

# Receptor binding specificities of influenza A viruses beyond the sialic acid linkage

Cindy Spruit

# **Receptor binding specificities of influenza A viruses beyond the sialic acid linkage**

## **Specificiteit van de receptorbinding van influenza A virussen verder dan de sialzuurkoppeling**

(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

maandag 13 mei 2024 des middags te 4.15 uur

door

**Cindy Maria Spruit**

geboren op 29 november 1994  
te Leiderdorp

**Promotor:**

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

**Copromotor:**

Dr. R.P. de Vries

**Beoordelingscommissie:**

Prof. dr. A. Gröne

Prof. dr. C.H. Hokke

Prof. dr. C.R. Parrish

Dr. M. Richard

Prof. dr. J.A. Stegeman

# Table of contents

<b>Chapter 1</b>	General introduction	4
<b>Chapter 2</b>	Contemporary human H3N2 influenza A viruses require a low threshold of suitable glycan receptors for efficient infection	24
<b>Chapter 3</b>	<i>N</i> -glycolylneuraminic acid in animal models for human influenza A virus	52
<b>Chapter 4</b>	<i>N</i> -glycolylneuraminic acid binding of avian and equine H7 influenza A viruses	70
<b>Chapter 5</b>	H7 influenza A viruses bind sialyl-LewisX, a potential intermediate receptor between species	94
<b>Chapter 6</b>	Synthesis of tri-antennary <i>N</i> -glycans terminated with sialyl-LewisX reveals the relevance of glycan complexity for influenza A virus receptor binding	122
<b>Chapter 7</b>	Summary and future outlook	150
<b>Appendices</b>	Nederlandse samenvatting List of publications Curriculum vitae Acknowledgements	166

All rights reserved. No part of this thesis may be reproduced, stored, or transmitted in any way or by any means without the prior permission of the author, or when applicable, of the publishers of the scientific papers. Permissions of the publishers were granted for the re-use of all figures and articles presented in this dissertation.

The printing of this dissertation was sponsored by Infection & Immunity Utrecht and the Utrecht Institute for Pharmaceutical Sciences. Printed by [www.proefschriften.nl](http://www.proefschriften.nl).



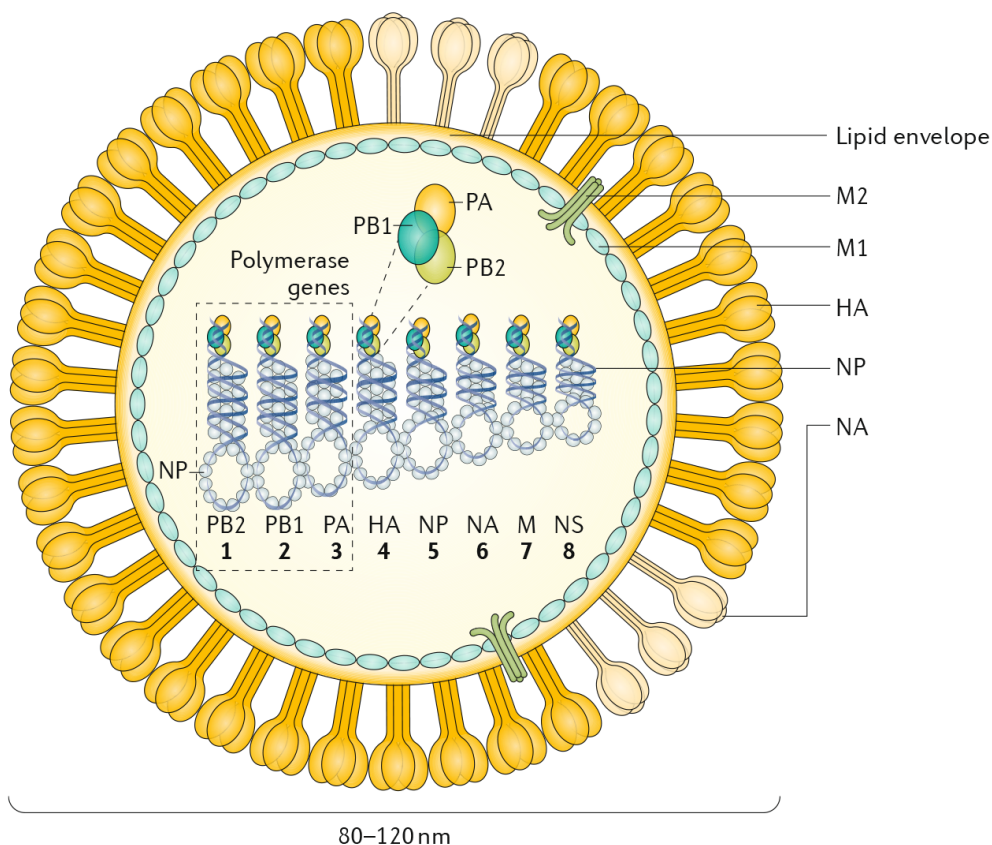
# Chapter 1

## **General introduction**

---

## Influenza A virus

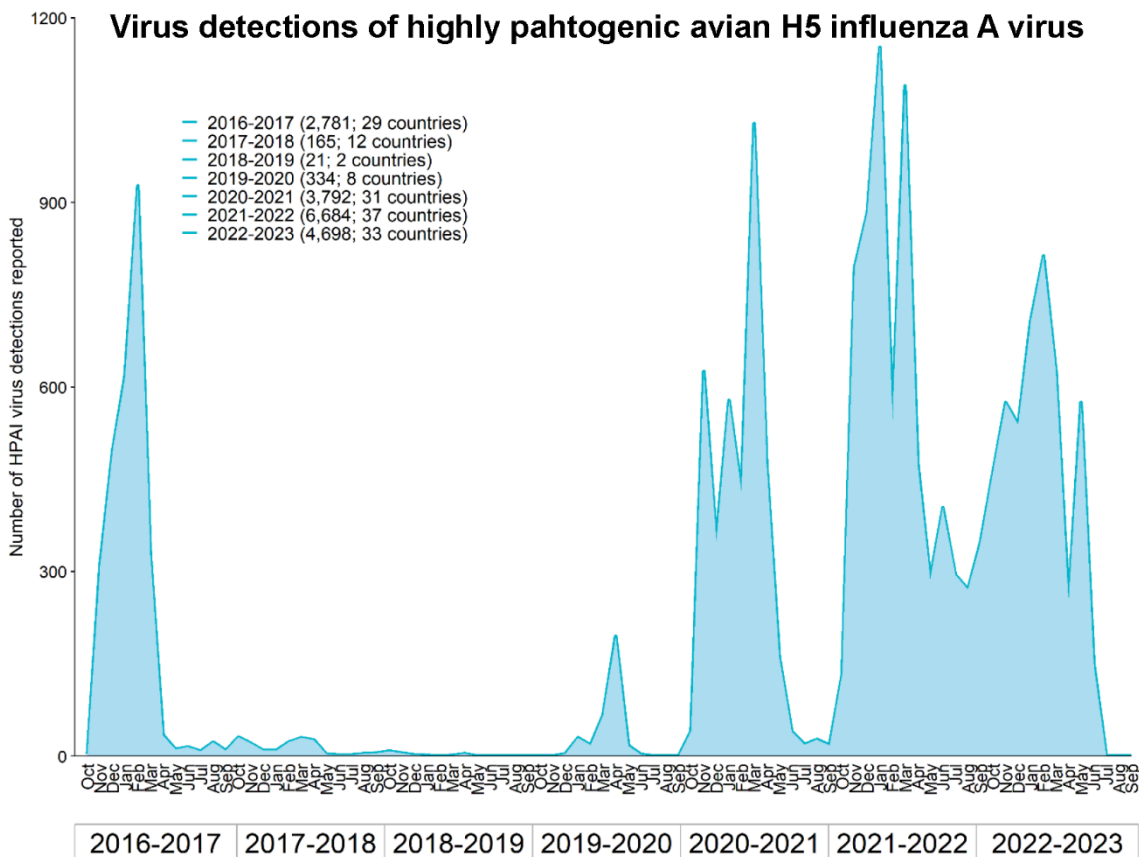
Influenza viruses belong to the *Orthomyxoviridae* family and are divided into four genera (A, B, C, and D). Whereas influenza B, C, and D viruses generally only present mild symptoms, the effects of influenza A viruses (IAVs) can be more severe [1]. Furthermore, IAVs are the most common of the influenza viruses and have the potential to cause pandemics in humans [2]. Virus particles are enveloped and proteins are encoded on eight single-stranded negative-sense RNA segments with a total length of 12-14kb (Fig. 1) [3]. IAVs encode for three subunits of the viral polymerase (PB1, PB2, and PA), a nucleoprotein (NP), a matrix protein (M1), a membrane protein (M2), two nonstructural proteins (NS1 and NS2), and the surface glycoproteins neuraminidase (NA) and hemagglutinin (HA) [4]. In IAVs, 18 antigenically distinct HA subtypes and 11 NA subtypes are distinguished [2, 3] that evolve continuously due to immune selection and adaptations to hosts, resulting in antigenic and functional drift [4]. Additionally, recombination events (antigenic shift), in which genetic fragments are exchanged between two virus particles, can occur when one individual is infected with two different IAVs simultaneously [1, 5].



**Fig 1. Representation of an influenza A virus particle.** The figure indicates the lipid envelope, eight negative-sense RNA fragments encoding the different IAV proteins, and proteins that are present in the virus particle. Figure from [2].

## Avian influenza A viruses

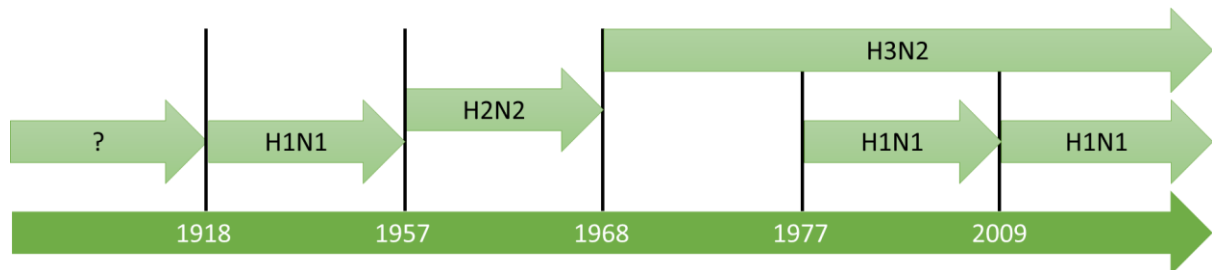
Waterfowl are the main host reservoir for IAV worldwide [3] and, therefore, recombination events can easily occur in these hosts, as has happened for some H5N1 and H7N9 viruses [6]. The coevolution of birds and IAV causes infections to be often asymptomatic in avian species [7]. However, low pathogenic avian influenza viruses of the H5 and H7 subtypes can become high pathogenic variants when the HA proteolytic cleavage site mutates towards multiple basic amino acids, with mortality rates approaching 100% [7, 8]. Besides infection in wild birds, the consequences of IAV infections in poultry (especially of H5, H7, and H9) can be severe. Since 2005, H5 and H7 IAVs have caused plentiful outbreaks in avian species (Fig. 2) and were responsible for the death of at least 422 million domestic birds [9]. Especially the viruses which arose from the A/goose/Guangdong/1/1996 (Gs/GD) lineage of H5 have been circulating in both wild birds and poultry since 1996 and are causing major problems [8]. These Gs/GD H5 viruses mutated towards the highly pathogenic 2.3.4.4b HA subclade which has been circulating since 2020 and caused over 4000 disease outbreaks. Besides the 2.3.4.4b viruses causing extremely high mortality in birds, significant transmission to mammals occurs [10-14]. Furthermore, occasionally other avian IAVs (H5, H6, H7, H9, and H10) are transmitted to humans [6] and avian H5, H7, and H9 IAVs have all caused epidemics in humans [1]. H5 and H7 avian IAVs have so far caused 2634 human cases with over 1000 fatalities [9].



**Fig 2. Overview of the number of highly pathogenic avian H5 influenza A virus detections in 37 countries in domestic and wild birds.** Modified from [14].

## Human influenza A viruses

In humans, seasonal influenza strains from the H3N2 and H1N1 subtypes are circulating [15]. Additionally, humans are sporadically infected with avian IAVs (H3, H5, H6, H7, H9, and H10) [13]. An estimated four million severe cases occur annually, resulting in approximately half a million deaths [1, 2]. The symptoms of IAV infection in humans include fever, headache, cough, sore throat, myalgia, nasal congestion, weakness, and loss of appetite [16]. In some cases, complications affecting the heart, central nervous system, and other organs are observed. Pneumonia due to IAV or a secondary bacterial infection in the lower respiratory tract can be lethal [2]. Four IAV pandemics (Fig. 3) of the subtypes H1N1 (1918), H2N2 (1957), H3N2 (1968), and H1N1 (2009) have struck [15, 17]. Additionally, in 1977, H1N1 viruses started to circulate in the human population without causing a pandemic [2]. Vaccines for IAV are updated every year to correct for the continuous antigenic drift of especially the viral surface glycoprotein HA, which is the most abundant protein on the virus surface and a major antigenic determinant [2, 6, 17, 18].



**Figure 3. Timeline of seasonal influenza A virus in humans.** Figure adapted from [2].

## Influenza A virus in other mammals

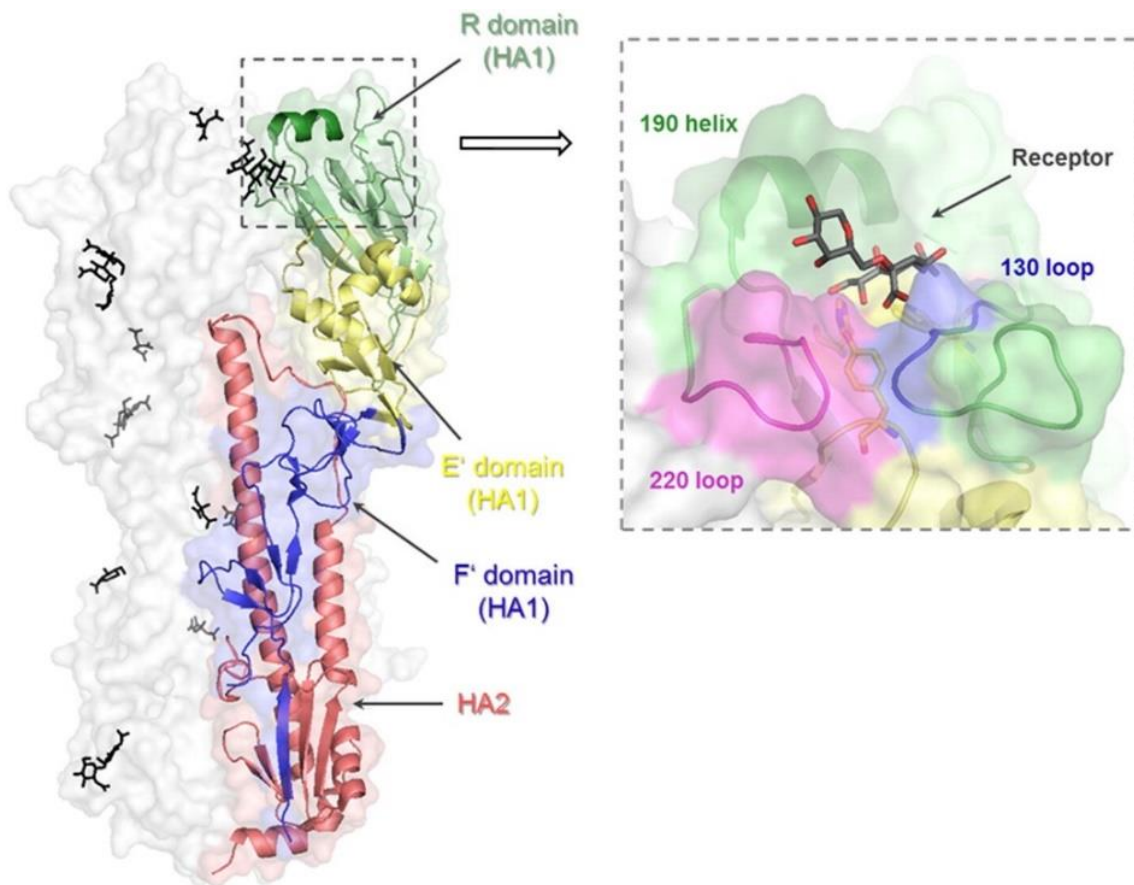
Although most studies investigate either human or avian viruses, IAVs also occur in different mammalian species (Fig. 4). In horses, only two subtypes of IAVs are reported, which are divergent from other IAVs. The H7N7 viruses circulated in the equine population until 1979 but are believed to be extinct since then. Currently, H3N8 viruses are circulating in horses worldwide. Equine influenza viruses usually cause mortality of around 1%, but some lineages have had mortalities up to 20% [19, 20].

The equine H3N8 viruses have also infected dogs. These viruses adapted to dogs afterward and continued to circulate for about 20 years without reinfections of horses [21-24]. Additionally, H3N2 viruses are endemic in dogs and, sporadically, dogs are infected with H1 and H5 IAVs [21-24]. Cats have been infected with H1, H3, H5, and H7 IAVs, but the viruses rarely circulate in cats [22, 23].

In swine, subtypes H1, H3, H4, and H9 are detected, of which many are endemic in various regions of the world. These viruses cause sickness in most animals, but mortality is generally low, although it can be 10-15% in naïve pigs [25, 26]. IAVs have been reported to transmit regularly between swine and humans, although little human-to-human transmission of swine IAVs has been reported. Additionally,



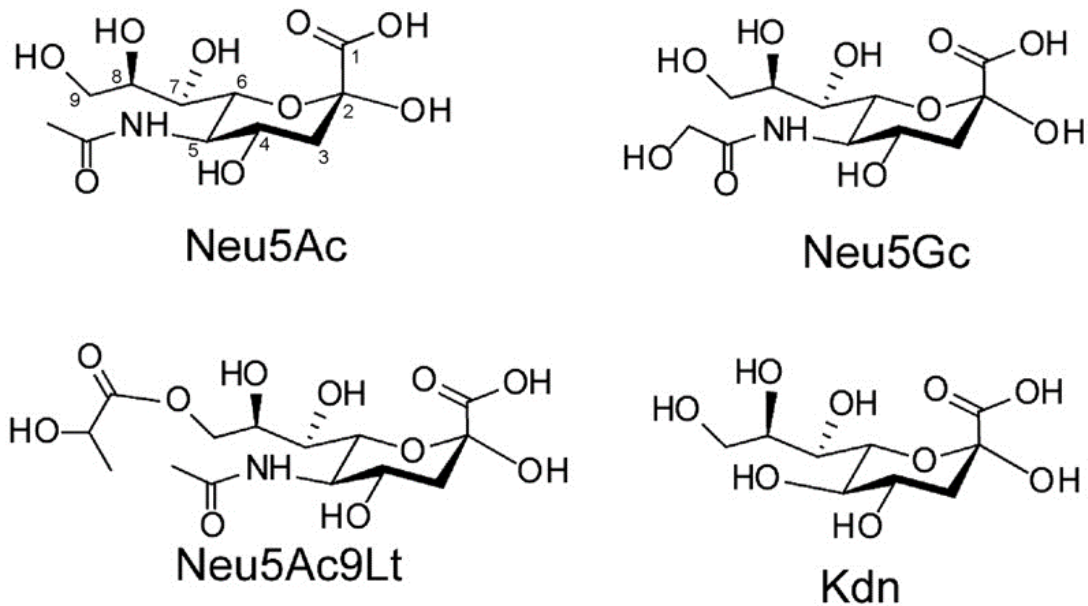
site (RBS) through hydrogen bonds and van der Waals contacts [37]. The RBS consists of three main domains, which are the 130-loop (residues 134–138), the 190-helix (residues 188–195), and the 220-loop (residues 221–228) (Fig. 5). The main highly conserved residues that interact with the receptor are Y98, W153, and H183 [18, 36, 37]. The major antigenic sites of IAV are present in the HA and overlap with the RBS. Immune evasion can be achieved by escape mutations and by mutations that enable additional *N*-glycosylation of the HA, thereby shielding the previously used antigenic sites. Therefore, the immune evasion, which is necessary for the virus to be maintained in the population, can change and affect the receptor binding [38–41].



**Fig 5. The receptor binding site of influenza A virus.** The receptor binding site is present in the hemagglutinin and consists of the 130-loop, 190-helix, and 220-loop. Figure from [36].

The receptors for IAVs are sialic acids (Sias), which are a group of derivate sugars with nine carbons with neuraminic acid as the core structure (Fig. 6) [42, 43]. Three main forms of Sias are distinguished by the modification of the 5-carbon position, which are *N*-acetylneuraminic acid (Neu5Ac), *N*-glycolylneuraminic acid (Neu5Gc), and 2-keto-3-deoxynononic acid (KDN). Additionally, the 4, 5, 7, 8, and 9-positions of the Sias can be modified with, among others, acetyl-, sulfo-, methyl-, lactam-, phospho-, and lactyl-groups, thereby forming more than 50 structurally distinct molecules [43, 44]. These Sias are capping glycans on glycoproteins and glycolipids, which are present on the cell surface [45, 46].





**Fig 6. Examples of different forms of sialic acids.** Modified from [47].

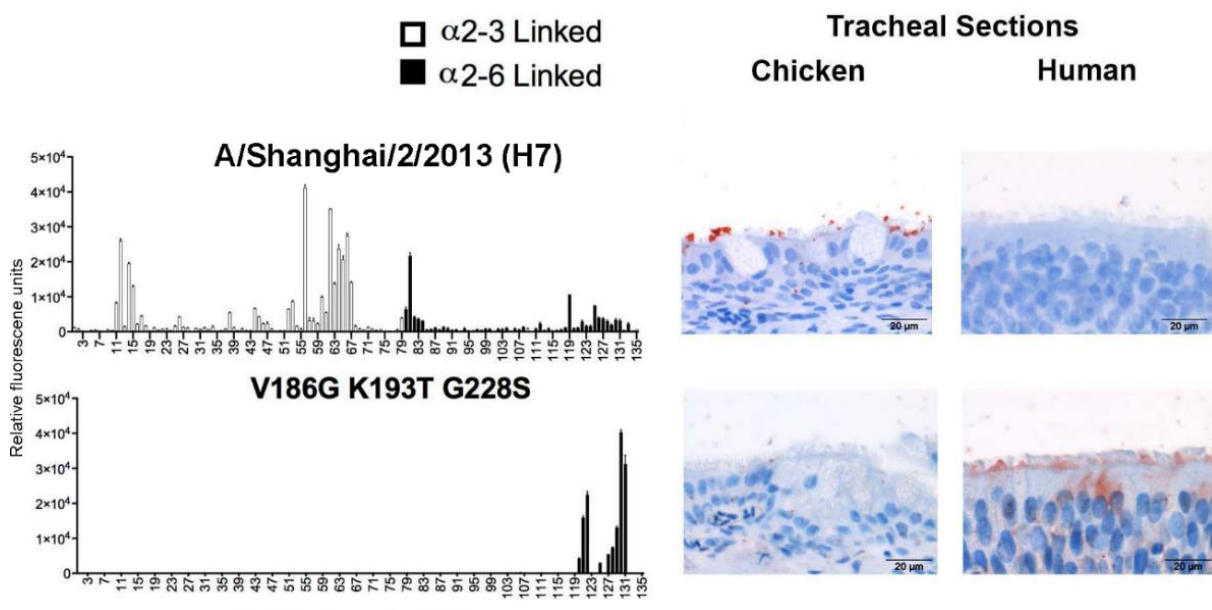
The vast majority of IAVs use terminal Neu5Ac as their receptor, linked to the penultimate galactose (Gal) on a glycan by either an  $\alpha$ 2,3- or  $\alpha$ 2,6-linkage. The presence of  $\alpha$ 2,3- and  $\alpha$ 2,6-linked Neu5Ac varies between and within species [42, 48, 49] and is summarized for mammals in **chapter 2**. The receptor binding of human and avian IAVs is well characterized and, therefore, here we focus on the receptor binding of those IAVs. Classically, the  $\alpha$ 2,3-linked Neu5Ac is the receptor of avian IAVs, while the  $\alpha$ 2,6-linked Neu5Ac is considered the receptor for human IAVs, corresponding to the most abundant Sias at the epithelial surfaces that are infected [2, 6, 18, 50]. Therefore, the ability of the HA to bind either  $\alpha$ 2,3- or  $\alpha$ 2,6-linked Neu5Ac is commonly considered a species barrier between avian and mammalian species [6, 42, 46]. However, receptor binding of IAVs is much more complex than just binding to  $\alpha$ 2,3- or  $\alpha$ 2,6-linked Neu5Ac since the complete glycan structure, although understudied, is of great relevance [42, 46, 51-53].

### **$\alpha$ 2,6-linked Neu5Ac as the receptor for human influenza A viruses**

Human IAVs are known to bind to  $\alpha$ 2,6-linked Neu5Ac, for which mutations E190D and D225G in H1 viruses and Q226L and G228S for H2 and H3 viruses are largely responsible [46, 54, 55]. Since its introduction in 1968, the receptor binding specificity of human H3N2 viruses has developed towards strictly binding  $\alpha$ 2,6-linked Neu5Ac, gradually losing affinity to  $\alpha$ 2,3-linked Neu5Ac [40, 56]. At the same time, these H3N2 viruses evolved to evade the human immune system. Since 1968, the number of *N*-glycosylation sites in the HA head, in which the RBS is located, has increased from three to eight per monomer [38]. Immune evasion is thought to be the reason that recent H3N2 viruses (especially of the 3C.2a subclade) evolved to binding a more restricted binding repertoire, only binding elongated glycans with multiple *N*-acetylglucosamine (LacNAc) repeating units terminating in an  $\alpha$ 2,6-linked Neu5Ac [38, 50, 56-62].

## $\alpha$ 2,3-linked Neu5Ac as the receptor for avian influenza A viruses

It is well recognized that avian IAVs bind to  $\alpha$ 2,3-linked Neu5Ac as their receptors. When studying the receptor binding of avian IAVs, most studies have concentrated on the risk of avian IAVs spreading to humans. This species barrier (from binding  $\alpha$ 2,3-linked to  $\alpha$ 2,6-linked Neu5Ac) can be overcome when the HA mutates at specific locations in or near the RBS, which enables or enhances interactions or removes an *N*-glycosylation site to free up binding sites [46, 63-67]. Some H5 and contemporary North American H7 avian viruses can already bind both  $\alpha$ 2,3- and  $\alpha$ 2,6-linked Neu5Ac [5, 68, 69]. Two well-known mutations to play a role in the receptor switching from  $\alpha$ 2,3-linked to  $\alpha$ 2,6-linked Neu5Ac are Q226L and G228S (H3 numbering) [46, 63, 64, 66, 67]. Additionally, many other mutations were found that can induce binding to  $\alpha$ 2,6-linked Neu5Ac by themselves or in combination, which is summarized elsewhere for avian H5 and H7 IAVs [36, 54, 70-73]. The specificity is most often changed towards a dual receptor specificity of both  $\alpha$ 2,3-linked and  $\alpha$ 2,6-linked Neu5Ac and rarely towards specific binding to  $\alpha$ 2,6-linked Neu5Ac and can be investigated using, for example, binding studies on glycan microarrays and tissue sections (Fig. 7).



**Fig 7. The binding specificity of wild-type and mutant (V186G, K193T, G228S) HA of A/Shanghai/2/2013 (H7) was investigated.** The mutations switched the receptor binding specificity from  $\alpha$ 2,3-linked Neu5Ac to  $\alpha$ 2,6-linked Neu5Ac, as shown on the glycan array (left) and on chicken (high in  $\alpha$ 2,3-linked Neu5Ac) and human (high in  $\alpha$ 2,6-linked Neu5Ac) tracheal tissues (right, red discoloration indicates HA binding). Modified from [74].

## Different glycan cores as potential influenza A virus receptors

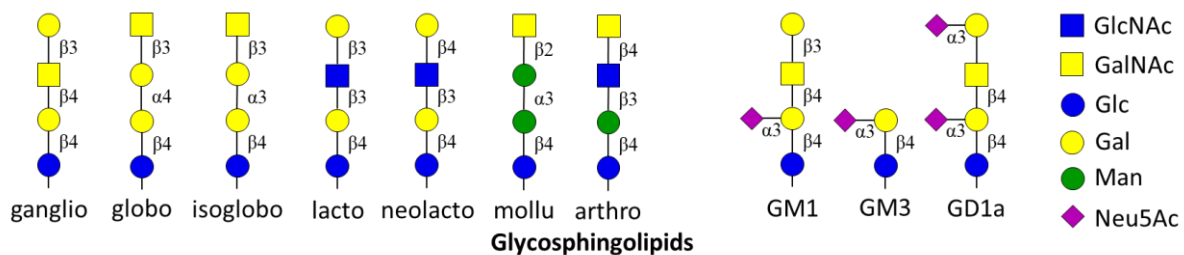
The specificity towards  $\alpha$ 2,3- or  $\alpha$ 2,6-linked Neu5Ac is an oversimplification of the zoonotic abilities of IAVs. The relevance of the complete glycan structure is increasingly recognized [42, 46, 51-53]. Sias can be presented on different glycan



cores, of which the three main cores are *N*-glycans, *O*-glycans, and glycolipids. Furthermore, vast variation occurs in the exact glycan structures, for example in the number and length of branches and the addition of sugar moieties or chemical groups. The exact role of the different glycans in IAV infection is currently unknown [67]. In this dissertation, we aim to further understand the role of the exact glycan structure on the receptor binding of IAVs.

### Glycosphingolipids

The first group of glycans are glycosphingolipids (GSLs), of which the first sugar moiety is either a galactose (Gal) or glucose (Glc) linked to a ceramide lipid, respectively forming GalCer and GlcCer [75]. Since GalCer structures are seldom extended, the GlcCer structures are most relevant for IAVs. The further extension of the GlcCer structures determines the series in which the GSLs are divided, either the ganglio-, globo-, isoglobo-, lacto-, neolacto-, molli-, or arthro-series (Fig. 8) [75]. GSLs are components of all vertebrate plasma membranes [76] and have been specifically analyzed in the respiratory tract of humans and ferrets, in which many elongated and sialylated structures were found [77, 78].



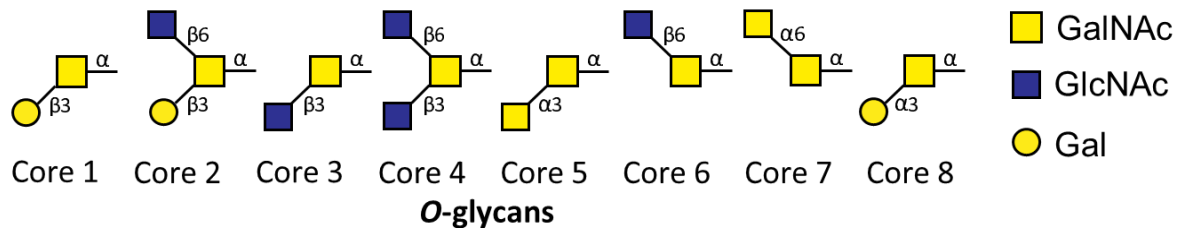
**Fig 8. Examples of glycosphingolipids.** Different glycosphingolipid series (ganglio-, globo-, isoglobo-, lacto-, neolacto-, molli-, and arthro-) and examples of sialylated glycosphingolipids (GM1, GM3, and GD1a) consisting of *N*-acetylglucosamine (GlcNAc), *N*-acetylgalactosamine (GalNAc), glucose (Glc), galactose (Gal), mannose (Man), and *N*-acetylneuraminic acid (Neu5Ac).

For IAV binding, sialylated GSLs are relevant, which can be extended with 10 or more sugar moieties (often Glc, Gal, GlcNAc (*N*-acetylglucosamine), and GalNAc (*N*-acetylgalactosamine)) and at least one Neu5Ac [76, 79, 80]. Examples of common sialylated GSLs are GM1, GM3, and GD1a (Fig. 8). To date, the role of GSLs in IAV infection remains unclear. IAVs are known to efficiently bind GSLs [53, 81-86], which may aid in the infection. The presence of GSLs is, however, not essential for IAV infection [87, 88].

### O-linked glycans

The second class of glycans are *O*-linked glycans, which are naturally linked to serine or threonine residues in a glycoprotein. All *O*-glycans are built from the first *N*-acetylgalactosamine (GalNAc) moiety, after which they can be extended into eight core structures [89] (Fig. 9). *O*-glycans are omnipresent and the exact *O*-glycan structures in the respiratory tract of swine, humans, and ferrets have been

investigated, observing a high abundance of sialylated core 1 and core 2 O-glycans [77, 90-92].

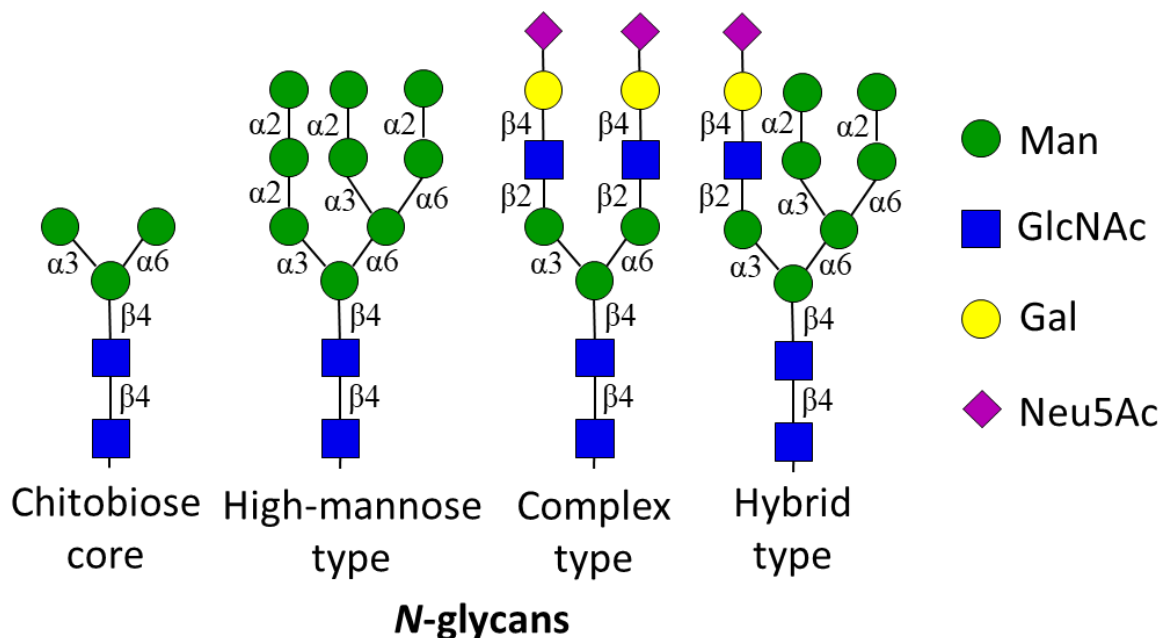


**Fig 9.** There are eight different O-glycan cores, consisting of N-acetylgalactosamine (GalNAc), N-acetylglucosamine (GlcNAc), and galactose (Gal) moieties.

Studies investigating O-glycans as potential receptors for IAV are underrepresented and often binding of IAVs to O-glycans on glycan microarrays is not thoroughly analyzed [59, 93, 94]. Nevertheless, O-glycans were suggested to be involved in IAV replication [95-97] and binding of IAVs to O-glycans has been observed [62, 91].

### N-linked glycans

The third class of glycans consists of N-glycans, which are linked to an asparagine (Asn/N) in a protein. All N-glycans contain a chitobiose core, which can then be extended into three types of N-glycans (Fig. 10) [89]. The N-glycans in the respiratory tract of humans, ferrets, pigs, and chickens have been analyzed and a wide variety of N-glycans (among which elongated glycans and glycan modifications) were found [58, 77, 78, 90-92, 98-100].

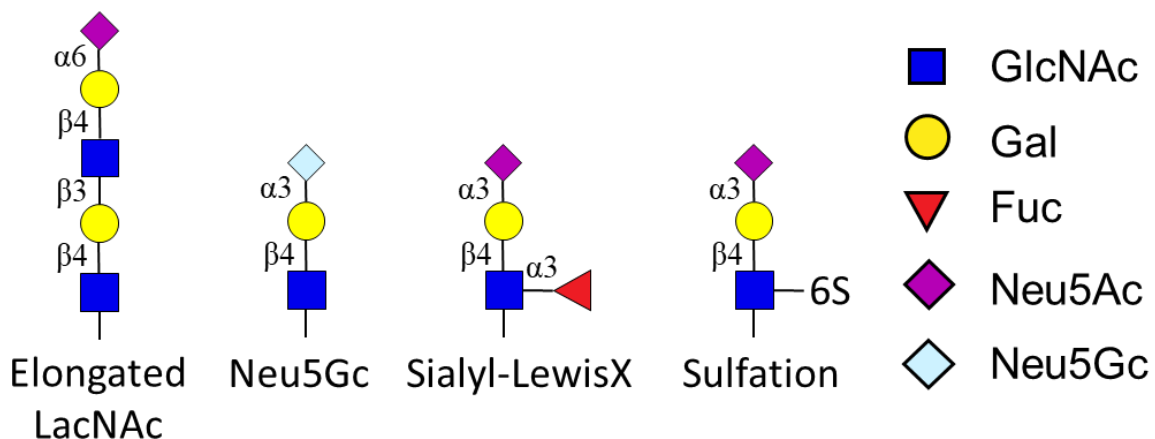


**Fig 10.** N-glycan core and types consist of mannose (Man), N-Acetylglucosamine (GlcNAc), galactose (Gal), and N-acetylneuraminic acid (Neu5Ac).

Since complex *N*-glycans are considered to be the most relevant receptors for IAVs [101, 102] and their isolation is relatively straightforward, the binding of IAVs to *N*-glycans has gained the most attention [103-105]. However, the influence of the exact glycan structures (for example, branching, additional sugar moieties, and different Sias) is largely unknown. Therefore, in **chapter 4**, we investigated the binding of IAVs to different Sias (Neu5Ac and Neu5Gc) presented on *N*-glycans. In **chapter 5 and 6**, we investigated the receptor binding to tri-antennary *N*-glycans. We showed that IAVs generally bind stronger to *N*-glycans than to linear glycans, and some *N*-glycan structures were preferred over others by certain IAVs, highlighting the importance of the exact glycan structure.

## The impact of LacNAc modifications on influenza A virus receptor binding

The different glycan cores can be extended with many different epitopes, which are potentially more important for IAV than differences in the core structures [46, 67]. Typically, IAV receptors consist of a Neu5Ac presented on the galactose (Gal) of a LacNAc repeating unit, which consists of a Gal  $\beta$ 1,4-linked to a *N*-acetylglucosamine (GlcNAc) moiety. A wide variety of modifications can occur on LacNAc repeating units, such as sulfation, fucosylation, or a combination thereof (Fig. 11). Furthermore, variation is possible in the number and length of the branches of especially *N*-glycans. Additionally, different forms of (modified) Sias can be displayed at the non-reducing end. The impact of all these variations on IAV binding is largely unknown, especially in the context of biologically relevant glycans such as *N*- or *O*-glycans. Therefore, in this dissertation, we aimed to increase our understanding of these modifications on IAV receptor binding.



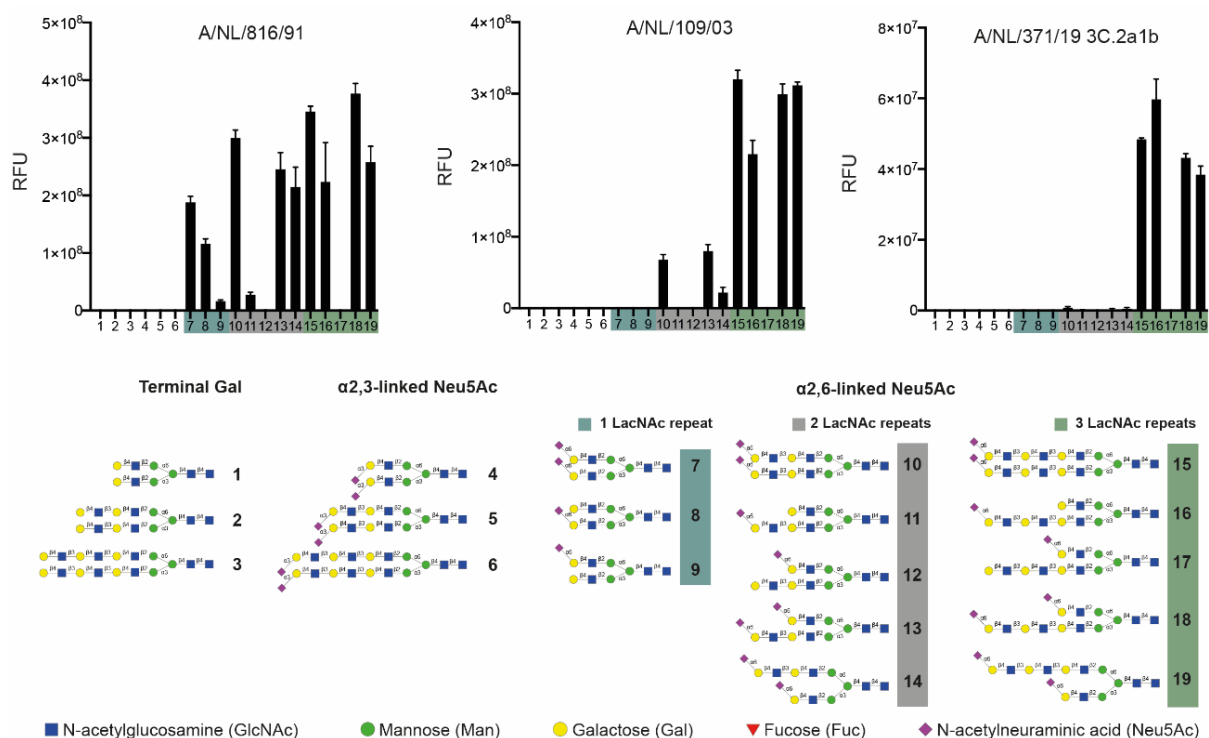
**Fig 11. Examples of terminal epitopes that can be presented on glycans.** The glycans consist of *N*-acetylglucosamine (GlcNAc), galactose (Gal), fucose (Fuc), *N*-acetylneuraminic acid (Neu5Ac), and *N*-glycolylneuraminic acid (Neu5Gc).

### Number of LacNAc repeating units

One relevant variation is the length of the branches on which the Neu5Ac is presented. The length of a glycan is determined by the number of LacNAc repeating units. Elongated *N*-glycans with at least three consecutive LacNAc

repeating units terminated with an  $\alpha$ 2,6-linked Neu5Ac were shown to be present in low quantities (at most 0.3% of the total *N*-glycans) in the respiratory tract of humans, ferrets, and pigs [58, 77, 78, 90, 91].

Glycans with multiple consecutive LacNAc repeating units were shown to be the preferred receptor for pandemic H1N1 viruses (starting from 2009) [105]. Moreover, contemporary H3N2 viruses only bind to glycans having multiple LacNAc repeating units (Fig. 12) [38, 50, 56-62]. The binding specificity for elongated glycans is most obvious for H3N2 IAVs of subclade 3C.2a that emerged in 2014 [106] since those require at least three subsequent LacNAc repeating units for binding [103]. This specificity for longer glycans is thought to be caused by a combination of residues in the 150-loop (155T, 156H, 159Y), 190-helix (186G, 194L), and 220-loop (220R, 222R, 225D) [38, 52]. However, the increased specificity came at a cost since the affinity for the receptors has been decreasing since 2000 [39, 107, 108]. The introduction of asparagine at position 133 or threonine at position 135 introduced an *N*-glycosylation site, resulting in a decreased binding affinity, potentially by partially shielding the RBS [39, 40]. Furthermore, the introduction of asparagine at position 225 of the HA decreased the binding by 10-fold, which appeared to be caused by the receptor sitting less deep in the HA binding pocket [107].



**Fig 12. Evolution of human H3N2 viruses (from 1991, 2003, and 2019) towards only binding glycans with at least three consecutive LacNAc repeating units.** Modified from [103].

The receptor binding properties of especially 3C.2a viruses make it increasingly difficult to isolate these viruses, thereby greatly hampering the further study of these viruses [39, 59, 109-111]. *N*-glycan analysis of the cells that are currently used

for virus isolation and propagation (MDCK and hCK cells) indicated a low abundance of glycans with at least three successive LacNAc repeating units terminating in an  $\alpha$ 2,6-linked Neu5Ac [112, 113], thereby posing the question of whether an increase of these receptors would also increase virus isolation and propagation. Therefore, in **chapter 2**, we modified cells to present a higher number of elongated glycans with at least three consecutive LacNAc repeating units.

### Sialic acid modifications

As described above, the sialic acids (Sia) are a group of derivate sugars with nine carbons with neuraminic acid as the core structure (Fig. 6) [42-44]. Additional modifications can be present at the 4-, 5-, 7-, 8-, and 9-positions. All types of modifications can be found in vertebrates [43, 114-116]. Most IAVs use terminal Neu5Ac as their receptor. KDN has so far not been reported to be bound by wild-type IAVs [117]. Only in rare cases, binding of IAV to further modified Sias was observed. Some H3 IAVs have been reported to bind 9-O-acetylated or 4-O-acetylated Neu5Ac [118-121], while other H1 and H3 viruses bound 9-O-lactoyl-acetylneuraminic acid (Neu5Ac9Lt) or Neu5Gc containing an additional methyl group (Neu5OMeGc) [47, 117].

Specific binding to Neu5Gc has only been described for extinct equine H7N7 viruses [104, 122], although binding to Neu5Gc, both in the  $\alpha$ 2,3-linked and  $\alpha$ 2,6-linked form, has also been described for IAVs from the subtypes H1, H2, H3, H7, and H11 [118, 123-128]. Neu5Gc can be formed in species that express an active form of the enzyme cytidine monophosphate-*N*-acetylneuraminic acid hydroxylase (CMAH), which facilitates the hydroxylation of Neu5Ac to convert it to Neu5Gc. The presence of Neu5Gc in different species is addressed in several publications [129-133]. In short, Neu5Gc is present in many species, but it is absent in for example humans, ferrets, European dogs, many monkey species, and seals. The presence of Neu5Gc in some hosts, but not others, suggests a potential species barrier for IAVs.

It is currently unknown which amino acids enable binding to Neu5Gc. Furthermore, the relevance of Neu5Gc in IAV infection remains largely unclear and is often not considered when choosing which animals models to use for studying IAVs. Therefore, we investigated the presence of Neu5Gc in animal models relevant for studying human IAVs in **chapter 3** and concluded that ferrets are the most appropriate animal model for studying human IAVs. Additionally, we found that Neu5Gc is not used as a decoy receptor and can therefore, potentially, acts as a functional receptor. Next, we investigated the molecular determinants of Neu5Gc-binding in **chapter 4** and we showed that mutation A135E enabled dual Neu5Ac and Neu5Gc receptor specificity, while adding four mutations enabled specific binding to Neu5Gc.

### Sulfation

One additional LacNAc modification that can occur is sulfation, in which the GlcNAc moiety is sulfated at the 6-position. Sulfated glycans are, so far, only reported to be present in the respiratory and intestinal tract of chickens and

turkeys, as well as the duck colon and pig lung [98, 99, 134]. The exact structure of these sulfated glycans is, however, often unknown since it is difficult to determine the exact position of the sulfate in a complex glycan structure.

Some IAVs of the subtypes H6 and H9 specifically bind sulfated glycans with an  $\alpha$ 2,3-linked Neu5Ac [66, 135, 136]. This specificity is thought to be mediated through amino acids at mainly position 190 (H3 numbering) of the HA [135], at which the absence of a negatively charged amino acid promotes binding to sulfated glycans [51, 136-138]. IAVs from other subtypes (except H8 and H15) are also known to bind to sulfated glycans with an  $\alpha$ 2,3-linked Neu5Ac to a certain degree [138-145]. Furthermore, IAVs from several subtypes (H1, H2, H3, H4, H5, H6, H7, H9, H13, H14, and H16) were reported to bind sulfated sialyl-LewisX (sLe<sup>x</sup>) epitopes, which have an additional  $\alpha$ 1,3-linked fucose to the GlcNAc of a LacNAc [38, 51, 139, 146, 147]. Additionally, glycans with an  $\alpha$ 2,6-linked Neu5Ac can be sulfated and were reported to be bound by some subtypes of IAVs (H1, H3, H7, H9, and H16) [61, 68, 90, 145, 148-151]. Nevertheless, most of these studies are performed with tri- and tetrasaccharides and not in the context of a glycan that occurs in nature, such as *N*-glycans and *O*-glycans. Therefore, it is currently unclear whether sulfated glycans are used as receptors for IAV and whether the presence of sulfated glycans affects infection.

### Sialyl-LewisX epitopes

Sialyl-LewisX (sLe<sup>x</sup>) epitopes are formed when a fucose is  $\alpha$ 1,3-linked to the GlcNAc of a LacNAc that is terminated in an  $\alpha$ 2,3-linked Neu5Ac (Fig. 11). In the trachea and intestine, relevant tissues for IAV infection, sLe<sup>x</sup> epitopes were found to be present in chickens, ducks, turkeys, pheasants, guinea fowls, pigeons, and dogs [98, 134, 142, 152-154]. No sLe<sup>x</sup> epitopes were observed in the respiratory tract of ferrets, swine (trachea and lung), and horses (trachea) [78, 90, 92, 99, 153]. In humans, no sLe<sup>x</sup> was found in the trachea [155], while most studies found sLe<sup>x</sup> to be present in the lung [58, 77, 91, 100]. Using an anti-sLe<sup>x</sup> antibody, we found the sLe<sup>x</sup> epitope to be present in the trachea and intestine of chickens, but not ducks, as described in **chapter 5**. Furthermore, in **chapter 6**, we show that sLe<sup>x</sup> epitopes are present on the tracheal tissue of chickens, turkeys, and guinea fowl, but not pigeons, ducks, horses, pigs, and humans. The exact structures of glycans containing sLe<sup>x</sup> epitopes remain, however, largely unclear due to a lack of techniques to analyze exact glycan structures [91, 92, 100, 156-158].

It was suggested that sLe<sup>x</sup> structures facilitate infection of some avian H7 IAVs [152, 159]. IAVs from all subtypes, except H15, have been shown to bind to some extent to sLe<sup>x</sup> epitopes [137, 139, 148, 151, 153, 160-163]. Some H3 and H5 IAVs bind specifically to sLe<sup>x</sup> structures [86, 154, 160, 161]. In H5 HAs, mutations (222Q/R and 227R) are known that enable binding sLe<sup>x</sup> [142, 160, 161], while in H7 viruses, sLe<sup>x</sup>-binding appears to be caused by 193K [152]. However, due to a lack of sLe<sup>x</sup> epitopes on complex *N*- or *O*-linked glycans on glycan arrays, sLe<sup>x</sup> specificities and the influence of the glycan structure have been understudied. Most of these studies were performed with the tetrasaccharide epitope of sLe<sup>x</sup> and occasionally

linear glycans ending in sLe<sup>x</sup> were used, both of which do not occur in nature as such. Therefore, the function of IAVs binding to sLe<sup>x</sup> is still largely unclear.

In this dissertation, we address this knowledge gap by investigating the binding of different IAVs to sLe<sup>x</sup> epitopes presented on biologically relevant glycan cores. In **chapter 5**, we use tri-antennary *N*-glycans presenting sLe<sup>x</sup> epitopes to investigate the receptor binding of H7 and H15 viruses. We showed that both subtypes strongly bind sLe<sup>x</sup> epitopes, which was previously unknown for H15 viruses, but binding to sLe<sup>x</sup> epitopes appears not to be essential for viral replication. Furthermore, binding to sLe<sup>x</sup> epitopes was only observed when the epitopes were presented on tri-antennary *N*-glycans and not on linear glycans, thereby highlighting the importance of the complete glycan structure for receptor binding. In **chapter 6**, we investigated the receptor binding of avian H5 IAVs to tri-antennary *N*-glycans on which the sLe<sup>x</sup> epitopes were presented on different arms. We showed that the binding to these glycans overshadows the binding to linear glycans and that some H5 viruses prefer to bind to glycans on which the sLe<sup>x</sup> epitope is presented on a specific arm, thereby again highlighting the relevance of the complete glycan structure for receptor binding. In conclusion, in this dissertation, we showed the relevance of presenting sLe<sup>x</sup> epitopes on biologically relevant glycan cores, since the complete glycan structure is relevant for the binding of IAVs.

## Aim and dissertation outline

In this dissertation, we aimed to investigate the fine receptor specificities of IAVs and the molecular determinants responsible for this binding, looking beyond the classically considered  $\alpha$ 2,3-linked and  $\alpha$ 2,6-linked Neu5Ac. We aimed to investigate biologically relevant *N*-glycan cores presenting a diverse repertoire of epitopes, such as elongated branches, multiple branches, Neu5Ac, Neu5Gc, and sLe<sup>x</sup> epitopes.

In **chapter 2**, we produced cells that presented elongated glycans and investigated the binding and infection of contemporary human H3N2 IAVs using these cells.

In **chapter 3**, we studied the presence of Neu5Gc,  $\alpha$ 2,3-linked, and  $\alpha$ 2,6-linked Sias in animal models used to study human IAVs. We also investigated the role of Neu5Gc in IAV infection.

In **chapter 4**, we explored which amino acids in avian H7 and H15 HAs are responsible for switching the receptor binding specificity from Neu5Ac to Neu5Gc.

In **chapter 5**, we examined the binding of Neu5Gc-adapted avian H7 viruses to sLe<sup>x</sup> epitopes and the effect of the Neu5Gc-adaptation on the infection of ducks and chickens.

In **chapter 6**, we studied the binding of avian H5 IAVs to tri-antennary *N*-glycans presenting sLe<sup>x</sup> epitopes on different arms. Additionally, we investigated the presence of sLe<sup>x</sup> epitopes on the tracheal tissues of several avian and mammalian species.



Finally, in **chapter 7**, we summarize the results of all research presented in this dissertation and provide a future outlook.

## References

1. Javanian, M., et al., *A brief review of influenza virus infection*. J Med Virol, 2021. 93(8): p. 4638-4646.
2. Kramer, F., et al., *Influenza*. Nat Rev Dis Primers, 2018. 4(1): p. 3.
3. Hutchinson, E.C., *Influenza virus*. Trends Microbiol, 2018. 26(9): p. 809-810.
4. Steinhauer, D.A. and J.J. Skehel, *Genetics of influenza viruses*. Annu Rev Genet, 2002. 36: p. 305-32.
5. Watanabe, Y., et al., *The changing nature of avian influenza A virus (H5N1)*. Trends Microbiol, 2012. 20(1): p. 11-20.
6. Webster, R.G. and E.A. Govorkova, *Continuing challenges in influenza*. Ann N Y Acad Sci, 2014. 1323(1): p. 115-39.
7. Causey, D. and S.V. Edwards, *Ecology of avian influenza virus in birds*. J Infect Dis, 2008. 197 Suppl 1(Suppl 1): p. S29-33.
8. Lee, D.H., M.F. Criado, and D.E. Swayne, *Pathobiological origins and evolutionary history of highly pathogenic avian influenza viruses*. Cold Spring Harb Perspect Med, 2021. 11(2).
9. Shi, J., et al., *Alarming situation of emerging H5 and H7 avian influenza and effective control strategies*. Emerg Microbes Infect, 2023. 12(1): p. 2155072.
10. Floyd, T., et al., *Encephalitis and death in wild mammals at a rehabilitation center after infection with highly pathogenic avian influenza A(H5N8) virus, United Kingdom*. Emerg Infect Dis, 2021. 27(11): p. 2856-2863.
11. Hiono, T., et al., *Virological, pathological, and glycovirological investigations of an Ezo red fox and a tanuki naturally infected with H5N1 high pathogenicity avian influenza viruses in Hokkaido, Japan*. Virology, 2023. 578: p. 35-44.
12. Agüero, M., et al., *Highly pathogenic avian influenza A(H5N1) virus infection in farmed minks, Spain, October 2022*. Eurosurveillance, 2023. 28(3): p. 2300001.
13. Abdelwhab, E.M. and T.C. Mettenleiter, *Zoonotic animal influenza virus and potential mixing vessel hosts*. Viruses, 2023. 15(4).
14. European Food Safety Authority, E.C.f.D.P., et al., *Avian influenza overview April - June 2023*. EFSA J, 2023. 21(7): p. e08191.
15. Monto, A.S. and K. Fukuda, *Lessons from influenza pandemics of the last 100 years*. Clin Infect Dis, 2020. 70(5): p. 951-957.
16. Eccles, R., *Understanding the symptoms of the common cold and influenza*. Lancet Infect Dis, 2005. 5(11): p. 718-25.
17. Palese, P., *Influenza: old and new threats*. Nat Med, 2004. 10(12 Suppl): p. S82-7.
18. Wu, N.C. and I.A. Wilson, *Influenza hemagglutinin structures and antibody recognition*. Cold Spring Harb Perspect Med, 2020. 10(8).
19. Chambers, T.M., *Equine influenza*. Cold Spring Harb Perspect Med, 2022. 12(1).
20. Oladunni, F.S., et al., *Equine influenza virus and vaccines*. Viruses, 2021. 13(8).
21. Parrish, C.R., P.R. Murcia, and E.C. Holmes, *Influenza virus reservoirs and intermediate hosts: dogs, horses, and new possibilities for influenza virus exposure of humans*. J Virol, 2015. 89(6): p. 2990-4.
22. Wasik, B.R., I.E.H. Voorhees, and C.R. Parrish, *Canine and feline influenza*. Cold Spring Harb Perspect Med, 2021. 11(1).
23. Borland, S., et al., *Influenza A virus infection in cats and dogs: a literature review in the light of the "One Health" concept*. Front Public Health, 2020. 8: p. 83.
24. Klivleyeva, N.G., et al., *Influenza A viruses circulating in dogs: A review of the scientific literature*. Open Vet J, 2022. 12(5): p. 676-687.
25. Reperant, L.A., G.F. Rimmelzwaan, and T. Kuiken, *Avian influenza viruses in mammals*. Rev Sci Tech, 2009. 28(1): p. 137-59.
26. Ma, W., *Swine influenza virus: current status and challenge*. Virus Res, 2020. 288: p. 198118.
27. Runstadler, J.A. and W. Puryear, *A brief introduction to influenza A virus in marine mammals*. Methods Mol Biol, 2020. 2123: p. 429-450.
28. Fereidouni, S., et al., *Influenza virus infection of marine mammals*. Ecohealth, 2016. 13(1): p. 161-70.
29. Shin, D.L., et al., *Highly pathogenic avian influenza A(H5N8) virus in gray seals, Baltic Sea*. Emerg Infect Dis, 2019. 25(12): p. 2295-2298.
30. Leguía, M., et al., *Highly pathogenic avian influenza A (H5N1) in marine mammals and seabirds in Peru*. Nat Commun, 2023. 14(1): p. 5489.
31. Puryear, W., et al., *Highly pathogenic avian influenza A(H5N1) virus outbreak in New England seals, United States*. Emerg Infect Dis, 2023. 29(4): p. 786-791.
32. Zhang, M., et al., *Influenza A virus-host specificity: an ongoing cross-talk between viral and host factors*. Front Microbiol, 2021. 12: p. 777885.
33. Ping, J., et al., *PB2 and hemagglutinin mutations are major determinants of host range and virulence in mouse-adapted influenza A virus*. J Virol, 2010. 84(20): p. 10606-18.
34. Neumann, G. and Y. Kawaoka, *Host range restriction and pathogenicity in the context of influenza pandemic*. Emerg Infect Dis, 2006. 12(6): p. 881-6.
35. Bisset, A.T. and G.F. Hoyne, *Evolution and adaptation of the avian H7N9 virus into the human host*. Microorganisms, 2020. 8(5).



36. Mair, C.M., et al., Receptor binding and pH stability - how influenza A virus hemagglutinin affects host-specific virus infection. *Biochim Biophys Acta*, 2014. 1838(4): p. 1153-68.
37. Skehel, J.J. and D.C. Wiley, Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. *Annu Rev Biochem*, 2000. 69: p. 531-69.
38. Yang, H., et al., Structure and receptor binding preferences of recombinant human A(H3N2) virus hemagglutinins. *Virology*, 2015. 477: p. 18-31.
39. Asaoka, N., et al., Low growth ability of recent influenza clinical isolates in MDCK cells is due to their low receptor binding affinities. *Microbes Infect*, 2006. 8(2): p. 511-9.
40. Liu, M., et al., H3N2 influenza A virus gradually adapts to human-type receptor binding and entry specificity after the start of the 1968 pandemic. *Proc Natl Acad Sci U S A*, 2023. 120(31): p. e2304992120.
41. Abe, Y., et al., Effect of the addition of oligosaccharides on the biological activities and antigenicity of influenza A/H3N2 virus hemagglutinin. *J Virol*, 2004. 78(18): p. 9605-11.
42. Ge, S. and Z. Wang, An overview of influenza A virus receptors. *Crit Rev Microbiol*, 2011. 37(2): p. 157-65.
43. Angata, T. and A. Varki, Chemical diversity in the sialic acids and related alpha-keto acids: an evolutionary perspective. *Chem Rev*, 2002. 102(2): p. 439-69.
44. Barnard, K.N., et al., Expression of 9-O- and 7,9-O-acetyl modified sialic acid in cells and their effects on influenza viruses. *mBio*, 2019. 10(6): p. e02490-19.
45. Nicholls, J.M., et al., Evolving complexities of influenza virus and its receptors. *Trends Microbiol*, 2008. 16(4): p. 149-57.
46. Long, J.S., et al., Host and viral determinants of influenza A virus species specificity. *Nat Rev Microbiol*, 2019. 17(2): p. 67-81.
47. Bradley, K.C., et al., Analysis of influenza virus hemagglutinin receptor binding mutants with limited receptor recognition properties and conditional replication characteristics. *J Virol*, 2011. 85(23): p. 12387-98.
48. Spruit, C.M., et al., N-glycolylneuraminic acid in animal models for human influenza A virus. *Viruses*, 2021. 13: 815(5).
49. Kumlin, U., et al., Sialic acid tissue distribution and influenza virus tropism. *Influenza Other Respir Viruses*, 2008. 2(5): p. 147-54.
50. Byrd-Leotis, L., et al., Antigenic pressure on H3N2 influenza virus drift strains imposes constraints on binding to sialylated receptors but not phosphorylated glycans. *J Virol*, 2019. 93(22).
51. Gambaryan, A., et al., Receptor specificity of influenza viruses from birds and mammals: new data on involvement of the inner fragments of the carbohydrate chain. *Virology*, 2005. 334(2): p. 276-83.
52. Unione, L., et al., Probing altered receptor specificities of antigenically drifting human H3N2 viruses by chemoenzymatic synthesis, NMR and modeling. *BiorXiv*, 2023.
53. Miller-Podraza, H., et al., A strain of human influenza A virus binds to extended but not short gangliosides as assayed by thin-layer chromatography overlay. *Glycobiology*, 2000. 10(10): p. 975-82.
54. Imai, M. and Y. Kawaoka, The role of receptor binding specificity in interspecies transmission of influenza viruses. *Curr Opin Virol*, 2012. 2(2): p. 160-7.
55. Connor, R.J., et al., Receptor specificity in human, avian, and equine H2 and H3 influenza virus isolates. *Virology*, 1994. 205(1): p. 17-23.
56. Gulati, S., et al., Human H3N2 influenza viruses isolated from 1968 to 2012 show varying preference for receptor substructures with no apparent consequences for disease or spread. *PLoS One*, 2013. 8(6): p. e66325.
57. Chandrasekaran, A., et al., Glycan topology determines human adaptation of avian H5N1 virus hemagglutinin. *Nat Biotechnol*, 2008. 26(1): p. 107-13.
58. Srivilajiroen, N., et al., N-glycan structures of human alveoli provide insight into influenza A virus infection and pathogenesis. *FEBS J*, 2018. 285(9): p. 1611-1634.
59. Peng, W., et al., Recent H3N2 viruses have evolved specificity for extended, branched human-type receptors, conferring potential for increased avidity. *Cell Host Microbe*, 2017. 21(1): p. 23-34.
60. Stevens, J., et al., Glycan microarray analysis of the hemagglutinins from modern and pandemic influenza viruses reveals different receptor specificities. *J Mol Biol*, 2006. 355(5): p. 1143-55.
61. Stevens, J., et al., Receptor specificity of influenza A H3N2 viruses isolated in mammalian cells and embryonated chicken eggs. *J Virol*, 2010. 84(16): p. 8287-99.
62. Nycholai, C.M., et al., Recognition of sialylated poly-N-acetyllactosamine chains on N- and O-linked glycans by human and avian influenza A virus hemagglutinins. *Angew Chem Int Ed Engl*, 2012. 51(20): p. 4860-3.
63. Ji, Y., et al., New insights into influenza A specificity: an evolution of paradigms. *Curr Opin Struct Biol*, 2017. 44: p. 219-231.
64. de Vries, R.P., et al., Hemagglutinin receptor specificity and structural analyses of respiratory droplet-transmissible H5N1 viruses. *J Virol*, 2014. 88(1): p. 768-73.
65. de Vries, R.P., et al., Only two residues are responsible for the dramatic difference in receptor binding between swine and new pandemic H1 hemagglutinin. *J Biol Chem*, 2011. 286(7): p. 5868-75.
66. de Vries, R.P., et al., A single mutation in Taiwanese H6N1 influenza hemagglutinin switches binding to human-type receptors. *EMBO Mol Med*, 2017. 9(9): p. 1314-1325.
67. de Graaf, M. and R.A. Fouchier, Role of receptor binding specificity in influenza A virus transmission and pathogenesis. *EMBO J*, 2014. 33(8): p. 823-41.
68. Belser, J.A., et al., Contemporary North American influenza H7 viruses possess human receptor specificity: Implications for virus transmissibility. *Proc Natl Acad Sci U S A*, 2008. 105(21): p. 7558-63.
69. Belser, J.A. and T.M. Tumpey, Mammalian models for the study of H7 virus pathogenesis and transmission. *Curr Top Microbiol Immunol*, 2014. 385: p. 275-305.
70. Eggink, D., et al., Phenotypic effects of substitutions within the receptor binding site of highly pathogenic avian influenza H5N1 virus observed during human infection. *J Virol*, 2020. 94(13).

71. Huang, P., et al., *Potential cross-species transmission of highly pathogenic avian influenza H5 subtype (HPAI H5) viruses to humans calls for the development of H5-specific and universal influenza vaccines*. *Cell Discov*, 2023. 9(1): p. 58.
72. Li, C. and H. Chen, *H7N9 Influenza Virus in China*. *Cold Spring Harb Perspect Med*, 2021. 11(8).
73. Jernigan, D.B. and N.J. Cox, *H7N9: preparing for the unexpected in influenza*. *Annu Rev Med*, 2015. 66: p. 361-71.
74. de Vries, R.P., et al., *Three mutations switch H7N9 influenza to human-type receptor specificity*. *PLoS Pathog*, 2017. 13(6): p. e1006390.
75. Schnaar, R.L., et al., *Glycosphingolipids*, in *Essentials of Glycobiology*, A. Varki, et al., Editors. 2022: Cold Spring Harbor (NY). p. 129-40.
76. Schnaar, R.L., *Glycobiology simplified: diverse roles of glycan recognition in inflammation*. *J Leukoc Biol*, 2016. 99(6): p. 825-38.
77. Jia, N., et al., *The human lung glycome reveals novel glycan ligands for influenza A virus*. *Sci Rep*, 2020. 10(1): p. 5320.
78. Jia, N., et al., *Glycomic characterization of respiratory tract tissues of ferrets: implications for its use in influenza virus infection studies*. *J Biol Chem*, 2014. 289(41): p. 28489-504.
79. Yu, R.K., et al., *Structures, biosynthesis, and functions of gangliosides—an overview*. *J Oleo Sci*, 2011. 60(10): p. 537-44.
80. Ledeen, R.W. and R.K. Yu, *Gangliosides: structure, isolation, and analysis*. *Methods Enzymol*, 1982. 83: p. 139-91.
81. Suzuki, Y., *Gangliosides as influenza virus receptors. Variation of influenza viruses and their recognition of the receptor sialo-sugar chains*. *Prog Lipid Res*, 1994. 33(4): p. 429-57.
82. Bergelson, L.D., et al., *Role of gangliosides in reception of influenza virus*. *Eur J Biochem*, 1982. 128(2-3): p. 467-74.
83. Gambaryan, A.S., J.S. Robertson, and M.N. Matrosovich, *Effects of egg-adaptation on the receptor-binding properties of human influenza A and B viruses*. *Virology*, 1999. 258(2): p. 232-9.
84. Matrosovich, M.N., et al., *Avian influenza A viruses differ from human viruses by recognition of sialyloligosaccharides and gangliosides and by a higher conservation of the HA receptor-binding site*. *Virology*, 1997. 233(1): p. 224-34.
85. Vrijens, P., et al., *Influenza virus entry via the GM3 ganglioside-mediated platelet-derived growth factor receptor beta signalling pathway*. *J Gen Virol*, 2019. 100(4): p. 583-601.
86. Meisen, I., et al., *The human H3N2 influenza viruses A/Victoria/3/75 and A/Hiroshima/52/2005 preferentially bind to alpha2-3-sialylated monosialogangliosides with fucosylated poly-N-acetylactosaminyl chains*. *Glycobiology*, 2012. 22(8): p. 1055-76.
87. Matrosovich, M., et al., *Gangliosides are not essential for influenza virus infection*. *Glycoconj J*, 2006. 23(1-2): p. 107-13.
88. Liang, C.Y., et al., *Avian influenza A viruses exhibit plasticity in sialylglycoconjugate receptor usage in human lung cells*. *J Virol*, 2023: p. e0090623.
89. Alley, W.R., Jr., B.F. Mann, and M.V. Novotny, *High-sensitivity analytical approaches for the structural characterization of glycoproteins*. *Chem Rev*, 2013. 113(4): p. 2668-732.
90. Bateman, A.C., et al., *Glycan analysis and influenza A virus infection of primary swine respiratory epithelial cells: the importance of NeuAc{alpha}2-6 glycans*. *J Biol Chem*, 2010. 285(44): p. 34016-26.
91. Walther, T., et al., *Glycomic analysis of human respiratory tract tissues and correlation with influenza virus infection*. *PLoS Pathog*, 2013. 9(3): p. e1003223.
92. Chan, R.W., et al., *Infection of swine ex vivo tissues with avian viruses including H7N9 and correlation with glycomic analysis*. *Influenza Other Respir Viruses*, 2013. 7(6): p. 1269-82.
93. Maines, T.R., et al., *Effect of receptor binding domain mutations on receptor binding and transmissibility of avian influenza H5N1 viruses*. *Virology*, 2011. 413(1): p. 139-47.
94. Stevens, J., et al., *Recent avian H5N1 viruses exhibit increased propensity for acquiring human receptor specificity*. *J Mol Biol*, 2008. 381(5): p. 1382-94.
95. Mayr, J., et al., *Unravelling the role of O-glycans in influenza A virus infection*. *Scientific Reports*, 2018. 8(1): p. 16382.
96. Nakamura, S., et al., *Influenza A virus-induced expression of a GalNAc transferase, GALNT3, via microRNAs is required for enhanced viral replication*. *J Virol*, 2016. 90(4): p. 1788-801.
97. Jayaraman, A., et al., *Decoding the distribution of glycan receptors for human-adapted influenza A viruses in ferret respiratory tract*. *PLoS One*, 2012. 7(2): p. e27517.
98. Suzuki, N., T. Abe, and S. Natsuka, *Structural analysis of N-glycans in chicken trachea and lung reveals potential receptors of chicken influenza viruses*. *Sci Rep*, 2022. 12(1): p. 2081.
99. Byrd-Leotis, L., et al., *Shotgun glycomics of pig lung identifies natural endogenous receptors for influenza viruses*. *Proc Natl Acad Sci U S A*, 2014. 111(22): p. E2241-50.
100. Ruhaak, L.R., et al., *Differential N-glycosylation patterns in lung adenocarcinoma tissue*. *J Proteome Res*, 2015. 14(11): p. 4538-49.
101. Chen, L.M., et al., *Receptor specificity of subtype H1 influenza A viruses isolated from swine and humans in the United States*. *Virology*, 2011. 412(2): p. 401-10.
102. Chu, V.C. and G.R. Whittaker, *Influenza virus entry and infection require host cell N-linked glycoprotein*. *Proc Natl Acad Sci U S A*, 2004. 101(52): p. 18153-8.
103. Broszeit, F., et al., *Glycan remodeled erythrocytes facilitate antigenic characterization of recent A/H3N2 influenza viruses*. *Nat Commun*, 2021. 12(1): p. 5449.
104. Broszeit, F., et al., *N-glycolylneuraminic acid as a receptor for influenza A viruses*. *Cell Rep*, 2019. 27(11): p. 3284-3294 e6.

105. Canales, A., et al., *Revealing the specificity of human H1 influenza A viruses to complex N-glycans*. JACS Au, 2023.
106. Allen, J.D. and T.M. Ross, *H3N2 influenza viruses in humans: viral mechanisms, evolution, and evaluation*. Hum Vaccin Immunother, 2018. 14(8): p. 1840-1847.
107. Lin, Y.P., et al., *Evolution of the receptor binding properties of the influenza A(H3N2) hemagglutinin*. Proc Natl Acad Sci U S A, 2012. 109(52): p. 21474-9.
108. Owen, R.E., et al., *Alterations in receptor binding properties of recent human influenza H3N2 viruses are associated with reduced natural killer cell lysis of infected cells*. J Virol, 2007. 81(20): p. 11170-8.
109. Chambers, B.S., et al., *Recent H3N2 influenza virus clinical isolates rapidly acquire hemagglutinin or neuraminidase mutations when propagated for antigenic analyses*. J Virol, 2014. 88(18): p. 10986-9.
110. Takada, K., et al., *A humanized MDCK cell line for the efficient isolation and propagation of human influenza viruses*. Nat Microbiol, 2019. 4(8): p. 1268-1273.
111. Oh, D.Y., et al., *MDCK-SIAT1 cells show improved isolation rates for recent human influenza viruses compared to conventional MDCK cells*. J Clin Microbiol, 2008. 46(7): p. 2189-94.
112. Byrd-Leotis, L., et al., *Sialylated and sulfated N-glycans in MDCK and engineered MDCK cells for influenza virus studies*. Sci Rep, 2022. 12(1): p. 12757.
113. Spruit, C.M., et al., *Contemporary human H3N2 influenza a viruses require a low threshold of suitable glycan receptors for efficient infection*. Glycobiology, 2023.
114. Wasik, B.R., K.N. Barnard, and C.R. Parrish, *Effects of sialic acid modifications on virus binding and infection*. Trends Microbiol, 2016. 24(12): p. 991-1001.
115. Aamelfot, M., et al., *The in situ distribution of glycoprotein-bound 4-O-Acetylated sialic acids in vertebrates*. Glycoconj J, 2014. 31(4): p. 327-35.
116. Rinninger, A., et al., *Localisation and distribution of O-acetylated N-acetylneuraminic acids, the endogenous substrates of the hemagglutinin-esterases of murine coronaviruses, in mouse tissue*. Glycoconj J, 2006. 23(1-2): p. 73-84.
117. Song, X., et al., *A sialylated glycan microarray reveals novel interactions of modified sialic acids with proteins and viruses*. J Biol Chem, 2011. 286(36): p. 31610-22.
118. Higa, H.H., G.N. Rogers, and J.C. Paulson, *Influenza virus hemagglutinins differentiate between receptor determinants bearing N-acetyl-, N-glycolyl-, and N,O-diacetylneuraminic acids*. Virology, 1985. 144(1): p. 279-82.
119. Sauter, N.K., et al., *Hemagglutinins from two influenza virus variants bind to sialic acid derivatives with millimolar dissociation constants: a 500-MHz proton nuclear magnetic resonance study*. Biochemistry, 1989. 28(21): p. 8388-96.
120. Pritchett, T.J. and J.C. Paulson, *Basis for the potent inhibition of influenza virus infection by equine and guinea pig alpha 2-macroglobulin*. J Biol Chem, 1989. 264(17): p. 9850-8.
121. Hanaoka, K., et al., *4-O-acetyl-N-acetylneuraminic acid in the N-linked carbohydrate structures of equine and guinea pig alpha 2-macroglobulins, potent inhibitors of influenza virus infection*. J Biol Chem, 1989. 264(17): p. 9842-9.
122. Gambaryan, A.S., et al., *Receptor-binding profiles of H7 subtype influenza viruses in different host species*. J Virol, 2012. 86(8): p. 4370-9.
123. Anders, E.M., et al., *Relationship between mitogenic activity of influenza viruses and the receptor-binding specificity of their hemagglutinin molecules*. J Virol, 1986. 60(2): p. 476-82.
124. Suzuki, T., et al., *Swine influenza virus strains recognize sialylsugar chains containing the molecular species of sialic acid predominantly present in the swine tracheal epithelium*. FEBS Lett, 1997. 404(2-3): p. 192-6.
125. Masuda, H., et al., *Substitution of amino acid residue in influenza A virus hemagglutinin affects recognition of sialyl-oligosaccharides containing N-glycolylneuraminic acid*. FEBS Lett, 1999. 464(1-2): p. 71-4.
126. Suzuki, Y., et al., *Sialic acid species as a determinant of the host range of influenza A viruses*. J Virol, 2000. 74(24): p. 11825-31.
127. Ito, T., et al., *Recognition of N-glycolylneuraminic acid linked to galactose by the alpha2,3 linkage is associated with intestinal replication of influenza A virus in ducks*. J Virol, 2000. 74(19): p. 9300-5.
128. Takahashi, T., et al., *Identification of amino acid residues of influenza A virus H3 HA contributing to the recognition of molecular species of sialic acid*. FEBS Lett, 2009. 583(19): p. 3171-4.
129. Nemanichvili, N., et al., *Wild and domestic animals variably display Neu5Ac and Neu5Gc sialic acids*. Glycobiology, 2022. 32(9): p. 791-802.
130. Peri, S., et al., *Phylogenetic distribution of CMP-Neu5Ac hydroxylase (CMAH), the enzyme synthesizing the proinflammatory human xenoantigen Neu5Gc*. Genome Biol Evol, 2018. 10(1): p. 207-219.
131. Yasue, S., et al., *Difference in form of sialic acid in red blood cell glycolipids of different breeds of dogs*. J Biochem, 1978. 83(4): p. 1101-7.
132. Schauer, R., et al., *Low incidence of N-glycolylneuraminic acid in birds and reptiles and its absence in the platypus*. Carbohydr Res, 2009. 344(12): p. 1494-500.
133. Ng, P.S., et al., *Ferrets exclusively synthesize Neu5Ac and express naturally humanized influenza A virus receptors*. Nat Commun, 2014. 5: p. 5750.
134. Kobayashi, D., et al., *Turkeys possess diverse Siaalpha2-3Gal glycans that facilitate their dual susceptibility to avian influenza viruses isolated from ducks and chickens*. Virus Res, 2022. 315: p. 198771.
135. Kikutani, Y., et al., *E190V substitution of H6 hemagglutinin is one of key factors for binding to sulfated sialylated glycan receptor and infection to chickens*. Microbiol Immunol, 2020. 64(4): p. 304-312.
136. Peacock, T.P., et al., *Genetic determinants of receptor-binding preference and zoonotic potential of H9N2 avian influenza viruses*. J Virol, 2021. 95(5).
137. Gambaryan, A.S., et al., *6-sulfo sialyl Lewis X is the common receptor determinant recognized by H5, H6, H7 and H9 influenza viruses of terrestrial poultry*. Virol J, 2008. 5: p. 85.

