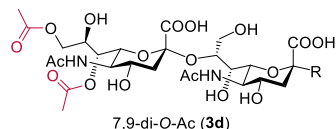
α 2,8-linked disialic acid



7-O-Ac (3f)

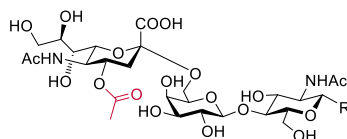


9-O-Ac (3g)

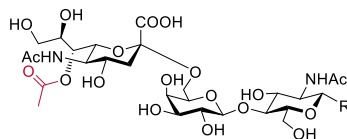


7,9-di-O-Ac (3d)

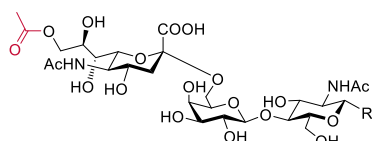
α 2,6-linked sialyl *N*-acetylglucosamine



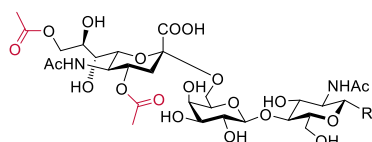
4-O-Ac (2e)



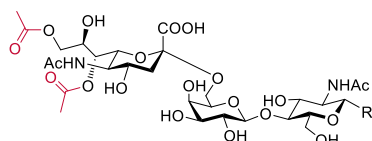
7-O-Ac (2f)



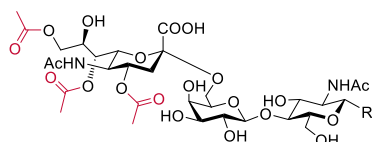
9-O-Ac (2g)



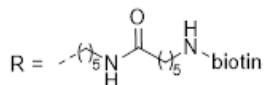
4,9-di-O-Ac (2c)



7,9-di-O-Ac (2d)

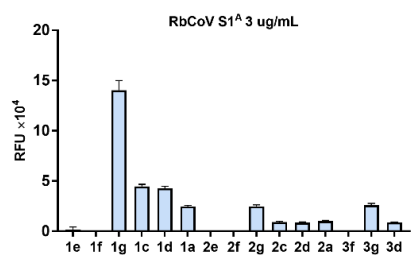
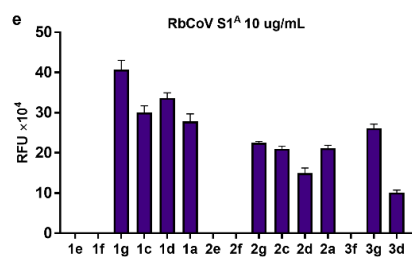
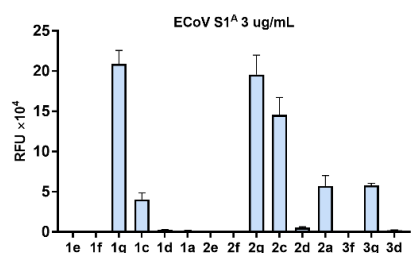
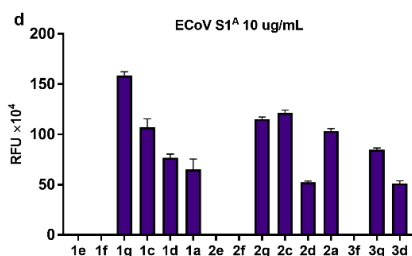
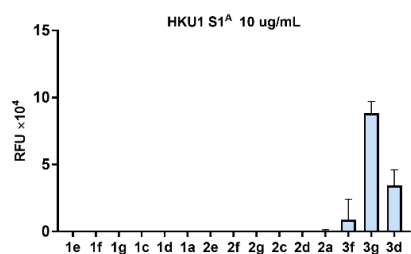
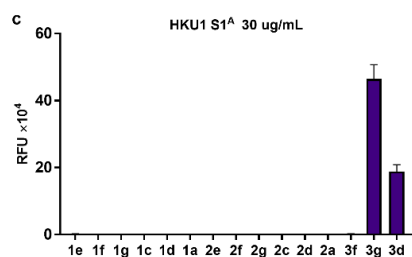
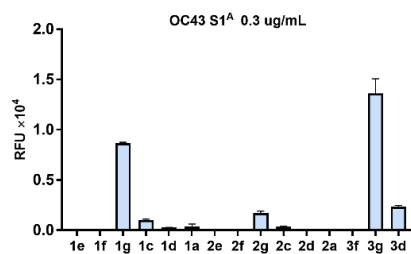
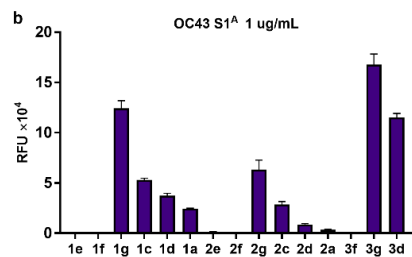
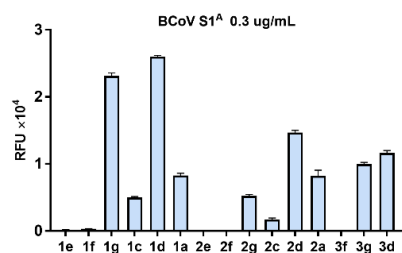
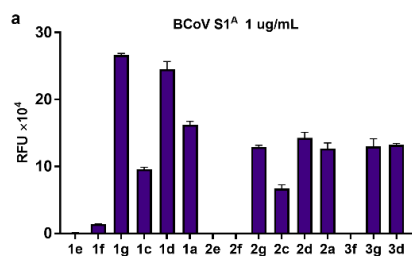


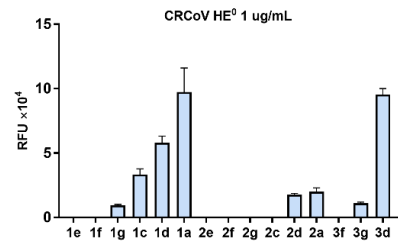
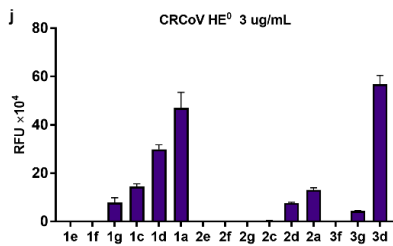
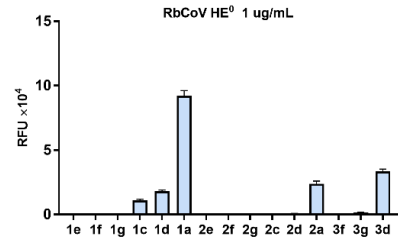
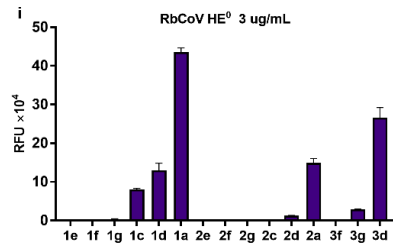
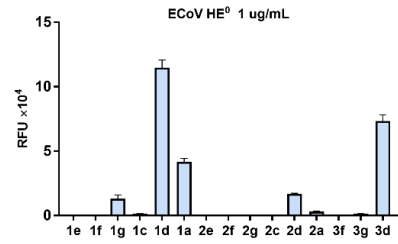
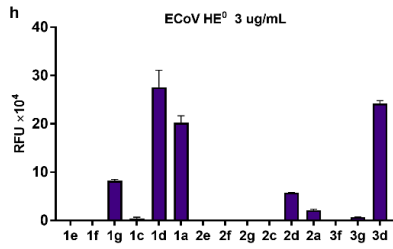
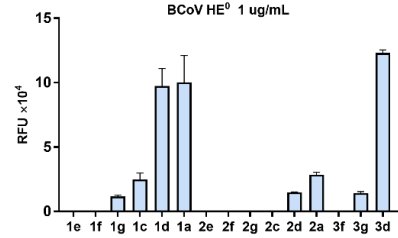
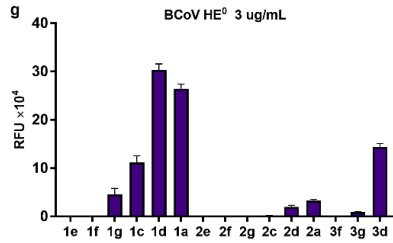
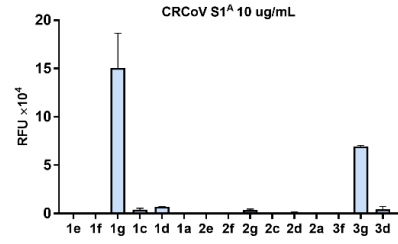
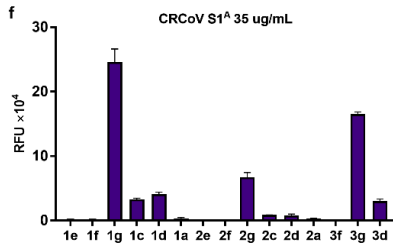
4,7,9-tri-O-Ac (2a)



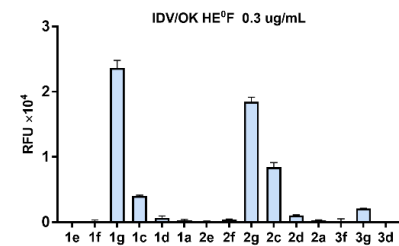
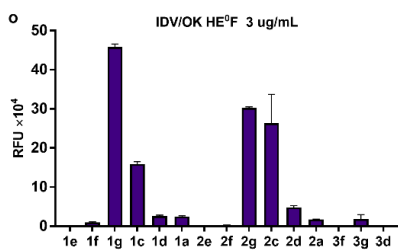
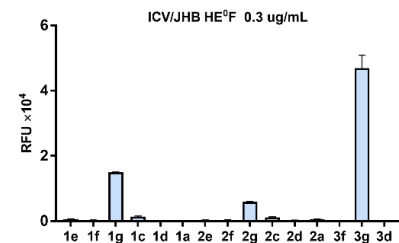
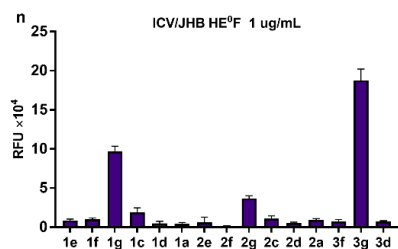
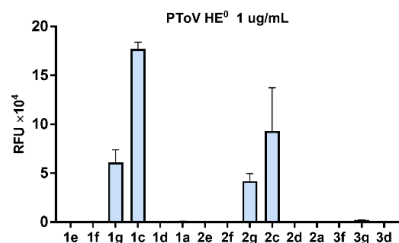
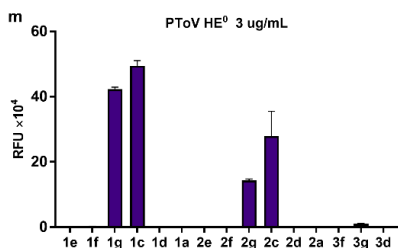
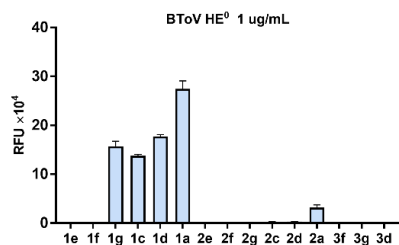
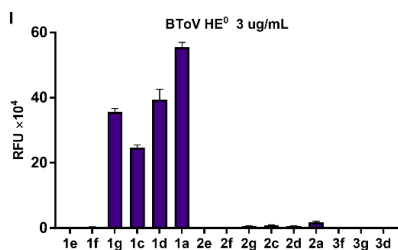
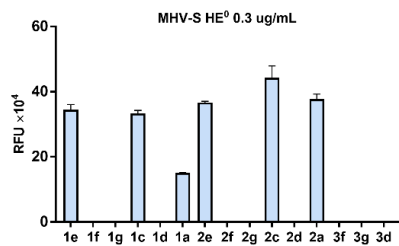
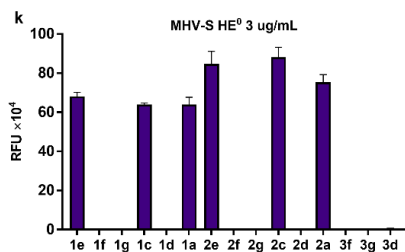
Supplementary Figure S10 | Glycan microarray data of different viral envelope glycoprotein concentrations (figures start on next page). **a-f**, spike (S) proteins of bovine coronavirus (BCoV), human coronavirus OC43 and HKU1, equine coronavirus (ECoV), Rabbit coronavirus (RbCoV) and canine respiratory coronavirus (CRCoV), respectively. **g-m**, HEs (esterase inactive form, denoted as HE⁰) of BCoV, ECoV, RbCoV, CRCoV, mouse hepatitis virus strain S (MHV-S), bovine torovirus (BToV) and porcine torovirus (PToV), respectively. **n** and **o**, Hemagglutinin-esterase-fusion proteins (esterase inactive form, denoted as HE⁰) of influenza C and D viruses, respectively. See Supplementary Table S1 for specification of viruses studied. All proteins are fused with human-Fc tag. Binding was detected using AlexaFluor-647-conjugated anti-human Fc antibody.

Chapter 2

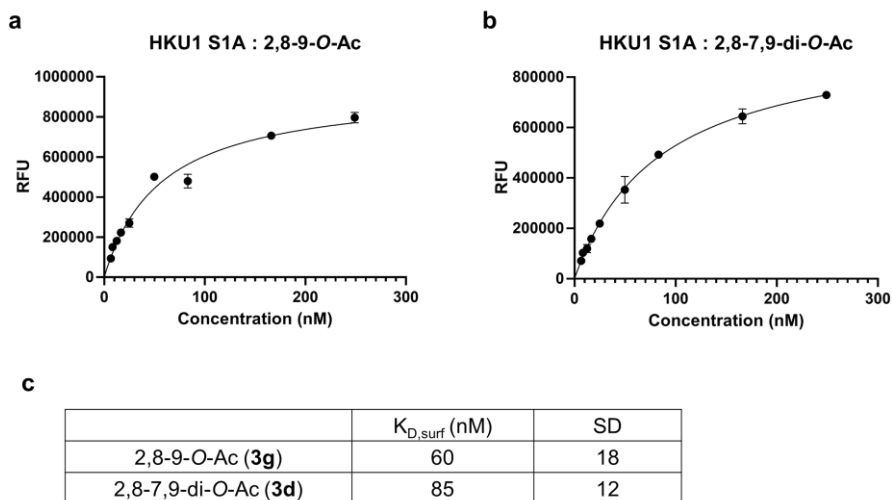




Chapter 2

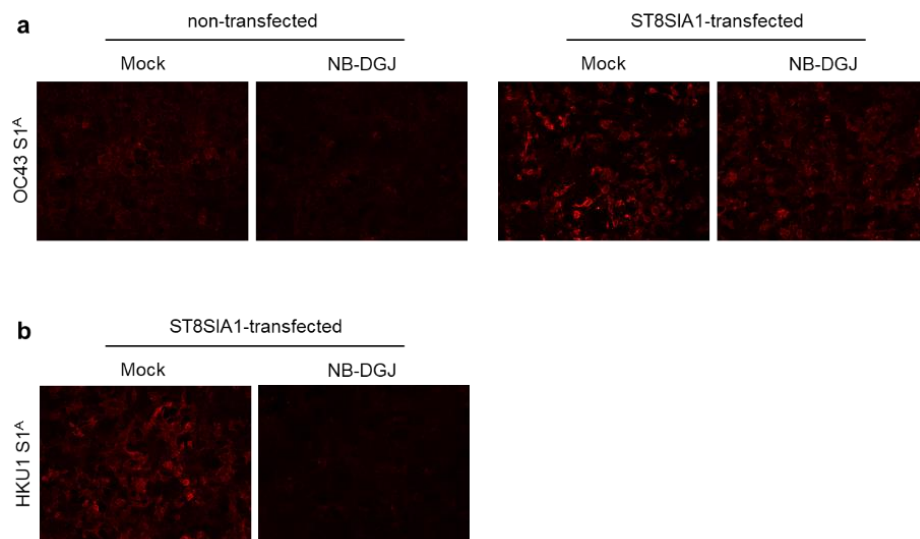


Supplementary Figure S11 | Surface dissociation constants ($K_{D,surf}$) of HKU1 S1^A. **a** and **b**, Representative curves for binding of HKU1 S1^A with 9-*O*-acetylated α 2,8-linked disialic acid (2,8-9-*O*-Ac, **3g**) and 7,9-di-*O*-acetylated α 2,8-linked disialic acid (2,8-7,9-di-*O*-Ac, **3d**), respectively. Curves were fitted using non-linear regression based on one-site specific binding model. **c**, Mean $K_{D,surf}$ values and standard deviation (SD) obtained from three individual technical replicates. These values were in agreement with the relative signal intensities reported in Figs. 4 and S12.



Chapter 2

Supplementary Figure S12 | Immunofluorescent staining of HEK293T cells treated with glycolipid biosynthesis inhibitor. **a**, OC43 S1^A staining of wildtype and ST8Sia1-transfected cells with or without *N*-butyl deoxygalactonojirimycin (NB-DGJ) treatment. **b**, HKU1 S1^A staining of ST8Sia1-transfected cells with or without NB-DGJ treatment. NB-DGJ is an inhibitor of glucosylceramide synthase and glucocerebrosidase and β -glucosidase 2, and suppresses glycolipids such as gangliosides. NB-DGJ was used at a concentration of 1 mM.



Part II. Experimental Section

1. Material and methods

1.1 Chemicals

Trimethylsilyl trifluoromethanesulfonate (TMSOTf, Acros Organics-209440500), sodium methoxide 30% solution in methanol (Sigma-Aldrich-71748), benzaldehyde dimethyl acetal (Sigma-Aldrich-226076), p-toluenesulfonic acid monohydrate (p-TsOH, Alfa Aesar-A14119), Camphor-10-sulfonic acid (β) (CSA, Sigma-Aldrich-147923) NaH 60% dispersion in mineral oil (Sigma-Aldrich-452912), benzyl bromide (BnBr, Acros Organics-105875000), trifluoroacetic acid (TFA, Biosolve Chemie-202341 or ACTU-ALL CHEMICALS-820041301), *N*-iodosuccinimide (NIS, Alfa Aesar-A14320), 2,3-Dichloro-5,6-dicyano-1,4-benzoquinone (DDQ, Acros Organics-113301000), potassium hydroxide (Sigma-Aldrich-757551), trimethylphosphine 1M solution in toluene (Sigma-Aldrich-324116), acetic anhydride (Sigma-Aldrich-242845), ammonium bicarbonate (Sigma-Alrich-09830), ammonium formate (Sigma-Alrich-25204), formic acid (J.T.Baker-0128), acetic acid (Biosolve Chemie-84528), palladium hydroxide Degussa type (Sigma-Aldrich-212911), cesium fluoride (CsF, Sigma-Alrich-198323), NHS-Lc-biotin (Thermo Scientific-21336), biotin (GL Biochem-10703), molecular sieves (MS, 3 angstrom, beads, 4-8 mesh, Sigma-Aldrich-208574). Acetonitrile, dichloromethane and *N,N*-dimethylformamide used for reactions were anhydrous and were taken from the solvent purifier (MB SPS 5). The rest of organic solvents for reactions were AR grade from Biosolve Chemie, and were directly used in reaction unless otherwise noted. The organic solvents in work-up steps were technical grade and were from VWR Chemicals. Deuterated solvents for NMR experiments were from Cambridge Isotope Laboratories. Molecular sieves (MS) were flame dried in a round bottom flask with Bunsen burner and were placed under high-vacuum to allow the activated molecular sieves to cool prior to use.

1.2 General analytical methods

TLC analysis was performed using precoated silica gel 60 F-254 plates (Merck). The plates were either visualized with UV lamp or directly stained with ceric ammonium molybdate (5 g $\text{Ce}(\text{SO}_4)_2$, 25 g $(\text{NH}_4)_6\text{Mo}_7\text{O}_{24} \cdot 4\text{H}_2\text{O}$, 50 mL conc. H_2SO_4 , 450 mL H_2O). Column chromatography was performed using flash silica gel (60 Å, from Silicycle, Canada). Water soluble compounds were purified using p-2 biogel (Biorad) and collected with BioFrac fraction collector (Biorad). ^1H and ^{13}C NMR spectroscopy was conducted on an Agilent 400-MR, VARIAN INOVA-500 or Bruker 600 UltraShield. MS-based reaction monitoring was performed on Shimadzu Kratos axima-CFR MALDI-ToF or Bruker micrOTOF-Q II ESI mass spectrometer. LC-MS traces were collected on Shimadzu LC-ESI-IT-TOF. HRMS data was collected on Agilent 6560 ion mobility Q-ToF MS.

1.3 Recombinant protein expression

List of recombinant proteins:

protein	amino acid residues	Uniprot ID	GenBank	remark
OC43 S	15-306	P36334		
HKU1 S	14-297		ADN03339.1	

Chapter 2

BCoV S	15-302	P15777		
BCoV HE	19-388	P15776		
BCoV HE ⁰	19-388	P15777		Ser 40 to Ala
ECoV S	15-302	Q6W355		
ECoV HE ⁰	18-387	Q6W356		Ser 39 to Ala
CRCoV S	15-302		ACL12996.1	
CRCoV HE ⁰	19-388		ACL12995.1	Ser 40 to Ala
RbCoV S	15-302	H9AA65		
RbCoV HE ⁰	19-388		AFE48816.1	Ser 40 to Ala
MHV-S HE	25-403	P31614		
MHV-S HE ⁰	25-404	P31615		Ser 45 to Ala
BToV HE ⁰	14-392	P0C0W0		Ser 37 to Ala
PToV HE ⁰	22-393		CAE01330.1	Ser 46 to Ala
ICV HE ⁰ F	14-446	P07975		Ser 71 to Thr
IDV HE ⁰ F	15-447	K9LG83		Ser 73 to Thr

Notes: HE⁰ denotes esterase-inactive variant. All proteins used in glycan microarray are tagged with human Fc (see below for sequence). OC43 and HKU1 S used in immunofluorescence assay and tissue staining are tagged with mouse Fc (see below for sequence).

Human Fc tag with linker

DPLVPRGSGGGGDPEPKSCDKTHTCPPCPAPELLGGPSVFLFPPKPKDTMLMSRTPETVCVVVDVSHEDPEVKFNWYVDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTTTPVLDSGDSFFLYSKLTVDKSRWQQGNVFCSCVMHEALHNHYTQKSLSLSPGK

Mouse Fc tag with linker

PLVPRGSGGGGDPEPRGPTIKCPPCKCPAPNLLGGPSVFIFPPKIKD^{VL}MISLSPIVTCVVVDVSEDDPDVQISWVFNNEVHTAQTQTHREDYNSTLRVVSALPIQH^{QD}WMSGKEFKCKVNNKDLPAPIERTISKPKGSVRAPQVYVLPPEEEMTKKQVTLTCMVTDFMPEDIYVEWTNNGKTELNYKNTEPVLDSGDSYFMYSKLRVEKKNWVERNSYSCSVVHEGLHNHHTTKSFSRTPGK

General protocols for protein expression. Human Fc-fused virolectins (S1^A-Fc proteins from BCoV, OC43, HKU1, ECoV, CRCoV and RbCoV; HE⁰-Fc proteins (active site mutated from Ser to Ala) from BCoV, MHV-S, ECoV, CRCoV, RbCoV, BToV and PToV; HEF-Fc s from ICV and IDV; and HE-Fc proteins from BCoV and MHV-S) were expressed following

the reported procedures (Ref. 10 in main text). For OC43 and HKU1 S1^A-mFc proteins used in immunofluorescence assay and tissue array, Fc fusion was with mouse IgG1. Briefly, synthetic human codon-optimized sequences of the Fc-fused virolectins were cloned into pCD5 or pCAGGS expression plasmids. The expression constructs encoding chimeric proteins in which S1^A domain of spike proteins or HE ectodomains were fused to the Fc domain of human IgG1, separated by a thrombin-cleavage site. For protein production, HEK293T cells ($\sim 2 \times 10^7$) were transiently transfected with 20 μ g plasmid DNA conjugated to 2 μ g polyethyleneimine (Polysciences). 16 h post transfection, the transfection mixture was replaced by 293 SFM II expression medium (Invitrogen) supplemented with 44.1 mM sodium bicarbonate, 11.1 mM glucose, Primatone RL-UF (3.0 g/L), penicillin (100 IU/mL), streptomycin (100 μ g/mL), 1% glutaMAX (Gibco), and 1.5% DMSO. Cell supernatants were harvested 6-7 days after transfection and clarified by consecutive centrifugation at 1200rpm, 4°C for 5 min and 4000 rpm, 4°C for 10 min. 150 μ l Protein A-Sepharose (v/v) was added per 50mL supernatant, followed by overnight incubation at 4°C and collection through a Poly-Prep chromatography column (Bio-rad). S1^A-Fc and HE-Fc proteins were eluted with 0.1M citric acid (volume equivalent with Protein A-Sepharose) and neutralized with 1/3 volume of 1M Tris-HCl, pH 8.8. For long-term storage, all proteins were dialyzed against PBS and store at -80°C until use.

1.4 Surface dissociation constants

Microarray experiments to measure surface dissociation constant ($K_{D,surf}$) values were carried out following the general microarray protocol employing protein concentration of 30, 20, 10, 6, 3, 2, 1.5, 1, 0.8, and 0.6 μ g/mL. The experiments were performed as three technical replicates. The fluorescent readings were plotted against the protein concentration and the data was fitted using the equation shown below, where F_{max} is the maximum fluorescence intensity, $[P]$ is the total protein concentration, and $K_{D,surf}$ is the surface equilibrium dissociation constant for surface carbohydrate and protein⁵.

$$F = \frac{F_{max}[P]}{[P] + K_{D,surf}}$$

1.5 Miscellaneous information

Enzymatic reactions were performed in MaxQ 4450 incubator (Thermo Scientific) at 37 °C with shaking. Room temperature was regulated at 21(\pm 1) °C. Chemical structures were made using ChemDraw Professional 16.0. NMR spectra were processed and analyzed using MestReNova. Column graphs, heatmap were plotted using Graphpad Prism 7. Silhouettes in the microarray heatmap are from PhyloPic. The human respiratory tract picture in graphical abstract is from WPClipart.com.

Chapter 2

2. Detailed synthesis

2.1 General synthetic procedures

General procedure A: HE-mediated de-O-acetylation. See also **Methods**. The common precursor **1a**, **2a** or **3a** was dissolved in 50 mM ammonium formate buffer (pH 6.4) in a 1.5 mL eppendorf tube to achieve 5 mM concentration. An HE stock solution was then added. Working concentration for both HEs were 10 μ g/mL. The reaction was incubated at 37°C for 2 hours. The progression of the reaction was followed using ESI-MS. Upon completion, the reaction was lyophilized and purified with reverse phase (RP) -HPLC (C18, acetonitrile/water 5% - 40%, 50 min, for the **1** and **2** series) or HILIC-HPLC (acetonitrile/water 95% - 65%, 60 min, for the **3** series).

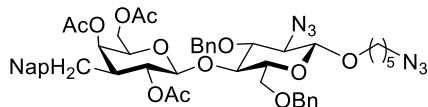
General procedure B: base-mediated acetyl migration. The 7-O-acetylated starting material was dissolved in 100 mM ammonium bicarbonate in D₂O at a concentration of 5 - 10 mM. This buffer gives pH(D) ~8. An aliquot was taken and loaded into an NMR tube. The starting point (*t*₀) proton-NMR spectrum was recorded with water suppression. The solution was placed in an incubator at 37 °C. The reaction was followed by recording the proton-NMR spectra at different time points. The disappearance of Sia H-7 signal around 5.15 ppm indicated C7-to-C9 migration. The solution was then lyophilized to give the 9-O-acetylated product.

General procedure C: hydrogenation, acetyl migration and biotinylation. See also **Methods**. In a double-necked flask, the starting material (**13**, **16** or **S15**) was dissolved in THF/water with containing acetic acid (5 eq., to starting materials). The flask was then filled with hydrogen, followed by addition of palladium hydroxide (Degussa type). Upon completion, the catalyst was filtered off over Celite and the filtrate was concentrated and then freeze-dried. The resulted sialoside carries acetyl esters at C-4, C-7, and C-8-positions. For disialic acid, the internal sialic acid carried an additional 4-O-Ac. The hydrogenated sugar was dissolved in DMSO-*d*₆ in presence of 3 eq. acetic acid-*d*₄ and 6 eq. cesium fluoride. An aliquot was taken for starting point proton NMR. The mixture was heated at 80 °C for 14-20 hours, after which the end point proton NMR was recorded. The shift of Sia H-8 from ~5 ppm to ~4 ppm indicated the reaction was completed. The mixture was then lyophilized. The resulting solid containing 4,7,9-tri-O-Ac sialic acid was directly taken up in DMSO, to which a solution of N-hydroxysuccinimide activated biotin or Lc-biotin was added. Upon completion indicated by ESI-MS, the solution was freeze-dried and the resulted crude mixture was passed through Biogel P-2 to give the biotinylated common precursor in pure form. The sugar was further purified with RP-HPLC (C18, acetonitrile/water 5% - 40%, 50 min, for **1a** and **2a**) or HILIC HPLC (acetonitrile/water 95% - 65%, 60 min, for the **3a**).

General procedure D: DDQ oxidation of 2-naphthylmethyl group^{6,7}. In a round bottom flask filled with argon, the starting material was dissolved in anhydrous DCM at the concentration of 10 mM. Molecular sieves were added. DDQ (1.2 – 2.0 eq. per 2-naphthylmethyl) was then added. The reaction was stirred at 21°C and monitored with TLC (toluene/acetonitrile 1:1). Upon completion, the mixture was poured into saturated sodium bicarbonate solution and washed. The organic phase was dried over sodium sulfate. After filtration, the solution was concentrated and loaded on to silica gel column for purification (toluene/acetonitrile 4:1 – 2:1 or 1% - 2% methanol in DCM).

2.2 Detailed procedures for chemoenzymatic synthesis

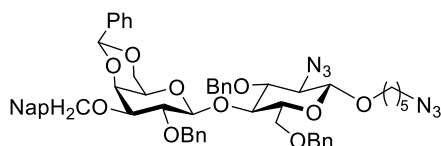
5-azidopentyl [2,4,6-tri-*O*-acetyl-3-*O*-(2-naphthylmethyl)- β -D-galactopyranosyl-(1 \rightarrow 4)]-2-azido-3,6-di-*O*-benzyl-2-deoxy- β -D-glucopyranoside (S4**)**



The glycosyl donor **S1**⁸ (5.3 g, 9.1 mmol) and acceptor **S2**^{9,10} (3 g, 6.0 mmol) were dissolved in anhydrous DCM (200 mL) in a 100 mL round bottom flask containing 4 Å molecular sieves and equipped with an argon balloon. The mixture was

stirred at 21°C for 1 h and was then cooled to -45°C, at which temperature TMSOTf (163 μ L, 0.9 mmol) was added slowly. The reaction was allowed to stir at this temperature for 1 h, followed by quenching with triethylamine. The mixture was then filtered through Celite and the filtrate was diluted with DCM, washed with sat. NaHCO₃ and dried over Na₂SO₄. After filtration, the solvent was removed in vacuo. The crude material was purified through silica gel column (petroleum ether: ethyl acetate 3:1) to give the product **S3** in 76% yield. ¹H NMR (400 MHz, Chloroform-*d*) δ 7.88 – 7.79 (m, 3H, Ar), 7.71 (d, *J* = 1.6 Hz, 1H, Ar), 7.53 – 7.10 (m, 14H, Ar), 5.44 (dd, *J* = 3.5, 1.1 Hz, 1H, H-4_{Gal}), 5.07 (dd, *J* = 10.0, 8.1 Hz, 1H, H-2_{Gal}), 4.98 (d, *J* = 10.7 Hz, 1H, CH₂Ar), 4.83 (d, *J* = 12.6 Hz, 1H, CH₂Ar), 4.74 (d, *J* = 10.6 Hz, 1H, CH₂Ar), 4.67 (d, *J* = 12.2 Hz, 1H, CH₂Ar), 4.51 (d, *J* = 12.6 Hz, 1H, CH₂Ar), 4.45 (d, *J* = 8.1 Hz, 1H, H-1_{Gal}), 4.37 (d, *J* = 12.2 Hz, 1H, CHAr), 4.18 (d, *J* = 7.6 Hz, 1H, H-1_{GlcN}), 4.06 (dd, *J* = 11.2, 7.4 Hz, 1H, H-6a_{GlcN}), 3.99 – 3.80 (m, 3H, H-4_{GlcN}, H-6a_{Gal}, H-6b_{GlcN}), 3.66 (t, *J* = 2.0 Hz, 2H, OCH₂), 3.56 – 3.42 (m, 2H, H-6b_{Gal}, H-5_{GlcN}), 3.38 – 3.18 (m, 6H, H-2_{GlcN}, H-3_{GlcN}, H-3_{Gal}, H-5_{Gal}, CH₂N₃), 2.12 (s, 3H, COCH₃), 2.03 (s, 3H, COCH₃), 1.95 (s, 3H, COCH₃), 1.70 – 1.58 (m, 4H, linker), 1.48 (ddd, *J* = 12.0, 9.3, 5.9 Hz, 2H, linker). HRMS-ESI (pos.) calculated for C₄₈H₅₆N₆O₁₃Na 947.3803, found 947.3804.

5-azidopentyl [2-*O*-benzyl-4,6-*O*-(*L*)-benzylidene-3-*O*-(2-naphthylmethyl)- β -D-galactopyranosyl-(1 \rightarrow 4)]-2-azido-3,6-di-*O*-benzyl-2-deoxy- β -D-glucopyranoside (S5**)**



In a 100 mL round bottom flask, the disaccharide **S4** (2 g, 2.2 mmol) was dissolved in methanol. Catalytic amount of sodium methoxide in methanol was added dropwise. The reaction was stirred at 21 °C for 2 h. Upon completion, acidic resin (Amberlite IR 120 hydrogen form) was

added with vigorous stirring. The resin was then filtered off and the solvent was removed in vacuo to give the de-*O*-acetylated intermediate, which was subjected to benzylidene protection without further purification.

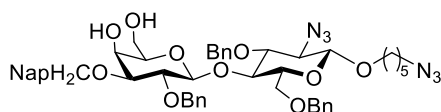
The de-*O*-acetylated disaccharide was dissolved in DMF in a 100 mL round bottom flask, to which benzaldehyde dimethyl acetal (2 eq. to **S4**) and catalytic *p*-toluenesulfonic acid were added. The reaction was stirred at 21 °C for 4 h and was then quenched with triethylamine. DMF was removed in vacuo. The residue was taken up in DCM and washed with sat. NaHCO₃. The organic phase was dried over Na₂SO₄, followed by filtration and removal of DCM to the crude product for the next step.

The resulted compound was directly dissolved in DMF in a 100 mL round bottom flask, followed by addition of benzyl bromide (3 eq. to **S4**), and was cooled to 0 °C, at which temperature sodium hydride (2 eq. to **S4**) was added. The mixture was then allowed to warm

Chapter 2

to 21 °C and stirred for 2 h. Excessive NaH was quenched with acetic acid upon completion. The mixture was then concentrated and resuspended in DCM and washed with sat. NaHCO₃. The organic phase was dried over Na₂SO₄, filtered and concentrated in vacuo. The residue was purified through silica gel column (toluene: ethyl acetate 9:1) to afford the product **S5** (1.7 g) in 82% yield over 3 steps. ¹H NMR (400 MHz, Chloroform-*d*) δ 7.86 – 7.65 (m, 5H, Ar), 7.57 – 7.40 (m, 8H, Ar), 7.38 – 7.17 (m, 13H, Ar), 5.46 (s, 1H, CH₂Ar), 5.21 (d, *J* = 10.5 Hz, 1H, CH₂Ar), 4.87 (m, 2H, CH₂Ar), 4.76 (m, 2H, CH₂Ar), 4.52 (d, *J* = 12.1 Hz, 1H, CH₂Ar), 4.42 (d, *J* = 7.9 Hz, 1H, H-1_{Gal}), 4.29 (d, *J* = 12.1 Hz, 1H, CH₂Ar), 4.24 – 4.14 (m, 2H, H-1_{GlcN}, H-6a_{Gal}), 4.04 (d, *J* = 3.6 Hz, 1H, H-4_{Gal}), 4.01 – 3.75 (m, 5H, H-4_{GlcN}, OCH₂, H-6a_{GlcN}, H-6b_{Gal}, H-2_{Gal}), 3.67 (dd, *J* = 11.0, 1.8 Hz, 1H, H-6b_{GlcN}), 3.51 (dd, *J* = 9.5, 6.5 Hz, 1H, OCH₂), 3.46 – 3.18 (m, 7H, H-3_{Gal}, H-3_{GlcN}, H-2_{GlcN}, H-5_{GlcN}, CH₂N₃), 2.92 (m, 1H, H-5_{Gal}), 1.71 – 1.54 (m, 4H, linker), 1.48 (m, 2H, linker). ¹³C NMR (101 MHz, Chloroform-*d*) δ = 138.75, 138.26, 138.23, 138.02, 135.75, 133.18, 132.99, 128.89, 128.84, 128.26, 128.17, 128.14, 128.12, 128.10, 127.81, 127.76, 127.68, 127.54, 127.50, 127.38, 126.51, 126.38, 126.15, 125.95, 125.72, 102.84, 102.01, 101.34, 81.17, 79.46, 78.74, 77.25, 75.62, 75.36, 75.17, 73.54, 72.94, 71.66, 69.67, 68.87, 67.91, 66.33, 66.05, 51.29, 29.00, 28.54, 23.17. HRMS-ESI (pos.) calculated for C₅₆H₆₀N₆O₁₀Na 999.4269, found 999.4271.

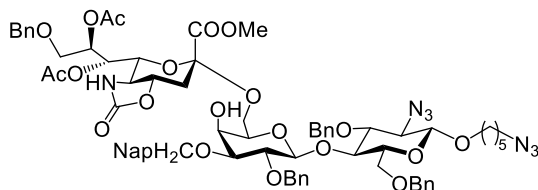
5-azidopentyl [2-*O*-benzyl-3-*O*-(2-naphthylmethyl)-β-*D*-galactopyranosyl-(1→4)]-2-azido-3,6-di-*O*-benzyl-2-deoxy-β-*D*-glucopyranoside (**6**)



To a 100 mL round bottom flask containing compound **S5** (1 g, 1.0 mmol) in DCM and water (99:1) was added trifluoroacetic acid (10% v/v).

The biphasic mixture was then vigorously stirred at 21 °C overnight and was directly poured into cold sat. NaHCO₃. The mixture was gently shaken in a separatory funnel until no gas was produced. The organic phase was collected and dried over Na₂SO₄. Solvent was removed in vacuo after filtration. Silica gel column purification (toluene: ethyl acetate 4:1) gave the diol **6** in 89% yield. ¹H NMR (400 MHz, Chloroform-*d*) δ 7.89 – 7.67 (m, 4H, Ar), 7.52 – 7.39 (m, 5H, Ar), 7.39 – 7.07 (m, 9H, Ar), 5.01 (d, *J* = 10.5 Hz, 1H, CH₂Ar), 4.91 – 4.69 (m, 5H, CH₂Ar), 4.54 (d, *J* = 12.1 Hz, 1H, CH₂Ar), 4.43 – 4.28 (m, 2H, CH₂Ar, H-1_{Gal}), 4.22 (d, *J* = 7.8 Hz, 1H, H-1_{GlcN}), 4.00 – 3.86 (m, 3H, H-4_{GlcN}, H-4_{Gal}, OCH₂), 3.80 (dd, *J* = 10.9, 4.0 Hz, 1H, H-6a_{GlcN}), 3.72 – 3.46 (m, 5H, H-6b_{GlcN}, H-2_{Gal}, H-6a,b_{Gal}, OCH₂), 3.46 – 3.29 (m, 4H, H-2_{GlcN}, H-5_{GlcN}, H-3_{Gal}, H-3_{GlcN}), 3.25 (t, *J* = 6.9 Hz, 2H, CH₂N₃), 3.18 – 3.07 (m, 1H, H-5_{Gal}), 1.63 (tt, *J* = 14.4, 6.7 Hz, 4H, linker), 1.53 – 1.38 (m, 2H, linker). ¹³C NMR (101 MHz, Chloroform-*d*) δ = 138.53, 138.24, 138.06, 135.18, 133.19, 133.04, 128.31, 128.27, 128.06, 127.86, 127.79, 127.77, 127.71, 127.66, 127.60, 127.53, 126.57, 126.24, 126.08, 125.68, 81.18, 80.74, 79.21, 76.53, 75.40, 75.32, 75.14, 73.88, 73.16, 72.15, 69.65, 67.83, 67.10, 65.93, 62.26, 51.32, 29.03, 28.55, 23.19. HRMS-ESI (pos.) calculated for C₄₉H₅₆N₆O₁₀Na 911.3956, found 911.3956.

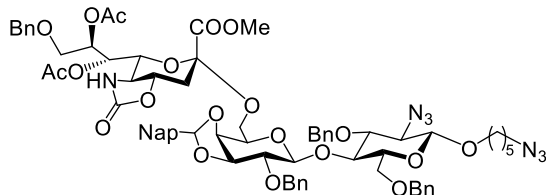
Methyl 5-acetamido-7,8-di-*O*-acetyl-9-*O*-benzyl-5-*N*,4-*O*-carbonyl-3,5-dideoxy-*D*-glycero-α-*D*-galacto-non-2-ulopyranosylonate-(2→6)-5-azidopentyl [2-*O*-benzyl-3-*O*-(2-naphthylmethyl)-β-*D*-galactopyranosyl-(1→4)]-2-azido-3,6-di-*O*-benzyl-2-deoxy-β-*D*-glucopyranoside (**S11**)



The diol acceptor **6** (1.1 g, 1.24 mmol) and sialyl donor **S10** (as α/β mixtures, 1.0 g, 1.48 mmol) were dissolved in anhydrous DCM (25 mL) a 100 mL round bottom flask containing 4 Å molecular sieves and equipped with argon balloon. The mixture was stirred

at 21 °C for 1 h and was then cooled to -45 °C, at which temperature TMSOTf (268 μ L, 1.48 mmol) was added dropwise. The reaction was stirred at this temperature for 10 min and was then quenched with triethylamine. The mixture was filtered through Celite, diluted in DCM and washed with sat. NaHCO_3 . The organic phase was dried over Na_2SO_4 and concentrated. The residue was loaded on to silica gel column for purification (toluene: ethyl acetate 4:1) to give the alpha isomer **S11** in 90% yield. ^1H NMR (400 MHz, Chloroform- d) δ 7.90 – 7.65 (m, 4H, Ar), 7.54 – 7.39 (m, 5H, Ar), 7.38 – 7.10 (m, 23H, Ar), 5.40 (d, 1H, 5-NH), 5.33 – 5.27 (m, 2H, CH_2Ar), 4.97 (d, J = 11.0 Hz, 1H, CH_2Ar), 4.90 (d, J = 11.9 Hz, 1H, CH_2Ar), 4.81 (d, J = 10.5 Hz, 2H, CH_2Ar), 4.78 – 4.69 (m, 2H, CH_2Ar), 4.58 – 4.48 (m, 3H, CH_2Ar), 4.44 (d, J = 7.7 Hz, 1H, H-1 $_{\text{Gal}}$), 4.39 (d, J = 12.3 Hz, 1H, CH_2Ar), 4.28 (d, J = 12.2 Hz, 1H, CH_2Ar), 4.25 – 4.15 (m, 2H, H-1 $_{\text{GlcN}}$, H-6 $_{\text{Sia}}$), 4.07 – 3.84 (m, 4H, H-4 $_{\text{Gal}}$, H-4 $_{\text{GlcN}}$, H-4 $_{\text{Sia}}$, OCH_2), 3.78 – 3.70 (m, 5H, H-6a,b $_{\text{GlcN}}$, COOCH_3), 3.64 – 3.58 (m, 3H, H-6a,b $_{\text{Gal}}$, H-2 $_{\text{Gal}}$), 3.53 – 3.47 (m, 3H, H-9a,b $_{\text{Sia}}$, OCH_2), 3.41 – 3.30 (m, 4H, H-3 $_{\text{Gal}}$, H-2 $_{\text{GlcN}}$, H-3 $_{\text{GlcN}}$, H-5 $_{\text{GlcN}}$), 3.28 – 3.25 (m, 3H, H-5 $_{\text{Gal}}$, CH_2N_3) 2.91 (m, 1H, H-5 $_{\text{Sia}}$) 2.73 (dd, J = 12.1, 3.7 Hz, 1H, H-3eq $_{\text{Sia}}$), 2.38 (d, J = 3.0 Hz, 1H, 4'-OH) 2.15 (s, 3H, COCH_3), 1.94 (dd, J = 12.7 Hz, 1H, H-3ax $_{\text{Sia}}$), 1.90 (s, 3H, COCH_3), 1.76 – 1.54 (m, 6H, linker), 1.55 – 1.40 (m, 2H, linker). ^{13}C NMR (101 MHz, Chloroform- d) δ = 171.07, 169.77, 167.91, 159.27, 138.61, 138.53, 138.15, 137.27, 135.39, 133.22, 133.00, 129.02, 128.48, 128.31, 128.29, 128.26, 128.21, 128.20, 128.08, 128.06, 127.93, 127.84, 127.80, 127.70, 127.59, 127.54, 127.50, 127.46, 126.39, 126.21, 126.05, 125.67, 125.28, 105.01, 102.41, 102.03, 100.18, 80.97, 80.87, 79.28, 77.21, 76.59, 76.10, 75.33, 75.14, 74.68, 73.78, 73.33, 73.14, 72.16, 71.87, 69.69, 69.03, 68.06, 67.26, 65.85, 65.25, 62.79, 58.00, 53.04, 51.34, 36.32, 29.06, 28.57, 23.21, 21.18, 20.64. HRMS-ESI (pos.) calculated for $\text{C}_{71}\text{H}_{81}\text{N}_7\text{O}_{20}\text{Na}$ 1374.5434, found 1374.5436.

5-azidopentyl (Methyl 5-acetamido-7,8-di-O-acetyl-9-O-benzyl-5-N,4-O-carbonyl-3,5-dideoxy-D-glycero- α -D-galacto-non-2-ulopyranosylonate)-(2 \rightarrow 6)-[2-O-benzyl-3,4-O-(2-naphthylmethylene)- β -D-galactopyranosyl-(1 \rightarrow 4)]-2-azido-3,6-di-O-benzyl-2-deoxy- β -D-glucopyranoside (S12**)**



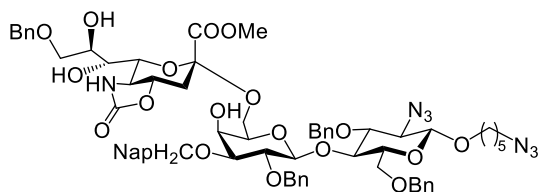
This step was performed according to *general procedure D*. To a 100 mL round bottom flask containing 4 Å molecular sieves and equipped with argon balloon was added the solution of **S11** (1.5 g, 1.11 mmol) in DCM (11 mL), followed addition of DDQ (277

mg, 1.22 mmol). The mixture was stirred for 2.5 h at 21 °C. Upon completion, DCM was added to dilute the mixture, which was then washed sat. NaHCO_3 until the color of organic phase turned light yellow. The DCM solution was collected, dried over Na_2SO_4 and concentrated. The residue was purified through silica gel column (toluene: ethyl acetate 6:1).

Chapter 2

The naphthylidene protected compound **S12** was obtained in 75% yield. ^1H NMR (400 MHz, Chloroform- d) δ 7.85 (td, J = 5.6, 4.9, 3.7 Hz, 4H, Ar), 7.57 – 7.42 (m, 6H, Ar), 7.39 – 7.06 (m, 22H, Ar), 6.06 (s, 1H, CHNap), 5.40 – 5.19 (m, 3H, CH_2Ar), 4.94 (d, J = 11.0 Hz, 1H, CH_2Ar), 4.85 – 4.73 (m, 2H, CH_2Ar), 4.69 (d, J = 11.7 Hz, 1H, CH_2Ar), 4.62 – 4.50 (m, 4H, H-1Gal, CH_2Ar), 4.49 – 4.34 (m, 3H, CH_2Ar , H-3Gal), 4.33 – 4.11 (m, 4H, CH_2Ar , H-1GlcN, H-4GlcN, H-6Sia), 4.04 – 3.79 (m, 6H, H-4GlcN, H-4Sia, OCH_2 , H-6aGal, H-6aGlcN), 3.75 (dd, J = 11.1, 1.8 Hz, 1H, H-6bGlcN), 3.72 – 3.58 (m, 2H, H-6bGal, H-5Gal), 3.54 – 3.47 (m, 3H, H-9aSia, OCH_2 , H-2Gal), 3.43 – 3.38 (m, 7H, H-9bSia, H-5GlcN, H-2GlcN, H-3GlcN, COOCH_3), 3.27 (t, J = 6.9 Hz, 2H, CH_2N_3), 2.93 (dd, 1H, H-5Sia), 2.82 (dd, J = 12.1, 3.7 Hz, 1H, H-3eqSia), 2.10 (s, 3H, COCH_3), 1.96 (dd, J = 13.3 Hz, 1H, H-3aXSia), 1.85 (s, 3H, COCH_3), 1.75 – 1.56 (m, 6H, linker), 1.55 – 1.40 (m, 2H, linker). ^{13}C NMR (101 MHz, Chloroform- d) δ = 171.12, 169.60, 167.91, 159.31, 138.41, 138.15, 137.97, 137.32, 135.78, 133.75, 132.84, 129.02, 128.52, 128.46, 128.38, 128.36, 128.30, 128.21, 128.19, 128.16, 128.07, 127.90, 127.89, 127.80, 127.77, 127.72, 127.66, 127.61, 127.57, 126.60, 126.41, 125.94, 125.28, 123.75, 103.52, 102.13, 101.78, 100.15, 81.10, 80.01, 77.76, 77.22, 76.13, 75.09, 74.86, 73.77, 73.45, 73.38, 73.33, 73.21, 71.50, 69.74, 68.99, 68.20, 68.08, 67.27, 65.89, 63.81, 58.06, 52.76, 51.34, 37.12, 29.07, 28.58, 23.22, 21.46, 21.13, 20.55. HRMS-ESI (pos.) calculated for $\text{C}_{71}\text{H}_{79}\text{N}_7\text{O}_{20}\text{Na}$ 1372.5278, found 1372.5276.

Methyl 5-acetamido-9-*O*-benzyl-5-*N*,4-*O*-carbonyl-3,5-dideoxy- α -D-galactono-2-ulopyranosylonate-(2 \rightarrow 6)-5-azidopentyl [2-*O*-benzyl-3-*O*-(2-naphthylmethyl)- β -D-galactopyranosyl-(1 \rightarrow 4)]-2-azido-3,6-di-*O*-benzyl-2-deoxy- β -D-glucopyranoside (**S17**)

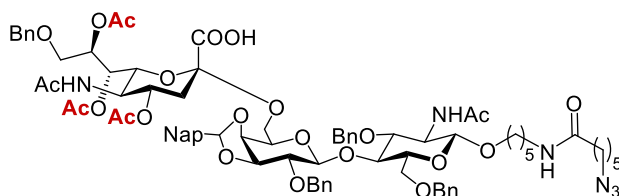


The sialyl donor **4** (415 mg, 0.66 mmol) and the diol acceptor **6** (450 mg, 0.51 mmol) and were dissolved in anhydrous DCM (23 mL) a 100 mL round bottom flask containing 4 Å molecular sieves and equipped with argon balloon. The mixture was stirred

at 21 °C for 1 h and was then cooled to -78°C. TMSOTf (120 μL , 0.66 mmol) was added dropwise. The reaction was stirred at -78°C for 5 min and was then quenched with triethylamine. The mixture was filtered through Celite, diluted in DCM and washed with sat. NaHCO_3 . The organic phase was dried over Na_2SO_4 and concentrated. The residue was taken up with DCM, followed by adding a few drops of water to make a biphasic mixture. TFA (10 - 15% v/v) was added, and the mixture was stirred vigorously at 21 °C for two hours. The mixture was then directly poured into cold sat. NaHCO_3 . The organic phase was collected when no more CO_2 was generated and dried over Na_2SO_4 . The concentrated solution was loaded on to a silica gel column for purification (toluene: ethyl acetate 3:1 - 2:1) to afford **S17** in 69% yield over two steps. ^1H NMR (400 MHz, Chloroform- d) δ 7.87 – 7.68 (m, 3H, Ar), 7.54 – 7.38 (m, 4H, Ar), 7.38 – 7.13 (m, 13H, Ar), 5.63 (s, 1H, NH), 4.94 (d, J = 10.4 Hz, 1H, CH_2Ar), 4.90 – 4.65 (m, 4H, CH_2Ar), 4.55 (d, J = 12.1 Hz, 1H, CH_2Ar), 4.51 – 4.41 (m, 2H, CH_2Ar), 4.41 – 4.30 (m, 2H, CH_2Ar , H-1Gal), 4.20 (d, J = 7.8 Hz, 1H, H-1GlcN), 4.02 – 3.85 (m, 5H, H-4GlcN, H-4Gal, H-6Sia, H-4Sia, OCH_2), 3.84 – 3.70 (m, 5H, COOCH_3 , H-6aGlcN, H-7Sia), 3.70 – 3.46 (m, 8H, H-9a,bSia, H-6bGlcN, H-6aGal, H-2Gal, H-6bGal, OCH_2 , H-8Sia), 3.46 – 3.17 (m, 7H, H-3GlcN, H-5Sia, H-3Gal, H-5GlcN, CH_2N_3 , H-5Gal), 2.76 (dd, J = 12.1, 3.7 Hz,

1H, H-3eq_{Sia}), 1.97 (dd, J = 12.7 Hz, 1H, H-3ax_{Sia}), 1.76 – 1.38 (m, 6H, linker). ¹³C NMR (101 MHz, cdcl₃) δ = 168.39, 159.61, 138.41, 138.39, 138.07, 137.87, 137.57, 135.13, 133.22, 133.07, 129.04, 128.50, 128.36, 128.34, 128.32, 128.23, 128.18, 127.92, 127.89, 127.84, 127.82, 127.74, 127.67, 127.63, 127.57, 126.52, 126.30, 126.15, 125.64, 125.30, 102.45, 102.09, 100.04, 81.22, 80.52, 79.23, 77.24, 77.12, 76.02, 75.39, 75.17, 75.13, 73.55, 73.21, 72.17, 72.15, 72.09, 71.17, 69.73, 69.69, 67.76, 65.75, 65.17, 61.83, 57.44, 53.34, 51.34, 35.30, 29.05, 28.57, 23.21. HRMS (pos.) calculated for C₆₇H₇₇N₇O₁₈NH₄ 1285.5669, found 1285.5676.

5-(6-azidohexanamido)pentyl (5-acetamido-4,7,8-tri-*O*-acetyl-9-*O*-benzyl-3,5-dideoxy-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid)-(2 \rightarrow 6)-[2-*O*-benzyl-3,4-*O*-(2-naphthylmethylene)- β -D-galactopyranosyl-(1 \rightarrow 4)]-2-acetamido-3,6-di-*O*-benzyl-2-deoxy- β -D-glucopyranoside (S15**)**



The trisaccharide **S12** was dissolved in dioxane: water 4:1, followed by addition of solid KOH (10 eq. to **S12**). The reaction was heated up to 60 °C and stirred at this temperature overnight. Upon completion

(by MALDI-MS), the reaction mixture was cooled to 0 °C and acetyl chloride (4 eq. to **S12**) in dioxane was added dropwise over 5 min, after which TLC (ethyl acetate: methanol: water 9:1:0.5) showed full N-acetylation. More KOH (5 eq. to **S12**) was added into the mixture. At 0 °C, 1 M trimethylphosphine solution in toluene (5 eq. to **S12**) was added dropwise. The mixture was stirred for 2 h until TLC showed full consumption of the bisazide. The suspension was concentrated in vacuo, taken up in DCM and transferred into a 50 mL Falcon tube. After centrifugation, the supernatant was then transferred into a new round bottom flask. The precipitate was washed 3 times with DCM and the wash solution was combined in the flask. The clear solution containing the bisamine **S13** was concentrated and subjected directly to the regioselective acylation with activated ester.

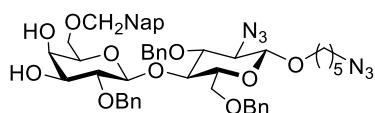
The bisamine compound **S13** from the one-pot deprotection was dissolved in DMF and the solution was cooled to 0 °C, followed by the addition of active ester **S14**¹¹ (1.5 eq. to **S12**) in DMF dropwise. TLC (ethyl acetate: ethanol: water 4:2:1) showed full conversion of the bisamine, upon which KOH (0.8 eq. to **S12**) was added to quench the leftover active ester. DMF was removed in vacuo to afford the crude product for the next step without purification.

The resulted residue was dissolved in pyridine, followed by the addition of acetic anhydride. The reaction was heated up to 50 °C, stirred overnight and then quenched with water. The solvent was removed by co-evaporating with toluene for 5 times. The residue was directly loaded on the silica gel column for purification (ethyl acetate: ethanol: acetic acid 19:1:0.1) to afford the product **S15** in 54% yield over 3 steps. ¹H NMR (400 MHz, Methanol-d₄) δ 7.93 – 7.71 (m, 4H, Ar), 7.50 – 7.37 (m, 3H, Ar), 7.36 – 7.05 (m, 20H, Ar), 6.04 (s, 1H, CHNap), 5.38 – 5.16 (m, 2H, H-7_{Sia}, H-8_{Sia}), 4.91 – 4.80 (m, 2H, H-4_{Sia}, CH₂Ar), 4.71 (d, J = 11.7 Hz, 1H, CH₂Ar), 4.64 – 4.52 (m, 2H, CH₂Ar, H-1_{Gal}), 4.48 – 4.41 (m, 2H, CH₂Ar, H-1_{GlcN}), 4.38 – 4.13 (m, 6H, H-3_{Gal}, CH₂Ar, H-4_{Gal}, H-4_{GlcN}), 4.01 – 3.27 (m, 14H, H-5_{Gal}, H-6a,b_{Gal}, H-6_{Sia}, H-9a, b_{Sia}, OCH₂, H-6a,b_{GlcN}, H-2_{Gal}, H-5_{GlcN}), 3.20 (t, J = 6.8 Hz, 2H,

Chapter 2

$\text{CH}_2\text{N-linker}$), 3.08 (p, $J = 6.5$ Hz, 2H, $\text{CH}_2\text{N-linker}$), 2.56 (dd, $J = 12.6, 4.7$ Hz, 1H, $\text{H-3eq}_{\text{Sia}}$), 2.11 (t, $J = 7.4$ Hz, 2H, $\text{CH}_2\text{C(O)NH-linker}$), 1.94 (s, 3H, COCH_3), 1.89 (s, 3H, COCH_3), 1.80 (s, 3H, COCH_3), 1.78 (s, 3H, COCH_3), 1.75 (s, 3H, COCH_3), 1.66 (t, $J = 12.4$ Hz, 1H, $\text{H-3ax}_{\text{Sia}}$), 1.60 – 1.25 (m, 12H, linker). ^{13}C NMR (101 MHz, cd_3od) $\delta = 176.03, 173.58, 173.33, 171.97, 171.83, 171.77, 170.68, 140.49, 139.87, 139.76, 139.32, 137.84, 135.31, 134.55, 129.66, 129.50, 129.49, 129.46, 129.43, 129.40, 129.37, 129.23, 129.19, 128.91, 128.88, 128.79, 128.65, 127.61, 127.41, 126.79, 125.13, 104.83, 103.26, 102.78, 100.66, 81.67, 81.40, 79.44, 78.19, 76.52, 75.51, 74.87, 74.47, 74.36, 74.30, 73.76, 73.24, 71.28, 70.80, 70.63, 70.02, 69.91, 69.77, 65.26, 58.47, 56.34, 52.45, 50.35, 40.46, 39.15, 37.07, 33.58, 30.39, 30.16, 29.77, 27.52, 26.71, 24.61, 23.24, 22.81, 21.49, 21.10, 20.94, 18.52$. HRMS-ESI (pos.) calculated for $\text{C}_{81}\text{H}_{98}\text{N}_6\text{O}_{23}\text{Na}$ 1545.6581, found 1545.6585.

5-azidopentyl [2-*O*-benzyl-6-*O*-(2-naphthylmethyl)- β -D-galactopyranosyl-(1 \rightarrow 4)]-2-azido-3,6-di-*O*-benzyl-2-deoxy- β -D-glucopyranoside (**5**)

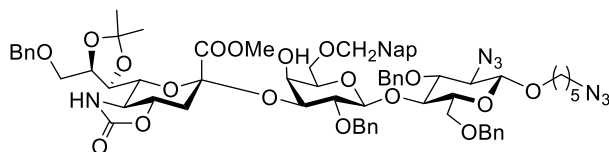


The Gal-imidate donor **S3**¹² (3.2 g, 5.5 mmol) and acceptor **S1** acceptor (2.3 g, 4.6 mmol) were dissolved in anhydrous DCM (50 mL) in a 250 mL round bottom flask containing 4 Å molecular sieves and equipped with an argon balloon. The mixture was stirred at 21 °C for 1 h and was then cooled to -45°C. TMSOTf (100 μL , 0.55 mmol) was added slowly at this temperature. The reaction was allowed to for 30 min. Upon completion indicated by TLC, the reaction was quenched with triethylamine. The mixture was then filtered through Celite and the filtrate was diluted with DCM, washed with sat. NaHCO_3 and dried over Na_2SO_4 . After filtration, the solvent was evaporated in vacuo. The residue was directly used in the next step without further purification.

In a 100 mL round bottom flask, the disaccharide was dissolved in methanol. Catalytic amount of sodium methoxide (30% w/v in methanol) was added dropwise. The reaction was stirred at 21 °C for 2 h. When TLC indicated completion, Amberlite IR 120 resin was added with vigorous stirring. The resin was then removed and the solution was concentrated in vacuo to give the de-*O*-acetylated product. The de-*O*-acetylated disaccharide was dissolved in acetone 2,2-dimethoxypropane (46 mL) in a 100 mL round bottom flask, to which and 0.2 eq. *p*-toluenesulfonic acid were added. The reaction was stirred at 21 °C for 1 h and was then quenched with triethylamine. The solution was concentrated in vacuo. The residue was taken up in DCM and washed with sat. NaHCO_3 . The organic phase was dried over Na_2SO_4 , followed by filtration and evaporation of DCM to give the crude material for the next step. The compound from isopropylidene protection was dissolved in DMF in a 100 mL round bottom flask, followed by addition of benzyl bromide (3 eq.), and was cooled to 0 °C, at which temperature sodium hydride (60%, 2 eq.) was added. The mixture was then allowed to warm to 21 °C and stirred for 2 h. Upon completion, excessive NaH was quenched with acetic acid. The mixture was then concentrated, taken up in DCM and washed with sat. NaHCO_3 . The organic phase was dried over Na_2SO_4 , filtered and concentrated in vacuo. The crude was then taken up in methanol/DCM (v/v 9:1, 50 mL), to which 0.5 eq. *p*-toluenesulfonic acid was added. The solution was then stirred at 21 °C overnight and was directly poured into cold sat. NaHCO_3 and washed. The organic phase was collected and dried over Na_2SO_4 . Solvent was evaporated in vacuo. Silica gel column purification (toluene: ethyl acetate 4:1) afforded the acceptor **S14** for sialylation in 54% yield over 5 steps. ^1H

NMR (400 MHz, Chloroform-*d*) δ 7.85 – 7.77 (m, 3H, Ar), 7.69 (s, 1H, Ar), 7.53 – 7.09 (m, 18H, Ar), 5.01 (d, *J* = 10.7 Hz, 1H, *CH*₂Ar), 4.82 (d, *J* = 11.6 Hz, 1H, *CH*₂Ar), 4.75 (d, *J* = 10.8 Hz, 1H, *CH*₂Ar), 4.68 (d, *J* = 11.6 Hz, 1H, *CH*₂Ar), 4.64 – 4.51 (m, 4H, *CH*₂Ar), 4.49 – 4.41 (m, 2H, *CH*₂Ar, Gal-H1), 4.25 – 4.18 (d, *J* = 6.9, 1H, GlcN-H1), 4.09 – 3.96 (m, 2H, GlcN-H4, Gal-H3), 3.96 – 3.87 (m, 1H, linker-OCH₂), 3.82 (dd, *J* = 11.1, 3.9 Hz, 1H, GlcN-H6a), 3.77 – 3.62 (m, 2H, GlcN-H6b, Gal-H6a), 3.58 – 3.31 (m, 8H, Gal-H6, linker-OCH₂, Gal-H2, Gal-H4, GlcN-H2, GlcN-H5, GlcN-H3, Gal-H5), 3.27 (t, *J* = 6.9, 2H, linker-CH₂N₃), 2.60 (s, 1H, OH), 2.47 (s, 1H, OH), 1.72 – 1.58 (m, 4H, linker), 1.49 (m, 2H, linker). ¹³C NMR (101 MHz, *cd*Cl₃) δ = 138.37, 138.31, 138.06, 135.38, 133.25, 133.01, 128.54, 128.34, 128.28, 128.25, 128.13, 127.90, 127.87, 127.70, 127.63, 127.53, 126.41, 126.14, 125.92, 125.54, 102.60, 102.05, 81.11, 80.03, 76.17, 75.21, 75.10, 74.95, 73.61, 73.48, 73.23, 73.02, 69.65, 68.83, 68.74, 67.96, 65.90, 51.34, 29.04, 28.57, 23.21. HRMS-ESI (pos.) calculated for C₄₉H₅₆N₆O₁₀NH₄ 906.4402, found 906.4421.

Methyl 5-acetamido-9-*O*-benzyl-5-*N*,4-*O*-carbonyl-3,5-dideoxy-7,8-*O*-isopropylidene- α -D-galacto-non-2-ulopyranosylonate(2 \rightarrow 3)-5-azidopentyl [2-*O*-benzyl-6-*O*-(2-naphthylmethyl)- β -D-galactopyranosyl-(1 \rightarrow 4)]-2-azido-3,6-di-*O*-benzyl-2-deoxy- β -D-glucopyranoside (S18)



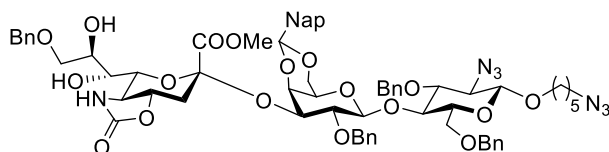
The diol acceptor **5** (1.0 g, 1.13 mmol) and sialyl donor **4** (850 mg, 1.35 mmol) were dissolved in anhydrous DCM (23 mL) a 100 mL round bottom flask containing 4 Å molecular sieves

and equipped with argon balloon. The mixture was stirred at 21 °C for 1 h and was then cooled to -78 °C, at which temperature TMSOTf (240 μ L, 1.35 mmol), diluted in DCM, was added dropwise. The reaction was stirred at this temperature for 10 min and was then quenched with triethylamine. The mixture was filtered through Celite, diluted in DCM and washed with sat. NaHCO₃. The organic phase was dried over Na₂SO₄ and concentrated. The residue was loaded on to silica gel column for purification (toluene: ethyl acetate 4:1) to give the α isomer **S18** in 89% yield. ¹H NMR (600 MHz, Chloroform-*d*) δ 7.82 – 7.71 (m, 3H, Ar), 7.67 (s, 1H, Ar), 7.49 – 7.42 (m, 5H, Ar), 7.40 – 7.12 (m, 18H, Ar), 5.03 – 4.97 (m, 2H, *CH*₂Ar, Sia-NH), 4.70 (d, *J* = 10.5 Hz, 2H, *CH*₂Ar), 4.64 (d, *J* = 11.3 Hz, 1H, *CH*₂Ar), 4.60 (d, *J* = 12.2 Hz, 1H, *CH*₂Ar), 4.55 (d, *J* = 12.1 Hz, 1H, *CH*₂Ar), 4.51 (m, 2H, *CH*₂Ar), 4.50 – 4.43 (m, 2H, Sia-H8, Gal-H1), 4.41 (d, *J* = 12.1 Hz, 1H, *CH*₂Ar), 4.19 (d, *J* = 7.3 Hz, 1H, GlcN-H1), 4.10 – 3.86 (m, 8H, Sia-H7, Gal-H3, Sia-H4, Sia-H6, Gal-H4, GlcN-H4, Sia-H9a, linker-OCH₂), 3.82 (dd, *J* = 10.4, 6.6 Hz, 1H, Sia-H9b), 3.79 – 3.72 (m, 4H, GlcN-H6a, Sia-COOCH₃), 3.68 (m, 2H, GlcN-H6b, Gal-H6a), 3.60 – 3.47 (m, 4H, Sia-H5, Gal-H2, Gal-H6b, linker-OCH₂), 3.44 – 3.38 (m, 1H, Gal-H5), 3.38 – 3.22 (m, 5H, GlcN-H2, GlcN-H3, GlcN-H5, linker-CH₂N₃), 2.77 (dd, *J* = 11.8, 3.3 Hz, 1H, Sia-H3a), 2.47 (s, 1H, OH), 2.19 (t, *J* = 12.6 Hz, 1H, Sia-H3), 1.70 – 1.59 (m, 4H, linker), 1.48 (h, *J* = 6.9 Hz, 2H, linker), 1.32 (s, 3H, isopropylidene), 1.25 (s, 3H, isopropylidene). ¹³C NMR (151 MHz, CDCl₃) δ = 168.81, 159.45, 138.30, 138.12, 137.89, 135.73, 133.26, 132.92, 128.53, 128.43, 128.40, 128.34, 128.30, 128.12, 128.06, 127.91, 127.88, 127.82, 127.79, 127.77, 127.66, 127.52, 127.43, 126.14, 126.03, 125.77, 125.57, 109.05, 102.42, 102.01, 100.30, 81.15, 78.06, 76.15, 76.10, 75.97, 75.56, 75.28, 75.13, 73.52, 73.45, 73.16, 72.87, 69.63, 68.85, 68.68, 68.10,

Chapter 2

67.96, 65.86, 58.12, 53.28, 51.35, 36.13, 29.06, 28.57, 26.74, 24.49, 23.22. HRMS-ESI (pos.) calculated for $C_{70}H_{81}N_7O_{18}NH_4$ 1325.5982, found 1325.6090.

