INFLUENZA VIRUS INTERACTIONS WITH MODIFIED SIALIC ACIDS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Karen Naomi Barnard

May 2020

INFLUENZA VIRUS INTERACTIONS WITH MODIFIED SIALIC ACIDS

Karen Naomi Barnard, Ph. D.

Cornell University, 2020

Influenza A viruses (IAV) are an important human pathogen causing 3 to 5 million cases of severe illness and 290,000 to 650,000 deaths globally each year (WHO, 2018). IAVs have broad host ranges and are able to infect a wide array of animal species including humans, pigs, horses, dogs, waterfowl, and domestic poultry. IAV use sialic acids (Sia) as the primary receptor for infection via the hemagglutinin (HA) and neuraminidase (NA) glycoproteins. Sia are found in large amounts both on the cell surface as part of the glycocalyx and in mucus that protects the respiratory and gastrointestinal tracts. Sia may be chemically modified (including 7,9-O-, 9-O-acetyl, 5-*N*-glycolyl) and are attached to glycan chains through different linkages, which vary between hosts and tissues. While the importance of Sia α2,3- and α2,6-linkages to IAV tropism and evolution have been well studied, the roles of modified Sia in IAV host adaptation are not well known. Modified Sia have been identified as inhibitors of NA and HA, but their role during infection is unclear.

This dissertation discusses the regulation of 7,9-*O*- and 9-*O*-acetyl modifications in cells through the action of the sialate 7,9-*O*-acetyltransferase, CasD1, and its complementary enzyme the sialate 7,9-*O*-acetylesterase, SIAE. Additionally, the distribution of 7,9-O-, 9-O-, and 4-*O*-acetyl, along with 5-*N*-glycolyl, modified Sia is examined for different mouse tissues, and secreted mucus and erythrocytes from IAV host species. The effects of these modifications on IAV HA and NA function is also determined, as well as their effect on virus infection. Finally, the outcome of passaging three IAV virus strains on different MDCK cell lineages, including those expressing different modified Sia, was examine using deep

sequencing to determine changes in viral populations over time.

This dissertation provides new information on the role of chemically modified Sia on IAV virus:host interactions, which had previously been overlooked in the IAV research field. Understanding the roles of modified Sia will give a more complete understanding of the interactions of these viruses with this complex carbohydrate receptor as they move between hosts, thus better informing our knowledge of IAV evolution and emergence.

BIOGRAPHICAL SKETCH

Karen Naomi Barnard attended Kodiak High School in Kodiak, Alaska where she graduated in 2009 as co-valedictorian. She then attended the University of Alaska Fairbanks as a biological sciences major. While there, she worked on independent research projects on plant genetics with Dr. Naoki Takebayashi, and on rabies virus host adaptation with Dr. Karsten Hueffer. Studying virus:host interactions and evolution in Dr. Hueffer's lab fascinated her, so after graduating from UAF in 2013 *summa cum laude* with a Bachelor of Science degree, Karen decided to pursue a PhD to continue her research on virus evolution. Karen enrolled in a PhD program at Cornell University in the Biological and Biomedical Sciences program. After rotating through various labs, Karen elected to join Dr. Colin Parrish's lab, where she focused her research on influenza A host adaptation and receptor interactions. Following the completion of her PhD, Karen will be continuing her studies of influenza A host interactions by joining Dr. Jesse Bloom's lab at the Fred Hutchinson Cancer Research Center at the University of Washington in Seattle, WA.

For my parents, Brenda and David Barnard, and my wonderful sister, Emily Hansen.

But especially for Anthony Stewart, for always supporting me.

ACKNOWLEDGMENTS

Thank you to everyone who helped support me through this great and difficult endeavor, I don't know how I could have done this without you. Thanks to my PI, Colin Parrish, for being an A+ mentor and teaching me how to be an independent researcher. Thanks also to my committee, Drs. Jeongmin Song, Matthew DeLisa, and David Russell for their advice and suggestions. Thanks to my wonderful lab mates and colleagues who made work entertaining and kept me sane: Heather Calloway, Ian Voorhees, Brynn Alford-Lawrence, Robert Lopez-Astacio, Brian Wasik, Simon Frueh, Wendy Weichart, and Becky Harmen. Thanks to everybody else at the Baker Institute for Animal Health for making such a supportive and fun work environment that made coming to lab enjoyable. And thanks to my Ithaca friends who let me escape lab for hikes, drinks, food, soccer, and dog parties: Matthew Pennington, Alyssa Wetterau, Jennifer Yordy, Justin Nicholatos, Rachel Evanowski, Mary Kate Koch, and everyone else.

Special thanks to my parents, Brenda and David Barnard, and my dear sister, Emily Hansen, for always keeping my spirits up and helping me to push through even the toughest times. Thanks to my very good and very dumb dog, Yukon, who gave me equal measures of joy and exasperation. Please do not break any more windows. And finally, thanks to Anthony Stewart for being the best part of graduate school, for being my cheerleader, shoulder to cry on, and best friend. I couldn't have done the final sprint without you.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE
Introduction

The first point of contact between many invading pathogens and their target hosts is carbohydrates present on mucosal tissues or on the surface of cells. Carbohydrate chains, or glycans, are large organic structures built up of various monosaccharides, and play a variety of well described roles in cell communication, homeostasis, and as physical barriers (1–4). While for some pathogens, this barrier function of carbohydrates can prevent infection, there are also many bacteria, viruses, and parasites that have evolved to utilize glycans as co-receptors or primary receptors to initiate infection. Glycans are highly variable, being expressed on the surface of cells attached to glycoproteins and glycolipids as part of the glycocalyx, as well as being present at mucosal surfaces on secreted glycoproteins like mucins. The diversity of glycans is determined by the building of monosaccharide chains, with one monosaccharide added at a time, leading to a spectrum of structural complexity from simple glycans chains of only a few monosaccharides to highly branched structures built up of hundreds of individual monosaccharide units. The building of these glycan chains is also not directly encoded by the genome, but is instead determined by the expression of an interwoven network of enzymes in the Golgi and endoplasmic reticulum, the expression and activity of which determines the glycans added to proteins in a process called glycosylation (5, 6). Different enzymes can also add chemical modifications to the monosaccharides that make up the glycan chains, leading to even further structurally variability. All of these enzymes involved in glycosylation can vary dramatically between cell types, tissues, and even between species leading to complex glycan environments that pathogens must navigate.

Many of these glycan chains on cell surfaces are terminated by a particular monosaccharide called sialic acid (Sia, **Fig. 1.1A**). Similar to glycan chains, Sia are also incredibly diverse with many different chemically modified forms existing in nature, varying

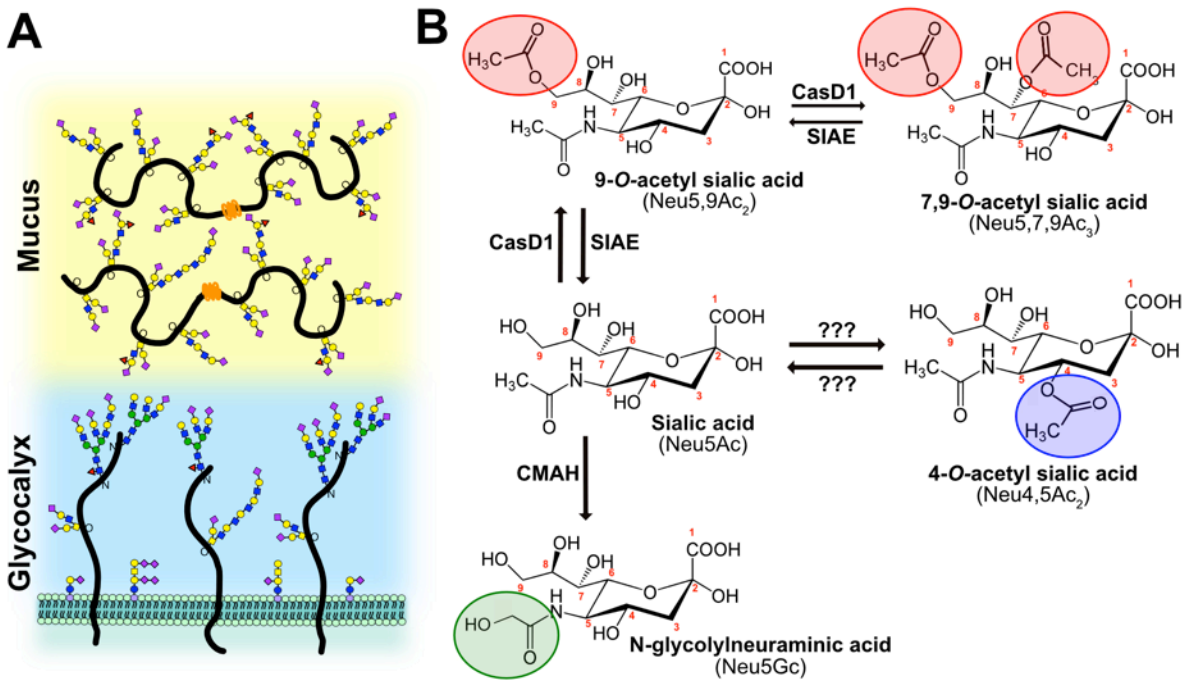


Figure 1.1

(**A**) Sialic acids (purple diamonds) terminate glycan chains on glycolipids and glycoproteins as part of the glycocalyx on the surface of cells. They can also terminate glycans on secreted glycoproteins, like mucins, that make up the protective mucosal barrier in gastrointestinal and respiratory tissue.
(**B**) Sialic acid (*N*-acetylneuraminic acid, Neu5Ac) can be modified by the addition of O-acetyl modifications at the C-4, 7, and 9 positions, or by the hydroxylation of the *N*-acetyl group at C-5 to form *N*-glycolylneuraminic acid (Neu5Gc) by the enzyme cytidine 5'-monophosphate-N-acetylneuraminic acid hydroxylase (CMAH). The sialate *O*-acetyltransferase, CasD1, adds acetyl groups at C-7 from which it migrates to the C-9 position ($Neu5,9Ac_2$) under physiological conditions. This can allow for an additional acetyl group to be added by CasD1 to C-7 ($Neu5,7,9Ac_3$). The sialate *O*-acetylesterase, SIAE, can remove these acetyl modifications, restoring the unmodified Neu5Ac form of sialic acid. *O*-acetyl modifications can also be added at the C-4 position by a specific 4-*O*-acetyltransferase ($Neu4,5Ac_2$) and removed by a 4-*O*-acetylesterase. However, the genes for these enzymes have not yet been identified.

between tissues and animal species (7, 8). As the common terminal sugar, Sia is often targeted by viruses, bacterial adhesion proteins and toxins, as well as parasites (4, 9). This includes important human pathogens such as influenza viruses, which are responsible for infecting between 9.3 million and 49 million people and killing 12,000 to 79,000 people annually in the United States alone (10, 11). Influenza A virus is of particular concern for human health as the emergence of novel virus strains from animal reservoirs has historically lead to global pandemics and loss of life, such as in 1918 and 2009 (12–14). Influenza A host tropism is determined by Sia diversity, with avian and mammalian strains each binding specifically to a different type of Sia (15, 16). However, with the great diversity of Sia that is known to vary between species, we do not know how many of the different Sia forms affect virus tropism and infection efficiency. In this dissertation, I will examine how some chemically modified forms of Sia effect influenza A virus infection efficiency and whether they play a role in host tropism.

1.1 Sialic Acids: simple sugars with built-in complexity

Before discussing the specifics of influenza viruses' interactions with Sia, it is important to understand the biology and chemistry of this monosaccharide. Sia is a nine-carbon monosaccharide made up of a core pyranose ring (C1-C6) and a glycerol side chain (C7-C9) with a *N*-acetyl chemical modification to C5 giving the core Sia form of *N*-acetylneuraminic acid (Neu5Ac) (7). While traditionally, Sia diversity is defined by the linkage of Sia to the underlying glycan chain, Sia can also have lactyl, glycolyl, methyl, acetyl, and sulfate chemical modifications added at the C4, C5, C7, C8, and/or C9 positions (17, 18) (**Fig. 1.1B**). These modifications can be added alone or in combination with each other, leading to the possibility of over 50 different unique forms of Sia that can vary between cells, tissues, and species. This multitude of forms allows for complex interactions with invading pathogens, as well as

commensal microbes as part of the microbiome (4, 19). These different chemical forms also play an important role in a diversity of functions that are integral for proper organismal development, immune responses, and normal cell homeostasis.

1.1a *O*-acetyl modifications to sialic acids

One of the most common chemical modifications to Sia is the addition of *O*-acetyl groups to the C7 and C9 positions of Sia (Neu5,7Ac_2, Neu5,9Ac_2, Neu5,7,9Ac_3) (17, 20). The addition of *O*-acetyl modifications to the C7 and C9 positions is mediated by the sialate *O*-acetyltransferase, Cas1 domain containing 1 (CasD1) and these modifications can removed by the complementary sialate *O*-acetylesterase (SIAE) (**Fig. 1.1B**) (21, 22). CasD1 is localized to the late Golgi, where it adds *O*-acetyl modifications to the activated form of Neu5Ac, CMP-Neu5Ac, at the C7 position where the modification can then migrate to the C8 and then C9 positions under physiological conditions, allowing the addition of a second *O*-acetyl modification at C7. The 7,9-*O*- or 9-*O*-acetylated Sia is then added to glycans on glycoproteins or glycolipids, where it is either retained in the Golgi or transported to the cell surface. The existence of a migrase enzyme aiding in the migration of the *O*-acetyl modification from C7 to C9 has been proposed but no enzyme has been yet identified (23, 24). The regulatory processes that control the number of acetyl groups added or their positions are not well defined, although clear differences in expression of 7,9-*O*-Ac and 9-*O*-Ac have been reported in mouse and human tissues and chicken embryos (25, 26).

The regulation and removal of 7,9-*O*- and 9-*O*-Ac is carried out by SIAE. The *SIAE* gene encodes two possible isoforms that vary in the presence of a proposed C-terminal localization tag in the lysosomal (Lse) form that is absent from the cytosolic form (Cse) (27–29). Lse has been found to localize to the Golgi and/or ER and on the surface of cells when over-expressed, with

the majority being secreted into the supernatant (27, 29). Immunofluorescence microscopy has found that Cse is found diffusely throughout the cytosol, where it is thought to remove 9-*O*- and 7-*O*-acetyl groups to recycle the Sia for reuse in glycosylation (29). However, the regulation of Lse and Cse expression has not been well defined, and species and tissue differences in expression remain to be fully characterized (28, 30).

O-acetyl modifications can also be added to the C4 position (Neu4,5Ac_2) in some species, such as horses, guinea pigs, mice, and some fish (31, 32). 4-*O*-Ac modifications are catalyzed by a separate sialate *O*-acetyltransferase which has only been functionally characterized and no gene encoding this enzyme has yet been identified (33, 34). The sialate 4-*O*-acetylesterase remains similarly uncharacterized, making 4-O-acetyl Sia possibly one of the least well understood of the Sia chemical modifications (35).

O-acetyl modifications have been found to change the function of Sia during interactions with both cellular and microbial proteins. 7,9-*O*- and 9-*O*-Ac play an important role in developmental, immunological, and cell-cell signaling processes (17, 35, 36). For example, binding of host lectins to Sia, including the Sialic acid-binding immunoglobulin-type lectins (Siglecs) that regulate many cell-cell signaling processes, can be affected by *O*-acetylation of Sia (4, 37, 38). This includes activation and differentiation of B- and T-cells which can be modulated by the presence of 9-*O*-acetyl Sia (39, 40). Incorrect regulation of 9-*O*- and 7,9-*O*-Ac through SIAE activity has been linked to autoimmune disorders through the development of auto-antibodies (41, 42). Regulation of 7,9-*O* and 9-*O*-acetyl levels by SIAE also appear to play important roles during early stages of embryonic development, spermatogenesis, and in different forms of cancer including acute lymphoblastic leukemia, colon cancer, and breast cancer (43–46).

In addition to regulating homeostatic processes in the host, *O*-acetylated Sia are also involved in many microbial interactions, both with commensal microbes and pathogens. Both pathogenic and commensal bacteria can utilize Sia as a carbon source, using sialidase enzymes to cleave Sia from cell surfaces and secreted glycoproteins (47). Some pathogenic bacteria will also incorporate harvested Sia into a protective capsular shell to prevent immune detection (48). However, *O*-acetyl modifications to Sia can block bacterial sialidase activity (49–51). To compensate, some bacteria also express esterases to removed *O*-acetyl modifications, increasing the efficiency of their sialidases (52, 53). Many viruses from both DNA and RNA-genome families use Sia as a co-receptor or primary receptor for infection (4, 9). Some viruses, including influenza viruses, can also utilize sialidase (neuraminidase, NA) enzymes to cleave Sia with *O*-acetylation blocking their activity (54, 55). *O*-acetyl modifications have also been shown to block the binding of reovirus to Sia (56). It is likely that *O*-acetyl modifications could also inhibit many other viruses from binding to Sia, although this remains an understudied interaction. However, *O*-acetyl modifications are not only inhibitory. There are also viruses that specifically target *O*-acetyl modified Sia as their primary receptor, including porcine torovirus (9-*O*-Ac), bovine coronavirus (7,9-*O*-Ac), infectious salmon anemia virus (4-*O*-Ac), mouse hepatitis virus strain-S (4-*O*-Ac), and influenza C and D (9-*O*-Ac) (31, 57–59). The exquisite specificity of these viral receptor-binding proteins for these modified Sia has also allowed them to be utilized as *in situ* probes for studying the localization and distribution of these modifications in different species and tissues, an advance that has greatly increased the ability to study these modified Sia forms (25, 26).

1.1b *N*-glycolyl modifications to sialic acids

While 7,9-*O*- and 9-*O*-Ac Sia have been found to be ubiquitously present in many vertebrate species, the *N*-glycolyl modification (Neu5Gc) has distinct expression across different linages of animals (60). Neu5Gc is formed by the modification of the acetyl group at C5 of the base Neu5Ac Sia form to a glycolyl group by the enzyme cytidine 5'-monophosphate-*N*-acetylneuraminic acid hydroxylase (CMAH) (**Fig. 1.1B**) (7). Unlike the *O*-acetyl modifications that can be reversibly removed by esterases, the addition of the glycolyl group at C5 is not reversible, making Neu5Gc a very stable modified form of Sia. Additionally, Sia modifying enzymes like CasD1 can utilize Neu5Gc as a substrate, although likely with much lower efficiency, allowing for further chemical modifications to be added to Neu5Gc (21, 61).

Neu5Gc has very distinct expression across different lineages of animals with some species having independently lost function of the *CMAH* gene and no longer synthesize Neu5Gc. This includes humans, new world monkeys, members of the *Mustilidae* family, western breeds of domesticated dogs, birds, and several species of marine mammals (8, 60, 62). One explanation for loss of CMAH could have been the selective drive to escape pathogen recognition of Neu5Gc as a receptor. For example, in humans, it is thought that CMAH activity was lost due to pressure from an ancestral *Plasmodium* species (63, 64). Indeed, many Sia-binding viruses that use humans as their specific host utilize unmodified Neu5Ac as their receptor while related zootropic viruses can utilize Neu5Gc. This includes polyomavirus-2, BK, and Merkel cell polyomaviruses in humans which all use Neu5Ac, while some polyomaviruses from other mammals (Simian Virus-40) use Neu5Gc (4, 9). Similarly, porcine and bovine rotaviruses use Neu5Gc as their receptor while human rotaviruses don't utilize Sia as a primary receptor at all (4, 9). Human-specific bacterial pathogens, too, seem to find Neu5Gc as a barrier to infection in other species.

Salmonella typhi, for example, only infects humans and utilizes a Sia-binding toxin during infection that binds specifically to Neu5Ac (65, 66). For human pathogens that rely on Neu5Ac, the presence of Neu5Gc in animal model systems, such as mice and rats, can be a difficulty that prevents these models from recapitulating the disease as seen in humans. For *S. typhi*, infection using a *CMAH*$^{-/-}$ mouse has been used as an alternative (67).

1.1c Sialic acids and mucins

Sia originally gained their name due to their discovery on salivary mucins (Greek for saliva is *sialos*) (7). Mucin proteins are some of the largest proteins encoded in the human genome, being ~200kDa in size with ~80% of the mass due to heavy glycosylation, including highly levels of sialylation (**Fig. 1.1A**) (68). Mucins are also key glycoproteins present in mucus barriers in the respiratory tract, gastrointestinal tract, reproductive tract, ocular surfaces, and in saliva (68–71). Mucus is also comprised of many other proteins including secreted antibodies, anti-microbial peptides, transferrin family proteins, and other cellular proteins secreted by immune cells, epithelial cells, and specialized secretory cells like goblet cells and sub-mucosal glands (72, 73). Mucins and other mucus proteins play important roles in protecting mucosal surfaces from invading pathogens such as viruses and bacteria (68, 74, 75). In the respiratory tract, for example, mucus can act as a trap for pathogens with the highly sialylated mucin proteins, like Muc5Ac and Muc5B, forming a tangled structure of polymers. Entangled pathogens are efficiently removed via the coordinating beating of airway cilia (71).

However, some pathogens have also developed methods to both exploit mucins and evade their protective effects. Sia on mucin proteins can function as an important carbon source and many bacteria, both commensal and pathogenic, encode sialidases (neuraminidases) for cleavage of Sia with some also expressing *O*-acetylesterases for removal of acetyl modifications

(47, 52, 53). For pathogenic bacteria, released Sia can also be taken up and incorporated into protective glycan-shield or capsules to mask them from immune recognition (48). Invading viruses have adopted several methods for penetrating through mucus to reach the epithelial cells of mucosal tissues. Some viruses, like herpesviruses, encode their own mucin-like proteins that allow them to move through mucus using charge interactions (76). However, the most common strategy used by many viruses, including orthomyxoviruses, coronaviruses, and paramyxoviruses, is to use neuraminidases or esterases to remove Sia decoy receptors to allow viruses to "chew" their way through the mucus layer (9, 77–79).

1.2 Influenza A, B, C, D: it's as easy as 123

Influenza viruses are enveloped viruses with negative sense, segmented RNA genomes and are part of the Orthomyxoviridae family, which includes four primary genera: influenza A (IAV), influenza B (IBV), influenza C (ICV), and the recently described influenza D (IDV). These four genera of viruses share many similar traits including the use of Sia as a primary receptor for infection of mucosal tissues, as well as similar replication strategies. However, there are some differences in their genomic structures, tropism, and infection methods that distinguish these diverse viruses from each other.

While many pathogens interact with Sia during infection, influenza viruses offer an important model for investigating the interaction of viruses with different Sia forms. IAV and IBV are both important human pathogens, and IAV in particular has periodic outbreaks of virus from zoonotic origins with the constant threat of an emergent pandemic virus, such as occurred in 1918, 1957, and 2009 (12, 14). IAV tropism is determined, in part, by different Sia linkages; however, the interaction between IAV and modified Sia has not been fully elucidated. Conversely, *O*-acetyl modified Sia are the primary receptor for both ICV and IDV (59, 80). Such

distinct Sia interactions across influenza viruses makes this family of viruses an ideal model for understanding the role of modified Sia in host tropism and infection efficiency.

1.2a Virus replication process

While there are some differences amongst the various orthomyxoviruses, which will be covered in the following section, their replication strategy follows a common theme (**Fig 1.2**). First, the virus glycoprotein (hemagglutinin [HA] in IAV and IBV, hemagglutinin-esterase [HE] in ICV and IDV) binds to Sia receptors on the surface of the cell, allowing for crosslinking of different sialylated cellular glycoproteins. This crosslinking triggers clatherin-mediated endocytosis of the virion into the cell inside of an endosome (81). As the endosome moves through the endolysosomal pathway, its internal pH drops. The viral M2 protein acts as a proton pump to lower the internal pH of the virion, which releases the viral genome from the M1 protein in preparation for fusion (82, 83). Once the endosome reaches a target acidic pH, the fusion glycoprotein of the virus then triggers allowing for fusion of the viral membrane with the endosomal membrane and subsequent release of the viral genome into the cytosol of the cell (82, 84). The virus genome, coated with NP proteins to block viral RNA sensors in the cytosol, is then trafficked to the nucleus via nuclear localization signals on NP to begin genome replication (82).

Because orthomyxoviruses have a negative-sense RNA genome, they must carry their own RNA-dependent RNA polymerase (RdRp) to copy their genome into positive-sense strands for protein production. The RdRp is comprised of three proteins: PB1, PB2, and PA. The RdRp appropriates the 5’ methylated cap from cellular mRNA through the endonuclease action of PB2, a process termed “cap snatching”, to prime transcription of viral mRNA (82, 85). Cap snatching not only allows the virus to prime its own RNA for transcription and efficient translation, but it

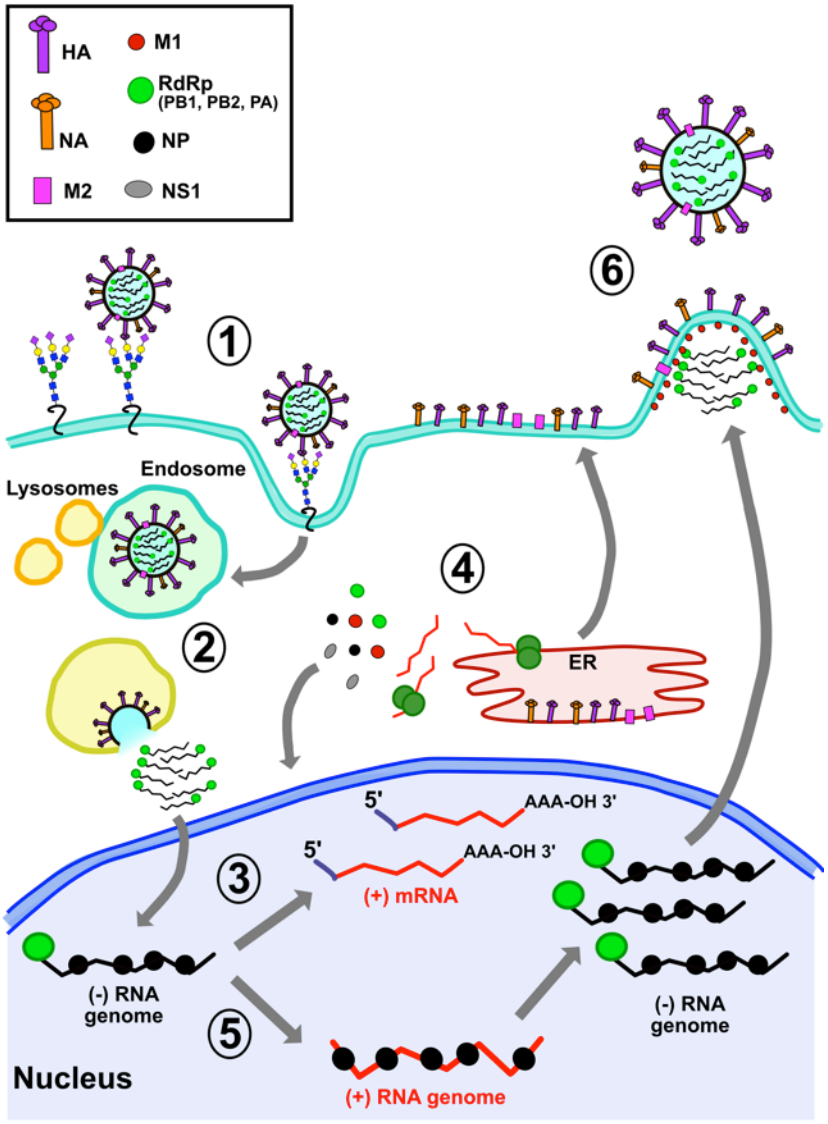


Figure 1.2

(**1**) The replication cycle for orthomyxoviruses, demonstrated with IAV here, starts by binding of the receptor-binding glycoprotein (HA, in purple) to the Sia receptor. This leads to uptake of the virus via clathrin-mediated endocytosis.
(**2**) As the endosome matures, it fuses with lysosomes that decrease the internal pH. The M2 proton pump (in pink), in turn, pumps H+ protons into the virion, which prepares the NP-coated RNA genomes for release into the cytoplasm. Once the endosome reaches an appropriate acidic level, the receptor-binding glycoprotein triggers fusion between the viral membrane and the endosomal membrane, releasing the genome segments into the cytosol.
(**3**) The nuclear localization signals on the NP-coated genomes (NP = black circles) lead to import of the gene segments into the nucleus. The viral RNA-dependent RNA polymerase (RdRp, green circle) copies the negative-sense RNA genomes into positive-sense mRNA. The RdRp utilizes cellular 5' methylated caps to prime copying of viral mRNA and also copies a poly(A) tail due to stuttering of the RdRp.

(**4**) Viral mRNA are transported to the cytosol where they are translated by host ribosomes for soluble proteins (NP in black, RdRp in green, NS1 in grey, M1 in red). Ribosomes on the ER translate the membrane-bound viral proteins (HA in purple, NA in orange, M2 in pink). NP and RdRP are transported to the nucleus while the membrane-associated proteins are transported to the cell surface.
(**5**) Once enough NP has been transported to the nucleus, it binds to the RNA genome segments, stabilizing the RdRp interaction with the negative-sense RNA genomes. This allows for complete positive-sense RNA templates to be copied for each gene segment. These templates are then used to make copies of each of the negative-sense RNA genome segments. The negative-sense RNA gene segments are then complexed with RdRp and NP to create the vRNP in preparation for transport to the cell surface for virion budding.
(**6**) The vRNPs associate with M1, HA, and NA at the cell surface membrane. Once all segments of the vRNPs are packaged, M1 drives the budding and release of the virion. The receptor-destroying glycoprotein (NA in orange) prevents the formation of viral aggregates.

also decreases cellular translation by destroying cellular mRNA transcripts, which can dampen immune responses and improve viral protein production (86). Viral mRNA is polyadenylated by stuttering of the RdRp on a section of poly(U) at the end of each gene segment (85). Each segment of the orthomyxovirus genome encodes at least one protein, although some segments, such as M and NS, encode two proteins that require host cellular mRNA splicing machinery for expression (82, 87). M2 and NEP, which are the spliced variants of the M and NS transcripts respectively, are expressed at much lower rates than the un-spliced M1 and NS transcripts (87). Other alternate transcripts have been reported for the PB1 segment (PB1-F2, N40) and PA segment (PA-X) although these do not occur in every IAV strain (82, 87). The positive-sense viral mRNA, with the appropriate 5'-caps and polyadenylation, are then transported to the cytosol for protein production.

Viral proteins are produced using ribosomes and go through the usual protein production pathways in both the cytosol, for soluble proteins, and the ER and Golgi for membrane-bound proteins. RdRp and NP are imported into the nucleus for further viral mRNA and genome production. Surface glycoproteins like HA, NA, M1, and M2 are trafficked through the ER/Golgi pathway to the cell surface to begin assembly of virions in preparation for budding. Other proteins, like NS1 and PA-X, suppress immune activation through multiple different activities, reviewed in (88).

Once enough NP protein is produced and transported to the nucleus, the RdRp is able to switch to production of full length negative-sense RNA genomes (82, 83, 85). NP helps to stabilize the RdRp interaction with viral negative-sense RNA, allowing the RdRp to produce a full positive-sense copy of the genome (85). This positive-sense template is then use to produce copies of the complete negative-sense RNA of each gene segment. These new negative-sense

RNA genomes are each bound by NP and RdRp components (PA, PB1, and PB2) to create the vRNP complex (82). The vRNPs are then exported from the nucleus and trafficked to the cell surface where NA, HA, M1, and M2 are ready to assemble the virion. Packaging of the eight vRNPs, one for each genome segment, occurs through interactions between NP and M1. It is not fully understood how the assembling virus is able to assure that one of each necessary segment is packaged, although incomplete packaging does occur (89). Once the genome is fully packaged, M1 triggers the pinching in of the cell surface membrane, which leads to a newly budded IAV virion being released into the extracellular space (82).

Released viruses can adhere back to the surface of cells or to other virions through the action of the receptor-binding glycoprotein. However, orthmyxoviruses also have a receptor-destroying enzyme (neuraminidase [NA] in IAV and IBV, the esterase domain of HE in ICV and IDV) that prevents the formation of viral aggregates and releases the budding virion completely so that it may go on to infect more cells (90). However, before the virus can infect other cells, its fusion glycoprotein (HA or HE) must be primed through proteolytic cleavage. For some viruses, this cleavage occurs during protein production in the ER/Golgi, but priming can also occur using extra-cellular proteases (91, 92). Differences in protease preferences and timing of glycoprotein priming have been tied to virulence in IAV between low-pathogenic and high-pathogenic avian strains (91).

1.2b Comparison of influenza A, B, C, and D viruses

While IAV and IBV are well-known human pathogens that are an important focus of public health initiatives and research (10), ICV and IDV are much less well studied. ICV and IDV are more similar to each other sharing similar genomic structures of having seven gene segments encoding nine known proteins (PB2, PB1, PA, HE, NP, NS, and M), in contrast to the

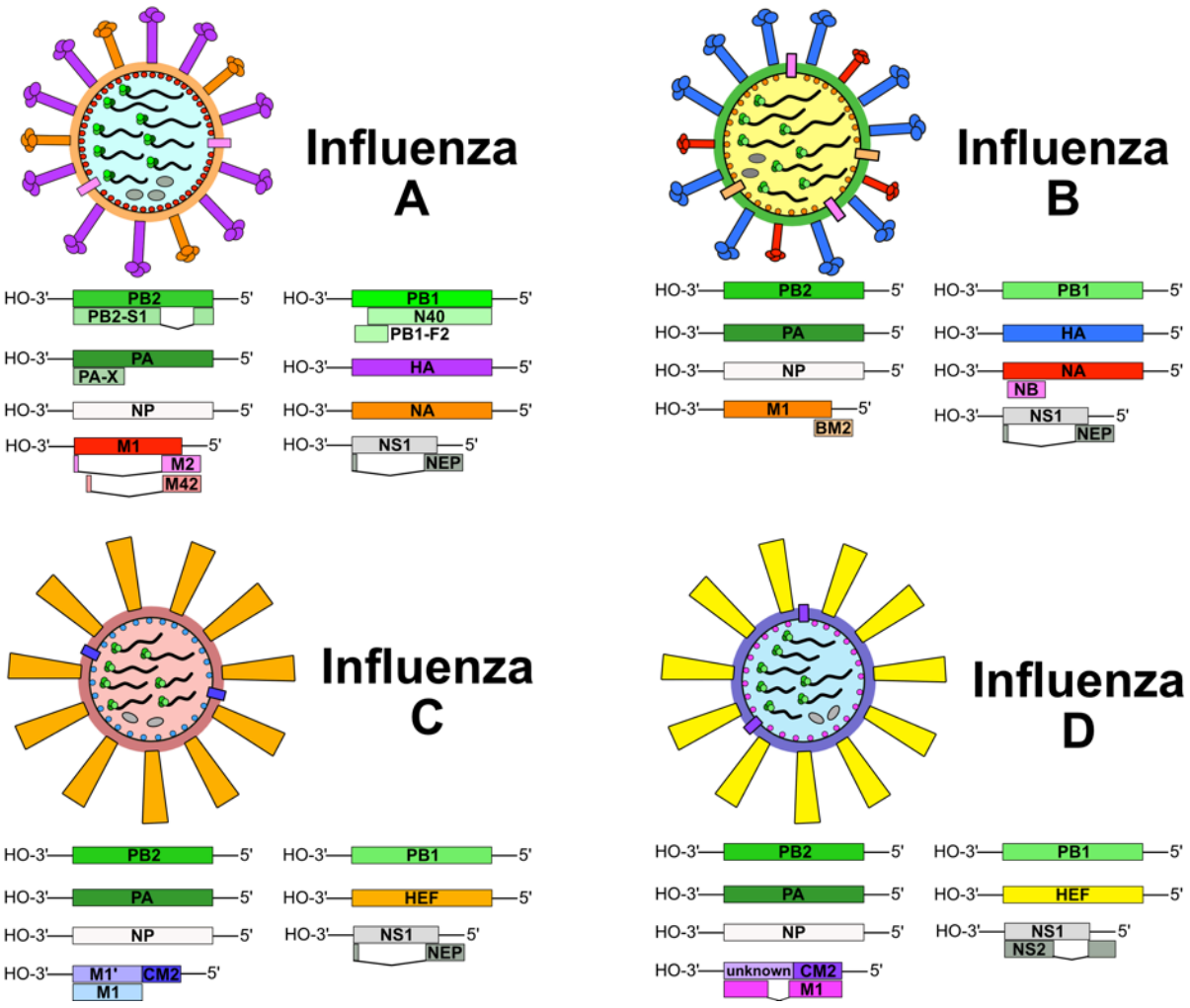


Figure 1.3

The genome structures of the four orthomyxoviruses. IAV and IBV each have eight gene segments while ICV and IDV have seven segments. However, despite sharing many of the same gene segments and core proteins, each orthomyxovirus varies in intron locations and splicing for some genes, such as on the M and NS gene segments (82, 93).

eight gene segments and ten proteins of IAV and IBV (PB2, PB1, PA, HA, NA, NP, NS, and M) (**Fig. 1.3**). One large difference is the surface glycoproteins expressed between these two groups. IAV and IBV express two glycoproteins, hemagglutinin (HA) which acts as the Sia binding and fusion protein, and neuraminidase (NA), which destroys the receptor by cleaving Sia from oligosaccharides. ICV and IDV have one surface glycoprotein called hemagglutinin-esterase (HE) that combines the functions of receptor binding, fusion, and receptor destroying. This difference in surface glycoprotein structures also underlies differences in receptor preferences between the two groups. IAV and IBV HA preferentially bind unmodified Neu5Ac forms of Sia with tissue and host specificity determined by the linkage of Sia to the glycan chain (15, 94). As a result, the NA receptor-destroying function also targets these specific Sia receptor types (95). ICV and IDV target 9-*O*-acetyl forms of Sia with the receptor-binding domain of the HE protein while the esterase domain functions similar to NA, by destroying the 9-O-Ac receptor by removing the acetyl modification (80, 93).