138. Peacock, T.P., et al., *Variability in H9N2 haemagglutinin receptor-binding preference and the pH of fusion*. *Emerg Microbes Infect*, 2017. 6(3): p. e11.
139. Verhagen, J.H., et al., *Host range of influenza A virus H1 to H16 in Eurasian ducks based on tissue and receptor binding studies*. *J Virol*, 2021. 95(6).
140. Xu, R., et al., *Preferential recognition of avian-like receptors in human influenza A H7N9 viruses*. *Science*, 2013. 342(6163): p. 1230-5.
141. Yang, H., et al., *Structural and molecular characterization of the hemagglutinin from the fifth-epidemic-wave A(H7N9) influenza viruses*. *J Virol*, 2018. 92(16).
142. Guo, H., et al., *Highly pathogenic influenza A(H5Nx) viruses with altered H5 receptor-binding specificity*. *Emerg Infect Dis*, 2017. 23(2): p. 220-231.
143. West, J., et al., *Characterization of changes in the hemagglutinin that accompanied the emergence of H3N2/1968 pandemic influenza viruses*. *PLoS Pathog*, 2021. 17(9): p. e1009566.
144. Collins, P.J., et al., *Recent evolution of equine influenza and the origin of canine influenza*. *Proc Natl Acad Sci U S A*, 2014. 111(30): p. 11175-80.
145. Wu, Y., et al., *Exploiting substrate specificities of 6-O-sulfotransferases to enzymatically synthesize keratan sulfate oligosaccharides*. *JACS Au*, 2023.
146. Boravleva, E., et al., *Molecular characteristics, receptor specificity, and pathogenicity of avian influenza viruses isolated from wild ducks in Russia*. *Int J Mol Sci*, 2022. 23(18).
147. Gambaryan, A.S., et al., *H5N1 chicken influenza viruses display a high binding affinity for Neu5Acalpha2-3Galbeta1-4(6-HSO3)GlcNAc-containing receptors*. *Virology*, 2004. 326(2): p. 310-6.
148. Ichimiya, T., et al., *Sulfated glycans containing NeuAcalpha2-3Gal facilitate the propagation of human H1N1 influenza A viruses in eggs*. *Virology*, 2021. 562: p. 29-39.
149. Hussein, I.T., et al., *New England harbor seal H3N8 influenza virus retains avian-like receptor specificity*. *Sci Rep*, 2016. 6: p. 21428.
150. Sauer, A.K., et al., *Characterization of the sialic acid binding activity of influenza A viruses using soluble variants of the H7 and H9 hemagglutinins*. *PLoS One*, 2014. 9(2): p. e89529.
151. Gambaryan, A.S., et al., *Receptor-binding properties of influenza viruses isolated from gulls*. *Virology*, 2018. 522: p. 37-45.
152. Guan, M., et al., *The sialyl Lewis X glycan receptor facilitates infection of subtype H7 avian influenza A viruses*. *J Virol*, 2022. 96(19): p. e0134422.
153. Wen, F., et al., *Mutation W222L at the receptor binding site of hemagglutinin could facilitate viral adaption from equine influenza A(H3N8) virus to dogs*. *J Virol*, 2018. 92(18).
154. Hiono, T., et al., *A chicken influenza virus recognizes fucosylated alpha2,3 sialoglycan receptors on the epithelial cells lining upper respiratory tracts of chickens*. *Virology*, 2014. 456-457: p. 131-8.
155. Allahverdian, S., K.R. Wojcik, and D.R. Dorscheid, *Airway epithelial wound repair: role of carbohydrate sialyl Lewisx*. *Am J Physiol Lung Cell Mol Physiol*, 2006. 291(4): p. L828-36.
156. Gizaw, S.T., et al., *Glycoblotting method allows for rapid and efficient glycome profiling of human Alzheimer's disease brain, serum and cerebrospinal fluid towards potential biomarker discovery*. *Biochim Biophys Acta*, 2016. 1860(8): p. 1716-27.
157. Li, Q., et al., *Comprehensive N-glycome profiling of cells and tissues for breast cancer diagnosis*. *J Proteome Res*, 2019. 18(6): p. 2559-2570.
158. Byrd-Leotis, L., et al., *Influenza binds phosphorylated glycans from human lung*. *Sci Adv*, 2019. 5(2): p. eaav2554.
159. Tan, M., et al., *Saliva as a source of reagent to study human susceptibility to avian influenza H7N9 virus infection*. *Emerg Microbes Infect*, 2018. 7(1): p. 156.
160. Hiono, T., et al., *Amino acid residues at positions 222 and 227 of the hemagglutinin together with the neuraminidase determine binding of H5 avian influenza viruses to sialyl Lewis X*. *Arch Virol*, 2016. 161(2): p. 307-16.
161. Gaide, N., et al., *Pathobiology of highly pathogenic H5 avian influenza viruses in naturally infected Galliformes and Anseriformes in France during winter 2015-2016*. *Vet Res*, 2022. 53(1): p. 11.
162. Heider, A., et al., *Alterations in hemagglutinin receptor-binding specificity accompany the emergence of highly pathogenic avian influenza viruses*. *J Virol*, 2015. 89(10): p. 5395-405.
163. Chen, C., et al., *Enzymatic modular synthesis and microarray assay of poly-N-acetylactosamine derivatives*. *Chem Commun (Camb)*, 2020. 56(55): p. 7549-7552.

# Chapter 2

## Contemporary human H3N2 influenza A viruses require a low threshold of suitable glycan receptors for efficient infection

*Cindy M. Spruit, Igor R. Sweet, Joshua C. L. Maliepaard, Theo Bestebroer, Pascal Lexmond, Boning Qiu, Mirjam J.A. Damen, Ron A. M. Fouchier, Karli R. Reiding, Joost Snijder, Sander Herfst, Geert-Jan Boons, Robert P. de Vries*

*Published: Glycobiology, October 2023, DOI: <https://doi.org/10.1093/glycob/cwad060>*

### Abstract

Recent human H3N2 influenza A viruses (IAV) have evolved to employ elongated glycans terminating in  $\alpha$ 2,6-linked sialic acid as their receptors. These glycans are displayed in low abundancies by (humanized) Madin-Darby Canine Kidney cells (MDCK and hCK) which are commonly employed to propagate IAV, resulting in low or no viral propagation. Here, we examined whether the overexpression of the glycosyltransferases B3GNT2 and B4GALT1, which are responsible for the elongation of poly-*N*-acetyllactosamines (LacNAc), would result in improved A/H3N2 propagation. Stable overexpression of B3GNT2 and B4GALT1 in MDCK and hCK cells was achieved by lentiviral integration and subsequent antibiotic selection and confirmed by qPCR and protein mass spectrometry experiments. Flow cytometry and glycan mass spectrometry experiments using the B3GNT2 and/or B4GALT1 knock-in cells demonstrated increased binding of viral hemagglutinins and the presence of a larger number of LacNAc repeating units, especially on hCK-B3GNT2 cells. An increase in the number of glycan receptors did, however, not result in a greater infection efficiency of recent human H3N2 viruses. Based on these results, we propose that H3N2 IAVs require a low number of suitable glycan receptors to infect cells and that an increase in the glycan receptor display above this threshold does not result in improved infection efficiency.

## Introduction

Influenza A viruses (IAV) of the H3N2 subtype cause seasonal epidemics, leading to illness, hospitalizations, and deaths in humans [1]. The crucial first step of infection is the binding of the viral hemagglutinin (HA) to a receptor on a cell, which are glycans terminating in  $\alpha$ 2,6-linked sialic acids (Sia) for human IAVs [2, 3]. Due to continuous immune evasion, antigenic drift of the surface proteins of IAVs takes place. This antigenic drift of IAVs has changed receptor specificities [4] and pandemic H1N1 viruses (starting from 2009) were shown to prefer binding to glycans with multiple consecutive LacNAc repeating units [5]. Furthermore, recent H3N2 viruses only bind to longer glycans having multiple consecutive oligo-*N*-acetyllactosamine (LacNAc) moieties terminating in an  $\alpha$ 2,6-linked Sia [2, 4, 6-12]. This specificity is most pronounced for H3N2 viruses of subclade 3C.2a, which require at least three subsequent LacNAc repeating units for binding [13].

These altered receptor specificities make it difficult to isolate and propagate H3N2 viruses, greatly hampering the further study of these viruses [9, 14-17]. Even when virus isolation is successful, viruses may have acquired adaptive mutations in the receptor binding site of HA, especially when isolated in eggs instead of MDCK (Madin-Darby Canine Kidney) cells, which are often used for the propagation of IAV [14, 16, 18-21]. MDCK cells have previously been modified to produce more  $\alpha$ 2,6-linked Sias by the overexpression of the enzyme ST6GAL1, resulting in MDCK-SIAT1 [22] and MDCK-AX4 [23] cells. These cells enabled the isolation of H3N2 viruses, especially of the 3C.2a and 3C.3a subclades, and resulted in higher titers of viral stocks [17, 24]. To allow the isolation of further evolved contemporary H3N2 viruses, with higher titers and fewer mutations, MDCK cells were further modified to eliminate  $\alpha$ 2,3-linked Sias while also overexpressing  $\alpha$ 2,6-linked Sias, resulting in “humanized” MDCK (hCK) cells [16].

Analysis of the *N*-glycans of MDCK, MDCK-SIAT1, and hCK cells indicated a low abundance of glycans with at least three successive LacNAc repeating units terminating in an  $\alpha$ 2,6-linked Sia (Fig. S1) [25]. The enzyme  $\beta$ 1,3-*N*-acetylglucosaminyltransferase (B3GNT2) is responsible for the addition of *N*-acetylglucosamine to glycans, while the galactose is transferred to the glycan by the enzyme  $\beta$ -1,4-galactosyltransferase 1 (B4GALT1). Previously, we successfully used these two enzymes to elongate LacNAc repeating units both in chemoenzymatic synthesis [26, 27] and on biological membrane surfaces of erythrocytes [13].

We previously showed that the supplementation of sialic acid-depleted MDCK cells with *N*-glycans bearing three consecutive LacNAc repeating units is essential for infection of contemporary H3N2 viruses, as *N*-glycans with 1 or 2 LacNAc repeating units did not support infection [9]. However, such an approach is not viable for many laboratories as the methods and means for glycolipid-engineering of cells are not widely available. Here, we genetically installed *N*-glycans having elongated LacNAc moieties on MDCK and hCK cells by the overexpression of B3GNT2 and B4GALT1, to obtain a cell line that can be used by many laboratories to study viral propagation. Surprisingly, although the B3GNT2/B4GALT1 knock-in

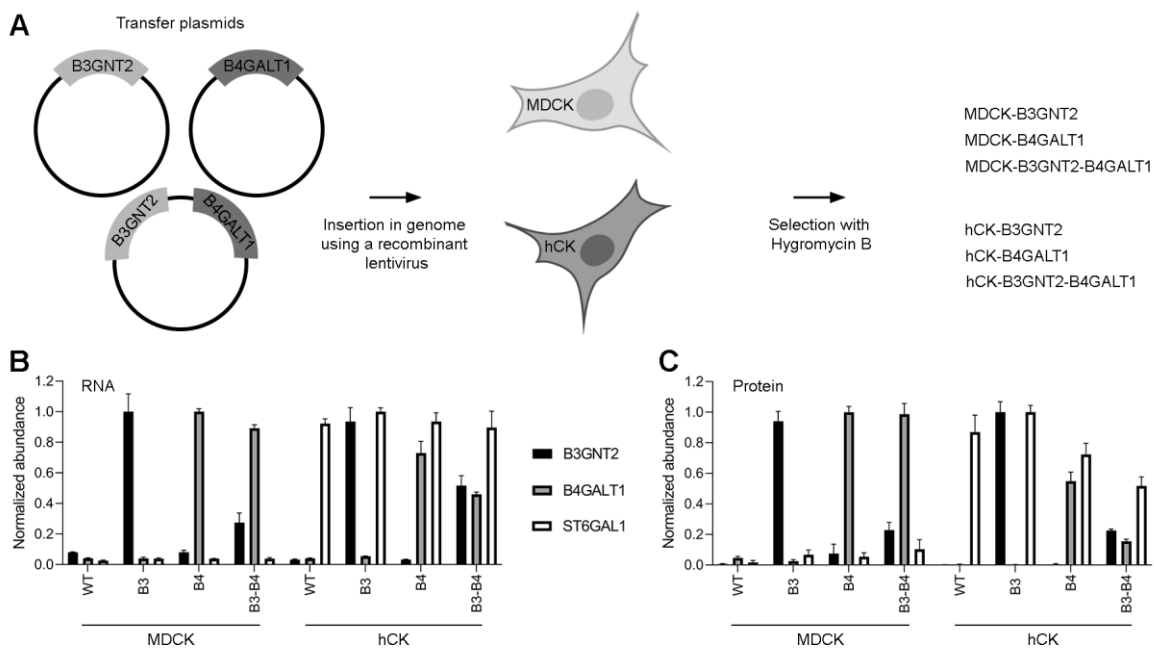


cells exhibited elevated binding of recent H3 HAs, the overexpression did not lead to improved virus isolation and infection efficiency. Several studies have indicated that a higher display of appropriate receptors leads to increased infectivity [16, 22, 23, 28], while others indicated that only low amounts of receptors are required for infection [29-31]. Based on our studies, we concluded that above a required threshold, a greater number of suitable glycan receptors for H3N2 IAVs does not result in increased infection efficiency.

## Results

### Generation of stable B3GNT2 and B4GALT1 knock-in MDCK and hCK cell lines

Recently, we and others have shown that poly-LacNAc containing *N*-glycans are critical for the binding of contemporary H3N2 viruses and pandemic H1N1 viruses [5, 9, 13]. Therefore, we used the glycosyltransferases B3GNT2 and B4GALT1 to increase the biosynthesis of LacNAc repeating units to produce extended *N*-glycans [13, 26, 27, 32]. We hypothesized that the overexpression of B3GNT2 and/or B4GALT1 in MDCK and hCK cells would produce appropriate glycan receptors for recent H3N2 (subclade 3C.2a) IAVs.



**Fig 1. Construction of MDCK and hCK cells that overexpress B3GNT2 and/or B4GALT1.** (A) MDCK and hCK cells were modified with recombinant lentiviruses containing transfer plasmids for the insertion of the *B3GNT2* and/or *B4GALT1* gene and the Hygromycin B resistance gene. The knock-in cells were selected with 300  $\mu$ g/ml Hygromycin B. (B) RT-qPCR was performed with primers that anneal to both the human and dog *B3GNT2*, *B4GALT1*, or *ST6GAL1* genes. The values relative to the dog *GAPDH* gene were used, which were then normalized to the highest value of each gene. Mean and SD ( $n=3$ ) are shown. (C) Peptide mass spectrometry of the B3GNT2, B4GALT1, and ST6GAL1 proteins. Only peptides unique to human proteins were selected. All samples were normalized against tubulin- $\beta$  and then normalized to the highest value of each protein. Mean and SD ( $n=3$ ) are shown.

To accomplish the overexpression of these genes in MDCK and hCK cells, lentiviral transfer plasmids encoding the human *B3GNT2* and/or *B4GALT1* genes, together with the *Hygromycin B* resistance gene, were constructed. The genes were expressed from one human EF-1 $\alpha$  promoter [33] and separated by P2A (and for double glycosyltransferase knock-ins also T2A) self-cleaving peptides. Lentiviruses were produced with a transfer plasmid and packaging plasmids, after which the viruses were used to transduce MDCK and hCK cells (Fig. 1A). Cells in which the genes were inserted in the genome were selected with Hygromycin B.

Stable overexpression of *B3GNT2* and *B4GALT1* was confirmed by RT-qPCR analysis on isolated cellular RNA. Primers for the *B3GNT2*, *B4GALT1*, and *ST6GAL1* genes were used, and the obtained values were normalized to the reference gene *GAPDH* (Fig. 1B). Overexpression of the control gene *ST6GAL1* was clearly shown in hCK but not MDCK cells. Furthermore, the overexpression of *B3GNT2* and *B4GALT1* was present in all glycoengineered cell lines in which these knock-ins were made. It should be noted that expression levels in the double knock-in cell lines showed lower expression of the glycosyltransferases, especially for *B3GNT2* in MDCK-*B3GNT2*-*B4GALT1* cells.

Thereafter, the protein levels of *B3GNT2*, *B4GALT1*, and *ST6GAL1* in cell lysates were measured using proteomic experiments based on liquid chromatography coupled to tandem mass spectrometry, using label-free quantitation relative to tubulin- $\beta$  expression (Fig. 1C). Only peptides unique for the human *B3GNT2*, *B4GALT1*, and *ST6GAL1* were selected. The proteomic data is comparable to the RT-qPCR data since elevated protein levels in the cell lines with knock-ins were observed. Collectively, the data showed that the stable overexpression of *B3GNT2* and *B4GALT1* in MDCK and hCK was successful.

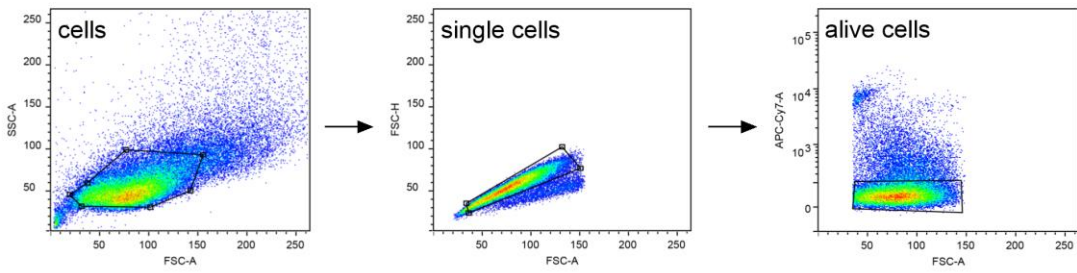
### **Flow cytometric characterization of *B3GNT2* and *B4GALT1* knock-in cells with plant and viral lectins**

Next attention was focused on whether the overexpression of *B3GNT2* and/or *B4GALT1* led to a display of a higher number of LacNAc repeating units on *N*-glycans. The glycans on the cell surface were first characterized using flow cytometry with standard lectins (Fig. 2B, specificities are summarized in Table SI). An alive, single-cell population was selected using a standard gating strategy, and mean fluorescence intensities were calculated over the cell population (Fig. 2A).

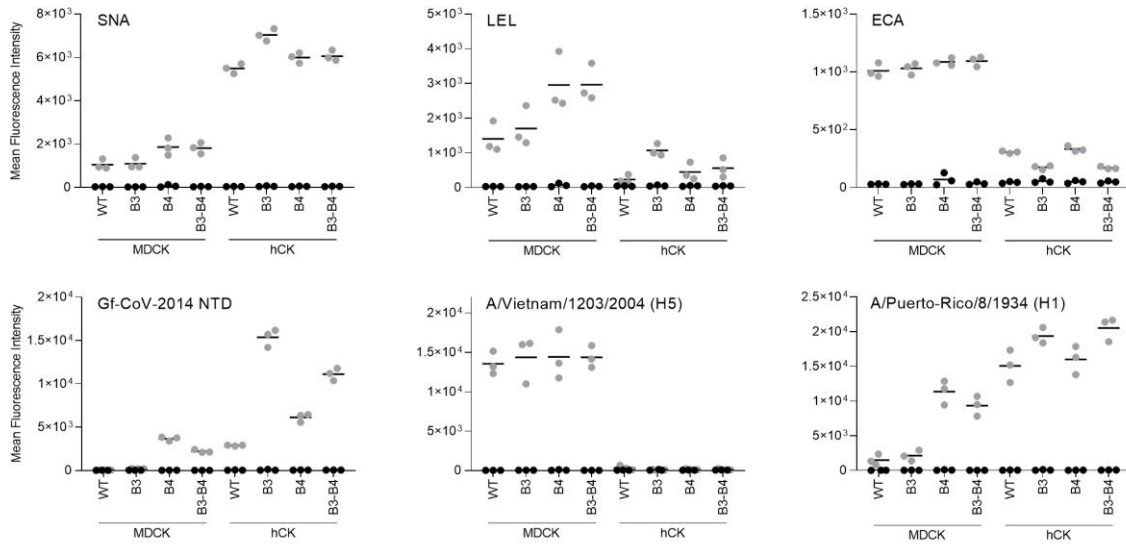
*Sambucus nigra* agglutinin (SNA) [34] was used to detect  $\alpha$ 2,6-linked Sias, which are present in higher quantities on hCK cells than MDCK cells due to the overexpression of *ST6GAL1* (Fig. 2B), which is consistent with earlier studies [25]. The *B3GNT2* and/or *B4GALT1* knock-ins caused minimal (less than 2-fold) differences in  $\alpha$ 2,6-linked Sia display, which is understandable since we did not interfere with the sialyltransferases. *Lycopersicon esculentum* lectin (LEL) recognizes elongated glycans [35] and we observed that glycans capped with  $\alpha$ 2,6-linked Sias need at least four consecutive LacNAc repeating units to be recognized, while glycans capped with  $\alpha$ 2,3-linked Sias are recognized when presented on two successive LacNAc repeating units (Fig. S2), which explains the lower signal for all hCK cells in



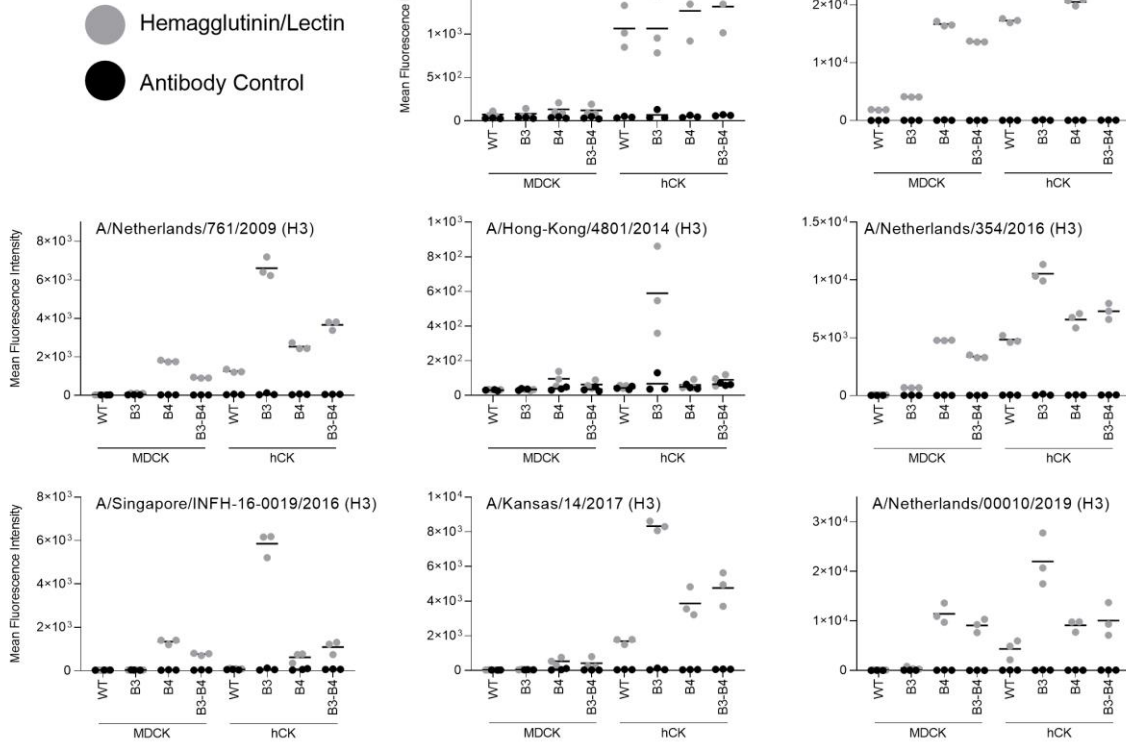
**A**



**B**



**C**



**Fig 2. Flow cytometric characterization of B3GNT2/B4GALT1 knock-in MDCK and hCK cells.** (A) The gating strategy that was used to select single, alive cells. (B) Flow cytometry measurements with lectins SNA (*Sambucus nigra* agglutinin, recognizes  $\alpha$ 2,6-Sia), LEL (*Lycopersicon esculentum* lectin, recognizes elongated glycans), and ECA (*Erythrina cristagalli* lectin, recognizes glycans without Sia) were performed. Furthermore, Gf-CoV-2014 NTD was used to detect elongated glycans, H5 HA of A/Vietnam/1203/2004 was used to detect  $\alpha$ 2,3-Sia, and H1 HA from A/Puerto-Rico/8/1934 was used as a standard influenza virus. Triplicate measurements were performed, of which the mean and all individual measurements are displayed. (C) A diverse set of H3 HAs was used to characterize the cells. Triplicate measurements were performed, of which the mean and all individual measurements are displayed. Titration curves of A/Hong-Kong/1/1968, A/Netherlands/109/2003, and A/Singapore/INFH-16-0019/2016 are shown in Fig. S4. Flow cytometric experiments with neuraminidase treatment of the cells are shown in Fig. S3.

general. An approximately 2-fold increase in the binding of LEL to MDCK-B4GALT1 and MDCK-B3GNT2-B4GALT1 compared to wild-type (WT) MDCK cells was observed, indicating that the LacNAc repeating units on MDCK cells are indeed elongated by the overexpression of mainly B4GALT1. Moreover, we observed an increase in LEL signal in the hCK-B3GNT2 cells compared to the hCK WT cells, indicating that the overexpression of B3GNT2 resulted in the formation of longer glycans on hCK cells. *Erythrina cristagalli* lectin (ECA) recognizes terminal galactose, and thus glycans lacking Sia capping [36]. The results using ECA indicated that all MDCK cells have a larger proportion of non-sialylated glycans compared to all hCK cells, which agrees with the overexpression of ST6GAL1 in all hCK cells. No major differences of more than 2-fold in the amount of non-sialylated glycans between WT and B3GNT2 and/or B4GALT1 knock-in cells were observed.

In addition to commonly employed plant lectins, viral proteins were used to examine the glycans displayed on the cells (Fig. 2B, specificities are summarized in Table S1). The N-terminal domain of  $\gamma$ CoV/AvCoV/guinea fowl/France/14032/2014 (Gf-CoV-2014 NTD) is known to bind elongated glycans [37]. MDCK-B4GALT1 and MDCK-B3GNT2-B4GALT1 cells showed an increased Gf-CoV-2014 NTD signal compared to MDCK WT cells. Furthermore, hCK WT cells appeared to have a higher number of LacNAc repeating units on glycans than MDCK WT cells. The hCK-B3GNT2 cells, and the hCK-B4GALT1 and hCK-B3GNT2-B4GALT1 cells to a lesser extent, showed a substantial increase (at least 2-fold) in Gf-CoV-2014 NTD binding compared to the hCK WT cells, indicating the presence of additional LacNAc repeating units on glycans. The HA of A/Vietnam/1203/2004 H5N1 (H5VN) is commonly used to probe the presence of terminal  $\alpha$ 2,3-linked Sias [36, 38, 39]. We observed a much lower amount of  $\alpha$ 2,3-linked Sias in all hCK cells compared to all MDCK cells, which is in agreement with the knock-outs of all  $\beta$ -galactoside  $\alpha$ -2,3 sialyltransferases that were made in the hCK cells previously [16] and findings in earlier studies [25]. The B3GNT2/B4GALT1 knock-ins did not alter the  $\alpha$ 2,3-linked Sia content. The HA of the human IAV A/Puerto-Rico/8/1934 H1N1 (PR8) H1 binds  $\alpha$ 2,6-linked Sias [40] and showed increased binding to all hCK cells compared to all MDCK cells, which is related to the overexpression of ST6GAL1 in hCK cells [16]. Increased binding of PR8 to MDCK-B4GALT1 and MDCK-B3GNT2-

B4GALT1 compared to MDCK WT cells was also observed, which is more pronounced than the results obtained with SNA.

### **hCK-B3GNT2 cells are preferentially bound by contemporary H3 HAs**

After initial characterization, an array of human H3 HAs was used for flow cytometric binding studies with B3GNT2 and B4GALT1 knock-in cells (Fig. 2C, specificities are summarized in Table SI). To cover a broad scope of receptor binding specificities, HAs from viruses from different years (1968-2019) and (sub)clades were chosen. Three HAs from the 3C.2a subclade (A/Singapore/INFH-16-0019/2016, A/Netherlands/00010/2019, and A/Hong-Kong/4801/2014) were chosen to assess the presence of glycans with elongated LacNAc structures [13] on the MDCK and hCK WT and B3GNT2/B4GALT1 knock-in cells.

These human H3 HAs prefer binding to terminal  $\alpha$ 2,6-linked Sias over terminal  $\alpha$ 2,3-linked Sias and therefore, in general, more binding is observed to hCK WT cells than to MDCK WT cells. The HA of A/Hong-Kong/1/1968 does not require multiple consecutive LacNAc repeating units for binding, but it does show a strong preference for glycans with three or four consecutive LacNAc repeating units compared to glycans with only one or two repeating units [9]. The preference of the HA of A/Hong-Kong/1/1968 for longer glycans is however not observed in our flow cytometry experiments, since the B3GNT2 and B4GALT1 knock-in cells did not show increased binding.

The human H3N2 IAVs A/Netherlands/109/2003 and A/Netherlands/761/2009 were previously shown to bind terminal  $\alpha$ 2,6-linked Sias presented on glycans with both two and three, but not one, consecutive LacNAc repeating units [13]. Increased binding of at least a 2-fold difference of the HA of A/Netherlands/109/2003 to all MDCK knock-in cells and all hCK cells compared to MDCK WT cells was observed. Similar binding patterns were observed for the HA of A/Netherlands/761/2009, although the binding to hCK-B3GNT2 cells was much more pronounced and no increased binding to MDCK-B3GNT2 cells was observed. Interestingly, the MDCK-B4GALT1 and MDCK-B3GNT2-B4GALT1 cells often showed comparable or higher binding to the recent H3 HAs than the hCK WT cells, while low levels of  $\alpha$ 2,6-linked Sias are present on all MDCK cells.

As described in the introduction, contemporary H3 IAVs are known to bind to terminal  $\alpha$ 2,6-linked Sias presented on glycans with multiple consecutive LacNAc repeating units, which is especially true for H3N2 viruses of subclade 3C.2a that require at least three subsequent LacNAc repeating units for binding. Increased binding to MDCK-B4GALT1, MDCK-B3GNT2-B4GALT1, and all hCK cell lines compared to MDCK WT cells was observed for the recent H3 HAs (2014-2019) A/Netherlands/354/2016, A/Singapore/INFH-16-0019/2016 (subclade 3C.2a), A/Kansas/14/2017 (3C.3a), and A/Netherlands/00010/2019 (subclade 3C.2a). The hCK-B3GNT2 cells that were already indicated to have the longest glycans by LEL and Gf-CoV-2014 NTD, also showed a substantial increase in binding of these recent H3 HAs compared to all other cell lines investigated. The strong binding to hCK-B3GNT2 cells of the HA of A/Hong-Kong/4801/2014 (subclade 3C.2a) was even more obvious, as other cell lines appear to be barely bound to this HA.

### Lectin binding to cells is concentration and sialic acid-dependent

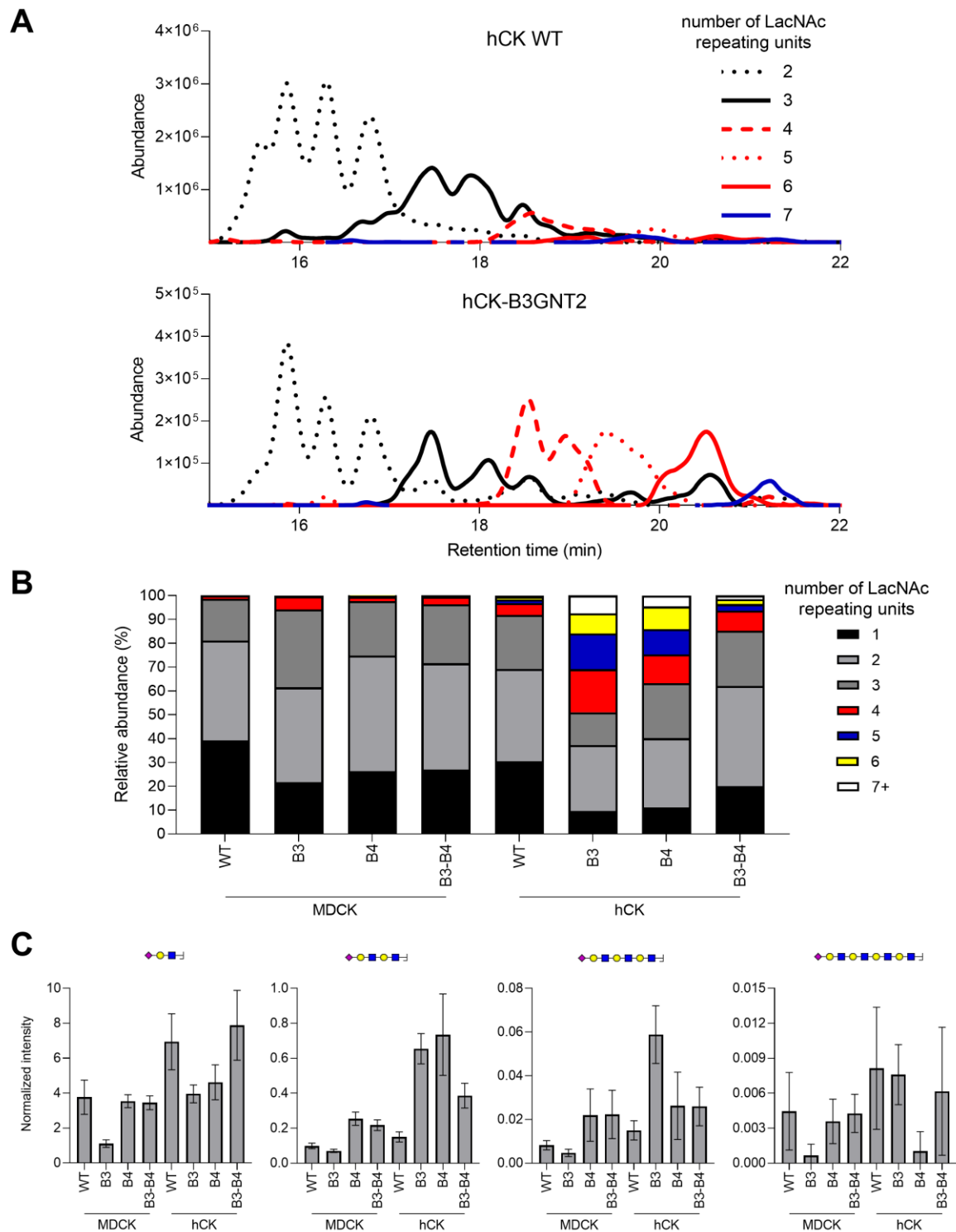
To investigate whether the binding of the H3 HAs in the flow cytometry experiments was indeed specific for Sias, we performed experiments with neuraminidase-treated cells (Fig. S3). As controls for the removal of  $\alpha$ 2,3-linked or  $\alpha$ 2,6-linked Sias, the H5 HA of A/Vietnam/1203/2004 [36, 38, 39] and SNA [34] were used. These lectins showed a decrease in the binding signal after neuraminidase treatment. When testing two H3 HAs with well-defined binding specificities, A/Netherlands/109/2003 and A/Netherlands/761/2009 [13], decreases in binding after neuraminidase treatment were also obtained, indicating that the binding of H3 HAs was indeed Sia-dependent

When examining the binding of the HAs of human H3N2 viruses to cells in Fig. 2C, there appeared to be no binding to the MDCK WT cells at all for A/Hong-Kong/1/1968, A/Netherlands/761/2009, A/Hong-Kong/4801/2014, A/Netherlands/354/2016, A/Singapore/INFH-16-0019/2016, A/Kansas/14/2017, and A/Netherlands/00010/2019, even though it is possible to propagate these viruses in MDCK WT cells. To investigate whether binding to MDCK cells occurred at all, a titration with H3 HAs was performed. As a positive control, A/Netherlands/109/2003 was used since binding was observed in Fig. 2C. Furthermore, the HAs of the well-defined A/Hong-Kong/1/1968 [9] and the subclade 3C.2a virus A/Singapore/INFH-16-0019/2016 were used. The titration indicated that there is indeed binding to MDCK WT cells of these HAs but to a much lesser extent than to hCK WT (or hCK-B3GNT2 cells) (Fig. S4). From the flow cytometric experiments, it appeared that the hCK-B3GNT2 cells present the highest number of consecutive LacNAc repeating units compared to the other cell lines investigated.

### Elongated N-glycans are detected on hCK-B3GNT2 cells

Since N-glycans are the most relevant receptors on cells for IAV [41], we further investigated the N-glycans of WT and B3GNT2/B4GALT1 knock-in MDCK and hCK by HILIC-IMS-QTOF positive mode mass spectrometry (MS) of released N-glycans. Compared to MDCK WT cells, all seven other cell lines showed a large reduction in the relative abundance of high-mannose glycans (Fig. S5A), to which IAV does not bind. Possibly, the presence of high-mannose glycans is influenced by the overexpression of other glycosyltransferases [42, 43], changes in glycan processing time or enzyme spatial localization [42], or a combination of multiple factors [44]. This increase in the relative abundance of complex and hybrid N-glycans may partially explain the improved binding phenotype of the H3 HAs to cell lines different than MDCK WT as observed in the flow cytometry experiments (Fig. 2C).

From the flow cytometry experiments, the hCK-B3GNT2 cells were expected to have the highest number of LacNAc repeating units compared to the other seven cell lines (Fig. 2C). The data obtained from the N-glycan HILIC-IMS-QTOF positive mode mass spectrometry experiments indeed showed a higher relative abundance of N-glycans with more than three LacNAc repeating units in hCK-B3GNT2 compared to hCK WT cells (Fig. 3A). The HILIC-IMS-QTOF positive mode MS data on the N-glycans with at least one LacNAc repeating unit were further



**Fig 3. N-glycan analysis of WT and B3GNT2/B4GALT1 knock-in MDCK and hCK cells using mass spectrometry.** The N-glycans from WT and B3GNT2/B4GALT1 knock-in MDCK and hCK cells were measured using HILIC-IMS-QTOF positive mode mass spectrometry. **(A)** Chromatograms of hCK WT and hCK-B3GNT2 cells were constructed for the glycans with at least two and at most seven LacNAc repeating units. The extracted-ion-counts for the ten most abundant glycan features per LacNAc repeating unit group were summed to yield a chromatogram. **(B)** The N-glycans found in HILIC-IMS-QTOF positive mode mass spectrometry with at least one LacNAc repeating unit were analyzed for the number of



LacNAc repeating units present and the relative abundance was calculated. Further analysis is presented in Fig. S5. Full glycan feature lists for each cell line are presented in Table SII-SIX. (C) Analysis of the *N*-glycans was additionally performed by LC-MS/MS, followed by analysis of the glycan oxonium ions (Table SX). Sialic acid capped (repeating) LacNAc oxonium ions with masses (from left to right) 657.2349, 1022.3671, 1387.4993, and 1752.6315 were identified and the amounts detected were normalized to the core fragments. Mean and standard errors (n=3) are shown. Further analysis is shown in Fig. S6 and annotated spectra are present in Fig. S7.

analyzed to determine the relative abundance of glycans with a different number of LacNAc repeating units in the eight different cell lines (Fig. 3B). Overall, relatively less glycans were detected with four or more LacNAc repeating units compared to an earlier study on MDCK and hCK WT cells [25], of which the meta-analysis is presented in Fig. S1A, although different methods were used in the previous study and not all peaks were assigned. Most importantly, the highest increase in the relative abundance of elongated *N*-glycans was observed in hCK-B3GNT2 cells (Fig. 3B), which agrees with the flow cytometric results (Fig. 2B-C).

For binding of contemporary H3N2 IAVs, *N*-glycans with at least three consecutive LacNAc repeating units should be capped with  $\alpha$ 2,6-linked Sias. While we were unable to determine the Sia linkage, we analyzed the percentage of sialylation of the *N*-glycans with at least one LacNAc repeating unit (Fig. S5B). In general, 75-96% of these *N*-glycans were sialylated, except for the glycans of MDCK-B3GNT2 cells (38% sialylated), which correlated with the low binding of H3 HAs as observed using flow cytometry (Fig. 2C). Notably, meta-analysis of a previous study [25] showed even lower degrees of sialylation of glycans from MDCK and hCK WT cells (65% and 79% respectively, Fig. S1B), although the methods used in this study were different. Whereas non-sialylated *N*-glycans occurred in all groups of LacNAc lengths in MDCK WT and B3GNT2/B4GALT1 knock-in cells, all *N*-glycans with at least four LacNAc repeating units (except for four glycans in total) on all hCK cell lines were sialylated (Table SII-SIX).

To further elucidate the *N*-glycan structures, MS/MS experiments were performed. As outcome, we integrated from MS/MS spectra the relative abundance of oxonium ions that could indicate glycan structural elements (Table SX). These included, among others, LacNAc repeating units from one to four units (respectively, *m/z* 366.1395, 731.2717, 1096.4039 and 1461.5361), as well as their variants with *N*-acetylneuraminic acid (NeuAc) sialylation (respectively, *m/z* 657.2349, 1022.3671, 1387.4993 and 1752.6315). While included in the analysis, no 2-keto-3-deoxynonic acid (KDN), *N*-glycolylneuraminic acid (NeuGc), acetylation, sulfation or phosphorylation were found. On the other hand, several ions were detected that could delineate  $\alpha$ -Gal and bisection, although these were low in intensity and could overlap with other compositions. We focused on the glycan fragments with one to four consecutive LacNAc repeating units, both with and without sialic acid (Fig. 3C, Fig. S6, Fig. S7), since contemporary H3N2 IAV require at least three consecutive LacNAc repeating units for binding.

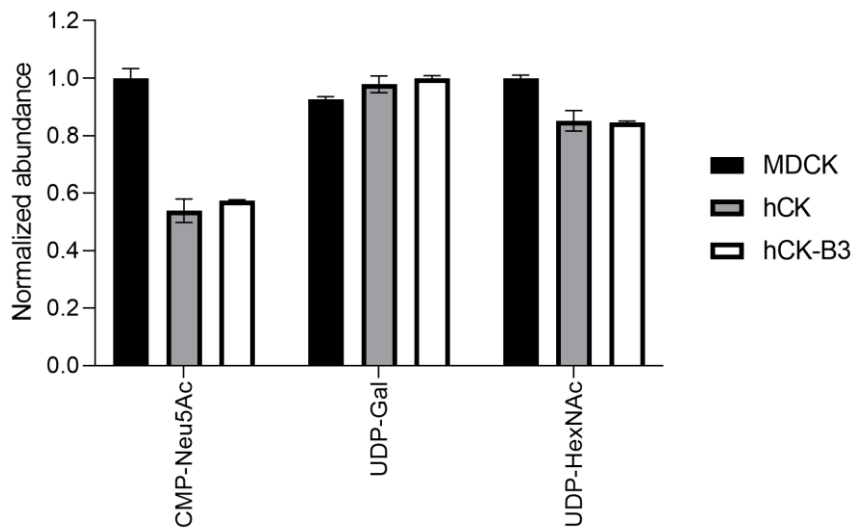
Although the knock-in cell lines MDCK-B4GALT1 and MDCK-B3GNT2-B4GALT1 did not show an increase in the number of LacNAc repeating units compared to MDCK WT in MS1 analysis (Fig. 3B), a clear increase in the glycan fragments with two and three consecutive LacNAc repeating units was observed compared to MDCK WT cells (Fig. 3C, Fig. S6). In MDCK-B3GNT2 and hCK WT cells, the relative increase of glycans with four LacNAc repeating units in the MS1 analysis was a few percent compared to MDCK WT cells (Fig. 3B). However, for the MDCK-B3GNT2 cells, this relative increase in consecutive LacNAc repeating units is not observed (Fig. 3C, Fig. S6). For the hCK WT cells, the MS1 and MS/MS analyses correlate, since a slight increase in the sialylated glycan fragments with one to four consecutive LacNAc repeating units (Fig. 3C) was observed. In contrast, in an earlier study [25], slightly more glycans with four or more LacNAc repeating units were found to be present on MDCK WT cells compared to hCK WT cells (meta-analysis presented in Fig. S1A). The relative abundance of LacNAc repeating units in our MS1 analysis was even higher in hCK-B3GNT2-B4GALT1 cells (Fig. 3B), which is also observed in the MS/MS analysis when looking at the sialylated glycan fragments with two and three LacNAc repeating units (Fig. 3C).

Surprisingly, hCK-B4GALT1 showed a substantial increase in the relative abundance of *N*-glycans with a higher number of LacNAc repeating units compared to hCK WT cells in the MS1 analysis, up to even nine LacNAcs (Fig. 3B, Table SVIII), which was not expected from the results of the flow cytometry experiments (Fig. 2). The high relative abundance of LacNAc repeating units in hCK-B4GALT1 cells (Fig. 3B) appears to be presented mainly as sialylated fragments with two instead of three or four consecutive LacNAc repeating units (Fig. 3C), which explains the lower binding of H3 HAs to these cells compared to hCK-B3GNT2 cells in the flow cytometry experiments (Fig. 2C). Most sialylated glycan fragments with three consecutive LacNAc repeating units, the required receptor for contemporary H3N2 IAV, were found on hCK-B3GNT2 (Fig. 3C), which again correlates to the increased binding of H3 HAs as observed in flow cytometry (Fig. 2C).

### **Sugar nucleotides are not a limiting factor in the biosynthesis of poly-LacNAc structures**

Changes in sugar nucleotide levels have been observed in cells after overexpression of B3GNT2 and B4GALT1 [32]. Although a clear increase in the number of consecutive LacNAc repeating units was observed in especially hCK-B3GNT2 cells, we wondered whether the further elongation of glycans may have been limited by depletion of sugar nucleotides. Therefore, the concentrations of sugar nucleotides in the cell lysates of MDCK WT, hCK WT, and hCK-B3GNT2 cells were measured by mass spectrometry (Fig. 4, with details in Fig. S8) [45]. The overexpression of ST6GAL1 in hCK WT and hCK-B3GNT2 cells compared to MDCK WT cells resulted in elevated levels of sialylation and thereby lower levels of available CMP-Neu5Ac, which was also demonstrated in the sugar nucleotide analysis. No other major differences or depletions of sugar nucleotides were observed in any of the cell lines. Most importantly, the sugar nucleotides that are required for the biosynthesis of LacNAc repeating units (UDP-galactose and UDP-

HexNAc) were not depleted in any of the cell lines, thus sugar nucleotide availability is likely not a limiting factor for the elongation of glycans.



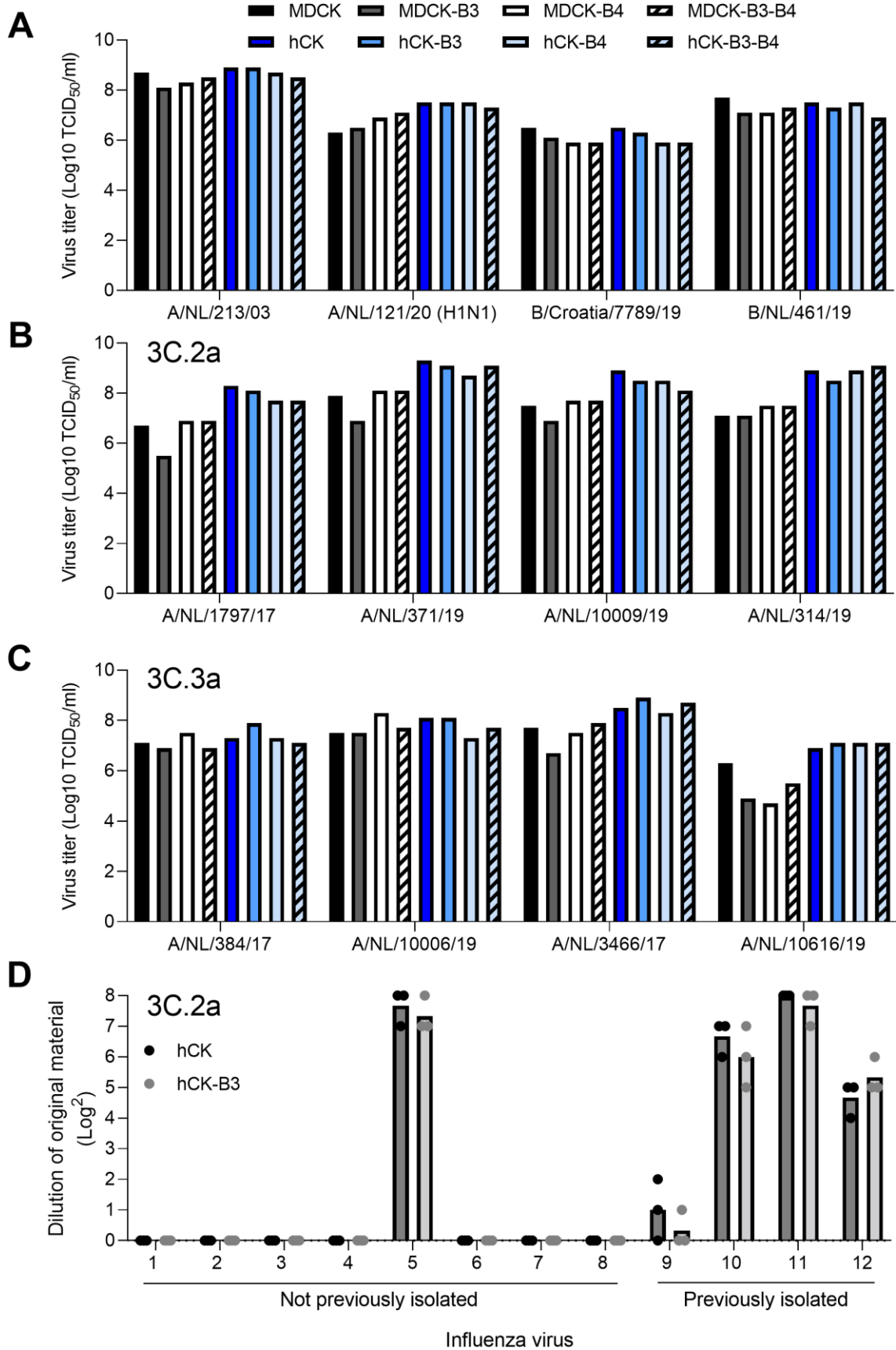
**Fig 4. Sugar nucleotide analysis of MDCK, hCK, and hCK-B3GNT2 cells.** The sugar nucleotides in the lysate of MDCK, hCK, and hCK-B3GNT2 cells were analyzed by mass spectrometry (n=2). The normalized abundance of CMP-Neu5Ac, UDP-Gal, and UDP-HexNAc are shown. Normalization was performed on the cell line with the highest amount of each sugar nucleotide. Detailed information about all measured sugar nucleotides is presented in Fig. S8.

### Improved binding of hemagglutinins to cells does not ensure higher virus titers and improved isolation of influenza viruses

MDCK WT, hCK WT, and B3GNT2 and/or B4GALT1 knock-in cells were inoculated with H3N2 viruses to investigate whether higher titers could be obtained in the knock-in cells. Four control viruses (H3N2 from 2003, H1N1, and influenza B, Fig. 5A) and eight recent (2017-2019) H3N2 viruses from the 3C.2a (Fig. 5B) and 3C.3a (Fig. 5C) subclades were used for inoculation. For the H3N2 virus from 2003, the H1N1 virus, and the influenza B viruses, no substantial difference was observed between the virus titers obtained in MDCK, hCK, or B3GNT2/B4GALT1 knock-in cells (Fig. 5A).

Recent 3C.2a viruses are known to only bind glycans having at least three consecutive LacNAc repeating units, while recent 3C.3a viruses also bind glycans with two consecutive LacNAc repeating units [13]. For the 3C.2a viruses, a considerable difference was visible between the titers in MDCK WT and hCK WT cells (Fig. 5B), which correlates with the increased binding as observed in the flow cytometric experiments (Fig. 2). Surprisingly, no substantial difference between the titers in hCK WT and hCK-B3GNT2 cells was observed, while the glycans on the latter cell line were extended as observed in the flow cytometry (Fig. 2) and glycan mass spectrometry experiments (Fig. 3). Furthermore, no difference was observed in the titers for the 3C.3a viruses (Fig. 5C), not even between MDCK and hCK cells. Therefore, we concluded that additional binding does not necessarily lead to a higher infection efficiency.





**Fig 5. Influenza virus inoculation of B3GNT2 and B4GALT1 knock-in MDCK and hCK cells.** End-point titrations with four control viruses and eight recent H3N2 IAVs (details in Table I) were performed, of which (A) four control viruses, (B) four 3C.2a viruses, and (C) four 3C.3a viruses. Infectious titers were determined either using a hemagglutination assay (A) or a nucleoprotein staining (B, C) when a hemagglutination assay was not possible. (D) An infection study using hCK and hCK-B3GNT2 cells with a twofold dilution of twelve H3N2 IAVs from the 3C.2a clade (details in Table II) which could previously either not be isolated in hCK cells (#1-8) or could be isolated in hCK cells (#9-12) was performed. Infection was assessed by the presence of cytopathic effects. Individual and mean values are shown.

While additional strains of H3N2 viruses can be isolated in hCK cells as compared to MDCK-SIAT1 and MDCK cells [16], we noticed that some H3N2 viruses could still not be isolated in hCK cells. To investigate whether cell lines with longer glycans (hCK-B3GNT2 cells) would facilitate the isolation of additional virus strains, we attempted to isolate twelve H3N2 viruses (clade 3C.2a) from original patient material in hCK and hCK-B3GNT2 cells (Fig. 5D). Since hCK cells are better suited for the isolation of these viruses [16] and the original patient material was only available in limited quantities, we did not attempt to also isolate the viruses from MDCK cells. Eight of the tested viruses could not be isolated previously since they did not replicate in hCK cells (Fig. 5D). All viruses that were previously isolated in hCK cells were again successfully isolated. However, the use of hCK-B3GNT2 cells did not result in more efficient isolation of those viruses. Of the viruses that could not be isolated previously, only A/Netherlands/173/2019 could now be isolated. However, isolation was achieved in both hCK and hCK-B3GNT2 cells. Therefore, a higher number of extended glycans did not improve the isolation of H3N2 IAVs from the 3C.2a clade.

## Discussion

Although we elongated the LacNAc repeating units on the glycans of MDCK and hCK cells, it did not result in a higher infection efficiency of recent H3N2 IAVs. It has been reported that a low but critical threshold of high-affinity receptors is required for infection [30, 31], though binding and infection are further assisted by the presence of high-abundance low-affinity receptors [29]. We previously showed that the presence of high-affinity receptors is indeed essential for infection of cells [9]. This implies that increased HA binding will lead to enhanced entry efficiency. It is therefore counterintuitive that presenting preferred ligands in copious amounts does not lead to increased infection. Strikingly, here we demonstrated that increased HA binding to cells does not necessarily result in more efficient infection.

On the other hand, several studies have indicated that increasing the number of preferred receptors will increase the infection efficiency of IAVs [16, 22, 23, 28], which was also shown with MDCK-SIAT1 [22], MDCK-AX4 [23], and hCK [16] cells. A possible explanation for this discrepancy may lie in the glycoproteins on which *N*-glycans, the presumed glycan receptors for IAVs [41], are presented. It has been suggested previously that only specific sialylated glycoproteins can be used as a receptor for IAV [41, 46], such as the voltage-dependent  $\text{Ca}^{2+}$  channel  $\text{Ca}_v1.2$  [47], NKp44 [48, 49], NKp46 [49-51], epidermal growth factor (EGFR) [52], and nucleolin [53]. Although we demonstrated that glycans on B3GNT2/B4GALT1 knock-in cells contained a higher number of LacNAc repeating units, we have not determined on which glycoproteins the elongated glycans are present. Possibly, the glycans that are used as a receptor and are present on specific glycoproteins can be modified in their *Sia* linkage, as done in MDCK-SIAT1, MDCK-AX4, and hCK cells, but not in the number of LacNAc repeating units. Additionally, the glycoproteins on which the glycans are presented may not be clustered enough [54] to support IAV infection. This would explain why the infection efficiency could not be increased by the elongation of LacNAc repeating units on MDCK and hCK cells.

Nevertheless, not all recent H3N2 IAVs could be isolated efficiently in hCK cells [16], as also shown in Fig. 5D. Possibly, no viable virus particles were present in the patient samples from which we attempted to isolate virus. Alternatively, we may be overlooking an identified [47-53] or an unidentified glycoprotein that is not present (in high enough quantities) on hCK cells. Furthermore, other types of glycans, such as phosphorylated [55] and sulfated glycans [2, 11, 56], possibly act as a receptor for IAV. Due to our sample preparation for the released glycan mass spectrometry analysis (acidic and basic conditions), we were unable to measure phosphorylated glycans. Potentially, some sulfated glycans could be partially retained. However, we did not detect sulfated glycans in the MS/MS analysis. To increase the isolation of recent H3N2 IAVs, it is of foremost importance to investigate the limiting factor in the infection efficiency of these viruses.

The overexpression of B3GNT2 and/or B4GALT1 is responsible for the elongation of glycans. Previously, overexpression of B4GALT1 was found to result in the elongation of glycans on CHO cells [32]. From the flow cytometry analysis, B4GALT1 appeared to be the limiting factor for the elongation of glycans in MDCK

cells, while the elongation was limited in hCK cells when B3GNT2 was not present in high enough amounts. Surprisingly, the elongation of glycans appeared to be inhibited in MDCK cells by the overexpression of B3GNT2, as can be observed when comparing MDCK-B3GNT2 cells with MDCK WT cells, but also when comparing MDCK-B3GNT2-B4GALT1 cells with MDCK-B4GALT1 cells. In the latter two, the overexpression of B4GALT1 is at the same level (Fig. 1C), but longer glycans are present on MDCK-B4GALT1 cells (Fig. 2B-C and 3C). In hCK cells, sialyltransferase expression is severely modified by the overexpression of ST6GAL1 and the knockout of all ST3GAL enzymes. Both the heavily overexpressed ST6GAL1 and B3GNT2 in hCK cells use terminal galactose as a substrate. The overexpression of B3GNT2 in hCK cells perhaps restores the balance between ST6GAL1 and B3GNT2, thereby allowing B3GNT2 to use the terminal galactose as a substrate again for the elongation of glycans before sialylation takes place. On the other hand, in MDCK cells, the balance may be skewed even more by the overexpression of B3GNT2, leading to the low sialylation of glycans on these cells (Fig. S5B). Furthermore, the cellular localization of the enzymes and potential enzyme complex formation may be of importance, which is not regulated with overexpression alone.

Our observations indicate that only few suitable glycan receptors are required for efficient infection. This is in line with our previous observations that an increase from 2.7% to 8.7% of sialylated glycans with at least three consecutive LacNAc repeating units on, respectively, unmodified and modified turkey erythrocytes allowed for the binding of contemporary H3N2 viruses [13]. Also in ferrets, an animal model that is often used to study human influenza viruses [39], the presence of glycans in the respiratory tract (lung, trachea, soft palate, nasal turbinate, nasal wash) was investigated. Elongated glycans were present solely as *N*-glycans, with a maximum of 9 LacNAc repeating units per glycan, but at most 0.17% of the detected glycans had at least three consecutive LacNAc repeating units terminating with Sia, which is required for H3N2 IAV binding [57]. Although the glycans in the human trachea have not been analyzed yet, data is available on other parts of the human airway system. Sensitive methods indicated the presence of extended *N*-glycans with up to 10 LacNAc repeating units in human lung tissue. However, at most 0.3% of the *N*-glycans were found to contain at least 3 consecutive LacNAc repeating units. The *N*-glycans in the bronchus and nasopharynx contained a lower number of LacNAc repeating units than in the lung [58]. Another study found *N*-glycans with up to 22 LacNAc repeating units in the lung. Even though the majority of the Sias were found to be  $\alpha$ 2,6-linked instead of  $\alpha$ 2,3-linked, the  $\alpha$ 2,6-linked Sias were mainly present on the shorter glycans [59], further supporting our conclusions that only minor amounts of suitable glycan receptors are required for efficient infection by IAVs. Therefore, we conclude that factors other than glycosylation must be investigated to address the suboptimal propagation of recent human H3N2 IAVs as an aid to their further study.

## Material and Methods

### Cell culturing and preparation of cell lysates

Cells were cultured in DMEM (Gibco) with 10% FCS (S7524, Sigma) and 1% penicillin and streptomycin (Sigma). All hCK cells [16], knock-in and WT, were maintained with an additional 10 µg/ml blasticidin and 2 µg/ml puromycin in the medium. B3GNT2 and B4GALT1 knock-in cells were maintained in medium containing an additional 300 µg/ml Hygromycin B, a concentration that was determined to kill MDCK (CCL-34) and hCK (a kind gift from Yoshihiro Kawaoka) without Hygromycin B resistance genes. Detaching of the (knock-in) MDCK and hCK cells was always done using 1X TrypLE Express Enzyme (12605010, Thermo Fisher Scientific), using 2 ml in a T75 flask, at a confluency of approximately 90%.

Cell lysates were obtained after first washing cell monolayers once using D-PBS (D5837, Sigma). Cells were subsequently harvested after incubation at 37°C for 20 min with 1X TrypLE Express Enzyme (2 ml in a T75 flask) at a confluency of approximately 90%. The cell suspension was centrifuged for 5 min at 250 rcf. The cell pellets were lysed by the addition of RIPA lysis buffer (20-188, Merck Millipore) supplemented with protease inhibitor (A32965, Thermo Fisher Scientific), which was vortexed for 20 sec. The suspension was incubated on ice for 30 min, after which it was centrifuged at 16500 rcf in a fixed-angle centrifuge at 4°C, after which the supernatant was used as cell lysate.

### Cloning of lentiviral transfer plasmids

Plasmid pCF525-EF1α-Hygro-P2A-mCherry-lenti [60] was a gift from Jennifer Doudna (Addgene plasmid # 115796) and was used as the backbone for the transfer plasmid. Three transfer plasmids were constructed (pCF-B3GNT2, pCF-B4GALT1, and pCF-B3GNT2-B4GALT1). The region between the P2A and WPRE was removed and replaced by either the *B3GNT2* or *B4GALT1*. When the genes of both glycosyltransferases were cloned into the plasmid they were connected with a T2A self-cleaving peptide. The *B3GNT2* and *B4GALT1* genes were always preceded by the signal sequence of the human GalT, which we copied from the EGFP-GalT plasmid (gift from Jennifer Lippincott-Schwartz, Addgene plasmid # 11929) [61]. The T2A self-cleaving peptide was amplified from plasmid tetO.Sox9.Puro [62], which was a gift from Henrik Ahlenius (Addgene plasmid # 117269). The *B3GNT2* and *B4GALT1* genes were amplified from plasmids B3GNT2-pGEn2-DES and B4GALT1-pGEn2-DES, which are a gift from Kelly Moremen and are available via . All segments were amplified with an overhang, using the primers indicated in Table SXI. Assembly of the plasmids was performed using Gibson assembly, after which they were sequenced to ensure correct amplification and assembly.

### Lentiviral integration of the *B3GNT2* and *B4GALT1* genes

Lentiviral particles were produced using HEK293T cells [63]. One of the transfer plasmids as described above, together with the packaging plasmids pMDLg/pRRE, pRSV-Rev, and pMD2.G, which were kind gifts from Didier Trono [64] (Addgene plasmids #12251, #12253, and #12259 respectively) were used. The day

before transduction, MDCK (CCL-34) and hCK cells were seeded in a 6-wells plate at a density of 100.000 cells per well. Transduction with 0.5-3  $\mu$ l of lentivirus was performed in the presence of 8  $\mu$ g/ml polybrene with 1 ml fresh medium per well. The medium was replaced with fresh medium containing 300  $\mu$ g/ml Hygromycin B at 18 hours after transduction. Cells were grown until no Hygromycin B sensitive cells were remaining. Cells were always maintained in the presence of 300  $\mu$ g/ml Hygromycin B.

### **RT-qPCR analysis on *B3GNT2*, *B4GALT1*, and *ST6GAL1* genes**

RNA extraction was performed using the GeneJET RNA purification kit (Thermo Fisher Scientific) according to the manufacturer's protocol, after which the DNA was treated with DNase I (#EN0251, Thermo Fisher Scientific). RT-qPCR was performed using the Luna universal one-step RT-qPCR kit (#E3005, New England Biolabs) according to the provided protocol, in which 10 ng of DNase I-treated RNA was used. Primers (Table SXII) for *B3GNT2*, *B4GALT1*, and *ST6GAL1* were designed to anneal both in the human and dog genome. Primers for GAPDH (household/reference gene) were designed using the dog genome. Experiments were performed in triplicate and Ct values of the RT-qPCR experiments on the glycosyltransferases were compared to the average Ct value of GAPDH of that specific cell line under the assumption that the amount of DNA doubles every cycle. The means and standard deviations of the amount of DNA relative to GAPDH were calculated.

### **Overexpression of *B3GNT2*, *B4GALT1*, and *ST6GAL1* on the protein level**

Cell lysates, obtained as described above, were further used for the label-free quantification of *B3GNT2*, *B4GALT1*, and *ST6GAL1* proteins. From the cell lysates, 10  $\mu$ g of protein was denatured, reduced, and alkylated by adding 100  $\mu$ l of a solution consisting of 150 mM Tris, 5mM TCEP (tris(2-carboxyethyl)phosphine), 30 mM chloroacetamide (CAA), and 1% sodium deoxycholate (SDC) at pH 8.5. Next, 100 ng lysyl endopeptidase (129-02541, Wako Chemicals GmbH) and 100 ng trypsin (T1426, Sigma) were added and the samples were incubated overnight at 37°C. The samples were then acidified by adding formic acid (FA) to a concentration of 0.5% before solid-phase extraction (SPE) sample clean-up, causing the SDC to precipitate. SPE clean-up was performed on an Oasis HBL u-elution plate.

After the SPE clean-up, the samples were dried with a vacuum centrifuge. Subsequently, the sample was reconstituted in 2% FA before analysis on the Orbitrap Exploris mass spectrometer (Thermo Scientific) connected to a UHPLC 3000 system (Thermo). Approximately 200 ng of reconstituted peptides were trapped on a pre-column and then separated on a 50 cm x 75  $\mu$ m Poroshell EC-C18 analytical column (2.7  $\mu$ m) temperature controlled at 40°C. Solvent A consisted of 0.1% FA, solvent B of 0.1% FA in 80% acetonitrile, and different combinations of solvent A and B were used in the next steps. Trapping was performed for 2 min in 9% solvent B. Peptides were separated by a 65 min gradient of 9–44 % buffer B followed by 44–99% B in 3 min, and 99% B for 4 min. MS data were obtained in a data-dependent acquisition mode. The full scans were

acquired in the  $m/z$  range of 350-1600 at the resolution of 60,000 ( $m/z$  400) with AGC target 3E6. The most intense precursor ions were automatically selected for higher-energy collisional dissociation (HCD) fragmentation performed at normalized collision energy 28, after accumulation to the target value of 1E5. MS/MS acquisition was performed at a resolution of 15,000. Protein identification was done with Byonic (Protein Metrics). A search was performed against the dog proteome (UP000002254\_9615) with the addition of the human B3GNT2, B4GALT1, and ST6GAL1 sequences. The search was performed with specific digestion C-terminal of R/K, allowing 3 missed cleavages, using precursor and fragment mass tolerances of 12 and 24 ppm, respectively. Carbamidomethylation of cysteine was set as a fixed modification and oxidation of the methionine or tryptophan as a variable modification. Peptides unique for the human B3GNT2, B4GALT1, and ST6GAL1 were manually selected and the MS1 peak areas were integrated with Skyline and normalized against the combined MS1 signals for identified peptides of tubulin- $\beta$  (E2RFJ7). The peptide library for Skyline was built by repeating the search with a focused database containing only the human B3GNT2, B4GALT1, and ST6GAL1 and tubulin- $\beta$  sequences. The mass spectrometry proteomic data have been deposited to the ProteomeXchange Consortium via the PRIDE [65] partner repository with the dataset identifier PXD037175.

### **Expression and purification of trimeric HA for binding studies**

Recombinant trimeric IAV hemagglutinin ectodomain proteins (HA) were cloned into the pCD5 expression vector as described previously [66, 67], in frame with a GCN4 trimerization motif (KQIEDKIEEIESKQKKIENEIARIKK), a superfolder GFP [40] or mOrange2 [68] and the Twin-Strep-tag (WSHPQFEKGGGSGGGSSWSHPQFEK); IBA, Germany). An example of such a plasmid is addgene plasmid #182546 [36], which contains the codon optimized HA sequence (synthesized by Genscript) of A/Vietnam/1203/2004 H5. The codon optimized genes of the other HAs (Table SXIII) were cloned into the plasmid using the restriction sites NheI and PacI. The trimeric HAs were expressed in HEK293S GnTI(-) cells with polyethyleneimine I (PEI) in a 1:8 ratio ( $\mu\text{g DNA}:\mu\text{g PEI}$ ) for the HAs as previously described [66], while a 1:12 ratio was used for the NTD of the guinea fowl CoV. The transfection mix was replaced after 6 hours by 293 SFM II suspension medium (Invitrogen, 11686029), supplemented with sodium bicarbonate (3.7 g/L), Primatone RL-UF (3.0 g/L, Kerry, NY, USA), glucose (2.0 g/L), glutamax (1%, Gibco), valproic acid (0.4 g/L) and DMSO (1.5%). Culture supernatants were harvested 5 days post-transfection and purified with sepharose strep-tactin beads (IBA Life Sciences, Germany) according to the manufacturer's instructions.

### **Flow cytometry studies**

Cells were harvested using TrypLE Express Enzyme as described above. After removal of the supernatant, cells were resuspended in PBS supplemented with 1% FCS (S7524, Sigma) and 2mM EDTA and kept at 4°C until the plate was measured in the flow cytometer. For experiments with  $\alpha$ 2-3,6,8,9 neuraminidase A (#P0722, New England Biolabs), neuraminidase (NA) was used 1:200 with 1,000,000 cells per ml in glycobuffer 1 (5 mM  $\text{CaCl}_2$ , 50 mM sodium acetate, in MQ water, at pH 5.5) for 16 hours at 37°C on a shaking platform in the dark, before incubation with the



lectin/HA mixes. In a round-bottom 96-wells plate (353910, Falcon), 150,000 cells were used. Per well, 100  $\mu$ l of PBS supplemented with 1  $\mu$ g of HA or biotinylated lectin (SNA (B1305), LEL (B1175), ECA (B1145), all from Vector Laboratories) was used, to achieve a final concentration of 10  $\mu$ g/ml. Hemagglutinins were precomplexed (on ice, 20 min) with 1.3  $\mu$ g monoclonal antibody detecting the Twin-Strep-tag and 0.325  $\mu$ g goat anti-human Alexa Fluor 488 (A11013, Invitrogen). Biotinylated lectins were precomplexed (on ice, 20 min) with 0.2  $\mu$ g streptavidin Alexa Fluor 488 (S32354, Invitrogen). For titration experiments, different amounts of HA, lectin, precomplexing antibodies, or streptavidin were used. Furthermore, eBioscience Fixable Viability Dye eFluor 780 (65-0865, Thermo Fisher Scientific) was diluted 1:2000 in the same mixture. Cells were incubated with the hemagglutinin/lectin mix for 30 minutes at 4°C in the dark. Cells were washed once with PBS supplemented with 1% FCS and 2 mM EDTA, after which the cells were fixed with 100  $\mu$ l of 1% paraformaldehyde in PBS for 10 minutes. Afterward, cells were washed once using PBS supplemented with 1% FCS and 2 mM EDTA, after which they were resuspended in 100  $\mu$ l of the same buffer. Flow cytometry was performed using the BD FACSCanto II (BD Biosciences) using appropriate laser voltages. Data were analyzed using FlowLogic (Inivai Technologies) and gated as described in Fig. 2A to consecutively select cells, single cells, and cells that are not dead. Mean fluorescence values of triplicates were averaged and standard deviations were calculated. Curves for titration experiments were smoothed using the standard settings.

### **Identification of N-glycans on cells by mass spectrometry**

Cell lysates of WT and B3GNT2/B4GALT1 knock-in MDCK and hCK cells were obtained as described above. The total protein concentration in the cell lysates was determined using a BCA assay. The glycans in 400  $\mu$ g of total protein were released by PNGaseF treatment. Proteins were first denatured in DTT/SDS (40 mM DTT, 0.5% v/v SDS) for 8 minutes at 95°C, after which they were cooled on ice. Subsequently, NP-40 (1% v/v) and glycobuffer G7 (50 mM sodium phosphate at pH 7.5) were added, together with 30  $\mu$ g of PNGaseF. The samples were incubated in a shaking incubator overnight at 37°C. Samples were centrifuged (4700 rcf, 3 min) to remove potential precipitate, after which they were loaded on separate C18 SPE cartridges (Avantor™ 7020-02 BAKERBOND™ SPE Octadecyl), which were beforehand conditioned with 1 ml acetonitrile (MeCN) and 1 ml MQ water. The flow-through was collected and the remaining glycans were eluted from the C18 cartridges with 1 ml of 5% MeCN and 0.05% trifluoroacetic acid (TFA) in MQ water. The MeCN and TFA in both samples were evaporated under a stream of nitrogen gas. Flow through and elution fractions were diluted into 500  $\mu$ l MQ water and combined, after which PGC SPE cartridges (Thermo Scientific™ HyperSep™ Hypercarb™ SPE cartridges) were used to further purify the samples according to procedures adapted from [69]. The PGC SPE cartridges were conditioned with 1 ml MeCN and 1 ml MQ water, after which the samples were loaded on the cartridges. The cartridges were washed with 1 ml 0.05% TFA in MQ water and 1 ml 5% MeCN with 0.05% TFA in MQ water. Samples were eluted with 50% MeCN and 0.1% TFA in MQ water and evaporated under a stream of nitrogen gas. The dried glycans were dissolved in 30  $\mu$ l MQ water and 6  $\mu$ l pure glacial

acetic acid and labeled using 5  $\mu$ l procainamide (105 mg/ml procainamide HCl in DMSO) and 5  $\mu$ l 2-picoline borane (107 mg/ml 2-Methylpyridine borane complex in DMSO) [70, 71]. The solution was vortexed thoroughly and incubated for 2 hours at 65°C, after which the samples were evaporated using the vacuum concentrator. The sample was dissolved in 300  $\mu$ l MQ water and vortexed until the pellets were dissolved, after which 5  $\mu$ l 25% (w/v) ammonia was added per sample to ensure a pH above 10. To remove the unused procainamide from the reaction mixture, liquid-liquid extraction with 500  $\mu$ l dichloromethane was performed three times, with centrifuge steps of 4700 rcf for 3 min in between. The dichloromethane was removed and residual dichloromethane in the aqueous layer was evaporated under a stream of nitrogen gas. The samples were dissolved in a total of 1 ml MQ water after which they were loaded onto PGC SPE cartridges (conditioned with 2 ml MeCN and 2 ml MQ water). The cartridges were washed with 2 ml MQ water and the glycans were eluted using 50% MeCN with 0.1% TFA in MQ water, after which the MeCN and TFA were evaporated under a stream of nitrogen gas, followed by lyophilization.

Before HILIC-IMS-QTOF analysis, the lyophilized samples were reconstituted in 15  $\mu$ l 70% MeCN in MQ water and centrifuged. The injected volume was 10  $\mu$ l. The HILIC-IMS-QTOF system was an Agilent 1260 Infinity LC coupled to a 6560 IM-QTOF mass spectrometer (Agilent Technologies, Santa Clara, USA). For HILIC separation, a SeQuant ZIC-cHILIC column (3  $\mu$ m, 100 Å; 150 x 2.1 mm) was used with a matching guard column (20 x 2.1 mm). The temperature of the column compartment was set at 40 °C. The mobile phase was composed of eluent A: 10 mM ammonium formate with 10 mM formic acid in MQ water, and eluent B: LC-MS grade MeCN. The initial eluent composition was 30% A at a flow rate of 0.2 ml/min, followed by a linear gradient to 50% A from 0 to 20 minutes. 50% A was held isocratically until 25 minutes. To re-establish initial conditions, the column was flushed with at least 10 column volumes of 30% A.

The IMS-QTOF was set to positive ion mode with a capillary voltage of 3500 V, nozzle voltage of 2000 V, and a fragmentor voltage of 360 V. The drying gas temperature was 300 °C with a flow rate of 8 l/min and the sheath gas temperature was 300 °C at 11 l/min. The nebulizer pressure was set at 40 psi. The ion mobility settings were set as follows: 18 IM transients per frame, an IM trap fill time of 3900  $\mu$ s and a release time of 250  $\mu$ s, the drift tube voltage was 1400 V, and the multiplexing pulsing sequence length was 4 bits.