5-azidopentyl (Methyl 5-acetamido-9-*O*-benzyl-5-*N*,4-*O*-carbonyl-3,5-dideoxy-*D*-glycero- α -*D*-galacto-non-2-ulopyranosylonate)-(2 \rightarrow 3)-[2-*O*-benzyl-4,6-*O*-(2-naphthylmethylene)- β -*D*-galactopyranosyl-(1 \rightarrow 4)]-2-azido-3,6-di-*O*-benzyl-2-deoxy- β -*D*-glucopyranoside (**9**)



To a 100 mL round bottom flask containing Compound **S18** (1 g, 1.0 mmol) in DCM:H₂O 19:1 (v/v) was added 10 – 15% TFA. The solution was then stirred at 21 °C for 2 hours and was

directly poured into cold sat. NaHCO₃ and washed. The organic phase was dried over Na₂SO₄. Solvent was evaporated in vacuo. The residue was transferred into a 100 mL round bottom flask for next step without further purification.

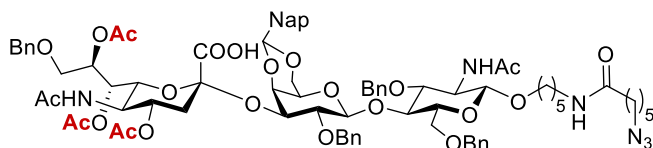
DDQ oxidation was performed following *general procedure D*. To the argon flushed 100 mL round bottom flask containing isopropylidene removed intermediate in dry DCM, 4 Å molecular sieves was added. The mixture was stirred at 21 °C for 1 hour, followed by the addition of DDQ (277 mg, 1.22 mmol). The brown suspension was stirred vigorously for 6 h at 21 °C. Upon completion, DCM was added to dilute the mixture, which was then washed sat. NaHCO₃ until the color of organic phase turned light yellow. The DCM solution was collected, dried over Na₂SO₄ and concentrated. The residue was purified through silica gel column (toluene: ethyl acetate 6:1). The naphthylmethylene protected compound **9** was obtained in 57% yield over 2 steps.

Later, it was found that performing TFA-mediated isopropylidene removal directly after sialylation reaction and the DDQ oxidation following the silica gel column gave a better yield (58% vs. 51%) from **5** to **9**. Moreover, the trisaccharide **8** with isopropylidene removed was easier to purify from the mixture.

¹H NMR (400 MHz, Chloroform-*d*) δ 7.94 (s, 1H, Ar), 7.84 – 7.67 (m, 2H, Ar), 7.60 (dt, *J* = 8.6, 1.4 Hz, 1H, Ar), 7.56 – 7.39 (m, 3H, Ar), 7.39 – 7.14 (m, 10H, Ar), 5.56 (d, *J* = 2.6 Hz, 2H, CHNap, NH), 5.19 (d, *J* = 10.5 Hz, 1H, CH₂Ar), 4.79 – 4.35 (m, 6H, CH₂Ar, H-1_{Gal}), 4.30 – 4.11 (m, 2H, H-6_{Gal}, H-1_{GlcN}), 4.06 – 3.82 (m, 6H, H-3_{Gal}, H-4_{GlcN}, OCH₂-linker, H-6_{Sia}, H-6_{GlcN}, H-4_{Sia}), 3.82 – 3.47 (m, 9H, COOCH₃, H-7_{Sia}, H-6_{bGlcN}, H-2_{Gal}, H-9_{a,bSia}, OCH₂-linker), 3.47 – 3.21 (m, 7H, H-8_{Sia}, H-2_{GlcN}, H-5_{GlcN}, H-5_{Sia}, H-3_{GlcN}, CH₂N₃), 3.15 (dd, *J* = 5.0, 1.1 Hz, 1H, OH), 3.06 – 2.93 (m, 2H, H-5_{Gal}, H-3_{eqSia}), 2.90 (d, *J* = 5.3 Hz, 1H, OH), 2.09 (t, *J* = 12.7 Hz, 1H, H-3_{axSia}), 1.77 – 1.37 (m, 6H, linker). ¹³C NMR (101 MHz, cdcl₃) δ = 168.89, 159.35, 138.38, 138.17, 138.05, 137.43, 135.22, 133.67, 132.83, 129.03, 128.89, 128.58, 128.43, 128.32, 128.25, 128.23, 128.22, 128.07, 128.01, 127.99, 127.81, 127.65, 127.63, 127.55, 127.54, 126.35, 126.07, 125.76, 125.29, 124.00, 102.74, 102.05, 101.41, 99.29, 81.10, 77.88, 77.53, 77.39, 77.33, 77.24, 77.22, 77.01, 76.70, 75.94, 75.70, 75.33, 75.25, 74.62, 73.60, 73.02, 72.41, 71.43, 70.54, 69.73, 69.63, 69.34, 68.85, 68.06, 66.18, 66.05, 63.03, 57.83, 53.39, 51.34, 47.18, 36.38, 31.70, 29.56, 29.04, 28.57,

25.43, 23.21, 22.62, 13.94, 11.78. HRMS-ESI (pos.) calculated for $C_{70}H_{79}N_7O_{18}NH_4$ 1283.5512, found 1283.5562.

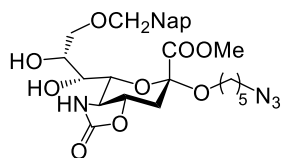
5-(6-azidohexanamido)pentyl (5-acetamido-4,7,8-tri-*O*-acetyl-9-*O*-benzyl-3,5-dideoxy-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid)-(2 \rightarrow 3)- [2-*O*-benzyl-4,6-*O*-(2-naphthylmethylene)- β -D-galactopyranosyl-(1 \rightarrow 4)]-2-acetamido-3,6-di-*O*-benzyl-2-deoxy- β -D-glucopyranoside (13**)**



The deprotective transformation was performed following the procedures described for **S18**, with modifications.

The titled compound **13** was obtained in 61% yield over 3 steps. 1H NMR (600 MHz, Methanol- d_4) δ 7.95 (s, 1H, Ar), 7.79 (d, J = 8.1 Hz, 1H, Ar), 7.73 (d, J = 8.6 Hz, 1H, Ar), 7.65 (d, J = 8.2 Hz, 1H, Ar), 7.61 (d, J = 8.5 Hz, 1H, Ar), 7.47 – 7.14 (m, 22H, Ar), 5.70 (s, 1H, CH_{Nap}), 5.58 (m, 1H, Sia-H8), 5.41 (dd, J = 9.5, 2.7 Hz, 1H, Sia-H7), 5.27 (d, J = 11.4 Hz, 1H, CH_2Ar), 5.00 – 4.89 (m, 2H, Sia-H4, CH_2Ar), 4.72 (m, 2H, Gal-H1, CH_2Ar), 4.62 (d, J = 11.4 Hz, 1H, CH_2Ar), 4.53 – 4.49 (m, 1H, Gal-H3), 4.45 – 4.24 (m, 8H, CH_2Ar , Gal-H4, Gal-H6a), 4.21 (d, J = 10.8 Hz, 1H, Sia-H6), 4.14 – 3.99 (m, 2H, Sia-H5, Gal-H6b), 3.95 (t, J = 9.1 Hz, 1H, GlcN-H4), 3.87 – 3.76 (m, 2H, GlcN-H6a, linker- OCH_2), 3.66 (m, 4H, GlcN-H2, Gal-H2, GlcN-H3, GlcN-H6b), 3.53 (dd, J = 12.4, 5.6 Hz, 1H, Sia-H9b), 3.44 (d, J = 9.2 Hz, 2H, GlcN-H5, linker- OCH_2), 3.37 (dd, J = 8.6, 4.1 Hz, 5H, Sia-H9b, Gal-H5), 3.25 (t, J = 7.0 Hz, 2H, linker- CH_2N_3), 3.12 (m, 2H, linker- $CH_2NHC(O)-$), 2.68 (dd, J = 13.1, 4.5 Hz, 1H, Sia-H3a), 2.16 (t, J = 7.3 Hz, 2H, linker- $C(O)CH_2-$), 2.11 (s, 3H, Ac), 1.96 (s, 3H, Ac), 1.89 – 1.78 (m, 10H, Ac, Sia-H3b), 1.68 – 1.33 (m, 12H, linker). ^{13}C NMR (151 MHz, CD_3OD) δ = 175.14, 172.83, 172.40, 171.04, 170.75, 170.45, 139.96, 139.73, 138.90, 138.39, 136.50, 134.30, 133.54, 129.00, 128.62, 128.54, 128.41, 128.18, 128.04, 127.88, 127.64, 126.52, 126.23, 126.04, 124.81, 103.10, 101.87, 101.62, 98.30, 81.73, 78.86, 77.80, 75.71, 75.56, 75.42, 74.72, 73.45, 72.57, 70.61, 69.68, 69.50, 69.36, 68.93, 68.68, 68.57, 66.88, 55.70, 51.55, 49.36, 49.11, 39.53, 38.09, 36.17, 29.43, 29.20, 28.87, 26.62, 25.80, 23.64, 22.35, 21.95, 20.87, 20.45, 20.03. HRMS-ESI (pos.) calculated for $C_{81}H_{98}N_6O_{23}Na$ 1545.6581, found 1545.6577.

Methyl (5-azidopentyl 5-amino-5-*N*,4-*O*-carbonyl-3,5-dideoxy-9-*O*-(2-naphthylmethyl)-D-glycero- α -D-galacto-2-nonulopyranosid)onate (7**)**

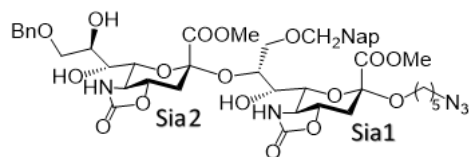


To a round bottom flask was added **S9** (1.5 g, 2.2 mmol) and 5-azidopentanol linker (341 mg, 2.6 mmol) in anhydrous DCM (44 mL) in presence of molecular sieves. The solution was stirred at 21 °C for 1 hour and was cooled to -78 °C. TMSOTf (437 μ L, 2.4 mmol) was added dropwise. After 5 minutes, the reaction was quenched with triethylamine, filtered and washed

Chapter 2

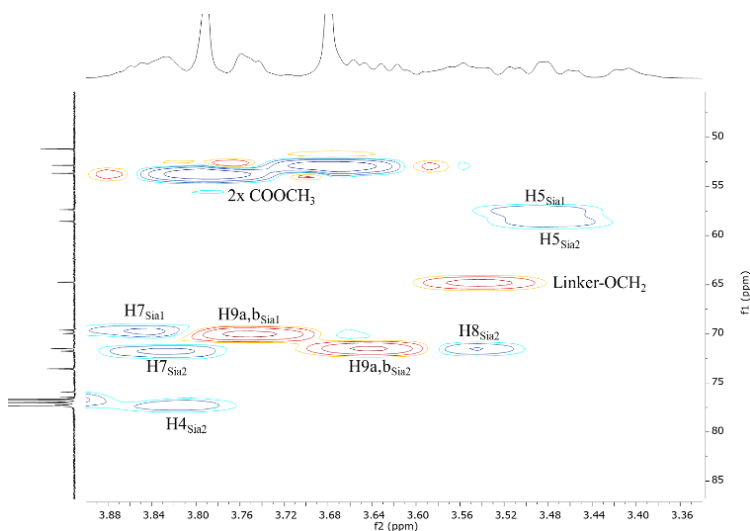
with saturated NaHCO_3 solution. The organic phase was dried over Na_2SO_4 and concentrated. The residue was taken up in DCM (40 mL), to which water (2 mL) and TFA (4 mL) was added at 0 °C. The reaction was then allowed to warm up to 21 °C and vigorously stirred for 2 hours. Upon completion, the biphasic mixture was slowly poured in saturated NaHCO_3 solution and washed 3 times. The organic layer was dried over Na_2SO_4 and concentrated in vacuo. The crude mixture was purified by silica gel (ethyl acetate/toluene 1:2 – 1:1) to give **5** (912 mg) in 74% yield over 2 steps. ^1H NMR (400 MHz, Chloroform- d) δ 7.91 – 7.69 (m, 3H, Nap), 7.56 – 7.38 (m, 2H, Nap), 5.82 (s, 1H, NH), 4.75 (s, 1H, CH_2Nap), 3.99 – 3.84 (m, 3H, H4, H7 and H6), 3.84 – 3.74 (m, 5H, COOCH_3 , H9a and b), 3.73 – 3.62 (m, 2H, H8, linker- OCH_2), 3.54 (dd, J = 11.2, 9.5 Hz, 1H, H5), 3.33 – 3.18 (m, 3H, linker- OCH_2 , linker- CH_2N_3), 3.01 – 2.80 (m, 2H, H3eq, OH), 2.09 (dd, J = 13.7, 11.9 Hz, 1H, H3ax), 1.63 – 1.47 (m, 4H, linker), 1.45 – 1.29 (m, 2H, linker). ^{13}C NMR (101 MHz, cdCl_3) δ = 168.88, 159.76, 135.01, 133.22, 133.06, 128.37, 127.87, 127.71, 126.73, 126.08, 125.66, 100.24, 77.50, 77.47, 77.32, 73.74, 72.39, 71.34, 69.84, 64.69, 57.68, 53.21, 51.26, 36.93, 28.92, 28.47, 23.15. HRMS-ESI (pos.) calculated for $\text{C}_{27}\text{H}_{34}\text{N}_4\text{O}_9\text{Na}$ 581.2223, found 581.2230.

Methyl (5-azidopentyl naphthylmethyl)-8-O-(methyl 5-amino-5-*N*,4-*O*-carbonyl-3,5-dideoxy-9-*O*-(2-naphthylmethyl)-8-*O*-(methyl 5-amino-9-*O*-benzyl-5-*N*,4-*O*-carbonyl-3,5-dideoxy-D-nonulopyranosylonate)-D-glycero- α -D-galacto-2-nonulopyranosid)onate (**14**)



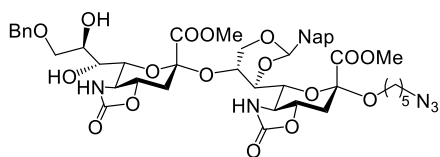
The sialyl donor **4** (1.3 g, 2.1 mmol) and acceptor **7** (640 mg, 1.1 mmol) were coupled under the same condition described for **S17**. The product **14** was obtained in 85% yield over 2 steps after silica gel purification (toluene/acetonitrile 4:1 – 3:1). ^1H NMR (400

MHz, Chloroform- d) δ 7.80 (d, J = 7.3 Hz, 3H, Ar), 7.73 (s, 1H, Ar), 7.53 – 7.37 (m, 3H, Ar), 7.37 – 7.21 (m, 5H, Ar), 6.12 (d, J = 13.9 Hz, 1H, NH), 5.88 (d, J = 21.2 Hz, 1H, NH), 4.77 – 4.59 (m, 2H, CH_2Ph), 4.53 (s, 2H, CH_2Nap), 4.32 (dt, J = 7.3, 3.8 Hz, 1H, H-8 $_{\text{Sia1}}$), 4.08 – 3.90 (m, 3H, H-4 $_{\text{Sia1}}$, H-6 $_{\text{Sia1}}$, H-6 $_{\text{Sia2}}$), 3.90 – 3.34 (m, 19H, see HSQC), 3.26 – 3.15 (m, 2H, linker- CH_2N_3), 3.11 (dt, J = 8.7, 6.3 Hz, 1H, linker- OCH_2), 2.98 (dd, J = 12.1, 3.6 Hz, 1H, H3eq $_{\text{Sia1}}$), 2.85 (dd, J = 12.0, 3.7 Hz, 1H, H3eq $_{\text{Sia2}}$), 2.22 (t, J = 12.6 Hz, 1H, H3ax $_{\text{Sia1}}$), 1.99 (t, J = 12.6 Hz, 1H,



H3ax_{Sia2}), 1.80 (m, 2H, OH), 1.59 – 1.17 (m, 6H, linker). ¹³C NMR (101 MHz, cdcl₃) δ = 168.55, 160.37, 159.88, 137.61, 134.90, 133.18, 133.04, 128.52, 128.37, 127.93, 127.85, 127.84, 127.73, 126.68, 126.36, 126.17, 125.63, 100.32, 100.25, 77.27, 77.21, 76.94, 76.48, 75.91, 73.59, 73.57, 71.77, 71.51, 69.98, 69.63, 64.75, 58.56, 57.37, 53.67, 52.86, 51.21, 37.27, 35.63, 28.91, 28.49, 23.17. HRMS-ESI (pos.) calculated for C₄₅H₅₅N₅O₁₇Na 960.3491, found 960.3495.

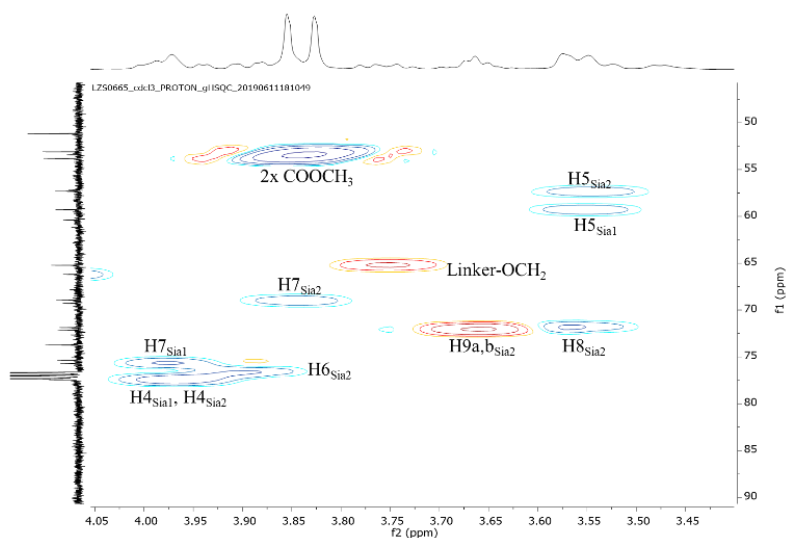
Methyl (5-azidopentyl 5-amino-5-*N*,4-*O*-carbonyl-3,5-dideoxy-7,9-*O*-(2-naphthylmethylene)-8-*O*-(methyl 5-amino-9-*O*-benzyl-5-*N*,4-*O*-carbonyl-3,5-dideoxy-nonulopyranosylonate)-D-glycero- α -D-galacto-2-nonulopyranosid)onate (15)



Compound **15** was prepared following the *general procedure D* starting from **14** (300 mg, 0.32 mmol). Isolated yield, 72%. ¹H NMR (400 MHz, Chloroform-*d*) δ 7.95 (s, 1H, Ar), 7.83 (m, 3H, Ar), 7.57 (dt, *J* = 8.5, 1.6 Hz, 1H, Ar), 7.52 – 7.43 (m, 2H, Ar), 7.39 – 7.20 (m, 5H, Ar), 6.19

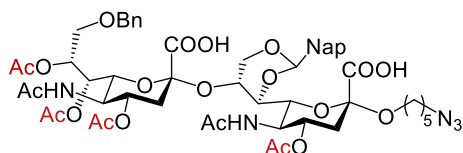
(m, 2H, NH, CHNap), 5.75 (s, 1H, NH), 4.56 (s, 2H, CH₂Ph), 4.36 (d, *J* = 11.9 Hz, 1H, H9a_{Sia1}), 4.31 – 4.19 (m, 2H, H9b_{Sia1}, H8_{Sia1}), 4.09 (s, 1H, H6_{Sia1}), 4.04 – 3.38 (m, 20H, see HSQC), 3.38 – 3.21 (m, 1H, linker-OCH₂), 3.19 – 3.04 (m, 3H, linker-CH₂N₃, H3eq_{Sia2}), 2.99 (dd, *J* = 12.0, 3.7 Hz, 1H, H3eq_{Sia1}), 2.15 (m, 2H, H3ax_{Sia1}, H3ax_{Sia2}), 1.62 – 1.15 (m, 6H, linker). ¹³C NMR (101 MHz, cdcl₃) δ = 168.98, 168.62, 160.54, 159.87, 137.51, 134.96, 133.58, 132.90, 128.56, 128.47, 128.32, 128.00, 127.89, 127.85, 127.72, 126.56, 126.35, 125.38, 123.42, 100.50, 100.28, 96.73, 77.44, 77.40, 77.21, 75.70, 75.35, 73.73, 72.13, 71.87, 69.28, 68.95, 66.19, 65.21, 60.39, 59.30, 57.33, 53.81, 53.10, 51.16, 37.54, 36.66, 28.98, 28.41, 23.14, 21.04, 18.56, 18.53, 14.18, 13.49.

HRMS-ESI
(pos.)
calculated for
C₄₅H₅₃N₅O₁₇H
936.3515,
found
936.3527.



Chapter 2

5-azidopentyl 5-acetamido-4-*O*-acetyl-3,5-dideoxy-7,9-*O*-(2-naphthylmethylene)-8-*O*-(5-acetamido-4,7,8-tri-*O*-acetyl-9-*O*-benzyl-3,5-dideoxy-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid)-D-glycero- α -D-galacto-2-nonulopyranosidonic acid (**16**)



Compound **15** was dissolved in dioxane/water (1:1 v/v), to which KOH (16 eq. to **15**) was added. The solution was heated at 50 °C for 2 hours. The reaction was cooled to 21 °C and concentrated in vacuo. The residue was suspended in DCM and cooled to 0 °C. Acetic

anhydride was added in portions (4 eq. to **15** each time) until more than two Ac additions was observed (monitored by ESI-MS). Presumably the amine groups undergo acetylation first and then the hydroxyls. Of note, this *N*-acetylation step is necessary. If the residue from saponification was directly subjected to Ac₂O/pyridine condition, re-introduction of *N,O*-carbonyl groups was observed by ESI-MS. The source of carbonyl may come from the in-situ formed anhydride species in Ac₂O/pyridine. This species may not form in DCM due to poor solubility of K₂CO₃ (resulted from 4*O*,5*N*-carbonyl deprotection). After *N*-acetylation in DCM, the mixture was concentrated and subjected to Ac₂O/pyridine mediated *O*-acetylation at 50 °C. Upon completion, water was added at 0 °C and the mixture was allowed to warm to 21 °C and stirred for 30 minutes. The solution was then concentrated in vacuo and the residue was purified by silica gel column (ethyl acetate/acetonitrile/water = 2:8:1) to give compound **16** in 57% yield over 3 steps. ¹H NMR spectra of **16** was with broadened peaks (methanol-d₄, CDCl₃ and DMSO-d₆), which precluded signal assignment and was hence directly subjected to next steps.

5-(6-(D-biotinamido)hexanamido)pentyl 5-acetamido-4,7,9-tri-*O*-acetyl-3,5-dideoxy-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid)-(2→3)- β -D-galactopyranosyl-(1→4)-2-acetamido-2-deoxy- β -D-glucopyranoside (1a**)** The titled compound was prepared from **13** following the *General procedure C* using NHS-biotin as the biotinylation reagent. See Table S2 for NMR analysis. HRMS (neg.) calculated for C₅₂H₈₄N₆O₂₅S-H 1223.5129, found 1223.5126.

5-(6-(D-biotinamido)hexanamido)pentyl 5-acetamido-4,9-di-*O*-acetyl-3,5-dideoxy-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid)-(2→3)- β -D-galactopyranosyl-(1→4)-2-acetamido-2-deoxy- β -D-glucopyranoside (1c**)** The titled compound was prepared from **1a** following the *General procedure A and B* using BCoV-HE as the de-*O*-acetylating enzyme. See Table S2 for NMR analysis. HRMS (neg.) calculated for C₅₀H₈₂N₆O₂₄S-H 1181.5023, found 1181.5034.

5-(6-(D-biotinamido)hexanamido)pentyl 5-acetamido-7,9-di-*O*-acetyl-3,5-dideoxy-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid)-(2→3)- β -D-galactopyranosyl-(1→4)-2-acetamido-2-deoxy- β -D-glucopyranoside (1d**)** The titled compound was prepared from **1a** following the *General procedure A* using MHV-S-HE as the de-*O*-acetylating enzyme. See Table S2 and Fig. S6a for NMR analysis. HRMS (neg.) calculated for C₅₀H₈₂N₆O₂₄S-H 1181.5023, found 1181.5029.

5-(6-(D-biotinamido)hexanamido)pentyl (5-acetamido-4-*O*-acetyl-3,5-dideoxy-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid)-(2 \rightarrow 3)- β -D-galactopyranosyl-(1 \rightarrow 4)-2-acetamido-2-deoxy- β -D-glucopyranoside (1e) The titled compound was prepared from **1c** following the *General procedure A* using BCoV-HE as the de-*O*-acetylating enzyme. See Table S2 for NMR analysis. HRMS (neg.) calculated for C₄₈H₈₀N₆O₂₃S-H 1139.4917, found 1139.4920.

5-(6-(D-biotinamido)hexanamido)pentyl (5-acetamido-7-*O*-acetyl-3,5-dideoxy-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid)-(2 \rightarrow 3)- β -D-galactopyranosyl-(1 \rightarrow 4)-2-acetamido-2-deoxy- β -D-glucopyranoside (1f) The titled compound was prepared from **1a** following the *General procedure A* using MHV-S- and BCoV-HEs as the de-*O*-acetylating enzymes. See Table S2 and Fig. S6b for NMR analysis. HRMS (neg.) calculated for C₄₈H₈₀N₆O₂₃S-H 1139.4917, found 1139.4921.

5-(6-(D-biotinamido)hexanamido)pentyl (5-acetamido-9-*O*-acetyl-3,5-dideoxy-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid)-(2 \rightarrow 3)- β -D-galactopyranosyl-(1 \rightarrow 4)-2-acetamido-2-deoxy- β -D-glucopyranoside (1g) The titled compound was prepared from **1f** following the *General procedure B*. See Table S2 for NMR analysis. HRMS (neg.) calculated for C₄₈H₈₀N₆O₂₃S-H 1139.4917, found 1139.4913.

5-(6-(D-biotinamido)hexanamido)pentyl (5-acetamido-4,7,9-tri-*O*-acetyl-3,5-dideoxy-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid)-(2 \rightarrow 6)- β -D-galactopyranosyl-(1 \rightarrow 4)-2-acetamido-2-deoxy- β -D-glucopyranoside (2a) The titled compound was prepared from **S15** following the *General procedure C* using NHS-biotin as the biotinylation reagent. See Table S2 for NMR analysis. HRMS (neg.) calculated for C₅₂H₈₄N₆O₂₅S-H 1223.5129, found 1223.5124.

5-(6-(D-biotinamido)hexanamido)pentyl (5-acetamido-4,9-di-*O*-acetyl-3,5-dideoxy-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid)-(2 \rightarrow 6)- β -D-galactopyranosyl-(1 \rightarrow 4)-2-acetamido-2-deoxy- β -D-glucopyranoside (2c) The titled compound was prepared from **2a** following the *General procedure A and B* using BCoV-HE as the de-*O*-acetylating enzyme. See Table S2 for NMR analysis. HRMS (neg.) calculated for C₅₀H₈₂N₆O₂₄S-H 1181.5023, found 1181.5030.

5-(6-(D-biotinamido)hexanamido)pentyl (5-acetamido-7,9-di-*O*-acetyl-3,5-dideoxy-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid)-(2 \rightarrow 6)- β -D-galactopyranosyl-(1 \rightarrow 4)-2-acetamido-2-deoxy- β -D-glucopyranoside (2d) The titled compound was prepared from **2a** following the *General procedure A* using MHV-S-HE as the de-*O*-

Chapter 2

acetylating enzyme. See Table S2 and Fig.S7a for NMR analysis. HRMS (neg.) calculated for $C_{50}H_{82}N_6O_{24}S-H$ 1181.5023, found 1181.5031.

5-(6-(D-biotinamido)hexanamido)pentyl (5-acetamido-4-O-acetyl-3,5-dideoxy-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid)-(2 \rightarrow 6)- β -D-galactopyranosyl-(1 \rightarrow 4)-2-acetamido-2-deoxy- β -D-glucopyranoside (2e) The titled compound was prepared from **2c** following the *General procedure A* using BCoV-HE as the de-O-acetylating enzyme. See Table S2 for NMR analysis. HRMS (neg.) calculated for $C_{48}H_{80}N_6O_{23}S-H$ 1139.4917, found 1139.4914.

5-(6-(D-biotinamido)hexanamido)pentyl (5-acetamido-7-O-acetyl-3,5-dideoxy-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid)-(2 \rightarrow 6)- β -D-galactopyranosyl-(1 \rightarrow 4)-2-acetamido-2-deoxy- β -D-glucopyranoside (2f) The titled compound was prepared from **2a** following the *General procedure A* using MHV-S- and BCoV-HEs as the de-O-acetylating enzymes. See Table S2 and Fig.7b for NMR analysis. HRMS (neg.) calculated for $C_{48}H_{80}N_6O_{23}S-H$ 1139.4917, found 1139.4920.

5-(6-(D-biotinamido)hexanamido)pentyl (5-acetamido-9-O-acetyl-3,5-dideoxy-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid)-(2 \rightarrow 6)- β -D-galactopyranosyl-(1 \rightarrow 4)-2-acetamido-2-deoxy- β -D-glucopyranoside (2g) The titled compound was prepared from **2f** following the *General procedure B*. See Table S2 for NMR analysis. HRMS (neg.) calculated for $C_{48}H_{80}N_6O_{23}S-H$ 1139.4917, found 1139.4916.

5-(6-(D-biotinamido)hexanamido)pentyl 5-acetamido-4-O-acetyl-3,5-dideoxy-8-O-(5-acetamido-4,7,9-tri-O-acetyl-3,5-dideoxy-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid)-D-glycero- α -D-galacto-2-nonulopyranosidonic acid (3a) The titled compound was prepared from **16** following the *General procedure C* using NHS-Lc-biotin as the biotinylation reagent. See Table S2 for NMR analysis. HRMS (neg.) calculated for $C_{51}H_{80}N_6O_{24}S-H$ 1191.4866, found 1191.4859.

5-(6-(D-biotinamido)hexanamido)pentyl 5-acetamido-3,5-dideoxy-8-O-(5-acetamido-7,9-di-O-acetyl-3,5-dideoxy-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid)-D-glycero- α -D-galacto-2-nonulopyranosidonic acid (3d) The titled compound was prepared from **3a** following the *General procedure A* using MHV-S-HE as the de-O-acetylating enzyme. See Table S2 and Fig.S8a for NMR analysis. HRMS (neg.) calculated for $C_{47}H_{76}N_6O_{22}S-H$ 1107.4655, found 1107.4649.

5-(6-(D-biotinamido)hexanamido)pentyl 5-acetamido-3,5-dideoxy-8-O-(5-acetamido-7-O-acetyl-3,5-dideoxy-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid)-D-glycero- α -D-galacto-2-nonulopyranosidonic acid (3f) The titled compound was prepared

from **3a** following the *General procedure A* using MHV-S- and BCoV-HEs as the de-*O*-acetylating enzymes. See Table S2 and Fig.S8b for NMR analysis. HRMS (neg.) calculated for C₄₅H₇₄N₆O₂₁S-H 1065.4549, found 1065.4543.

5-(6-(D-biotinamido)hexanamido)pentyl 5-acetamido-3,5-dideoxy-8-*O*-(5-acetamido-9-*O*-acetyl-3,5-dideoxy-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid)-D-glycero- α -D-galacto-2-nonulopyranosidonic acid (3g) The titled compound was prepared from **3f** following the *General procedure B* See Table S2 for NMR analysis. HRMS (neg.) calculated for C₄₅H₇₄N₆O₂₁S-H 1065.4549, found 1065.4546.

Chapter 2

3. Additional references

1. Schultze, B., Gross, H. J., Brossmer, R. & Herrler, G. The S protein of bovine coronavirus is a hemagglutinin recognizing 9-*O*-acetylated sialic acid as a receptor determinant *J Virol* **65**, 6232-6237 (1991).
2. Langereis, M. A. et al. Structural basis for ligand and substrate recognition by torovirus hemagglutinin esterases *Proc Natl Acad Sci* **106**, 15897-15902 (2009).
3. Rogers, G. N., Herrler, G., Paulson, J. C. & Klenk, H.-D. Influenza C virus uses 9-*O*-acetyl-*N*-acetylneuraminic acid as a high affinity receptor determinant for attachment to cells *J Biol Chem* **261**, 5947-5951 (1986).
4. Song, H. et al. An open receptor-binding cavity of hemagglutinin-esterase-fusion glycoprotein from newly-identified influenza D virus: Basis for its broad cell tropism *Plos Pathog* **12**, e1005411 (2016).
5. Liang, P.-H., Wang, S.-K. & Wong, C.-H. Quantitative analysis of carbohydrate-protein interactions using glycan microarrays: Determination of surface and solution dissociation constants. *J Am Chem Soc* **129**, 11177-11184, (2007).
6. Hamagami, H. et al. 6-azido-6-deoxy-*L*-idose as a hetero-bifunctional spacer for the synthesis of azido-containing chemical probes. *Chem Eur J* **22**, 12884-12890 (2016).
7. Morley, T. J. & Withers, S. G. Chemoenzymatic synthesis and enzymatic analysis of 8-modified cytidine monophosphate-sialic acid and sialyl lactose derivatives. *J Am Chem Soc* **132**, 9430-9437 (2010).
8. Xia, J., Alderfer, J. L., Locke, R. D., Piskorz, C. F. & Matta, K. L. , Complex oligosaccharide investigations: Synthesis of an octasaccharide incorporating the dimeric Le(X) structure of PSGL-1. *J Org Chem* **68**, 2752-2759 (2003).
9. Van Der Vorm, S., Overkleeft, H. S., Van Der Marel, G. A. & Codee, J. D. C. Stereoselectivity of conformationally restricted glucosazide donors *J Org Chem* **82**, 4793-4811 (2017).
10. Orgueira, H. A. et al. Modular synthesis of heparin oligosaccharides. *Chem Eur J* **9**, 140-169 (2003).
11. Vong, K., Tam, I. S., Yan, X. & Auclair, K. Inhibitors of aminoglycoside resistance activated in cells. *ACS Chem Biol* **7**, 470-475 (2012).
12. Liao, W., Locke, R. D. & Matta, K. L. Selectin ligands: 2,3,4-tri-*O*-acetyl-6-*O*-(2-naphthyl)methyl (Nap) α -*D*-galactopyranosyl imidate as a novel glycosyl donor for the efficient total synthesis of branched mucin core 2-structure containing the NeuAca2,3(SO₃Na-6)Gal β 1,3GalNAca sequence. *Chem Comm* 369-370 (2000).

CHAPTER 3 | Synthetic *O*-Acetyl-*N*-glycolylneuraminic Acid Oligosaccharides Reveal Host-Associated Binding Patterns of Coronaviral Glycoproteins

Zeshi Li¹, Lin Liu³, Luca Unione¹, Yifei Lang², Raoul J. de Groot², Geert-Jan Boons^{1,3,4,5}

Affiliations

¹Department of Chemical Biology and Drug Discovery, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, 3584 CG Utrecht, The Netherlands.

²Virology Division, Department of Biomolecular Health Sciences, Faculty of Veterinary Medicine, Utrecht University, 3584 CL Utrecht, the Netherlands.

³Complex Carbohydrate Research Center, University of Georgia, 315 Riverbend Road, Athens, GA 30602, USA.

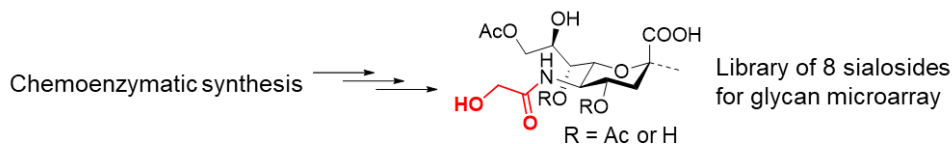
⁴Bijvoet Center for Biomolecular Research, Utrecht University, Utrecht, The Netherlands.



⁵Chemistry Department, University of Georgia, Athens, GA 30602, USA.

Chapter 3

Abstract

A panel of *O*-acetylated *N*-glycolylneuraminic acid oligosaccharides was prepared by diversification of common synthetic precursors by regioselective de-*O*-acetylation by coronaviral hemagglutinin-esterase (HE) combined with C-7 to C-9 acetyl ester migration. The resulting compound library was printed on streptavidin-coated glass slides to give a glycan microarray to investigate receptor binding specificities of viral envelope glycoproteins, including spike proteins and HEs from animal and human coronaviruses. It was found that the binding patterns of these viral proteins of *N*-glycolylated sialosides differ considerably from previously synthesized *N*-acetylated counterparts. Generally, the spike proteins tolerate *N*-glycolyl modification but selectivities differ among viruses targeting different hosts. On the other hand, the lectin domain of the corresponding HEs showed a substantial decrease or loss of binding of *N*-glycolylated sialosides. MD simulations indicate that glycolyl recognition by HE is mediated by polar residues in a loop region (109-119) that interacts with the 5-*N*-glycolyl moiety. Collectively, the results of this study indicate that coronaviruses have adjusted their receptor fine specificities to adapt to the sialoglycome of the corresponding host species.



Coronaviral glycoprotein	Spike					Hemagglutinin-esterase			
Host species	 		 				  		
Neu5Gc binding	+++	++	+			+++	+		

Evolutionary pressure exerted by pathogenic microbes have diverged glycan biosynthesis and as a result, glycomes of even closely related species may differ considerably.^{1, 2} This is well-documented by sialic acids; a diverse family of negatively charged, nine-carbon keto-sugar distributed across all domains of life. In vertebrates, *N*-acetylneuraminic acid (Neu5Ac) and *N*-glycolylneuraminic acid (Neu5Gc) are the most common forms of sialic acids.³ The latter modification is biosynthesized by hydroxylation of the *N*-acetyl moiety at C5 cytidine-5'-monophosphoryl Neu5Ac (CMP-Neu5Ac) by the enzyme CMP-*N*-acetylneuraminic acid hydroxylase (CMAH).⁴ Various sialyl transferases can utilize the resulting CMP-Neu5Gc to biosynthesize Neu5Gc containing glycoconjugates. In humans, the gene encoding CMAH has been inactivated and as a result cannot biosynthesize CMP-Neu5Gc. Although Neu5Gc is considered a non-human sialic acid, it can be incorporated into its sialoglycans by utilizing exogenous Neu5Gc from dietary sources.⁵

Neu5Ac and Neu5Gc commonly occur at termini of complex glycans and can be attached to a penultimate sugar in several glycosidic forms: α 2,3-linked to galactose (Gal), α 2,6-linked to Gal, *N*-acetylgalactosamine (GalNAc) or *N*-acetylglucosamine (GlcNAc), or α 2,8-linked to another sialic acid. Neu5Ac and Neu5Gc can also undergo hydroxyl modifications including mono-, di-, tri- and even tetra-*O*-acetylation at C4, C7, C8 and/or C9 position(s) giving rise to over ten molecular variants (Fig. 1). The biological roles of specific variants have been difficult to dissect partly due to their lability and a lack of well-defined compounds.

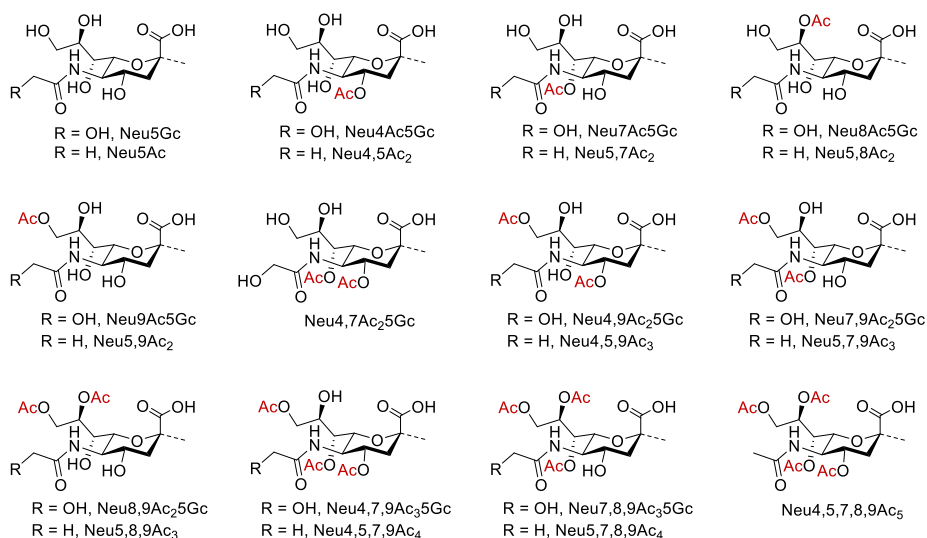


Figure 1 | Structures of *N*-acetyl and *N*-glycolyl neuraminic acid and their naturally occurring *O*-acetylated variants. Abbreviations: Ac, acetyl; Gc, glycolyl. Up till now, only Neu9Ac5Gc and Neu4Ac5Gc have been incorporated in oligosaccharides.

O-acetylated variants of sialic acid serve as receptors for many viruses including embecoviruses (family Coronaviridae), toroviruses (Tobaniviridae), influenza C and D viruses and infectious salmon anemia virus (Orthomyxoviridae).⁶ Receptor binding preferences critically determine viral host range and tissue/organ tropism. Insight in receptor

preferences of these viruses will provide important knowledge of the factors driving and hampering cross-species transmission, thereby facilitating prediction and prevention of emerging zoonotic outbreaks.⁷ Recently, we demonstrated that the spike of OC43 and HKU1, which both are human corona viruses of zoonotic origin that can cause upper respiratory tract infections, have convergently evolved to selectively engage with 9-*O*-acetylated α 2,8-linked disialosides for cell entry.⁸ Furthermore, it was found that the spike of related animal viruses exhibits distinctive selectivities for various *O*-acetylation forms and glycosidic linkages. It remained elusive whether 5-*N*-glycolylation can modulate virus-sialic acid interactions. Given that Neu5Gc repertoires differ considerably among mammalian species,⁹ it is conceivable that viruses targeting specific hosts may exhibit differential preferences for acetylated sialosides having a 5-*N*-glycolyl moiety.¹⁰

Herein, we describe the chemoenzymatic synthesis of a library of *N*-glycolyl-*O*-acetylneuraminic acid oligosaccharides by hemagglutinin-esterase (HE) mediated regioselective de-*O*-acetylation in combination with controlled base-mediated C7 to C-9 acetyl ester migration of common synthetic 4,7,9-tri-*O*-acetylated sialosides. It gave fourteen *O*-acetylated sialosides from two common precursors in high overall yields. The resulting compounds and previously synthesized Neu5Ac counterparts⁸ were printed as a glycan microarray that was probed for binding of recombinant HEs and spike proteins from various beta-coronaviruses that target different host species. It was found that the lectin domain of HE of all examined viruses showed a reduced- or no affinity for the glycolyl-modified structures. The spike of the corresponding viruses did bind Neu5Gc modified-glycans, but the binding patterns differed from those observed of Neu5Ac containing glycans.

Previously, we described a strategy for the preparation of a panel of *O*-acetylated Neu5Ac-containing oligosaccharides starting from α 2,3- and 2,6-linked common precursors **1a** and **2a** (Fig. 1a, respectively, having acetyl esters at C-4, 5, 7 and 9. These compound could be diversified by treatments with HEs from bovine and murine coronaviruses (BCoV and MHV-S, respectively), which cleave 9- and 4-*O*-acetyl esters with exquisite regioselectivity, respectively, in combination with controlled base-induced acetyl ester migration from C7 to C9 (Fig. 2a). We envisaged that if HEs can also cleave glycolyl-modified *O*-acetylate sialosides, the methodology can be extended to the preparation of an almost complete panel of Neu5Gc sialoglycans with *O*-acetylation patterns mirroring that of previously prepared Neu5Ac counterparts. The resulting glycan library would cover most of the commonly found *O*-acetylated sialic acids in humans and other mammals allowing in-depth characterization of virus-receptor interactions. To implement the chemoenzymatic strategy for preparing *O*-acetylated Neu5Gc derivatives, we chemically synthesized common precursors **3a** and **4a** by adapting the late-stage *N*- and *O*-acylation steps employed for the synthesized sialosides **1a** and **2a**, (Fig. 2b), which were then subjected to HE treatments and controlled 7-to-9 *O*-acetyl migration.

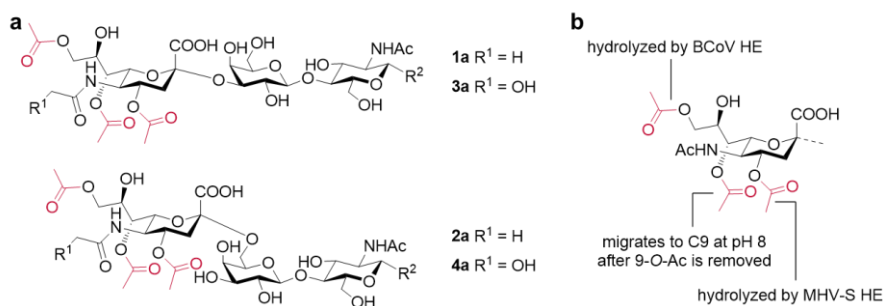


Figure 2 | General synthetic strategy for Neu5Gc oligosaccharides. a, Remodeling of a Neu5Ac common precursor with enzymes and a mildly basic condition. **b,** α 2,3- (top) and α 2,6- (bottom) linked Neu5Gc glycoside common precursors for diversification.

The preparation of *O*-acetylated sialo-glycans comes with number of challenges. These included stereo-control of α -selective chemical sialylation and the unstable nature of acetyl esters on sialic acid. In particular for *N*-glycosylated sialosides, an additional challenge lies in C5 amine derivatization, which requires orthogonal protecting groups for the secondary C2-amine of glucosamine and the C5-amine of neuraminic acid. Building blocks **5**, **7** and **9** (Fig. 3a) were employed for the chemical synthesis of trisaccharides **10** as previously reported,⁸ which through a number of manipulations was expected to provide common precursor **3a**. The 5-*N*,4-*O*-carbonyl of sialyl donor **5** ensured high reactivity and facilitated α -glycosidic bond formation.¹¹ The galactosyl donors **7** could be coupled with acceptor **9** to give a β -linked disaccharide that after deacetylation provided a highly reactivity glycosyl acceptor for regioselective glycosylations with **5**. After sialylation, the naphthyl-2-methyl (Nap) protecting group allowed protection of the C-4 neighboring alcohol *via* oxidative ring closure forming an acetal.¹² The 2-amino moiety of glucosamine **9** is

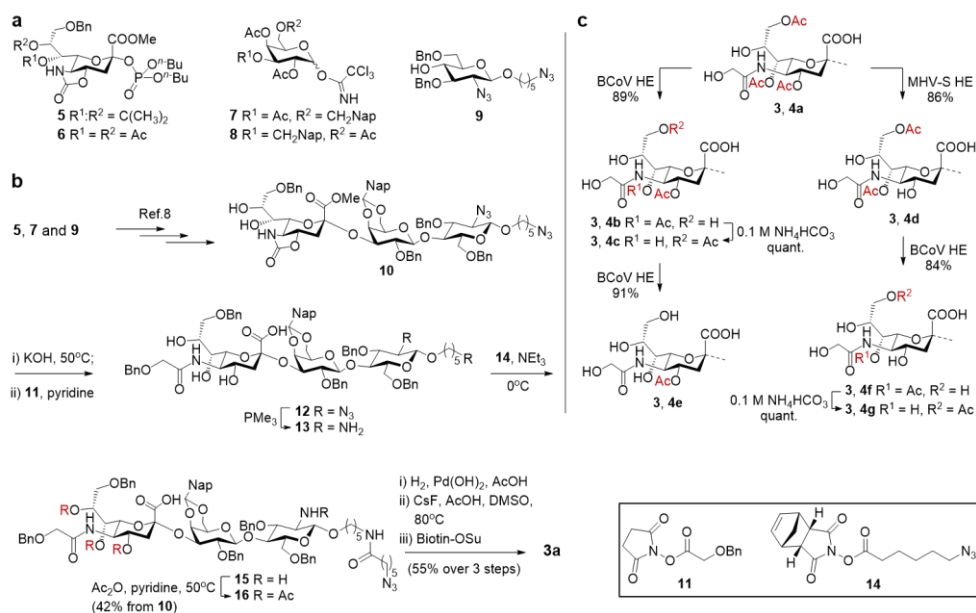


Figure 3 | Chemoenzymatic synthesis of Neu5Gc glycosides. a, Glycosyl donors and acceptor for chemical assembly of trisaccharide common precursors. Building blocks **5** and **7** were used for the synthesis of α 2,3-linked sialoside **3a**, and **6** and **8** for α 2,6-linked trisaccharide **4a**. **b,** Chemical assembly of α 2,3-linked Neu5Gc glycoside common precursor **3a** (4,7,9-tri-*O*-Ac). See Ref. 8 for synthesis of **10**. See supplementary Information for synthesis of α 2,6-linked common precursor **4a**. **c,** HE-catalyzed regioselective de-*O*-acetylation of **3a** and **4a**. Complete enzymatic conversion was observed for all steps. Yields were determined after HILIC purification, hence were not quantitative.

Chapter 3

masked as an azide which allows selective deprotection of the 5-*N*,4-*O*-carbonyl of sialic acid for selective installation of a glycolyl moiety.

Treatment of **10** with aqueous KOH resulted in the removal of base-sensitive protecting groups including the 4,5-carbamate thereby exposing the C5-amine of sialic acid (Fig. 3b). After desalting by LH-20 size exclusion column chromatography, the free amine of the resulting compound was acylated with active ester **11** to install a properly protected glycolyl moiety resulting in the formation of compound **12**. Reduction of the azide of **12** using Staudinger conditions (\rightarrow **13**), followed by regioselective acylation of the amine of the anomeric spacer using azide-modified activated ester **14** (**13** \rightarrow **15**), and finally acetylation of the remaining amine and hydroxyls using acetic anhydride in pyridine provided trisaccharide **16** in an overall yield of 42% (five steps). Hydrogenation of **16** over Pd(OH)₂ led to the removal of the benzyl ethers and reduction of the azide of the anomeric linker to an amine. The resulting compound was subjected to controlled C8 to C9 acetyl ester migration using cesium fluoride and acetic acid in DMSO as an aprotic, buffered system.¹³ This was followed by selective biotinylation of the amine of the anomeric linker using biotin succinimide to give, after purification by P-2 biogel size exclusion column chromatography and HPLC using a HILIC column, α 2,3-linked 5-*N*-glycolylated common precursor **3a**. In a similar fashion, α 2,6-linked Neu5Gc glycoside **4a** was assembled from the building blocks **6**, **8** and **9** (Supplementary Fig. S1).

B_{CoV} and MHV-S HEs can hydrolyze with exquisite regiospecificity 9- and 4-*O*-acetyl esters of glycosides of Neu5Ac, respectively.^{14, 15} However, the activities of these two enzymes for *O*-acetylated Neu5Gc have not been systematically investigated. To test if *O*-acetylated Neu5Gc is a substrate for the two enzymes, the 4,7,9-tri-*O*-acetylated common precursors **3a** and **4a** were incubated with B_{CoV} and MHV-S HEs (Fig. 3c). The former HE readily cleaved the 9-*O*-acetyl ester to give **3b** and **4b**, whereas the latter enzyme specifically removed 4-*O*-Ac to provide **3d** and **4d**. The enzymatic transformations were monitored by electrospray ionization mass spectrometry (ESI-MS) where loss of only one acetyl moiety (42 units) indicated regio-specific ester hydrolysis. The products were purified by high-performance hydrophilic interaction liquid chromatography (HILIC), and the structures confirmed by 1D and 2D NMR experiments. In particular, shift of sialic acid protons at C9 (for **3b** and **4b**) and C4 (for **3d** and **4d**) to higher field confirmed deacetylation of the corresponding hydroxyls. 4,7-di-*O*-acetylated sialosides **3b** and **4b** were treated with mild base (100 mM ammonium bicarbonate in D₂O) resulting in the migration of the C7 acetyl ester to C-8 thereby providing 4,9-diacetylated **3c** and **4c**. ¹H NMR with water peak suppression, showed the disappearance of C7-proton above 5.0 ppm confirming migration had occurred. Upon further treatment with B_{CoV} HE, **3c** and **4c** were converted into **3e** and **4e**, respectively. The 7-mono-*O*-acetylated **3f** and **4f** were obtained by cleaving the 9-*O*-acetyl of **3d** and **4d** using B_{CoV} HE. In 100 mM ammonium bicarbonate, the C7-acetyl ester of **3f** and **4f** readily migrated to C9 to afford **3g** and **4g**.

Having prepared a panel of *O*-acetylate Neu5Gc glycosides, attention was focused on examining receptor specificities of *Embecovirus*-derived envelope glycoproteins, spike and HE. These viruses bind to *O*-acetylated sialosides *via* their spike (S) for virion uptake.¹⁶ Destruction of decoy receptors and virion release is mediated by HE that is comprised of a lectin domain that recognizes *O*-acetylated sialosides and esterase domain that hydrolyses acetyl esters of sialosides thereby destroying decoy receptors.¹⁷ The lectin domain has as an important function to enhance the catalytic efficiency of HE for multivalently presented substrates.

Thus, the *O*-acetylated Neu5Gc glycosides (**3a-g** and **4a-g**) were printed on streptavidin-coated slides. The Neu5Ac derivatives having the same *O*-acetylation pattern were incorporated in the array for comparison. Receptor binding domains (S1^A) of the spike of human coronavirus OC43, bovine coronavirus (BCoV), rabbit coronavirus (RbCoV), equine coronavirus (ECoV) and canine respiratory coronavirus (CRCoV) were expressed as Fc-fusion proteins in HEK293T cells.⁸ In addition, enzymatically inactive forms of the ectodomains of corresponding HEs were also expressed as Fc-fusion proteins. The viral proteins were exposed to the array and detection of binding was accomplished using an anti-Fc antibody modified with AlexaFluor-647.

HEs of all examined viruses showed a decrease or loss of binding of the glycolyl-modified sialosides. In the case of BCoV, the additional hydroxyl group at the C5 substituent resulted in a moderate reduction in binding. On the other hand, the HEs of CRCoV and RbCoV showed a dramatic reduction in binding for all 5-*N*-glycolylated structures and signals for the 7,9-di- and 4,7,9-triacetylated forms were detected slightly above baseline. In the case of ECoV HE, no detectable binding of glycolyl modified structures was observed. HE of human coronaviruses have lost the ability to bind sialosides¹⁸ and therefore were not examined.

Unlike the HEs, the spike-derived proteins displayed variable responses to 5-*N*-glycolylation. In the case of RbCoV and BCoV spikes, 5-*N*-glycolylation resulted in an increase in binding of 4,9-di-*O*- and 4,7,9-tri-*O*-acetylated Neu5Gc sialosides. Conversely, the spikes of human and canine viruses, OC43 and CRCoV, respectively, gave considerably lower responses for 5-*N*-glycolylated structures. The 4,9-di-*O*-acetylated form was an exception, which displayed a striking increase in binding compared to the corresponding Neu5Ac glycosides. Interestingly, the effect of 5-*N*-glycolylation on ECoV spike was glycosidic linkage-selective. In the case of the α 2,3-linked sialosides, 5-*N*-glycolylation caused a substantial reduction of binding, whereas the same modification was better tolerated for the corresponding α 2,6-linked glycosides. The binding-decreasing effect of 5-*N*-glycolylation was particularly striking for 9-monoacetylated sialosides.

Recently, we provided an atomic-level interpretation of the recognition of Neu5Ac by the HEs of several host species by using a combination of NMR experiments and MD simulation.¹⁹ It demonstrated the importance of complementarity of the acetyl moieties of the sialosides and hydrophobic pockets of HE for binding. Thus, it is not surprising that the polar hydroxyl of glycolyl may unfavorably interfere in these hydrophobic interactions. The microarray data indicate that the lectin domain of the HEs of ECoV and RbCoV poorly recognize glycolyl modified sialosides whereas this modification is tolerated by the HE from BCoV. To offer a structural explanation of these observations, MD simulations (for details see SI section MD simulation) were performed on the three HEs in complex with the 4,7,9-tri-acetylated form of the 5-*N*-glycolylated sialoside **3a**.

The MD results indicate that residues in the loop region (amino acid residues 109-119) of the sialic acid binding domain are responsible for the observed differences in Neu5Ac and Neu5Gc binding. In particular, residues 114-116 face the glycolyl amide of the ligand, and in the case of BCoV, these residues are polar entailing a TTS sequon. The MD simulations showed that Thr114 and Thr115 can establish polar interactions with the amide at C5 of the sialoside by either directly engaging with the carbonyl oxygen or through water molecules. In turn, Ser116 makes a hydrogen bonding interaction with the hydroxy group of glycolyl (Figure 5a). These intermolecular interactions approximate the loop to the ligand (Figure 5d). In the case of ECoV, residue 115 is Ala and, thus, cannot participate to the intermolecular polar interaction. As consequence, the Ser 116 is not sufficiently close to make an additional

Chapter 3

hydrogen bond with the hydroxyl (Figure 5b). Although residues 114 and 115 are Thr and Ser, respectively for RbCoV, and thus can engage the amide of the Neu5Gc in a polar interaction as for BCoV, residue 116 is a Phe, which cannot establish a hydrogen bond with the hydroxy group of the Gc moiety (Fig. 5c). As result, in both RbCoV and ECoV HEs, the loop is farther from the ligand than that of BCoV HE. (Figure 5d). Taken together, the *in silico* studies indicate residues in the 109-119 loop modulate 5-*N*-glycolyl recognition.

The differences in receptor specificity of Spike and lectin domain of HE of the same virus may be due to the fact they evolved to preferentially recognize different extracellular matrix and cell surface glycoconjugates. The spike has as a main function to bind to cell surface sialylated glycolipid receptors to promote virion uptake.⁸ The lectin domain of HE facilitates initial attachment by binding to *O*-acetylated glycans present in the mucus of the host.¹⁸ It regulates sialate-*O*-acetyl esterase activity by increasing enzyme activity of clustered receptors. As a result, decoy receptors are destroyed facilitating virion migration through mucus. The spike protein may also participate in attaching to multivalent decoy receptors, and for BCoV and OC43 it has been shown that HE and spike function as a two-component system for dynamic receptor interactions.²⁰

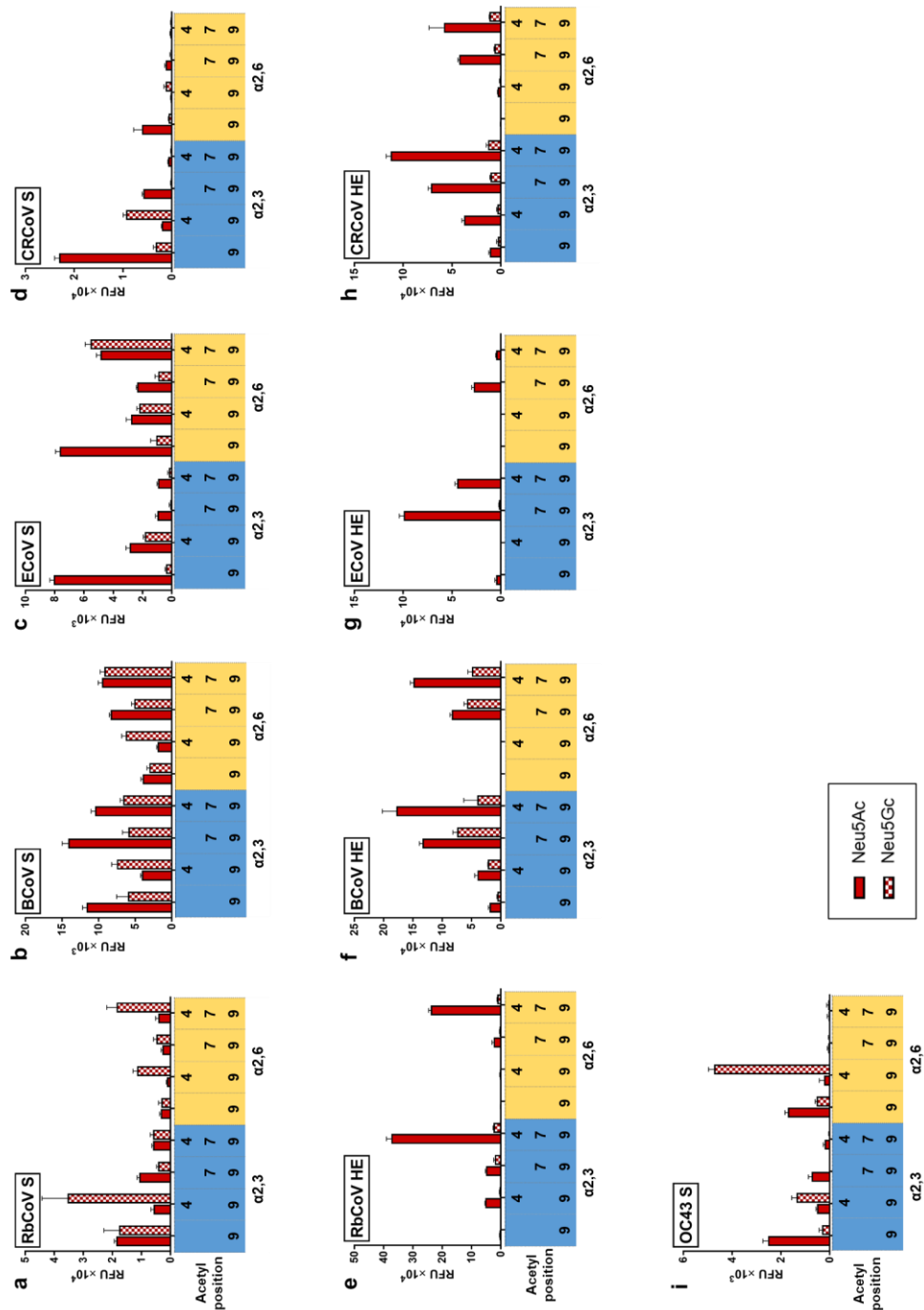
As a result of adaptation to their host sialoglycomes, which can differ dramatically between various species,²¹ binding selectivities of spike and HE are fine-tuned to achieve viral fitness. When Neu5Gc receptor is present in smaller quantities than Neu5Ac counterpart, the degradation of the former receptor by the HE may be reduced accordingly. This can be achieved *via* decreasing the avidity of the lectin domain for Neu5Gc, while the esterase domain remains catalytically active. In bovine, both 5-*N*-glycolylated sialoglycolipids (cell surface receptors) and mucin (decoy receptors) are highly abundant,^{21, 22} which is in line with the observation that both spike and HE of BCoV tolerate the 5-*N*-glycolyl modification. In rabbits, Neu5Gc gangliosides are highly expressed across multiple tissues.²³ Consistently, the spike protein showed excellent binding to Neu5Gc structures. Mucins of rabbits appear to have lower level of 5-*N*-glycolylation compared to 5-*N*-acetylation, which may explain the reduced tolerance of the HE lectin domain for *O*-acetylated Neu5Gc sialosides. Likewise, in equine, Neu5Gc has been found largely on cell surface glycoconjugates, and it appears that secreted mucin harbor greatly reduced level of Neu5Gc.^{21, 24} This is consistent with the observation that the spike moderately tolerated Neu5Gc sialosides, whereas this is not the case for HE. It is likely that RbCoV and ECoV have tuned-down the hydrolytic activities for 9-*O*-acetylated Neu5Gc glycans in response to the smaller quantities present in the mucus of these animals to facilitate initial attachment. An extreme example of decreasing catalytic efficiency by compromising lectin domain binding is the HE of OC43, which harbors a non-functional lectin domain and as a result hydrolyzes multivalent 9-*O*-acetylated sialosides much slower than the HE of ancestral BCoV.¹⁸

BCoV, OC43 and CRCoV spikes and HEs exhibit much lower affinities for Neu5Gc. Recently, it was observed that a 5-*N*-glycolyl moiety is not tolerated by human influenza C virus hemagglutinin-esterase-fusion protein,²⁵ but is recognized by distantly related influenza D virus, which infect ruminants or swine. The loss or reduction of Neu5Gc binding is likely a result of viral adaptation to sialoglycome of human and canine,⁹ in which Neu5Gc biosynthesis is abrogated. 4,9-di-*O*-acetylated Neu5Gc was well recognized by the OC43 spike, which was surprising because humans do not harbor 4-*O*-acetylated sialosides.²⁶ It is possible that the recognition of this sialoside is a remnant from an ancestral virus infecting a different host species. The observations warrant structural studies of spike proteins in complex with *O*-acetylated sialic acid receptors other than the 9-monoacetylated form. Furthermore, the results of these studies call for more detailed characterization of *O*-

acetylated sialosides in the extracellular matrix and on the cell surface to further rationalize receptor specificities of spike and lectin domain of HE of the various viruses.

In conclusion, we have found that that HEs of BCoV and MHV-S can readily cleave *O*-acetylated sialosides opening the possibility for HE-mediated chemoenzymatic preparation of *O*-acetylated Neu5Gc oligosaccharides. The resulting compounds were printed as glycan microarray for the in-depth analysis of receptor binding preference of a panel of coronaviral S proteins and HEs. It was found the spike proteins exhibit distinct but promiscuities binding patterns for *O*-acetylated, *N*-glycolylated sialosides. The lectin domain of the examined HE exhibited moderate to large reduction in binding. The observations indicate the sialic acid populations, or sialome, across the corresponding hosts may vary considerably in composition, and accordingly, the receptor binding specificities of the viral glycoproteins have been fine-tuned for optimal interaction with specific sialoside forms for the virus to breach the species barrier and become adapted in the new hosts. We reveal by MD simulations that glycolyl recognition is modulated by polar residues in a loop region (109-119) close to the 5-*N*-glycolyl moiety. In addition to host-guest interactions, Neu5Gc plays a significant role in chronic inflammation and cancer progression, likely by inducing anti-Neu5Gc antibodies.²⁷ This repertoire remains to be characterized in more depth with a full spectrum of Neu5Gc variants, particularly as antibodies against *O*-acetylated sialic acid have been reported.^{28, 29} We envision that the *O*-acetylated Neu5Gc sialoside library described here will facilitate fingerprinting of the anti-Neu5Gc response, allowing dissecting pathological relevance of Neu5Gc sialoglycans.

Figure 4 | Microarray analysis of coronaviral spike proteins and HEs. **a-d** and **i**, Column graphs for spike proteins (S1^A domain) from RbCoV, BCoV, ECoV, CRCoV and OC43, respectively. **e-h**, Column graphs for HEs from RbCoV, BCoV, ECoV and CRCoV. OC43 HE has a non-functional lectin domain and does not bind *O*-acetylated sialic acid. Protein concentration: RbCoV S1^A 1 µg/mL, BCoV S1^A 0.3 µg/mL, ECoV S1^A 0.3 µg/mL, CRCoV S1^A 3 µg/mL, OC43 S1^A 0.3 µg/mL, RbCoV HE 3 µg/mL, BCoV HE 1 µg/mL, ECoV HE 1 µg/mL, CRCoV HE 1 µg/mL. All proteins are tagged with human Fc and detected with AlexaFluor647-conjugated goat anti-human IgG antibody. Solid fill indicates Neu5Ac binding and pattern fill indicates Neu5Gc binding.



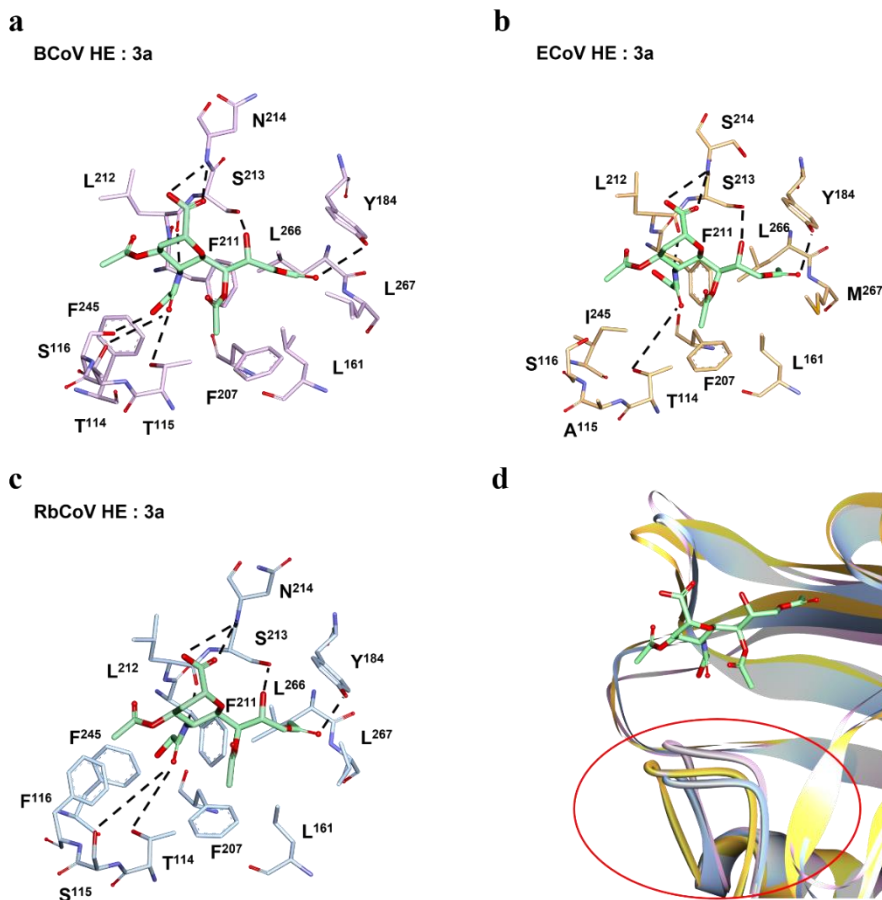


Figure 5 | Structural basis of HE-receptor interaction. Stick representation of the atomic details of **a)** BCoV HE:3a, **b)** ECoV HE:3a and **c)** RbCoV HE:3a, interactions as determined by MD simulations. Dashed black lines represent hydrogen bond interactions. **d)** Cartoon representation of the HEs-receptor interaction, with a focus onto the 109-119 loop. Ligand is in green stick, proteins are in pink (BCoV), yellow (ECoV) and cyan (RbCoV) ribbon. Note the proximity of the loop to the receptor, which is closer for BCoV but not for ECoV and RbCoV.