In addition to these differences in genome structure and glycoprotein functions, the different influenza viruses also have different host tropism. ICV and IBV are both primarily human pathogens, although outbreaks of ICV and IBV in pigs and IBV in seals have been reported (96–98). While IBV, similar to IAV, causes common seasonal epidemics in humans with some times severe disease, ICV primarily causes mild upper respiratory infections in people. IDV, on the other hand, is the only influenza genera that hasn't been found to infect humans and instead has been isolated from pigs and cows only (59, 93). However, due to the ability of the other influenza viruses to infect humans, and its recent discovery, IDV is a topic of research and surveillance to determine its risk of zoonotic emergence (99). While IBV, ICV, and

IDV have fairly narrow mammalian host ranges, IAV is unique in having extremely broad host tropism across mammalian and avian species that warrants a more detailed discussion.

1.3 Duck, duck, human: influenza A tropism and evolution

The natural reservoir for IAV is in waterfowl, such as geese and ducks, where it causes gastrointestinal infection and is spread via the fecal-oral route (15, 16). However, IAV has adapted to a range of different avian species, including songbirds and domestic poultry, and mammalian species, including humans, pigs, horses, dogs, and seals (with the occasional spillover into other mammalian species such as cats) (**Fig. 1.4**) (100, 101). IAV subtypes are determined by the sequence of their HA and NA genome segments, which are denoted in the variant name by the H number and N number (example, H1N1 or H3N2). There are currently 18 known HA variants and 11 NA variants, which could theoretically provide up to 198 different IAV subtypes, although not all combinations of HA and NA have been reported in nature (100). To give further complexity, each genome segment of IAV behaves independently during genome packaging such that in a cell co-infected with two or more IAV viruses, trading of genome segments or reassortment can occur amongst any of the eight segments (89). Thus, even between two H1N1 viruses for example, the other six segments can vary widely in their lineage. The result is a near infinite number of possible segment combinations, resulting in viruses with vastly different host tropisms and pathogenesis.

1.3a Determinants of influenza A host tropism

One determinate of IAV host tropism is the linkage type of Sia to the underlying glycan chain. As mentioned previously, Sia can vary in their linkage including α2,3-, α2,6-, and α2,8-linkage to galactose. Both α2,3- and α2,6-linkages are found in most tissues in vertebrate species, while α2-8-linked Sia is primarily found associated with neurons and the central nervous system

(7). However, the ratio and distribution of α2,3- to α2,6-linked Sia can vary between species, and it is this variability that underlies the tropism differences in IAV (**Fig. 1.4**). For example, the HA of avian strains of IAV binds to α2,3-linked Sia which is also the dominant Sia linkage found in the GI tract of birds where the virus replicates. Conversely, human strains of IAV have an HA that prefers α2,6-linked Sia, as there is a larger proportion of this linkage type in the upper-respiratory tract which is the primary virus replication site in humans (15, 16, 100). However, humans do have α2,3-linked Sia in the lower portions of the respiratory tract, including the lungs, allowing for potential receptors for avian strains of IAV (16). It has been proposed that human infections of avian IAV strains can occur if the virus is inhaled deeply enough in the respiratory tract to bind to these α2,3-linked Sia receptors (102). However, these avian IAV strains have not yet shown an ability to spread efficiently person to person, possibly due to a combination of α2,6-linked Sia utilization in the upper-respiratory tract and other factors such as differences in virion stability and the presences of other host factors (103).

Pigs are thought to be a "mixing vessel" as they express both α2,3- and α2,6-linked Sia throughout their respiratory tract, allowing for infection by both human, avian, and swine IAV strains (16, 100). Co-infection of these different strains allows for reassortment of gene segments between these different viruses and emergence of novel IAV strains (104). This is likely how the 2009 H1N1 pandemic virus arose, as it resulted from a triple re-assortment virus of gene segments from human, swine, and avian IAV strains (12). Unlike birds and humans, pigs have an intact *CMAH* gene and produce Neu5Gc in their respiratory tract and mucus (60, 105). Some previous research has shown that Neu5Gc in pig mucus can inhibit IAV NA cleavage (106). However, the effect of Neu5Gc on virus adaptation hasn't been well characterized. Other IAV host species include dogs and horses. Previous research has shown that IAV strains from both

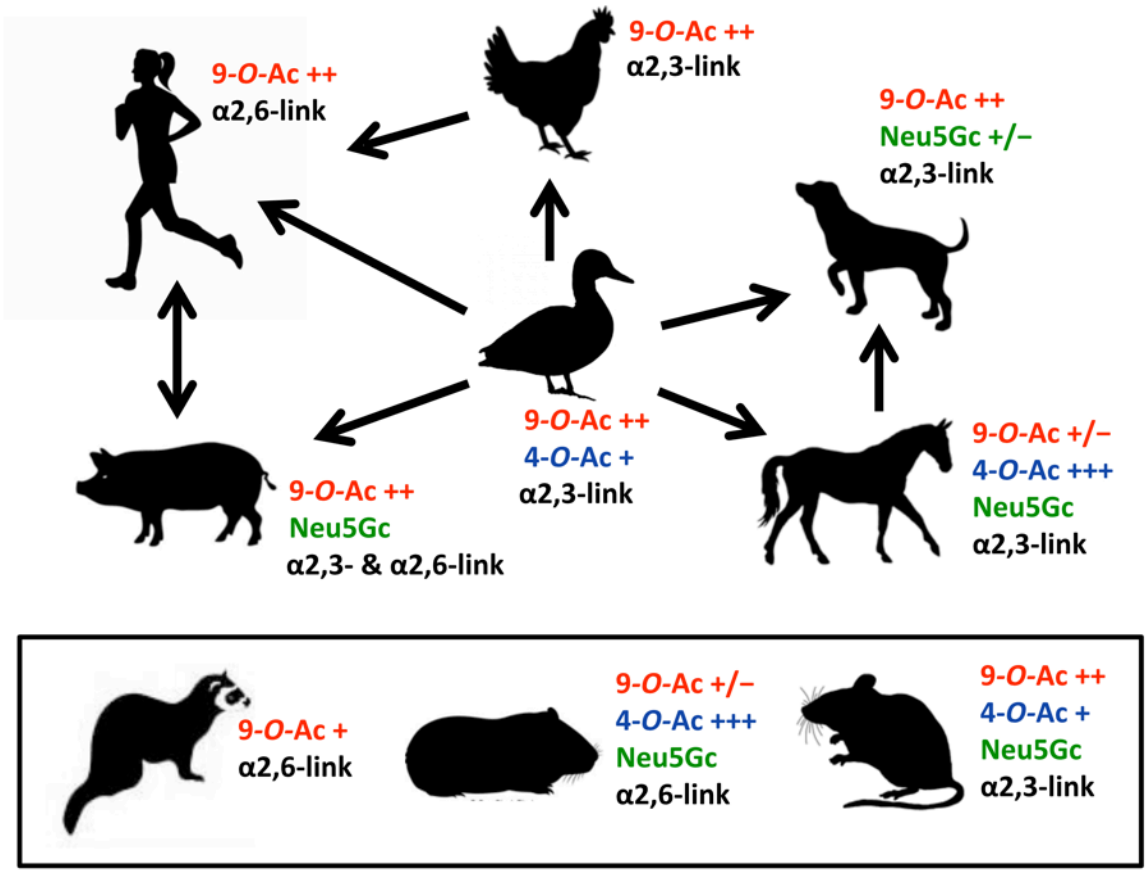


Figure 1.4

Influenza A virus has a broad host range with each host species having different Sia environments. Natural IAV hosts are at the top with arrows denoting jumping of virus strains from one host into another. Species used in IAV research are in the box at the bottom.

dogs and horses prefer to bind α2,3-linked Sia, similar to avian IAV strains (107). This could explain why both avian H3N2 and equine H3N8 were able to emerge in dogs during the last 15 years (108).

While receptor binding is an important host determinate, it is only the first hurdle a virus must overcome when jumping into a new host. The pH stability of HA represents an additional barrier to infection, as it determines the timing of HA protein unfolding and fusion with the endosomal membrane (84, 109). If HA is too unstable at low pH, this can lead to premature unfolding of the HA protein which neutralizes the virus before it can properly fuse. Conversely, if HA is too stable then it can lead to degradation of the virion as the endosome matures into a lysosome. Different species generally have different pH ranges for ideal timing of HA unfolding for successful fusion, with human strains having an ideal range of 5.0 to 5.6, while swine and avian strains tend to have higher pH activation values (84, 110). pH stability can also have important impacts on transmissibility and environmental stability of virus, as mutations in the HA stem have been shown to effect these characteristics in animal models (111, 112).

Another well-studied factor in host specificity, particularly between avian and mammalian IAV strains, is virus polymerase activity. One example is the interaction between PB2 and the host-factor importin-α, which is necessary for efficient import of the viral vRNP complex into the cell nucleus (113). Avian-adapted viruses utilize importin-α3 when infecting mammalian cells, while mammalian-adapted viruses utilize importin-α7 (103, 113, 114). Mutations in PB2 that allow for avian viruses to switch to using importin-α7 leads to more efficient infection and transmission, indicating that interaction with host-specific factors during replication can act as a species barrier. Similarly, interactions between the viral RdRP complex of PB1, PB2, and PA are also effected by differences in the host factor ANP32A, which varies

between mammalian and avian hosts (115, 116). Other mutations in avian IAV PB2, PA, and NP have been proposed to improve interactions with mammalian proteins, cellular RNA binding, and efficient replications at mammalian respiratory tract temperatures showing that species barriers are usually not singular but are instead a combination of complex interactions between viral proteins, cell receptors, and the host cellular environment (103).

1.3b Influenza A HA and NA balance

While it may seem counter productive to carry an enzyme that removes the necessary receptor for viral infection, NA is highly important for efficient infection by allowing for penetration of mucus and prevention of viral aggregate formation during virion budding (117). The efficiency of IAV infection relies, however, on HA and NA functions being in balance. This need for balance can be seen during experimental conditions, such as when IAV is passed in the presence of an NA inhibitor, such as Zanamivir and Oseltamivir, where without the use of NA, virus replication is greatly reduced (118). Interestingly, viruses that gain resistance to NA inhibitors tend to favor compensatory mutations in HA to decrease Sia binding affinity over the more obvious tactic of selecting for mutations in NA to prevent binding of the drug (119). Similar mutations to decrease HA binding affinity are also seen in recombinant viruses without NA or with NA mutants that lack a stalk domain (95, 120). On the opposite end of this spectrum, mutations in HA in recombinant virus that increase HA binding affinity lead to an increased dependence on NA cleavage (120). Thus, HA binding and NA are kept in balance with each other in these experimental systems.

Natural evolution of IAV also shows the need for HA/NA balance, in particular during host switching events. The best documented is the matching of HA binding specificity with NA cleavage between avian and mammalian viruses. As birds are the natural host of IAV, the

cleavage preference for NA in avian strains is against α2,3-linked Sia which matches the HA binding preference (90). For example, when the avian H2N2 IAV emerged in humans in 1957, first the HA adapted to binding α2,6-linked Sia, but its NA retained an α2,3-linked Sia cleavage preference (120, 121). Over time, and following a reassortment event during which the HA segment was swapped for an H3, the NA segment gained efficient α2,6-linked cleavage, while also retaining the ability to cleave α2,3 linked Sia. A similar case for the acquisition of α2,6-linked NA cleavage function to match HA binding was also seen for human-adapted H1N1 viruses (120).

As detailed previously, mammalian species express different combinations of Sia including linkage types and chemical modifications. IAV hosts like humans, birds, and dogs lack a functional *CMAH* gene and express predominantly Neu5Ac Sia forms while pigs and horses retain CMAH function and can have high levels of Neu5Gc in their respiratory tissues (60, 122). Other chemically modified forms are also differentially expressed across IAV host species including 7,9-*O*- and 9-*O*-Ac which are present at variable levels in respiratory tissues between species, and 4-*O*-Ac which is highly expressed in horse tissues (25, 31, 32). Previous research has found that these chemically modified forms can be inhibitory for HA binding and NA cleavage, although this remains an understudied aspect of IAV biology (32, 54, 55, 123, 124). It is not know if these Sia modifications can act as a barrier for virus moving between species or if they exert any selective pressure on emerging viruses.

1.4 Thesis Goals

The driving question of this dissertation is to determine how modified Sia might affect IAV infection efficiency through interactions with HA and NA. If both *O*-acetyl and *N*-glycolyl modifications to Sia are inhibitory to IAV HA and NA, it follows that viruses would be selected

to either avoid these modifications interacting with HA or NA, or adapt to utilizing them as potential receptors.

Chapter 2 looks at the regulation of 7,9-*O*- and 9-*O*-acetyl modified Sia in cell culture using CRISPR-Cas9 manipulation of CasD1 and SIAE expression. The effects of cell-surface expressed 7,9-*O*- and 9-*O*-Ac on IAV, IBV, ICV, and IDV infection were determined.

Chapter 3 examines the display of *O*-acetyl and Neu5Gc Sia forms across mouse mucosal tissues, as well as in secreted mucins and erythrocytes of IAV host species. The effects of modified Sia on HA binding and NA cleavage efficiency are also determined using functional assays.

Chapter 4 utilizes new advances in Illumina MiSeq deep sequencing to examine whole virus population changes of IAV over time during passage on different MDCK cell lines. MDCK cells have been the gold standard cells for IAV growth *in vitro*, including acting as a tractable model for expressing different Sia forms including linkages and Neu5Gc synthesis. However, MDCK cells are also innately heterogeneous and are comprised of two different cell types that vary in their permissiveness to IAV infection. In this chapter, selection of minority variants in three IAV strains serial passed on MDCK-wild type cells, MDCK-SiaT1, MDCK-CMAH, MDCK-Type I, and MDCK-Type II cells were determined.

Chapter 5 summaries the findings of chapter 2-4 and discusses how this research fits into the broader IAV field, as well as future directions for the work described here.

REFERENCES

1. Dube DH, Bertozzi CR. 2005. Glycans in cancer and inflammation–potential for therapeutics and diagnostics. Nature reviews Drug discovery 4:477–488.

2. Bhide GP, Colley KJ. 2017. Sialylation of N-glycans: mechanism, cellular compartmentalization and function. Histochem Cell Biol 147:149–174.

3. Brusés JL, Rutishauser U. 2001. Roles, regulation, and mechanism of polysialic acid function during neural development. Biochimie 83:635–643.

4. Lehmann F, Tiralongo E, Tiralongo J. 2006. Sialic acid-specific lectins: Occurrence, specificity and function. Cellular and Molecular Life Sciences 63:1331–1354.

5. Colley KJ, Varki A, Kinoshita T. 2015. Cellular Organization of Glycosylation, p. . *In* Varki, A, Cummings, RD, Esko, JD, Stanley, P, Hart, GW, Aebi, M, Darvill, AG, Kinoshita, T, Packer, NH, Prestegard, JH, Schnaar, RL, Seeberger, PH (eds.), Essentials of Glycobiology, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY).

6. Rini J, Esko J, Varki A. 2009. Glycosyltransferases and Glycan-processing Enzymes, p. 1–7. *In* Varki, A, Cummings, RD, Esko, J (eds.), Essentials of Glycoboilogy. 2nd edition., 2nd ed. Cold Spring Harbor Laboratory Press, Cold Springs Harbor, NY.

7. Varki A, Schauer R. 2009. Chapter 14 Sialic AcidsEssentials of Glycobiology - second edition, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Springs Harbor, NY.

8. Varki NM, Varki A. 2007. Diversity in cell surface sialic acid presentations: implications for biology and disease. Laboratory investigation; a journal of technical methods and pathology 87:851–857.

9. Wasik BR, Barnard KN, Parrish CR. 2016. Effects of Sialic Acid Modifications on Virus Binding and Infection. Trends in Microbiology xx:1–11.

10. CDC. 2019. Estimated Influenza Illnesses, Medical visits, Hospitalizations, and Deaths in the United States — 2017–2018 influenza season.

11. World Health Organization. 2014. WHO Influenza (Seasonal) Media centre 211:2–5.

12. York I, Donis RO. 2013. The 2009 Pandemic Influena Virus: Where Did It Come from, Where Is It Now, and Where Is It Going? Current topics in microbiology and immunology 370:241–257.

13. Gambaryan AS, Matrosovich MN. 2015. What adaptive changes in hemagglutinin and neuraminidase are necessary for emergence of pandemic influenza virus from its avian precursor? Biochemistry Mosc 80:872–880.

14. Wilks S, de Graaf M, Smith DJ, Burke DF. 2012. A review of influenza haemagglutinin receptor binding as it relates to pandemic properties. Vaccine 30:4369–4376.

15. Kumlin U, Olofsson S, Dimock K, Arnberg N. 2008. Sialic acid tissue distribution and influenza virus tropism. Influenza Other Respir Viruses 2:147–154.

16. de Graaf M, Fouchier R a M. 2014. Role of receptor binding specificity in influenza A virus transmission and pathogenesis. The EMBO journal 33:823–41.

17. Mandal C, Schwartz-Albiez R, Vlasak R. 2012. Functions and Biosynthesis of O-acetylated sialic acids. Top Curr Chem.

18. Varki A. 1992. Diversity in the sialic acids. Glycobiology 2:25–40.

19. Van Breedam W, Pöhlmann S, Favoreel HW, de Groot RJ, Nauwynck HJ. 2014. Bitter-sweet symphony: Glycan-lectin interactions in virus biology. FEMS Microbiology Reviews 38:598–632.

20. Schauer R, Schmid H, Pommerencke J, Iwersen M, Kohla G. 2001. Metabolism and Role of O-Acetylated Sialic Acids, p. 325–342. *In* The Molecular Immunology of Complex Carbohydrates-2. Spring US.

21. Baumann A-MTM, Bakkers MJ, Buettner FF, Hartmann M, Grove M, Langereis MA, de Groot RJ, Muhlenhoff M. 2015. 9-O-Acetylation of sialic acids is catalysed by CASD1 via a covalent acetyl-enzyme intermediate. Nature communications 6:7673.

22. Arming S, Wipfler D, Mayr J, Merling A, Vilas U, Schauer R, Schwartz-Albiez R, Vlasak R. 2011. The human Cas1 protein: A sialic acid-specific O-acetyltransferase? Glycobiology 21:553–564.

23. Kamerling JP, Schauer R, Shukla AK, Stoll S, Halbeek H, Vliegenthart J. 1987. Migration of O-acetyl groups in N,O-acetylneuraminic acids. European Journal of Biochemistry 162:601–607.

24. Vandamme-Feldhaus V r., Schauer R. 1998. Characterization of the Enzymatic 7- O-Acetylation of Sialic Acids and Evidence for Enzymatic O-Acetyl Migration from C-7 to C-9 in Bovine Submandibular Gland. Journal of Biochemistry 124:111–121.

25. Wasik BR, Barnard KN, Ossiboff RJ, Khedri Z, Feng KH, Perez DR, Varki A, Parrish CR. 2017. Distribution of O-acetylated sialic acids among the tissues of influenza hosts. mSphere 2:e00379–16.

26. Langereis MA, Bakkers MJ, Deng L, Vered P-K, Vervoort SJ, Hulswit RJ, van Vliet AL, Gerwig GJ, de Poot SA, Boot W, van Ederen AM, Heesters BA, van der Loos CM, van Kuppeveld FJ, Yu H, Huizinga EG, Chen X, Varki A, Kamerling JP, de Groot RJ. 2015. Complexity and Diversity of the Mammalian Sialome Revealed by Nidovirus Virolectins. Cell reports 11:1966–1978.

27. Ravasio V, Damiati E, Zizioli D, Orizio F, Giacopuzzi E, Manzoni M, Bresciani R, Borsani G, Monti E. 2017. Genomic and biochemical characterization of sialic acid acetylesterase (siae) in zebrafish. Glycobiology 1–9.

28. Orizio F, Damiati E, Giacopuzzi E, Benaglia G, Pianta S, Schauer R, Schwartz-Albiez R, Borsani G, Bresciani R, Monti E. 2015. Human sialic acid acetyl esterase: Towards a better understanding of a puzzling enzyme. Glycobiology 25:992–1006.

29. Takematsu H, Diaz S, Stoddart A, Zhang Y, Varki A. 1999. Lysosomal and cytosolic sialic acid 9-O-acetylesterase activities can be encoded by one gene via differential usage of a signal peptide-encoding exon at the N terminus. Journal of Biological Chemistry 274:25623–25631.

30. Zhu H, Chan HC, Zhou Z, Li J, Zhu H, Yin L, Xu M, Cheng L, Sha J. 2004. A gene encoding sialic-acid-specific 9-O-acetylesterase found in human adult testis. Journal of Biomedicine and Biotechnology 2004:130–136.

31. Aamelfot M, Dale OB, Weli SC, Koppang EO, Falk K. 2014. The in situ distribution of glycoprotein-bound {4-O-Acetylated} sialic acids in vertebrates. Glycoconjugate journal 31:327–335.

32. Hanaoka K, Pritchett TJ, Takasaki S, Kochibe N, Sabesan S, Paulson JC, Kobata A. 1989. 4-O-Acetyl-N-acetylneuraminic acid in the N-linked carbohydrate structures of equine and guinea pig a2-macroglobulins, potent inhibitors of influenza virus infection. Journal of Biological Chemistry 264:9842–9849.

33. Iwersen M, Dora H, Kohla G, Gasa S, Schauer R. 2003. Solubilisation and properties of the {sialate-4-O-acetyltransferase} from guinea pig liver. Biological chemistry 384:1035–1047.

34. Iwersen M, Vandamme-feldhaus V, Schauer R, Institut B, Kiel D-. 1998. Enzymatic 4-O-acetylation of N-acetylneuraminic acid in guinea-pig liver. Glycoconjugate Journal 904:895–904.

35. Schauer R. 2009. Sialic acids as regulators of molecular and cellular interactions. Curr Opin Struct Biol 19:507–514.

36. Schauer R, Srinivasan GV, Wipfler D, Kniep B, Schwartz-Albiez R. 2011. O-Acetylated Sialic Acids and Their Role in Immune Defense, p. 525–548. *In* The Molecular Immunology of Complex Carbohydrates-3. Springer Science + Business Media, LLC.

37. Khatua B, Roy S, Mandal C. 2013. Sialic acids siglec interaction: a unique strategy to circumvent innate immune response by pathogens. Indian J Med Res 138:648–662.

38. Bochner BS, Zimmermann N. 2015. Role of siglecs and related glycan-binding proteins in immune responses and immunoregulation. J Allergy Clin Immunol 135:598–608.

39. Cariappa a., Takematsu H, Liu H, Diaz S, Haider K, Boboila C, Kalloo G, Connole M, Shi HN, Varki N, Varki A, Pillai S. 2009. B cell antigen receptor signal strength and peripheral B cell development are regulated by a 9-O-acetyl sialic acid esterase. Journal of Experimental Medicine 206:125–138.

40. Wipfler D, Srinivasan GV, Sadick H, Kniep B, Arming S, Willhauck-Fleckenstein M, Vlasak R, Schauer R, Schwartz-Albiez R. 2011. Differentially regulated expression of 9-O-acetyl GD3 (CD60b) and 7-O-acetyl-GD3 (CD60c) during differentiation and maturation of human T and B lymphocytes. Glycobiology 21:1161–1172.

41. Pillai S, Cariappa A, Pirnie SP. 2009. Esterases and autoimmunity: the sialic acid acetylesterase pathway and the regulation of peripheral B cell tolerance. Trends in Immunology 30:488–493.

42. Mahajan VS, Pillai S. 2016. Sialic acids and autoimmune disease. Immunol Rev 269:145–161.

43. Ma F, Wu D, Deng L, Secrest P, Zhao J, Varki N, Lindheim S, Gagneux P. 2012. Sialidases on mammalian sperm mediate deciduous sialylation during capacitation. Journal of Biological Chemistry 287:38073–38079.

44. Cavdarli S, Dewald JH, Yamakawa N, Guérardel Y, Terme M, Le Doussal J-M, Delannoy P, Groux-Degroote S. 2019. Identification of 9-O-acetyl-N-acetylneuraminic acid (Neu5,9Ac2) as main O-acetylated sialic acid species of GD2 in breast cancer cells. Glycoconjugate Journal.

45. Shen Y, Kohla G, Lrhorfi AL, Sipos B, Kalthoff H, Gerwig GJ, Kamerling JP, Schauer R, Tiralongo J. 2004. O-acetylation and {de-O-acetylation} of sialic acids in human colorectal carcinoma. European journal of biochemistry / {FEBS} 271:281–290.

46. Mandal C, Srinivasan GV, Chowdhury S, Chandra S, Mandal C, Schauer R, Mandal C. 2009. High level of sialate-O-acetyltransferase activity in lymphoblasts of childhood acute lymphoblastic leukaemia (ALL): Enzyme characterization and correlation with disease status. Glycoconjugate Journal 26:57–73.

47. Lewis AL, Lewis WG. 2012. Host sialoglycans and bacterial sialidases: a mucosal perspective. Cellular Microbiology 14:1174–1182.

48. Lewis AL, Hensler ME, Varki A, Nizet V. 2006. The Group B Streptococcal Sialic Acid *O* -Acetyltransferase Is Encoded by *neuD* , a Conserved Component of Bacterial Sialic Acid Biosynthetic Gene Clusters. Journal of Biological Chemistry 281:11186–11192.

49. Corfield AP, Veh RW, Wember M, Michalski JC, Schauer R. 1981. The release of *N* -acetyl- and *N* -glycolloyl-neuraminic acid from soluble complex carbohydrates and erythrocytes by bacterial, viral and mammalian sialidases. Biochemical Journal 197:293–299.

50. Kleineidam RG, Furuhata K, Ogura H, Schauer R. 1990. 4-Methylumbelliferyl-alpha-glycosides of partially O-acetylated N-acetylneuraminic acids as substrates of bacterial and viral sialidases. Biol Chem Hoppe Seyler 371:715–719.

51. Li W, Xiao A, Li Y, Yu H, Chen X. 2017. Chemoenzymatic synthesis of Neu5Ac9NAc-containing α2–3- and α2–6-linked sialosides and their use for sialidase substrate specificity studies. Carbohydrate Research 451.

52. Kahya HF, Andrew PW, Yesilkaya H. 2017. Deacetylation of sialic acid by esterases potentiates pneumococcal neuraminidase activity for mucin utilization, colonization and virulence. PLoS Pathogens 13:1–21.

53. Robinson LS, Lewis WG, Lewis AL. 2017. The sialate *O* -acetylesterase EstA from gut *Bacteroidetes* species enables sialidase-mediated cross-species foraging of 9- *O* -acetylated sialoglycans. Journal of Biological Chemistry 292:11861–11872.

54. Muñoz-Barroso I, García-Sastre a, Villar E, Manuguerra JC, Hannoun C, Cabezas J a. 1992. Increased influenza A virus sialidase activity with N-acetyl-9-O-acetylneuraminic acid-containing substrates resulting from influenza C virus O-acetylesterase action. Virus research 25:145–153.

55. Higa H, Rogers G, Paulson C. 1985. Influenza virus hemagglutinins differentiate between receptor determinants bearing N-acetyl-, N-glycollyl-, and N,O-diacetylneuraminic acids. Virology 144:279–282.

56. Gentsch JR, Pacitti AF. 1987. Differential interaction of reovirus type 3 with sialylated receptor components on animal cells. Virology 161:245–248.

57. Langereis M a., Zeng Q, Heesters B, Huizinga EG, de Groot RJ. 2012. The murine coronavirus hemagglutinin-esterase receptor-binding site: A major shift in ligand specificity through modest changes in architecture. PLoS Pathogens 8:1–8.

58. Langereis M a, Zeng Q, Gerwig GJ, Frey B, von Itzstein M, Kamerling JP, de Groot RJ, Huizinga EG. 2009. Structural basis for ligand and substrate recognition by torovirus hemagglutinin esterases. Proceedings of the National Academy of Sciences of the United States of America 106:15897–15902.

59. Hause BM, Collin E a., Liu R, Huang B, Sheng Z, Lu W, Wang D, Nelson E a., Li F. 2014. Characterization of a novel influenza virus in cattle and swine: Proposal for a new genus in the Orthomyxoviridae family. mBio 5:1–10.

60. Peri S, Kulkarni A, Feyertag F, Berninsone PM, Alvarez-Ponce D. 2018. Phylogenetic Distribution of CMP-Neu5Ac Hydroxylase (CMAH), the Enzyme Synthetizing the Proinflammatory Human Xenoantigen Neu5Gc. Genome Biol Evol 10:207–219.

61. Hedlund M, Tangvoranuntakul P, Takematsu H, Long JM, Housley GD, Kozutsumi Y, Suzuki A, Wynshaw-Boris A, Ryan AF, Gallo RL, Varki N, Varki A. 2007. N-Glycolylneuraminic Acid Deficiency in Mice: Implications for Human Biology and Evolution. Molecular and Cellular Biology 27:4340–4346.

62. Hashimoto Y, Yamakawa T, Tanabe Y. 1984. Further studies on the red cell glycolipids of various breeds of dogs. A possible assumption about the origin of Japanese dogs. Journal of biochemistry 96:1777–1782.

63. Martin MJ, Rayner JC, Gagneux P, Barnwell JW, Varki A. 2005. Evolution of human-chimpanzee differences in malaria susceptibility: relationship to human genetic loss of N-glycolylneuraminic acid. Proceedings of the National Academy of Sciences of the United States of America 102:12819–12824.

64. Proto WR, Siegel SV, Dankwa S, Liu W, Kemp A, Marsden S, Zenonos ZA, Unwin S, Sharp PM, Wright GJ, Hahn BH, Duraisingh MT, Rayner JC. 2019. Adaptation of Plasmodium falciparum to humans involved the loss of an ape-specific erythrocyte invasion ligand. Nat Commun 10:4512.

65. Minami A, Ishibashi S, Ikeda K, Ishitsubo E, Hori T, Tokiwa H, Taguchi R, Ieno D, Otsubo T, Matsuda Y, Sai S, Inada M, Suzuki T. 2013. Catalytic preference of Salmonella typhimurium LT2 sialidase for N-acetylneuraminic acid residues over N-glycolylneuraminic acid residues. FEBS Open Bio 3:231–236.

66. Deng L, Song J, Gao X, Wang J, Yu H, Chen X, Varki N, Naito-Matsui Y, Galán JE, Varki A. 2014. Host adaptation of a bacterial toxin from the human pathogen Salmonella Typhi. Cell 159:1290–1299.

67. Yang Y-A, Chong A, Song J. 2018. Why Is Eradicating Typhoid Fever So Challenging: Implications for Vaccine and Therapeutic Design. Vaccines 6:45.

68. Zanin M, Baviskar P, Webster R, Webby R. 2016. The Interaction between Respiratory Pathogens and Mucus. Cell Host & Microbe 19:159–168.

69. Gabriel MO, Grünheid T, Zentner A. 2005. Glycosylation pattern and cell attachment-inhibiting property of human salivary mucins. J Periodontol 76:1175–1181.

70. Slomiany BL, Murty VL, Piotrowski J, Slomiany A. 1996. Salivary mucins in oral mucosal defense. Gen Pharmacol 27:761–771.

71. Knowles MR, Boucher RC. 2002. Mucus clearance as a primary innate defense mechanism for mammalian airways. Journal of Clinical Investigation 109:571–577.

72. Joo NS, Evans IAT, Cho H-J, Park I-H, Engelhardt JF, Wine JJ. 2015. Proteomic Analysis of Pure Human Airway Gland Mucus Reveals a Large Component of Protective Proteins. PLOS ONE 10:e0116756.

73. Casado B, Pannell LK, Iadarola P, Baraniuk JN. 2005. Identification of human nasal mucous proteins using proteomics. Proteomics 5:2949–2959.

74. Deplancke B, Gaskins HR. 2001. Microbial modulation of innate defense: goblet cells and the intestinal mucus layer. The American journal of clinical nutrition 73:1131S–1141S.

75. Sicard J-F, Le Bihan G, Vogeleer P, Jacques M, Harel J. 2017. Interactions of Intestinal Bacteria with Components of the Intestinal Mucus. Frontiers in Cellular and Infection Microbiology 7.

76. Altgärde N, Eriksson C, Peerboom N, Phan-Xuan T, Moeller S, Schnabelrauch M, Svedhem S, Trybala E, Bergström T, Bally M. 2015. Mucin-like Region of Herpes Simplex Virus Type 1 Attachment Protein Glycoprotein C (gC) Modulates the Virus-Glycosaminoglycan Interaction. Journal of Biological Chemistry 290:21473–21485.

77. Cohen M, Zhang X-Q, Senaati HP, Chen H-W, Varki NM, Schooley RT, Gagneux P. 2013. Influenza A penetrates host mucus by cleaving sialic acids with neuraminidase. Virology journal 10:321.

78. Huang X, Dong W, Milewska A, Golda A, Qi Y, Zhu QK, Marasco W a, Baric RS, Sims AC, Pyrc K, Li W, Sui J. 2015. Human Coronavirus HKU1 Spike Protein Uses O-Acetylated Sialic Acid as an Attachment Receptor Determinant and Employs Hemagglutinin-Esterase Protein as a Receptor-Destroying Enzyme. Journal of virology 89:7202.

79. De Groot RJ. 2006. Structure, function and evolution of the hemagglutinin-esterase proteins of corona- and toroviruses. Glycoconjugate Journal 23:59–72.

80. Wang M, Veit M. 2016. Hemagglutinin-esterase-fusion (HEF) protein of influenza C virus. Protein & Cell 7:28–45.

81. Lakadamyali M, Rust MJ, Zhuang X. 2004. Endocytosis of influenza viruses. Microbes Infect 6:929–936.

82. Dou D, Revol R, Östbye H, Wang H, Daniels R. 2018. Influenza A Virus Cell Entry, Replication, Virion Assembly and Movement. Frontiers in Immunology 9.

83. Samji T. 2009. Influenza A: understanding the viral life cycle. Yale J Biol Med 82:153–159.

84. Di Lella S, Herrmann A, Mair CM. 2016. Modulation of the pH Stability of Influenza Virus Hemagglutinin: A Host Cell Adaptation Strategy. Biophysical Journal 110:2293–2301.

85. Neumann G, Brownlee GG, Fodor E, Kawaoka Y. 2004. Orthomyxovirus replication, transcription, and polyadenylation. Curr Top Microbiol Immunol 283:121–143.

86. De Vlugt C, Sikora D, Pelchat M. 2018. Insight into Influenza: A Virus Cap-Snatching. Viruses 10.

87. Dubois J, Terrier O, Rosa-Calatrava M. 2014. Influenza Viruses and mRNA Splicing: Doing More with Less. mBio 5.

88. Nogales A, Martinez-Sobrido L, Topham DJ, DeDiego ML. 2018. Modulation of Innate Immune Responses by the Influenza A NS1 and PA-X Proteins. Viruses 10.

89. Lakdawala SS, Fodor E, Subbarao K. 2016. Moving On Out: Transport and Packaging of Influenza Viral RNA into Virions. Annu Rev Virol 3:411–427.

90. Byrd-Leotis L, Cummings RD, Steinhauer DA. 2017. The Interplay between the Host Receptor and Influenza Virus Hemagglutinin and Neuraminidase. International Journal of Molecular Sciences 18:1541.

91. Böttcher-Friebertshäuser E, Garten W, Matrosovich M, Klenk HD. 2014. The hemagglutinin: a determinant of pathogenicity. Curr Top Microbiol Immunol 385:3–34.

92. Bertram S, Glowacka I, Steffen I, Kühl A, Pöhlmann S. 2010. Novel insights into proteolytic cleavage of influenza virus hemagglutinin. Rev Med Virol 20:298–310.

93. Su S, Fu X, Li G, Kerlin F, Veit M. 2017. Novel Influenza D virus: Epidemiology, pathology, evolution and biological characteristics. Virulence 8:1580–1591.

94. Velkov T. 2013. The specificity of the influenza B virus hemagglutinin receptor binding pocket: What does it bind to? Journal of Molecular Recognition 26:439–449.

95. Gaymard A, Le Briand N, Frobert E, Lina B, Escuret V. 2016. Functional balance between neuraminidase and haemagglutinin in influenza viruses. Clinical Microbiology and Infection 22:975–983.

96. Bodewes R, Morick D, de Mutsert G, Osinga N, Bestebroer T, van der Vliet S, Smits SL, Kuiken T, Rimmelzwaan GF, Fouchier RAM, Osterhaus ADME. 2013. Recurring influenza B virus infections in seals. Emerging Infect Dis 19:511–512.