IM-MS data was calibrated to reference signals of  $m/z$  121.050873 and 922.009798 using the IM-MS reprocessor utility of the Agilent Masshunter software. The mass-calibrated data was then demultiplexed using the PNNL preprocessor software using a 5-point moving average smoothing and interpolation of 3 drift bins. To find potential glycan hits in the processed data, the 'find features' (IMFE) option of the Agilent IM-MS browser was used with the following criteria: 'Glycans' isotope model, limited charge state to 5 and an ion intensity above 500. The found features were filtered by  $m/z$  range of 300 – 3200 and an abundance of over 500 (a.u.) where abundance for a feature was defined as 'max ion volume' (the peak area of the most abundant ion for that feature).

After exporting the list of filtered features, glycans with a mass below 1129 Da (the mass of an *N*-glycan core) were removed. The ExPASy GlycoMod tool [72] was used to search for glycan structures (monoisotopic mass values, 5 ppm mass tolerance, neutral, derivatized *N*-linked oligosaccharides, procainamide (mass 235.168462) as reducing terminal derivative, looking for underivatized monosaccharide residues (Hexose, HexNAc, Deoxyhexose, and NeuAc)). Although we initially also screened for NeuGc, it did not provide plausible glycan composition hits. Instead, most possible NeuGc hits were isobaric Deoxyhexose+NeuGc = Hexose+NeuAc compositions, where the Hexose+NeuAc composition was deemed more realistic. Indeed, commonly observed NeuGcLacNAc (673.2298 *m/z*) in-source oxonium fragments were absent, while NeuAcLacNAc (657.2298 *m/z*) was found in high abundance. This is further supported by our previous work in which we showed that NeuGc is absent on MDCK cells [36]. Thus, we removed NeuGc from the analysis to simplify the further processing of the data. For features with multiple potential monosaccharide combinations, the most realistic glycan in the biological context was chosen. The abundance of glycan features with the same mass, composition, and a maximum difference of 0.1 min in the retention time were combined as one isomer. Full glycan composition feature lists for the different cell lines are presented in Table SII-SIX.

Analysis of the number of LacNAc repeating units was performed on the complex and hybrid *N*-glycans with at least one LacNAc repeating unit. A glycan with one LacNAc repeating unit was defined as a glycan with 4 hexoses and a minimum of 3 HexNAcs or 3 HexNAcs and at least 4 hexoses. A glycan with two LacNAc repeating units was defined as a glycan with 5 hexoses and a minimum of 4 HexNAcs or 4 HexNAcs and at least 5 hexoses. This pattern was continued for the higher numbers of LacNAc repeating units. The total absolute abundance of all selected glycans was added up, after which the relative abundance of a given number of LacNAc repeating units was calculated from this total. Additionally, the percentage of these glycans with at least one Sia was calculated.

Chromatograms of the *N*-glycans with two to seven LacNAc repeating units, calculated as described above, from hCK WT and hCK-B3GNT2 cells were constructed using Agilent's Masshunter Qualitative Analysis 10.0 software (Fig. 3A). The shown chromatograms are the summed extracted-ion-count (EIC) for the ten most abundant glycan features per LacNAc repeating unit group. The EIC for a glycan was set as the observed *m/z* value with a symmetrical 10 ppm expansion. Different ionization states of the same glycan that were found as a separate feature by the feature-finding software were also included in the summed EIC chromatogram.

The MS/MS experiments were performed on a Thermo Scientific Exploris 480 connected to a Thermo Scientific Ultimate 3000 UPLC system. Solvent A consisted of 0.1% FA (formic acid), solvent B of 0.1% FA in 80% acetonitrile. A 50 cm, 75µm ID, 2.4 µm reprosil column was used with a 60 min gradient of 2% B at 0-1 min, 44% B at 39 min, 55% B at 44 min, 99% B at 45-50 min, 2% B at 50-60 min. After an initial MS scan, and MS/MS scan was triggered for ions with an intensity of more than 2e5

and a charge state from 2+ to 8+. The MS/MS scan used HCD at 15% normalized collision energy (NCE). All scans were performed with the instrument in peptide application mode.

Proteowizard MSconvert (version 3.0.21328-404bcf1) was used to convert Thermo raw files to MGF format using MGF as output format, 64-bit binary encoding precision and with the following options selected: write index, zlib compression and TPP compatibility. No filters were used when converting raw files to MGF format. To search MGF files for spectra containing glycan oxonium ions an internally developed tool named Peaksuite (v1.10.1) was used with an ion delta of 20 ppm, noise filter of 0% and using a list of oxonium  $m/z$  values as mass targets (Table SX). Scans without any detected peaks were removed. Python 3.2.2 was used for data curation based on precursor  $m/z$  (10 ppm), retention time (17-24 min) and intensities of oxonium ions that originated from the glycan core ( $m/z$  441.2707, 587.3286, 644.3501, 790.4080, 806.4029, and 952.4608). The sum intensity threshold of the core oxonium ions was set to  $1e4$ . Python 3.2.2. was also used for calculating the relative intensities of oxonium ions normalized versus the sum intensities of the core oxonium ions.

### **Sugar nucleotide analysis**

Cells were grown to 60-70% confluency in a 6-wells plate, after which the medium was removed and the cells were washed twice with wash buffer (75 mM ammonium carbonate in MQ water, pH 7.4 (corrected with glacial acetic acid), at 4°C). The cells were then treated with 700  $\mu$ l of extraction buffer (40% acetonitrile, 40% methanol, 20% MQ water, at 4°C) per well for 2 minutes, after which the supernatant was transferred to a vial. This extraction step is repeated for 3 minutes, after which the two extracts were pooled and centrifuged at 18000 rcf for 3 min. The supernatant was taken and dried in the vacuum concentrator. Samples were frozen at -80°C until analysis using an ion-pair UHPLC-QqQ 1290-6490 Agilent mass spectrometer by Glycomscan BV (Oss, the Netherlands) [45].

### **Virus titration on B3GNT2/B4GALT1 knock-in MDCK and hCK cells**

Virus titers in the virus stocks in Table I were determined using end-point titration in MDCK cells and inoculated cell cultures were tested for agglutination activity using turkey red blood cells as an indicator of virus replication in the cells. For recent (2017-2019) H3N2 viruses, no binding to erythrocytes was observed and therefore virus titers were determined using a nucleoprotein (NP) staining. The NP staining was performed on the inoculated cells that were fixed with acetone for at least 20 minutes at -20°C. Primary mouse anti-NP antibody (HB65, 2 mg/ml) was diluted 1:3000 and the secondary goat anti-mouse IgG HRP antibody (A16702, 1 mg/ml, Thermo Fisher Scientific) was used at a dilution of 1:30000, after which 50  $\mu$ l per well was used for both solutions. True Blue substrate (KPL) was then added to visualize positive wells using an ImmunoSpot Analyzer (CTL Europe, Bonn, Germany). Based on the negative control values and the highest positive values per plate, the cut-off for positivity was determined. Infectious titers were calculated from five replicates using the Spearman-Kärber method [73].

**Table I. Details of IAVs used in the experiment shown in Fig. 5A-C.** These viruses include one older H3N2 virus from 2003, one H1N1 virus, and two influenza B viruses. The viruses were passaged in MDCK, MDCK-SIAT1 [22], and/or hCK cells.

Virus	Virus details	Passage history
A/Netherlands/1797/2017	Subtype H3N2, clade 3C.2a1	SIAT2hCK2
A/Netherlands/371/2019	Subtype H3N2, clade 3C.2a1	SIAT2MDCK1hCK1
A/Netherlands/10009/2019	Subtype H3N2, clade 3C.2a1b	SIAT1hCK3
A/Netherlands/314/2019	Subtype H3N2, clade 3C.2a1b	SIAT2hCK2
A/Netherlands/384/2017	Subtype H3N2, clade 3C.3a	SIAT2hCK4
A/Netherlands/10006/2019	Subtype H3N2, clade 3C.3a	SIAT1hCK3
A/Netherlands/3466/2017	Subtype H3N2, clade 3C.2a2	SIAT2hCK2
A/Netherlands/10616/2019	Subtype H3N2, clade 3C.2a2	SIAT1hCK2
A/Netherlands/213/2003	Subtype H3N2	MDCK2
A/Netherlands/121/2020	Subtype H1N1	hCK2
B/Croatia/7789/2019	Lineage Victoria	MDCKxMDCK3
B/Netherlands/461/2019	Lineage Yamagata	SIAT5hCK1

### Inoculation of hCK and hCK-B3GNT2 cells with influenza viruses

To evaluate whether IAVs that could not be isolated previously in hCK cells would replicate in hCK-B3GNT2 cells, hCK and hCK-B3GNT2 cells were seeded at a density of 20,000 cells per well in 96-wells plates at 24 hours before inoculation. The original patient materials (100  $\mu$ l), which are influenza virus positive swabs collected as part of the Dutch influenza virus surveillance program, (details of viruses in Table II) were diluted in 700  $\mu$ l infection medium (EMEM (Cambrex, Heerhugowaard, The Netherlands) supplemented with 100 U/ml penicillin, 100  $\mu$ g/ml streptomycin, 2mM glutamine, 1.5mg/ml sodium bicarbonate (Cambrex), 10mM Hepes (Cambrex), nonessential amino acids (MP Biomedicals) and 20  $\mu$ g/ml trypsin (Cambrex)), after which a two-fold dilution series was made. After three days, the presence or absence of cytopathic effects in each well was scored and the mean (n=3) of the number of infected wells was calculated.

**Table II. Details of H3N2 IAVs used in the experiment shown in Fig. 5D.** The exact virus is indicated as well as the clade and HA mutations if applicable. Previous attempts of isolating these viruses in hCK cells were either unsuccessful (-) or successful (+).

Number	Virus	Clade (+ HA mutations)	Cultured previously
1	A/Netherlands/3425/2017	3C.2a	-
2	A/Netherlands/2362/2018	3C.2a1b	-
3	A/Netherlands/2380/2018	3C.2a1b	-
4	A/Netherlands/010/2019	3C.2a1b	-
5	A/Netherlands/173/2019	3C.2a1b	-
6	A/Netherlands/1268/2019	3C.2a4	-
7	A/Netherlands/1735/2019	3C.2a1b + T135K	-
8	A/Netherlands/1747/2019	3C.2a1b + T135K	-
9	A/Netherlands/1439/2019	3C.2a1b	+
10	A/Netherlands/1734/2019	3C.2a1b + T135K	+
11	A/Netherlands/054/2020	3C.2a1b + T131K	+
12	A/Netherlands/008/2021	3C.2a1b.2a2	+

## Acknowledgments

We thank professor Yoshihiro Kawaoka for providing the hCK cells. We would like to thank Frederik Broszeit for producing the glycans that are used in the glycan microarray in Fig. S2. Balthasar Heesters is thanked for his advice on the flow cytometry experiments and analysis. We thank Gerlof Bosman for the production of PNGaseF. Monique van Scherpenzeel is thanked for the sugar nucleotide analysis. We also thank Roosmarijn van der Woude and Alinda Berends for their technical assistance.

## Funding

This work was supported by European Commission [ERC Starting Grant 802780 to R.P.dV.]; and the and the Royal Dutch Academy of Sciences [Beijerinck Premium to R.P.dV.]; and the Dutch Research Council [NWO Gravitation 2013 BOO, Institute for Chemical Immunology (ICI; 024.002.009) to J.S.]; and NIAID/NIH Centers of Excellence [Influenza Research and Response contract 75N93021C00014 to R.A.M.F and S.H.].

## Data Availability Statement

The mass spectrometry proteomic data (for the quantification of quantification of B3GNT2, B4GALT1, and ST6GAL1 proteins) have been deposited to the ProteomeXchange Consortium via the PRIDE [65] partner repository with the dataset identifier PXD037175.

The released *N*-glycan raw data (for the glycomic analyses of the cell lines) have been deposited to the GlycoPOST repository [74] under identifier GPST000345.

The supplementary data to this chapter is available via <https://doi.org/10.1093/glycob/cwad060>.

## Author contributions

Production of lentiviruses, C.M.S, B.Q.; cell culturing, C.M.S.; genetic modification of cells, C.M.S.; RT-qPCR, C.M.S.; peptide mass spectrometry, M.J.A.D., J.S.; production of hemagglutinins, C.M.S.; flow cytometry, C.M.S.; sample preparation for mass spectrometry, C.M.S, I.R.S.; mass spectrometry measurements, I.R.S, J.C.L.M., K.R.R.; data analysis of mass spectrometry experiments, C.M.S., I.R.S., J.C.L.M., K.R.R.; sample preparation for sugar nucleotide analysis, C.M.S.; sugar nucleotide analysis, external (Monique van Scherpenzeel); end-point titrations and infection studies, T.B., P.L.; data visualization, C.M.S.; conceptualization, C.M.S., R.P.dV.; supervision, R.P.dV., R.A.M.F., K.R.R., J.S., S.H., G.J.B.; funding acquisition, R.P.dV., J.S., R.A.M.F., S.H., G.J.B.; writing—original draft preparation, C.M.S.; project administration, R.P.dV.; writing—review and editing, all authors.

## References

1. Jester, B.J., T.M. Uyeki, and D.B. Jernigan, *Fifty years of influenza A(H3N2) following the pandemic of 1968*. *Am J Public Health*, 2020. 110(5): p. 669-676.
2. Stevens, J., et al., *Glycan microarray analysis of the hemagglutinins from modern and pandemic influenza viruses reveals different receptor specificities*. *J Mol Biol*, 2006. 355(5): p. 1143-55.
3. Rogers, G.N. and J.C. Paulson, *Receptor determinants of human and animal influenza virus isolates: differences in receptor specificity of the H3 hemagglutinin based on species of origin*. *Virology*, 1983. 127(2): p. 361-73.
4. Yang, H., et al., *Structure and receptor binding preferences of recombinant human A(H3N2) virus hemagglutinins*. *Virology*, 2015. 477: p. 18-31.
5. Canales, A., et al., *Revealing the specificity of human H1 influenza A viruses to complex N-glycans*. *JACS Au*, 2023.
6. Gulati, S., et al., *Human H3N2 influenza viruses isolated from 1968 to 2012 show varying preference for receptor substructures with no apparent consequences for disease or spread*. *PLoS One*, 2013. 8(6): p. e66325.
7. Chandrasekaran, A., et al., *Glycan topology determines human adaptation of avian H5N1 virus hemagglutinin*. *Nat Biotechnol*, 2008. 26(1): p. 107-13.
8. Srivilajaroen, N., et al., *N-glycan structures of human alveoli provide insight into influenza A virus infection and pathogenesis*. *FEBS J*, 2018. 285(9): p. 1611-1634.
9. Peng, W., et al., *Recent H3N2 viruses have evolved specificity for extended, branched human-type receptors, conferring potential for increased avidity*. *Cell Host Microbe*, 2017. 21(1): p. 23-34.
10. Byrd-Leotis, L., et al., *Antigenic pressure on H3N2 influenza virus drift strains imposes constraints on binding to sialylated receptors but not phosphorylated glycans*. *J Virol*, 2019. 93(22).
11. Stevens, J., et al., *Receptor specificity of influenza A H3N2 viruses isolated in mammalian cells and embryonated chicken eggs*. *J Virol*, 2010. 84(16): p. 8287-99.
12. Nycholat, C.M., et al., *Recognition of sialylated poly-N-acetylglucosamine chains on N- and O-linked glycans by human and avian influenza A virus hemagglutinins*. *Angew Chem Int Ed Engl*, 2012. 51(20): p. 4860-3.
13. Broszeit, F., et al., *Glycan remodeled erythrocytes facilitate antigenic characterization of recent A/H3N2 influenza viruses*. *Nat Commun*, 2021. 12(1): p. 5449.
14. Chambers, B.S., et al., *Recent H3N2 influenza virus clinical isolates rapidly acquire hemagglutinin or neuraminidase mutations when propagated for antigenic analyses*. *J Virol*, 2014. 88(18): p. 10986-9.
15. Asaoka, N., et al., *Low growth ability of recent influenza clinical isolates in MDCK cells is due to their low receptor binding affinities*. *Microbes Infect*, 2006. 8(2): p. 511-9.
16. Takada, K., et al., *A humanized MDCK cell line for the efficient isolation and propagation of human influenza viruses*. *Nat Microbiol*, 2019. 4(8): p. 1268-1273.
17. Oh, D.Y., et al., *MDCK-SIAT1 cells show improved isolation rates for recent human influenza viruses compared to conventional MDCK cells*. *J Clin Microbiol*, 2008. 46(7): p. 2189-94.
18. Lee, H.K., et al., *Comparison of mutation patterns in full-genome A/H3N2 influenza sequences obtained directly from clinical samples and the same samples after a single MDCK passage*. *PLoS One*, 2013. 8(11): p. e79252.
19. Medeiros, R., et al., *Hemagglutinin residues of recent human A(H3N2) influenza viruses that contribute to the inability to agglutinate chicken erythrocytes*. *Virology*, 2001. 289(1): p. 74-85.
20. Allen, J.D. and T.M. Ross, *H3N2 influenza viruses in humans: viral mechanisms, evolution, and evaluation*. *Hum Vaccin Immunother*, 2018. 14(8): p. 1840-1847.
21. Peck, H., et al., *Enhanced isolation of influenza viruses in qualified cells improves the probability of well-matched vaccines*. *NPJ Vaccines*, 2021. 6(1): p. 149.



22. Matrosovich, M., et al., Overexpression of the alpha-2,6-sialyltransferase in MDCK cells increases influenza virus sensitivity to neuraminidase inhibitors. *J Virol*, 2003. 77(15): p. 8418-25.
23. Hatakeyama, S., et al., Enhanced expression of an alpha2,6-linked sialic acid on MDCK cells improves isolation of human influenza viruses and evaluation of their sensitivity to a neuraminidase inhibitor. *J Clin Microbiol*, 2005. 43(8): p. 4139-46.
24. Lin, Y., et al., The characteristics and antigenic properties of recently emerged subclade 3C.3a and 3C.2a human influenza A(H3N2) viruses passaged in MDCK cells. *Influenza Other Respir Viruses*, 2017. 11(3): p. 263-274.
25. Byrd-Leotis, L., et al., Sialylated and sulfated N-glycans in MDCK and engineered MDCK cells for influenza virus studies. *Sci Rep*, 2022. 12(1): p. 12757.
26. Li, T., et al., An automated platform for the enzyme-mediated assembly of complex oligosaccharides. *Nat Chem*, 2019. 11(3): p. 229-236.
27. Liu, L., et al., Streamlining the chemoenzymatic synthesis of complex N-glycans by a stop and go strategy. *Nat Chem*, 2019. 11(2): p. 161-169.
28. Lin, Y.P., et al., Evolution of the receptor binding properties of the influenza A(H3N2) hemagglutinin. *Proc Natl Acad Sci U S A*, 2012. 109(52): p. 21474-9.
29. Liu, M., et al., Human-type sialic acid receptors contribute to avian influenza A virus binding and entry by hetero-multivalent interactions. *Nat Commun*, 2022. 13(1): p. 4054.
30. Rimmelzwaan, G.F., et al., Attachment of infectious influenza A viruses of various subtypes to live mammalian and avian cells as measured by flow cytometry. *Virus Res*, 2007. 129(2): p. 175-81.
31. Kumari, K., et al., Receptor binding specificity of recent human H3N2 influenza viruses. *Virology*, 2007. 4: p. 42.
32. Wang, Q., et al., Metabolic engineering challenges of extending N-glycan pathways in Chinese hamster ovary cells. *Metab Eng*, 2020. 61: p. 301-314.
33. Wang, X., et al., The EF-1alpha promoter maintains high-level transgene expression from episomal vectors in transfected CHO-K1 cells. *J Cell Mol Med*, 2017. 21(11): p. 3044-3054.
34. Shibuya, N., et al., The elderberry (*Sambucus nigra* L.) bark lectin recognizes the Neu5Ac(alpha 2-6)Gal/GalNAc sequence. *J Biol Chem*, 1987. 262(4): p. 1596-601.
35. Sweeney, J.G., et al., Loss of GCNT2/II-branched glycans enhances melanoma growth and survival. *Nat Commun*, 2018. 9(1): p. 3368.
36. Broszeit, F., et al., N-glycolylneuraminic acid as a receptor for influenza A viruses. *Cell Rep*, 2019. 27(11): p. 3284-3294 e6.
37. Bouwman, K.M., et al., Guinea fowl coronavirus diversity has phenotypic consequences for glycan and tissue binding. *J Virol*, 2019. 93(10).
38. Spruit, C.M., et al., N-glycolylneuraminic acid binding of avian and equine H7 influenza A viruses. *J Virol*, 2022. 96(5): p. e0212021.
39. Spruit, C.M., et al., N-glycolylneuraminic acid in animal models for human influenza A virus. *Viruses*, 2021. 13(5).
40. Nemanichvili, N., et al., Fluorescent trimeric hemagglutinins reveal multivalent receptor binding properties. *J Mol Biol*, 2019. 431(4): p. 842-856.
41. Chu, V.C. and G.R. Whittaker, Influenza virus entry and infection require host cell N-linked glycoprotein. *Proc Natl Acad Sci U S A*, 2004. 101(52): p. 18153-8.
42. Hossler, P., B.C. Mulukutla, and W.S. Hu, Systems analysis of N-glycan processing in mammalian cells. *PLoS One*, 2007. 2(8): p. e713.
43. Ren, W.W., et al., Glycoengineering of HEK293 cells to produce high-mannose-type N-glycan structures. *J Biochem*, 2019. 166(3): p. 245-258.
44. Mastrangeli, R., et al., The Formidable Challenge of Controlling High Mannose-Type N-Glycans in Therapeutic mAbs. *Trends Biotechnol*, 2020. 38(10): p. 1154-1168.
45. van Scherpenzeel, M., et al., Dynamic tracing of sugar metabolism reveals the mechanisms of action of synthetic sugar analogs. *Glycobiology*, 2022. 32(3): p. 239-250.
46. Karakus, U., M.O. Pohl, and S. Stertz, Breaking the convention: sialoglycan variants, coreceptors, and alternative receptors for influenza A virus entry. *J Virol*, 2020. 94(4).
47. Fujioka, Y., et al., A sialylated voltage-dependent Ca(2+) channel binds hemagglutinin and mediates influenza A virus entry into mammalian cells. *Cell Host Microbe*, 2018. 23(6): p. 809-818 e5.
48. Ho, J.W., et al., H5-type influenza virus hemagglutinin is functionally recognized by the natural killer-activating receptor NKp44. *J Virol*, 2008. 82(4): p. 2028-32.
49. Arnon, T.I., et al., The mechanisms controlling the recognition of tumor- and virus-infected cells by NKp46. *Blood*, 2004. 103(2): p. 664-72.
50. Achdout, H., et al., Killing of avian and swine influenza virus by natural killer cells. *J Virol*, 2010. 84(8): p. 3993-4001.
51. Mandelboim, O., et al., Recognition of haemagglutinins on virus-infected cells by NKp46 activates lysis by human NK cells. *Nature*, 2001. 409(6823): p. 1055-60.
52. Eierhoff, T., et al., The epidermal growth factor receptor (EGFR) promotes uptake of influenza A viruses (IAV) into host cells. *PLoS Pathog*, 2010. 6(9): p. e1001099.
53. Chan, C.M., et al., Hemagglutinin of influenza A virus binds specifically to cell surface nucleolin and plays a role in virus internalization. *Virology*, 2016. 494: p. 78-88.
54. Sieben, C., et al., Influenza A viruses use multivalent sialic acid clusters for cell binding and receptor activation. *PLoS Pathog*, 2020. 16(7): p. e1008656.
55. Byrd-Leotis, L., et al., Influenza binds phosphorylated glycans from human lung. *Sci Adv*, 2019. 5(2): p. eaav2554.
56. Gambaryan, A., et al., Receptor specificity of influenza viruses from birds and mammals: new data on involvement of the inner fragments of the carbohydrate chain. *Virology*, 2005. 334(2): p. 276-83.

57. Jia, N., et al., *Glycomic characterization of respiratory tract tissues of ferrets: implications for its use in influenza virus infection studies*. J Biol Chem, 2014. 289(41): p. 28489-504.
58. Walther, T., et al., *Glycomic analysis of human respiratory tract tissues and correlation with influenza virus infection*. PLoS Pathog, 2013. 9(3): p. e1003223.
59. Jia, N., et al., *The human lung glycome reveals novel glycan ligands for influenza A virus*. Sci Rep, 2020. 10(1): p. 5320.
60. Watters, K.E., et al., *Systematic discovery of natural CRISPR-Cas12a inhibitors*. Science, 2018. 362(6411): p. 236-239.
61. Cole, N.B., et al., *Diffusional mobility of golgi proteins in membranes of living cells*. Science, 1996. 273(5276): p. 797-801.
62. Canals, I., et al., *Rapid and efficient induction of functional astrocytes from human pluripotent stem cells*. Nat Methods, 2018. 15(9): p. 693-696.
63. Qiu, B., R.J. de Vries, and M. Caiazzo, *Direct cell reprogramming of mouse fibroblasts into functional astrocytes using lentiviral overexpression of the transcription factors NFIA, NFIB, and SOX9*. Methods Mol Biol, 2021. 2352: p. 31-43.
64. Dull, T., et al., *A third-generation lentivirus vector with a conditional packaging system*. J Virol, 1998. 72(11): p. 8463-71.
65. Perez-Riverol, Y., et al., *The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences*. Nucleic Acids Res, 2022. 50(D1): p. D543-D552.
66. de Vries, R.P., et al., *The influenza A virus hemagglutinin glycosylation state affects receptor-binding specificity*. Virology, 2010. 403(1): p. 17-25.
67. Zeng, Q., et al., *Structure of coronavirus hemagglutinin-esterase offers insight into corona and influenza virus evolution*. Proc Natl Acad Sci U S A, 2008. 105(26): p. 9065-9.
68. Shaner, N.C., et al., *Improving the photostability of bright monomeric orange and red fluorescent proteins*. Nat Methods, 2008. 5(6): p. 545-51.
69. Packer, N.H., et al., *A general approach to desalting oligosaccharides released from glycoproteins*. Glycoconj J, 1998. 15(8): p. 737-47.
70. Keser, T., et al., *Comparison of 2-Aminobenzamide, Procainamide and RapiFluor-MS as Derivatizing Agents for High-Throughput HILIC-UPLC-FLR-MS N-glycan Analysis*. Front Chem, 2018. 6: p. 324.
71. Ruhaak, L.R., et al., *2-picoline-borane: a non-toxic reducing agent for oligosaccharide labeling by reductive amination*. Proteomics, 2010. 10(12): p. 2330-6.
72. Cooper, C.A., E. Gasteiger, and N.H. Packer, *GlycoMod - a software tool for determining glycosylation compositions from mass spectrometric data*. Proteomics, 2001. 1(2): p. 340-9.
73. Kärber, G., *Beitrag zur kollektiven Behandlung pharmakologischer Reihenversuche*. Naunyn-Schmiedeberg's Archiv für experimentelle Pathologie und Pharmakologie, 1931. 162(4): p. 480-483.
74. Watanabe, Y., et al., *GlycoPOST realizes FAIR principles for glycomics mass spectrometry data*. Nucleic Acids Res, 2021. 49(D1): p. D1523-D1528.

# Chapter 3

## ***N*-glycolylneuraminic acid in animal models for human influenza A virus**

Cindy M. Spruit, Nikoloz Nemanichvili, Masatoshi Okamoto, Hiromu Takematsu, Geert-Jan Boons, and Robert P. de Vries

Published: *Viruses*, May 2021, DOI: <https://doi.org/10.3390/v13050815>

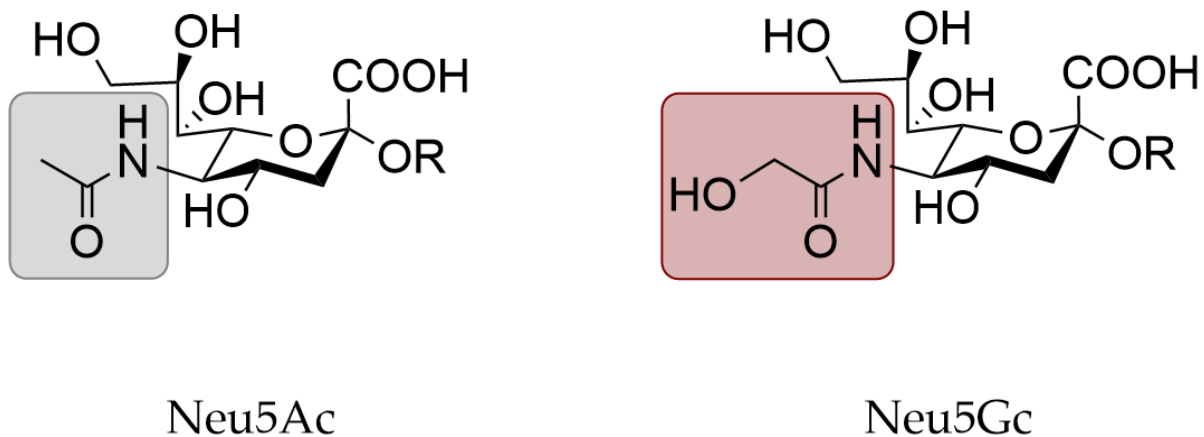
### **Abstract**

The first step in influenza virus infection is the binding of hemagglutinin to sialic acid-containing glycans present on the cell surface. Over 50 different sialic acid modifications are known, of which *N*-acetylneuraminic acid (Neu5Ac) and *N*-glycolylneuraminic acid (Neu5Gc) are the two main species. Animal models with  $\alpha$ 2,6 linked Neu5Ac in the upper respiratory tract, similar to humans, are preferred to enable and mimic infection with unadapted human influenza A viruses. Animal models that are currently most often used to study human influenza are mice and ferrets. Additionally, guinea pigs, cotton rats, Syrian hamsters, tree shrews, domestic swine, and non-human primates (macaques and marmosets) are discussed. The presence of Neu5Gc and the distribution of sialic acid linkages in the most commonly used models is summarized and experimentally determined. We also evaluated the role of Neu5Gc in infection using Neu5Gc binding viruses and cytidine monophosphate-*N*-acetylneuraminic acid hydroxylase (CMAH)  $-/-$  knockout mice, which lack Neu5Gc, and concluded that Neu5Gc is unlikely to be a decoy receptor. This article provides a base for choosing an appropriate animal model. Although mice are one of the most favored models, they are hardly naturally susceptible to infection with human influenza viruses, possibly because they express mainly  $\alpha$ 2,3 linked sialic acids with both Neu5Ac and Neu5Gc modifications. We suggest using ferrets, which resemble humans closely in the sialic acid content, both in the linkages and the lack of Neu5Gc, lung organization, susceptibility, and disease pathogenesis.

## Introduction

Infection of humans by influenza A viruses starts at the epithelium cells in the upper respiratory tract, where the hemagglutinin (HA) on the outside of a virus particle binds to glycans with a terminal sialic acid. The terminal sialic acids can be linked through an  $\alpha$ 2,3 or  $\alpha$ 2,6 bond to the penultimate galactose. In humans, mainly  $\alpha$ 2,6 linked sialic acids are present in the upper respiratory tract. The expression of  $\alpha$ 2,3 and  $\alpha$ 2,6 linked sialic acids varies between species and tissues [1].

Another variable in the influenza receptor is the type of sialic acid, of which *N*-acetylneuraminic acid (Neu5Ac) and *N*-glycolylneuraminic acid (Neu5Gc) are the main species (Figure 1). The majority of influenza A viruses use a glycan with a terminal Neu5Ac as their receptor, although some strains use Neu5Gc instead [2, 3]. Importantly, only Neu5Ac is present in humans [4-6]. Human influenza A viruses specifically bind  $\alpha$ 2,6 linked sialic acids.



**Fig 1. Structure of *N*-acetylneuraminic acid (Neu5Ac) and *N*-glycolylneuraminic acid (Neu5Gc).**

Neu5Gc can be produced in animals that express an active form of the enzyme cytidine monophosphate-*N*-acetylneuraminic acid hydroxylase (CMAH), which facilitates the hydroxylation of Neu5Ac to turn it into Neu5Gc. However, the gene encoding CMAH, mainly expressed in mammalian species, has been lost partially or completely in several events during evolution [4]. Possibly, the negative selection of Neu5Gc was induced by lethal pathogens binding to Neu5Gc. Therefore, the loss of Neu5Gc protected individuals from infection with these pathogens [7]. The presence of an intact CMAH gene does not automatically lead to high expression of Neu5Gc in all tissues [8]. The expression of Neu5Gc in species used as animal models for human influenza has received little attention so far and no clear overview of this expression is available.

Proper animal models are essential for fundamental and applied research on human influenza viruses, vaccines, and antivirals. Often considered factors for choosing an animal model are the experimental costs, disease pathogenesis, and susceptibility. Currently, the sialic acid linkage and especially the Neu5Gc content are often overlooked. Human influenza viruses mainly bind  $\alpha$ 2,6 linked Neu5Ac, while many animal models express Neu5Gc. The lack of correct sialic acid

receptors in animal models could skew the results of a study since adaptation of a virus may be required before a successful infection is possible. The animal models mostly used to study human influenza are ferrets (*Mustela putorius furo*) and mice (*Mus musculus*). While ferrets mainly express the human receptor ( $\alpha 2,6$  linked Neu5Ac), mice also express Neu5Gc and the sialic acid linkages in the respiratory tract differ from humans [1, 6, 9-19]. Other animal models that are discussed in this article are cotton rats (*Sigmodon* species), Syrian hamsters (*Mesocricetus auratus*), guinea pigs (*Cavia porcellus*), domestic swine (*Sus scrofa domesticus*), macaques (*Macaca*), and marmosets (*Callitrichidae*). Of these animals, domestic swine are naturally infected by human influenza viruses [1].

In this article, we summarize the current knowledge on Neu5Gc expression in animal models for human influenza and supplement this with protein histochemistry stains on lung tissues. The role of Neu5Gc in influenza virus infection is still unclear. Furthermore, studies on Neu5Gc specific influenza viruses or Neu5Ac specific viruses in animal models that are rich in Neu5Gc are scarce. Therefore, we also studied the infection of H5N1 viruses that bind either Neu5Ac or Neu5Gc in CMAH  $-/-$  knockout mice [20]. Interestingly, although the Neu5Gc specific H5N1 virus retained Neu5Ac binding, no significant difference in pathogenesis was observed between wild-type (WT) and CMAH  $-/-$  mice. The Neu5Gc binding H5N1 virus was less infectious compared to the WT virus but this was not affected by the presence of Neu5Gc in mice. This article provides a base for choosing suitable animal models to study human influenza.

## Results

### Animal models for human influenza

The most used animal models for human influenza are mice [21-25], ferrets [21-23, 25-29], and guinea pigs [22, 25]. Other small animal models that are used are cotton rats [25, 30-32] and Syrian hamsters [23, 25], and recently, tree shrews have been suggested [33, 34]. Commonly used larger animal models are domestic swine [35] and non-human primates (mainly macaques and marmosets) [25, 32, 36-39].

When comparing experimental costs, housing requirements, space usage, and handling conditions for small animal models, mice and cotton rats are less demanding than guinea pigs, ferrets, tree shrews, and Syrian hamsters [21, 23, 25, 40-43]. Swine and non-human primates require more space and more complicated husbandry and handling conditions, associated with higher experimental costs compared to small animal models [25, 40].

For the analysis of infection and immunological reactions, high availability of immunological reagents is beneficial, such as available for mice, cotton rats, swine, and non-human primates [21, 25, 44]. For ferrets, guinea pigs, and Syrian hamsters [25, 42, 45, 46], only a few immunological reagents are available, which complicates the analysis of experiments.

### Physical characteristics of animal models

Comparable disease pathogenesis (sneezing, nasal discharge, lethargy, fever, weight loss, and viral shedding) in humans and animal models simplifies the analysis of the influenza virus infection. From all animal models, ferrets best represent the symptoms of human influenza virus infection, whereas mice, non-human primates, swine, cotton rats, and tree shrews moderately reflect the symptoms. On the other hand, guinea pigs and Syrian hamsters only reflect human symptoms to a lower extent [25, 34, 40, 47].

The similarity in disease pathogenesis does not seem to be correlated to the resemblance between human and animal model lung structures. In the human lung, the major vessels in the lung and the bronchi are linked by extensive interlobular and intralobular connective tissue. Human lungs are most closely resembled by lungs of non-human primates in terms of structure, physiology, and mucosal immune mechanisms [48], but many (ethical) issues are associated with experiments with non-human primates. Swine and ferret lungs very closely resemble human lungs as well [27, 44, 47, 49]. Furthermore, similarities in the lung organization are found in most mammals, among which guinea pigs, cotton rats, tree shrews, and Syrian hamsters [25, 50-52]. However, mice lungs have different branching, organization of bronchi, and distribution of connective tissue [44, 53].

To study the clearance and development of influenza virus infection, it is useful to have similar immune responses to influenza infection. The immune system of ferrets has not been studied in detail [22, 45]. The immune systems of guinea pigs, Syrian hamsters, tree shrews, and swine, compared to murine species (mice and rats), are more similar to the human immune system [35, 41, 42, 44, 54-56]. Non-human primates have high genetic homology to humans and more than 90% amino acid similarity is found in the cytokines of humans and non-human primates [38, 57, 58], but differences in the sialic acid-recognizing immunoglobulin-like lectins are observed [59]. Interestingly, Neu5Gc regulates and influences several immunological pathways. Upon activation of T cells, upregulation of Neu5Ac and  $\alpha$ 2,3 linked sialic acids and downregulation of Neu5Gc and  $\alpha$ 2,6 linked sialic acids was observed [60, 61]. Furthermore, T cells were more reactive to stimulation in the absence than in the presence of Neu5Gc [61, 62]. Additionally, Neu5Gc suppresses the B cell responses and proliferation [20]. Therefore, the presence of Neu5Gc may be important in the comparison of the immune responses of different animal models. Apart from the influence of sialic acids on the immune system, humans can already have pre-existing immunity to a broad spectrum of influenza strains. This pre-existing immunity is often lacking in animal models, although pre-immune mice and ferret models are available [63-67].

The susceptibility of animal models to influenza viruses that are not adapted is an important factor for choosing an animal model, since an adapted virus may have different properties than the original virus. Swine are natural hosts of influenza viruses and are infected by the same subtypes of influenza as humans [1]. Furthermore, some non-human primates in nature are also infected with human influenza A viruses [37, 68]. Additionally, ferrets, guinea pigs, cotton rats, Syrian hamsters, and tree shrews can be infected with human influenza viruses without

adaptation [21, 23, 25, 30, 34, 40, 45, 46]. When considering the two most often used small animal models for human influenza, ferrets are susceptible to influenza without adaptation, while the susceptibility of mice depends on both the influenza subtype and mice strains that are used. Efforts have been made to provide different strains of mice with a broad genetic background [69], of which some strains were shown to be more susceptible to H3N2 infection [70]. In general, DBA/2 mice seem to be more susceptible than C57BL/6 mice to human influenza infection [24, 71-75]. Furthermore, genetic modification of mice in for example the IFITM3 [76], Mx1 [77], or Casp1 [78] genes can increase the susceptibility. Depending on the strain of influenza used in the research, it is important to choose an appropriate and susceptible murine strain that does not require adaptation of the virus.

### **Expression of *N*-glycolylneuraminic acid in animal models**

The susceptibility of animal models is for the most part caused by the receptor specificity of the hemagglutinin and the sialic acids present at the site of infection. To mimic influenza virus infection in humans closely, similar sialic acids and sialic acid linkages in the respiratory tract of animal models are required. In humans, mainly  $\alpha$ 2,6 linked Neu5Ac is present in the upper respiratory tract, which is bound by human influenza A viruses [1]. Importantly, only Neu5Ac is present in humans, since the CMAH gene responsible for producing Neu5Gc (Figure 1) is not intact [4-6].

The linkage of sialic acids to the penultimate galactose residues in the respiratory tract of animal models has already been extensively reviewed by others and is summarized in Table 1. In short, tree shrews and ferrets have a human-like distribution of sialic acid linkages, with  $\alpha$ 2,6 linked sialic acids in the upper respiratory tract and  $\alpha$ 2,3 linked sialic acids in the lower respiratory tract. Swine mainly express  $\alpha$ 2,6 linked sialic acids and guinea pigs, rats, and Syrian hamsters express both  $\alpha$ 2,3 and  $\alpha$ 2,6 linked sialic acids. Non-human primates (macaque and marmoset) and mice mainly express  $\alpha$ 2,3 linked sialic acids. So far, no overview is available describing the presence of Neu5Gc in animal models for human influenza. Therefore, we summarized the presence of both intact CMAH genes and the expression of Neu5Gc (Table 1) in animal models for human influenza, since the presence of CMAH does not automatically lead to high expression of Neu5Gc [8]. For completeness, we added dogs and horses to the table, since these species highlight the differences in sialic acid expression between species, although they are hardly used as models for research on human influenza A viruses. Neu5Gc is present in all animal models except the ferret and marmoset. Information on the percentage of Neu5Gc in the respiratory tract with respect to the total sialic acid content was, to our knowledge, only available for swine and varied from 9-53%. In dogs, the presence of Neu5Gc depends on the breed and the consensus is that Neu5Gc is present in some Japanese breeds, while Neu5Gc lacks in all other breeds [79]. At the moment, it is unclear whether an intact CMAH gene is present in all dog breeds. For the other investigated species, the presence of an intact CMAH gene correlates with the expression of Neu5Gc.

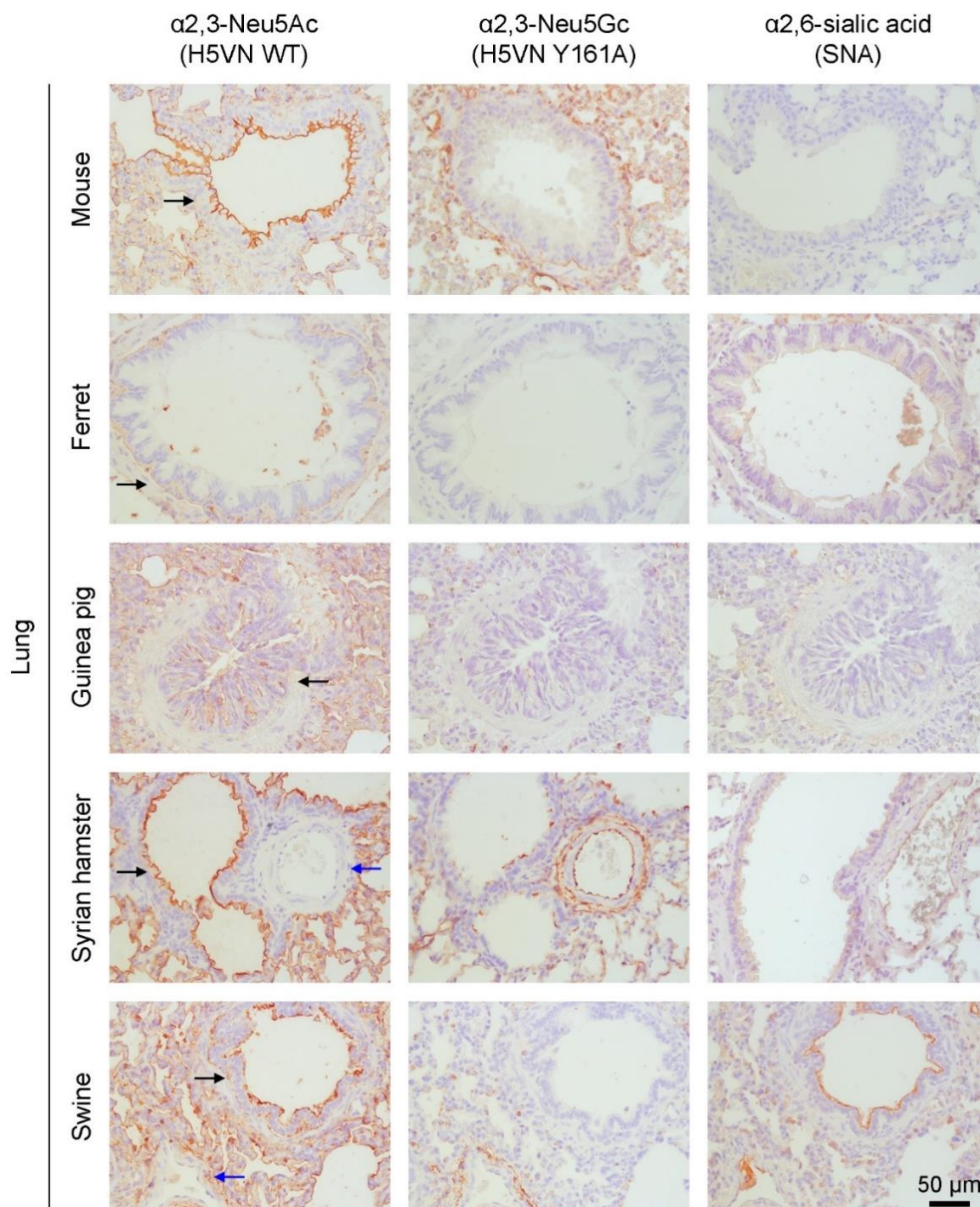


**Table 1. Overview of sialic acid linkages in the respiratory tract, the presence of Neu5Gc, and the presence of an intact CMAH gene in different animal models. + : high levels, - : low levels, -- : not detected.**

Animal model	Sialic acid linkage in the respiratory tract		Neu5Gc detected	Intact CMAH gene [4]
	$\alpha$ 2,3	$\alpha$ 2,6		
Mouse	+ [1]	- [1, 16]	Yes [6, 9-15, 80, 81] Yes (Figure 2)	Yes
Ferret	+ [17] - (upper respiratory tract) [1] + (lung) [1]	+ [17] + (upper respiratory tract) [1] - (lung) [1]	No [17-19] No (Figure 2)	No
Guinea pig	+ (nasal tract, trachea) [82] + (lung) [82]	+ (nasal tract, trachea) [82] - (lung) [82]	Yes [9, 83] Yes (Figure 2)	Yes
Rat	+ (trachea) [84] -- (lung) [84]	+ (trachea) [84] - (lung) [84]	Yes [9, 85]	Yes
Syrian hamster	+ [23]	+ [23]	Yes (Figure 2)	Yes
Tree shrew	+ [86] - (nasal tract, trachea) [34] + (lung) [34]	+ [86] + (nasal tract, trachea) [34] - (lung) [34]	Yes [86]	Yes
Swine	- [1, 87] + (nasal tract) [88] - (trachea) [88] + (lung) [88, 89]	+ [1, 87] - (nasal tract) [88] + (trachea) [88] + (lung) [88, 89]	Yes [87, 89-93] Yes, 53% [94] Yes, 9-14% [95] Yes (Figure 2)	Yes
Marmoset	+ [36]	-- [36]	No [96]	No
Macaque	Unknown	Unknown	Yes [9, 97]	Yes
Dog	+ [98]	- [3, 98]	No [79, 99, 100] Yes [79, 80, 101, 102]	Yes
Horse	+ [1, 98, 103]	-- [3, 98, 103]	Yes [3, 8, 100]	Yes

The table is supplemented with protein histochemistry of the Neu5Gc content of the lungs of mice, ferrets, guinea pigs, Syrian hamsters (which was not investigated for Neu5Gc content before), and domestic swine (Figure 2). We used the WT and Y161A mutant HA of A/Vietnam/1203/2004 (H5N1) which have previously been used to specifically detect  $\alpha$ 2,3 linked Neu5Ac and  $\alpha$ 2,3 linked Neu5Gc, respectively [3]. To clarify the presence of  $\alpha$ 2,6 linked sialic acids, *Sambucus nigra* agglutinin (SNA) was used [3]. We found that Neu5Ac is expressed in all investigated lungs. In most lungs,  $\alpha$ 2,3 linked Neu5Ac was predominantly expressed in the bronchioles and alveoli, except for the ferret in which only alveolar staining was observed. Neu5Gc expression is detected throughout the tissues for all species except the ferret. Interestingly, Neu5Ac but not Neu5Gc is expressed in the bronchioles of the Syrian hamster and swine, while the inverse is true for endothelial cells in blood vessels. Thus, although Neu5Gc is present in species with an intact CMAH, expression is mainly seen in the alveoli, which are located deep in the lung. Moreover, Neu5Gc is present in the vascular system which is hardly infected by influenza A viruses. Mice hardly expressed  $\alpha$ 2,6 linked sialic acids, as expected. High expression of  $\alpha$ 2,6 linked sialic acids is observed in ferrets and swine, moderate expression in guinea pigs and Syrian hamsters. Taken together, the human respiratory tract is best resembled, in terms of sialic acids, by ferrets, which express mainly  $\alpha$ 2,6 linked Neu5Ac. On the other hand, mice mainly

express  $\alpha 2,3$  linked sialic acids, of which a part is Neu5Gc instead of Neu5Ac. We wondered if abrogating Neu5Gc from mice using CMAH  $-/-$  knockout mice would make mice more susceptible to influenza A virus infection as most of those viruses exclusively bind Neu5Ac.



**Fig 2. Visualization of the presence of  $\alpha 2,3$  linked Neu5Ac (WT HA of A/Vietnam/1203/2004 (H5N1)),  $\alpha 2,3$  linked Neu5Gc (Y161A mutant HA of A/Vietnam/1203/2004 (H5N1)), and  $\alpha 2,6$  linked sialic acids (SNA) using AEC staining on lung tissue of mouse (C57BL/6), ferret, guinea pig, Syrian hamster, and domestic swine.** Brown staining indicates binding of the HAs to the tissue and blue indicates the cells. Selected bronchioles (black arrows) and blood vessels (blue arrows) are indicated. Images are representative of two independent experiments.

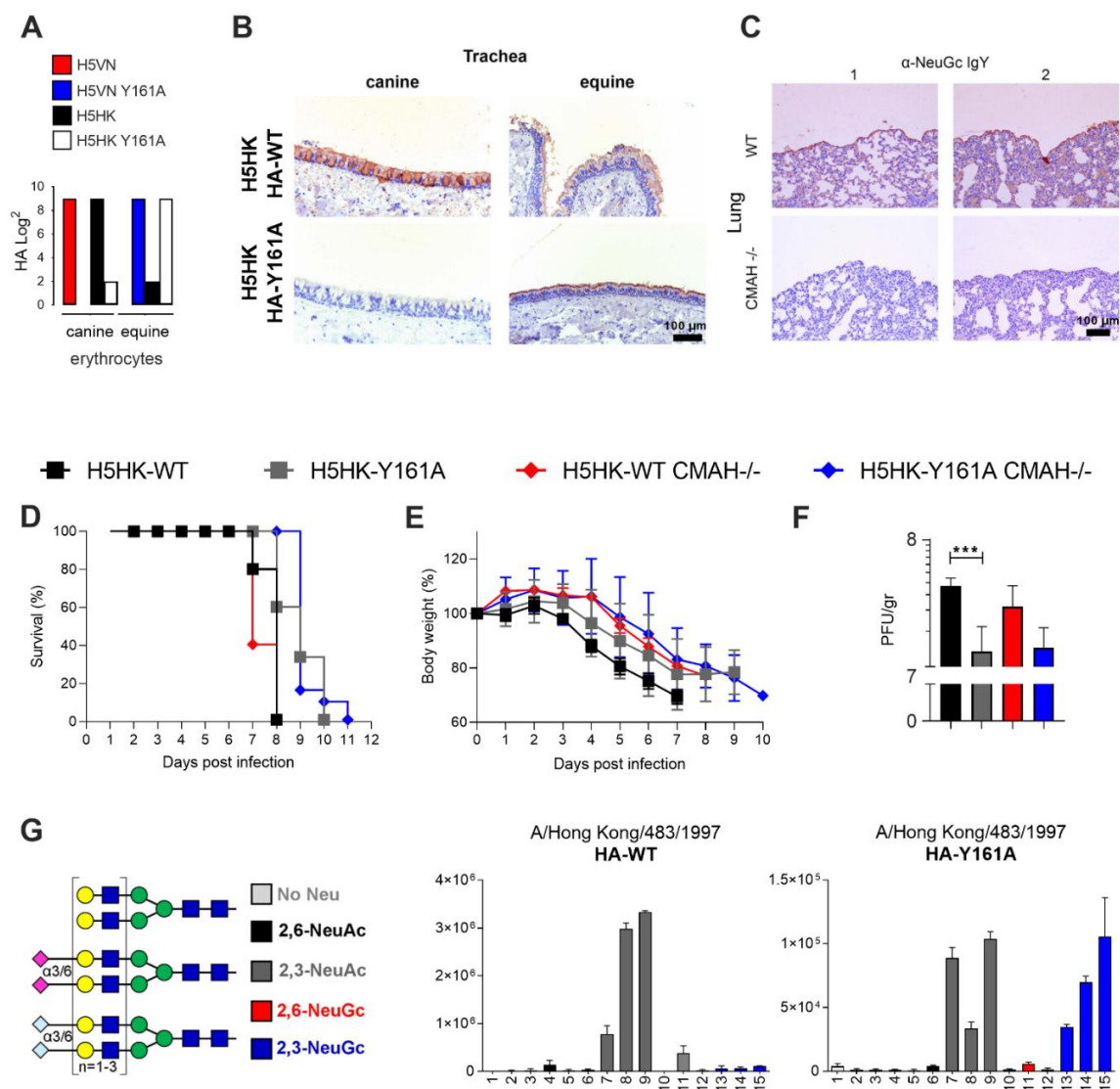
### Neu5Gc as a functional receptor

The effect of Neu5Gc on influenza virus infection is not clear yet. Although Neu5Gc was previously suggested to act as a decoy receptor for influenza viruses [87, 104], many others suggest that Neu5Gc is not a decoy receptor and Neu5Gc may even be a functional receptor. Firstly, an influenza A strain that bound Neu5Ac showed no reduced infectivity for CMAH-transfected cells with a higher Neu5Gc content [104]. Secondly, a mutant H3N8 virus that preferred binding to Neu5Ac over Neu5Gc only moderately infected ponies, which have a high Neu5Gc content, while another mutant virus that preferred binding to Neu5Gc infected all ponies in the experiment [103]. Thirdly, equine H7 viruses specifically bind Neu5Gc [2, 3] and some other influenza A viruses bind Neu5Gc as well [2, 94, 105]. Some of these Neu5Gc binding viruses infect swine, with a significant amount of Neu5Gc in their respiratory tract. There is however no indication that these viruses persisted in the domestic swine population.

To investigate the effect of Neu5Gc on influenza virus infection, the CMAH  $-/-$  knockout mice model [15, 20, 62] was used, which is not able to produce Neu5Gc and is, therefore, more similar to humans than WT mice. The CMAH  $-/-$  model is also thought to produce a stronger immune response to influenza virus infection than WT mice [62]. Previously, we showed with a glycan array and hemagglutination assays that H5N1 WT virus and HA of A/Vietnam/1203/2004 (H5VN) solely bind Neu5Ac, while the Y161A mutant only binds Neu5Gc [3]. Here, we demonstrate that the WT HA of a closely related H5N1 virus, A/Hong Kong/483/1997 (H5HK), mainly binds to canine erythrocytes (Western breed), containing only Neu5Ac. The H5HK Y161A mutant mainly binds to equine erythrocytes (Figure 3A), containing primarily Neu5Gc, as indicated in Table 1. Like H5VN, the H5HK WT HA binds both equine and canine trachea, while the Y161A mutant only binds the equine trachea (Figure 3B). Therefore, the results of the hemagglutination assay and tissue staining indicated that both H5VN Y161A and H5HK Y161A are specifically binding Neu5Gc.

To further investigate the effect of Neu5Ac and Neu5Gc binding viruses, we produced a Y161A mutant of the H5HK virus. Infection of this virus was examined in WT (C57BL/6) and CMAH  $-/-$  knockout mice. Anti-Neu5Gc stains confirmed the absence of Neu5Gc in CMAH  $-/-$  mice and the presence in WT mice (Figure 3C), and the presence of Neu5Ac was shown in Figure 2. Upon infection, it became clear that the WT and CMAH  $-/-$  knockout mice were equally well infected, as shown by the survival curves (Figure 3D), bodyweight curves (Figure 3E), and the measurement of the PFU per gram in the lungs at three days post infection (Figure 3F). Thus, abrogating Neu5Gc did not facilitate increased virus infection and replication of a Neu5Ac binding virus and therefore it is also unlikely that Neu5Gc is used as a decoy receptor. Possibly, enough Neu5Ac was present in both the WT and CMAH  $-/-$  mice to support effective infection of the Neu5Ac binding virus. The H5HK Y161A virus, binding Neu5Gc, has significantly lower pathogenicity (WT mice in Figure 3F) compared to the WT virus ( $p=0.0007$ ). Interestingly, this virus still infected the CMAH  $-/-$  knockout mice, which was not expected based on the Neu5Gc specificity. Sequencing of the inoculated and isolated viruses (at 3 dpi)





**Fig 3. H5 WT and HA-Y161A binding properties and infectivity in WT (C57BL/6) and CMAH<sup>-/-</sup> mice.** (A) A hemagglutination assay ( $n=3$ ) with canine and equine erythrocytes was performed using HAs of WT and Y161A mutants of - A/Vietnam/1203/2004 (H5VN) and A/Hong Kong/483/1997 (H5HK). (B) Tissue binding of the HA of H5HK WT and mutant Y161A on canine and equine trachea is visualized with AEC staining (representative for three independent assays), the scale bar represents 100  $\mu\text{m}$ . (C) Mice lung (WT and CMAH<sup>-/-</sup> knockout) are stained in duplicate with anti-Neu5Gc IgY and binding is visualized with AEC staining. WT (C57BL/6) and CMAH<sup>-/-</sup> knockout mice ( $n=5$ ) were infected with H5HK WT and HA-Y161A mutant virus and (D) the survival and (E) body weight curves are shown, together with the (F) PFU per gram at 3 dpi in the lungs ( $n=5$ ), \*\*\* indicates a p-value of 0,0007 in an unpaired two-tailed t-test. (G) Synthetic glycans printed on the microarray ( $n=6$ ), either without sialic acid (structures 1-3, light gray), with  $\alpha$ 2,6-linked Neu5Ac (4-6, black),  $\alpha$ 2,3-linked Neu5Ac (7-9, dark gray),  $\alpha$ 2,6-linked Neu5Gc (10-12, red) or  $\alpha$ 2,3-linked Neu5Gc (13-15, blue).

confirmed that before and after inoculation, amino acid 161 of the HA of the WT virus is a tyrosine (encoded by TAC) and this amino acid in the HA-Y161A virus is an alanine (encoded by GCC).

After this observation *in vivo*, we decided to do a final test on the binding properties of the WT and mutant H5HK HA on the glycan array. We found that the WT H5HK HA only binds Neu5Ac, as expected. However, the Y161A mutant binds both Neu5Ac and Neu5Gc on the glycan array (Figure 3G), whereas we expected to see only binding to Neu5Gc as with H5VN Y161A. This may explain that the H5HK Y161A virus still infected CMAH <sup>-/-</sup> mice through their Neu5Ac receptors.

We now know that Neu5Gc and Neu5Ac binding on tissues and erythrocytes (chicken and horse) not always fully corresponds to glycan array data, possibly because different glycans are present in the assays [106]. We did however expect larger differences between viruses with Neu5Ac and dual Neu5Ac and Neu5Gc specificity since previously it was shown that influenza viruses with dual  $\alpha$ 2,6 and  $\alpha$ 2,3 linkage specificities transmitted less efficiently among ferrets than viruses with a specificity for a single linkage type of sialic acids [107-109].

## Discussion

### Relevance of Neu5Gc in animal models for human influenza

Considering space usage, ethical considerations, and experimental costs, small animal models (mouse, ferret, guinea pig, rat, tree shrew, or Syrian hamster) are preferred over non-human primates and swine for experiments with human influenza viruses. Non-human primates are also not preferred because Old World monkeys, which are mostly used for influenza research [37, 38], among which the macaques, have an intact CMAH gene and express Neu5Gc [96] and mainly  $\alpha$ 2,3 linked sialic acids. They are thus very different from humans in the sialic acid content and susceptibility to human influenza A viruses.

Guinea pigs are commonly used small animal models that express both Neu5Ac and Neu5Gc and mainly  $\alpha$ 2,3 linked sialic acids in the respiratory tract. Guinea pigs can be infected by human influenza viruses, but human symptoms of influenza infection are only reflected to a low extent and just a few immunological reagents are available. Recently, Syrian hamsters have been extensively used to investigate SARS-CoV-2 [110, 111] which suggests that these animals closely resemble human respiratory tract properties. Both  $\alpha$ 2,3 and  $\alpha$ 2,6 linked sialic acids are present in their respiratory tract and Syrian hamsters can be infected by unadapted human influenza viruses. However, as we showed here for the first time, Neu5Gc is expressed in their lungs, although not on the epithelium of the bronchioles.

In fact, we observe quite different expression profiles of Neu5Ac versus Neu5Gc not only in the Syrian hamster but also in other influenza A virus animal models. This could be the virtue of using an H5VN Y161A mutant lectin with specificity to a subset of Neu5Gc capped *N*-glycans to which we initially screened it for [3]. On the other hand, the commonly used anti-Neu5Gc antibody is raised against a glycolipid structure [12, 112]. It would be interesting to compare several Neu5Gc specific lectins to elucidate cell- and organ-specific Neu5Gc expression profiles.

Mice are one of the most favored models because they are widely available, relatively inexpensive, and easy to keep. However, mice are hardly naturally

susceptible to infection with human influenza viruses, possibly because they express mainly  $\alpha 2,3$  linked sialic acids with both Neu5Ac and Neu5Gc modifications. Previously, mice with modified glycosylation patterns, namely ST6GAL1 knockout mice which are not able to produce  $\alpha 2,6$  linked sialic acids, were used to investigate the infectivity of human influenza viruses. The ST6GAL1 knockout model still supported replication of human influenza viruses, although these viruses were mouse-adapted and therefore probably bound to  $\alpha 2,3$  linked sialic acids [113]. Here, we did not observe a difference in infectivity in CMAH  $-/-$  knockout and WT mice upon infection by H5N1 viruses binding Neu5Ac or with a dual Neu5Ac and Neu5Gc receptor specificity. In contrast, older literature clearly shows the importance of Neu5Gc specificity in experimental animals [103].

Although the ferret model is complicated and expensive, it is still the best model to mimic influenza infection in humans. Ferrets resemble humans very closely in terms of lung structure, susceptibility, and disease pathogenesis [25, 40, 45]. Importantly, the sialic acid content is very similar to that of humans, as they lack Neu5Gc and have a high  $\alpha 2,6$  linked Neu5Ac expression in the respiratory tract. Indeed the importance of  $\alpha 2,6$  versus  $\alpha 2,3$  linked sialic linkage specificity has been reported extensively [114-118].

### **Remaining questions on the role of Neu5Gc and Neu5Gc binding**

In this article, we described the presence of Neu5Gc in the animal models that are most commonly used to study human influenza. However, a quantification of the amounts of Neu5Gc present in the nasal tract, upper respiratory tract, lower respiratory tract, and lungs is still lacking. More information on this topic would provide insights into the suitability of different animal models for human influenza research.

Since Neu5Gc is present in many animal models but is lacking in humans, it is important to investigate the role of Neu5Gc in influenza virus infection. We concluded that Neu5Gc is most likely not used as a decoy receptor. However, the exact role of Neu5Gc, and whether Neu5Gc is used as a functional receptor, remains to be elucidated. Further research comparing animal models with and without Neu5Gc and studies using Neu5Gc binding viruses would be beneficial to improve the understanding of the role of Neu5Gc in influenza virus infection.

## **Materials and Methods**

### **Virus and cells**

WT and mutant HA-Y161A influenza viruses A/Hong Kong/483/1997 (H5N1) were propagated in 10-day-old embryonated chicken eggs at 35°C for 48 h. The collected allantoic fluid was stored at -80°C. Madin-Darby canine kidney (MDCK) cells were cultured at 37°C with 5% CO<sub>2</sub> in minimum essential medium (MEM) (Nissui Pharmaceutical, Tokyo, Japan) supplemented with 10% non-immobilized fetal calf serum (FCS; SAFC Biosciences, Lenexa, KS, USA), 0.3 mg/mL L-glutamine (Wako Chemicals, Tokyo, Japan), 0.1 mg/mL streptomycin (Meiji Seika Pharma, Tokyo, Japan), 100 U/mL penicillin G (Meiji Seika Pharma, Tokyo, Japan), and 8 µg/mL gentamicin (Takata Pharmaceutical, Saitama, Japan).



## Expression and purification of HA

The pCD5 expression vector was used to clone HA-encoding cDNAs in frame with DNA sequences coding for a secretion signal sequence, a Twin-Strep (WSHPQFEKGGGSGGGWSHPQFEK); IBA, Germany), a GCN4 leucine zipper trimerization domain (RMKQIEDKIEEIESKQKKIENEIARIKK) [119], and a superfolder green fluorescent protein (GFP) [120] or mOrange2 [121]. H5 HA encoding cDNAs of A/Hong Kong/483/1997 (both WT and mutant Y161A) or A/Vietnam/1203/2004 (both WT and mutant Y161A) were cloned into the pCD5 expression vector as described previously [122, 123]. The HAs were expressed in HEK293S GnTI(-) cells and purified from cell culture supernatants as described earlier [122]. In short, transfection was performed using the pCD5 expression vectors and polyethyleneimine I. At 6 h after transfection, the transfection mixtures were replaced by 293 SFM II expression medium (Gibco) supplemented with DMSO (1.5%), 1% glutaMAX (Gibco), sodium bicarbonate (3.7 g/L), Primatone RL-UF (3.0 g/L), glucose (2.0 g/L), and valproic acid (0.4 g/L). The cell culture supernatants were collected at 5 to 6 days after transfection and HAs were purified using Strep-Tactin sepharose beads (IBA, Germany) according to the manufacturer's instructions.

## Protein histochemistry

Sections of formalin-fixed, paraffin-embedded horse (*Equus ferus caballus*), guinea pig (*Cavia porcellus*), domestic swine (*Sus scrofa domesticus*), ferret (*Mustela putorius furo*), and dog (*Canis lupus familiaris*) tissues were obtained from the Division of Pathology, Department of Biomolecular Health Sciences, Faculty of Veterinary Medicine, Utrecht University, the Netherlands. Syrian hamster (*Mesocricetus auratus*) lung tissue was a kind gift of Barry Rockx from the department of Viroscience, Erasmus Medical Center, the Netherlands. The lungs of five-week-old C57BL/6 (Japan SLC, Shizuoka, Japan) and CMAH knockout (-/-) mice (*Mus musculus*) [20] were also fixed in formalin and embedded in paraffin. Protein histochemistry was performed as previously described [124, 125]. Shortly, 4 µm tissue sections were deparaffinized and rehydrated. Antigens were retrieved by heating the slides for 10 min in 10 mM sodium citrate at pH 6.0. Subsequently, slides were treated with 1% hydrogen peroxide in MeOH for 30 min to inactivate endogenous peroxidases. The slides were blocked overnight at 4°C with PBS supplemented with 3% BSA (w/v) (for protein stains) or carbo-free blocking solution (SP-5040-125; Vector Laboratories, Burlingame, CA, USA) (for *Sambucus nigra* agglutinin (SNA) stains). The slides were then stained using 100 µl of 5 µg/ml HAs after pre-complexing the proteins on ice for 20 min with mouse anti-streptag-HRP and goat anti-mouse IgG-HRP (626520; Thermo Fisher) in PBS in respectively a 4:2:1 molar ratio. For anti-Neu5Gc stains, 1:300 diluted Anti-Neu5Gc IgY (Biolegend, San Diego, CA, USA) and 1:100 diluted Rabbit anti-chicken IgY-HRP (31401; Thermo Fisher) were used. For SNA stains, 100 µl of 5 µg/ml biotinylated SNA (B1305; Vector Laboratories, Burlingame, CA, USA) in PBS was used to stain the slides for 30 minutes. Afterward, the Vectastain ABC kit (PK-4000; Vector Laboratories, Burlingame, CA, USA) was applied according to the manufacturer's protocol. After washing with PBS, slides were counterstained using hematoxylin and the HA

binding was visualized using 3-amino-9-ethylcarbazole (AEC) (Sigma-Aldrich, Germany).

### **Hemagglutination assay**

Hemagglutination assays with pre-complexed HAs, at a starting concentration of 10 µg/mL, were performed with 0.5% canine (beagle, that lacks Neu5Gc [99]) and equine erythrocytes as described previously [122]. Shortly, precomplexation was performed with HAs, anti-streptag, and goat anti-mouse antibodies in a 4:2:1 molar ratio respectively.

### **Glycan microarray binding of HA**

The previously presented glycan microarray [3] was used. HAs were pre-complexed with mouse anti-streptag-HRP (horseradish peroxidase) and goat anti-mouse-Alexa555 antibodies in a 4:2:1 molar ratio respectively in 50 µL PBS with 0.1% Tween-20 and incubated for 15 min on ice. The mixture was incubated on the glycan array slides for 90 min in a humidified chamber after which the slides were rinsed successively with PBS supplemented with 0.1% Tween-20, PBS, and deionized water. The slides were dried by centrifugation and immediately scanned as described previously [3]. The data (six replicates) was processed by removing the highest and lowest replicate after which the mean value and standard deviation were calculated over the four remaining replicates.

### **Virus challenge in mice**

Five-week-old C57BL/6 (Japan SLC, Shizuoka, Japan) and CMAH knockout (-/-) mice [20] (n = 20 mice/each strain) were acclimatized for one week before virus challenge. Mice were inoculated intranasally with the WT or mutant HA-Y161A A/Hong Kong/483/1997 (H5N1) viruses (n = 10 mice/each strain) using 10 times the 50% mouse lethal dose (MLD<sub>50</sub>) in 30 µL per mouse under anesthesia into C57BL/6 and CMAH knockout mice. The MLD<sub>50</sub> was determined in WT C57BL/6 mice with WT A/Hong Kong/483/1997 (H5N1) virus. The infectivity titer (10 MLD<sub>50</sub> = 10<sup>3.0</sup> 50% egg infective dose) of the inoculum was adjusted with PBS. For 14 days following inoculation, mice (n = 5/each strain) were observed daily for body weight, clinical signs, and survival. Viruses were fully sequenced and evaluated for changes in amino acid 161 of the hemagglutinin before inoculation and at 3 days post infection (dpi).

### **Preparation of lung homogenates**

At 3 dpi, mouse lungs were collected after euthanasia (n = 5/each strain). Lungs were homogenized with 2 mL of MEM supplemented with 10,000 U/mL penicillin G, 10 mg/mL streptomycin, 0.3 mg/mL gentamicin, 250 U/mL Nystatin (Sigma-Aldrich, St. Louis, MO, USA), and 0.5% bovine serum albumin fraction V (Roche, Basel, Switzerland). The homogenized lung tissue was centrifuged for 5 min at 4°C and 8000 rpm. The supernatant was collected and stored at -80°C until quantification of virus titers.

### **Virus titration of lung homogenates**

Plaque assays were performed as described previously [126]. Briefly, the mouse lung homogenates were diluted in MEM without FCS, applied in tenfold dilutions onto confluent monolayers of MDCK cells, and incubated at 35°C with 5% CO<sub>2</sub>. After 1 h, the unbound virus was removed by discarding the supernatant and washing the cells with PBS. The cells were then overlaid with MEM containing 5 µg/mL acetylated trypsin (Sigma-Aldrich) and 1% Bacto Agar (Becton, Dickinson and Company, Franklin Lakes, NJ, USA). After incubation for 48 h at 35°C, the cells were stained with 0.005% neutral red. After incubation for another 24 h at 35°C, the number of plaques was counted. The number of plaque-forming units (PFU) was calculated as the product of the reciprocal value of the highest virus dilution and the number of plaques in the dilution.

### **Acknowledgments**

We would like to thank the Department of Equine Sciences and the Department of Clinical Sciences of Companion Animals for supplying erythrocytes. We thank Andrea Gröne and Hélène Verheije from the Division of Pathology, Department of Biomolecular Health Sciences, Faculty of Veterinary Medicine of Utrecht University for providing paraffin-embedded tissues. We thank Barry Rockx from the department of Viroscience, Erasmus Medical Center, the Netherlands for providing the hamster lung tissues. We thank Richard Wubbolts for allowing us to use the microscopes of the Centre for Cell Imaging at Utrecht University. We would like to thank professor Yoshihiro Sakoda (Hokkaido University, Laboratory of Microbiology, Department of Disease Control, Graduate School of Veterinary Medicine) for helping with the manuscript.

### **Funding**

R.P.dV. is a recipient of an ERC Starting Grant from the European Commission (802780) and a Beijerinck Premium of the Royal Dutch Academy of Sciences.

### **Institutional Review Board Statement**

Animal experiments were approved by the Institutional Animal Care and Use Committee of the Faculty of Veterinary Medicine, Hokkaido University (18-0006), and all experiments were carried out per the guidelines of this committee. All applicable international, national, and/or institutional guidelines for the care and use of animals were followed. The tissues used for this study were obtained from animal necropsies performed at the Veterinary Pathologic Diagnostic Center, Faculty of Veterinary Medicine, Utrecht University. These necropsies were performed for diagnostic and educational purposes and therefore no permission from the Committee on the Ethics of Animal Experiment was required.

### **Author Contributions**

Protein histochemistry, C.M.S., N.N., hemagglutination assay, R.P.dV.; glycan microarray studies, R.P.dV.; mice studies, M.O., expression and production of hemagglutinins, C.M.S.; conceptualization, C.M.S. and R.P.dV.; methodology,

C.M.S., N.N., M.O.; formal analysis, C.M.S., R.P.dV.; investigation, C.M.S., N.N., M.O.; resources, H.T. G.J.B.; writing—original draft preparation, C.M.S., R.P.dV.; writing—review and editing, all authors.; visualization, C.M.S., R.P.dV.; supervision, R.P.dV.; project administration, R.P.dV.; funding acquisition, G.J.B., R.P.dV. All authors have read and agreed to the published version of the manuscript.

## References

1. Ge, S. and Z. Wang, *An overview of influenza A virus receptors*. Crit Rev Microbiol, 2011. 37(2): p. 157-65.
2. Gambaryan, A.S., et al., *Receptor-binding profiles of H7 subtype influenza viruses in different host species*. J Virol, 2012. 86(8): p. 4370-9.
3. Broszeit, F., et al., *N-glycolylneuraminic acid as a receptor for influenza A viruses*. Cell Rep, 2019. 27(11): p. 3284-3294 e6.
4. Peri, S., et al., *Phylogenetic distribution of CMP-Neu5Ac hydroxylase (CMAH), the enzyme synthesizing the proinflammatory human xenoantigen Neu5Gc*. Genome Biol Evol, 2018. 10(1): p. 207-219.
5. Altman, M.O. and P. Gagneux, *Absence of Neu5Gc and presence of anti-Neu5Gc antibodies in humans—an evolutionary perspective*. Front Immunol, 2019. 10: p. 789.
6. Yoshida, Y., et al., *Quantitative analysis of total serum glycome in human and mouse*. Proteomics, 2016. 16(21): p. 2747-2758.
7. Varki, A., *Loss of N-glycolylneuraminic acid in humans: mechanisms, consequences, and implications for hominid evolution*. Am J Phys Anthropol, 2001. Suppl 33: p. 54-69.
8. Jahan, M., et al., *The non-human glycan, N-glycolylneuraminic acid (Neu5Gc), is not expressed in all organs and skeletal muscles of nine animal species*. Food Chem, 2021. 343: p. 128439.
9. Raju, T.S., et al., *Species-specific variation in glycosylation of IgG: evidence for the species-specific sialylation and branch-specific galactosylation and importance for engineering recombinant glycoprotein therapeutics*. Glycobiology, 2000. 10(5): p. 477-86.
10. Bleckmann, C., et al., *O-glycosylation pattern of CD24 from mouse brain*. Biol Chem, 2009. 390(7): p. 627-45.
11. Gizaw, S.T., et al., *A comprehensive glycome profiling of Huntington's disease transgenic mice*. Biochim Biophys Acta, 2015. 1850(9): p. 1704-18.
12. Diaz, S.L., et al., *Sensitive and specific detection of the non-human sialic acid N-glycolylneuraminic acid in human tissues and biotherapeutic products*. PLoS One, 2009. 4(1): p. e4241.
13. Reiding, K.R., et al., *Murine plasma N-Glycosylation traits associated with sex and strain*. J Proteome Res, 2016. 15(10): p. 3489-3499.
14. Ghaderi, D., et al., *Sexual selection by female immunity against paternal antigens can fix loss of function alleles*. Proc Natl Acad Sci U S A, 2011. 108(43): p. 17743-8.
15. Hedlund, M., et al., *N-glycolylneuraminic acid deficiency in mice: implications for human biology and evolution*. Mol Cell Biol, 2007. 27(12): p. 4340-6.
16. Gagneux, P., et al., *Human-specific regulation of  $\alpha$ 2-6-linked sialic acids*. J Biol Chem, 2003. 278(48): p. 48245-50.
17. Jia, N., et al., *Glycomic characterization of respiratory tract tissues of ferrets: implications for its use in influenza virus infection studies*. J Biol Chem, 2014. 289(41): p. 28489-504.
18. Ng, P.S., et al., *Ferrets exclusively synthesize Neu5Ac and express naturally humanized influenza A virus receptors*. Nat Commun, 2014. 5: p. 5750.
19. Zou, G., et al., *Comprehensive analysis of N-glycans in IgG purified from ferrets with or without influenza A virus infection*. J Biol Chem, 2018. 293(50): p. 19277-19289.
20. Naito, Y., et al., *Germinal center marker GL7 probes activation-dependent repression of N-glycolylneuraminic acid, a sialic acid species involved in the negative modulation of B-cell activation*. Mol Cell Biol, 2007. 27(8): p. 3008-22.
21. Radigan, K.A., et al., *Modeling human influenza infection in the laboratory*. Infect Drug Resist, 2015. 8: p. 311-20.
22. Thangavel, R.R. and N.M. Bouvier, *Animal models for influenza virus pathogenesis, transmission, and immunology*. J Immunol Methods, 2014. 410: p. 60-79.
23. Iwatsuki-Horimoto, K., et al., *Syrian hamster as an animal model for the study of human influenza virus infection*. J Virol, 2018. 92(4).
24. Samet, S.J. and S.M. Tompkins, *Influenza pathogenesis in genetically defined resistant and susceptible murine strains*. Yale J Biol Med, 2017. 90(3): p. 471-479.
25. Mifsud, E.J., C.M. Tai, and A.C. Hurt, *Animal models used to assess influenza antivirals*. Expert Opin Drug Discov, 2018. 13(12): p. 1131-1139.
26. Albrecht, R.A., et al., *Moving forward: recent developments for the ferret biomedical research model*. mBio, 2018. 9(4).
27. Belser, J.A., et al., *A guide for the use of the ferret model for influenza virus infection*. Am J Pathol, 2020. 190(1): p. 11-24.
28. Belser, J.A., et al., *Complexities in ferret influenza virus pathogenesis and transmission models*. Microbiol Mol Biol Rev, 2016. 80(3): p. 733-44.
29. Maher, J.A. and J. DeStefano, *The ferret: an animal model to study influenza virus*. Lab Anim (NY), 2004. 33(9): p. 50-3.
30. Ottolini, M.G., et al., *The cotton rat provides a useful small-animal model for the study of influenza virus pathogenesis*. J Gen Virol, 2005. 86(Pt 10): p. 2823-2830.