Chapter 3

References

1. Varki, A. Nothing in Glycobiology Makes Sense, except in the Light of Evolution. *Cell* 2006, **126**(5): 841-845.
2. Corfield, A. P.; Berry, M. Glycan Variation and Evolution in the Eukaryotes. *Trends Biochem Sci* 2015, **40**(7): 351-359.
3. Schauer, R.; Kamerling, J. P. Exploration of the Sialic Acid World. *Adv Carbohydr Chem Biochem* 2018, **75**: 1-213.
4. Shaw, L.; Schauer, R. The Biosynthesis of *N*-glycolylneuraminic Acid Occurs by Hydroxylation of the CMP-Glycoside of *N*-Acetylneuraminic Acid. *Biol Chem Hoppe Seyler* 1988, **369**(6): 477-486.
5. Byres, E.; Paton, A. W.; Paton, J. C.; Löfing, J. C.; Smith, D. F.; Wilce, M. C. J., *et al.* Incorporation of a Non-Human Glycan Mediates Human Susceptibility to a Bacterial Toxin. *Nature* 2008, **456**(7222): 648-652.
6. Wasik, B. R.; Barnard, K. N.; Parrish, C. R. Effects of Sialic Acid Modifications on Virus Binding and Infection. *Trends Microbiol* 2016, **24**(12): 991-1001.
7. Morse, S. S.; Mazet, J. A. K.; Woolhouse, M.; Parrish, C. R.; Carroll, D.; Karesh, W. B., *et al.* Prediction and Prevention of the Next Pandemic Zoonosis. *Lancet* 2012, **380**: 1956-1965.
8. Li, Z.; Lang, Y.; Liu, L.; Bunyatov, M. I.; Sarmiento, A. I.; de Groot, R. J., *et al.* Synthetic *O*-Acetylated Sialosides Facilitate Functional Receptor Identification for Human Respiratory Viruses. *Nat Chem* 2021, **13**(5): 496-503.
9. Peri, S.; Kulkarni, A.; Feyertag, F.; Berninsone, P. M.; Alvarez-Ponce, D. Phylogenetic Distribution of CMP-Neu5Ac Hydroxylase (CMAH), the Enzyme Synthetizing the Proinflammatory Human Xenoantigen Neu5Gc. *Genome Biology and Evolution* 2018, **10**(1): 207-219.
10. Long, J. S.; Mistry, B.; Haslam, S. M.; Barclay, W. S. Host and Viral Determinants of Influenza A Virus Species Specificity. *Nat Rev Microbiol* 2018, **17**(2): 67-81.
11. Chuang, H. Y.; Ren, C. T.; Chao, C. A.; Wu, C. Y.; Shivatare, S. S.; Cheng, T. J., *et al.* Synthesis and Vaccine Evaluation of the Tumor-Associated Carbohydrate Antigen RM2 from Prostate Cancer. *J Am Chem Soc* 2013, **135**(30): 11140-11150.
12. Hamagami, H.; Kumazoe, M.; Yamaguchi, Y.; Fuse, S.; Tachibana, H.; Tanaka, H. 6-Azido-6-Deoxy-*L*-Idose as a Hetero-Bifunctional Spacer for the Synthesis of Azido-Containing Chemical Probes. *Chem Eur J* 2016, **22**(36): 12884-12890.
13. Tamai, H.; Ando, H.; Ishida, H.; Kiso, M. First Synthesis of a Pentasaccharide Moiety of Ganglioside GAA-7 Containing Unusually Modified Sialic Acids through the Use of *N*-Troc-Sialic Acid Derivative as a Key Unit. *Org Lett* 2012, **14**(24): 6342-6345.
14. Zeng, Q.; Langereis, M. A.; van Vliet, A. L.; Huizinga, E. G.; de Groot, R. J. Structure of Coronavirus Hemagglutinin-Esterase Offers Insight into Corona and Influenza Virus Evolution. *Proc Natl Acad Sci USA* 2008, **105**(26): 9065-9069.
15. Regl, G.; Kaser, A.; Iwersen, M.; Schmid, H.; Kohla, G.; Strobl, B., *et al.* The Hemagglutinin-Esterase of Mouse Hepatitis Virus Strain S is a Sialate-4-*O*-Acetylesterase. *J Virol* 1999, **73**(6): 4721.
16. Li, F. Structure, Function, and Evolution of Coronavirus Spike Proteins. *Annu Rev Virol* 2016, **3**(1): 237-261.
17. Smits, S. L.; Gerwig, G. J.; van Vliet, A. L. W.; Lissenberg, A.; Briza, P.; Kamerling, J. P., *et al.* Nidovirus Sialate-*O*-Acetylesterases - Evolution and Substrate Specificity of Coronaviral and Toroviral Receptor-Destroying Enzymes. *J Biol Chem* 2005, **280**(8): 6933-6941.
18. Bakkers, M. J.; Lang, Y.; Feitsma, L. J.; Hulswit, R. J.; de Poot, S. A.; van Vliet, A. L., *et al.* Betacoronavirus Adaptation to Humans Involved Progressive Loss of Hemagglutinin-Esterase Lectin Activity. *Cell Host Microbe* 2017, **21**(3): 356-366.
19. Li, Z.; Unione, L.; Liu, L.; Lang, Y.; de Vries, R. P.; de Groot, R. J., *et al.* Synthetic *O*-Acetylated Sialosides and their Acetamido-Deoxy Analogues as Probes for Coronaviral Hemagglutinin-esterase Recognition. *J Am Chem Soc* 2021.

20. Lang, Y.; Li, W.; Li, Z.; Koerhuis, D.; van den Burg, A. C. S.; Rozemuller, E., *et al.* Coronavirus Hemagglutinin-Esterase and Spike Proteins Coevolve for Functional Balance and Optimal Virion Avidity. *Proc Natl Acad Sci U S A* 2020, **117**(41): 25759-25770.
21. Barnard, K. N.; Alford-Lawrence, B. K.; Buchholz, D. W.; Wasik, B. R.; LaClair, J. R.; Yu, H., *et al.* Modified Sialic Acids on Mucus and Erythrocytes Inhibit Influenza A Virus Hemagglutinin and Neuraminidase Functions. *J Virol* 2020, **94**(9): e01567-01519.
22. Davies, L. R. L.; Pearce, O. M. T.; Tessier, M. B.; Assar, S.; Smutova, V.; Pajunen, M., *et al.* Metabolism of Vertebrate Amino Sugars with *N*-Glycolyl Groups: Resistance of α 2,8-Linked *N*-Glycolylneuraminic Acid to Enzymatic Cleavage. *J Biol Chem* 2012, **287**(34): 28917-28931.
23. Iwamori, M.; Nagai, Y. GM3 Ganglioside in Various Tissues of rabbit. Tissue-Specific Distribution of *N*-Glycolylneuraminic Acid-Containing GM3. *J Biochem* 1978, **84**(6): 1609-1615.
24. Dorvignit, D.; Boligan, K. F.; Relova-Hernandez, E.; Clavell, M.; Lopez, A.; Labrada, M., *et al.* Antitumor Effects of the GM3(Neu5Gc) Ganglioside-Specific Humanized Antibody 14F7hT against Cmah-Transfected Cancer Cells. *Sci Rep* 2019, **9**(1): 9921.
25. Liu, R.; Sreenivasan, C.; Yu, H.; Sheng, Z.; Newkirk, S. J.; An, W., *et al.* Influenza D Virus Diverges from Its Related Influenza C Virus in the Recognition of 9-*O*-Acetylated *N*-Acetyl- or *N*-Glycolyl-Neuraminic Acid-containing Glycan Receptors. *Virology* 2020, **545**: 16-23.
26. Langereis, M. A.; Bakkers, M. J.; Deng, L.; Padler-Karavani, V.; Vervoort, S. J.; Hulswit, R. J., *et al.* Complexity and Diversity of the Mammalian Sialome Revealed by Nidovirus Virolectins. *Cell Rep* 2015, **11**(12): 1966-1978.
27. Samraj, A. N.; Pearce, O. M.; Laubli, H.; Crittenden, A. N.; Bergfeld, A. K.; Banda, K., *et al.* A Red Meat-Derived Glycan Promotes Inflammation and Cancer Progression. *Proc Natl Acad Sci U S A* 2015, **112**(2): 542-547.
28. Bandyopadhyay, S.; Mukherjee, K.; Chatterjee, M.; Bhattacharya, D. K.; Mandal, C. Detection of Immune-Complexed 9-*O*-Acetylated Sialoglycoconjugates in the Sera of Patients with Pediatric Acute Lymphoblastic Leukemia. *J Immunol Method* 2005, **297**(1): 13-26.
29. Ritter, G.; Ritter-Boosfeld, E.; Adluri, R.; Calves, M.; Ren, S.; Yu, R. K., *et al.* Analysis of the Antibody Response to Immunization with Purified *O*-Acetyl GD3 Gangliosides in Patients with Malignant Melanoma. *Intl J Cancer* 1995, **62**(6): 668-672.

Supplementary Information for

Synthetic *O*-Acetyl-*N*-glycolylneuraminic Acid Oligosaccharides Reveal Host-Associated Binding Patterns of Coronaviral Glycoproteins

Part I. Material and methods

Chemicals

Unless otherwise noted, all chemicals used in this work were the same as the previous report (see Ref.8 main text). 2-benzyloxyacetic acid *N*-hydroxy succinimide ester was prepared from 2-benzyloxyacetyl chloride (301019, Sigma-Aldrich).

General analytical procedures

TLC analysis was performed using precoated silica gel 60 F-254 plates (Merck). The plates were either visualized under a UV lamp or stained with ceric ammonium molybdate (5 g $\text{Ce}(\text{SO}_4)_2$, 25 g $(\text{NH}_4)_6\text{Mo}_7\text{O}_{24}\cdot 4\text{H}_2\text{O}$, 50 mL conc. H_2SO_4 , 450 mL H_2O). Column chromatography was used to purify synthetic intermediates and was performed using flash silica gel (60 Å, from Silicycle, Canada). Deprotected compounds were purified using p-2 biogel (Biorad) and collected with BioFrac fraction collector (Biorad). The p-2 purified product was first run on analytical HILIC-HPLC system (XBridge HILIC 5 μm , 4.6 \times 250mm) prior to preparative scale purification (XBridge BEH Prep Amide 5 μm , 10 \times 250mm). ^1H - and ^{13}C -NMR spectroscopy was conducted on an Agilent 400-MR or Bruker 600 UltraShield. Reaction monitoring using ESI-MS was performed on Bruker micrOTOF-Q II ESI mass spectrometer. HRMS data was collected on Agilent 6560 ion mobility Q-ToF MS.

General synthetic procedures

General procedure I. Hydrogenation, acetyl migration and biotinylation. The procedure was similar to the previous work (Ref. 8 in the main text), with slight modification. In a double-necked flask, the starting material (e.g. **16**) was dissolved in THF/water (1:1 v/v) with containing acetic acid (5 eq., to starting materials). Palladium hydroxide (Degussa type) was added to the flask while the solution is in hydrogen atmosphere. The reaction was monitored with ESI-MS (direct infusion, negative mode). Upon completion, the catalyst was filtered off with Celite and the filtrate was concentrated in vacuo (water bath temperature 20 °C) to remove organic solvent, and then freeze-dried. The resulted sialoside carries *O*-Ac moieties at C-4, C-7, and C-8-positions. The deprotected sugar was dissolved in DMSO-*d*₆, followed by addition of 3 eq. acetic acid-*d*₄ and 6 eq. cesium fluoride. An aliquot (~500 μL) was taken for starting point ^1H -NMR. The mixture was heated at 80 °C for 14-20 hours, after which the end point ^1H -NMR was recorded. The shift of Sia H-8 from ~5 ppm to ~4 ppm indicated the reaction was completed. The mixture was then lyophilized. The residue, which is 4,7,9-tri-*O*-acetylated sialoside, was directly taken up in DMSO, to which a solution of biotin (for **1** and **2**) or Lc-biotin (for **3**) *N*-hydroxysuccinimide ester was added. Upon completion monitored by ESI-MS (direct infusion, negative mode), the solution was freeze-dried and the crude mixture was first desalted using biogel p-2 (5% n-butanol in MilliQ water, 4 °C), and

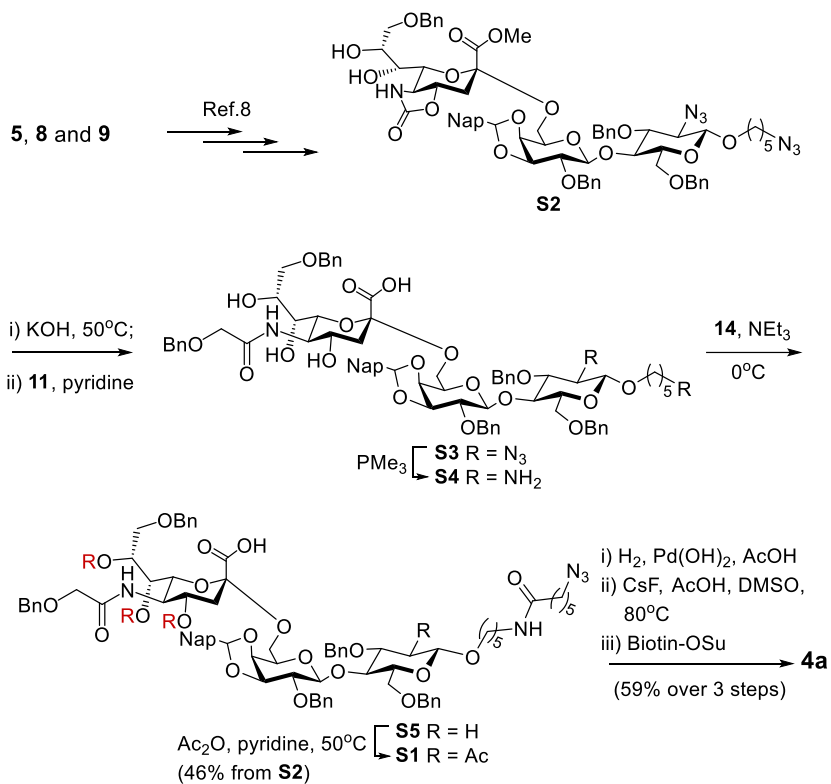
then further purified with HILIC (acetonitrile/water 95% - 65%, 0.1% formic acid, 2.4mL/min, 50 min) to give the biotinylated common precursor sialosides (**1a**, **2a** and **3a**).

General procedure II. HE-mediated de-O-acetylation. The procedure was similar to the previous work (see Ref.8 main text). Briefly, ~2 mg α 2,3-, 2,6- or 2,8-linked common precursor (**1a**, **2a** or **3a**, respectively) was dissolved in freshly prepared ammonium formate buffer (50 mM, pH 6.4) in a 1.5 mL Eppendorf tube (5 mM substrate concentration). The stock solution of HE was then added to achieve a working concentration of 10 μ g/mL for both HEs. The reaction was monitored using ESI-MS (direct infusion, negative mode). Upon completion, the solution was freeze-dried and purified with HILIC (acetonitrile/water 95% - 65%, 0.1% formic acid, 2.4mL/min, 50 min).

General procedure III. Base-mediated acetyl ester migration. The procedure was similar to the previous work (see Ref.8 main text). Briefly, 2-4 mg 7-*O*-acetylated or 4,7-di-*O*-acetylated substrate was directly taken up in 100 mM ammonium bicarbonate in D₂O at the concentration of 2.5-5 mM, which gave a pD(H) of ~8. The solution was transferred into an NMR tube. The starting point (t_0) ¹H-NMR spectrum was recorded with water suppression immediately after making the reaction solution. The NMR tube was placed in an incubator at 37 °C. The reaction was monitored by recording the ¹H-NMR spectra at different time points. The disappearance of the signal around 5.15 ppm, which is given rise to by H-7 of Neu5Gc, indicated *O*7-to-*O*9 migration. The solution was then freeze-dried for 2-3 times to completely remove ammonium bicarbonate.

Chapter 3

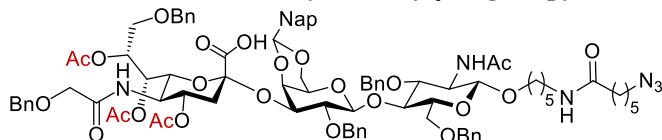
Part II. Detailed synthetic procedures



Supplementary Figure S1. Synthesis of α 2,6 sialoside **4a**.

Synthetic procedure and compound characterization

5-(6-azidohexanamido)pentyl (4,7,8-tri-*O*-acetyl-9-*O*-benzyl-5-(2-(benzyloxy)acetamido)-3,5-dideoxy-*D*-glycero- α -*D*-galacto-non-2-ulopyranosylonic acid)-(2 \rightarrow 3)- [2-*O*-benzyl-4,6-*O*-(2-naphthylmethylene)- β -*D*-galactopyranosyl-(1 \rightarrow 4)]-2-acetamido-3,6-di-*O*-benzyl-2-deoxy- β -*D*-glucopyranoside (**16**)



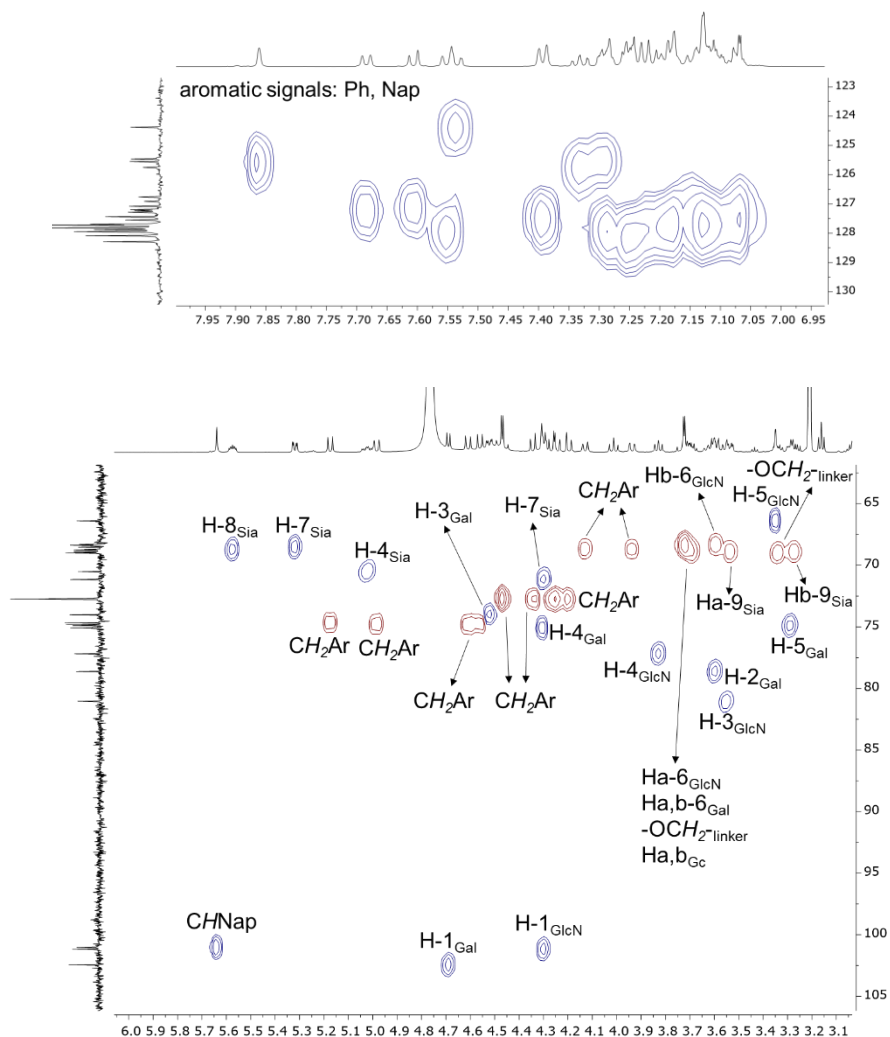
Compound **16** was prepared following the reported procedure in Ref.8 main text with modification at sialate-5-*N*-acylation step.

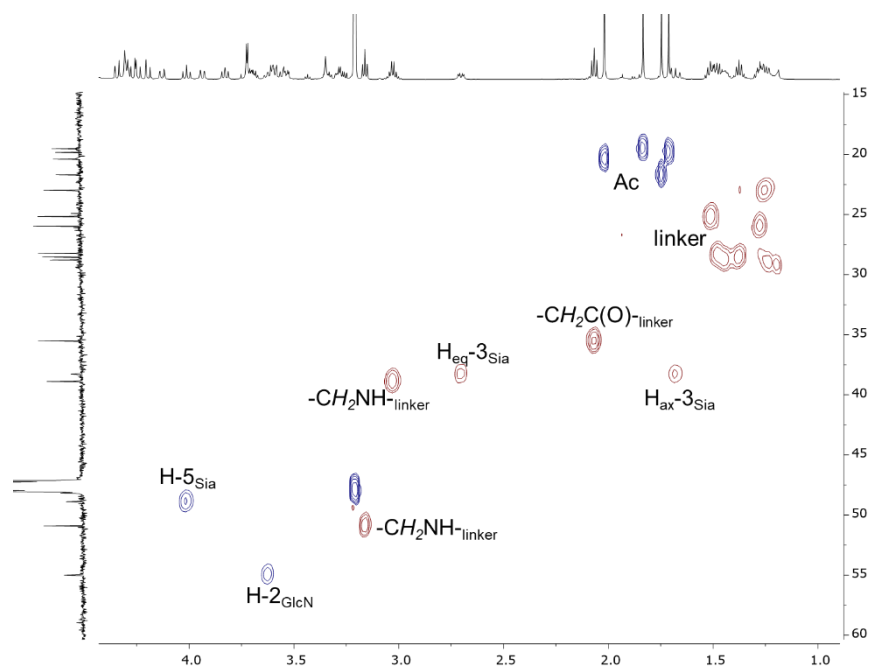
In a 50 mL round bottom flask, the starting compound **10** (340 mg, 0.26 mmol) was suspended in dioxane/water (1:1 v/v, 10 mL total volume), followed by adding KOH (solid, 16 eq. to **10**, 237 mg, 4.2 mmol). The suspension turned to clear solution, which was heated at 50 °C for 90 minutes. The solution was neutralized (at 0 °C) with 1 M HCl to achieve a pH of 8-9, which was then concentrated in vacuo. As an alternative to remove excess base in the system, the residue solid can be passed through LH-20 column.

The saponified intermediate was directly taken up with 5 mL pyridine, to which 2-benzyloxyacetic acid *N*-hydroxy succinimide (NHS) ester (3 eq. to **10**, 205 mg, 0.78 mmol) was added. The reaction was monitored with ESI-MS (direct infusion, negative mode). Upon completion of *N*-acylation reaction, the solution was concentrated in vacuo, and co-evaporated with toluene for several times to remove residual pyridine. The crude solid was taken up in 5 mL THF/water (2:1 v/v), followed by the addition of trimethylphosphine (1 M solution in toluene, 4 eq. to **10**, 1 mL, 1.0 mmol) and KOH (4 eq. to **10**, 58 mg, 1.0 mmol) at 0 °C. The Staudinger reaction was allowed to warm up to 21 °C and stirred for 2 hours. Upon completion, the suspension was concentrated with reduced pressure. 10 mL DCM was then added to the crude mixture to make a suspension, which was sonicated and transferred into a 50 mL Falcon tube. The mixture was centrifuged at 4500 rpm for 10 minutes. The supernatant was transferred to a new 50 mL round bottom flask. The precipitate was washed with 5 mL DCM twice and the wash solution was combined in the same flask. Crude product **13** was dissolved in 25 mL DMF/water (4:1, v/v), to which a solution of NBI-activated 6-azidohecanoic acid ester (1.2 eq. to **10**, 99 mg, 0.31 mmol) in DMF was added dropwise at 0 °C. When TLC (ethyl acetate/methanol/water 9:1:0.2) showed full consumption of **13**, a few drops of 1 M NaOH (aq.) was added to quench excess active ester. The solution was concentrated in vacuo. The crude product with anomeric linker *N*-acylated was dissolved in 8 mL pyridine. To this solution, 4 mL acetic anhydride was added on ice. The mixture was then heated at 50 °C for ~20 hours to fully acetylate the hydroxy groups. The reaction was monitored with ESI-MS (direct infusion, negative mode). Upon completion, the solution was cooled to 21 °C, quenched with water and concentrated under reduced pressure. The crude product was purified with a quick silica gel column (ethyl acetate/methanol 9:1) followed preparative scale C4 RP-HPLC (solvent A: 50% MeOH in water, solvent B: acetonitrile; gradient: 5-80% solvent B in 50 min) to give **16** in 42% isolated yield over five steps. ¹H NMR (600 MHz, Methanol-*d*₄) δ 7.86 (s, 1H), 7.68 (d, *J* = 8.1 Hz, 1H), 7.61 (d, *J* = 8.5 Hz, 1H), 7.54 (t, *J* = 8.9 Hz, 2H), 7.39 (d, *J* = 7.4 Hz, 2H), 7.36 – 6.98 (m, 21H), 5.64 (s, 1H), 5.57 (dt, *J* = 8.7, 4.9 Hz, 1H), 5.32 (dd, *J* = 8.9, 2.4 Hz, 1H), 5.17 (d, *J* = 11.6 Hz, 1H), 5.07 – 4.95 (m, 2H), 4.69 (d, *J* = 7.6 Hz, 2H), 4.61 (d, *J* = 12.2 Hz, 1H), 4.56 (d, *J* = 11.7 Hz, 2H), 4.52 (dd, *J* = 9.9, 3.3 Hz, 2H), 4.50 – 4.43 (m, 3H), 4.34 (d, *J* = 11.8 Hz, 1H), 4.32 – 4.22 (m, 4H), 4.20 (d, *J* = 11.7 Hz, 1H), 4.13 (d, *J* = 11.4 Hz, 1H), 4.01 (t, *J* = 10.5 Hz, 1H), 3.94 (d,

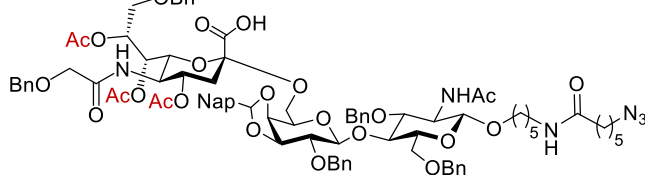
Chapter 3

$J = 11.1$ Hz, 1H), 3.87 – 3.80 (m, 1H), 3.77 – 3.66 (m, 4H), 3.38 – 3.31 (m, 2H), 3.31 – 3.24 (m, 2H), 3.16 (t, $J = 6.8$ Hz, 2H), 3.03 (h, $J = 6.4$ Hz, 2H), 2.70 (dd, $J = 12.0, 4.6$ Hz, 1H), 2.07 (t, $J = 7.4$ Hz, 2H), 2.02 (s, 3H), 1.83 (s, 3H), 1.79 – 1.61 (m, 6H), 1.57 – 1.14 (m, 13H). ^{13}C NMR (151 MHz, MeOD) $\delta = 174.51, 171.90, 171.73, 170.40, 170.36, 170.32, 139.85, 139.12, 138.36, 137.88, 137.38, 136.05, 133.62, 132.87, 128.29, 128.08, 127.95, 127.88, 127.82, 127.72, 127.69, 127.55, 127.44, 127.28, 127.23, 127.19, 127.08, 126.91, 126.76, 125.75, 125.55, 125.46, 124.38, 102.44, 101.17, 101.02, 81.04, 78.62, 77.19, 75.09, 74.90, 74.79, 74.66, 74.02, 72.75, 72.71, 71.15, 70.52, 69.01, 68.92, 68.76, 68.71, 68.53, 68.37, 68.30, 66.40, 55.00, 50.91, 48.89, 48.17, 38.89, 38.28, 35.52, 28.78, 28.53, 28.22, 25.98, 25.16, 22.98, 21.69, 20.39, 19.83, 19.52. See HSQC spectra below for NMR signal assignment. HRMS (neg.) calculated for $\text{C}_{88}\text{H}_{103}\text{N}_6\text{O}_{24}^-$ $[\text{M}-\text{H}]^-$ 1627.7024, found 1627.7020.$



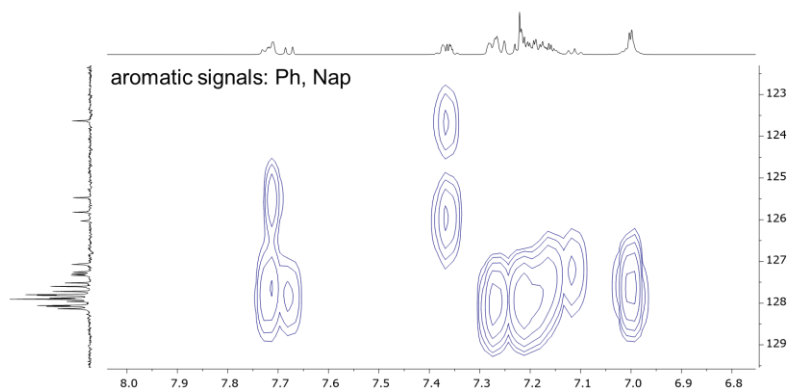


5-(6-azidohexanamido)pentyl (4,7,8-tri-*O*-acetyl-9-*O*-benzyl-5-(2-(benzyloxy)acetamido)-3,5-dideoxy-*D*-glycero- α -*D*-galacto-non-2-ulopyranosylonic acid)-(2 \rightarrow 6)-[2-*O*-benzyl-3,4-*O*-(2-naphthylmethylene)- β -*D*-galactopyranosyl-(1 \rightarrow 4)]-2-acetamido-3,6-di-*O*-benzyl-2-deoxy- β -*D*-glucopyranoside (S1)

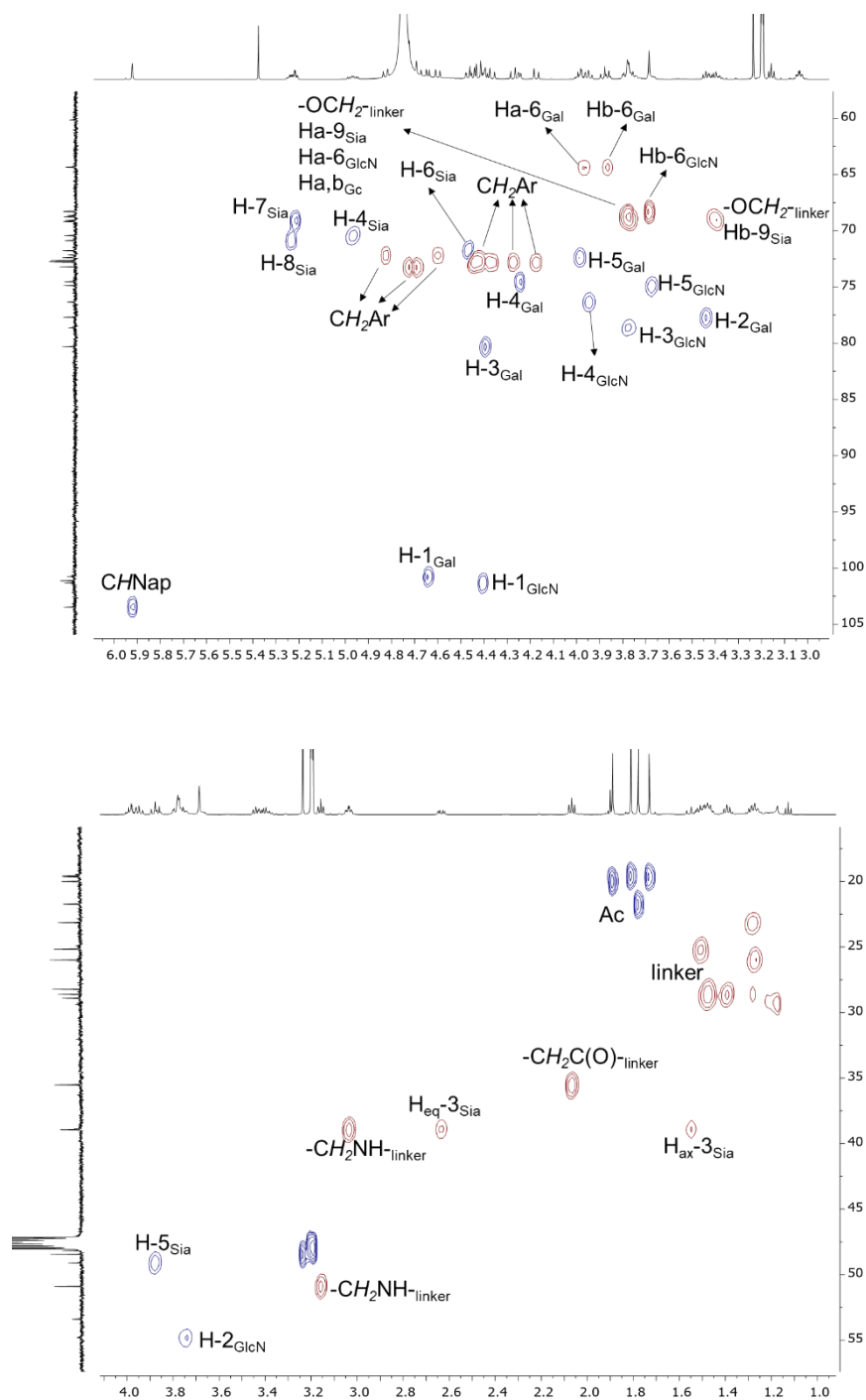


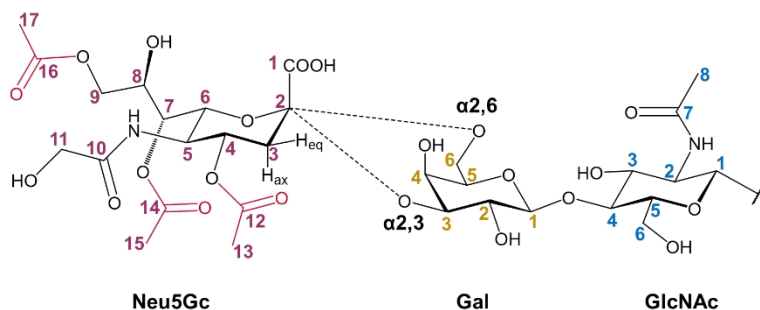
Compound **S1** was prepared following the same procedure described for **16**. Isolated yield, 46% over five steps. ^1H NMR (600 MHz, Methanol- d_4) δ 5.92 (s, 1H), 5.22 (ddd, J = 11.2, 6.3, 2.8

Hz, 2H), 4.97 (td, J = 11.9, 4.9 Hz, 1H), 4.83 (d, J = 11.1 Hz, 1H), 4.69 (s, 1H), 4.64 (d, J = 7.8 Hz, 1H), 4.60 (d, J = 11.1 Hz, 1H), 4.52 – 4.34 (m, 5H), 4.27 (d, J = 11.7 Hz, 1H), 4.25 (d, J = 6.5 Hz, 1H), 4.17 (d, J = 11.7 Hz, 1H), 4.06 – 3.91 (m, 3H), 3.91 – 3.83 (m, 1H), 3.76 (qd, J = 17.4, 15.0, 5.2 Hz, 4H), 3.69 (s, 2H), 3.50 – 3.33 (m, 3H), 3.16 (t, J = 6.8 Hz, 1H), 3.04 (dt, J = 7.2, 3.5 Hz, 2H), 2.63 (dd, J = 11.9, 4.8 Hz, 1H), 2.07 (t, J = 7.5 Hz, 2H), 1.89 (s, 3H), 1.81 (s, 2H), 1.78 (s, 2H), 1.73 (s, 2H), 1.61 – 1.44 (m, 5H), 1.39 (p, J = 7.4 Hz, 2H), 1.34 – 1.23 (m, 3H), 1.13 (t, J = 7.1 Hz, 2H). ^{13}C NMR (151 MHz, MeOD) δ = 174.50, 172.27, 171.82, 171.69, 170.79, 170.67, 170.30, 168.91, 138.87, 138.39, 138.31, 137.95, 137.36, 136.27, 133.81, 132.97, 128.14, 128.08, 128.06, 127.96, 127.91, 127.88, 127.82, 127.80, 127.72, 127.60, 127.52, 127.33, 127.30, 127.25, 127.07, 127.06, 126.03, 125.82, 125.47, 123.63, 103.45, 101.31, 101.12, 100.75, 80.31, 78.62, 77.69, 76.33, 74.88, 74.53, 73.22, 72.80, 72.78, 72.66, 72.39, 72.16, 71.78, 70.88, 70.43, 69.19, 69.09, 68.76, 68.64, 68.29, 64.35, 60.15, 53.40, 50.91, 49.11, 48.45, 48.17, 38.95, 35.53, 28.90, 28.62, 28.23, 25.99, 25.17, 23.14, 21.73, 20.00, 19.63, 19.57. HRMS (neg.) calculated for $\text{C}_{88}\text{H}_{103}\text{N}_6\text{O}_{24}$

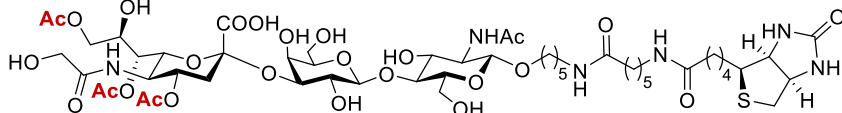


$[\text{M}-\text{H}]^-$ 1627.7024, found 1627.6981.





5-(6-(D-biotinamido)hexanamido)pentyl (4,7,9-tri-*O*-acetyl-3,5-dideoxy-5-(2-hydroxy)acetamido-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid)-(2 \rightarrow 3)- β -D-galactopyranosyl-(1 \rightarrow 4)-2-acetamido-2-deoxy- β -D-glucopyranoside (3a**)**

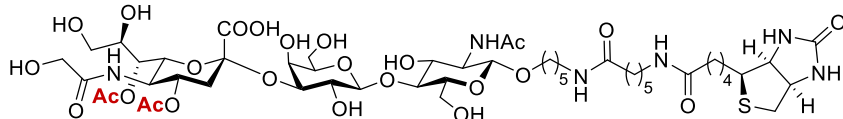


Compound **3a** was prepared from **16** following the **General procedure I** using NHS-biotin as the biotinylating reagent. See table below for NMR signal assignment. HRMS (neg.) calculated for $C_{52}H_{83}N_6O_{25}S^-$ 1239.5078, found 1239.5072.

Neu4,7,9Ac35Gc- α 2,3-Gal- β 1,4-GlcNAc (1a)							
Neu5Gc				Gal			
proton	δ_H (ppm)	carbon	δ_C (ppm)	proton	δ_H (ppm)	carbon	δ_C (ppm)
		C-1	173.0	H-1	4.57	C-1	102.6
		C-2	99.98	H-2	3.60	C-2	69.22
Heq-3	2.84	C-3	36.7	H-3	4.15	C-3	75.98
Hax-3	1.98			H-4	3.98	C-4	67.61
H-4	5.02	C-4	70.45	H-5	3.72	C-5	75.23
H-5	4.02	C-5	48.62	Ha-6	3.70-3.80	C-6	61.00
H-6	4.2	C-6	71.4	Hb-6			
H-7	5.15	C-7	69.19				
H-8	4.28	C-8	67.45				
Ha-9	4.11-4.22	C-9	64.43	GlcNAc			
Hb-9				proton	δ_H (ppm)	carbon	δ_C (ppm)
		C-10	175.6	H-1	4.53	C-1	101.1
Ha-11	3.93-4.09	C-11	61.06	H-2	3.74	C-2	55.11
Hb-11				H-3	3.59	C-3	74.92
		C-12	173.0	H-4	3.74	C-4	78.46
H-13	2.06	C-13	20.29	H-5	3.71	C-5	72.49
		C-14	172.6	Ha-6	4.01	C-6	60.15

H-15	2.14	C-15	20.29	Hb-6	3.87		
		C-16	174.2			C-7	174.3
H-17	2.14	C-17	20.29	H-8	2.04	C-8	22.24

5-(6-(D-biotinamido)hexanamido)pentyl (4,7-di-*O*-acetyl-3,5-dideoxy-5-(2-hydroxy)acetamido-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid)-(2 \rightarrow 3)- β -D-galactopyranosyl-(1 \rightarrow 4)-2-acetamido-2-deoxy- β -D-glucopyranoside (3b**)**

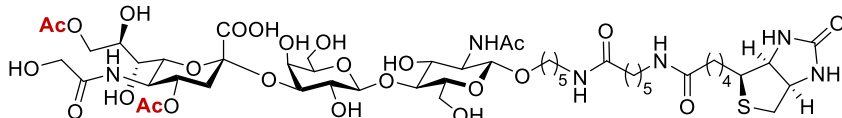


Compound **3b** was prepared from **3a** following the **General procedure II** using BCov-HE as the de-*O*-acetylating enzyme. See table below for NMR signal assignment. HRMS (neg.) calculated for C₅₀H₈₁N₆O₂₄S⁻ 1197.4972, found 1197.4979.

Neu4,7Ac ₂ 5Gc- α 2,3-Gal- β 1,4-GlcNAc (1b)							
Neu5Gc				Gal			
proton	δ_H (ppm)	carbon	δ_C (ppm)	proton	δ_H (ppm)	carbon	δ_C (ppm)
		C-1	172.9	H-1	4.59	C-1	102.7
		C-2	100.0	H-2	3.61	C-2	68.32
Heq-3	2.85	C-3	36.76	H-3	4.17	C-3	76.09
Hax-3	1.98			H-4	4.00	C-4	67.76
H-4	5.03	C-4	70.53	H-5	3.74	C-5	75.29
H-5	4.03	C-5	48.07	Ha-6	3.72-3.78	C-6	60.96
H-6	4.19	C-6	71.6	Hb-6			
H-7	5.09	C-7	69.53				
H-8	4.05	C-8	69.96				
Ha-9	3.71	C-9	61.76	GlcNAc			
Hb-9	3.54			proton	δ_H (ppm)	carbon	δ_C (ppm)
		C-10	175.4	H-1	4.53	C-1	101.1
Ha-11	3.98-4.02	C-11	60.88	H-2	3.74	C-2	55.18
Hb-11				H-3	3.59	C-3	74.98
		C-12	173.0	H-4	3.74	C-4	78.49
H-13	2.06	C-13	20.38	H-5	3.73	C-5	72.58
		C-14	172.7	Ha-6	4.01	C-6	60.14
H-15	2.15	C-15	20.46	Hb-6	3.88		
		C-16	-			C-7	174.4
H-17	-	C-17	-	H-8	2.05	C-8	22.34

Chapter 3

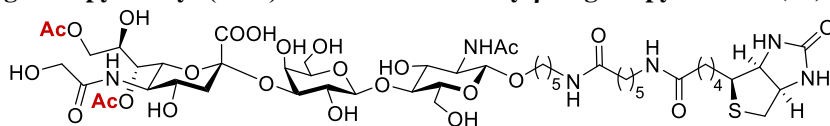
5-(6-(D-biotinamido)hexanamido)pentyl (4,9-di-*O*-acetyl-3,5-dideoxy-5-(2-hydroxy)acetamido-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid)-(2 \rightarrow 3)- β -D-galactopyranosyl-(1 \rightarrow 4)-2-acetamido-2-deoxy- β -D-glucopyranoside (**3c**)



Compound **3c** was prepared from **3b** following the **General procedure III**. See table below for NMR signal assignment. HRMS (neg.) calculated for $C_{50}H_{81}N_6O_{24}S^-$ 1197.4972, found 1197.4976.

Neu4,9Ac25Gc- α 2,3-Gal- β 1,4-GlcNAc (1c)							
Neu5Gc				Gal			
proton	δ_H (ppm)	carbon	δ_C (ppm)	proton	δ_H (ppm)	carbon	δ_C (ppm)
		C-1	173.6	H-1	4.57	C-1	102.8
		C-2	99.56	H-2	3.59	C-2	69.34
Heq-3	2.82	C-3	36.88	H-3	4.15	C-3	75.81
Hax-3	1.95			H-4	3.97	C-4	67.55
H-4	5.05	C-4	70.94	H-5	3.73	C-5	75.29
H-5	4.19	C-5	49.03	Ha-6	3.71-3.78	C-6	61.09
H-6	3.98	C-6	72.16	Hb-6			
H-7	3.70	C-7	68.04				
H-8	4.14	C-8	69.44				
Ha-9	4.04-4.10	C-9	60.95	GlcNAc			
Hb-9				proton	δ_H (ppm)	carbon	δ_C (ppm)
		C-10	174.7	H-1	4.53	C-1	101.3
Ha-11	4.43	C-11	65.74	H-2	3.73	C-2	55.13
Hb-11	4.21			H-3	3.59	C-3	74.98
		C-12	173.6	H-4	3.73	C-4	78.5
H-13	2.08	C-13	20.31	H-5	3.72	C-5	72.67
		C-14	-	Ha-6	4.00	C-6	60.26
H-15	-	C-15	-	Hb-6	3.86		
		C-16	175.0			C-7	174.5
H-17	2.15	C-17	20.38	H-8	2.04	C-8	22.18

5-(6-(D-biotinamido)hexanamido)pentyl (7,9-di-*O*-acetyl-3,5-dideoxy-5-(2-hydroxy)acetamido-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid)-(2 \rightarrow 3)- β -D-galactopyranosyl-(1 \rightarrow 4)-2-acetamido-2-deoxy- β -D-glucopyranoside (**3d**)



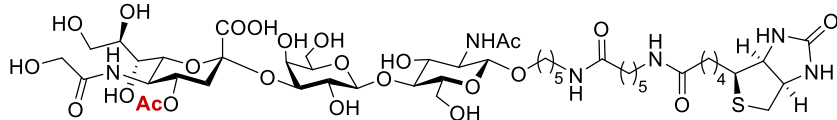
Compound **3d** was prepared from **3a** following

the **General procedure II** using MHV-S-HE as the de-*O*-acetylating enzyme. See table below for NMR signal assignment. HRMS (neg.) calculated for $C_{50}H_{81}N_6O_{24}S^-$ 1197.4972, found 1197.4977.

Neu7,9Ac ₂ 5Gc- α 2,3-Gal- β 1,4-GlcNAc (1d)							
Neu5Gc				Gal			
proton	δ_H (ppm)	carbon	δ_C (ppm)	proton	δ_H (ppm)	carbon	δ_C (ppm)
		C-1	173.4	H-1	4.57	C-1	102.7
		C-2	100	H-2	3.60	C-2	69.25
Heq-3	2.82	C-3	39.70	H-3	4.14	C-3	75.92
Hax-3	1.85			H-4	3.98	C-4	67.57
H-4	3.75	C-4	67.99	H-5	3.72	C-5	75.27
H-5	3.86	C-5	50.91	Ha-6	3.69-3.82	C-6	61.03
H-6	4.04	C-6	71.87	Hb-6			
H-7	5.14	C-7	69.35				
H-8	4.26	C-8	67.67				
Ha-9	4.10-4.23	C-9	64.47	GlcNAc			
Hb-9				proton	δ_H (ppm)	carbon	δ_C (ppm)
		C-10	175.6	H-1	4.53	C-1	101.2
Ha-11	4.03-4.16	C-11	61.04	H-2		C-2	
Hb-11				H-3	3.59	C-3	74.95
		C-12	-	H-4	3.74	C-4	78.47
H-13	-	C-13	-	H-5	3.72	C-5	72.51
		C-14	172.7	Ha-6	4.02	C-6	60.17
H-15	2.14	C-15	20.34	Hb-6	3.87		
		C-16	175.3			C-7	174.3
H-17	2.14	C-17	20.34	H-8	2.04	C-8	22.21

Chapter 3

5-(6-(D-biotinamido)hexanamido)pentyl (4-*O*-acetyl-3,5-dideoxy-5-(2-hydroxy)acetamido-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid)-(2 \rightarrow 3)- β -D-galactopyranosyl-(1 \rightarrow 4)-2-acetamido-2-deoxy- β -D-glucopyranoside (3e**)**

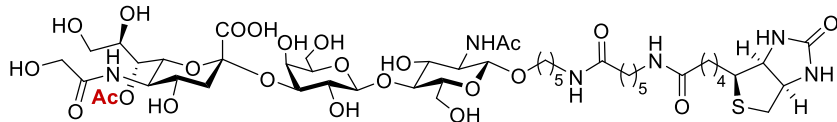


Compound **3e** was prepared from **3c** following

the **General procedure II** using BCoV-HE as the de-*O*-acetylating enzyme. See table below for NMR signal assignment. HRMS (neg.) calculated for C₄₈H₇₉N₆O₂₃S⁻ 1155.4866, found 1155.4859.

Neu4Ac5Gc- α 2,3-Gal- β 1,4-GlcNAc (1e)							
Neu5Gc				Gal			
proton	δ_H (ppm)	carbon	δ_C (ppm)	proton	δ_H (ppm)	carbon	δ_C (ppm)
		C-1	173.5	H-1	4.57	C-1	102.6
		C-2	99.73	H-2	3.59	C-2	69.28
Heq-3	2.81	C-3	36.8	H-3	4.16	C-3	75.58
Hax-3	1.95			H-4	3.98	C-4	67.52
H-4	5.06	C-4	70.94	H-5	3.73	C-5	75.19
H-5	4.18	C-5	49.04	Ha-6	3.72-3.76	C-6	61.02
H-6	3.93	C-6	71.63	Hb-6			
H-7	3.64	C-7	67.92				
H-8	3.92	C-8	71.67				
Ha-9	3.88	C-9	62.55	GlcNAc			
Hb-9	3.66			proton	δ_H (ppm)	carbon	δ_C (ppm)
		C-10	175.3	H-1	4.53	C-1	101.1
Ha-11	4.02-4.07	C-11	60.92	H-2	3.73	C-2	55.06
Hb-11				H-3	3.59	C-3	74.86
		C-12	172.9	H-4	3.73	C-4	78.41
H-13	2.09	C-13	20.36	H-5	3.72	C-5	72.57
		C-14	-	Ha-6	4.00	C-6	60.15
H-15	-	C-15	-	Hb-6	3.86		
		C-16	-			C-7	174.4
H-17	-	C-17	-	H-8	2.04	C-8	22.23

5-(6-(D-biotinamido)hexanamido)pentyl (7-*O*-acetyl-3,5-dideoxy-5-(2-hydroxy)acetamido-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid)-(2 \rightarrow 3)- β -D-galactopyranosyl-(1 \rightarrow 4)-2-acetamido-2-deoxy- β -D-glucopyranoside (3f**)**



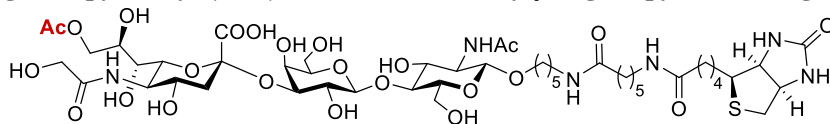
Compound **3f** was prepared from **3a** following

the **General procedure II** using MHV-S- and BCoV-HEs as the de-*O*-acetylating enzymes. See table below for NMR signal assignment. HRMS (neg.) calculated for C₄₈H₇₉N₆O₂₃S⁻ 1155.4866, found 1155.4868.

Neu7Ac5Gc- α 2,3-Gal- β 1,4-GlcNAc (1f)							
Neu5Gc				Gal			
proton	δ_H (ppm)	carbon	δ_C (ppm)	proton	δ_H (ppm)	carbon	δ_C (ppm)
		C-1	173.5	H-1	4.58	C-1	102.7
		C-2	100.2	H-2	3.61	C-2	69.31
Heq-3	3.00	C-3	39.69	H-3	4.15	C-3	75.99
Hax-3	1.85			H-4	4.00	C-4	67.69
H-4	3.75	C-4	68.04	H-5	3.74	C-5	75.27
H-5	3.86	C-5	50.96	Ha-6	3.71-3.79	C-6	61.00
H-6	4.03	C-6	71.99	Hb-6			
H-7	5.06	C-7	69.59				
H-8	4.03	C-8	70.14				
Ha-9	3.69	C-9	61.74	GlcNAc			
Hb-9	3.52			proton	δ_H (ppm)	carbon	δ_C (ppm)
		C-10	175.5	H-1	4.53	C-1	101.16
Ha-11	4.01-4.10	C-11	60.93	H-2	3.73	C-2	55.14
Hb-11				H-3	3.59	C-3	74.91
		C-12	-	H-4	3.74	C-4	78.48
H-13	-	C-13	-	H-5	3.72	C-5	72.56
		C-14	173.0	Ha-6	4.01	C-6	60.18
H-15	2.15	C-15	20.48	Hb-6	3.88		
		C-16	-			C-7	174.4
H-17	-	C-17	-	H-8	2.04	C-8	22.26

Chapter 3

5-(6-(D-biotinamido)hexanamido)pentyl (9-O-acetyl-3,5-dideoxy-5-(2-hydroxy)acetamido-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid)-(2 \rightarrow 3)- β -D-galactopyranosyl-(1 \rightarrow 4)-2-acetamido-2-deoxy- β -D-glucopyranoside (3g**)**

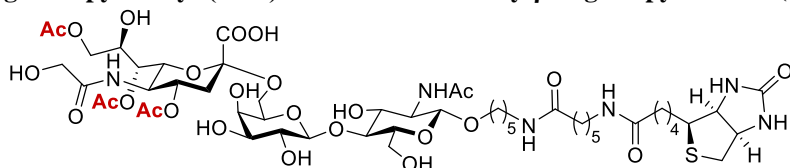


Compound **3g** was prepared from **3f** following

the **General procedure III**. See table below for NMR signal assignment. HRMS (neg.) calculated for $C_{48}H_{79}N_6O_{23}S^-$ 1155.4866, found 1155.4860.

Neu9Ac5Gc- α 2,3-Gal- β 1,4-GlcNAc (1g)							
Neu5Gc				Gal			
proton	δ_H (ppm)	carbon	δ_C (ppm)	proton	δ_H (ppm)	carbon	δ_C (ppm)
		C-1	173.6	H-1	4.57	C-1	102.6
		C-2	99.75	H-2	3.59	C-2	69.26
Heq-3	2.79	C-3	39.81	H-3	4.14	C-3	75.61
Hax-3	1.83			H-4	3.97	C-4	67.45
H-4	3.81	C-4	68.16	H-5	3.72	C-5	75.22
H-5	3.97	C-5	51.36	Ha-6	3.72-3.78	C-6	61.02
H-6	3.80	C-6	72.49	Hb-6			
H-7	3.67	C-7	68.12				
H-8	4.12	C-8	69.4				
Ha-9	4.12-4.16	C-9	61.01	GlcNAc			
Hb-9				proton	δ_H (ppm)	carbon	δ_C (ppm)
		C-10	174.4	H-1	4.53	C-1	101.1
Ha-11	4.43	C-11	65.71	H-2	3.74	C-2	55.07
Hb-11	4.20			H-3	3.59	C-3	74.89
		C-12	-	H-4	3.73	C-4	78.35
H-13	-	C-13	-	H-5	3.72	C-5	72.56
		C-14	-	Ha-6	4.00	C-6	60.12
H-15	-	C-15	-	Hb-6	3.86		
		C-16	175.0			C-7	174.0
H-17	2.15	C-17	20.32	H-8	2.04	C-8	22.24

5-(6-(D-biotinamido)hexanamido)pentyl (4,7,9-tri-*O*-acetyl-3,5-dideoxy-5-(2-hydroxy)acetamido-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid)-(2 \rightarrow 6)- β -D-galactopyranosyl-(1 \rightarrow 4)-2-acetamido-2-deoxy- β -D-glucopyranoside (4a)



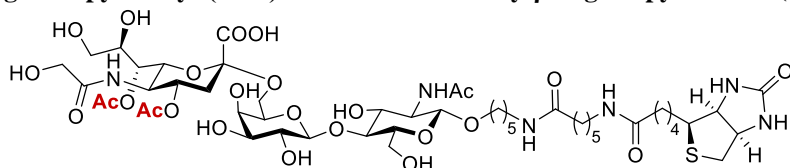
Compound **4a** was prepared from **S1** following the **General procedure I**

using NHS-biotin as the biotinyating reagent. See table below for NMR signal assignment. HRMS (neg.) calculated for $C_{52}H_{83}N_6O_{25}S^-$ 1239.5078, found 1239.5083.

Neu4,7,9Ac ₃ 5Gc- α 2,6-Gal- β 1,4-GlcNAc (2a)							
Neu5Gc				Gal			
proton	δ_H (ppm)	carbon	δ_C (ppm)	proton	δ_H (ppm)	carbon	δ_C (ppm)
		C-1	172.5	H-1	4.48	C-1	103.3
		C-2	100.4	H-2	3.56	C-2	70.88
Heq-3	2.80	C-3	37.5	H-3	3.68	C-3	72.54
Hax-3	1.86			H-4	3.96	C-4	68.38
H-4	5.00	C-4	70.37	H-5	3.77	C-5	72.47
H-5	4.04	C-5	48.7	Ha-6	3.99	C-6	63.54
H-6	4.22	C-6	71.32	Hb-6	3.64		
H-7	5.20	C-7	69.19				
H-8	4.28	C-8	67.3				
Ha-9	4.17	C-9	64.13	GlcNAc			
Hb-9	4.14			proton	δ_H (ppm)	carbon	δ_C (ppm)
		C-10	174.5	H-1	4.56	C-1	101.07
Ha-11	3.94-4.01	C-11	60.81	H-2	3.74	C-2	55.04
Hb-11				H-3	3.62	C-3	74.66
		C-12	173.1	H-4	3.67	C-4	80.25
H-13	2.07	C-13	20.39	H-5	3.84	C-5	73.63
		C-14	172.8	Ha-6	4.00	C-6	60.45
H-15	2.15	C-15	20.51	Hb-6	3.86		
		C-16	172.9			C-7	174.2
H-17	2.14	C-17	20.33	H-8	2.08	C-8	22.4

Chapter 3

5-(6-(D-biotinamido)hexanamido)pentyl (4,7-di-*O*-acetyl-3,5-dideoxy-5-(2-hydroxy)acetamido-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid)-(2 \rightarrow 6)- β -D-galactopyranosyl-(1 \rightarrow 4)-2-acetamido-2-deoxy- β -D-glucopyranoside (**4b**)

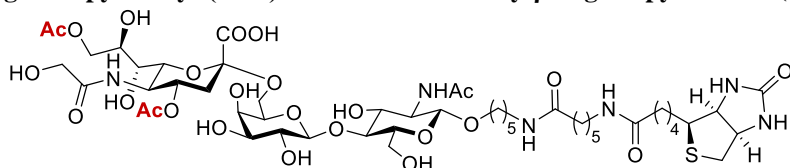


Compound **4b** was prepared from **4a** following the **General procedure II**

using BCoV-HE as the de-*O*-acetylating enzyme. See table below for NMR signal assignment. HRMS (neg.) calculated for $C_{50}H_{81}N_6O_{24}S^-$ 1197.4972, found 1197.4974.

Neu4,7Ac ₂ 5Gc- α 2,6-Gal- β 1,4-GlcNAc (2b)							
Neu5Gc				Gal			
proton	δ_H (ppm)	carbon	δ_C (ppm)	proton	δ_H (ppm)	carbon	δ_C (ppm)
		C-1	172.6	H-1	4.48	C-1	103.4
		C-2	100.5	H-2	3.56	C-2	70.99
Heq-3	2.80	C-3	37.52	H-3	3.68	C-3	72.62
Hax-3	1.85			H-4	3.96	C-4	68.44
H-4	4.99	C-4	70.51	H-5	3.84	C-5	73.67
H-5	4.03	C-5	48.78	Ha-6	3.99	C-6	63.56
H-6	4.21	C-6	71.57	Hb-6	3.64		
H-7	5.12	C-7	69.66				
H-8	4.04	C-8	69.88				
Ha-9	3.69	C-9	61.74	GlcNAc			
Hb-9	3.5			proton	δ_H (ppm)	carbon	δ_C (ppm)
		C-10	175.1	H-1	4.56	C-1	101.2
Ha-11	3.97-4.02	C-11	60.91	H-2	3.73	C-2	55.1
Hb-11				H-3	3.62	C-3	74.75
		C-12	172.9	H-4	3.67	C-4	80.35
H-13	2.06	C-13	20.4	H-5	3.77	C-5	72.58
		C-14	172.6	Ha-6	3.85	C-6	60.50
H-15	2.15	C-15	20.56	Hb-6	4.00		
		C-16	-			C-7	174.1
H-17	-	C-17	-	H-8	2.09	C-8	22.38

5-(6-(D-biotinamido)hexanamido)pentyl (4,9-di-O-acetyl-3,5-dideoxy-5-(2-hydroxy)acetamido-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid)-(2 \rightarrow 6)- β -D-galactopyranosyl-(1 \rightarrow 4)-2-acetamido-2-deoxy- β -D-glucopyranoside (4c)



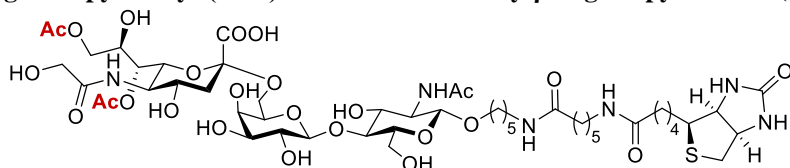
Compound **4c** was prepared from **4a** following the **General procedure III**.

See table below for NMR signal assignment. HRMS (neg.) calculated for $C_{50}H_{81}N_6O_{24}S^-$ 1197.4972, found 1197.4974.

Neu4,9Ac ₂ 5Gc- α 2,6-Gal- β 1,4-GlcNAc (2c)							
Neu5Gc				Gal			
proton	δ_H (ppm)	carbon	δ_C (ppm)	proton	δ_H (ppm)	carbon	δ_C (ppm)
		C-1	173.1	H-1	4.46	C-1	103.8
		C-2	100.1	H-2	3.56	C-2	70.89
Heq-3	2.72	C-3	37.32	H-3	3.69	C-3	72.61
Hax-3	1.86			H-4	3.94	C-4	68.48
H-4	5.01	C-4	70.79	H-5	3.85	C-5	73.89
H-5	4.15	C-5	48.27	Ha-6	4.03	C-6	63.64
H-6	4.04	C-6	71.93	Hb-6	3.55		
H-7	3.66	C-7	68.35				
H-8	4.14	C-8	69.23				
Ha-9	4.04-4.08	C-9	61.01	GlcNAc			
Hb-9				proton	δ_H (ppm)	carbon	δ_C (ppm)
		C-10	174.5	H-1	4.56	C-1	101.1
Ha-11	4.43	C-11	65.72	H-2	3.75	C-2	54.93
Hb-11	4.22			H-3	3.62	C-3	74.6
		C-12	174.6	H-4	3.62	C-4	81.45
H-13	2.09	C-13	20.51	H-5	3.77	C-5	72.59
		C-14	-	Ha-6	3.99	C-6	60.61
H-15	-	C-15	-	Hb-6	3.84		
		C-16	174.6			C-7	174.5
H-17	2.14	C-17	20.4	H-8	2.11	C-8	22.47

Chapter 3

5-(6-(D-biotinamido)hexanamido)pentyl (7,9-di-O-acetyl-3,5-dideoxy-5-(2-hydroxy)acetamido-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid)-(2 \rightarrow 6)- β -D-galactopyranosyl-(1 \rightarrow 4)-2-acetamido-2-deoxy- β -D-glucopyranoside (4d)

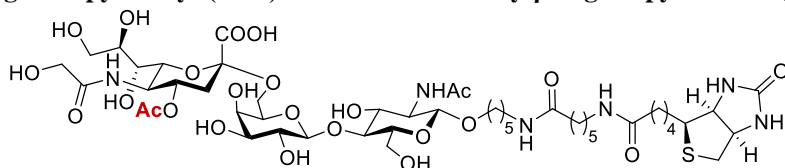


Compound **4d** was prepared from **4a** following the **General procedure II**

using MHV-S-HE as the de-O-acetylating enzyme. See table below for NMR signal assignment. HRMS (neg.) calculated for $C_{50}H_{81}N_6O_{24}S^-$ 1197.4972, found 1197.4967.

Neu7,9Ac ₂ 5Gc- α 2,6-Gal- β 1,4-GlcNAc (2d)							
Neu5Gc				Gal			
proton	δ_H (ppm)	carbon	δ_C (ppm)	proton	δ_H (ppm)	carbon	δ_C (ppm)
		C-1	173.1	H-1	4.48	C-1	103.4
		C-2	100.9	H-2	3.56	C-2	71.00
Heq-3	2.78	C-3	40.42	H-3	3.69	C-3	72.6
Hax-3	1.75			H-4	3.96	C-4	68.51
H-4	3.74	C-4	68.18	H-5	3.83	C-5	73.68
H-5	3.86	C-5	50.97	Ha-6	3.98	C-6	63.53
H-6	4.06	C-6	71.75	Hb-6	3.65		
H-7	5.18	C-7	69.33				
H-8	4.26	C-8	67.5				
Ha-9	4.16	C-9	64.22	GlcNAc			
Hb-9	4.11			proton	δ_H (ppm)	carbon	δ_C (ppm)
		C-10	175.5	H-1	4.56	C-1	101.2
Ha-11	4.06	C-11	61.07	H-2	3.75	C-2	55.16
Hb-11	4.01			H-3	3.62	C-3	74.83
		C-12	-	H-4	3.69	C-4	80.07
H-13	-	C-13	-	H-5	3.75	C-5	72.51
		C-14	173.2	Ha-6	4.00	C-6	60.46
H-15	2.15	C-15	20.28	Hb-6	3.86		
		C-16	175.3			C-7	174.7
H-17	2.14	C-17	20.28	H-8	2.06	C-8	22.25

5-(6-(D-biotinamido)hexanamido)pentyl (4-*O*-acetyl-3,5-dideoxy-5-(2-hydroxy)acetamido-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid)-(2 \rightarrow 6)- β -D-galactopyranosyl-(1 \rightarrow 4)-2-acetamido-2-deoxy- β -D-glucopyranoside (4e)



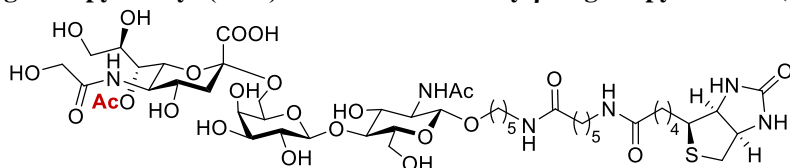
Compound **4e**
was prepared
from **4c**
following
**General
procedure II**

using BCoV-HE as the de-*O*-acetylating enzyme. See table below for NMR signal assignment. HRMS (neg.) calculated for C₄₈H₇₉N₆O₂₃S⁻ 1155.4866, found 1155.4873.

Neu4Ac5Gc- α 2,6-Gal- β 1,4-GlcNAc (2e)							
Neu5Gc				Gal			
proton	δ_H (ppm)	carbon	δ_C (ppm)	proton	δ_H (ppm)	carbon	δ_C (ppm)
		C-1	173.4	H-1	4.44	C-1	103.7
		C-2	100.01	H-2	3.54	C-2	70.74
Heq-3	2.70	C-3	37.2	H-3	3.67	C-3	72.51
Hax-3	1.85			H-4	3.92	C-4	68.32
H-4	5.00	C-4	70.69	H-5	3.83	C-5	73.74
H-5	4.12	C-5	49.21	Ha-6	4.03	C-6	63.41
H-6	3.92	C-6	71.78	Hb-6	3.53		
H-7	3.58	C-7	68.25				
H-8	3.99	C-8	71.95				
Ha-9	3.88	C-9	62.62	GlcNAc			
Hb-9	3.65			proton	δ_H (ppm)	carbon	δ_C (ppm)
		C-10	175.7	H-1	4.55	C-1	101.0
Ha-11	4.02-4.08	C-11	60.88	H-2	3.73	C-2	54.78
Hb-11				H-3	3.61	C-3	74.45
		C-12	173.8	H-4	3.61	C-4	81.39
H-13	2.08	C-13	20.47	H-5	3.75	C-5	72.50
		C-14	-	Ha-6	3.97	C-6	60.48
H-15	-	C-15	-	Hb-6	3.82		
		C-16	-			C-7	174.3
H-17	-	C-17	-	H-8	2.09	C-8	22.39

Chapter 3

5-(6-(D-biotinamido)hexanamido)pentyl (7-O-acetyl-3,5-dideoxy-5-(2-hydroxy)acetamido-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid)-(2 \rightarrow 6)- β -D-galactopyranosyl-(1 \rightarrow 4)-2-acetamido-2-deoxy- β -D-glucopyranoside (4f)

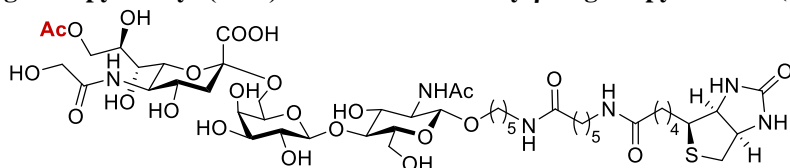


Compound **4f** was prepared from **4a** following the **General procedure II**

using MHV-S- and BCoV-HEs as the de-*O*-acetylating enzymes. See table below for NMR signal assignment. HRMS (neg.) calculated for C₄₈H₇₉N₆O₂₃S⁻ 1155.4866, found 1155.4860.