97. Ran Z, Shen H, Lang Y, Kolb EA, Turan N, Zhu L, Ma J, Bawa B, Liu Q, Liu H, Quast M, Sexton G, Krammer F, Hause BM, Christopher-Hennings J, Nelson EA, Richt J, Li F, Ma W. 2015. Domestic pigs are susceptible to infection with influenza B viruses. J Virol 89:4818–4826.

98. Salem E, Cook EAJ, Lbacha HA, Oliva J, Awoume F, Aplogan GL, Hymann EC, Muloi D, Deem SL, Alali S, Zouagui Z, Fèvre EM, Meyer G, Ducatez MF. 2017. Serologic Evidence for Influenza C and D Virus among Ruminants and Camelids, Africa, 1991-2015. Emerging Infect Dis 23:1556–1559.

99. Song H, Qi J, Khedri Z, Diaz S, Yu H, Chen X, Varki A, Shi Y, Gao GF. 2016. An Open Receptor-Binding Cavity of Hemagglutinin-Esterase-Fusion Glycoprotein from Newly-Identified Influenza D Virus: Basis for Its Broad Cell Tropism. PLOS Pathogens 12:e1005411.

100. Yoon S-W, Webby RJ, Webster RG. 2014. Evolution and ecology of influenza A viruses. Curr Top Microbiol Immunol 385:359–375.

101. Parrish CR, Murcia PR, Holmes EC. 2015. Influenza virus reservoirs and intermediate hosts: dogs, horses, and new possibilities for influenza virus exposure of humans. Journal of virology 89:2990–4.

102. Shinya K, Ebina M, Yamada S, Ono M, Kasai N, Kawaoka Y. 2006. Avian flu: influenza virus receptors in the human airway. Nature 440:435–436.

103. Herfst S, Imai M, Kawaoka Y, Fouchier R a. M. 2014. Avian influenza virus transmission to mammals. Curr Top Microbiol Immunol 385:137–155.

104. Kahn RE, Ma W, Richt JA. 2014. Swine and influenza: a challenge to one health research. Curr Top Microbiol Immunol 385:205–218.

105. Wang R-G, Ruan M, Zhang R-J, Chen L, Li X-X, Fang B, Li C, Ren X-Y, Liu J-Y, Xiong Q, Zhang L-N, Jin Y, Li L, Li R, Wang Y, Yang H-Y, Dai Y-F. 2018. Antigenicity of tissues and organs from GGTA1/CMAH/β4GalNT2 triple gene knockout pigs. J Biomed Res.

106. Xu G, Suzuki T, Maejima Y, Mizoguchi T, Tsuchiya M, Kiso M, Hasegawa A, Suzuki Y. 1995. Sialidase of swine influenza A viruses: variation of the recognition specificities for sialyl linkages and for the molecular species of sialic acid with the year of isolation. Glycoconjugate journal 12:156–161.

107. Feng KH, Gonzalez G, Deng L, Yu H, Tse VL, Huang L, Huang K, Wasik BR, Zhou B, Wentworth DE, Holmes EC, Chen X, Varki A, Murcia PR, Parrish CR. 2015. Equine and canine influenza H3N8 viruses show minimal biological differences despite phylogenetic divergence. Journal of Virology 89:JVI.00521–15.

108. Parrish CR, Voorhees IEH. 2019. H3N8 and H3N2 Canine Influenza Viruses: Understanding These New Viruses in Dogs. Vet Clin North Am Small Anim Pract 49:643–649.

109. Hamilton BS, Whittaker GR, Daniel S. 2012. Influenza virus-mediated membrane fusion: determinants of hemagglutinin fusogenic activity and experimental approaches for assessing virus fusion. Viruses 4:1144–1168.

110. Galloway SE, Reed ML, Russell CJ, Steinhauer DA. 2013. Influenza HA subtypes demonstrate divergent phenotypes for cleavage activation and pH of fusion: implications for host range and adaptation. PLoS Pathog 9:e1003151.

111. Russell CJ, Hu M, Okda FA. 2018. Influenza Hemagglutinin Protein Stability, Activation, and Pandemic Risk. Trends in Microbiology 26:841–853.

112. Shelton H, Roberts KL, Molesti E, Temperton N, Barclay WS. 2013. Mutations in haemagglutinin that affect receptor binding and pH stability increase replication of a PR8 influenza virus with H5 HA in the upper respiratory tract of ferrets and may contribute to transmissibility. J Gen Virol 94:1220–1229.

113. Soh YS, Moncla LH, Eguia R, Bedford T, Bloom JD. 2019. Comprehensive mapping of adaptation of the avian influenza polymerase protein PB2 to humans. eLife 8.

114. Manz B, Schwemmle M, Brunotte L. 2013. Adaptation of Avian Influenza A Virus Polymerase in Mammals To Overcome the Host Species Barrier. Journal of Virology 87:7200–7209.

115. Baker SF, Ledwith MP, Mehle A. 2018. Differential Splicing of ANP32A in Birds Alters Its Ability to Stimulate RNA Synthesis by Restricted Influenza Polymerase. Cell Reports 24:2581-2588.e4.

116. Long JS, Giotis ES, Moncorgé O, Frise R, Mistry B, James J, Morisson M, Iqbal M, Vignal A, Skinner MA, Barclay WS. 2016. Species difference in ANP32A underlies influenza A virus polymerase host restriction. Nature 529:101–104.

117. Jagadesh A, Salam AAA, Mudgal PP, Arunkumar G. 2016. Influenza virus neuraminidase (NA): a target for antivirals and vaccines. Arch Virol 161:2087–2094.

118. Blick TJ, Tiong T, Sahasrabudhe a, Varghese JN, Colman PM, Hart GJ, Bethell RC, McKimm-Breschkin JL. 1995. Generation and characterization of an influenza virus neuraminidase variant with decreased sensitivity to the neuraminidase-specific inhibitor 4-guanidino-Neu5Ac2en. Virology 214:475–484.

119. Gimsa U, Grötzinger I, Gimsa J. 1996. Two evolutionary strategies of influenza viruses to escape host non-specific inhibitors: alteration of hemagglutinin or neuraminidase specificity. Virus Res 42:127–135.

120. Wagner R, Matrosovich M, Klenk HD. 2002. Functional balance between haemagglutinin and neuraminidase in influenza virus infections. Reviews in Medical Virology 12:159–166.

121. Kosik I, Yewdell JW. 2019. Influenza Hemagglutinin and Neuraminidase: Yin¯Yang Proteins Coevolving to Thwart Immunity. Viruses 11.

122. Scocco P, Pedini V. 2008. Localization of influenza virus sialoreceptors in equine respiratory tract. Histology and Histopathology 23:973–978.

123. Broszeit F, Tzarum N, Zhu X, Nemanichvili N, Eggink D, Leenders T, Li Z, Liu L, Wolfert MA, Papanikolaou A, Martínez-Romero C, Gagarinov IA, Yu W, García-Sastre A, Wennekes T, Okamatsu M, Verheije MH, Wilson IA, Boons G-J, de Vries RP. 2019. N-Glycolylneuraminic Acid as a Receptor for Influenza A Viruses. Cell Reports 27:3284-3294.e6.

124. Rogers GN, Pritchett TJ, Lane JL, Paulson JC. 1983. Differential sensitivity of human, avian, and equine influenza A viruses to a glycoprotein inhibitor of infection: selection of receptor specific variants. Virology 131:394–408.

CHAPTER TWO

Expression of 9-*O*- and 7,9-*O*-acetyl modified sialic acid in cells and their effects on influenza viruses

Karen N. Barnard, Brian R. Wasik, Justin R. LaClair, David W. Buchholz, Wendy S. Weichert, Brynn K. Alford-Lawrence, Hector C. Aguilar, Colin R. Parrish. 2019. Expression of 9-*O*- and 7,9-*O*-acetyl modified sialic acid in cells and their effects on influenza viruses. *mBio,* 10(6): e02490-19.

2.1 ABSTRACT

Sialic acids (Sia) are widely displayed on the surfaces of cells and tissues. Sia come in a variety of chemically modified forms, including those with acetyl modifications at the C7, C8, and C9 positions. Here, we analyzed the distribution and amounts of these acetyl modifications in different human and canine cells. As Sia or their variant forms are receptors for influenza A, B, C, and D viruses, we examined the effects of these modifications on virus infections. We confirmed that 9-*O*-acetyl and 7,9-*O*-acetyl modified Sia are widely but variably expressed across cell lines from both humans and canines. While they were expressed on the cell surface of canine MDCK cell lines, they were located primarily within the Golgi compartment of human HEK-293 and A549 cells. The *O*-acetyl modified Sia were expressed at low levels of 1-2% of total Sia in these cell lines. We knocked out and over-expressed the sialate *O*-acetyltransferase gene (CasD1), and knocked out the sialate *O*-acetylesterase gene (SIAE) using CRISPR/Cas9 editing. Knocking out CasD1 removed 7,9-*O*- and 9-*O*-acetyl Sia expression, confirming previous reports. However, over-expression of CasD1 and knockout of SIAE gave only modest increases in 9-*O*-acetyl levels in cells and no change in 7,9-*O*-acetyl levels, indicating that there are complex regulations of these modifications. These modifications were essential for influenza C and D infection, but had no obvious effect on influenza A and B infection.

2.2 IMPORTANCE

Sialic acids are key glycans that are involved in many different normal cellular functions, as well as being receptors for many pathogens. However, Sia come in diverse chemically modified forms. Here we examined and manipulated the expression of 7,9-*O*- and 9-*O*-acetyl modified Sia on cells commonly used in influenza virus and other research by engineering the enzymes that produce or remove the acetyl groups.

2.3 INTRODUCTION

Sialic acids (Sia) are a family of nine-carbon monosaccharides expressed mainly in vertebrates that serve as terminal residues of carbohydrate chains on cell membrane glycoproteins and glycolipids, as well as on secreted glycoproteins at all mucosal surfaces (**Fig 2.1A**) (1, 2). Sia are key mediators of many cell and tissue functions, where they are bound by cellular receptors such as selectins and siglecs (3, 4). Their ubiquitous presence on cells, tissues, and mucosal surfaces also make Sia a key point of contact for both commensal microbes and invading pathogens, including viruses, bacteria, and parasites (3).

Sia are highly diverse as there are more than 50 different chemically distinct variants formed from the basic structure of *N*-acetylneuraminic acid (Neu5Ac) by the addition of chemical groups at various positions on the pyranose ring or the glycerol side chain. These modifications include acetyl, sulfo, methyl, and lactyl groups, among others, which may be present individually or in many combinations (1). As many different enzymes and pathways introduce these modifications, complex mixtures of Sia forms may be present, with significant variation in both the levels and specific combinations of each modification (1, 2, 5).

Common chemical additions include *O*-acetyl modifications to the C-4, 7, 8, and/or 9 positions, resulting in a variety of combinations including Neu4,5Ac, Neu5,9Ac_2, Neu5,7,9Ac_3 Sia. The addition of *O*-acetyl modifications to the C7 and/or C9 positions is mediated by the sialate *O*-acetyltransferase enzyme, Cas1 domain containing 1 (CasD1) (**Fig. 2.1B**). CasD1 appears to add acetyl groups to the C7 position of Sia, from which it may migrate to the C8 and C9 position under physiological conditions, allowing the possible addition of another acetyl group to C7 (6, 7). A migrase enzyme has been proposed to aid in the transfer of the acetyl group from C7 to C9, however a specific enzyme has yet to be identified (5, 8, 9). CasD1 is localized in

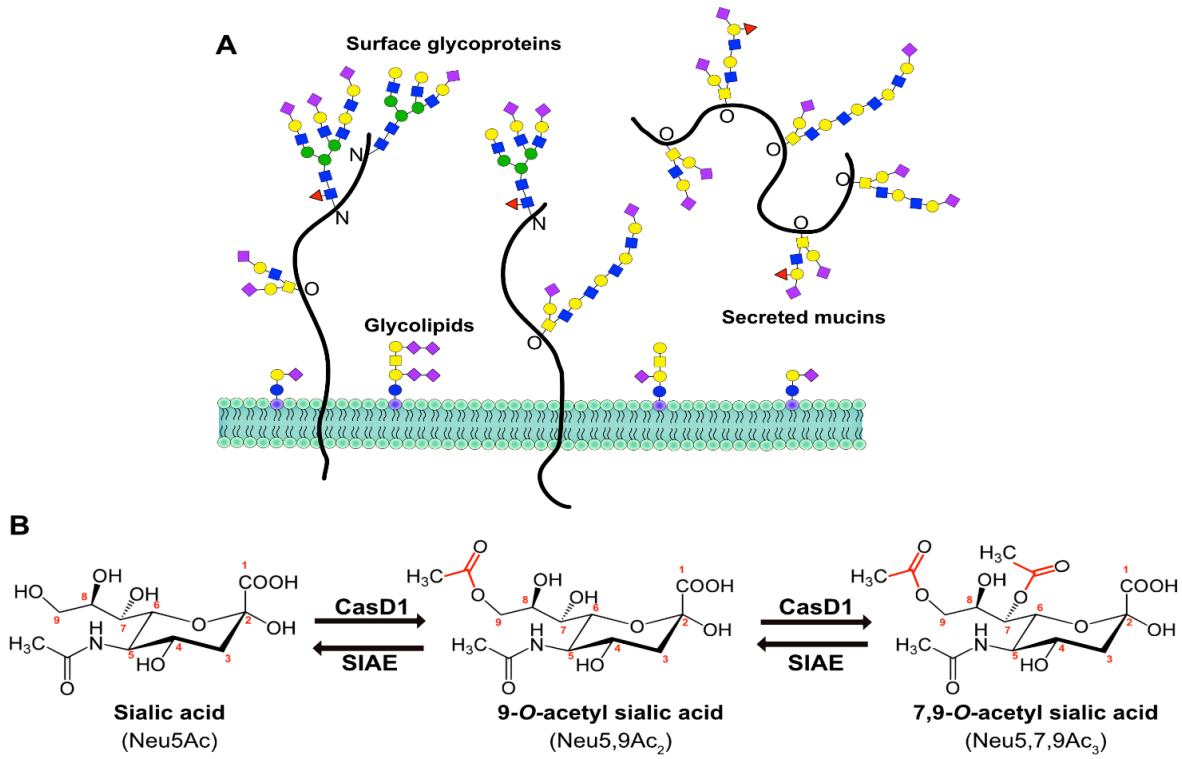


Figure 2.1

A) Sialic acids (purple diamonds) terminate glycan chains on glycolipids and glycoproteins as part of the glycocalyx on the surface of cells. They can also terminate glycans on secreted glycoproteins, like mucins, that make up the protective mucosal barrier in gastrointestinal and respiratory tissue. **B**) The sialate *O*-acetyltransferase, CasD1, adds acetyl groups to sialic acid (N-acetylneuraminic acid, Neu5Ac) at the C-7 from which it migrates to the C-9 position ($Neu5,9Ac_2$) under physiological conditions. This can allow for an additional acetyl group to be added by CasD1 to C-7 ($Neu5,7,9Ac_3$). The sialate *O*-acetylesterase, SIAE, can remove these acetyl modifications, restoring the unmodified Neu5Ac form of sialic acid.

the late Golgi, so acetyl modifications are added during the later stages of protein glycosylation. The regulatory processes that control the number of acetyl groups added or their positions have not been well defined, although clear differences in expression of 7,9-*O*-Ac and 9-*O*-Ac have been reported in mouse and human cells and tissues, in chicken embryos, and in the tissues of some other animals (10, 11). CasD1 uses acetyl-CoA and likely has a preference for CMP-Neu5Ac as a substrate, so that it is less active on CMP-Neu5Gc (7).

At least one sialate *O*-acetylesterase (SIAE) enzyme regulates the display of 7,9-*O*- and 9-*O*-Ac in many cells and tissues (**Fig. 2.1B**). The SIAE gene encodes two isoforms that vary in the presence of a proposed C-terminal localization tag in the lysosomal (Lse) form that is absent from the cytosolic form (Cse) (12). However, the processes that regulate the expression and activity of these two isoforms remain poorly defined. Lse has been found to localize to the Golgi and/or ER and on the surface of cells when over-expressed, with the majority being secreted into the supernatant (12). Antibody staining shows that Cse is found diffusely throughout the cytosol, where it is thought to remove 9-*O*- and 7-*O*-acetyl groups to recycle the Sia for reuse in glycosylation (12). Despite the reports of distinct protein expression in mouse tissues, bioinformatic analysis of RNA expression in human cells and tissues show mRNA corresponding to the Lse form of SIAE and none responding to the Cse form (13). It is therefore still unclear how these two isoforms are regulated in humans or other animals, whether their expression relates to CasD1, or what controls the levels and locations of 9-*O*- and 7,9-*O*-Ac expression.

Sia *O*-acetylation and de-acetylation play important roles in many different biological processes, particularly in development, cancer, and immunology. For example, *O*-acetyl modifications to Sia may alter the binding of host lectins, including the Sialic acid-binding

immunoglobulin-type lectins (Siglecs) (3, 5, 14). Siglecs are regulators of many different cell functions and developmental processes – examples include B- and T-cells where the presence of 9-*O*-acetyl Sia modulates immune cell activation and differentiation (15). In B-cells, negative regulation of B-cell receptor (BCR) activation by Siglec CD22 is mediated by binding to Neu5Ac-terminated glycan chains on the BCR, which can be blocked by *O*-acetyl modifications (16, 17). The presence of 9-*O*-Ac can also reduce the activity of sialidases, including human neuraminidases (18). Incorrect regulation of 9-*O*-Ac, 7,9-*O*-Ac, and SIAE activity has been linked to autoimmune disorders through the development of auto-antibodies (17, 19). 9-*O*-Ac and 7,9-*O*-Ac and their regulation by SIAE also appear to play important roles during early stages of embryonic development, spermatogenesis, and in different forms of cancer including acute lymphoblastic leukemia, colon cancer, and breast cancer (20–23).

Effects of different Sia modifications have also been suggested for the binding of pathogens or on the activities of their sialidases (neuraminidases). However, in general these are still not well documented, with the exception of those that use the modified forms as receptors. Influenza A (IAV), influenza B (IBV), influenza C (ICV) and influenza D (IDV) viruses use Sia as their primary receptors for host recognition and cell entry, but with different effects of Sia modification. IAV and IBV interact with Sia through two surface glycoproteins, hemagglutinin (HA) and neuraminidase (NA). For both IAV and IBV, HA binds Sia to initiate the endocytic uptake of the virus by the cell, leading to fusion between the viral envelope and the endosomal membrane at low pH (24, 25). NA is a sialidase which cleaves Sia off glycan chains when it is present in mucus or on the surface of cells, allowing the virus to penetrate to the epithelial cells, and also preventing aggregation of newly produced virus after budding (26, 27). In contrast to IAV and IBV, ICV and IDV have one surface glycoprotein, the hemagglutinin-esterase fusion

protein (HEF), which serves similar purposes to both HA and NA. HEF binds specifically to 9-*O*-acetyl Sia to initiate uptake of the virus into cells, while the esterase domain removes 9-*O*-acetyl modifications, releasing the virus from mucus, and disassembling virus aggregates after budding (28–31). While the role of *O*-acetyl modified Sia for ICV and IDV infections are well documented, how these modifications affect IAV is not well characterized. In addition, while previous preliminary studies have suggested that the presence of 9-*O*-Ac on cells may be inhibitory for both NA activity and HA binding of IAV (32, 33), the details are unclear.

Here we examine and more closely define the expression and distribution of 9-*O*-acetyl and 7,9-*O*-acetyl Sia on cells in culture, define the effects of SIAE and CasD1 on display of these Sia modifications, and perform an initial examination of the effects of these modified Sia on infection by IAV, IBV, ICV, and IDV. We used CRISPR-Cas9 for gene engineering of the enzymes that add or remove the 9-*O*- and 7,9-*O*-acetyl groups from Sia, and combine these with the recently developed viral protein-derived probes that specifically recognize modified Sia. By merging these with HPLC-based quantification of the different Sia forms, we provide a more detailed understanding of the expression and localization of these modifications in cells, and examine their effects on host-virus interactions as examples.

2.4 RESULTS

Expression of 7,9-*O*- and 9-*O*-acetyl Sia in cells. There is currently only sporadic information about the expression of modified Sia on commonly used cell lines, or an understanding of how that compares to the expression in animal tissues. We examined cell lines that are widely used in many experimental systems: A549 human type II alveolar epithelial cells, HEK-293 human kidney derived cells (possibly embryonic adrenal precursor cells (34)), and MDCK canine kidney epithelial cells. Additionally, we tested MDCK-type I and type II cells that

were previously sub-cloned by others from the ATCC MDCK line (MDCK-NBL2) and which have been extensively characterized (35–37). We used probes derived from porcine torovirus (PToV) and bovine coronavirus (BCoV) hemagglutinin-esterase proteins (HE) that were fused to human IgG1 Fc and which had the esterase active site inactivated (HE-Fc). The PToV HE-Fc probe recognizes 9-*O*-Ac, while BCoV HE-Fc recognizes primarily 7,9-*O*-Ac although with a low affinity for 9-*O*-Ac (10, 11). By immunofluorescence microscopy, the different forms were present at variable levels, with between 10 and 70% of the cells of each type showing staining under standard culture conditions (**Fig. 2.2A,B**). MDCK-NBL2 and MDCK-type I cells showed both strong surface and internal staining for 7,9-*O*- and 9-*O*-Ac forms, while both were mostly found in intracellular locations in A549 and HEK-293 cells, with an occasional cell showing bright surface staining. MDCK-type II cells showed staining only for 9-*O*-Ac and none for 7,9-*O*-Ac, indicating that these modifications are regulated independently. In HEK-293 and A549 cells, both 7,9-*O*- and 9-*O*-Ac appeared to be localized within the Golgi compartment as confirmed by co-staining with the Golgi marker GM130 (**Fig. 2.3**). Similar localization differences between some cell lines have been seen previously using the ICV HEF as a probe (38). In our studies, there was inherent variability within populations in terms of both level of staining and localization. For example, in MDCK-NBL2 not all cells were positive for 9-*O*-Ac, while in MDCK-type I some cells retained more 9-*O*-Ac and 7,9-*O*-Ac internally while others displayed more of the modified Sia on their surface (**Fig. 2.2A**). This heterogeneity was consistent between different passages of each cell line.

The expression levels of those modified Sia variants were also quantified by HPLC analysis using DMB labeling and fluorescence detection (39). The cells were treated at 80˚C for 3h in the presence of 2M acetic acid, and therefore the data likely represents the total Sia present

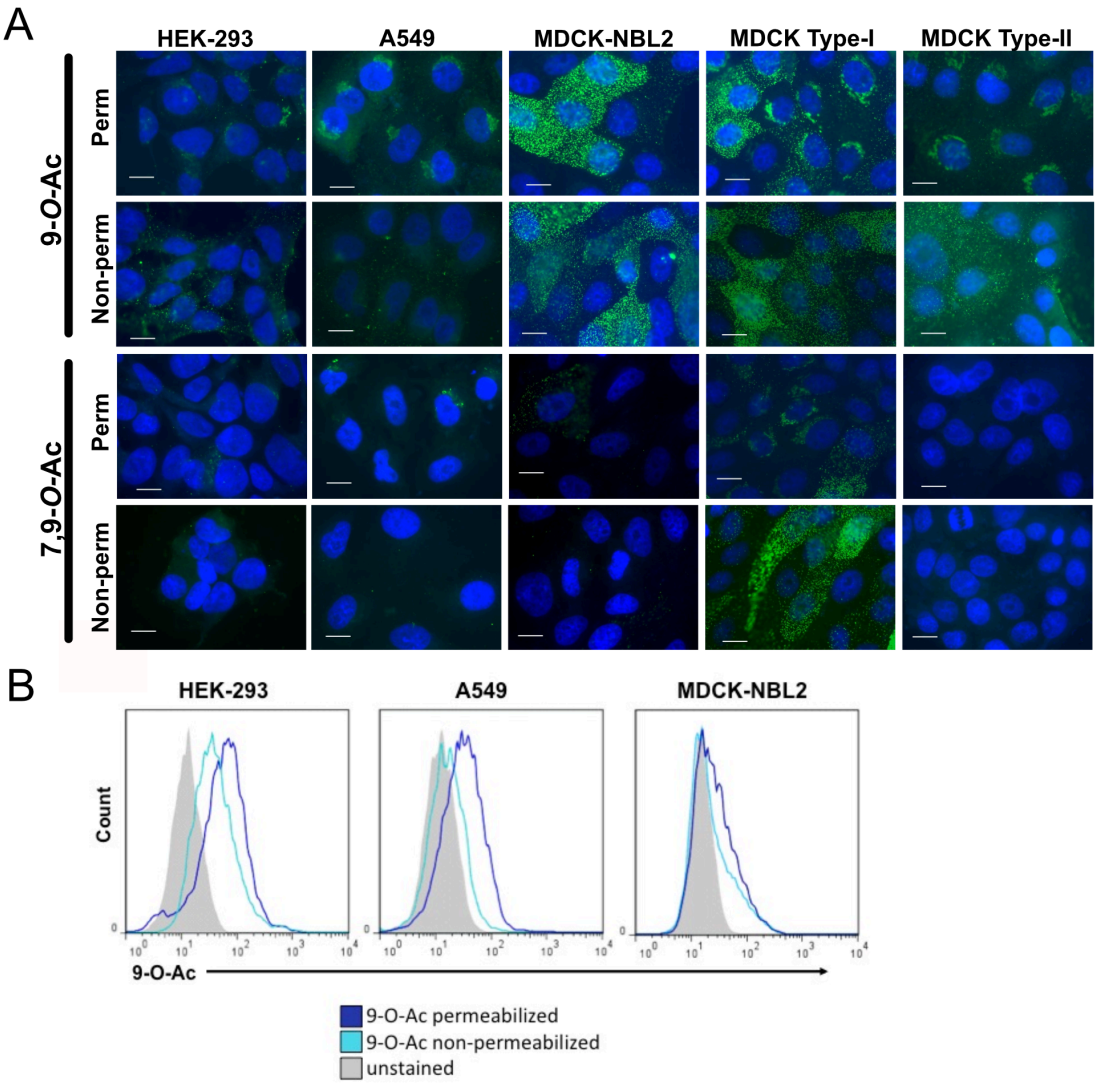


Figure 2.2

Surface and internal expression of 9-*O*-Ac and 7,9-*O*-Ac on different cell lines.
A) Fluorescent staining of human HEK-293, A549, and canine MDCK AATC line (NBL2), MDCK type I, and MDCK type II. Cells were probed with HE-Fcs probes derived from BCoV and PToV, which recognize 9-*O*-Ac and 7,9-*O*-Ac respectively. Cells were permeabilized (perm) using Carbo-Free blocking reagent with 0.001% Tween-20 while non-permeabilized cells (non-perm) received only Carbo-Free block. All cells imaged at 60×, nuclei stained with DAPI. Scale bar = 10 µm. **B**) Representative flow cytometry graphs showing distribution of positive staining for HEK-293, A549 and MDCK-NBL2 cell lines. BCoV and PToV HE-Fcs probes were used and permeabilization (perm) and non-permeabilization (non-perm) methods were as in IFA staining.

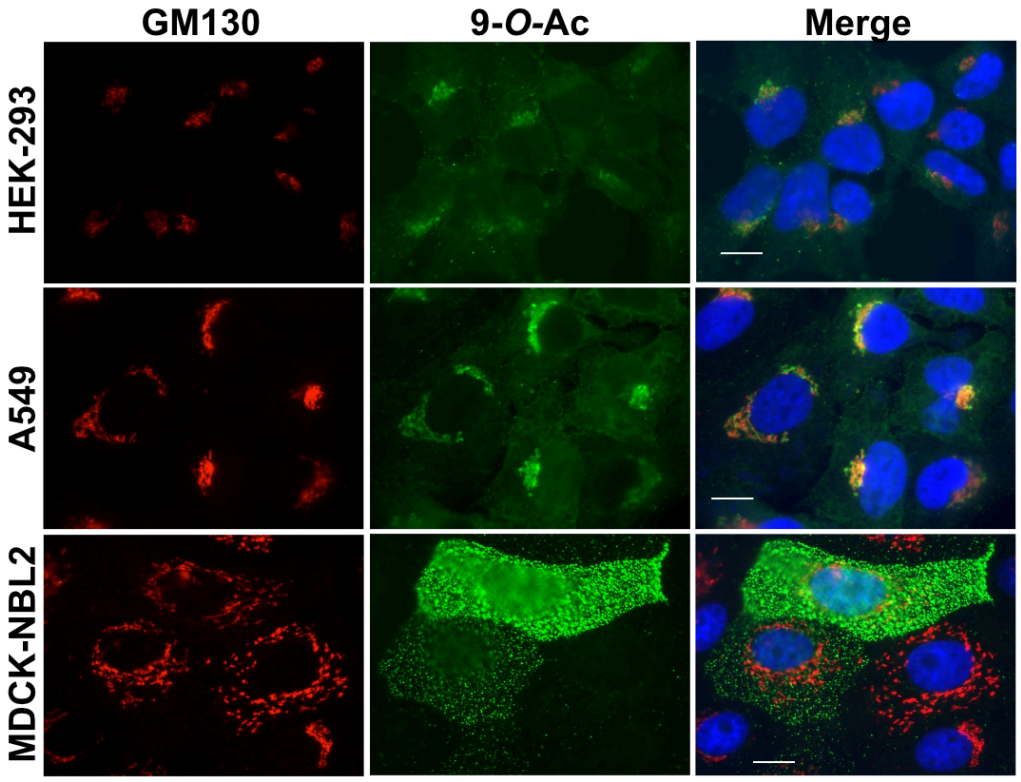


Figure 2.3

Staining of HEK-293, A549, and MDCK-NBL2 cells with PToV HE-Fc for 9-O-Ac, and co-staining for the Golgi marker GM130. Cells were permeabilized using 0.001% Tween-20 and imaged at 60x magnification. Scale bar = 10 μm.

in the cells. Even though some cell lines showed strong staining by the HE-Fc probes, levels of 9-*O*-Ac were 1-2% of the total collected Sia from all cells while levels of 7,9-*O*-Ac were below the level of detection in all cell lines (**Fig. 2.2; Table 2.1; Fig. 2.4**). This includes cells from other species including cat, mouse, horse, and swine. Unique among the cells tested, A549 cells are secretory and express significant levels of mucin (including MUC1 and MUC5B) (40). While mucin proteins are not the only secreted protein found in mucus, we chose to focus on them due to their high levels of glycosylation (80% of weight) and proposed importance as a barrier to pathogens (41–43). We found that A549 cells were able to secrete MUC5B into conditioned media (**Fig. 2.5A**), although we found that the amounts secreted were variable between collections. Analysis of conditioned media from A549 cells showed approximately 2% of Sia was 9-*O*-Ac on secreted proteins with no 7,9-*O*-Ac detected, indicating that at least in these cells, secreted proteins were not enriched for *O*-acetylated Sia (**Fig. 2.5B**). However, it would be worth examining proteins secreted by primary cells from humans or that are present in different tissues, including mucins, to determine whether the secreted proteins from A549 are representative of respiratory mucus.

The low levels of 9-*O*- and 7,9-*O*-Ac detected on HEK-293, A549, and MDCK cells by HPLC could be due to the heterogeneity of the population, as this method measures total Sia for all cells in the population. However, it is likely that even on cells expressing higher levels of 7,9-O- and 9-O-Ac, these modified Sia make up a small proportion of the total Sia present in the cell or glycocalyx. Previous studies have reported that podoplanin (GP40 in canine cells) is a primary carrier of 9-*O*-Ac on MDCK cells and therefore acts as the main receptor for ICV (44, 45). However, when we co-stained MDCK cells with an anti-podoplanin antibody and PToV HE-Fc, we saw little correlation between podoplanin and 9-*O*-Ac staining (**Fig. 2.6**).

Table 2.1

Total sialic acids were collected from different cell lines via mild acid hydrolysis and analyzed using HPLC. Percentages are out of total sialic acid collected. Representative chromatograms for standard, HEK-293, A549, and MDCK-NBL2 are presented in **Figure 2.4**. n/d = not detected. * primary swine nasal (SiNEC) and tracheal (SiTEC) epithelial cells, courtesy of Dr. Stacey Schultz-Cherry.

Cell Type	Species	%Neu5Gc	% Neu5Ac	% 9-*O*-Ac	% 7,9-*O*-Ac
A549	Human	2.21	96.40	1.39	n/d
HEK-293	Human	2.26	96.72	1.02	n/d
MDCK-NBL2	Canine	0.79	97.89	1.32	n/d
MDCK Type I	Canine	0.66	97.67	1.67	n/d
MDCK Type II	Canine	0.76	98.13	1.11	n/d
NLFK	Cat	2.00	96.87	1.13	n/d
EqKc3	Horse	2.93	97.07	n/d	n/d
NBL6	Horse	2.52	96.38	1.1	n/d
L cells	Mouse	3.87	96.13	n/d	n/d
A72	Canine	2.04	95.85	2.11	n/d
SiNEC	Swine*	4.69	93.56	1.75	n/d
SiTEC	Swine*	1.97	96.54	1.49	n/d

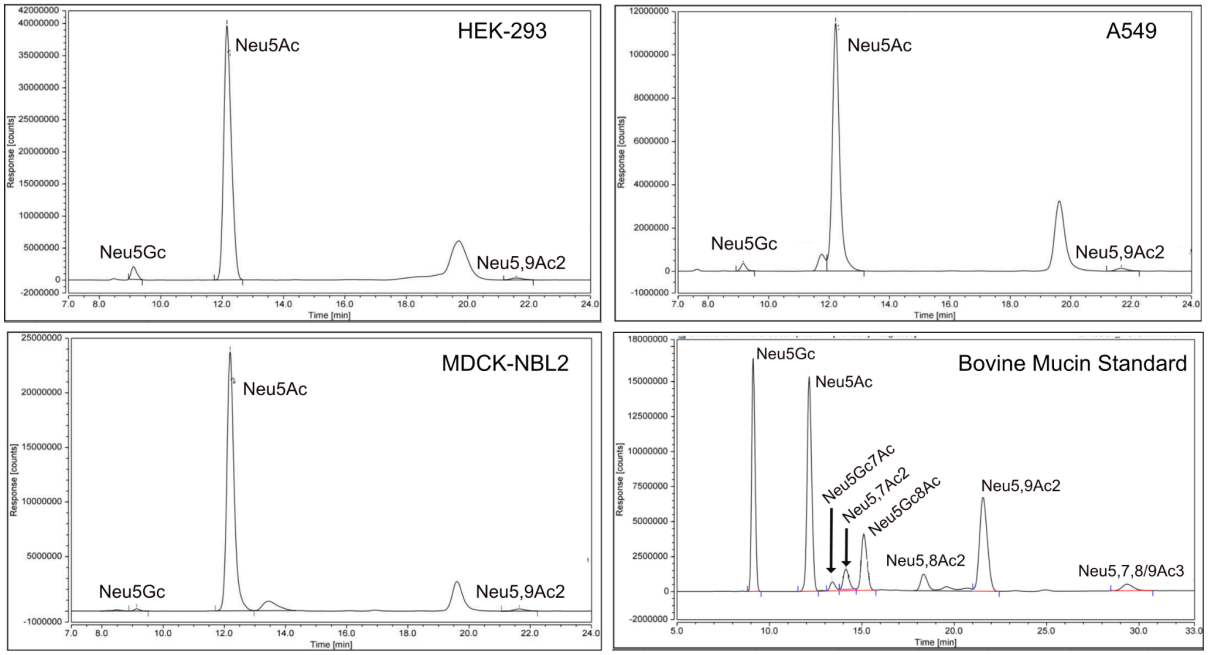


Figure 2.4

HPLC chromatograms showing wild-type HEK-293 cells, A549 cells, MDCK-NLB2 cells, and bovine mucin standard of *O*-acetyl modifications. Total sialic acids were collected from cell lines and standards via mild acid hydrolysis. Neu5Gc on cells is likely derived from the fetal bovine serum used in the growth media, which is taken up by cells and displayed on the cell surface. Humans and canines do not have a functional CMAH gene to synthesize Neu5Gc endogenously.

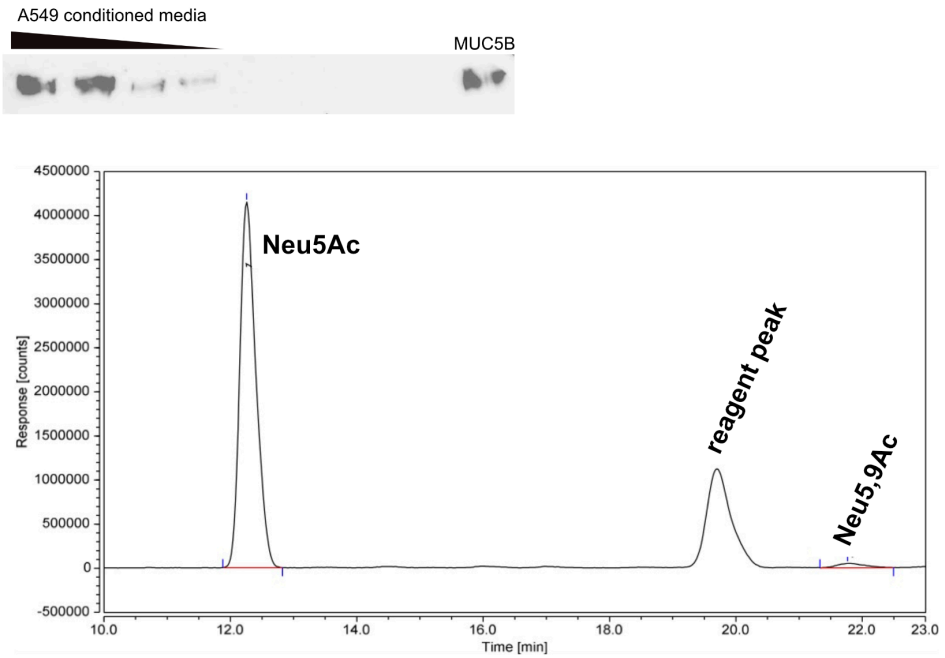


Figure 2.5

A549 conditioned media was analyzed for the presence of secreted mucins and total sialic acid content.
A) Conditioned media from A549 was concentrated using a 30kd filter column and then titrated on a western blot for Muc5B expression compared to a purified human Muc5B (a gift from Stefan Ruhl, University of Buffalo). **B)** A representative chromatogram of total sialic acid collected from A549 conditioned media using HPLC analysis. The percent of different sialic acid forms found in total sialic acid collected is summarized in the table.