31. Green, M.G., D. Huey, and S. Niewiesk, *The cotton rat (Sigmodon hispidus) as an animal model for respiratory tract infections with human pathogens*. Lab Anim (NY), 2013. 42(5): p. 170-6.
32. Bouvier, N.M., *Animal models for influenza virus transmission studies: a historical perspective*. Curr Opin Virol, 2015. 13: p. 101-8.
33. Li, R., et al., *Tree shrew as a new animal model to study the pathogenesis of avian influenza (H9N2) virus infection*. Emerg Microbes Infect, 2018. 7(1): p. 166.
34. Yang, Z.F., et al., *The tree shrew provides a useful alternative model for the study of influenza H1N1 virus*. Virol J, 2013. 10: p. 111.
35. Holzer, B., et al., *T and B cell immune responses to influenza viruses in pigs*. Front Immunol, 2019. 10: p. 98.
36. Iwatsuki-Horimoto, K., et al., *The marmoset as an animal model of influenza: infection with A(H1N1)pdm09 and highly pathogenic A(H5N1) viruses via the conventional or tracheal spray route*. Front Microbiol, 2018. 9: p. 844.
37. Davis, A.S., J.K. Taubenberger, and M. Bray, *The use of nonhuman primates in research on seasonal, pandemic and avian influenza, 1893-2014*. Antiviral Res, 2015. 117: p. 75-98.
38. Manickam, C., et al., *Cytokine-mediated tissue injury in non-human primate models of viral infections*. Front Immunol, 2018. 9: p. 2862.
39. Moncla, L.H., et al., *A novel nonhuman primate model for influenza transmission*. PLoS One, 2013. 8(11): p. e78750.
40. Oh, D.Y. and A.C. Hurt, *Using the ferret as an animal model for investigating influenza antiviral effectiveness*. Front Microbiol, 2016. 7: p. 80.
41. Padilla-Carlin, D.J., D.N. McMurray, and A.J. Hickey, *The guinea pig as a model of infectious diseases*. Comp Med, 2008. 58(4): p. 324-40.
42. Miao, J., et al., *Syrian hamster as an animal model for the study on infectious diseases*. Front Immunol, 2019. 10: p. 2329.
43. Yao, Y.G., *Creating animal models, why not use the Chinese tree shrew (Tupaia belangeri chinensis)?* Zool Res, 2017. 38(3): p. 118-126.
44. Meurens, F., et al., *The pig: a model for human infectious diseases*. Trends Microbiol, 2012. 20(1): p. 50-7.
45. Wong, J., et al., *Improving immunological insights into the ferret model of human viral infectious disease*. Influenza Other Respir Viruses, 2019. 13(6): p. 535-546.
46. Tripp, R.A. and S.M. Tompkins, *Animal models for evaluation of influenza vaccines*. Curr Top Microbiol Immunol, 2009. 333: p. 397-412.
47. Rajao, D.S. and A.L. Vincent, *Swine as a model for influenza A virus infection and immunity*. ILAR J, 2015. 56(1): p. 44-52.
48. Miller, L.A., et al., *Nonhuman primate models of respiratory disease: past, present, and future*. ILAR J, 2017. 58(2): p. 269-280.
49. Belser, J.A., J.M. Katz, and T.M. Tumpey, *The ferret as a model organism to study influenza A virus infection*. Dis Model Mech, 2011. 4(5): p. 575-9.
50. Kennedy, A.R., et al., *Morphometric and histological analysis of the lungs of Syrian golden hamsters*. J Anat, 1978. 125(Pt 3): p. 527-53.
51. Ye, L., et al., *Tree shrew as a new animal model for the study of lung cancer*. Oncol Lett, 2016. 11(3): p. 2091-2095.
52. Canning, B.J. and Y. Chou, *Using guinea pigs in studies relevant to asthma and COPD*. Pulm Pharmacol Ther, 2008. 21(5): p. 702-20.
53. Pan, H., et al., *Comprehensive anatomic ontologies for lung development: a comparison of alveolar formation and maturation within mouse and human lung*. J Biomed Semantics, 2019. 10(1): p. 18.
54. Zhang, J., et al., *Characteristics of the tree shrew humoral immune system*. Mol Immunol, 2020. 127: p. 175-185.
55. Tchilian, E. and B. Holzer, *Harnessing local immunity for an effective universal swine influenza vaccine*. Viruses, 2017. 9(5).
56. Chua, S., et al., *Alternative experimental models for studying influenza proteins, host-virus interactions and anti-influenza drugs*. Pharmaceuticals, 2019. 12(4).
57. Messaoudi, I., et al., *Nonhuman primate models of human immunology*. Antioxid Redox Signal, 2011. 14(2): p. 261-73.
58. Barreiro, L.B., et al., *Functional comparison of innate immune signaling pathways in primates*. PLoS Genet, 2010. 6(12): p. e1001249.
59. Varki, N.M., et al., *Biomedical differences between human and nonhuman hominids: potential roles for uniquely human aspects of sialic acid biology*. Annu Rev Pathol, 2011. 6: p. 365-93.
60. Redelinghuys, P., et al., *Early murine T-lymphocyte activation is accompanied by a switch from N-glycolyl- to N-acetyl-neuraminic acid and generation of ligands for siglec-E*. J Biol Chem, 2011. 286(40): p. 34522-32.
61. Naito-Matsui, Y., et al., *Functional evaluation of activation-dependent alterations in the sialoglycan composition of T cells*. J Biol Chem, 2014. 289(3): p. 1564-79.
62. Buchlis, G., et al., *Enhanced T cell function in a mouse model of human glycosylation*. J Immunol, 2013. 191(1): p. 228-37.
63. Kreijtz, J.H., et al., *Primary influenza A virus infection induces cross-protective immunity against a lethal infection with a heterosubtypic virus strain in mice*. Vaccine, 2007. 25(4): p. 612-20.
64. Nachbagauer, R., et al., *Defining the antibody cross-reactome directed against the influenza virus surface glycoproteins*. Nat Immunol, 2017. 18(4): p. 464-473.
65. Allen, J.D., et al., *Elicitation of protective antibodies against 20 years of future H3N2 cocirculating influenza virus variants in ferrets preimmune to historical H3N2 influenza viruses*. J Virol, 2019. 93(3).
66. Music, N., et al., *Repeated vaccination against matched H3N2 influenza virus gives less protection than single vaccination in ferrets*. NPJ Vaccines, 2019. 4: p. 28.

67. Skarlupka, A.L. and T.M. Ross, *Immune imprinting in the influenza ferret model*. *Vaccines*, 2020. 8(2).
68. Karlsson, E.A., et al., *Influenza virus infection in nonhuman primates*. *Emerg Infect Dis*, 2012. 18(10): p. 1672-5.
69. Collaborative Cross, C., *The genome architecture of the Collaborative Cross mouse genetic reference population*. *Genetics*, 2012. 190(2): p. 389-401.
70. Leist, S.R., et al., *Influenza H3N2 infection of the collaborative cross founder strains reveals highly divergent host responses and identifies a unique phenotype in CAST/EiJ mice*. *BMC Genomics*, 2016. 17: p. 143.
71. Boon, A.C., et al., *Host genetic variation affects resistance to infection with a highly pathogenic H5N1 influenza A virus in mice*. *J Virol*, 2009. 83(20): p. 10417-26.
72. Boon, A.C., et al., *Cross-reactive neutralizing antibodies directed against pandemic H1N1 2009 virus are protective in a highly sensitive DBA/2 mouse influenza model*. *J Virol*, 2010. 84(15): p. 7662-7.
73. Pica, N., et al., *The DBA.2 mouse is susceptible to disease following infection with a broad, but limited, range of influenza A and B viruses*. *J Virol*, 2011. 85(23): p. 12825-9.
74. Srivastava, B., et al., *Host genetic background strongly influences the response to influenza a virus infections*. *PLoS One*, 2009. 4(3): p. e4857.
75. Trammell, R.A., T.A. Liberati, and L.A. Toth, *Host genetic background and the innate inflammatory response of lung to influenza virus*. *Microbes Infect*, 2012. 14(1): p. 50-8.
76. Everitt, A.R., et al., *IFITM3 restricts the morbidity and mortality associated with influenza*. *Nature*, 2012. 484(7395): p. 519-23.
77. Shin, D.L., et al., *Protection from severe influenza virus infections in mice carrying the Mx1 influenza virus resistance gene strongly depends on genetic background*. *J Virol*, 2015. 89(19): p. 9998-10009.
78. Huang, C.H., et al., *Caspase-1 deficient mice are more susceptible to influenza A virus infection with PA variation*. *J Infect Dis*, 2013. 208(11): p. 1898-905.
79. Schauer, R., *Sialic acids as link to Japanese scientists*. *Proc Jpn Acad Ser B Phys Biol Sci*, 2016. 92(4): p. 109-20.
80. Löffling, J., et al., *Canine and feline parvoviruses preferentially recognize the non-human cell surface sialic acid N-glycolylneuraminic acid*. *Virology*, 2013. 440(1): p. 89-96.
81. Martin, P.T., et al., *A comparative study of N-glycolylneuraminic acid (Neu5Gc) and cytotoxic T cell (CT) carbohydrate expression in normal and dystrophin-deficient dog and human skeletal muscle*. *PLoS One*, 2014. 9(2): p. e88226.
82. Sun, Y., et al., *Guinea pig model for evaluating the potential public health risk of swine and avian influenza viruses*. *PLoS One*, 2010. 5(11): p. e15537.
83. Iwersen, M., V. Vandamme-Feldhaus, and R. Schauer, *Enzymatic 4-O-acetylation of N-acetylneuraminic acid in guinea-pig liver*. *Glycoconj J*, 1998. 15(9): p. 895-904.
84. Blanco, J.C., et al., *Receptor characterization and susceptibility of cotton rats to avian and 2009 pandemic influenza virus strains*. *J Virol*, 2013. 87(4): p. 2036-45.
85. Gao, W.N., et al., *Microfluidic chip-LC/MS-based glycomic analysis revealed distinct N-glycan profile of rat serum*. *Sci Rep*, 2015. 5: p. 12844.
86. Yau, L.F., et al., *Comprehensive glycomic profiling of respiratory tract tissues of tree shrews by TiO<sub>2</sub>-PGC chip mass spectrometry*. *J Proteome Res*, 2020. 19(4): p. 1470-1480.
87. Bateman, A.C., et al., *Glycan analysis and influenza A virus infection of primary swine respiratory epithelial cells: the importance of NeuAca2-6 glycans*. *J Biol Chem*, 2010. 285(44): p. 34016-26.
88. Iwatsuki-Horimoto, K., et al., *The microminipig as an animal model for influenza A virus infection*. *J Virol*, 2017. 91(2).
89. Byrd-Leotis, L., et al., *Shotgun glycomics of pig lung identifies natural endogenous receptors for influenza viruses*. *Proc Natl Acad Sci U S A*, 2014. 111(22): p. E2241-50.
90. Tao, N., et al., *Structural determination and daily variations of porcine milk oligosaccharides*. *J Agric Food Chem*, 2010. 58(8): p. 4653-9.
91. Cooper, D.K., *Modifying the sugar icing on the transplantation cake*. *Glycobiology*, 2016. 26(6): p. 571-81.
92. Kim, Y.G., et al., *Identification of alpha-Gal and non-Gal epitopes in pig corneal endothelial cells and keratocytes by using mass spectrometry*. *Curr Eye Res*, 2009. 34(10): p. 877-95.
93. Burlak, C., et al., *N-linked glycan profiling of GGTA1/CMAH knockout pigs identifies new potential carbohydrate xenoantigens*. *Xenotransplantation*, 2013. 20(5): p. 277-91.
94. Suzuki, T., et al., *Swine influenza virus strains recognize sialylsugar chains containing the molecular species of sialic acid predominantly present in the swine tracheal epithelium*. *FEBS Lett*, 1997. 404(2-3): p. 192-6.
95. Sriwilaijaroen, N., et al., *N-glycans from porcine trachea and lung: predominant NeuAca2-6Gal could be a selective pressure for influenza variants in favor of human-type receptor*. *PLoS One*, 2011. 6(2): p. e16302.
96. Springer, S.A., S.L. Diaz, and P. Gagneux, *Parallel evolution of a self-signal: humans and new world monkeys independently lost the cell surface sugar Neu5Gc*. *Immunogenetics*, 2014. 66(11): p. 671-4.
97. Li, Q., et al., *Carbohydrate antigen expression and anti-pig antibodies in New World capuchin monkeys: Relevance to studies of xenotransplantation*. *Xenotransplantation*, 2019. 26(3): p. e12498.
98. Daly, J.M., et al., *Transmission of equine influenza virus to English foxhounds*. *Emerg Infect Dis*, 2008. 14(3): p. 461-4.
99. Yasue, S., et al., *Difference in form of sialic acid in red blood cell glycolipids of different breeds of dogs*. *J Biochem*, 1978. 83(4): p. 1101-7.
100. Barnard, K.N., et al., *Modified sialic acids on mucus and erythrocytes inhibit influenza A virus hemagglutinin and neuraminidase functions*. *J Virol*, 2020. 94(9).
101. Hashimoto, Y., T. Yamakawa, and Y. Tanabe, *Further studies on the red cell glycolipids of various breeds of dogs. A possible assumption about the origin of Japanese dogs*. *J Biochem*, 1984. 96(6): p. 1777-82.
102. Wen, F., et al., *Mutation W222L at the receptor binding site of hemagglutinin could facilitate viral adaption from equine influenza A (H3N8) virus to dogs*. *J Virol*, 2018. 92(18).



103. Suzuki, Y., et al., *Sialic acid species as a determinant of the host range of influenza A viruses*. J Virol, 2000. 74(24): p. 11825-31.
104. Takahashi, T., et al., *N-glycolylneuraminic acid on human epithelial cells prevents entry of influenza A viruses that possess N-glycolylneuraminic acid binding ability*. J Virol, 2014. 88(15): p. 8445-56.
105. Higa, H.H., G.N. Rogers, and J.C. Paulson, *Influenza virus hemagglutinins differentiate between receptor determinants bearing N-acetyl-, N-glycolyl-, and N,O-diacetylneuraminic acids*. Virology, 1985. 144(1): p. 279-82.
106. Spruit, C.M., et al., *N-glycolylneuraminic acid binding of avian H7 influenza A viruses*. bioRxiv, 2020.
107. Tumpey, T.M., et al., *A two-amino acid change in the hemagglutinin of the 1918 influenza virus abolishes transmission*. Science, 2007. 315(5812): p. 655-9.
108. Maines, T.R., et al., *Lack of transmission of H5N1 avian-human reassortant influenza viruses in a ferret model*. Proc Natl Acad Sci U S A, 2006. 103(32): p. 12121-6.
109. Gambaryan, A., et al., *Evolution of the receptor binding phenotype of influenza A (H5) viruses*. Virology, 2006. 344(2): p. 432-8.
110. Imai, M., et al., *Syrian hamsters as a small animal model for SARS-CoV-2 infection and countermeasure development*. Proc Natl Acad Sci U S A, 2020. 117(28): p. 16587-16595.
111. Rosenke, K., et al., *Defining the Syrian hamster as a highly susceptible preclinical model for SARS-CoV-2 infection*. Emerg Microbes Infect, 2020. 9(1): p. 2673-2684.
112. Padler-Karavani, V., et al., *Diversity in specificity, abundance, and composition of anti-Neu5Gc antibodies in normal humans: potential implications for disease*. Glycobiology, 2008. 18(10): p. 818-30.
113. Glaser, L., et al., *Effective replication of human influenza viruses in mice lacking a major  $\alpha$ 2,6 sialyltransferase*. Virus Res, 2007. 126(1-2): p. 9-18.
114. Long, J.S., et al., *Host and viral determinants of influenza A virus species specificity*. Nat Rev Microbiol, 2019. 17(2): p. 67-81.
115. Ji, Y., et al., *New insights into influenza A specificity: an evolution of paradigms*. Curr Opin Struct Biol, 2017. 44: p. 219-231.
116. de Vries, R.P., et al., *Hemagglutinin receptor specificity and structural analyses of respiratory droplet-transmissible H5N1 viruses*. J Virol, 2014. 88(1): p. 768-73.
117. de Vries, R.P., et al., *Only two residues are responsible for the dramatic difference in receptor binding between swine and new pandemic H1 hemagglutinin*. J Biol Chem, 2011. 286(7): p. 5868-75.
118. de Vries, R.P., et al., *A single mutation in Taiwanese H6N1 influenza hemagglutinin switches binding to human-type receptors*. EMBO Mol Med, 2017. 9(9): p. 1314-1325.
119. Harbury, P.B., et al., *A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants*. Science, 1993. 262(5138): p. 1401-7.
120. Nemanichvili, N., et al., *Fluorescent trimeric hemagglutinins reveal multivalent receptor binding properties*. J Mol Biol, 2019. 431(4): p. 842-856.
121. Shaner, N.C., et al., *Improving the photostability of bright monomeric orange and red fluorescent proteins*. Nat Methods, 2008. 5(6): p. 545-51.
122. de Vries, R.P., et al., *The influenza A virus hemagglutinin glycosylation state affects receptor-binding specificity*. Virology, 2010. 403(1): p. 17-25.
123. Zeng, Q., et al., *Structure of coronavirus hemagglutinin-esterase offers insight into corona and influenza virus evolution*. Proc Natl Acad Sci U S A, 2008. 105(26): p. 9065-9.
124. Bouwman, K.M., et al., *Three amino acid changes in avian coronavirus spike protein allow binding to kidney tissue*. J Virol, 2020. 94(2).
125. Wickramasinghe, I.N., et al., *Binding of avian coronavirus spike proteins to host factors reflects virus tropism and pathogenicity*. J Virol, 2011. 85(17): p. 8903-12.
126. Hiono, T., et al., *Experimental infection of highly and low pathogenic avian influenza viruses to chickens, ducks, tree sparrows, jungle crows, and black rats for the evaluation of their roles in virus transmission*. Vet Microbiol, 2016. 182: p. 108-15.



# Chapter 4

## ***N*-glycolylneuraminic acid binding of avian and equine H7 influenza A viruses**

Cindy M. Spruit, Xueyong Zhu, Ilhan Tomris, María Ríos Carrasco, Alvin X. Han, Frederik Broszeit, Roosmarijn van der Woude, Kim M. Bouwman, Michel M. T. Luu, Keita Matsuno, Yoshihiro Sakoda, Colin A. Russell, Ian A. Wilson, Geert-Jan Boons, and Robert P. de Vries

Published: *Journal of Virology*, March 2022, DOI: <https://doi.org/10.1128/jvi.02120-21>

### **Abstract**

Influenza A viruses (IAV) initiate infection by binding to glycans with terminal sialic acids on the cell surface. Hosts of IAV variably express two major forms of sialic acid, *N*-acetylneuraminic acid (NeuAc) and *N*-glycolylneuraminic acid (NeuGc). NeuGc is produced in most mammals including horses and pigs, but is absent in humans, ferrets, and birds. The only known naturally occurring IAVs that exclusively bind NeuGc are extinct highly pathogenic equine H7N7 viruses. We determined the crystal structure of a representative equine H7 hemagglutinin (HA) in complex with NeuGc and observed high similarity in the receptor-binding domain with an avian H7 HA. To determine the molecular basis for NeuAc and NeuGc specificity, we performed systematic mutational analyses, based on the structural insights, on two distant avian H7 HAs and an H15 HA. We found that mutation A135E is key for binding  $\alpha$ 2,3-linked NeuGc but does not abolish NeuAc binding. Additional mutations S128T, I130V, T189A, and K193R converted the specificity from NeuAc to NeuGc. We investigated the residues at positions 128, 130, 135, 189, and 193 in a phylogenetic analysis of avian and equine H7 HAs. This revealed a clear distinction between equine and avian residues. The highest variability was observed at key position 135, of which only the equine glutamic acid led to NeuGc binding. These results demonstrate that genetically distinct H7 and H15 HAs can be switched from NeuAc to NeuGc binding and vice versa after introduction of several mutations, providing insights into the adaptation of H7 viruses to NeuGc receptors.

## Introduction

Influenza A viruses (IAV) can infect a broad range of animals, including mammalian and avian species. Infection is initiated when the hemagglutinin (HA) on the outside of a virus particle binds to glycans with terminal sialic acid on the cell surface. The vast majority of IAVs use a glycan with a terminal *N*-acetylneuraminic acid (NeuAc) as their receptor, although some strains use *N*-glycolylneuraminic acid (NeuGc) instead. Sialic acids are bound in the receptor-binding site (RBS) of the HA, consisting of conserved residues (Y98, W153, H183, and Y195) and structural features (130-, 150-, and 220-loops and 190-helix) [1]. Amino acid mutations in or near the RBS can change HA binding specificity, as shown extensively for HAs binding to either  $\alpha$ 2,3-linked or  $\alpha$ 2,6-linked NeuAc [2-6].

The ability of viruses to bind either  $\alpha$ 2,3-linked or  $\alpha$ 2,6-linked sialic acids is a host determinant. Binding to either NeuAc or NeuGc could likewise affect the host range. NeuGc is only present in species that express an active form of the enzyme cytidine monophosphate-*N*-acetyl neuraminic acid hydroxylase (CMAH), which facilitates the hydroxylation of NeuAc to convert it to NeuGc. The gene encoding CMAH, mainly expressed in mammalian species, has been partially or completely lost at several distinct events during evolution [7], causing NeuGc to be absent in, among others, humans, ferrets, European dogs, and avian species [8-11]. In species that generate NeuGc, its percentage of the total sialic acid content varies. For instance, pig trachea contains an equal amount of NeuAc and NeuGc, while 90% of the sialic acids on equine trachea and erythrocytes is NeuGc [12-15]. The loss of active CMAH enzymes may have been initiated by evolutionary pressure from lethal pathogens binding to NeuGc, thereby protecting individuals with low levels of NeuGc [16]. Thereupon, IAV may have co-evolved with host species to bind NeuAc instead of NeuGc. Birds, which do not express NeuGc, are the reservoir for IAV.

The high NeuGc content in horses may explain why equine H7N7 viruses are the only known IAV that bind  $\alpha$ 2,3-linked NeuGc [17, 18]. These highly pathogenic equine H7 viruses have not been isolated since 1978 and are, therefore, thought to be extinct [19, 20]. Unlike equine H7 strains, avian and human H7 viruses, as well as related avian H15 viruses, bind NeuAc [18, 21, 22]. At the moment, it is still unclear whether NeuGc could have been the archaic receptor of IAV, where the NeuGc binding of equine H7 viruses originated from, and what the molecular determinants of NeuGc specificity are.

Here, we investigated the receptor binding specificities of equine and avian H7 and H15 HAs to identify the origin of the NeuGc receptor binding of equine H7 viruses. The crystal structure of the HA of A/Equine/New York/49/73 H7N7 in complex with its ligand NeuGc was elucidated. Inspired by the similarities between this structure and the crystal structure of the HA of A/Turkey/Italy/214845/02 H7N3, we performed targeted mutagenesis of avian H7 and H15 HAs and the equine H7 HA. Several combinations of mutations were found that switched H7 and H15 HAs from binding NeuAc to NeuGc and vice versa. Our results demonstrate a phenotypical relationship between avian and equine H7 and H15 HA receptor

binding despite the substantial genetic distance between these subtypes and provide insights into the use of NeuGc as a, potentially archaic, receptor for IAV.

## Results

### Crystal structure of an equine H7 HA in complex with receptor analog 3'-GcLN and its similarity to an avian H7 HA

We previously reported the crystal structure of the HA of the highly pathogenic A/Equine/New York/49/73 H7N7 (H7eq) without a ligand (PDB: 6N5A [17]). To understand the structural basis for NeuGc specificity of H7eq, we determined the crystal structure of H7eq in complex with its natural ligand 3'-GcLN (NeuGca2-3Gal $\beta$ 1-4GlcNAc) at 2.05 Å resolution (Fig 1A and Table 1, PDB: 7T1V). The electron density for the ligand 3'-GcLN could be fitted well for all three monosaccharides (Fig 1B). H7eq binds 3'-GcLN mainly through NeuGc-1, but the interactions extend over the 220-loop with hydrogen bonds between Gal-2 and the main-chain carbonyl oxygen of G225 and between GlcNAc-3 and the side chain of Q222 (Fig 1A).

**Table 1. Data collection and refinement statistics of H7eq in complex with 3'-GcLN.**

<sup>a</sup> Parentheses denote outer-shell statistics. <sup>b</sup>  $R_{sym} = \sum_{hkl} \sum_i |I_{hkl,i} - \langle I_{hkl} \rangle| / \sum_{hkl} \sum_i I_{hkl,i}$  and  $R_{pim} = \sum_{hkl} [1/(N-1)]^{1/2} \sum_i |I_{hkl,i} - \langle I_{hkl} \rangle| / \sum_{hkl} \sum_i I_{hkl,i}$ , where  $I_{hkl,i}$  is the scaled intensity of the  $i$ th measurement of reflection  $h, k, l$ ,  $\langle I_{hkl} \rangle$  is the average intensity for that reflection, and  $N$  is the redundancy. <sup>c</sup> No. molecules for complexes refers to number of HA protomers per asymmetric unit (ASU). <sup>d</sup>  $R_{cryst} = \sum_{hkl} |F_o - F_c| / \sum_{hkl} |F_o|$ , where  $F_o$  and  $F_c$  are the observed and calculated structure factors. <sup>e</sup>  $R_{free}$  was calculated as for  $R_{cryst}$ , but on 5% of data excluded before refinement. <sup>f</sup> The values are the percentage of residues in the favored and outliers regions analyzed by MolProbity [23].

Data Collection statistics		Refinement statistics	
X-ray source	APS 23ID-D	Resolution (Å) <sup>a</sup>	48.87-2.05
Space group	P3	Reflections in refinement	110,267
Unit cell (Å)	$a = b = 112.8, c = 130.2$	Refined residues	1,455
Resolution (Å) <sup>a</sup>	48.87-2.05 (2.09-2.05)	Refined waters	810
Unique reflections	116,095 (5,442)	$R_{cryst}$ <sup>d</sup>	0.217
Redundancy <sup>a</sup>	9.4 (5.2)	$R_{free}$ <sup>e</sup>	0.245
Average $I/\sigma(I)$ <sup>a</sup>	16.1 (1.0)	B-values (Å <sup>2</sup> )	
		• Protein	70
		• RBS subdomain (Residues 117-265 of HA1) of chain A, C, E	33, 40, 64
		• Ligand of chain A, C, E	30, 53, 100
		• Waters	48
		• Wilson B-values (Å <sup>2</sup> )	28
Completeness <sup>a</sup>	99.5 (93.2)	Ramachandran values (%) <sup>f</sup>	96.3, 0
$R_{sym}$ <sup>a,b</sup>	0.10 (0.77)	r.m.s.d. bond (Å)	0.009
$R_{pim}$ <sup>a,b</sup>	0.03 (0.34)	r.m.s.d. angle (deg.)	1.43
CC <sub>1/2</sub> <sup>a</sup>	0.99 (0.70)	PDB code	7T1V
No. molecules per ASU <sup>c</sup>	3		



The RBS structures of H7eq and the low pathogenic A/Turkey/Italy/214845/02 H7N3 (H7tu) (PDB: 4SBI) [24] appeared to be very similar (Fig 1C-D), although the turkey strain binds NeuAc instead of NeuGc and was isolated almost 30 years after the equine strain. Nevertheless, 85% of HA1 residues are identical and the amino acid sequences around the RBS differ at 13 positions (Fig 1E and full alignment in Fig S1). The NeuGc-Gal bond of 3'-GcLN in the H7eq complex adopts a *cis* conformation, which is consistent with our previous findings in the structure of 3'-GcLN in complex with the A/Vietnam/1203/2004 H5N1 Y161A mutant that shifts receptor specificity from NeuAc to NeuGc [17]. On the contrary, the NeuAc-Gal bond in the avian receptor analog 3'-SLN (NeuAca2-3Gal $\beta$ 1-4GlcNAc) in complex with H7tu adopts a *trans* NeuAc-Gal bond (Fig 1C-D).

In the H7eq 3'-GcLN structure, the 1-hydroxyl group of NeuGc-1 forms a hydrogen bond with the main-chain nitrogen of E135 and the E135 side chain makes a salt bridge with R144 (Fig 1A). The amino acid at position 193 is known to be an important determinant of receptor specificity [25-28]. In the complex structure of H7eq with 3'-GcLN, R193 forms a hydrogen bond with NeuGc-1 (Fig 1A). In comparison, K193 in H7tu, with a shorter side-chain, is not in hydrogen bond distance with the NeuAc-1 of 3'-SLN (Fig 1C).

Despite these similarities in RBS structures, we found that H7tu bound solely to  $\alpha$ 2,3-linked NeuAc on the glycan array (Fig 2B), whereas H7eq exclusively bound to  $\alpha$ 2,3-linked NeuGc [17]. To decipher which residues determine NeuGc and NeuAc receptor specificity, targeted mutagenesis was performed on H7tu by replacing residues in the RBS with H7eq-like amino acids.

### **Amino acid 135 is essential for binding *N*-glycolylneuraminic acid**

To identify which amino acids are critical for NeuGc binding, we mutated the HA of H7tu towards H7eq at eight locations (128, 130, 135, 144, 159+160, 189, 193, and 219). Using the previously published glycan microarray containing glycans with terminal NeuAc or NeuGc (Fig 2A, [17]), we assessed the binding specificities of these recombinantly expressed HA mutants.

In the 130-loop, mutations S128T and I130V did not induce clear changes in the NeuAc/NeuGc specificity (Fig 2C-D). The amino acid at position 135 of H7 HAs is naturally diverse and has been associated with the adaptation of viruses between avian species and humans during a zoonotic outbreak of an H7N9 virus [29]. We observed that mutating position 135 (A135E) resulted in a gain of binding of the H7tu HA to NeuGc while maintaining binding to NeuAc (Fig 2E). Residue 143 has previously been suggested to be relevant for NeuGc recognition in H3 viruses [30]. In H7eq, R144 forms salt bridges with the 130-loop residue E135, but mutation G144R alone in the H7tu HA did not change the binding specificity (Fig 2F). In the 150-loop, a highly conserved tyrosine is present at position 161 in all HA subtypes except H7, H10, H12, H15, H17, and H18 [26, 31, 32]. Previously, it was demonstrated that a Y161A mutation changed the binding properties of an H5 HA from NeuAc to NeuGc [17, 31]. However, introducing Y161A in other HA subtypes (H1, H2, and H4) did not change binding specificity (Fig 3). We made mutations A159G and A160V simultaneously in the H7tu HA but, unlike the Y161A mutation in H5, we did

not observe NeuGc binding with this double mutation (Fig 2G). In the 190-helix, residue 189 is next to E190, which hydrogen bonds to the ligand, in both H7eq and H7tu (Fig 1A and 1C). Mutation T189A in the H7tu HA did not change receptor specificity when introduced on its own (Fig 2H). Residue 193 is important for ligand recognition (Fig 1A) [25-28]. Introducing K193R into the H7tu HA seemed to abolish all binding to the glycan array (Fig 2I), even when illuminating the glycan array with higher laser power. Despite residue 219 being very close to the 220-loop, mutation A219P did not change the binding properties of the H7tu HA from NeuAc to NeuGc (Fig 2J). In summary, while most mutations performed on the H7tu HA did not affect binding specificity, the introduction of mutation K193R abolished glycan-binding and A135E seemed to be key for binding NeuGc while maintaining binding to NeuAc.

### **Various combinations of mutations switch binding from NeuAc to NeuGc**

Starting from the key mutation A135E, we continued mutagenesis in the recombinantly expressed HAs by adding mutations at the previously stated positions (Fig 4A). Mutating more amino acids in the 130-loop, at positions 128 (S128T) or 130 (I130V), appeared to abolish NeuAc binding while maintaining binding to NeuGc on the glycan microarray. The combination of mutation A135E and mutation G144R, A159G+A160V, T189A, or A219P did not change binding specificity compared to mutation A135E alone since both NeuAc and NeuGc were still bound. Whereas almost all binding was abolished when introducing mutation K193R by itself (Fig 2I), adding mutation A135E restored binding to both NeuAc and NeuGc.

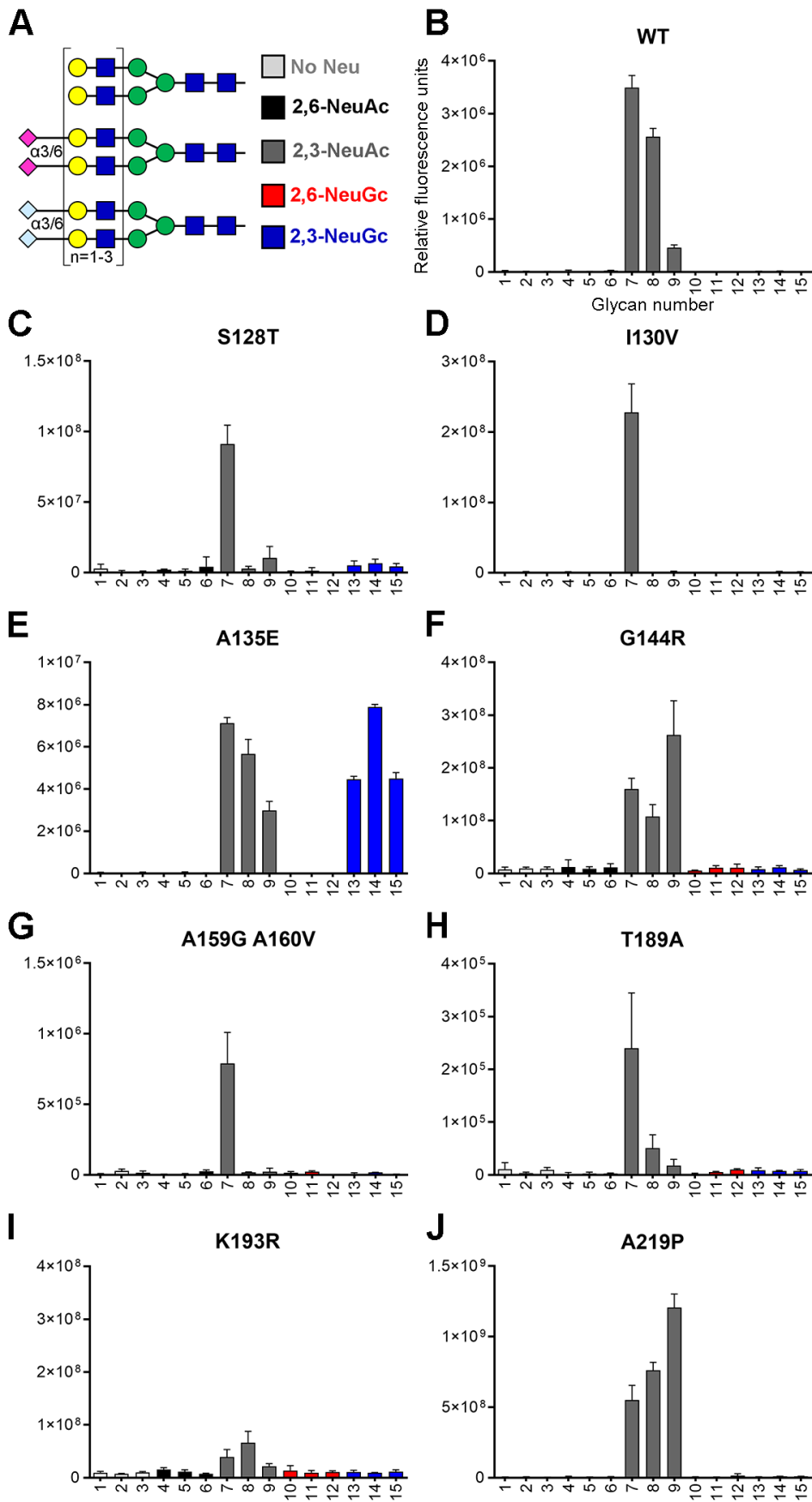
Since mutations A135E and K193R both affected the receptor-binding properties, we further combined these two mutations with mutations that did not change binding specificity so far (Fig 4B). We found that adding mutation G144R or A159G+A160V abolished binding to the array. Adding the T189A mutation in the A135E+K193R background, switched the H7tu HA to binding mainly NeuGc. The addition of mutation A219P did not affect the binding specificity since both NeuAc and NeuGc were still bound almost equally. In short, we were able to modify the H7tu HA for binding NeuGc specifically on the glycan microarray by combinations of mutations A135E+S128T, A135E+I130V, or A135E+T189A+K193R. These results show that residues in the 130-loop or 190-helix modify the specificity towards NeuGc.

### **No binding specificity to avian or equine erythrocytes and tracheal epithelium was observed for avian H7 mutants that bind NeuGc on the glycan microarray**

A glycan microarray, as used in this study, is a sophisticated tool to investigate the binding of proteins to synthetic glycans of which we know the exact structure. However, not all natural host-glycans can be present on the array and, therefore, it is necessary to investigate the binding specificities of HAs to host cells and tissues. Avian species lack a functional CMAH and therefore do not have NeuGc. On the contrary, on equine erythrocytes and tracheal tissue approximately 90% of the sialic acids are NeuGc [12, 14, 15] and to our knowledge, there are no species with a higher percentage of NeuGc. Therefore, we performed a

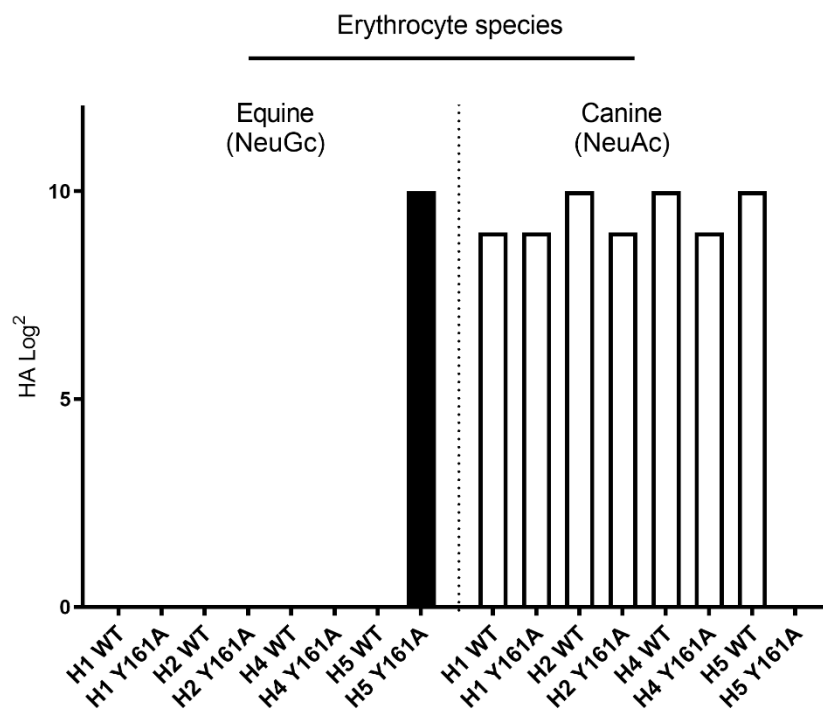


# A/Turkey/Italy/214845/02 H7N3





**Fig 2. Evaluation of the binding specificities of single mutants of the HA of A/Turkey/Italy/214845/02 H7N3.** (A) Synthetic glycans printed on the microarray ( $n=6$ ), either without sialic acid (structures 1-3, light gray), with  $\alpha 2,6$ -linked NeuAc (4-6, black),  $\alpha 2,3$ -linked NeuAc (7-9, dark gray),  $\alpha 2,6$ -linked NeuGc (10-12, red) or  $\alpha 2,3$ -linked NeuGc (13-15, blue). Structures 1, 4, 7, 10, and 13 contain one LacNAc repeat, while structures 2, 5, 8, 11, and 14 have two repeats and structures 3, 6, 9, 12, and 15 contain three repeats [17]. The glycan microarray, which is representative for two independent assays, was used to determine the receptor specificity of recombinantly expressed HA of (B) H7tu wild-type (WT), (C) mutant S128T, (D) I130V, (E) A135E, (F) G144R, (G) A159G+A160V, (H) T189A, (I) K193R, and (J) A219P.



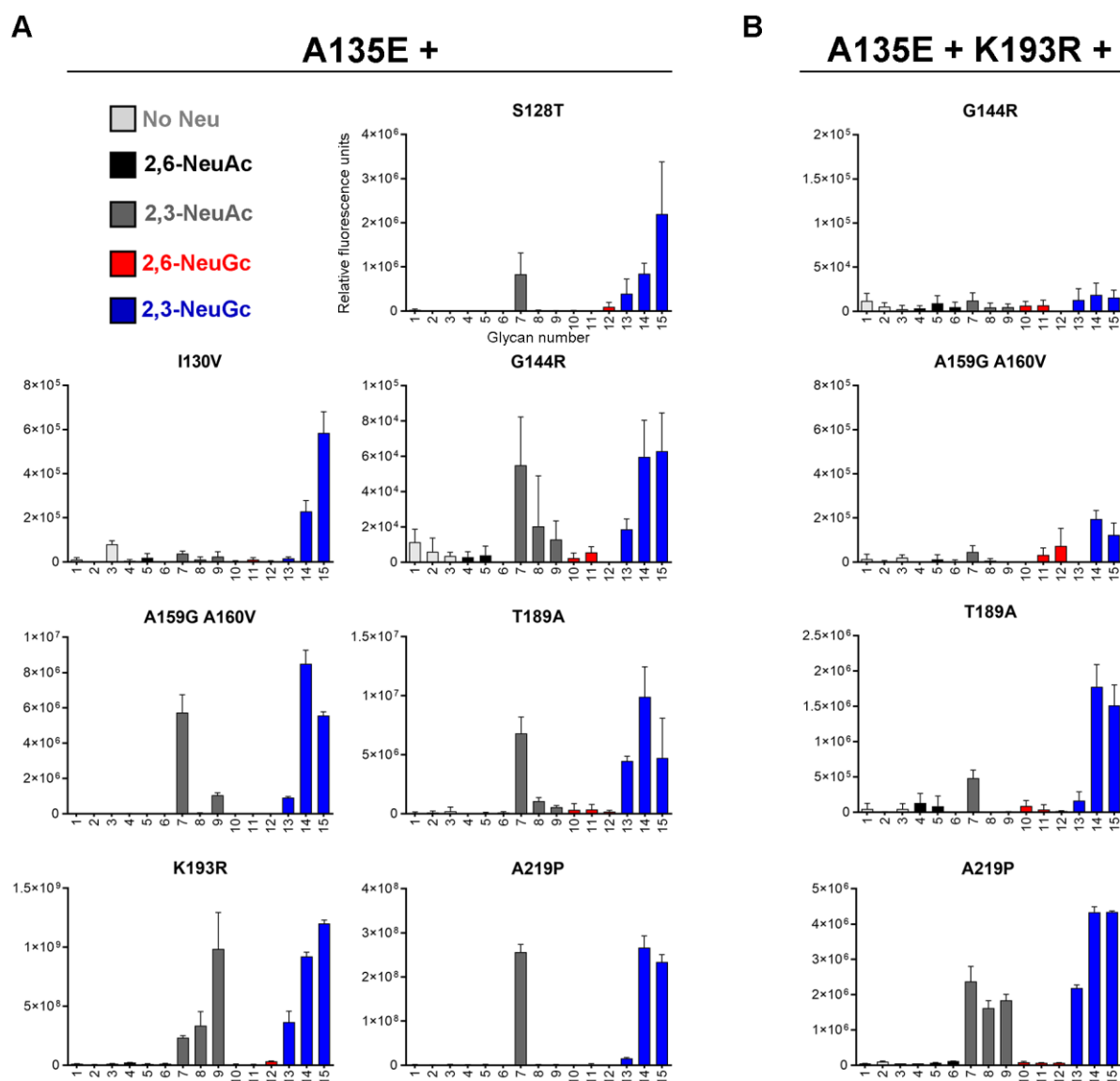
**Fig 3. Hemagglutination assay with Y161A HA mutants on equine and canine erythrocytes.** Equine erythrocytes contain approximately 90% NeuGc [12, 14, 15] and canine species cannot produce NeuGc [10]. Recombinantly expressed HAs of wild-type and Y161A mutants of A/Duck/Hokkaido/111/2009 H1N5, A/Duck/Hokkaido/95/2001 H2N2, A/Duck/Hokkaido/138/2007 H4N6, and A/Vietnam/1203/2004 H5N1 were investigated.

hemagglutination assay with avian and equine erythrocytes and tissue staining on tracheal epithelium, which is the natural location of infection, of the same species.

As controls for the presence of NeuAc and NeuGc on erythrocytes and tracheal epithelium, we used our previously studied WT and Y161A mutant HAs of A/Vietnam/1203/2004 H5N1 (H5VN) which specifically bind  $\alpha 2,3$ -linked NeuAc and  $\alpha 2,3$ -linked NeuGc, respectively. To our knowledge, these HAs are the only available controls for specific NeuAc and NeuGc binding [17]. The virus particles from these viruses however do not show exclusive specificity for NeuAc or NeuGc [17] and are therefore not appropriate to show the presence of these sialic acids.

While binding only to NeuAc on the glycan array, the wild-type (WT) H7tu HA agglutinated both chicken erythrocytes, which contain only NeuAc [7, 8], and

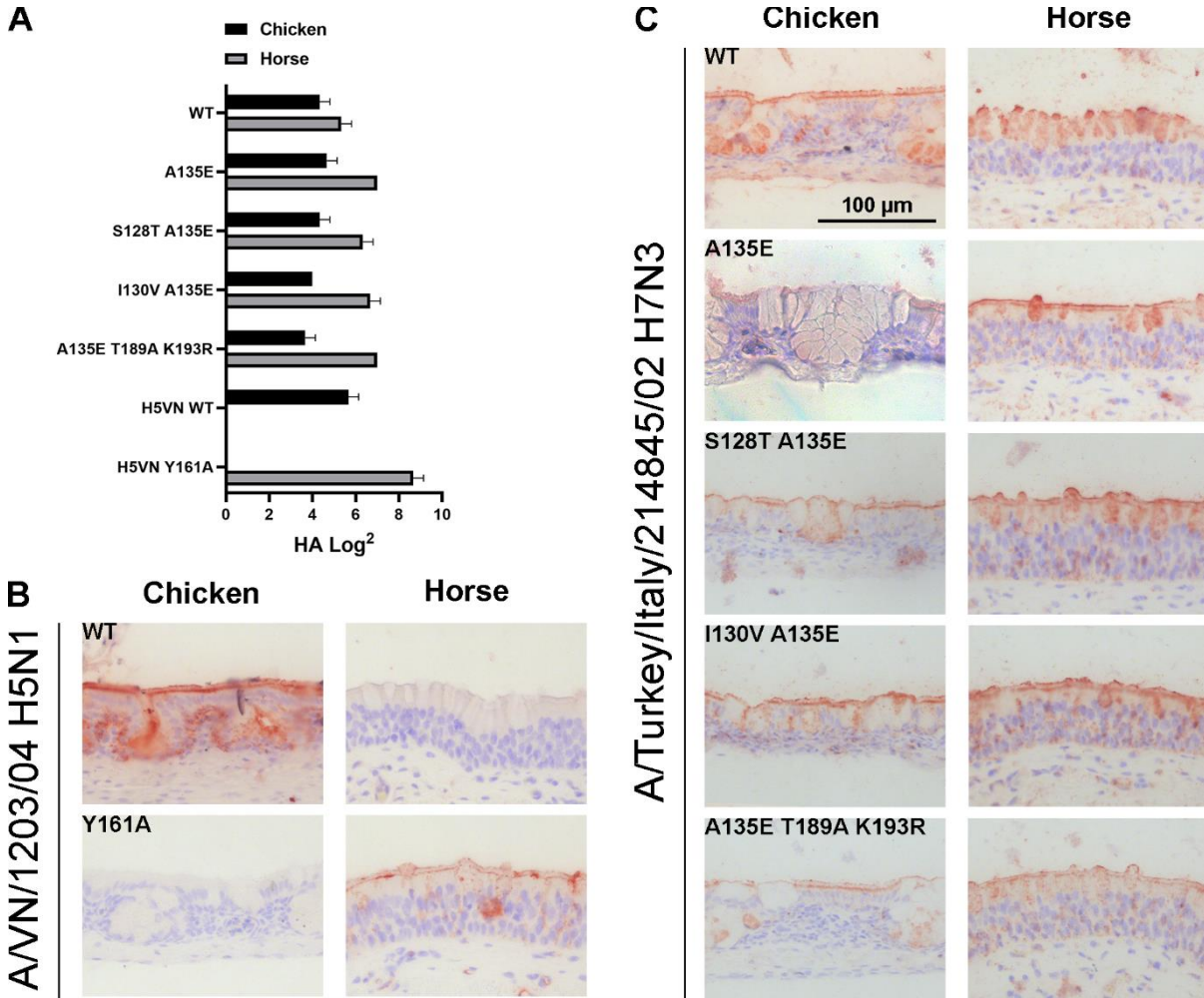
### A/Turkey/Italy/214845/02 H7N3



**Fig 4. Evaluation of the binding specificities of double and triple mutants of the HA of A/Turkey/Italy/214845/02 H7N3.** The glycan microarray as described in Fig 2A was used, containing glycans with terminal NeuAc or NeuGc or without sialic acid. Representative binding specificities for two independent assays for (A) mutant HAs containing mutation A135E and an additional mutation (S128T, I130V, G144R, A159G+A160V, T189A, K193R, or A219P) and (B) mutant HAs containing mutations A135E, K193R, and an additional mutation (G144R, A159G+A160V, T189A, or A219P) are shown.

horse erythrocytes, which contain mainly NeuGc and a small portion of NeuAc [12, 14, 15] (Fig 5A). Therefore, a loss of binding to chicken erythrocytes would indicate a loss of NeuAc-binding. However, both types of erythrocytes were still bound by HAs with combinations of all investigated mutants (A135E, A135E+S128T, A135E+I130V, and A135E+T189A+K193R, Fig 5A) and therefore no conclusions

concerning NeuGc specificity could be drawn from these hemagglutination experiments. Similarly, WT and all mutants of H7tu bound both horse and chicken tracheal tissue (Fig 5C). The results demonstrated that there are some differences in glycan binding between the glycan array, hemagglutination assay, and tissue staining.



**Fig 5. Binding specificities of (mutant) HA of A/Turkey/Italy/214845/02 H7N3 to chicken and horse erythrocytes and tracheal epithelium.** (A) A hemagglutination assay (n=3, mean and SD shown) with chicken and horse erythrocytes was performed using H7tu WT and mutant HAs (A135E, A135E+S128T, A135E+I130V, and A135E+T189A+K193R). AEC staining is used to visualize tissue binding. Tissue staining of chicken and horse tracheal epithelium is performed with (B) WT and Y161A mutant HA of A/Vietnam/1203/2004 H5N1 as a positive and negative control and (C) H7tu WT or mutant HAs as in panel A.

**NeuGc binding specificity can also be achieved in another avian H7 strain**

To investigate whether the mutations that were found to switch the HA of A/Turkey/Italy/214845/02 H7N3 (a virus from the Eurasian lineage) towards NeuGc binding are universal among H7 strains, we analyzed the HA of another avian strain from the North American lineage (A/Chicken/Jalisco/12283/2012 H7N3, highly pathogenic). Alignment of the HA sequences showed that the two strains differ at four amino acid positions (158, 188, 208, and 214) in the otherwise very similar RBS

(Fig 6A, full alignment in Fig S1). In glycan array analysis the WT HA of A/Chicken/Jalisco/12283/12 H7N3 bound NeuAc (Fig 6B). Solely introducing mutation A135E enabled NeuGc binding and seemed to abolish some binding to NeuAc. Furthermore, NeuGc binding specificity on the glycan array was achieved by combining mutation A135E with mutations I130V or T189A+K193R. A combination of A135E and S128T resulted in a loss of glycan binding on the array.

The combinations of mutations A135E, A135E+S128T, and A135E+I130V did not change the binding specificity of the HA in either the hemagglutination assay using chicken or horse erythrocytes (Fig 6C) or on equine and avian trachea (Fig 6D), as both species were still bound, similar to observations in H7tu. The combination of mutations A135E+T189A+K193R did not change the binding specificity in the hemagglutination assay either, but binding to both chicken and horse tracheal tissue was lost. Nevertheless, based on the glycan array analysis, we conclude that distant avian H7 HAs from different lineages can acquire NeuGc binding through identical amino acid changes.

### **Reciprocal mutations in the equine H7 HA allows binding to NeuAc**

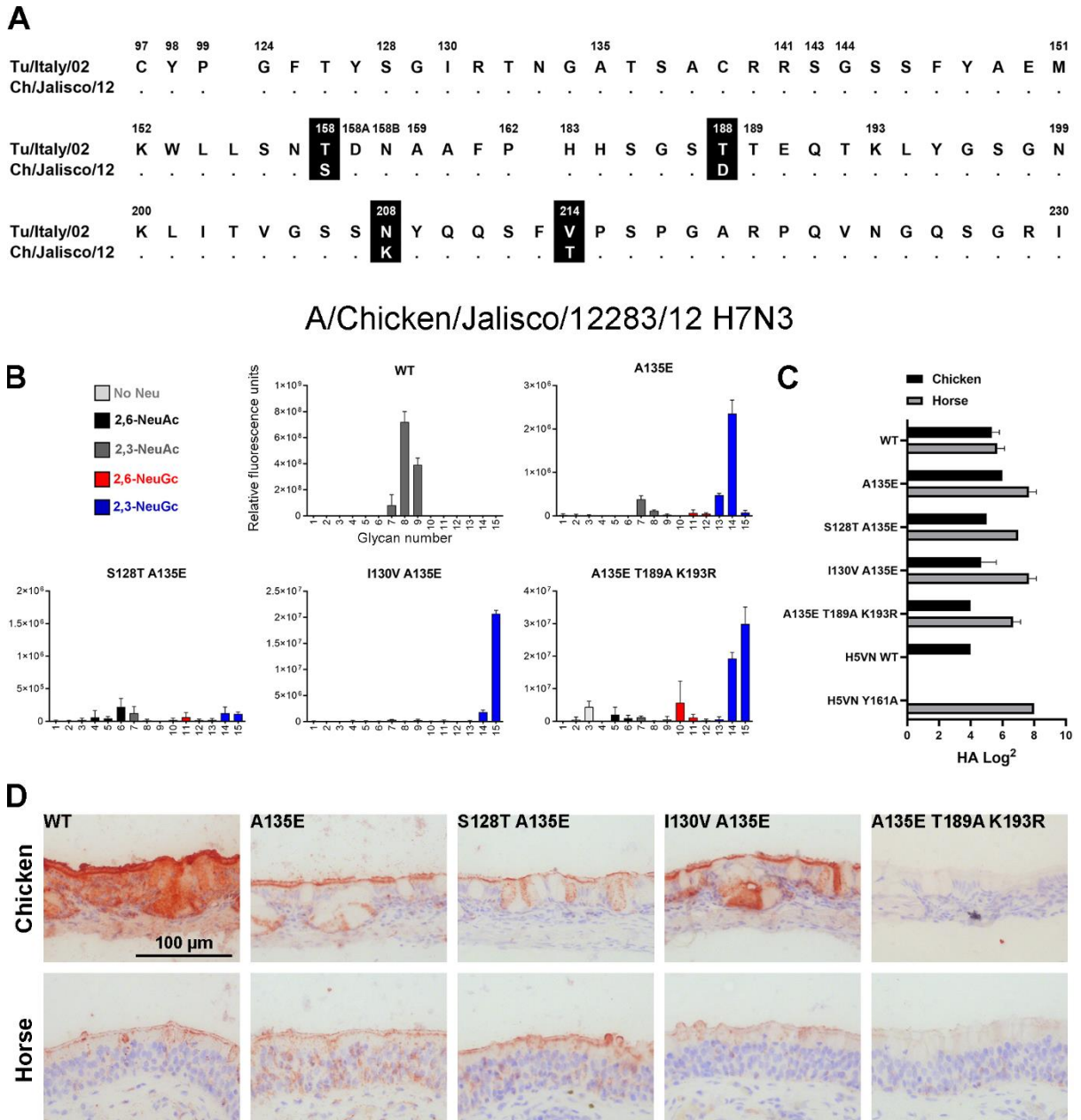
Since we did not observe exclusive NeuGc binding with the double and triple mutants, we combined the five mutations S128T, I130V, A135E, T189A, and K193R in the H7tu HA. With this mutant, similar results were achieved as previously, with NeuGc binding on the glycan array (Fig 7A) and binding to both horse and chicken erythrocytes (Fig 7B) and tracheal epithelium (Fig 7C).

The WT H7eq HA was previously shown to bind NeuGc on the glycan array [17] and here we showed that this WT HA also specifically bound horse erythrocytes (Fig 7E) and tracheal epithelium (Fig 7F). After the introduction of the five reciprocal mutations (T128S, V130I, E135A, A189T, R189K), the equine HA still seemed specific for NeuGc on the glycan array (Fig 7D) but started to bind to chicken erythrocytes and tracheal epithelium, while binding to equine tracheal epithelium was decreased. Since chickens are unable to produce NeuGc, this switch in binding indicated that the mutant H7Eq HA gained the ability to bind to NeuAc.

### **Equine and avian H7 strains are evolutionarily distant**

The fact that avian H7 HAs can be mutated towards binding NeuGc and equine H7 HAs towards binding NeuAc suggests that equine and avian H7 strains are phenotypically related. To investigate the genetic relationship between H7 strains, we reconstructed a maximum likelihood (ML) phylogenetic tree using HA sequences of equine H7 strains and their closest related Eurasian avian H7 strains (Fig 8A-E and S3). All equine strains cluster under a single monophyletic clade. Strains A/FPV/Dutch/1927 H7N7 and A/Fowl/Weybridge/1934 H7N7 appeared to be the closest related avian strains to the equine viruses. We investigated the natural variation in amino acids at positions for which binding specificity changed (128, 130, 135, 189, and 193).

For each selected amino acid position, we annotated the ML tree based on the variation in residues (Fig 8A-E). The predicted most recent common ancestor

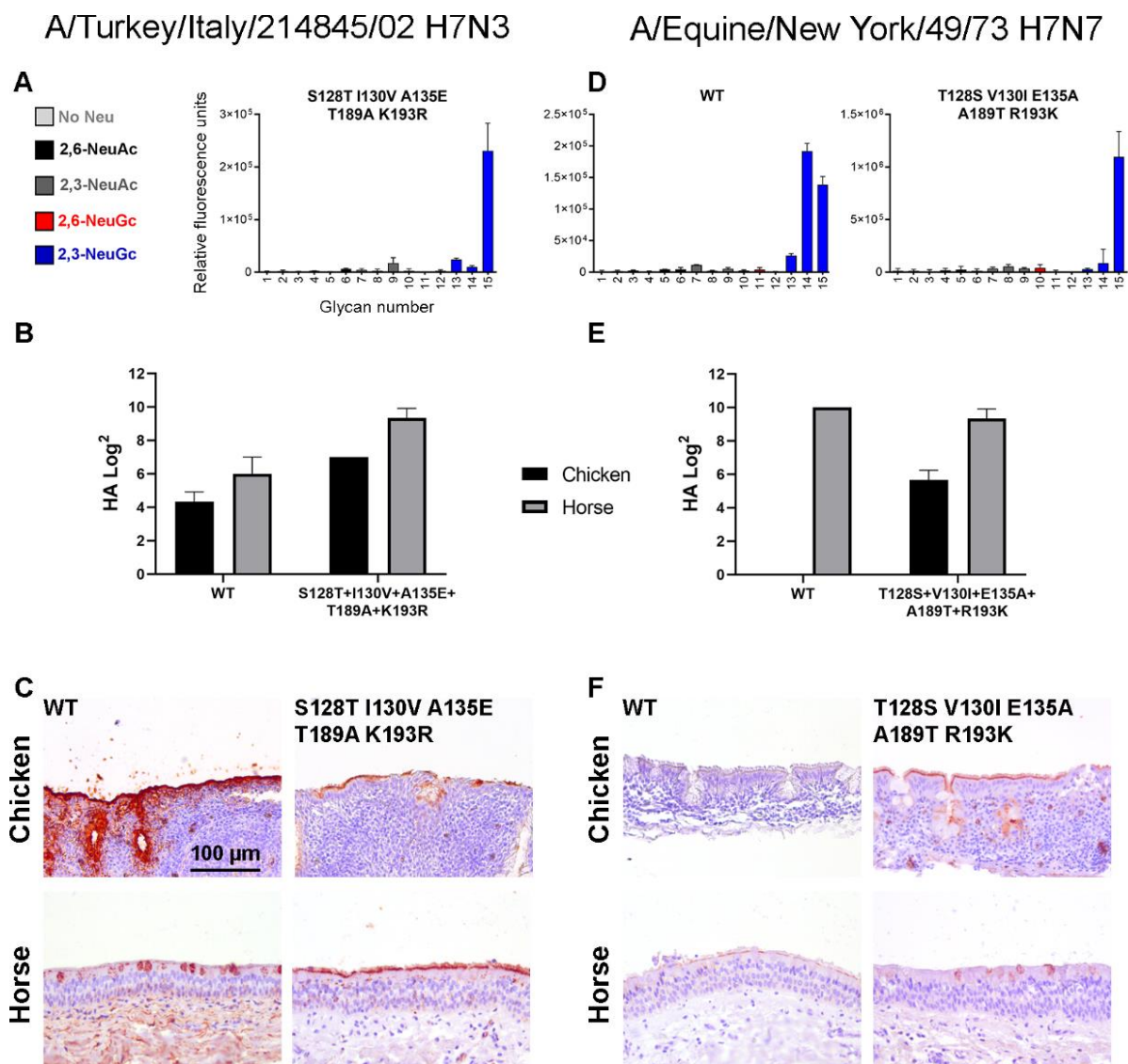


**Fig 6. Evaluation of the binding specificities of the (mutant) HA of A/Chicken/Jalisco/12283/12 H7N3.** (A) Alignment of the RBS of the HAs of A/Turkey/Italy/214845/02 H7N3 and A/Chicken/Jalisco/12283/12 H7N3, with the amino acid positions indicated above the alignment and dots indicating identical amino acids. A full alignment of the HAs is present in Fig S1. (B) The binding specificities of WT HA of A/Chicken/Jalisco/12283/12 H7N3 and mutant A135E, S128T+A135E, I130V+A135E, and A135E+T189A+K193R were evaluated on the glycan microarray as described in Fig 2A. (C) Binding specificities of WT and mutant HAs were furthermore tested in a hemagglutination assay on chicken and horse erythrocytes (n=3, mean and SD shown). (D) The binding of the WT and mutant HAs to chicken and horse tracheal epithelium (controls shown in Fig 5B) is visualized using AEC staining.

(MRCA) at all positions contained avian-like amino acids. Of the five investigated amino acid positions, the highest variability in amino acids is present at key position 135, although we observed a clear distinction between a glutamic acid in the equine strains and a variation of alanine, valine, and threonine in the avian strains



(Fig 8A). At position 128, there is an obvious distinction between the threonine in equine strains and mainly serine in the avian strains (Fig 8B). Again, a clear difference was observed at position 130 between avian strains (isoleucine) and equine strains (valine). Surprisingly, a closely related avian strain, A/Turkey/England/1963 H7N3, also contains a valine at position 130, just like the equine strains (Fig 8C). At position 189, all but one of the equine strains contain an alanine, while there is a variety of mainly alanine and threonine present in the avian strains (Fig 8D). At position 193, nearly all equine strains contain an arginine, whereas most avian strains, except for a small clade of viruses from chickens in Pakistan, contain a lysine (Fig 8E). We conclude that there is a clear distinction between the amino acids in the equine and avian strains, with the highest variability in residues being present at position 135, which we investigated further using targeted mutagenesis.



**Fig 7. The effect of the combination of the 5 mutations at positions 128, 130, 135, 189, and 193 in A/Turkey/Italy/214845/02 H7N3 and A/Equine/New York/43/73 H7N7 HA.** The binding specificity of the mutant HA (S128T, I130V, A135E, T189A, K193R) of A/Turkey/Italy/214845/02 H7N3 was evaluated using (A) the glycan microarray as described in Fig 2A, (B) hemagglutination assay with chicken and horse erythrocytes (n=3,

mean and SD shown), H5 NeuAc and NeuGc controls are shown in Fig 9, (C) and immunohistochemistry using chicken and horse tracheal epithelium (controls shown in Fig 5B), visualized using AEC staining. Likewise, the binding specificities of the WT and mutant (T128S, V130I, E135A, A189T, R189K) HA of A/Equine/New York/43/73 H7N7 were evaluated on using (D) the glycan microarray in which the WT was previously investigated already [17], (E) hemagglutination assay, (F) and tissue staining.

Four different amino acids (glutamic acid, alanine, valine, and threonine) are naturally present at position 135 of avian and equine H7 viruses (Fig 8A). When the alanine is encoded by either GCG or GCA, changing a single base pair will change the amino acid to glutamic acid, valine, or threonine. We introduced all these residues at position 135 of H7tu to investigate whether the acquirement of NeuGc binding was specific for the glutamic acid. This was indeed the case as the introduction of a threonine or a valine at position 135 did not promote binding to NeuGc (Fig 8F).

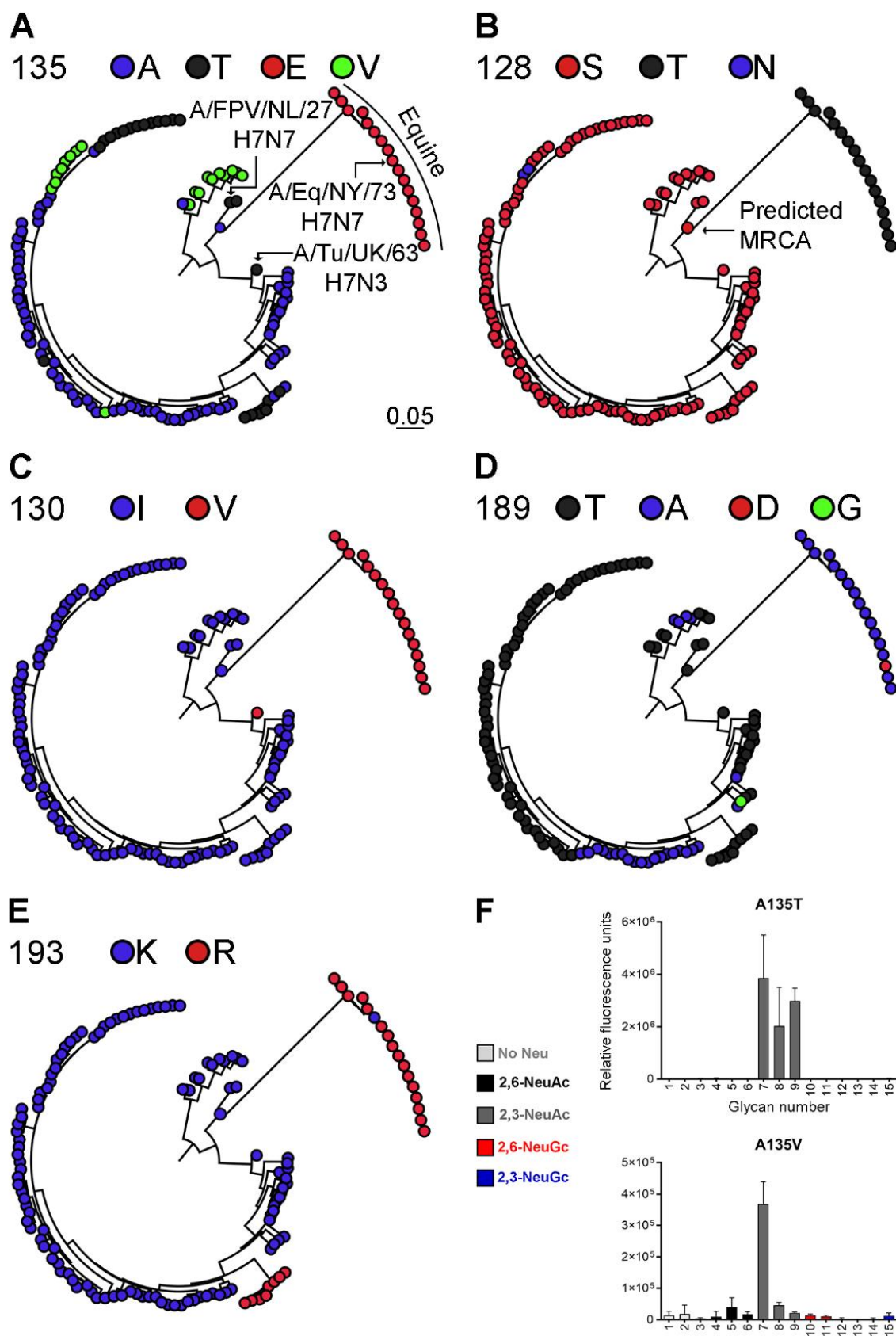
### H15 HA can also be switched to NeuGc binding

To further investigate the conservation of the switch to NeuGc binding due to mutations at positions 128, 130, 135, 189, and 193 in other subtypes, we investigated the receptor binding of the HA of the low pathogenic A/Duck/Australia/341/1983 H15N8 virus. H15 and H7 viruses are related and are present in one subgroup together with H10 viruses [21]. Nevertheless, there are 22 amino acid differences between the RBS of H7tu and this H15 (Fig 9A, full alignment in Fig S1), which is much more than the four different residues between the two distant avian H7 strains that we investigated (Fig 6A).

Like other avian H15 viruses [22, 33], we found that this WT H15 HA also bound  $\alpha$ 2,3-linked NeuAc on the glycan array (Fig 9B). Furthermore, the WT H15 HA bound both horse and chicken erythrocytes in the hemagglutination assay (Fig 9C), as was also observed for the two studied avian H7 HAs. However, the WT H15 HA only bound chicken and not horse tracheal epithelium (Fig 9D), which is different from the investigated avian H7 HAs.

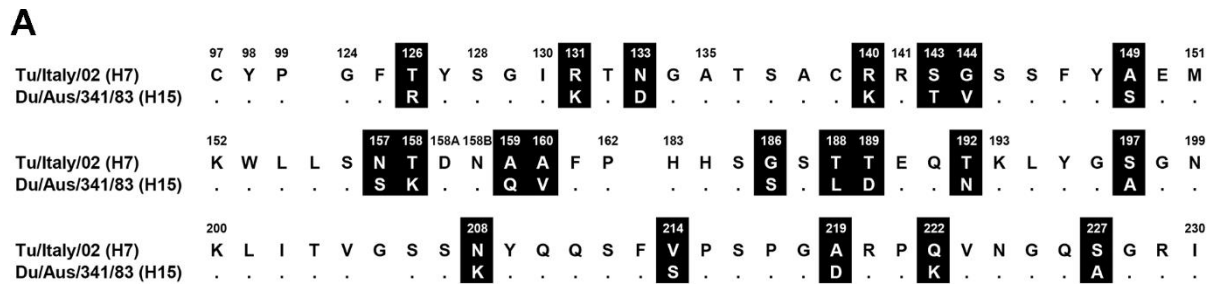
As soon as the key mutation A135E was introduced, the HA showed similar binding patterns as the avian H7 HAs with binding to both NeuAc and NeuGc and both chicken and horse erythrocytes and tissue (Fig 9B-D). The combination of A135E+S128T+I130V did not change the binding specificities compared to only A135E. When we however combined mutations A135E+D189A+K193R or all five mutations, the H15 HA became specific for NeuGc on the glycan array. Like in the avian H7 HAs, this was not observed in the hemagglutination assay and tissue staining.



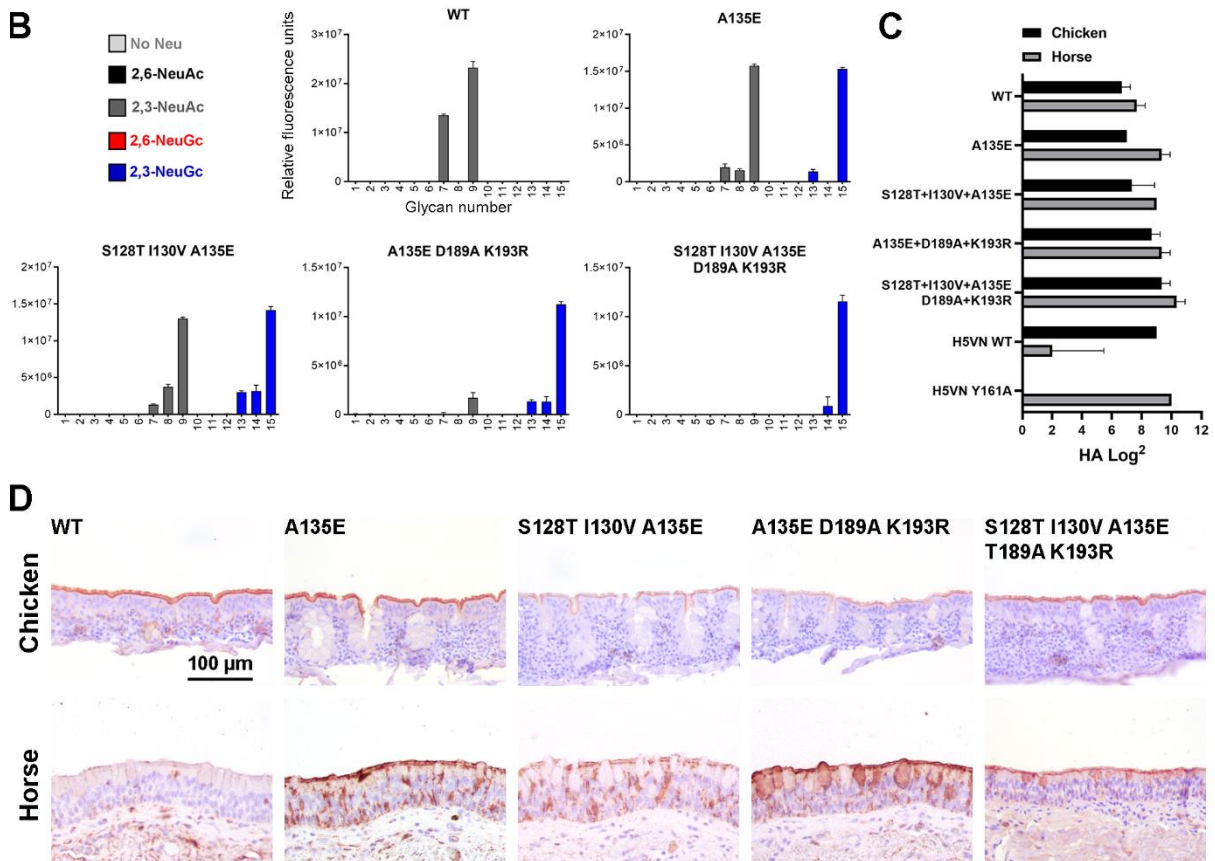


**Fig 8. Phylogenetic tree of equine and avian H7 HA sequences and evaluation of the binding specificities of mutants of the HA of A/Turkey/Italy/214845/02 H7N3 at amino acid position 135. Phylogenetic trees of equine and avian H7 IAV strains from the Eurasian**

lineage were reconstructed. The equine H7 strains cluster as a single monophyletic clade. The avian strains that are closest related to the equine strains (A/FPV/Dutch/1927 H7N7 and A/Turkey/England/1963 H7N3) are indicated, as well as A/Equine/New York/43/73 H7N7. The annotated phylogenetic tree with all strain names is displayed in Fig S2. The variation in amino acids at positions (A) 135 (alanine, threonine, glutamic acid, valine), (B) 128 (serine, threonine, asparagine), (C) 130 (isoleucine, valine), (D) 189 (threonine, alanine, aspartic acid, glycine) and (E) 193 (lysine, arginine) is shown. For all positions, the amino acid of the predicted most recent common ancestor (MRCA) is shown. (F) Representative binding specificities on the glycan microarray (as in Fig 2A) for H7tu mutants A135T and A135V are shown.



A/Duck/Australia/341/1983 H15N8



**Fig 9. Evaluation of the binding specificities of the (mutant) HA of A/Duck/Australia/341/1983 H15N8.** (A) Alignment of the RBS of the HAs of A/Turkey/Italy/214845/02 H7N3 and A/Duck/Australia/341/1983 H15N8, with the amino acid positions indicated above the alignment and dots indicating identical amino acids.

A full alignment of the HAs is present in Fig S1. **(B)** The binding specificities of WT HA of A/Duck/Australia/341/1983 H15N8 and mutant A135E, S128T+I130V+A135E, A135E+T189A+K193R, and S128T+I130V+A135E+T189A+K193R were evaluated on the glycan microarray as described in Fig 2A. **(C)** Binding specificities of WT and mutant HAs were furthermore tested in a hemagglutination assay on chicken and horse erythrocytes ( $n=3$ , mean and SD shown). **(D)** Binding of the WT and mutant HAs to chicken and horse tracheal epithelium (controls shown in Fig 5B) is visualized using AEC staining.