Neu7Ac5Gc- α 2,6-Gal- β 1,4-GlcNAc (2f)							
Neu5Gc				Gal			
proton	δ_H (ppm)	carbon	δ_C (ppm)	proton	δ_H (ppm)	carbon	δ_C (ppm)
		C-1	172.9	H-1	4.56	C-1	101.1
		C-2	100.6	H-2	3.56	C-2	70.95
Heq-3	2.78	C-3	40.39	H-3	3.68	C-3	72.53
Hax-3	1.75			H-4	3.96	C-4	68.44
H-4	3.73	C-4	68.19	H-5	3.84	C-5	73.68
H-5	3.85	C-5	50.99	Ha-6	3.98	C-6	63.46
H-6	4.05	C-6	71.82	Hb-6	3.65		
H-7	5.09	C-7	69.69				
H-8	4.02	C-8	69.95				
Ha-9	3.68	C-9	61.7	GlcNAc			
Hb-9	3.49			proton	δ_H (ppm)	carbon	δ_C (ppm)
		C-10	175.3	H-1	4.48	C-1	103.30
Ha-11	4.02-4.08	C-11	61.00	H-2	3.75	C-2	55.11
Hb-11				H-3	3.62	C-3	74.75
		C-12	-	H-4	3.69	C-4	79.97
H-13	-	C-13	-	H-5	3.75	C-5	72.45
		C-14	173.0	Ha-6	4.00	C-6	60.45
H-15	2.15	C-15	20.58	Hb-6	3.85		
		C-16	-			C-7	174.3
H-17	-	C-17	-	H-8	2.06	C-8	22.31

5-(6-(D-biotinamido)hexanamido)pentyl (9-O-acetyl-3,5-dideoxy-5-(2-hydroxy)acetamido-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid)-(2 \rightarrow 6)- β -D-galactopyranosyl-(1 \rightarrow 4)-2-acetamido-2-deoxy- β -D-glucopyranoside (4g)



Compound **4g** was prepared from **4f** following the **General procedure III**.

See table below for NMR signal assignment. HRMS (neg.) calculated for $C_{48}H_{79}N_6O_{23}S^-$ 1155.4866, found 1155.4871.

Neu9Ac5Gc- α 2,6-Gal- β 1,4-GlcNAc (2g)							
Neu5Gc				Gal			
proton	δ_H (ppm)	carbon	δ_C (ppm)	proton	δ_H (ppm)	carbon	δ_C (ppm)
		C-1	173.5	H-1	4.46	C-1	103.4
		C-2	100.1	H-2	3.55	C-2	70.79
Heq-3	2.69	C-3	40.24	H-3	3.68	C-3	72.45
Hax-3	1.74			H-4	3.94	C-4	68.43
H-4	3.62	C-4	68.37	H-5	3.84	C-5	73.77
H-5	3.92	C-5	51.58	Ha-6	4.00	C-6	63.52
H-6	3.87	C-6	72.17	Hb-6	3.56		
H-7	3.77	C-7	68.04				
H-8	4.12	C-8	69.2				
Ha-9	4.11-4.14	C-9	61.03	GlcNAc			
Hb-9				proton	δ_H (ppm)	carbon	δ_C (ppm)
		C-10	174.5	H-1	4.56	C-1	100.8
Ha-11	4.42	C-11	65.66	H-2	3.74	C-2	54.97
Hb-11	4.21			H-3	3.62	C-3	74.59
		C-12	-	H-4	3.65	C-4	80.7
H-13	-	C-13	-	H-5	3.75	C-5	72.42
		C-14	-	Ha-6	3.99	C-6	60.41
H-15	-	C-15	-	Hb-6	3.84		
		C-16	176.0			C-7	174.7
H-17	2.14	C-17	20.39	H-8	2.07	C-8	22.43

Chapter 3

CHAPTER 4 | Synthetic *O*-Acetylated Sialosides and their Acetamido-deoxy Analogs as Probes for Coronaviral Hemagglutinin-Esterase Recognition

Zeshi Li^{1,6}, Luca Unione^{1,6}, Lin Liu², Yifei Lang³, Robert P. de Vries¹, Raoul J. de Groot³, Geert-Jan Boons^{1,2,4,5*}

Affiliations

¹Department of Chemical Biology and Drug Discovery, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, 3584 CG Utrecht, The Netherlands. E-mail: g.j.p.h.boons@uu.nl

²Complex Carbohydrate Research Center, University of Georgia, 315 Riverbend Road, Athens, GA 30602, USA. E-mail: gjboons@ccrc.uga.edu

³Virology Division, Department of Biomolecular Health Sciences, Faculty of Veterinary Medicine, Utrecht University, 3584 CL Utrecht, the Netherlands.

⁴Bijvoet Center for Biomolecular Research, Utrecht University, Utrecht, The Netherlands.

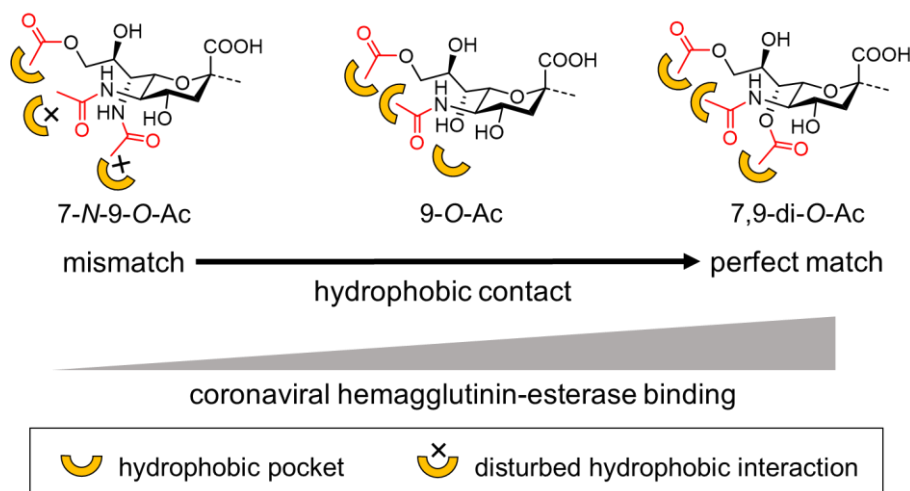
⁵Chemistry Department, University of Georgia, Athens, GA 30602, USA.

⁶These authors contributed equally: Zeshi Li, Luca Unione.

Chapter 4

Abstract

O-acetylation is a common modification of sialic acids that can occur at carbon 4-, 7-, 8-, and/or 9. Acetylated sialosides are employed as receptors by several betacoronaviruses and toroviruses, and influenza C and D viruses. The molecular basis by which these viruses recognize specific *O*-acetylated sialosides is poorly understood, and it is unknown how viruses have evolved to recognize specific *O*-acetylated sialosides expressed by their host. Here, we describe a chemoenzymatic approach that can readily provide sialoglycan analogs in which acetyl esters at C4 and/or C7 are replaced by stabilizing acetamide moieties. The analogs and their natural counterparts were used to examine ligand requirements of the lectin domain of coronaviral hemagglutinin-esterases (HEs). It revealed that HEs from viruses targeting different host species exhibit different requirements for *O*-acetylation. It also showed that ester-to-amide perturbation results in decreased or loss of binding. STD-NMR and molecular modeling of the complexes of the HE of BCoV with the acetamido analogs and natural counterparts revealed that binding is governed by the complementarity between acetyl moieties of the sialosides and hydrophobic patches of the lectin. The precise spatial arrangement of these elements is important, and an ester-to-amide perturbation results in substantial loss of binding. Molecular Dynamics simulations with HEs from coronaviruses infecting other species indicate that these viruses have adapted their HE specificities by the incorporation of hydrophobic or hydrophilic elements to modulates acetyl ester recognition.



Introduction

Sialic acids are a class of keto-acid sugars having a nine-carbon backbone that exhibit extraordinary structural diversity.¹ They are often found at termini of glycans of animal tissues with *N*-acetylneuraminic acid (Neu5Ac) and *N*-glycolylneuraminic acid (Neu5Gc) being the most common forms. Neu5Ac is usually attached to galactose (Gal) through α 2,3- or α 2,6-linkage, *N*-acetylgalactosamine (GalNAc) in α 2,6-fashion or to another Neu5Ac via an α 2,8-linkage. Further structural diversity can arise from modifications of hydroxyls. *O*-acetylation is the most common form, which can occur at the 4-, 7-, 8-, and/or 9-positions giving rise to up to nine partially *O*-acetylated variants (Fig. 1). The expression of these glycotopes is regulated in a tissue- and cell-specific manner via a hitherto unknown mechanism.² A growing body of literature has linked *O*-acetylated sialic acids (*O*-Ac-Sias) to a wide range of biological processes such as B- and T-cell immunity,³⁻⁴ cancer drug resistance,⁵ and bacterial foraging.⁶⁻⁷ These biomolecules are also employed as receptors by several betacoronaviruses and toroviruses (subfamily *Orthocoronavirinae* and family *Tobaniviridae*, respectively), and influenza C and D viruses.⁸ Different host species express different repertoires of *O*-acetylated sialosides which appears to hamper cross species transmission.

Recently, we reported a chemo-enzymatic methodology that can provide sialosides having different patterns of *O*-acetylation.⁹ It is based on the chemical synthesis of α 2,3-, 2,6- and 2,8-linked sialosides carrying acetyl esters at *C*-4, *C*-7 and *C*-9, which could readily be diversified by treatment with the hemagglutinin-esterases (HEs) from bovine coronavirus (BCoV) and hepatitis virus strain S (MHV-S) which have 9-*O*- and 4-*O*-acetyl esterase activity, respectively, and in combination with controlled acetyl ester migration from *C*-7 to *C*-9, could provide almost any sialate-acetylation pattern. The resulting collection of *O*-acetylated sialoglycans was printed as a microarray, which was employed to examine receptor specificities of a group of viruses that engage with *O*-acetylated sialosides of host cells for infection.

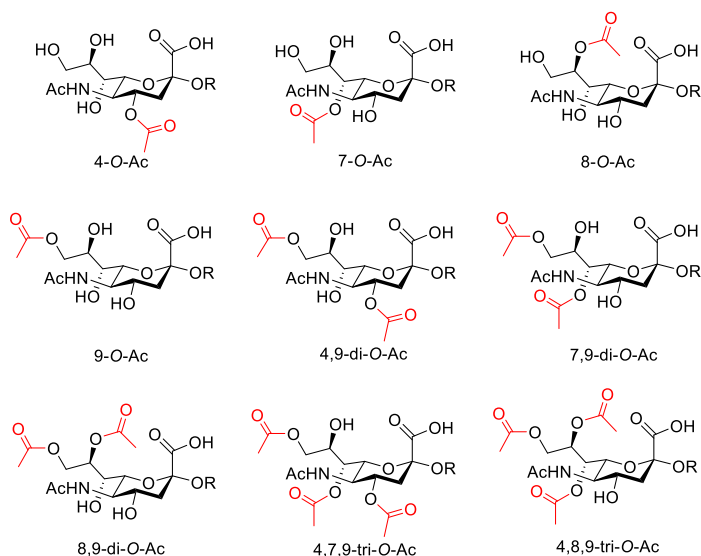


Figure 1. Naturally occurring *O*-acetylated variants of sialic acid.

Chapter 4

For various bioactive molecules, such as acetylcholine¹⁰ and procaine,¹¹ it has been demonstrated that an acetyl ester can be replaced by an acetamido moiety without loss of recognition and resulting in improved stability. We and others have shown that C9- and C4-acetamido sialosyl analogs are resistant to esterase degradation, overcoming the chemical instability while retaining binding to the virolectins derived from nidoviruses.¹²⁻¹⁴ Therefore, we were compelled to develop a chemo-enzymatic methodology that can provide easy access to the 7- and 4-deoxy-*N*-acetyl sialosyl analogues **1** - **3** that are linked in a α 2,3- or α 2,6-fashion to *N*-acetyl lactosamine (Fig. 2). The resulting compounds and corresponding *O*-acetyl-modified sialosides (**4** - **8**) were printed as a glycan microarray, which was probed for recognition by the lectin domain of HEs and spike proteins from bovine, canine, equine, rabbit and murine coronaviruses and the distantly related bovine torovirus. The array binding studies showed that an 7-*O*-acetyl ester enhances binding for a number of HEs that have an obligatory requirement for 9-*O*-acetylation. Furthermore, it revealed that 7- and 4-*N*-acetamides are not well tolerated by the proteins. Although the acetamide-containing analogs were not well recognized, they proved to be valuable probes for protein recognition studies. Proton saturation-transfer difference (¹H-STD)-NMR experiments and molecular modeling of the complex of the HE of BCoV with **1a** and its natural counterparts **7a** revealed that the HE receptor preference is critically governed by the complementarity between acetyl moieties of the sialosides and hydrophobic patches of the lectin with stringent requirements for precise spatial arrangement of these elements. Molecular Dynamics (MD) simulations with HEs from coronaviruses infecting other species indicate that these viruses have adapted their HE specificities for specific forms of *O*-acetylated sialoglycans by the incorporation of hydrophobic or hydrophilic elements to modulate acetyl ester recognition.

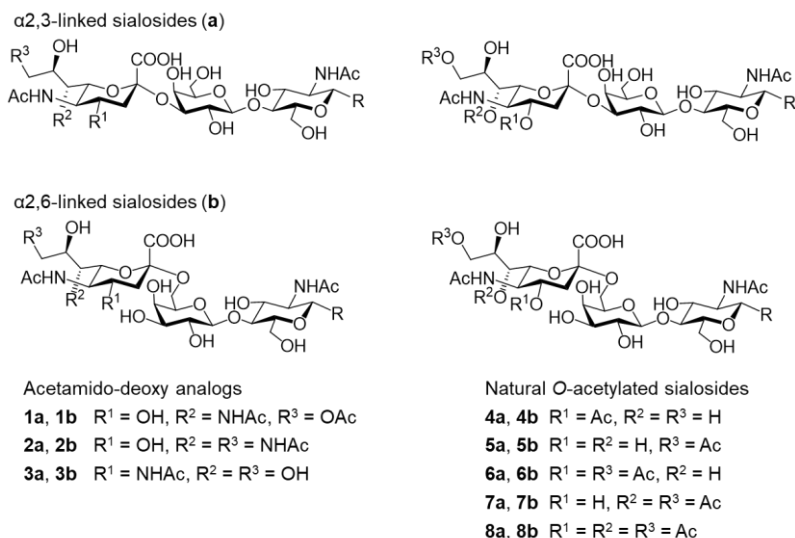
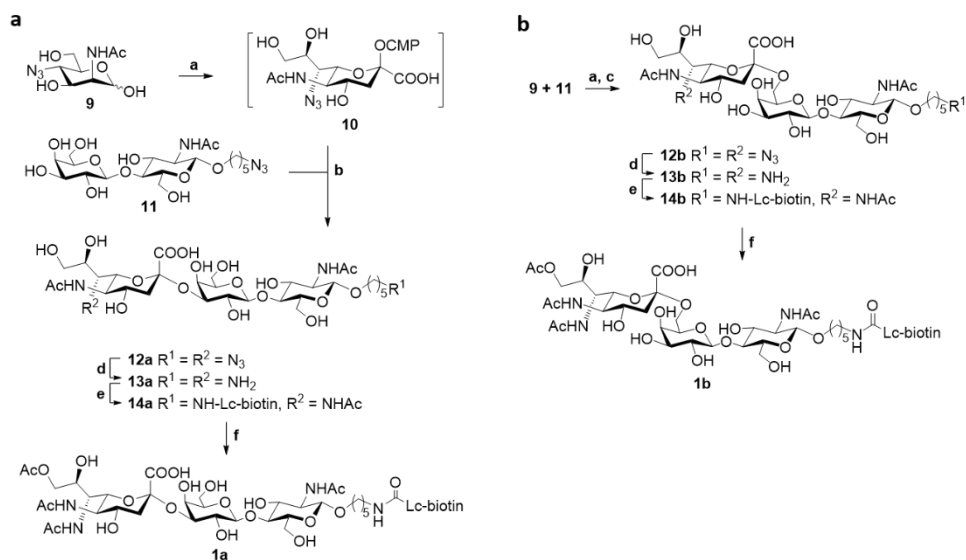


Figure 2. Synthetic *O*-acetylated sialosides and their acetamido-deoxy analogs used in this study. The substituent at the reducing end is not shown for clarity.

Results and discussion

Synthesis of sialosides modified by 7- and 4-acetamido-deoxy moieties. We devised a chemoenzymatic strategy for the preparation of **1a**, **1b**, **2a** and **2b** based on the enzymatic conversion of 4-azido-*N*-acetyl-mannosamine (**9**) into the corresponding CMP-sialic acid derivative, which could be transferred by an appropriate sialyltransferase to the C-3' or C-6' position of spacer containing *N*-acetyl lactosamine (**11**) to give **12b** and **12a**. Subsequent regioselective chemical manipulations provided the target compounds (Scheme 1). Thus, condensation of 4-azido ManNAc **1** with pyruvate in the presence of the aldolase from *E. coli* gave 7-azido-sialic acid which was immediately converted into CMP-Neu5Ac7N₃ (**10**) by reaction with cytidine monophosphate (CMP) catalyzed by CMP-sialic acid synthetase from *Neisseria meningitidis*.¹⁵ The *in-situ* formed donor could readily be employed by the human α 2,6-sialyltransferase, ST6Gal1, and sialylation of spacer-containing LacNAc **11** resulted in the efficient formation of **12b**. Human ST3Gal4, which biosynthesizes the α 2,3-linked isomer, could not employ CMP-Neu5Ac7N₃ as a donor and even after an prolonged incubation time, no product formation was detected. These observations indicate that the human sialyltransferases differ in substrate promiscuity and exhibit fine specificities for differentially modified CMP-sialic acid donors. Fortunately, the bacterial sialyltransferase PMST1¹⁶ could readily employ CMP-Neu5Ac7N₃ as a donor and sialylation of **11** gave 7-azido azido-containing α 2,3-linked sialyl LacNAc **12a**.

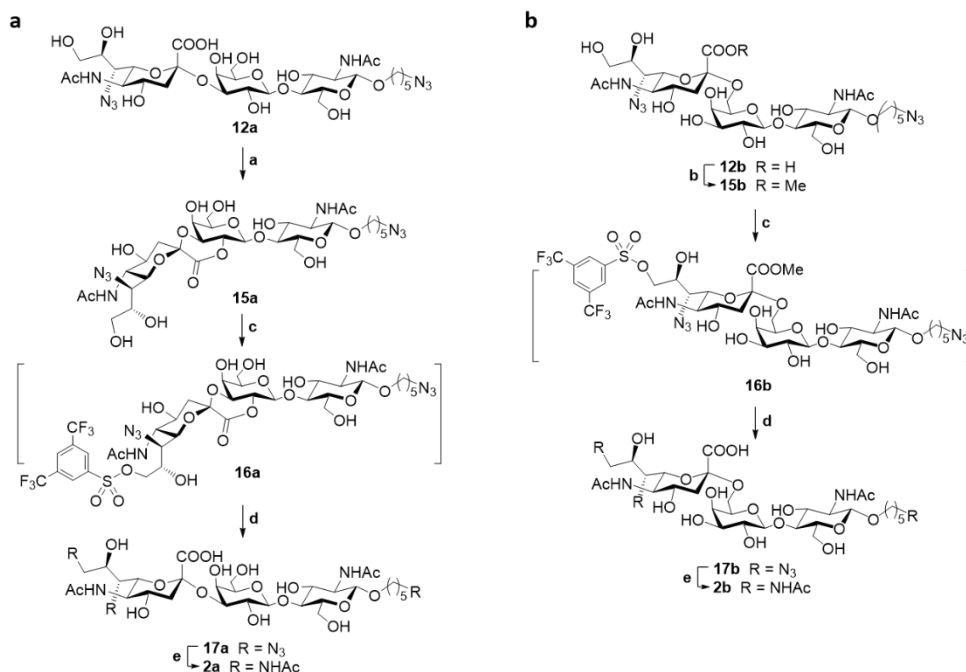


Scheme 1. Chemoenzymatic synthesis of 7-*N*,9-*O*-acetylated α 2,3 (Panel **a**) and 2,6-linked (Panel **b**) sialosides. Reaction conditions: a, sodium pyruvate, cytidine 5'-triphosphate (CTP, sodium salt), sialic acid aldolase, CMP-sialic acid synthetase from *N. meningitidis*, 20 mM MgCl₂, 0.2 M Tris·HCl buffer, 37°C. b, sialyltransferase PMST1-M144D; isolated yield 77% for **12a**. c, human ST6Gal1; isolated yield, 67% for **12b**. d, H₂, Pd(OH)₂, H₂O, 21°C. e, NHS-Lc-biotin (see Supporting Information for structure), triethylamine, DMF/H₂O (4:1, v/v), 0°C, then Ac₂O; isolated yields, 87% for **14a** and 74% for **14b**, over two steps. f, trimethylorthoacetate, toluenesulfonic acid monohydrate, DMSO; isolated yields, 72% for **1a**, 67% for **1b**.

Chapter 4

The azido moieties of **12a** and **12b** were reduced by hydrogenation using palladium hydroxide to give bis-amine intermediates **13a** and **13b**. The amine on the anomeric linker of the latter compounds was selectively modified with Lc-biotin using the corresponding succinimide ester, which was followed by acetylation of the C7 amine of Neu5Ac using acetic anhydride in the presence of triethylamine, to provide **14a** and **14b** in 87% and 74% over all yields, respectively, without intermediate purification steps.

Compounds **14a** and **14b** were convenient intermediates for the preparation of the target compounds **1a** and **1b**, which have a sialoside modified by a C-7 acetamido moiety and a C-9 acetyl ester. We reason that the C9-hydroxyl of the glycerol chain of Neu5Ac would be most reactive, and as expected reaction of **14a** and **14b** with trimethylorthoacetate¹⁷ in the presence of toluenesulfonic acid in DMSO resulted in the modification of C8-C9 diol of the glycerol side chain to give an intermediate orthoesters that upon hydrolysis provided **1a** and **1b**, respectively. The compounds were purified by HPLC over a HILIC column to remove residual unreacted C9 non-acetylated starting material. The regioselectivity of the reaction was confirmed by NMR, which demonstrated that the geminal hydrogens at C9 of Neu5Ac had substantially shifted downfield.

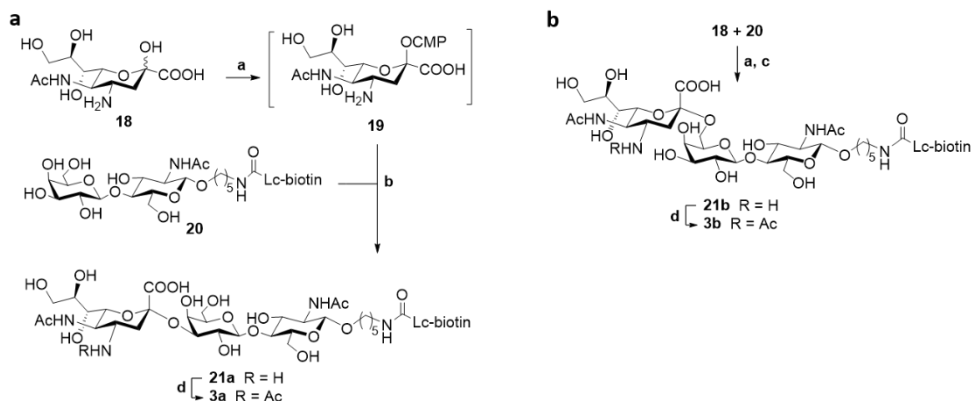


Scheme 2. Chemoenzymatic synthesis of 7,9-di-*N*-acetylated α 2,3 (Panel **a**) and 2,6-linked (Panel **b**) sialosides. Reaction conditions: a, EDC-HCl, HOBT, DMF, 0–21°C. b, EDC-HCl, HOBT, MeOH, 0–21°C. c, Bu₂SnCl₂, DIPEA, 3,5-bis(trifluoromethyl)benzenesulfonyl chloride, THF/DMF (8:1, v/v), 0–37°C. d, sodium azide, DMF, 60°C, then 1M NaOH, 21°C; isolated yields, 73% for **17a**, 60% for **17b**, over three steps. e, i) H₂, Pd(OH)₂, H₂O, 21°C; ii) Ac₂O, triethylamine, DMF/H₂O (4:1, v/v), 0°C; isolated yields, 84% for **2a**, 88% for **2b**, over two steps.

The 7-azido sialyl LacNAc **12a** and **12b** were also the starting materials for the 7,9-di-*N*-acetylated analogs **2a** and **2b** using an alternative derivatization procedure (Scheme 2).

Thus, treatment of **12a** with EDC·HCl and HOBt in DMF resulted in the formation of lactone **15a**. The methyl ester **15b** was formed by treatment of **12b** with methanol in the presence of EDC·HCl and HOBt.¹⁸ The 9-hydroxyl of Neu5Ac of **15a** and **15b** was modified as a 3,5-bis(trifluoromethyl)benzenesulfonate by treatment with the corresponding sulfonyl chloride in the presence of Bu₂SnCl₂ and DIPEA to give **16a** and **16b**, respectively.¹⁹ Probably, the reaction proceeds through a dibutyl stannylene acetal which preferably forms at the C8-C9 diol of the glycerol chain of Neu5Ac, and by limiting the stoichiometry of sulfonyl chloride, **16a** and **16b** are selectively formed. Displacement of the sulfonate **16a** and **16b** by sodium azide followed by hydrolysis of the lactone and methyl esters afforded compounds **17a** and **17b**. The azido moiety of the latter compounds was reduced to an amine by hydrogenation, which was followed by *N*-acetylation using acetic anhydride to give **2a** and **2b** in high overall yield. The successful preparation of the modified sialosides demonstrate that orthoester-mediated *O*-acylation and organotin-catalyzed sulfonylation provide attractive avenues for the regioselective derivatization of a vicinal diol of a highly complex poly-hydroxylated compound.

Next, attention was focused on the preparation of compounds **3a** and **3b**, which have an unnatural acetamido moiety at C4 of sialic acid (Scheme 3). Thus, the 4-amino derivative of Neu5Ac (**18**, see Scheme S1 for the synthesis) was converted into the corresponding of CMP-Neu5Ac **19** by treatment with CTP in the presence of the CMP-sialic acid synthetase. The *in-situ* generated donor could be transferred to biotin modified LacNAc derivative **20** by the bacterial sialyltransferase PMST1 to provide, after acetylation using acetic anhydride, α2,3-sialosyl analog **3a**. The corresponding 2,6-sialoside **3b** was prepared by a similar route employing a mutant of PMST1 possessing α2,6-sialylating activity.²⁰ Interestingly, the two human enzymes could not transfer CMP-Neu5Ac modified by a C4 amine. This observation indicates that such modifications are not tolerated by human sialyltransferases, which may be associated to an absence of 4-*O*-acetylated sialosides in humans.² Several studies have shown that derivatization of sialosides at C4 can increase binding of immunoregulatory lectins such as Siglecs²¹ and viral glycoproteins including neuraminidases of influenza A viruses.²² The approaches described here provide convenient synthetic routes to such compounds.



Scheme 3. Chemoenzymatic synthesis of 4-*N*-acetylated sialosides α2,3 (Panel **a**) and 2,6-linked (Panel **b**) sialosides. Reaction conditions: a, CTP sodium salt, sialic acid aldolase, CMP-sialic acid synthetase from *N. meningitidis*, 20 mM MgCl₂, 0.2 M Tris·HCl buffer, 37°C. b, PMST1-M144D. c, PMST1-P34H/M144L. d, Ac₂O, triethylamine, DMF/H₂O (4:1, v/v), 0°C; isolated yields, 71% for **3a**, 52% for **3b**, over two steps.

Chapter 4

Microarray analysis of HE binding. Having successfully prepared a range of sialoglycan analogs, attention was focused on the construction of a microarray to probe interactions with hemagglutinin-esterases (HE) of Embecoviruses from different host species. Embecoviruses, which is a subgenus of the coronaviridae, have a broad host range, and can infect human, bovine, canine, equine, murine and rabbits. In addition to their spike (S) protein, embecoviruses encode a hemagglutinin-esterase (HE)²³ that recognizes *O*-acetylated sialic acid *via* a lectin domain, and hydrolyze sialate-*O*-acetyl esters *via* an esterase domain. Fine-tuned binding preferences of HEs for different forms of *O*-acetylated sialic acids have been observed for embecoviruses targeting distinct host species,⁹ which likely is a result of adaptation to the *O*-acetyl sialoglycome of the host.

Compounds **1-3a,b** and a panel of previously prepared *O*-acetylated sialosides **4-8a,b** (Fig. 2) were employed to create a microarray to examine the effect of *O*-to-*N* substitution on protein recognition. The sialosides were printed on streptavidin coated glass slides, and the resulting microarray was exposed to Fc-fusion proteins of inactive esterase (HEs) domains from bovine, equine, canine and rabbit betacoronaviruses (BCoV, ECoV, CRCoV and RbCoV, respectively) bovine toroviruses (BToV),^{9, 24-26} and murine hepatitis virus strain S (MHV-S).¹² Detection of binding was accomplished using an anti-Fc antibody tagged with AlexaFluor-647.

BCoV, RbCoV, ECoV, CRCoV and BToV HEs recognized 2,3-linked sialoside **7a** which has acetyl esters at C7 and C9. Replacement of the acetyl ester at C7 by an acetamide to give compound **1a** abolished binding (Fig. 3a-e). Similar observations were made for 2,6-linked sialosides (**7b** vs. **1b**). MHV-S HE has an obligatory requirement for 4-*O*-acetylation.¹² The ester-to-amide substitution at C4 (compounds **3a** and **3b**) was tolerated by MHV-S HE but resulted in substantially weaker binding compared to the parent compounds **4a** and **4b** (Fig. 3f). These results indicate that unlike *N*-acetylation at C9,¹³ acetamides at C-7 and C-4 are not appropriate isosteres of acetyl esters for HE recognition.

The microarray data showed that the different HEs have different requirements for 7-*O*-acetylation. In the case of BCoV, RbCoV ECoV and CRCoV HEs, 7-*O*-acetylation (as in **7a** and **b**) increased binding while BToV HE was unaffected by this modification. It also revealed multifaceted roles of 4-*O*-acetylation in the regulation of HE recognition for the closely related embecoviruses BCoV, CRCoV, RbCoV and ECoV. The HEs of the first two viruses tolerated this modification, and exhibited a similar pattern of receptor recognition. On the other hand, the presence of a 4-*O*-acetyl ester substantially increased binding for RbCoV HE and α 2,3-linked 4,7,9-tri-*O*-acetylated Neu5Ac derivative **8a** was the strongest binder (Fig. 3b). In contrast, the same modification decreased binding of ECoV HE (Fig. 3c, **6a** vs. **5a**). However, in this case the unfavorable effect of 4-*O*-acetylation could be alleviated by the presence of a 7-*O*-acetyl ester (**8a** vs. **6a**). The glycan array was also probed for spike proteins from BCoV, RbCoV, ECoV and CRCoV (Fig. S1g-n). While the spike of BCoV bound to 9-mono and 7,9-di-*O*-acetylated sialosides equally well, it exhibited only weak binding when C7 is modified by an acetamide moiety. For the spike proteins from RbCoV, ECoV and CRCoV, the presence of an acetyl ester decreases the binding, while compounds having a 7-acetamide substitution did not produce any responsiveness. Probing the microarray at a lower protein concentration (1 μ g/mL, Fig. S1a-f) showed in each case a dose response effect indicating that saturation of binding had not been reached.

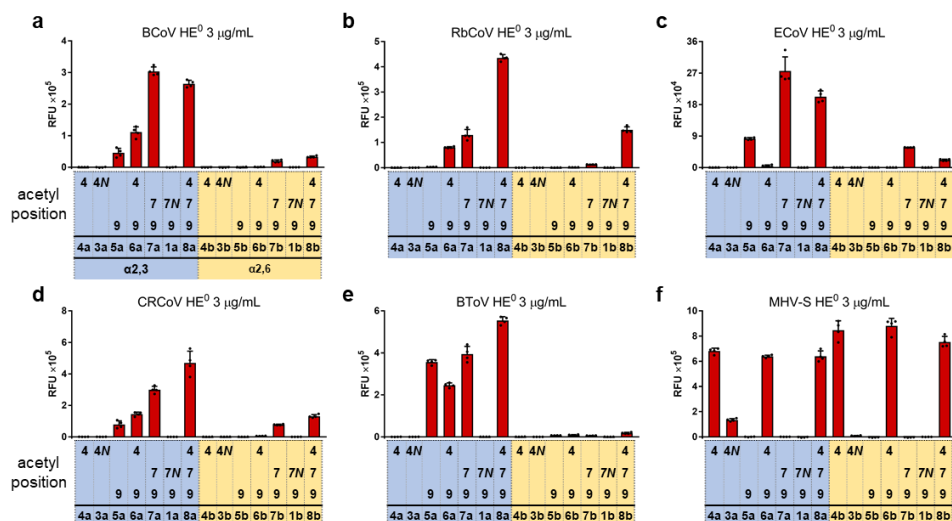


Figure 3. Glycan microarray analysis of BCoV (a), RbCoV (b), ECoV (c), CRCoV (d), BToV (e) and MHV-S (f) HEs binding. HE⁰ denotes the esterase-inactive mutant. Sialosides having an α2,3-linkage are shaded in blue and α2,6 in yellow. 4N and 7N in the tables below the bar graphs indicate the corresponding acetamido-deoxy analogs. Relative fluorescence unit (RFU) values are with background subtracted. Heights of columns show the background-subtracted average RFUs of four replicate spots. Error bars represent standard deviation RFUs. Data points for *O*-acetylated sialosides (4-8a and b) were from our previous publication,⁹ which were obtained on the same slides with the acetamido-deoxy analogs (1 and 3a and b). See Fig. S1g-n in the SI for spike binding. All binding experiments were performed at 4 °C to avoid potential *O*-acetyl migration.

Structural basis of BCoV HE-receptor interactions. HEs of bovine, canine, equine and rabbit coronaviruses require an acetyl ester at C9 for binding but exhibit different requirements for additional acetyl esters. Understanding of *O*-Ac-Sia engagement by viral receptor binding proteins at a structural level is largely limited to 9-mono-acetylated sialic acid. An X-ray crystal structure of BCoV HE in complex with sialic acid carrying acetyl ester at C4 and C-9 (pdb code 3CL5) has provided some insight in the recognition of this class of glycan.²⁴ A deep pocket (P₁) defined by Tyr¹⁸⁴, Leu²⁶⁶ and Leu²⁶⁷ accommodate the 9-*O*-acetyl ester and the glycerol side chain. The 5-*N*-acetyl moiety is placed in a wide cavity (P₂) defined by Phe²⁴⁵, Leu²¹² and Phe²¹¹, which is sufficiently large to host an additional *O*-acetyl ester at C4. The complex is stabilized by hydrogen bonds between the C5-amide and Leu²¹², the C1-carboxylate and Asn²¹⁴, and the C8 oxygen and Ser²¹³.

The structural data does not provide a rationale for the substantial higher affinity for sialosides having an acetyl ester at C7. Furthermore, there are no reported structures of the HEs from canine, equine and rabbit coronaviruses, and it is not understood at a structural level how these closely related proteins have adapted to different forms of *O*-acetylated sialosides. Here, we employed NMR experiments, homology modeling and molecular dynamics (MD) simulations to understand, at a molecular level, receptor specificities of HEs for various coronaviruses.

Examination of the receptor binding site of the lectin domain of *apo* BCoV HE (pdb code 3CL4) indicates the presence of an additional hydrophobic pocket (P₃), defined by

Chapter 4

Phe²⁰⁷, Thr¹¹⁴ and Leu¹⁶¹, which may accommodate a C-7 acetyl ester (Fig. 4a) and may be responsible for the higher affinity of BCoV HE for 7,9-di-*O*-acetylated sialosides. To validate this hypothesis, Saturation Transferance Difference (¹H-STD NMR) and MD simulations were performed for the complex of BCoV HE with 7,9-di-*O*-acetylated sialoside **7a** and **7b**. Furthermore, the ester-to-amide perturbation strategy^{27,28} was employed to demonstrate a stringent requirement of proper positioning of hydrophobic elements for HE:sialic acid complex formation.

The interaction of BCoV HE with the α 2,3 linked 7,9-di-*O*-acetylated sialoside (**7a**), the α 2,6 linked 7,9-di-*O*-acetylated sialoside (**7b**) and the 7-*N*,9-*O*-acetylated analog (**1a**) was studied by ¹H-STD-NMR spectroscopy to provide atomic details of the interaction between the acetylated sialosides and the HE. The ¹H-STD-NMR experiments were performed using a BCoV HE hydrolase-inactive form⁹ at 30 μ M concentration in 50 mM PBS, in D₂O (pD 7.2) using lectin/ligand ratios of 1:50. The temperature was set to 298 K. STD experiments were performed at 600 MHz Bruker spectrometer, using standard Bruker

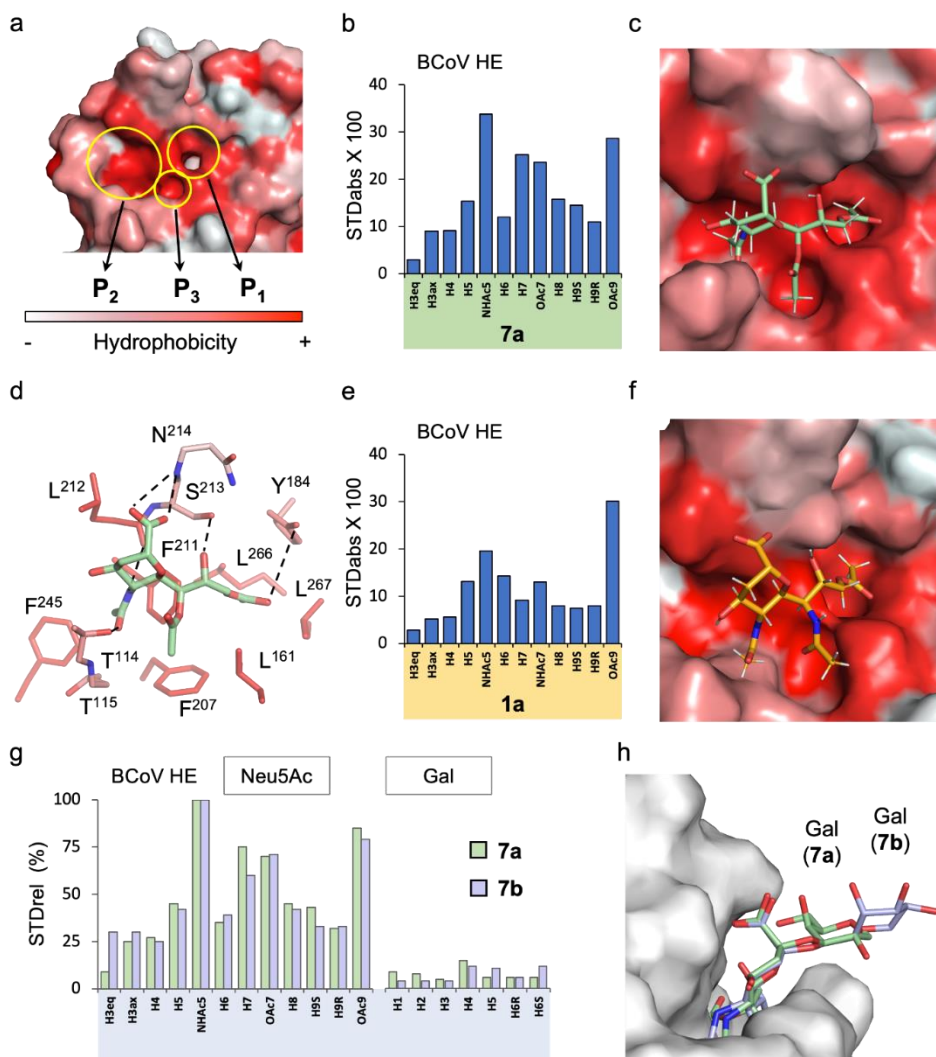


Figure 4. Structural basis of BCoV HE:receptor interaction. **a)** Surface representation of the BCoV HE apo form (pdb code 3CL4).²⁴ Protein surface is colored according to the amino acids hydrophobicity. The three hydrophobic pockets are indicated with yellow circles. **b)** ¹H-STD NMR profile of compound **7a** in complex with BCoV HE. Heights of columns represent absolute STD (STDabs) values. **c)** Surface and stick representation of the BCoV HE lectin site in complex with compound **7a** as modeled by MD simulations. The Gal and GlcNAc residues are not shown for clarity. **d)** Atomic details of BCoV HE:**7a** interactions as determined by ¹H-STD NMR and MD simulations. **e)** ¹H-STD NMR profile of the acetamido-deoxy analog compound **1a** in complex with BCoV HE. **f)** Surface and stick representation of the BCoV HE lectin site in complex with compound **1a** as determined by ¹H-STD NMR. Note the different binding pose with respect to the natural compound **7a**. The main differences reside in the packing of Neu5Ac-7-*O*-Ac and 5-NHAc groups. **g)** Comparison of the relative (%) ¹H-STD NMR profiles obtained for compounds **7a** and **7b**. The largest value given rise to by 5-NHAc of **7a** is set as 100% **h)** Superimposition of the MD derived structures for BCoV HE in complex with **7a** (green) and **7b** (blue). Note the proximity of the galactose ring to the protein for **7a** with respect to **7b**. GlcNAc residues are hidden for clarity.

pulse sequences without water suppression nor protein spin-lock filter. Protein saturation was achieved with a Gaussian-shaped pulse of 40 ms during a saturation time of 2s. The on-resonance frequency was set at aliphatic regions (0.76 ppm) and the off-resonance frequency at 100 ppm (for further details, see Fig. S1a and General Procedure section in the SI).

For compound **7a** (Fig. 4b and S2b), the strongest STD signals arose from the methyl groups of 5-*N*-, 9-*O*-, 7-*O*-acetyl moieties and from the H7 of sialic acid, indicating that ligand binding is mainly governed by the acetyl moieties of the sialoside. Moderate STD signals were observed for the H5, H6, H8, and H9_S of the sialoside, which indicate that these atoms face the protein surface. Weak STD signals were detected for H3eq, H3ax, H4 and H9_R, indicating that these atoms are solvent exposed. Even weaker STD signals were detected for the underlying galactoside (Gal) while those of the *N*-acetyl glucosamine (GlcNAc) residue were not present, indicating that the underlying glycan makes only marginal contacts with the protein.

¹H-STD NMR experiments with α 2,6-linked compound **7b** gave similar results as for the α 2,3 linked analog **7a** (Fig. 4g and S2d). The intensity of the STD signals of the sialic acid residue were essentially the same as for **7a** while those of the galactose residue were also weak. Nonetheless, differences were observed and for compound **7b** the strongest signals of Gal arose from H6S, H5 and H4 whereas for compound **7a** the strongest signals corresponded to H4, H1 and H2. These results indicate differences in the positioning of the Gal residues between the two binding complexes, where in case of **7b** the C1-C4 side of the galactosyl ring points to the protein surface whereas for **7a** it is the C4-C6 side.

Next, the molecular mechanisms of poor binding of the ester-to-amide substitution at C7 was investigated. The ¹H-STD NMR spectrum of 7-*N*-acetyl analogue **1a** with BCoV HE showed much weaker STD intensities for the methyl groups of the C7 and C5 substituents compared to **7a** (Fig. 4e and S2c). The sialic acid backbone hydrogens (H3-H9) were less affected whereas the STD signal for the C9 acetyl ester was unchanged. These observations indicates that despite the C9 acetyl ester is properly anchored in the P1 pocket, the acetamide at C7 hinders optimal interactions with the P2 and P3 pockets, which likely results in much weaker binding.

In silico studies of BCoV, RbCoV and ECoV receptor preferences. The differences in receptor preference of BCoV, RbCoV and ECoV HEs are remarkable given the high sequence similarity of their lectin domains (96%). BCoV HE binds 7,9-di- (**7a**) and 4,7,9-tri-*O*-acetylated (**8a**) Neu5Ac derivatives equally well, whereas RbCoV HE showed a strong preference for the 4,7,9-tri-*O*-acetylated form. In contrast to the RbCoV and BCoV, C4 acetylation reduced binding to ECoV HE virolectin (Fig. 3). We performed 500 ns Molecular dynamics simulations of complexes of the HEs with relevant sialosides in a 10 Å octahedral box of explicit TIP3P waters and compared the resulting structures to uncover the molecular basis of receptor preferences.

BCoV HE. Molecular dynamics (MD) simulations were performed to obtain a structural model of BCoV HE complexed with **7a** and **7b**. The crystallographic structure of BCoV HE in complex with 4,9-O-Ac-Sia is available (pdb code 3CL5). Therefore, the 7,9-di-*O*-acetylated sialosides were superimposed onto the corresponding monosaccharide in the crystal structure, and the resulting binding poses minimized and subjected to MD simulation in explicit water using the AMBER molecular simulation program.²⁹ The GLYCAM force field was employed for the glycan,³⁰ and the 14SB force field³¹ for the protein using periodic boundary conditions in explicit water.³² The resulting structures (Fig. 4c) are in excellent agreement with the STD NMR derived epitope mapping and recapitulated observations made by X-ray crystallography. The 9-*O*-acetyl methyl group and the glycerol chain docked into the P₁ pocket defined by Tyr¹⁸⁴, Leu²⁶⁶ and Leu²⁶⁷ (see Fig. S3-S5 for analyses of key hydrophobic interaction). The 5-*N*-acetyl fits into the wide P₂ cavity defined by Phe²⁴⁵, Leu²¹² and Phe²¹¹ (Fig. S6). Furthermore, the sialic acid makes polar interactions through a hydrogen bonding network with residues Thr¹¹⁵, Tyr¹⁸⁴, Leu²¹², Ser²¹³ and Asn²¹⁴ (Fig. 4d, Table S1 and Fig. S9). As anticipated, the MD simulations showed a third hydrophobic cavity (P₃), defined by Phe²⁰⁷, Thr¹¹⁴ and Leu¹⁶¹, which can accommodate the additional acetyl ester at C7 (Fig. S7 and S8, summarized in Fig. 6a). Remarkably, the latter residues are in two loop regions that contribute to the architecture of the glycan binding pocket and define the receptor specificity. Our structural model reveals a precise alignment of the hydrophobic patch in BCoV HE with the acetyl moieties of the sialoglycans. Comparison of the MD trajectory of BCoV HE in complex with **7a** or **7b** (Fig. S10) revealed that in case of α 2,3-isomer the Gal moiety contributes to protein binding through transient intermolecular hydrogen bond interactions that mainly involve either Asn²¹⁴ or Lys¹⁶³. In case of the α 2,6-isomer, the Gal ring is mainly solvent exposed due to a longer and more flexible glycosidic linkage (Fig. 4h). Thus, the different modes of protein-carbohydrate contacts, possibly arising from distinct topologies of the glycans, may account for differences in avidity of HEs for α 2,3- and 2,6-linked sialosides.³³

A similar computational approach was used to obtain a model for the complex of the 7-*N*,9-*O*-acetylated analog **1a** with BCoV HE. The MD trajectory could not be ascribed to a unique binding mode. Therefore, the NMR results were used to extract from the MD simulation a three-dimensional model that fits the experimental data. The resulting structure of the BCoV HE:**1a** complex (Fig. 4f) showed that the polar amide function at the C7 provokes an unfavorable contact with the hydrophobic patch (P₃) of the protein precluding proper interactions with the side chains of Leu¹⁶¹ and Thr¹¹⁴. To minimize these unfavorable interactions, the glycerol chain tilted and the Neu5Ac ring moved away from the protein binding pocket, while keeping the C9 methyl group into the P1 pocket. In this binding pose, the acetamide moiety at C5 pushes away Thr¹¹⁵ resulting in an open cavity at the sialic acid binding site. This mode of binding is consistent with the observed ¹H-STD NMR signal,

which showed a substantial decrease in STD signal intensities especially for H7 and the CH₃ groups of the acetyl esters at C7 and C5, while that of C9 was unperturbed. Interestingly, the

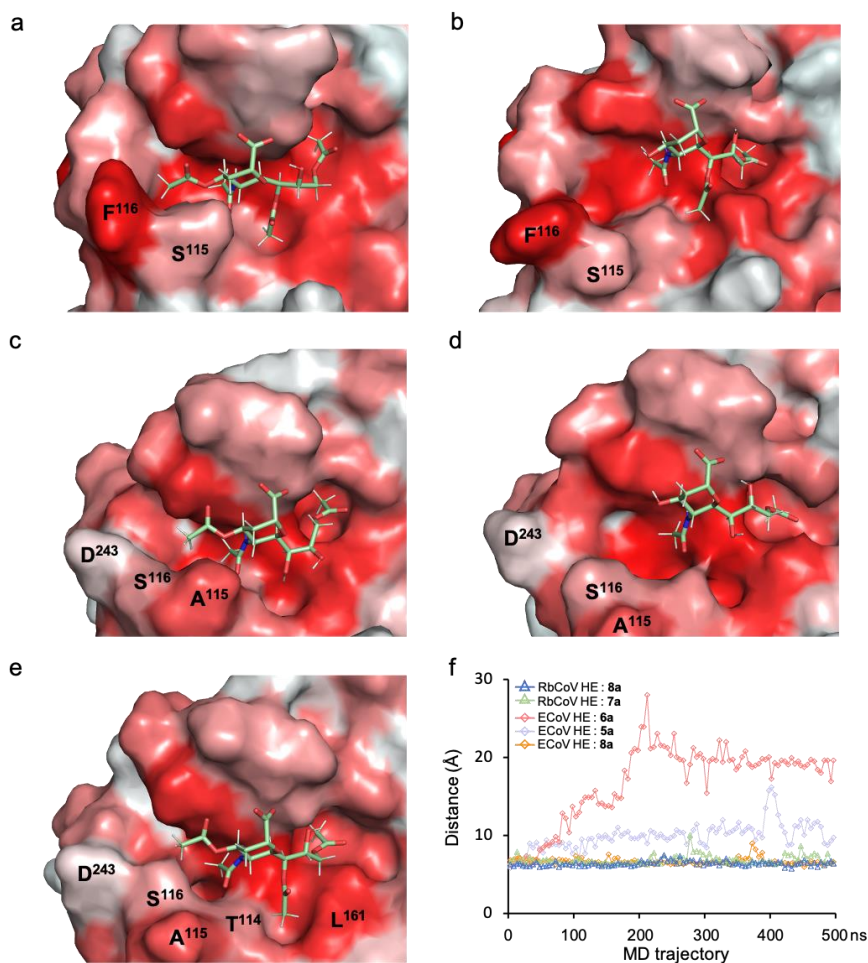


Figure 5. Structural models of RbCoV and ECoV HE:receptors interaction. **a)** Structure of the RbCoV HE lectin domain in complex with compound **8a**. Note the proximity of the F116 side-chain to the Neu5Ac-4-*O*-Ac-methyl group favorable for CH- π interactions. **b)** Structure of the RbCoV HE lectin domain in complex with compound **7a**. **c)** Structure of the ECoV HE lectin domain in complex with compound **6a**. **d)** Structure of the ECoV HE in complex with compound **5a**. **e)** Structure of the ECoV HE lectin domain in complex with **8a**. All the structures correspond to selected frame from MD simulations. Gal and GlcNAc residues are hidden for clarity. **f)** Stability of the complexes along the MD simulations. The curves represent the distance between the centers of mass of Neu5Ac residue and Phe²¹¹ along the MD trajectory. Phe²¹¹ is the key residue composing the lectin binding site, and was arbitrarily chosen among the others. The MD simulations of ECoV HE with **6a** and **5a** show a weak binding. In contrast, the short and constant distance of ECoV HE with **8a** indicates for a stable binding pose. Similarly, in RbCoV HE the graphic shows a more stable pose for the tri-*O*-acetylated compound **8a** than in case of the di-*O*-acetylated compound **7a**.

Chapter 4

increase in the STD signal of H6 agrees with the proposed binding pose that places this atom closer to the protein surface.

The data demonstrate that an acetamide at C-7 of Neu5Ac is not a proper isostere of the corresponding acetyl ester. It has a slightly increased size and is more hydrophilic and conformationally rigid.²⁸ Such perturbations greatly reduced the binding of the 7-*N*-acetylated analog **1a**, highlighting that HE:*O*-Ac-Sia interactions are governed by the precise positioning of sialate-acetyl groups in the hydrophobic pockets of the protein.

RbCoV HE. An X-ray crystal structure of RbCoV has not been reported, and therefore a homology model was generated using the crystal structure of BCoV HE as a template.³⁴⁻³⁶ The resulting Apo protein was subjected to a minimization protocol employing the AMBER suite of programs. Next, the *O*-acetylated sialosides **7a** and **8a** were placed into the ligand binding pocket by molecular alignment with the BCoV HE crystal structure and the generated complexes were subjected to MD simulations as described above.

For both ligands, the MD simulations showed stable binding poses in which the acetyl esters at positions 9, 5 and 7 established hydrophobic interactions with the P₁, P₂ and P₃ pockets, respectively. However, key differences were observed. For the tri-acetylated **8a**, the hydroxyl of the side chain of Ser¹¹⁵ engaged with the acetamide at C5 *via* a hydrogen bond with the carbonyl oxygen (Table S2 and Fig. S11), while the aromatic ring of Phe¹¹⁶ established a CH- π interaction with the acetyl ester at C4 (Fig. 5a and S12). On the other hand, the complex with **7a** (see Table S3 and Fig. S13 for hydrogen bonding interactions) placed the polar hydroxyl group at C4 toward the hydrophobic Phe¹¹⁶, and this unfavorable contact pushed the Phe¹¹⁶ and the contiguous Ser¹¹⁵ residue apart, leaving a wide, open cavity and, thus, precluding hydrogen bond interaction between the Ser¹¹⁵ and the acetamide at the C5 of the sialoside (Fig. 5b). These results demonstrate that complementarity of the three acetyl moieties at C4, C7 and C9 and the hydrophobic pockets of RbCoV HE are critical for ligand specificity (summarized in Fig. 6b and c).

ECoV HE. Finally, we explored structural elements important for ECoV HE binding. Mono-, di- and tri-*O*-acetylated sialosides **5a**, **6a** and **8a** were docked into the binding site of ECoV HE obtained by homology modeling (Fig. 5c-e, respectively). MD simulations revealed important differences in the ECoV HE binding site when compared to that of BCoV and RbCoV HEs due to key mutated residues in the loop in close proximity to C4 of Neu5Ac. Among these, Thr²⁴³Asp introduces a negative charge, which disrupts protein-ligand contacts with **6a** due to a close proximity to the hydrophobic acetyl ester at C4 of Neu5Ac. Moreover, in ECoV HE the Ala¹¹⁵, which substitutes either Thr or Ser in BCoV and RbCoV respectively, cannot engage with the acetamide at C5 through hydrogen bonding, reducing the number of intermolecular interactions. In agreement with the array data, MD simulations of the ECoV HE with either **6a** and **5a** resulted in loose poses (Fig. 5f, see Table S4 and S5 and Fig. S14 and S15 in SI for hydrogen bonding interactions). The MD trajectory of compound **8a** showed tighter binding mainly due to hydrophobic interactions of the acetyl ester at C7 with the side chains of Leu¹⁶¹ and Thr¹¹⁴ (Fig. 5f, see Table S6 and Fig. S16 in SI for hydrogen bonding interactions and Fig. S17-S19 for analyses of key hydrophobic interactions). Thus, for ECoV HE the binding depends even more on interactions between the *O*-acetyl moieties and the surrounding hydrophobic side chains. Consequently, the binding enhancing effect of 7-*O*-acetylation became more profound, as was demonstrated in microarray and the MD simulation of ECoV HE in complex with 4,7,9-tri-*O*-acetylated sialoside (Fig. 5e), in which the disadvantageous effect of C4-acetylation was overruled by the presence of a 7-*O*-acetyl and its interaction with Leu¹⁶¹ and Thr¹¹⁴ (summarized in Fig. 6d-f).

Fig. 6 summarises hydrogen bonding, hydrophobic contacts, CH- π interactions and unfavorable interactions of BCoV, RbCoV and ECoV with various *O*-acetylated sialosides and provide a rationale for their binding preferences.

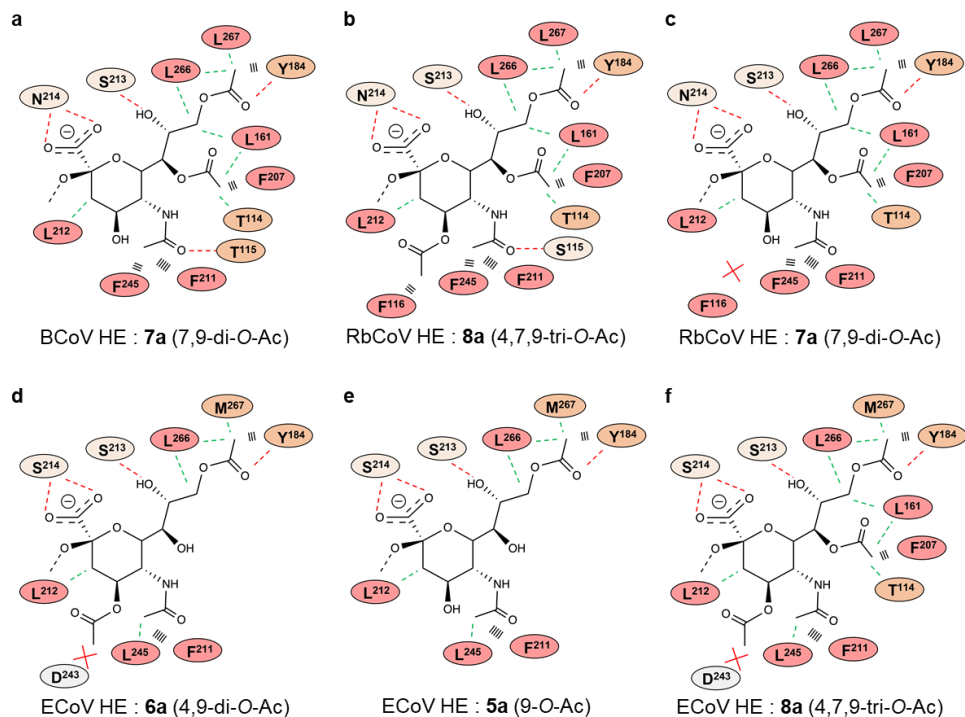


Figure 6. Schematic representation of the interactions between different acetylated sialosides and the lectin domain of the HEs from BCoV (**a**), RbCoV (**b** and **c**) and ECoV (**d-f**). Red dashed lines indicate hydrogen bonding interactions, green dashed lines for hydrophobic contacts, black parallel lines for CH- π interactions and red crosses for unfavorable interactions.

Conclusion

Coronaviruses (CoVs) pose great zoonotic threats, and in particular viruses of the subgenus *Embecovirus* have a remarkable ability to cross species barriers.³⁷ These viruses attach to *O*-acetylated sialoglycans *via* a spike protein for cell entry and membrane fusion. They also express a hemagglutinin-esterase (HE), which acts as a receptor-destroying enzyme. HE harbors a lectin domain that contribute to virion attachment and enhances sialate-*O*-acetylsterase activity of clustered sialosides. The receptor specificity of the lectin domain of HE is critically associated with host specificity,^{9,26} however, the molecular basis for such preferences is poorly understood. Here, we present an integrated approach to investigate the molecular mechanisms by which HEs recognize specific *O*-acetylated sialosides. It employs a range of *O*-acetylated sialoglycans and analogs in which acetyl esters at C7 and C4 are replaced by acetamides. The latter compounds were prepared by a

Chapter 4

chemoenzymatic strategy, in which the acetamides were installed by highly regioselective and late-stage chemical modifications. Glycan microarray binding studies demonstrated the modified sialoglycans exhibit much lower affinities compared to their natural counterparts. The ester-to-amide replacement strategy made it possible to examine the fidelity of recognition of *O*-acetylated sialosides. In particular, ¹H-STD-NMR and MD simulations demonstrated that the fine specificities of different *O*-acetylation forms of closely related proteins arise from hydrophobic patches that can accommodate sialate-*O*-acetyl moieties. Optimal binding is only achieved when all sialate-*O*-acetyl moieties align with the hydrophobic patches of the proteins. A small perturbation in such an alignment, as introduced by an ester-to-amide replacement, greatly reduces binding. HEs from CoVs infecting different host species exhibit different preferences for sialate-*O*-acetyl patterns. Our findings provide insights how receptor binding selectivities leads to viral adaptation to the host's sialoglycome.³⁷ It paves the way to specificity-based engineering of these viral glycoproteins for *in-situ* detection of a select population of *O*-Ac-Sias.^{2, 38} It also offers opportunities to design inhibitors for HEs which may pose antiviral properties. The *O*-to-*N* substitution may offer valuable tools to examine molecular recognition by other proteins such as ficolins³⁹ and anti-*O*-Ac-Sia antibodies.⁴⁰ It is known that immunoregulatory lectins, such as Siglecs,²¹ and viral glycoproteins including neuraminidases of influenza A viruses,⁴¹ tolerate modifications at C4 of a sialosides which provides opportunities to develop high affinity ligands. The synthetic approaches described here provide avenues for the development of such compounds.

Acknowledgments

This work was supported by TOP-PUNT Grant 718.015.003 of the Netherlands Organization for Scientific Research (G.-J.B.); the Human Frontier Science Program Organization (HFSP) grant LT000747/2018-C (L.U.); ECHO Grant 711.011.006 of the Council for Chemical Sciences of the Netherlands Organization for Scientific Research (R.J.d.G.); and China Scholarship Council 2014-03250042 (Y.L.).

Author contribution

G.-J. B. and Z. L. conceived the study. G.-J. B. supervised the study. Z. L. performed chemoenzymatic synthesis and characterization of all compounds. L. U. performed STD-NMR and computational studies. L. L. performed microarray experiments. L. L., Z. L. and G.-J. B. analyzed microarray data. R. P. d. V. expressed glycosyltransferases for chemoenzymatic synthesis. Y. L. and R. J. d. G. provided viral glycoproteins used in the microarray and NMR experiments and critical insights in structural studies. G.-J. B., Z. L. and L. U. co-wrote the paper. All authors discussed and revised the manuscript.

References

1. Schauer, R.; Kamerling, J. P., Exploration of the sialic acid world. *Adv. Carbohydr. Chem. Biochem.* **2018**, *75*, 1-213.
2. Langereis, M. A.; Bakkers, M. J.; Deng, L.; Padler-Karavani, V.; Vervoort, S. J.; Hulswit, R. J.; van Vliet, A. L.; Gerwig, G. J.; de Poot, S. A.; Boot, W.; van Ederen, A. M.; Heesters, B. A.; van der Loos, C. M.; van Kuppeveld, F. J.; Yu, H.; Huizinga, E. G.; Chen, X.; Varki, A.; Kamerling, J. P.; de Groot, R. J., Complexity and diversity of the mammalian sialome revealed by nidovirus virolectins. *Cell Rep.* **2015**, *11* (12), 1966-78.

3. Cariappa, A.; Takematsu, H.; Liu, H.; Diaz, S.; Haider, K.; Boboila, C.; Kalloo, G.; Connole, M.; Shi, H. N.; Varki, N.; Varki, A.; Pillai, S., B cell antigen receptor signal strength and peripheral B cell development are regulated by a 9-*O*-acetyl sialic acid esterase. *J. Exp. Med.* **2008**, *206* (1), 125-138.
4. Wipfler, D.; Srinivasan, G. V.; Sadick, H.; Kniep, B.; Arming, S.; Willhauck-Fleckenstein, M.; Vlasak, R.; Schauer, R.; Schwartz-Albiez, R., Differentially regulated expression of 9-*O*-acetyl GD3 (CD60b) and 7-*O*-acetyl-GD3 (CD60c) during differentiation and maturation of human T and B lymphocytes. *Glycobiology* **2011**, *21* (9), 1161-72.
5. Parameswaran, R.; Lim, M.; Arutyunyan, A.; Abdel-Azim, H.; Hurtz, C.; Lau, K.; Muschen, M.; Yu, R. K.; von Itzstein, M.; Heisterkamp, N.; Groffen, J., *O*-acetylated *N*-acetylneuraminic acid as a novel target for therapy in human pre-B acute lymphoblastic leukemia. *J. Exp. Med.* **2013**, *210* (4), 805-819.
6. Robinson, L. S.; Lewis, W. G.; Lewis, A. L., The sialate *O*-acetyltransferase EstA from gut Bacteroidetes species enables sialidase-mediated cross-species foraging of 9-*O*-acetylated sialoglycans. *J. Biol. Chem.* **2017**, *292* (28), 11861-11872.
7. Nguyen, T.; Lee, S.; Yang, Y. A.; Ahn, C.; Sim, J. H.; Kei, T. G.; Barnard, K. N.; Yu, H.; Millano, S. K.; Chen, X.; Parrish, C. R.; Song, J., The role of 9-*O*-acetylated glycan receptor moieties in the typhoid toxin binding and intoxication. *PLoS Pathog.* **2020**, *16* (2), e1008336.
8. Wasik, B. R.; Barnard, K. N.; Parrish, C. R., Effects of sialic acid modifications on virus binding and infection. *Trends Microbiol.* **2016**.
9. Li, Z.; Lang, Y.; Liu, L.; Bunyatov, M. I.; Sarmiento, A. I.; de Groot, R. J.; Boons, G.-J., Synthetic *O*-acetylated sialosides facilitate functional receptor identification for human respiratory viruses. *Nat. Chem.* **2021**, *13* (5), 496-503.
10. Barlow, R. B.; Bremner, J. B.; Soh, K. S., The Effects of Replacing Ester by Amide on the Biological Properties of Compounds Related to Acetylcholine. *Br. J. Pharmacol.* **1978**, *62* (1), 39-50.
11. Neto, F. R.; Sperelakis, N., Effects of lidocaine, procaine, procainamide and quinidine on electrophysiological properties of cultured embryonic chick hearts. *Br. J. Pharmacol.* **1985**, *86* (4), 817-826.
12. Bakkens, M. J.; Zeng, Q.; Feitsma, L. J.; Hulswit, R. J.; Li, Z.; Westerbeke, A.; van Kuppeveld, F. J.; Boons, G. J.; Langereis, M. A.; Huizinga, E. G.; de Groot, R. J., Coronavirus receptor switch explained from the stereochemistry of protein-carbohydrate interactions and a single mutation. *Proc. Natl. Acad. Sci. USA* **2016**, *113* (22), E3111-9.
13. Khedri, Z.; Xiao, A.; Yu, H.; Landig, C. S.; Li, W. Q.; Diaz, S.; Wasik, B. R.; Parrish, C. R.; Wang, L. P.; Varki, A.; Chen, X., A chemical biology solution to problems with studying biologically important but unstable 9-*O*-acetyl sialic acids. *ACS Chem. Biol.* **2017**, *12* (1), 214-224.
14. Ji, Y.; Sasmal, A.; Li, W.; Oh, L.; Srivastava, S.; Hargett, A. A.; Wasik, B. R.; Yu, H.; Diaz, S.; Choudhury, B.; Parrish, C. R.; Freedberg, D. I.; Wang, L.-P.; Varki, A.; Chen, X., Reversible *O*-Acetyl Migration within the Sialic Acid Side Chain and Its Influence on Protein Recognition. *ACS Chem. Biol.* **2021**, *16*, 1951-1960.
15. Khedri, Z.; Li, Y.; Muthana, S.; Muthana, M. M.; Hsiao, C. W.; Yu, H.; Chen, X., Chemoenzymatic synthesis of sialosides containing C7-modified sialic acids and their application in sialidase substrate specificity studies. *Carbohydr. Res.* **2014**, *389*, 100-11.
16. Sugiarto, G.; Lau, K.; Qu, J.; Li, Y.; Lim, S.; Mu, S.; Ames, J. B.; Fisher, A. J.; Chen, X., A sialyltransferase mutant with decreased donor hydrolysis and reduced sialidase activities for directly sialylating LewisX. *ACS Chem. Biol.* **2012**, *7* (7), 1232-40.
17. Rauvolfova, J.; Venot, A.; Boons, G. J., Chemo-enzymatic synthesis of C-9 acetylated sialosides. *Carbohydr. Res.* **2008**, *343* (10-11), 1605-11.
18. Reiding, K. R.; Blank, D.; Kuijper, D. M.; Deelder, A. M.; Wührer, M., High-throughput profiling of protein *N*-glycosylation by MALDI-TOF-MS employing linkage-specific sialic acid esterification. *Anal. Chem.* **2014**, *86* (12), 5784-93.
19. Muramatsu, W., Chemo- and regioselective monosulfonylation of nonprotected carbohydrates catalyzed by organotin dichloride under mild conditions. *J. Org. Chem.* **2012**, *77* (18), 8083-91.