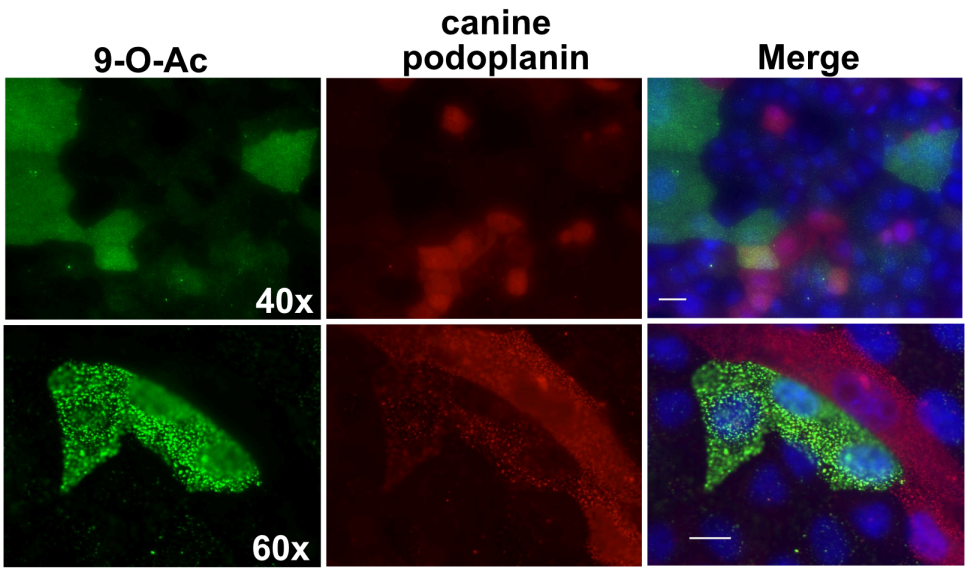


Figure 2.6

MDCK-NBL2 cells were co-stained for the presence of 9-*O*-Ac using PToV HE-Fc (green) and canine podoplanin (red) using an anti-podoplanin antibody (courtesy of Dr. Yukinari Kato, Tohoku University). Cells were imaged at 40× and 60× magnification as indicated. Scale bars = 10 μm.

Production of CasD1 knock out and over-expressing cells.

Knock out and over-expression of CasD1. To better understand the control of expression of 7,9-*O*- and 9-*O*-Ac, glyco-engineered cell lines were created by manipulating the expression of CasD1 and SIAE genes. To do this, we knocked out CasD1 via CRISPR-Cas9 editing, or over-expressed CasD1 via transfection of an expression plasmid. Knock-out variants of CasD1 (ΔCasD1) were prepared from MDCK-NBL2, A549, and HEK-293 cells using CRISPR-Cas9 targeting of early exons in the CasD1 gene (**Fig. 2.7A**). ΔCasD1 clones were confirmed by examining for 7,9-*O*- and 9-*O*-Ac display, and then by PCR and sequencing of the genomic region surrounding the deletion in modified Sia negative clones (**Fig. 2.7A**). For all cell types, ΔCasD1 variants showed loss of both 7,9-*O*- and 9-*O*-Ac display when probed with the different specific HE-Fc probes via flow cytometry and immunofluorescence microscopy (**Fig. 2.7B, C, D**). This agrees with previous findings that loss of CasD1 leads to loss of both 7,9-*O*- and 9-*O*-Ac modifications in haploid HAP1 cells (7). HPLC analysis of total Sia showed a significant decrease in 9-*O*-Ac expression compared to wild-type (WT) cells, however very low levels (<1%) were still detectable, despite a lack of staining with the HE-Fc probes (**Fig. 2.7E**). This could be due to exogenous Sia from the fetal bovine serum in the growth media, as is the case for Neu5Gc which is also detectable at similarly low levels on cells even though the gene for Neu5Gc synthesis, CMAH, is not functional in humans or canines (**Fig 2.4**) (46, 47).

Due to the heterogeneity and low expression of 7,9-*O*- and 9-*O*-Ac in WT cells, we sought to engineer cells with more homogeneous and higher levels of these modifications. We over-expressed the human CasD1 (CasD1-OX) in the ΔCasD1 variants of MDCK-NBL2, A549, and HEK-293 cells, expecting that expression of CasD1 under a strong promoter in the ΔCasD1 background would increase the consistency of the synthesis and display of 9-*O*- and 7,9-*O*-acetyl

Sia relative to WT. CasD1-OX cells showed significantly higher levels of CasD1 mRNA compared to WT cells, indicating that the CasD1 expression plasmid was being transcribed at high levels (**Fig. 2.7E**). However, only a modest increase in average 9-*O*- and 7,9-*O*-Ac expression across the population in HEK-293 cells was seen by flow cytometry, while fluorescent microscopy showed heterogeneity of expression in the population (**Fig. 2.7B, C, D**). The transfected MDCK and A549 CasD1-OX cells showed recovery of 9-*O*- and 7,9-*O*-Ac synthesis when analyzed by flow cytometry and immunofluorescence microscopy, but levels were not as high as seen in WT cells (**Fig. 2.7B, C, D**). HPLC analysis confirmed the expression levels for HEK-293, A549, and MDCK cells (**Fig. 2.7F**). Additionally, heterogeneity was still seen in these populations for 9-*O*-Ac. This suggests that 7,9-*O*- and 9-*O*-Ac expression is not directly regulated by the levels of CasD1 gene expression, but may be affected by post-translational regulation of CasD1 activity or removal of the modifications by SIAE or other enzymes. In addition, these modifications are not regulated the same across individual cells as evidenced by the population heterogeneity. SIAE transcripts were consistently expressed in HEK-293 WT, ΔCasD1, and CasD1-OX cell populations by qPCR analysis, although SIAE mRNA levels in CasD1-OX cells were lower than those seen in WT (**Fig. 2.7E**).

SIAE knock out cells and display of modified Sia. To determine the role of SIAE activity in regulating 7,9-*O*- and 9-*O*-Ac display, SIAE knockout (ΔSIAE) cells were generated from WT HEK-293 and A549 cells by targeting early exons in the SIAE gene (**Fig. 2.8A**). Complete knockout of SIAE was confirmed both by genotyping to show deletions in both alleles and by loss of mRNA by qPCR (**Fig. 2.8A, D**). ΔSIAE cells showed a small but significant increase in surface display and a large increase in internal display of 9-*O*-Ac based on flow cytometry, but did not appear to show any changes in surface or internal display of 7,9-*O*-Ac

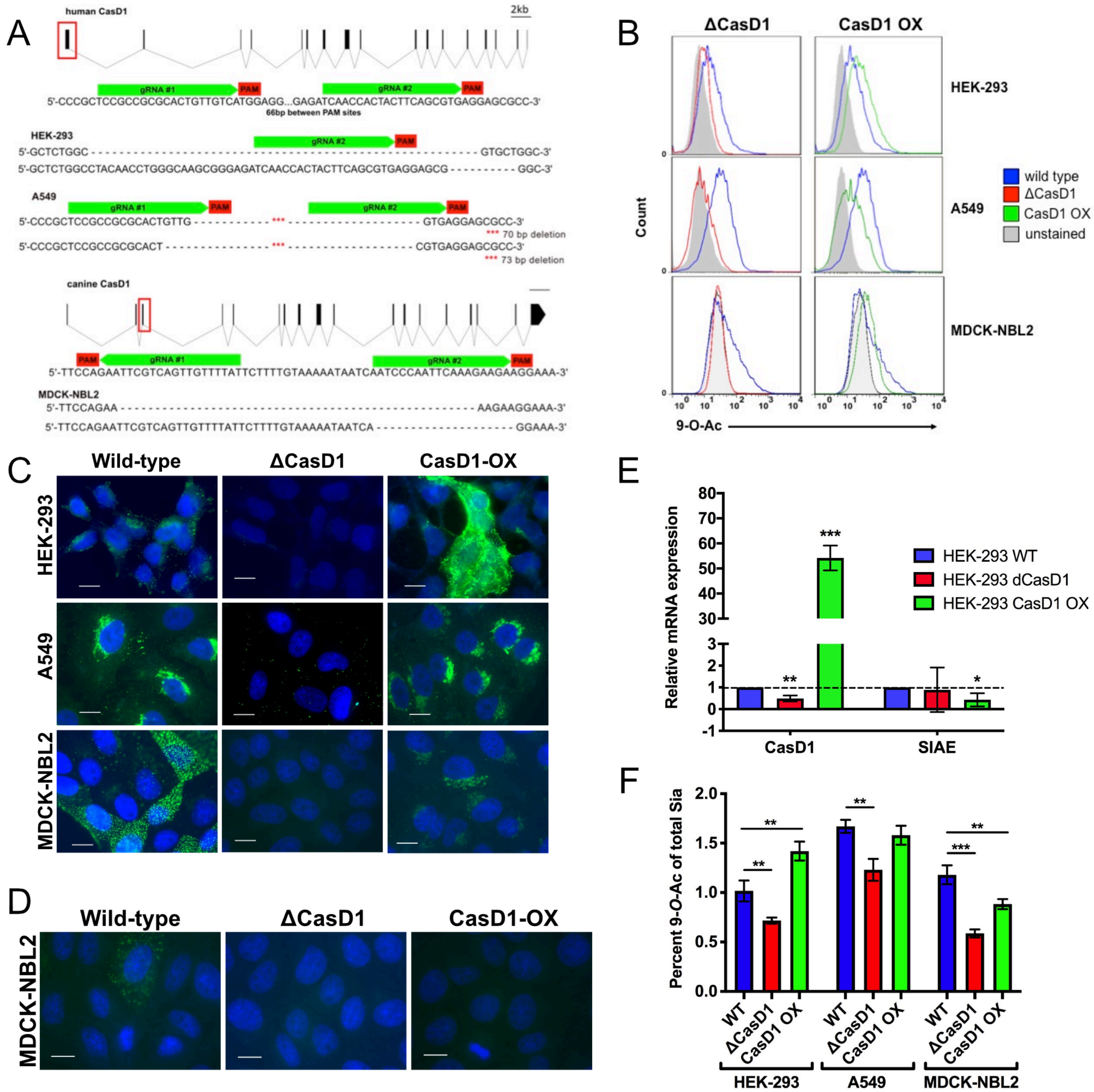


Figure 2.7

Editing expression of CasD1 in A549, HEK-293, MDCK-NLB2 cells.
A) Schematic of edits in the CasD1 gene and genotypes of edited cells for HEK-293, A549, and MDCK-NBL2 cells. **B**) Phenotype of edited cells by flow cytometry using 9-*O*-Ac probe (PToV HE-Fc). Cells were permeabilized using 0.001% Tween-20 to determine internal expression, as most modified Sia is retained internally. The graph is representative of three independent experiments. **C**) Immunofluorescence microscopy images of the different engineered cells stained with PToV HE-Fc to detect 9-*O*-Ac. Cells were permeabilized to reveal both surface and internal expression. Cells imaged at 60× magnification. Scale bar = 10 μm. **D**) Staining of MDCK WT, ΔCasD1, and CasD1-OX showing representative display of 7,9-*O*-Ac using the BCoV HE-Fc probe. Cells were permeabilized to reveal both surface and internal expression.

Cells imaged at 60× magnification. Scale bar = 10 μm. **E**) qPCR of CasD1 and SIAE expression in HEK-293 WT, ΔCasD1, CasD1-OX cells. CasD1 still shows mRNA due to mismatch between qPCR primers and edit sites, mRNA is still present but doesn't produce functional protein. Expression relative to house-keeping gene, GAPDH. Data analyzed using t-test in PRISM software. **F**) HPLC data for total Sia collected from cells by mild acid hydrolysis for WT, ΔCasD1, and CasD1 OX in HEK-293, A549, and MDCK cells. Data analyzed by t-test using PRISM software.
* = p-value ≤0.05; ** = p-value ≤0.01; *** = p-value ≤0.001.

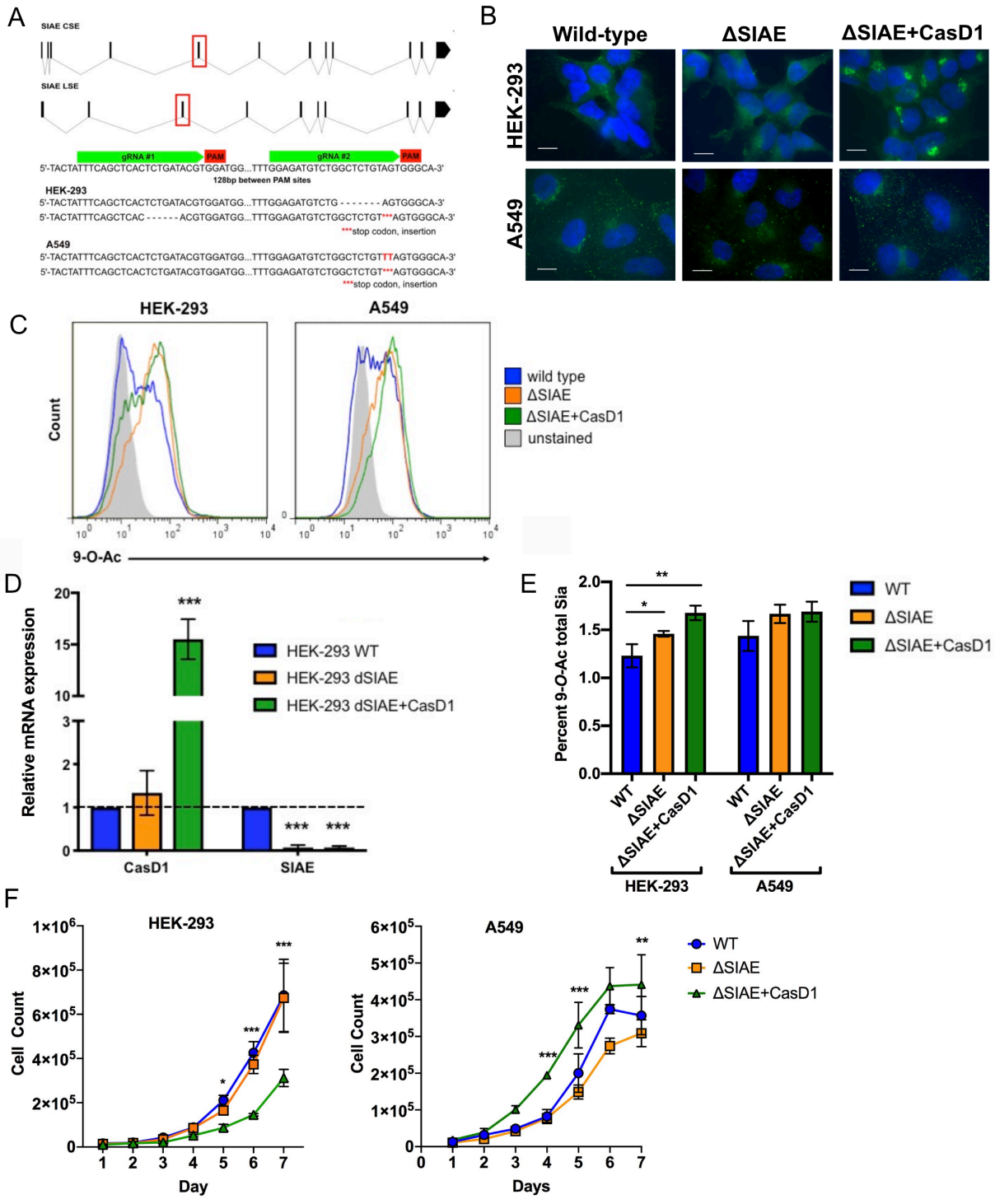


Figure 2.8

Editing the expression of SIAE in HEK-293 and A549 cells.
A) Schematic of edits in the SIAE gene, which was targeted to remove both isotypes of SIAE. The genotypes of edited cells show frame-shifts that lead to stop codons in all cases. **B**) Staining

of the different engineered cells with PToV HE-Fc to detect 9-*O*-Ac. Cells were permeabilized using 0.001% tween-20 to determine surface and internal expression. Cells imaged at 60x. Scale bar = 10 μm. **C**) Flow cytometry using PToV HE-Fc showing relative display of 9-*O*-Ac. Cells were permeabilized using 0.001% tween-20 to show both surface and internal expression. Graph is representative of three independent experiments. **D**) qPCR of relative SIAE and CasD1 mRNA expression in HEK-293 WT, ΔSIAE, and ΔSIAE+CasD1 cells compared to GAPDH. SIAE qPCR primers overlap with edit site. Data analyzed by t-test using PRISM software. **E**) HPLC data for total Sia collected from cells by mild acid hydrolysis for WT, ΔSIAE, and ΔSIAE+CasD1 in HEK-293 and A549 cells. Data analyzed by t-test using PRISM software. **F**) Growth curve WT, ΔSIAE, and ΔSIAE+CasD1 in HEK-293 and A549 cells. Cells were counted every 24 hours. Experiment was performed in triplicate. Data analyzed by 2-way Anova using PRISM software.
* = p-value ≤0.05; ** = p-value ≤0.01; *** = p-value ≤0.001.

(**Fig. 2.8B, C**). HPLC analysis also showed a small increase of 9-*O*-Ac in HEK-293 cells, but no 7,9-*O*-Ac was detected in either cell line (**Fig. 2.8E**). When the overexpression CasD1 plasmid was transfected into ΔSIAE cells (ΔSIAE+CasD1), cells showed an increase in both surface and internal 9-*O*-Ac, but no increase in surface or internal display of 7,9-*O*-Ac by either flow cytometry or immunofluorescence microscopy (**Fig. 2.8B,C**). HPLC confirmed the flow cytometry results by showing a small increase of 9-*O*-Ac levels in HEK-293 cells, similar to those seen in CasD1-OX (**Figs. 2.8E**). A549 cells showed a small but non-significant increase in 9-*O*-Ac levels compared to WT by HPLC analysis. Interestingly, HEK-293 ΔSIAE+CasD1 cells grew more slowly than either ΔSIAE or WT cells (**Fig. 2.8F**). However, A549 ΔSIAE+CasD1 showed increased growth rates compared to ΔSIAE or WT cells, while ΔSIAE cells had slightly lower growth rates compared to WT. This suggests that 7,9-*O*- and 9-*O*-Ac may affect cell metabolism and growth rates, and that these effects may be cell type specific. Overall, these results show that SIAE regulates levels of 9-*O*-Ac and 7,9-*O*-Ac, but that knocking out SIAE only leads to small increases in these modifications. However, there appear to be mechanisms regulating the surface display of 9-*O*-Ac and 7,9-*O*-Ac, as most of the modified Sia were specifically retained in the Golgi on human HEK-293 and A549 cells, in comparison to WT MDCK cells which appear to display most modified Sia on their surface.

Effects of 7,9-*O*-Ac and 9-*O*-Ac on influenza A, B, C, and D virus infection. To test the effects of 9-*O*-Ac and 7,9-*O*-Ac on virus infection, WT, ΔCasD1, and CasD1-OX HEK-293 cells were inoculated with human H1N1 (A/California/04/2009) and human H3N2 (A/Victoria/361/2011) IAV strains, and found no significant difference in infection efficiency in any of these cells (**Fig. 2.9A).** IAV strains, and IBV strains B/Colorado/06/2017 and B/Memphis/1/2018, also showed equal infection efficiency on MDCK WT compared to ΔCasD1

and CasD1-OX cells (**Fig. 2.9B,D**). B/Memphis did show a difference in relative infected cell counts between ΔCasD1 and CasD1-OX, although B/Colorado did not. However, ICV strains C/Ann Arbor/1/50, C/Taylor/1233/1947, and C/Victoria/1/2011, as well as IDV strains D/bovine/MS/C00020N/2014 and D/swine/OK/1334/2011, showed no infectivity in MDCK ΔCasD1 (**Fig. 2.9C,D**). The C/Ann Arbor and C/Victoria recovered a low level of infectivity in the MDCK CasD1-OX cells, while the D/bovine, and D/swine strains had a higher level of infectivity recovered in the CasD1-OX cells compared to ICV, but still were much lower than in MDCK WT. The inability of the ICV and IDV strains to fully recover infectivity may be due to the low surface levels of 9-*O*-Ac Sia on MDCK CasD1-OX cells. While the CasD1-OX cells had detectable surface levels of 9-*O*-Ac, these levels were lower than in MDCK WT and could be below the receptor levels required for efficient infection by these viruses. This suggests that ICV and IDV are able to utilize the levels of modified Sia on MDCK WT as their primary receptors for binding and infection, but cannot infect when they are removed or below a certain necessary threshold.

Effects of 7,9-*O*- or 9-*O*-acetyl Sia modifications on HA binding and NA activity. Since the low levels of 7,9-*O*- and 9-*O*-AC on the surface of cells did not effect IAV infection in our assays, we sought to determine the effects of these modifications on HA binding and NA cleavage using mouse erythrocytes which have ~45% of their total Sia is *O*-acetylated, primarily 7,9-*O*- and 9-*O*-Ac (**Fig 2.10A**). While mouse erythrocytes contain primarily α2,3-linked Sia, they also contain some α2,6-linked Sia that could be bound by human IAV strains (**Fig. 2.10A**). Mouse erythrocytes were therefore used as a substrate for both binding (hemagglutination) assays and neuraminidase cleavage assays. The human IAV strains human H1N1 (A/California/04/2009) and human H3N2 (A/Victoria/361/2011) strains were mixed with either

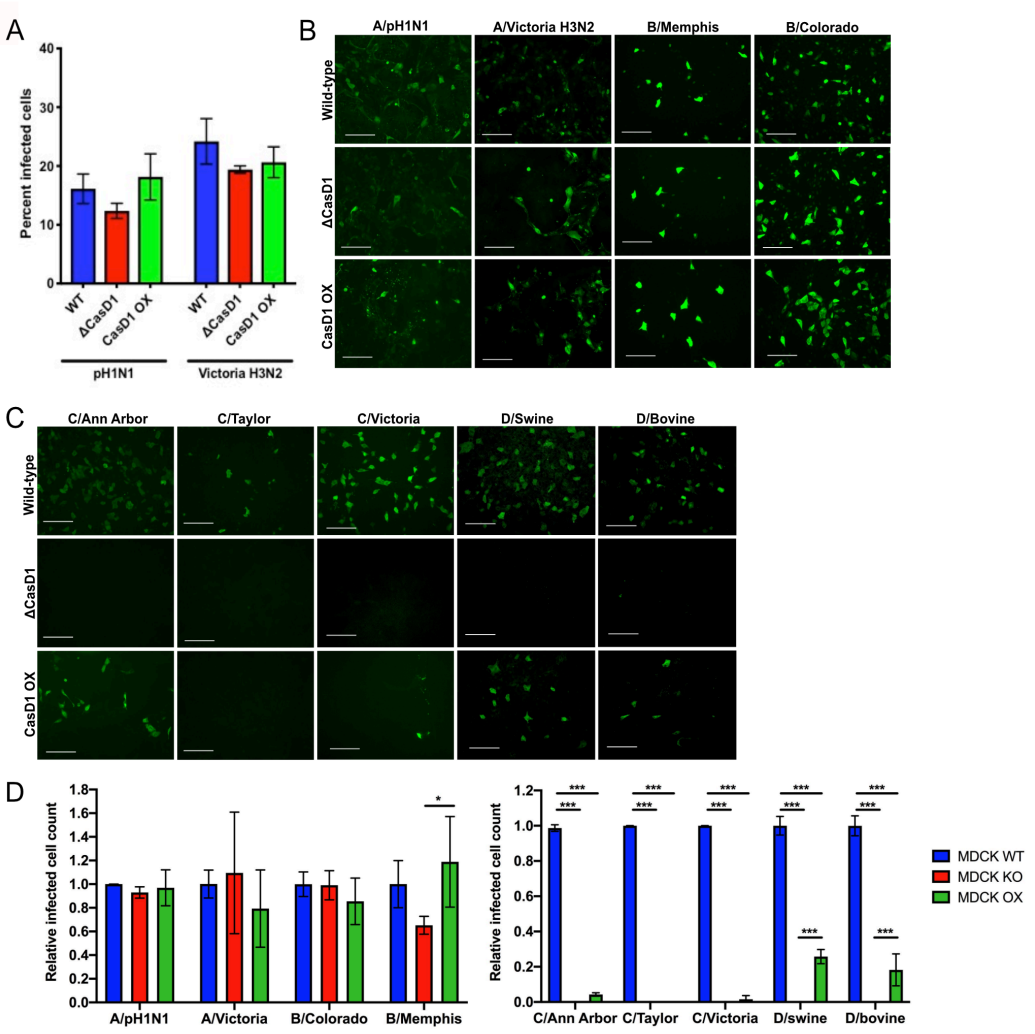
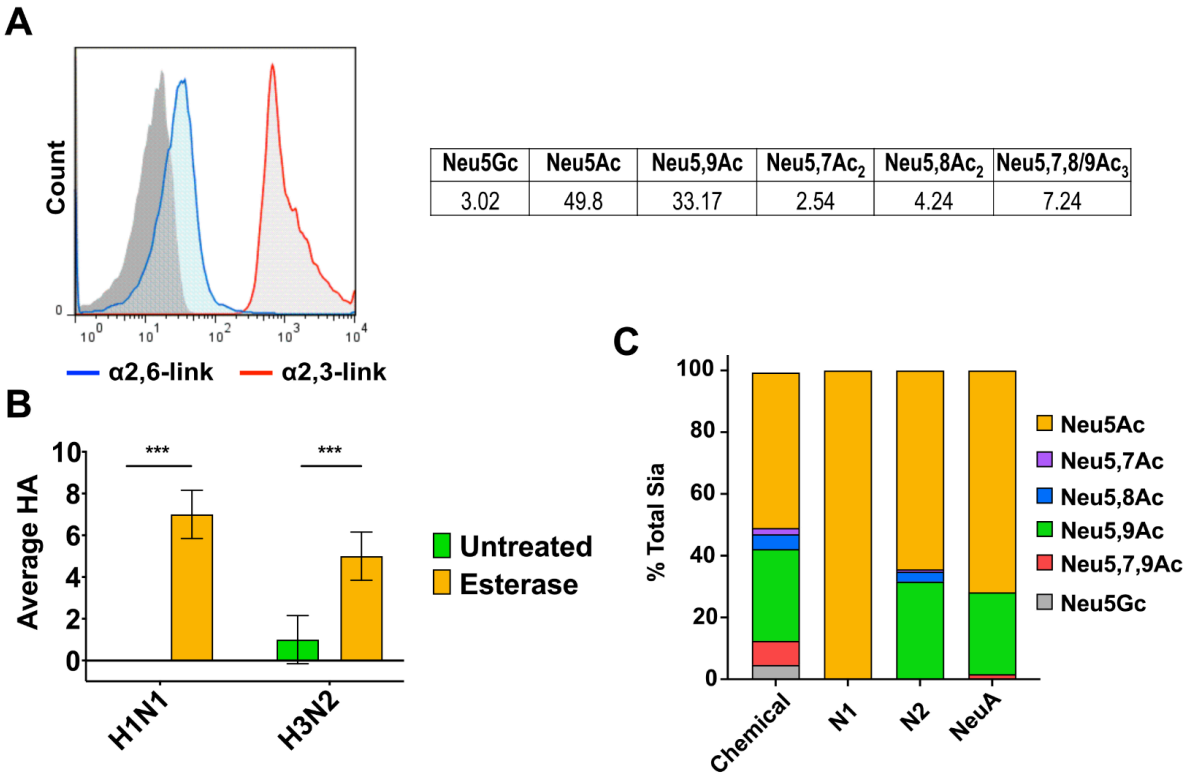


Figure 2.9

Infection of WT, ΔCasD1, and CasD1-OX cells with IAV, IBV, ICV, and IDV.
A) HEK-293 WT, ΔCasD1, CasD1-OX cells were inoculated at MOI 0.1 with IAV strains pH1N1 (A/California/04/2009) and Victoria H3N2 (A/Victoria/361/2011). Cells were fixed at 24 hours and infected cells per field were counted. Experiment was performed in triplicate. Data analyzed by 2-way Anova using PRISM software. **B**) MDCK WT, ΔCasD1, and CasD1-OX cells were inoculated at high MOI with IAV strains pH1N1 (A/California/04/2009) and Victoria H3N2 (A/Victoria/361/2011), and IBV strains B/Memphis/1/2018 and B/Colorado/06/2017 for 48 hr, then imaged at 10x magnification. Scale bar = 100 µm. Representative images of three independent experiments. **C**) MDCK WT, ΔCasD1, and CasD1-OX cells were inoculated at high MOI with ICV strains C/Ann Arbor/1/50, C/Taylor/1233/1947, and C/Victoria/1/2011, and IDV strains D/bovine/MS/C000020N/2014 and D/swine/OK/1334/2011 for 48 hr, then imaged at 10x magnification. Scale bar = 100 µm. Representative image of three independent experiments. **D**) Quantification of relative infected cell counts for (B) and (C) as determined through image analysis with Image J. Data analyzed by 2-way Anova using PRISM software.
* = p-value ≤0.05; ** = p-value ≤0.01; *** = p-value ≤0.001.

untreated mouse erythrocytes or mouse erythrocytes pre-treated with esterase active BCV HE-Fc to remove all acetyl modifications (**Fig. 2.10B**). Significantly greater hemagglutination was seen for both viruses on the esterase-treated mouse erythrocytes compared to untreated erythrocytes. It should be noted that the hemagglutination was still quite low, likely due to the lower levels of α2,6-linked Sia on mouse erythrocytes.

To determine the effect of 7,9-*O*- and 9-*O*-Ac on NA cleavage, mouse erythrocytes were incubated with soluble NA, in the form of NA-expressing virus like particles (VLPs) (48). Freed Sia was then collected, analyzed using HPLC, and the profile of Sia released by the NA VLPs was compared to Sia composition on mouse erythrocytes released by chemical hydrolysis (**Fig. 2.10C**). The profiles for both N1 and N2 VLPs showed a preferential release of unmodified Neu5Ac compared to modified Sia forms, shown in the increased proportion of this Sia in the N1 and N2 profiles. Compared to the chemical release profile, the N1 VLPs did not release any detectable *O*-acetylated Sia, while the N2 VLPs were able to release 9-*O*-Ac (Neu5,9Ac_2), but released significantly less 7-*O*- and 8-*O*-Ac, and were unable to release any 7,9-*O*-Ac (Neu5,7,9Ac_3). Similar to N2 VLPs, commercial NeuA from *Arthrobacter ureafaciens,* used as an activity control, also had a bias towards unmodified Neu5Ac, with decreased activity against 7-*O*-*,* 8-*O*-*,* and 7,9-*O*-Ac. These results indicate that mono *O*-acetyl modifications, including 7-*O*-, 8-*O*-, and 9-*O*-Ac, reduced the Sia susceptibility to NA cleavage and that di-acetyl 7,9-*O*-Ac was the most resistant to NA cleavage. Additionally, there was variability between the ability of the NA VLPs and NeuA to cleave the different Sia forms. While the presence of these modified Sia on cells was too low to effect IAV infection, it is likely that higher levels present on erythrocytes, or potentially in secreted proteins in mucus, would inhibit infection or release of virus.



Neu5Gc	Neu5Ac	Neu5,9Ac	Neu5,7Ac$_2$	Neu5,8Ac$_2$	Neu5,7,8/9Ac$_3$
3.02	49.8	33.17	2.54	4.24	7.24

Figure 2.10

The effects of 7,9-*O*- and 9-*O*-Ac on IAV HA binding and NA cleavage
A) Surface lectin staining for α2,6-linked Sia and α2,3-linked Sia on mouse erythrocytes via flow cytometry and a table showing total Sia analysis of mouse erythrocytes using HPLC. Data for each Sia variant is given as a percent of total Sia, averaged across three independent samples. **B**) Hemagglutination of human H1N1 and H3N2 IAV strains on untreated or esterase treated mouse erythrocytes. Data is averaged across three independent experiments, with data analyzed by 2-way Anova using PRISM software. **C**) Profiles of total Sia freed by either chemical treatment, N1 VLPs, N2 VLPs, or commercial NeuA sialidase as determined by HPLC. Data is averaged across two independent experiments.
* = p-value ≤0.05; ** = p-value ≤0.01; *** = p-value ≤0.001.

2.5 DISCUSSION

Both 9-*O*-Ac and 7,9-*O*-Ac Sia are widely expressed within tissues and on mucosal surfaces of many animals, but with significant variation in the amounts present in different cells, tissues, and animals (10, 11). These modified Sia are present in secreted mucus on mucosal surfaces, including GI and respiratory tissues, where they can potentially control many interactions with both normal flora and pathogens. However, there is still little known about the details of their display levels, cell association, and the ways in which their synthesis is regulated either in cells or on mucosal surfaces. While there have been suggestions that 9-*O*- and 7,9-*O*-Ac might influence IAV and IBV infection by interfering with HA binding or NA activities, direct evidence for their effects is sparse. In contrast, ICV and IDV are known to use 9-*O*-Ac as their primary receptor for cell binding and infection. Here we use a number of new tools to define the cell-specific expression of 9-*O*- and 7,9-*O*-Ac and provide a preliminary test of their effects on IAV, IBV, ICV, and IDV infection.

We confirmed that 9-*O*- and 7,9-*O*-Ac are expressed on cells in culture and that expression varies between cell lines, as has been previously reported (38). However, when present these modified Sia made up only 1 to 2% of the total Sia when analyzed by HPLC. Cells showed distinct population heterogeneity in 7,9-O- and 9-*O*-Ac display and localization that was observed over many passages examined. Previous studies using ICV HEF probes found similar staining patterns on some cell lines and also showed that after sorting into high and low staining populations, both populations returned to previous levels of heterogeneity within a few passages (11, 38). Here we saw that the modified Sia are retained in the Golgi of HEK-293 and A549 cells, although an occasional cell displayed these modified Sia on the surface. It is not clear why these modified Sia are localized in the Golgi, although perhaps the modifications could block the

onward trafficking of glycoproteins to the cell surface. In contrast to the human cells lines, MDCK cells showed many cells expressing 7,9-*O*-Ac and 9-*O*-Ac on the cell surface, raising the possibility that trafficking could be cell-type or species specific. A summary of the synthesis and localization of 7,9-*O*- and 9-*O*-Ac is shown in **Figure 2.11**.

To look more closely at the expression and roles of 7,9-*O*- and 9-*O*-Ac modifications and their effects on viral infections, we prepared glyco-engineered cell lines that lacked these modifications or that expressed higher levels of CasD1. A deletion and frame shift in CasD1 completely removed both 9-*O*- and 7,9-*O*-Ac expression, confirming this enzyme was responsible for creating both modifications, likely through addition to the C-7 position, from which it migrates to the C-9 position (5, 8, 9). Adding CasD1 back into the cells by plasmid transfection restored modified Sia expression, but none of the cell clones isolated showed universally higher levels of modified Sia synthesis and population heterogeneity was still seen. The CasD1-transfected HEK-293 cells showed the greatest increase, while for A549 and MDCK cells expression was similar to or lower than WT cells. However, even in the HEK-293 CasD1-OX cells, 9-*O*-Ac still only accounted for ~1.5% of total Sia and 7,9-O-Ac was not detected. This indicates that the levels of these modifications are not only controlled by the expression of CasD1, and could be regulated by other processes. Additionally, there is clearly a differential regulation of 7,9-*O*-Ac compared to 9-*O*-Ac, as over-expressing CasD1 did not lead to an increase in 7,9-*O*-Ac, as it was not detectable by HPLC and only very low fluorescence via staining with BCoV HE-Fc.