## Discussion

To elucidate the molecular determinants for NeuGc binding, we determined the crystal structure of the HA of A/Equine/New York/49/73 H7N7 in complex with its receptor analog 3'-GcLN (NeuGca2-3Gal $\beta$ 1-4GlcNAc). The overall RBS structures of H7eq and A/Turkey/Italy/214845/02 H7N3 were shown to be similar. To examine the critical amino acids for NeuGc binding, we performed mutational analysis on two distant avian H7 HAs and an avian H15 HA that specifically bound NeuAc. Previously, we demonstrated that HAs can bind either NeuAc or NeuGc [17]. Here, we demonstrate that avian H7 and H15 HAs can bind both NeuAc and NeuGc, by the introduction of A135E.

We previously studied the NeuAc-specific HA of A/Vietnam/1203/2004 H5N1 (H5VN) and its Y161A mutant that is specific for NeuGc, which showed complete specificity on the glycan microarray, in the hemagglutination assay [17], and on tracheal epithelium tissue (Fig 5C). Similarly, we found that WT H1, H2, and H4 HAs bind dog, but not horse, erythrocytes in the hemagglutination assay (Fig 3). Contrarily, we observed that WT avian H7 HAs bound both chicken (NeuAc [7, 8]) and horse (mainly NeuGc [12, 14, 15]) erythrocytes and tracheal tissue while binding specifically to NeuAc on the glycan microarray. These findings then distinguish these H7 viruses from other subtypes of IAV. Possibly, the presence of NeuAc on horse erythrocytes and tracheal tissue, although estimated to be less than 10% of the total sialic acids [12, 14, 15], was sufficient to be bound by the WT avian H7 HAs. Additionally, the residual NeuAc-binding capacity of the mutant avian H7 HAs may explain the binding to chicken erythrocytes and tissue. Importantly, not all compounds that are naturally present in the host are represented on the glycan microarray. Therefore, missing glycans on the array may explain the binding of avian H7 HAs to horse and chicken erythrocytes and tissue.

The natural variety in glycans in nature is massive and it is impossible to synthesize all of these glycans for our glycan microarray. For example, glycans can be elongated, either symmetrically or asymmetrically, be tri- or tetra-antennary, and contain one or multiple terminal sialic acids. Furthermore, the addition of fucose at different positions on LacNAc structures gives rise to different Lewis antigens and additional sulfate or O-acetyl groups can be present, adding another layer of complexity. Fucosylated (Lewis X) and sulfated glycans are present in the human lung [34, 35] and sulfated glycans have been observed in porcine lungs [36]. For equine and avian species, the glycans in the respiratory tract have not been studied in detail yet. One study describes the presence of sialyl Lewis X structures

in the respiratory tract of chickens [37]. Little is known about the glycans present on erythrocytes of different species, apart from two studies that describe that very few glycans with fucoses are present on chicken and mouse erythrocytes [38, 39]. IAV of different subtypes (H1, H3, H4, H5, H6, H7, H9, H13, H14) are known that (specifically) bind, or do not bind at all, to fucosylated and sulfated glycans [40-47]. Most relevant, avian, human, and seal H7 HAs also prefer to bind sulfated sialyl Lewis X structures [18, 41, 44]. In conclusion, fine receptor binding specificities regarding fucosylation and sulfation, which are observed in many different IAVs, may be present for the avian and equine H7 and H15 HAs besides NeuAc and NeuGc that are studied here.

It has been suggested that recognition of NeuGc by IAV is essential for viral replication in horses [12]. The most prevalent IAVs among horses currently and in the past are H3N8 and H7N7 viruses [48]. While equine H7N7 viruses have been shown to prefer binding to NeuGc [17, 18], the currently circulating equine H3N8 viruses bind to NeuAc [18], indicating that these viruses may not be under pressure to adapt to NeuGc binding since still small amounts of NeuAc are present in horses. These equine H3N8 viruses often infect dogs as well [48, 49], which are not able to make glycans containing NeuGc due to the lack of a functional CMAH. Furthermore, NeuAc binding could be advantageous for IAV to maintain circulation in NeuAc-rich reservoirs. Additionally, if NeuGc binding were required for replication of IAV in horses, more equine IAVs that bind NeuGc would be expected. Therefore, it seems unlikely that NeuGc recognition by IAV is essential for replication in horses.

Equine and avian H7 strains are estimated, by phylogenetic analysis, to have separated in the mid to late 1800s and separation between H7 and H15 viruses is estimated to have taken place in the early 1800s [50]. Nevertheless, here we demonstrated that avian H7 and H15 HAs, although genetically distinct from equine H7 viruses, are able to bind NeuGc after the introduction of mutation A135E. Further NeuGc specificity was obtained when adding mutations at positions 128, 130, 189, and 193. Reciprocal mutations in an equine H7 HA likewise resulted in the ability to bind NeuAc. These findings suggest that avian and horse H7 and H15 IAV are phenotypically related.

We further showed that a broad range of IAVs can bind NeuGc with the introduction of a few mutations. Furthermore, CMAH genes have been inactivated at several distinct events in evolution [7], causing the loss of NeuGc expression in different species over time. This loss of CMAH activity was potentially triggered by evolutionary pressure from lethal pathogens binding to NeuGc [16]. Therefore, we have taken the opportunity to consider NeuGc as a potential archaic receptor of IAV.

## Material and Methods

### Expression, crystallization, and structural determination of the equine H7 HA in complex with receptor analog 3'-GcLN

The HA ectodomain of A/Equine/New York/49/73 H7N7 (GenBank ID LC414434) was cloned and expressed as described previously [17]. Briefly, cDNA corresponding to residues 11 to 327 of HA1 and 1 to 179 of HA2 (H3 numbering) was cloned into a pFastbac vector. The HA was expressed in Hi5 insect cells as described [51], after which it was purified, the trimerization domain and His<sub>6</sub>-tag were removed, and the HA was concentrated to 6 mg/ml in 20 mM Tris, pH 8.0, 150 mM NaCl.

Crystals of the H7eq HA were obtained at 20°C using the vapor diffusion sitting drop method against a reservoir solution containing 32% (w/v) polyethylene glycol 400 and 0.1 M CAPS at pH 10. The complex of HA protein with 3'-GcLN was obtained by soaking HA crystals in a reservoir that contained 3'-GcLN to a final concentration of 10 mM for 1 hour at 20°C. The crystals were flash cooled in liquid nitrogen, without additional cryoprotectant, before x-ray data collection at the Advanced Photon Source (APS) (Table 1). Data integration and scaling were performed using HKL2000 [52]. Molecular replacement using Phaser [53] was used to solve the H7eq complex structure, for which an apo H7eq HA structure (PDB: 6N5A) was utilized as the search model. REFMAC5 [54] was used for structure refinement and modeling was done with COOT [55]. The final refinement statistics are outlined in Table 1.

### Expression and purification of HA for binding studies

HA encoding cDNAs of A/Turkey/Italy/214845/02 H7N3 [24] (synthesized and codon-optimized by GenScript), A/Chicken/Jalisco/12283/12 H7N3 (a kind gift from Florian Krammer, Mt Sinai Medical School), A/Duck/Australia/341/1983 H15N8 (a kind gift from Keita Matsuno), A/Equine/New York/49/73 H7N7 (a kind gift from Keita Matsuno), and A/Vietnam/1203/2004 H5N1 (synthesized and codon-optimized by GenScript) were cloned into the pCD5 expression vector as described previously [56, 57]. The pCD5 expression vector was adapted to clone the HA-encoding cDNAs in frame with DNA sequences coding for a secretion signal sequence, the Twin-Strep (WSHPQFEKGGGSGGGWSHPQFEK); IBA, Germany), a GCN4 trimerization domain (RMKQIEDKIEEIESKQKKIENEIARIKK), and a superfolder GFP [58] or mOrange2 [59]. Mutations in HAs were generated by site-directed mutagenesis. The HAs were purified from cell culture supernatants after expression in HEK293S GnTI(-) cells as described previously [56]. In short, transfection was performed using the pCD5 expression vectors and polyethyleneimine I. The transfection mixtures were replaced at 6 h post-transfection by 293 SFM II expression medium (Gibco), supplemented with sodium bicarbonate (3.7 g/L), Primatone RL-UF (3.0 g/L, Kerry, NY, USA), glucose (2.0 g/L), glutaMAX (1%, Gibco), valproic acid (0.4 g/L) and DMSO (1.5%). At 5 to 6 days after transfection, tissue culture supernatants were collected and Strep-Tactin sepharose beads (IBA, Germany) were used to purify the HA proteins according to the manufacturer's instructions.



### Glycan microarray binding of HA proteins

The glycan microarray as earlier presented [17] was utilized. HAs were pre-complexed with mouse anti-streptag-HRP and goat anti-mouse-Alexa555 antibodies in a 4:2:1 molar ratio respectively in 50  $\mu$ L PBS with 0.1% Tween-20. The mixture was incubated on ice for 15 min and afterward incubated on the surface of the array for 90 min in a humidified chamber. Then, slides were rinsed successively with PBS-T (0.1% Tween-20), PBS, and deionized water. The arrays were dried by centrifugation and immediately scanned as described previously [17]. Processing of the six replicates was performed by removing the highest and lowest replicate and subsequently calculating the mean value and standard deviation over the four remaining replicates. Supplemental dataset 1 presents a full dataset of the glycan microarray experiments.

### Hemagglutination assay

Hemagglutination assays were performed with pre-complexed HAs, as described for the glycan microarray, on 1.0% erythrocytes as previously described [56] with a starting concentration of 10  $\mu$ g/ml of HA for avian H7, H15, and H5 HAs. For equine H7 HAs, a starting concentration of 20  $\mu$ g/ml HA was used. Erythrocytes were provided by the Department of Equine Sciences and the Department of Farm Animal Health of the Faculty of Veterinary Medicine, Utrecht University, the Netherlands. The blood was taken from adult animals that are in the educational program of the Faculty of Veterinary Medicine. Complete datasets of the hemagglutination assays are present in supplemental dataset 2.

### Protein histochemistry

Sections of formalin-fixed, paraffin-embedded chicken (*Gallus gallus domesticus*) and equine (*Equus ferus caballus*) trachea were obtained from the Division of Pathology, Department of Biomolecular Health Sciences, Faculty of Veterinary Medicine of Utrecht University, the Netherlands. Tissues from three different horses and chickens were used in the assays to account for biological variation between individuals. In the figures, representative images of at least two individual experiments are shown. Protein histochemistry was performed as previously described [60, 61]. In short, tissue sections of 4  $\mu$ m were deparaffinized and rehydrated, after which antigens were retrieved by heating the slides in 10 mM sodium citrate (pH 6.0) for 10 min. Endogenous peroxidase was inactivated using 1% hydrogen peroxide in MeOH for 30 min at RT. Tissues were blocked overnight at 4°C using 3% BSA (w/v) in PBS and subsequently stained for 90 minutes using pre-complexed HAs as previously described for the glycan microarray. For avian H7 and H5 HA, 5  $\mu$ g/ml HA was used. For H15 HA, we used 2.5  $\mu$ g/ml HA and for equine H7 HA, we used 10  $\mu$ g/ml HA. After washing with PBS, binding was visualized using 3-amino-9-ethylcarbazole (AEC) (Sigma-Aldrich, Germany) and slides were counterstained using hematoxylin.

### Phylogenetic trees

All available, high-quality HA nucleotide sequences (i.e. sequence length is >90% of full-length HA gene segment and has <1% of ambiguous base) of avian H7Nx



and equine H7N7 influenza viruses dated between 1905 and 2005 from the NCBI Genbank database were downloaded (N=944). The maximum-likelihood phylogenetic tree was reconstructed using IQ-TREE [62] using the optimal nucleotide substitution model (i.e. GTR+F+R3) based on the Bayesian Information Criterion as determined by ModelFinder [63]. Ancestral sequences were reconstructed using treeTime [64].

### **Data analysis and statistical analysis**

The data in this article were analyzed and visualized using GraphPad Prism 9.2.0.

### **Acknowledgments**

We would like to thank the Department of Equine Sciences and the Department of Farm Animal Health of Utrecht University for supplying erythrocytes. We thank Andrea Gröne and Hélène Verheije from the Division of Pathology, Department of Biomolecular Health Sciences, Faculty of Veterinary Medicine of Utrecht University for providing paraffin-embedded tissues. We like to thank Eva Klaver and Nigel Kroone for their technical assistance.

### **Funding**

R.P.dV. is a recipient of an ERC Starting Grant from the European Commission (802780) and a Beijerinck Premium of the Royal Dutch Academy of Sciences. C.A.R. and A.X.H. are supported by an ERC Consolidator Grant from the European Commission (818353). Synthesis and microarray analyses were funded by a grant from the Netherlands Organization for Scientific Research (NWO TOPPUNT 718.015.003) to G.-J.B.. This work was funded in part by the Bill and Melinda Gates Foundation (OPP1170236) to I.A.W. X-ray data were collected at the beamline 23ID-D 9GM/CA CAT). The use of the APS was supported by the U.S. Department of Energy (DOE), Basic Energy Sciences, Office of Science, under contract DE-AC02-06CH11357.

### **Data deposition**

The atomic coordinates and structure factors of the HA of A/Equine/New York/49/73 H7N7 in complex with 3'-GcLN have been deposited in the Protein Data Bank (PDB) under accession code 7T1V.

The supplementary data to this chapter is available via <https://doi.org/10.1128/jvi.02120-21>.

### **Author contributions**

Protein crystallography, X.Z., I.A.W.; mutagenesis of hemagglutinins, C.M.S., M.R.C., R.vdW., M.M.T.L.; expression and production of hemagglutinins, C.M.S., I.T., M.R.C., R.vdW.; hemagglutination assays, C.M.S., R.P.dV., M.R.C.; protein histochemistry, C.M.S., K.M.B.; production of glycans, F.B.; preparation and analysis of glycan microarrays, C.M.S.; execution of glycan microarray experiments, R.P.dV.; alignments, C.M.S.; phylogenetic analysis; A.X.H., C.A.R.; data analysis and visualization, C.M.S.; conceptualization, C.M.S., R.P.dV.; supervision, R.P.dV., G.J.B.,

I.A.W., C.A.R., Y.S., K.M.; funding acquisition, R.P.dV., G.J.B., C.A.R., A.X.H.; writing—original draft preparation, C.M.S.; project administration, R.P.dV.; writing—review and editing, all authors.

## References

1. Skehel, J.J. and D.C. Wiley, *Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin*. *Annu Rev Biochem*, 2000. 69: p. 531-69.
2. Long, J.S., et al., *Host and viral determinants of influenza A virus species specificity*. *Nat Rev Microbiol*, 2019. 17(2): p. 67-81.
3. Ji, Y., et al., *New insights into influenza A specificity: an evolution of paradigms*. *Curr Opin Struct Biol*, 2017. 44: p. 219-231.
4. de Vries, R.P., et al., *Hemagglutinin receptor specificity and structural analyses of respiratory droplet-transmissible H5N1 viruses*. *J Virol*, 2014. 88(1): p. 768-73.
5. de Vries, R.P., et al., *Only two residues are responsible for the dramatic difference in receptor binding between swine and new pandemic H1 hemagglutinin*. *J Biol Chem*, 2011. 286(7): p. 5868-75.
6. de Vries, R.P., et al., *A single mutation in Taiwanese H6N1 influenza hemagglutinin switches binding to human-type receptors*. *EMBO Mol Med*, 2017. 9(9): p. 1314-1325.
7. Peri, S., et al., *Phylogenetic distribution of CMP-Neu5Ac hydroxylase (CMAH), the enzyme synthesizing the proinflammatory human xenantigen Neu5Gc*. *Genome Biol Evol*, 2018. 10(1): p. 207-219.
8. Schauer, R., et al., *Low incidence of N-glycolylneuraminic acid in birds and reptiles and its absence in the platypus*. *Carbohydr Res*, 2009. 344(12): p. 1494-500.
9. Ng, P.S., et al., *Ferrets exclusively synthesize Neu5Ac and express naturally humanized influenza A virus receptors*. *Nat Commun*, 2014. 5: p. 5750.
10. Yasue, S., et al., *Difference in form of sialic acid in red blood cell glycolipids of different breeds of dogs*. *J Biochem*, 1978. 83(4): p. 1101-7.
11. Spruit, C.M., et al., *N-glycolylneuraminic acid in animal models for human influenza A virus*. *Viruses*, 2021. 13: 815(5).
12. Suzuki, Y., et al., *Sialic acid species as a determinant of the host range of influenza A viruses*. *J Virol*, 2000. 74(24): p. 11825-31.
13. Suzuki, T., et al., *Swine influenza virus strains recognize sialylsugar chains containing the molecular species of sialic acid predominantly present in the swine tracheal epithelium*. *FEBS Lett*, 1997. 404(2-3): p. 192-6.
14. Suzuki, Y., M. Matsunaga, and M. Matsumoto, *N-acetylneuraminylactosylceramide, GM3-NeuAc, a new influenza A virus receptor which mediates the adsorption-fusion process of viral infection. Binding specificity of influenza virus A/Aichi/2/68 (H3N2) to membrane-associated GM3 with different molecular species of sialic acid*. *J Biol Chem*, 1985. 260(3): p. 1362-5.
15. Barnard, K.N., et al., *Modified sialic acids on mucus and erythrocytes inhibit influenza A virus hemagglutinin and neuraminidase functions*. *J Virol*, 2020. 94: e01567-19.(9).
16. Varki, A., *Loss of N-glycolylneuraminic acid in humans: mechanisms, consequences, and implications for hominid evolution*. *Am J Phys Anthropol*, 2001. Suppl 33: p. 54-69.
17. Broszeit, F., et al., *N-glycolylneuraminic acid as a receptor for influenza A viruses*. *Cell Rep*, 2019. 27(11): p. 3284-3294.e6.
18. Gambaryan, A.S., et al., *Receptor-binding profiles of H7 subtype influenza viruses in different host species*. *J Virol*, 2012. 86(8): p. 4370-9.
19. Webster, R.G., et al., *Evolution and ecology of influenza A viruses*. *Microbiol Rev*, 1992. 56(1): p. 152-79.
20. Webster, R.G., *Are equine 1 influenza viruses still present in horses?* *Equine Vet J*, 1993. 25(6): p. 537-8.
21. Liu, S., et al., *Panorama phylogenetic diversity and distribution of Type A influenza virus*. *PLoS One*, 2009. 4(3): p. e5022.
22. Tzarum, N., et al., *Unique structural features of influenza virus H15 hemagglutinin*. *J Virol*, 2017. 91(12).
23. Chen, V.B., et al., *MolProbity: all-atom structure validation for macromolecular crystallography*. *Acta Crystallogr D Biol Crystallogr*, 2010. 66(Pt 1): p. 12-21.
24. Russell, R.J., et al., *H1 and H7 influenza haemagglutinin structures extend a structural classification of haemagglutinin subtypes*. *Virology*, 2004. 325(2): p. 287-96.
25. de Vries, R.P., et al., *Three mutations switch H7N9 influenza to human-type receptor specificity*. *PLoS Pathog*, 2017. 13(6): p. e1006390.
26. Tzarum, N., et al., *The 150-loop restricts the host specificity of human H10N8 influenza virus*. *Cell Rep*, 2017. 19(2): p. 235-245.
27. Peng, W., et al., *Enhanced human-type receptor binding by ferret-transmissible H5N1 with a K193T mutation*. *J Virol*, 2018. 92(10): p. e02016-17.
28. Medeiros, R., et al., *Binding of the hemagglutinin from human or equine influenza H3 viruses to the receptor is altered by substitutions at residue 193*. *Arch Virol*, 2004. 149(8): p. 1663-71.
29. Watanabe, T., et al., *Characterization of H7N9 influenza A viruses isolated from humans*. *Nature*, 2013. 501(7468): p. 551-5.
30. Masuda, H., et al., *Substitution of amino acid residue in influenza A virus hemagglutinin affects recognition of sialyl-oligosaccharides containing N-glycolylneuraminic acid*. *FEBS Lett*, 1999. 464(1-2): p. 71-4.
31. Wang, M., et al., *Residue Y161 of influenza virus hemagglutinin is involved in viral recognition of sialylated complexes from different hosts*. *J Virol*, 2012. 86(8): p. 4455-62.
32. Wen, F., et al., *A Y161F hemagglutinin substitution increases thermostability and improves yields of 2009 H1N1 influenza A virus in cells*. *J Virol*, 2018. 92: e01621-17(2).
33. Yang, H., et al., *Molecular characterization and three-dimensional structures of avian H8, H11, H14, H15 and swine H4 influenza virus hemagglutinins*. *Heliyon*, 2020. 6(6): p. e04068.
34. Jia, N., et al., *The human lung glycome reveals novel glycan ligands for influenza A virus*. *Sci Rep*, 2020. 10(1): p. 5320.

35. Sriwilaijaroen, N., et al., *N-glycan structures of human alveoli provide insight into influenza A virus infection and pathogenesis*. FEBS J, 2018. 285(9): p. 1611-1634.
36. Byrd-Leotis, L., et al., *Shotgun glycomics of pig lung identifies natural endogenous receptors for influenza viruses*. Proc Natl Acad Sci U S A, 2014. 111(22): p. E2241-50.
37. Hiono, T., et al., *A chicken influenza virus recognizes fucosylated alpha2,3 sialoglycan receptors on the epithelial cells lining upper respiratory tracts of chickens*. Virology, 2014. 456-457: p. 131-8.
38. Hua, S., et al., *Isomer-specific LC/MS and LC/MS/MS profiling of the mouse serum N-glycome revealing a number of novel sialylated N-glycans*. Anal Chem, 2013. 85(9): p. 4636-43.
39. Aich, U., et al., *Glycomics-based analysis of chicken red blood cells provides insight into the selectivity of the viral agglutination assay*. FEBS J, 2011. 278(10): p. 1699-712.
40. Gambaryan, A., et al., *Evolution of the receptor binding phenotype of influenza A (H5) viruses*. Virology, 2006. 344(2): p. 432-8.
41. Gambaryan, A.S., et al., *6-sulfo sialyl Lewis X is the common receptor determinant recognized by H5, H6, H7 and H9 influenza viruses of terrestrial poultry*. Virol J, 2008. 5: p. 85.
42. Bateman, A.C., et al., *Glycan analysis and influenza A virus infection of primary swine respiratory epithelial cells: the importance of NeuAc{alpha}2-6 glycans*. J Biol Chem, 2010. 285(44): p. 34016-26.
43. Stevens, J., et al., *Receptor specificity of influenza A H3N2 viruses isolated in mammalian cells and embryonated chicken eggs*. J Virol, 2010. 84(16): p. 8287-99.
44. Gambaryan, A., et al., *Receptor specificity of influenza viruses from birds and mammals: new data on involvement of the inner fragments of the carbohydrate chain*. Virology, 2005. 334(2): p. 276-83.
45. Stevens, J., et al., *Glycan microarray analysis of the hemagglutinins from modern and pandemic influenza viruses reveals different receptor specificities*. J Mol Biol, 2006. 355(5): p. 1143-55.
46. Hiono, T., et al., *Amino acid residues at positions 222 and 227 of the hemagglutinin together with the neuraminidase determine binding of H5 avian influenza viruses to sialyl Lewis X*. Arch Virol, 2016. 161(2): p. 307-16.
47. Wen, F., et al., *Mutation W222L at the receptor binding site of hemagglutinin could facilitate viral adaption from equine influenza A(H3N8) virus to dogs*. J Virol, 2018. 92(18).
48. Murcia, P.R., J.L. Wood, and E.C. Holmes, *Genome-scale evolution and phylodynamics of equine H3N8 influenza A virus*. J Virol, 2011. 85(11): p. 5312-22.
49. Collins, P.J., et al., *Recent evolution of equine influenza and the origin of canine influenza*. Proc Natl Acad Sci U S A, 2014. 111(30): p. 11175-80.
50. Worobey, M., G.Z. Han, and A. Rambaut, *A synchronized global sweep of the internal genes of modern avian influenza virus*. Nature, 2014. 508(7495): p. 254-7.
51. Stevens, J., et al., *Glycan microarray technologies: tools to survey host specificity of influenza viruses*. Nat Rev Microbiol, 2006. 4(11): p. 857-64.
52. Otwinowski, Z. and W. Minor, *Processing of X-ray diffraction data collected in oscillation mode*. Methods Enzymol, 1997. 276: p. 307-26.
53. McCoy, A.J., et al., *Likelihood-enhanced fast translation functions*. Acta Crystallogr D Biol Crystallogr, 2005. 61(Pt 4): p. 458-64.
54. Murshudov, G.N., et al., *REFMAC5 for the refinement of macromolecular crystal structures*. Acta Crystallogr D Biol Crystallogr, 2011. 67(Pt 4): p. 355-67.
55. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics*. Acta Crystallogr D Biol Crystallogr, 2004. 60(Pt 12 Pt 1): p. 2126-32.
56. de Vries, R.P., et al., *The influenza A virus hemagglutinin glycosylation state affects receptor-binding specificity*. Virology, 2010. 403(1): p. 17-25.
57. Zeng, Q., et al., *Structure of coronavirus hemagglutinin-esterase offers insight into corona and influenza virus evolution*. Proc Natl Acad Sci U S A, 2008. 105(26): p. 9065-9.
58. Nemanichvili, N., et al., *Fluorescent trimeric hemagglutinins reveal multivalent receptor binding properties*. J Mol Biol, 2019. 431(4): p. 842-856.
59. Shaner, N.C., et al., *Improving the photostability of bright monomeric orange and red fluorescent proteins*. Nat Methods, 2008. 5(6): p. 545-51.
60. Bouwman, K.M., et al., *Three amino acid changes in avian coronavirus spike protein allow binding to kidney tissue*. J Virol, 2020. 94(2).
61. Wickramasinghe, I.N., et al., *Binding of avian coronavirus spike proteins to host factors reflects virus tropism and pathogenicity*. J Virol, 2011. 85(17): p. 8903-12.
62. Minh, B.Q., et al., *IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era*. Mol Biol Evol, 2020.
63. Kalyaanamoorthy, S., et al., *ModelFinder: fast model selection for accurate phylogenetic estimates*. Nat Methods, 2017. 14(6): p. 587-589.
64. Sagulenko, P., V. Puller, and R.A. Neher, *TreeTime: maximum-likelihood phylodynamic analysis*. Virus Evol, 2018. 4(1): p. vex042.

# Chapter 5

## H7 influenza A viruses bind sialyl-LewisX, a potential intermediate receptor between species

Cindy M. Spruit, Diana I. Palme, Tiehai Li, María Ríos Carrasco, Alba Gabarroca García, Igor R. Sweet, Maryna Kuryshko, Joshua C. L. Maliepaard, Karli R. Reiding, David Scheibner, Geert-Jan Boons, Elsayed M. Abdelwhab, and Robert P. de Vries

Published: *Journal of Virology*, March 2024, DOI: <https://doi.org/10.1128/jvi.01941-23>

### Abstract

Influenza A viruses (IAVs) can overcome species barriers by adaptation of the receptor binding site of the hemagglutinin (HA). To initiate infection, HAs bind to glycan receptors with terminal sialic acids, which are either *N*-acetylneuraminic acid (NeuAc) or *N*-glycolylneuraminic acid (NeuGc), the latter is mainly found in horses and pigs but not in birds and humans. We investigated the influence of previously identified equine NeuGc-adapting mutations (S128T, I130V, A135E, T189A, and K193R) in avian H7 IAVs *in vitro* and *in vivo*. We observed that these mutations negatively affected viral replication in chicken cells, but not in duck cells, and positively affected replication in horse cells. *In vivo*, the mutations reduced virus virulence and mortality in chickens. Ducks excreted high viral loads for a longer time than chickens, although they appeared clinically healthy. To elucidate why chickens and ducks were infected by these viruses despite the absence of NeuGc, we re-evaluated the receptor binding of H7 HAs using glycan microarray and flow cytometry studies. This revealed that mutated avian H7 HAs also bound to  $\alpha$ 2,3-linked NeuAc and sialyl-LewisX, which have an additional fucose moiety in their terminal epitope, explaining why infection of ducks and chickens was possible. Interestingly, the  $\alpha$ 2,3-linked NeuAc and sialyl-LewisX epitopes were only bound when presented on tri-antennary *N*-glycans, emphasizing the importance of investigating the fine receptor specificities of IAVs. In conclusion, the binding of NeuGc-adapted H7 IAV to sialyl-LewisX enables viral replication and shedding by chickens and ducks, potentially facilitating interspecies transmission of equine-adapted H7 IAVs.

## Introduction

Influenza A viruses (IAVs) are a member of the virus family *Orthomyxoviridae* and their proteins are encoded on eight single-stranded negative-sense RNA segments with a total length of 12-14kb [1]. The enveloped virion of IAVs is coated with the surface proteins hemagglutinin (HA) and neuraminidase, which allow the classification into different subtypes (HxNx). IAVs infect a variety of avian and mammalian species, including humans, pigs, and horses [2]. The natural reservoirs for IAVs are wild waterfowl, but transmission from ducks to other susceptible avian and mammalian species is frequent [3, 4]. Avian IAV infection in wild birds is often asymptomatic, due to the coevolution of IAV and wild birds [5]. However, in poultry low pathogenicity avian influenza viruses (LPAIV) can evolve into high pathogenicity avian influenza virus (HPAIV) causing mortality rates up to 100% in infected flocks. One of the key determinants for the virulence and pathogenicity of HPAIV is the acquisition of a multibasic cleavage site in the HA, which is most common in H5 and H7 IAVs [6].

High pathogenicity H7 IAVs are occasionally transmitted to humans and other mammalian species [7-9]. Equine H7N7 influenza A viruses contain a multibasic cleavage site and are suspected to have originated from an avian H7 ancestor virus from an HPAIV outbreak in poultry [10]. Furthermore, reassortant viruses with the equine H7N7 HA and other genes from a chicken H5N2 IAV were shown to be lethal in chickens [11]. Nowadays, equine H7N7 viruses are thought to be extinct, leaving equine H3N8 as the only active circulating equine influenza virus [12, 13]. The presence of H7 IAVs in different species emphasizes the relevance of further investigating these viruses and their interspecies transmission.

Overcoming host species barriers and establishing species-specific influenza strains involves the accumulation of point mutations during adaptation [14-17]. The main host species barrier of IAVs is the receptor binding specificity of HA to terminal sialic acid (Sia) epitopes on the host cell surface, which is important for virus uptake into the cell [18]. Receptor binding of HAs is strain-specific and has co-evolved with receptors found in the respiratory and/or intestinal tract of susceptible host species. Therefore, avian influenza viruses (AIV) bind preferentially to  $\alpha$ 2,3-linked Sia, whereas human-adapted strains prefer  $\alpha$ 2,6-linked Sia receptors [7, 19-21].

Besides the glycosidic linkage, the host cell receptor's type of terminating Sia and the underlying glycan structures are important factors in IAV receptor binding properties and host range [14, 22-25]. Unlike the majority of IAV, which predominantly bind to glycans with a terminal *N*-acetylneuraminic acid (NeuAc), equine H7N7 IAV predominantly bind to the *N*-glycolylneuraminic acid (NeuGc) [26, 27]. Levels of NeuGc are variably present in the respiratory tract of most mammalian species, especially horses and pigs. However, no NeuGc is expressed in, among others, birds, humans, and ferrets [28-32]. Previously, we identified five mutations S128T, I130V, A135E, T189A, and K193R in the receptor binding site (RBS), based on an equine H7N7 virus, that switched avian H7 IAVs from binding NeuAc to NeuGc [33].



In this study, we examined the impact of the equine NeuGc-adapting mutations (S128T, I130V, A135E, T189A, and K193R) on avian H7 IAVs both *in vitro* and *in vivo*, with particular emphasis on economically important poultry (chickens) and natural reservoir bird (ducks). The mutated viruses showed reduced replication in chicken cells, however, the replication in duck cells remained unaffected. On the other hand, viral replication in horse cells was increased. *In vivo*, the NeuGc-adapted viruses showed reduced mortality and virulence in infected chickens compared to the WT HPAIV, but the viral distribution between chicken organs was mostly unaffected by the mutations. However, virus shedding was higher in cloacal swab samples and ducks excreted high viral loads for a longer time than chickens, although they did not show symptoms of disease. Avian wild-type and mutant H7 hemagglutinins bound to both  $\alpha$ 2,3-linked NeuAc and sialyl-LewisX epitopes (an  $\alpha$ 2,3-linked NeuAc presented on an N-acetylglucosamine (LacNAc) with an additional fucose moiety  $\alpha$ 1,3-linked to the N-acetylglucosamine of the LacNAc), but only when presented in complex N-glycans. These findings improve the understanding of equine-specific adaptations in avian H7 receptor binding, viral replication, and pathogenicity while assessing the potential of interspecies transmission of these viruses.

## Results

### NeuGc-specific mutations have differential effects in chicken, duck, and horse cells

Previously, we investigated the molecular determinants for binding of avian H7 IAVs to NeuGc and found that five amino acids that are abundant in equine H7 viruses (I28T, I30V, I35E, I89A, and I93R) were responsible for binding to NeuGc, a common sialic acid in horses [33]. Curiously, we observed that these mutations in the H7 HA of A/turkey/Italy/214845/02 switched the receptor binding specificity from  $\alpha$ 2,3-linked NeuAc to  $\alpha$ 2,3-linked NeuGc, but did not cause a loss of binding to chicken trachea and erythrocytes, which do not contain NeuGc [29, 31]. This observation raised the question of whether the infection capabilities of avian viruses with these equine NeuGc-specific mutations would be affected.

To investigate whether the NeuGc-specific mutations would affect the viral fitness of an avian virus *in vitro*, we rescued A/chicken/Germany/R28/2003 as a wild-type (WT) H7N7 HPAIV (designated H7N7\_avHA) and a mutant of this virus carrying the five NeuGc-specific mutations in the HA (designated H7N7\_5eqHA). The RBS of H7N7\_avHA is identical to the RBS of A/turkey/Italy/214845/02, which we used in our previous publication [33]. Sequence analysis of avian and equine H7 sequences showed that all five amino acid residues are highly conserved in equine H7 IAVs (97-100%), but also naturally occur in some of the analyzed avian H7 HA sequences (Table S1).

The impact of the equine-specific residues on cell-to-cell spread and viral growth kinetics was investigated in various cell lines (Fig. 1A-E). The viruses' ability to spread from one cell to another was analyzed in a plaque assay using MDCKII cells, which are the most commonly used cells for IAV replication assays [34]. The NeuGc-specific residues significantly reduced the intercellular spread at 72 hours post-

infection (hpi) in MDCKII cells (Fig. 1A) and the replication in these cells (Fig. 1B). The NeuGc-specific mutations significantly increased viral replication in equine lung cells (PLU-R) and equine epidermal cells (E.Derm) 24 hpi (Fig. 1C). In contrast, H7N7\_5eqHA replication in chicken fibroblasts, primary CEK cells, and SPF eggs was significantly reduced compared to WT H7N7\_avHA (Fig. 1D). However, no significant differences between replication of H7N7\_avHA and H7N7\_5eqHA were observed in duck embryo fibroblast cells 24 hpi (Fig. 1E). These findings indicate that the equine-specific amino acid residues in the RBS of H7 HPAIV reduced cell-to-cell spread and viral replication in a host-dependent manner. They increased replication of the avian virus in horse cells and reduced replication in chicken cells, but replication in duck cells was unaffected.

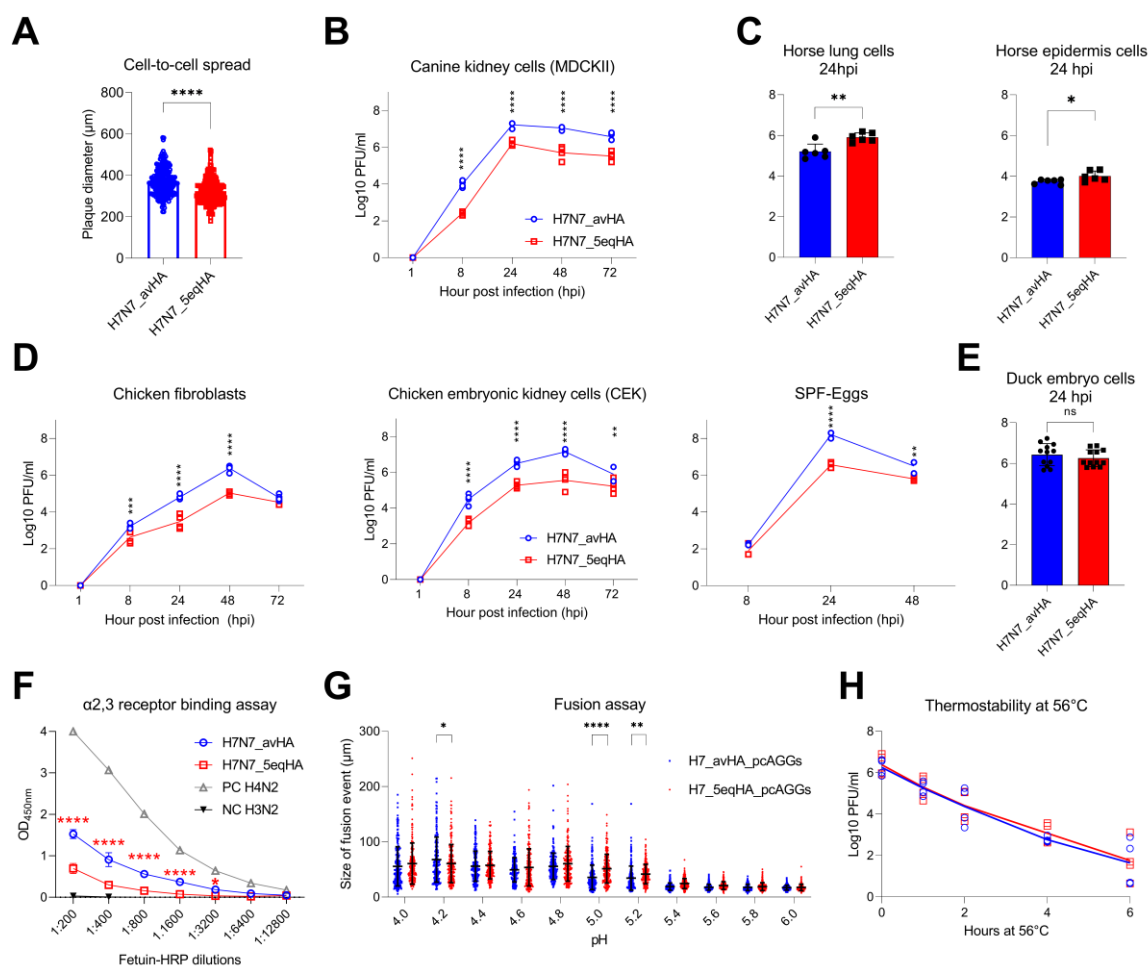
### **NeuGc-specific mutations reduced the binding affinity to $\alpha$ 2,3-NeuAc without a significant impact on the pH-dependent HA activation and thermostability**

*In vitro* characterization of generated H7N7 viruses in cell culture revealed that NeuGc-specific mutations affected viral replication and spread in a host-dependent manner (Fig. 1A-E). To ascertain whether these mutations have an influence on the viruses' biological properties and thus on viral replication, the receptor binding properties, thermostability, and pH activation were tested.

To assess whether the introduction of the NeuGc-specific mutations in the RBS changed the binding affinity to  $\alpha$ 2,3-NeuAc, a solid-phase assay using  $\alpha$ 2,3-linked fetuin as a substrate was performed (Fig. 1F) as previously described [26, 35]. An avian H4N2 virus was used as a positive control for  $\alpha$ 2,3-NeuAc binding and a human H3N2 IAV was used as a negative control. We observed that H7N7\_5eqHA had a significantly lower binding affinity to  $\alpha$ 2,3-NeuAc than the WT H7N7\_avHA virus (Fig. 1F), which is in accordance with our previously obtained results [33]. In addition to affecting receptor binding properties, mutations in HA1 may affect pH activation of hemagglutinin and subsequently affect the replication of viruses such as AIV H5N1 [53], although there is limited knowledge of how it affects AIV H7N7. Therefore, we assessed the potential influence of the five mutations on the pH-dependent fusion-HA activation by measuring the diameter of cell-to-cell fusion after transfection of avian cells (QM9) with protein expression vectors (pCAGGS) carrying HA from H7N7\_avHA (H7\_avHA\_pcAGGs) or H7N7\_5eqHA (H7\_5eqHA\_pcAGGs) (Fig. 1G). Both hemagglutinins were activated at a broad range of pH values from 4.0 to 6.0. However, the fusion efficiency of H7\_5eqHA\_pcAGGs in QM9 cells at a pH of 5.0 and 5.2 was significantly higher than that of H7\_avHA\_pcAGGs. Interestingly, the pH-dependent activation of H7\_avHA\_pcAGGs was found to be significantly higher at a pH value of 4.2 than H7\_5eqHA\_pcAGGs. The size of fusion events at other pH values was comparable (Fig. 1G).

The thermostability of the HA is known to be linked to virulence in different influenza strains [36]. Therefore, we evaluated the thermostability of the two viruses at 56°C for 2, 4, and 6 hours, a standard treatment for enveloped viruses [37]. Both viruses lost infectivity at comparable levels indicating that the introduction of equine-specific amino acids did not affect the thermostability of the HPAIV (Fig. 1H). In conclusion, the NeuGc-specific mutations reduced the replication of this H7N7

HPAIV in the chicken cells probably due to reduced binding affinity to the 2,3-NeuAc without significantly impacting the HA pH-dependent activation and thermostability of the viruses.



**Fig 1. *In vitro* characterization of WT (H7N7\_avHA) and mutant (H7N7\_5eqHA) A/chicken/Germany/R28/2003 H7N7 viruses.** (A) Cell-to-cell spread was investigated by measuring the diameter of about 100 plaques in MDCKII cells. (B) Viral replication at indicated time points was assessed in MDCKII, (C) horse lung and horse epidermal cells, (D) chicken fibroblasts (DF-1), primary chicken cells (CEK), SPF embryonated chicken eggs (ECE), and (E) in duck embryo fibroblast cells. (F) The receptor binding affinity to  $\alpha$ 2,3-linked NeuAc was measured using  $\alpha$ 2,3-Sia fetuin substrate. Human H3N2 virus (specific for  $\alpha$ 2,6-linked NeuAc) was used as a negative control (NC). An avian H4N2 virus (specific for  $\alpha$ 2,3-linked NeuAc) was used as a positive control (PC). Shown are representative results calculated as means and standard deviations of three independent experiments, each was run in duplicates. (G) pH-dependent activation of HA in a fusion assay was measured after transfection of quail cells (QM-9) with pCAGGS protein expression vector containing the HA of H7N7\_avHA or H7N7\_5eqHA. Cells were simultaneously transfected with pCAGGS carrying eGFP to facilitate the evaluation of the assay. Cell fusion was triggered 24 hpi with PBS of different pH values for two minutes. The diameter of syncytia was measured using Eclipse Ti-S with software NIS-Elements, version 4.0; Nikon. (H) The thermostability of viruses was measured in duplicates and repeated twice. The reduction in virus infectivity at indicated time points was assessed by titration of heated viruses using a plaque test in MDCKII cells and expressed as plaque-forming units per ml ( $\text{Log}_{10}$  PFU/ml). All results are expressed as means and standard deviations of at least two independent

experiments run in duplicates. Asterisks indicate statistical significance based on p values: \* ≤ 0.05, \*\* ≤ 0.01, \*\*\* ≤ 0.001, \*\*\*\* ≤ 0.0001.

### The NeuGc-specific residues significantly reduced virulence in infected chickens, but had no impact on virus virulence in ducks

Since the NeuGc-specific mutations had an impact on receptor binding, cell-to-cell transmission, and viral replication in chicken, but not duck cells (Fig. 1), these mutations potentially also affect the viruses *in vivo*. Therefore, we performed an animal experiment in chickens, from which the H7N7\_avHA was originally isolated and which are economically crucial hosts, and ducks as the natural reservoir species of AIVs.

Nine SPF chickens and ten Pekin ducks were infected by intravenous (IV) or intramuscular (IM) injection, respectively with H7N7\_avHA or H7N7\_5eqHA to assess the viral pathogenicity index (PI) according to the WOAHA standard [38]. All ducks infected IM with H7N7\_avHA or H7N7\_5eqHA survived the animal experiment and showed no clinical disorders (Table 1). Nevertheless, all ducks seroconverted indicating a successful infection. The intramuscular pathogenicity index (IMPI) for H7N7\_avHA and H7N7\_5eqHA in ducks was determined to be 0.0. Conversely, chickens that were IV-infected with H7N7\_avHA died within a mean time of death (MDT) of 4.6 days post-infection (dpi). All chickens displayed clinical signs of infection and an intravenous pathogenicity index (IVPI) of 2.4 was calculated. H7N7\_avHA is therefore classified as an HPAIV according to the WOAHA classification (IVPI > 1.2 indicates an AIV as HPAIV). Interestingly, the introduction of the five equine mutations into the avian HA reduced mortality to 3 out of 9 IV-infected chickens with H7N7\_5eqHA, but all chickens exhibited transient mild to moderate clinical signs. Notably, IV-infected chickens with H7N7\_5eqHA died earlier compared to those IV-infected with H7N7\_avHA, with an MDT of 3.0 days (Table 1), although the differences in MDT of both groups were not statistically significant. Nevertheless, the IVPI of H7N7\_5eqHA inoculated chickens was determined to be 1.5, which is still classified as HPAIV. All remaining chickens subsequently developed antibodies.

**Table 1. The pathogenicity indices of H7N7 viruses after the injection of chickens and ducks.** The table shows the mortality, morbidity, and seroconversion of intravenous infected (IV) chickens and intramuscular infected (IM) Pekin ducks. In addition, the calculated mean time of death (MDT) in days post-infection (dpi) and intravenous (IVPI) or intramuscular (IMPI) pathogenicity indices are shown. \*n.a. = not applicable.

Virus	Animal	Group	Mortality	Mean time of death (MDT)	Morbidity	Pathogenicity index (PI)	Seropositive birds/total birds
H7N7_avHA	Chickens	IV infected	9/9	4.6	9/9	IVPI: 2.4	n.a
	Ducks	IM infected	0/10	n.a.*	0/10	IMPI: 0.0	10/10
H7N7_5eqHA	Chickens	IV infected	3/10	3.0	10/10	IVPI: 1.5	7/7
	Ducks	IM infected	0/10	n.a.*	0/10	IMPI: 0.0	10/10

In the second animal experiment, we wanted to mimic the natural course of infection. Therefore, ten chickens and ten ducks were inoculated by the ocular-nasal (ON) route. Furthermore, at 1 dpi five chickens and five ducks were

added to each group to assess chicken-to-chicken or duck-to-duck transmission. All chickens primarily inoculated with H7N7\_avHA died within an MDT of 5.7 dpi, with an average CS of 1.7 (Fig. 2A, Table 2). Only one contact chicken died on the eighth dpi in the avian H7N7 (H7N7\_avHA) infected group. However, all contact chickens displayed signs of morbidity (Table 2). Six out of ten chickens infected with H7N7\_5eqHA died with an MDT of 4.5 days. In this group, three out of five contact chickens died with MDT of 7.3 days, and the two remaining chickens showed transient mild to moderate clinical signs. The average PI was 1.4 for the primarily inoculated chickens (Fig. 2A, Table 2). Conversely, and similar to the IM-injected ducks, neither the ON-inoculated nor the contact ducks in either group displayed any signs of illness. All ducks survived until the end of the animal trial (Table 2, Fig. 2B). Seropositive results were recorded for all remaining chickens and ducks using an anti-NP ELISA (Table 2). These findings confirm the results from the IVPI and IMPI-infected birds, as H7N7\_5eqHA is less lethal in chickens than H7N7\_avHA and ducks are not affected at all. Taken together, and regardless of the infection routes, the equine-adaptive mutations reduced HPAIV H7N7 virulence in chickens, while ducks were clinically resistant to both viruses.

**Table 2. *In vivo* data from oculonasal infected chickens and Pekin ducks.** The table shows the sick (morbidity) and dead (mortality) animals per group, as well as the results from sera analysis in a competitive NP ELISA. Calculated mean time to death (MDT) in days post-infection as well as the clinical score (CS) are shown. \*n.a. = not applicable.

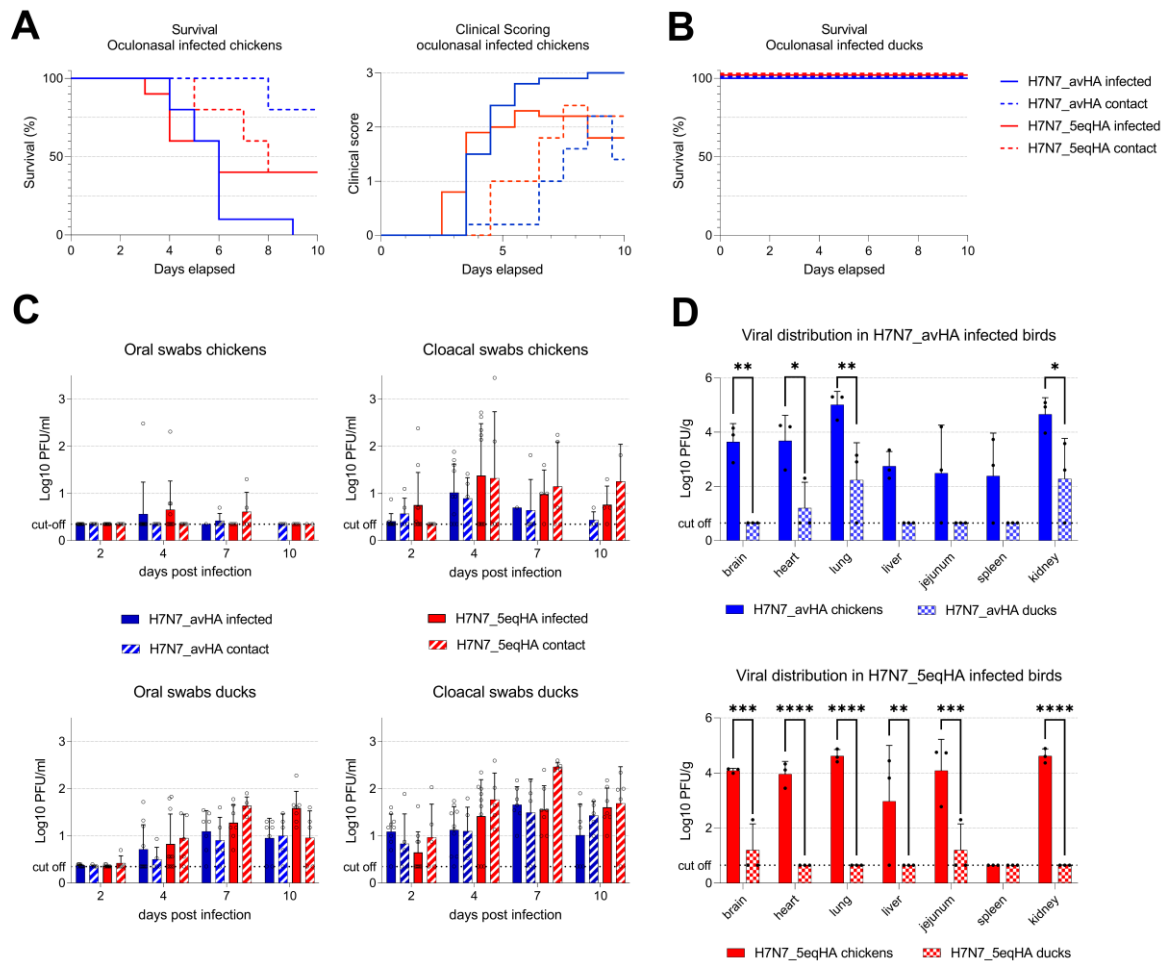
Virus	Animal	Group	Mortality	Mean time of death (MDT)	Morbidity	Clinical Score (CS)	Seropositive birds/total birds
H7N7_avHA	Chickens	ON infected	10/10	5.7	10/10	1.7	n.a
		ON Contact	1/5	8.0	5/5	0.6	4/5
	Ducks	ON infected	0/10	n.a.*	0/10	0.0	7/7
		ON Contact	0/5	n.a.*	0/5	0.0	5/5
H7N7_5eqHA	Chickens	ON infected	6/10	4.5	10/10	1.4	4/4
		ON Contact	3/5	7.3	5/5	1.0	2/2
	Ducks	ON infected	0/10	n.a.*	0/10	0.0	7/7
		ON Contact	0/5	n.a.*	0/5	0.0	5/5

### The NeuGc-specific residues did not affect virus replication or excretion in chickens, but ducks are potentially silent spreaders of H7N7 viruses

We further determined the effect of the five equine mutations on viral loads in swab and organ samples obtained from ON-inoculated birds and their contacts. Oral and cloacal swabs collected 2, 4, 7, and 10 dpi were tested using plaque assay. In chickens and ducks, the level of virus shedding from the cloacal route was higher compared to the oral route, although the differences were not statistically significant (Fig. 2C). Interestingly, ducks excreted high viral loads by the oral and fecal routes for a longer time than chickens. Moreover, the viral distribution in different organs (brain, heart, lungs, liver, jejunum, spleen, and kidneys) obtained 4 dpi from three birds of each group was broader in chickens than in ducks. In H7N7\_avHA-infected chickens, the viral load was significantly higher in the brain, heart, liver, and kidneys compared to infected ducks (Fig. 2D).

Similarly, in H7N7\_5eqHA-infected chickens, the viral load was significantly higher in all organs except the spleen compared to ducks.

In conclusion, we observed increased levels of viral shedding via the cloacal routes and prolonged shedding of viruses in ducks compared to chickens. The viral distribution in chicken organs was broader than in ducks, which may explain the high mortality in chickens. These findings raise concerns about the potential spread of H7N7 viruses, particularly by ducks as silent spreaders. The high and prolonged shedding of the H7 viruses in ducks, even without exhibiting clinical symptoms, pose a potential risk for their reintroduction to hosts that have a high presence of NeuGc, like pigs and horses.



**Fig 2. In vivo characterization of WT (H7N7\_avHA) and mutant (H7N7\_5eqHA) A/chicken/Germany/R28/2003 H7N7 virus.** (A) Survival and clinical score of ON infected chickens throughout the animal experiment. (B) Survival curve of ON infected Pekin ducks. (C) Analysis of oral and cloacal swab samples taken from chickens and ducks in plaque tests expressed as Log<sub>10</sub> PFU/ml. (D) The viral distribution in duck and chicken organs was analyzed in plaque tests and expressed as PFU/gram. Asterisks indicate statistical significance based on p values \* $\leq$ 0.05, \*\* $\leq$ 0.01, \*\*\* $\leq$ 0.001, \*\*\*\* $\leq$ 0.0001. ns = not significant. Dashed lines indicate the predicted detection limit of the plaque assay (cut-off).



### Avian H7 IAVs bind both $\alpha$ 2,3-linked NeuAc and sialyl-LewisX epitopes

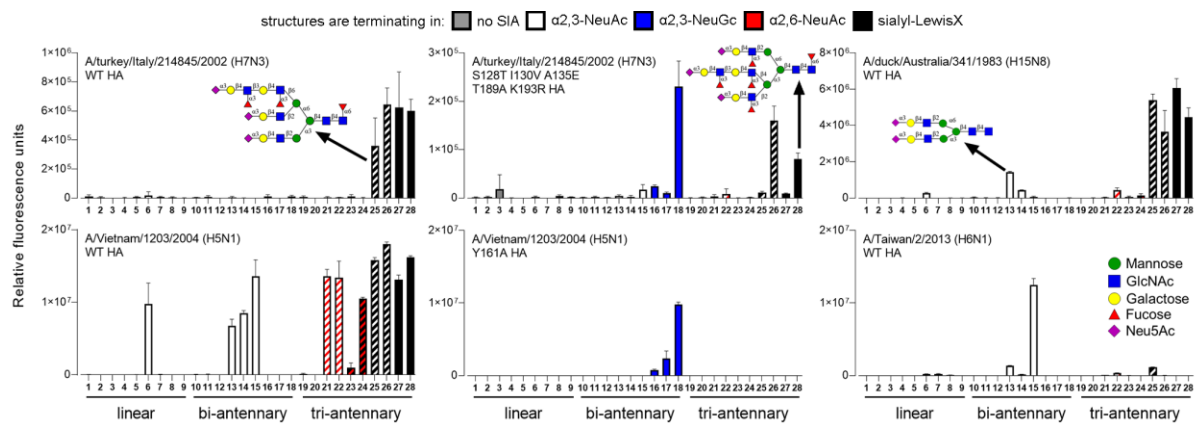
The animal experiments showed that both chickens and ducks were infected by the NeuGc-specific H7 viruses (Fig. 2), although both species are known to not express NeuGc [29, 31]. This strongly suggests that in our previous research, in which we observed NeuGc-specificity of this mutant (S128T, I130V, A135E, T189A, and K193R) H7 HA [33], we overlooked the binding to one of the many other glycans that may be present in nature. We hypothesized that sialyl-LewisX (sLe<sup>x</sup>) epitopes are important as they have recently been shown to be involved in H7 IAV infection [39, 40] and are bound by nearly all IAV subtypes [41-49]. The sLe<sup>x</sup> epitope consists of an  $\alpha$ 2,3-linked NeuAc presented on an *N*-acetylglucosamine (GlcNAc) with an additional fucose moiety  $\alpha$ 1,3-linked to the *N*-acetylglucosamine (GlcNAc) of the LacNAc. The sLe<sup>x</sup> epitopes are present in some species and tissues, such as chicken trachea and colon, guinea fowl trachea, turkey respiratory tract, and human lung [45, 50-55]. Most research investigating binding to the sLe<sup>x</sup> epitope has been performed using a tetrasaccharide sLe<sup>x</sup> epitope, due to a lack of biologically relevant glycans.

Here, we investigated which exact glycans are bound by WT and mutant avian H7 HAs to explain how chickens and ducks are infected by NeuGc-specific H7 viruses. Since the complex glycan structure can influence receptor binding [14, 22-25], we here focused on biologically relevant complex *N*-glycans presenting  $\alpha$ 2,3-linked NeuAc,  $\alpha$ 2,3-linked NeuGc,  $\alpha$ 2,6-linked NeuAc, and sLe<sup>x</sup> epitopes (Fig. S1). These glycans were printed on glass slides and after incubation of the slides with the HAs and fluorescent secondary antibodies, the glycan-HA binding was evaluated.

Whereas the WT HA of A/turkey/Italy/214845/02 (H7tu) previously only showed binding to glycans terminating in  $\alpha$ 2,3-linked NeuAc [33], here we observed a strong preference for tri-antennary *N*-glycans presenting at least one sLe<sup>x</sup> epitope (glycans **25-28**, Fig. 3). Both glycans solely presenting sLe<sup>x</sup> epitopes (**27, 28**) and glycans presenting a sLe<sup>x</sup> on one arm and  $\alpha$ 2,3-linked NeuAc on the other two arms (**25, 26**) were bound. For the latter, it cannot be distinguished whether binding is caused by the sLe<sup>x</sup> or the  $\alpha$ 2,3-linked NeuAc. Interestingly, binding to sLe<sup>x</sup> epitopes presented on linear glycans (**7-9**) was not observed. Furthermore, steric hindrance due to the presence of  $\alpha$ 2,6-linked NeuAc on two arms, besides the sLe<sup>x</sup> on one arm, appeared to be present, as glycan **23** and **24** were not bound. We also showed that the mutant H7tu HA bound both glycans terminating in  $\alpha$ 2,3-linked NeuGc, as well as sLe<sup>x</sup>-presenting tri-antennary *N*-glycans (Fig. 3). In conclusion, the observed binding to sLe<sup>x</sup> (Fig. 3) showed that the previously studied mutant avian H7 HAs were not strictly specific for NeuGc and may explain why these HAs bound to chicken tissue and erythrocytes previously [33]. This sLe<sup>x</sup>-binding possibly also explains why ducks and chickens could be infected by the NeuGc-adapted avian H7 virus (Fig. 2).

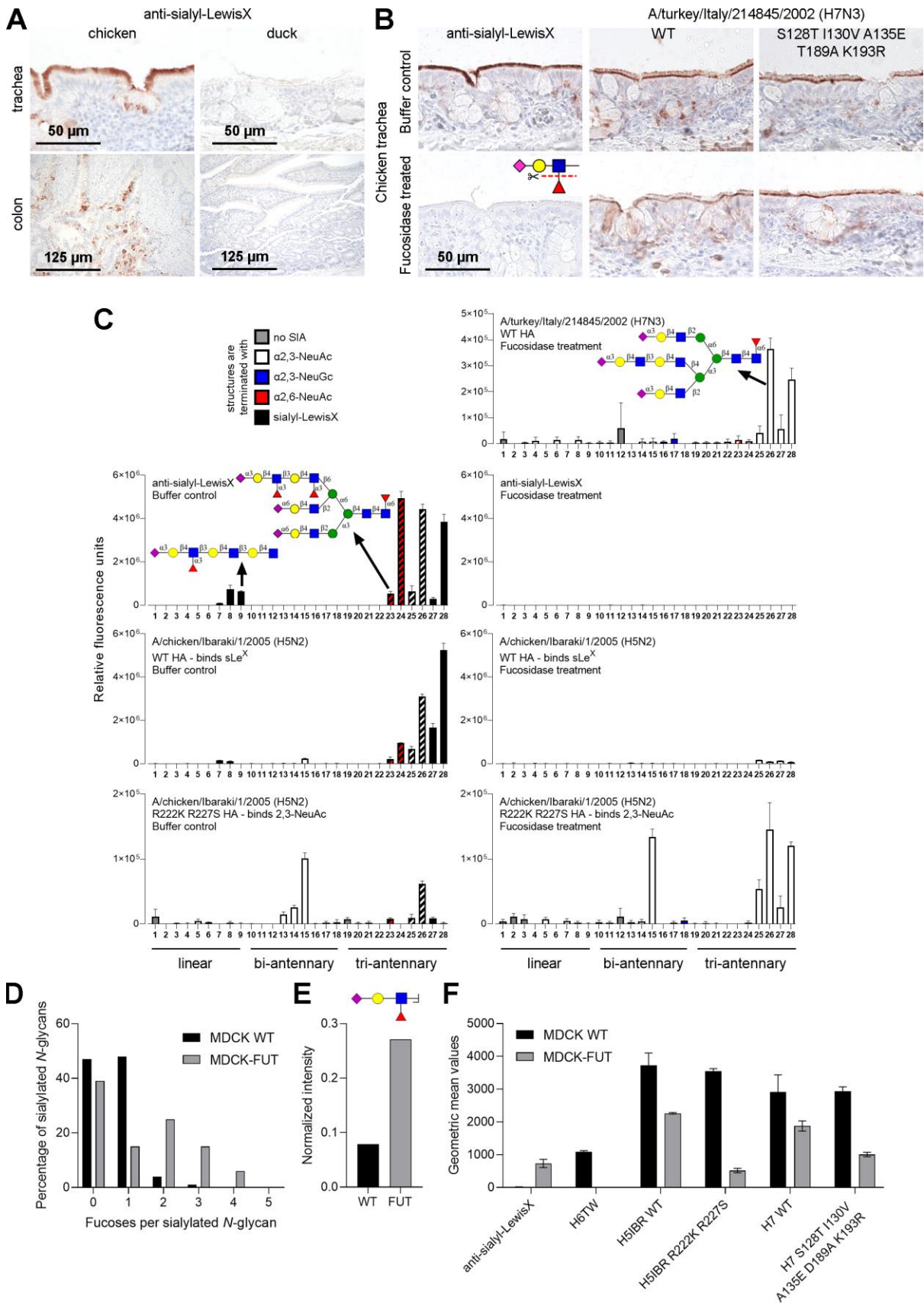
Furthermore, we examined the receptor specificity of the H15 HA of A/duck/Australia/341/1983, the closest related subtype to H7. The H15 HA showed a similar binding phenotype to the H7tu HA and bound  $\alpha$ 2,3-linked NeuAc and sLe<sup>x</sup>, while structures presenting  $\alpha$ 2,6-linked NeuAc on the other arms were not

bound (Fig. 3). Interestingly, H15 HAs were previously not known to bind to sLe<sup>x</sup> epitopes. The steric hindrance due to  $\alpha$ 2,6-linked NeuAc appeared to be specific for the H7 and H15 HAs, since the WT H5 HA from A/Vietnam/1203/2004 (H5VN) showed binding to all tri-antennary N-glycans presenting at least one sLe<sup>x</sup> epitope, regardless of the terminal epitopes presented on the other arms (Fig. 3). Consistent with the results from H7 and H15 HAs, the WT H5VN HA only bound to sLe<sup>x</sup> epitopes when presented on tri-antennary N-glycans, and not linear glycans, possibly due to a multivalency effect because of the high density of binding epitopes in one glycan. Furthermore, an HA that we previously used as a control for specific binding of NeuGc (the Y161A HA mutant A/Vietnam/1203/2004 H5N1) bound specifically to NeuGc and not sLe<sup>x</sup> (Fig. 3). Interestingly, the H6 HA from A/Taiwan/2/2013 was strictly specific for  $\alpha$ 2,3-linked NeuAc and did not bind to glycans that present other epitopes on the other arms (25 and 26). The results showed that the fine receptor specificity is highly dependent on the IAV and the exact complex glycan structure.



**Figure 3. Avian H7 HAs bind to sialyl-LewisX epitopes.** Synthetic glycans were used to assess the receptor binding of the IAV HAs (A/turkey/Italy/214845/2002 H7, A/duck/Australia/341/1983 H15, A/Vietnam/1203/2004 H5, and A/Taiwan/2/2013 H6). The glycans were terminating in galactose (no SIA, grey),  $\alpha$ 2,3-linked NeuAc (white),  $\alpha$ 2,3-linked NeuGc (blue),  $\alpha$ 2,6-linked NeuAc (red), or sialyl-LewisX (black). Bars with two colors indicate glycans terminating in different epitopes on different arms. Fig. S1 presents all structures that are present on the array. Bars represent the mean  $\pm$  SD (n=4).

Nevertheless, the binding of the H7 HAs to sLe<sup>x</sup> epitopes does not explain the infection in ducks, since ducks are generally assumed not to present NeuGc nor sLe<sup>x</sup> epitopes [29, 31, 50]. A tissue stain using anti-sialyl-LewisX antibodies revealed that, indeed, no sLe<sup>x</sup> epitopes were found on duck colon and tracheal tissues (Fig. 4A). Therefore, we aimed to investigate whether the binding of these avian H7 HAs was truly dependent on sLe<sup>x</sup> epitopes. We performed assays using a fucosidase E1\_10125 from *Ruminococcus gnavus* E1, which specifically cleaves the fucose moiety from the sLe<sup>x</sup> epitope [56] (Fig. 4B). The anti-sialyl-LewisX antibody showed that no more sLe<sup>x</sup> epitopes remained on the chicken trachea after fucosidase treatment. However, the binding of both WT and mutant H7tu (Fig. 4B) HA to chicken trachea remains unchanged after fucosidase treatment, indicating that  $\alpha$ 2,3-linked NeuAc is bound in a tissue section.



**Fig 4. Avian H7 HAs bind both  $\alpha$ 2,3-linked NeuAc and sialyl-LewisX epitopes.** (A) The presence of sialyl-LewisX epitopes on chicken and duck trachea and colon was investigated using anti-sialyl-LewisX antibodies. (B) The binding of anti-sialyl-LewisX antibodies and H7 HAs to chicken tracheal tissue (with and without fucosidase treatment)

was assessed. Tissue binding was visualized using AEC staining. **(C)** Synthetic glycans (with and without fucosidase treatment) were used to assess the binding of the anti-sialyl-LewisX antibody, the WT H7 HA of A/turkey/Italy/214845/02, and H5 HAs of A/chicken/Ibaraki/1/2005. **(D)** The *N*-glycans of MDCK WT and MDCK-FUT cells were investigated using HILIC-IMS-QTOF positive mode mass spectrometry. The number of fucoses per sialylated *N*-glycan was analyzed for both cell types. Further analysis is presented in Table S2. **(E)** The *N*-glycans were further analyzed using LC-MS/MS. The oxonium ions of *m/z* 803.2928 (sLe<sup>x</sup>) were identified and normalized to the core fragments. Mean and standard errors (*n*=3) are shown. **(F)** Flow cytometry measurements were performed to assess the binding anti-sialyl-LewisX antibodies and HAs to MDCK WT and MDCK-FUT cells. Triplicate measurements were performed, of which the mean and standard deviation are displayed.

To investigate which exact glycans may have been involved in this binding to the chicken trachea, we treated the glycan microarray with fucosidase (Fig. 4C). Using an anti-sLe<sup>x</sup> antibody that bound all sLe<sup>x</sup>-containing structures on the glycan microarray as a control, we indeed showed that all sLe<sup>x</sup> epitopes were removed after fucosidase treatment of the array. This suggests that no sLe<sup>x</sup> epitopes remained on the chicken trachea either after fucosidase treatment (Fig. 4B). As additional controls, we used the WT and R222K R227S mutant H5 HAs of A/chicken/Ibaraki/1/2005 (H5IBR), which are, respectively, specifically binding to sLe<sup>x</sup> structures and  $\alpha$ 2,3-linked NeuAc [43]. After fucosidase treatment, these WT and mutant H5IBR HA controls showed, respectively, no binding to sLe<sup>x</sup> glycans and increased binding to sLe<sup>x</sup> glycans (that are now converted to  $\alpha$ 2,3-linked NeuAc) (Fig. 4C). Interestingly, the H7tu WT HA still bound to structures **25** to **28** (of which the sLe<sup>x</sup> is converted to  $\alpha$ 2,3-linked NeuAc) after fucosidase treatment, but not bi-antennary *N*-glycans presenting  $\alpha$ 2,3-linked NeuAc (**13-15**, Fig. S1, Fig. 4C), suggesting that tri-antennary *N*-glycans are preferred as receptors for H7tu. In conclusion, both  $\alpha$ 2,3-linked NeuAc and sLe<sup>x</sup> epitopes presented on the tri-antennary *N*-glycans are bound efficiently by the H7tu WT HA.

To further investigate binding to sLe<sup>x</sup> glycans, we used MDCK WT and MDCK-FUT cells. The latter overexpress the chicken fucosyltransferase genes *FUT3*, *FUT5*, and *FUT6* [50], which was expected to increase the amount of sLe<sup>x</sup> epitopes that are presented on the cells. To investigate whether indeed increased amounts of sLe<sup>x</sup> epitopes were presented on the cells, we employed mass spectrometry (MS) methods. We first investigated the released *N*-glycans from MDCK WT and FUT cells by HILIC-IMS-QTOF positive mode MS and found that MDCK-FUT cells presented a higher number of fucoses on sialylated *N*-glycans than MDCK WT cells (Fig. 4D, Table S2). To further investigate whether the fucoses were present in sLe<sup>x</sup> epitopes, we analyzed the *N*-glycans using fragmentation in LC-MS/MS, which indeed showed a higher relative abundance of sLe<sup>x</sup> fragments (oxonium ions of *m/z* 803.2928) on the MDCK-FUT cells (Fig. 4E). We then continued to use these MDCK WT and FUT cells in flow cytometry analysis. The controls (anti-sLe<sup>x</sup> and the HA of A/Taiwan/2/2013 H6N1, which is specific for  $\alpha$ 2,3-linked NeuAc (Fig. 3)) showed that the amount of sLe<sup>x</sup> on MDCK WT cells and the amount of  $\alpha$ 2,3-linked NeuAc on MDCK-FUT cells was very low (Fig. 4F). Surprisingly, the H5IBR HAs (WT and mutant) that were assumed to be specific for sLe<sup>x</sup> and  $\alpha$ 2,3-linked NeuAc,

respectively, bound well to both cell types. Similar to the result in the glycan microarray and tissue stains, the H7 (and H15) WT and mutant HAs bound to both cell types. In conclusion, both  $\alpha$ 2,3-linked NeuAc and sLe<sup>x</sup> epitopes appear to be bound efficiently by the H7tu, but binding depends on the exact glycan structure.

## Discussion

Here, we studied the effects of equine NeuGc-adapting mutations (S128T, I130V, A135E, T189A, and K193R) in avian H7 IAVs *in vitro* and *in vivo*. These viruses are potentially candidates for interspecies transmission between avian and mammalian species expressing NeuGc receptors. The insertion of equine NeuGc-adapting mutations resulted in stable and viable viruses and increased viral replication in horse cells. While the mutations reduced viral replication in chicken and dog cells, interestingly the replication in duck cells was not affected. *In vivo*, the NeuGc-adapting mutations not only reduced the pathogenicity index in intravenously infected chickens but also mortality and morbidity in oculonasal-infected chickens. In ducks, on the other hand, neither virus caused signs of illness or increased mortality. Nevertheless, ducks shed high amounts of virus for a longer time compared to chickens. Here, NeuGc-adapting mutations were not disadvantageous in viral shedding compared to the WT HPAIV. The NeuGc-adapted H7 was additionally found to bind  $\alpha$ 2,3-linked NeuAc and sialyl-LewisX (sLe<sup>x</sup>) epitopes, but only when these epitopes were presented on tri-antennary N-glycans. Binding to these epitopes explains why ducks and chickens could be infected and emphasizes the risk of interspecies transmission of H7 IAVs.

Although sLe<sup>x</sup> epitopes were identified as potential receptors for the studied H7 viruses, it is currently unclear whether sLe<sup>x</sup> is used in IAV infections as a functional receptor, or whether it has other functions. It has been suggested that the presence of sLe<sup>x</sup> facilitates H7 IAV infection. If sLe<sup>x</sup> binding is important in IAV infection, this may cause a species barrier or act as an intermediate receptor since some species and tissues, such as the chicken trachea and colon, guinea fowl trachea, turkey respiratory tract, and human lung present sLe<sup>x</sup> epitopes [45, 50-55].

The molecular determinants for the binding of H7 viruses to sLe<sup>x</sup> epitopes are currently unknown. Since we showed that not all IAVs bind to sLe<sup>x</sup> epitopes, such as the H6 HA of A/Taiwan/2/2013 (Fig. 3), likely certain amino acids are responsible for the binding to sLe<sup>x</sup>. For H5 viruses, it was shown that mutations K222R/Q and S227R in the HA convert from binding to  $\alpha$ 2,3-linked NeuAc to sLe<sup>x</sup> [43, 49, 51]. Especially, the lysine at position 222 was shown to sterically hinder binding to sLe<sup>x</sup> epitopes [57], while a glutamine or arginine at that position enables, potentially through a hydrogen bond, sLe<sup>x</sup> binding. Indeed, the H7 viruses that were investigated here contain a glutamine (Q) at position 222, which is highly conserved in H7 viruses [51] and partially explains the binding to sLe<sup>x</sup>. However, position 227 is also a glutamine in the investigated H7 viruses, of which the effect on sLe<sup>x</sup> binding is currently unknown. Elsewhere, the presence of a lysine at position 193 is reported to be important for sLe<sup>x</sup> binding and, indeed, the H7tu



contains a 193K [40]. Additionally, amino acids at other positions, which have not been investigated yet, may also affect the binding to sLe<sup>x</sup> epitopes.

Previously, IAVs from all subtypes, except H15, were shown to bind sLe<sup>x</sup> epitopes [41-49]. Using tri-antennary *N*-glycans presenting sLe<sup>x</sup> epitopes, we here showed that also H15 IAVs, the closest related subtype to H7, are capable of binding sLe<sup>x</sup> epitopes. Furthermore, the investigated avian H7 HAs were previously not known to bind to sLe<sup>x</sup> epitopes, as binding was only observed when the epitopes were presented on tri-antennary *N*-glycans and not linear glycans. Additionally, the WT H7 HA appeared to bind stronger to  $\alpha$ 2,3-linked NeuAc when presented on tri-antennary *N*-glycans (sLe<sup>x</sup> after fucosidase treatment) than bi-antennary *N*-glycans or linear glycans. The H7tu also appeared to bind stronger to tri-antennary *N*-glycans presenting the  $\alpha$ 2,3-linked NeuAc on an elongated MGAT4 arm (**26** and **28**) instead of an elongated MGAT5 arm (**25** and **27**) (Fig. 4C, Fig. S1), although this was not consistent throughout the replicates. Using these sLe<sup>x</sup>-presenting *N*-glycans in combination with other HAs may reveal the binding of more IAVs to sLe<sup>x</sup> and the role of these epitopes in IAV infection. These observations highlight the relevance of looking beyond the terminal epitope and considering the fine receptor specificity when investigating IAV receptor binding.

The distribution and types of Sias are species-specific and variable throughout the respiratory tract of IAV-susceptible species [17]. HA specificity is often adapted to the particular Sia receptors present in the host [14]. Thus, interspecies transmission and establishment in a new host requires a successful adaptation of HA binding specificity to the new host environment as seen in equine H7N7 viruses originating from avian H7 viruses [10]. Although equine H7N7 IAVs are thought to be extinct [12, 13], the amino acid residues coding for the NeuGc binding specificity persist in avian H7 sequences (Table S1), enabling a potential re-emergence of NeuGc-binding viruses. The avian H7 IAVs with equine-adapted mutations that we investigated not only bound to equine-specific NeuGc-receptors but were also able to replicate and infect avian hosts (Fig. 2). The viruses with equine-adapted NeuGc-specific mutations may not be as effective in avian  $\alpha$ 2,3-linked NeuAc receptor binding and viral replication as WT avian virus but still show infection *in vivo*. Furthermore, reassortant viruses with an equine H7N7 HA and other genes from a chicken H5N2 IAV were shown to be lethal in chickens [11]. This suggests a potential for transmission of equine-adapted viruses with NeuGc binding specificity back to avian species like chickens or ducks, for example, due to the close proximity of these species in farms. Our observations highlight the relevance of considering the fine receptor binding specificity when investigating the effect of species-specific adaptations in the RBS of HA and their potential in interspecies transmission events.



## Materials & Methods

### Cell culturing and preparation of cell lysates

The Biobank of the Friedrich-Loeffler-Institut (FLI), Greifswald Insel-Riems, Germany provided the following cell cultures for the *in vitro* characterization of the viruses: human embryonic kidney cells (HEK293T), Douglas Foster-1 cells (DF-1), Madin-Darby Canine Kidney type II cells (MDCKII), quail muscle 9 cells (QM-9), horse epidermal cells (E.Derm), horse lung cells (PLU-R), and duck embryo cells (SEF-R). In addition, 11d-old specific-pathogen-free embryonated chicken eggs SPF-ECE (Valo BioMedia, Germany) and chicken embryonic kidney cells (CEK) isolated from 18d-old SPF-ECE were used to perform replication kinetics [58].