Chapter 4

20. McArthur, J. B.; Yu, H.; Zeng, J.; Chen, X., Converting *Pasteurella multocida* α 2–3-sialyltransferase 1 (PmST1) to a regioselective α 2–6-sialyltransferase by saturation mutagenesis and regioselective screening. *Org. Biomol. Chem.* **2017**, *15* (7), 1700–1709.
21. Kelm, S.; Madge, P.; Islam, T.; Bennett, R.; Koliwer-Brandl, H.; Waespy, M.; von Itzstein, M.; Haselhorst, T., C-4 modified sialosides enhance binding to siglec-2 (CD22): Towards potent siglec inhibitors for immunoglycotherapy. *Angew. Chem. Int. Ed.* **2013**, *52* (13), 3616–3620.
22. Weinhold, E. G.; Knowles, J. R., Design and evaluation of a tightly binding fluorescent ligand for influenza A hemagglutinin. *J. Am. Chem. Soc.* **1992**, *114* (24), 9270–9275.
23. de Groot, R. J., Structure, function and evolution of the hemagglutinin-esterase proteins of corona- and toroviruses. *Glycoconj. J.* **2006**, *23* (1–2), 59–72.
24. Zeng, Q.; Langereis, M. A.; van Vliet, A. L.; Huizinga, E. G.; de Groot, R. J., Structure of coronavirus hemagglutinin-esterase offers insight into corona and influenza virus evolution. *Proc. Natl. Acad. Sci. USA* **2008**, *105* (26), 9065–9.
25. Langereis, M. A.; Zeng, Q.; Gerwig, G. J.; Frey, B.; von Itzstein, M.; Kamerling, J. P.; de Groot, R. J.; Huizinga, E. G., Structural basis for ligand and substrate recognition by torovirus hemagglutinin esterases. *Proc. Natl. Acad. Sci. USA* **2009**, *106* (37), 15897–902.
26. Bakkers, M. J.; Lang, Y.; Feitsma, L. J.; Hulswit, R. J.; de Poot, S. A.; van Vliet, A. L.; Margine, I.; de Groot-Mijnes, J. D.; van Kuppeveld, F. J.; Langereis, M. A.; Huizinga, E. G.; de Groot, R. J., Betacoronavirus Adaptation to Humans Involved Progressive Loss of Hemagglutinin-Esterase Lectin Activity. *Cell Host Microbe* **2017**, *21* (3), 356–366.
27. Deechongkit, S.; Dawson, P. E.; Kelly, J. W., Toward Assessing the Position-Dependent Contributions of Backbone Hydrogen Bonding to β -Sheet Folding Thermodynamics Employing Amide-to-Ester Perturbations. *J. Am. Chem. Soc.* **2004**, *126* (51), 16762–16771.
28. Gao, J.; Bosco, D. A.; Powers, E. T.; Kelly, J. W., Localized thermodynamic coupling between hydrogen bonding and microenvironment polarity substantially stabilizes proteins. *Nat. Struct. Mol. Biol.* **2009**, *16* (7), 684–690.
29. D.A. Case, R. M. B., D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke.; A.W. Goetz, N. H., S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C.; Lin, T. L., R. Luo, B. Madej, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, H.T. Nguyen, I.; Omelyan, A. O., D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J. Swails.; R.C. Walker, J. W., R.M. Wolf, X. Wu, L. Xiao, P.A. Kollman, AMBER 2016. *University of California, San Francisco*.
30. Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J., GLYCAM06: A generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem.* **2008**, *29* (4), 622–655.
31. Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C., ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696–3713.
32. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79* (2), 926–935.
33. Chandrasekaran, A.; Srinivasan, A.; Raman, R.; Viswanathan, K.; Raguram, S.; Tumpey, T. M.; Sasisekharan, V.; Sasisekharan, R., Glycan topology determines human adaptation of avian H5N1 virus hemagglutinin. *Nat. Biotechnol.* **2008**, *26* (1), 107–13.
34. Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T., SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, *46* (W1), W296–W303.
35. Guex, N.; Peitsch, M. C.; Schwede, T., Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis* **2009**, *30* (S162–S173).
36. Bienert, S.; Waterhouse, A.; de Beer, T. A.; Tauriello, G.; Studer, G.; Bordoli, L.; Schwede, T., The SWISS-MODEL Repository-new features and functionality. *Nucleic Acids Res.* **2017**, *45* (D1), D313–D319.
37. Lang, Y.; Li, W.; Li, Z.; Koerhuis, D.; van den Burg, A. C. S.; Rozemuller, E.; Bosch, B.-J.; van Kuppeveld, F. J. M.; Boons, G.-J.; Huizinga, E. G.; van der Schaar, H. M.; de Groot, R. J.,

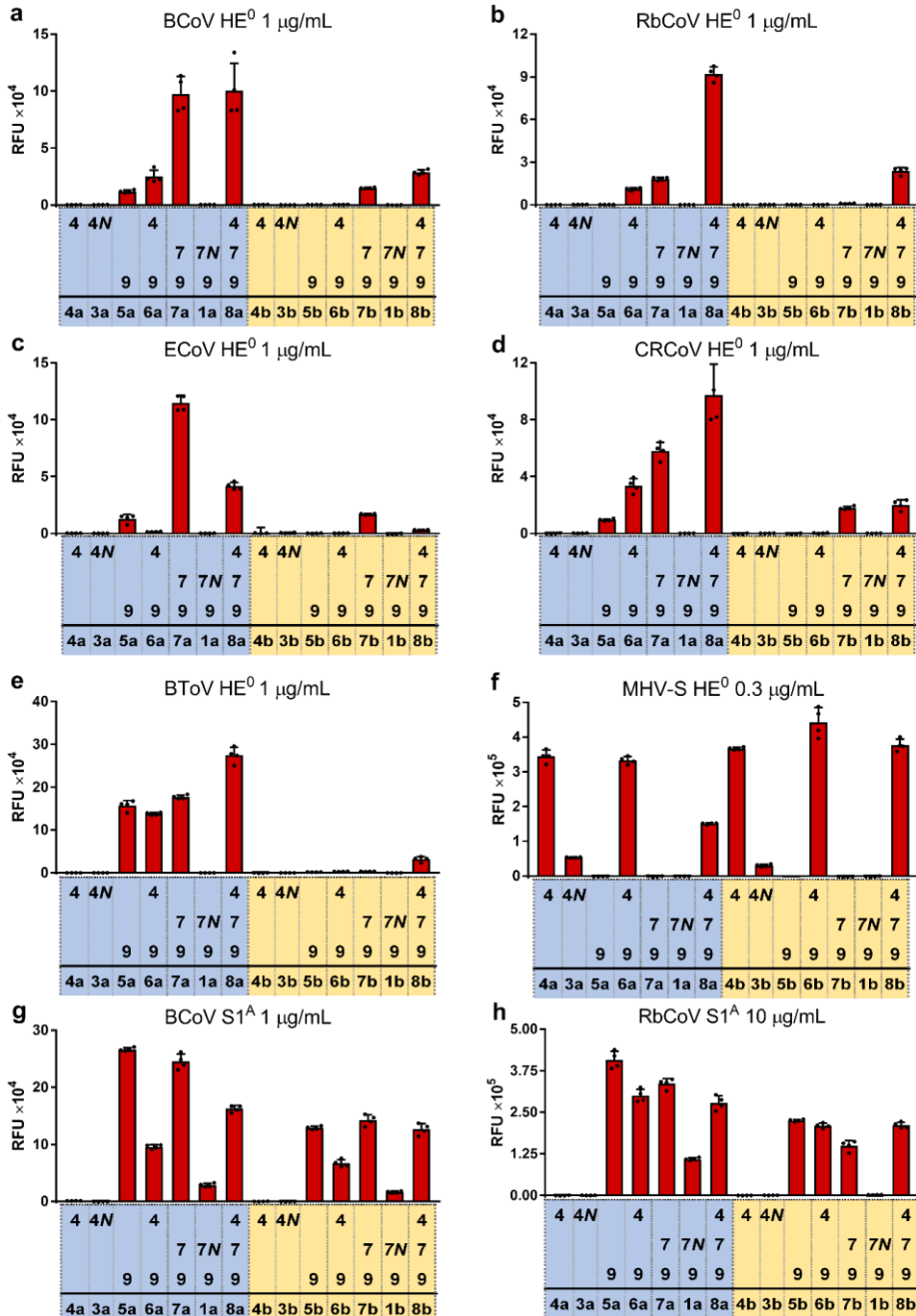
- Coronavirus hemagglutinin-esterase and spike proteins coevolve for functional balance and optimal virion avidity. *Proc. Natl. Acad. Sci. USA* **2020**, 202006299.
38. Wasik, B. R.; Barnard, K. N.; Ossiboff, R. J.; Khedri, Z.; Feng, K. H.; Yu, H.; Chen, X.; Perez, D. R.; Varki, A.; Parrish, C. R., Distribution of O-Acetylated Sialic Acids among Target Host Tissues for Influenza Virus. *mSphere* **2017**, 2 (5), e00379-16.
 39. Gout, E.; Garlatti, V.; Smith, D. F.; Lacroix, M.; Dumestre-Perard, C.; Lunardi, T.; Martin, L.; Cesbron, J. Y.; Arlaud, G. J.; Gaboriaud, C.; Thielens, N. M., Carbohydrate recognition properties of human ficolins: glycan array screening reveals the sialic acid binding specificity of M-ficolin. *J. Biol. Chem.* **2010**, 285 (9), 6612-22.
 40. Kniep, B.; Peter-Katalinić, J.; Flegel, W.; Northoff, H.; Rieber, E. P., CDw 60 antibodies bind to acetylated forms of ganglioside GD3. *Biochem. Biophys. Res. Commun.* **1992**, 187 (3), 1343-1349.
 41. Kiefel, M. J.; von Itzstein, M., Recent Advances in the Synthesis of Sialic Acid Derivatives and Sialylmimetics as Biological Probes. *Chem. Rev.* **2002**, 102 (2), 471-490.

Chapter 4

Supplementary Information for

Synthetic *O*-Acetylated Sialosides and their Acetamido-deoxy Analogues as Probes for Coronaviral Hemagglutinin-Esterase Recognition

Part I. Supplementary Figures and Tables



(Figure continues on the next page)

Chapter 4

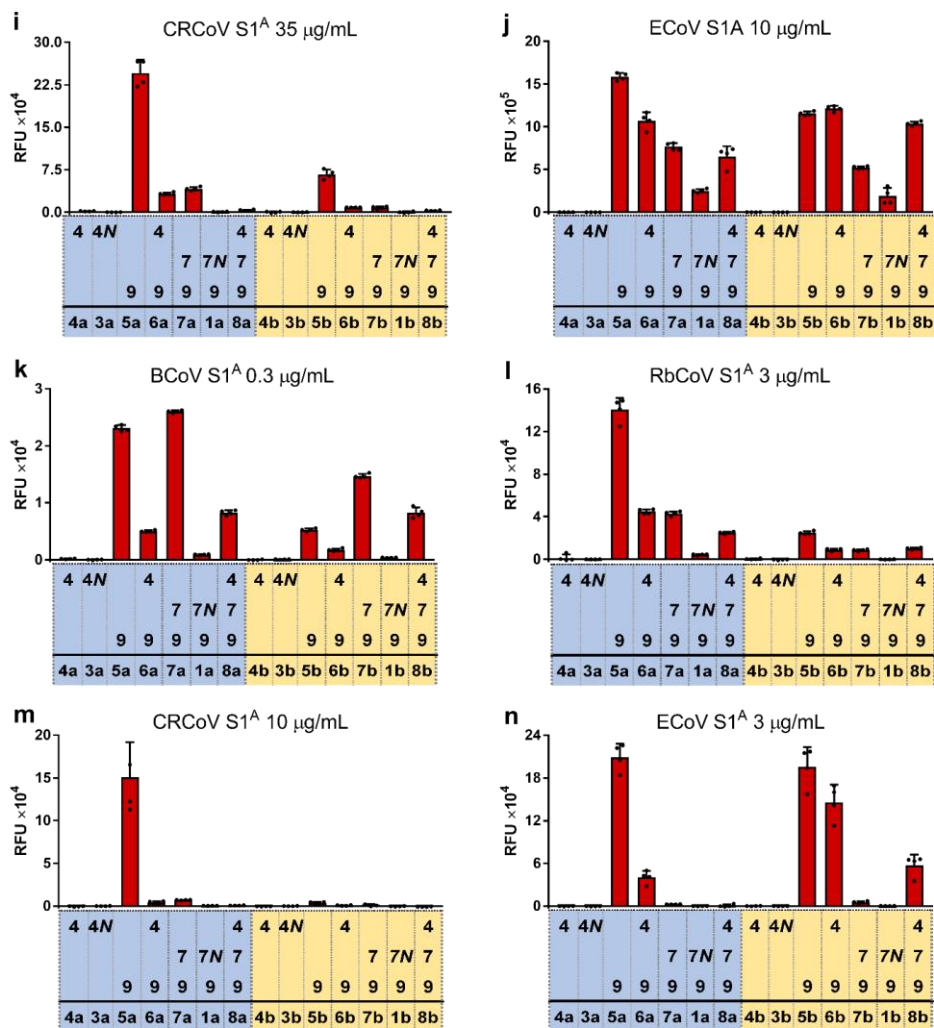
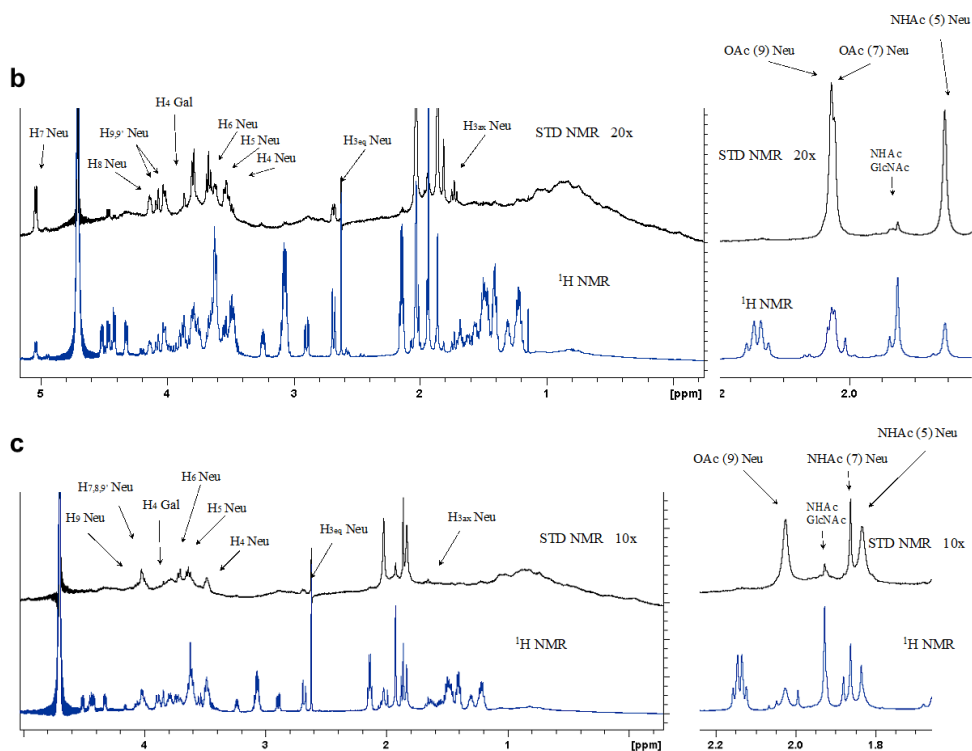


Figure S1. Microarray results of HE and S1^A domain of spike proteins (figures start on the next page). The identities of proteins and the concentration used in the screen are displayed above the column graphs. Columns indicate mean RFU values after background subtraction. Error bars indicate stand deviation. Dots in each compound indicate signal intensities from four replicate spots on microarray.

a

	H3	C3	H4	C4	H5	C5	H6	C6	H7	C7	H8	C8	H9	C9
7a	2.77													
	1.81	39.53	3.61	68.33	3.76	51.13	3.88	71.92	5.12	69.17	4.22	67.54	4.18-4.09	64.33
7b	2.73													
	1.71	40.22	3.59	68.43	3.76	51.15	3.90	71.76	5.16	69.11	4.22	67.33	4.14-4.07	64.02
1a	2.77													
	1.76	39.89	3.59	68.86	3.74	51.63	3.82	71.64	4.12	48.85	4.12	68.82	4.17-4.11	64.92

	CH ₃ 5-N-Ac	CH ₃ 5-N-Ac	CH ₃ 7-O(N)-Ac	CH ₃ 7-O(N)-Ac	CH ₃ 9-O-Ac	CH ₃ 9-O-Ac
7a	1.94	22.19	2.12	20.39	2.11	20.39
7b	1.93	21.91	2.12	20.26	2.11	20.26
1a	1.95	21.92	2.03	21.92	2.14	20.07



(Figure continues on the next page)

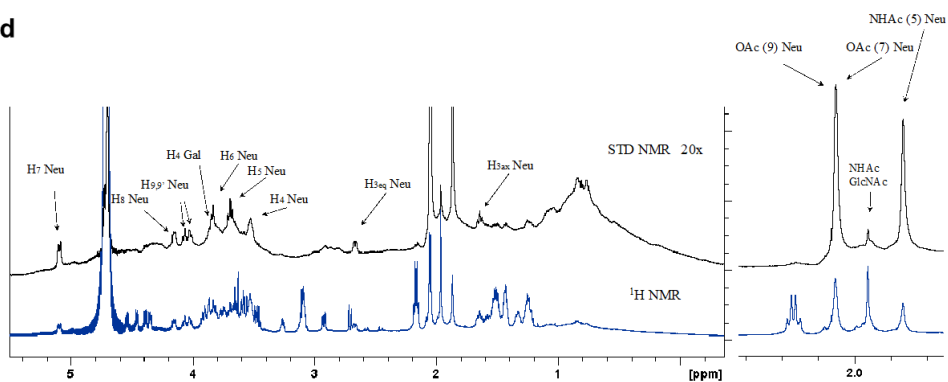
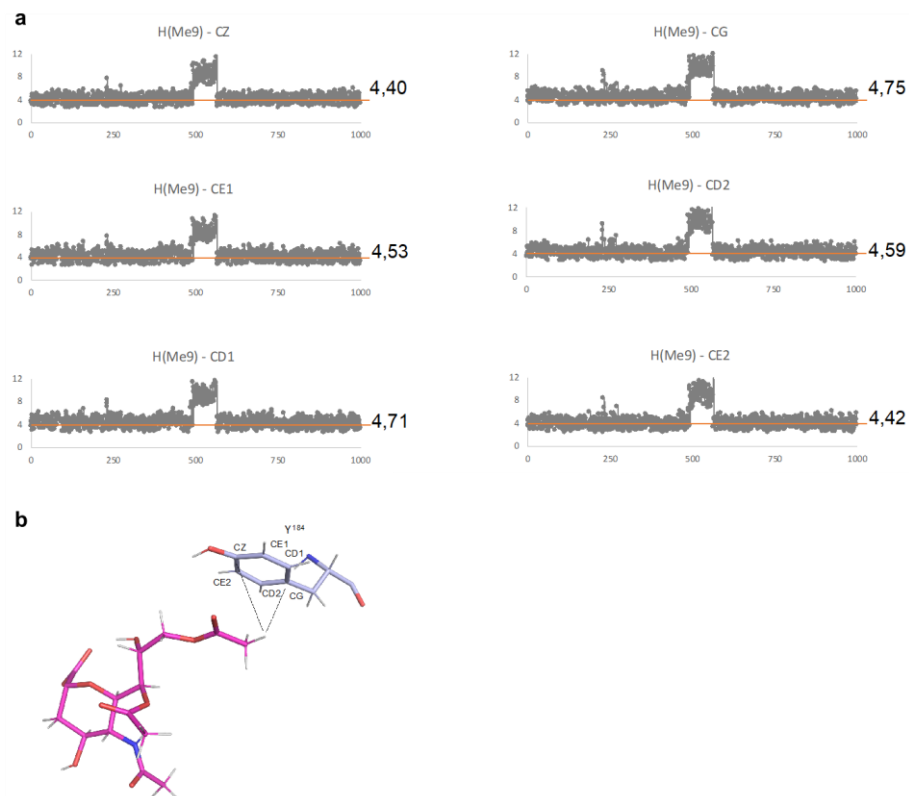
d

Figure S2. ^1H -STD-NMR spectra. **a**, Assignments of Neu5Ac residue for **7a**, **1a** and **7b**. ^1H -STD experiments of **7a** (**b**), **1a** (**c**), and **7b** (**d**) in complex with hydrolase-inactive bovine coronavirus hemagglutinin-esterase (BCoV HE).

Figure S3. Analysis of key hydrophobic intermolecular interactions between BCoV HE and **7a**. **a**, Trajectories of interatomic distances between Y¹⁸⁴ and the methyl of sialate-9-*O*-acetyl along the MD simulation for BCoV HE and **7a**. Grey dots indicate distance between the hydrogen atoms of methyl group in the acetyl ester function at C9 of the Neu5Ac and the carbon atoms of the aromatic ring of the Y¹⁸⁴ side chain at each frame. The threshold of 4 Å, under which effective CH- π interaction occurs, is marked by the orange line. The average distances (in angstroms) are specified on the side of each graph. **b**, Stick representation of the methyl- π interactions between BCoV HE Y¹⁸⁴ and **7a**. The model corresponds to the last frame from the MD simulation.



Chapter 4

Figure S4. Analysis of key hydrophobic intermolecular interactions between BCoV HE and **7a**. **a**, Trajectories of interatomic distances between L²⁶⁶ and the methyl of sialate-9-*O*-acetyl along the MD simulation for BCoV HE and **7a**. Grey dots indicate distance between the hydrogen atoms of the methyl group of the acetyl ester function at C9 of the Neu5Ac and the carbon atoms of the carbon atoms of the L²⁶⁶ side chain at each frame. The threshold of 4 Å, under which effective van der Waals interaction occurs, is marked by the orange line. The average distances (in angstroms) are specified on the side of each graph. **b**, Stick representation of the hydrophobic interactions between BCoV HE L²⁶⁶ and **7a**. The model corresponds to the last frame from the MD simulation.

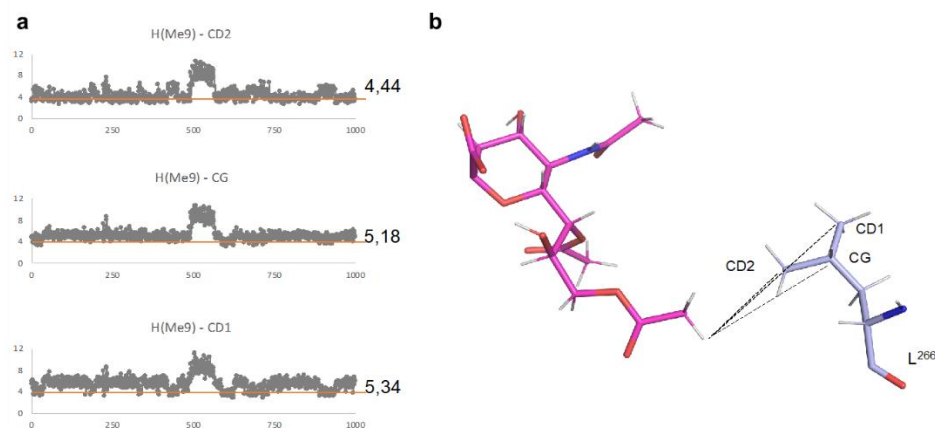
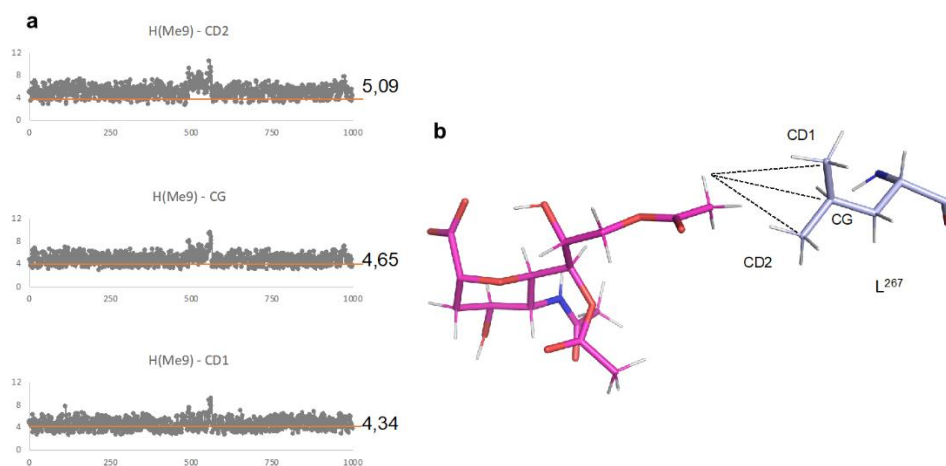


Figure S5. Analysis of key hydrophobic intermolecular interactions between BCoV HE and **7a**. **a**, Trajectories of interatomic distances between L²⁶⁷ and the methyl of sialate-9-*O*-acetyl along the MD simulation for BCoV HE and **7a**. Grey dots indicate distance between the hydrogen atoms of the methyl group of the acetyl ester function at C9 of the Neu5Ac and the carbon atoms of the carbon atoms of the L²⁶⁷ side chain at each frame. The threshold of 4 Å, under which effective van der Waals interaction occurs, is marked by the orange line. The average distances (in angstroms) are specified on the side of each graph. **b**, Stick representation of the hydrophobic interactions between BCoV HE L²⁶⁷ and **7a**. The model corresponds to the last frame from the MD simulation.



Chapter 4

Figure S6. Analysis of key hydrophobic intermolecular interactions between BCoV HE and **7a**. **a**, Trajectories of interatomic distances between F²¹¹ and the methyl of sialate-5-*N*-acetyl along the MD simulation for BCoV HE and **7a**. Grey dots indicate distance between the hydrogen atoms of methyl group in the acetamide function at C5 of the Neu5Ac and the carbon atoms of the aromatic ring of the F²¹¹ side chain at each frame. The threshold of 4 Å, under which effective CH- π interaction occurs, is marked by the orange line. The average distances (in angstroms) are specified on the side of each graph. **b**, Stick representation of the methyl CH- π interactions between BCoV HE F²¹¹ and **7a**. The model corresponds to the last frame from the MD simulation.

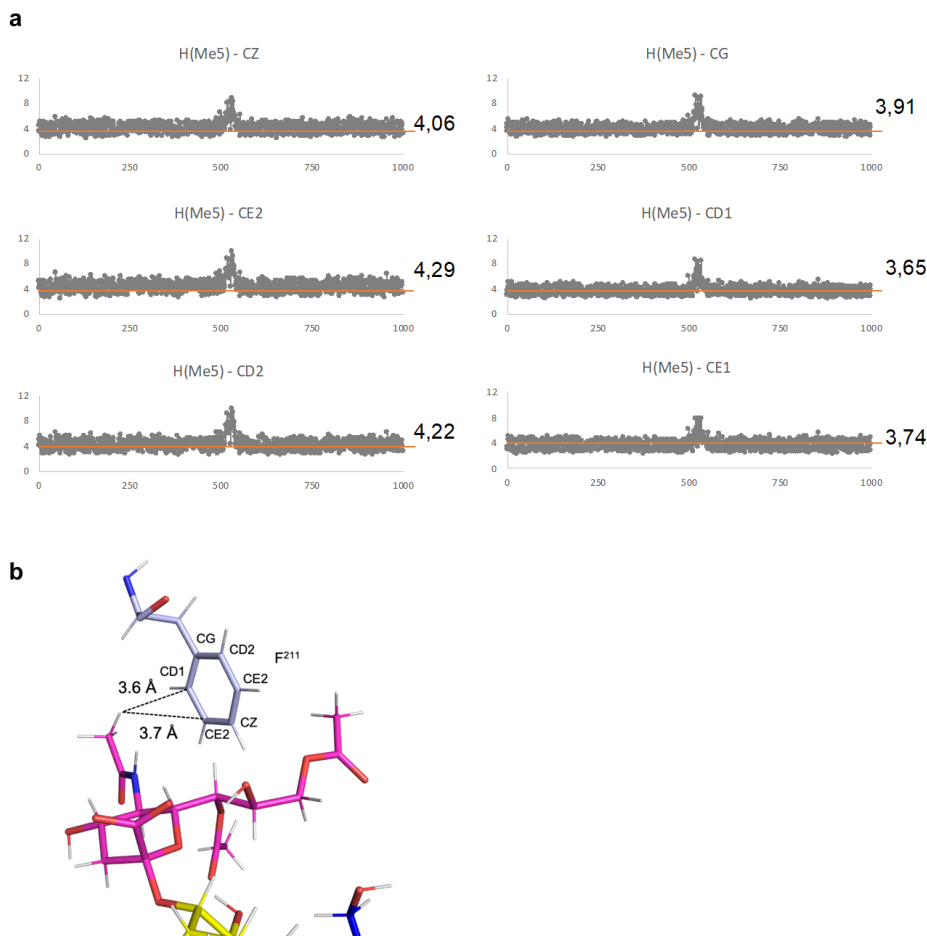
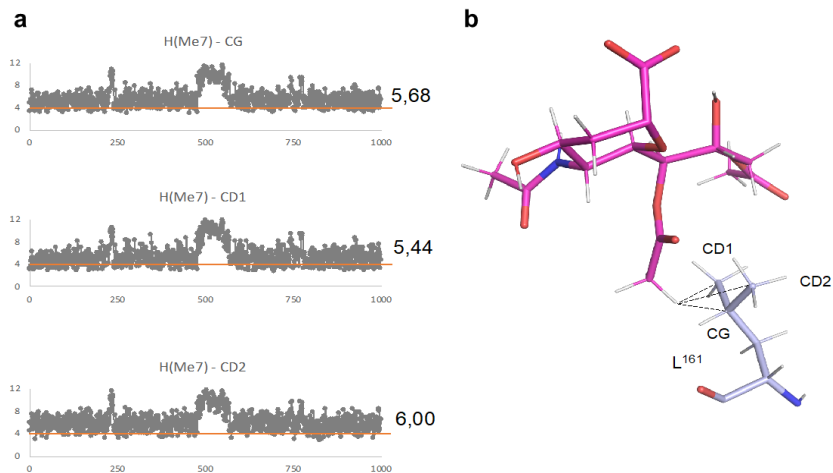


Figure S7. Analysis of key hydrophobic intermolecular interactions between BCoV HE and **7a**. **a**, Trajectories of interatomic distances between L¹⁶¹ and the methyl of sialate-7-*O*-acetyl along the MD simulation for BCoV HE and **7a**. Grey dots indicate distance between the hydrogen atoms of the methyl group of the acetyl ester function at C7 of the Neu5Ac and the carbon atoms of the carbon atoms of the L¹⁶¹ side chain at each frame. The threshold of 4 Å, under which effective van der Waals interaction occurs, is marked by the orange line. The average distances (in angstroms) are specified on the side of each graph. **b**, Stick representation of the hydrophobic interactions between BCoV HE L¹⁶¹ and **7a**. The model corresponds to the last frame from the MD simulation.



Chapter 4

Figure S8. Analysis of key hydrophobic intermolecular interactions between BCoV HE and **7a**. **a**, Trajectories of interatomic distances between T¹¹⁴ and the methyl of sialate-7-*O*-acetyl along the MD simulation for BCoV HE and **7a**. Grey dots indicate distance between the hydrogen atoms of the methyl group of the acetyl ester function at C7 of the Neu5Ac and the carbon atoms of the carbon atoms of the T¹¹⁴ side chain at each frame. The threshold of 4 Å, under which effective van der Waals interaction occurs, is marked by the orange line. The average distances (in angstroms) are specified on the side of each graph. **b**, Stick representation of the hydrophobic interactions between BCoV HE T¹¹⁴ and **7a**. The model corresponds to the last frame from the MD simulation.

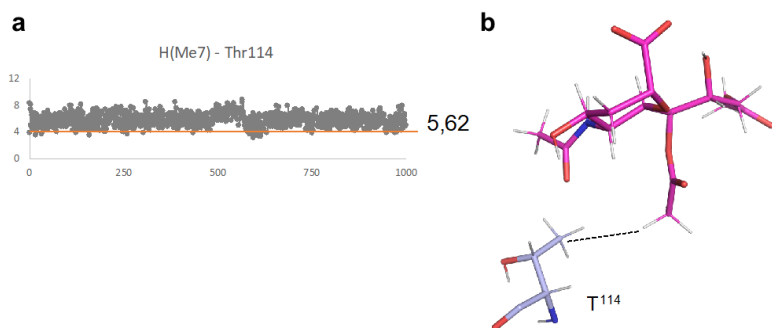
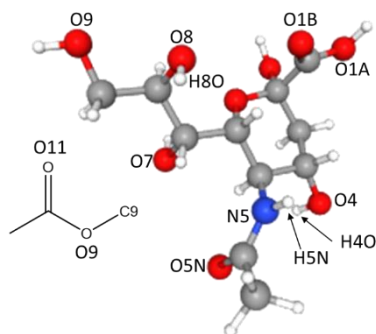


Table S1. Analysis of intermolecular hydrogen bonding (HB) interactions along the MD trajectory of simulation for BCoV HE in complex with 7,9-di-*O*-acetylated α 2,3-linked sialoside **7a** (Table shown on the next page). The presence of HB interactions was calculated based on interatomic distances and angles. The average values for these parameters are listed in the table. HB interactions that populated more than 1% of the 1000 frames are reported. The numbering and nomenclature of amino acid atoms complies with PDB and the latter can be found

http://www.imgt.org/IMGTEducation/Aide-memoire/_UK/aminoacids/formuleAA/.

See figure below for nomenclature of interacting atoms for sialic acid residue. Red, oxygen; blue, nitrogen; grey, carbon; white, hydrogen. Column titles: **Acceptor**, H-bonding acceptor atom; **DonorH**, H-bonding donor hydrogen; **Donor**, the atom where donor hydrogen is attached; **Frames**, number of frames where H-bonding occurs; **Frac**, fraction of frames in a total of 1000 where H-bonding occurs; **AvgDist**, average distance between a donor and an acceptor; **AvgAng**, average angle formed by the three atoms above. For example, Sia@O8 means the oxygen atom at C8 position of sialic acid residue.



Acceptor	DonorH	Donor	Frames	Frac	AvgDist	AvgAng
Sia@O8	Ser_213@HG	Ser_213@OG	811	81%	2.7	163.1
Leu_212@O	Sia@H5N	Sia@N5	728	73%	2.8	155.9
Sia@O5N	Thr_115@H	Thr_115@N	435	43%	2.9	160.7
Sia@O1B	Asn_214@H	Asn_214@N	242	24%	2.9	163.7
Sia@O4	Thr_115@HG1	Thr_115@OG1	224	22%	2.8	158.6
Sia@O5N	Thr_115@HG1	Thr_115@OG1	218	22%	2.8	162.9
Sia@O1A	Asn_214@H	Asn_214@N	207	21%	2.9	163.1
Thr_115@OG1	Sia@H4O	Sia@O4	199	20%	2.8	159.5
Sia@O1A	Ser_213@HG	Ser_213@OG	51	5%	2.7	164.7
Sia@O1B	Ser_213@HG	Ser_213@OG	48	5%	2.7	163.2
Sia@O11	Tyr_184@HH	Tyr_184@OH	13	1.3%	2.8	153.8

Chapter 4

Figure S9. Stick representation of the BCoV HE lectin domain in complex with **7a**. The model corresponds to the last frame from the molecular dynamic simulations. The galactose (Gal) and *N*-acetylglucosamine (GlcNAc) residues are hidden for clarity. Amino acid residues that participates in HB interactions with **7a** are shown (grey-blue sticks) and correspond to those listed in the **Table S1**.

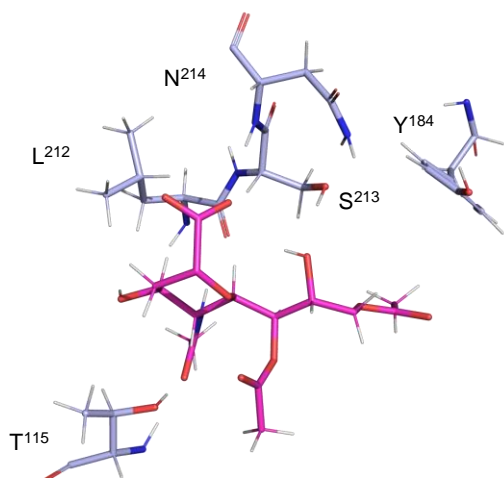
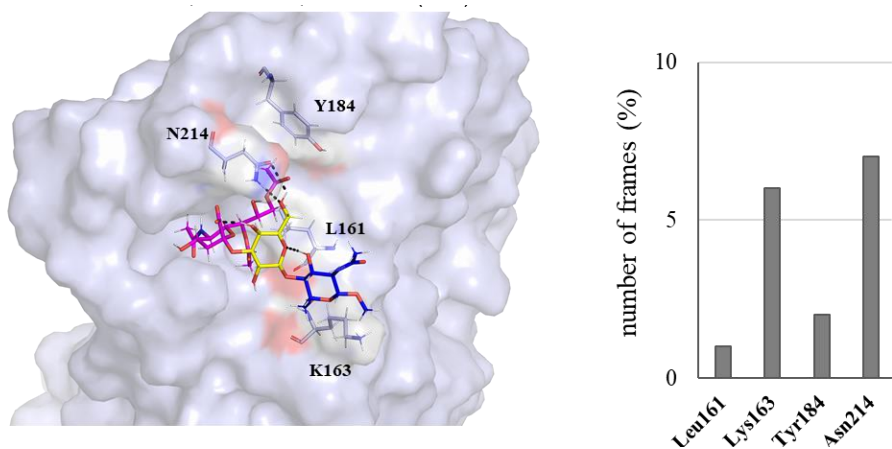
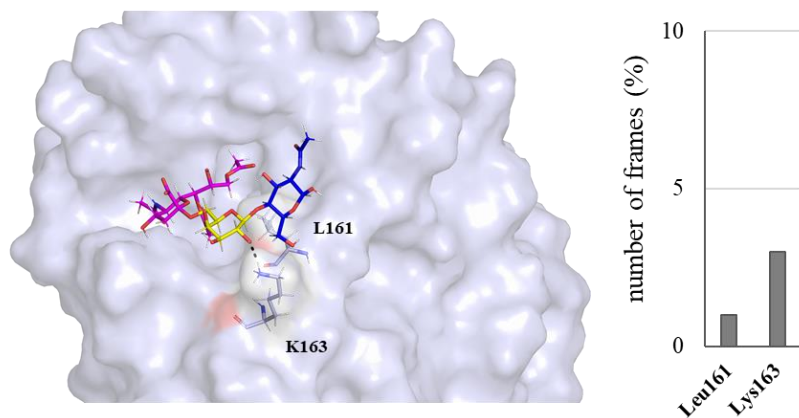


Figure S10. Transient interaction between Gal residue with BCoV HE. **a**, BCoV HE in complex with α 2,3-linked 7,9-di-*O*-acetylated sialoside **7a**. **b**, BCoV HE in complex with α 2,6-linked 7,9-di-*O*-acetylated sialoside **7b**. Bar graphs on the right show Gal-interacting residues in simulation.

a



b

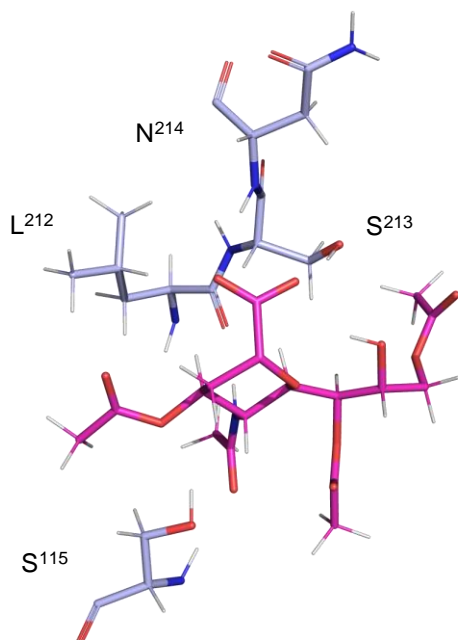


Chapter 4

Table S2. Analysis of HB interactions along the MD trajectory of simulation for RbCoV HE in complex with 4,7,9-tri-*O*-acetylated α 2,3-linked sialoside **8a**. The presence of HB interactions was calculated based on interatomic distances and angles. The average values for these parameters are listed in the table. HB interactions that populated more than 1% of the 1000 frames are reported. See legends of Table S1 for explanation of atom numbering and abbreviations of column titles.

Acceptor	DonorH	Donor	Frames	Frac	AvgDist	AvgAng
Sia@O8	Ser_213@HG	Ser_213@OG	632	63%	2.7	163.0
Leu_212@O	Sia@H5N	Sia@N5	626	63%	2.8	155.5
Sia@O5N	Ser_115@HG	Ser_115@OG	235	23%	2.8	163.9
Sia@O5N	Ser_115@H	Ser_115@N	188	18%	2.9	160.5
Sia@O1A	Asn_214@H	Asn_114@N	158	16%	2.9	161.9
Sia@O1B	Asn_214@H	Asn_214@N	148	15%	2.9	161.1
Leu_161@O	Gal@H6O	Gal@O6	86	9%	2.8	151.8
Ser_213@OG	Sia@H8O	Sia@O8	60	6%	2.8	156.3
Gal@O6	Lys_163@HZ2	Lys_163@NZ	34	3%	2.8	151.8
Gal@O6	Lys_163@HZ1	Lys_163@NZ	32	3%	2.8	152.9
Gal@O6	Lys_163@HZ3	Lys_163@NZ	31	3%	2.9	155.1
Sia@O1B	Ser_213@HG	Ser_213@OG	30	3%	2.7	161.8
Sia@O1A	Ser_213@HG	Ser_213@OG	28	3%	2.7	163.5
Gal@O2	Ser_115@HG	Ser_115@OG	19	2%	2.8	160.53
GlcNAc@O2	Gly_165@H	Gly_165@N	16	2%	2.9	152.1

Figure S11. Stick representation of the RbCoV HE lectin domain in complex with **8a**. The model corresponds to the last frame from the molecular dynamic simulations. The Gal and GlcNAc residues are hidden for clarity. Amino acid residues that participates in HB interactions with **8a** are indicated and correspond to those listed in the **Table S2**.



Chapter 4

Figure S12. Analysis of hydrophobic intermolecular interactions between RbCoV HE and **8a**. **a**, Trajectories of interatomic distances between F¹¹⁶ and the methyl of sialate-4-*O*-acetyl along the MD simulation for RbCoV HE and **8a**. Grey dots indicate distance between the hydrogen atoms of methyl group in the acetyl ester function at C4 of the Neu5Ac and the carbon atoms of the aromatic ring of the F¹¹⁶ side chain at each frame. The threshold of 4 Å, under which effective CH- π interaction occurs, is marked by the orange line. The average distances (in angstroms) are specified on the side of each graph. **b**, Stick representation of the methyl- π interactions between RbCoV HE F¹¹⁶ and **8a**. The model corresponds to the last frame from the MD simulation.

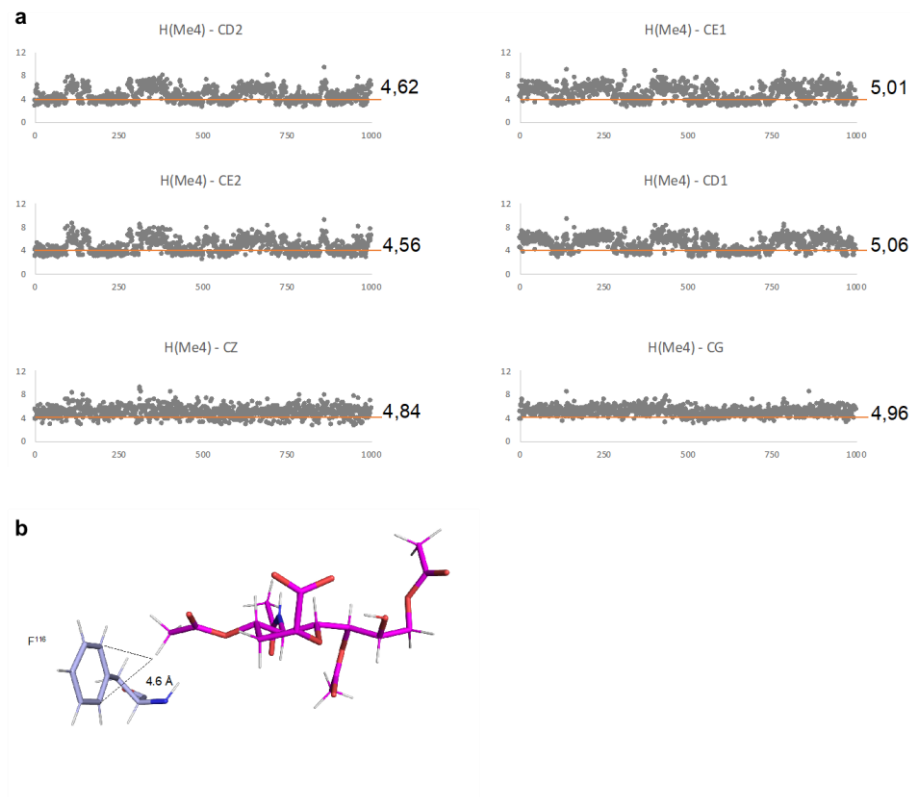


Table S3. Analysis of HB interactions along the MD trajectory of simulation for RbCoV HE in complex with 7,9-di-*O*-acetylated α 2,3-linked sialoside **7a**. The presence of HB interactions was calculated based on interatomic distances and angles. The average values for these parameters are listed in the table. HB interactions that populated more than 1% of the 1000 frames are reported. See legends of Table S1 for explanation of atom numbering and abbreviations of column titles.

Acceptor	DonorH	Donor	Frames	Frac	AvgDist	AvgAng
Leu_212@O	Sia@H5N	Sia@N5	432	43%	2.9	161.4
Sia@O1B	Asn_214@H	Asn_214@N	264	26%	2.8	157.4
Sia@O1A	Ser_213@HG	Ser_213@OG	229	23%	2.6	164.3
Sia@O8	Ser_213@HG	Ser_213@OG	226	23%	2.7	161.6
Sia@O1A	Asn_214@H	Asn_214@N	183	18%	2.9	156.6
Sia@O1B	Ser_213@HG	Ser_213@OG	175	17%	2.7	165.1
Sia@O8	Tyr_184@HH	Tyr_184@OH	170	17%	2.8	161.4
Sia@O11	Arg_177@HH1 2	Arg_177@NH1	121	12%	2.8	155.3
Sia@O11	Arg_177@HH2 2	Arg_177@NH2	103	10%	2.8	152.7
Ser_213@OG	Sia@H8O	Sia@O8	22	2.2%	2.8	157.9
Sia@O11	Asn_214@HD2 2	Asn_214@ND2	17	1.7%	2.9	147.8
Tyr_184@OH	Sia@H8O	Sia@O8	15	1.5%	2.8	158.3
Asn_214@OD1	Sia@H8O	Sia@O8	14	1.4%	2.8	158.2
Sia@O1B	Asn_214@HD2 2	Asn_214@ND2	11	1.1%	2.9	150.6
Sia@O1B	Thr_215@H	Thr_215@N	11	1.1%	2.9	167.6
Sia@O11	Arg_177@HH2 1	Arg_177@NH2	10	1.0%	2.8	147.4
Sia@O1A	Thr_215@H	Thr_215@N	10	1.0%	2.9	162.9

Chapter 4

Figure S13. Stick representation of the RbCoV HE lectin domain in complex with **7a**. The model corresponds to the last frame from the molecular dynamic simulations. The Gal and GlcNAc residues are hidden for clarity. Amino acid residues that participates in HB interactions with **7a** are indicated and correspond to those listed in the **Table S3**.

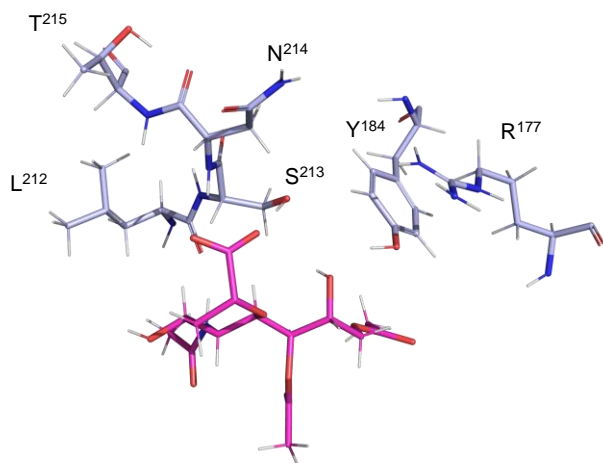
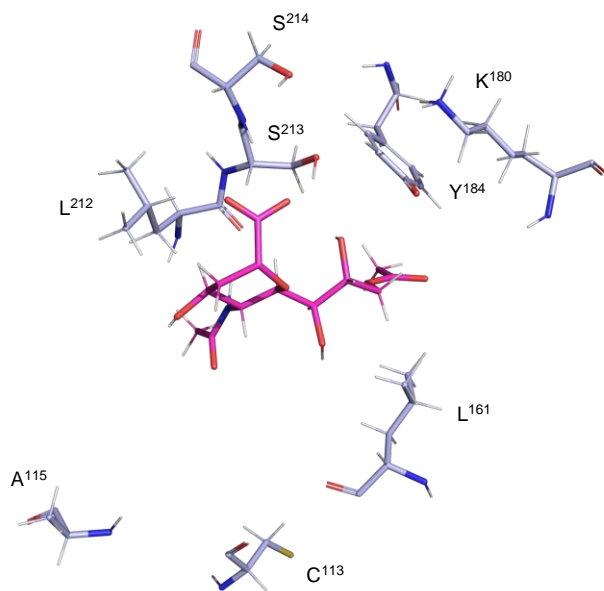


Table S4 (shown on the next page). Analysis of HB interactions along the MD trajectory of simulation for ECoV HE in complex with 9-*O*-acetylated α 2,3-linked sialoside **5a** (Table shown on the next page). The presence of HB interactions was calculated based on interatomic distances and angles. The average values for these parameters are listed in the table. HB interactions that populated more than 1% of the 1000 frames are reported. See legends of Table S1 for explanation of atom numbering and abbreviations of column titles.

Chapter 4

Acceptor	DonorH	Donor	Frames	Frac	AvgDist	AvgAng
Cys_113@O	Sia@H4O	Sia@O4	334	33%	2.8	157.3
Leu_161@O	Sia@H7O	Sia@O7	282	28%	2.8	156.6
Leu_161@O	Sia@H2O	Sia@O2	110	11%	2.8	154.7
Sia@O4	Ser_213@HG	Ser_213@OG	92	9.2%	2.7	163.3
Sia@O4	Gln_164@HE2 2	Gln_164@NE2	92	9.2%	2.9	156.6
Lys_163@O	Gal@H4O	Gal@O4	91	9.1%	2.8	155.0
Sia@O8	Ser_213@HG	Ser_213@OG	55	5.5%	2.7	162.7
Leu_212@O	Sia@H5N	Sia@N5	38	3.8%	2.9	154.6
Sia@O1A	Ser_213@HG	Ser_213@OG	35	3.5%	2.7	162.4
Sia@O11	Tyr_184@HH	Tyr_184@OH	35	3.5%	2.8	158.6
Ser_213@OG	Sia@H4O	Sia@O4	28	2.8%	2.8	161.9
Sia@O11	Lys_177@H	Lys_177@N	23	2.3%	2.9	156.7
Sia@O11	Ser_213@HG	Ser_213@OG	18	1.8%	2.7	162.3
Sia@O1A	Tyr_184@HH	Tyr_184@OH	17	1.7%	2.7	158.5
Cys_113@O	Sia@H5N	Sia@N5	17	1.7%	2.9	153.9
Sia@O4	Ala_115@H	Ala_115@N	16	1.6%	2.8	156.3
Lys_163@O	GlcNAc@H6O	GlcNAc@O6	13	1.3%	2.8	161.4
Sia@O4	Ser_214@HG	Ser_214@OG	1.3	1.3%	2.8	160.6
GlcNAc@O6	Lys_163@HZ1	Lys_163@NZ	1.3	1.3%	2.9	158.1
Sia@O1A	Ser_214@H	Ser_214@N	1.3	1.3%	2.9	163.5

Figure S14. Stick representation of the ECoV HE lectin domain in complex with **5a**. The model corresponds to the last frame from the molecular dynamic simulations. The Gal and GlcNAc residues are hidden for clarity. Amino acid residues that participates in HB interactions with **5a** are indicated and correspond to those listed in the **Table S4**.



Chapter 4

Table S5 (shown on the next page). Analysis of HB interactions along the MD trajectory of simulation for ECoV HE in complex with 4,9-di-*O*-acetylated α 2,3-linked sialoside **6a**. The presence of HB interactions was calculated based on interatomic distances and angles. The average values for these parameters are listed in the table. HB interactions that populated more than 1% of the 1000 frames are reported. See legends of Table S1 for explanation of atom numbering and abbreviations of column titles.

Acceptor	DonorH	Donor	Frames	Frac	AvgDist	AvgAng
Sia@O1B	Ser_213@HG	Ser_213@OG	132	13%	2.7	162.1
Sia@O7	Thr_159@HG1	Thr_159@OG1	120	12%	2.8	157.0
Ala_158@O	GlcNAc@H3O	GlcNAc@O3	119	12%	2.8	158.8
Sia@O11	Lys_177@H	Lys_177@N	84	8.4%	2.9	156.5
Leu_212@O	Sia@H5N	Sia@N5	77	7.7%	2.8	160.0
Asn_160@OD1	GlcNAc@H6O	GlcNAc@O6	76	7.6%	2.7	161.3
GlcNAc@O2N	Ala_158@H	Ala_158@N	72	7.2%	2.9	159.7
Sia@O1B	Ser_214@HG	Ser_214@OG	46	4.6%	2.7	162.2
Sia@O1A	Ser_214@H	Ser_214@N	38	3.8%	2.8	158.3
Ala_158@O	Gal@H6O	Gal@O6	37	3.7%	2.8	158.8
Sia@O11	Ser_215@H	Ser_215@N	32	3.2%	2.9	156.4
Sia@O1A	Ser_213@HG	Ser_213@OG	26	2.6%	2.8	158.3
Asn_160@O	Sia@H7O	Sia@O7	21	2.1%	2.8	158.5
GlcNAc@O6	Asn_160@HD2 2	Asn_160@ND2	19	1.9%	2.8	163.2
Gal@O6	Thr_159@HG1	Thr_159@OG1	17	1.7%	2.8	165.9
Sia@O5N	Ala_115@H	Ala_115@N	17	1.7%	2.9	160.8
GlcNAc@O2N	Tyr_150@HH	Tyr_150@OH	16	1.6%	2.7	158.6
Sia@O11	Tyr_184@HH	Tyr_184@OH	14	1.4%	2.7	157.0
GlcNAc@O6	Thr_159@HG1	Thr_159@OG1	13	1.3%	2.8	154.9
GlcNAc@O6	Asn_160@H	Asn_160@N	12	1.2%	2.9	154.3

Chapter 4

Figure S15. Stick representation of the ECoV HE lectin domain in complex with **6a**. The model corresponds to the last frame from the molecular dynamic simulations. The Gal and GlcNAc residues are hidden for clarity. Amino acid residues that participates in HB interactions with **6a** are indicated and correspond to those listed in the **Table S5**.

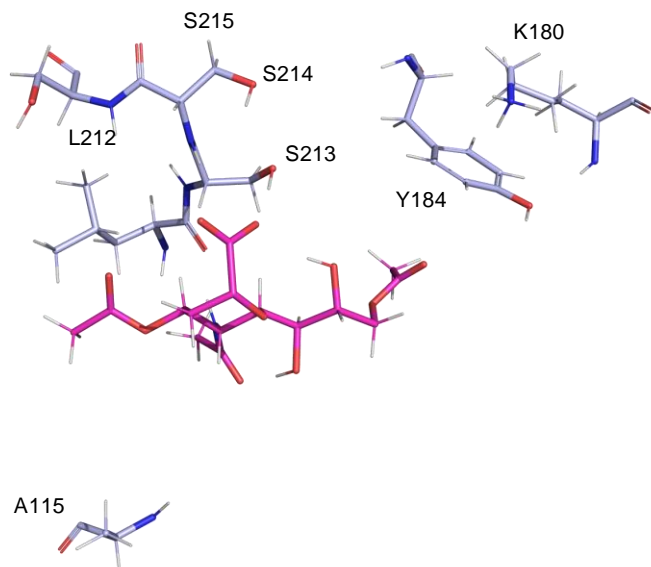


Table S6. Analysis of HB interactions along the MD trajectory of simulation for ECoV HE in complex with 4,7,9-tri-*O*-acetylated α 2,3-linked sialoside **8a**. The presence of HB interactions was calculated based on interatomic distances and angles. The average values for these parameters are listed in the table. HB interactions that populated more than 1% of the 1000 frames are reported. See legends of Table S1 for explanation of atom numbering and abbreviations of column titles.

Acceptor	DonorH	Donor	Frames	Frac	AvgDist	AvgAng
Leu_212@O	Sia@H5N	Sia@N5	800	80%	2.8	160.2
Sia@O8	Ser_213@HG	Ser_213@OG	343	34%	2.7	162.4
Sia@O1A	Ser_214@H	Ser_214@N	269	27%	2.9	158.9
Sia@O1B	Ser_213@HG	Ser_213@OG	266	27%	2.7	163.7
Sia@O1A	Ser_213@HG	Ser_213@OG	234	23%	2.7	163.2
Sia@O1B	Ser_214@H	Ser_214@N	205	20%	2.9	158.5
Sia@O5N	Ala_115@H	Ala_115@N	170	17%	2.8	162.3
Sia@O11	Ser_215@H	Ser_215@N	45	4.5%	2.9	154.3
Ser_213@OG	Sia@H8O	Sia@O8	34	3.4%	2.8	155.7
Sia@O8	Tyr_184@HH	Tyr_184@OH1	32	3.2%	2.8	160.0
Leu_161@O	Gal@H6O	Gal@O6	23	2.3%	2.8	157.0
Lys_163@O	GlcNAc@H6O	GlcNAc@O6	20	2.0%	2.8	155.4
Sia@O11	Tyr_184@HH	Tyr_184@OH	17	1.7%	2.8	152.9
Tyr_184@OH	Sia@H8O	Sia@O8	13	1.3%	2.9	160.7

Chapter 4

Figure S16. Stick representation of the ECoV HE lectin domain in complex with **8a**. The model corresponds to the last frame from the molecular dynamic simulations. The Gal and GlcNAc residues are hidden for clarity. Amino acid residues that participates in HB interactions with **8a** are indicated and correspond to those listed in the **Table S6**.

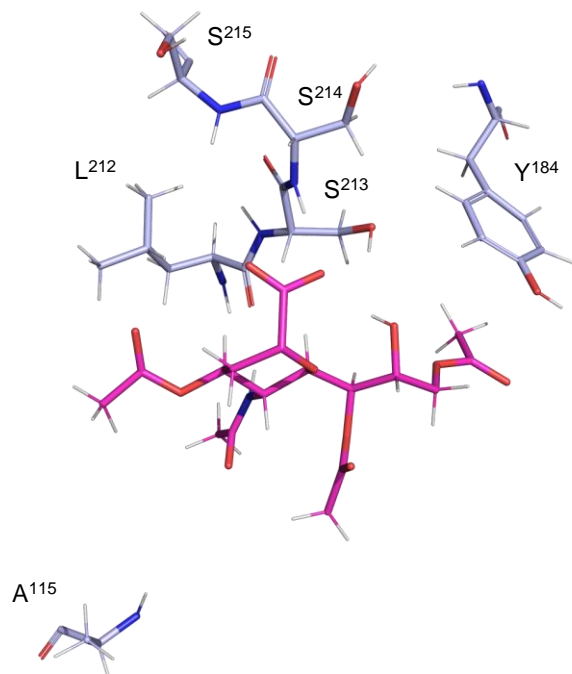
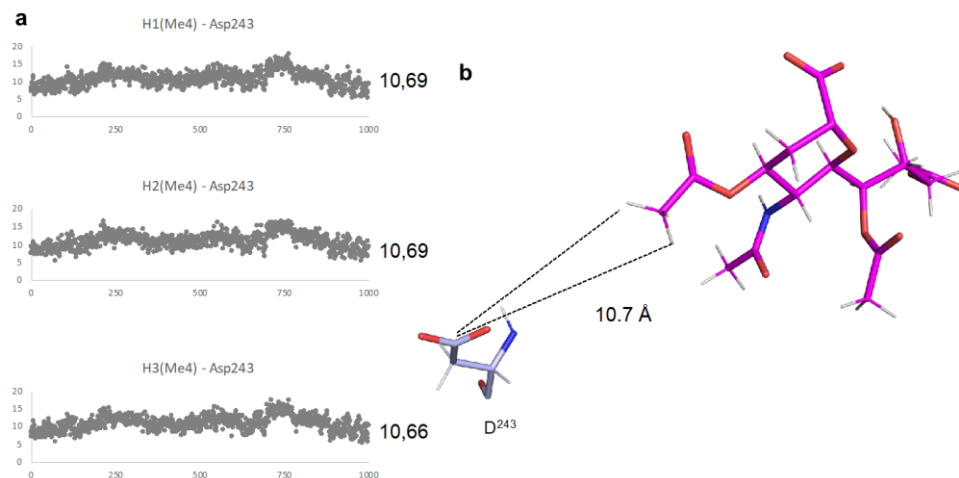


Figure S17. Reduced hydrophobic intermolecular interactions between ECoV HE and **8a**. **a**, Trajectories of interatomic distances between D²⁴³ and the methyl of sialate-4-*O*-acetyl along the MD simulation for ECoV HE and **8a**. Grey dots indicate distance between the hydrogen atoms of methyl group in the acetyl ester function at C4 of the Neu5Ac and the carboxylate carbon atom of the D²⁴³ at each frame. **b**, Stick representation of **8a** in the binding site of ECoV HE. The hydrophobic interaction with 4-*O*-acetyl is absent. The model corresponds to the last frame from the MD simulation.



Chapter 4

Figure S18. Reduced hydrophobic intermolecular interactions between ECoV HE and **8a**. **a**, Trajectories of interatomic distances between I²⁴⁵ and the methyl of sialate-5-*N*-acetyl along the MD simulation for ECoV HE and **8a**. Grey dots indicate distance between the hydrogen atoms of methyl group in the acetamide function at C5 of the Neu5Ac and the carbon atom of the I²⁴⁵ at each frame. Note the large distance between the two moieties. **b**, Stick representation of **8a** in the binding site of ECoV HE. The model corresponds to the last frame from the MD simulation.

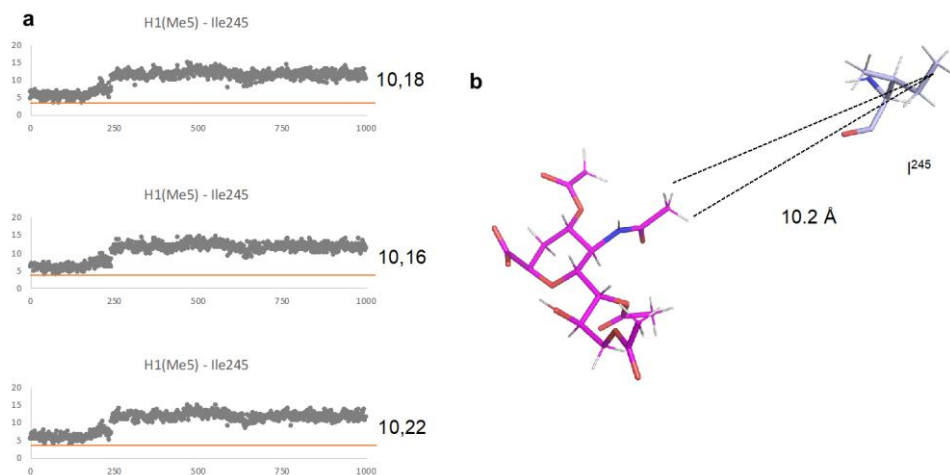
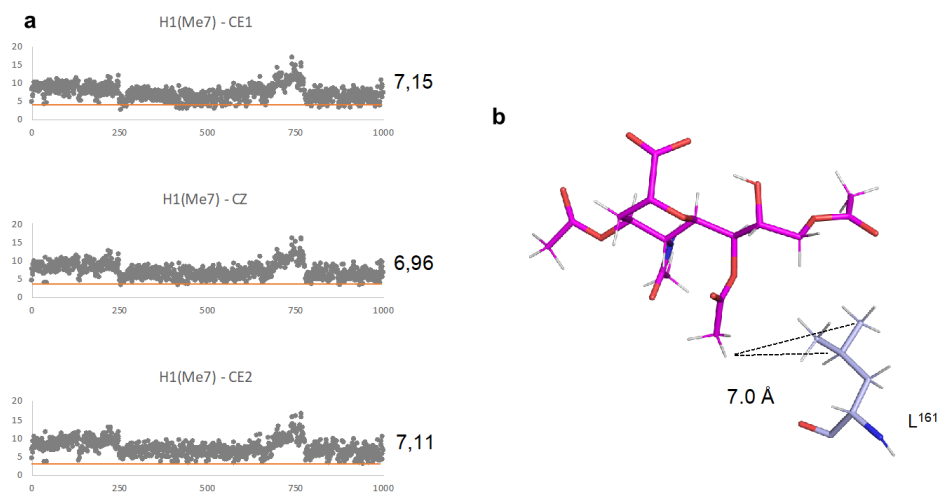


Figure S19. Reduced intermolecular interactions between ECoV HE and **8a**. **a**, Trajectories of interatomic distances between L¹⁶¹ and the methyl of sialate-7-*O*-acetyl along the MD simulation for ECoV HE and **8a**. Grey dots indicate distance between the hydrogen atoms of methyl group in the acetyl ester function at C7 of the Neu5Ac and the carbon atom of the L¹⁶¹ at each frame. Note the larger distance between the two moieties compared with that in BCoV and RbCoV binding. **b**, Stick representation of **8a** in the binding site of ECoV HE. The model corresponds to the last frame from the MD simulation.



Chapter 4

Part II. Experimental Section

1. Materials and methods

Unless otherwise stated, all reagents for chemical reactions were purchased from Merck Sigma-Aldrich. Solvents for reactions were taken from the solvent purifier (MB SPS 5) or purchased from Biosolve Chemie. Technical grade organic solvents in work-ups were from VWR Chemicals. Deuterated solvents for NMR experiments were purchased from Cambridge Isotope Laboratories. Sialic acid aldolase was purchased from TOYOBO Enzymes and was reconstituted into stock solution according to manufacturer's guide. ST6Gal-I, PMST1-M144D and PMST1-P34H/M144L were expressed according to reported protocols (Refs 13 and 17 in main text and ¹). Pd₂,6ST were purchased from Chemily, LLC. Precoated silica gel 60 F-254 plates (Merck) were used for TLC analysis. Column chromatography was performed using flash silica gel (60 Å, from Silicycle, Canada). NMR spectra were collected on an Agilent 400-MR, VARIAN or Bruker 600 UltraShield. Shimadzu Kratos axima-CFR MALDI-ToF or Bruker micrOTOF-Q II ESI mass spectrometer were used for MS-based reaction monitoring. Agilent 6560 ion mobility Q-ToF MS was used to collect HRMS data. Enzymatic reactions are conducted using MaxQ 4450 incubator (Thermo Scientific) at 37 °C with shaking. Chemical structures were drawn using ChemDraw Professional 16.0. NMR spectra were processed and analyzed using MestReNova. Microarray and computational results were graphed using Graphpad Prism 7.

2. General procedures

General procedure A, sialylation using human or bacterial sialyltransferases. The procedure for enzymatic sialylation was adapted from the previous report with slight modification. For the conversion from ManNAc, the glycosyl acceptor **11** (working concentration 10 mM), the six-carbon substrate **9** (15 mM), cytidine triphosphate sodium salt (15 mM) and sodium pyruvate (50 mM) were dissolved in 200 mM Tris buffer pH 8.9. To this solution, 200mM (10x) magnesium chloride solution was added to achieve a working concentration of 20 mM. The pH of the resulted solution was adjusted to 7.4 with 1M NaOH before adding enzymes. The aldolase (working concentration ~0.8 mg/mL), synthetase (0.5 – 1 mg/mL) and sialyltransferases (0.25 – 0.5 mg/mL for ST6Gal1 and 1 – 2 mg/mL for PMST1) were added in a sequential manner. The mixture was incubated at 37 °C and was monitored with TLC (ethyl acetate/methanol/water 4/2/1) and ESI-MS (negative mode) until disappearance of **11** was observed. Upon completion, cold ethanol was added and the mixture was place at 4 °C for two hours. After centrifugation, the supernatant was transferred to a clean round bottom flask. The precipitate was washed 3 times with cold ethanol and all wash solution was combine in the same round bottom flask. The solution was concentrated in vacuo and was loaded onto biogel p2 for purification. For conversion from 4-amino Neu5Ac, sodium pyruvate and aldolase were omitted. 2 – 4 mg/mL PMST1 and its α 2,6-transferase mutant were used.

General procedure B, hydrogenation-biotinylation-acetylation (performed post-sialylation). This general procedure can apply to **12a**, **12b**, **17a** and **17b**, except that for **17a** and **17b**, biotinylation step was omitted. In a round bottom flask equipped with hydrogen balloon, the azido starting materials were dissolved in water. Catalytic (5 – 10% wt) palladium hydroxide was added to into the solution, which was then vigorously stirred. TLC and ESI-MS were used to follow the reaction. Upon completion, the mixture was filtered through Celite pad to remove the catalyst. The filtrate was lyophilized. The residue was taken

up with DMF/water (4:1), into which 3 eq. triethylamine (NEt_3) was added. 1.2 eq. NHS-Lc-biotin was dissolved in DMF and added dropwise at 0 °C. Upon full consumption of bis-amine compound (monitored with ESI-MS), 10 eq. NEt_3 and 5 eq. acetic anhydride was added. The reaction was followed with TLC and ESI-MS. The completed reaction mixture was concentrated in vacuo. The over-acetylated species, which carries *O*-acetyl groups, could be hydrolysed with 0.5M NaOH. The strongly basic solution was neutralized by addition of acidic resin. After filtration, the solution was concentrated and loaded onto biogel p2 for purification.