One candidate for control of these modifications is SIAE. When we knocked SIAE out of HEK-293 and A549 cells, we saw an increase in 9-*O*-Ac that was still retained in the Golgi but

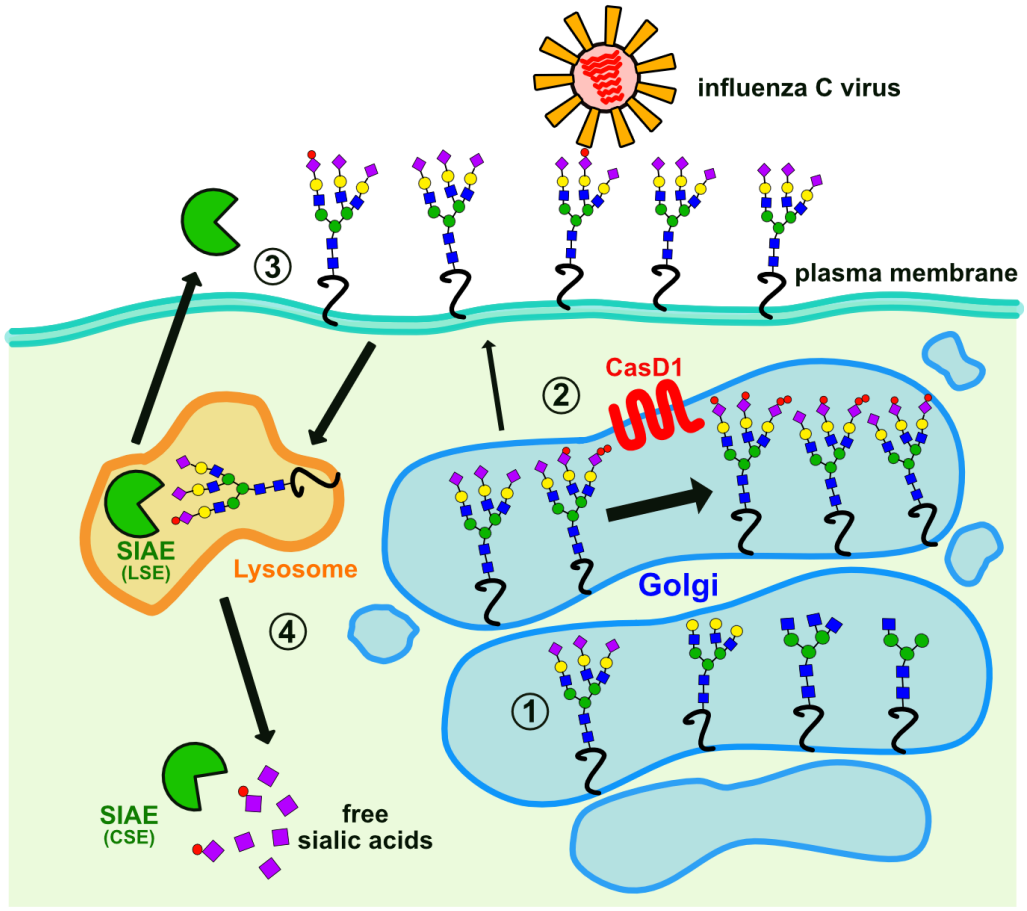


Figure 2.11

Summary of proposed 7,9-*O*- and 9-*O*-acetyl sialic acid production and trafficking in cells.
(1) Sia (purple diamond) is added to the growing glycan chain in the Golgi by sialyltransferases using CMP-Neu5Ac or CMP-Neu5Gc substrates which are synthesized in the nucleus by the addition of cytodine monophosphate (CMP) to Neu5Ac or Neu5Gc, and are specifically imported into the Golgi.
(2) CMP-Neu5Ac or CMP-Neu5Gc Sia are modified by CasD1, adding one or two acetyl groups to form 9-*O*-Ac or 7,9-*O*-Ac, respectively (red circles) before being added to glycan chains. The majority of glycoproteins with these modifications are retained in the Golgi (large arrow) of many cells including the HEK-293 and A549 cells examined here, while only some, and mostly 9-*O*-Ac, are transported to the cell surface (small arrow).
(3) Surface displayed *O*-acetyl Sia can interact with pathogens, cell receptors, or lectins. For example, influenza C virus uses 9-*O*-Ac as its receptor. Secreted forms of SIAE may also remove the *O*-acetyl modifications, altering these lectin-ligand interactions.
(4) When glycoproteins are recycled from the cell surface, the lysosomal form of SIAE (LSE) can remove *O*-acetyl modifications from Sia. Free Sia are exported to the cytosol, where the cytosolic form of SIAE (CSE) can also remove any remaining *O*-acetyl modifications. Unmodified Sia can then be "activated" in the nucleus by the addition of CMP and transported to the Golgi for addition to new glycan chains.

no increase in 7,9-*O*-Ac. Expressing CasD1 from a plasmid in ΔSIAE HEK-293 and A549 cells resulted in an additional small increase in 9-*O*-Ac that was still Golgi associated, but with little increase in expression on the cell surface. Similar to the CasD1-OX cells, no increase in 7,9-*O*-Ac was seen in either ΔSIAE and ΔSIAE+CasD1 cell lines by HPLC analysis. Both ΔSIAE+CasD1 lines had altered growth rates compared to WT cells: HEK-293 ΔSIAE+CasD1 had delayed growth rates, while A549 ΔSIAE+CasD1 had increased growth rates. This suggests that dysregulation of SIAE and CasD1 by gene manipulation affects cell metabolism and growth in a cell-type specific manner, possibly through the build-up of glycoproteins in the Golgi or through dysregulation of the sialic acid recycling pathway. These effects on cell growth could have implications for cancer and organismal development, as these variant Sia are involved in both these processes (15–17, 20–23, 49). Further research is needed to determine how 7,9-*O*- and 9-*O*-Ac expression is regulated, how they are transported within the ER and Golgi, and their roles in cell growth. Of particular interest will be disentangling the individual regulation of 7,9-*O*-Ac from 9-*O*-Ac, as there does seem to be differences in how they are regulated both in our cells and in previously reported expression in animal tissues (10, 11).

Studies with two strains of IAV and two strains of IBV showed no differences in infection efficiency in WT HEK-293 or MDCK-NBL2 cells compared to their ΔCasD1 or CasD1-OX variants. This is unsurprising, as >95% of the Sia is un-acetylated Neu5Ac which can be utilized by IAV and IBV as a receptor for binding and entry. While the low levels present on cells were not enough to effect either IAV and IBV infection, we did determine that higher levels, such as those found on mouse erythrocytes (~45% *O*-Ac), were able to block IAV HA binding and decrease IAV NA cleavage efficiency in a strain dependent manner. It is possible, then, that if higher levels are present on secreted proteins in mucus, they could inhibit infection

prior to the virus reaching the cell surface. Further analysis of mucus *O*-acetylation in IAV hosts would be necessary to determine if 7,9-*O*- and 9-*O*-Ac could contribute to the barrier function of mucus in respiratory tissues (41).

The 1-2% of 9-*O*-Ac on the surface of WT MDCK-NBL2 cells was sufficient for ICV and IDV virus binding and entry, and that cell susceptibility was lost when CasD1 was inactivated. Interestingly, there does still seem to be necessary minimum threshold of 9-*O*-Ac presence needed for some ICV and IDV strains to infect cells, as C/Ann Arbor and C/Victoria were able to infect some MDCK-CasD1 OX cells, at a lower level than MDCK WT, while C/Taylor strain was not. Similarly, IDV also showed lower levels of infectivity in MDCK-CasD1 OX cells compared to MDCK-WT. It is likely that similar results would be seen for other viruses that use these modified Sia as a receptor, including human coronaviruses OC43 and HKU1 (50).

In summary, we have shown that these modifications are present in different cell lines sourced from different species of animals, but they make up a small minority of the total Sia present. In addition, these modifications vary considerably in their localization and have an inherent heterogeneity within cell populations. While the presence of both 7,9-*O*- and 9-*O*-Ac were dependent on the activity of CasD1, the relative proportions, levels of expression, and localization appear to be controlled by more complex mechanisms than simply the expression of CasD1 and SIAE. How this regulated expression affects cell homeostasis is unknown, but it is likely relevant during development, immune responses, and in cancers that show dysregulation of 7,9-*O*- and 9-*O*-Ac expression. For viruses such as ICV and IDV that rely on 9-*O*-Ac for infection, the low levels seen on cell surfaces are sufficient for infection. We found that 7,9-*O*- and 9-*O*-Ac do have inhibitory activity on HA binding and NA cleavage when present at higher levels. While these modifications are present on cell surfaces at levels too low to affect IAV and

IBV infection, they are expressed at much higher levels in mucosal tissues and in secreted mucus proteins of many animals, which may provide a more effective barrier (10, 11, 51–53). We are currently examining these processes for secreted mucus and other sources in different animals.

2.6 MATERIALS AND METHODS

Cells and virus. HEK-293, A549, MDCK-NBL2, MDCK type I, and MDCK type II cells were grown in Dulbecco's Modified Eagle Medium (DMEM) with 5% fetal bovine serum and 50 μg/ml gentamicin. HEK-293, A549, and MDCK-NBL2 cells were obtained from ATCC. MDCK type I and type II cells were gifts from Dr. William Young (University of Kentucky) (35, 37, 54, 55). SiNEC and SiTEC cells were gifts from Dr. Stacey Shultz-Cherry (St. Jude Children's Hospital). Influenza A virus strains pH1N1 (A/California/04/2009) and Victoria H3N2 (A/Victoria/361/2011) were rescued from reverse genetics plasmids using established protocols (56, 57). Rescued viruses were grown to low passage on MDCK-NBL2 cells using infection media containing DMEM, 0.03% BSA, and 1 μg/ml TPCK-treated trypsin. Influenza B strains B/Colorado/06/2017 and B/Memphis/1/2018, and influenza D strains D/bovine/MS/C00020N/2014 and D/swine/OK/1334/2011 were gifts from Dr. Richard Webby (St. Jude Children's hospital) and were grown in infection media containing DMEM, 0.03% BSA, and 1 μg/ml TPCK-treated trypsin. Influenza C virus C/Ann Arbor/1/50 and C/Taylor/1233/1947 were a gift from Dr. Andrew Pekosz (Johns Hopkins University), C/Victoria/1/2011 was a gift from Dr. Richard Webby (St. Jude Children's Hospital). All ICV strains were grown on MDCK-NBL2 cells using infection media containing DMEM, 0.03% BSA, and 5 μg/ml TPCK-treated trypsin.

Immunofluorescence microscopy and flow cytometry. Cells were stained using probes derived from viral hemagglutinin esterase proteins fused to human IgG1 Fc (HE-Fc). The

porcine torovirus strain 4 HE-Fc (PToV HE-Fc) primarily recognizes 9-*O*-Ac and the bovine coronavirus Mebus strain HE-Fc (BCoV HE-Fc) recognizes 7,9-*O*-Ac and shows low levels of binding to 9-*O*-Ac (10, 11). For immunofluorescence microscopy, cells were seeded onto glass coverslips and incubated overnight at 37°C and 5% CO_2. Coverslips were fixed in 4% paraformaldehyde (PFA) for 15 min. Coverslips were incubated with Carbo-Free Blocking Solution (Vector Laboratories) for 1 hr at room temperature, with optional permeabilization with 0.001% Tween-20. To stain, HE-Fc probes were pre-complexed with Alexa-488 labeled anti-human IgG antibody for 1 hr at 4°C then diluted in Carbo-Free blocking solution to a final concentration of 5 μg/ml HE-Fc and 1:500 of secondary antibody. Cells were stained with HE-Fc/anti-IgG complex for 1 hr at room temperature. Coverslips were mounted using Prolong Antifade-Gold with DAPI (Invitrogen). Cells were imaged using a Nikon TE300 fluorescent microscope. For flow cytometry, cells were seeded onto non-adherent cell culture dishes and incubated overnight at 37°C and 5% CO_2. Cells were collected using ice-cold PBS (HEK-293 and A549 cells) or Accutase (Sigma, MDCK cells) to retain surface glycans, then fixed in 4% PFA for 15 min. Cells were blocked as above. HE-Fc probes were prepared as above with final concentrations of 5 μg/ml HE-Fc probe and 1:1200 of anti-IgG. A Millipore Guava EasyCyte Plus flow cyotometer (EMD Millipore, Billerica, MA) was used to collect data, analysis using FlowJo software (TreeStar, Ashland, OR). Statistical analyses were performed in PRISM software (GraphPad, version 8).

Cell line mutations and characterization. Two methods for utilizing CRISPR-Cas9 were used. For A549 and MDCK cells, paired Cas9 plasmids (PX459, Addgene plasmid #62988) targeted adjacent sites in early exons of CasD1 as diagrammed in Figure 4A. Plasmids were transfected using TransIT-X2 (Mirus Bio LLC) (8). For HEK-293 cells, nickase Cas9 plasmids

(PX462, Addgene plasmid #62987) were used instead. Transfected cells were selected with puromycin and single cell clones screened with PToV-P4 HE-Fc to identify non-staining variants. Edited sequences were confirmed by PCR amplification of the targeted regions, and sequencing the PCR product for each allele. Knock-out cell lines were used to prepare over-expression cell lines by transfection of a pcDNA3.1(-) plasmid expressing the complete human CasD1 cDNA open reading frame synthesized by Bio Basic (Markham, Ontario, Canada). Transfected cells were selected with G418 and single cell clones screened by staining with PToV-P4 HE-Fc to identify 9-*O*-Ac positive cell lines. Editing of the SIAE gene followed a similar protocol for CRISPR-Cas9 as above, and the gene regions targeted are shown in Figure 5A. After transfection and selection, cells were cloned and single-cell clones were screened by direct PCR amplification of the target gene region and analysis of the PCR product size for the edited form of both alleles. Full sequencing of each allele and qPCR were performed to confirm deletion of the gene.

Quantification of Sia variants. The Sia composition of cells were determined by incubating with 2M acetic acid at 80°C for 3 hr, filtration through a Microcon 10 kD centrifugal filter (Millipore), and drying in a SpeedVac vacuum concentrator. Released Sia were derivatized with 1,2-diamino-4, 5-methylenedioxybenzene (DMB, Sigma Aldrich) for 2.5 hr at 50°C (39). HPLC analysis was performed using a Dionex UltiMate 3000 system with an Acclaim C18 column (ThermoFisher) under isocratic elution in 7% methanol, 7% acetonitrile, and 86% water. Sia standards included bovine sub-maxillary mucin and commercial standards for Neu5Ac and Neu5Gc (Sigma Aldrich). Statistical analyses were performed in PRISM software (GraphPad, version 8).

Characterization of A549 conditioned media. Conditioned media from A549 cells was prepared by washing a fully confluent flask of cells to remove any serum, and allowing the cells to grow in serum-free media for 5-7 days. Conditioned media was collected, dialyzed with three volumes of PBS, and concentrated using a 30 kD centrifugal filter (Pall Corporation). Protein concentration was determined using a Qubit 4 fluorometer (Invitrogen). To determine Muc5B presence, two-fold dilutions of conditioned media were compared to purified human Muc5B using a 8% SDS-PAGE gel and probed with an anti-human Muc5B antibody (both purified protein and antibody were gifts from Dr. Stefan Ruhl, University of Buffalo). Conditioned media was analyzed for total Sia using the HPLC methods listed above.

qPCR of SIAE and CasD1 expression. RNA from cells was extracted using EZNA Total RNA Kit I (Omega Bio-Tek) and cDNA was synthesized using SuperScript II Reverse Transcriptase (Invitrogen) standard protocol using oligo$(dT)_{12\text{-}18}$ primers (Invitrogen). CasD1 and SIAE specific primers were designed using Geneious (Biomatters, Ltd.). For CasD1, adequate primers could not be targeted around the CRISPR-Cas9 edit site due to high G/C content. qPCR was performed on a Applied Biosystems StepOnePlus Real-Time PCR System using Fast SYBR Green (Bio-Rad). Data was analyzed using StepOne software (Applied Biosystems, version 2.1) and PRISM statistical analysis software (GraphPad, version 8).

Virus infection assays. For infection in HEK-293 cells with influenza A virus, cells were seeded on coverslips and inoculated with an MOI of 0.1. Coverslips of inoculated cells were fixed at 24 hrs and stained for virus using an anti-NP antibody and co-stained with DAPI. Percent-infected cells were determined by imaging coverslips and counting infected cells per field. Statistical analyses were performed in PRISM software (GraphPad, version 8). For infection in MDCK cells with influenza A, B, C, and D, cells were seeded into 96 well plates and

inoculated for 48 hrs. Cells were then fixed and stained using a mouse anti-influenza A NP antibody, a mouse anti-influenza B NP antibody (Abcam), a poly-clonal mouse anti-influenza C antibody (a gift from Dr. Peter Palese, Icahn School of Medicine at Mount Sinai), or a poly-clonal rabbit anti-influenza D antibody (a gift from Dr. Feng Li, South Dakota State University). Cells were imaged using a Nikon TE300 fluorescent microscope. Images were analyzed for relative infected cell counts using Image J (NIH and LOCI, University of Wisconsin).

Mouse erythrocyte hemagglutination and lectin staining. Mouse erythrocytes were collected from C57BL/6 mice by euthanizing the mice and immediately collecting blood by heart puncture into Alsevier's solution. Erythrocytes were washed in PBS three times and diluted to 5% v/v in PBS. To determine Sia linkage type, 5% washed erythrocytes were blocked for 1 hr using 1x Carbo Free blocking buffer (Vector Laboratories) and then stained for 1 hr with fluoroscein-labeled plant lectins SNA, which binds α2,6-linked Sia, and MAA I, which binds α2,3-linked Sia (Vector Laboratories). Stained erythrocytes were then analyzed using flow cytometry as described above**.** For hemagglutination experiments, 5% mouse erythrocytes were left untreated or treated with 30 ug/ml BCoV HE-Fc for 18 hrs at 37°C, then diluted to 0.75% v/v in PBS for HA assays. Briefly HA assays were performed by 2-fold serially diluting virus in duplicate per treatment in a V-bottom 96 well plate. Treated or untreated mouse erythrocytes were added to virus and allowed to hemagglutinate at 4°C to prevent NA cleavage.

Generation of NA VLPs & NA cleavage assay. NA sequences were obtained from GenBank (N1: ACP44181, N2: AGC70842). Sequences were tagged, codon optimized, and ordered through Biomatik in the PcDNA3.1(+) vector. To produce VLPs, HEK-293T cells were seeded in 15cm plates and transfected when 80% confluent. Cells were transfected with 4 µl of Polyethylenimine (PEI) (Polysciences (cat# 23966-2)) at 1mg/ml concentration for every 1 µg

plasmid DNA stock of in 9ml of Opti-MEM. Eight hours post transfection, 6ml of pre-warmed Opti-MEM was added. Supernatant was collected 72 hours post transfection and purified using ultracentrifugation (110,000 xg, 1.5 hrs, 4°C) through a 20% sucrose cushion. Pellet was re-suspended in PBS and stored at 4°C. To determine NA cleavage on mouse erythrocytes, 5% v/v mouse erythrocytes in PBS were treated with 1:100 NA VLPs for 4 hours at 37°C. Free Sia was collected and prepared for HPLC analysis as above.

2.7 ACKNOWLEDGEMENTS

We thank Wendy Weichert for expert technical support. Brynn Lawrence for HA-Fc production.

SUPPORT

Supported in part by CRIP (Center of Research in Influenza Pathogenesis), an NIAID funded Center of Excellence in Influenza Research and Surveillance (CEIRS) contract HHSN272201400008C to DRP and CRP, NIH grant R01 GM080533 to CRP, and NIH Common Fund Grant (U01CA199792) to AV.

REFERENCES

1. Varki A, Schauer R. 2009. Chapter 14 Sialic AcidsEssentials of Glycobiology - second edition, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Springs Harbor, NY.

2. Varki NM, Varki A. 2007. Diversity in cell surface sialic acid presentations: implications for biology and disease. Laboratory investigation; a journal of technical methods and pathology 87:851–857.

3. Lehmann F, Tiralongo E, Tiralongo J. 2006. Sialic acid-specific lectins: Occurrence, specificity and function. Cellular and Molecular Life Sciences 63:1331–1354.

4. Varki A. 2007. Glycan-based interactions involving vertebrate sialic-acid-recognizing proteins. Nature 446:1023–9.

5. Mandal C, Schwartz-Albiez R, Vlasak R. 2012. Functions and Biosynthesis of O-acetylated sialic acids. Top Curr Chem.

6. Arming S, Wipfler D, Mayr J, Merling A, Vilas U, Schauer R, Schwartz-Albiez R, Vlasak R. 2011. The human Cas1 protein: A sialic acid-specific O-acetyltransferase? Glycobiology 21:553–564.

7. Baumann A-MTM, Bakkers MJ, Buettner FF, Hartmann M, Grove M, Langereis MA, de Groot RJ, Muhlenhoff M. 2015. 9-O-Acetylation of sialic acids is catalysed by CASD1 via a covalent acetyl-enzyme intermediate. Nature communications 6:7673.

8. Vandamme-Feldhaus V, Schauer R. 1998. Characterization of the enzymatic 7-O-acetylation of sialic acids and evidence for enzymatic O-acetyl migration from C-7 to C-9 in bovine submandibular gland. J Biochem 124:111–121.

9. Kamerling JP, Schauer R, Shukla AK, Stoll S, Halbeek H, Vliegenthart J. 1987. Migration of O-acetyl groups in N,O-acetylneuraminic acids. European Journal of Biochemistry 162:601–607.

10. Wasik BR, Barnard KN, Ossiboff RJ, Khedri Z, Feng KH, Perez DR, Varki A, Parrish CR. 2017. Distribution of O-acetylated sialic acids among the tissues of influenza hosts. mSphere 2:e00379–16.

11. Langereis MA, Bakkers MJ, Deng L, Vered P-K, Vervoort SJ, Hulswit RJ, van Vliet AL, Gerwig GJ, de Poot SA, Boot W, van Ederen AM, Heesters BA, van der Loos CM, van Kuppeveld FJ, Yu H, Huizinga EG, Chen X, Varki A, Kamerling JP, de Groot RJ. 2015. Complexity and Diversity of the Mammalian Sialome Revealed by Nidovirus Virolectins. Cell reports 11:1966–1978.

12. Takematsu H, Diaz S, Stoddart A, Zhang Y, Varki A. 1999. Lysosomal and cytosolic sialic acid 9-O-acetylesterase activities can be encoded by one gene via differential usage of a signal peptide-encoding exon at the N terminus. Journal of Biological Chemistry 274:25623–25631.

13. Orizio F, Damiati E, Giacopuzzi E, Benaglia G, Pianta S, Schauer R, Schwartz-Albiez R, Borsani G, Bresciani R, Monti E. 2015. Human sialic acid acetyl esterase: Towards a better understanding of a puzzling enzyme. Glycobiology 25:992–1006.

14. Schauer R, Srinivasan GV, Wipfler D, Kniep B, Schwartz-Albiez R. 2011. O-Acetylated Sialic Acids and Their Role in Immune Defense, p. 525–548. *In* The Molecular Immunology of Complex Carbohydrates-3. Springer Science + Business Media, LLC.

15. Wipfler D, Srinivasan GV, Sadick H, Kniep B, Arming S, Willhauck-Fleckenstein M, Vlasak R, Schauer R, Schwartz-Albiez R. 2011. Differentially regulated expression of 9-O-acetyl GD3 (CD60b) and 7-O-acetyl-GD3 (CD60c) during differentiation and maturation of human T and B lymphocytes. Glycobiology 21:1161–1172.

16. Cariappa a., Takematsu H, Liu H, Diaz S, Haider K, Boboila C, Kalloo G, Connole M, Shi HN, Varki N, Varki A, Pillai S. 2009. B cell antigen receptor signal strength and peripheral B cell development are regulated by a 9-O-acetyl sialic acid esterase. Journal of Experimental Medicine 206:125–138.

17. Pillai S, Cariappa A, Pirnie SP. 2009. Esterases and autoimmunity: the sialic acid acetylesterase pathway and the regulation of peripheral B cell tolerance. Trends Immunol 30:488–493.

18. Hunter CD, Khanna N, Richards MR, Rezaei Darestani R, Zou C, Klassen JS, Cairo CW. 2018. Human Neuraminidase Isoenzymes Show Variable Activities for 9-O-Acetyl-sialoside Substrates. ACS Chem Biol.

19. Mahajan VS, Pillai S. 2016. Sialic acids and autoimmune disease. Immunol Rev 269:145–161.

20. Muchmore E a, Varki NM, Fukuda M, Varki a. 1987. Developmental regulation of sialic acid modifications in rat and human colon. The FASEB journal : official publication of the Federation of American Societies for Experimental Biology 1:229–235.

21. Shen Y, Kohla G, Lrhorfi AL, Sipos B, Kalthoff H, Gerwig GJ, Kamerling JP, Schauer R, Tiralongo J. 2004. O-acetylation and de-O-acetylation of sialic acids in human colorectal carcinoma. Eur J Biochem 271:281–290.

22. Mandal C, Srinivasan GV, Chowdhury S, Chandra S, Mandal C, Schauer R, Mandal C. 2009. High level of sialate-O-acetyltransferase activity in lymphoblasts of childhood acute lymphoblastic leukaemia (ALL): Enzyme characterization and correlation with disease status. Glycoconjugate Journal 26:57–73.

23. Cavdarli S, Dewald JH, Yamakawa N, Guérardel Y, Terme M, Le Doussal J-M, Delannoy P, Groux-Degroote S. 2019. Identification of 9-O-acetyl-N-acetylneuraminic acid (Neu5,9Ac2) as main O-acetylated sialic acid species of GD2 in breast cancer cells. Glycoconjugate Journal.

24. de Graaf M, Fouchier R a M. 2014. Role of receptor binding specificity in influenza A virus transmission and pathogenesis. The EMBO journal 33:823–41.

25. Velkov T. 2013. The specificity of the influenza B virus hemagglutinin receptor binding pocket: What does it bind to? Journal of Molecular Recognition 26:439–449.

26. Cohen M, Zhang X-Q, Senaati HP, Chen H-W, Varki NM, Schooley RT, Gagneux P. 2013. Influenza A penetrates host mucus by cleaving sialic acids with neuraminidase. Virology journal 10:321.

27. Gamblin SJ, Skehel JJ. 2010. Influenza hemagglutinin and neuraminidase membrane glycoproteins. J Biol Chem 285:28403–28409.

28. Martin LT, Verhagen a, Varki a. 2003. Recombinant influenza C hemagglutinin-esterase as a probe for sialic acid 9-O-acetylation. Methods Enzymol 363:489–498.

29. Su S, Fu X, Li G, Kerlin F, Veit M. 2017. Novel Influenza D virus: Epidemiology, pathology, evolution and biological characteristics. Virulence 8:1580–1591.

30. Wang M, Veit M. 2016. Hemagglutinin-esterase-fusion (HEF) protein of influenza C virus. Protein & Cell 7:28–45.

31. Song H, Qi J, Khedri Z, Diaz S, Yu H, Chen X, Varki A, Shi Y, Gao GF. 2016. An Open Receptor-Binding Cavity of Hemagglutinin-Esterase-Fusion Glycoprotein from Newly-Identified Influenza D Virus: Basis for Its Broad Cell Tropism. PLOS Pathogens 12:e1005411.

32. Muñoz-Barroso I, García-Sastre a, Villar E, Manuguerra JC, Hannoun C, Cabezas J a. 1992. Increased influenza A virus sialidase activity with N-acetyl-9-O-acetylneuraminic acid-containing substrates resulting from influenza C virus O-acetylesterase action. Virus research 25:145–153.

33. Higa H, Rogers G, Paulson C. 1985. Influenza virus hemagglutinins differentiate between receptor determinants bearing N-acetyl-, N-glycollyl-, and N,O-diacetylneuraminic acids. Virology 144:279–282.

34. Lin Y-C, Boone M, Meuris L, Lemmens I, Van Roy N, Soete A, Reumers J, Moisse M, Plaisance S, Drmanac R, Chen J, Speleman F, Lambrechts D, Van de Peer Y, Tavernier J, Callewaert N. 2014. Genome dynamics of the human embryonic kidney 293 lineage in response to cell biology manipulations. Nat Commun 5:4767.

35. Dukes JD, Whitley P, Chalmers AD. 2011. The MDCK variety pack: choosing the right strain. BMC Cell Biology 12:43.

36. Nakazato Y, Suzuki H, Saruta T. 1989. Characterization of subclones of Madin-Darby canine kidney renal epithelial cell line. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research 1014:57–65.

37. Nichols GE, Lovejoy JC, Borgman CA, Sanders JM, Young WW. 1986. Isolation and characterization of two types of MDCK epithelial cell clones based on glycosphingolipid pattern. Biochim Biophys Acta 887:1–12.

38. Dumermuth E, Beuret N, Spiess M, Crottet P. 2002. Ubiquitous 9-O-acetylation of sialoglycoproteins restricted to the Golgi complex. J Biol Chem 277:18687–18693.

39. Varki A, Diaz S. 1984. The release and purification of sialic acids from glycoconjugates: Methods to minimize the loss and migration of O-acetyl groups. Analytical Biochemistry 137:236–247.

40. Berger JT, Voynow JA, Peters KW, Rose MC. 1999. Respiratory carcinoma cell lines. MUC genes and glycoconjugates. Am J Respir Cell Mol Biol 20:500–510.

41. Zanin M, Baviskar P, Webster R, Webby R. 2016. The Interaction between Respiratory Pathogens and Mucus. Cell Host & Microbe 19:159–168.

42. Casado B, Pannell LK, Iadarola P, Baraniuk JN. 2005. Identification of human nasal mucous proteins using proteomics. Proteomics 5:2949–2959.

43. Joo NS, Evans IAT, Cho H-J, Park I-H, Engelhardt JF, Wine JJ. 2015. Proteomic Analysis of Pure Human Airway Gland Mucus Reveals a Large Component of Protective Proteins. PLOS ONE 10:e0116756.

44. Zimmer G, Klenk H-D, Herrler G. 1995. Identification of a 40-kDa Cell Surface Sialoglycoprotein with the Characteristics of a Major Influenza C Virus Receptor in a Madin-Darby Canine Kidney Cell Line. Journal of Biological Chemistry 270:17815–17822.

45. Zimmer G, Lottspeich F, Maisner A, Klenk HD, Herrler G. 1997. Molecular characterization of gp40, a mucin-type glycoprotein from the apical plasma membrane of Madin-Darby canine kidney cells (type I). Biochem J 326 (Pt 1):99–108.

46. Sateesh Peri, Asmita Kulkarni, Felix Feyertag, Patricia M Berninsone, David Alvarez-Ponce. 2017. Phylogenetic distribution of CMP-Neu5Ac hydroxylase (CMAH), the enzyme synthetizing the pro-inflammatory human xeno-antigen Neu5Gc. Genome Biology and Evolution evx251.

47. Bardor M, Nguyen DH, Diaz S, Varki A. 2005. Mechanism of uptake and incorporation of the non-human sialic acid N-glycolylneuraminic acid into human cells. Journal of Biological Chemistry 280:4228–4237.

48. Thompson CM, Petiot E, Mullick A, Aucoin MG, Henry O, Kamen AA. 2015. Critical assessment of influenza VLP production in Sf9 and HEK293 expression systems. BMC Biotechnology 15.

49. Varki A, Hooshmand F, Diaz S, Varki NM, Hedrick SM. 1991. Developmental abnormalities in transgenic mice expressing a sialic acid-specific {9-O-acetylesterase.}. Cell 65:65–74.

50. Hulswit RJG, Lang Y, Bakkers MJG, Li W, Li Z, Schouten A, Ophorst B, van Kuppeveld FJM, Boons G-J, Bosch B-J, Huizinga EG, de Groot RJ. 2019. Human coronaviruses OC43 and HKU1 bind to 9- *O* -acetylated sialic acids via a conserved receptor-binding site in spike protein domain A. Proceedings of the National Academy of Sciences 116:2681–2690.

51. Corfield AP, Wagner SA, Safe A, Mountford RA, Clamp JR, Kamerling JP, Vliegenthart JF, Schauer R. 1993. Sialic acids in human gastric aspirates: detection of 9-O-lactyl- and 9-O-acetyl-N-acetylneuraminic acids and a decrease in total sialic acid concentration with age. Clin Sci 84:573–579.

52. Robinson LS, Lewis WG, Lewis AL. 2017. The sialate *O* -acetylesterase EstA from gut *Bacteroidetes* species enables sialidase-mediated cross-species foraging of 9- *O* -acetylated sialoglycans. Journal of Biological Chemistry 292:11861–11872.

53. Corfield AP, Donapaty SR, Carrington SD, Hicks SJ, Schauer R, Kohla G. 2005. Identification of 9-O-acetyl-N-acetylneuraminic acid in normal canine pre-ocular tear film secreted mucins and its depletion in Keratoconjunctivitis sicca. Glycoconj J 22:409–416.

54. Barker G, Simmons NL. 1981. Identification of two strains of cultured canine renal epithelial cells (MDCK cells) which display entirely different physiological properties. Quarterly Journal of Experimental Physiology 66:61–72.

55. Nakazato Y, Suzuki H, Saruta T. 1989. Characterization of subclones of Madin-Darby canine kidney renal epithelial cell line. Biochimica et Biophysica Acta (BBA) - Molecular Cell Research 1014:57–65.

56. Hoffmann E, Neumann G, Kawaoka Y, Hobom G, Webster RG. 2000. A DNA transfection system for generation of influenza A virus from eight plasmids. Proceedings of the National Academy of Sciences of the United States of America 97:6108–6113.

57. Hoffmann E, Krauss S, Perez D, Webby R, Webster RG. 2002. Eight-plasmid system for rapid generation of influenza virus vaccines. Vaccine 20:3165–3170.

CHAPTER THREE

Modified sialic acids on mucus and erythrocytes inhibit influenza A HA and NA functions

Karen N. Barnard, Brynn K. Alford-Lawrence, David W. Buchholz, Brian R. Wasik, Justin R. LaClair, Hai Yu, Rebekah Honce, Stefan Ruhl, Petar Pajic, Erin K. Daugherity, Xi Chen, Stacey L. Schultz-Cherry, Hector C. Aguilar, Ajit Varki, Colin R. Parrish. 2019. Modified sialic acids on mucus and erythrocytes inhibit influenza A HA and NA functions. *Journal of Virology. In revision.*

3.1 ABSTRACT

Sialic acids (Sia) are the primary receptors for influenza viruses, and are widely displayed on cell surfaces and in secreted mucus. Sia may be present in variant forms that include *O*-acetyl modifications at C4, C7, C8, and C9 positions, and *N*-acetyl or *N*-glycolyl at C5. They can also vary in their linkages, including α2-3 or α2-6-linkages. Here, we analyzed the distribution of modified Sia in cells and tissues of wild-type mice, or in mice lacking the cytidine 5'-monophosphate-*N*-acetylneuraminic acid hydroxylase (CMAH) enzyme that synthesizes *N*-glycolyl modifications (Neu5Gc). We also examined the variation of Sia forms on erythrocytes and saliva from different animals. To determine the effect of Sia modifications on influenza A virus (IAV) infection, we tested for effects on hemagglutinin (HA) binding and neuraminidase (NA) cleavage. We confirmed that 9-*O*-acetyl, 7,9-*O*-acetyl, 4-*O*-acetyl, and Neu5Gc modifications are widely but variably expressed in mouse tissues, with the highest levels detected in the respiratory and gastrointestinal tracts. Secreted mucins in saliva and surface proteins of erythrocytes showed a great degree of variability in display of modified Sia between different species. IAV HA from different virus strains showed consistently reduced binding to both Neu5Gc and *O*-acetyl modified Sia; however, while IAV NA were inhibited by Neu5Gc and *O*-acetyl modifications, there was significant variability between NA types. The modifications of Sia in mucus may therefore have potent effects on the functions of IAV, and may affect both pathogens and the normal flora of different mucosal sites.

3.2 IMPORTANCE

Sialic acids (Sia) are involved in many different cellular functions and are receptors for many pathogens. Sia come in many chemically modified forms but we lack a clear understanding of how they alter the interactions with microbes. Here we examine the expression of modified Sia

in mouse tissues, on secreted mucus in saliva, and on erythrocytes, including those from IAV host species and animals used in IAV research. These Sia forms varied considerably between different animals, and their inhibitory effects on IAV NA and HA activities and on bacterial sialidases (neuraminidases) suggest a host-variable protective role in secreted mucus.

3.3 INTRODUCTION

Sialic acids (Sia) are a family of nine-carbon monosaccharides that often serve as terminal residues of carbohydrate chains. They are present at high levels on cell membrane glycoproteins and glycolipids, as well as on secreted glycoproteins and mucus at all mucosal surfaces (**Fig. 3.1A**) (1, 2). Sia are key mediators of many normal cell and tissue functions through a wide variety of highly regulated cell-cell interactions during both development and homeostatic processes, where they may be bound by cellular receptors and members of the selectin family (3, 4). Their ubiquitous presence on cells, tissues, and mucosal surfaces also make Sia a key point of contact for commensal microbes and for invading pathogens including viruses, bacteria, and parasites (3, 5, 6).