MDCKII and PLU-R cells were cultured in minimal essential medium (MEM) with 10% fetal calf serum (FCS) containing Hank's, Earls salts, NaHCO<sub>3</sub>, sodium pyruvate, and non-essential amino acids. For HEK293T, DF-1, QM-9, and SEF-R Iscove's Modified Dulbecco's medium (IMDM) with 10% FCS, Ham's F12 nutrient mix, and glutamine was used. E.Derm and CEK cells were cultured in Eagle's MEM and different concentrations of NaHCO<sub>3</sub>. HEK293S GnTI(-) cells were cultured in DMEM with 10% FCS. All cells were cultured at 37°C with 5% CO<sub>2</sub>.

MDCK WT (CCL-34) and MDCK-FUT [50] (a kind gift from Takahiro Hiono) cells were cultured in DMEM (Gibco) with 10% FCS (S7524, Sigma) and 1% penicillin and streptomycin (Sigma). MDCK-FUT cells were maintained with an additional 500 µg/ml G418 sulfate. MDCK-FUT cells overexpress the chicken fucosyltransferase genes *FUT3*, *FUT5*, and *FUT6* [50]. Cells were cultured at 37°C with 5% CO<sub>2</sub>. Detaching of the cells was always done using 1X TrypLE Express Enzyme (12605010, Thermo Fisher Scientific), using 2 ml in a T75 flask, at a confluency of approximately 90%. Cell lysates were obtained using TrypLE Express Enzyme and RIPA lysis buffer as described previously [59].

### Viruses

The influenza viruses were obtained from different cooperation partners: A/chicken/Germany/R28/2003 (H7N7) (designated H7N7\_avHA) was provided by Timm C. Harder, head of the reference laboratory for avian influenza virus, Friedrich-Loeffler-Institut (FLI), Greifswald Insel-Riems, Germany. AIV A/quail/California/D113023808/2012 (H4N2) was supplied by Beate Crossley from the California Animal Health and Food Safety Laboratory System, Department of Medicine and Epidemiology, University of California, Davis, United States. Stephan Pleschka and Ahmed Mostafa from Justus-Liebig-University, Gießen, Germany provided the human isolate A/Victoria/1975 (H3N2).

The avian sequence containing the five equine mutations S128T, I130V, A135E, T189A, and K193R (H3 numbering) was ordered from GenScript and inserted into the HA of A/chicken/Germany/R28/2003 in a pHWSccdB vector by restriction enzyme cloning. The IAV carrying these 5 mutations (designated H7N7\_5eqHA) was generated in the backbone A/chicken/Germany/R28/2003 using reverse genetics. The virus was rescued in HEK293T cells, propagated in 9- to 11-d-old SPF eggs, and pooled for further use. Sequence analysis of different isolated viral pools

revealed a stable establishment of the five mutations in the HA without a reversion to the WT sequence.

### **α2,3-linked NeuAc receptor binding affinity assay**

The binding of H7N7\_avHA and H7N7\_5eqHA to avian α2,3-NeuAc sialic acid receptor types was determined in a solid-phase binding assay using fetuin as a substrate as previously described [60, 61]. The majority of sialic acids in fetuin are NeuAc and the low amount of NeuGc is neglectable for this assay [62]. Briefly, 96 well-plates were coated with 200 μl of 10 μg/ml fetuin from fetal bovine serum (Merck, F3004) in 2x PBS overnight at 4°C. Fetuin-coated plates were washed with distilled water, dried at RT, and coated with 5 log<sup>2</sup> HA units of H7N7\_avHA, H7N7\_5eqHA, A/quail/California/D113023808/2012 (H4N2) (positive control), or A/Victoria/1975 (H3N2) (negative control) in TBS at 4°C overnight. Viruses were tested in duplicates. Afterwards, plates were washed with PBS and then blocked for 1h RT using 0.1% de-sialylated BSA in 2x PBS. BSA was de-sialylated by incubating 5% BSA in 2x PBS + 0.02% penicillin-streptavidin with 1 unit of *Vibrio cholerae* neuraminidase for 24h at 60°C. The plate was washed with washing solution containing 2x PBS + 0.01% TWEEN80. Horseradish peroxidase (HRP) labeled fetuin was diluted 1:2 in 2x PBS with 0.02% TWEEN80 + 0.1% de-sialylated BSA and dilutions from 1:200 to 1:12,800 were added to the plate after washing and incubated for 1h at 4°C. After an additional washing step using washing solution, 100 μl peroxidase substrate (Rockland; Lot# 24241) was added at RT. After 30 min the reaction was stopped using 50mM H<sub>2</sub>SO<sub>4</sub> and the optical density was measured at 450 nm with an ELISA reader (Tecan).

### **Plaque test and cell-to-cell spread**

Plaque tests were performed using MDCKII for virus titration. Confluent MDCKII cells were incubated with diluted or undiluted samples for 1h at 37°C with 5% CO<sub>2</sub>. After infection, cells were washed twice using sterile PBS and overlaid with 1.8% bacto-agar in Dulbecco's modified Eagle's medium (DMEM) containing 5% bovine serum albumin (BSA). After incubation for 72 hours, plaque assays were fixed using a 10% formaldehyde solution with 0.1% crystal violet for 24h. After removal of the agar, the viral titers were calculated as PFU/ml or PFU/g, or the size of the plaques was measured under the microscope using Nikon NIS-Elements software.

### **Replication kinetics**

Different cell cultures were infected with H7N7\_avHA and H7N7\_5eqHA with a multiplicity of infection (MOI) of 0.001 for 1h. After washing with PBS, the cells were incubated at 37°C with 5% CO<sub>2</sub>. Cells and supernatants were harvested at indicated time points and stored at -80°C until the PFU/ml were assessed using plaque tests. The viral replication in SPF-ECE was tested in 11d-old eggs. Eggs were inoculated with 100 PFU/0.1 ml of each virus and incubated at 37°C with 56% humidity. Allantoic fluids were harvested at 8, 24, and 48 hpi and the PFU/ml was determined using plaque tests.

### **Fusion assay to measure pH-dependent HA activation**

The HA segments of H7N7\_avHA and H7N7\_5eqHA were cloned into the protein expression vector pCAGGS by restriction enzyme cloning (H7N7\_avHA\_pcAGGs and H7N7\_5eqHA\_pcAGGs respectively). A fusion assay was performed to assess the influence of the equine mutations on the pH activation of the HA as previously described [35]. Briefly, confluent QM-9 cells were transfected in a 24-well plate with 500 ng of H7N7\_avHA\_pcAGGs and H7N7\_5eqHA\_pcAGGs and 100 ng GFP\_pcAGGs per well using 1  $\mu$ l/ $\mu$ g DNA Lipofectamine 2000 (Thermofisher). PBS buffers were prepared with pH values ranging from 4.0 to 6.0 (0.2 steps). Transfected cells were incubated for 16h at 37°C with 5% CO<sub>2</sub> and each well was washed with a different pH-adjusted PBS buffer for 10 min at RT after incubation. Cells were incubated for another 4h at 37°C with 5% CO<sub>2</sub> and then fixed with 4% paraformaldehyde for 30 min at RT. The sizes of the fusion events were measured using a microscope and Nikon NIS-Elements software.

### **Thermostability**

The thermostability of H7N7\_avHA and H7N7\_5eqHA viruses was tested in a thermostability assay at 56°C. Allantoic fluid aliquots containing 10<sup>5</sup> PFU/ml of viruses were prepared in tubes and incubated at 56°C. Samples were taken at 0, 1, 2, 4, and 6 hours post incubation. The infectivity of heat-exposed viruses was assessed in a plaque test on MDCKII cells. The PFU/ml are shown as mean values from two independent experimental rounds.

### **Animal experiments**

For the assessment of the pathogenicity index, nine chickens were infected intravenously (IV) and ten Pekin ducks were infected intramuscular (IM) with both recombinant viruses. Daily clinical scoring and the calculation of the pathogenicity index were performed according to the standard protocol of the world organization for animal health (WOAH). In addition, ten Pekin ducks and ten chickens were inoculated via the oculonasal route with 10<sup>5</sup> PFU/ml. One day post-infection (dpi) five contact birds were added to each group to assess bird-to-bird transmission. In addition to the daily clinical scoring, oral and cloacal swab samples were taken on 2, 4, 7, and 10 dpi. The oral and cloacal swab samples were stored in MEM (H) + MEM (E) + NEA medium with BSA containing enrofloxacin, lincomycin, and gentamicin at -70°C until further use. The swab samples were titrated in plaque tests on MDCKII cells to assess the PFU/ml in the collected swabs. Organ samples were collected 4 dpi from three freshly dead or euthanized birds to assess viral distribution. Organ samples from the brain, heart, lung, liver, jejunum, spleen, and kidney were lysed in PBS using a tissue lyser and the PFU/g was assessed using plaque tests on MDCKII cells. The surviving birds were culled on day ten of the trial and blood was collected for serum samples. The serum was tested for influenza A NP antibody using a competitive ELISA kit (ID Screen Influenza A Antibody Competition Multispecies; IDvet).

### Expression and purification of HA for binding studies

HA encoding cDNAs of A/turkey/Italy/214845/02 H7N3 [63] (synthesized and codon-optimized by GenScript), A/duck/Australia/341/1983 H15N8 (a kind gift from Keita Matsuno), A/Vietnam/1203/2004 H5N1 (synthesized and codon-optimized by GenScript), A/chicken/Ibaraki/1/2005 H5N2 [43], and A/Taiwan/2/2013 H6N1 were cloned into the pCD5 expression vector as described previously [64]. The pCD5 expression vector was adapted to clone the HA-encoding cDNAs in frame with DNA sequences coding for a secretion signal sequence, the Twin-Strep (WSHPQFEKGGGSGGGWSHPQFEK; IBA, Germany), a GCN4 trimerization domain (RMKQIEDKIEEIESKQKKIENEIARIKK), and a superfolder GFP [65] or mOrange2 [66]. Mutations in HAs were generated by site-directed mutagenesis (primers are available upon request). The HAs were purified from cell culture supernatants after expression in HEK293S GnTI(-) cells as described previously [64]. In short, transfection was performed using the pCD5 expression vectors and polyethyleneimine I. The transfection mixtures were replaced at 6 h post-transfection by 293 SFM II expression medium (Gibco), supplemented with sodium bicarbonate (3.7 g/L), Primatone RL-UF (3.0 g/L, Kerry, NY, USA), glucose (2.0 g/L), glutaMAX (1%, Gibco), valproic acid (0.4 g/L) and DMSO (1.5%). At 5 to 6 days after transfection, tissue culture supernatants were collected and Strep-Tactin sepharose beads (IBA, Germany) were used to purify the HA proteins according to the manufacturer's instructions.

### Glycan microarray binding of HA proteins

A selection of a glycan array that was presented elsewhere [67] was used and the full list of glycans is presented in Fig. S1. When fucosidase treatment was performed, cells were treated overnight at 37°C with 200 µg/ml fucosidase in fucosidase buffer (50 mM citrate buffer + 5 mM CaCl<sub>2</sub>, pH 6.0). Anti-sialyl-LewisX antibodies (mouse IgM, #551344, clone CSLEX1, BD Biosciences) at 50 µg/mL in 40 µL PBS-T were applied to the subarrays for 90 min. Subsequently, the arrays were incubated with a mixture of goat anti-mouse IgM HRP (10 µg/mL; #1021-05 Southern Biotech) and donkey anti-goat antibody labeled with AlexaFluor555 (5 µg/mL; A21432, Invitrogen) in 40 µL PBS-T for 1 h. HAs were pre-complexed with human anti-streptag and goat anti-human-AlexaFluor555 antibodies in a 4:2:1 molar ratio, respectively in 50 µL PBS-T on ice for 15 min. The mixtures were added to the subarrays for 90 min in a humidified chamber. Wash steps after each incubation (e.g. enzyme treatment, HA, or antibody incubation) involved six successive washes of the whole slides with either twice PBS-T, twice PBS, and twice deionized water. The arrays were dried by centrifugation and immediately scanned as described previously [27]. Processing of the six replicates was performed by removing the highest and lowest replicates and subsequently calculating the mean value and standard deviation over the four remaining replicates.

### Fucosidase production

The protein sequence of fucosidase E1\_10125 from *Ruminococcus gnavus* E1 [56] was obtained from the RCSB Protein Data Bank under accession number 6TR3. The

nucleotide sequence was obtained from the closest hit in a protein-protein search in BLAST, of which the nucleotide sequence was corrected to obtain the exact same amino acid sequence. This open reading frame was ordered at GenScript and cloned into backbone pET23A in frame with a His-tag on the C-terminus of the open reading frame of the fucosidase. Cloning was performed in JM109 *Escherichia coli* (Promega). The plasmid is deposited at addgene under the name: pET23A-Fucosidase-E1\_10125-Ruminococcus\_gnavus-His (#207665). Expression of the fucosidase was performed in BL21 (DE3) *E. coli* (New England Biolabs). An overnight culture (90 ml per liter culture) was inoculated in 2xYT medium (Serva) supplemented with 50 µg/ml ampicillin (13398.02, Serva). Bacteria were grown at 37°C while shaking at 200 rpm until OD 0.8-1.0, after which fucosidase production was induced with 1 mM isopropyl β-d-1-thiogalactopyranoside (R0309, Invitrogen). Afterward, the bacteria were grown for 21 h at room temperature while shaking at 200 rpm. Cell pellets were obtained by centrifugation in a swing-out centrifuge for 30 min at 4°C at 629 rcf. The pellets were resuspended in 50ml lysis buffer (100mM Tris-HCl, pH 8.0, 0.1% TritonX-100) per liter bacterial culture, which was supplemented with 1 gram per liter bacterial culture lysozyme (62971, Merck), 25 µl per liter bacterial culture DNase (EN0521, Thermo Fisher Scientific), and 25 µl per liter bacterial culture DNase buffer. The mixtures were incubated for 50 minutes at 37°C while shaking at 200 rpm. The cells were additionally lysed by sonication (Bandelin, Sonopuls, needle MS73) at 50% amplitude, three times for one minute at 10s intervals. The lysates were centrifuged for 1.5 h at 4°C at 629 rcf until the supernatant was clear. The supernatant was filtered through a 0.45 µm filter (431220, Corning) and incubated for 16h with Ni-NTA beads at 4°C while rotating. The beads were washed using a washing buffer (0.5M NaCl, 20mM Tris-HCl, pH 7.5) after which the fucosidase was eluted using the same buffer supplemented with 10-200mM imidazol. The elutions were concentrated and the buffer was exchanged to fucosidase buffer (50 mM citrate buffer + 5 mM CaCl<sub>2</sub>, pH 6.0) using a centrifugal concentrator with a molecular weight cutoff of 10 kDa (Vivaspin 6, VS0602, Sartorius). The presence of the fucosidase (62 kDa) was evaluated by running an SDS-PAGE gel (after denaturing for 15 min at 95°C with the addition of denaturing buffer (NP0009, Invitrogen)) with consequent staining using Coomassie blue dye.

### Protein histochemistry

Sections of formalin-fixed, paraffin-embedded chicken (*Gallus gallus domesticus*) were obtained from the Division of Pathology, Department of Biomolecular Health Sciences, Faculty of Veterinary Medicine of Utrecht University, the Netherlands. Sections of Pekin ducks were obtained from the animal experiment that is described above. Protein histochemistry was performed as previously described [68, 69]. In short, tissue sections of 4 µm were deparaffinized and rehydrated, after which antigens were retrieved by heating the slides in 10 mM sodium citrate (pH 6.0) for 10 min. Endogenous peroxidase was inactivated using 1% hydrogen peroxide in MeOH for 30 min at RT. Slides were treated overnight at 37°C with 150 µg/ml fucosidase in the fucosidase buffer (50 mM citrate buffer + 5 mM CaCl<sub>2</sub>, pH 6.0) or buffer only. Subsequently, slides were washed with PBS with 0.1% Tween-20 (PBS-T). Tissues were blocked for 90 min at RT using 3% BSA (w/v) in PBS. Anti-sialyl-

LewisX antibodies (mouse IgM, #551344, clone CSLEX1, BD Biosciences) were diluted 1:1000 in PBS and precomplexed with goat anti-mouse IgM-HRP (#1021-05, Southern Biotech) in a 1:100 dilution on ice for 20 min. The slides were stained for 90 min with pre-complexed HAs as previously described for the glycan microarray or the anti-sialyl-LewisX mixtures. For WT H5 HA, we used 1.5 µg/ml HA, for H5 R222K R227S HA, we used 3 µg/ml HA, for H7 HAs, we used 2 µg/ml HA, and for H15 HA, we used 1 µg/ml HA. After washing with PBS, binding was visualized using 3-amino-9-ethylcarbazole (AEC) (Sigma-Aldrich, Germany) and slides were counterstained using hematoxylin.

### Identification of N-glycans on cells by mass spectrometry

Cell lysates were obtained as described above. Preparation of mass spectrometry samples and measuring of these samples was performed as described previously [59]. Briefly, glycans from cell lysates were released using PNGaseF treatment, labeled with procainamide 2-picoline borane, and purified in multiple steps. The samples were measured both using positive mode HILIC-IMS-QTOF analysis and MS/MS (using a Thermo Scientific Exploris 480 connected to a Thermo Scientific Ultimate 3000 UPLC system).

IM-MS data was calibrated to reference signals of  $m/z$  121.050873 and 922.009798 using the IM-MS reprocessor utility of the Agilent Masshunter software. The mass-calibrated data was then demultiplexed using the PNNL preprocessor software using a 5-point moving average smoothing and interpolation of 3 drift bins. To find potential glycan hits in the processed data, the 'find features' (IMFE) option of the Agilent IM-MS browser was used with the following criteria: 'Glycans' isotope model, limited charge state to 5 and an ion intensity above 500. The found features were filtered by  $m/z$  range of 300 – 3200 and an abundance of over 500 (a.u.) where abundance for a feature was defined as 'max ion volume' (the peak area of the most abundant ion for that feature). After exporting the list of filtered features, glycans with a mass below 1129 Da (the mass of an N-glycan core) were removed. The ExPASy GlycoMod tool [70] was used to search for glycan structures (monoisotopic mass values, 5 ppm mass tolerance, neutral, derivatized N-linked oligosaccharides, procainamide (mass 235.168462) as reducing terminal derivative, looking for underivatized monosaccharide residues (Hexose, HexNAc, Deoxyhexose, and NeuAc)). For features with multiple potential monosaccharide combinations, the most realistic glycan in the biological context was chosen. The abundance of glycan features with the same mass, composition, and a maximum difference of 0.1 min in the retention time were combined as one isomer. A full glycan composition feature list for MDCK-FUT cells is presented in Table S2. Analysis of the number of fucoses per sialylated glycan was performed on the complex and hybrid N-glycans with at least one sialic acid.

For MS/MS data, Proteowizard MSconvert (version 3.0.21328-404bcf1) was used to convert Thermo raw files to MGF format using MGF as output format, 64-bit binary encoding precision and with the following options selected: write index, zlib compression and TPP compatibility. No filters were used when converting raw files to MGF format. To search MGF files for spectra containing glycan oxonium ions an internally developed tool named Peaksuite (v1.10.1) was used with an ion delta of



20 ppm, noise filter of 0% and using a list of oxonium  $m/z$  values as mass targets (Table SX from [59]). Scans without any detected peaks were removed. Python 3.2.2 was used for data curation based on precursor  $m/z$  (10 ppm), retention time (17-24 min) and intensities of oxonium ions that originated from the glycan core ( $m/z$  441.2707, 587.3286, 644.3501, 790.4080, 806.4029, and 952.4608). The sum intensity threshold of the core oxonium ions was set to  $1e4$ . Python 3.2.2 was also used for calculating the relative intensities of oxonium ions corresponding to sLe<sup>x</sup> ( $m/z$  803.2928) normalized versus the sum intensities of the core oxonium ions.

### Flow cytometry studies

Detaching of the cells (MDCK WT and MDCK-FUT) was performed with 1X TrypLE Express Enzyme (12605010, Thermo Fisher Scientific), using 2 ml in a T75 flask, at a confluency of approximately 90%. Cell pellets were obtained by centrifugation for 5 min at 250 rcf, after which the cells were resuspended in PBS supplemented with 1% FCS (S7524, Sigma) and 2mM EDTA and kept at 4°C until the plate was measured in the flow cytometer. In a round-bottom 96-well plate (353910, Falcon), 50,000 cells per well were used. Per well, 1 µg of precomplexed HA or precomplexed anti-sialyl-LewisX antibody (CD15S, clone CSLEX1, #551344, BD Biosciences) was added to PBS supplemented with 1% FCS and 2mM EDTA to achieve a final concentration of 20 µg/ml. Precomplexation of HAs was performed with 1.3 µg monoclonal antibody detecting the Twin-Strep-tag and 0.325 µg goat anti-human Alexa Fluor 488 (A11013, Invitrogen) and incubated on ice for 20 min. Precomplexation of the anti-sialyl-LewisX antibody was performed with a 1:50 dilution of goat anti-mouse IgM-HRP (#1021-05, Southern Biotech) and 0.65 µg donkey anti-goat Alexa Fluor 555 (A21432, Invitrogen) and incubated on ice for 20 min. Furthermore, eBioscience Fixable Viability Dye eFluor 780 (65-0865, Thermo Fisher Scientific) was diluted 1:2000 in the same mixture. Cells were incubated with the hemagglutinin/antibody mix for 30 minutes at 4°C in the dark. Cells were washed once with PBS supplemented with 1% FCS and 2 mM EDTA, after which the cells were fixed with 100 µl of 1% paraformaldehyde in PBS for 10 minutes. Afterward, cells were washed twice using PBS supplemented with 1% FCS and 2 mM EDTA, after which they were resuspended in 100 µl of the same buffer. Flow cytometry was performed using the BD FACSCanto II (BD Biosciences) using appropriate laser voltages. Alexa Fluor 488 signal (HAs) was measured using the FITC filter and Alexa Fluor 555 signal (anti-sialyl-LewisX signal) was measured using the PE filter. Data were analyzed using FlowLogic (Inivai Technologies) and gated for cells, single cells, and alive cells as described previously [59]. The mean value and standard deviation were calculated over triplicate measurements.

### Sequence analysis

The prevalence of the five equine amino acid residues on positions 128, 130, 135, 189, and 193 (numbering according to H3 HA) was analyzed in all equine and avian H7 sequences with a minimum length of 1,600 bp. The protein sequences were downloaded from GISAID (date of download: 31.08.2023) and aligned using MAFFT package in Geneious software. Only the receptor binding site of the different sequences was compared. In total 3,402 avian H7 and 24 equine H7 sequences were analyzed.

## Data analysis and statistical analysis

The data in this article were (statistically) analyzed and visualized using GraphPad Prism 9.2.0.

## Acknowledgments

We would like to thank Dajana Helke, Nadine Bock, Luise Hohensee, and Bastiaan Vroege for laboratory technical assistance. Furthermore, we thank the animal caretakers and the department of experimental animal facilities and biorisk management at the FLI-Riems for their support in the animal experiments. Timm C. Harder, Ahmed Mostafa, Stephan Pleschka, and Beate Crossley are thanked for providing the viruses, and Stefan Finke for providing the pCAGGS plasmid used in this study. We thank the authors who submitted nucleotide sequences in the GISAID that were used for analysis in this study.

## Funding

This research was made possible by funding from ICRAD, an ERA-NET co-funded under the European Union's Horizon 2020 research and innovation programme (<https://ec.europa.eu/programmes/horizon2020/en>), under Grant Agreement n°862605 (Flu-Switch) to R.P. de Vries and E.M. Abdelwhab. This work was also supported by grants from the Deutsche Forschungsgemeinschaft (DFG, AB 567, to E.M. Abdelwhab), European Commission (ERC Starting Grant 802780 to R.P. de Vries), and the Mizutani Foundation for Glycoscience (Research Grant 2023 to R.P. de Vries).

## Data deposition

The released *N*-glycan raw data (for the glycomic analyses of the cell lines) have been deposited to the GlycoPOST repository (Watanabe, Y., Aoki-Kinoshita, K.F., et al. 2021) under the identifier GPST000345.

## Ethics statement

All performed animal experiments were approved and monitored by the commissioner for animal welfare at the Friedrich-Loeffler-Institut (FLI), representing the Institutional Animal Care and Use Committee (IACUC). The animal trial was performed under biosafety level 3 (BSL3) conditions in the animal facilities of FLI according to the German Regulations for Animal Welfare. The experiments were approved by the authorized ethics committee of the State Office of Agriculture, Food Safety, and Fishery in Mecklenburg-Western Pomerania (LALLF M-V) under registration number 7221.3–1.1-051-12. The specific-pathogen-free (SPF) embryonated chicken eggs (ECE) used for this publication were purchased from Valo BioMedia (Germany). The storage and handling of SPF-ECE were performed according to the guidelines of the World Organization for Animal Health (WOAH).

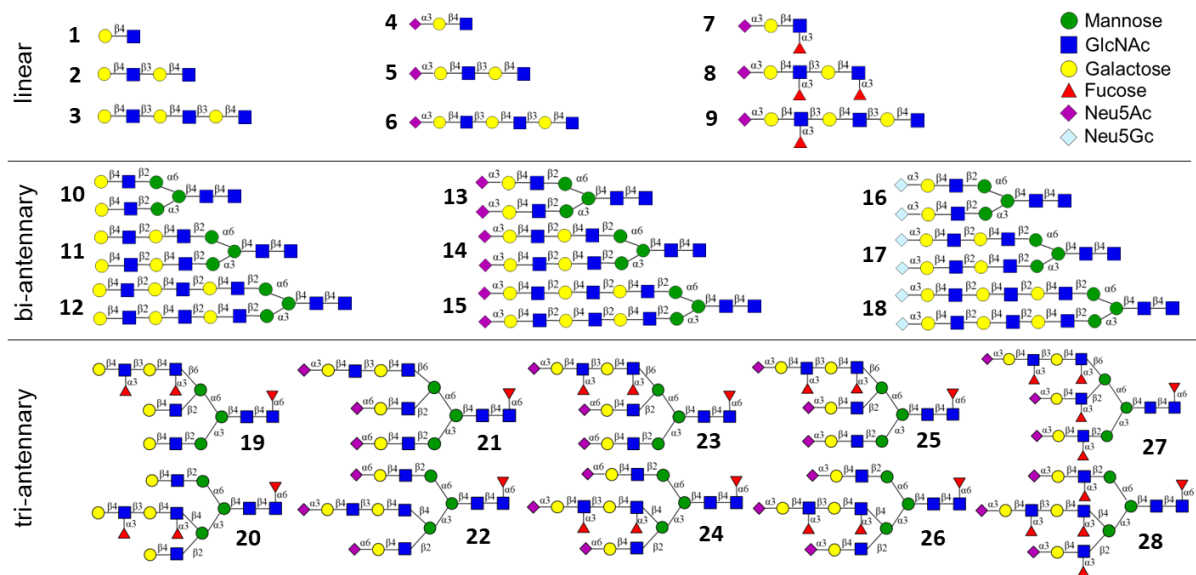
## Author contributions

Cell culturing, C.M.S., A.G.G., D.I.P.; fucosidase production, C.M.S.; protein histochemistry, C.M.S., A.G.G.; expression and production of hemagglutinins,

C.M.S., M.R.C., A.G.G.; preparation and analysis of glycan microarrays, C.M.S.; execution of glycan microarray experiments, R.P.dV.; sample preparation for mass spectrometry, C.M.S, I.R.S.; mass spectrometry measurements, I.R.S, J.C.L.M, K.R.R.; data analysis of mass spectrometry experiments, C.M.S., J.C.L.M.; flow cytometry, C.M.S., A.G.G., design of the mutations in the virus, C.M.S.; *in vitro* characterization of viruses, D.I.P., M.K., D.S., E.M.A.; animal experiments, D.I.P., M.K., D.S., E.M.A.; synthesis of glycans, T.L.; sequence analysis, D.I.P.; data analysis and visualization, C.M.S., D.I.P.; conceptualization, C.M.S., R.P.dV., D.I.P, E.M.A.; supervision, R.P.dV., E.M.A., K.R.R., funding acquisition, R.P.dV., E.M.A., G.J.B.; writing—original draft preparation, C.M.S., D.I.P.; project administration, R.P.dV., E.M.A.; writing—review and editing, all authors.

## Supplementary information

Figure S1. Glycans presented on the glycan microarray.



**Table S1. Prevalence of different amino acid residues in avian and equine H7 HA sequences.** Sequences with a size of at least 1,600bp were downloaded from GISAID on 31.08.2023 and analyzed using Geneious software.

HA position	Amino acid	Avian H7 (n=3,402)		Equine H7 (n=24)	
		No.	%	No.	%
128	S	2,887	84.86	-	-
	T	2	0.06	24	100
	N	487	14.32	-	-
	D	20	0.59	-	-
	G	3	0.09	-	-
	I	1	0.03	-	-
	Y	1	0.03	-	-
130	R	1	0.03	-	-
	I	3,356	98.65	-	-
	V	8	0.24	24	100
135	M	38	1.12	-	-
	A	2,574	75.66	-	-
	E	8	0.24	24	100
	V	522	15.34	-	-
	T	295	8.67	-	-
	G	1	0.03	-	-
	K	1	0.03	-	-
189	S	1	0.03	-	-
	T	1,752	51.50	-	-
	A	1,297	38.12	24	100
	S	168	4.94	-	-
	D	93	2.73	-	-
	E	43	1.26	-	-
	N	41	1.21	-	-
	I	5	0.15	-	-
	G	1	0.03	-	-
193	K	1	0.03	-	-
	K	3,360	98.77	1	4.17
	R	37	1.09	23	95.83
	M	3	0.09	-	-
	N	1	0.03	-	-
	Q	1	0.03	-	-



**Table S2. Relative abundance of N-glycans of MDCK FUT cells measured using HILIC-IMS-QTOF positive mode mass spectrometry.** The data from MDCK WT cells is published in Table SII of [59].

Relative abundance (%)	Retention time (min)	Mass (Da)	Hex	Hex NAc	Fuc	NeuAc	Relative abundance (%)	Retention time (min)	Mass (Da)	Hex	Hex NAc	Fuc	NeuAc
4.197	18.4	2501.02	5	5	3	0	0.169	21.9	3539.40	8	7	4	0
3.305	13.8	1681.72	3	4	1	0	0.169	18.0	2501.01	5	5	3	0
2.942	16.1	1615.66	6	2	0	0	0.164	18.3	2542.04	4	6	3	0
2.922	19.7	2101.82	9	2	0	0	0.162	18.0	2499.00	3	3	4	2
2.754	14.5	1453.60	5	2	0	0	0.160	18.2	3157.25	6	6	3	1
2.740	17.7	2297.94	5	4	3	0	0.160	16.2	2109.83	6	3	0	1
2.694	20.0	3012.21	6	6	4	0	0.160	18.4	2808.11	6	5	2	1
2.334	19.0	3157.24	6	6	3	1	0.158	18.2	3594.40	6	6	4	2
2.331	16.7	2192.91	4	5	2	0	0.157	14.5	1640.69	4	3	1	0
2.114	16.6	2807.08	6	5	0	2	0.156	16.5	1761.72	6	2	1	0
2.030	16.7	2442.98	5	4	2	1	0.153	21.5	3976.55	8	7	5	1
2.001	17.5	2646.06	5	5	2	1	0.152	18.7	3341.25	10	3	2	2
1.993	14.7	1884.80	3	5	1	0	0.151	14.8	1681.71	3	4	1	0
1.569	17.7	2354.96	5	5	2	0	0.147	18.4	2135.88	4	4	3	0
1.524	16.9	2151.88	5	4	2	0	0.146	13.8	1535.66	3	4	0	0
1.371	18.7	1939.77	8	2	0	0	0.146	13.6	1972.81	3	4	1	1
1.359	16.8	3098.18	6	5	0	3	0.145	19.4	2904.10	10	3	1	1
1.350	17.3	3098.19	6	5	0	3	0.144	17.5	2192.91	4	5	2	0
1.274	14.5	1681.72	3	4	1	0	0.141	16.7	1989.83	4	4	2	0

1.260	16.1	1453.61	5	2	0	0		0.141	19.6	2297.94	5	4	3	0
1.050	15.5	2150.86	5	4	0	1		0.141	13.3	1535.66	3	4	0	0
1.034	17.4	2379.99	3	6	3	0		0.137	19.0	2501.01	5	5	3	0
0.992	13.8	1478.63	3	3	1	0		0.136	18.4	1939.76	8	2	0	0
0.990	19.4	2663.07	6	5	3	0		0.135	19.0	2686.05	4	6	0	2
0.944	17.5	2338.97	4	5	3	0		0.134	17.6	1615.66	6	2	0	0
0.887	15.3	2441.96	5	4	0	2		0.131	13.9	1827.78	3	4	2	0
0.884	16.3	2515.99	6	5	0	1		0.131	17.4	1615.66	6	2	0	0
0.881	15.8	2087.88	3	6	1	0		0.130	20.0	2354.96	5	5	2	0
0.863	18.7	3303.30	6	6	4	1		0.129	20.3	3815.50	9	9	2	0
0.837	16.9	2515.99	6	5	0	1		0.127	14.4	2077.84	4	3	2	1
0.795	19.4	2866.15	6	6	3	0		0.124	15.3	1884.80	3	5	1	0
0.772	20.5	3158.26	6	6	5	0		0.124	16.7	2500.00	5	5	1	1
0.757	14.1	1640.69	4	3	1	0		0.122	21.3	2809.13	6	5	4	0
0.735	17.5	2151.88	5	4	2	0		0.121	20.5	3196.21	10	3	3	1
0.731	20.0	2501.02	5	5	3	0		0.119	16.6	2483.99	4	5	2	1
0.730	17.4	1777.71	7	2	0	0		0.115	18.2	2808.11	6	5	2	1
0.728	15.8	2046.85	4	5	1	0		0.113	14.9	2280.92	4	4	2	1
0.725	20.9	3174.26	7	6	4	0		0.113	21.2	3415.29	11	4	2	1
0.694	15.1	1884.80	3	5	1	0		0.109	16.7	2805.08	4	3	1	4
0.692	19.6	2809.13	6	5	4	0		0.108	18.1	3245.26	6	5	3	2
0.688	20.0	3050.16	10	3	2	1		0.106	19.6	3011.20	6	6	2	1
0.620	17.2	2807.09	6	5	0	2		0.103	16.7	2515.99	6	5	0	1
0.583	16.3	1948.80	5	3	2	0		0.103	19.0	2646.06	5	5	2	1
0.575	20.0	2866.15	6	6	3	0		0.102	16.2	1802.75	5	3	1	0
0.574	15.9	2296.92	5	4	1	1		0.101	17.6	2395.98	4	6	2	0
0.572	21.7	3523.40	7	7	5	0		0.098	19.4	2354.96	5	5	2	0
0.567	17.6	1777.71	7	2	0	0		0.096	20.0	3503.30	11	3	2	2
0.534	21.2	3377.34	7	7	4	0		0.096	19.3	2850.15	5	6	4	0
0.518	15.9	1989.83	4	4	2	0		0.095	18.0	2558.04	5	6	2	0
0.515	16.0	2807.09	6	5	0	2		0.094	15.9	1802.74	5	3	1	0
0.507	16.1	2005.82	5	4	1	0		0.093	18.3	2296.93	5	4	1	1
0.494	18.4	2354.96	5	5	2	0		0.093	17.0	2005.82	5	4	1	0
0.469	15.4	2093.84	5	3	1	1		0.092	20.5	2647.07	5	5	4	0
0.468	18.5	2954.16	6	5	3	1		0.091	16.3	2290.95	3	7	1	0
0.444	15.3	2087.88	3	6	1	0		0.091	17.1	2338.96	4	5	3	0
0.435	20.5	2263.87	10	2	0	0		0.091	20.6	3253.24	10	4	2	1
0.424	15.8	1884.80	3	5	1	0		0.090	16.6	2046.85	4	5	1	0
0.402	16.2	2484.00	4	5	2	1		0.090	17.2	2208.90	5	5	1	0
0.400	18.0	3389.28	6	5	0	4		0.086	17.7	1786.74	4	3	2	0
0.388	15.2	1786.75	4	3	2	0		0.084	20.5	2501.02	5	5	3	0
0.382	15.7	1859.77	5	4	0	0		0.082	18.3	3011.19	6	6	2	1
0.379	16.4	2233.93	3	6	2	0		0.081	16.6	1623.68	3	3	0	1
0.379	21.3	3320.32	7	6	5	0		0.080	17.5	2280.92	4	4	2	1
0.373	14.7	1640.69	4	3	1	0		0.079	20.0	2954.16	6	5	3	1
0.369	16.4	2296.92	5	4	1	1		0.077	19.4	2501.01	5	5	3	0
0.366	18.7	2704.09	5	6	3	0		0.075	15.4	2337.94	4	5	1	1
0.361	14.7	2150.86	5	4	0	1		0.074	12.9	1478.64	3	3	1	0
0.353	18.4	1989.82	4	4	2	0		0.074	15.2	2134.86	4	4	1	1
0.349	17.5	2134.86	4	4	1	1		0.072	17.4	2233.93	3	6	2	0
0.349	17.2	2792.11	5	5	3	1		0.070	20.7	2663.07	6	5	3	0
0.341	20.8	3212.21	11	3	2	1		0.068	17.8	2849.14	5	6	2	1
0.337	21.3	3358.27	11	3	3	1		0.067	17.1	2167.88	6	4	1	0
0.335	15.0	2134.87	4	4	1	1		0.067	18.1	2808.11	6	5	2	1
0.330	17.3	2441.96	5	4	0	2		0.066	15.5	1989.83	4	4	2	0
0.322	17.3	1964.80	6	3	1	0		0.066	16.7	2296.92	5	4	1	1
0.322	20.6	3215.28	6	7	4	0		0.065	15.6	2337.94	4	5	1	1
0.320	19.0	3011.19	6	6	2	1		0.064	14.5	2175.89	3	5	1	1
0.317	19.0	3195.19	10	3	1	2		0.063	16.4	1818.74	6	3	0	0
0.315	17.1	2233.93	3	6	2	0		0.062	18.3	2558.04	5	6	2	0
0.312	18.5	3448.34	6	6	3	2		0.062	14.5	1535.66	3	4	0	0
0.308	14.9	1599.67	5	2	1	0		0.060	19.9	2809.12	6	5	4	0
0.308	16.2	2525.03	3	6	2	1		0.059	16.8	1623.68	3	3	0	1
0.278	22.4	3685.45	8	7	5	0		0.058	18.6	2313.93	6	4	2	0
0.273	17.5	2500.00	5	5	1	1		0.058	16.6	2441.95	5	4	0	2
0.270	16.0	2441.96	5	4	0	2		0.057	13.7	2077.85	4	3	2	1
0.269	16.8	2441.95	5	4	0	2		0.056	16.2	2249.93	4	6	1	0
0.267	20.4	3012.21	6	6	4	0		0.056	16.6	1948.80	5	3	2	0
0.265	16.5	2255.90	6	3	1	1		0.055	16.3	2588.02	5	4	1	2
0.259	15.1	1681.72	3	4	1	0		0.054	13.7	1494.64	4	3	0	0
0.257	15.0	1843.77	4	4	1	0		0.053	16.2	2513.98	4	3	1	3
0.256	17.3	2224.90	6	5	0	0		0.053	16.0	1623.68	3	3	0	1
0.253	15.7	2030.85	3	5	2	0		0.052	16.3	1332.58	3	3	0	0
0.251	17.5	2110.85	6	3	2	0		0.050	18.1	2662.05	6	5	1	1
0.247	21.7	3561.35	11	4	3	1		0.050	16.3	2150.86	5	4	0	1
0.240	20.1	3465.36	7	6	4	1		0.047	12.6	1478.64	3	3	1	0

0.236	15.5	2484.00	4	5	2	1		0.047	14.6	1988.81	4	4	0	1
0.234	17.1	2395.99	4	6	2	0		0.046	22.1	3520.32	12	3	3	1
0.234	14.5	1786.75	4	3	2	0		0.046	19.7	1777.71	7	2	0	0
0.229	20.0	2647.07	5	5	4	0		0.046	15.0	1947.78	5	3	0	1
0.228	17.1	2192.90	4	5	2	0		0.045	20.9	2425.92	11	2	0	0
0.228	21.6	3012.20	6	6	4	0		0.044	16.7	1931.79	4	3	1	1
0.226	18.1	2354.96	5	5	2	0		0.044	14.8	2030.85	3	5	2	0
0.226	14.2	1681.72	3	4	1	0		0.044	19.4	2889.11	9	3	4	0
0.224	17.4	2417.94	7	3	1	1		0.043	21.6	2587.98	12	2	0	0
0.217	17.9	3098.19	6	5	0	3		0.043	16.9	1859.76	5	4	0	0
0.216	13.8	1738.74	3	5	0	0		0.043	17.0	2937.15	5	5	2	2
0.213	14.1	1931.79	4	3	1	1		0.043	17.5	1989.83	4	4	2	0
0.212	14.9	1786.75	4	3	2	0		0.043	19.7	1939.76	8	2	0	0
0.202	16.8	2208.90	5	5	1	0		0.041	17.6	1453.61	5	2	0	0
0.202	18.7	1777.71	7	2	0	0		0.039	16.1	2255.89	6	3	1	1
0.200	20.6	3668.44	7	7	4	1		0.039	14.3	1738.74	3	5	0	0
0.200	16.5	2151.87	5	4	2	0		0.038	16.3	1859.76	5	4	0	0
0.198	18.7	3158.25	8	8	0	0		0.037	15.0	1656.69	5	3	0	0
0.196	15.4	1802.74	5	3	1	0		0.036	14.1	2077.84	4	3	2	1
0.195	15.9	2337.94	4	5	1	1		0.033	15.0	2296.92	5	4	1	1
0.195	19.0	2663.07	6	5	3	0		0.032	16.5	1599.66	5	2	1	0
0.195	18.1	2313.93	6	4	2	0		0.031	16.4	2249.93	4	6	1	0
0.191	16.3	2046.85	4	5	1	0		0.030	16.8	2513.98	4	3	1	3
0.188	18.8	2459.99	6	4	3	0		0.028	14.4	1988.81	4	4	0	1
0.187	13.7	1785.73	4	3	0	1		0.028	17.1	2647.06	5	5	4	0
0.185	17.3	2297.93	5	4	3	0		0.027	17.7	2605.03	6	4	2	1
0.183	14.6	1827.78	3	4	2	0		0.024	15.3	1785.72	4	3	0	1
0.180	16.5	2005.82	5	4	1	0		0.024	16.2	2337.95	4	5	1	1
0.179	19.8	3319.30	7	6	3	1		0.024	16.2	2562.98	7	3	0	2
0.176	15.6	1843.77	4	4	1	0		0.024	18.7	1615.66	6	2	0	0
0.175	15.0	1827.78	3	4	2	0		0.022	16.2	2378.98	3	6	1	1
0.172	19.0	2647.07	5	5	4	0		0.021	15.5	1941.82	3	6	0	0
0.172	17.0	2150.86	5	4	0	1		0.021	16.7	1332.58	3	3	0	0
0.171	21.7	3567.36	6	3	4	4		0.019	14.7	1494.63	4	3	0	0
0.170	18.7	2517.01	6	5	2	0		0.018	17.1	2805.08	4	3	1	4
0.170	14.3	1884.80	3	5	1	0								



## References

1. Hutchinson, E.C., *Influenza virus*. Trends Microbiol, 2018. 26(9): p. 809-810.
2. Abdelwhab, E.M. and T.C. Mettenleiter, *Zoonotic animal influenza virus and potential mixing vessel hosts*. Viruses, 2023. 15(4).
3. Fouchier, R.A. and V.J. Munster, *Epidemiology of low pathogenic avian influenza viruses in wild birds*. Rev Sci Tech, 2009. 28(1): p. 49-58.
4. Olsen, B., et al., *Global patterns of influenza A virus in wild birds*. Science, 2006. 312(5772): p. 384-8.
5. Causey, D. and S.V. Edwards, *Ecology of avian influenza virus in birds*. J Infect Dis, 2008. 197 Suppl 1(Suppl 1): p. S29-33.
6. Lee, D.H., M.F. Criado, and D.E. Swayne, *Pathobiological origins and evolutionary history of highly pathogenic avian influenza viruses*. Cold Spring Harb Perspect Med, 2021. 11(2).
7. Webster, R.G. and E.A. Govorkova, *Continuing challenges in influenza*. Ann N Y Acad Sci, 2014. 1323(1): p. 115-39.
8. Shi, J., et al., *Alarming situation of emerging H5 and H7 avian influenza and effective control strategies*. Emerg Microbes Infect, 2023. 12(1): p. 2155072.
9. Javanian, M., et al., *A brief review of influenza virus infection*. J Med Virol, 2021. 93(8): p. 4638-4646.
10. Chambers, T.M., *A brief introduction to equine influenza and equine influenza viruses*. Methods Mol Biol, 2020. 2123: p. 355-360.
11. Banbura, M.W., et al., *Reassortants with equine 1 (H7N7) influenza virus hemagglutinin in an avian influenza virus genetic background are pathogenic in chickens*. Virology, 1991. 184(1): p. 469-71.
12. Webster, R.G., et al., *Evolution and ecology of influenza A viruses*. Microbiol Rev, 1992. 56(1): p. 152-79.
13. Webster, R.G., *Are equine 1 influenza viruses still present in horses?* Equine Vet J, 1993. 25(6): p. 537-8.
14. Long, J.S., et al., *Host and viral determinants of influenza A virus species specificity*. Nat Rev Microbiol, 2019. 17(2): p. 67-81.
15. Imai, M. and Y. Kawaoka, *The role of receptor binding specificity in interspecies transmission of influenza viruses*. Curr Opin Virol, 2012. 2(2): p. 160-7.
16. Ji, Y., et al., *New insights into influenza A specificity: an evolution of paradigms*. Curr Opin Struct Biol, 2017. 44: p. 219-231.
17. de Graaf, M. and R.A. Fouchier, *Role of receptor binding specificity in influenza A virus transmission and pathogenesis*. EMBO J, 2014. 33(8): p. 823-41.
18. Mair, C.M., et al., *Receptor binding and pH stability - how influenza A virus hemagglutinin affects host-specific virus infection*. Biochim Biophys Acta, 2014. 1838(4): p. 1153-68.
19. Krammer, F., et al., *Influenza*. Nat Rev Dis Primers, 2018. 4(1): p. 3.
20. Byrd-Leotis, L., et al., *Antigenic pressure on H3N2 influenza virus drift strains imposes constraints on binding to sialylated receptors but not phosphorylated glycans*. J Virol, 2019. 93(22).
21. Wu, N.C. and I.A. Wilson, *Influenza hemagglutinin structures and antibody recognition*. Cold Spring Harb Perspect Med, 2020. 10(8).
22. Ge, S. and Z. Wang, *An overview of influenza A virus receptors*. Crit Rev Microbiol, 2011. 37(2): p. 157-65.
23. Gambaryan, A., et al., *Receptor specificity of influenza viruses from birds and mammals: new data on involvement of the inner fragments of the carbohydrate chain*. Virology, 2005. 334(2): p. 276-83.
24. Unione, L., et al., *Probing altered receptor specificities of antigenically drifting human H3N2 viruses by chemoenzymatic synthesis, NMR and modeling*. BiorXiv, 2023.
25. Miller-Podraza, H., et al., *A strain of human influenza A virus binds to extended but not short gangliosides as assayed by thin-layer chromatography overlay*. Glycobiology, 2000. 10(10): p. 975-82.
26. Gambaryan, A.S., et al., *Receptor-binding profiles of H7 subtype influenza viruses in different host species*. J Virol, 2012. 86(8): p. 4370-9.
27. Broszeit, F., et al., *N-glycolylneuraminic acid as a receptor for influenza A viruses*. Cell Rep, 2019. 27(11): p. 3284-3294 e6.
28. Yasue, S., et al., *Difference in form of sialic acid in red blood cell glycolipids of different breeds of dogs*. J Biochem, 1978. 83(4): p. 1101-7.
29. Schauer, R., et al., *Low incidence of N-glycolylneuraminic acid in birds and reptiles and its absence in the platypus*. Carbohydr Res, 2009. 344(12): p. 1494-500.
30. Ng, P.S., et al., *Ferrets exclusively synthesize Neu5Ac and express naturally humanized influenza A virus receptors*. Nat Commun, 2014. 5: p. 5750.
31. Perí, S., et al., *Phylogenetic distribution of CMP-Neu5Ac hydroxylase (CMAH), the enzyme synthesizing the proinflammatory human xenoantigen Neu5Gc*. Genome Biol Evol, 2018. 10(1): p. 207-219.
32. Nemanichvili, N., et al., *Wild and domestic animals variably display Neu5Ac and Neu5Gc sialic acids*. Glycobiology, 2022. 32(9): p. 791-802.
33. Spruit, C.M., et al., *N-glycolylneuraminic acid binding of avian and equine H7 influenza A viruses*. J Virol, 2022. 96(5): p. e0212021.
34. Allen, J.D. and T.M. Ross, *H3N2 influenza viruses in humans: viral mechanisms, evolution, and evaluation*. Hum Vaccin Immunother, 2018. 14(8): p. 1840-1847.
35. Gischke, M., et al., *The role of glycosylation in the N-terminus of the hemagglutinin of a unique H4N2 with a natural polybasic cleavage site in virus fitness in vitro and in vivo*. Virulence, 2021. 12(1): p. 666-678.
36. Russell, C.J., *Hemagglutinin stability and its impact on influenza A virus infectivity, pathogenicity, and transmissibility in avians, mice, swine, seals, ferrets, and humans*. Viruses, 2021. 13(5).
37. Park, S.L., et al., *Virus-specific thermostability and heat inactivation profiles of alphaviruses*. J Virol Methods, 2016. 234: p. 152-5.

38. Scheibner, D., et al., Genetic determinants for virulence and transmission of the panzootic avian influenza virus H5N8 clade 2.3.4.4 in pekin ducks. *J Virol*, 2022. 96(13): p. e0014922.
39. Tan, M., et al., Saliva as a source of reagent to study human susceptibility to avian influenza H7N9 virus infection. *Emerg Microbes Infect*, 2018. 7(1): p. 156.
40. Guan, M., et al., The sialyl Lewis X glycan receptor facilitates infection of subtype H7 avian influenza A viruses. *J Virol*, 2022. 96(19): p. e0134422.
41. Gambaryan, A.S., et al., 6-sulfo sialyl Lewis X is the common receptor determinant recognized by H5, H6, H7 and H9 influenza viruses of terrestrial poultry. *Virol J*, 2008. 5: p. 85.
42. Heider, A., et al., Alterations in hemagglutinin receptor-binding specificity accompany the emergence of highly pathogenic avian influenza viruses. *J Virol*, 2015. 89(10): p. 5395-405.
43. Hiono, T., et al., Amino acid residues at positions 222 and 227 of the hemagglutinin together with the neuraminidase determine binding of H5 avian influenza viruses to sialyl Lewis X. *Arch Virol*, 2016. 161(2): p. 307-16.
44. Gambaryan, A.S., et al., Receptor-binding properties of influenza viruses isolated from gulls. *Virology*, 2018. 522: p. 37-45.
45. Wen, F., et al., Mutation W222L at the receptor binding site of hemagglutinin could facilitate viral adaption from equine influenza A(H3N8) virus to dogs. *J Virol*, 2018. 92(18).
46. Chen, C., et al., Enzymatic modular synthesis and microarray assay of poly-N-acetyllactosamine derivatives. *Chem Commun (Camb)*, 2020. 56(55): p. 7549-7552.
47. Ichimiya, T., et al., Sulfated glycans containing NeuAalpha2-3Gal facilitate the propagation of human H1N1 influenza A viruses in eggs. *Virology*, 2021. 562: p. 29-39.
48. Verhagen, J.H., et al., Host range of influenza A virus H1 to H16 in Eurasian ducks based on tissue and receptor binding studies. *J Virol*, 2021. 95(6).
49. Gaide, N., et al., Pathobiology of highly pathogenic H5 avian influenza viruses in naturally infected Galliformes and Anseriformes in France during winter 2015-2016. *Vet Res*, 2022. 53(1): p. 11.
50. Hiono, T., et al., A chicken influenza virus recognizes fucosylated alpha2,3 sialoglycan receptors on the epithelial cells lining upper respiratory tracts of chickens. *Virology*, 2014. 456-457: p. 131-8.
51. Guo, H., et al., Highly pathogenic influenza A(H5Nx) viruses with altered H5 receptor-binding specificity. *Emerg Infect Dis*, 2017. 23(2): p. 220-231.
52. Kobayashi, D., et al., Turkeys possess diverse Siaalpha2-3Gal glycans that facilitate their dual susceptibility to avian influenza viruses isolated from ducks and chickens. *Virus Res*, 2022. 315: p. 198771.
53. Suzuki, N., T. Abe, and S. Natsuka, Structural analysis of N-glycans in chicken trachea and lung reveals potential receptors of chicken influenza viruses. *Sci Rep*, 2022. 12(1): p. 2081.
54. Walther, T., et al., Glycomic analysis of human respiratory tract tissues and correlation with influenza virus infection. *PLoS Pathog*, 2013. 9(3): p. e1003223.
55. Sriwilaijaroen, N., et al., N-glycan structures of human alveoli provide insight into influenza A virus infection and pathogenesis. *FEBS J*, 2018. 285(9): p. 1611-1634.
56. Wu, H., et al., Fucosidases from the human gut symbiont *Ruminococcus gnavus*. *Cell Mol Life Sci*, 2021. 78(2): p. 675-693.
57. Xiong, X., et al., Recognition of sulphated and fucosylated receptor sialosides by A/Vietnam/1194/2004 (H5N1) influenza virus. *Virus Res*, 2013. 178(1): p. 12-4.
58. Hennion, R.M. and G. Hill, The preparation of chicken kidney cell cultures for virus propagation. *Methods Mol Biol*, 2015. 1282: p. 57-62.
59. Spruit, C.M., et al., Contemporary human H3N2 influenza A viruses require a low threshold of suitable glycan receptors for efficient infection. *Glycobiology*, 2023.
60. Matrosovich, M.N. and A.S. Gambaryan, Solid-phase assays of receptor-binding specificity. *Methods Mol Biol*, 2012. 865: p. 71-94.
61. Zaraket, H., et al., Increased acid stability of the hemagglutinin protein enhances H5N1 influenza virus growth in the upper respiratory tract but is insufficient for transmission in ferrets. *J Virol*, 2013. 87(17): p. 9911-22.
62. Spiro, R.G., Studies on fetuin, a glycoprotein of fetal serum. I. Isolation, chemical composition, and physicochemical properties. *J Biol Chem*, 1960. 235(10): p. 2860-9.
63. Russell, R.J., et al., H1 and H7 influenza haemagglutinin structures extend a structural classification of haemagglutinin subtypes. *Virology*, 2004. 325(2): p. 287-96.
64. de Vries, R.P., et al., The influenza A virus hemagglutinin glycosylation state affects receptor-binding specificity. *Virology*, 2010. 403(1): p. 17-25.
65. Nemanichvili, N., et al., Fluorescent trimeric hemagglutinins reveal multivalent receptor binding properties. *J Mol Biol*, 2019. 431(4): p. 842-856.
66. Shaner, N.C., et al., Improving the photostability of bright monomeric orange and red fluorescent proteins. *Nat Methods*, 2008. 5(6): p. 545-51.
67. Li, T., et al., Synthesis of tri-antennary N-glycans terminated with sialyl-LewisX reveals the relevance of glycan complexity for influenza A viruses receptor binding. manuscript in preparation, 2024.
68. Wickramasinghe, I.N., et al., Binding of avian coronavirus spike proteins to host factors reflects virus tropism and pathogenicity. *J Virol*, 2011. 85(17): p. 8903-12.
69. Bouwman, K.M., et al., Three amino acid changes in avian coronavirus spike protein allow binding to kidney tissue. *J Virol*, 2020. 94(2).
70. Cooper, C.A., E. Gasteiger, and N.H. Packer, GlycoMod - a software tool for determining glycosylation compositions from mass spectrometric data. *Proteomics*, 2001. 1(2): p. 340-9.

# Chapter 6

## Synthesis of tri-antennary *N*-glycans terminated with sialyl-LewisX reveals the relevance of glycan complexity for influenza A virus receptor binding

*Cindy M. Spruit, Tiehai Li, Na Wei, Lin Liu, Margreet Wolfert, Robert P. de Vries, and Geert-Jan Boons*

### Abstract

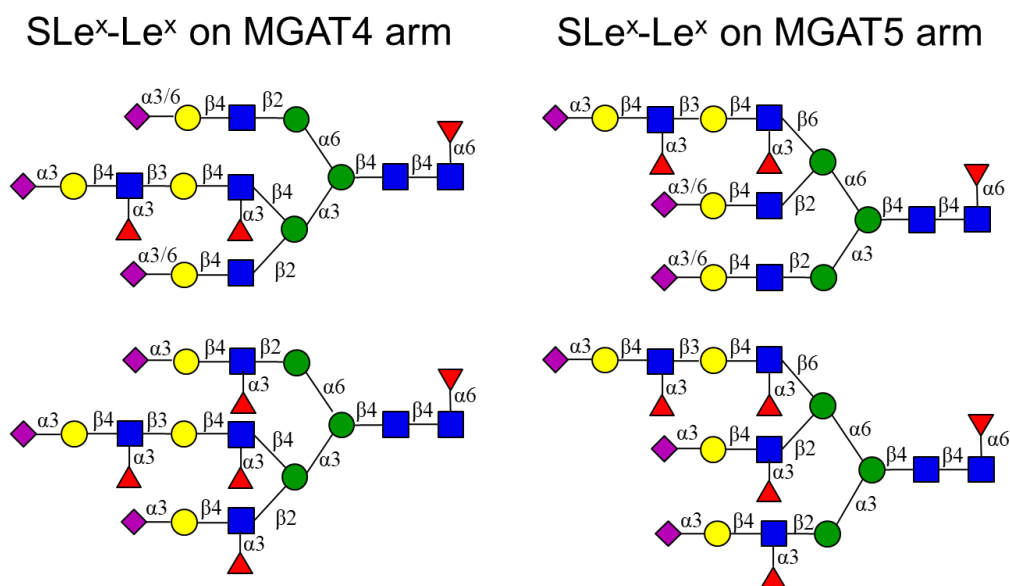
Sialyl-LewisX (SLe<sup>x</sup>) epitopes are widely involved in the immune system, human fertilization, cancer, and bacterial and viral diseases. However, the influence of the complex glycan structures, which present SLe<sup>x</sup> epitopes, on binding is largely unknown. We report a chemoenzymatic strategy for the preparation of a panel of asymmetrical tri-antennary *N*-glycans presenting SLe<sup>x</sup>-Le<sup>x</sup> epitopes. In glycan microarray binding studies, we showed that E-selectin bound equally well to linear glycans presenting SLe<sup>x</sup>(-Le<sup>x</sup>) epitopes and tri-antennary *N*-glycans on which the SLe<sup>x</sup>-Le<sup>x</sup> was presented on either the MGAT4 or MGAT5 arm. On the other hand, we showed a strong preference of H5 influenza A virus (IAV) hemagglutinins (HA) for the tri-antennary *N*-glycan structures when compared to linear structures terminating in SLe<sup>x</sup>. Furthermore, some H5 HAs showed increased binding when the SLe<sup>x</sup> epitope was presented on the MGAT4 arm, in contrast to the MGAT5 arm. SLe<sup>x</sup> is displayed in the respiratory tract of several avian species, demonstrating the relevance of investigating the binding of, among others IAVs, to complex *N*-glycans presenting SLe<sup>x</sup>.

## Introduction

ABO(H) and Lewis antigens were initially discovered on erythrocytes and therefore designated as histo-blood group antigens. These carbohydrate structures are, however, widely distributed and found on many other cell types including epithelial- and vascular endothelial cells, sensory neurons, and oocytes, where they are involved in a multitude of biological processes [1, 2].

Sialyl Lewis<sup>x</sup> (SLe<sup>x</sup>), which is composed of a type-2 *N*-acetyl lactosamine (Gal $\beta$ 1-4GlcNAc, LacNAc) moiety that is  $\alpha$ (1,3)-fucosylated at GlcNAc and 2,3-sialylated at galactose (Gal), is a Lewis antigen that plays important roles in cell adhesion. For example, E- and P-selectin are C-type lectins expressed by inflamed endothelial cells that can bind SLe<sup>x</sup>-containing cell surface glycoconjugates of circulating leukocytes resulting in cell rolling and vascular extravasation [1]. The overexpression of SLe<sup>x</sup> by cancer cells promotes metastasis by binding to selectins resulting in tumor cell adhesion and extravasation [3]. SLe<sup>x</sup> containing glycoconjugates can also serve as (co-)receptor for several pathogenic bacteria such as *Helicobacter pylori* and *Anaplasma phagocytophilum*, and viruses including noroviruses, human parainfluenza viruses, mumps virus, Sendai virus, MERS-CoV, caliciviruses [4-6]. Influenza A viruses (IAV) from all subtypes, except H15, can also employ SLe<sup>x</sup> containing glycans as receptors [7, 8].

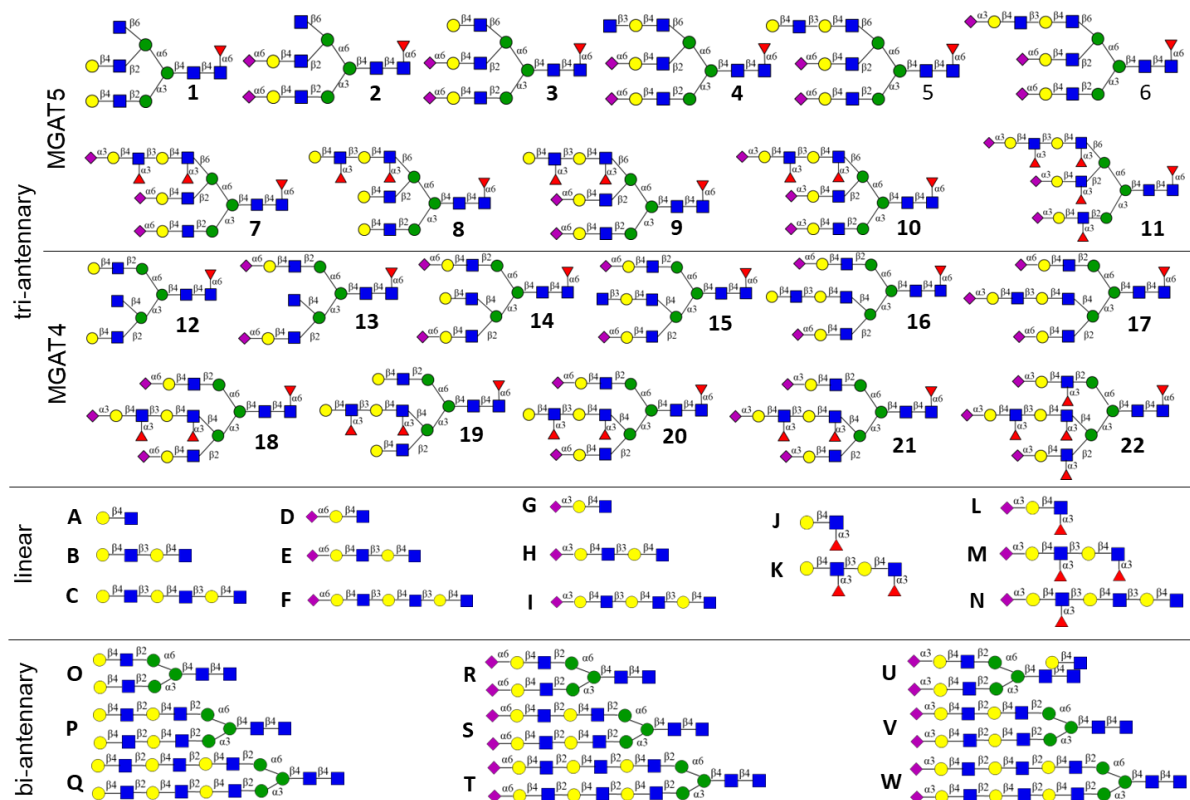
Lewis antigens, such as SLe<sup>x</sup>, are usually part of *N*- and *O*-glycans and glycolipids [9, 10]. The glycan moieties of these biomolecules have usually highly complex architectures. There is data to indicate that the complex architecture of glycans can modulate recognition of minimal epitopes such as SLe<sup>x</sup> [11, 12]. *N*-glycans have a common core pentasaccharide that can be modified by a family of *N*-acetylglucosaminyltransferases (mannosyl-glycoprotein *N*-acetylglucosaminyltransferases or MGATs), to give oligosaccharides having



**Fig 1. Putative ligands for E-selectin.** Glycans similar to the upper left structure, presenting SLe<sup>x</sup>-Le<sup>x</sup> on the extended MGAT4 arm and  $\alpha$ 2,3- or  $\alpha$ 2,6-linked sialic acids on the other antennae, have been suggested as high-affinity ligands for E-selectin.

various numbers and patterns of branching *N*-acetylglucosamine (GlcNAc) moieties that can carry Lewis antigens such as SLe<sup>x</sup>. Affinity purification of *N*-glycans by E-selectin followed by structural elucidation of the isolated compounds indicated that it preferentially binds to multi-antennary *N*-glycan presenting a SLe<sup>x</sup>-Le<sup>x</sup> epitope on the MGAT4 antenna (Fig. 1) [12]. The isolated ligands were further decorated by  $\alpha$ 2,3- or  $\alpha$ 2,6-linked sialosides at the other antennae [12]. Other observations have indicated that E-selectin preferentially binds to glycans presenting multiple SLe<sup>x</sup> moieties [12-14].

Here, we report a chemoenzymatic strategy for the preparation of a panel of asymmetrical tri-antennary *N*-glycans **1-22** (Fig. 2, Fig. 3) that include putative high-affinity ligands for E-selectin and derivatives thereof. It is composed of two series of isomeric compounds (**1-11** and **12-22**) that have a complex epitope, including SLe<sup>x</sup>-Le<sup>x</sup>, on either the MGAT4 or MGAT5 antenna. Additionally, glycans were synthesized with either one or multiple SLe<sup>x</sup> epitopes. It was expected that the series of compounds would be well-suited to establish the importance of the presentation of SLe<sup>x</sup>-Le<sup>x</sup> on the MGAT4 or MGAT5 antenna and multiple SLe<sup>x</sup> epitopes on the same glycan for recognition by glycan-binding proteins such as E-selectin. The synthetic tri-antennary glycans and several reference compounds (Fig. 2) were printed as a glycan microarray that was probed for recognition by several plant lectins, DC-SIGN, galectins, and E-selectin. It was found that the molecular complexity of the glycans does not influence the recognition of these



**Fig 2.** The glycans that were printed on the glycan microarray presented in this study. Structures 1-22 were synthesized in this study (Fig. 3) and linear and bi-antennary structures A-W were used as reference compounds.

proteins. The microarray was also used to examine receptor specificities of hemagglutinins of several avian H5 IAVs. In these cases, the complexity of the tri-antennary *N*-glycan modulates binding. Not only fucosylation of a 2,3-sialyl LacNAc moiety to give SLe<sup>x</sup> but also the length of a LacNAc chain and presentation at a specific antenna of an *N*-glycan influences receptor binding. Our findings demonstrate the importance of considering the complex glycan structure in binding studies of IAVs and other processes and diseases.

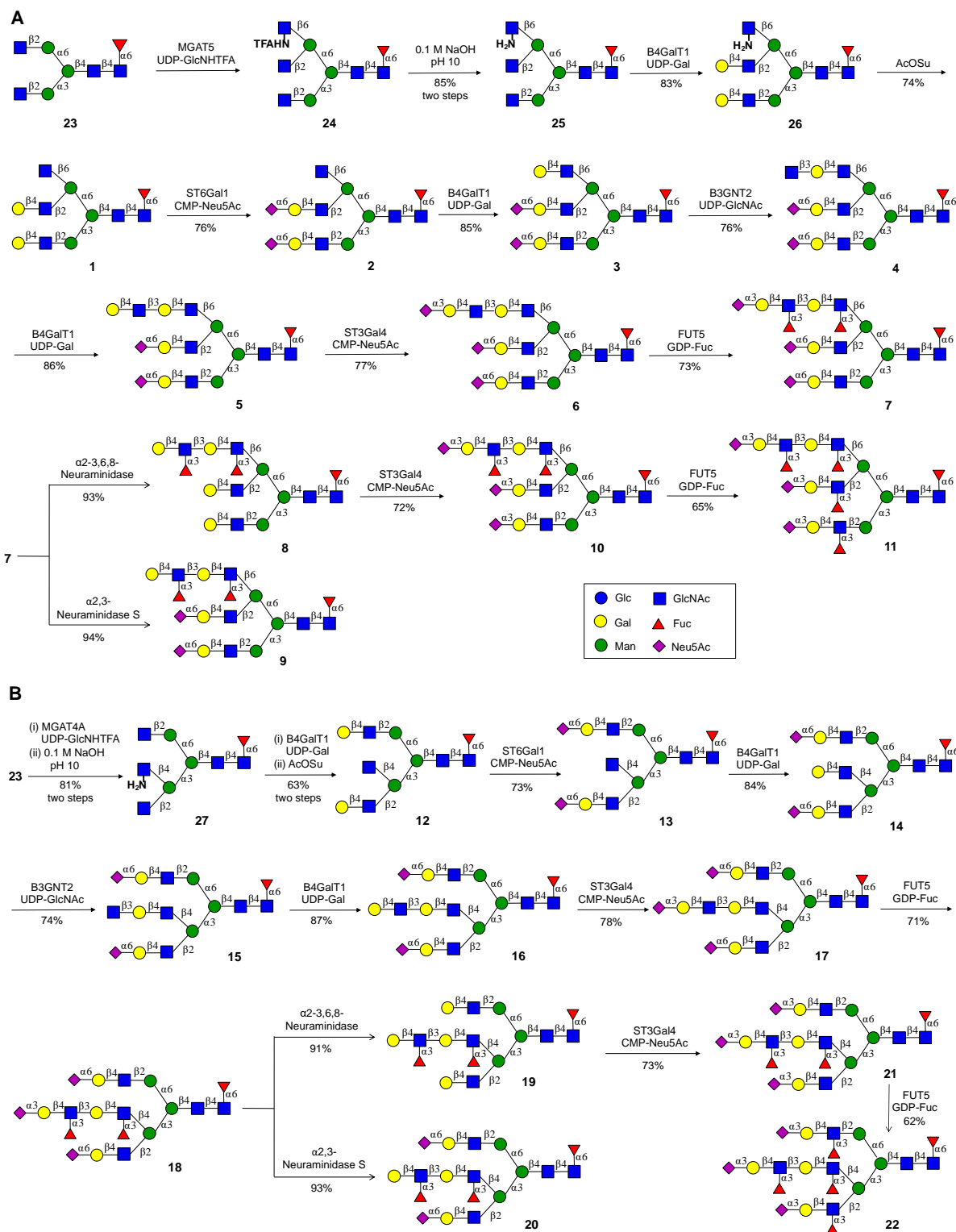
## Results

### Chemoenzymatic Synthesis of isomeric tri-antennary *N*-glycans

The asymmetrical branched tri-antennary *N*-glycans **1-22** were prepared by a chemoenzymatic approach that takes advantage of the inherent substrate preferences of glycosyltransferases. Furthermore, it makes use of a so-called stop-and-go strategy [15] to temporarily disable a specific antenna from enzymatic extension. The preparation of the panel of compounds started from biantennary *N*-glycan **23** that could easily be prepared from a sialoglycopeptide (SGP) isolated from egg yolk powder (Fig. 3) [16, 17]. This compound is a substrate for the enzymes MGAT4 and MGAT5 making it possible to install an additional branching GlcNAc moiety to give tri-antennary glycans. The target compounds have asymmetric architectures differing in epitopes at the MGAT4 or 5 and MGAT1 and 2 antennae. Therefore, a stop-and-go synthetic methodology [15] was employed that allows unique enzymatic modification of the MGAT4 or 5 antenna of tri-antennary glycans. The MGAT4 and 5 enzymes can utilize 5'-diphospho-*N*-trifluoroacetylglucosamine (UDP-GlcNHTFA) as a substrate to give GlcNHTFA-containing tri-antennary glycans. Treatment with sodium hydroxide to remove the TFA moiety was expected to give GlcNH<sub>2</sub> containing glycans **25** and **27**. The GlcNH<sub>2</sub> moiety of the latter derivatives will temporarily block extension by glycosyltransferases, and therefore the GlcNAc moieties of the MGAT1 and 2 arm can be extended by appropriate glycosyltransferases to install specific epitopes. At an appropriate stage in the synthesis, the GlcNH<sub>2</sub> moiety can be acetylated to give natural GlcNAc allowing selective elaboration of the MGAT4 or MGAT5 antenna into complex structures such as SLe<sup>x</sup>-Le<sup>x</sup>. A key strategic consideration was that the MGAT1 and MGAT2 antennae were, in the first instance, modified as 2,6-sialo-LacNAc to give compounds such as **2** and **13**. The latter structures are resistant to modification by most mammalian glycosyltransferases including  $\alpha$ 1,3-fucosyltransferases such as FUT5. Therefore, a 2,3-sialyl-di-LacNAc moiety could be installed at the MGAT4 or 5 antenna that could selectively be fucosylated by FUT5 to form a SLe<sup>x</sup>-Le<sup>x</sup> epitope. Cleavage of the sialosides by a specific- or broad-acting sialidase followed by re-sialylation and further fucosylation was expected to diversify the compounds to give a library of structural analogs.

SGP isolated from egg yolk powder was subsequently treated with a neuraminidase from *Clostridium perfringens* and galactosidase from *Aspergillus niger* to remove the sialosides and galactosides, respectively, to give a G0 structure. The G0 structure was subjected to  $\alpha$ -1,6-fucosyltransferase (FUT8) to install a core fucoside and then to peptide *N*-glycosidase F (PNGaseF) to remove





**Fig 3. Chemoenzymatic synthesis of a panel of isomeric tri-antennary glycans having a complex epitope at either the MGAT4 or 5 antenna.**

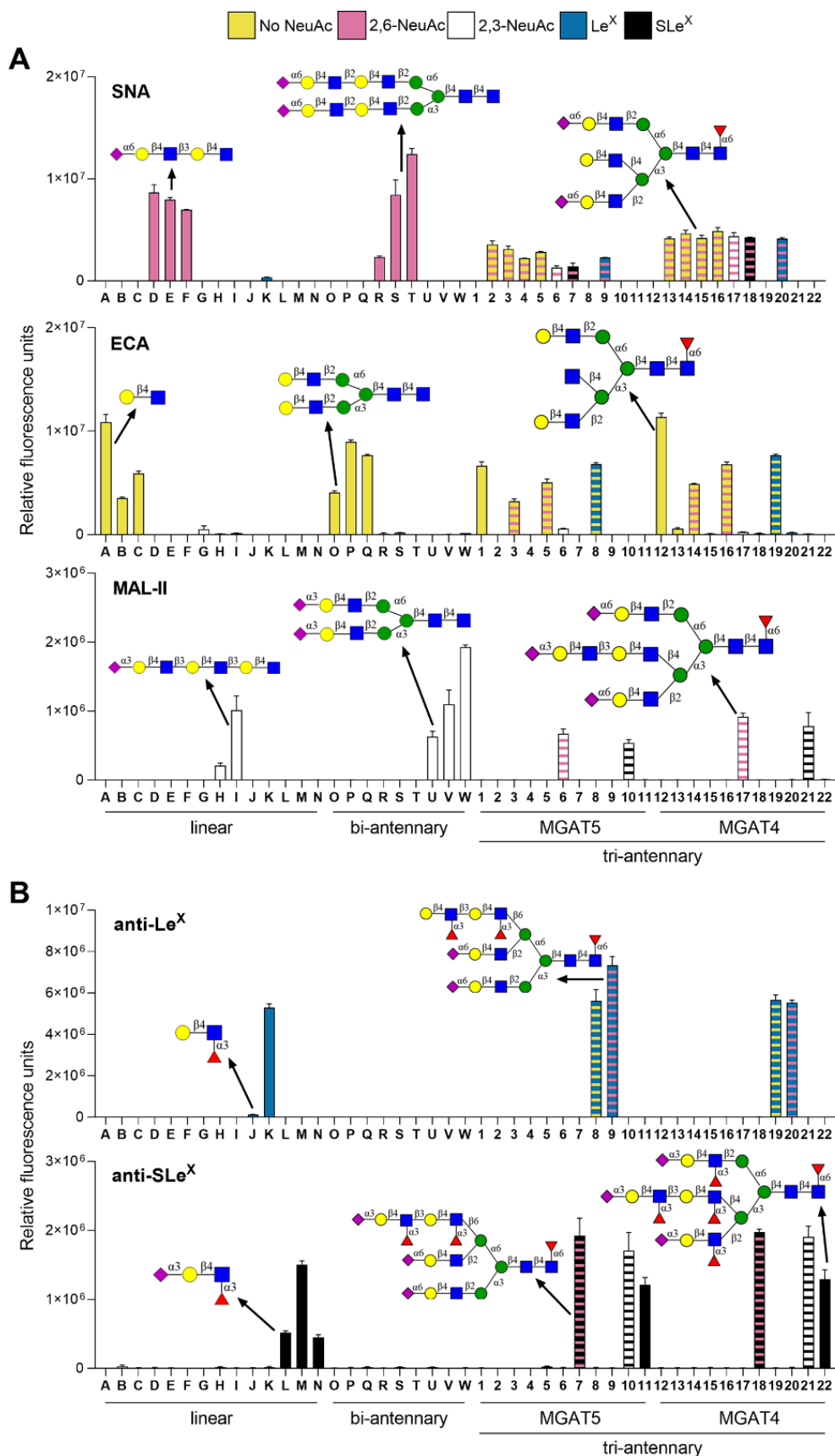
the peptide to provide symmetrical bi-antennary *N*-glycan **23**. The latter compound was exposed to MGAT5 in the presence of UDP-GlcNHTFA to install a GlcNHTFA moiety at the C-6 arm to afford tri-antennary *N*-glycan **24**. The TFA

group of the latter compound could readily be removed by aqueous NaOH (0.1 M, pH 10) to give **25** having a GlcNH<sub>2</sub> moiety at the MGAT5 antenna. As expected, the GlcNAc moieties at the MGAT1 and 2 antennae could readily be galactosylated by  $\beta$ 1,4-galactosyltransferase 1 (B4GalT1) and uridine-5'-diphosphogalactose (UDP-Gal) whereas the GlcNH<sub>2</sub> was resistant to modification by this enzyme to provide bis-galactosylated glycan **26** [15]. Next, the free amine of GlcNH<sub>2</sub> moiety was acetylated by *N*-acetylsuccinimide (AcOSu) to give the tri-antennary glycan **1** having a natural GlcNAc residue at the C-6 antenna. The LacNAc moieties of the C-2 and C-2' antennae of **1** were selectively sialylated with beta-galactoside alpha-2,6-sialyltransferase 1 (ST6Gal1) and cytidine-5'-monophospho-*N*-acetylneuraminic acid (CMP-Neu5Ac) to afford bis-sialoside **2**. The terminal GlcNAc residue of **2** was galactosylated by B4GalT1 and UDP-Gal to give **3** bearing a LacNAc moiety at the C-6 arm. An additional LacNAc moiety was installed at this antenna to give **5** by treatment with  $\beta$ 1,3-*N*-acetylglucosaminyltransferase 2 (B3GNT2) and uridine 5'-diphospho-*N*-acetylglucosamine (UDP-GlcNAc) (**4**) and then B4GalT1 and UDP-Gal. The terminal galactoside of **5** was modified as a 2,3-sialoside using ST3 beta-galactoside alpha-2,3-sialyltransferase 4 (ST3Gal4) in the presence of CMP-Neu5Ac to provide **6**. The 2,3-sialyl-di-LacNAc moiety at the MGAT5 antenna of compound **6** is a substrate for 4-galactosyl-*N*-acetylglucosaminide 3- $\alpha$ -L-fucosyltransferase (FUT5) and treatment of this compound with the enzyme in the presence of guanosine 5'-diphospho- $\beta$ -L-fucose (GDP-Fuc) afforded **7** having a SLe<sup>x</sup>-Le<sup>x</sup> moiety at the MGAT5 arm. All enzymatic transformations were monitored by LC-MS using hydrophilic interaction liquid chromatography (HILIC) and in the case a reaction did not proceed to completion, additional glycosyltransferases and sugar nucleotides were added. The compounds were purified by size exclusion column chromatography over Biogel P4 and fully characterized by multi-dimensional NMR and MS.