General procedure C, orthoester-mediated regioselective *O*-acetylation. This procedure applies to **14a** and **14b**. The starting material was dissolved in DMSO to achieve a concentration of 25 mM. 1 eq. toluenesulfonic acid monohydrate and 7.5 eq. trimethyl orthoacetate were added. This solution was then warmed up to 40 °C. The reaction was monitored with ESI-MS. When ~90% of the starting material was consumed, the reaction was quenched with 10 times volume water containing 1 eq. NEt_3 and was freeze-dried. The residue was purified sequentially with biogel p2 and HILIC-HPLC.

General procedure D, sulfonylation-azide substitution. This procedure applies to **15a** and **15b**. The methyl ester or lactone was dissolved in DMF (0.25 M). The reagents were first made into solutions in THF: A. dibutyltin dichloride, 0.1 M; B. diisopropylethylamine (DIPEA) 0.2 M; C. 3,5-bis(trifluoromethyl)benzenesulfonyl chloride (TfBsCl), 0.2 M. Solution A and B were added first into starting material DMF solution. The mixture was stirred at room temperature for 10 minutes, after which solution C was added. The final concentrations for each component were as follows: **15a** or **15b**, 30 mM; dibutyltin dichloride, 3 mM; DIPEA, 60 mM; TfBsCl, 45 mM. The reaction was warmed up to 40 °C and was monitored with TLC (dichloromethane:methanol 6:1). Upon completion, the reaction was quenched with sat. NH_4Cl , extracted with ethyl acetate. The organic phase was washed with brine and dried over Na_2SO_4 . After filtration, the solution was concentrated in vacuo and was directly subjected to the next reaction without further purification. The residue from sulfonylation was taken up with DMF, to which 5 eq. NaN_3 was added. The mixture was heated up to 60 °C and stirred vigorously for 6 hours. The reaction can be monitored with TLC, for this $\text{S}_\text{N}2$ reaction results in loss of UV absorption. Upon completion, the mixture was concentrated in vacuo and was then taken up with 0.2 M NaOH to hydrolyse the methyl ester or the lactone. The solution was then neutralized with acidic resin and freeze-dried. The residue was loaded onto biogel p2 for purification.

General procedure for microarray analysis. This protocol complies with MIRAGE Glycan Array Guideline v 1.0². The protocol was the same as reported in Ref. 9 in the main text. The streptavidin coated glass slides used were SuperStreptavidin Microarray Substrate Slides from ArrayIt Inc. The compounds were printed using a Scienion sciFLEXARRAYER S3 non-contact microarray equipped with a Scienion PDC80 nozzle (Scienion Inc). The oligosaccharides were first dissolved in MilliQ water at the concentration of 1 mM to make the stock solution, which should be placed on ice while in use, and at -80 °C for long-term storage. The printing solution of biotinylated sialosides was freshly made each time at 100 μM from by diluting the stock solution in the ice cold printing buffer (10 mM PBS buffer, pH 7). The compounds were printed in replicates of 6 with spot volume ~ 400 pL, at 20 °C and 50% humidity. After printing, the slides should be stored at 4 °C. For binding studies, the slides were blocked with TSM binding buffer (20 mM Tris-HCl, pH 7, 150 mM NaCl, 2 mM CaCl_2 and 2 mM MgCl_2 , 0.05% Tween-20, 1% BSA) for 1 h at 4 °C. The human-Fc tagged proteins were pre-mixed with goat anti-human IgG antibody (Alexa Fluor 647

Chapter 4

conjugated, 109-605-008, Jackson ImmunoResearch) in a 1:1 ratio in TSM binding buffer and incubated at 4°C for 1 h. The slides were incubated with the binding solution containing pre-complexed virolectin-secondary antibody at 4 °C for 1 h. After incubation, the slides were washed with **1)** TSM washing buffer (20 mM Tris·HCl, pH 7, 150 mM NaCl, 2 mM CaCl₂ and 2 mM MgCl₂, 0.05% Tween-20), **2)** TSM buffer (20 mM Tris·HCl, pH 7, 150 mM NaCl, 2 mM CaCl₂ and 2 mM MgCl₂) and **3)** water, and were finally spun dry. The slides were scanned using a GenePix 4000B microarray scanner (Molecular Devices) at the appropriate excitation wavelength. The power for scanning was optimized to ensure no over-saturation would occur. The image was analyzed using GenePix Pro 7 software (version 7.2.29.2, Molecular Devices). The data processing Macro run in Microsoft Excel is available at <http://zenodo.org/record/5146251>.

General procedure for ¹H Saturation transfer difference (¹H-STD) NMR. The samples for ¹H-STD NMR experiments were prepared using the prototype virolectin BCoV HE (hydrolase-inactive form) at 30 μM concentration in 50 mM PBS, in D₂O (pD 7.2) using lectin/ligand ratios of 1:50. The temperature was set to 298 K. STD experiments were performed at 600 MHz Bruker spectrometer, using standard Bruker pulse sequences without water suppression nor protein spin-lock filter. Protein saturation was achieved with a Gaussian-shaped pulse of 40 ms during a saturation time of 2s. The relaxation delay was set to 3s. The on-resonance frequency was set at aliphatic regions (0.76 ppm) and the off-resonance frequency at 100 ppm. Although a peak-broadening effect was observed for the signals of oligosaccharides in the presence of HEs, it did not affect the chemical shifts, and allowed unambiguous signal assignments. Blank STD experiments of the ligands and protein alone were acquired in the same conditions.

Recombinant hemagglutinin-esterase (HE) expression. Procedures for expression of HEs were the same as Ref. 9 in the main text. All HEs used in this study were expressed in the esterase-inactive forms. In brief, codon-optimized sequences of the Fc-fused HEs were cloned into pCD5 or pCAGGS plasmids. In the construct, HE ectodomains were fused to the Fc domain of human IgG1, separated by a thrombin-cleavage site. HEK293T cells (~2*10⁷) were transfected with 20 μg plasmid DNA conjugated to 2 μg polyethyleneimine (Polysciences). The transfection mixture was replaced by 293 SFM II expression medium (Invitrogen) supplemented with 44.1 mM sodium bicarbonate, 11.1 mM glucose, Primatone RL-UF (3.0 g/L), penicillin (100 IU/mL), streptomycin (100 μg/mL), 1% glutaMAX (Gibco), and 1.5% DMSO 16 h after transfection. The supernatants were harvested 6-7 days after transfection, and subjected to consecutive centrifugation at 1200rpm, 4°C for 5 min and 4000 rpm, 4°C for 10 min to obtain a clear solution. The solutions were incubated with Protein A-Sepharose (v/v, 150 μl per 50 mL supernatant) overnight at 4°C, and were then collected through a Poly-Prep chromatography column (Bio-rad). HE-Fc proteins were eluted from resin with 0.1M citric acid and neutralized with 1/3 volume of 1M Tris-HCl, pH 8.8. All proteins were dialyzed against PBS and store at -80°C for long term storage until use. Coronaviral S1^A-Fc proteins were expressed following exactly the same procedure, except for the S1^A domains from the corresponding spike proteins were fused with human Fc.

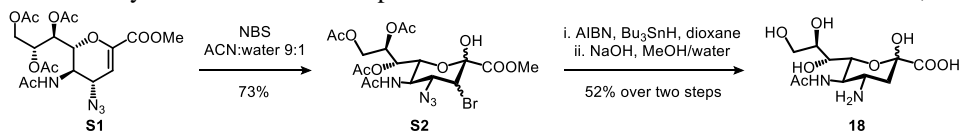
In silico studies. Initial geometries of the ligands were built in the Glycam web (<http://glycam.org>). The initial pdb coordinates for BCoV HE were derived from the crystal structure Protein Database (PDB) 3cl5, whereas those for RbCoV and ECoV HEs were generated as homology pdbs using SWISS-MODEL. The Neu5Ac residue of each glycan was superimposed onto the corresponding sugar in the deposited 3cl5 structure. The resulting

binding poses were used as starting points for molecular dynamics (MD) simulations. The MD simulations were performed using the Amber16 program with the ff14SB force field parameters for protein and GLYCAM_06hj-1 for the saccharides. Thereafter, the starting 3D geometries were placed into a 10 Å octahedral box of explicit TIP3P waters, and counterions were added to maintain electroneutrality. Two consecutive minimization stages were performed involving (1) only the water molecules and ions and (2) the whole system with a higher number of cycles, using the steepest descent algorithm. The system was subjected to two rapid molecular dynamic simulations (heating and equilibration) before starting the real dynamic simulation. The equilibrated structures were the starting points for the final MD simulations at constant temperature (300 K) and pressure (1 atm). 500 ns Molecular dynamics simulations without constraints were recorded, using an NPT ensemble with periodic boundary conditions, a cutoff of 10 Å, and the particle mesh Ewald method. A total of 50,000,000 molecular dynamics steps were run with a time step of 1 fs per step. Coordinates and energy values were recorded every 50,000 steps (10 ps) for a total of 1,000 MD models. A detailed analysis of the intermolecular hydrogen bonds (HB) and hydrophobic interactions was performed along the MD trajectory using the cpptraj module included in Amber-Tools 16 package.

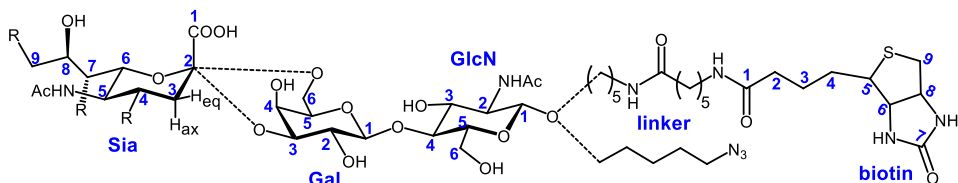
Chapter 4

3. Detailed synthesis and characterization

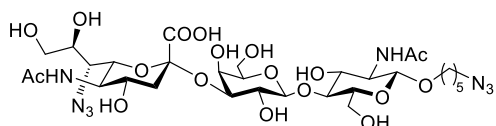
Scheme S1. Synthetic routes from reported intermediate **S1** to **18**. Abbreviations: NBS, *N*-



bromosuccinimide; ACN, acetonitrile; AIBN, 2,2'-Azobis(2-methylpropionitrile); Bu₃SnH, tri-*n*-butyltin hydride; NaOH, sodium hydroxide; MeOH, methanol.

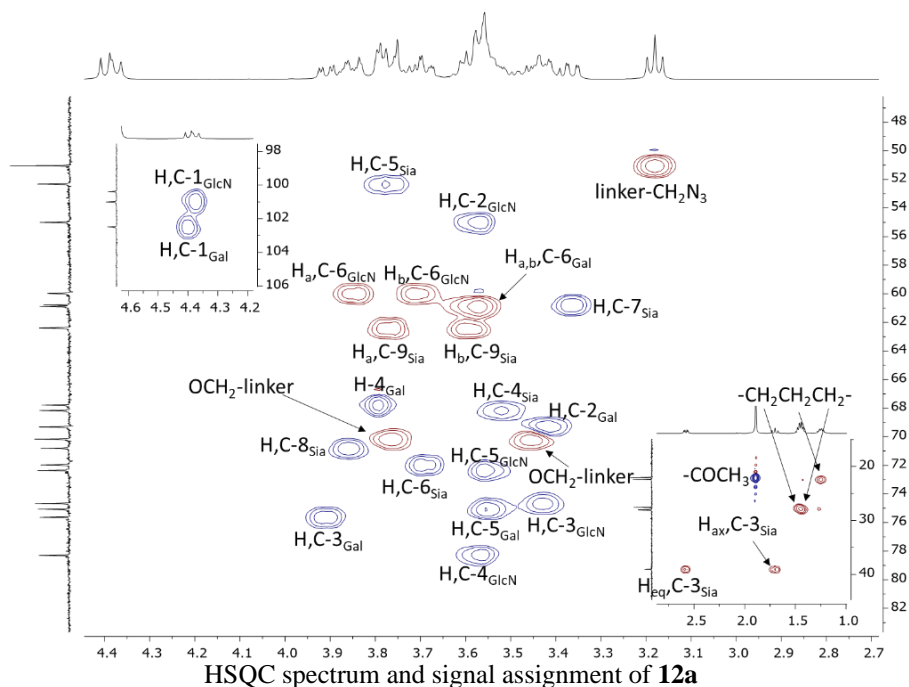


5-azidopentyl (5-acetamido-7-azido-3,5,7-trideoxy-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid)-(2 \rightarrow 3)- β -D-galactopyranosyl-(1 \rightarrow 4)-2-acetamido-2-deoxy- β -D-glucopyranoside (12a)



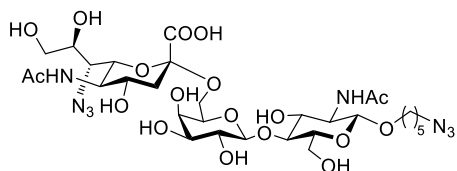
Compound **12a** was prepared from 4-azido ManNAc **9** and LacNAc **11** following the General procedure A, in which bacterial enzyme PMST1 was employed. Isolated yield: 77%.

^1H NMR (400 MHz, Deuterium Oxide) δ 4.56 (dd, J = 10.2, 7.7 Hz, 2H), 4.15 – 3.47 (m, 21H), 3.35 (t, J = 6.8 Hz, 2H), 2.75 (dd, J = 12.5, 4.6 Hz, 1H), 1.87 (t, J = 12.1 Hz, 1H), 1.61 (m, 4H), 1.50 – 1.32 (m, 2H). ^{13}C NMR (101 MHz, d_2O) δ = 174.41, 174.31, 173.43, 102.48, 101.00, 100.38, 78.30, 75.63, 75.07, 74.67, 72.36, 71.97, 70.81, 70.17, 69.31, 68.16, 67.78, 62.39, 60.89, 60.78, 59.98, 55.01, 52.32, 51.04, 39.04, 28.04, 27.56, 22.39, 22.06. See HSQC below for signal assignment. HRMS (neg.) calculated for $\text{C}_{30}\text{H}_{49}\text{N}_8\text{O}_{18}^-$ $[\text{M}-\text{H}]^-$ 809.3165, found 809.3154.



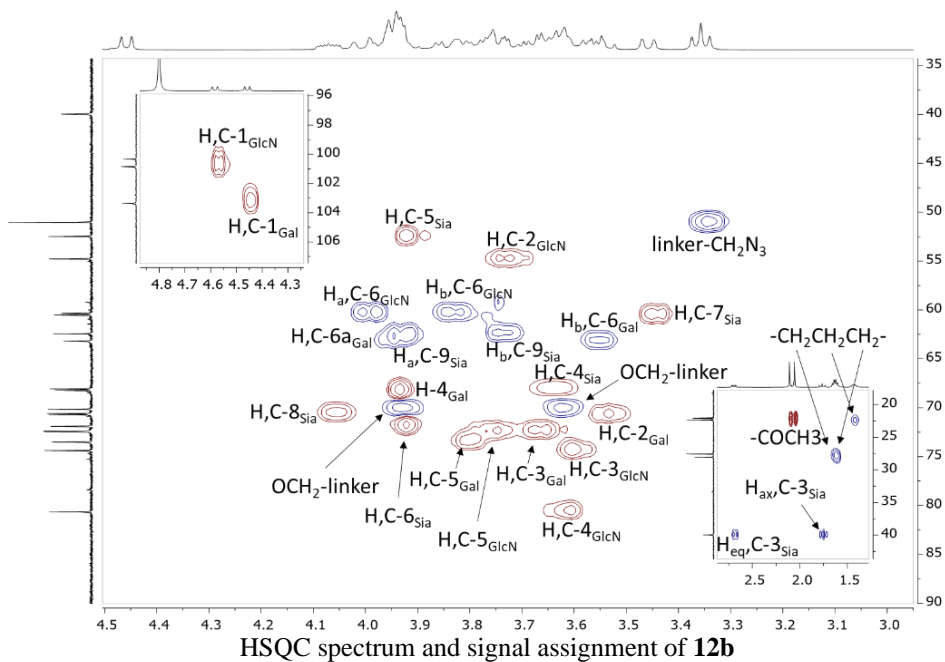
Chapter 4

5-azidopentyl (5-acetamido-7-azido-3,5,7-trideoxy-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid)-(2 \rightarrow 6)- β -D-galactopyranosyl-(1 \rightarrow 4)-2-acetamido-2-deoxy- β -D-glucopyranoside (**12b**)

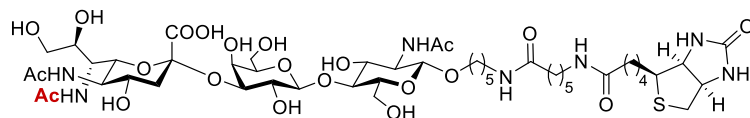


Compound **12b** was prepared from 4-azido ManNAc **9** and LacNAc **11** following the General procedure A, in which human enzyme ST6Gal1 was employed. Isolated yield: 67%. ^1H NMR (400 MHz, Deuterium Oxide) δ 4.58 (d, J = 7.9 Hz, 1H), 4.46 (d, J = 7.8 Hz, 1H), 4.12 – 3.50 (m, 22H), 3.50 – 3.42 (m, 1H), 3.36

(t, J = 6.8 Hz, 2H), 2.69 (dd, J = 12.5, 4.6 Hz, 1H), 2.11 (s, 3H), 2.06 (s, 3H), 1.76 (t, J = 12.2 Hz, 1H), 1.63 (m, 4H), 1.43 (m, 2H). ^{13}C NMR (101 MHz, d_2O) δ = 174.29, 174.24, 173.22, 103.36, 100.84, 100.33, 80.67, 74.39, 73.52, 72.45, 72.40, 71.91, 70.73, 70.61, 70.16, 68.24, 68.08, 63.19, 62.44, 60.52, 60.35, 54.78, 52.48, 51.04, 39.97, 28.06, 27.57, 22.40, 22.16, 22.07. See HSQC below for signal assignment. HRMS (neg.) calculated for $\text{C}_{30}\text{H}_{49}\text{N}_8\text{O}_{18}^-$ [$\text{M}-\text{H}$] $^-$ 809.3165, found 809.3160.

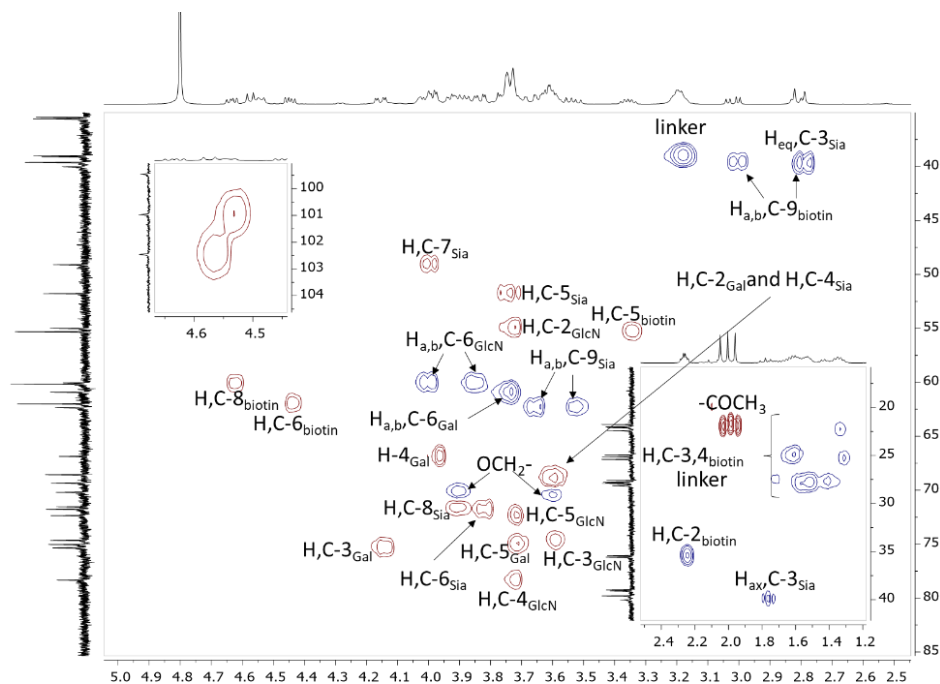


5-(6-(D-biotinamido)hexanamido)pentyl (5,7-bisacetamido-3,5,7-trideoxy-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid)-(2 \rightarrow 3)- β -D-galactopyranosyl-(1 \rightarrow 4)-2-acetamido-2-deoxy- β -D-glucopyranoside (14a)



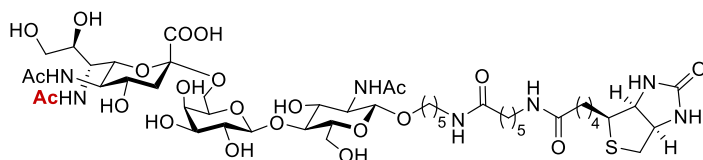
Compound **14a** was prepared from **12a** (8 mg, 10 μ mol) following the General procedure

B. Yield, 87% over 2 steps. ^1H NMR (400 MHz, d_2O) δ = 4.63 (dd, 1H, J = 4.8, 7.9 Hz), 4.57 (d, 1H, 7.9 Hz), 4.54 (d, 1H, 7.2 Hz), 4.45 (dd, 1H, J = 4.5, 8.0 Hz), 4.15 (dd, 1H, J = 3.0, 9.9 Hz), 4.03-3.51 (m, 21H), 3.36 (m, 1H), 3.21-3.17 (m, 4H), 3.02 (dd, 1H, 5.0, 13.1 Hz), 2.81 (dd, 1H), 2.80 (d, 1H, J = 12.8 Hz), 2.26 (m, 4H), 2.05 (s, 3H), 2.00 (s, 3H), 1.96 (s, 3H), 1.81-1.29 (m, 19H). ^{13}C NMR (101 MHz, d_2O) δ = 176.60, 176.49, 174.22, 173.91, 173.78, 173.73, 165.23, 102.46, 100.97, 99.46, 78.34, 75.35, 75.04, 74.65, 72.35, 71.77, 71.54, 70.20, 69.35, 68.57, 66.87, 62.34, 62.00, 60.92, 60.16, 55.31, 55.00, 51.78, 49.10, 40.02, 39.62, 39.10, 39.00, 35.58, 35.44, 28.14, 27.91, 27.86, 27.77, 27.59, 25.40, 25.14, 24.96, 22.42, 22.11, 22.02, 21.77. See HSQC below for signal assignment. HRMS (neg.) calculated for $\text{C}_{48}\text{H}_{80}\text{N}_7\text{O}_{22}\text{S}^-$ $[\text{M}-\text{H}]^-$ 1138.5077, found 1138.5112.

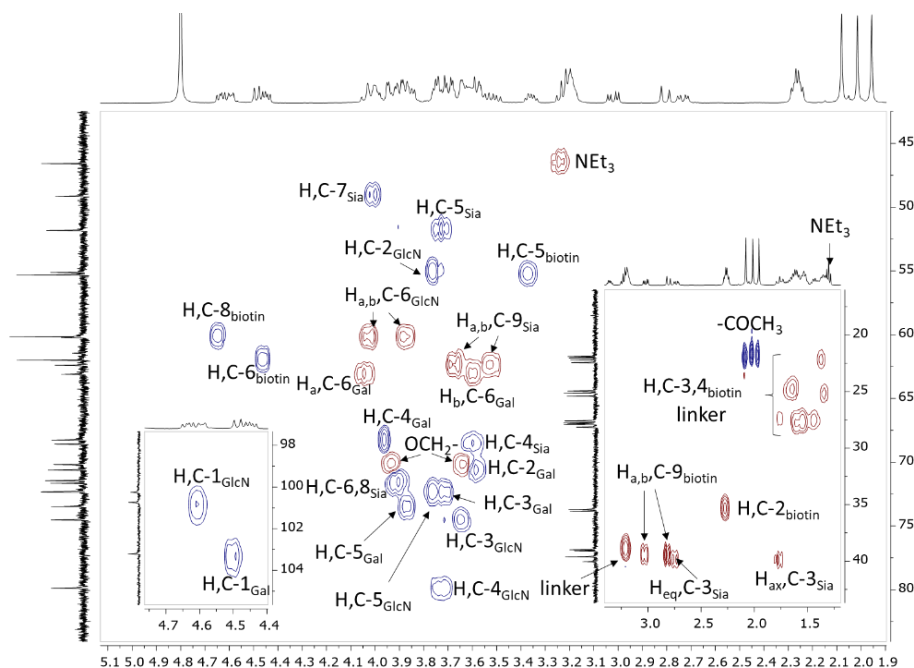


HSQC spectrum and signal assignment of **14a**

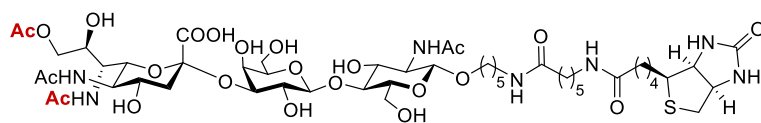
5-(6-(D-biotinamido)hexanamido)pentyl (5,7-bisacetamido-3,5,7-trideoxy-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid)-(2 \rightarrow 6)- β -D-galactopyranosyl-(1 \rightarrow 4)-2-acetamido-2-deoxy- β -D-glucopyranoside (14b)



MHz, d2o) δ = 4.63 (dd, 1H, J = 4.8, 7.8 Hz), 4.59 (d, 1H, J = 7.7 Hz), 4.49 (d, 1H, J = 7.9 Hz), 4.45 (dd, 1H, J = 4.4, 7.8 Hz), 4.06-3.48 (m, 22H), 3.36 (m, 1H), 3.25-3.20 (m, 4H), 3.02 (dd, 1H, J = 5.0, 12.7 Hz), 2.81 (d, 1H, J = 13.0 Hz), 2.73 (dd, 1H, J = 4.6, 12.5 Hz), 2.26 (m, 4H), 2.08 (s, 3H), 2.02 (s, 3H), 1.96 (s, 3H), 1.76-1.28 (m, 19H). ¹³C NMR (101 MHz, d2o) δ = 176.59, 176.49, 174.20, 173.67, 173.64, 173.28, 165.22, 103.21, 100.74, 100.25, 79.88, 74.53, 73.47, 72.32, 71.65, 71.45, 70.61, 70.19, 68.59, 68.26, 63.10, 62.40, 62.00, 60.16, 55.31, 55.12, 51.82, 49.15, 46.58, 40.05, 39.62, 39.10, 39.00, 35.58, 35.44, 28.15, 27.91, 27.87, 27.77, 27.60, 25.41, 25.14, 24.96, 22.44, 22.21, 22.03, 21.91. See HSQC below for signal assignment. HRMS (neg.) calculated for C₄₈H₈₀N₇O₂₂S⁻ [M-H]⁻ 1138.5077, found 1138.5099.

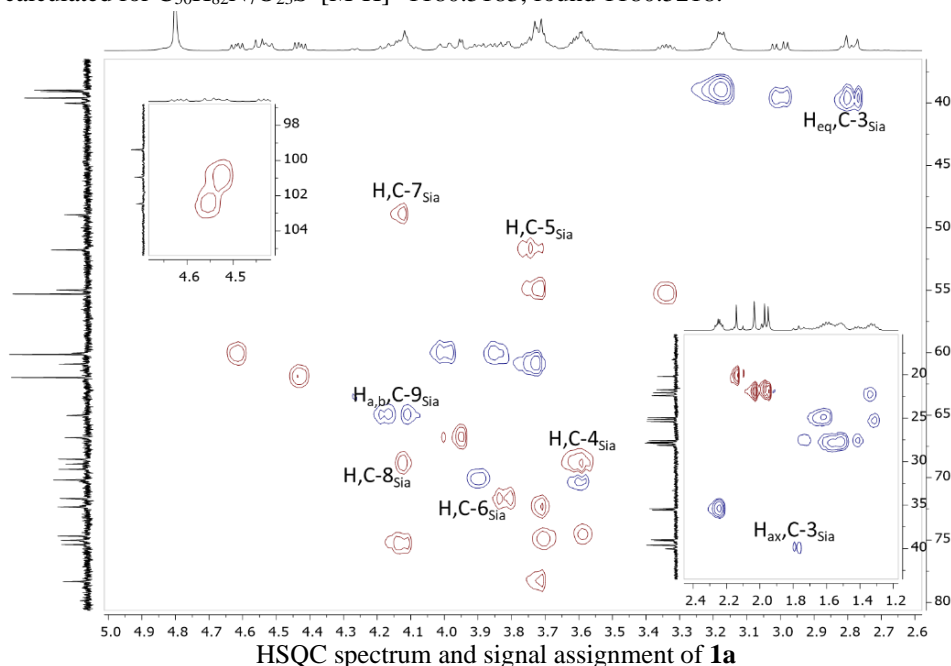
HSQC spectrum and signal assignment of **14b**

5-(6-(D-biotinamido)hexanamido)pentyl (5,7-bisacetamido-9-*O*-acetyl-3,5,7-trideoxy-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid)-(2 \rightarrow 3)- β -D-galactopyranosyl-(1 \rightarrow 4)-2-acetamido-2-deoxy- β -D-glucopyranoside (1a**)**



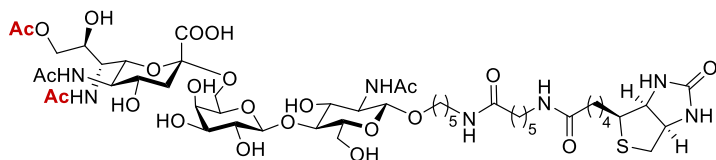
Compound **1a** was prepared from **14a** (13.3 mg, 15.7 μ mol) following the

General procedure C. HILIC-HPLC, 81% ACN, 0.1% formic acid. Yield, 72% (partial material loss on HPLC). ^1H NMR (400 MHz, d_2O) δ = 4.62 (dd, 1H, J = 4.8, 7.9 Hz), 4.55 (d, 1H, 7.8 Hz), 4.52 (d, 1H, 7.4 Hz), 4.43 (dd, 1H, J = 4.5, 7.9 Hz), 4.19-3.56 (m, 22H), 3.34 (m, 1H), 3.22-3.13 (m, 4H), 3.00 (dd, 1H, J = 5.0, 13.0 Hz), 2.80 (dd, 1H), 2.79 (d, 1H, J = 13.0 Hz), 2.25 (m, 4H), 2.14 (s, 3H), 2.03 (s, 3H), 1.97 (s, 3H), 1.95 (s, 3H), 1.80-1.28 (m, 19H). ^{13}C NMR (101 MHz, d_2O) δ = 176.58, 176.47, 174.20, 174.10, 173.95, 173.70, 173.48, 165.21, 102.45, 100.95, 99.39, 78.32, 75.36, 75.02, 74.68, 72.35, 71.67, 70.18, 69.33, 68.91, 68.52, 66.81, 65.02, 61.98, 60.90, 60.14, 55.30, 54.98, 51.76, 48.95, 40.02, 39.60, 39.08, 38.98, 35.56, 35.42, 28.11, 27.89, 27.84, 27.76, 27.58, 25.38, 25.12, 24.94, 22.40, 22.09, 22.04, 21.73, 20.20. See HSQC below for signal assignment. Due to marginal change in signals from other residues, only signals from Sia residue were assigned. HRMS (neg.) calculated for $\text{C}_{50}\text{H}_{82}\text{N}_7\text{O}_{23}\text{S}^-$ $[\text{M}-\text{H}]^-$ 1180.5183, found 1180.5218.



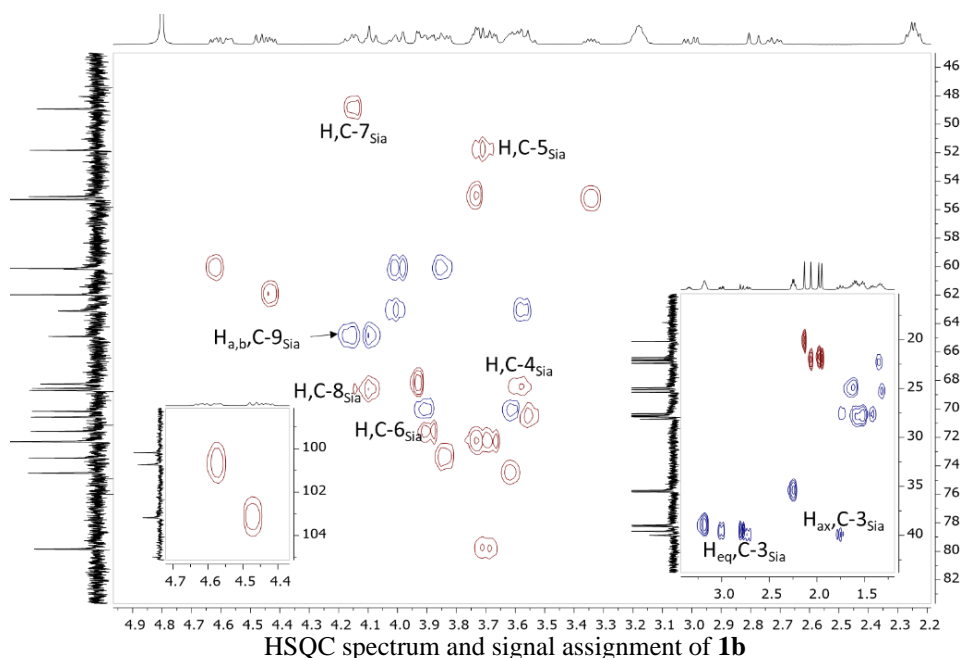
Chapter 4

5-(6-(D-biotinamido)hexanamido)pentyl (5,7-bisacetamido-9-*O*-acetyl-3,5,7-trideoxy-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid)-(2 \rightarrow 6)- β -D-galactopyranosyl-(1 \rightarrow 4)-2-acetamido-2-deoxy- β -D-glucopyranoside (**1b**)

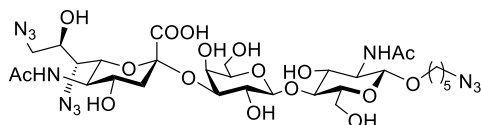


Compound **1b** was prepared from **14b** (12.3 mg, 14.5 μ mol) following the General procedure C. HILIC-HPLC, 81% ACN, 0.1%

formic acid. Yield, 67% (partial material loss on HPLC). ^1H NMR (400 MHz, d_2O) δ = 4.62 (dd, 1H, J = 5.1, 7.9 Hz), 4.57 (d, 1H, J = 7.2 Hz), 4.47 (d, 1H, J = 7.9 Hz), 4.43 (dd, 1H, J = 4.4, 7.8 Hz), 4.17-3.53 (m, 22H), 3.34 (m, 1H), 3.24-3.12 (m, 4H), 3.00 (dd, 1H, J = 5.0, 13.1 Hz), 2.79 (d, 1H, 13.0 Hz), 2.72 (dd, 1H, J = 4.4, 12.4 Hz), 2.25 (m, 4H), 2.13 (s, 3H), 2.06 (s, 3H), 1.98 (s, 3H), 1.95 (s, 3H), 1.79-1.28 (m, 19H). ^{13}C NMR (101 MHz, d_2O) δ = 176.59, 176.48, 174.18, 174.15, 173.72, 173.35, 173.16, 165.21, 103.18, 100.73, 100.18, 79.85, 74.51, 73.45, 72.30, 71.57, 70.57, 70.17, 68.70, 68.55, 68.25, 64.91, 63.12, 61.98, 60.14, 55.29, 55.08, 51.83, 48.93, 40.00, 39.60, 39.08, 38.98, 35.56, 35.42, 28.12, 27.89, 27.84, 27.75, 27.58, 25.38, 25.12, 24.94, 22.42, 22.19, 22.05, 21.85, 20.20. See HSQC below for signal assignment. Due to marginal change in signals from other residues, only signals from Sia residue were assigned. HRMS (neg.) calculated for $\text{C}_{50}\text{H}_{82}\text{N}_7\text{O}_{23}\text{S}^-$ [$\text{M}-\text{H}$] $^-$ 1180.5183, found 1180.5180.

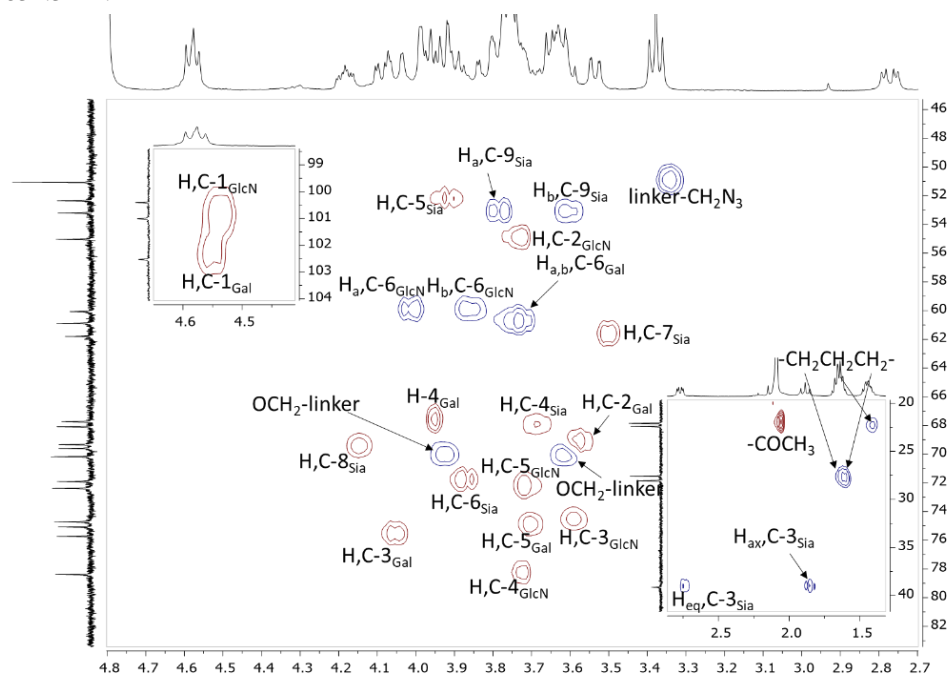


5-azidopentyl (5-acetamido-7,9-bisazido-3,5,7,9-tetradeoxy-D-glycero- α -D-galactonon-2-ulopyranosylonic acid)-(2 \rightarrow 3)- β -D-galactopyranosyl-(1 \rightarrow 4)-2-acetamido-2-deoxy- β -D-glucopyranoside (17a)



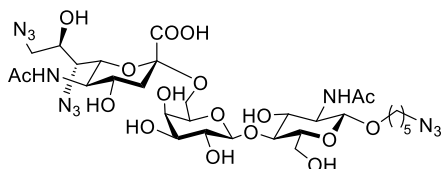
Compound **17a** was prepared from **12a** (21 mg, 26 μ mol) following the General procedure D. Yield, 73% over 3 steps. ^1H NMR (400 MHz, Deuterium Oxide) δ 4.58 (dd, J = 7.8, 5.7 Hz, 2H), 4.24 – 3.46 (m, 23H), 3.38 (t, J = 6.8 Hz, 2H), 2.77 (dd, J =

12.5, 4.6 Hz, 1H), 2.09 (d, J = 3.3 Hz, 6H), 1.89 (t, J = 12.2 Hz, 1H), 1.72 – 1.57 (m, 4H), 1.53 – 1.37 (m, 2H). ^{13}C NMR (101 MHz, d_2O) δ = 174.45, 174.31, 173.34, 102.53, 101.01, 100.41, 78.38, 75.73, 75.07, 74.72, 72.38, 71.91, 70.18, 69.61, 69.34, 68.09, 67.74, 61.80, 60.89, 60.07, 55.03, 53.20, 52.35, 51.06, 39.14, 28.06, 27.58, 22.41, 22.09. See HSQC below for signal assignment. HRMS (neg.) calculated for $\text{C}_{30}\text{H}_{48}\text{N}_{11}\text{O}_{17}^-$ $[\text{M}-\text{H}]^-$ 834.3230, found 834.3242.

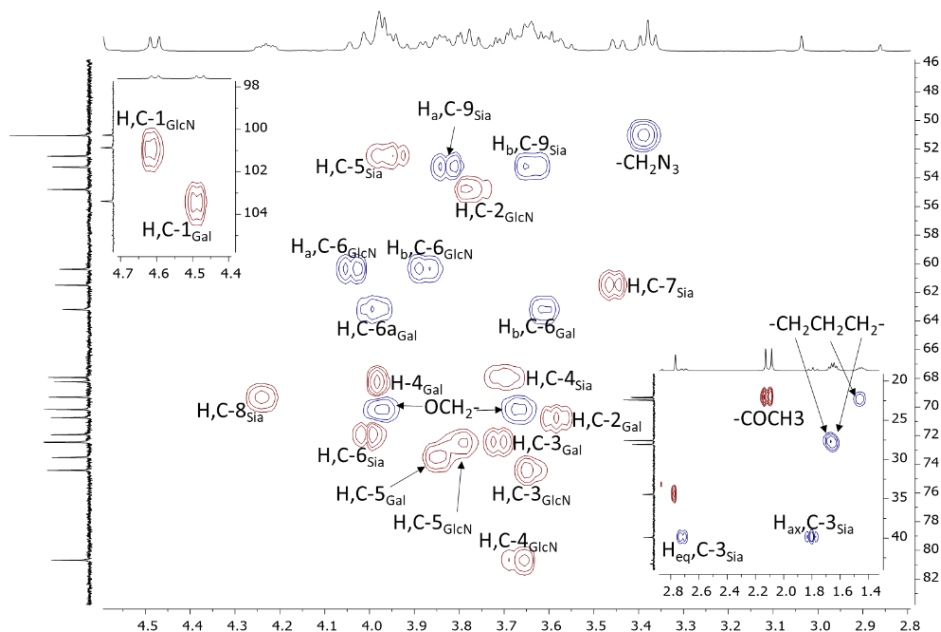


HSQC spectrum and signal assignment of **17a**

5-azidopentyl (5-acetamido-7,9-bisazido-3,5,7,9-tetradeoxy-D-glycero- α -D-galactonon-2-ulopyranosylonic acid)-(2 \rightarrow 6)- β -D-galactopyranosyl-(1 \rightarrow 4)-2-acetamido-2-deoxy- β -D-glucopyranoside (17b)

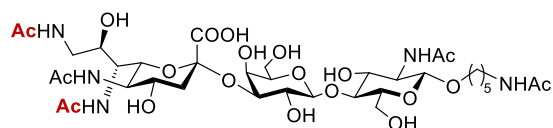


834.3230, found 834.3238.



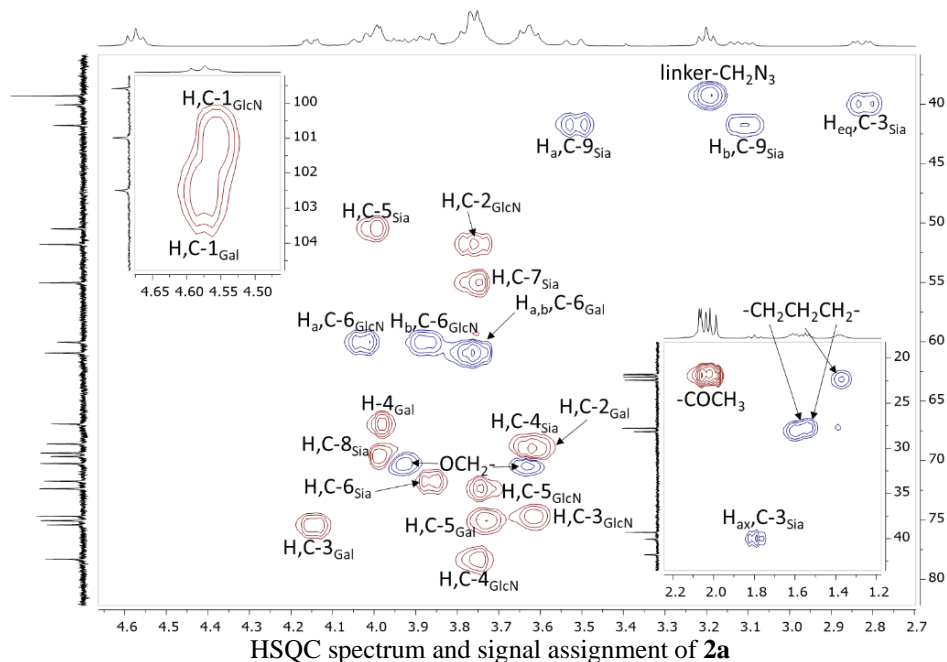
190

5-acetamidopentyl (5,7,9-trisacetamido-3,5,7,9-tetradecoxy-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid)-(2 \rightarrow 3)- β -D-galactopyranosyl-(1 \rightarrow 4)-2-acetamido-2-deoxy- β -D-glucopyranoside (2a)

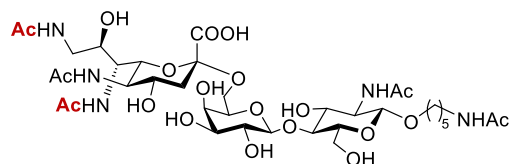


Compound **2a** was prepared from **17a** (11 mg, 13 μ mol) following the General procedure B, except for that the biotinylation step was skipped. Yield, 84% over 2 steps. ^1H NMR (400 MHz, Deuterium Oxide) δ 4.57

(m, 2H), 4.15 (dd, J = 9.9, 3.1 Hz, 1H), 4.10 – 3.57 (m, 19H), 3.52 (dd, J = 14.2, 2.2 Hz, 1H), 3.16 (dt, J = 30.3, 6.8 Hz, 3H), 2.83 (dd, J = 12.4, 4.5 Hz, 1H), 2.19 – 1.91 (m, 16H), 1.80 (t, J = 12.1 Hz, 1H), 1.58 (m, 4H), 1.38 (m, 2H). ^{13}C NMR (101 MHz, d_2O) δ = 174.28, 174.12, 173.94, 173.83, 173.75, 173.72, 102.49, 100.99, 99.58, 78.31, 75.42, 75.06, 74.69, 72.37, 71.74, 70.25, 69.66, 69.37, 68.60, 66.93, 60.95, 60.04, 55.03, 51.81, 50.50, 41.79, 40.05, 39.30, 28.15, 27.77, 22.44, 22.09, 21.86, 21.78. See HSQC below for signal assignment. HRMS (neg.) calculated for $\text{C}_{36}\text{H}_{60}\text{N}_5\text{O}_{20}^-$ $[\text{M}-\text{H}]^-$ 882.3832, found 882.3834.

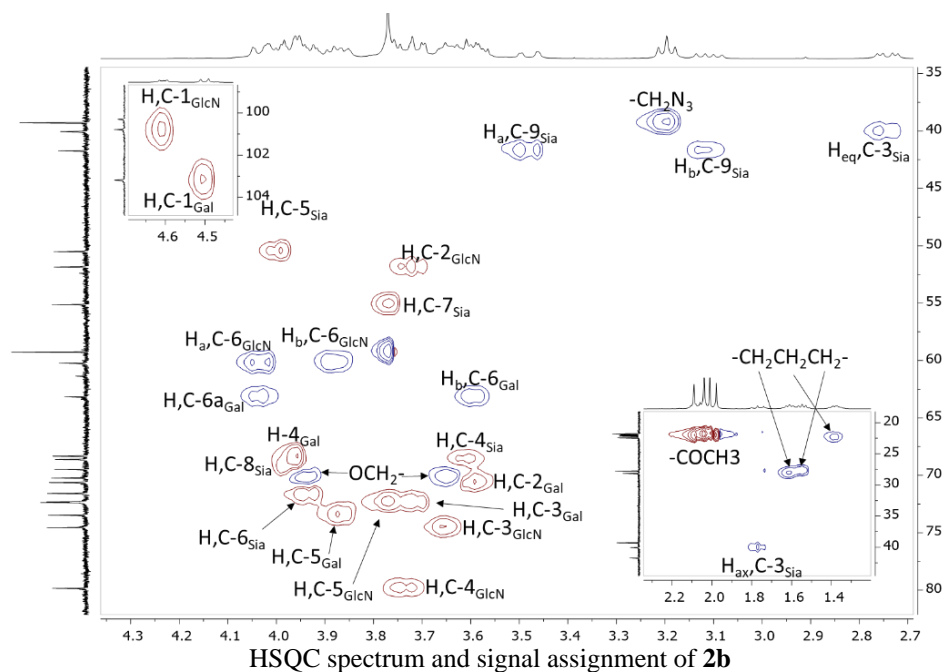


5-acetamidopentyl (5,7,9-trisacetamido-3,5,7,9-tetradecoxy-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid)-(2 \rightarrow 6)- β -D-galactopyranosyl-(1 \rightarrow 4)-2-acetamido-2-deoxy- β -D-glucopyranoside (2b)

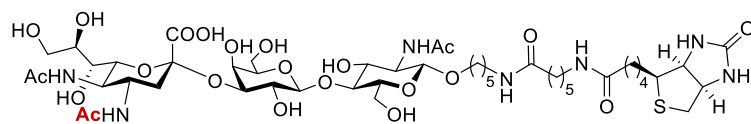


Compound **2b** was prepared from **17b** (8 mg, 9.6 μ mol) following the General procedure B, except for that the biotinylation step was skipped. Yield, 88% over 2 steps. ^1H NMR (400 MHz, Deuterium Oxide) δ 4.66 – 4.57 (m, 1H), 4.50 (d, J = 7.8 Hz, 1H), 4.14 – 3.52 (m,

23H), 3.48 (dd, $J = 14.1, 2.3$ Hz, 1H), 3.16 (dt, $J = 31.5, 6.9$ Hz, 3H), 2.74 (dd, $J = 12.5, 4.6$ Hz, 1H), 2.25 – 1.87 (m, 15H), 1.82 – 1.74 (m, 1H), 1.57 (m, 4H), 1.38 (m, 2H). ^{13}C NMR (101 MHz, d_2o) $\delta = 174.26, 174.16, 173.83, 173.73, 173.72, 173.61, 103.19, 100.79, 100.30, 79.85, 74.55, 73.47, 72.36, 72.32, 71.58, 70.63, 70.23, 69.49, 68.62, 68.33, 63.17, 60.23, 59.26, 55.13, 51.85, 50.51, 41.74, 40.06, 39.29, 28.16, 27.77, 22.45, 22.19, 22.07, 21.97, 21.84, 21.77$. See HSQC below for signal assignment. HRMS (neg.) calculated for $\text{C}_{36}\text{H}_{60}\text{N}_5\text{O}_{20}^-$ [M-H] $^-$: 882.3832, found 882.3869.

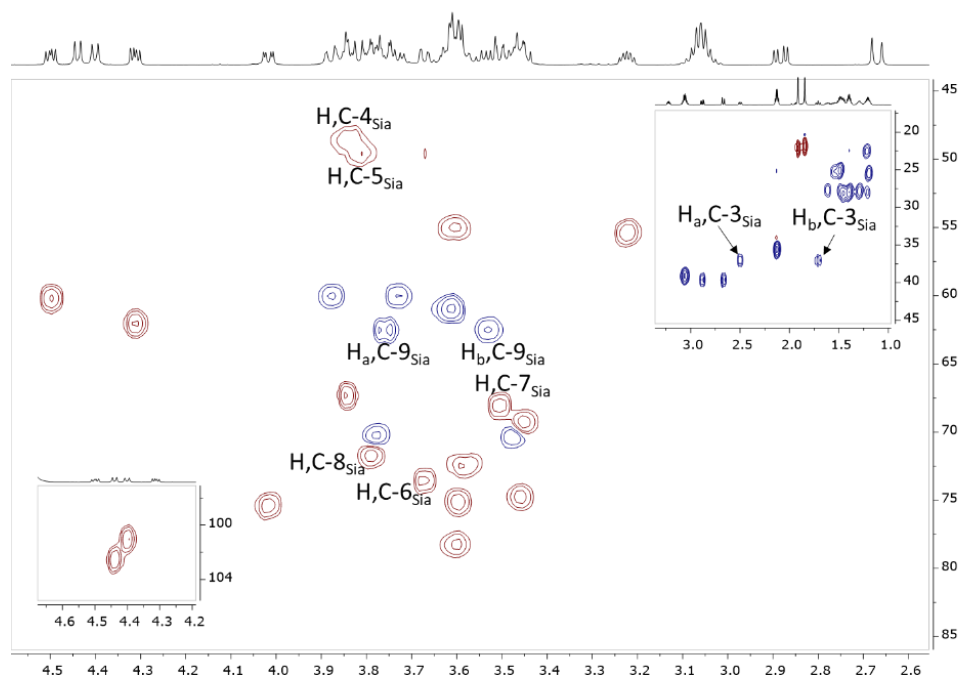


5-(6-(D-biotinamido)hexanamido)pentyl (4,5-bisacetamido-3,4,5-trideoxy-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid)-(2 \rightarrow 3)- β -D-galactopyranosyl-(1 \rightarrow 4)-2-acetamido-2-deoxy- β -D-glucopyranoside (3a)



Compound **3a** was prepared by treating **21a** with 4 eq. acetic anhydride in the

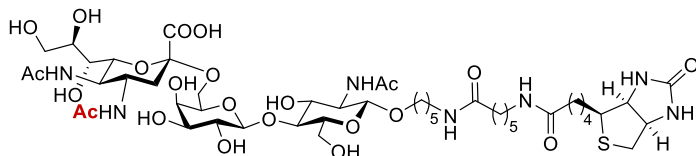
presence of 8 eq. triethylamine in DMF/water (4:1 v/v), followed by adding water and lyophilization. The residue was loaded onto C18 RP-HPLC for purification. Yield, 71% from **20**. Compound **21a** was obtained by coupling 4-amino Neu5Ac **18** with LacNAc-Lc-biotin **20** as described the General procedure A, in which bacterial enzyme PMST1 was employed. Biogel p2-purified **21a** was present in a mixture with CMP-4-amino-Neu5Ac, and was directly subjected to *N*-acetylation to afford **3a**. See HSQC below for signal assignment. HRMS (neg.) calculated for $C_{48}H_{80}N_7O_{22}S^-$ $[M-H]^-$ 1138.5077, found 1138.5084.



HSQC spectrum and signal assignment of **3a**

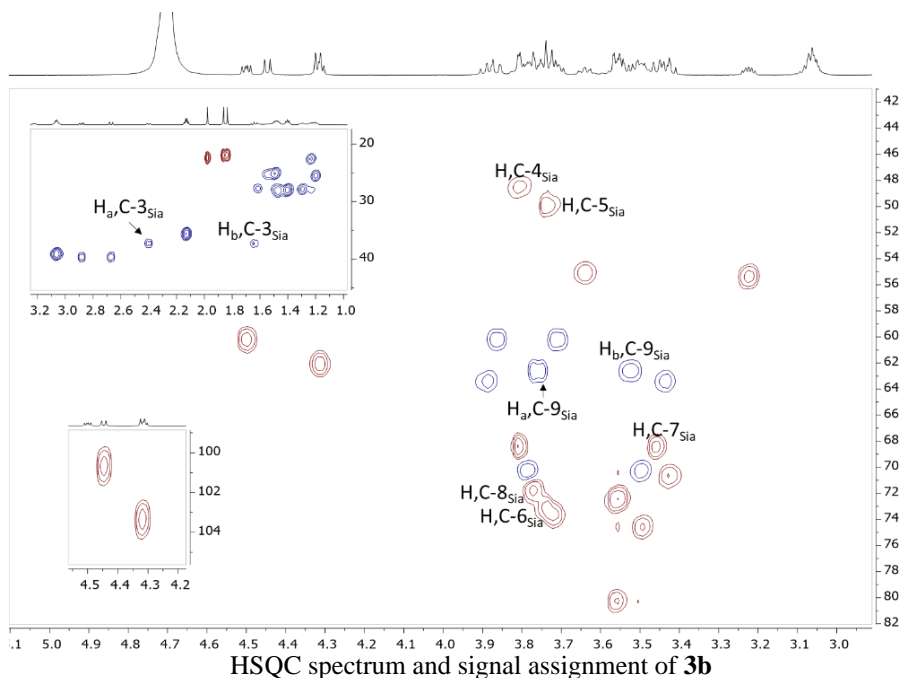
Chapter 4

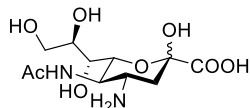
5-(6-(D-biotinamido)hexanamido)pentyl (4,5-bisacetamido-3,4,5-trideoxy-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid)-(2 \rightarrow 6)- β -D-galactopyranosyl-(1 \rightarrow 4)-2-acetamido-2-deoxy- β -D-glucopyranoside (3b**)**



Compound **3b** was prepared following a likewise procedure as describe for **3a**, except for that the sialyltransferase employed was the

PMST1 mutant with α 2,6-transferring activity. Yield, 52% from **20**. Neither ST6Gal1 nor the other bacterial α 2,6-sialyltransferase Pd2,6ST worked. See HSQC below for signal assignment. HRMS (neg.) calculated for $C_{48}H_{80}N_7O_{22}S^-$ $[M-H]^-$ 1138.5077, found 1138.5114.



5-acetamido-4-amino-3,4,5-trideoxy-D-glycero- α -D-galacto-non-2-ulopyranosylonic acid (18**)**

Compound **18** was synthesized following what we reported in Ref. 23 in the main text with modification. **S1** (320 mg, 0.7 mmol) was dissolved in 7 mL acetonitrile/water (9:1 v/v) and was cooled to 0 °C. *N*-bromosuccinimide (3 eq. 374 mg, 2.1 mmol) was added.

The reaction was allowed to warm up to room temperature and stirred for 4 hours. The mixture was diluted with ethyl acetate and washed with 5% sodium thiosulfate and then brine. The organic phase was dried over Na₂SO₄ and was then filtered and concentrated in vacuo. The residue was passed through a short pad of silica gel (petroleum ether:ethyl acetate 3:1, then elute with acetonitrile straight) to give **S2** in 73% yield. This rough purification step could be skipped, which would result in lower yield in the next steps.

Compound **S2** (120 mg, 0.2 mmol) was taken up in anhydrous 4 mL dioxane in a dry round bottom flask equipped with argon balloon. The flask was pre-filled with argon. Tributyltin hydride (4 eq., 232 mg, 0.8 mmol, 215 μ L) was added. The mixture was heated up to 80 °C, at which catalytic (0.1 eq.) AIBN in dioxane (0.5 M) was added dropwise. The reaction was stirred vigorously for 3 hours. After cooling down to room temperature, the solution was concentrated in vacuo and was then taken up with acetonitrile and washed with hexane for 3 times to remove organotin reagent. The acetonitrile phase was concentrated and loaded on to silica gel column (ethyl acetate straight, then ethyl acetate:methanol 4:1 – 2:1). The resulted amine compound was then hydrolysed with NaOH in water/methanol. The excessive base was neutralized with 1M HCl. The mixture was concentrated in vacuo and then loaded onto biogel p2 for desalting, providing **18** in 52% yield over 2 steps. Due to inseparable impurities, the compound is directly subjected to enzymatic reactions, and was characterized at the stage of **3a** and **b**.

Chapter 4

4. Additional references

1. Moremen, K. W., Ramiah, A., Stuart, M., Steel, J., Meng, L., Forouhar, F., *et al.* Expression system for structural and functional studies of human glycosylation enzymes. *Nat. Chem. Biol.* 2018, **14**(2): 156-162.
2. Liu, Y., McBride, R., Stoll, M., Palma, A. S., Silva, L., Agravat, S., *et al.* The minimum information required for a glycomics experiment (MIRAGE) project: improving the standards for reporting glycan microarray-based data. *Glycobiology* 2017, **27**(4): 280-284.

CHAPTER 5 | Conjugation of a Toll-like Receptor Agonist to Glycans Preserves Neutralization Epitopes of HIV Native-like Envelope Trimer

Zeshi Li¹, Ronald Derking², Wen-Hsin Lee³, Gerlof P. Bosman¹, Andrew B. Ward³, Rogier W. Sanders^{2,4}, Geert-Jan Boons^{1,5,6,7,*}

Affiliations

¹Department of Chemical Biology and Drug Discovery, Utrecht Institute for Pharmaceutical Sciences, Utrecht University, 3584 CG Utrecht, The Netherlands. E-mail: g.j.p.h.boons@uu.nl

²Department of Medical Microbiology, Amsterdam UMC, University of Amsterdam, Amsterdam Institute for Infection and Immunity, 1105AZ Amsterdam, The Netherlands.

³Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA, 92037, USA.

⁴Department of Microbiology and Immunology, Weill Medical College of Cornell University, New York, NY 10021, USA.

⁵Complex Carbohydrate Research Center, University of Georgia, Athens, GA 30602, USA. E-mail: gjboons@ccrc.uga.edu

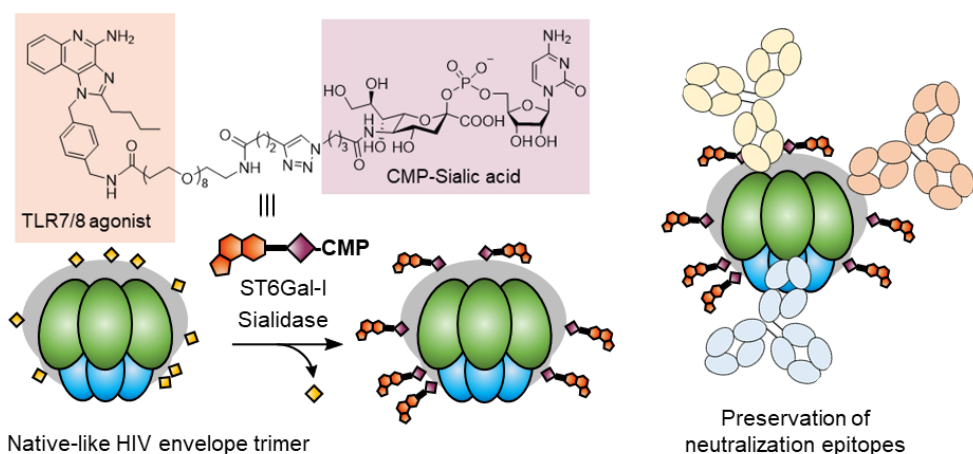
⁶Bijvoet Center for Biomolecular Research, Utrecht University, Utrecht, The Netherlands.

⁷Chemistry Department, University of Georgia, Athens, GA 30602, USA.

Chapter 5

Abstract

Induction of broadly and durably protective response has been the ultimate goal of prophylactic vaccine development. The use of Toll-like receptor agonists has been increasingly pursued to augment such immune responses. Attachment of these immunomodulating agents to immunogens offers several advantages over combined administration thereof as separate entities. However, there is a lack of general method to conjugate TLR agonists to complex immunogens, such as viral glycoproteins, without compromising important neutralization epitopes. Herein we describe a one-step enzymatic conjugation approach which allows for covalent attachment of small molecule immune activators to asparagine (*N*)-linked glycans of a native-like HIV envelope trimer immunogen. The drug-to-protein ratio can be readily quantified chromatographically. Binding studies with several broadly neutralizing antibodies and electron microscopic characterization reveal that the glycan-targeted covalent modification of the glycoprotein immunogen minimally affects its neutralization epitopes.



Introduction

Subunit vaccines, in which only one or few microbial component(s) are administered, have greatly contributed to vaccine safety. For example, microbial toxins generated through recombinant DNA technology and altered to reduce toxicity, are successfully used as vaccines for a number of pathogens. In addition, polysaccharide–protein conjugate vaccines have been developed as vaccines for *H. influenzae*, *N. meningitidis* and *S. pneumoniae*.^{1–3}

Although the subunit approach has many attractive features, it comes at the expense of decreased immunogenicity. This limitation is particularly problematic for immuno-compromised patients and the elderly who often suffer from immuno-senescence leading to decreased immune response to vaccination. In addition, it has been difficult to develop vaccines for diseases such as human immunodeficiency virus infection and acquired immunodeficiency syndrome (HIV/AIDS), malaria, *Mycobacterium tuberculosis*, *Plasmodium falciparum*, and hepatitis C virus (HCV). There are no licensed vaccines for nosocomial bacterial infections, such as *Pseudomonas aeruginosa*, *E. coli* and *Staphylococcus aureus*. It is clear new vaccine paradigms are needed to allow for more potently activating the immune system.

Tremendous progress has been made in understanding the molecular mechanisms that control immune cell activation.^{4,5} For example, receptors of the innate immune system have been identified that can recognize pathogen-associated molecular patterns (PAMPs), which are generally conserved components of pathogens, such as nucleic acids, cell wall components, and flagellin that are not produced by humans. Recognition of PAMPs leads to activation of the innate immune system, which in addition to providing early protection, also shapes adaptive immune responses. This insight has led to the rationale development of adjuvants that can potentiate immune responses induced by subunit vaccines.^{6–8}

The conjugation of PAMPs to an antigen of interest provides an attractive means to increase vaccine potency. In this approach, antigen and immune-potentiator are delivered to the same antigen presenting cell resulting in more efficient immune activation. It makes it possible to reduce the adjuvant and antigen dose, thereby minimizing the risk of adverse effects. We and others pioneered this approach and demonstrated that a glycosylated cell surface associated mucin 1 (MUC1)-derived glycopeptide covalently linked to a Toll-like receptor (TLR) agonist can elicit potent humoral and cellular immune responses and is efficacious in reversing tolerance and generating a therapeutic response.⁹ Synthetic approaches make it possible to attach various small molecule adjuvants such as TLR2 and TLR7/8 agonists to synthetic antigens.¹⁰ Furthermore, recombinant DNA technologies have been used to make fusion proteins of antigen and a protein-based immune-potentiator such as flagellin,¹¹ which is a TLR5 agonist, and HSP70¹² and type-III repeat extra domain A from fibronectin (EDA)¹³, which are TLR4 ligands.

Modification of protein antigens with small molecule immune potentiators have been challenging, and no robust methods have been reported yet. Small molecule TLR agonists are a fast-growing class of immune potentiators that are being explored as vaccine adjuvants.¹⁴ Thus robust methods need to be developed to attach such compounds to protein antigens without compromising protein and antigenic integrity. UV cross-linking has been employed to link a TLR7/8 agonist (TLR7/8a) to the HIV-1 gag protein, which resulted in improved T-cell immunity.¹⁵ In another study, a TLR7 ligand was attached to the model protein carrier, mouse serum albumin, through a 2-step procedure involving NHS activation of the protein followed by hydrazone mediated ligation.¹⁶ The resulting conjugate induced more potent *in-vitro* and *in-vivo* cytokine production. A TLR9 agonist (CpG) was attached to nitrophenol-modified chicken gamma globulin as a model immunogen, which led to a

Chapter 5

higher number of follicular T helper cells in germinal centers and improved class-switching, affinity maturation and memory responses.¹⁷ In the study above, nitrophenol-modified chicken gamma globulin was biotinylated with biotin-sulfosuccinimidyl ester which was mixed with biotin-CpG1826 and then conjugated to streptavidin.

Although these studies have demonstrated the attractiveness of attaching a small molecule agonist to a protein antigen, the employed conjugation approaches are not suitable for many complex glycoprotein antigens. In this respect, conventional protein modification often exploits the nucleophilicity of lysine, arginine or cysteine side chains (Fig. 1a), and modification of these residues can compromise the integrity of important epitopes. For example, it has been shown that a two-step conjugation of a TLR7/8a to HIV-1 gp120 *via* lysine/arginine led to substantially decreased binding by broadly neutralizing antibodies (bNAbs).¹⁸ Furthermore, these conjugation approaches often involve non-physiological pH and/or use of organic solvents, which can cause incompatibilities with sensitive antigens.

Herein, we describe a one-step enzymatic conjugation approach for the covalent attachment of small molecule immune activators to asparagine-linked glycans of viral glycoprotein immunogens. It involves pre-attachment of a small molecule TLR7/8a to an azido-modified β -linked cytidine-5'-monophosphoryl sialic acid (CMP-Neu5AcN₃) derivative, which can then be transferred to *N*-glycans of a viral glycoprotein by human

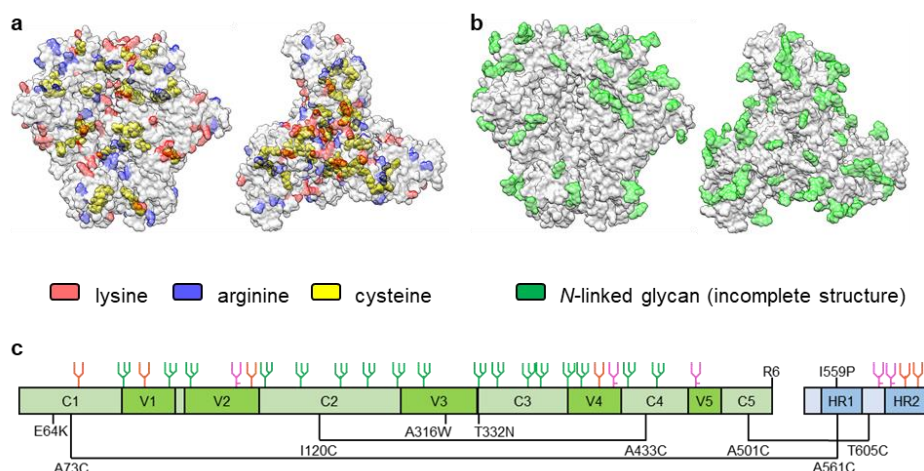


Figure 1 | Potential sites for modification on a native-like HIV envelope trimer. **a**, Lysine (red), arginine (blue) and cysteine (yellow) residues highlighted in the SOSIP native-like Env trimer (PDB 6V0R). Side view is on the left and top view is on the right. Protein surface is shown at 70% transparency for cysteine residues (spherical) to be visible. Glycan structures are omitted. **b**, *N*-linked glycans (incomplete structure, green, surface representation) present on the same protein. **c**, Linear schematic of native-like Env trimer BG505 SOSIP v5.2¹⁹ used in this study. Gp120 is shaded in green and gp41 in blue. Solid lines linking two cysteine residues indicate disulfide bonds. Glycan sites are color-coded based on the glycoforms identified in site-specific glycomic analysis.²⁰⁻²² Sites shown in green are dominated by oligomannose-type glycans (100 - 80% oligomannose), in purple by hybrid-/complex-type glycans (39 - 0% oligomannose); the sites where both types are present in similar abundance (79 - 40% oligomannose) are shown in orange.

sialyltransferase ST6Gal-I. We demonstrate that the method can modify a fully-glycosylated, native-like HIV envelope (Env) trimer without loss of recognition by bNAbs.