Sia are a highly diverse family of molecules that may be present as more than 50 structurally and chemically distinct modified variants. These are formed from the basic structure of the *N*-acetylneuraminic acid (Neu5Ac) by the addition of chemical groups at various positions on the pyranose ring or the glycerol side chain. Those modifications may include *N*-glycolyl and/or *O*-linked acetyl, sulfo, methyl, and lactyl groups, among others (1, 2, 7). Many different enzymes and pathways introduce these chemical modifications and some can be removed by regulatory enzymes. The different modified Sia are often themselves substrates for modifying enzymes and transferases, so that each modification may alter the synthesis of other modified forms. This therefore leads to complex patterns and mixtures of modified Sia forms, with

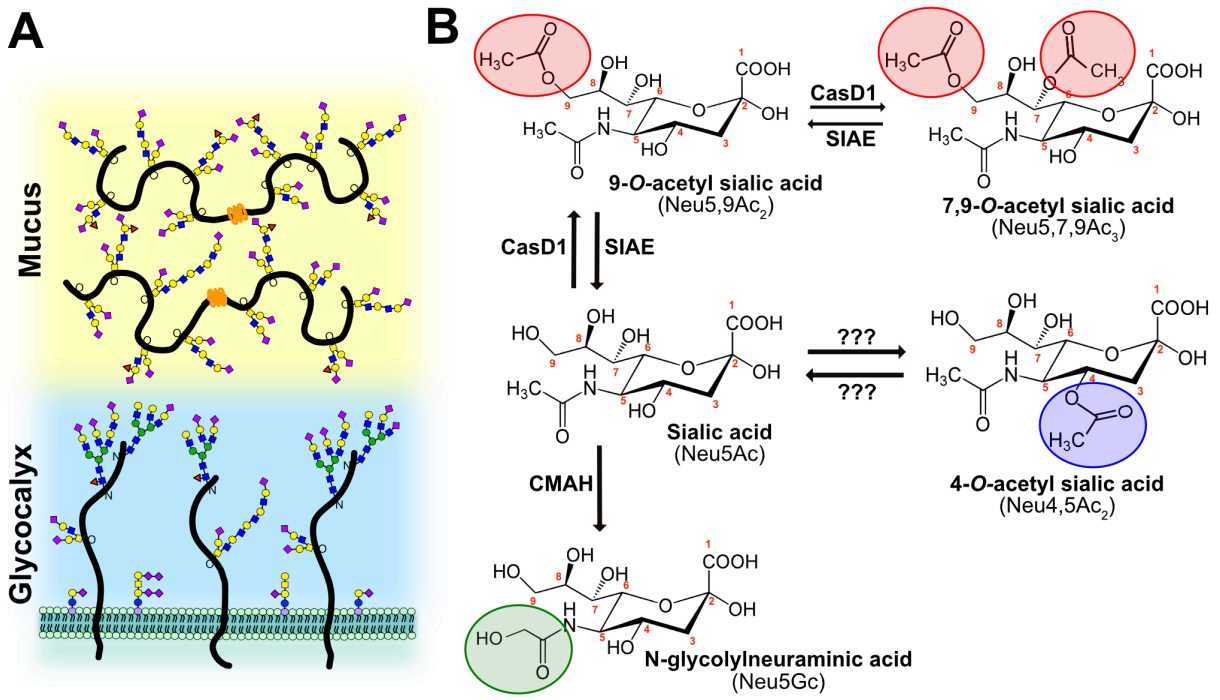


Figure 3.1

A) Sialic acids (purple diamonds) terminate glycan chains on glycolipids and glycoproteins as part of the glycocalyx on the surface of cells. They can also terminate glycans on secreted glycoproteins, like mucins, that are an important component of the protective mucosal barrier in gastrointestinal and respiratory tissue. **B**) Sialic acid (*N*-acetylneuraminic acid, Neu5Ac) can be modified by the addition of O-acetyl modifications at the C-4, 7, and 9 positions, or by the hydroxylation of the *N*-acetyl group at C-5 to form *N*-glycolylneuraminic acid (Neu5Gc) by the enzyme cytidine 5'-monophosphate-N-acetylneuraminic acid hydroxylase (CMAH). The sialate *O*-acetyltransferase, CasD1, adds acetyl groups at C-7 from which it migrates to the C-9 position ($Neu5,9Ac_2$) under physiological conditions. This can allow for an additional acetyl group to be added by CasD1 to C-7 ($Neu5,7,9Ac_3$). The sialate *O*-acetylesterase, SIAE, can remove these acetyl modifications, restoring the unmodified Neu5Ac form of sialic acid. *O*-acetyl modifications can also be added at the C-4 position by a specific 4-*O*-acetyltransferase ($Neu4,5Ac_2$) and removed by a 4-*O*-acetylesterase. However, the genes for these enzymes have not yet been identified.

significant variation in both the levels and specific combinations of modifications in different hosts, tissues, and under different physiological conditions (4, 8–11). The role of these modified Sia forms on cellular functions and interactions with microbes has been understudied due to a lack of reagents. The recent development of specific probes for some Sia variants (12, 13) combined with existing methods for studying glycans has allowed us to finally determine the expression of modified Sia between species and tissues, as well as beginning to unravel their effects on host:pathogen interactions.

Sialic acid modifications. While there exist in nature over 50 different chemical forms of Sia, the most common chemical additions seen in vertebrates include ester-bonded *O*-acetyl (*O*-Ac) modifications to C-4, 7, 8, and/or 9 positions, resulting in a variety of combinations of Sia forms including Neu4,5Ac_2, Neu5,9Ac_2, Neu5,7,9Ac_3 Sia, as well as their *N*-glycolylneuraminic acid (Neu5Gc) analogs with the same *O*-acetyl modifications (**Fig. 3.1B**). Neu5Gc is produced from Neu5Ac by the activity of cytidine 5'-monophosphate-N-acetylneuraminic acid hydroxylase (CMAH) in the cytoplasm of cells, and this enzyme is missing or inactive in some animals, including humans (9). The addition of *O*-Ac to the C7 and/or C9 positions is mediated by the sialylate *O*-acetyltransferase enzyme, Cas1 domain containing 1 (CasD1), which has been suggested to add an *O*-acetyl group to the C7 position, from which it would migrate to the C8 and C9 position under physiological conditions, allowing the possibility of the addition of another *O*-acetyl group to C7 (7, 14, 15). The regulatory processes that control the number or positions of acetyl groups have not been well defined, although distinct differences in expression of 7,9-*O*-Ac and 9-*O*-Ac Sia have been reported in mouse and human tissues, chicken embryos, and in some other animals (7). CasD1 uses acetyl-CoA to modify Sia in the activated CMP-Sia form before it is added to the glycan chain, and likely has a preference for CMP-Neu5Ac as a

substrate and is less active on CMP-Neu5Gc (15). The sialate *O*-acetylesterase (SIAE) enzyme can remove the 7,9-*O*- and 9-*O*-Ac modifications, although its activities and roles are not well understood (16–19). The 4-*O*-Ac Sia is produced in some tissues of many animals by a distinct sialate 4-*O*-acetyltransferase that is also likely expressed in the Golgi compartment; however, the gene for this enzyme has also not yet been identified (20–22).

The 7, 8, and/or 9-*O*-Ac Sia appear to be present at low levels - a few percent or less - in the cell-associated Sia on many cultured cells, but may be present at higher levels (10 to 50%) in Sia on the secreted mucus of various animals and on erythrocyte-associated glycans (13, 23). However, the expression, distribution, and regulations of these modified Sia are not well documented, nor do we understand their impact on pathogens, host homeostasis, and normal microbiota at different mucosal sites (12, 13, 24, 25).

It is important to remember that while *O*-acetyl and Neu5Gc are the most common chemically modified forms of Sia and the focus of this study, there is a greater diversity of Sia that exist in nature. These other chemically modified forms, including *O*-sulfo, *O*-lactyl, and *O*-methyl modifications, likely play important roles in development and cell-cell signaling (1). Both *O*-sulfo and *O*-methyl modifications have been detected in invertebrate species such as sea stars and sea urchins, as well as some mouse tissues (26–28). *O*-lactyl Sia has been detected in human gastric aspirate along with *O*-acetyl modified Sia (29). However, similar to *O*-acetyl and Neu5Gc modifications, these other Sia variants have also been understudied due to a lack of reagents; thus, little is known about their specific regulation, expression, and function in different animal species and tissues.

Modified sialic acid and pathogen interactions. Many pathogens interact with Sia on host cells at various stages in their infection cycles, including various viruses, bacteria, and

parasites (3, 5). The densely expressed Sia within various mucus layers on mucosal surfaces also act to bind incoming pathogens and likely regulate both the release and transmission of pathogens (30, 31). Many pathogens therefore express proteins that attach to Sia, as well as expressing receptor-modifying enzymes such as sialidases (neuraminidases) that remove the Sia from the underlying glycan. Bacterial adhesins and toxins may recognize Sia on the surface of cells, and many bacteria can also use Sia as a metabolic carbon source after release through the activity of neuraminidases, and uptake into the cell by Sia transporters (3, 32–35). These bacterial interactions with Sia are potentially affected by chemical modifications (34, 36). Both enveloped and non-enveloped viruses may also bind Sia as primary receptors or co-receptors for cell recognition and infection, although only the enveloped viruses appear to express neuraminidases or sialate *O*-acetyl esterases, possibly to reduce aggregation of viral particles during budding (5, 37). For some viruses, Sia modifications are required for infection as viral proteins specifically bind to modified Sia – examples include human coronavirus OC43 and HKU1, and influenza C and D viruses, which all require 9-*O*-Ac Sia for cell infection (38, 39).

Significant effects of different Sia modifications on the binding of pathogens or the activities of their sialidases (neuraminidases) have been suggested, but in general these are still not well understood. Influenza A viruses (IAV) use Sia as primary receptors for host recognition and cell entry through the activity of two surface glycoproteins that interact with Sia, hemagglutinin (HA) and neuraminidase (NA). HA is a trimeric protein that binds Sia to initiate the endocytic uptake of the virus by the cell, leading to fusion between the viral envelope and the endosomal membrane after exposure of the virus to low pH (40). NA is a sialidase which cleaves Sia from the mucus, cell surface, and from viral glycoproteins, allowing the virus to penetrate through mucus to the epithelial cells and reducing the aggregation of virions after budding from

the surface of cells (41). Previous studies have shown that modifications such as 7,9-*O*- and 9-*O*-Ac Sia are expressed on cells or in tissues of many IAV host species and there is some evidence that these modifications may be inhibitory for NA activity and HA binding (42, 43). However, the specific effects of 7,9-*O*- and 9-*O*-Ac on the binding of Sia by HA or the cleavage of Sia by IAV NA have not been examined in detail, and it is unclear whether these changes influence infection efficiency or viral shedding. For example, modification of HA binding might influence the attachment of virus to cells or to mucus, while inhibition of NA cleavage of *O*-acetyl modified Sia may lead to virus being trapped in the mucus and cleared, reducing the efficiency of infection.

The difference between Neu5Gc and Neu5Ac has been found to influence the tropism of several different viruses, as well as some bacterial toxins (3, 5). Indeed, it has been proposed that the loss of the *CMAH* gene in humans was an adaptive response to pathogen pressures (44–46). Neu5Gc is highly expressed in some tissues of IAV natural host species, including pigs and horses, and is also present in the tissues of mice and guinea pigs, which are frequently used as animals models (9, 47). Neu5Gc has been seen to prevent binding of the HAs of some IAV, particularly in human-adapted strains (43, 48), but the effects on NA have not been well characterized. Nevertheless, examination of swine IAV isolates found distinct strain differences in their ability to cleave Neu5Gc by sialidase activity which were generally lower than against Neu5Ac (49).

In this study, we define the expression of 7,9-*O*-Ac, 9-*O*-Ac, and Neu5Gc modified Sia in the mucus, saliva, and on erythrocytes of different IAV host animals, as an example to highlight the variability in expression of different Sia modifications between species. We also examine the display of modified Sia on the tissues and secreted mucus of mice, an important model species

not only for IAV research, but many viruses and bacterial pathogens. Finally, we test the effects of these modifications on HA binding and NA activity of different strains of IAV, as well as their potential to alter virus infection. While we have focused on IAV in these studies, defining the expression of modified Sia between different hosts and in mice has importance to many viruses that infect mucosal tissues. Therefore the findings here are broadly relevant and underline the need for understanding the expression of modified Sia between species and their potential role in virus tropism and infection.

3.4 RESULTS.

Distribution of modified Sia in mouse tissues. Previous research on display of modified Sia in animal tissues and cultured cells have shown varying distributions of 7,9-*O*-Ac, 9-*O*-Ac, and 4-*O*-Ac Sia depending on the animal and the tissue examined (12, 13). Mice are an important model species for biomedical research, and some tissues have previously been screened for *O*-Ac display using probes derived from viral glycoproteins (virolectins) and by other methods (12, 13). We examined the distribution of modified Sia in a variety of tissues of wild-type (WT) C57/BL6 mice. Both 9-*O*- and 7,9-*O*-Ac were found throughout the lung and trachea as well as in the tracheal sub-mucosal glands that produce most of the mucus (**Fig. 3.2A**). These modified Sia were also found throughout the GI tract, with staining associated with epithelial cells, goblet cells, and associated mucus layers of the gastrointestinal tissues, including the stomach, small intestine, and colon (**Fig. 3.2B**). Interestingly, 9-*O*-Ac appeared to be present in higher amounts in most tissues, including salivary gland and esophagus, while the 7,9-*O*-Ac staining was minimal. However, 7,9-*O*-Ac did stain stronger than 9-*O*-Ac in stomach-associated mucus, and on tracheal epithelial cells. This seems to indicate that while the same enzymes (CasD1 and SIAE) are considered to control the presence of these modifications, the expression of 9-*O*- and

7,9-*O*-Ac are differentially regulated in individual tissues. The 4-*O*-Ac Sia showed high levels of probe binding in the colon, primarily in the mucus, and there was also some expression on the mucosal surfaces in the stomach, small intestine (jejunum) and trachea, as well as on cells within the red pulp of the spleen (**Fig. 3.2A, B; Fig. 3.3**).

The virolectin probes used are sensitive, but did not reveal the quantity of each of the modified Sia forms present. To determine the relative amounts of different Sia forms, we tested tissues from mice using 1,2-diamino-4,5-methylenedioxybenzene (DMB) labeling of the Sia and analysis with high performance liquid chromatography (HPLC), under conditions that reveal the amounts of Neu5Ac and Neu5Gc, and preserve most of the *O*-acetylation of the Sia (50). The WT mice showed varying levels of Neu5Gc (**Fig. 3.4A; Table 3.1**), while, as expected, the $CMAH^{-/-}$ mice showed only Neu5Ac in all tissues, similar to amounts reported previously for some of those tissues (51) (**Table 3.2**). The lack of Neu5Gc in $CMAH^{-/-}$ mice was also seen in GI tissues, indicating that Neu5Gc from dietary sources was not detectably being taken up by these mice, as is seen in humans who eat a diet containing that Neu5Gc (52). In the WT mice, tissues showed a great deal of variability in Neu5Gc expression. Most tissues had around 50–60% Neu5Gc; however, some tissues, including the liver, had higher levels of over 70%, while the brain and salivary glands had only 10%. All *O*-acetyl Sia variants combined comprised between 2 and 16% of the Sia in most tissues, with the majority being 9-*O*-Ac (1–9% of total Sia) (**Fig. 3.4B; Table 3.1**). There was about 1.5 to 3 fold higher levels of *O*-acetyl Sia in most tissues of the CMAH knock-out mice (**Fig. 3.4C; Table 3.2**), as has been reported previously (51). The levels of 4-*O*-Ac Sia were generally low, making up ~2% of the Sia in small intestine (duodenum) and ~1% in spleen, testes, and esophagus. Mouse colon samples showed the highest levels of total *O*-acetylation, with ~17% of Sia having one or more *O*-acetyl modification, again

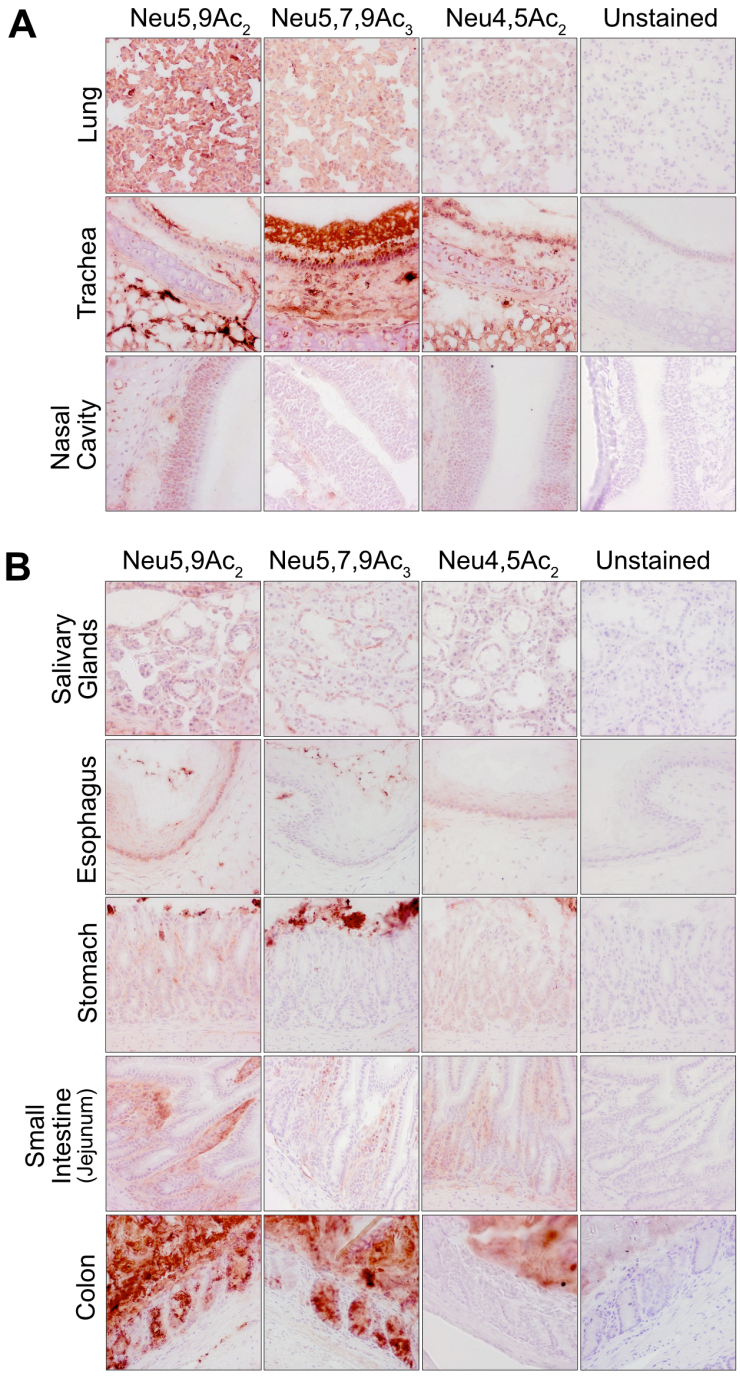


Figure 3.2

Expression of *O*-acetylated Sia varies between tissues in wild-type C57BL/6 mice. Frozen tissue sections from respiratory tissues (**A**) and gastrointestinal tissues (**B**) were stained using virolectins derived from the hemagglutinin esterases (HE-Fc) of various nidoviruses with high specificity for the different *O*-acetyl modified Sia forms. Sections were counterstained with hematoxylin and imaged at 40x magnification.

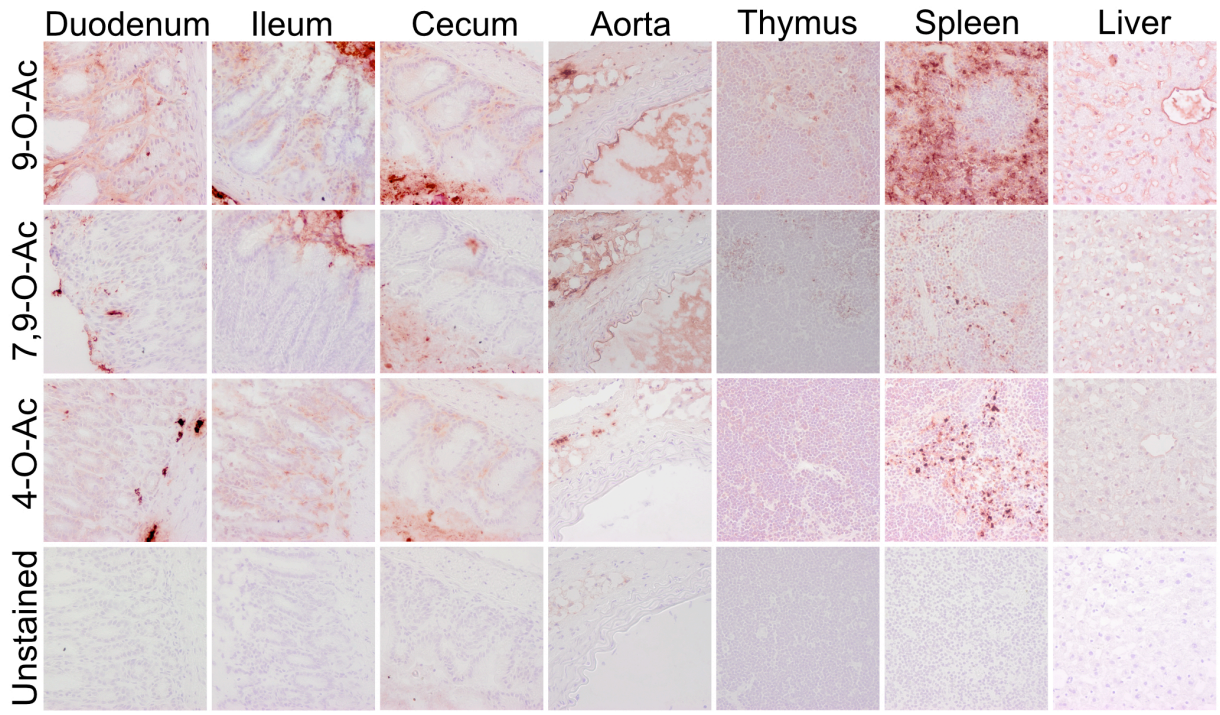


Figure 3.3

Expression of *O*-acetylated Sia varies between tissues in wild-type C57BL/6 mice
Frozen tissue sections were stained using virolectins derived from the hemagglutinin esterases (HE-Fc) of various nidoviruses with high specificity for the different *O*-acetyl modified Sia forms. Sections were counterstained with hematoxylin and imaged at 40x magnification**.**

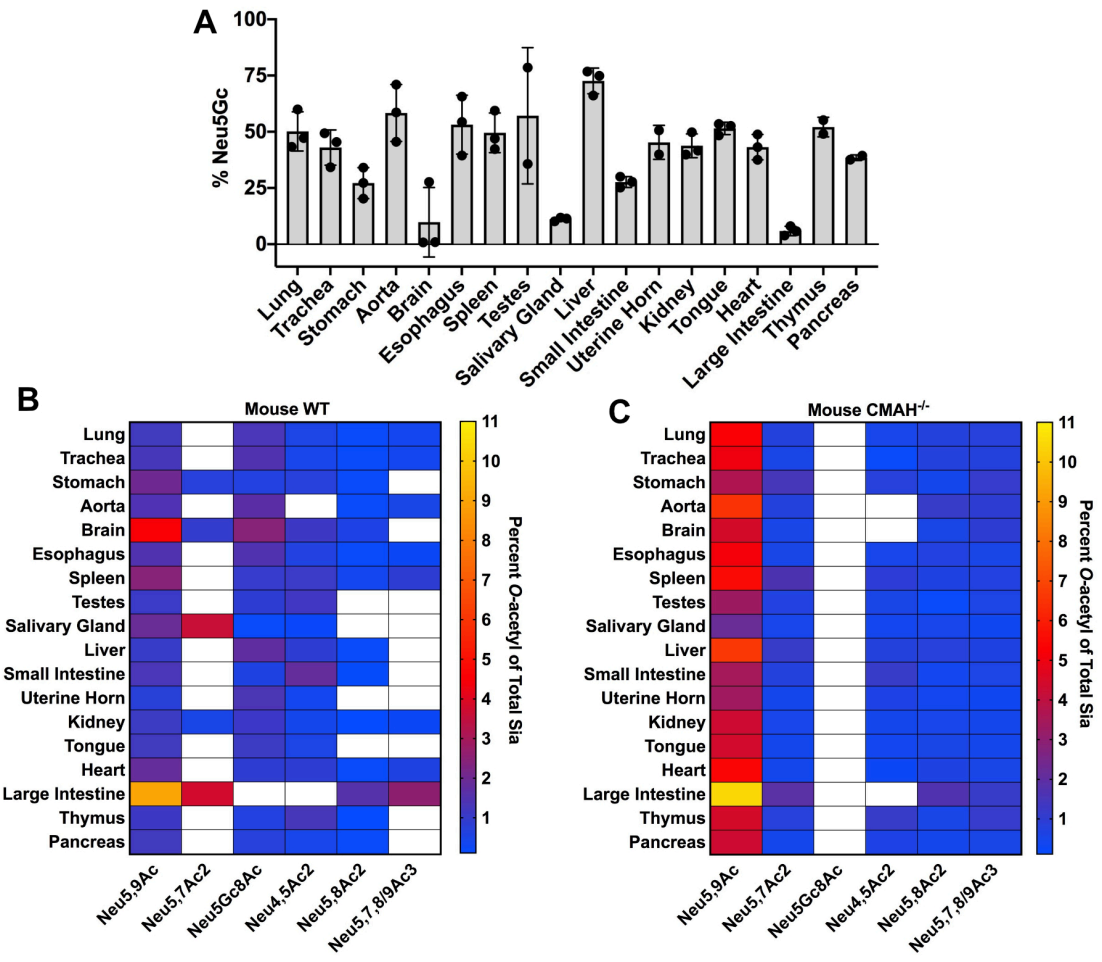


Figure 3.4

Neu5Gc and *O*-acetyl Sia modifications vary by tissue in wild-type C57BL/6 mice, and absence of Neu5Gc in CMAH$^{-/-}$ mice leads to an increase in *O*-acetylation across tissues. Total Sia was measured from tissue samples using HPLC analysis to determine relative Sia quantities. (**A**) Neu5Gc levels were measured across tissues in wild-type (WT) mice, showing highly specific expression. CMAH$^{-/-}$ had undetectable levels of Neu5Gc. (**B, C**) Percent *O*-acetyl modified Sia in different tissues from WT (**B**) and CMAH $^{-/-}$ (**C**) mice are given as a heat map showing variation across tissues. White squares indicate when a Sia form is below detection. Values are given as the percentage of total Sia collected from tissue samples. Sample size for each tissue was three individual mice (n=3) of each mouse strain with average values for total sialic acid content given in **Tables 3.1 and 3.2**.

Table 3.1

Average relative Sia quantities as determined by HPLC analysis for each tissue tested in WT C57BL/6 mice (n=3). Table shows the proportion of total Sia for each variant given as a percentage, with the sum of all *O*-acetyl forms combined given in the far right column. If a Sia form was below detection, this is indicated by b/d

Tissue	Neu5Gc	Neu5Ac	Neu5,9Ac	Neu5,7Ac$_2$	Neu5Gc8Ac	Neu4,5Ac$_2$	Neu5,8Ac$_2$	Neu5,7,8/9Ac$_3$	O-Ac sum
Lung	50.2	47.6	1.2	b/d	1.3	0.5	0.2	0.4	3.6
Trachea	43.0	53.5	1.3	b/d	1.5	0.4	0.2	0.4	3.7
Stomach	27.2	68.8	2.0	0.7	0.6	0.7	0.2	b/d	4.3
Aorta	58.4	38.2	1.4	b/d	1.7	b/d	0.2	b/d	3.3
Brain	9.8	83.1	4.5	1.0	2.5	1.1	0.5	b/d	9.5
Esophagus	53.1	43.2	1.5	b/d	1.4	0.6	0.2	0.2	3.9
Spleen	49.6	45.2	2.4	b/d	1.0	1.1	0.3	0.8	5.6
Testes	57.1	40.4	1.1	b/d	0.8	1.2	b/d	b/d	3.0
Salivary Gland	11.1	85.5	1.9	3.6	0.2	0.2	0.1	b/d	6.0
Liver	72.6	23.8	1.0	b/d	1.7	0.8	0.2	b/d	3.7
Small Intestine (duodenum)	27.6	68.7	1.3	b/d	0.6	1.8	0.1	b/d	3.8
Uterine Horn	45.3	52.4	0.8	b/d	1.4	0.4	0.1	b/d	2.6
Kidney	43.8	53.5	1.1	0.5	1.0	0.4	0.1	0.2	3.4
Tongue	51.5	45.8	1.2	b/d	1.1	0.5	n/d	b/d	2.8
Heart	43.2	53.3	1.8	b/d	0.9	0.8	0.2	0.6	4.3
Large Intestine (colon)	5.9	77.2	9.0	3.8	b/d	b/d	1.6	2.6	16.9
Thymus	52.1	44.8	1.1	b/d	0.6	1.3	0.1	b/d	3.1
Pancreas	38.5	59.1	1.2	b/d	0.7	0.4	0.2	b/d	2.5

.

Table 3.2

Average relative Sia quantities as determined by HPLC analysis for each tissue tested in $CMAH^{-/-}$ C57BL/6 mice (n=3). Table shows the proportion of total Sia for each variant given as a percentage, with the sum of all *O*-acetyl forms combined given in the far right column. If a Sia form was below detection, this is indicated by b/d.

Tissue	Neu5Gc	Neu5Ac	Neu5,9Ac	Neu5,7Ac$_2$	Neu5Gc8Ac	Neu4,5Ac$_2$	Neu5,8Ac$_2$	Neu5,7,8/9Ac$_3$	O-Ac sum
Lung	b/d	92.7	5.6	1.5	b/d	0.5	0.8	1.0	9.4
Trachea	b/d	93.4	5.8	1.1	b/d	0.2	0.7	0.7	8.5
Stomach	b/d	92.6	3.4	4.5	b/d	0.8	0.4	0.9	10.0
Aorta	b/d	90.5	5.4	2.3	b/d	b/d	1.2	1.1	10.0
Brain	b/d	94.6	5.1	1.1	b/d	b/d	0.6	b/d	6.7
Esophagus	b/d	92.8	4.4	1.5	b/d	0.5	0.7	0.6	7.6
Spleen	b/d	90.4	4.9	4.9	b/d	0.9	0.7	0.9	12.2
Testes	b/d	95.7	5.2	0.7	b/d	0.6	0.2	b/d	6.7
Salivary Gland	b/d	96.5	2.0	1.5	b/d	0.4	0.5	0.3	4.7
Liver	b/d	90.1	5.3	3.6	b/d	0.8	0.8	0.7	11.2
Small Intestine (duodenum)	b/d	94.3	3.8	2.2	b/d	1.2	0.3	0.4	8.0
Uterine Horn	b/d	94.9	4.0	0.9	b/d	0.6	0.4	0.3	6.2
Kidney	b/d	93.8	5.0	1.5	b/d	0.4	0.5	0.5	7.9
Tongue	b/d	94.2	4.3	1.1	b/d	0.4	0.6	0.4	6.7
Heart	b/d	92.7	4.7	1.1	b/d	0.3	0.6	0.6	7.3
Large Intestine (colon)	b/d	84.2	10.5	5.7	b/d	0.1	1.7	1.8	19.8
Thymus	b/d	92.6	4.3	2.4	b/d	1.2	0.5	0.4	8.8
Pancreas	b/d	94.1	3.9	1.3	b/d	0.6	0.5	0.5	6.8

primarily 9-*O*-Ac. Given the patterns seen using the HE-Fc virolectin staining, the 7,9-*O*- and 9-*O*-Ac forms must be present at high levels within or on certain cell sub-populations, as well as within mucus or mucus-secreting cells. For example, the high levels of 7,9-*O*- and 9-*O*-Ac found in the mouse colon were most likely associated with secreted mucus as most of those modified Sia were present in goblet cells (**Fig. 3.2B**). However, the differences seen between probe binding and the ratios of the different modified Sia within the total Sia underscore the importance of quantifying the different forms.

Analysis of modified Sia in saliva, mucus, and on erythrocytes. It has been previously reported that human colonic mucin is highly enriched in 9-*O*-Ac Sia, which may regulate the activity of some sialidases and Sia transporters of the gut microflora (11, 36, 53, 54). Strong staining for 9-*O*-Ac in human respiratory tissues, and also within the submucosal glands of human respiratory tissue have also been reported, indicating that mucus from these glands could be enriched in *O*-acetylated Sia (53, 54). To determine if human respiratory mucus was enriched in 7,9-*O*- and 9-*O*-Ac, secreted mucus from primary normal human bronchial epithelial cells (NHBE) as well as conditioned media from human alveolar basal epithelial adenocarcinoma A549 cells, were analyzed by HPLC to determine Sia composition. We found that the secreted proteins in mucus from NHBE cells and A549 cells conditioned media contained primarily unmodified Neu5Ac with ~1–2% of 9-*O*-Ac and no detectable levels of 7,9-*O*-Ac (**Table 3.3**). This indicates that secreted mucus from these respiratory cells in culture are not enriched for *O*-acetyl modifications, which differs from previous reports for colonic mucin (11, 53).

To look more broadly at the possible range of modified Sia present in secreted mucus in different animals, we examined saliva from a number of influenza host species, including human, pig, horse, and dog (**Fig. 3.5A,B**). While the proteins in saliva differ from those found in

respiratory mucus, they do contain some of the same heavily glycosylated proteins including mucins like MUC5B (55–57). Human saliva was similar to the secreted mucus from NHBE and A549 cells in containing primarily Neu5Ac with little 9-*O*-Ac Sia, and the composition of dog saliva showed a similar profile. However, most other animals showed far more diversity in their Sia profiles, with both mice and horses having enrichment for several different *O*-acetyl modifications. Laboratory mouse saliva contained a combined ~17% *O*-acetylated Sia in the forms of 7-*O*-*, 8*-*O*-, and 9-*O*-Ac, while horse saliva contained ~10% 4-*O*-Ac as well as ~19% of other *O*-acetyl Sia variants combined. Pig saliva were unique among the IAV hosts examined in having ~90% of total Sia being of the Neu5Gc form. The diversity of modifications seen in mouse, horse, and pig saliva may have a strong influence on any Sia-binding pathogens, including influenza viruses, as well as on commensal bacterial communities in different species.

Erythrocytes (red blood cells, RBCs) express high levels of sialylated surface molecules, primarily on glycophorins, and are used in IAV research to study the interactions of HA binding specificity, determining viral titer through the hemagglutination assay, and inhibition of hemagglutination by antibodies (HAI assay) (58, 59). It has long been known that IAV varies in hemagglutination of RBCs from different species, at least in part due to differences in the Sia linkages present. The structures of HAs with Sia bound often suggest that modification of the C4, 5, 7, and/or 9 positions would influence IAV interactions with Sia. We found that chicken and guinea pig RBCs, which are often used to titer IAV virus and as the standard substrate for HAI assays, contained almost exclusively unmodified Neu5Ac, as did those from humans and dogs (**Fig. 3.5C,D**). In contrast, pig, horse, cow, and sheep RBCs contain high proportions of Neu5Gc, along with varying amounts of *O*-acetyl modifications. The high levels of Neu5Gc Sia present on the RBCs of these species had been previously reported, although not directly

Table 3.3

A549 conditioned media and collected mucus from NHBE cultures were analyzed for total sialic acid content using HPLC analysis. Table shows the proportion of total Sia for each variant given as a percentage, with the sum of all *O*-acetyl forms combined given in the far right column. If a Sia form was below detection, this is indicated by b/d. Percentages are an average of multiple conditioned media collections from A549 cells (n=4) and multiple NHBE donors (n=4).