Compound **7** was diversified to provide additional derivatives. Thus, treatment of **7** with an  $\alpha$ 2-3,6,8 sialidase from *Clostridium perfringens* removed all sialosides to give **8**. Alternatively, the use of 2,3-neuraminidase S resulted in the selective removal of the 2,3-sialoside to provide compound **9** that has a Le<sup>x</sup>-Le<sup>x</sup> moiety at the MGAT5 antenna. Compound **8** was exposed to ST3Gal4 and CMP-Neu5Ac and as expected the LacNAc moieties at the MGAT1 and 2 antennae were easily sialylated. A prolonged reaction time and excess of CMP-Neu5Ac were required to modify the Le<sup>x</sup> moiety with a 2,3-sialoside, giving **10**. The reaction mixture was purified by size-exclusion chromatography using Bio-Gel P-4, which gave the desired glycan **10** and a small amount of partially sialylated material that was removed by semi-preparative HPLC using a HILIC column (XBridge Amide 5  $\mu$ m, 10 mm  $\times$  250 mm, Waters). Detailed NMR and MS analysis confirmed the purity and structural integrity of target glycan **10**. Compound **10** was further modified using FUT5 and GDP-Fuc to afford glycan **11**.

A range of isomeric compounds (**12-22**, Fig. 3B) was prepared by modifying symmetrical glycan **23** with  $\alpha$ 1,3-mannosyl-glycoprotein  $\beta$ 1,4-*N*-acetylglucosaminyltransferase (MGAT4) and GlcNTFA followed by the removal of

the TFA moiety (**27**) and further enzymatic and chemical manipulations to give glycans **12-22**.



**Fig 4. Microarray binding of plant lectins and antibodies to synthetic glycans.** For the tri-antennary *N*-glycans, either the MGAT4 or MGAT5 arm was elongated to two LacNAc repeating units. Glycans were terminating without a NeuAc (yellow), or with  $\alpha$ 2,6-linked NeuAc (pink),  $\alpha$ 2,3-linked NeuAc (white), Le<sup>x</sup> (blue), or SLe<sup>x</sup> (black). Bars with two colors indicate glycans terminating in different epitopes on different arms. The glycan microarray was used to probe the presence of (A)  $\alpha$ 2,6-linked sialic acids (SNA, *Sambucus nigra* agglutinin), non-sialylated epitopes (ECA, *Erythrina cristagalli* lectin),  $\alpha$ 2,3-sialylated glycans (MAL-II, *Maackia amurensis* lectin II), as well as (B) Le<sup>x</sup> and SLe<sup>x</sup> using antibodies. See Fig. 2 for the structures of A-W and 1-22. Bars represent the mean  $\pm$  SD (n=4).

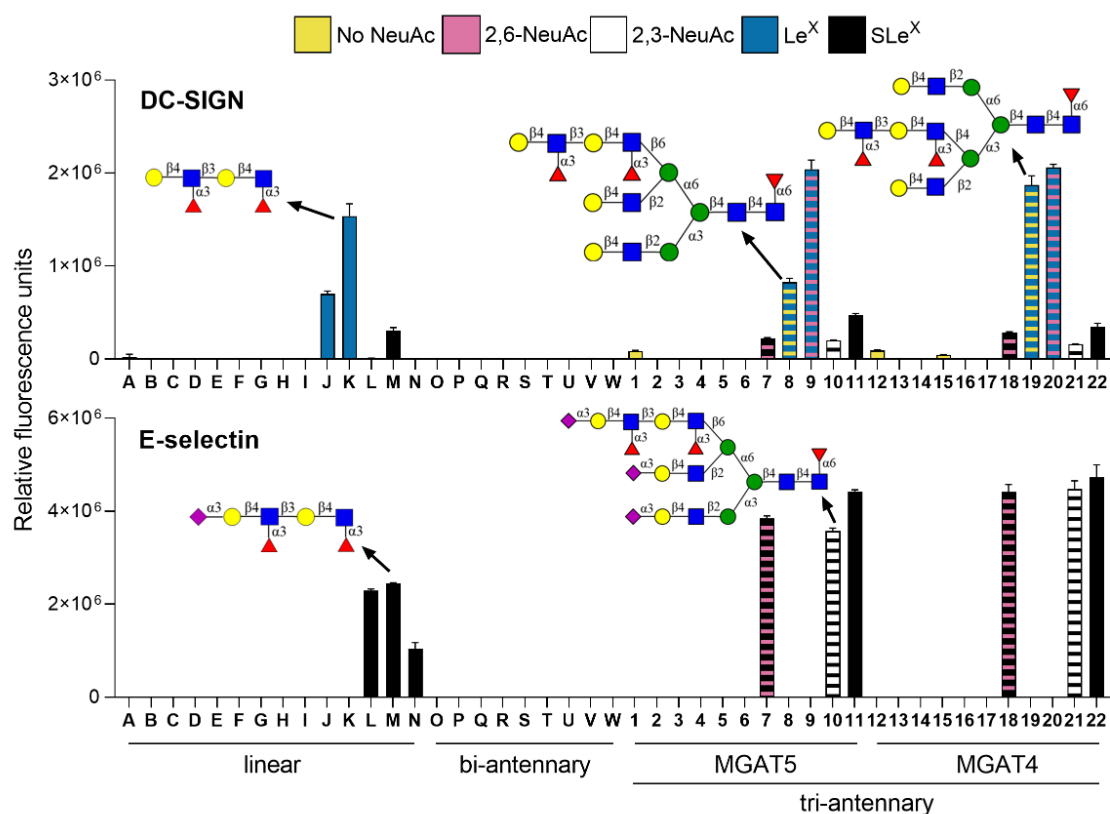
### Glycan array development and validation by lectins and antibodies

Next, attention was focused on the development of a glycan microarray using the newly synthesized tri-antennary glycans and number control compounds including linear glycans having various forms of sialylation and fucosylation and bi-antennary glycans terminating in 2,3- and 2,6-sialosides relevant for IAV binding. The tri-antennary glycans have a reducing end and therefore were treated with 2-[(methylamino)oxy]ethanamine to install an amino-containing anomeric linker. After purification by P4 size exclusion column chromatography, the resulting derivatives were printed on *N*-hydroxysuccinimide (NHS)-activated glass slides. The microarray was interrogated with several fluorescently labeled plant lectins to validate proper immobilization (Fig. 4A). *Sambucus nigra* lectin (SNA), which binds to  $\alpha$ 2,6-linked sialic acids [18, 19] recognized all linear, bi-antennary, and tri-antennary glycans having such a moiety. *Erythrina cristagalli* agglutinin (ECA) binds glycans having a terminal LacNAc residue [20] and all compounds having such an entity were detected. As expected, fucosylation and sialylation of LacNAc moieties abolished binding by this lectin. The presence of Le<sup>x</sup> and SLe<sup>x</sup> was further confirmed by using monoclonal antibodies recognizing these epitopes (Fig. 4B). The anti-Le<sup>x</sup> antibodies did not bind Le<sup>x</sup> (J), however, Le<sup>x</sup>-Le<sup>x</sup> (K) and compounds presenting this epitope on the MGAT4 (19, 20) or 5 (8, 9) antenna were bound with similar affinities. Binding studies with additional plant lectins (AAL, LEL, and SBA) further confirmed proper printing of the glycans and the presence of fucosylated glycans, elongated glycans, and glycans with a terminal galactose, respectively (Fig. S1).

### E-selectin does not show a preference for linear glycans or SLe<sup>x</sup>-Le<sup>x</sup> presented on the MGAT4 and 5 antenna

The dendritic cell-specific intracellular adhesion molecule-3-grabbing nonintegrin (DC-SIGN) is a C-type lectin presented on the surface of antigen-presenting cells, where it functions as a pathogen recognition receptor [21]. It can also interact with endogenous glycoproteins by binding to epitopes such as Le<sup>x</sup> [22]. All structures having a Le<sup>x</sup> epitope showed similar responsiveness (Fig. 5., J, K, 8, 9, 19, and 20) indicating that this glycan binding protein does not have a preference for a specific presentation of the minimal epitope. SLe<sup>x</sup> containing glycans were poorly bound by this glycan binding protein (M, 7, 10, 11, 18, 21, and 22).

Next, the binding specificity of E-selectin for the different SLe<sup>x</sup>-containing structures was investigated (Fig. 5). Previous studies indicated that E-selectin has a preference for *N*-glycans having a SLe<sup>x</sup>-Le<sup>x</sup> epitope displayed on the MGAT4 antenna [12]. However, glycans having SLe<sup>x</sup>-Le<sup>x</sup> at the MGAT4 or MGAT5 arm of a tri-antennary *N*-glycan (**7**, **10**, **11**, **18**, **21**, and **22**) or as part of linear structures (**M**) gave similar responsiveness. Furthermore, this lectin bound SLe<sup>x</sup> and SLe<sup>x</sup>-Le<sup>x</sup> equally well (**L**, **M**, and **N**). It has also been suggested that E-selectin preferentially binds to multi-antennary glycans displaying several SLe<sup>x</sup> moieties [13, 23]. The array data does not show such a preference since tri-antennary *N*-glycans having three SLe<sup>x</sup> moieties (**11** and **22**) gave similar responses compared to tri-antennary glycans having only one such moiety (**7**, **10**, **18**, **21**). Thus, the array data indicate that the presentation of SLe<sup>x</sup> in the context of a complex glycan does not modulate E-selectin binding. We also investigated the binding properties of various galectins (1, 3, 4, 8, 9) and collectin-12 and similar binding patterns were observed for linear and tri-antennary glycans (Fig. S2).



**Fig 5. Microarray binding of DC-SIGN and E-selectin.** For the tri-antennary *N*-glycans, either the MGAT4 or MGAT5 arm was elongated to two LacNAc repeating units. Glycans were terminating without a NeuAc (yellow), or with  $\alpha$ 2,6-linked NeuAc (pink),  $\alpha$ 2,3-linked NeuAc (white), Le<sup>x</sup> (blue), or SLe<sup>x</sup> (black). Bars with two colors indicate glycans terminating in different epitopes on different arms. See Fig. 2 for the structures of A-W and 1-22. Bars represent the mean  $\pm$  SD ( $n=4$ ).

## Receptor specificity of H5 influenza A viruses is modulated by the complex architecture of *N*-glycans

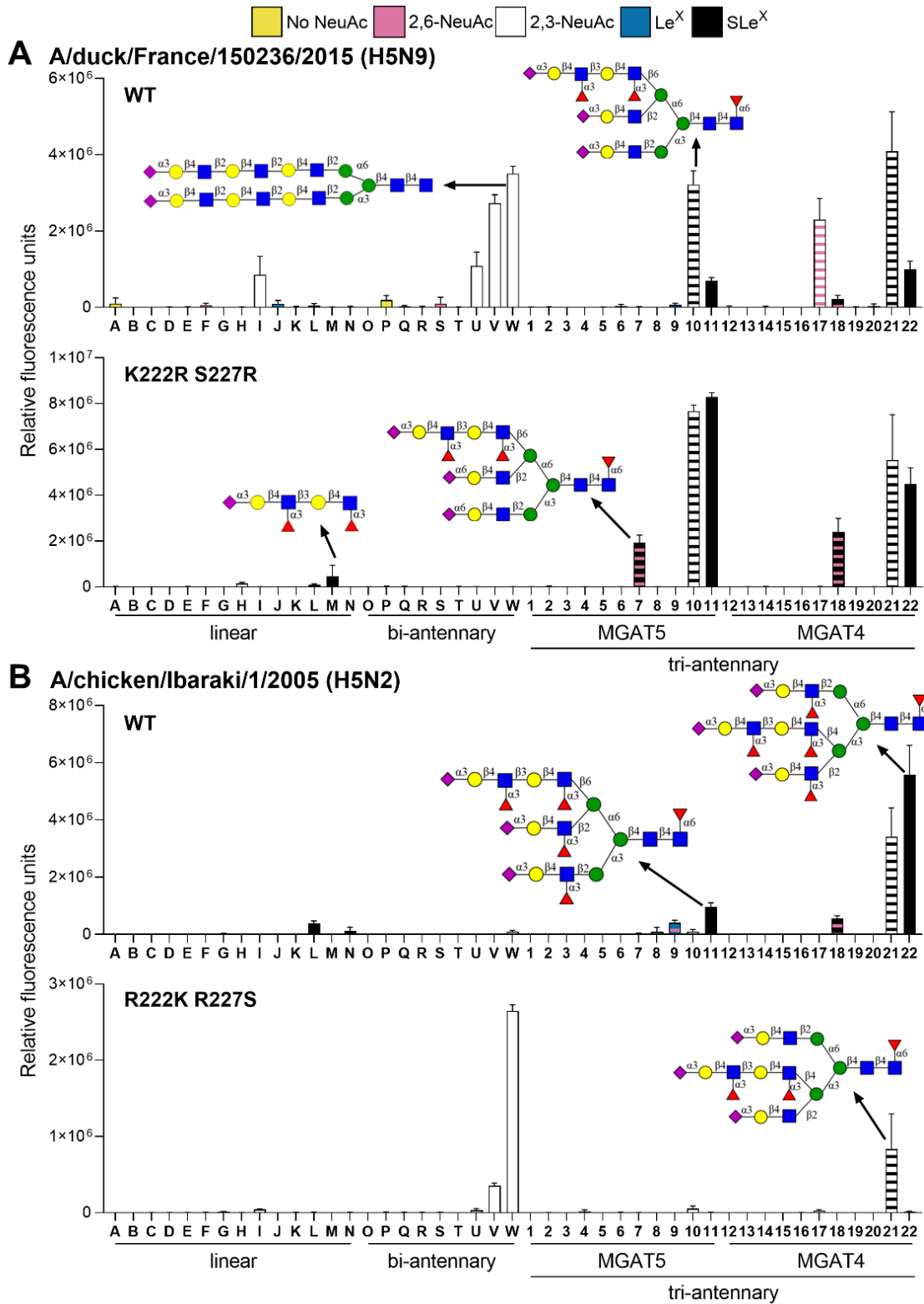
Influenza A viruses (IAV) employ sialic acid as receptors for cell entry. Sialic acids can be linked at the penultimate galactoside of glycans by either an  $\alpha$ 2,3- or  $\alpha$ 2,6-anomeric linkage, which are, respectively, considered as the avian- and human-type receptors [4]. The presence of  $\alpha$ 2,3- and  $\alpha$ 2,6-linked sialic acids in the upper airway and intestinal tissues differs between species [11] and represents a barrier for cross-species infection. Receptor binding of IAVs is, however, much more complex than the 2,3- vs. 2,6-sialoside paradigm [11]. For example, chicken IAVs exhibit a binding preference for, among others, SLe<sup>x</sup> whereas duck-adapted IAVs employ non-fucosylated  $\alpha$ 2,3-sialosides as their receptors [24]. Phylogenetic and functional analyses have indicated that amino acids at positions 222 and 227 of H5 HA are responsible for this switch in specificity. In particular, K222 and S227 are highly conserved among duck-adapted AIVs, while K222R and S227R substitutions have been observed in chicken-adapted IAVs [24-26].

A highly pathogenic avian IAV (H5Nx) emerged in 2015 in the South-West of France involving more than 70 outbreaks in commercial poultry flocks [24]. Previous glycan array analysis indicated that recombinant HA derived from wild-type (WT) virus (A/duck/France/150236/2015) can bind tri-LacNAc modified by  $\alpha$ 2,3-Neu5Ac, which is consistent with a duck-adapted virus [24]. K222R and S227R mutations switched the binding preference to SLe<sup>x</sup> and resulted in increased adhesion to the chicken trachea [24].

Here, we examined the receptor specificity of the recombinant HA (rHA) of A/duck/France/150236/2015 and the double mutant (K222R and S227R) using a much broader range of glycans including the previously employed linear compound (**C**, **F**, **I**, **N**), bi-antennary glycan having LacNAc moieties of different length capped by a 2,3- or 2,6-sialosides and the newly synthesized tri-antennary glycans (Fig. 6A). As expected, the WT HA preferentially bound 2,3-sialyl LacNAc containing compounds. The length of the LacNAc chain appeared to be important for binding and in the case of the linear glycans only compound **I**, which has a tri-LacNAc backbone, showed binding. In the case of the bi-antennary glycans, increasing the length of the LacNAc chain resulted in stronger binding (**U**, **V**, **W**). However, compound **U**, which has two  $\alpha$ (2,3)-sialyl moieties at a single LacNAc moiety showed binding whereas  $\alpha$ (2,3)-sialyl-LacNAc alone (**G**) did not give any responsiveness, highlighting that presentation of the epitope as part of an *N*-glycan substantially increases binding. Interestingly, most tri-antennary glycans having an  $\alpha$ (2,3)-sialyl-LacNAc moiety also showed strong binding (**10**, **17**, and **21**). Interestingly, compound **6**, an isomeric derivative of **17**, was not bound by the WT HA, indicating that the presentation of the SLe<sup>x</sup> epitope at the MGAT4 or 5 antennae may be important for recognition. Some binding was observed for tri-antennary glycans having only terminal SLe<sup>x</sup>-Le<sup>x</sup> and sLe<sup>x</sup> moieties (**11** and **22**) whereas a similar linear glycan did not show responsiveness (**M**).

The K222R and S227R mutant HA of A/duck/France/150236/2015 bound preferentially to SLe<sup>x</sup> containing glycans and it was found that in this case display of the epitope as part of a complex glycan also modulates binding (Fig. 6A). For





**Fig 6. Microarray binding of influenza A H5 HAs towards linear glycans and bi- and tri-antennary N-glycans.** (A) Binding of the recombinantly expressed IAV H5 HAs of A/duck/France/150236/2015 (WT, binds  $\alpha$ 2,3-linked NeuAc, and mutant K222R+S227R, binds SLe<sup>x</sup>) and (B) A/chicken/Ibaraki/1/2005 (WT, binds SLe<sup>x</sup>, and mutant R222K+R227S, binds  $\alpha$ 2,3-linked NeuAc). For the tri-antennary N-glycans, either the MGAT4 or MGAT5 arm was elongated to two LacNAc repeating units. Glycans were terminating without a NeuAc (yellow), or with  $\alpha$ 2,6-linked NeuAc (pink),  $\alpha$ 2,3-linked NeuAc (white), Le<sup>x</sup> (blue), or

SLe<sup>x</sup> (black). Bars with two colors indicate glycans terminating in different epitopes on different arms. See Fig. 2 for the structures of A-W and 1-22. Bars represent the mean  $\pm$  SD (n=4).

example, the linear SLe<sup>x</sup> containing glycans (**L**, **M**, **N**) showed weak or no responsiveness whereas the tri-antennary glycans (**7**, **10**, **11**, **18**, **21**, and **22**) bound strongly. Furthermore, compound **7**, which has 2,6-LacNAc moieties at the MGAT1 and 2 antennae, bound somewhat weaker compared to compounds **10** and **11**, having a 2,3-sialyl-LacNAc or SLe<sup>x</sup> epitope at these positions.

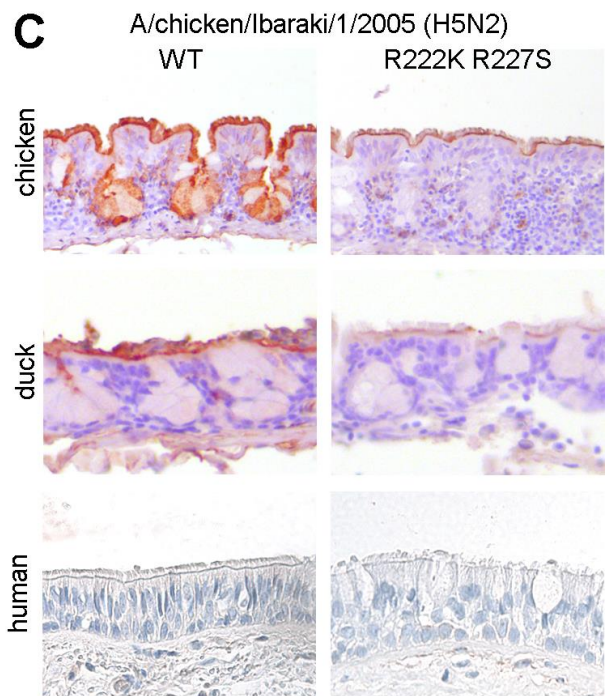
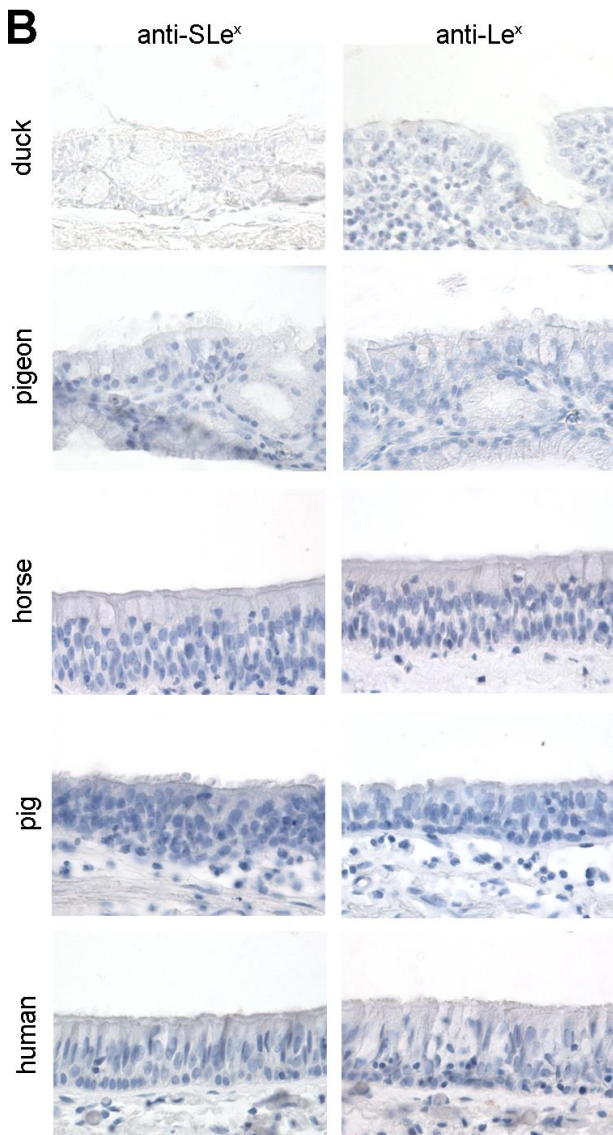
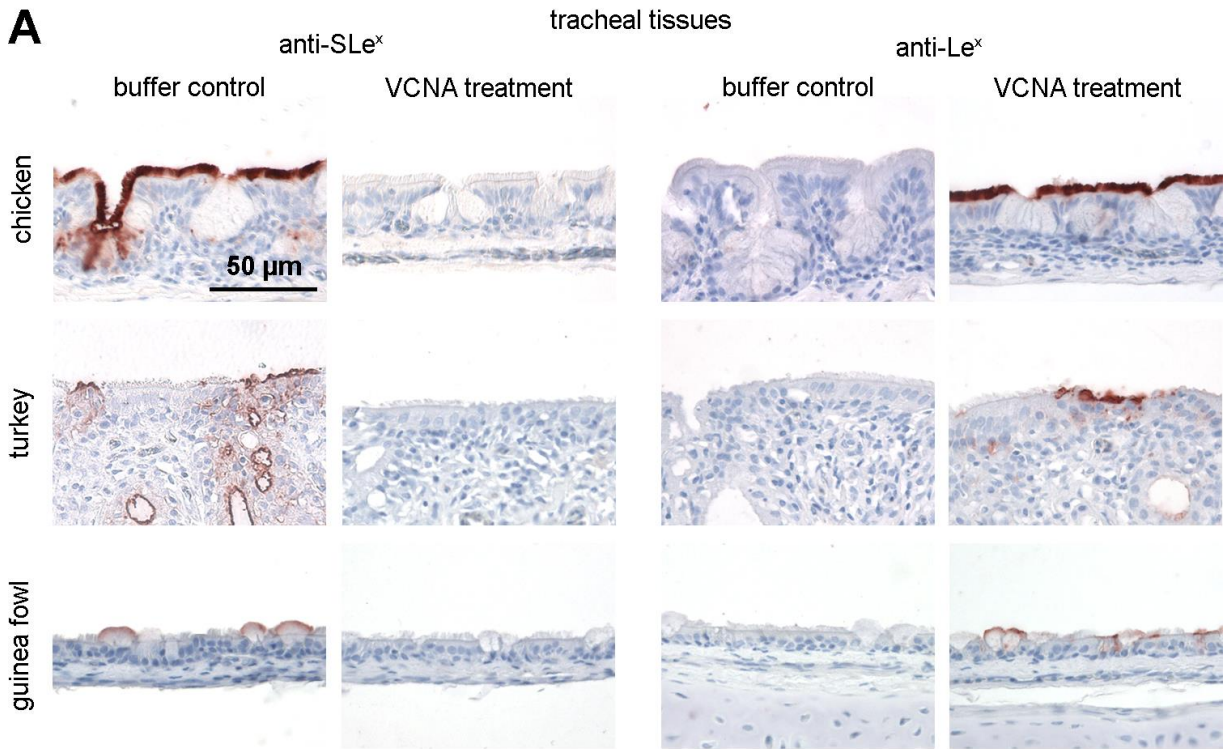
Next, we examined the receptor specificity of rHA of A/chicken/Ibaraki/1/2005 H5N2 (Fig. 6B), which has arginine at positions 222 and 227 of the HA and was previously shown to bind linear SLe<sup>x</sup> containing glycans [25]. When the receptor specificity of this rHA was probed using the new glycan microarray, it was found that it has a strong preference for tri-antennary glycans having a SLe<sup>x</sup> moiety (**11**, **21**, **22**). Unlike the mutant HA of A/duck/France/150236/2015, the WT HA of A/Chicken/Ibaraki/1/2005 preferentially bound to glycans having a SLe<sup>x</sup>-Le<sup>x</sup> epitope on the MGAT4 antenna (**21** and **22** vs. **11**) indicating fine receptor specificity. Compounds **21** and **22** gave similar responsiveness indicating that the additional SLe<sup>x</sup> moieties at the MGAT1 and MGAT2 do not contribute to binding. The R222K R227S mutant HA of A/chicken/Ibaraki/1/2005 H5N2 showed a switch in receptor specificity and only bound to  $\alpha$ 2,3-Neu5Ac containing glycans **V**, **W**, and **21** (Fig. 6B). Bi-antennary glycan **W** having a 2,3-sialoside at a tri-LacNAc structure showed the greatest responsiveness.

The relevance of the tri-antennary structures was further shown by the binding of whole influenza A viruses (Fig. S3). A/Indonesia/5/2005 is known to bind  $\alpha$ 2,3-Neu5Ac containing glycans [27]. The binding of other H5 IAVs to SLe<sup>x</sup> structures has so far only been shown in the context of tetrasaccharides [28, 29] or linear glycans [24, 25]. Here, we showed that A/Indonesia/5/2005 (H5N1), which contains the typical amino acids for binding  $\alpha$ 2,3-Neu5Ac (222K and 227S), bound equally well to *N*-glycans terminating in either  $\alpha$ 2,3-linked Neu5Ac and SLe<sup>x</sup> (Fig. S3). No preference was observed for linear glycans or bi- or tri-antennary *N*-glycans.

The human H3N2 IAV A/Netherlands/816/1991 is known to bind  $\alpha$ 2,6-linked sialic acids [30] and all structures with at least one  $\alpha$ 2,6-linked sialic acid were bound (Fig. S3). No binding preference was observed for either linear glycans or bi- or tri-antennary *N*-glycans. Collectively the data shows that the receptor specificity of IAVs is complex and that not only fucosylation of a 2,3-sialyl LacNAc moiety to give SLe<sup>x</sup> but also the length of a LacNAc chain and presentation at a specific antenna can modulate receptor binding.

### **SLe<sup>x</sup> epitopes are displayed on tracheal tissues of several avian, but not mammalian, species**

There is limited information about the presence and location of SLe<sup>x</sup> epitopes in the respiratory tracts of IAV hosts. Therefore, we investigated whether SLe<sup>x</sup> structures are present on the trachea, which is the natural location of IAV infection,



**Fig 7. Binding of anti-SLe<sup>x</sup> antibodies and influenza A H5 HAs to tracheal tissues of different species.** (A, B) The presence of (S)Le<sup>x</sup> on the tracheal tissue of chicken, turkey, guinea fowl, duck, pigeon, horse, pig, and human was investigated using antibodies against (S)Le<sup>x</sup>, precomplexed with anti-mouse IgM-HRP antibodies. Tissues that were positive for SLe<sup>x</sup> (A) were additionally treated with *Vibrio cholerae* neuraminidase (VCNA) to remove sialic acids. (C) The binding of the WT (binds SLe<sup>x</sup>) and mutant (R222K+R227S, binds  $\alpha$ 2,3-linked NeuAc) H5 HA of A/chicken/Ibaraki/1/2005 to chicken, duck, and human trachea was investigated. Binding (red discoloration) was visualized using AEC and cell nuclei (blue) were visualized using hematoxylin.

of a variety of avian and mammalian species which are relevant for IAV infections. Histological staining of paraffin-embedded trachea sections was performed using monoclonal antibodies recognizing SLe<sup>x</sup> and Le<sup>x</sup>. The binding properties of these antibodies were characterized by the glycan microarray presented here, which showed it can bind a wide variety of glycans displaying the epitope (Fig. 4B). The antibodies were precomplexed with secondary antibodies coupled to horseradish peroxidase for detection. As a control for SLe<sup>x</sup>-positive tissues, slides were pretreated with a neuraminidase (VCNA) to remove all sialosides and in this way converting SLe<sup>x</sup> into Le<sup>x</sup>.

In agreement with previous reports, we found SLe<sup>x</sup> to be present on chicken [5, 26, 31, 32] and turkey trachea [33] (upper two panels of Fig. 7A). Additionally, we observed SLe<sup>x</sup> on the trachea of guinea fowl (Fig. 7A). Interestingly, we observed that the location of SLe<sup>x</sup> on tracheal tissues varied between species. In the chicken trachea, the cilia on the epithelial cells and some mucus-producing goblet cells were positive, while in the turkey trachea, only goblet cells presented SLe<sup>x</sup>. In the guinea fowl trachea, only goblet cells on the apical membrane presented SLe<sup>x</sup>. IAV infection of hosts occurs through epithelial cells [8] and, therefore, the SLe<sup>x</sup> epitopes on the chicken cilia are potentially used as receptors. On the other hand, mucus forms a barrier against IAVs [34] and the presence of SLe<sup>x</sup> on mucus-producing goblet cells from turkeys and guinea fowl may indicate that these epitopes are used as a barrier that prevents infection. A neuraminidase treatment (VCNA) of the tissues indeed resulted in a loss of SLe<sup>x</sup>, whereas stronger staining was observed for the anti-Le<sup>x</sup> antibody (Fig. 7A). Le<sup>x</sup> epitopes were not detected in any of the species without VCNA treatment. On the other hand, SLe<sup>x</sup> was absent from the duck [31] and pigeon trachea (Fig. 7B). Thus, SLe<sup>x</sup> is not universally presented by avian species.

We did not observe SLe<sup>x</sup> in the trachea of the examined mammalian species (horse, pig, human) (Fig. 7B), which is in agreement with earlier observations [32, 35, 36]. Furthermore, the avian WT and mutant (R222K R227S) HAs of A/chicken/Ibaraki/1/2005, bound to both chicken and duck trachea (Fig. 7C), despite the narrow receptor binding specificities of these HAs (Fig. 6B), which is consistent with earlier observations [24]. However, the WT and mutant H5 HAs failed to bind the human trachea (Fig. 7C), indicating that an increased binding specificity towards SLe<sup>x</sup> epitopes does not automatically result in increased zoonotic potential.



## Discussion

We describe here, for the first time, the preparation of a panel of isomeric glycans that have either a complex epitope on the MGAT4 or MGAT5 antenna, thereby providing opportunities to investigate the importance of glycan isomerism for biological recognition. The selection of the target compounds was guided by compounds that were identified as putative high-affinity ligands of E-selectin that have a SLe<sup>x</sup>-Le<sup>x</sup> epitope at the MGAT4 antenna. A previous synthetic approach to synthesize *N*-glycans having a SLe<sup>x</sup>-Le<sup>x</sup> epitope was based on the time-consuming chemical synthesis of a tri-antennary glycan having different moieties at the MGAT4 and 5 arms followed by selective modification by glycosyltransferases [37]. Here, a much more efficient stop-and-go strategy [15] is employed in which a bi-antennary glycan derived from a glycopeptide isolated from egg yolk powder is converted into a tri-antennary glycan using MGAT4 or 5 and UDP-GlcNTFA as a glycosyl donor. Removal of the TFA moiety by aqueous sodium hydroxide gave a GlcNH<sub>2</sub> residue that temporarily blocked enzymatic modification by a panel of glycosyltransferases. Acetylation of GlcNH<sub>2</sub> to give natural GlcNHAc allowed selective extension of the MGAT4 or 5 antenna by a complex epitope such as SLe<sup>x</sup>-Le<sup>x</sup>. It is demonstrated that the compounds could be diversified to give a panel of structural analogs by removal of the sialosides by either a broad-acting or 2,3-selective sialidase followed by resialylation or fucosylation. The tri-antennary glycans and several reference compounds were printed as a glycan microarray that was used to examine ligand requirements of lectins, antibodies, and several glycan-binding proteins. E-selectin bound to all compounds displaying a SLe<sup>x</sup> or SLe<sup>x</sup>-Le<sup>x</sup> epitope at the MGAT4 or MGAT5 antenna and simpler linear derivatives gave similar responsiveness. Furthermore, an *N*-glycan having one SLe<sup>x</sup> epitope gave similar binding as an *N*-glycan displaying three of these epitopes. These observations indicate that the presentation of the minimal epitope in the context of a complex glycan does not modulate the binding of E-selectin. Previous affinity purification of glycans using E-selectin indicated a multi-antennary glycan as a high-affinity ligand [12]. Furthermore, multivalent SLe<sup>x</sup> derivatives have proven to be more potent inhibitors of E-selectin [13, 23]. In these cases, the glycans were used as soluble ligands and E-selectin was immobilized on a resin or presented on a cell surface. Binding affinities and possibly also selectivities can greatly differ when glycans are presented in solution or on a surface and this may account for possible differences in selectivity.

The new glycan array was also used to examine receptor specificities of several avian influenza viruses. It is known that sulfation and fucosylation of 2,3-sialyl-LacNAc can be an important determinant for receptor binding specificity of avian IAVs and for example, those adapted to duck preferentially bind 2,3-sialyl-LacNAc whereas chicken IAVs preferentially bind to SLe<sup>x</sup> [8, 24-26]. We found that the complexity of glycans has a major impact on binding selectivities of rHAs and whole viruses and not only fucosylation but also the length of a LacNAc chain and presentation at a specific antenna determines binding. For example, we showed fine receptor specificity of H5 HA of A/chicken/Ibaraki/1/2005 which has a preference for SLe<sup>x</sup> at the MGAT4 antenna of an *N*-glycan. Duck-adapted IAVs

generally have K222 and S227 at their HAs while chicken AIVs have K222R and S227R substitutions, which correlate with binding to 2,3-sialyl-LacNAc and SLe<sup>x</sup>-containing glycans. The array studies using several mutant HAs confirmed this selectivity. Previously, receptor binding properties may not have been properly assigned due to a lack of complex glycans on a used microarray.

Overexpression of SLe<sup>x</sup> on MDCK cells was previously shown to increase the virus replication, suggesting that SLe<sup>x</sup> is an important functional receptor for IAV [31, 38]. However, the investigated viruses also grew in SLe<sup>x</sup>-deficient cells [31, 38], suggesting that SLe<sup>x</sup> is not essential [26]. Furthermore, SLe<sup>x</sup> was found to be present on O-glycans in mucus [39], therefore, the use of SLe<sup>x</sup> as a decoy receptor has been suggested [34]. The strong binding of IAVs to specific SLe<sup>x</sup>-terminated glycans indicates the necessity to further investigate the function of SLe<sup>x</sup> in IAV infection.

Although histological experiments have shown that various avian species including chicken, duck, turkey, pheasant, guinea fowl, and pigeon have SLe<sup>x</sup> containing glycans in their respiratory and intestinal tract, [26, 31-33, 38] these studies do not provide information the context in which SLe<sup>x</sup> is presented. Mass spectrometry (MS) can provide more detailed structural information, and it has been shown that glycans obtain from chicken trachea contain SLe<sup>x</sup> moieties [40]. MS provides, however, often only compositions of glycans but not detailed structures [35]. The synthetic glycans synthesized here can be used as standards to develop MS methods to determine the exact structures of (fragments of) glycans [17]. Future synthetic efforts should be guided by glycan structures found on the respiratory tissues of humans and relevant animals.



## Materials and methods

### General protocols for chemoenzymatic reactions

The tri-antennary *N*-glycans were synthesized with the general protocols described below. The analysis of the synthesized glycans consisted of LC-MS and NMR characterization.

#### Procedure for the installation of $\beta$ 1,6 GlcNTFA using MGAT5

To a solution of *N*-glycan acceptor (5-10 mM) and UDP-GlcNTFA (1.5 eq) in a sodium cacodylate buffer solution (100 mM, pH 6.8) containing  $\text{MnCl}_2$  (10 mM) and BSA (1% total volume, stock solution = 10 mg/mL) was added calf intestinal alkaline phosphatase (CIAP, 1% total volume, 1000 U/mL) and MGAT5 (50  $\mu\text{g}/\mu\text{mol}$  acceptor). The reaction mixture was incubated at 37 °C. Reaction progress was monitored by MALDI-TOF MS, if starting material remained after 18 h, another portion of MGAT5 and UDP-GlcNTFA (0.5 eq) were added until no starting material could be detected. The reaction mixture was centrifuged, and the resulting supernatant was loaded on Bio-Gel P4 column (eluent:  $\text{H}_2\text{O}$ ). The product containing fractions were combined and lyophilized to afford the desired product.

#### Procedure for the installation of $\beta$ 1,4 GlcNTFA using MGAT4A

To a solution of *N*-glycan acceptor (5-10 mM) and UDP-GlcNTFA (1.5 eq) in a Tris buffer solution (100 mM, pH 7.0) containing  $\text{MnCl}_2$  (10 mM) and BSA (1% total volume, stock solution = 10 mg/mL) was added calf intestinal alkaline phosphatase (CIAP, 1% total volume, 1000 U/mL) and MGAT4A (300  $\mu\text{g}/\mu\text{mol}$  acceptor). The reaction mixture was incubated at 37 °C. Reaction progress was monitored by MALDI-TOF MS, if starting material remained after 18 h, another portion of MGAT4 and UDP-GlcNTFA (0.5 eq) were added until no starting material could be detected. The reaction mixture was centrifuged, and the resulting supernatant was loaded on Bio-Gel P4 column (eluent:  $\text{H}_2\text{O}$ ). Product containing fractions were combined and lyophilized to afford the desired product.

#### General procedure for the removal of TFA protecting group

The GlcNTFA moiety of *N*-glycan (10 mM) was converted to  $\text{GlcNH}_2$  by the addition of an aqueous solution of 100 mM NaOH. The reaction mixture was incubated at room temperature. Reaction progress was monitored by MALDI-TOF MS. Once the reaction was completed, 1 M acetic acid was added to the mixture to adjust the pH to 7. The resulting mixture was loaded on Bio-Gel P2 column (eluent: 0.1  $\text{NH}_4\text{HCO}_3$ ). Product containing fractions were combined and lyophilized to afford the desired product.

#### General procedure for the installation of $\beta$ 1,4 Gal using B4GalT1

To a solution of *N*-glycan acceptor (5-10 mM) and UDP-Gal (1.2 eq per Gal) in a Tris buffer solution (100 mM, pH 7.5) containing  $\text{MnCl}_2$  (10 mM) and BSA (1% total volume, stock solution = 10 mg/mL) was added calf intestinal alkaline phosphatase (CIAP, 1% total volume, 1000 U/mL) and B4GalT1 (40  $\mu\text{g}/\mu\text{mol}$  acceptor). The reaction mixture was incubated at 37 °C. Reaction progress was monitored by MALDI-TOF MS or ESI-MS, if starting material remained after 16 h,

another portion of B4GalT1 and UDP-Gal (0.3 eq per Gal) were added until no starting material could be detected. The reaction mixture was centrifuged, and the resulting supernatant was loaded on Bio-Gel P4 column (eluent: 0.1 NH<sub>4</sub>HCO<sub>3</sub>). Product containing fractions were combined and lyophilized to afford the desired product.

#### General procedure for acetylation of amine

The GlcNH<sub>2</sub> moiety of *N*-glycan (10 mM) was converted to natural GlcNAc by the addition of AcOSu (10 eq) in a basic solution (pH 8). After the reaction mixture was agitated for 30 min at room temperature, the pH of the solution was checked again and adjusted to 7-9 by adding 1 M NaOH(aq.). The reaction was kept at room temperature until full acetylation was observed by ESI-MS. If the starting amine was detected, additional AcOSu (5 eq) was added until complete conversion was observed. The reaction mixture was centrifuged, and the resulting supernatant was loaded on Bio-Gel P4 column (eluent: 0.1 NH<sub>4</sub>HCO<sub>3</sub>). Product containing fractions were combined and lyophilized to afford the desired product.

#### General procedure for the installation of $\alpha$ 2,6 Neu5Ac using ST6Gal1

To a solution of *N*-glycan acceptor (5-10 mM) and CMP-Neu5Ac (1.5 eq per Neu5Ac) in a sodium cacodylate buffer solution (100 mM, pH 7.2) containing BSA (1% total volume, stock solution = 10 mg/mL) was added calf intestinal alkaline phosphatase (CIAP, 1% total volume, 1000 U/mL) and ST6Gal1 (80  $\mu$ g/ $\mu$ mol acceptor), and the reaction mixture was incubated at 37 °C. Reaction progress was monitored by ESI-MS, if starting material remained after 16 h, another portion of ST6Gal1 and CMP-Neu5Ac (0.5 eq per Neu5Ac) were added until no starting material could be detected. The reaction mixture was centrifuged, and the resulting supernatant was loaded on Bio-Gel P4 column (eluent: 0.1 NH<sub>4</sub>HCO<sub>3</sub>). Product containing fractions were combined and lyophilized to afford the desired product.

#### General procedure for the installation of $\beta$ 1,3 GlcNAc using B3GNT2

To a solution of *N*-glycan acceptor (5-10 mM) and UDP-GlcNAc (1.5 eq) in a HEPES buffer solution (100 mM, pH 7.2) containing KCl (25 mM), MgCl<sub>2</sub> (2 mM), DTT (1 mM), and BSA (1% total volume, stock solution = 10 mg/mL) was added calf intestinal alkaline phosphatase (CIAP, 1% total volume, 1000 U/mL) and B3GNT2 (100  $\mu$ g/ $\mu$ mol acceptor), and the reaction mixture was incubated at 37 °C. Reaction progress was monitored by ESI-MS, if starting material remained after 18 h, another portion of B3GNT2 and UDP-GlcNAc (0.3 eq) were added until no starting material could be detected. The reaction mixture was centrifuged, and the resulting supernatant was loaded on Bio-Gel P4 column (eluent: 0.1 NH<sub>4</sub>HCO<sub>3</sub>). Product containing fractions were combined and lyophilized to afford the desired product.

#### General procedure for the installation of $\alpha$ 2,3 Neu5Ac using ST3Gal4

To a solution of *N*-glycan acceptor (5-10 mM) and CMP-Neu5Ac (1.5 eq) in a sodium cacodylate buffer solution (100 mM, pH 7.2) containing BSA (1% total volume, stock solution = 10 mg/mL) was added calf intestinal alkaline phosphatase (CIAP, 1% total volume, 1000 U/mL) and ST3Gal4 (100  $\mu$ g/ $\mu$ mol

acceptor), and the reaction mixture was incubated at 37 °C. Reaction progress was monitored by ESI-MS, if starting material remained after 18 h, another portion of ST3Gal4 and CMP-Neu5Ac (0.5 eq per Neu5Ac) were added until no starting material could be detected. The reaction mixture was centrifuged, and the resulting supernatant was loaded on Bio-Gel P4 column (eluent: 0.1 NH<sub>4</sub>HCO<sub>3</sub>). Product containing fractions were combined and lyophilized to afford the desired product.

#### General procedure for the installation of $\alpha$ 1,3 Fuc using FUT5

To a solution of *N*-glycan acceptor (5-10 mM) and GDP-Fuc (1.5 eq per Fuc) in a Tris buffer solution (100 mM, pH 7.5) containing MnCl<sub>2</sub> (10 mM) and BSA (1% total volume, stock solution = 10 mg/mL) was added calf intestinal alkaline phosphatase (CIAP, 1% total volume, 1000 U/mL) and FUT5 (80  $\mu$ g/ $\mu$ mol acceptor). The reaction mixture was incubated at 37 °C. Reaction progress was monitored by ESI-MS, if starting material remained after 18 h, another portion of FUT5 and GDP-Gal (0.5 eq per Fuc) were added until no starting material could be detected. The reaction mixture was centrifuged, and the resulting supernatant was loaded on Bio-Gel P4 column (eluent: 0.1 NH<sub>4</sub>HCO<sub>3</sub>). Product containing fractions were combined and lyophilized to afford the desired product.

#### General procedure for the removal of terminal Neu5Ac using $\alpha$ 2-3,6,8 neuraminidase

To a solution of *N*-glycan (5-10 mM) in sodium acetate buffer solution (50 mM, pH 5.5) containing CaCl<sub>2</sub> (5 mM) was added  $\alpha$ 2-3,6,8 neuraminidase (2500 U/mL, from *Clostridium perfringens*, #P0720S, New England Biolabs), and the mixture was incubated at 37 °C. Reaction progress was monitored by ESI-MS, if starting material remained after 16 h, another portion of neuraminidase was added until no starting material could be detected. Once the reaction was completed, the neuraminidase was inactivated at 65 °C for 10 min. The reaction mixture was centrifuged, and the resulting supernatant was loaded on Bio-Gel P4 column (eluent: 0.1 NH<sub>4</sub>HCO<sub>3</sub>) to afford the desired product.

#### General procedure for selective removal of terminal $\alpha$ 2,3-Neu5Ac using $\alpha$ 2,3-neuraminidase S

To a solution of *N*-glycan (5-10 mM) in sodium acetate buffer solution (50 mM, pH 5.5) containing CaCl<sub>2</sub> (5 mM) was added  $\alpha$ 2,3-neuraminidase S (8000 U/mL, from *Streptococcus pneumoniae*, #0743S, New England Biolabs), and the mixture was incubated at 37 °C. Reaction progress was monitored by ESI-MS, if starting material remained after 16 h, another portion of neuraminidase was added until no starting material could be detected. Once the reaction was completed, the neuraminidase was inactivated at 65 °C for 10 min. The reaction mixture was centrifuged, and the resulting supernatant was loaded on Bio-Gel P4 column (eluent: 0.1 NH<sub>4</sub>HCO<sub>3</sub>) to afford the desired product.

#### Procedure for ST3Gal4-catalyzed reaction for $\alpha$ 2,3-sialylation of the Le<sup>x</sup> moiety of *N*-glycan

To a solution of *N*-glycan acceptor **8** or **19** (5-10 mM) and CMP-Neu5Ac (6 eq) in a sodium cacodylate buffer solution (100 mM, pH 7.2) containing BSA (1% total volume, stock solution = 10 mg/mL) was added calf intestinal alkaline

phosphatase (CIAP, 1% total volume, 1000 U/mL) and ST3Gal4 (300  $\mu\text{g}/\mu\text{mol}$  acceptor), and the reaction mixture was incubated at 37 °C for 24 h. ESI-MS analysis showed that most of the starting material was remaining. Additional ST3Gal4 and CMP-Neu5Ac (3 eq) were added and incubated at 37 °C. Reaction progress was further monitored by ESI-MS, if starting material remained after 24 h, another portion of ST3Gal4 and CMP-Neu5Ac (1.5 eq) were added. After 4 days, the reaction mixture was centrifuged, and the resulting supernatant was loaded on Bio-Gel P4 column (eluent: 0.1  $\text{NH}_4\text{HCO}_3$ ). Product containing fractions were combined and lyophilized for further purification by HPLC to afford the desired product.

#### Procedure for FUT5-catalyzed reaction for $\alpha$ 1,3-fucosylation to synthesize 11 or 22

To a solution of *N*-glycan acceptor **10** or **21** (5-10 mM) and GDP-Fuc (3 eq) in a Tris buffer solution (100 mM, pH 7.5) containing  $\text{MnCl}_2$  (10 mM) and BSA (1% total volume, stock solution = 10 mg/mL) was added calf intestinal alkaline phosphatase (CIAP, 1% total volume, 1000 U/mL) and FUT5 (200  $\mu\text{g}/\mu\text{mol}$  acceptor), and the reaction mixture was incubated at 37 °C for 24 h. ESI-MS analysis showed most of the starting material was remaining. Additional FUT5 and GDP-Fuc (2 eq) were added and incubated at 37 °C for 36 h. Reaction progress was further monitored by ESI-MS, if starting material remained after 24 h, another portion of FUT5 and GDP-Fuc (1 eq) were added. After 3 days, the reaction mixture was centrifuged, and the resulting supernatant was loaded on Bio-Gel P4 column (eluent: 0.1  $\text{NH}_4\text{HCO}_3$ ). Product containing fractions were combined and lyophilized for further purification by HPLC to afford the desired product.

#### General procedure for installation of a bi-functional spacer to reduce the terminal of glycans

Glycans **1-22** (10-20 mM) and linker 2-[(methylamino)oxy]ethanamine (100 eq) were dissolved in aqueous NaOAc buffer (0.1 M, 50-200  $\mu\text{L}$ ), and the final pH of the reaction mixture was adjusted to 4.0-5.0. The reaction was incubated at 37 °C for 3 days. The resulting mixture was loaded on a Carbograp cartridge for rapid solid phase extraction (SPE) to remove excess spacer and salt (eluent: 10 mM  $\text{NH}_4\text{HCO}_3$ ), followed by releasing product with a mixture solution of  $\text{CH}_3\text{CN}$  and 10 mM  $\text{NH}_4\text{HCO}_3$  (v/v 1:1). Product containing fractions were combined and lyophilized to afford glycans with an amino-containing linker at reducing terminal [41].

#### **Expression and purification of HA for binding studies**

HA encoding cDNAs of A/duck/France/150236/2015 (H5N9) and A/chicken/Ibaraki/1/2005 (H5N2) were cloned into the pCD5 expression vector as described previously [42]. HAs were mutated using site-directed mutagenesis and checked using sequencing. The HA-encoding cDNAs were cloned in frame with a secretion signal sequence, the Twin-Strep (WSHPQFEKGGGSGGGWSHPQFEK); IBA, Germany), a GCN4 trimerization domain (RMKQIEDKIEEIESKQKKIENEIARIKK), and a superfolder GFP [43]. The HAs were expressed in HEK293S GnTI(-) cells and subsequently purified from the cell culture supernatants as described previously [42]. In short, transfection was performed using the pCD5 plasmid DNA and polyethyleneimine I. The transfection mixtures were replaced at 6 h post-

transfection by 293 SFM II expression medium (Gibco), supplemented with sodium bicarbonate (3.7 g/L), Primatone RL-UF (3.0 g/L, Kerry, NY, USA), glucose (2.0 g/L), glutaMAX (1%, Gibco), valproic acid (0.4 g/L) and DMSO (1.5%). At 5 to 6 days after transfection, cell culture supernatants were collected and Strep-Tactin sepharose beads (IBA, Germany) were used to purify the HA proteins according to the manufacturer's instructions.

### Glycan microarray binding studies

Microarrays were constructed as described before [22, 44, 45]. Briefly, the synthetic glycans (100  $\mu$ M in 250 mM sodium phosphate buffer, pH 8.5) were printed as replicates of 6 on activated glass slides (Nexterion Slide H, Schott Inc) with 24 subarrays per slide by piezoelectric non-contact printing (sciFLEXARRAYER S3, Scienion Inc). After overnight incubation, the remaining activated esters were quenched with 50 mM ethanolamine in 100 mM TRIS, pH 9.0. Washed and dried slides were stored in a desiccator at RT. Biotinylated lectins (from Vector Laboratories) SNA (*Sambucus nigra* agglutinin, 20  $\mu$ g/mL; B-1305), ECA (*Erythrina cristagalli* agglutinin, 20  $\mu$ g/mL; B-1145), MAL-II (*Maackia amurensis* lectin II, 20  $\mu$ g/mL; B-1265), AAL (*Aleuria aurantia* lectin, 5  $\mu$ g/mL; B-1395), LEL (*Lycopersicon esculentum* lectin, 10  $\mu$ g/mL; B-1175), and SBA (soybean agglutinin, 20  $\mu$ g/mL; B-1015) were premixed with streptavidin-conjugated AlexaFluor555 (2  $\mu$ g/mL; S32355, Invitrogen) or AlexaFluor635 (5  $\mu$ g/mL; S32364, Molecular Probes), after which the mixtures were incubated on the subarrays in TSM binding buffer (20 mM Tris Cl, pH 7.4, 150 mM NaCl, 2 mM CaCl<sub>2</sub>, 2 mM MgCl<sub>2</sub>, 0.05% Tween, 1% BSA) for 1 h. Similarly as described for the plant lectins, recombinant human DC-SIGN--Fc Chimera (10  $\mu$ g/mL; R&D systems, 161-DC) and recombinant human E-selectin--Fc Chimera (2  $\mu$ g/mL; R&D systems, 724-ES) were assayed premixed with anti-IgG Fc-biotin (5  $\mu$ g/mL; ThermoFisher Scientific, A18833) and streptavidin-AlexaFluor635 (5  $\mu$ g/mL) in TSM binding buffer. Mixtures of recombinant human galectins-1, -3, -4, -8, and -9 from R&D systems (3  $\mu$ g/mL; 1152-GA, 3  $\mu$ g/mL; 1154-GA, 3  $\mu$ g/mL; 1227-GA, 10  $\mu$ g/mL; 1305-GA, and 1  $\mu$ g/mL; 2045-GA, respectively), their corresponding biotinylated anti-human galectin goat IgG from R&D systems (3  $\mu$ g/mL; BAF1152, BAF1154, BAF1227, BAF1305, and BAF2045, respectively), and streptavidin-AlexaFluor635 (5  $\mu$ g/mL) in TSM binding buffer were incubated for 2 h on the microarray. Recombinant human collectin-12 with an N-terminal 9-His tag (10  $\mu$ g/mL; CL-P1/COLEC12, R&D systems, 2690-CL) was assayed premixed with 6x-HisTag monoclonal antibody conjugated to AlexaFluor647 (5  $\mu$ g/mL; ThermoFisher Scientific, MA1-135-A647) in TSM binding buffer with 1 h incubation. Anti-sialyl-LewisX and anti-LewisX antibodies (mouse IgM, respectively #551344 (clone CSLEX1) and #555400 (clone HI98 or HIM1), BD Biosciences) at 50  $\mu$ g/mL in 40  $\mu$ L PBS-T were applied to the subarrays for 90 min. Subsequently, the arrays were incubated with a mixture of goat anti-mouse IgM HRP (10  $\mu$ g/mL; #1021-05 Southern Biotech) and donkey anti-goat antibody labeled with AlexaFluor555 (5  $\mu$ g/mL; A21432, Invitrogen) in 40  $\mu$ L PBS-T for 1 h. H5 HAs were pre-complexed with human anti-streptag and goat anti-human-AlexaFluor555 antibodies in a 4:2:1 molar ratio, respectively in 50  $\mu$ L PBS with 0.1% Tween-20 (PBS-T) on ice for 15 min. Next, the mixtures were added to the subarrays for 90 min. Virus isolates from A/Indonesia/05/2005 H5N1 or A/Netherlands/816/1991 H3N2 (25  $\mu$ L) were diluted

with 25  $\mu\text{L}$  PBS-T supplemented with 400 nM oseltamivir and incubated on the subarrays for 1 h. Afterward, the arrays were incubated with CR8020 (for H3N2) or CR6261 (for H5N1) influenza hemagglutinin stem-specific antibody (5  $\mu\text{g}/\text{mL}$ ) in 100  $\mu\text{L}$  PBS-T. Next, a secondary goat anti-human AlexaFluor647 antibody (2  $\mu\text{g}/\text{mL}$  in 100  $\mu\text{L}$  PBS-T) was incubated on the arrays for 1 h. Wash steps after each incubation, irrespective of the antibody, lectin, virus, or HA, involved 4 successive washes of the whole slides with either PBS-T – PBS – 2x deionized water (for LEL, HAs, viruses) or TSM wash buffer (20 mM Tris Cl, pH 7.4, 150 mM NaCl, 2 mM  $\text{CaCl}_2$ , 2 mM  $\text{MgCl}_2$ , 0.05% Tween-20) - TSM buffer (20 mM Tris Cl, pH 7.4, 150 mM NaCl, 2 mM  $\text{CaCl}_2$ , 2 mM  $\text{MgCl}_2$ ) - 2x deionized water (for the other lectins and recombinant proteins) with 5 min soak times. Unless stated otherwise, all steps were performed at RT. The slides were dried by centrifugation and immediately scanned for fluorescence on a microarray slide scanner (Innopsys) as described previously. The six replicates were processed by removing the highest and lowest value, after which the mean value and standard deviation over the four remaining replicates was calculated.

### Histochemical tissue staining

Sections of formalin-fixed, paraffin-embedded animal tracheal tissues were obtained from the Division of Pathology, Department of Biomolecular Health Sciences, Faculty of Veterinary Medicine of Utrecht University, the Netherlands. Sections of formalin-fixed, paraffin-embedded human tracheal tissues were obtained from the UMC Utrecht, Department of Pathology, Utrecht, the Netherlands (TCBio-number 22-599). Tissue sections of 4  $\mu\text{m}$  were deparaffinized and rehydrated, after which antigens were retrieved by heating the slides in 10 mM sodium citrate (pH 6.0) for 10 min. Endogenous peroxidase was inactivated using 1% hydrogen peroxide in MeOH for 30 min at RT. When a neuraminidase treatment was performed, slides were incubated overnight at 37 °C with *Vibrio cholerae* neuraminidase (VCNA, #11080725001, Roche) diluted 1:50 in a solution of 10 mM potassium acetate and 0.1% Triton X-100 at pH 4.2. Non-treated control slides in experiments with neuraminidase were incubated with buffer only. After washing with PBS-T, tissues were blocked using BSA (3% w/v) in PBS at 4 °C for at least 90 min. Anti-sialyl-LewisX antibodies or anti-LewisX antibodies (mouse IgM, respectively #551344 (clone CSLEX1) and #555400 (clone HI98 or HIM1), BD Biosciences) were diluted 1:1000 in PBS and precomplexed with goat anti-mouse IgM-HRP (#1021-05, Southern Biotech) in a 1:100 dilution on ice for 20 min. H5 HAs (5  $\mu\text{g}/\text{ml}$ ) were pre-complexed with human anti-streptag and goat anti-human-HRP (#31410, Thermo Fisher Scientific) in a 4:2:1 molar ratio in PBS and incubated on ice for 20 min. Afterwards, the slides were stained for 90 min with these mixtures. After washing with PBS, binding was visualized using 3-amino-9-ethylcarbazole (AEC) (Sigma-Aldrich) and slides were counterstained using hematoxylin.

### Data and statistical analysis

The microarray data were processed with Mapix microarray image analysis software V.8.1.0 (Innopsys) and further analyzed using our home written Microsoft Excel macro [46]. Data were fitted using Prism Version 10.1.1 (GraphPad Software, Inc).



## **Acknowledgements**

We thank Andrea Gröne and Erik Weerts from the Division of Pathology, Department of Biomolecular Health Sciences, Faculty of Veterinary Medicine of Utrecht University, the Netherlands for providing paraffin-embedded tissues and assistance with the assignment of structures in tissues.

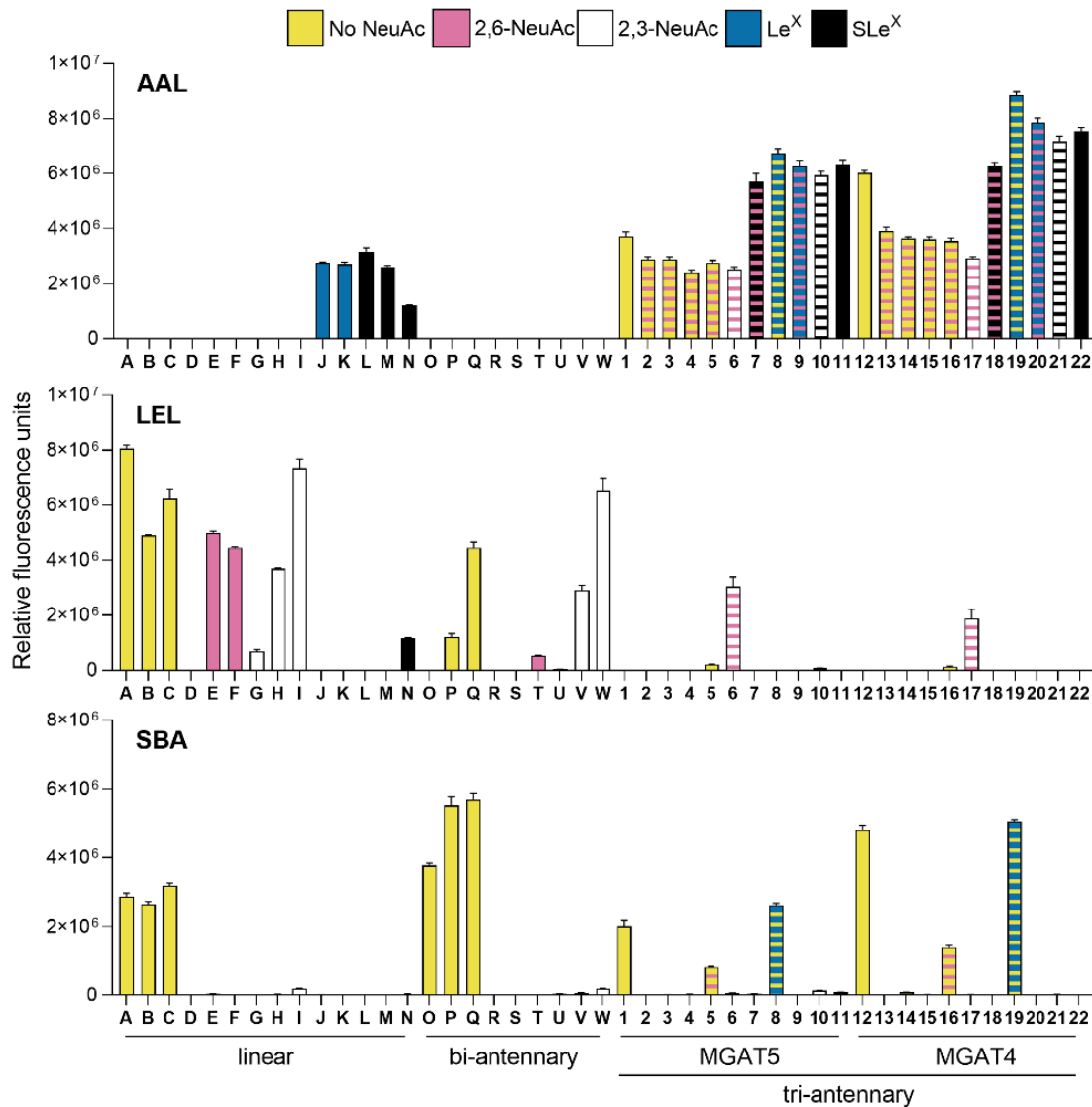
## **Funding**

R.P.dV. is a recipient of an ERC Starting Grant from the European Commission (802780) and a Beijerinck Premium of the Royal Dutch Academy of Sciences.

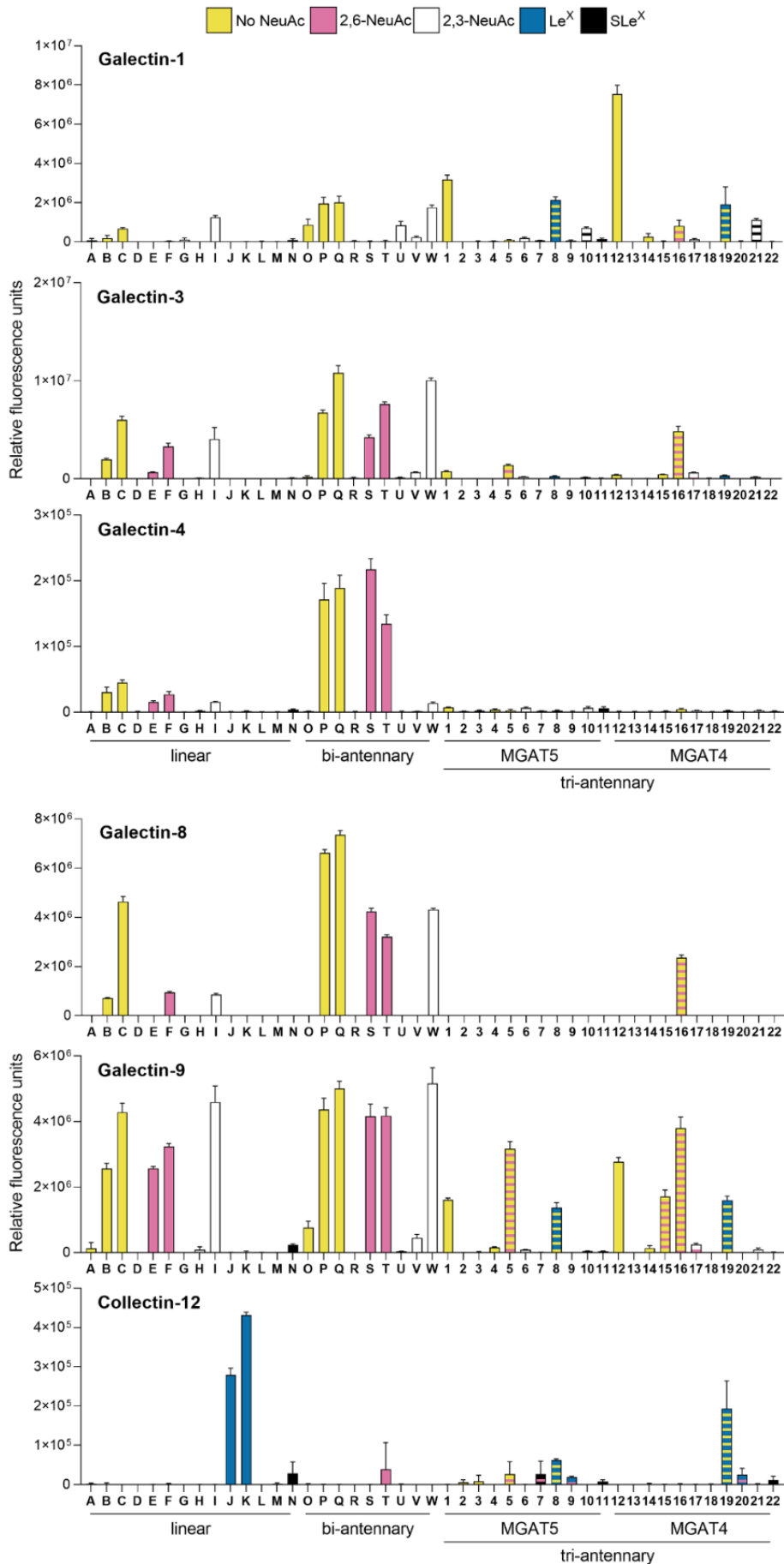
## **Author contributions**

Glycan synthesis, T.L., N.W., L.L.; glycan microarray experiments, R.P.dV., M.W., L.L.; analysis of glycan microarray experiments, C.M.S.; protein histochemistry, C.M.S.; expression and production of hemagglutinins, C.M.S.; data visualization, C.M.S.; conceptualization, C.M.S., R.P.dV., T.L, G.J.B.; supervision, R.P.dV., G.J.B., funding acquisition, R.P.dV., G.J.B.; writing—original draft preparation, C.M.S.; project administration, R.P.dV., G.J.B.; writing—review and editing, all authors.

Supplementary figures

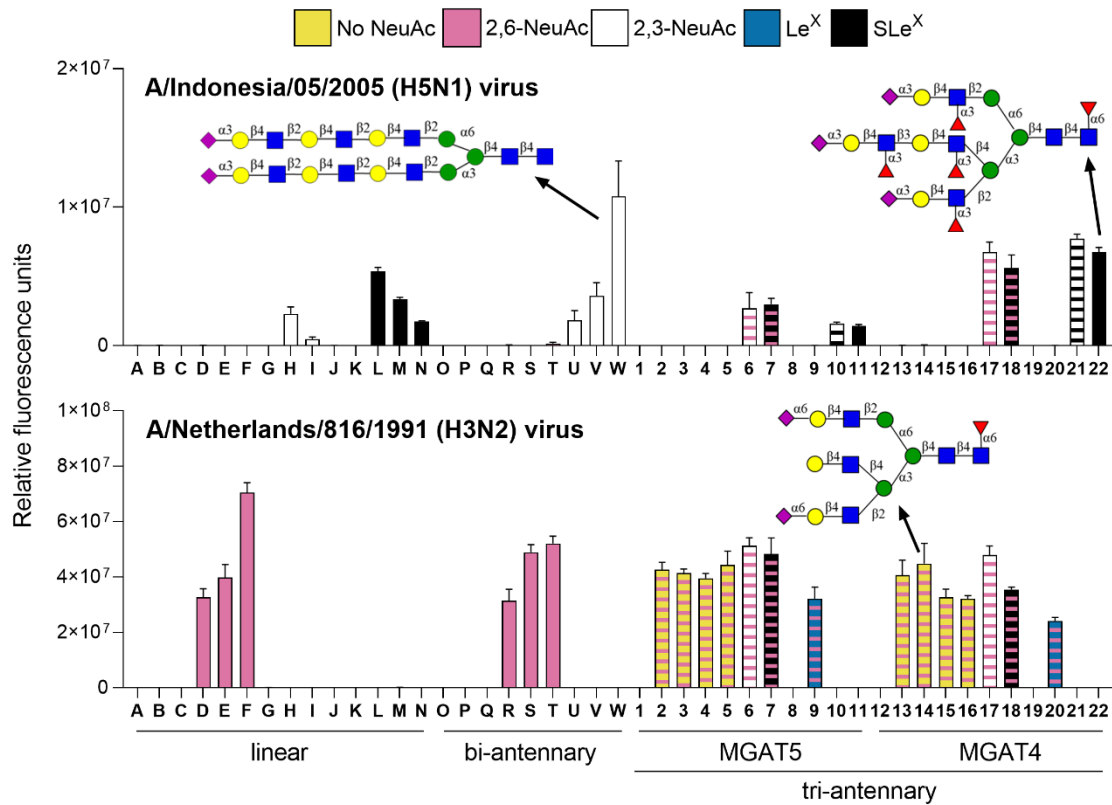


**Fig S1. Microarray binding of plant lectins to synthetic glycans.** For the tri-antennary *N*-glycans, either the MGAT4 or MGAT5 arm was elongated to two LacNAc repeating units. Glycans were terminating without a NeuAc (yellow), or with  $\alpha$ 2,6-linked NeuAc (pink),  $\alpha$ 2,3-linked NeuAc (white), Le<sup>x</sup> (blue), or SLe<sup>x</sup> (black). Bars with two colors indicate glycans terminating in different epitopes on different arms. The glycan microarray was used to probe the presence of fucosylated glycans (*Aleuria aurantia* lectin, AAL), elongated glycans (*Lycopersicon esculentum* lectin, LEL), and glycans with a terminal galactose (soybean agglutinin, SBA). See Fig. 2 for the structures of A-W and 1-22. Bars represent the mean  $\pm$  SD (n=4).



**Fig S2. Microarray binding of galectins (1, 3, 4, 8, 9) and collectin-12 to synthetic glycans.**

For the tri-antennary *N*-glycans, either the MGAT4 or MGAT5 arm was elongated to two LacNAc repeating units. Glycans were terminating without a NeuAc (yellow), or with  $\alpha$ 2,6-linked NeuAc (pink),  $\alpha$ 2,3-linked NeuAc (white), Le<sup>X</sup> (blue), or SLe<sup>X</sup> (black). Bars with two colors indicate glycans terminating in different epitopes on different arms. See Fig. 2 for the structures of A-W and 1-22. Bars represent the mean  $\pm$  SD (n=4).



**Fig S3. Microarray binding of avian H5N1 (A/Indonesia/05/2005, specific for  $\alpha$ 2,3-linked sialic acids) and human H3N2 (A/Netherlands/816/1991, specific for  $\alpha$ 2,6-linked sialic acids) whole influenza A viruses.** For the tri-antennary *N*-glycans, either the MGAT4 or MGAT5 arm was elongated to two LacNAc repeating units. Glycans were terminating without a NeuAc (yellow), or with  $\alpha$ 2,6-linked NeuAc (pink),  $\alpha$ 2,3-linked NeuAc (white), Le<sup>X</sup> (blue), or SLe<sup>X</sup> (black). Bars with two colors indicate glycans terminating in different epitopes on different arms. See Fig. 2 for the structures of A-W and 1-22. Bars represent the mean  $\pm$  SD (n=4).