Mature native HIV-1 Env spike proteins form a trimer of heterodimers (gp120 and gp41), which is rather unstable and readily dissociates. The intact trimeric state of the spike is critical for presenting neutralizing epitopes and reducing the exposure of non-neutralizing epitopes which may lead to non-relevant immune responses.²³ An attractive strategy to stabilize the hexameric state of HIV Env spike, which is termed “SOSIP” (Fig. 1c), is based on introducing several inter-subunit disulfide bonds between Cys⁵⁰¹ and Cys⁶⁰⁵, an Ile⁵⁵⁹Pro mutation to stabilize the protein complex in a prefusion state, and an improved furin (R6) cleavage site to facilitate maturation.²³⁻²⁶ Furthermore, additional mutations and disulfide bonds can be introduced to further stabilize the complex.¹⁹ SOSIP trimers are attractive vaccine candidates and we aimed to modify such a glycoprotein with a TLR7/8 agonist.

The HIV-1 Env trimer contains up to 90 *N*-linked glycans making up greater than half of its molecular weight (Fig. 1b and c). The *N*-glycan sites of the gp120 portion carries largely under-processed oligomannosides, whereas the gp41 portion harbors a higher fraction of Golgi-processed complex- and hybrid-type *N*-glycans. We envision that the sialic acid residues at the termini of the latter two glycan types may serve as attachment sites for desired functionalities. These glycan termini are at a relatively large distance from the protein surface, and their modification was expected to only marginally effect Env recognition by most bNAbs. An increasing number of bNAbs against multiple viral variants have been discovered which are valuable for guiding vaccine design.²⁷⁻²⁹ However, induction of such antibodies *via* immunization for broad and durable protection has not been realized yet.

Previously, we demonstrated that the human sialyltransferase, ST6Gal-I, tolerates modifications at the C5-position of CMP-Neu5Ac derivatives.³⁰ Thus, it was anticipated that CMP-Neu5Ac derivative **1** modified at C-5 with a TLR7/8a can be transferred by ST6Gal I to *N*-glycans having terminal galactosyl residues. Furthermore, we have observed that a sialoside modified at C5 is resistant to *C. perfringens* sialidase treatment. Thus, removal of endogenous sialic acid present on a glycoprotein and the installation of a C5-modified counterpart can be performed as a one-pot procedure (Fig. 2).

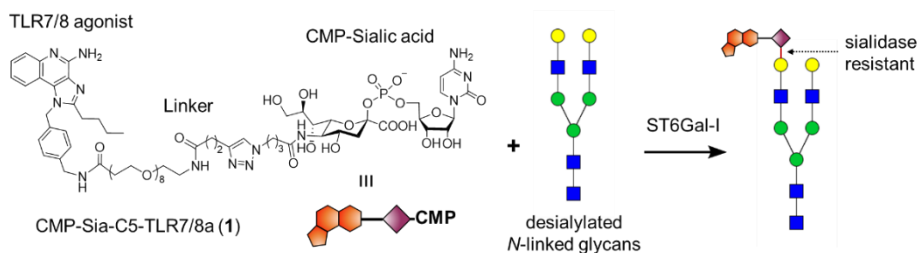


Figure 2 | Strategy for protein modification. The conjugating agent **1** is composed of three parts: (1) a TLR7/8 agonist, (2) an oligo(ethylene glycol) spacer to increase aqueous solubility, and (3) a CMP-activated carrier sialic acid to be transferred by ST6Gal-I. Once attached to an *N*-glycan, the glycosidic bond is insensitive to sialidase degradation.

Results

CMP-Neu5Ac derivative **1** was prepared by copper-catalyzed click chemistry of the alkyne modified TLR7/8a **4** with CMP5Neu5Ac derivative **5** having an azido-moiety at C-5.

Chapter 5

Thus, the TLR7/8a-alkyne portion (**4**), was prepared by acylation of the benzylic amine of TLR7/8a agonist **2** with *N*-hydroxysuccinimide-activated, oligo(ethylene glycol)-spaced 4-pentynoyl moiety **3**. The oligo-ethylene glycol spacer was incorporated to increase water solubility. It has been reported that modification of the TLR7/8a **2** at the benzylic amine position with various functionalities does not compromise activity,^{31, 32} making this drug molecule an appropriate candidate for bioconjugation. The TLR7/8a-alkyne **4** was attached to CMP-Neu5AcN₃ (**5**) using *in-situ* generated Cu(I) from CuSO₄ and sodium ascorbate, and THPTA as an aqueously soluble ligand to give **1** and **7**, respectively.³³ The CMP-Sia-linked TLR/8a **1** is completely water soluble, which eliminates the need for organic co-solvent at the glycoprotein conjugation step. We also attempted to prepare compound **1** by condensation of **7** with CTP in the presence of *N. meningitidis* CMP-Sia synthetase. However, the later enzyme does not tolerate large modifications at C-5 of Neu5Ac and no product formation was observed.

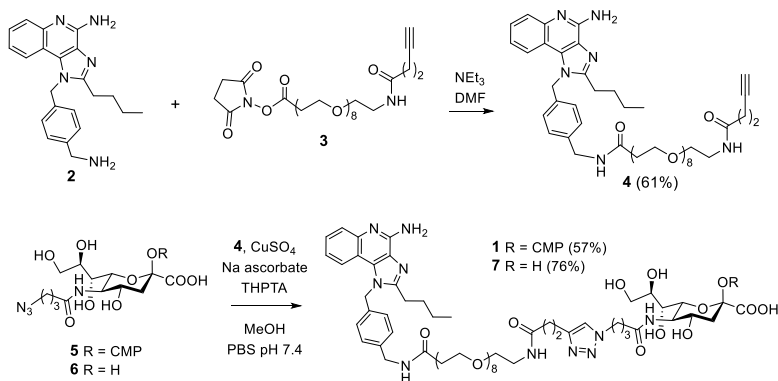


Figure 3 | Preparation of conjugating agent **1 and non-CMP-activated **7**.** Abbreviations: NEt₃, triethylamine; DMF, *N,N*-dimethylformamide; THPTA, tris(3-hydroxypropyl)triazolymethylamine; MeOH, methanol; PBS, phosphate buffer saline.

Next, attention was focused on the modification of the native-like BG505 SOSIP v5.2 Env trimer¹⁹ with compound **1**. Thus, the trimer was incubated with **1** in the presence of the sialidase from *C. perfringens* and ST6Gal-I for 18 h. The trimer has a substantially larger molecular weight compared to ST6GalI and sialidase, and as a result could readily be purified by Superdex 200 size-exclusion chromatography to give pure chimeric SOSIP trimer-TLR7/8a conjugate (SOSIP-TLR7/8a) as determined by the chromatogram and native gel electrophoresis (Fig. S1a and b). The conjugated protein carrying TLR7/8a could not be quantified by nanodrop due to interference of UV absorption by the drug. Instead, a bicinchoninic acid (BCA) assay was used to determine the protein concentration (Fig. S1c). Native gel electrophoresis, which showed SOSIP-TLR7/8a remained intact post-modification, does not have the resolving power to differentiate the TLR7/8a-conjugated and non-modified protein, due to the relatively small difference in molecular weight. To confirm successful modification, the SOSIP-TLR7/8a conjugate and the non-modified counterpart was treated with 50 mM trifluoroacetic acid (TFA) at 80°C to release sialic acids (Fig. 4c). Control experiments with pure water heated at the same temperature were also included. The resulting mixtures were subjected to high performance liquid chromatography coupled with a UV detector and a mass spectrometry (HPLC-UV-MS). Only TFA-treated SOSIP-TLR7/8a gave a signal in the chromatogram with a maximum absorption at 320 nm, and with a mass

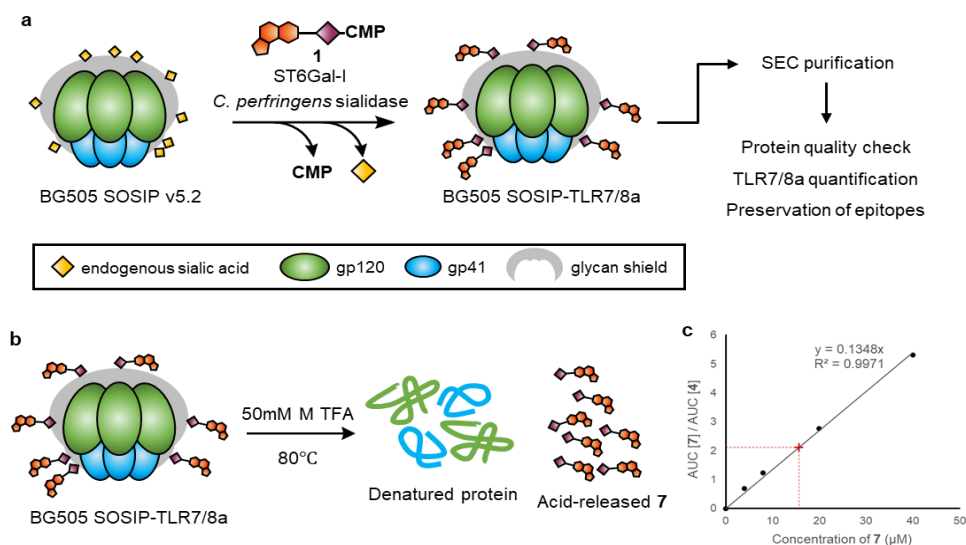


Figure 4 | SOSIP-trimer-drug conjugation and characterization. **a**, One-step de-sialylation and re-sialylation using sialyltransferase ST6Gal-I and sialidase from *C. perfringens*. Post-modification mixture was purified by size-exclusion chromatography (SEC, see Supplementary Fig. S1 for chromatogram). An aliquot of the purified protein was taken for BCA analysis to determine the concentration. **b**, Reference-free 2D class averages of negative stain electron microscopy images of purified SOSIP-TLR7/8a. A total of 11,134 particles were analyzed, all of which are in the native-like trimer form. **c**, Trifluoroacetic acid (TFA)-mediated sialic acid cleavage for quantification. Released sialic acid-TLR7/8a was subjected HPLC-UV-MS analysis. **d**, Standard curve of released sialic acid-TLR7/8a species 7. See Supporting Information for LC chromatograms of compound 7 at different concentrations. The red cross sign on the fitted trendline indicates observed normalized (against 4) area under curve (AUC) values for sialic acid-TLR7/8a released from BG505 SOSIP-TLR7/8a protein.

of the corresponding HPLC peak matching exactly that of chemically synthesized sialic acid-TLR7/8a 7 (Fig. S2).

To determine TLR7/8a-to-Env-trimer ratio, we used chemically synthesized compound 7, which has the same structure as the TFA-cleaved sialic acid-TLR7/8a species, to make a standard curve. Solutions of 7 in water at varying concentrations (40 to 4 μM) were prepared and subjected to HPLC-UV analysis. Compound 4 was used at a concentration of 5 μM as an internal standard, because its retention time on HPLC differs considerably from 7, yet having the same maximum absorption wavelength at 320 nm. The standard curve was obtained by plotting the normalized peak area of compound 7 in the chromatogram against its concentration, which made it possible to get the average number of the attached drug molecules after TFA-mediated sialic acid release from SOSIP-TLR7/8a (Fig. 4d). The average attachment number was 13 (see Supplementary Fig. S3 for calculation), which falls well within the theoretical numbers of total complex *N*-glycan sites on an Env trimer (see Supplementary Text for calculation).

Next, ELISA was used to examine if the protein modification preserves critical neutralization epitopes on the Env trimer (Fig. 5). Thus, the TLR7/8a-conjugate and non-modified trimers carrying His-tag were immobilized on Ni-NTA-coated plate. Primary bNABs were added following a blocking step. BNABs targeting multiple classes of

Chapter 5

neutralization epitopes, including ones that bind glycans were investigated and binding was detected using a goat-anti-human secondary antibody conjugated with HRP. As shown in panel **a**, the CD4 binding site bNAb VRC01³⁴ was not affected by the modification. 2G12³⁵ (panel **b**) and PGT128³¹ (panel **c**) recognize Asn332-centered high-mannose glycans. The former binds glycans exclusively, whereas the latter also interacts with the V3 loop in addition to high-mannose glycans. Neither of the bNAbs were affected by TLR7/8a conjugation. In the cases of trimer apex bNAb PGT145^{36, 37} (panel **d**) and gp120-gp41 interface bNAb PGT151³⁸ (panel **e**), potency was decreased by a modest 2-3 folds, indicating the modification was well-tolerated by these bNAbs.

Importantly, the modification of a complex immunogen like the HIV-1 Env trimer should preserve its structural integrity. We used negative stain electron microscopy (NS-EM) to characterize the conjugated product. We found after TLR7/8a conjugation, the Env trimers remained native-like without detectable malformed trimers (Fig. 5f). This is in consistency with the remained binding by PGT145, as the trimeric state of Env is necessary. Taken together, these results indicate the conjugation method is compatible with highly complex glycoprotein immunogens and is able to maintain the structural integrity and neutralization epitopes.

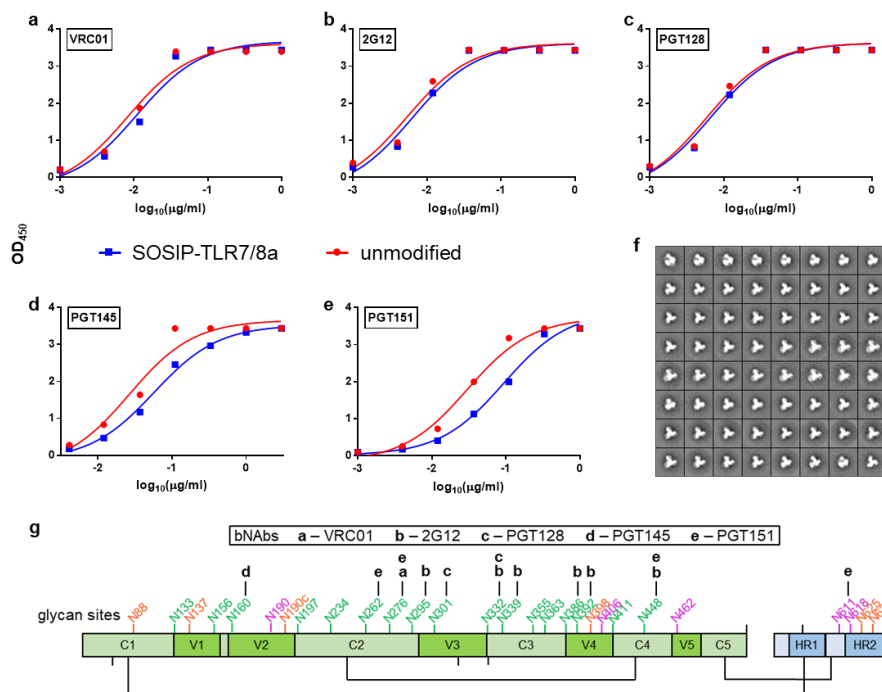


Figure 5 | Glycan conjugation preserves neutralization epitopes. **a – e**, ELISA using the antigenic surface coated with TLR7/8a-modified or native HIV Env trimer, probed with monoclonal antibodies VRC01 (**a**), 2G12 (**b**), PGT128 (**c**), PGT145 (**d**) and PGT151 (**e**). **f**, NS-EM 2D reference-free class averages for TLR7/8a-conjugated Env trimer. All 11,134 particles were found to be native-like trimers. **g**, Linear schematic of SOSIP Env trimer showing glycan-bNAb interactions. Glycans involved in binding are labeled with corresponding bNAb names. Glycan sites are color-coded in the same fashion as Fig. 1c. Amino acid labels are hidden for clarity (see Fig. 1c).

Discussion

There is data to support that the modification of a viral protein, such as HIV-1 Env trimers, with a TLR agonist may increase immunogenicity,¹⁸ and co-administration of TLR7 agonist with HIV envelope immunogens can induce durable humoral responses.³⁹ However, the challenge is to establish a conjugation methodology that allows the introduction of such an entity without compromising the glycoprotein and underlying bNAb epitopes. Here we describe a mild efficient one-step procedure to modify glycoprotein immunogens, such as HIV-1 Env trimers, by a small molecule adjuvant. The drug-to-protein ratio could readily be quantified by cleaving the modified sialic acid from the modified protein and subjecting the resulting sample to chromatographic analysis. Unlike lysine or cysteine-based covalent attachment, the modification occurs at termini of glycans, which are well-spaced away from the protein surface. BNABs against multiple important neutralization epitopes were not or minimally affected by the modification, which indicates promise as a vaccine candidate, which will be the focus of future studies.

Previously, gp120 was modified by an unnatural monosaccharide by metabolic labeling and click chemistry technology for imaging purposes.⁴⁰ It resulted in glycoproteins having low labeling efficiency ranges from 8-17%. Remarkably, the one-step modification using exogenously administered ST6Gal-I achieved an average incorporation of 43% per complex *N*-glycosylation site (see Supplementary Text for calculation).

The slight reduction in binding of trimer apex-specific bNAb PGT145 and gp120-pg41 interface targeting bNAb PGT151 is in concordance with previous reports. These two bNABs make contacts with a mixed population of *N*-glycans. Specifically, PGT145 interacts asymmetrically with the glycans at Asn¹⁶⁰ of two gp120 subunits and exhibits a preference for oligomannose type glycan, whereas complex type glycans including α 2,6-sialylated glycoform has a negative effect on the recognition due to possible steric clashes.^{37, 41} Asn¹⁶⁰ indeed harbors a small fraction of Golgi-processed complex-type *N*-glycans, which can be sialylated, in addition to the dominating oligomannose glycans.^{20, 22} PGT151 does prefer tri- and tetra-antennary complex type glycans present on gp41.^{38, 42} However, α 2,6-sialylation has been reported to negatively affect the antibody recognition of the epitope,⁴¹ which is consistent with our observations. Thus, decreased binding is probably caused by an altered glycoform on the Env protein upon sialylation, rather than the introduction of an unnatural moiety which is appropriately spaced away from the site of antibody engagement.

Future studies will focus on the evaluation of immunogenicity of the conjugated Env trimer for their ability to activate innate immune cells such as dendritic cells and macrophages, as well as to induce neutralizing antibodies against HIVs. We expect the approach can be extended to other viral glycoprotein immunogens for vaccine development, such as for influenza A virus hemagglutinin and SARS-CoV-2 spike protein, both harboring a considerable number of complex/hybrid type *N*-glycans sites for conjugation.^{43, 44}

Chapter 5

References

1. Schneerson, R.; Barrera, O.; Sutton, A.; Robbins, J. B. Preparation, Characterization, and Immunogenicity of Haemophilus Influenzae Type B Polysaccharide-Protein Conjugates. *J Exp Med* 1980, **152**(2): 361-376.
2. Sadarangani, M.; Pollard, A. J. Serogroup B Meningococcal Vaccines - An Unfinished Story. *Lancet Infect Dis* 2010, **10**(2): 112-124.
3. Durando, P.; Faust, S. N.; Fletcher, M.; Krizova, P.; Torres, A.; Welte, T. Experience with Pneumococcal Polysaccharide Conjugate Vaccine (Conjugated to CRM197 Carrier Protein) in Children and Adults. *Clin Microbiol Infect* 2013, **19**: 1-9.
4. Kawai, T.; Akira, S. The Role of Pattern-Recognition Receptors in Innate Immunity: Update on Toll-Like Receptors. *Nat Immunol* 2010, **11**(5): 373-384.
5. Underhill, D. M. Collaboration between the Innate Immune Receptors Dectin-1, TLRs, and NODs. *Immunol Rev* 2007, **219**(1): 75-87.
6. Kasturi, S. P.; Skountzou, I.; Albrecht, R. A.; Koutsouanos, D.; Hua, T.; Nakaya, H. I., *et al.* Programming The Magnitude and Persistence of Antibody Responses with Innate Immunity. *Nature* 2011, **470**(7335): 543-547.
7. Moyle, P. M.; Toth, I. Modern Subunit Vaccines: Development, Components, and Research Opportunities. *Chem Med Chem* 2013, **8**(3): 360-376.
8. Hou, B.; Saudan, P.; Ott, G.; Wheeler, Matthew L.; Ji, M.; Kuzmich, L., *et al.* Selective Utilization of Toll-like Receptor and MyD88 Signaling in B Cells for Enhancement of the Antiviral Germinal Center Response. *Immunity* 2011, **34**(3): 375-384.
9. Ingale, S.; Wolfert, M. A.; Gaekwad, J.; Buskas, T.; Boons, G.-J. Robust Immune Responses Elicited by A Fully Synthetic Three-Component Vaccine. *Nat Chem Biol* 2007, **3**(10): 663-667.
10. Ignacio, B. J.; Albin, T. J.; Esser-Kahn, A. P.; Verdoes, M. Toll-like Receptor Agonist Conjugation: A Chemical Perspective. *Bioconjug Chem* 2018, **29**(3): 587-603.
11. Hajam, I. A.; Dar, P. A.; Shahnawaz, I.; Jaume, J. C.; Lee, J. H. Bacterial Flagellin—A Potent Immunomodulatory Agent. *Exp Mol Med* 2017, **49**(9): e373-e373.
12. Fang, H.; Wu, Y.; Huang, X.; Wang, W.; Ang, B.; Cao, X., *et al.* Toll-like Receptor 4 (TLR4) Is Essential for Hsp70-like Protein 1 (HSP70L1) to Activate Dendritic Cells and Induce Th1 Response. *J Biol Chem* 2011, **286**(35): 30393-30400.
13. Lasarte, J. J.; Casares, N.; Gorraiz, M.; Hervás-Stubbs, S.; Arribillaga, L.; Mansilla, C., *et al.* The Extra Domain A from Fibronectin Targets Antigens to TLR4-Expressing Cells and Induces Cytotoxic T Cell Responses In Vivo. *J Immunol* 2007, **178**(2): 748.
14. Duthie, M. S.; Windish, H. P.; Fox, C. B.; Reed, S. G. Use of Defined TLR Ligands as Adjuvants within Human Vaccines. *Immunol Rev* 2011, **239**(1): 178-196.
15. Wille-Reece, U.; Flynn, B. J.; Loré, K.; Koup, R. A.; Kedl, R. M.; Mattapallil, J. J., *et al.* HIV Gag Protein Conjugated To A Toll-Like Receptor 7/8 Agonist Improves The Magnitude And Quality Of Th1 And CD8 T Cell Responses In Nonhuman Primates. *Proc Natl Acad Sci USA* 2005, **102**(42): 15190.
16. Wu, C. C.; Hayashi, T.; Takabayashi, K.; Sabet, M.; Smeets, D. F.; Guiney, D. D., *et al.* Immunotherapeutic Activity of a Conjugate of a Toll-like Receptor 7 Ligand. *Proc Natl Acad Sci USA* 2007, **104**(10): 3990-3995.
17. Rookhuizen, D. C.; DeFranco, A. L. Toll-Like Receptor 9 Signaling Acts on Multiple Elements of The Germinal Center to Enhance Antibody Responses. *Proc Natl Acad Sci USA* 2014, **111**(31): E3224.
18. Feng, Y.; Forsell, M. N.; Flynn, B.; Adams, W.; Lore, K.; Seder, R., *et al.* Chemical Cross-Linking of HIV-1 Env for Direct TLR7/8 Ligand Conjugation Compromises Recognition of Conserved Antigenic Determinants. *Virology* 2013, **446**(1-2): 56-65.
19. Torrents de la Peña, A.; Julien, J.-P.; de Taeye, S. W.; Garces, F.; Guttman, M.; Ozorowski, G., *et al.* Improving the Immunogenicity of Native-like HIV-1 Envelope Trimers by Hyperstabilization. *Cell Rep* 2017, **20**(8): 1805-1817.
20. Cao, L.; Diedrich, J. K.; Kulp, D. W.; Pauthner, M.; He, L.; Park, S. R., *et al.* Global Site-Specific N-Glycosylation Analysis of HIV Envelope Glycoprotein. *Nat Commun* 2017, **8**: 14954.

21. Struwe, W. B.; Chertova, E.; Allen, J. D.; Seabright, G. E.; Watanabe, Y.; Harvey, D. J., *et al.* Site-Specific Glycosylation of Virion-Derived HIV-1 Env Is Mimicked by a Soluble Trimeric Immunogen. *Cell Rep* 2018, **24**(8): 1958-1966.
22. Behrens, A. J.; Vasiljevic, S.; Pritchard, L. K.; Harvey, D. J.; Andev, R. S.; Krumm, S. A., *et al.* Composition and Antigenic Effects of Individual Glycan Sites of a Trimeric HIV-1 Envelope Glycoprotein. *Cell Rep* 2016, **14**(11): 2695-2706.
23. Sanders, R. W.; Derking, R.; Cupo, A.; Julien, J. P.; Yasmeen, A.; de Val, N., *et al.* A Next-Generation Cleaved, Soluble HIV-1 Env Trimer, BG505 SOSIP.664 Gp140, Expresses Multiple Epitopes for Broadly Neutralizing But Not Non-Neutralizing Antibodies. *PLoS Pathog* 2013, **9**(9): e1003618.
24. Sanders R. W.; Vesanen, M.; Schuelke, N.; Master, A.; Schiffner, L.; Kalyanaraman, R., *et al.* Stabilization of the Soluble, Cleaved, Trimeric Form of the Envelope Glycoprotein Complex of Human Immunodeficiency Virus Type 1. *J Virol* 2002, **76**(17): 8875-8889.
25. Binley James, M.; Sanders Rogier, W.; Clas, B.; Schuelke, N.; Master, A.; Guo, Y., *et al.* A Recombinant Human Immunodeficiency Virus Type 1 Envelope Glycoprotein Complex Stabilized by an Intermolecular Disulfide Bond between the gp120 and gp41 Subunits Is an Antigenic Mimic of the Trimeric Virion-Associated Structure. *J Virol* 2000, **74**(2): 627-643.
26. Sanders, R. W.; van Gils, M. J.; Derking, R.; Sok, D.; Ketas, T. J.; Burger, J. A., *et al.* HIV-1 Neutralizing Antibodies Induced by Native-Like Envelope Trimers. *Science* 2015, **349**(6244): aac4223.
27. Corti, D.; Lanzavecchia, A. Broadly Neutralizing Antiviral Antibodies. *Annu Rev Immunol* 2013, **31**(1): 705-742.
28. Burton, D. R.; Hangartner, L. Broadly Neutralizing Antibodies to HIV and Their Role in Vaccine Design. *Annu Rev Immunol* 2016, **34**: 635-659.
29. Kwong, P. D.; Mascola, J. R. HIV-1 Vaccines Based on Antibody Identification, B Cell Ontogeny, and Epitope Structure. *Immunity* 2018, **48**(5): 855-871.
30. Sun, T.; Yu, S.-H.; Zhao, P.; Meng, L.; Moremen, K. W.; Wells, L., *et al.* One-Step Selective Exoenzymatic Labeling (SEEL) Strategy for the Biotinylation and Identification of Glycoproteins of Living Cells. *J Am Chem Soc* 2016, **138**(36): 11575-11582.
31. Shukla, N. M.; Salunke, D. B.; Balakrishna, R.; Mutz, C. A.; Malladi, S. S.; David, S. A. Potent Adjuvanticity of a Pure TLR7-Agonistic Imidazoquinoline Dendrimer. *PLoS One* 2012, **7**(8): e43612.
32. Shukla, N. M.; Mutz, C. A.; Ukani, R.; Warshakoon, H. J.; Moore, D. S.; David, S. A. Syntheses of Fluorescent Imidazoquinoline Conjugates as Probes of Toll-Like Receptor 7. *Bioorg Med Chem Lett* 2010, **20**(22): 6384-6386.
33. Hong, V.; Steinmetz, N. F.; Manchester, M.; Finn, M. G. Labeling Live Cells by Copper-Catalyzed Alkyne-Azide Click Chemistry. *Bioconj Chem* 2010, **21**(10): 1912-1916.
34. Zhou, T.; Georgiev, I.; Wu, X.; Yang, Z.-Y.; Dai, K.; Finzi, A., *et al.* Structural Basis for Broad and Potent Neutralization of HIV-1 by Antibody VRC01. *Science* 2010, **329**(5993): 811.
35. Calarese, D. A.; Scanlan, C. N.; Zwick, M. B.; Deechongkit, S.; Mimura, Y.; Kunert, R., *et al.* Antibody Domain Exchange Is an Immunological Solution to Carbohydrate Cluster Recognition. *Science* 2003, **300**(5628): 2065.
36. Walker, L. M.; Huber, M.; Doores, K. J.; Falkowska, E.; Pejchal, R.; Julien, J.-P., *et al.* Broad Neutralization Coverage of HIV by Multiple Highly Potent Antibodies. *Nature* 2011, **477**(7365): 466-470.
37. Lee, J. H.; Andrabi, R.; Su, C.-Y.; Yasmeen, A.; Julien, J.-P.; Kong, L., *et al.* A Broadly Neutralizing Antibody Targets the Dynamic HIV Envelope Trimer Apex via a Long, Rigidified, and Anionic β -Hairpin Structure. *Immunity* 2017, **46**(4): 690-702.
38. Falkowska, E.; Le, K. M.; Ramos, A.; Doores, K. J.; Lee, J. H.; Blattner, C., *et al.* Broadly Neutralizing HIV Antibodies Define a Glycan-Dependent Epitope on the Prefusion Conformation of gp41 on Cleaved Envelope Trimers. *Immunity* 2014, **40**(5): 657-668.
39. Kasturi Sudhir, P.; Rasheed Mohammed Ata, U.; Havenar-Daughton, C.; Pham, M.; Legere, T.; Sher Zarphreen, J., *et al.* 3M-052, A Synthetic TLR-7/8 Agonist, Induces Durable HIV-1

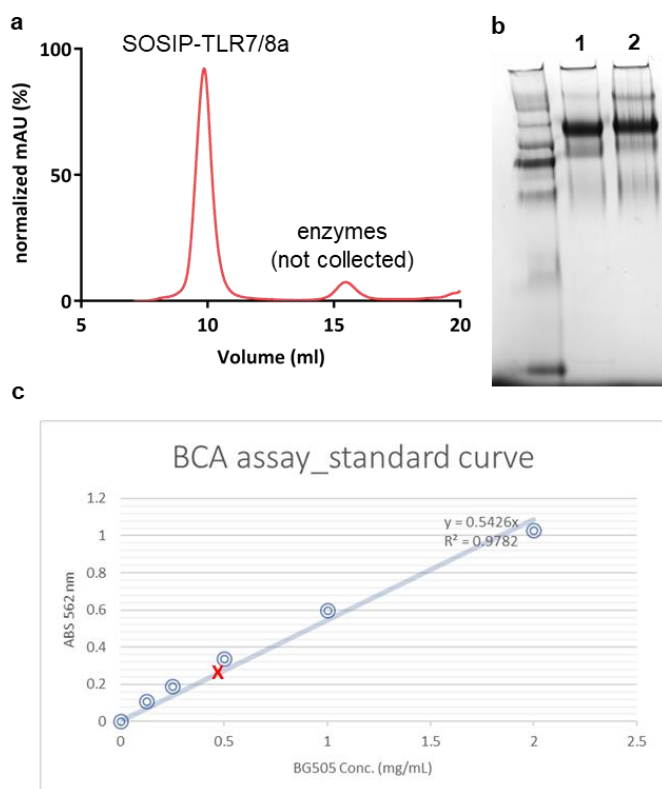
Chapter 5

- Envelope-Specific Plasma Cells And Humoral Immunity In Nonhuman Primates. *Sci Immunol* 2020, **5**(48): eabb1025.
40. Sun, L.; Ishihara, M.; Middleton, D. R.; Tiemeyer, M.; Avci, F. Y. Metabolic Labeling of HIV-1 Envelope Glycoprotein gp120 to Elucidate the Effect of gp120 Glycosylation on Antigen Uptake. *J Biol Chem* 2018, **293**(39): 15178-15194.
 41. Crooks, E. T.; Grimley, S. L.; Cully, M.; Osawa, K.; Dekkers, G.; Saunders, K., *et al.* Glycoengineering HIV-1 Env Creates 'Supercharged' and 'Hybrid' Glycans to Increase Neutralizing Antibody Potency, Breadth and Saturation. *PLoS Pathog* 2018, **14**(5): e1007024.
 42. Blattner, C.; Lee, Jeong H.; Sliepen, K.; Derking, R.; Falkowska, E.; de la Peña, Alba T., *et al.* Structural Delineation of a Quaternary, Cleavage-Dependent Epitope at the gp41-gp120 Interface on Intact HIV-1 Env Trimers. *Immunity* 2014, **40**(5): 669-680.
 43. Watanabe, Y.; Allen, J. D.; Wrapp, D.; McLellan, J. S.; Crispin, M. Site-Specific Glycan Analysis of the SARS-CoV-2 Spike. *Science* 2020, **369**(6501): 330.
 44. Cruz, E.; Cain, J.; Crossett, B.; Kayser, V. Site-Specific Glycosylation Profile of Influenza A (H1N1) Hemagglutinin through Tandem Mass Spectrometry. *Hum Vaccines Immunother* 2018, **14**(3): 508-517.

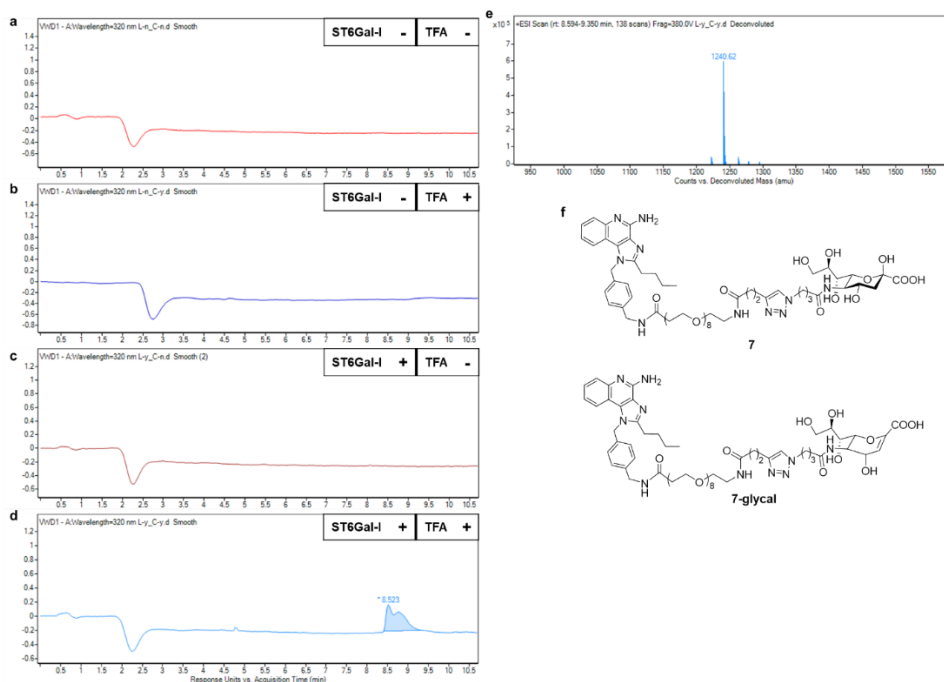
Supplementary Information for

Conjugation of A Toll-like Receptor Agonist to Glycans Preserves Neutralization Epitopes of HIV Native-like Envelope Trimer

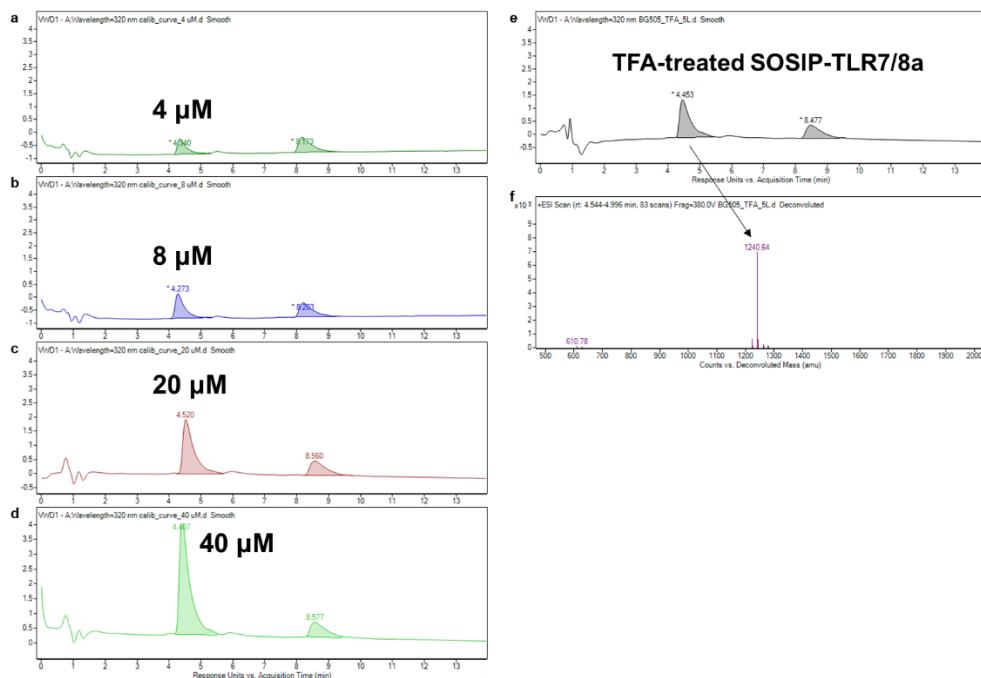
Part I. Supplementary Figures and Text



Supplementary Figure S1 | Protein purification and quality control. **a**, Size exclusion chromatogram of TLR7/8a-conjugated Env trimer. **b**, Non-denaturing Native-PAGE of the HIV envelope trimer before (Lane No. 1) and after (Lane No. 2) conjugation.¹ **c**, BCA assay using unmodified native-like HIV envelope trimer BG505 SOSIP v5.2 to plot the standard curve. The red cross sign indicates the ABS value obtained for the stock solution of SEC-purified modified trimer SOSIP-TLR7/8a. The ABS₅₆₂ value for the solution is 0.26 (baseline subtracted). Concentration of the stock solution is therefore 0.48 mg/mL.



Supplementary Figure S2 | Confirmation of TLR7/8 agonist conjugation using chromatographic analysis. **a-d**, HPLC chromatograms of native-like HIV envelope trimer before or after modification, with (+) or without (-) 50 mM trifluoroacetic acid treatment. In **a** and **b**, the glycoprotein was incubated with compound **1** and sialidase but in the absence of sialyltransferase ST6Gal-I. In **a** and **c**, samples were heated at 80°C in MilliQ water. 40 µg of protein was used for each group. Acquity UPLC peptide BEH C18 column was used, and eluted with a gradient of 0 – 50% acetonitrile at 0.3 mL/min flow rate within a period of 15 min. UV detector was set at 320 nm. **e**, Deconvoluted mass spectrum of the peak eluted around 8.5 min in panel **d** as a confirmation of presence of cleaved C5-TLR7/8a-conjugated sialic acid. Positive mode was used. **f**, There is a minor cleaved species with mass unit of 1222, which likely was given rise to by glycal (loss of water from free-reducing sialic acid). However, the abundance of this peak is marginal compared with the major species and does not interfere with the quantification demonstrated in Fig. S3.



Supplementary Figure S3 | Preparation of standard curves for compound 7. **a-d**, Chemically synthesized **7** has exactly the same structure as does acid-cleaved TLR-7/8 agonist modified sialic acid. The former was dissolved at different concentrations (4-40 μM) in a MilliQ water solution containing 5 μM compound **4** as internal standard for normalization. These solutions were subjected to HPLC-UV analysis, using Acquity UPLC peptide BEH C18 column, eluted with a gradient of 20-32% acetonitrile in a period of 12 min. UV detector was set at 320 nm. Area under curve (AUC, curve smoothened) was integrated manually using the spectrum-processing software MassHunter Qualitative Navigator. These values were normalized via dividing the AUCs for 4-5 min peaks (compound **7**) by those for 8-9 min peaks (compound **7**). **e** and **f**, UV@320 nm chromatogram of 4.8 μg TFA-treated SOSIP-TLR7/8a and deconvoluted mass spectrum of 4-5 min peak, respectively. Post-sialic acid release, the freeze-dried sample was reconstituted in 10 μL MilliQ containing the internal standard, a ratio of 2.1 (indicated with a red cross in Fig. 4c) was obtained for trifluoroacetic acid (TFA)-treated SOSIP-TLR7/8a, which gives a concentration of 15.6 μM . An estimated molecular weight of 400 kD was used for the protein, which gave a molarity of 1.2 μM . Thus, the average attachment number is $15.6/1.2 = 13$.

Chapter 5

Supplementary Text | Calculation of total complex *N*-glycosylation sites for a native-like HIV envelope trimer.

There are 24 N-glycan sites on a gp120 subunit and 4 on gp41. A native-like envelope trimer has hence a total of 84 N-glycan sites. The majority of glycan sites harbor a mixed population of glycoform types, which can be categorized into oligomannose/hybrid type (M/H) and complex type. The former is characterized by the sensitivity to degradation by glycosidase Endo H. Combining enzymatic treatment with sophisticated LC-MS/MS analysis, the M/H-type glycan content on a native-like envelope trimer has been determined, and has a value of 64% (Ref. 20 in main text). The rest 36% can be ascribed to complex type. Thus, the theoretical total number of complex glycan sites is $84 \times 36\% = 30$. This means, on average, there are 30 glycans that could be potentially modified with the TLR7/8 agonist. Alternatively, a theoretical complex glycan sites can be calculated from an independent study (Ref. 22 in main text), where the fraction of complex glycans were provided for each glycan site. The sum value of the fraction (9.0) for a total of 28 sites is the theoretical complex glycan number on one copy of protomer. Hence the total number of complex glycan sites is $9 \times 3 = 27$, which is in agreement with the former calculated number.

Incorporation rate of the TLR7/8a into the protein in Discussion section was calculated by dividing the average attachment number (13) by the theoretical total number of complex glycan sites (30), presuming each glycan can only be modified with one drug due to the inherent N-glycan branch selectivity of ST6Gal-I. The labeling efficiency of metabolic glycan tagging mentioned earlier in the same paragraph, as reported in Ref. 40 in main text, was the ratio between the MS abundancy of an azide-tagged glycopeptide ions to the sum abundancy of the same glycopeptide with and without azide-tag.

Part II. Experimental Section

Materials And Methods

Unless otherwise stated, all reagents for chemical reactions were purchased from Merck Sigma-Aldrich. Solvents for reactions were taken from the solvent purifier (MB SPS 5) or purchased from Biosolve Chemie. Technical grade organic solvents in work-ups were from VWR Chemicals. Deuterated solvents for NMR experiments were purchased from Cambridge Isotope Laboratories. ST6Gal-I and were expressed and purified according to reported protocols.² Sialidase from *C. perfringens* was purchased from Merck Sigma-Aldrich (11585886001). CTP and CMP-Sia synthetase NmCSS for were purchased from Chemily, LLC. RP-HPLC analysis was performed using the column Acquity UPLC peptide BEH C18 column (300Å, 1.7 µm, 1 mm × 50 mm) from Waters. NMR spectra were collected on an Agilent 400-MR, VARIAN or Bruker 600 UltraShield. Bruker micrOTOF-Q II ESI mass spectrometer were used for MS-based reaction monitoring. Agilent 6560 ion mobility Q-ToF MS couple with HPLC system was used for drug attachment quantification and HRMS data. Enzymatic reactions are conducted using MaxQ 4450 incubator (Thermo Scientific) at 37 °C with shaking or in a water bath set at 37 °C. **ELISA**: Ni-NTA HisSorb™ plates (Qiagen 35061), 1-Step™ Turbo TMB-ELISA Substrate Solution (ThermoFisher Scientific, 34022). Non-fat milk (Biorad, 1706404). **Native-PAGE**: NuPAGE 4-12% Bis-Tris Gel (Novex NP0321BOX), 20x NativePAGE Running Buffer (Invitrogen BN2001), 20x NativePAGE Cathode Buffer Additive (Invitrogen BN2002), Coomassie Brilliant Blue G-250 (Fluka 27815), 100% Ultrapure Glycerol (Invitrogen 15514-011), HMW-Native Protein Mixture (GE Healthcare 17-0445-01). Negative stain electron microscopy (NS-EM) characterization was performed as previously described (Ref. 25 in main text).

Protein expression. BG505 SOSIP v5.2 were expressed and purified following Ref. 25 in main text. In brief, HEK293T cells were cultured in Dulbecco's modified Eagle's medium (DMEM) containing 10% fetal calf serum (FCS), penicillin (100 U/ml), streptomycin (100 µg/ml), Glutamax (Invitrogen), non-essential amino acids (0.1 mM), sodium pyruvate (0.1 mM) and HEPES (0.1 mM). For gp140 trimer production, HEK293T cells were seeded at a density of 5.5×10^4 /ml in a Corning Hyperflask. The cells reach a density of 1.0×10^6 /ml after 3 days. To transfect the cells, PEI-MAX (1.0 mg/ml) in water was mixed with expression plasmids for Env and furin in OPTI-MEM. For one Corning Hyperflask, 600 µg of Env plasmid, 150 µg of Furin plasmid and 3 mg of PEI-MAX were added in 550 ml of growth media. Culture supernatants were harvested 72 h after transfection. The transfection supernatants were vacuum filtered through 0.2-µm filters and then passed (0.5–1 ml/min flow rate) over a PGT145 Sepharose column. The eluted proteins were buffer exchanged into 75 mM NaCl, 10 mM Tris, pH 8.0, using Snakeskin dialysis tubing (10K MWCO, Thermo Scientific). The proteins after dialysis were concentrated using Vivaspin columns with a 30-kDa cut off (GE Healthcare). The concentrated Env proteins were further purified using size exclusion chromatography (SEC) on a Superdex 200 26/60 column (GE Healthcare). The trimer fractions were collected and pooled. For NS-EM, BG505 SOSIP v8 was used.

Negative-stain electron microscopy. Trimer protein was diluted to 0.02mg/ml in 1x Tris-buffered saline, 3µl applied to a 400mesh Cu grid, blotted with filter paper, and stained with 2% uranyl formate. Micrographs were collected at 52,000 magnification on a ThermoFisher Tecnai Spirit microscope operating at 120kV with a TVIPS TemCam F416 CMOS camera using Leginon automated image collection software.³ Particles were picked using DogPicker⁴

Chapter 5

and 2D classification was done using iterative multivariate statistical analysis (MSA)/multireference alignment (MRA).⁵

Protein structures are visualized in UCSF Chimera. Chemical structures were drawn using ChemDraw Professional 16.0. Diagrams were drawn with Microsoft PowerPoint. NMR spectra were processed and analyzed using MestReNova. ELISA results were graphed using Graphpad Prism 7. Standard curves were plotted using Microsoft Excel.

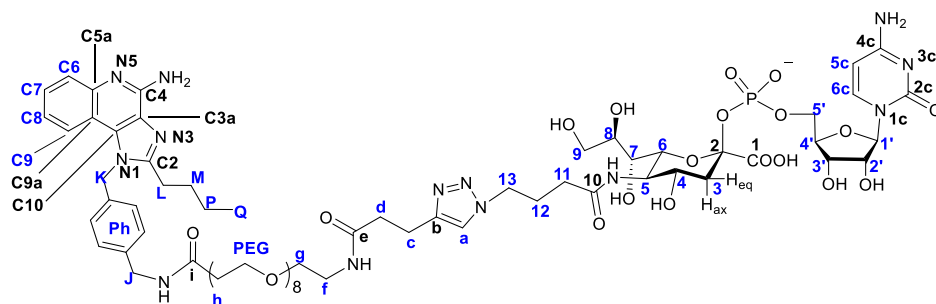
General Procedures for Protein Modification and Quantification

For chromatographic analysis demonstrated in Fig. S2, 40-50 μ g protein was used, whereas a normal scale reaction is performed with 1-2 mg protein. 50 μ g ST6Gal-I per mg of BG505 SOSIP v5.2, 0.1 U/mg sialidase from *C. perfringens*, 500 μ M compound **1** were used. Enzymatic reaction was performed at a working concentration of 1-2 mg/mL BG505 SOSIP v5.2, and were placed in a 37°C incubator or water bath for 16-20 hours.

For small scale modification, the solution was directly transferred to a 500 μ L spin filter with a molecular weight cut-off (MWCO) of 100 kD. The upper vial was topped up to the maximum volume. The spin filter was centrifuged at 6000 rpm for 20 min (if the volume in the upper vial does not go below 50 μ L, higher speed or longer time should be used) at 4°C for seven times to ensure excess CMP-Sia derivative **1** was removed to the extent below detection limit. The spin-concentrated solution (25-50 μ L) was transferred to a new vial for TFA-mediated sialic acid cleavage.

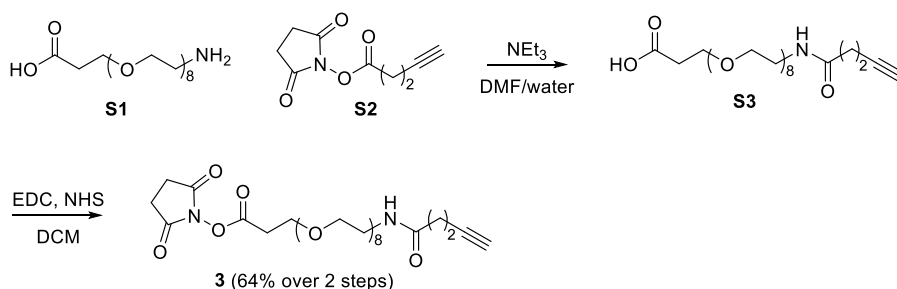
For normal scale, the labeling solution was subjected to size exclusion chromatography (Superdex Increase 200, elute with 20 mM Tris/HCl 75mM NaCl pH 8) for purification. The fractions were combined and concentrated using the spin filter (100 kD MWCO). Small aliquots were taken for BCA assay (Pierce™ BCA Protein Assay Kit 23227) and TFA treatment and LC-UV-MS attachment number determination. The aliquot (~5 μ g) was diluted to ~100 μ L with MilliQ and an equivolume of 100 mM TFA was added. The solution was heated at 80°C for 1 hour and then dried with speed-vac. The residue was constituted with 10 μ L of MilliQ which contains 5 μ M compound **4** as internal standard and was injected in one go into Acquity UPLC peptide BEH C18 column (20-32%). The peaks corresponding to **1** (~4.5 min) and **4** (~8.5 min) were integrated manually. The same chromatographic condition was applied to the chemically synthesized **7** at different concentrations (4, 8, 20, 40 μ M) for making the standard curve.

Compound Numbering For NMR Assignments



Numbering was shown using compound **1** and the rest of the compounds follow the same numbering scheme. In compound **4**, position a is an alkyne instead of a double bond in triazole moiety. The signals for positions shown in blue are assigned.

4. Detailed synthetic procedures

Synthesis of compound **4**

TLR7/8a as benzylamine derivative **2** was prepared following reported procedures⁶. Alkyne-containing linker **3** was prepared from carboxylic acid-PEG₈-amine **S1**. Specifically, in a round bottom flask, the starting material (330 mg, 0.74 mmol) was dissolved in 5 mL DMF/water (4:1). To this solution was added triethylamine (2 eq., 1.5 mmol, 0.2 mL) and *N*-hydroxysuccinimide-activated 4-pentynoic acid **S2** (1.3 eq., 0.96 mmol, 200 mg). The reaction was stirred at 21°C for 1 hour. The solution was concentrated in vacuo, and co-evaporated with toluene for 3 times. The residue was loaded on a LH-20 column to remove excess reagents. The acylated linker **S3** was taken up in (8 mL) DCM, to which *N*-hydroxysuccinimide (NHS, 2 eq. to **S1**, 1.5 mmol, 169 mg) was added. *N*-(3-Dimethylaminopropyl)-*N*'-ethylcarbodiimide (EDC)·HCl was added at 0°C. The reaction was then allowed to warm up to 21°C and stirred for 1 hour. The reaction was then diluted with DCM and washed ice cold 1N HCl solution and brine, and dried over sodium sulfate. After filtration, the organic solution was concentrated in vacuo to give compound **3** as a crude. This was used directly to acylate compound **2** without further purification.

Compound **2** as TFA salt (13 mg, 28 μmol) was dissolved in 0.3 mL DMF in a 15 mL Falcon tube, to which triethylamine (3 eq., 0.84 μmol, 12 μL) and **3**, which was prepared as a 0.5 M solution in DMF, was added (1.5 eq. 42 μmol, 80 μL). The reaction was stirred at 21°C for 1 hour. DMF was removed in vacuo by speed-vac with heating at 45°C. The residue was

Chapter 5

directly subjected to preparative reverse phase HPLC purification using a C8 column (20-65% acetonitrile/water, 12 mL/min, 50 min). Of note, the crude compound was taken up in methanol for injection. 28 mg compound **4** was obtained in 61% isolated yield. See table below for ^1H - and ^{13}C -NMR assignment and HRMS.

Compound 4: <i>N</i> -(4-((4-amino-2-butyl-1 <i>H</i> -imidazo[4,5- <i>c</i>]quinolin-1-yl)methyl)benzyl)-1-(pent-4-ynamido)-3,6,9,12,15,18,21,24-octaoxaheptacosan-27-amide					
MeOD- <i>d</i> 4, 600 MHz					
	δH (ppm)	δC (ppm)		δH (ppm)	δC (ppm)
C6	7.77	117.5	a	2.28	68.49
C7	7.39	124.3	c	2.46	13.49
C8	7.66	128.8	d	2.4	33.82
C9	7.99	120.8	f	3.36	38.28
K	5.59	47.68	g	3.6-3.4 ^a	69-70
J	4.38	41.32	h	2.46	35.59
L	3.02	25.66	Q	0.96	12.00
M	1.87	28.18	Ph	7.32	127.2
P	1.48	21.23		7.05	124.6
HRMS (ESI positive mode), calculated for C ₄₆ H ₆₇ N ₆ O ₁₀ [M+H] ⁺ 863.4919, found 863.4929.					
^a Peak buried in the PEG signal region.					

Synthesis of compounds **1** and **7**

Compound **4** (12 mg, 14 μmol) was dissolved in 1 mL methanol, to which 1 mL aqueous solution of compound **5** (1.0 eq. 21 μmol , 15mg) or 1 mL methanol solution of **6** (1.5 eq. 21 μmol , 8 mg) was added. When water was added to the methanol containing **4**, the solution turned cloudy due to low water solubility of **4**. The following reagents for click reaction were pre-mixed in water as a solution: CuSO_4 (0.5 eq. 1.7 mg as $\text{CuSO}_4 \cdot 5\text{H}_2\text{O}$, 7 μmol), THPTA (0.8 eq. 4.9 mg, 11 μmol), sodium ascorbate (1.0 eq. 2.7 mg, 14 μmol). This solution has a volume of 0.5 mL, and was immediately added to the solutions containing **5** or **6**. Upon addition of the click reagents, the solution of **4** mixed with **5** rapidly turned from cloudy to clear. These two solutions were stirred at 21°C for an hour. The reactions were monitored using ESI-MS (negative mode for **4**+**5** and positive mode for **4**+**6**). Methanol was removed by blowing nitrogen gently to the solution. Then the light blue water solutions were freeze-dried. The residue compounds were loaded onto biogel p2 for purification (elution: 50 mM ammonium bicarbonate, 4°C for compound **1** and 21°C for compound **7**). 12 mg compound **1** and 13 mg compound **7** were obtained. Isolated yields: 57 and 76%, respectively. See tables below for ^1H - and ^{13}C -NMR assignment and HRMS.

Compound 7 : 5-(<i>N</i> -(4-((4-amino-2-butyl-1 <i>H</i> -imidazo[4,5- <i>c</i>]quinolin-1-yl)methyl)benzyl)-1-(3-(1-(4-(methylamino)-4-oxobutyl)-1 <i>H</i> -1,2,3-triazol-4-yl)propanamido)-3,6,9,12,15,18,21,24-octaosaheptacosan-27-amido)-3,5-di-deoxy-D-glycero-D-galacto-2-nonulosonic acid
--

D ₂ O, 400 MHz					
	δ H (ppm)	δ C (ppm)		δ H (ppm)	δ C (ppm)
C6	7.61	119.6	3eq	2.20	39.53
C7	7.3	125.0	3ax	1.82	
C8	7.58	129.6	4	4.01	67.14
C9	7.76	121.0	5	3.93	52.2
K	5.73	48.48	6	3.97	70.13
J	4.32	42.34	7	3.61-3.26 ^a	68-70
L	2.94	26.28	8	3.74	70.21
M	1.67	28.84	9	3.82	62.2
P	1.33	21.61			
Q	0.83	12.93	11	2.23	32.21
Ph	7.24	127.7	12	2.12	25.63
	6.97	125.6	13	4.34	49.21
a	7.71	121.2	f	3.28	38.8
c	2.89	20.98	g	3.61-3.26 ^a	68-70
d	2.52	34.93	h	2.47	36.12
HRMS (ESI positive mode), calculated for C ₅₉ H ₉₀ N ₁₀ O ₂₁ [M+2H] ²⁺ 1242.6383, found 621.3196 (z=2).					
^a Peak buried in the PEG signal region.					

Compound 1: Cytidine-5'-monophospho-5-(<i>N</i> -(4-((4-amino-2-butyl-1 <i>H</i> -imidazo[4,5- <i>c</i>]quinolin-1-yl)methyl)benzyl)-1-(3-(1-(4-(methylamino)-4-oxobutyl)-1 <i>H</i> -1,2,3-triazol-4-yl)propanamido)-3,6,9,12,15,18,21,24-octaaxaheptacosan-27-amido)-3,5-di-deoxy-D-glycero-D-galacto-2-nonulosonic acid				
D ₂ O, 600 MHz				
	δ H (ppm)	δ C (ppm)	multiplet	<i>J</i> (Hz)
C6	7.63	120.6	d	9.0
C7	7.16	123.5	dd	7.6, 9.0
C8	7.49	128.4	dd	7.6, 9.0
C9	7.69	123.0	d	- ^a
K	5.61	48.28	s	
J	4.33	42.39	s	
L	2.89	26.36	t	-
M	1.62	29.23	m	-
P	1.33	21.76	m	-
Q	0.84	13.01	t	7.4
Ph	7.23	127.7	d	8.0
	6.93	125.8	d	8.0

Chapter 5

a	7.61	123.7	s	
c	2.91	20.99	t	-
d	2.54	35.01	t	7.5
f	3.30	38.89	t	-
g	3.60-3.16 ^b	68-70	-	-
h	2.48	36.25	t	5.5
3eq	2.52	41.33	dd	4.6, 12.0
3ax	1.67		d(t) ^c	12.0, -
4	4.09	66.85	ddd	4.6, 10.7, 11.1
5	3.99	51.82	dd	10.4, 11.1
6	4.16	71.78	dd	10.6, -
7	3.60-3.16 ^b	68-70	-	-
8	3.96	69.63	ddd	2.2, 6.5, -
9a	3.88	62.97	dd	2.1, 12.0
9b	3.64		dd	6.4, 12.0
11	2.26	32.36	m	-
12	2.15	25.78	m	-
13	4.35	49.27	t	6.6
5c	6.06	96.47	d	7.6
6c	7.94	141.5	d	7.6
1'	5.93	89.03	d	4.3
2'	4.27	74.36	dd	-
3'	4.33	69.25	dd	-
4'	4.23	82.91	dt	-
5'a, b	4.25	64.84	t	-
HRMS (ESI negative mode), calculated for C ₆₈ H ₉₈ N ₁₃ O ₂₆ [M-2H] ²⁻ 1543.6484, found 771.8248 (z=2).				
^a "-" Indicates coupling constant is not determined due to signal overlap.				
^b Peak buried in the PEG signal region.				
^c Axial H-3 of sialic acid becomes a dt peak upon CMP attachment.				

Additional references

1. Sliepen, K. *et al.* Structure and immunogenicity of a stabilized HIV-1 envelope trimer based on a group-M consensus sequence. *Nat Commun* 2019, **10**(1): 2355.
2. Moremen, K. W. *et al.* Expression system for structural and functional studies of human glycosylation enzymes. *Nat Chem Biol* 2018, **14**(2): 156.
3. Suloway, C. *et al.* Automated molecular microscopy: The new Legimon system. *J Struct Biol* 2005, **151**, 41-60.
4. Voss, N. R. *et al.* DoG Picker and TiltPicker: software tools to facilitate particle selection in single particle electron microscopy. *J Struct Biol* 2009, **166**, 205–213.
5. Ogura, T. *et al.* Topology representing network enables highly accurate classification of protein images taken by cryo electron-microscope without masking. *J Struct Biol* 2003, **143**:185–200.
6. Shukla, N. M. *et al.* Syntheses of fluorescent imidazoquinoline conjugates as probes of Toll-like receptor 7. *Bioorg Med Chem Lett* 2010, **20**(22): 6384.

Chapter 5

CHAPTER 6 | Discussion & Summary

Chapter 6

General Discussion

6.1 Working with sensitive molecules

The chemical lability of sialate-*O*-acetyl ester is well documented in literature and has been a major technical hurdle in the analysis and preparation of these carbohydrates¹⁻³. In Chapter 2 and 3, a chemoenzymatic strategy was described, which allowed the synthesis of almost any *O*-acetylated form of Neu5Ac and Neu5Gc oligosaccharides, respectively. During the course of the synthesis and product purification, potential *O*-acetyl hydrolysis or migration was avoided. Comprehensive 2D-NMR analysis, including HSQC and HMBC, demonstrating homogeneity and purity of each compound. Advantage of this unprecedented collection of oligosaccharides was taken to assess their stability profile not only to ensure reproducibility of the microarray experiments, but also to provide insights for future studies aiming at analyzing these molecules present in complex mixtures.

In this thesis the stability assessment on the α 2,3-linked 7-*O*-acetylated sialoside (**VI-1**, Fig. 1a) in 10 mM PBS buffer (pH 7), using LC-MS equipped with a HILIC column is demonstrated. This acetylation form is among the most prone to migration to give 9-*O*-acetylated product (**VI-2**). It was found that the compound remained stable in the solution on ice for an extended period of time (Fig. 1b).

Routine procedures for analyzing *O*-acetylated sialoglycome has remained underdeveloped^{4,5}. Given the sensitive nature of sialate-*O*-acetyl esters, the analysis of these molecules deserves caution. The synthetic library of *O*-acetylated sialosides not only provides a valuable set of analytical standards for LC/GC-based analysis of acid-released, DMB-derivatized Sias from a complex mixture to ensure unambiguous identification of *O*-acetylated sialic acid species, but also allows optimization of conditions for sample processing (e.g. glycan release) and chemical derivatization (e.g. permethylation) to fulfill sensitive and accurate glycomic analysis of *O*-acetylated sialosides, which in turn could facilitate deciphering the *O*-acetylated sialoglycome.

6.2 Going for higher complexity

CD60c, a mysterious ganglioside. In Chapter 4, the assembly of the ganglioside GD3 and GD2 glycan portions is demonstrated, which suggests that the chemoenzymatic methodology could be potentially integrated into the synthesis of highly complex molecules. Indeed, we have been pursuing the total synthesis of the full glycolipid ganglioside GD3 carrying *O*-acetyl ester at C7 (Fig. 3a), a hitherto functionally elusive molecule. Is it merely a transient biosynthetic precursor of 9-*O*-acetylated GD3 (CD60b), or does it retain 7-*O*-acetylation for a sufficiently long time and independently perform a different biological function? Although a growing body of literature has suggested the latter. These questions have nonetheless remained unanswered—and even impossible to answer—owing to the limited availability and the labile nature of this molecule.

To make 7-*O*-acetylated GD3, we planned to perform the lipid coupling at final stage of synthesis using the endoglycoceramidase-II (EGC-II) glycosynthase⁶. BCoV and MHV-S HE-mediated regioselective de-*O*-acetylation may be carried out on the intact glycolipid. The former enzymatic conversion requires an α -glycosyl fluoride, which we sought to prepare chemically. We have assembled the GD3 tetrasaccharide **VI-14** with a phenylthio moiety at reducing end (Fig. 3b) and have performed test reaction using a model compound, lactose disaccharide **VI-16**, and found the fluorination reaction mediated by Barluenga reagent⁷ afforded glycosyl fluoride in an α -only configuration, as was confirmed with ¹H- and HSQC

NMR experiments. It is expected the condition could be readily applied to tetrasaccharide **VI-14**.

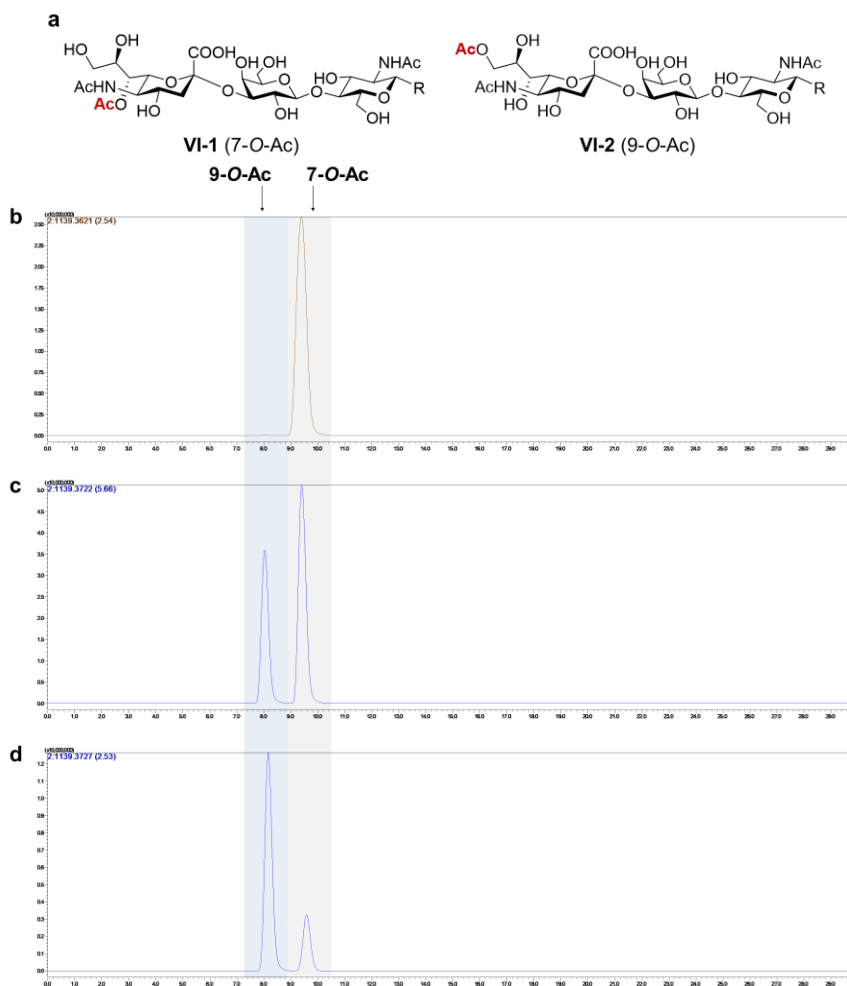


Figure 1 | Assessment of 7-*O*-acetyl stability in solution. **a**, Structures of 7-*O*-acetylated α2,3-linked sialyl LacNAc (left, **VI-1**) and 9-*O*-acetylated isomer (right, **VI-2**). **b**, LC-MS trace of freshly prepared printing solution of **VI-1** (100 μM in 10 mM PBS pH 7) immediately subjected to analysis. Acetyl migration from C7 to C9 was not observed. Detected mass: 1139.3636 (M-H). **c**, LC-MS trace of the same printing solution deliberately placed at room temperature (21°C) overnight before analysis. Detected mass: 1139.3694 (M-H). **d**, LC-MS trace of base-treated **VI-1**. The stock solution of **VI-1** was deliberately put under acetyl migration condition (pH 8, 37°C) in 10 mM PBS buffer for 2 hours. Rapid acetyl migration from C7 to C9 (**VI-2**) was observed under this condition. Detected mass: 1139.3681 (M-H). Mass was not calibrated.

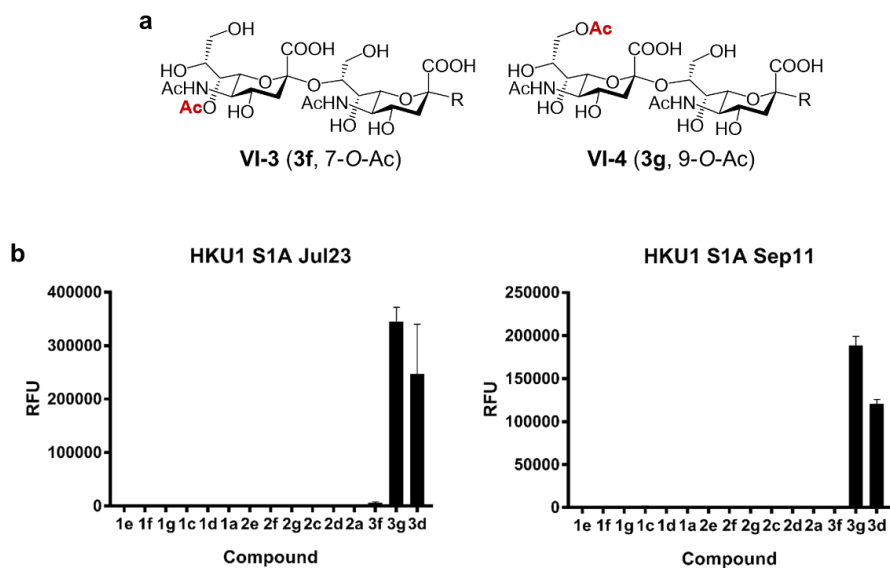


Figure 2 | Assessment of 7-*O*-acetyl stability on-slide. **a**, Structures of 7-*O*-acetylated disialic acid (left, **VI-3**, compound **3f** in the graph) and 9-*O*-acetylated isomer (right, **VI-4**, compound **3g** in the graph). **b**, Two independent screening experiments with HKU1 S1^A protein were performed on different dates on slides printed in the same batch. The compound library was printed onto streptavidin slides on July 16th, 2020. Printed slides were store at 4 °C in a closed cassette. Left, Data set of the screening on July 23rd, 2020, HKU1 S1^A protein concentration: 10 µg/mL. Right, Data set of the screening on September 11th, 2020, HKU1 S1^A protein concentration: 3 µg/mL. Compound **VI-3** remained negative in HKU1 S1^A binding experiment supporting that 7-*O*-acetylation was preserved during slide storage (>1 month) and no migration to C-9 had taken place.