Cells	Source	Neu5Gc	Neu5Ac	Neu5,9Ac	Neu5,7Ac$_2$	Neu5,8Ac$_2$	Neu5,7,8/9Ac$_3$
A549	Conditioned media*	b/d	98.16	1.84	b/d	b/d	b/d
NHBE	mucus	b/d	98.6	1.40	b/d	b/d	b/d

*: collected from A549 bronchial epithelial cells conditioned media, contains mucins

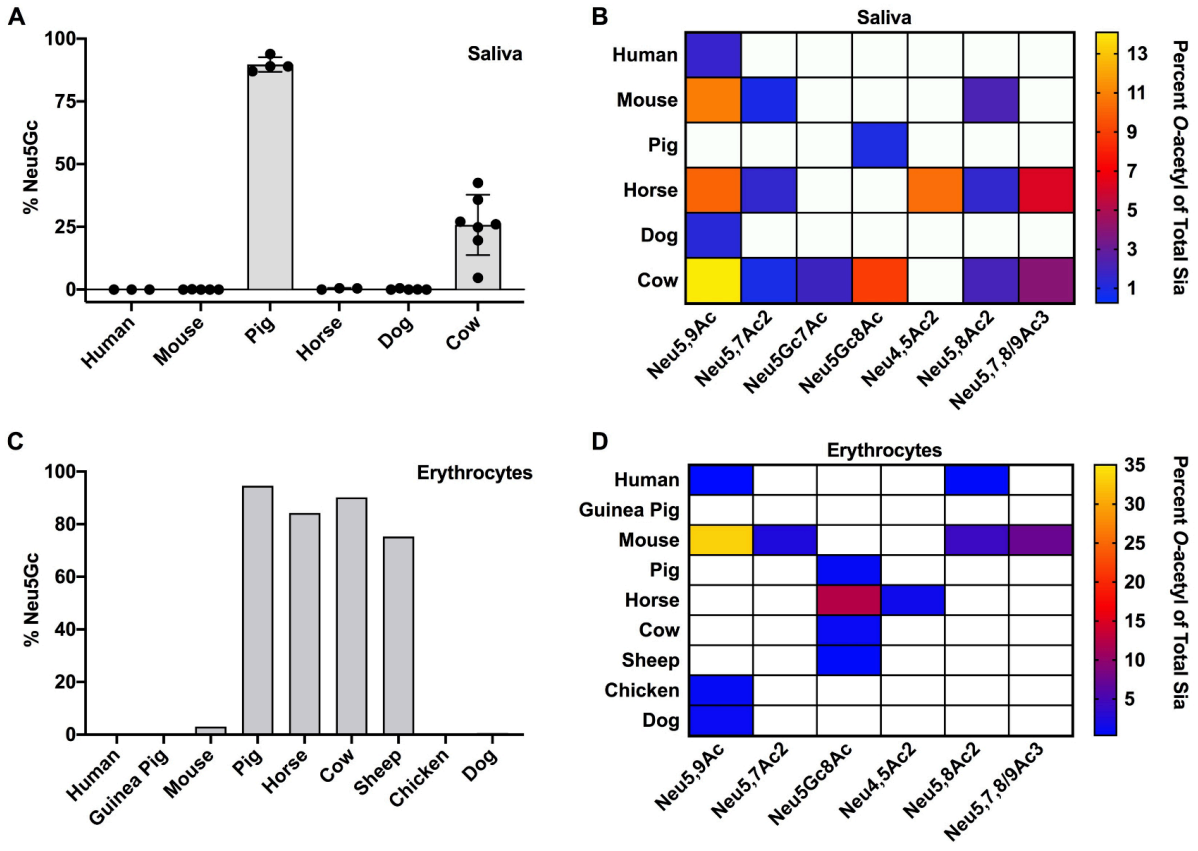


Figure 3.5

Total Sia of saliva (**A, B**) and erythrocytes (**C, D**) was collected via acid hydrolysis and analyzed using HPLC. *O*-Acetyl Sia percentages are given as a heat map with white squares indicating when a Sia form is below detection. Values are given as the percentage of total Sia collected from tissue samples. Saliva sample sizes (n = number of individuals of each species): human (n=3), mouse (n=5), pig (n=4), horse (n=3), dog (n=5), cow (n=7). Erythrocyte sample sizes (n = number of individuals of each species): human (n=2), guinea pig (n=1), mouse (n=2), pig (n=1), horse (n=1), cow (n=1), sheep (n=1), chicken (n=1), dog (n=2).

quantified as presented here (23, 60). The mouse erythrocytes tested here (from C57BL/6 mice) had the greatest diversity of modifications, with little Neu5Gc, around 50% Neu5Ac, and ~50% of the Sia modified by 7, 8 and/or 9-*O*-acetylation, as previously reported (44).

Effects of modified Sia on NA cleavage. Neuraminidases (sialidases) expressed by bacteria and viruses cleave Sia from oligosaccharides and glycoconjugates, and some have been shown to be affected by various Sia modifications (42, 48, 60–62). But little is known about the effects of modified Sia on IAV NAs from different strains. We examined the effects of Sia modifications on cleavage by several different IAV NAs, as well as on the activity of *Arthrobacter ureafaciens* neuraminidase (NeuA) which was included as a positive control for sialidase activity. Substrates used had high levels of 7,9-*O*- and 9-*O*-Ac Neu5Ac (bovine sub-maxillary mucin, BSM), Neu5Gc (horse RBCs), or unmodified Neu5Ac (chicken RBCs). IAV NA from a variety of different IAV strains (N1, N2, N3, N7, and N9) were expressed alone in cells (**Fig. 3.6A**) and recovered as purified VLPs (**Fig. 3.6B**) (63). VLPs are composed of cell membranes and NA glycoproteins only with very few host-derived surface proteins. The sources for the NA proteins were from both human isolates (N1, N7, and N9) and avian isolates (N2 and N3). These NA VLPs were first tested for enzymatic activity using a standard NA cleavage assay using methylumbelliferyl *N*-acetylneuraminide (MuNANA) as the substrate and a dilution of 1:100 was chosen for further cleavage analysis as this gave an equal and high level of enzymatic activity for all VLPs (**Fig. 3.6C**) (48). The NA-expressing VLPs were incubated with BSM or with RBCs, and the released Sia were collected and analyzed by HPLC. For BSM, HPLC profiles of total Sia were created to compare NA cleavage preferences (**Fig. 3.7A**). These profiles showed the Sia forms that were susceptible to NA cleavage and release while the non-released forms were considered to be resistant to NA. These HPLC profiles were then compared

to the total Sia released chemically by acid hydrolysis, an unbiased method that removes all Sia forms present in the original sample (50). All of the viral NA and the bacterial NeuA showed the highest level of cleavage for unmodified Neu5Ac compared to any of the modified forms, as more Neu5Ac was present in the released profiles compared to chemical release. There was substantial variation in the cleavage activity against the modified Sia by the different viral NAs. N1 and N7 showed the lowest activities against any modified Sia, N3 and N9 had intermediate activities, while N2 and NeuA were active against the greatest number of modified forms, and most closely matching the chemical release profile. There was lower activity for mono-*O*-acetylated Sia (7-*O*-, *8*-*O*-, and 9-*O*-Ac) by N1, N3, N7, and N9, while all NAs tested had lower activity against the di-*O*-acetylated Sia (7,8/9-*O*-Ac_2) and mono-*O*-acetylated Neu5Gc forms. All the viral and bacterial NAs apart from N2 had several-fold lower activities on Neu5Gc alone, as seen in the smaller proportion of that Sia form released compared to the chemical release profile.

To further test the ability of these NA VLPs to cleave Neu5Gc compared to Neu5Ac, the amounts of Sia released from either horse RBCs (84% Neu5Gc) or chicken RBCs (99% Neu5Ac) were compared. All NAs showed significantly lower levels of Neu5Gc Sia released from horse RBCs compared to the amounts of Neu5Ac released from chicken RBCs (**Fig. 3.7B**). When compared directly, NA VLPs removed 5–12% of Neu5Gc compared to their activities on Neu5Ac (**Fig. 3.7C**). Again, variability was seen between NA VLPs here, with N7 having the lowest activity against Neu5Gc compared to Neu5Ac and N9 having the most. There was also variability in cleavage activity between NA from different strains as well, as seen in the variable amount of Neu5Ac removed by the NA VLPs in **Figure 3.7B**, but it is unclear whether this difference is due to the intrinsic activities of the NAs when expressed as VLPs or to innate differences in the specific activities of each NA enzyme, or both. It is clear, however, that *O*-

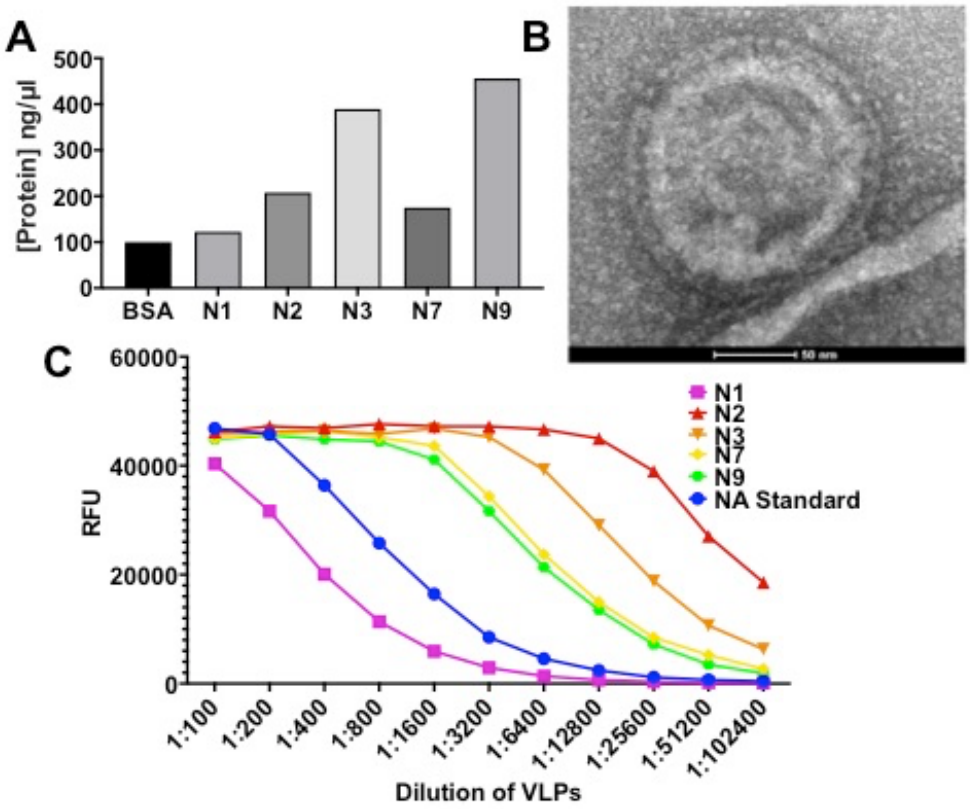


Figure 3.6

NA VLPs produced in HEK-293 cells are enzymatically functional. **A**) NA levels after control background subtraction from Coomassie Blue stain of NA protein expression in VLPs. Data represents the single preparation of VLPs used in these experiments. **B**) TEM micrograph of a VLP expressing N2. **C**) Comparison of NA enzymatic activity using a MuNANA assay between different NA serotypes. One representative experiment is shown (n = 3). NA standard was the commercial bacterial sialidase, NeuA.

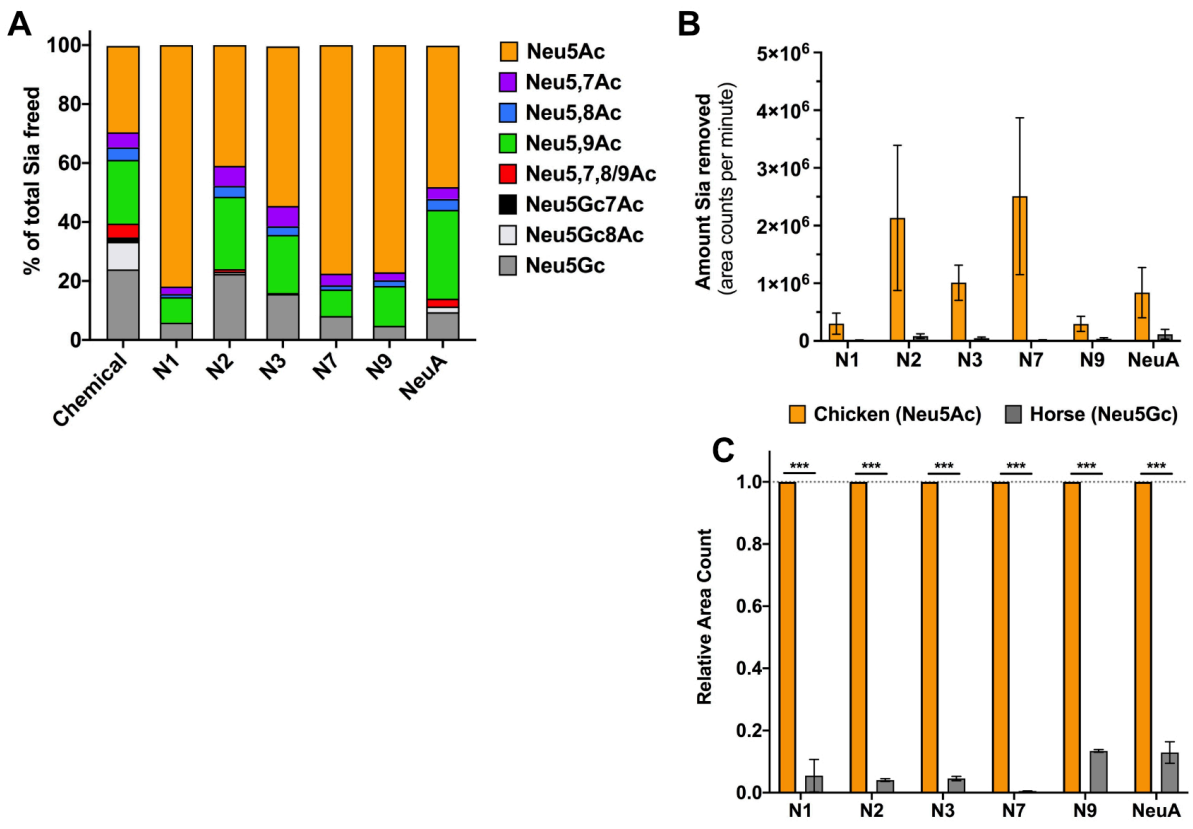


Figure 3.7

NA VLPs preferentially cleave unmodified Neu5Ac Sia and NA activity is inhibited by *O*-acetyl and Neu5Gc modifications. **A**) Bovine sub-maxillary mucin was treated with 1:100 NA VLPs or *Arthrobacter ureafaciens* NA (NeuA) for 4 hrs at 37°C and freed Sia was collected and analyzed using HPLC. The profiles of freed Sia were then compared to the profile of Sia removed chemically, a more unbiased approach. Profiles shown are the average of two independent experiments. **B, C**) Chicken erythrocytes (Neu5Ac) or horse erythrocytes (Neu5Gc) were treated with 1:100 NA VLPs for 4 hrs at 37°C and freed Sia was collected and total Sia removed was determined using HPLC. (**B**) Average area counts per minute (area under curve of chromatogram) were used as a measure of the absolute amount of Sia removal to compare Sia released between chicken and horse erythrocytes. (**C**) Relative area counts compared between chicken and horse erythrocytes. Data shown is average of two independent experiments. Data analyzed by t-test using PRISM software. * = p-value ≤0.05; ** = p-value ≤0.01; *** = p-value ≤0.001.

acetyl and Neu5Gc modifications inhibit the activities of many different IAV NA and of NeuA, and that multiple modifications, such as di-*O*-acetyl modifications, or Neu5Gc that is also *O*-acetylated, were even more inhibitory, showing high resistance to most of the viral and bacterial NAs tested here.

Effects of modified Sia on HA binding. The initiation of IAV infection requires HA glycoprotein binding to Sia to allow the virus to be taken up into the cell, and there appears to be a direct relationship between Sia binding affinity and infection (64). To clearly determine the effect of Sia modifications on HA binding, we examined the binding of soluble HA fused to a human IgG1 Fc (HA-Fc) to synthetic, biotinylated α2–6 linked sialosides of Neu5Ac, Neu5Ac9NAc, and Neu5Gc (65). Neu5Ac9NAc was used instead of 9-*O*-acetyl Neu5Ac (Neu5,9Ac_2) due to the increased stability of the 9-*N*-Ac group (66, 67). Briefly, ELISA-grade 96 well plates were coated with HA-Fc derived from California/04/2009 H1N1 and Aichi/2/1968 H3N2 strains, then incubated with the synthetic biotinylated sialosides. Binding of the biotinylated sialosides to the HA-Fc was detected using a streptavidin-linked HRP probe as previously described (68, 69). Compared to Neu5Ac sialosides, both California/04/2009 H1 and Aichi/1968 H3 HA-Fcs had decreased binding to Neu5Gc and Neu5Ac9NAc (**Fig. 3.8A**). The addition of the *N*-acetyl group at C9 blocked most binding, while Neu5Gc showed only a low level of binding. We saw the same binding dynamics for other H1 and H3 HA-Fc, but the SNA lectin, which recognizes α2–6-linked Sia, bound equally well to Neu5Ac and Neu5Gc, but not to Neu5Ac9NAc (**Fig. 3.8B**). This shows that the presence of *O*- (and in this case *N*-) acetyl modifications can inhibit many Sia-binding proteins.

Effects of modified Sia on influenza A infection. While the low surface expression of 9-*O*- and 7,9-*O*-Ac on cells does not reduce IAV infection, viruses will also encounter modified

Sia in mucus, which in many hosts and tissues has larger amount of these modifications. To determine how these Sia modifications can affect IAV infection, untreated BSM or BSM treated with esterase to remove 9-O-Ac and 7,9-O-Ac, were incubated with A/California/04/2009 (pH1N1), A/Puerto Rico/8/1934 (PR8 H1N1), and A/Victoria/361/2011 (Victoria H3N2) prior to infection of cells (**Fig 3.9A**). Both untreated BSM and esterase-treated BSM were inhibitory towards all three IAV strains, a trend towards higher inhibition by esterase-treated BSM, and significantly more inhibition for the PR8 H1N1 strain. This suggests that removing the 7,9-*O*- and 9-*O*-Ac from the Sia may have increased the virus binding to the mucin and inhibition of viral infection.

Sia in sera have long been known to bind to influenza viruses, so that sera are often treated with neuraminidase as a "receptor destroying enzyme" prior to use in serological tests (70, 71). To specifically compare the effects of added Neu5Gc or Neu5Ac on the efficiency of infection, the same three IAV strains were incubated with mouse serum from either wild-type mice (>80% Neu5Gc) or $CMAH^{-/-}$ mice (100% Neu5Ac) prior to inoculation of cells (**Fig. 3.9B**). In this case no specific trend was detected for the three viruses tested. Victoria H3N2 showed almost complete inhibition of infection by both sera, while PR8 H1N1 had a lower level of inhibition. Only pH1N1 showed a significant difference in response to the serums, with the WT mouse serum having a much stronger inhibitory effect compared to the $CMAH^{-/-}$ serum. This suggests that Neu5Gc vs. Neu5Ac inhibition may vary by virus strain, and that likely depends on some combination of HA binding specificity and NA activity.

3.5 DISCUSSION

Modified Sia are widely expressed within tissues and on mucosal surfaces of many animals, but with significant variation in the amounts of each modified Sia present (12, 13).

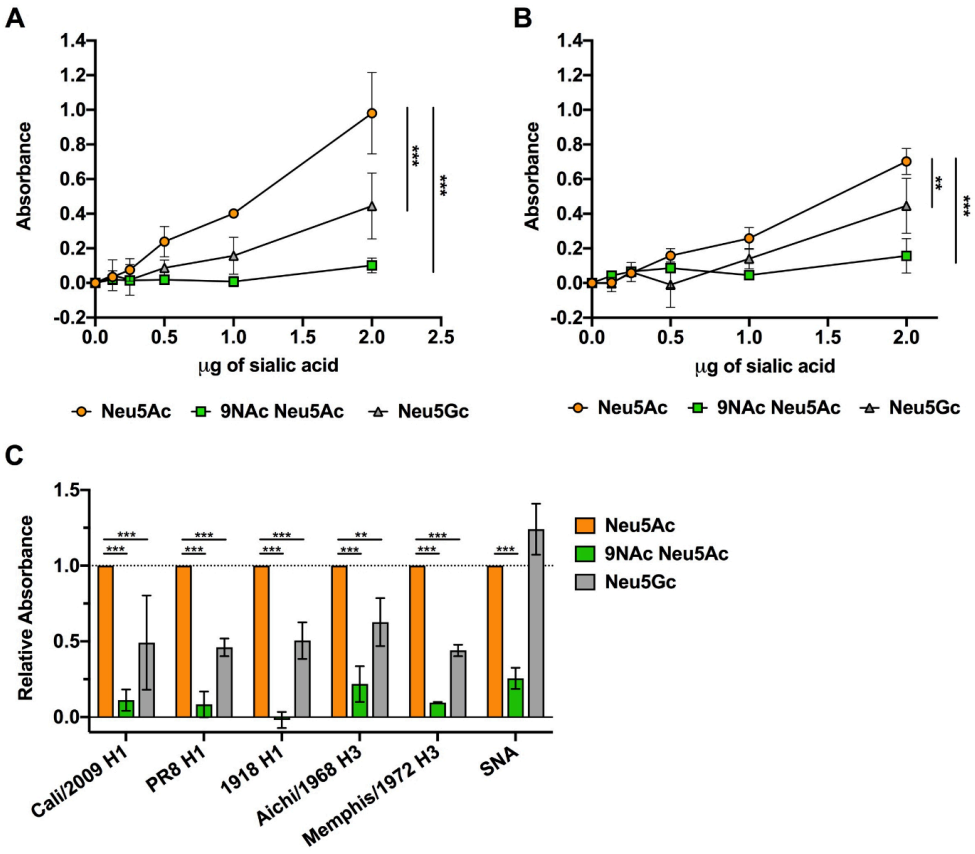


Figure 3.8

Soluble HA-Fc binding to synthetic sialosides showed decreased binding to modified Sia in an ELLA assay. **A, B**) Soluble HA constructs were developed by expressing HA proteins from different IAV strains fused to a human IgG1 Fc (HA-Fc). HA-Fc binding to synthetic α2–6-linked sialosides was assessed using an ELLA assay. Titration curves of sialoside binding by HA-Fc for (**A**) A/California/04/2009 H1N1 and (**B**) A/Aichi/2/1968 H3N2 were measured via colorimetric measurement. **C**) Sialoside binding for different H1 and H3 HA-Fc were determined using 2 μg of sialic acid. Lectin from *Sambucus nigra* (SNA), which specifically binds α2–6-linked Sia, was also included as a control. Data is shown as relative to HA-Fc binding to unmodified Neu5Ac.
Data analyzed by 2-way Anova using PRISM software. Data is average of three independent experiments.
* = p-value ≤0.05; ** = p-value ≤0.01; *** = p-value ≤0.001.

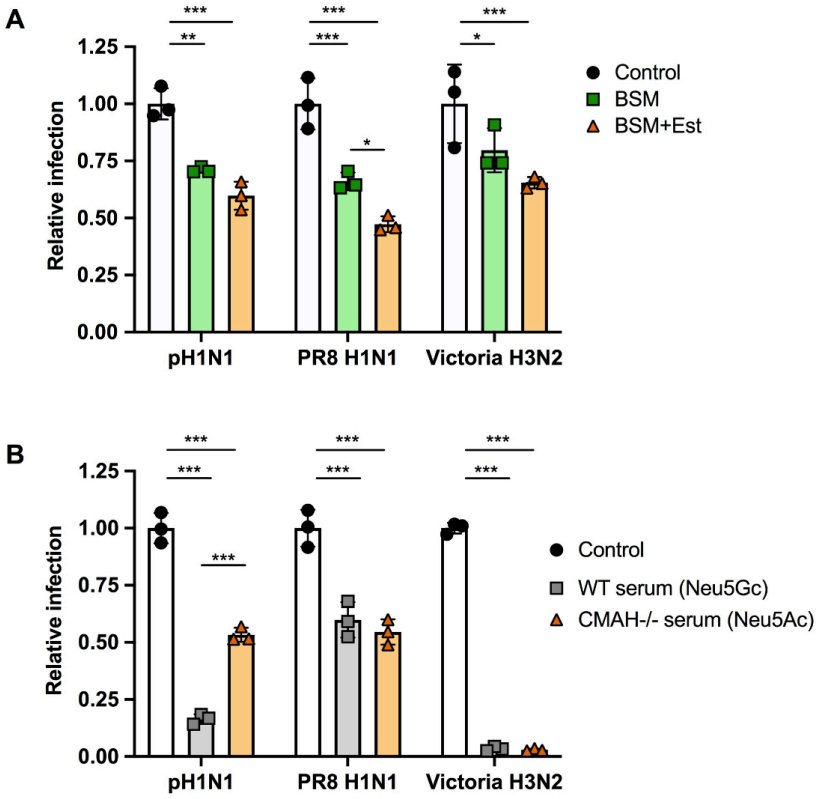


Figure 3.9

Virus infection is inhibited by mucin with greater inhibition when *O*-acetyl groups are removed. Virus infection is also inhibited by serum, with no clear difference between Neu5Ac and Neu5Gc presence.
A) A/California/04/2009 (pH1N1), A/Puerto Rico/8/1934 (PR8, H1N1), and A/Victoria/361/2011 (Victoria, H3N2) were mixed with 20 µg of BSM or BSM pre-treated with esterase active bovine coronavirus (BCoV HE-Fc) to remove *O*-acetyl modifications. This mixture was then used to infect cells at an MOI of 0.5 for 10 hrs. Infectivity was determined by flow cytometry analysis for NP positive cells.
B) A/California/04/2009 (pH1N1), A/Puerto Rico/8/1934 (PR8, H1N1), and A/Victoria/361/2011 (Victoria, H3N2) were mixed with serum from either WT mice (Neu5Gc) or $CMAH^{-/-}$ mice (Neu5Ac). This mixture was then used to infect cells at an MOI of 0.5 for 10 hrs. Infectivity was determined by flow cytometry analysis for NP positive cells.
Data analyzed by 2-way Anova using PRISM software.
* = p-value ≤0.05; ** = p-value ≤0.01; *** = p-value ≤0.001.

Modified Sia are present at high levels on mucosal surfaces, including in GI and respiratory tissues, indicating that they are likely involved in tissue- and host-specific interactions with both pathogens and normal flora. While there have been suggestions that 9-*O*-Ac, 7,9-*O*-Ac, and Neu5Gc Sia might influence IAV infection by interfering with HA binding or NA activities, there is not a lot of direct evidence for their effects. The goal of this study was to re-examine and define the tissue-specific expression of the 9-*O*-Ac, 7,9-*O*-Ac, and Neu5Gc Sia in mice and in secreted mucus from different IAV host species. We also defined the effect of these Sia modifications on HA binding, NA activity, and infection for different IAV strains. While we focused on IAV in these studies, the findings here are broadly relevant and underline the need for understanding the expression of modified Sia between species and their potential role in tropism and infection of many viruses that interact with or utilize Sia during infection.

Mouse tissue distribution. In previous studies we have shown that there is variation in the expression of the 9-*O*-, 7,9-*O*- and 4-*O*-Ac Sia (13). To better understand the type of variation seen in other tissues, we examined the distribution of these forms, as well as Neu5Gc, in the different tissues of wild-type C57BL/6 mice, as well as for $CMAH^{-/-}$ mice. Using the HE-Fc virolectin probes, *O*-acetyl modified Sia were found only in specific tissues and on certain cell subpopulations. Higher staining was seen for cells and mucus in the respiratory and GI tracts, with trachea and colon having particularly strong staining for *O*-acetyl Sia on epithelial cells and associated mucus layers. The percentages of *O*-acetyl modified Sia in the tissues of mice varied widely, with high levels of 7,9-*O*-Ac_2 and 9-*O*-Ac being found on erythrocytes (~47%) and the colon (~19%). The levels in most other tissues varied between 2 and 9%, although these would likely be associated with higher levels on the smaller subsets of cells that were positive for staining with the probes. Neu5Gc was also present in many wild-type mouse

tissues with expression varying widely (between 10 and 80%). This is consistent with staining for Neu5Gc previously reported in some mouse tissues (51). 4-*O*-Ac Sia was found in small quantities in only a few tissues, with the highest being found in small intestine (~2%). This is consistent with previous findings of 4-*O*-Ac levels in mouse brain and liver quantified by GC/MS, although gut associated levels of 4-*O*-Ac were much higher in that study, possibly due to differences in tissue preparation (72). Considering the high level of apparent 4-*O*-Ac specific staining for the mucus in the colon, further investigation of this modification in mouse colonic mucin would be warranted.

It is interesting to note that the levels of *O*-acetyl modifications were more prevalent on Neu5Ac Sia than on Neu5Gc, and that when Neu5Gc was removed in the $CMAH^{-/-}$ mice, the levels of *O*-acetyl modified Sia were higher across most tissues tested. This is consistent with previous analysis of some $CMAH^{-/-}$ tissues by immunohistochemical staining (51). However, Neu5Gc also had *O*-acetyl variants (Neu5Gc7Ac and Neu5Gc8Ac) that were present at low levels in mouse tissues, as well as in bovine mucin, which are consistent with previous reports (28, 73). It has been previously shown that the *O*-acetyltransferase, CasD1, prefers CMP-Neu5Ac as a substrate over CMP-Neu5Gc, but it appears to still have some activity on CMP-Neu5Gc (15). However, the regulation of the *O*-acetyl modification placement on both Neu5Ac and Neu5Gc, and possible differences between these two Sia forms, remains to be fully elucidated.

Variation on erythrocytes, in saliva, and on other mucins. Erythrocytes are an important tool in virology for analyzing virus binding to Sia in hemagglutination assays, as well as for determining virus titer, and antibody inhibition (HAI assays) (58, 59). It has long been known that IAV varies in hemagglutination of RBCs from different species, at least in part due

to differences in the Sia linkages present. The structures of HAs with Sia bound often suggest that modification of the C4, 5, 7, and/or 9 positions would influence IAV interactions with Sia. We found the Sia modifications on erythrocytes vary greatly between species, with mice have a high degree of *O*-acetyl modified Sia while horses, cows, sheep, and pigs have primarily Neu5Gc Sia. The reasons for such variation in Sia modifications are likely complex and could involve cell-cell signaling between erythrocytes and endothelial and immune cells, as well as interactions with pathogens. It has recently been suggested that loss of Neu5Gc expression in humans could be related to pressure from an ancestral strain of *Plasmodium falciparum* (44, 46).

To define the variation of modified Sia present in secreted mucus across different animals, we examined saliva as it contains many of the same heavily glycosylated proteins and mucins, including MUC5B, that are present in respiratory mucus (55–57). We found a great deal of variability in both the amounts and types of modified Sia present. Some animals (horses, mice, and cows), had larger amounts of *O*-acetyl Sia in their saliva, while human and dogs had primarily unmodified Neu5Ac forms. Pigs are a natural IAV host, and their saliva contained primarily Neu5Gc, so that pig saliva might inhibit IAV infection. Horses had around 10% 4-*O*-Ac Sia present in their saliva. The 4-*O*-Ac modification has been proposed to be a potent inhibitor of many types of viral and bacterial neuraminidases, and to be the γ-inhibitor of horse serum, where it may be present at high levels on the α-2 macroglobulin protein (61, 70, 74). However, as the gene for the 4-*O*-sialyl acetyltransferase has not yet been identified, little is known about its synthesis, expression, or regulation (21, 22).

It has previously been reported that human colonic mucus is enriched for 7,9-*O*- and 9-*O*-Ac (11, 53). However, human saliva and secreted mucus from respiratory cells contained mostly unmodified Neu5Ac, suggesting different expression of modified Sia in the mucus of the

respiratory and gastrointestinal tissues. This may be due to the particular functional characteristics of the microbiome in the GI tract compared to other mucosal sites, where 7,9-*O*- and 9-*O*-Ac on colonic mucus may decrease bacterial degradation of sialylated glycans, perhaps improving mucus integrity (75, 76). Microbiome interactions may also explain the increased *O*-acetyl modifications found on both cow and mouse saliva. Cows are ruminants and as part of their digestion, regurgitate partially digested food from their microbe-rich rumen into their mouth to continue chewing as a cud to extract more nutrients. Somewhat similarly, mice are coprophagic and consume feces, along with associated microbes, to re-digest again to improve absorption of nutrients. This would give both cows and mice the potential for more complex mucus and microbe interactions in both the oral cavity and the gut, thus having *O*-acetyl modifications on sialylated glycans on saliva mucus proteins could be advantageous by preventing degradation. However, further research is needed on the roles of these modifications for the interactions between oral and colonic mucus with the microbiome.

Effects of modified Sia on influenza viruses. IAV use Sia as their primary receptor for host infection, and the specific linkages of Sia to the underlying glycan chain have long been known to influence host tropism (40). We examined the effects of *O*-acetyl and Neu5Gc modifications on IAV HA binding, NA cleavage, and on infection, and found differences among the IAV strains examined. All the different IAV NA tested showed preferential removal of Neu5Ac over any modified Sia form. Cleavage by N1, N7, and N9 were strongly inhibited by mono-acetylated Sia, while N2 and N3 were less affected. All NA had much lower activity against Neu5Gc, di-acetylated Sia, and particularly against Neu5Gc forms with additional *O*-acetylations. This confirms previous reports showing or suggesting inhibition of NA cleavage in H1N1 and H3N2 strains (42, 48), but revealing that there is wide variation of effects on different

NA types. These differences did not seem to follow broad structural groupings between Group 1 NA proteins (N1, N4, N5, N8) and Group 2 NA proteins (N2, N3, N6, N7, N9) (77). However, one possible source of variation could be the host tropism of the parent viruses for the NA proteins used in the VLPs. Both N2 and N3 originate from avian isolates and while N1, N7, and N9 came from human isolates, which could explain differences seen in response to mono-acetylated Sia. Differences in NA cleavage ability against Neu5Gc had previously been reported for swine IAV isolates (49) but comparisons between NA proteins against *O*-acetylated Sia forms have been restricted to human isolates (42). The structure of NA inhibitors, such as Oseltamivir and Zanamivir, resemble Sia with side groups added along the glycerol side chain, analogous to C7 through C9, and at the C5 and C4 positions (78). The location of these inhibitory side groups resembles the chemical modifications of NeuGc and 7,9-*O*- and 9-*O*-Ac Sia variants, which explains why these natural Sia forms are NA inhibitors. It would be of interest to examine NA cleavage efficiency against these natural modified Sia forms from IAV strains that were adapted to other hosts with Neu5Gc expression, such as horses and pigs, as well as NA proteins with resistance to synthetic NA inhibitors.

Acetylation and Neu5Gc modifications were also inhibitory for HA binding, with soluble HA-Fc sourced from different H1N1 and H3N2 IAV strains showing significantly lower binding to these forms compared to unmodified Neu5Ac in an ELLA assay. Previous research has shown that Neu5Gc is not bound by most natural IAV isolates, with the exception of an equine H7N7 strain. Comparisons between this equine H7N7 strain and a lab-generated H5 Y161A mutant that bound Neu5Gc, it was found that the structural changes in the 130-loop caused by the loss of hydrophobic interactions of the H1 Y161A mutation were also shared in the H7 structure (48). Thus, it appears that the receptor-binding pocket of most IAV HA proteins cannot accommodate

the Neu5Gc modification due to steric hindrance of the 130-loop. Structural analysis of HA proteins with *O*-acetyl variants have not been studied to this same level of detail; however, current structural models of HA bound to Neu5Ac show important hydrogen-bond interactions between the C9 of Sia and amino acids at position 98, 190, and 228 in the receptor binding pocket (79). It is likely that the *O*-acetyl modification at C9 would block Sia from binding correctly in the receptor-binding pocket.

IAV infects cells at mucosal surfaces including in the gastrointestinal tract of birds and the respiratory tract of mammalian hosts. Therefore, it is likely that viruses interact directly with mucus both to initiate infection and during viral release. Using untreated or esterase treated BSM (to remove 7,9-*O*- and 9-*O*-Ac) it was seen that the unmodified Neu5Ac was most inhibitory, suggesting that *O*-acetyl Sia reduces HA binding to the mucin, allowing binding to the unmodified Sia on the cell surface. Esterase treatment, therefore, allows more efficient binding to the BSM and greater inhibition of IAV infection. For some viruses there is likely a complementarity between the inhibitory effects of 7,9-*O*- and 9-*O*-Ac on HA and NA, where lower HA binding allows the viruses to avoid binding to Sia forms that NA cannot remove efficiently. This effect has been reported for virus grown in the presence of other NA inhibitors (64, 80), and is seen in the balance between the activities of the HA and NA for the α2–3 and α2–6-linked Sia (80, 81). Incubation of virus with the wild-type mouse serum (>80% Neu5Gc) compared to the serum of $CMAH^{-/-}$ mice (100% Neu5Ac) also showed varying effects. Inhibition by WT serum was seen only for pH1N1 virus, while Victoria H3N2 and PR8 H1N1 were inhibited by both sera. This could confirm previous findings that the density of Sia on these serum proteins is the strongest inhibitory factor rather than the type of modified Sia present, as shown for incubation with horse serum (82). However, the variable results for different viruses

may indicate variation in sensitivity to Neu5Gc between strains that requires further investigation. A recent paper showed that repeat passaging of human and canine strains of IAV in wild-type mice compared to $CMAH^{-/-}$ mice saw no mutations in HA or NA, nor any differences in disease severity between mouse strains (83). It is possible that the amount of unmodified Sia present in mouse respiratory tissue is sufficient for infection, with virus avoiding binding to Neu5Gc. However, a natural isolate of equine H7N7 has previously been described to preferentially binding Neu5Gc which seems to indicate that adaptation to bind this Sia form is possible (48). Therefore, it would also be worthwhile comparing the interactions of IAV strains adapted to different host species with Neu5Gc, particularly species with higher levels of modified Sia present in their respiratory tracts such as horses and pigs.

In summary we have shown that both the *O*-acetyl and Neu5Gc modifications present on secreted glycoproteins in mucus and saliva, as well as on erythrocytes, vary greatly between different species. Some of these modifications inhibit HA binding and NA cleavage, but with a significant variability between IAV strains. While the presence of these modifications can inhibit infection, how they affect virus host tropism and evolution is likely complex and still not fully understood.

3.6 MATERIALS AND METHODS

Cells and virus. Canine MDCK-NBL2 (ATCC) and A549 (ATCC) cells were grown in Dulbecco's Modified Eagle Medium (DMEM) with 5% fetal bovine serum and 50 μg/ml gentamicin. Influenza A virus strains pH1N1 (A/California/04/2009), Victoria H3N2 (A/Victoria/361/2011), and PR8 (A/Puerto Rico/8/1934) were rescued from reverse genetics plasmids using previously established protocols (84). Rescued virus were grown to low passage

on MDCK-NBL2 cells using infection media containing DMEM, 0.03% BSA, and 1ug/ml TPCK-treated trypsin.