## References

1. Jin, F. and F. Wang, The physiological and pathological roles and applications of sialyl Lewis x, a common carbohydrate ligand of the three selectins. *Glycoconj J*, 2020. 37(2): p. 277-291.
2. Vestweber, D., Ligand-specificity of the selectins. *J Cell Biochem*, 1996. 61(4): p. 585-91.
3. Kannagi, R., et al., Carbohydrate-mediated cell adhesion in cancer metastasis and angiogenesis. *Cancer Sci*, 2004. 95(5): p. 377-84.
4. Thompson, A.J., R.P. de Vries, and J.C. Paulson, Virus recognition of glycan receptors. *Curr Opin Virol*, 2019. 34: p. 117-129.
5. Kobayashi, M., et al., Carbohydrate-dependent defense mechanisms against *Helicobacter pylori* infection. *Curr Drug Metab*, 2009. 10(1): p. 29-40.
6. Seidman, D., et al., Essential domains of *Anaplasma phagocytophilum* invasins utilized to infect mammalian host cells. *PLoS Pathog*, 2015. 11(2): p. e1004669.
7. Verhagen, J.H., et al., Host range of influenza A virus H1 to H16 in Eurasian ducks based on tissue and receptor binding studies. *J Virol*, 2021. 95(6).
8. Zhao, C. and J. Pu, Influence of host sialic acid receptors structure on the host specificity of influenza viruses. *Viruses*, 2022. 14(10).
9. Phang, R. and C.H. Lin, Synthesis of type-I and type-II LacNAc-repeating oligosaccharides as the backbones of tumor-associated Lewis antigens. *Front Immunol*, 2022. 13: p. 858894.
10. Jia, N., et al., The human lung glycome reveals novel glycan ligands for influenza A virus. *Sci Rep*, 2020. 10(1): p. 5320.
11. Ge, S. and Z. Wang, An overview of influenza A virus receptors. *Crit Rev Microbiol*, 2011. 37(2): p. 157-65.
12. Patel, T.P., et al., Isolation and characterization of natural protein-associated carbohydrate ligands for E-selectin. *Biochemistry*, 1994. 33(49): p. 14815-24.
13. Thomas, V.H., Y. Yang, and K.G. Rice, In vivo ligand specificity of E-selectin binding to multivalent sialyl Lewis X N-linked oligosaccharides. *J Biol Chem*, 1999. 274(27): p. 19035-40.
14. Thomas, L.J., et al., Production of a complement inhibitor possessing sialyl Lewis X moieties by in vitro glycosylation technology. *Glycobiology*, 2004. 14(10): p. 883-93.
15. Liu, L., et al., Streamlining the chemoenzymatic synthesis of complex N-glycans by a stop and go strategy. *Nat Chem*, 2019. 11(2): p. 161-169.
16. Liu, L., et al., Improved isolation and characterization procedure of sialylglycopeptide from egg yolk powder. *Carbohydr Res*, 2017. 452: p. 122-128.
17. Sastre Torano, J., et al., Identification of isomeric N-glycans by conformer distribution fingerprinting using ion mobility mass spectrometry. *Chemistry*, 2021. 27(6): p. 2149-2154.
18. Broszeit, F., et al., N-glycolylneuraminic acid as a receptor for influenza A viruses. *Cell Rep*, 2019. 27(11): p. 3284-3294.e6.
19. Shibuya, N., et al., The elderberry (*Sambucus nigra* L.) bark lectin recognizes the Neu5Ac(alpha 2-6)Gal/GalNAc sequence. *J Biol Chem*, 1987. 262(4): p. 1596-601.
20. Iglesias, J.L., H. Lis, and N. Sharon, Purification and properties of a D-galactose/N-acetyl-D-galactosamine-specific lectin from *Erythrina cristagalli*. *Eur J Biochem*, 1982. 123(2): p. 247-52.
21. van Die, I., et al., The dendritic cell-specific C-type lectin DC-SIGN is a receptor for *Schistosoma mansoni* egg antigens and recognizes the glycan antigen Lewis x. *Glycobiology*, 2003. 13(6): p. 471-8.
22. Gagarinov, I.A., et al., Protecting-group-controlled enzymatic glycosylation of oligo-N-acetylglucosamine derivatives. *Angew Chem Int Ed Engl*, 2019. 58(31): p. 10547-10552.
23. Smith, K.D., et al., The heterogeneity of the glycosylation of alpha-1-acid glycoprotein between the sera and synovial fluid in rheumatoid arthritis. *Biomed Chromatogr*, 2002. 16(4): p. 261-6.
24. Gaide, N., et al., Pathobiology of highly pathogenic H5 avian influenza viruses in naturally infected Galliformes and Anseriformes in France during winter 2015-2016. *Vet Res*, 2022. 53(1): p. 11.
25. Hiono, T., et al., Amino acid residues at positions 222 and 227 of the hemagglutinin together with the neuraminidase determine binding of H5 avian influenza viruses to sialyl Lewis X. *Arch Virol*, 2016. 161(2): p. 307-16.
26. Guo, H., et al., Highly pathogenic influenza A(H5Nx) viruses with altered H5 receptor-binding specificity. *Emerg Infect Dis*, 2017. 23(2): p. 220-231.
27. Eggink, D., et al., Phenotypic effects of substitutions within the receptor binding site of highly pathogenic avian influenza H5N1 virus observed during human infection. *J Virol*, 2020. 94(13).
28. Linster, M., et al., Identification, characterization, and natural selection of mutations driving airborne transmission of A/H5N1 virus. *Cell*, 2014. 157(2): p. 329-339.
29. Heider, A., et al., Alterations in hemagglutinin receptor-binding specificity accompany the emergence of highly pathogenic avian influenza viruses. *J Virol*, 2015. 89(10): p. 5395-405.
30. Broszeit, F., et al., Glycan remodeled erythrocytes facilitate antigenic characterization of recent A/H3N2 influenza viruses. *Nat Commun*, 2021. 12(1): p. 5449.
31. Hiono, T., et al., A chicken influenza virus recognizes fucosylated alpha2,3 sialoglycan receptors on the epithelial cells lining upper respiratory tracts of chickens. *Virology*, 2014. 456-457: p. 131-8.
32. Wen, F., et al., Mutation W222L at the receptor binding site of hemagglutinin could facilitate viral adaption from equine influenza A(H3N8) virus to dogs. *J Virol*, 2018. 92(18).
33. Kobayashi, D., et al., Turkeys possess diverse Sialpha2-3Gal glycans that facilitate their dual susceptibility to avian influenza viruses isolated from ducks and chickens. *Virus Res*, 2022. 315: p. 198771.
34. Wallace, L.E., et al., Respiratory mucus as a virus-host range determinant. *Trends Microbiol*, 2021. 29(11): p. 983-992.

35. Chan, R.W., et al., Infection of swine ex vivo tissues with avian viruses including H7N9 and correlation with glycomic analysis. *Influenza Other Respir Viruses*, 2013. 7(6): p. 1269-82.
36. Allahverdian, S., K.R. Wojcik, and D.R. Dorscheid, Airway epithelial wound repair: role of carbohydrate sialyl Lewisx. *Am J Physiol Lung Cell Mol Physiol*, 2006. 291(4): p. L828-36.
37. Chinoy, Z.S., et al., Chemoenzymatic synthesis of asymmetrical multi-antennary N-glycans to dissect glycan-mediated interactions between human sperm and oocytes. *Chemistry*, 2018. 24(31): p. 7970-7975.
38. Guan, M., et al., The sialyl Lewis X glycan receptor facilitates infection of subtype H7 avian influenza A viruses. *J Virol*, 2022. 96(19): p. e0134422.
39. Xia, B., et al., Altered O-glycosylation and sulfation of airway mucins associated with cystic fibrosis. *Glycobiology*, 2005. 15(8): p. 747-75.
40. Suzuki, N., T. Abe, and S. Natsuka, Structural analysis of N-glycans in chicken trachea and lung reveals potential receptors of chicken influenza viruses. *Sci Rep*, 2022. 12(1): p. 2081.
41. Bohorov, O., et al., Arraying glycomics: a novel bi-functional spacer for one-step microscale derivatization of free reducing glycans. *Glycobiology*, 2006. 16(12): p. 21C-27C.
42. de Vries, R.P., et al., The influenza A virus hemagglutinin glycosylation state affects receptor-binding specificity. *Virology*, 2010. 403(1): p. 17-25.
43. Nemanichvili, N., et al., Fluorescent trimeric hemagglutinins reveal multivalent receptor binding properties. *J Mol Biol*, 2019. 431(4): p. 842-856.
44. Prudden, A.R., et al., Synthesis of asymmetrical multiantennary human milk oligosaccharides. *Proc Natl Acad Sci U S A*, 2017. 114(27): p. 6954-6959.
45. Zong, C., et al., Heparan sulfate microarray reveals that heparan sulfate-protein binding exhibits different ligand requirements. *J Am Chem Soc*, 2017. 139(28): p. 9534-9543.
46. Liu, L., enthalpyliu/carbohydrate-microarray-processing: carbohydrate microarray processing (v1.0). Zenodo, 2017.



# Chapter 7

## Summary and future outlook

---

### Summary

In **chapter 2**, we genetically modified cells to present elongated glycans using the genes of the glycosyltransferases B3GNT2 and B4GALT1. We used these cells to investigate the binding and infection of contemporary human H3N2 influenza A viruses (IAVs) that use elongated glycans terminating in  $\alpha$ 2,6-linked *N*-acetylneuraminic acid (Neu5Ac) as their receptors. Flow cytometry analysis revealed that human H3 hemagglutinins (HAs) showed increased binding to the cells that presented elongated glycans. However, the increased number of elongated glycans on cells did not result in a greater infection efficiency of recent human H3N2 viruses. Therefore, we propose that H3N2 IAVs require a low number of suitable glycan receptors to infect cells and that an increase in the glycan receptor display above this threshold does not result in improved infection efficiency.

In **chapter 3**, we investigated the presence of *N*-glycolylneuraminic acid (Neu5Gc),  $\alpha$ 2,3-linked, and  $\alpha$ 2,6-linked Sias in animal models (mice, ferrets, guinea pigs, cotton rats, Syrian hamsters, tree shrews, domestic swine, and non-human primates) used to study human IAVs. We suggest using ferrets as animal models for human IAVs, since they closely resemble humans in the sialic acid content, both in the sialic acid linkages in different positions of the respiratory tract and the lack of Neu5Gc, lung organization, susceptibility, and disease pathogenesis. Additionally, we evaluated the role of Neu5Gc in infection using Neu5Gc binding viruses and CMAH  $-/-$  knockout mice and concluded that Neu5Gc is unlikely to be a decoy receptor.

In **chapter 4**, we then studied the molecular determinants for H7 and H15 IAVs binding to Neu5Gc using HAs from extinct highly pathogenic equine H7N7 viruses that exclusively bind Neu5Gc. The resolved crystal structure of the equine H7 HA in complex with Neu5Gc showed a high similarity with the receptor-binding site (RBS) of an avian H7 HA. In two distant avian H7 HAs and an H15 HA, we showed that mutation A135E is key for binding  $\alpha$ 2,3-linked Neu5Gc but does not abolish NeuAc binding. Additional mutations S128T, I130V, T189A, and K193R converted the specificity from NeuAc to Neu5Gc. Phylogenetic analysis revealed a clear distinction between equine and avian residues at positions 128, 130, 135, 189, and 193. The results provide insights into the adaptation of H7 viruses to Neu5Gc receptors.

In **chapter 5**, we investigated the influence of the five previously identified equine Neu5Gc-adapting mutations in avian H7 IAVs *in vitro* and *in vivo*. We showed that these mutations negatively affected virus replication in chicken cells, but not in duck cells, and positively affected replication in horse cells. Furthermore, the mutations reduced virus virulence and mortality in chickens but did not influence the low virulence of this virus in ducks. To explain why chickens and ducks were infected by these Neu5Gc-specific viruses despite the absence of Neu5Gc, we re-evaluated the receptor binding of H7 HAs. This showed that the mutated avian H7 HAs also bound to  $\alpha$ 2,3-linked NeuAc and sLe<sup>x</sup> epitopes, explaining why infection of ducks and chickens was possible. Interestingly, the  $\alpha$ 2,3-linked NeuAc and sLe<sup>x</sup> epitopes were only bound when presented on tri-antennary *N*-glycans.

In **chapter 6**, we showed that H5 HAs prefer to bind to tri-antennary *N*-glycan structures when compared to linear structures terminating in sLe<sup>x</sup>. Furthermore, some H5 HAs showed increased binding when the sLe<sup>x</sup> epitope was presented on an elongated MGAT4 arm, in contrast to an elongated MGAT5 arm. Additionally, we showed that sLe<sup>x</sup> epitopes are present on the trachea of several avian species, but not horses, pigs, and humans.

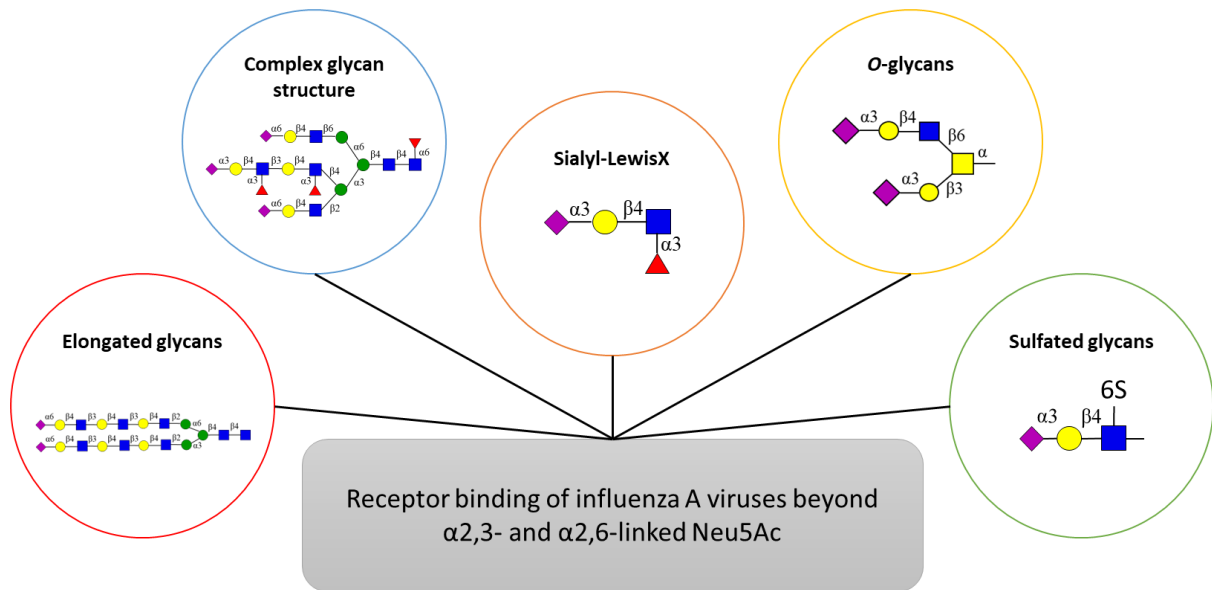
## **Continuation of the research on the receptor binding of influenza A virus beyond $\alpha$ 2,3- and $\alpha$ 2,6-linked Neu5Ac**

In this dissertation, we investigated the receptor binding of IAVs beyond  $\alpha$ 2,3- and  $\alpha$ 2,6-linked Neu5Ac. The relevance of the complexity of the glycan structure on IAV receptor binding is indeed increasingly recognized [1, 2]. Based on the obtained results, further studies into the receptor binding of IAV to understudied epitopes and biologically relevant complex glycans are recommended to improve our understanding of IAV pathogenesis (examples in Fig. 1). Furthermore, understanding the molecular determinants of these interactions may enable us to improve vaccine selection, design, and production, and could provide insights that are beneficial for the creation of antivirals.

### **What are the molecular determinants of binding to (elongated) complex *N*-glycans?**

In **chapter 2**, we used human H3N2 IAVs that specifically bind to glycans with at least three consecutive LacNAc repeating units. However, the molecular determinants of this interaction are currently unknown. Previously, amino acids in the 150-loop (155T, 156H, 159Y), 190-helix (186G, 193S, 194L), and 220-loop (220R, 222R, 225D) were suggested to accommodate the specific binding to elongated glycans [3, 4]. These amino acids can be used as a starting point in future research to gain a molecular understanding of how elongated glycans are bound. Furthermore, comparing the amino acid sequences and crystal structures of IAV HAs that are known to bind short or long glycans, as well as molecular modelling of glycan-HA complexes [4], can indicate relevant residues for this specific binding phenotype.

One of the ways to investigate receptor specificities towards elongated complex *N*-glycans is using mutated HAs or viruses on the glycan microarray [3], which



**Fig 1. Overview of some possible glycans and epitopes that can be investigated to increase the understanding of the receptor binding of influenza A viruses beyond  $\alpha$ 2,3- and  $\alpha$ 2,6-linked Neu5Ac.**

provides information about which exact glycans are bound. The expansion of the glycans presented on the glycan array, such as with the tri-antennary glycans that were described in **chapter 6** and recently synthesized asymmetrical *N*-glycans [5], may provide valuable information on the receptor binding of IAVs, such as arm preference and previously unknown receptors (such as sLe<sup>x</sup> epitopes in **chapter 5**). Although the synthesis of these glycans is a huge effort, recent insights into using chemoenzymatic synthesis methods [5-8] will simplify the synthesis of complex glycans.

At a lower high-throughput level, more specific information about HA-glycan interactions at the level of specific sugar moieties can be obtained using NMR methods [4]. Furthermore, investigation of the receptor binding in a more natural context is possible with glycan-modified erythrocytes [9] or hCK-B3GNT2 cells (**chapter 2**) that both display elongated glycans, although the exact glycans involved in the interactions cannot be distinguished when using cells.

In **chapter 5 and 6**, we showed that IAVs bind stronger to  $\alpha$ 2,3-linked Neu5Ac and sLe<sup>x</sup> when presented on tri-antennary *N*-glycans instead of linear glycans. This is possibly due to the specific glycan structure or a multivalency effect from the receptor side since multiple binding epitopes are presented on the same glycan. Furthermore, we found some HAs that preferred binding to epitopes only when presented on specific arms. It has been recently shown that human H3N2 viruses prefer to bind to the  $\alpha$ 1,3-antenna of *N*-glycans [5, 9]. These results emphasize the need to closely consider the exact glycan structures when investigating IAV receptor binding, which will aid the risk assessment of IAVs in different species, depending on the presence and abundance of the epitopes in relevant target cells in IAV hosts.

### Which influenza A viruses bind to sialyl-LewisX epitopes?

In **chapter 5**, we showed that H7 and H15 viruses bind to sLe<sup>x</sup> epitopes when presented on tri-antennary *N*-glycans. Additional preliminary results showed that most, but not all, avian IAVs bind to sLe<sup>x</sup> when presented on tri-antennary *N*-glycans. Therefore, the presence of sLe<sup>x</sup> epitopes on tissues of the respiratory and intestinal tract, locations relevant for IAV infection, should be considered as a potential risk for IAV infections. Since we observed that sLe<sup>x</sup> binding is not universal for IAVs (**chapter 5 and 6**), and that sLe<sup>x</sup> epitopes are not present in all species (**chapter 5 and 6**), specificity for this epitope could be considered as a potential species barrier.

Since sLe<sup>x</sup> epitopes are also presented on *O*-glycans [10] and little is known about the binding of IAVs to *O*-glycans, investigating the binding of IAVs to *O*-glycans that present sLe<sup>x</sup> could generate novel insights. Preliminary results showed that sLe<sup>x</sup> epitopes presented on core 2 *O*-glycans are bound by multiple subtypes of avian IAVs and that the binding is often enhanced by 6-sulfation of the glycans. However, sLe<sup>x</sup> presented on core 3 is only bound by avian H5 IAVs of the 2.3.4.4 subclade, again emphasizing that the exact glycan structure is relevant for receptor binding. Further investigation of IAV binding to sLe<sup>x</sup> epitopes presented on different *O*-glycan cores and the presence of these glycans in IAV hosts will increase the understanding of IAV pathogenesis.

So far, the only known molecular determinants for binding to sLe<sup>x</sup> epitopes are HA residues 222Q/R and 227R in H5 viruses [11-13] and 193K in H7 viruses [14]. It is currently unknown whether these mutations also enable interactions with sLe<sup>x</sup> in other subtypes, or whether other amino acids are involved in binding to sLe<sup>x</sup>. Investigation of the molecular determinants of sLe<sup>x</sup> binding will support the risk analysis of newly appearing and zoonotic IAVs.

### What is the role of *O*-glycans in influenza A virus infection?

*O*-glycans are underrepresented in IAV research, largely because *O*-glycans are absent on the glycan microarrays that are currently available. Nevertheless, binding of human and avian HAs to *O*-glycan cores 2, 3, and 4 extended with sialylated poly-LacNAc structures [15] as well as binding of human, avian, and swine IAs to simple *O*-glycans [16] has been shown previously. Additionally, an H1N1 virus showed more binding to core 1 *O*-glycans than an H3N2 virus [17]. More detailed NMR and molecular dynamics studies investigated the residues and sugar moieties that were involved in the binding of core 1 *O*-glycans to avian H5 IAVs [17].

Interestingly, it has recently been shown that IAVs can replicate in cells with truncated *N*-glycans, *O*-glycans, and glycolipids. Notably, the truncations affected human, but not avian, IAV replication and avian IAVs were suggested to utilize less prevalent shorter *O*-glycans (sialyl T and sialyl Tn) which were remaining on the cells [18]. In contrast, studies using cells without complex and hybrid *N*-glycans showed reduced IAV replication when additionally inhibiting *O*-glycan formation [17]. Others suggested *O*-glycans to be involved in the IAV life cycle, since a knockdown of GALNT3, an enzyme responsible for initiating *O*-glycan

synthesis, reduced IAV replication [19]. The presence of sialylated O-glycans in the submucosal glands and lower respiratory tract of ferrets [20] supports the hypothesis that O-glycans are involved in the IAV life cycle since IAV antigens are predominantly found in these locations [21]. Nonetheless, it is currently unknown whether all IAVs bind to O-glycans and what the exact function of O-glycans in the IAV life cycle is and, therefore, research investigating the interactions between IAVs and O-glycans is recommended.

Preliminary results suggest that many different subtypes of IAVs broadly bind to core 2 and core 3 O-glycans. Interestingly, Neu5Ac was also bound when it was not presented on the traditional LacNAc repeating unit, and the presence of multiple Neu5Ac enhanced the binding to O-glycans. The current strategies for chemically synthesizing complex O-glycans are demanding and time-consuming [22-26]. Therefore, the development of enzymatic synthesis routes would simplify the synthesis of O-glycans with different cores and extensions, which could in turn be used to evaluate IAV receptor binding.

### **Are sulfated glycans receptors for influenza A viruses?**

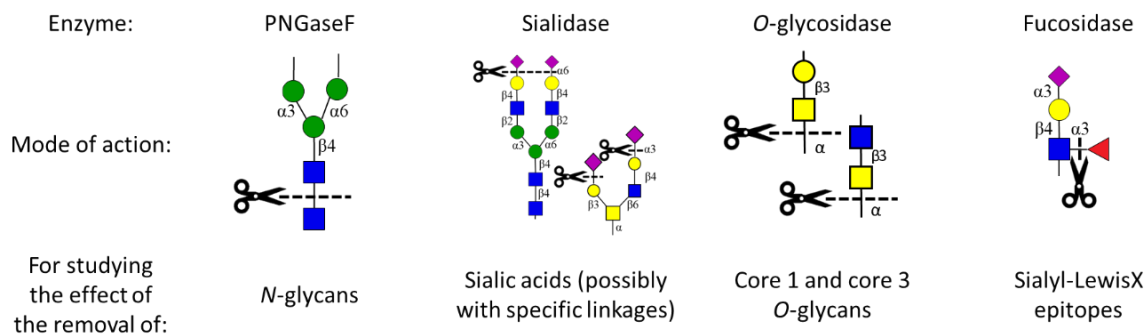
A glycan modification that is also understudied in biologically relevant glycans, in the context of IAV receptor binding, is sulfation. So far, the great majority of studies has been performed using sulfated tri- or tetrasaccharides [27-29]. Recently, binding of H3, H4, H5, and H6 IAVs to sulfated N- and O-glycans has been shown [30], emphasizing the relevance of further investigating IAV binding to biologically relevant glycans. Additional preliminary results suggested that 6-sulfated O-glycans may be universal receptors for IAVs. It is unknown whether the broad binding of sulfated epitopes is due to specific sulfated epitopes or the negative charge of the epitopes in general. The synthesis of biologically relevant sulfated N-glycans and O-glycans, supported by recently developed synthesis methods [30], could aid in studying the relevance of such glycans for IAVs.

### **Using enzymatic treatments on tissue slides to investigate receptor binding**

The glycan microarray provides the most accurate method for investigating which exact structures are bound. However, the natural variety of glycans is massive and, therefore, not all glycans can be synthesized and presented on the glycan microarray. Furthermore, glycans are not presented in their natural context and the density on the glycan microarray and the linkers between the surface and the glycans may affect binding. The most relevant, possibly very variable, glycans are present in the tracheal or intestinal tract of different species (used in **chapter 3 through 6**), which are the natural locations of infection. While an observation of binding to these tissues is informative, it does not provide information about which exact glycans or groups of glycans are bound.

To further elucidate the structures that are involved in these interactions, epitopes or complete groups of glycans should be selectively removed from the tissues. The binding of viruses or HAs can be compared before and after treatment to gain knowledge about relevant glycans. Some enzymes that can be used for this aim are well known, such as PNGaseF for the removal of N-glycans, different endo-enzymes to partially remove N-glycans, sialidases for the removal of sialic acids

(universal or specific for certain linkages), and O-glycosidases for the removal of core 1 and core 3 O-glycans [31] (Fig. 2). Nevertheless, the function of some of these enzymes is not optimal and the prices are currently high. Furthermore, enzymes to study many other glycan interactions are lacking. Therefore, it is important to continue the efforts of discovering and characterizing new and improved enzymes and making them affordable for a wide public. Recently, novel enzymes have been discovered, such as specific sulfatases [32, 33], fucosidases as shown in **chapter 5** [34], and influenza C virus hemagglutinin-esterases that remove 9-O-acetylation from Sias [35]. Not only will these enzymes aid in identifying the relevant glycan structures in tissue binding studies, but they can also be used in enzymatic synthesis to further expand the number of glycans that is available on the glycan array.



**Fig 2. Examples of enzymes that can remove specific groups of glycans or epitopes.**

### Studying receptor binding using (glycan-modified) cellular arrays

Alternatively, (living) cells, which also present broad arrays of glycans in a natural context, can be used to further investigate receptor binding, as we did in **chapter 2 and 5**. Cells may reveal enhanced specificities due to the presence of glycoconjugates, interactions with adjacent glycans, and their presentation on the cell membrane [36]. Cells that are commonly used to study IAV are Madin-Darby canine kidney (MDCK) cells because those are the most reliable, sensitive, and easy-to-handle cell line for IAV studies. Nevertheless, other cell types can be used to study IAVs, such as A549, BHK-21, HEK293, SPJL, LLC-MK2, and Vero cells [37-39].

Like with the tissue slides, enzymes can be used on cells for specifically cleaving a group of glycans or glycan epitopes (Fig. 2). The use of enzymes allows for the use of any cell type for binding studies, provided that the glycans that are targeted in the study are present on the cells. On the other hand, the use of enzymes relies on the amount of enzyme, reaction conditions, and reaction time. Therefore, incorrect usage of enzymes may lead to incomplete removal of glycans or epitopes. Furthermore, enzyme treatments are only recommended on fixed cells, since living cells renew the glycans that are present on the cell surface and therefore undo the action of the enzymes.



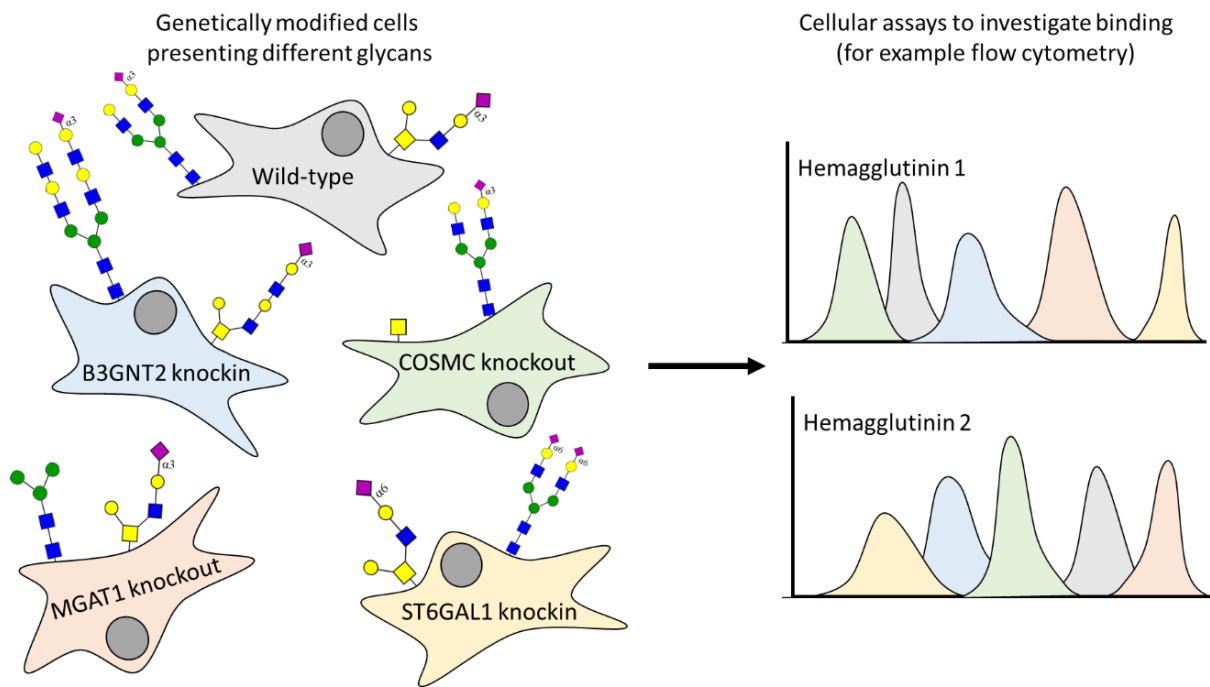
Alternatively, the synthesis of specific groups of glycans on cells can be constrained by adding inhibitors to cells while they are growing [40], such as O-glycans inhibitors [41], N-glycan inhibitors [42, 43], or glycolipid inhibitors [44-46]. Inhibitors will continually inhibit the glycan synthesis in cells and, therefore, living cells can be used in assays. However, the inhibitors must be added at each passage of cells, which can be considered as a disadvantage. Additionally, some inhibitors may be (to some extent) toxic to cells and the full impact of many of the inhibitors on cells is unknown. Furthermore, full inhibition of glycan synthesis depends on the quality of the inhibitor, the amount and concentration of inhibitor added, and the exact number of cells present. Therefore, using inhibitors may lead to inconsistencies.

Genetic modification of cells would modify the glycans present on the cell surface more consistently. Cells have been modified to present sLe<sup>x</sup> [14, 47], sulfated (O-)glycans [39, 48], Neu5Gc [49], truncated N-glycans, O-glycans, or glycolipids [18], or extended glycans (**chapter 2**, [50]). Furthermore, specific lectin-resistant cells have been created [51] and systematically knocking in or out genes of glycosyltransferases generated a range of cells with specific glycosylation patterns that covers a large part of the structural diversity of the (human) glycome [36]. To improve the available repertoire of cells to investigate specific research questions, it would be beneficial to continue to genetically modify cells by the overexpression or knock-out of genes involved in the production of different glycans, such as genes involved in producing multi-antennary glycans, specific O-glycan cores, O-acetylation, and sulfation. Glycan-modified cells are self-renewing and do not require the addition of, and therefore dependence on, external compounds. Thus, these cells are consistent in the presentation of the glycans on their surface, especially when clones are used. However, the synthesis pathways of glycans are not fully elucidated yet and different glycosyltransferases compete for substrates [36, 50]. Therefore, knockins and knockouts of genes can lead to the formation of unexpected glycans. Before the genetically modified cells can be used confidently in IAV binding studies, they require extensive characterization of the cell surface glycans using binding and mass spectrometry studies.

An array of cells presenting different glycans, obtained using enzymes, inhibitors, or genetic modification, can be used in cellular assays such as flow cytometry and fluorescent staining of cells with viruses or HAs. These methods create a cellular glycan array in which the IAV binding to different glycans can be compared (Fig. 3) [36]. To a smaller extent, we also used cellular glycan arrays in **chapter 2 and 5**. These cellular assays can also be performed with fixed and frozen cell stocks [36], further simplifying the analysis. The cellular glycan array would complement the printed glycan microarray and provide more information about IAV receptor binding in a natural context.

### **What is the relation between IAV binding and internalization?**

After receptor binding occurs, which we thoroughly investigated in this dissertation in **chapter 3 through 6**, the next step is the internalization of the virus particles into



**Fig 3. Example of a cellular glycan array.** Genetically modified cells are used (B3GNT2 knockin to elongate glycans, COSMC knockout to truncate O-glycans, MGAT1 knockout to truncate N-glycans, and ST6GAL1 knockin to present more  $\alpha$ 2,6-linked sialic acids). These cells present different glycans and can be used in cellular assays, such as flow cytometry, to investigate the receptor binding of different hemagglutinins.

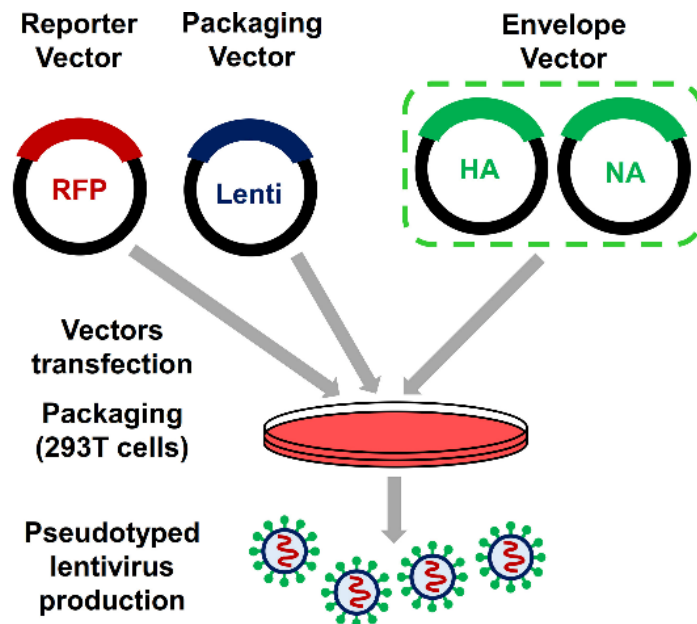
the cell. Some report that an increase in the available glycan receptors on cells leads to more efficient infection of cells [52, 53] and that a lower number of available receptors leads to less infection of cells [49]. Furthermore, a lower binding avidity correlated with decreased infection efficiency [53]. Additionally, low-affinity receptors were shown to contribute to IAV binding and entry [54]. Others suggest that an increase in the number of suitable glycan receptors on the cell surface does not lead to increased infection efficiencies [18, 50, 55-57], possibly because of a threshold of receptors that is required for efficient IAV infection as suggested in **chapter 2**.

Alternatively, binding may not always lead directly to internalization since some epitopes are suggested to be decoy receptors [58-60]. Besides glycan receptor binding, NA functionalities (binding or sialidase activity), specific (protein) receptors for internalization, or specific signaling pathways [57, 61-63] may be involved in the internalization of a virus into a cell. Further research into the relation between IAV receptor binding and virus uptake is necessary to understand the significance of IAVs binding to specific glycan receptors.

### **Pseudoviruses to investigate high-risk IAV infection**

Investigating the infection of IAVs requires viruses or pseudoviruses and a library of cells that present different glycans, as presented above. Pseudoviruses are replication-deficient and therefore safer to use, which is recommended for especially highly pathogenic IAVs or when mutating HAs to investigate the

molecular determinants of binding to a specific receptor. Pseudoviruses are generally virus particles of other viruses in which the envelope proteins are replaced by IAV HA and/or NA proteins and can be produced using different systems, such as retroviruses (Fig. 4) [64, 65], vesicular stomatitis viruses [66], adenoviruses [67], and nanoparticles [68]. When developing a pseudovirus system, ideally, the HAs would be easily interchangeable, to quickly generate an array of IAV pseudotypes. Setting up this system would enable looking beyond the receptor binding of IAVs and investigating the effect of the receptor binding on cellular internalization.

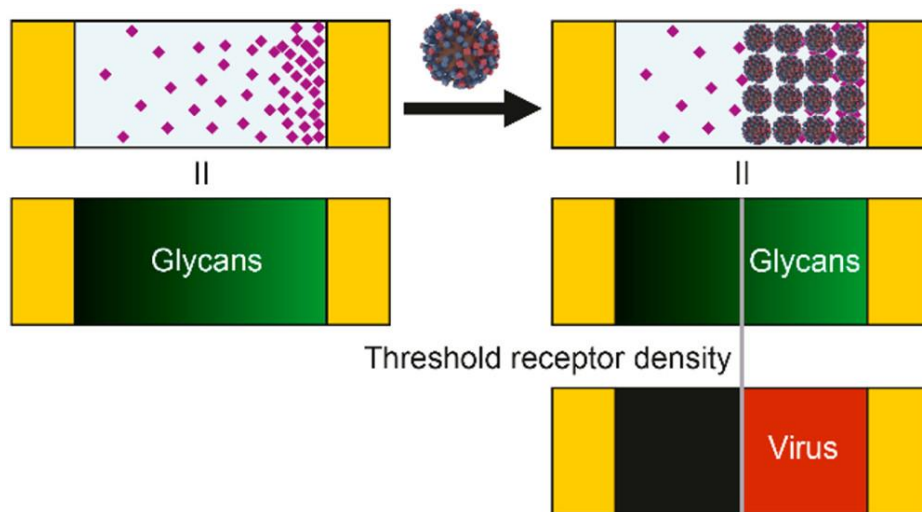


**Fig 4. Example of the production of lentiviral pseudoviruses with influenza A virus HA and NA surface proteins.** Modified from [65].

## How many glycans or HAs are required for IAV binding?

In **chapter 2** we suggested that a low amount of suitable glycan receptors is required for efficient infection. Similarly, recently, it was shown that IAVs can infect cells presenting only minor amounts of truncated sialylated N-glycans (0.3%), O-glycans (7%), and glycolipids [18]. While infection of these cells with truncated glycans was reduced by 95% for human H1 IAVs, the infection of avian H5 IAVs was only reduced by 20%, indicating that human IAVs may require a higher density of glycan receptors on the cell surface for infection and binding [18]. It is broadly acknowledged that multivalent interactions are involved in IAV receptor binding [69, 70]. Previously, it was estimated that eight interactions between an HA and a glycan receptor were required for efficient binding of a virus particle [71]. However, this number was found to depend on the glycan structures that were presented [71]. Apart from this article, little is known about the exact number of glycans or epitopes that is required for efficient binding. Therefore, the question remains exactly how many glycans and HA trimers are involved in binding. Furthermore, it is unclear whether it is more important to present multiple glycans in close proximity or multiple binding epitopes in the same glycan.

Further investigation of the required number of glycans could be done by presenting glycans on a surface in different gradients or dilutions, such as in ELISA or on a supported lipid bilayer [71] (Fig. 5). It would be interesting to investigate whether a different number of glycans is required for efficient (pseudo)virus binding when different glycans, such as *O*-glycans, simple *N*-glycans presenting only one Sia, and multi-antennary *N*-glycans presenting multiple Sias, are used. Additionally, it has been suggested that the presence of low-affinity receptors can enhance the binding to high-affinity receptors [54], and therefore investigating a combination of different receptors would also be of interest. Furthermore, investigating a broad range of IAVs would indicate whether all IAVs require a similar number of glycan receptors for binding.



**Fig 5. Schematic representation of an experiment in which glycans are presented in a gradient on a surface and virus binding is probed.** From [71].

## Which exact glycans are present in IAV hosts?

Once we know which and how many glycans are required for IAV binding (**chapter 2 through 6**) and infection, it would be useful to understand where these glycans are present in the required density in the respiratory and intestinal tract, which are the natural locations of IAV infections. A better understanding of the glycans at these locations would indicate potentially relevant receptors and thereby aid studies investigating IAV receptor binding. Furthermore, a clearer overview of the glycans present in biologically relevant locations would also provide a good basis for choosing appropriate animal models, like we showed for Neu5Gc in **chapter 3**. Lastly, the presence of specific glycans in some species, but not others, would indicate potential IAV species barriers. Currently, it is unclear which exact glycan structures are presented in the respiratory and intestinal tract and the glycans that are known to be present there are not quantified. Therefore, further research into the presence of glycans in different hosts and tissues would aid in future IAV research.

### Investigating the presence of glycan epitopes using antibodies, lectins, and hemagglutinins

A rough indication of the glycan epitopes can be obtained using lectins or antibodies in immunohistochemistry. Traditionally, lectins like SNA, MAL-I, and MAL-II have been used to investigate terminal epitopes that are relevant for IAV [72-74]. However, the specificities of these lectins are often not fully elucidated. Furthermore, antibodies against for example sLe<sup>x</sup> (**chapter 5 and 6**) or sulfated sLe<sup>x</sup> epitopes can be used [75]. Nevertheless, antibodies against some epitopes, like sLe<sup>x</sup> presented on a core 3 O-glycan, are not available. Efforts into the development of antibodies against currently non-detectable epitopes would benefit the studies of these epitopes. Alternatively, IAV HAs with a restricted specificity can be used to investigate the presence of specific epitopes, such as the HA of A/Taiwan/2/2013 H6N1 ( $\alpha$ 2,3-linked NeuAc) (**chapter 5**), A/chicken/Tainan/V156/1999 H6N1 ( $\alpha$ 2,3-linked NeuAc with a sulfated LacNAc) [76], or the Y161A mutant of A/Vietnam/1203/2004 H5N1 (Neu5Gc) (**chapter 5**) [77].

To efficiently investigate the presence of glycan epitopes in multiple species, tissue microarrays could be used, which present multiple tissues from one host and/or similar tissues from multiple hosts on one microscope slide [78]. Histochemical staining experiments provide a first indication of the epitopes that are present but will not provide information about the exact glycan structures and the number of these structures that are presented.

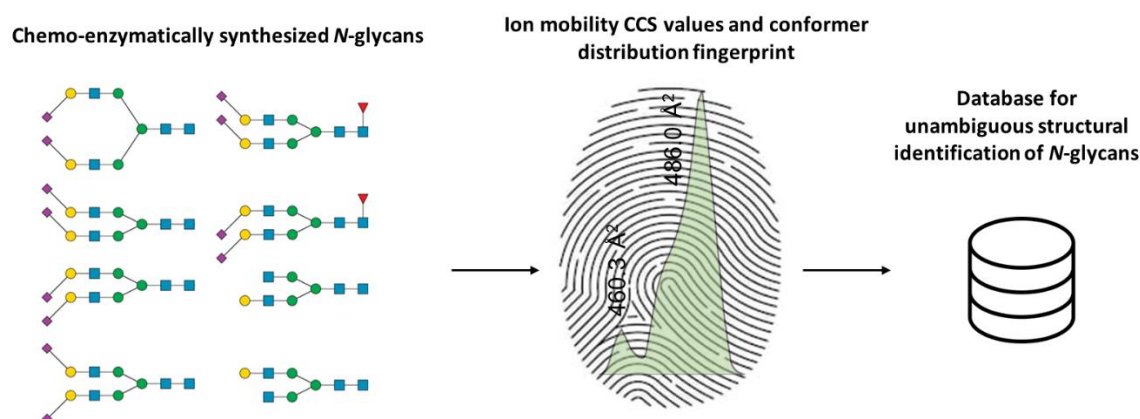
### Improving mass spectrometry methods to identify and quantify exact glycan structures

Since we showed that the exact glycan structures affect IAV binding (**chapter 5 and 6**), it is important to elucidate the exact structures of N-glycans, O-glycans, and glycolipids that are present in the respiratory and intestinal tract. Often, only the composition of glycans, without elucidating the exact structure, is obtained using mass spectrometry (MS) experiments [16, 79-83]. When exact glycan structures can be obtained from MS data, due to time constraints, often only a few structures are fully elucidated [84, 85]. MS analysis also allows for the relative or absolute quantification of glycans [86, 87]. Improving MS methods to achieve a higher resolution will benefit the quantitative and qualitative analysis of N-glycans, O-glycans, and glycolipids.

Currently, different methods are available for MS sample preparation and analysis [86, 87] and new methods are continuously developed. Recently, a method was developed for O-glycan release that maintains labile modifications such as sialic acids, sulfates, and acetyl esters [88]. It is important to realize that different possibilities for glycan release, labeling, purification, separation, and detection each come with advantages and disadvantages and many methods are biased towards a certain group or type of glycans. Furthermore, increasing the number of processing steps may result in reduced recovery of compounds and the loss of specific groups. Therefore, appropriate methods tailored to the research question

should be chosen and efforts should be continued to improve the available array of methods.

Recently, methods for improved O-glycan detection have been developed by introducing a precursor in cells and analyzing the O-glycans that were built on these precursors [89]. Furthermore, a combination of MS and chromatographic tools allowed for the characterization of glycan isomers [90]. Alternatively, isomers (such as glycans with differential length of branches, sialic acid linkages, or positions of sialic acid O-acetylation) can be identified by combining ion-mobility mass spectrometry (IMS-MS) with well-defined synthetic glycan standards (Fig. 6) [91-93]. IMS-MS provides the collision-cross-section (CCS) of an ion, which depends on the size and shape of the ion. Comparison of the CCS values of glycan standards and released glycans from biological samples reveals the exact glycan isomers present in these samples. The CCS values of glycan fragment ions of standards can also be determined, which facilitates the exact identification of other unknown glycans for which standards are not available. This will enable the *de novo* sequencing of glycans without requiring the labor-intensive synthesis of glycan standards. It is expected that the further development of these and other methods will lead to approaches for exact structure identification of glycans from complex (tissue) samples.

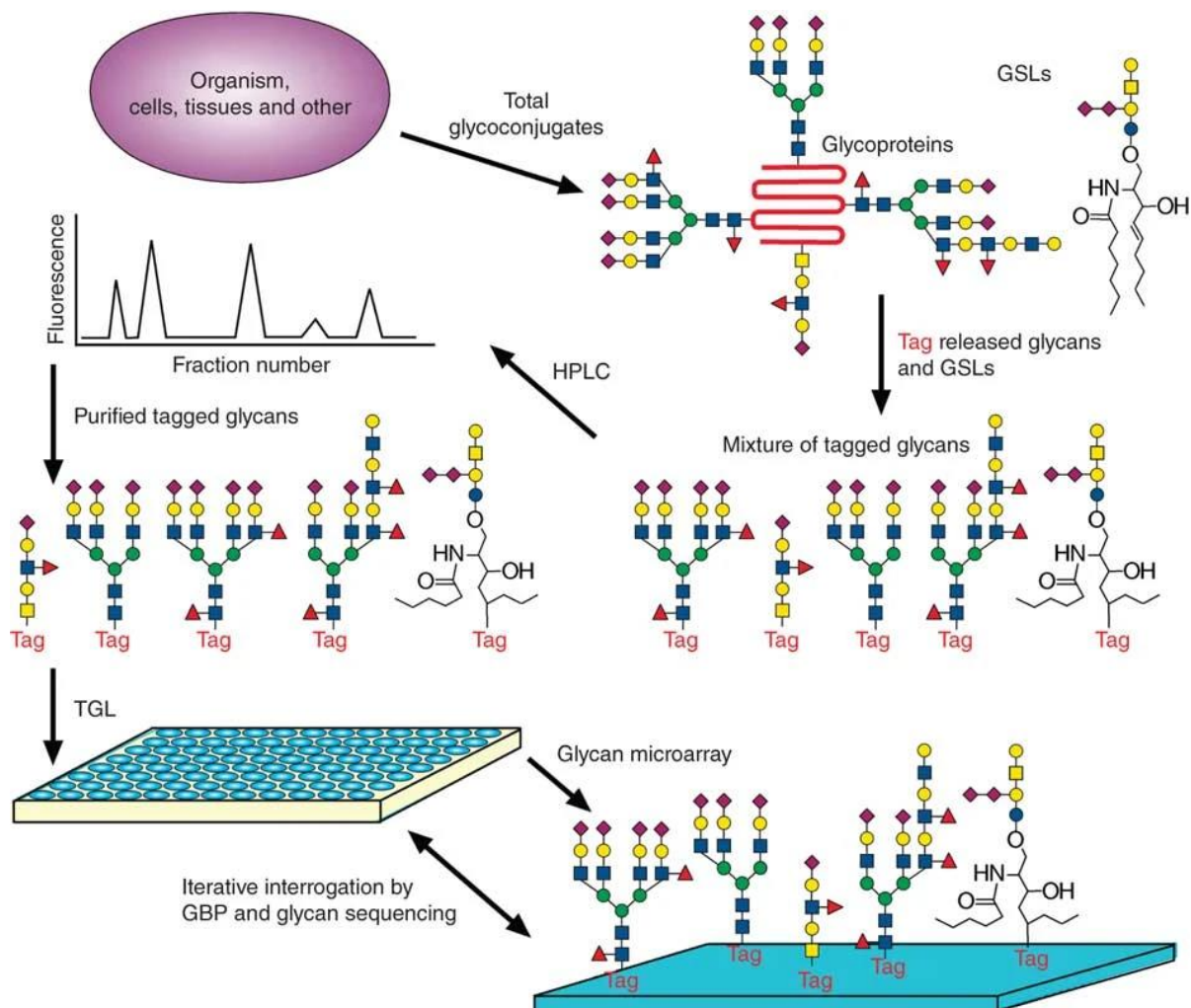


**Fig 6. Principle of ion-mobility mass spectrometry with well-defined synthetic glycans.** The synthesized glycans are used to obtain collision-cross-section (CCS) values and conformer distribution fingerprints of complete glycans and glycan fragments, which are saved in a database. The CCS values from released glycans from other samples can be compared to the CCS values in the database to elucidate the exact isomer structures that are present in the samples. From [92].

Alternatively, shotgun glycomics only investigates a selection of relevant IAV-binding glycans that are presented on tissues, which eliminates the time-consuming analysis of structures that are irrelevant to IAVs. In shotgun glycomics, all glycans (O-glycans, N-glycans, and glycolipids) are released from biologically relevant tissues, after which they are labeled, separated, and printed on a glycan microarray (Fig. 7). The glycan microarray is used to investigate the binding of IAVs or HAs. Only the structures of glycans that are bound by HAs or IAVs are further



elucidated using MS [94-96]. Nevertheless, shotgun glycomics usually does not allow for the exact identification of the glycan isomers that are involved in the interactions with IAV, unless it is combined with the IMS-MS methods that were mentioned above. Therefore, the glycans that are identified using shotgun glycomics can be used as a guideline for the chemo-enzymatically synthesis of different glycan isomers. These glycan isomers can be used to further investigate the exact receptors of IAVs.



**Figure 7. The principle of shotgun glycomics.** Glycans (O-glycans, N-glycans, and glycosphingolipids (GSL)) are released from glycoproteins, labeled with a tag, separated, and printed on a glycan microarray. Binding of HAs or viruses to the glycan microarray is investigated, after which the structures of glycans that are bound are further elucidated. From [94].

## References

1. Ge, S. and Z. Wang, An overview of influenza A virus receptors. *Crit Rev Microbiol*, 2011. 37(2): p. 157-65.
2. Long, J.S., et al., Host and viral determinants of influenza A virus species specificity. *Nat Rev Microbiol*, 2019. 17(2): p. 67-81.
3. Yang, H., et al., Structure and receptor binding preferences of recombinant human A(H3N2) virus hemagglutinins. *Virology*, 2015. 477: p. 18-31.
4. Unione, L., et al., Probing altered receptor specificities of antigenically drifting human H3N2 viruses by chemoenzymatic synthesis, NMR and modeling. *BiorXiv*, 2023.
5. Ma, S., et al., Asymmetrical bi-antennary glycans prepared by a stop-and-go strategy reveal receptor binding evolution of human influenza A viruses. *BiorXiv*, 2023.
6. Liu, L., et al., Streamlining the chemoenzymatic synthesis of complex N-glycans by a stop and go strategy. *Nat Chem*, 2019. 11(2): p. 161-169.
7. Zhao, X., et al., Recent chemical and chemoenzymatic strategies to complex-type N-glycans. *Front Chem*, 2022. 10: p. 880128.
8. Chao, Q., et al., Recent progress in chemo-enzymatic methods for the synthesis of N-glycans. *Front Chem*, 2020. 8: p. 513.
9. Broszeit, F., et al., Glycan remodeled erythrocytes facilitate antigenic characterization of recent A/H3N2 influenza viruses. *Nat Commun*, 2021. 12(1): p. 5449.
10. Morelle, W., et al., FAB-MS characterization of sialyl Lewis x determinants on poly lactosamine chains of human airway mucins secreted by patients suffering from cystic fibrosis or chronic bronchitis. *Glycoconj J*, 2001. 18(9): p. 699-708.
11. Hiono, T., et al., Amino acid residues at positions 222 and 227 of the hemagglutinin together with the neuraminidase determine binding of H5 avian influenza viruses to sialyl Lewis X. *Arch Virol*, 2016. 161(2): p. 307-16.
12. Guo, H., et al., Highly pathogenic influenza A(H5Nx) viruses with altered H5 receptor-binding specificity. *Emerg Infect Dis*, 2017. 23(2): p. 220-231.
13. Gaide, N., et al., Pathobiology of highly pathogenic H5 avian influenza viruses in naturally infected Galliformes and Anseriformes in France during winter 2015-2016. *Vet Res*, 2022. 53(1): p. 11.
14. Guan, M., et al., The sialyl Lewis X glycan receptor facilitates infection of subtype H7 avian influenza A viruses. *J Virol*, 2022. 96(19): p. e0134422.
15. Nycholat, C.M., et al., Recognition of sialylated poly-N-acetyl lactosamine chains on N- and O-linked glycans by human and avian influenza A virus hemagglutinins. *Angew Chem Int Ed Engl*, 2012. 51(20): p. 4860-3.
16. Walther, T., et al., Glycomic analysis of human respiratory tract tissues and correlation with influenza virus infection. *PLoS Pathog*, 2013. 9(3): p. e1003223.
17. Mayr, J., et al., Unravelling the role of O-glycans in influenza A virus infection. *Scientific Reports*, 2018. 8(1): p. 16382.
18. Liang, C.Y., et al., Avian influenza A viruses exhibit plasticity in sialylglycoconjugate receptor usage in human lung cells. *J Virol*, 2023: p. e0090623.
19. Nakamura, S., et al., Influenza A virus-induced expression of a GalNAc transferase, GALNT3, via microRNAs is required for enhanced viral replication. *J Virol*, 2016. 90(4): p. 1788-801.
20. Jayaraman, A., et al., Decoding the distribution of glycan receptors for human-adapted influenza A viruses in ferret respiratory tract. *PLoS One*, 2012. 7(2): p. e27517.
21. Memoli, M.J., et al., Multidrug-resistant 2009 pandemic influenza A(H1N1) viruses maintain fitness and transmissibility in ferrets. *J Infect Dis*, 2011. 203(3): p. 348-57.
22. Gutiérrez Gallego, R., et al., Enzymatic synthesis of the core-2 sialyl Lewis X O-glycan on the tumor-associated MUC1a' peptide. *Biochimie*, 2003. 85(3-4): p. 275-86.
23. Wang, S., et al., Chemoenzymatic modular assembly of O-GalNAc glycans for functional glycomics. *Nat Commun*, 2021. 12(1): p. 3573.
24. Xu, Z., et al., Diversity-oriented chemoenzymatic synthesis of sulfated and nonsulfated core 2 O-GalNAc glycans. *J Org Chem*, 2021. 86(15): p. 10819-10828.
25. Ma, W., et al., Integrated chemoenzymatic approach to streamline the assembly of complex glycopeptides in the liquid phase. *J Am Chem Soc*, 2022. 144(20): p. 9057-9065.
26. Chinoy, Z.S., K.W. Moremen, and F. Friscourt, A clickable bioorthogonal sydnone-aglycone for the facile preparation of a core 1 O-glycan-array. *European J Org Chem*, 2022. 2022(27): p. e202200271.
27. Gambaryan, A.S., et al., H5N1 chicken influenza viruses display a high binding affinity for Neu5Acalpha2-3Galbeta1-4(6-HSO3)GlcNAc-containing receptors. *Virology*, 2004. 326(2): p. 310-6.
28. Verhagen, J.H., et al., Host range of influenza A virus H1 to H16 in Eurasian ducks based on tissue and receptor binding studies. *J Virol*, 2021. 95(6).
29. de Vries, R.P., et al., A single mutation in Taiwanese H6N1 influenza hemagglutinin switches binding to human-type receptors. *EMBO Mol Med*, 2017. 9(9): p. 1314-1325.
30. Wu, Y., et al., Exploiting substrate specificities of 6-O-sulfotransferases to enzymatically synthesize keratan sulfate oligosaccharides. *JACS Au*, 2023.
31. Freeze, H.H. and C. Kranz, Endoglycosidase and glycoamidase release of N-linked glycans. *Curr Protoc Mol Biol*, 2010. Chapter 17: p. Unit 17 13A.
32. Luis, A.S., et al., Sulfated glycan recognition by carbohydrate sulfatases of the human gut microbiota. *Nat Chem Biol*, 2022. 18(8): p. 841-849.
33. Parenti, G., G. Meroni, and A. Ballabio, The sulfatase gene family. *Curr Opin Genet Dev*, 1997. 7(3): p. 386-91.
34. Wu, N.C. and I.A. Wilson, Influenza hemagglutinin structures and antibody recognition. *Cold Spring Harb Perspect Med*, 2020. 10(8).

35. Wang, M. and M. Veit, Hemagglutinin-esterase-fusion (HEF) protein of influenza C virus. *Protein Cell*, 2016. 7(1): p. 28-45.
36. Narimatsu, Y., et al., An atlas of human glycosylation pathways enables display of the human glycome by gene engineered cells. *Mol Cell*, 2019. 75(2): p. 394-407 e5.
37. Allen, J.D. and T.M. Ross, H3N2 influenza viruses in humans: viral mechanisms, evolution, and evaluation. *Hum Vaccin Immunother*, 2018. 14(8): p. 1840-1847.
38. Asaoka, N., et al., Low growth ability of recent influenza clinical isolates in MDCK cells is due to their low receptor binding affinities. *Microbes Infect*, 2006. 8(2): p. 511-9.
39. Sun, L., et al., Installation of O-glycan sulfation capacities in human HEK293 cells for display of sulfated mucins. *J Biol Chem*, 2022. 298(2): p. 101382.
40. Brown, J.R., B.E. Crawford, and J.D. Esko, Glycan antagonists and inhibitors: a fount for drug discovery. *Crit Rev Biochem Mol Biol*, 2007. 42(6): p. 481-515.
41. Wang, S.S., et al., Efficient inhibition of O-glycan biosynthesis using the hexosamine analog Ac(5)GalNTGc. *Cell Chem Biol*, 2021. 28(5): p. 699-710 e5.
42. Elbein, A.D., Inhibitors of the biosynthesis and processing of N-linked oligosaccharide chains. *Annu Rev Biochem*, 1987. 56: p. 497-534.
43. Mahoney, W.C. and D. Duksin, Separation of tunicamycin homologues by reversed-phase high-performance liquid chromatography. *J Chromatogr*, 1980. 198(4): p. 506-10.
44. McEachern, K.A., et al., A specific and potent inhibitor of glucosylceramide synthase for substrate inhibition therapy of Gaucher disease. *Mol Genet Metab*, 2007. 91(3): p. 259-67.
45. Zhao, H., et al., Inhibiting glycosphingolipid synthesis improves glycemic control and insulin sensitivity in animal models of type 2 diabetes. *Diabetes*, 2007. 56(5): p. 1210-8.
46. Ashe, K.M., et al., Efficacy of enzyme and substrate reduction therapy with a novel antagonist of glucosylceramide synthase for fabry disease. *Mol Med*, 2015. 21(1): p. 389-99.
47. Hiono, T., et al., A chicken influenza virus recognizes fucosylated alpha2,3 sialoglycan receptors on the epithelial cells lining upper respiratory tracts of chickens. *Virology*, 2014. 456-457: p. 131-8.
48. Ichimiya, T., et al., Sulfated glycans containing NeuAcalpha2-3Gal facilitate the propagation of human H1N1 influenza A viruses in eggs. *Virology*, 2021. 562: p. 29-39.
49. Chien, Y.A., et al., Single particle analysis of H3N2 influenza entry differentiates the impact of the sialic acids (Neu5Ac and Neu5Gc) on virus binding and membrane fusion. *J Virol*, 2023. 97(3): p. e0146322.
50. Kikuchi, C., et al., Glyco-engineered MDCK cells display preferred receptors of H3N2 influenza absent in eggs used for vaccines. *Nat Commun*, 2023. 14(1): p. 6178.
51. Patnaik, S.K. and P. Stanley, Lectin-resistant CHO glycosylation mutants. *Methods Enzymol*, 2006. 416: p. 159-82.
52. Takada, K., et al., A humanized MDCK cell line for the efficient isolation and propagation of human influenza viruses. *Nat Microbiol*, 2019. 4(8): p. 1268-1273.
53. Lin, Y.P., et al., Evolution of the receptor binding properties of the influenza A(H3N2) hemagglutinin. *Proc Natl Acad Sci U S A*, 2012. 109(52): p. 21474-9.
54. Liu, M., et al., Human-type sialic acid receptors contribute to avian influenza A virus binding and entry by hetero-multivalent interactions. *Nat Commun*, 2022. 13(1): p. 4054.
55. Kumari, K., et al., Receptor binding specificity of recent human H3N2 influenza viruses. *Virol J*, 2007. 4: p. 42.
56. Rimmelzwaan, G.F., et al., Attachment of infectious influenza A viruses of various subtypes to live mammalian and avian cells as measured by flow cytometry. *Virus Res*, 2007. 129(2): p. 175-81.
57. Mair, C.M., et al., Receptor binding and pH stability - how influenza A virus hemagglutinin affects host-specific virus infection. *Biochim Biophys Acta*, 2014. 1838(4): p. 1153-68.
58. Bateman, A.C., et al., Glycan analysis and influenza A virus infection of primary swine respiratory epithelial cells: the importance of NeuAc{alpha}2-6 glycans. *J Biol Chem*, 2010. 285(44): p. 34016-26.
59. Takahashi, T., et al., N-glycolylneuraminic acid on human epithelial cells prevents entry of influenza A viruses that possess N-glycolylneuraminic acid binding ability. *J Virol*, 2014. 88(15): p. 8445-56.
60. Wallace, L.E., et al., Respiratory mucus as a virus-host range determinant. *Trends Microbiol*, 2021. 29(11): p. 983-992.
61. Gulati, S., et al., Human H3N2 influenza viruses isolated from 1968 to 2012 show varying preference for receptor substructures with no apparent consequences for disease or spread. *PLoS One*, 2013. 8(6): p. e66325.
62. Chu, V.C. and G.R. Whittaker, Influenza virus entry and infection require host cell N-linked glycoprotein. *Proc Natl Acad Sci U S A*, 2004. 101(52): p. 18153-8.
63. Karakus, U., M.O. Pohl, and S. Stertz, Breaking the convention: sialoglycan variants, coreceptors, and alternative receptors for influenza A virus entry. *J Virol*, 2020. 94(4).
64. Wang, W., et al., Establishment of retroviral pseudotypes with influenza hemagglutinins from H1, H3, and H5 subtypes for sensitive and specific detection of neutralizing antibodies. *J Virol Methods*, 2008. 153(2): p. 111-9.
65. Huang, S.W., et al., Assessing the application of a pseudovirus system for emerging SARS-CoV-2 and re-emerging avian influenza virus H5 subtypes in vaccine development. *Biomed J*, 2020. 43(4): p. 375-387.
66. Cheresiz, S.V., et al., A vesicular stomatitis pseudovirus expressing the surface glycoproteins of influenza A virus. *Arch Virol*, 2014. 159(10): p. 2651-8.
67. Powell, T.J., et al., Pseudotyped influenza A virus as a vaccine for the induction of heterotypic immunity. *J Virol*, 2012. 86(24): p. 13397-406.
68. Xia, M., et al., Bioengineered pseudovirus nanoparticles displaying the HA1 antigens of influenza viruses for enhanced immunogenicity. *Nano Res*, 2022. 15(5): p. 4181-4190.
69. Overeem, N.J., E. van der Vries, and J. Huskens, A dynamic, supramolecular view on the multivalent interaction between influenza virus and host cell. *Small*, 2021. 17(13): p. e2007214.

70. Lu, W. and R.J. Pieters, *Carbohydrate-protein interactions and multivalency: implications for the inhibition of influenza A virus infections*. *Expert Opin Drug Discov*, 2019. 14(4): p. 387-395.
71. Overeem, N.J., et al., *Hierarchical multivalent effects control influenza host specificity*. *ACS Cent Sci*, 2020. 6(12): p. 2311-2318.
72. Shen, C.L., et al., *The infection of chicken tracheal epithelial cells with a H6N1 avian influenza virus*. *PLoS One*, 2011. 6(5): p. e18894.
73. Van Poucke, S.G., et al., *Replication of avian, human and swine influenza viruses in porcine respiratory explants and association with sialic acid distribution*. *Virology*, 2010. 7: p. 38.
74. Qin, Y., et al., *Age- and sex-associated differences in the glycopatterns of human salivary glycoproteins and their roles against influenza A virus*. *J Proteome Res*, 2013. 12(6): p. 2742-54.
75. Liu, W., et al., *A novel monoclonal antibody against 6-sulfo sialyl Lewis x glycans attenuates murine allergic rhinitis by suppressing Th2 immune responses*. *Sci Rep*, 2023. 13(1): p. 15740.
76. Kikutani, Y., et al., *E190V substitution of H6 hemagglutinin is one of key factors for binding to sulfated sialylated glycan receptor and infection to chickens*. *Microbiol Immunol*, 2020. 64(4): p. 304-312.
77. Broszeit, F., et al., *N-glycolylneuraminic acid as a receptor for influenza A viruses*. *Cell Rep*, 2019. 27(11): p. 3284-3294 e6.
78. Nemanichvili, N., et al., *Tissue microarrays to visualize influenza D attachment to host receptors in the respiratory tract of farm animals*. *Viruses*, 2021. 13(4).
79. Gizaw, S.T., et al., *Glycoblotting method allows for rapid and efficient glycome profiling of human Alzheimer's disease brain, serum and cerebrospinal fluid towards potential biomarker discovery*. *Biochim Biophys Acta*, 2016. 1860(8): p. 1716-27.
80. Li, Q., et al., *Comprehensive N-glycome profiling of cells and tissues for breast cancer diagnosis*. *J Proteome Res*, 2019. 18(6): p. 2559-2570.
81. Ruzhak, L.R., et al., *Differential N-glycosylation patterns in lung adenocarcinoma tissue*. *J Proteome Res*, 2015. 14(11): p. 4538-49.
82. Byrd-Leotis, L., et al., *Influenza binds phosphorylated glycans from human lung*. *Sci Adv*, 2019. 5(2): p. eaav2554.
83. Chan, R.W., et al., *Infection of swine ex vivo tissues with avian viruses including H7N9 and correlation with glycomic analysis*. *Influenza Other Respir Viruses*, 2013. 7(6): p. 1269-82.
84. Jia, N., et al., *The human lung glycome reveals novel glycan ligands for influenza A virus*. *Sci Rep*, 2020. 10(1): p. 5320.
85. Suzuki, N., T. Abe, and S. Natsuka, *Structural analysis of N-glycans in chicken trachea and lung reveals potential receptors of chicken influenza viruses*. *Sci Rep*, 2022. 12(1): p. 2081.
86. Lageveen-Kammeijer, G.S.M., et al., *High sensitivity glycomics in biomedicine*. *Mass Spectrom Rev*, 2021. 41(6): p. 1014-1039.
87. Cao, W.Q., et al., *Novel methods in glycomics: a 2019 update*. *Expert Rev Proteomics*, 2020. 17(1): p. 11-25.
88. Vos, G.M., et al., *Oxidative release of O-glycans under neutral conditions for analysis of glycoconjugates having base-sensitive substituents*. *Anal Chem*, 2023. 95(23): p. 8825-8833.
89. Zhang, Q., et al., *Synthesis and characterization of versatile O-glycan precursors for cellular O-glycomics*. *ACS Synth Biol*, 2019. 8(11): p. 2507-2513.
90. Peng, W., et al., *MS-based glycomics and glycoproteomics methods enabling isomeric characterization*. *Mass Spectrom Rev*, 2021. 42(2): p. 577-616.
91. Sastre Torano, J., et al., *Ion-mobility spectrometry can assign exact fucosyl positions in glycans and prevent misinterpretation of mass-spectrometry data after gas-phase rearrangement*. *Angew Chem Int Ed Engl*, 2019. 58(49): p. 17616-17620.
92. Sastre Torano, J., et al., *Identification of isomeric N-glycans by conformer distribution fingerprinting using ion mobility mass spectrometry*. *Chemistry*, 2020. 27(6): p. 2149-2154.
93. Vos, G.M., et al., *Sialic acid O-acetylation patterns and glycosidic linkage type determination by ion mobility-mass spectrometry*. *Nat Commun*, 2023. 14(1): p. 6795.
94. Song, X., et al., *Shotgun glycomics: a microarray strategy for functional glycomics*. *Nat Methods*, 2011. 8(1): p. 85-90.
95. Byrd-Leotis, L., et al., *Shotgun glycomics of pig lung identifies natural endogenous receptors for influenza viruses*. *Proc Natl Acad Sci U S A*, 2014. 111(22): p. E2241-50.
96. Bui, D.T., et al., *Mass spectrometry-based shotgun glycomics for discovery of natural ligands of glycan-binding proteins*. *Curr Opin Struct Biol*, 2022. 77: p. 102448.

# Appendices

## Nederlandse samenvatting

Influenza A virussen (griepvirussen) zorgen ervoor dat elk jaar miljoenen mensen en dieren ziek worden en overlijden. Als je ziek wordt, maken virussen nieuwe virusdeeltjes in jouw cellen. Om in deze cellen te kunnen komen, moeten virussen eerst binden aan structuren op de buitenkant van een cel. Deze structuren (de receptoren) zijn ketens van suikermoleculen (glycanen). Glycanen kunnen veel variëren, bijvoorbeeld in de lengte, de opbouw met verschillende suikermoleculen en de manier waarop deze suikermoleculen aan elkaar verbonden zijn. In elk dier zien de glycanen er net iets anders uit.

De overdracht van griepvirussen van vogels, een soort waarin veel griepvirussen voorkomen, naar mensen brengt een groot risico met zich mee. Daarom zijn er traditioneel veel studies gedaan naar de overdracht van griepvirussen tussen vogels en mensen. Griepvirussen binden aan sialzuren, specifieke suikermoleculen die aanwezig zijn op het uiteinde van glycanen. De glycanen in vogels en mensen zijn anders doordat de sialzuren op een andere manier aan de glycanen zijn verbonden. In vogels komt de  $\alpha 2,3$ -koppeling veel voor, terwijl dat in mensen juist de  $\alpha 2,6$ -koppeling is, wat ervoor zorgt dat het virus zich aan moet passen om de andere soort ziek te kunnen maken.

Echter is er nog veel onbekend over de invloed van de gehele complexe glycaanstructuur op de receptorbinding van griepvirussen. In dit proefschrift hebben we onderzoek gedaan naar de binding van griepvirussen aan hun receptoren, waarbij we verder hebben gekeken dan het traditionele onderscheid tussen de  $\alpha 2,3$ - en  $\alpha 2,6$ -koppelingen. Met deze resultaten kunnen we griepvirussen beter begrijpen en een betere risicoanalyse maken over de overdracht van griepvirussen tussen verschillende soorten dieren en mensen.

In **hoofdstuk 2** hebben we menselijke griepvirussen bestudeerd. Er zijn veel verschillende griepvirussen en deze zijn ingedeeld in groepen (subtypes). Het subtype wordt bepaald door de eiwitten die aanwezig zijn op de buitenkant van de virusdeeltjes, namelijk de hemagglutinine (HA) en neuraminidase (NA). Menselijke griepvirussen van het subtype H3N2 hebben zich in de loop der jaren aangepast om alleen lange glycanen te binden. Echter zijn deze lange glycanen niet aanwezig op de cellen die gebruikt worden om deze virussen te produceren voor onderzoek. Als er toch geprobeerd wordt om de virussen te produceren in deze cellen, leidt dit ertoe dat de virussen zich ofwel aanpassen aan de cellen door andere glycanen te binden die aanwezig zijn, of dat het onmogelijk is om grote hoeveelheden van deze virussen te verkrijgen. Beide gevallen zijn problematisch voor het bestuderen van deze virussen, aangezien hiervoor grote hoeveelheden van de originele vorm van het virus noodzakelijk zijn. In dit oofdstuk



hebben we cellen genetisch aangepast zodat zij langere glycanen presenteren op hun membraan. De humane H3N2 virussen, die lange glycanen nodig hebben om te binden, lieten meer binding zien aan deze cellen. Echter zorgden de langere glycanen er niet voor dat deze virussen de cellen beter konden infecteren. Daarom concludeerden we dat humane H3N2 virussen slechts een laag aantal geschikte glycanen nodig hebben en dat meer glycanen boven de drempelwaarde niet bijdragen aan een betere infectie.

In **hoofdstuk 3** hebben we onderzocht welke proefdieren het meest geschikt zijn voor het bestuderen van menselijke griepvirussen, waarbij we hebben gekeken naar de siaalzuren die aanwezig zijn in de verschillende dieren. De twee meest voorkomende siaalzuren zijn N-acetylneuraminezuur (Neu5Ac), het siaalzuur dat door de meeste griepvirussen wordt gebonden, en N-glycolylneuraminezuur (Neu5Gc), dat maar door enkele griepvirussen wordt gebonden en niet in alle dieren aanwezig is. We hebben de aanwezigheid van Neu5Gc,  $\alpha$ 2,3- en  $\alpha$ 2,6-gekoppelde siaalzuren in beeld gebracht en zijn tot de conclusie gekomen dat fretten qua siaalzuren het meeste op mensen lijken. Daarnaast zijn de longen, de vatbaarheid, en het ziektebeeld van fretten ook het meest vergelijkbaar met mensen. In dit onderzoek hebben we ook de rol van Neu5Gc in de infectie van griepvirussen bestudeerd en zijn tot de conclusie gekomen dat Neu5Gc mogelijk gebruikt wordt als een receptor van griepvirussen.

In **hoofdstuk 4** hebben we twee virussen onderzocht die specifiek binden aan een andere soort siaalzuur (Neu5Ac en Neu5Gc) maar erg op elkaar lijken. We hebben onderzocht welke verschillen tussen de virussen ervoor zorgen dat Neu5Ac of Neu5Gc wordt gebonden. Het eiwit hemagglutinine (HA), dat zich bevindt aan de buitenkant van het griepvirusdeeltje, is verantwoordelijk voor de binding van het griepvirus aan siaalzuren. We hebben gevonden dat één mutatie in de HA (van virussen van het subtype H7 en H15) ervoor zorgde dat zowel Neu5Ac als Neu5Gc gebonden werd. Met vijf mutaties in de HA werd Neu5Ac niet meer gebonden en was de HA specifiek voor Neu5Gc.

In **hoofdstuk 5** hebben we onderzocht wat de invloed is van de Neu5Gc-mutaties (uit het vorige hoofdstuk) op de virusgroei van H7 griepvirussen in cellen en dieren. De mutaties zorgden ervoor dat virussen beter groeiden in paardencellen, die veel Neu5Gc bevatten. Daarnaast was de virusgroei lager in kippen, die geen Neu5Gc bevatten. Echter was de virusgroei onveranderd in eendencellen, die ook geen Neu5Gc bevatten. De mutaties zorgden ervoor dat het virus minder ziekmakend en dodelijk was in kippen. Daarentegen zorgden de mutaties er ook voor dat eenden meer virus verspreidden, alhoewel eenden niet ziek werden van het virus. Hierna hebben we onderzocht hoe eenden en kippen, die geen Neu5Gc hebben, toch geïnfecteerd werden door een Neu5Gc-specifiek virus. De H7 griepvirussen bleken te binden aan zowel Neu5Gc, Neu5Ac en sialyl-LewisX, een structuur op het uiteinde van een glycaan met een extra suikergroep. Neu5Ac en sialyl-LewisX werden alleen gebonden als zij gepresenteerd werden op een specifieke complexe glycaanstructuur.

In **hoofdstuk 6** laten we zien dat H5 vogelgriepvirussen sterker binden aan sialyl-LewisX wanneer deze gepresenteerd wordt op complexe glycanen in plaats van



lineaire glycanen. Lineaire glycanen worden vaak gebruikt in onderzoek maar komen niet in mensen en dieren voor. Complexe glycanen komen in mensen en dieren voor en kunnen door het griepvirus gebruikt worden voor infectie. De luchtweg is de natuurlijke infectielocatie van het griepvirus. In dit hoofdstuk hebben we ook laten zien dat sialyl-LewisX aanwezig is in de luchtweg van verschillende vogels, maar niet in de luchtweg van zoogdieren.

## List of publications

Publications presented in this dissertation:

Li T\*, **Spruit CM\***, Wei N, Liu L, Wolfert MA, de Vries RP, Boons GJ. 2024. Chemoenzymatic synthesis of tri-antennary *N*-glycans terminating in sialyl-LewisX reveals the importance of glycan complexity for influenza A virus receptor binding. Manuscript submitted.

**Spruit CM\***, Palme DI\*, Li T, Ríos Carrasco M, Gabarroca García A, Sweet IR, Kuryshko M, Maliepaard JCL, Reiding KR, Scheibner D, Boons GJ, Abdelwhab EM, de Vries EM. 2024. Complex *N*-glycans are important for interspecies transmission of H7 influenza A viruses. *Journal of Virology*.

**Spruit CM**, Sweet IR, Maliepaard JCL, Bestebroer T, Lexmond P, Qiu B, Damen MJA, Fouchier RAM, Reiding KR, Snijder J, Herfst S, Boons GJ, de Vries RP. 2023. Contemporary human H3N2 influenza A viruses require a low threshold of suitable glycan receptors for efficient infection. *Glycobiology*.

**Spruit CM**, Zhu X, Tomris I, Rios-Carrasco M, Han AX, Broszeit F, van der Woude R, Bouwman KM, Luu MMT, Matsuno K, Sakoda Y, Russell CA, Wilson IA, Boons GJ, de Vries RP. 2022. *N*-glycolylneuraminic acid binding of avian and equine H7 influenza A viruses. *Journal of Virology*.

**Spruit CM**, Nemanichvili N, Okamatsu M, Takematsu H, Boons GJ, de Vries RP. 2021. *N*-glycolylneuraminic acid in animal models for human influenza A virus. *Viruses*.

Other publications:

Gräwe A, **Spruit CM**, de Vries RP, Merx M. 2023. Bioluminescent detection of viral surface proteins using branched multivalent protein switches. *RCS Chemical Biology*

Canales A, Sastre J, Orduna JM, **Spruit CM**, Perez-Castells J, Dominguez G, Bouwman KM, van der Woude R, Canada FJ, Nycholat CM, Paulson JC, Boons GJ, Jiménez-Barbero J, de Vries RP. 2023. Revealing the specificity of human H1 influenza A viruses to complex *N*-glycans. *JACS Au*.

Nemanichvili N, **Spruit CM**, Berends AJ, Grone A, Rijks JM, Verheije MH, de Vries RP. 2022. Wild and domestic animals variably display Neu5Ac and Neu5Gc sialic acids. *Glycobiology*.

Bermudez-Mendez E, Katrukha EA, **Spruit CM**, Kortekaas J, Wichgers Schreur PJ. 2021. Visualizing the ribonucleoprotein content of single bunyavirus virions reveals more efficient genome packaging in the arthropod host. *Nature Communications Biology*.

Kasurinen J, **Spruit CM**, Wicklund A, Pajunen MI, Skurnik M. 2021. Screening of bacteriophage encoded toxic proteins with a next generation sequencing-based assay. *Viruses*.

**Spruit CM**, Wicklund A, Wan X, Skurnik M, Pajunen MI. 2020. Discovery of three toxic proteins of *Klebsiella* phage fHe-Kpn01. *Viruses*.

Mohanraj U, Wan X, **Spruit CM**, Skurnik M, Pajunen MI. 2019. A toxicity screening approach to identify bacteriophage-encoded anti-microbial proteins. *Viruses*.

\* Authors contributed equally