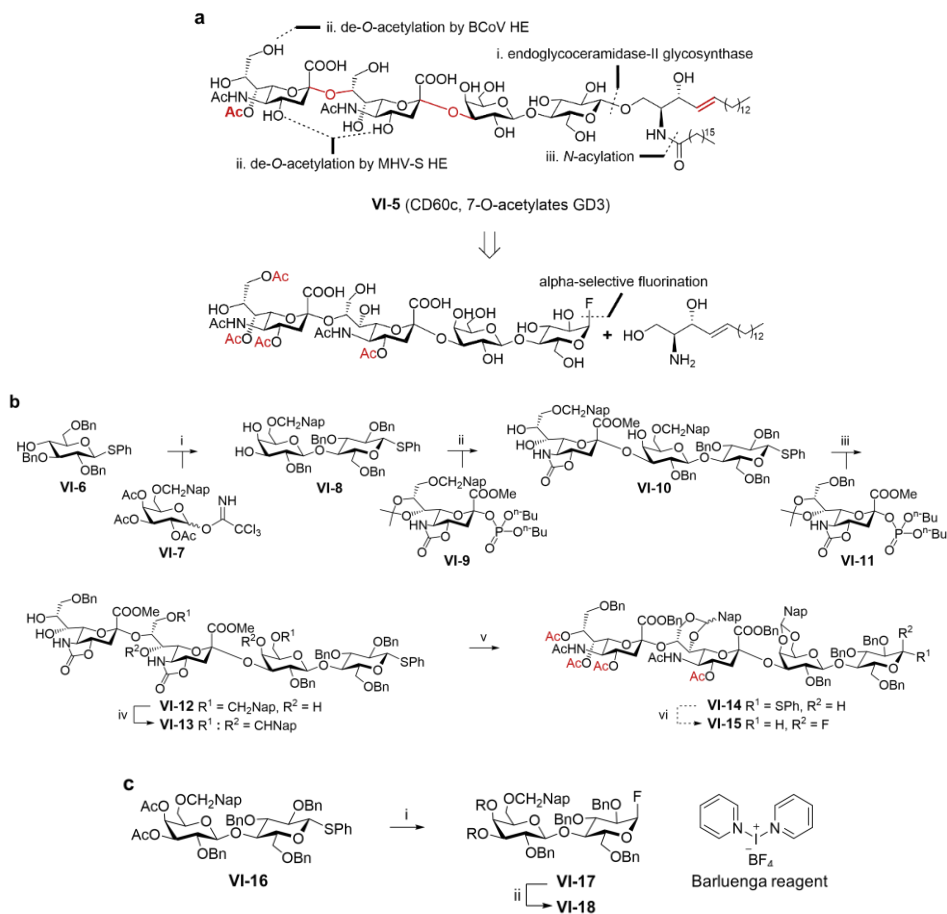


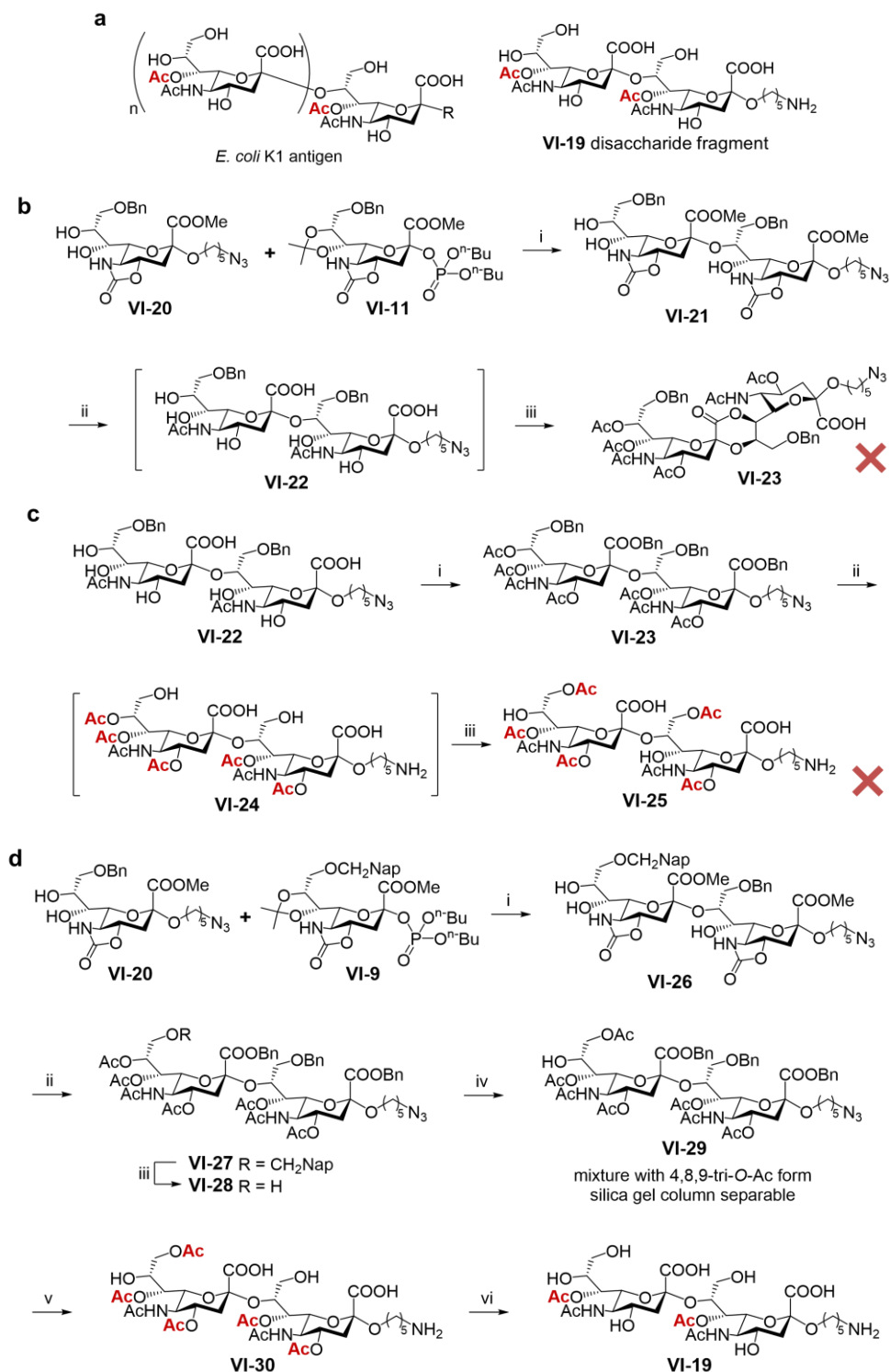
Figure 3 | Total synthesis of CD60c. **a**, Synthetic strategy for CD60c glycolipid. **b**, Assembly of glycan portion as glycosyl α -fluoride. Reagents and conditions: *i*. a) 1.2 eq. **VI-7**, TMSOTf, DCM, -45°C , 4 Å MS; b) MeONa, MeOH; c) 2,2-dimethoxypropane, CSA, acetonitrile; d) BnBr, NaH, DMF; e) CSA, MeOH. *ii*. a) 1.5 eq. **VI-9**, TMSOTf, DCM, -78°C , 4 Å MS; b) TFA, water, DCM. *iii*. a) 2 eq. **VI-11**, TMSOTf, DCM, -78°C , 4 Å MS; b) TFA, water, DCM. *iv*. DDQ, DCM, 4 Å MS. *v*. a) KOH, water/1,4-dioxane, 50°C ; b) Ac_2O , DCM; c) BnBr, CsCO_3 , DMF; d) Ac_2O , pyridine, 50°C . *vi*. Barluenga reagent, DCM. **c**, α -Selective fluorination of lactose disaccharide. Reagents and conditions: *i*. Barluenga reagent, DCM. *ii*. KOH, water/1,4-dioxane. Abbreviations: Ac, acetyl; Bn, benzyl; Bu, butyl; CSA, camphorsulfonic acid; DCM, dichloromethane; DDQ, 2,3-dichloro-5,6-dicyano-1,4-benzoquinone; DMF, dimethylformamide; Me, methyl; MS, molecular sieves; Nap, 2-naphthyl; Ph, phenyl; TFA, trifluoroacetic acid; TMSOTf, trimethylsilyl triflate.

Chapter 6

A form variant of bacterial capsular polysaccharide. Another potential target is (a fragment of) capsular antigen from *E. coli* K1. The K1 capsular polysaccharide contains a linear homopolymer of α 2,8-linked Neu5Ac which can undergo sub-stoichiometric 7-*O*-acetylation by the action of bacterial sialate-*O*-acetyltransferase encoded by *neuO*⁸. Under physiological conditions, some 7-*O*-acetyl groups spontaneously migrate to C9, creating a microheterogeneous mixture composed of differentially *O*-acetylated forms of capsular polysialic acid. The so-called form variation of *E. coli* K1 capsule bears a hitherto unknown function, and has been linked with increased virulence⁹. Literature has suggested that it may modulate interaction with host cationic anti-microbial peptides (CAMPs) via altering the hydrophobicity of the polysialic acid chain.

Initial attempts have been made to synthesize the K1 antigen fragment disaccharide **VI-19** (Fig. 4a). Although **VI-19** differs from the disialic acid described in Chapter 2 (compound **3f**) by only an additional 7-*O*-acetyl at the reducing end Neu5Ac residue, byproduct formation dominating the reaction at certain steps (Fig. 4b and c) underway to the final target suggests the synthetic route may deviate from that for **3f**. Indeed, we have made several modifications to the original route (see Chapter 2) to adapt the chemistry to **VI-19** synthesis (Fig. 4d). These variations include: 1) use of 9-*O*-Nap sialyl donor **VI-9** to attach non-reducing sialic acid; 2) protection of carboxylate as benzyl esters **VI-27**; and 3) *O*-acetyl migration prior to hydrogenation (**VI-27** to **-29**). We have tested the activity of MHV-S and BCoV HEs on the deprotected precursor **VI-30**, and observed it was readily converted to the desired 7,7'-di-*O*-acetylated disialic acid **VI-19**. Having demonstrated the chemistry on a disaccharide fragment, it opens opportunities to assemble longer chains of oligosialic acid and study kinetics of *O*-acetyl migration and the influence of 7- and 9-*O*-acetylation on oligosialic acid helix formation.

Figure 4 | Synthesis of *E. coli* K1 antigen fragment. **a**, Structure of K1 antigen (left) and the synthetic target, a disaccharide fragment (right). The degree of polymerization (*n*) of the former is larger than 100. **b**, Lactone formation upon *O*-acetylation. Reagents and conditions: *i*. a) 2 eq. **VI-11**, TMSOTf, DCM, -78°C, 4 Å MS; b) TFA, water, DCM. *ii*. a) KOH, water/1,4-dioxane, 50°C; b) Ac₂O, DCM. *iii*. Ac₂O, pyridine, 50°C. **c**, Unwanted C7 acetyl migration on reducing end Neu5Ac. Reagents and conditions: *i*. a) BnBr, CsCO₃, DMF; b) Ac₂O, pyridine, 50°C. *ii*. H₂, Pd(OH)₂, AcOH, THF/water. *iii*. CsF, AcOH, DMSO, 80°C. **d**, Optimized synthetic scheme for **VI-19**. Reagents and conditions: *i*. a) 2 eq. **VI-9**, TMSOTf, DCM, -78°C; b) TFA, water, DCM. *ii*. a) KOH, water/1,4-dioxane, 50°C; b) Ac₂O, DCM; c) BnBr, CsCO₃, DMF; d) Ac₂O, pyridine, 50°C. *iii*. DDQ, water, DCM. *iv*. CsF, AcOH, DMSO, 80°C. *v*. H₂, Pd(OH)₂, AcOH, THF/water. *vi*. BCoV and MHV-S HE, 50 mM HCOONH₄, 37°C. Abbreviations: DMSO, dimethylsulfoxide; THF, tetrahydrofuran.



Chapter 6

6.3 New tools for sialoglycobiology

Upgrading the toolbox for *in situ* O-Ac-Sia detection. Chapter 2 and 3 describes the in-depth analysis of receptor selectivities of a comprehensive panel of viral glycoproteins. It was found that the binding of these proteins was not only affected by *O*-acetylation pattern, but also structures beyond the terminal Sia moiety, in which glycosidic linkage plays a critical role. It was identified that some virolectins are promiscuous and accept many acetylation forms and sialosidic linkages, such as BCoV S1^A protein; while others are rather specific for certain *O*-acetylated forms or glycosidic linkage (Fig. 5a). For example, HKU1 S1^A binds exclusively to α 2,8-linked sialosides (Fig. 5b); PToV HE only recognizes 9- or 4,9-di-*O*-acetylation of α 2,3 and 2,6, but not 2,8 linkages (Fig. 5c); and BToV HE almost binds only α 2,3-linked sialosides but tolerates any *O*-acetyl esters in addition to 9-*O*-acetyl (Fig. 5d).

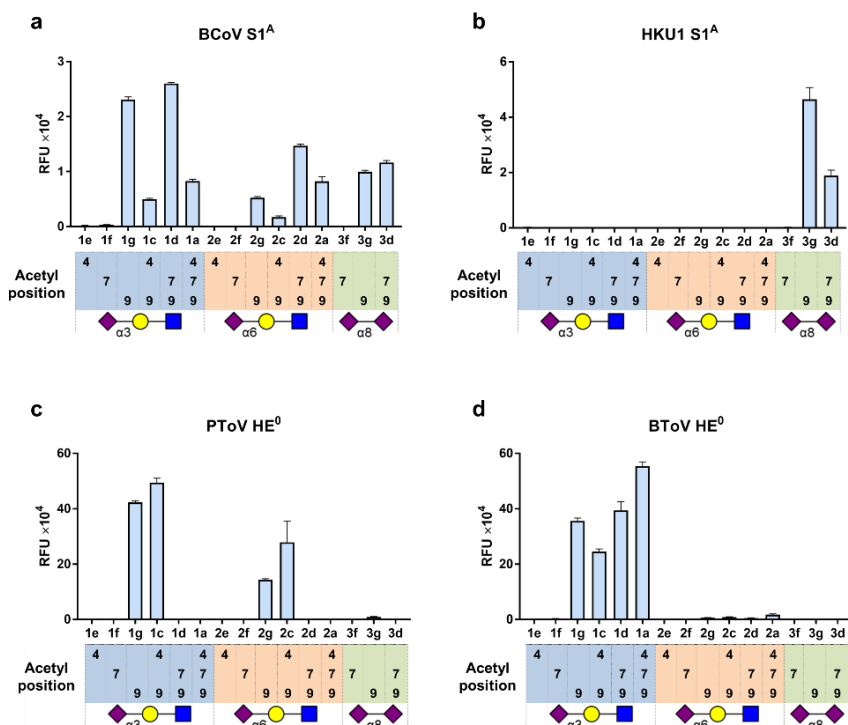


Figure 5 | Microarray results from Chapter 2. The results were taken from Supplementary Information of Chapter 2 with modifications. Abbreviations: BCoV, bovine coronavirus; PToV, porcine torovirus; BToV, bovine torovirus; HE⁰, hemagglutinin-esterase with inactive hydrolase.

These interesting properties make them useful tools for detecting specific sialic acid species *in situ* in various cell lines or tissue sections. The toolbox can be potentially upgraded by protein engineering to increase or loosen their binding specificities for a subset of *O*-acetylated sialosides. Demonstrated with HEs from closely related different animal coronaviruses in Chapter 4 and in a previous report¹⁰, alteration of even a single residue in the lectin pocket is sufficient to produce mutant HEs which differ in their *O*-acetyl form preference from the parent HE. The availability of a comprehensive library of homogeneous,

structurally well-defined oligosaccharides will facilitate elucidation of structural basis for *O*-acetyl form/glycosidic linkage selectivities of these viral glycoproteins to guide mutation. With the microarray platform, the engineered protein candidates can be rapidly screened for selection purposes. It is expected the structure-guided specificity-directed approach could afford virolectin-based *O*-Ac-Sia probes with tailor-made glycan specificities to enable investigation of tissue/cell-specific sialome.

Almost a perfect negative control. Given the instability of sialate-*O*-acetyl groups, we were motivated to investigate whether acetamide functionality could serve as a stabilizing bioisostere at C7 or C4 of sialic acid in Chapter 4. In this chapter, a new chemoenzymatic strategy was developed to afford unnatural 7- and 4-deoxy-acetamido sialoside analogs. The nitrogen atom was introduced at the stage of their corresponding synthetic precursors. The modified positions are C4 position of ManNAc as an azide for 7-*N*-acetyl analog and to the C4 of Neu5Ac as an amine for 4-*N*-acetyl analog. However, unlike the 9-*N*-acetylated analogs reported by Chen group¹¹, 7-*N*-acetyl failed to mimic the ester counterpart and abolished the recognition of several virolectins even when 9-*O*-acetyl is present. 4-*N*-acetyl was tolerated to some extent by MHV-S HE, despite the decreased binding. These results suggest that *O*-to-*N* substitution of sialic acid is not a universal solution to achieve stabilization without compromising protein recognition, which necessitates the use of natural *O*-acetylated sialosides to probe protein-carbohydrate interaction at the current stage.

The unfavorable effect of 7-*N*-acetyl moiety appears to arise from crosstalk with C5-acetamide, which limits probable conformations of the sialic acid skeleton. The conformational rigidity was observed in ¹H-NMR spectrum of 7-*N*-9-*O*-diacetylated sialoside in DMSO-*d*₆, where the methyl signal of the C7 and C5 acetyl moieties were significantly broadened at ambient temperature, but the peak broadening was less profound at elevated temperatures (Fig. 6a). Conversely, this effect was absent in the natural 7,9-di-*O*-acetylated sialoside (Fig. 6b). Thus, the 7-deoxy-acetamido analog may be considered almost an ideal negative control of the 7-*O*-acetyl esters: the two have almost the same composition and are structurally similar. Yet the former does not show binding to any tested proteins. The pair of structurally closely related compounds may be utilized to examine whether the binding of a given receptor protein is dependent merely on the presence of acetyl groups alone or those present on a Sia backbone. A specific example is provided for this application in the following section.

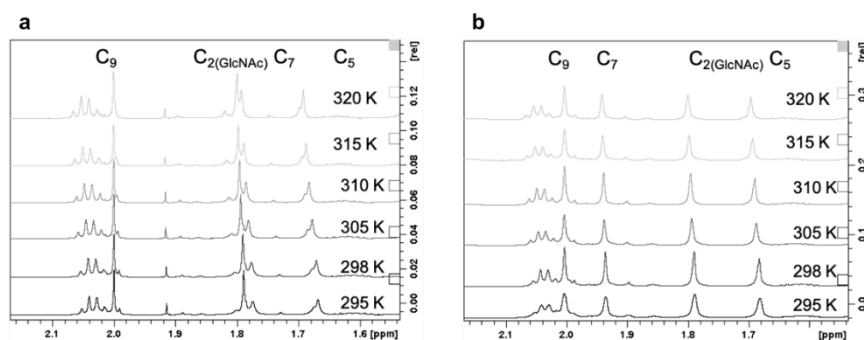


Figure 6 | Variable-temperature ¹H-NMR of α2,3-linked 7-*N*-9-*O*-Ac (a) and 7,9-di-*O*-Ac (b) sialosides in DMSO-*d*₆. Peak broadening was observed for the acetyl methyl groups of Sia C7 and C5, but not for those of C2 (GlcNAc) or C9.

Chapter 6

Hunting for endogenous *O*-Ac-Sia-specific lectins. *O*-acetylation of sialic acid has long been described as a post-synthetic modification of sialoglycans to block certain interaction. For example, 9-*O*-acetylation of the outermost Neu5Ac residue of gangliosides GD3 can block the glycolipid from disrupting the transmembrane potential of mitochondria, thereby dampening the death signals, leading to neutralization of the pro-apoptotic effect of GD3^{12, 13}. Sia present on B cell receptor (BCR) is a ligand for inhibitory immune receptor Siglec-2, and is essential for the latter receptor to associate with BCR. 9-*O*-acetylation of the BCR Sia residues effectively block such interaction, leading to lowered threshold for BCR-dependent B cell activation^{14, 15}. Given the wide tissue distribution of *O*-Ac-Sias^{16, 17}, we asked if there could be any endogenous binding partners, or lectins, which specifically recognize *O*-Ac-Sias as ligands. In other words, instead of blocking interactions, can *O*-acetylation of Sia specifically promote endogenous protein-carbohydrate interactions?

One interesting candidate is ficolin, which refers to a family of lectins consisting of three isoforms, namely ficolin-1, -2 and -3¹⁸. Ficolins have been proposed to serve as a pattern recognition receptor in innate immunity and recognize acetylated compounds¹⁹. They are implicated in the lectin pathway of complement activation and anti-microbial responses. Surprisingly, preliminary microarray results showed ficolin-2 bound multiply *O*-acetylated sialosides with high affinity and specificity (Fig. 7). Interestingly, similar to the cases of virolectins, ficolin-2 showed little, if any, binding to 7-deoxy-acetamido analogs. These results suggest the heightened affinity was not merely resulted from increased number of acetyl groups, but from better aligned hydrophobic elements in Sia with the lectin binding pocket, which warrants further structural studies. We infer that multiply *O*-acetylated sialosides may be a biologically relevant endogenous ligand for ficolin-2, despite the protein is somewhat promiscuous to other acetylated compounds. In humans, 7,9-di-*O*-acetylated sialic acids have been detected at considerable abundance across several tissues, which may be prioritized for investigating the function of ficolin-2.

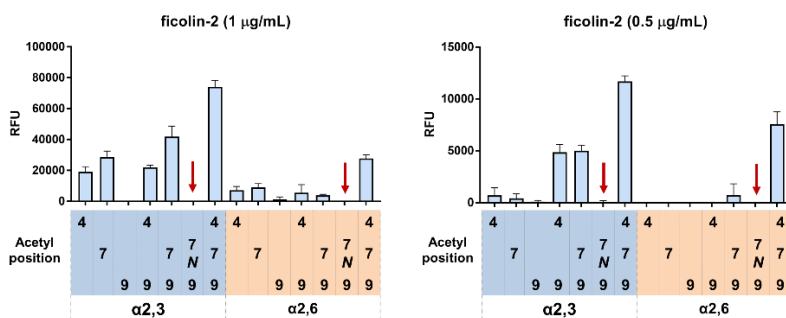


Figure 7 | Microarray results of ficolin-2. Microarray screening were performed using two concentrations of the protein (1 µg/mL, left; and 0.5 µg/mL, right). Ficolin-2 is Fc-tagged and is detected using AlexaFluor-conjugated mouse anti-human Fc secondary antibody. Negative binding results using 7-*N*-9-*O*-Ac sialosides are indicated with red arrows. Note that binding with the natural counterparts 7,9-di-*O*-Ac sialosides were detected at higher intensities.

6.4 Glycolipids as an essential receptor.

The S protein and HE as a two-component system perform balanced function in carbohydrate receptor binding and destroying in the mucus. Such highly dynamic, reversible

interaction with carbohydrates would become unfavorable as virus particles land on host cells. Instead, a tighter association with the cell surface would more likely to facilitate the uptake of the virion. In this sense, glycolipids extend only ~ 15 Å from cell surface and are almost an ideal receptor which may be accessed by the 20-nm S protein, while kept at some distance from the 8-nm HE. Such spatial segregation of receptor binding and destroying proteins could ensure a stable association of the virion with the cell surface.

It was shown in Chapter 2 when HEK293T cells were transfected with ST8Sia1 to produce $\alpha 2,8$ -linked disialogangliosides, they promoted OC43 and HKU1 S-pseudotyped VSV infection, whereas the sialate-*O*-acetyltransferase (CASD1)-knockout cells was resistant to infection. VSVs pseudotyped with S proteins having a S1^A deficient in *O*-Ac-Sia binding also failed to infect. These results indicate $\alpha 2,8$ -linked disialogangliosides can facilitate viral entry in a sialate-9-*O*-acetyl-dependent fashion and that the viral entry requires a *O*-Ac-Sia-binding-competent S1^A domain. At the current stage, we do not exclude a proteinaceous receptor may come in action upon carbohydrate engagement to mediate viral entry^{20, 21}. Also, it still remains an open question whether $\alpha 2,8$ -linked disialylated glycolipid alone, presented in a multivalent fashion, is sufficient trigger fusion machinery of the protease-cleaved S protein at an appropriate pH (neutral or acidic), despite that previous report demonstrated 9-*O*-acetylated sialic acid monosaccharide in solution present at millimolar concentration failed to do so²². Further structural studies employing Cryo-EM on these glycoproteins are likely to provide an unambiguous answer to this question.

It is of equal importance to establish essential receptors for entry for the remaining animal embecoviruses, including BCoV, ECoV, RbCoV and CRCoV. Although it is tempting to hypothesize that 9-*O*-acetylated disialogangliosides such as GD3 may also promote entry for other virus, given $\alpha 2,8$ -linked disialosides are bound by the spike proteins from these viruses, questions remain, however, whether the disialoside glycotope is indeed present in the relevant tissues for viral replication in the animals, and whether GD3, for example, is the very glycolipid which promote virion uptake in natural infection in these animals. It may seem too early to conclude that $\alpha 2,8$ -linked disialoglycolipid is a universal receptor for all embecoviruses, especially as we have observed the S proteins derived from viruses targeting distinct hosts exhibited differential toleration for acetyl ester modification in addition to 9-*O*-Ac, C5 hydroxylation (Neu5Gc, see Chapter 3). These results suggest that each embecovirus may target a different subset (not mutually exclusive) of gangliosides. Future research should focus on analysis of sialoglycome or sialoglycolipidome of the target tissues in the host to elucidate specific composition of *O*-acetylated sialic acid pool with molecular details on *O*-acetylation form, C5 modification and sialosidic linkage.

6.5 Lessons on linker design from rabbit immunization

In Chapter 6, we have utilized sialic acid as a handle for bioconjugation and managed to develop a general method to modify native-like HIV1 envelope (Env) glycoprotein with an immune-activator drug, a TLR7/8 agonist. The average number of drug attachment was readily quantified using acid-mediated Sia cleavage combined with LC-MS analysis. This glyco-modification of the immunogen, unlike direct lysine conjugation which occurs on the surface of the protein²³, does not compromise important neutralization epitopes, because the sites of attachment are spaced away from the protein surface, thus minimizing interference with epitope recognition by a wide range of broadly neutralizing antibodies. The bioconjugation method can be potentially applied to other viral glycoproteins such as the spike proteins of SARS-CoV-2 and the hemagglutinin or neuraminidase of influenza A virus.

Chapter 6

The ability of this chimeric immunogen to induce neutralizing antibodies is evaluated with rabbit immunization. In contrast to our expectation, the anti-sera from immunized rabbits showed a decrease in pseudovirus neutralization (Fig. 8a). Anti-sera from two out of five rabbits gave comparable ID₅₀ (see Fig. 8 captions for definition) to the non-conjugated Env-immunized group. Three out of five rabbits gave little, if any, neutralizing response. One possible explanation is that the neo-epitopes, such as the polyethylene glycol spacer, the triazole moiety and/or the drug molecule introduced to the intrinsically weakly immunogenic HIV Env had become immunodominant, and thereby distracted the response against the neutralization epitopes on Env^{24, 25}. A biotinylated trisaccharide containing exactly the same neo-epitopes introduced to Env has been prepared and tested with the rabbit anti-sera to examine the “off-target” response. We have found through ELISA that a considerable portion of the antisera were against the neo-epitopes, overriding that against the Env itself (Fig. 8b).

These results suggest that linker design can be an important factor for the outcome of immunization and warrants careful considerations. To tackle the issue of potential immunogenic nature of the introduced unnatural structures, a different type of linker with decreased hydrophobicity may be developed²⁶. On the other hand, immunization regimen may also be further optimized: instead of using the conjugated Env in the entire course of vaccination, limiting the administration of chimeric immunogen only at priming stage or right before final boost with native Envs could potentially reduce the off-target response against the unnatural linker region.

Although rabbits are normally the first-in-line model in which the immunogenicity of an HIV-1 Env is evaluated, a fair question to ask is whether they represent a proper model to evaluate TLR7/8 agonists²⁷, provided that toll-like receptors are considerably different in terms of their functions and expression patterns between rabbits and humans, as well as mice. As a matter of fact, many of the TLR agonists developed are tested in human cell cultures expressing human or mouse TLRs and in vivo using mice. Thus, the chimeric immunogen had better be subjected to a series of in vitro assays prior to immunization to examine if the covalent attachment of a TLR7/8 agonist could indeed have beneficial effects, for example, in activating dendritic cells and B cells expressing the cognate antigen receptor²³. Subsequently, Ig-knock-in mice may be used for immunization²⁸.

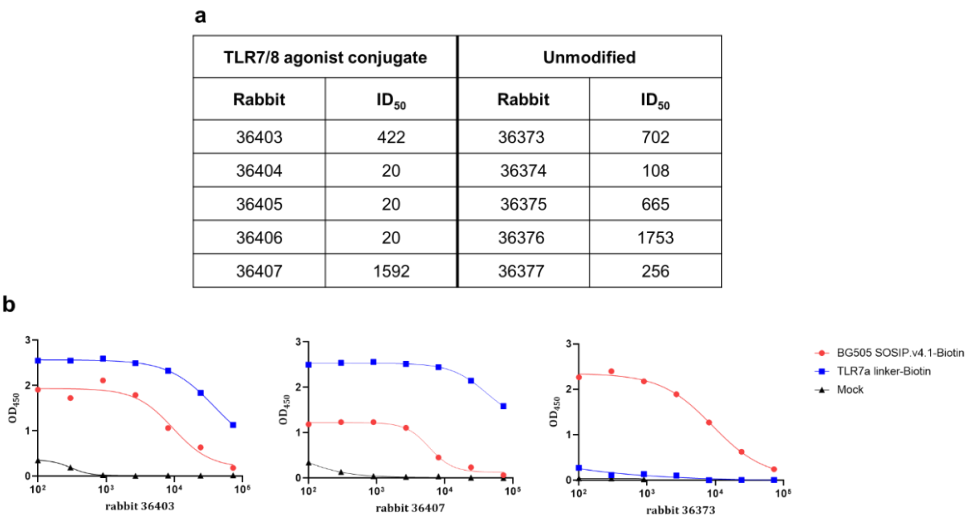


Figure 8 | Rabbit immunization studies. **a**, Neutralization assay of antisera from rabbits immunized with unmodified native-like Env trimer (right) and that conjugated with TLR7/8 agonist (left). Neutralization is quantified by the antiserum dilution factors reducing pseudoviral infectivity by 50% (half-maximal inhibitory dilution factor, or ID₅₀). **b**, ELISA of off-target response. Biotinylated Env trimer (BG505 SOSIP v4.1) and TLR7/8 agonist containing trisaccharide (TLR7a-linker) were immobilized on streptavidin-coated plate. Rabbits 36403 and 36407 were immunized with TLR7/8 agonist-conjugated Env trimer, 36373 with unmodified Env trimer.

6.6 Outlook

Synthetic glycochemistry has been an enabling technology for studying protein-carbohydrate interaction with molecular precision by providing structurally well-defined probes in high purity, which in turn sheds light upon biological functions of carbohydrates. In Chapter 2, 3 and 4, we sought to understand the biological significance of a mysterious class of molecules, *O*-acetylated sialic acid, in viral infection. To this end, a highly efficient platform was developed for high-throughput analysis of viral receptor specificities, which facilitated the identification of gangliosides as essential receptors for a group of human respiratory viruses. Key to these findings was the development of a general chemoenzymatic method for preparation of an unprecedented collection of *O*-acetylated sialosides, which represents a major advancement in glycochemistry, overcoming a long-standing challenge in carbohydrate synthesis. The methodology was demonstrated for simple di- and trisaccharides. Future research will focus on applying this strategy to a more diversified collection of underlying structures and to more complex molecules, such as *E. coli* K1 antigen and ganglioside GD3 variants.

In Chapter 5, we switched our focus onto a different project, in which we sought to improve the immunogenicity of native-like HIV envelope trimer by covalently conjugating a toll-like receptor 7 agonist to the glycoprotein via sialic acid. The one-step enzymatic conjugation is epitope-friendly, as the sites of conjugation are spaced away from protein surface. Initial rabbit immunization suggested the linker portion may require further structural optimization to eliminate the off-target effect. Future work should focus on preparation of CMP-sialic acid derivatives possessing a zwitterionic or cleavable linker.

References

1. Kamerling JP, Schauer R, Shukla AK, Stoll S, Vanhalbeek H, Vliegthart JFG. Migration of O-Acetyl Groups in N,O-Acetylneuraminic Acids. *Eur J Biochem* 1987, **162**(3): 601-607.
2. Chen X, Varki A. Advances in the Biology and Chemistry of Sialic Acids. *ACS Chem Biol* 2010, **5**(2): 163-176.
3. Park SS, Gervay-Hague J. Synthesis of partially O-acetylated N-acetylneuraminic acid using regioselective silyl exchange technology. *Org Lett* 2014, **16**(19): 5044-5047.
4. Zhang Q, Wang Y, Zheng Q, Li J. Analysis of O-Acetylated Sialic Acids in Dried Blood Spots. *Anal Chem* 2019, **91**(4): 2744-2751.
5. North SJ, Jang-Lee J, Harrison R, Canis K, Ismail MN, Trollope A, *et al.* Mass Spectrometric Analysis of Mutant Mice. 2010, **478**: 27-77.
6. Vaughan MD, Johnson K, DeFrees S, Tang X, Warren RAJ, Withers SG. Glycosynthase-Mediated Synthesis of Glycosphingolipids. *J Am Chem Soc* 2006, **128**(19): 6300-6301.
7. Huang K-T, Winssinger N. IPy2BF₄-Mediated Glycosylation and Glycosyl Fluoride Formation. *Eur J Org Chem* 2007, **2007**(12): 1887-1890.

Chapter 6

8. Deszo EL, Steenbergen SM, Freedberg DI, Vimr ER. *Escherichia coli* K1 polysialic acid O-acetyltransferase gene, neuO, and the mechanism of capsule form variation involving a mobile contingency locus. *Proc Natl Acad Sci USA* 2005, **102**(15): 5564.
9. King MR, Steenbergen SM, Vimr ER. Going for baroque at the *Escherichia coli* K1 cell surface. *Trends Microbiol* 2007, **15**(5): 196-202.
10. Bakkers MJ, Zeng Q, Feitsma LJ, Hulswit RJ, Li Z, Westerbeke A, *et al.* Coronavirus receptor switch explained from the stereochemistry of protein-carbohydrate interactions and a single mutation. *Proc Natl Acad Sci U S A* 2016, **113**(22): E3111-3119.
11. Khedri Z, Xiao A, Yu H, Landig CS, Li WQ, Diaz S, *et al.* A Chemical Biology Solution to Problems with Studying Biologically Important but Unstable 9-O-Acetyl Sialic Acids. *ACS Chem Biol* 2017, **12**(1): 214-224.
12. De Maria R, Lenti L, Malisan F, Agostino F, Tomassini B, Zeuner A, *et al.* Requirement for GD3 Ganglioside in CD95- and Ceramide-Induced Apoptosis. *Science* 1997, **277**(5332): 1652.
13. Malisan F, Franchi L, Tomassini B, Ventura N, Condò I, Rippo MR, *et al.* Acetylation Suppresses the Proapoptotic Activity of GD3 Ganglioside. *J Exp Med* 2002, **196**(12): 1535-1541.
14. Cariappa A, Takematsu H, Liu H, Diaz S, Haider K, Boboila C, *et al.* B cell antigen receptor signal strength and peripheral B cell development are regulated by a 9-O-acetyl sialic acid esterase. *J Exp Med* 2008, **206**(1): 125-138.
15. Han S, Collins BE, Bengtson P, Paulson JC. Homomultimeric complexes of CD22 in B cells revealed by protein-glycan cross-linking. *Nat Chem Biol* 2005, **1**(2): 93-97.
16. Langereis MA, Bakkers MJ, Deng L, Padler-Karavani V, Vervoort SJ, Hulswit RJ, *et al.* Complexity and Diversity of the Mammalian Sialome Revealed by Nidovirus Virolectins. *Cell Rep* 2015, **11**(12): 1966-1978.
17. Wasik BR, Barnard KN, Ossiboff RJ, Khedri Z, Feng KH, Yu H, *et al.* Distribution of O-Acetylated Sialic Acids among Target Host Tissues for Influenza Virus. *mSphere* 2017, **2**(5): e00379-00316.
18. Endo Y, Matsushita M, Fujita T. The role of ficolins in the lectin pathway of innate immunity. *The International Journal of Biochemistry & Cell Biology* 2011, **43**(5): 705-712.
19. Gout E, Garlatti V, Smith DF, Lacroix M, Dumestre-Perard C, Lunardi T, *et al.* Carbohydrate recognition properties of human ficolins: glycan array screening reveals the sialic acid binding specificity of M-ficolin. *J Biol Chem* 2010, **285**(9): 6612-6622.
20. Chan CM, Lau SKP, Woo PCY, Tse H, Zheng B-J, Chen L, *et al.* Identification of Major Histocompatibility Complex Class I C Molecule as an Attachment Factor That Facilitates Coronavirus HKU1 Spike-Mediated Infection. *J Virol* 2009, **83**(2): 1026.
21. Szczepanski A, Owczarek K, Bzowska M, Gula K, Drebot I, Ochman M, *et al.* Canine Respiratory Coronavirus, Bovine Coronavirus, and Human Coronavirus OC43: Receptors and Attachment Factors. *Viruses*; 2019.
22. Tortorici MA, Walls AC, Lang Y, Wang C, Li Z, Koerhuis D, *et al.* Structural basis for human coronavirus attachment to sialic acid receptors. *Nat Struct Mol Biol* 2019, **26**(6): 481-489.
23. Feng Y, Forsell MN, Flynn B, Adams W, Lore K, Seder R, *et al.* Chemical cross-linking of HIV-1 Env for direct TLR7/8 ligand conjugation compromises recognition of conserved antigenic determinants. *Virology* 2013, **446**(1-2): 56-65.
24. Wan X, Zhang J, Yu W, Shen L, Ji S, Hu T. Effect of protein immunogenicity and PEG size and branching on the anti-PEG immune response to PEGylated proteins. *Process Biochem* 2017, **52**: 183-191.
25. Nguyen DN, Xu B, Stanfield RL, Bailey JK, Horiya S, Temme JS, *et al.* Oligomannose Glycopeptide Conjugates Elicit Antibodies Targeting the Glycan Core Rather than Its Extremities. *ACS Central Science* 2019, **5**(2): 237-249.
26. Li B, Yuan Z, Hung H-C, Ma J, Jain P, Tsao C, *et al.* Revealing the Immunogenic Risk of Polymers. *Angew Chem Int Ed* 2018, **57**(42): 13873-13876.
27. Lai C-Y, Liu Y-L, Yu G-Y, Maa M-C, Leu T-H, Xu C, *et al.* TLR7/8 agonists activate a mild immune response in rabbits through TLR8 but not TLR7. *Vaccine* 2014, **32**(43): 5593-5599.
28. Verkoczy L, Alt FW, Tian M. Human Ig knockin mice to study the development and regulation of HIV-1 broadly neutralizing antibodies. *Immunol Rev* 2017, **275**(1): 89-107.

Nederlandse Samenvatting

Introductie

O-Acetyl siaalzuur. Siaalzuur (Sia) is een negatief geladen, negen koolstof keto-suiker aanwezig op de uiteinden van complexe koolhydraten. N-Acetylneuraminezuur (Neu5Ac) en N-glycolylneuraminezuur (Neu5Gc) zijn de vaakst voorkomende typen Sia. Neu5Ac en Neu5Gc ondergaan veelvuldige post-synthetische modificaties waarvan O-acetylactie het meest prevalent is. Het is mogelijk om tot 22 moleculaire varianten te produceren door de toevoeging van een of meerdere acetyl esters op de C4, C7, C8 en/of C9 positie. Onderzoek naar O-geacetylerde Sias heft zich voornamelijk beperkt tot 9-O-acetylactie. Er zijn aanwijzingen dat deze modificatie een belangrijke rol speelt in het aangeboren en adaptieve immuunsysteem, bacterieel foerageren en bacteriële pathogenese, virale infectie en vele soorten kanker. Ondanks de grote prevalentie blijft de biologische significantie van de overige acetylacties onduidelijk. Deze klasse koolhydraten is door chemische en enzymatische instabiliteit slechts beperkt beschikbaar, en incompleet inzicht in de biosynthese maakt genetische manipulatie zeer uitdagend.

O-geacetyleerd siaalzuur als virusreceptoren. Vele virussen binden initieel aan O-geacetylerde siaalzuren, waarna het virus cellen kan binnendringen. Om enkelen te noemen; coronavirussen zoals OC43 en HKU1 (*Betacoronavirus*, *embecovirus*) en influenza C virus infiltreren op die manier in mensen. Hoewel de effecten na infectie verschillen van individu tot individu, zijn de symptomen over het algemeen mild in volwassenen, maar kunnen in pasgeborenen, ouderen en personen met een verzwakt immuunsysteem tot ernstige complicaties leiden. In andere diersoorten worden vergelijkbare mechanismen gevonden. Voor zowel het rundercoronavirus – waarvan men vermoedt dat het de voorouder is van OC43 – als het influenza D virus – dat mogelijk gerelateerd is aan influenza C – geldt dat zij binden aan O-geacetylerde siaalzuren. Runder- en zwijntorovirussen idem dito. Typerend voor coronavirussen is dat de spike-eiwitten betrokken zijn bij binding aan O-geacetylerde sialosides. Daarnaast spelen homodimerische hemagglutinine-esterases (HE) op het oppervlak van het virus, bestaande uit een binding- en esterasedomein, een rol in de verwoesting van receptoren alsook de uitscheiding van virusdeeltjes. Torovirussen bezitten ook HE, terwijl influenza C en D een multifunctioneel hemagglutinine-esterase fusie-eiwit met zich meedragen. Dit fusie-eiwit verzekert via het hemagglutinine domein voor binding, de esterase zorgt voor receptorinactivatie en het fusie-eiwit voor fusie met het membraan van de gastheer. Voor de eerdergenoemde virussen geldt dat precieze vereisten, op moleculair niveau, voor initiële binding verder onderzoek behoeft. Onttrafeling van de exacte mechaniek zou een essentieel inzicht geven in hoe virussen zich tussen verschillende soorten kunnen verspreiden.

HIV envelop glyco-eiwit gebaseerde immunogeen. Tot dusver is inductie van neutraliserende antilichamen tegen virussen zoals HIV via vaccinatie nog niet gerealiseerd. Onderzoek in zowel dier als mens naar de mechaniek achter de activatie van aangeboren immuniteit via toll-like receptoren (TLR)en wijst uit dat zij geschikte kandidaten zijn om B- en T-cel activatie te bewerkstelligen. Tegelijkertijd is gebleken dat het behalen van een voldoende biologische beschikbaarheid van toegediende TLR-agonisten een uitdaging is. Enkele publicaties meldden dat vóór toediening, in-vitro covalente koppeling van de TLR-agonisten aan een immunogeen kan leiden tot het gewenste effect op het terugdringen van systemische ontstekingen. Echter, dergelijke benaderingen zijn in context van virale glyco-eiwitten die dienen als immunogeen nog onvoldoende onderzocht, deels door ontoereikend inzicht in hoe deze koppeling uit te voeren zonder verlies van functionaliteiten.

Chapter 6

HIV envelop spike-eiwitten bevatten 30 N-glycanen per molecuul, die voor de helft bijdragen aan de totale massa. Op het oppervlak van gp120 eiwit, worden met name oligomannose-type N-glycanen teruggevonden, terwijl het gp41 eiwit, voornamelijk complex/hydride-type N-glycanen draagt. De natuurlijke siaalzuren van de complex/hydride-type N-glycanen op het gp41 eiwit kunnen vervangen worden door gemodificeerde siaalzuren die dienen als bioorthogonale handvaten. Aangezien de uiteindes van de N-glycanen redelijk ver van het oppervlak verwijderd zijn, wordt aangenomen dat breed-neutraliserende antilichamen weinig hinder zullen ondervinden van modificaties aan die termini.

Overzicht van de thesis

In **hoofdstuk 2** wordt de ontwikkeling beschreven van een platform voor de diepgaande analyse van de receptorbinding van een groep virussen waarvan interactie met O-acetylsiaalzuur is beschreven. Deze groep omvat meerdere betacoronavirussen, torovirussen en influenza C en D virussen. Dit platform faciliteerde de identificatie van gangliosiden als functionele receptor voor twee humane coronavirussen. Essentieel in dit platform was de implementatie van een nieuwe synthetische methodologie voor de preparatie van een vrijwel complete set O-acetylsiaalzuur variaties. In deze methode zijn drie gemeenschappelijke uitgangsstoffen met C4, C7 en C9 O-acetylesters chemisch opgebouwd met een α 2,3, 2,6 of 2,8 koppeling. Ook kunnen per uitgangsstof nog zeven types O-acetyl siaalzuur worden geproduceerd door middel van HEs (hemagglutinin-esterase) van twee betacoronavirussen, welke op specifieke posities acetylesters verwijderen. In **hoofdstuk 3** wordt de chemoenzymatische methodologie uitgebreid naar de synthese van een set oligosacchariden die 5-N-glycolylneuraminezuuren bevat. Deze set weerspiegelt de eerder geproduceerde 5-N-acetylsiaalzuur set. Deze stoffen zijn gebruikt om de specifieke repons van betacoronavirale spike en HE eiwitten op 5-N-glycolylatie te bestuderen.

In **hoofdstuk 4** wordt de ontwikkeling van een chemo-enzymatische strategie voor de preparatie van 7- en 4-deoxy-N-acetamido analogen van O-acetylsiaalzuren beschreven. Het doel van de deze analogen is de chemische stabilisatie van de acetyl groep. De binding van deze 7- en 4-deoxy-N-acetyleerde analogen met coronavirale HE is vergeleken met de natuurlijk voorkomende O-acetylsiaalzuren op een glycaan microarray. De bindingstudie wees uit dat een O-naar-N substitutie een negatief effect had op HE binding. Een mogelijke verklaring voor dit effect is de veranderde positionering van het hydrophobe element in de suiker tijdens de interactie met het HE bindingsdomein. De 7-deoxy-N-acetamido analoog is vervolgens gebruikt als controle in een studie om de moleculaire basis van de HE:O-Ac-Sia interactie vast te stellen met STD-NMR in combinatie met moleculaire modellering. De resultaten duiden op een generieke herkenningwijze, gedeeld door verscheidene O-Ac-Sia bindende eiwitten.

In **hoofdstuk 5** wordt een dergelijke procedure, waarin gemodificeerd siaalzuur dient als bioorthogonaal handvat, verder beschreven. HIV-1 envelop trimer BG505.SOSIP v5.2 wordt gebruikt als dragereiwit. Een combinatie van de humane siaaltransferase ST6Gal1, een sialidase (*C. perfringens*) en een suikernucleotide met vooraf gekoppeld TLR7 zorgt ervoor dat de complex/hybride-type N-glycanen worden gemodificeerd met de agonist. Het aantal gekoppelde agonisten per dragereiwit wordt bepaald aan de hand van LC-MS. Maskering van epitopen, zoals gebeurt bij lysine-koppeling van agonisten, wordt uitgesloten met behulp van ELISA.

Acknowledgements (致谢)

I would like to express my deep gratitude to **Prof. Geert-Jan Boons** for the guidance and advice along the road of my PhD study, for the tolerance towards my rookie mistakes, and for the vision in scientific research that has led the team to standing at the frontier of carbohydrate science.

I am very grateful to **Raoul de Groot**, who has led me into the realm of virology. You have provided tremendous support and valuable insights into our collaborative project.

I want to thank **Prof. Rogier Sanders** for offering important insights and greatly facilitating the exploratory project of anti-HIV-1 vaccine development. I truly appreciate your high efficiency and have always been amazed by the work produced from your lab and how they are advancing the field.

The work presented in this thesis would not have been possible without teamwork. Therefore, I would like to sincerely thank those who have made significant contributions to these projects. I want to start by thanking **Margreet Wolfert** for keeping Boons Lab organized, coordinating the paper submission and helping with microarray experiments. You make sure our lab runs in a well-managed and efficient manner. I would like to extend my gratitude to the lab technicians and scientific staff, **Arwin Brouwer**, **Javier Sastre Terano**, **John Kruijtzter**, and **Justyna Dobruchowska**, who constantly keep the instruments in order and have been always willing to offer their kind help.

这本博士论文里面的工作如果没有**郎一飞 (Yifei Lang)** 的鼎力相助, 很难想象我该如何完成撰写。犹记得当时我们是如何把午饭桌上的灵机一动慢慢发展成为一篇令我们引以为豪的工作, 也记得投稿、拒稿、再投, 到接收时那坐过山车一般的心路历程。而往事不止可以回味, 让我们更以此共勉, 继续前行。I got to know **Frederik Broszeit** back when we were at CCRC and came to CBDD around the same time. You were particularly good at operating all those instruments and were very kind to show me how to use them. I enjoyed working with you, having lunch together, and occasionally complaining (through which we tried to find solutions) about research together. **Mehman Bunyatov**, my buddy of chemical sialylation, I think we have finally mastered that reaction after we documented so many failures. Thank you for completing the sialylation diary with me all these years. **刘林 (Lin Liu)** 师兄是我心目中科研工作者的榜样。我不仅佩服你的博闻强识, 更惊叹你在钻研科学问题时的耐心和见解。感谢你对 microarray 实验的鼎力相助。I would like to thank **Luca Unione**, who have taken many research projects to the next level by employing sophisticated NMR techniques. Most amazingly, you have even turned an inactive compound into a useful tool in structural biology, which allowed us to gain valuable insights into virus glycan interaction. I owe many thanks to **Gerlof Bosman (林荷平)**, who assisted in mass spectrometric characterization of proteins, and were producing all kinds of enzymes like a factory. While helping others, you have been daily supervisor of not

one but several students. I was impressed by how you are able to multi-task and plan things well ahead. I would like to thank **Ronald Derking** for always being responsive and extremely efficient. It was truly a pleasure to work with you. I am grateful to **Rob de Vries**, who provided many enzymes and other proteins to facilitate our projects. Apart from being helpful in the lab, Rob knows a lot about meat and how to cook them.

Besides those who made significant contributions to the projects in this thesis, I would like to thank the people that made my work and life at CBDD and Utrecht more 'colourful'. Thank you all for being there! I have known **Apoorva Srivastava** and **Ivan Gagarinov** for long since 2013 and followed their steps to the Netherlands. I enjoyed particularly every late-night chat in the lab when we just started in Utrecht. Even now I still remember how we turned the lab then into a bit more organized place. I would like to thank **Gael Vos**, your creativity and daring ideas have always amazed me. It was always enjoyable to talk science and exchange ideas with you. **Ingrid 't Hart**, thank you for being a great lab mate in our fume hood bay and the awesome job as a borrel committee. **Rosanne van Beek**, I would like to thank you for being the 'biology consultant' over the past few years and always being helpful. I am thankful to **Jelle Fok**, who continues with the anti-viral vaccine project. Good luck with the next-gen linker design!

很荣幸可以在乌特勒支结识一群有趣的小伙伴们。我刚刚来到 CBDD 时, 犹记得张浩 (**Hao Zhang**) 像个老大哥一样带着我熟悉校园各个地方。而印象最深的就是聚会时看着你和欣爷 (**Xin Yang**) 互损, 而一旁的小于 (**Guangyun Yu**)、施杰 (**Jie Shi**) 和文静姐 (**Wenjing Lu**) 则在一旁围观憨笑或是默不作声。Dowson、龙爷 (**Minglong Liu**)、凤姐 (**Lifeng Sun**)、孟爷 (**Xianke Meng**)、燕燕 (**Yanyan Liu**)、张志勇 (**Zhiyong Zhang**) 的到来使“中国队”在 CBDD 的日渐壮大, 而在 CBDD 之外, 也有幸认识了涛哥 (**Wentao Li**)、赵珊 (**Shan Zhao**)、陈婕 (**Jie Chen**) 和付东隆 (**Donglong Fu**), 仿佛整个乌特勒支也跟着渐渐热闹起来。如今大家或去或留, 各奔东西, 不免感叹时光飞逝。感谢你们的陪伴, 愿多年后祖国再度相聚, 把酒 (包含软饮) 言欢!

I am thankful to the current and former CBDD members for all their help and the fun we had together: Prof. Roland Pieters, Tom Wennekes, Bernd Stahl, Balthasar Heesters, Seino Jongkees, Dirk Rijkers, Ed Moret, Luuk Stel, Cyril Balsollier, Enrico Verpalen, Pieter de Saint Aulaire, Tim Leenders, Yvette Luijkx, Vito Thijssen, Ryoji Yoshisada, Hanna de Jong, Xuan Wei, Diksha Haskar, Reshmi Mukherjee, Suhela Sharif, Francesco Palmieri, Ana Gimeno Cardells, Ilhan Tomris, Helena Ehren, Jun Ong, Erianna Alvarado Melendez, Krishna Desai, Yunfei Wu, Yujie Ma, Liangwei Zhang, Xiufen Liu, Margherita Duca, Elena Loi, Igor Sweet, Cindy Spruit, Lemeng Chao, Pouya Zaree, Elif Uslu, Roosmarijn van der Woude, Linda Quarles van Ufford, Nishant Sewgobind, Oier Aizpurua-Olaizola, Nuria Martinez Saez, Victor Somovilla Busto, Nino Trattinig, Julia Weber, Babara Steigenberger, and Torben Heise.

Before this thesis was completed, I already started my post-doc research at TUDelft and LUMC. I would like to thank **Prof. Chirlmin Joo**, **Vered Raz** and **Thom Sharp** for offering me the chance to embark on an exciting journey to shed some light upon the mysterious glycoRNA. As I join the team, many have provided significant help and support. **Sungchul Kim** and **Adam Pomorski** have led me into the RNA world. Hope you guys will lead successful career in academia. **Bhagyashree Joshi** is the cell imaging Guru. It has been truly a pleasure to work with you on the cell surface RNA. Hope we will manage to take it to the next level. **Mike Filius** and **Raman van Wee** invited me into the protein sequencing team. Let's give those proteins a pinch of sugar. I would like to thank **Carolien Bastiaanssen** and **Ilja Westerlaken** for their clear instruction on using the instruments, and the orientation in the BN maze. I am also thankful to other group members for their help and advice, including Kijun Kim, Sung Hyun Kim, Koushik Sreenivasa, Carlos de Lannoy, Ivo Severins, Dong Hoon Shin, Margreet Docter and Jan Wignand at TUDelft; and Cyriel Olie, Milad Shademan, Danish Khan, Tooba Abbassi and Salma El Abdellaoui at LUMC.

I also would like to thank the colleagues/collaborators outside CBDD. It was my pleasure to help and to work with you, Mark Bakkers, Hongbo Guo, Meiling Dai, Wenjuan du, Xander de Haan, Eric de Vries, Berend Jan Bosch, Frank van Kuppelveld, Nico Overeem, Juriaan Huskens, Alexandra Tortorici, and David Veesler. Also, I would like to thank those who kindly offered their help along my PhD: Kwinten Sliepen, Richard Scheltema, 干劲(Jin Gan), and 娄博(Bo Lou).

I started my PhD at University of Georgia, and would like to express my gratitude to the colleagues over there at CCRC: Chantelle Capicciotti, Chengli Zong, Tiantian Sun, Tiehai Li, Ning Ding, Frederic Friscourt, Anthony Prudden, Robert Chapman, Apoorva Joshi, Josette Wilkes, Maria Moure Garcia, Nitin Supekar, Manish Hudlikar, Yi Gu, Pradeep Chopra, Rachel Bainum, and all current and former members of Boons Group.

在佐治亚或已离开的朋友们, 洪冬喆, 祺哥, 雷依, 王瑶, 巍钟, 炜哥, 卫刚, 潇月, 莉姐, 赵博远, 秀茹姐和庆勇哥, 感谢有你们的陪伴, 让雅典这座略显单调的小镇变得充实多彩。另外, 想特别感谢一下我在海牙的邻居, 李天昂和沈听然夫妇, 真是应了那句“远亲不如近邻”, 你们对我和朱林立的帮助渗透在生活的各个方面。愿云宝茁壮成长, 有一个快乐幸福的童年。

亲爱的爷爷奶奶, 外婆, 岳父岳母, 爸爸妈妈, 还有各位长沙、北京、太原、上海的长辈、亲人们, 海外求学多年, 未能常伴您们左右, 感谢您们对朱林立和我无私的支持、理解与包容!

致谢写到了最重要的位置, 在此我特别想感谢我的妻子, 朱林立。首先要感谢你对我在学术方面的支持, 特别是你教会了我如何与数据打交道, 你提供的解决方法总能把耗时数小时的数据处理缩短到几分钟。我更要感谢你的陪伴, 以及对我的包容、理解和启发, 尤其是在课题一筹莫展之时, 我难免会深陷于自己的思维漩涡,

无法自拔。而你总能将我成功抽离这种淤滞的心境，让我从容面对科研的困境，走出死胡同。最后，我想对我们即将降生的孩子说，我们很高兴成为你的父母，希望未来的你拥有健康的身体，独立的思想和自由的灵魂。

February, 2022

于荷兰 海牙 (den Haag, the Netherlands)

Curriculum Vitae

Updated 10/2021

Zeshi Li (Jack)

lizeshi.nl@gmail.com | +31 6 24 83 42 66

LIST OF PUBLICATIONS AND MANUSCRIPTS

1. **Z. Li***, Y. Lang*, L. Liu, M. I. Bunyatov, A. Isaza Sarmiento, R. J. de Groot, G.-J. Boons (2021) Synthetic O-acetylated sialosides facilitate identification of functional receptor for human respiratory viruses. *Nat. Chem.* 13 (5): 496-503. (* equal contribution)
2. **Z. Li***, L. Unione*, L. Liu, Y. Lang, R. J. de Groot, G.-J. Boons (2022) Synthetic O-acetylated sialosides and their acetamido-deoxy analogs as probes for coronaviral hemagglutinin-esterase recognition. Status: Accepted, *J. Am. Chem. Soc.*
3. **Z. Li**, R. Derking, G. P. Bosman, R. W. Sanders, G.-J. Boons (2021) Glycoconjugation of toll-like receptor-7 agonist to native-like HIV-1 envelope trimer preserves neutralization epitopes. Status: Submitted.
4. **Z. Li**, Y. Lang, L. Liu, R. J. de Groot, G.-J. Boons (2021) Synthetic O-acetyl-N-glycolylneuraminic acid oligosaccharides reveal host-associated binding patterns of coronaviral glycoproteins. Status: Submitted.
5. Y. Lang*, **Z. Li***, A. Schouten, F. J. M. van Kuppeveld, G.-J. Boons, E. G. Huizinga, R. J. de Groot (2020) Structural basis for recognition of different O-acetylated sialic acid derivatives by bovine coronavirus hemagglutinin-esterase. Status: In preparation. (* equal contribution)
6. Y. Lang, W. Li, **Z. Li**, D. Koerhuis, A. C. S. van den Burg, E. Rozemuller, B.-J. Bosch, F. J. M. van Kuppeveld, G.-J. Boons, E. G. Huizinga, H. M. van der Schaar, R. J. de Groot (2020) Coronavirus hemagglutinin-esterase and spike proteins co-evolve for functional balance and optimal virion avidity. *PNAS*, 117(41): 25759-25770
7. N. J. Overeem, P. H. E. Hamming, O. C. Grant, D. Di Iorio, M. Tieke, M. Candelaria Bertolino, **Z. Li**, G. Vos, R. P. de Vries, R. J. Woods, N. B. Tito, G.-J. Boons, E. van der Vries, J. Huskens (2020) Hierarchical Multivalent Effects Control Influenza Host Specificity. *ACS Cent. Sci.* 6(12): 2311-2318.
8. F. Broszeit, O. Aizpurua-Olaizola, **Z. Li**, R. van Beek, G.-J. Boons, R. P. de Vries (2020) Glycomic analysis of erythrocytes reveals extended, poly-lactosamine containing N-glycans as a unique characteristic of guinea pig erythrocytes. Status: In preparation.
9. M. A. Tortorici, A. C. Walls, Y. Lang, C. Wang, **Z. Li**, D. Koerhuis, G.-J. Boons, B.-J. Bosch, F. A. Rey, R. J. de Groot, D. Veasler (2019) Structural basis for human coronavirus attachment to sialic acid receptors. *Nat. Struct. Mol. Biol.* 26: 481–489.
10. W. Du, H. Guo, V. S. Nijman, J. Doedt, E. van der Vries, J. van der Lee, **Z. Li**, G.-J. Boons, F. J. M. van Kuppeveld, E. de Vries, M. Matrosovich, C. A. M. de Haan (2019) The 2nd sialic acid-binding site of influenza A virus neuraminidase is an important determinant of the hemagglutinin-neuraminidase-receptor balance. *Plos Pathog.* 15(6): e1007860.

11. F. Broszeit, N. Tzarum, X. Zhu, N. Nemanichvili, D. Eggink, T. Leenders, **Z. Li**, L. Liu, M. A. Wolfert, A. Papanikolaou, C. Martínez-Romero, I. A. Gagarinov, W. Yu, A. García-Sastre, T. Wennekes, M. Okamatsu, M. H. Verheije, I. A. Wilson, G.-J. Boons, R. P. de Vries (2019) N-glycolylneuraminic acid as a receptor for influenza A viruses. *Cell Rep.* 27(11): 3284-3294.e6.
12. R. J. G. Hulswit, Y. Lang, M. J. G. Bakkers, W. Li, **Z. Li**, A. Schouten, B. Ophorst, F. J. M. van Kuppeveld, G.-J. Boons, B.-J. Bosch, E. G. Huizinga, R. J. de Groot (2019) Human coronaviruses OC43 and HKU1 bind to 9-O-acetylated sialic acids via a conserved receptor-binding site in spike protein domain A. *PNAS*, 116(7): 2681-90.
13. W. Du, M. Dai, **Z. Li**, G.-J. Boons, B. Peeters, F. J. M. van Kuppeveld, E. de Vries, C. A. M. de Haan (2018) Substrate binding by the 2nd sialic acid-binding site of influenza A virus N1 neuraminidase contributes to enzymatic activity. *J. Virol.* 92(20): e01243-18.
14. M. J. G. Bakkers, Q. Zeng, L. J. Feitsma, R. J. G. Hulswit, **Z. Li**, A. Westerbeke, F. J. M. van Kuppeveld, G.-J. Boons, M. A. Langereis, E. G. Huizinga, and R. J. de Groot (2016) Coronavirus receptor switch explained from the stereochemistry of protein-carbohydrate interactions and a single mutation. *PNAS*, 113(22): E3111-9.

EDUCATION AND EMPLOYMENT

post-doctoral	Department of Human Genetics, Leiden University Medical Center, Leiden, the Netherlands. Supervisor: Dr. Vered Raz Department of Bionanoscience, Delft University of Technology, Delft, the Netherlands. Co-supervisor: Prof. Dr. Chirlmin Joo	01/2021 – current
Ph.D.	Department of Chemical Biology and Drug Discovery (CBDD), Bijvoet Center for Biomolecular Research, Utrecht University, Utrecht, the Netherlands. Promoter: Prof. Dr. Geert-Jan Boons Department of Chemistry, University of Georgia (UGA), Athens, GA, USA. Transferred to Utrecht University. Advisor: Prof. Dr. Geert-Jan Boons	04/2016 – 12/2020 09/2013 – 03/2016
B.Sc.	Department of Pharmaceutical Sciences, Peking University (PKU), Beijing, China. Advisor: Prof. Dr. Xinshan Ye	09/2009 – 06/2013

RESEARCH EXPERIENCES

Research projects:

HG, LUMC / BN, TUDelft

01/2021 – current

An integrated approach to study cell-surface (glyco)RNA.

Contribution: conceived the study, RNA-seq library preparation and NGS data analysis, identified a probe for detecting cell surface RNAs.

CBDD, Utrecht University

04/2016 – 12/2020

1. *Development of novel synthetic strategy for O-acetylated sialosides and the use thereof to study human respiratory and related animal viruses.*
Contribution: conceived the study, synthesized compounds, analyzed compounds using mass spectrometry (MS) and NMR, performed glycan microarray studies, wrote the paper.
2. *Chemoenzymatic synthesis of stabilized N-acetylated sialoside analogs as probes for acetylated sialic acid binding lectins.*
Contribution: conceived the study, synthesized compounds, analyzed compounds using MS and NMR, performed glycan microarray studies, advised molecular dynamic studies, wrote the paper.
3. *Chemoenzymatic synthesis of N-glycosylated sialosides as probes for human respiratory and related animal viruses.*
Contribution: conceived the study, synthesized compounds, analyzed compounds using MS and NMR, performed glycan microarray studies, wrote the paper.
4. *Chemoenzymatic synthesis of ganglioside glycans GD2 and GD3.*
Contribution: synthesized compounds, analyzed compounds using mass MS and NMR.
5. *Development of hybrid immunogen for envelope trimer-based HIV-1 vaccine by TLR7 agonist glycoconjugation.*
Contribution: conceived the study, synthesized compounds, analyzed compounds using MS and NMR, performed protein conjugation, purification, drug quantification, native-PAGE and ELISA, wrote the paper.
6. *Investigating the role of phase variation of E. coli. K1 capsular antigens.*
Contribution: conceived the study, synthesized compounds, characterized compounds using MS and NMR, supervised junior PhD.
7. *Designing selective inhibitors for human sialyltransferases.*
Contribution: conceived the study, advised junior PhD.
8. *Investigating role of ABO blood antigens in SARS-CoV-2 infection.*
Contribution: conceived the study.

Dept. Chemistry, UGA, USA

09/2013 – 03/2016

Chemoenzymatic synthesis of native and unnaturally-modified sialosides as probes for influenza A viruses.

Contribution: synthesized compounds, analyzed compounds using MS and NMR, performed binding studies.

Dept. Pharmaceutical Sciences, PKU, China

09/2009 – 06/2013

Thesis title: *Development of novel amidoxy- and imidoxy-type leaving groups for chemical glycosylation.*

Other scientific experiences:

CBDD, Utrecht University

04/2016 – current

1. Developed an executable for automated extraction of the data output from GlycoMOD tool (ExPASy) for processing large-scale glycomic data.
2. Assisted in ERC Advanced grant proposal writing for Prof. Geert-Jan Boons.
3. Reviewed a manuscript submitted to *Journal of the American Chemical Society*.

LABORATORY SKILLS

Mammalian cell culturing

Organic synthesis

RNA biochemistry

Glycoprotein modification

RNA-seq library preparation

Liquid chromatography

RNA-seq data analysis

NMR analysis

Confocal imaging

Glycan microarray

Flow cytometry

Glycomic analysis

TEACHING AND COACHING EXPERIENCES

HG, LUMC

02/2021 – current

Master student supervision

Provided daily supervision and advice for research projects.

CBDD, Utrecht University

09/2018 – 12/2020

Junior PhD student Coaching

Provided daily supervision and advice for research projects.

Dept. Chemistry, UGA

09/2013 – 06/2014

Teaching Assistant | Course: CHEM 2212 - Mod Org Chem II

Supervised weekly laboratory sessions, proctored and graded exams, provided office hours for undergraduate students.

AWARDS AND HONORS

Chinese Government Award for Outstanding Self-Financed Students Abroad (2020)

SEMINARS AND POSTER COMMUNICATIONS

- | | |
|---------|---|
| 06/2019 | <i>Chemoenzymatic syntheses of hydrolysis/migration-resistant N-acetylated sialoside analogs.</i> (poster)
Carbohydrate Gordon Research Conference, Hong Kong, China |
| 10/2018 | <i>A platform for streamlined characterization of coronaviruses.</i> (7-minute talk)
Department Day presentation competition, Utrecht, the Netherlands |
| 06/2018 | <i>A platform for streamlined characterization of coronaviruses.</i> (20-minute talk)
David de Wied Colloquium, Utrecht University, Utrecht, the Netherlands |
| 05/2018 | <i>A platform for streamlined characterization of coronaviruses using synthetic acetylated sialosides.</i> (poster)
1 st International Symposium on Glycovirolgy, Schöntal, Germany |
| 04/2018 | <i>Chemoenzymatic synthesis of acetylated sialosides for streamlined analysis of coronaviruses.</i> (20-minute talk)
24 th Bijvoet Symposium, Utrecht, the Netherlands |
| 03/2015 | <i>Antibody-drug conjugates in cancer therapy.</i> (45-minute talk)
Chemistry Department Seminar, University of Georgia, Athens, GA, USA |

LANGUAGES

English (proficient) and Mandarin (native)