Erythrocytes, mucus, and saliva. Chicken, cow, sheep, guinea pig, and pig erythrocytes were sourced from Lampire Biological Laboratories (Pipersville, PA). Horse and mouse blood were sourced from The Baker Institute for Animal Health (Cornell University, Ithaca, NY). Dog blood was sourced from the Cornell Veterinary Hospital Diagnostic Center (Ithaca, NY). All erythrocytes were washed in PBS three times and diluted to 5% v/v in PBS. Bovine submaxilliary mucin was purchased from Sigma-Aldrich (St. Louis, MO). Animal studies were all subject to approved protocols from the Cornell Institutional Animal Care and Use Committee.

Saliva from humans was collected by passive drooling following the protocol approved by the University at Buffalo Human Subjects IRB board (study # 030–505616). Informed consent was obtained from all human participants. Saliva from animals was collected by suction using commercially available devices containing absorbent sponges in a syringe-like receptacle (Super-SAL and Micro-SAL, Oasis Diagnostics, Vancouver, WA). Saliva from mice (laboratory strain C57BL/10SnJ) was kindly provided by Jill Kramer (University at Buffalo) using a collection procedure as previously described (85). Saliva from dogs, cows, horses, and pigs, was provided by Erin Daugherity and Luce E. Guanzini (Cornell University). Animals were not allowed to eat or drink prior to the collection to ensure the oral cavity was free of food and other debris. The collection was performed using a commercially available device (Micro-SAL). Large animals were gently restrained and a larger collection device (Super-SAL) was placed under the tongue for up to three minutes, or until fully soaked. Saliva from castrated domestic pigs were also provided by Anja Globig (Friedrich-Loeffler-Institut, Insel Riems - Greifswald, Germany). Saliva from dogs was also kindly provided by Barbara McCabe (Buffalo, NY).

Normal human bronchial epithelial cells (Lonza; cat#CC-2540S) were seeded onto human placental collagen-coated permeable transwell inserts at a density of $2.5x10^4$ cells per well and cultured in bronchial epithelial cell growth basal medium (BEBM) supplemented with bronchial epithelial cell growth SingleQuots (BEGM) in both the apical and basal compartments. After reaching confluence at approximately 7 days post seeding, apical media was aspirated, and basal media replaced with air-liquid interface media containing half DMEM, half BEBM, plus the BEGM SingleQuots to complete differentiation. Apical surfaces of NHBE cells were washed twice with phosphate buffered saline (PBS) to collect mucins for HPLC analysis.

Conditioned media from A549 cells was prepared by washing a fully confluent flask of cells to remove any serum and allowing the cells to grow in serum-free media for 5-7 days. Conditioned media was collected, dialyzed with three volumes of PBS, and concentrated using a 30 kD centrifugal filter (Pall Corporation). Protein concentration was determined using a Qubit 4 fluorometer (Invitrogen).

Immunohistochemistry of mouse tissues. Expression of O-acetyl modified Sia in various tissues of mice was examined by preparing frozen sections of optimal cutting temperature compound (OCT)-embedded tissue. After a 30 min fixation in 10% buffered formalin, sections were incubated with recombinant virolectins made by expressing nidovirus HE glycoprotein fused to the Fc region of human IgG1, as described by others (12). Nidovirus HEs are specific for *O*-acetyl Sia modifications: MHV-S for 4-*O*-Ac, BCoV-Mebus for 7,9-*O*-Ac (and low recognition of 9-*O*-Ac), PToV-P4 for 9-*O*-Ac. Virolectins were then detected using a biotin-conjugated αFc region secondary antibody followed by incubation with the Vectastain ABC reagent and NovaRed substrate (Vector). Sections were counterstained with hematoxylin.

Quantification of sialic acid by HPLC The sialic acid composition from tissues, mucin, and erythrocytes were analyzed by HPLC analysis as previously described (50, 86). In brief, Sia from 20-30 mg of tissue, 50 μg of mucin or saliva, or 100 μl of washed 5% v/v erythrocytes were release using 2M acetic acid at 80°C for 3 hr followed by filtration through a 10kD centrifugal filter (Microcon) and dried using a vacuum concentrator (SpeedVac). Released Sia were labeled with 1,2-diamino-4, 5-methylenedioxybenzene (DMB, Sigma Aldrich) for 2.5 hr at 50°C. HPLC analysis was performed using a Dionex UltiMate 3000 system with an Acclaim C18 column (ThermoFisher) under isocratic elution in 7% methanol, 7% acetonitrile, and 86% water. Sia standards were bovine submaxillary mucin, normal horse serum, and commercial standards for Neu5Ac and Neu5Gc (Sigma Aldrich). Pre-treatment of samples with 30 μg/ml esterase-active BCoV HE-Fc overnight at 37°C removed 7,9-*O*- and 9-*O*-acetyl modifications. Final data analysis was completed using PRISM software (GraphPad, version 8).

Biotinylated a2–6-linked sialosides. Biotinylated α2–6-linked sialosides Sia α2–6LacNAc-biotin containing Neu5Ac, Neu5Gc, or Neu5Ac9NAc as the sialic acid form were synthesized from LacNAc-biotin (87) as the acceptor substrate and Neu5Ac, ManNGc (88), or ManNAc6NAc (67) as the donor precursor using a one-pot multienzyme sialylation system similar to that described previously (89).

IAV HA affinity for 9-*O*-Ac modified Sia. HA-Fc constructs were produced as previously described (90). HA-Fc binding to sialosides was performed as previously described (68, 69). In brief, ELISA-grade 96 well plates (ThermoFisher Scientific) were coated with 5 μg of HA-Fc for overnight at 4°C. Plates were then washed 3× with PBS and blocked using 1x Carbo Free Blocking Buffer (Vector Labs, Burlingame, CA) for 1 hr. After blocking, plates were washed once with PBS and treated with sialosides diluted in PBS for 3 hr at room temperature,

then washed 3× with PBS. Plates were then incubated with HRP-streptavidin complex (Vector Labs) for 45 min, washed 3× with PBS, then incubated with 3,3',5,5'-tetramethyl benzidine (TMB, Thermo Fisher Scientific). TMB development was stopped with 2M sulfuric acid and then analyzed using a colorimetric plate reader (Multiskan EX, ThermoFisher Scientific). Data analysis was completed in PRISM software (GraphPad, version 8).

Generation of NA VLPs. NA sequences were obtained from GenBank (A/Ohio/07/2009 N1: ACP44181; A/mallard/Ohio/11OS2045/2011 N2: AGC70842; A/chicken/Murree/NARC-01/1995 N3: ACL11962, A/Netherlands/219/2003 N7: AAR11367, A/Yunnan/0129/2017 N9: ARG43209). Sequences were codon optimized, tagged, and ordered through Biomatik in the pcDNA3.1(+) vector. To produce VLPs, HEK-293T cells were seeded in 15cm plates and transfected when 80% confluent. Cells were transfected with 4 µl of Polyethylenimine (PEI) (Polysciences cat# 23966-2) at 1 mg/ml concentration for every 1 µg plasmid DNA stock in 9ml of Opti-MEM. Eight hours post transfection, 6ml of pre-warmed Opti-MEM was added. Supernatant was collected 72 hours post transfection and purified using ultracentrifugation (110,000xg, 1.5 hours, 4°C) through a 20% sucrose cushion, then the pellet re-suspended in PBS and stored at 4°C.

NA cleavage assay with NA VLPs. Bovine sub-maxillary mucin (BSM) or erythrocytes from horses and chickens were used as a substrate to determine NA activity of the different NA VLPs. Briefly, 50 µg of BSM or 5% v/v washed erythrocytes in PBS were treated with 1:100 NA VLPs or 1:100 *Arthrobacter ureafaciens* NA (NeuA, New England BioLabs) for 4 hours at 37°C. These dilutions of VLPs were chosen based the enzymatic activity as determined by a standard NA cleavage assay using methylumbelliferyl *N*-acetylneuraminide (MuNANA) as the substrate (48). The conditions for NA cleavage were chosen to allow for enough Sia to be

released for HPLC analysis based on enzymatic activity of the NA VLPs and NeuA control. Free Sia was collected and prepared for HPLC analysis as above.

Mucin and serum inhibition of infection. MDCK cells were seeded to ~80% confluency in 12 well plates with cells allowed to settle for 6 hrs. Virus was diluted in PBS to give MOI of 0.5 and then mixed with 20 μg untreated BSM or esterase treated BSM and incubated for 45 min at room temp. For mouse serum inhibition, virus was mixed with 200 μg wild-type or $CMAH^{-/-}$ serum instead. Serum- or mucin-treated virus was then added to washed MDCK cells and incubated for 1 hr with tilting to prevent cell drying. Inoculum was then removed, media added and cells were incubated for 10 hrs. Cells were then harvested, stained with an anti-influenza A NP antibody, and analyzed for infection using a Millipore Guava EasyCyte Plus flow cyotometer (EMD Millipore, Billerica, MA) with analysis using FlowJo software (TreeStar, Ashland, OR). Statistical analyses were performed in PRISM software (GraphPad, version 8).

3.7 ACKNOWLEDGEMENTS

We thank Wendy Weichert for expert technical support. We also thank Lubov Neznanova and Jill M. Kramer (University of Buffalo), along with Anja Globig (Friedrich Loeffler Institute, Germany), for sharing samples of saliva. Also thanks to Jessica Waltemyer (Cornell University).

SUPPORT

Supported by CRIP (Center of Research in Influenza Pathogenesis), an NIAID funded Center of Excellence in Influenza Research and Surveillance (CEIRS) contract HHSN272201400008C to Colin Parrish, NIH grants R01 GM080533 to Colin Parrish and R01AI130684 to Xi Chen and Ajit Varki, NIH Common Fund Grant (U01CA199792) to Ajit Varki, as well as NIDCR R01DE019807 and NIH Common Fund Grant (U01CA221244) to Stefan Ruhl.

REFERENCES

1. Varki A, Schauer R. 2009. Chapter 14 Sialic AcidsEssentials of Glycobiology - second edition, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Springs Harbor, NY.

2. Varki NM, Varki A. 2007. Diversity in cell surface sialic acid presentations: implications for biology and disease. Laboratory investigation; a journal of technical methods and pathology 87:851–857.

3. Lehmann F, Tiralongo E, Tiralongo J. 2006. Sialic acid-specific lectins: Occurrence, specificity and function. Cellular and Molecular Life Sciences 63:1331–1354.

4. Varki A. 2007. Glycan-based interactions involving vertebrate sialic-acid-recognizing proteins. Nature 446:1023–9.

5. Wasik BR, Barnard KN, Parrish CR. 2016. Effects of Sialic Acid Modifications on Virus Binding and Infection. Trends in Microbiology xx:1–11.

6. Van Breedam W, Pöhlmann S, Favoreel HW, de Groot RJ, Nauwynck HJ. 2014. Bitter-sweet symphony: Glycan-lectin interactions in virus biology. FEMS Microbiology Reviews 38:598–632.

7. Mandal C, Schwartz-Albiez R, Vlasak R. 2012. Functions and Biosynthesis of O-acetylated sialic acids. Top Curr Chem.

8. Hirose K, Amano M, Hashimoto R, Lee YC, Nishimura SI. 2011. Insight into glycan diversity and evolutionary lineage based on comparative avio-N-glycomics and sialic acid analysis of 88 egg whites of galloanserae. Biochemistry 50:4757–4774.

9. Peri S, Kulkarni A, Feyertag F, Berninsone PM, Alvarez-Ponce D. 2018. Phylogenetic Distribution of CMP-Neu5Ac Hydroxylase (CMAH), the Enzyme Synthetizing the Proinflammatory Human Xenoantigen Neu5Gc. Genome Biol Evol 10:207–219.

10. Aamelfot M, Dale OB, Weli SC, Koppang EO, Falk K. 2014. The in situ distribution of glycoprotein-bound {4-O-Acetylated} sialic acids in vertebrates. Glycoconjugate journal 31:327–335.

11. Muchmore E a, Varki NM, Fukuda M, Varki a. 1987. Developmental regulation of sialic acid modifications in rat and human colon. The FASEB journal : official publication of the Federation of American Societies for Experimental Biology 1:229–235.

12. Langereis MA, Bakkers MJ, Deng L, Vered P-K, Vervoort SJ, Hulswit RJ, van Vliet AL, Gerwig GJ, de Poot SA, Boot W, van Ederen AM, Heesters BA, van der Loos CM, van Kuppeveld FJ, Yu H, Huizinga EG, Chen X, Varki A, Kamerling JP, de Groot RJ. 2015. Complexity and Diversity of the Mammalian Sialome Revealed by Nidovirus Virolectins. Cell reports 11:1966–1978.

13. Wasik BR, Barnard KN, Ossiboff RJ, Khedri Z, Feng KH, Perez DR, Varki A, Parrish CR. 2017. Distribution of O-acetylated sialic acids among the tissues of influenza hosts. mSphere 2:e00379–16.

14. Arming S, Wipfler D, Mayr J, Merling A, Vilas U, Schauer R, Schwartz-Albiez R, Vlasak R. 2011. The human Cas1 protein: A sialic acid-specific O-acetyltransferase? Glycobiology 21:553–564.

15. Baumann A-MTM, Bakkers MJ, Buettner FF, Hartmann M, Grove M, Langereis MA, de Groot RJ, Muhlenhoff M. 2015. 9-O-Acetylation of sialic acids is catalysed by CASD1 via a covalent acetyl-enzyme intermediate. Nature communications 6:7673.

16. Ravasio V, Damiati E, Zizioli D, Orizio F, Giacopuzzi E, Manzoni M, Bresciani R, Borsani G, Monti E. 2017. Genomic and biochemical characterization of sialic acid acetylesterase (siae) in zebrafish. Glycobiology 1–9.

17. Orizio F, Damiati E, Giacopuzzi E, Benaglia G, Pianta S, Schauer R, Schwartz-Albiez R, Borsani G, Bresciani R, Monti E. 2015. Human sialic acid acetyl esterase: Towards a better understanding of a puzzling enzyme. Glycobiology 25:992–1006.

18. Takematsu H, Diaz S, Stoddart A, Zhang Y, Varki A. 1999. Lysosomal and cytosolic sialic acid 9-O-acetylesterase activities can be encoded by one gene via differential usage of a signal peptide-encoding exon at the N terminus. Journal of Biological Chemistry 274:25623–25631.

19. Zhu H, Chan HC, Zhou Z, Li J, Zhu H, Yin L, Xu M, Cheng L, Sha J. 2004. A gene encoding sialic-acid-specific 9-O-acetylesterase found in human adult testis. Journal of Biomedicine and Biotechnology 2004:130–136.

20. Schauer R, Shukla AK. 2008. Isolation and properties of two sialate-O-acetylesterases from horse liver with 4- and 9-O-acetyl specificities. Glycoconjugate Journal 25:625–632.

21. Iwersen M, Dora H, Kohla G, Gasa S, Schauer R. 2003. Solubilisation and properties of the {sialate-4-O-acetyltransferase} from guinea pig liver. Biological chemistry 384:1035–1047.

22. Iwersen M, Vandamme-feldhaus V, Schauer R, Institut B, Kiel D-. 1998. Enzymatic 4-O-acetylation of N-acetylneuraminic acid in guinea-pig liver. Glycoconjugate Journal 904:895–904.

23. Gentsch JR, Pacitti AF. 1987. Differential interaction of reovirus type 3 with sialylated receptor components on animal cells. Virology 161:245–248.

24. Sateesh Peri, Asmita Kulkarni, Felix Feyertag, Patricia M Berninsone, David Alvarez-Ponce. 2017. Phylogenetic distribution of CMP-Neu5Ac hydroxylase (CMAH), the

enzyme synthetizing the pro-inflammatory human xeno-antigen Neu5Gc. Genome Biology and Evolution evx251.

25. Klein A, Krishna M, Varki NM, Varki A. 1994. 9-O-Acetylated sialic acids have widespread but selective expression: Analysis using a chimeric dual-function probe derived from influenza C hemagglutinin-esterase. Proceedings of the National Academy of Sciences of the United States of America 91:7782–7786.

26. Yu H, Cao H, Tiwari VK, Li Y, Chen X. 2011. Chemoenzymatic synthesis of C8-modified sialic acids and related α2–3- and α2–6-linked sialosides. Bioorganic & Medicinal Chemistry Letters 21:5037–5040.

27. Ravindranath RMH, Basilrose SRM. 2005. Localization of sulfated sialic acids in the dentinal tubules during tooth formation in mice. Acta Histochemica 107:43–56.

28. Morimoto N, Nakano M, Kinoshita M, Kawabata A, Morita M, Oda Y, Kuroda R, Kakehi K. 2001. Specific Distribution of Sialic Acids in Animal Tissues As Examined by LC–ESI-MS after Derivatization with 1,2-Diamino-4,5-Methylenedioxybenzene. Analytical Chemistry 73:5422–5428.

29. Corfield AP, Wagner SA, Safe A, Mountford RA, Clamp JR, Kamerling JP, Vliegenthart JFG, Schauer R. 1993. Sialic Acids in Human Gastric Aspirates: Detection of 9- *O* -Lactyl- and 9- *O* -Acetyl- *N* -Acetylneuraminic Acids and a Decrease in Total Sialic Acid Concentration with Age. Clinical Science 84:573–579.

30. Deplancke B, Gaskins HR. 2001. Microbial modulation of innate defense: goblet cells and the intestinal mucus layer. The American journal of clinical nutrition 73:1131S–1141S.

31. Knowles MR, Boucher RC. 2002. Mucus clearance as a primary innate defense mechanism for mammalian airways. Journal of Clinical Investigation 109:571–577.

32. Severi E, Hood DW, Thomas GH. 2007. Sialic acid utilization by bacterial pathogens. Microbiology 153:2817–2822.

33. Thomas GH. 2016. Sialic acid acquisition in bacteria-one substrate, many transporters. Biochemical Society Transactions 44:760–765.

34. Phansopa C, Kozak RP, Liew LP, Frey AM, Farmilo T, Parker JL, Kelly DJ, Emery RJ, Thomson RI, Royle L, Gardner RA, Spencer DIR, Stafford GP. 2015. Characterization of a sialate-O-acetylesterase (NanS) from the oral pathogen Tannerella forsythia that enhances sialic acid release by NanH, its cognate sialidase. Biochem J 472:157–167.

35. Lewis AL, Lewis WG. 2012. Host sialoglycans and bacterial sialidases: a mucosal perspective. Cellular Microbiology 14:1174–1182.

36. Robinson LS, Lewis WG, Lewis AL. 2017. The sialate *O* -acetylesterase EstA from gut *Bacteroidetes* species enables sialidase-mediated cross-species foraging of 9- *O* -acetylated sialoglycans. Journal of Biological Chemistry 292:11861–11872.

37. Matrosovich M, Herrler G, Klenk HD. 2015. Sialic Acid Receptors of Viruses, p. 1–28. *In* Top Curr Chem.

38. Hulswit RJG, Lang Y, Bakkers MJG, Li W, Li Z, Schouten A, Ophorst B, van Kuppeveld FJM, Boons G-J, Bosch B-J, Huizinga EG, de Groot RJ. 2019. Human coronaviruses OC43 and HKU1 bind to 9- *O* -acetylated sialic acids via a conserved receptor-binding site in spike protein domain A. Proceedings of the National Academy of Sciences 116:2681–2690.

39. Huang X, Dong W, Milewska A, Golda A, Qi Y, Zhu QK, Marasco W a, Baric RS, Sims AC, Pyrc K, Li W, Sui J. 2015. Human Coronavirus HKU1 Spike Protein Uses O-Acetylated Sialic Acid as an Attachment Receptor Determinant and Employs Hemagglutinin-Esterase Protein as a Receptor-Destroying Enzyme. Journal of virology 89:7202.

40. de Graaf M, Fouchier R a M. 2014. Role of receptor binding specificity in influenza A virus transmission and pathogenesis. The EMBO journal 33:823–41.

41. Cohen M, Zhang X-Q, Senaati HP, Chen H-W, Varki NM, Schooley RT, Gagneux P. 2013. Influenza A penetrates host mucus by cleaving sialic acids with neuraminidase. Virology journal 10:321.

42. Muñoz-Barroso I, García-Sastre a, Villar E, Manuguerra JC, Hannoun C, Cabezas J a. 1992. Increased influenza A virus sialidase activity with N-acetyl-9-O-acetylneuraminic acid-containing substrates resulting from influenza C virus O-acetylesterase action. Virus research 25:145–153.

43. Higa H, Rogers G, Paulson C. 1985. Influenza virus hemagglutinins differentiate between receptor determinants bearing N-acetyl-, N-glycollyl-, and N,O-diacetylneuraminic acids. Virology 144:279–282.

44. Klotz FW, Orlandi P a., Reuter G, Cohen SJ, Haynes JD, Schauer R, Howard RJ, Palese P, Miller LH. 1992. Binding of Plasmodium falciparum 175-kilodalton erythrocyte binding antigen and invasion of murine erythrocytes requires N-acetylneuraminic acid but not its O-acetylated form. Molecular and Biochemical Parasitology 51:49–54.

45. Varki A. 2001. Loss of N-glycolylneuraminic acid in humans: Mechanisms, consequences, and implications for hominid evolution. American journal of physical anthropology Suppl 33:54–69.

46. Proto WR, Siegel SV, Dankwa S, Liu W, Kemp A, Marsden S, Zenonos ZA, Unwin S, Sharp PM, Wright GJ, Hahn BH, Duraisingh MT, Rayner JC. 2019. Adaptation of Plasmodium falciparum to humans involved the loss of an ape-specific erythrocyte invasion ligand. Nat Commun 10:4512.

47. Malykh YN, Shaw L, Schauer R. 1998. The role of CMP-N-acetylneuraminic acid hydroxylase in determining the level of N-glycolylneuraminic acid in porcine tissues. Glycoconjugate Journal 15:885–893.

48. Broszeit F, Tzarum N, Zhu X, Nemanichvili N, Eggink D, Leenders T, Li Z, Liu L, Wolfert MA, Papanikolaou A, Martínez-Romero C, Gagarinov IA, Yu W, García-Sastre A, Wennekes T, Okamatsu M, Verheije MH, Wilson IA, Boons G-J, de Vries RP. 2019. N-Glycolylneuraminic Acid as a Receptor for Influenza A Viruses. Cell Reports 27:3284-3294.e6.

49. Xu G, Suzuki T, Maejima Y, Mizoguchi T, Tsuchiya M, Kiso M, Hasegawa A, Suzuki Y. 1995. Sialidase of swine influenza A viruses: variation of the recognition specificities for sialyl linkages and for the molecular species of sialic acid with the year of isolation. Glycoconjugate journal 12:156–161.

50. Varki A, Diaz S. 1984. The release and purification of sialic acids from glycoconjugates: Methods to minimize the loss and migration of O-acetyl groups. Analytical Biochemistry 137:236–247.

51. Hedlund M, Tangvoranuntakul P, Takematsu H, Long JM, Housley GD, Kozutsumi Y, Suzuki A, Wynshaw-Boris A, Ryan AF, Gallo RL, Varki N, Varki A. 2007. N-Glycolylneuraminic Acid Deficiency in Mice: Implications for Human Biology and Evolution. Molecular and Cellular Biology 27:4340–4346.

52. Lu Q, Padler-Karavani V, Yu H, Chen X, Wu S-L, Varki A, Hancock WS. 2012. LC-MS analysis of polyclonal human anti-Neu5Gc xeno-autoantibodies immunoglobulin G Subclass and partial sequence using multistep intravenous immunoglobulin affinity purification and multienzymatic digestion. Anal Chem 84:2761–2768.

53. Shen Y, Tiralongo J. 2004. Regulation of sialic acid O-acetylation in human colon mucosa. Biological Chemistry 385:145–152.

54. Campbell F, Appleton MAC, Fuller CE, Greeff MP, Hallgrimsson J, Katoh R, Ng OLI, Satir A, Williams GT, Williams ED. 1994. Racial variation in the O-acetylation phenotype of human colonic mucosa. The Journal of Pathology 174:169–174.

55. Cross BW, Ruhl S. 2018. Glycan recognition at the saliva – oral microbiome interface. Cellular Immunology 333:19–33.

56. Holmén JM, Karlsson NG, Abdullah LH, Randell SH, Sheehan JK, Hansson GC, Davis CW. 2004. Mucins and their O-Glycans from human bronchial epithelial cell cultures. American journal of physiology Lung cellular and molecular physiology 287:L824–L834.

57. Joo NS, Evans IAT, Cho H-J, Park I-H, Engelhardt JF, Wine JJ. 2015. Proteomic Analysis of Pure Human Airway Gland Mucus Reveals a Large Component of Protective Proteins. PLOS ONE 10:e0116756.

58. Pretini V, Koenen MH, Kaestner L, Fens MHAM, Schiffelers RM, Bartels M, Van Wijk R. 2019. Red Blood Cells: Chasing Interactions. Frontiers in Physiology 10.

59. Ustinov NB, Zavyalova EG, Smirnova IG, Kopylov AM. 2017. The power and limitations of influenza virus hemagglutinin assays. Biochemistry (Moscow) 82:1234–1248.

60. Corfield AP, Veh RW, Wember M, Michalski JC, Schauer R. 1981. The release of *N* -acetyl- and *N* -glycolloyl-neuraminic acid from soluble complex carbohydrates and erythrocytes by bacterial, viral and mammalian sialidases. Biochemical Journal 197:293–299.

61. Hanaoka K, Pritchett TJ, Takasaki S, Kochibe N, Sabesan S, Paulson JC, Kobata A. 1989. 4-O-Acetyl-N-acetylneuraminic acid in the N-linked carbohydrate structures of equine and guinea pig a2-macroglobulins, potent inhibitors of influenza virus infection. Journal of Biological Chemistry 264:9842–9849.

62. Hunter CD, Khanna N, Richards MR, Rezaei Darestani R, Zou C, Klassen JS, Cairo CW. 2018. Human Neuraminidase Isoenzymes Show Variable Activities for 9- *O* -Acetyl-sialoside Substrates. ACS Chemical Biology 13:922–932.

63. Thompson CM, Petiot E, Mullick A, Aucoin MG, Henry O, Kamen AA. 2015. Critical assessment of influenza VLP production in Sf9 and HEK293 expression systems. BMC Biotechnology 15.

64. Mitnaul LJ, Matrosovich MN, Castrucci MR, Tuzikov AB, Bovin NV, Kobasa D, Kawaoka Y. 2000. Balanced hemagglutinin and neuraminidase activities are critical for efficient replication of influenza A virus. J Virol 74:6015–6020.

65. Chokhawala HA, Yu H, Chen X. 2007. High-throughput substrate specificity studies of sialidases by using chemoenzymatically synthesized sialoside libraries. ChemBioChem 8:194–201.

66. Li W, Xiao A, Li Y, Yu H, Chen X. 2017. Chemoenzymatic synthesis of Neu5Ac9NAc-containing α2–3- and α2–6-linked sialosides and their use for sialidase substrate specificity studies. Carbohydrate Research 451.

67. Khedri Z, Xiao A, Yu H, Landig CS, Li W, Diaz S, Wasik BR, Parrish CR, Wang L-P, Varki A, Chen X. 2016. A Chemical Biology Solution to Problems with Studying Biologically Important but Unstable 9-O-Acetyl Sialic Acids. ACS Chemical Biology acschembio.6b00928.

68. Kumari K, Gulati S, Smith DF, Gulati U, Cummings RD, Air GM. 2007. Receptor binding specificity of recent human H3N2 influenza viruses. Virology Journal 4:42.

69. Wu W, Air GM. 2004. Binding of influenza viruses to sialic acids: reassortant viruses with A/NWS/33 hemagglutinin bind to α2,8-linked sialic acid. Virology 325:340–350.

70. Matrosovich MN, Gambaryan AS, Chumakov MP. 1992. Influenza Viruses Differ in Recognition of 4-0-Acetyl Substitution of Sialic Acid Receptor Determinant 858:854–858.

71. Matrosovich M, Gao P, Kawaoka Y. 1998. Molecular Mechanisms of Serum Resistance of Human Influenza H3N2 Virus and Their Involvement in Virus Adaptation in a New Host 72:6373–6380.

72. Rinninger A, Richet C, Pons A, Kohla G, Schauer R, Bauer H-C, Zanetta J-P, Vlasak R. 2006. Localisation and distribution of O-acetylated N-acetylneuraminic acids, the endogenous substrates of the hemagglutinin-esterases of murine coronaviruses, in mouse tissue. Glycoconjugate Journal 23:73–84.

73. Reuter G, Pfeil R, Stoll S, Schauer R, Kamerling JP, Versluis C, Vliegenthart JF. 1983. Identification of new sialic acids derived from glycoprotein of bovine submandibular gland. European journal of biochemistry / FEBS 134:139–143.

74. Pepper DS. 1968. The Sialic Acids of Horse Serum with Special Reference to Their Virus Inhibitory Properties. Biochimica et biophysica acta 156:317–325.

75. Dickson RP, Erb-Downward JR, Martinez FJ, Huffnagle GB. 2016. The Microbiome and the Respiratory Tract. Annu Rev Physiol 78:481–504.

76. Sicard J-F, Le Bihan G, Vogeleer P, Jacques M, Harel J. 2017. Interactions of Intestinal Bacteria with Components of the Intestinal Mucus. Frontiers in Cellular and Infection Microbiology 7.

77. McAuley JL, Gilbertson BP, Trifkovic S, Brown LE, McKimm-Breschkin JL. 2019. Influenza Virus Neuraminidase Structure and Functions. Frontiers in Microbiology 10.

78. Laborda P, Wang S-Y, Voglmeir J. 2016. Influenza Neuraminidase Inhibitors: Synthetic Approaches, Derivatives and Biological Activity. Molecules 21:1513.

79. Lin YP, Xionga X, Whartona SA, Martinc SR, Coombsa PJ, Vachieria SG, Christodouloub E, Walkerb PA, Liua J, Skehela JJ, Gamblinb SJ, Haya AJ, Danielsa RS, McCauleya JW. 2013. Evolution of the receptor binding properties of the influenza A (H3N2) hemagglutinin. Proceedings of the National Academy of Sciences 110:2676–2676.

80. Wagner R, Matrosovich M, Klenk HD. 2002. Functional balance between haemagglutinin and neuraminidase in influenza virus infections. Reviews in Medical Virology 12:159–166.

81. Guo H, Rabouw H, Slomp A, Dai M, van der Vegt F, van Lent JWM, McBride R, Paulson JC, de Groot RJ, van Kuppeveld FJM, de Vries E, de Haan CAM. 2018. Kinetic analysis of the influenza A virus HA/NA balance reveals contribution of NA to virus-receptor binding and NA-dependent rolling on receptor-containing surfaces. PLOS Pathogens 14:e1007233.

82. Pritchett TJ, Paulson JC. 1989. Basis for the potent inhibition of influenza virus infection by equine and guinea pig alpha 2-macroglobulin. Journal of Biological Chemistry 264:9850–9858.

83. Wasik BR, Voorhees IEH, Barnard KN, Alford-Lawrence BK, Weichert WS, Hood G, Nogales A, Martínez-Sobrido L, Holmes EC, Parrish CR. 2019. Influenza Viruses in Mice: Deep Sequencing Analysis of Serial Passage and Effects of Sialic Acid Structural Variation. Journal of Virology 93.

84. Hoffmann E, Krauss S, Perez D, Webby R, Webster RG. 2002. Eight-plasmid system for rapid generation of influenza virus vaccines. Vaccine 20:3165–3170.

85. Kiripolsky J, McCabe LG, Gaile DP, Kramer JM. 2017. Myd88 is required for disease development in a primary Sjögren's syndrome mouse model. Journal of Leukocyte Biology 102:1411–1420.

86. Cao C, Wang WJ, Huang YY, Yao HL, Conway LP, Liu L, Voglmeir J. 2017. Determination of Sialic Acids in Liver and Milk Samples of Wild-type and CMAH Knock-out Mice. Journal of Visualized Experiments.

87. Chokhawala HA, Huang S, Lau K, Yu H, Cheng J, Thon V, Hurtado-Ziola N, Guerrero JA, Varki A, Chen X. 2008. Combinatorial chemoenzymatic synthesis and high-throughput screening of sialosides. ACS Chem Biol 3:567–576.

88. Yu H, Yu H, Karpel R, Chen X. 2004. Chemoenzymatic synthesis of CMP-sialic acid derivatives by a one-pot two-enzyme system: comparison of substrate flexibility of three microbial CMP-sialic acid synthetases. Bioorg Med Chem 12:6427–6435.

89. Yu H, Chokhawala HA, Huang S, Chen X. 2006. One-pot three-enzyme chemoenzymatic approach to the synthesis of sialosides containing natural and non-natural functionalities. Nat Protoc 1:2485–2492.

90. Feng KH, Gonzalez G, Deng L, Yu H, Tse VL, Huang L, Huang K, Wasik BR, Zhou B, Wentworth DE, Holmes EC, Chen X, Varki A, Murcia PR, Parrish CR. 2015. Equine and canine influenza H3N8 viruses show minimal biological differences despite phylogenetic divergence. Journal of Virology 89:JVI.00521–15.

CHAPTER FOUR

Influenza A viruses serially passaged in different MDCK cell lines show few sequence variations across genomes except in HA

Karen N. Barnard, Brian R. Wasik, Brynn K. Alford-Lawrence, Jessica J. Hayward, Wendy Weichert, Ian E.H. Voorhees, Edward C. Holmes, Colin R. Parrish. Influenza A viruses serially passaged in different MDCK cell lineages show few sequence variations across genomes except in HA. *Manuscript in preparation*.

4.1 ABSTRACT

Deep sequence analysis provides an opportunity to follow the emergence and dynamics of virus mutations in real time. Viruses are grown in cell culture for research and for vaccine development. However, cell lines used to grow virus are often not derived from the same host species or tissue that the virus naturally replicates in. The selective pressures of culturing virus *in vitro* is still only partially understood. We passaged human H3N2, H1N1pandemic, and canine H3N2 influenza A viruses (IAV) in different lineages of MDCK cells. MDCK cells are naturally heterogeneous, and they are a tractable model for engineering sialic acid (Sia) receptor expression. We passaged viruses in different lineages of MDCK cells, including those engineered to express different forms of Sia receptor, including α2,3- and α2,6-linkages or expression of *N*-glycolylneuraminic acid (Neu5Gc) or *N*-acetylneuraminic acid (Neu5Ac) forms. While the different MDCK variants were genetically identical, they varied in IAV susceptibility. MDCK Type II cells had lower infection efficiency and virus production, and were highly reliant on protease presence in infection media. When viruses were passaged in the different cells they showed only small numbers of consensus-level mutations, most of which were in the HA gene. Both human IAV showed selection for single nucleotide minority variants in the HA stem across cell types, and more variants arose at low frequency in the receptor binding site in virus passaged in cells with Neu5Gc. Canine H3N2 also showed minority variants near the receptor-binding site in cells with Neu5Gc and those expressing α2,6-linkages.

4.2 IMPORTANCE

The adaptability of viruses is a fundamental property that allows their sustained success in nature in the face of varying host environments and immune responses that attempt to control their replication and spread. Growth of viruses in culture is widely used for their study and for

preparing vaccines, yet the selections that cell passaging imposes on viruses is often poorly understood. Here we use deep sequence analysis to define in detail how three different influenza A viruses respond to passaging in different lineages of canine MDCK cells which are commonly used for their growth, and in MDCK cells engineered to express different forms of their cell surface receptor, sialic acid. The results show that only a few changes in the virus population sequences become widespread, and these are primarily in the HA gene.

4.3 INTRODUCTION

High levels of genetic variation are a characteristic feature of RNA viruses. Evolution in the face of antibody immunity, anti-viral drug treatment, and emergence into new host species are necessary for viruses to continue to successfully propagate and transmit over extended periods of time. While the natural and experimental evolution of viruses has been described many times, the underlying processes of sequence variation, genetic drift, and selection that drive evolution of viral genomic sequences at a whole population scale are still not well understood. In this study, we define the detailed dynamics of virus evolution in an *in vitro* model, allowing us to better understand key aspects of the selection pressures that act on virus populations, particularly host cell and receptor adaptation.

The growth of viruses in culture is widely used to assess their properties and many "wild type" viruses used in research and virus strains used in vaccine development have been extensively passaged in culture. Yet the selection pressures that cell culture imposes on these viruses are often poorly understood. Often, viruses are grown in cell lines that are derived from different species and tissues that the virus does not infect during a natural infection. Influenza A viruses (IAV) are a segmented, negative-sense RNA virus that replicates in respiratory tissues in mammals and in the gastrointestinal tracts of birds (1, 2). However, Madin-Darby Canine Kidney

(MDCK) cells have been considered the gold standard cell line for the culturing of IAV since the mid-1970s (3). These cells were isolated from the kidney of a dog and established in culture and have features of proximal tubule epithelial cells (4, 5). Their use for growing IAV likely stems from their ability to grow high titers of many avian and human IAV strains, to remain attached to the growth substrate in the presence of the trypsin used in infection media to cleave and prime the hemagglutinin (HA) glycoprotein for infection, and human viruses appear to show relatively few mutations after passage in MDCK cells, although some selection of HA mutations may occur (3, 6). In contrast, growth of human IAV isolates in embryonated chicken eggs, long used for isolation and growth of many viruses, often results in the selection of mutations in HA associated with the adaptation of viruses to the predominant α2,3-linkage of the sialic acid (Sia) receptor present, which is often referred to as the "avian receptor" form (7, 8). These HA mutations that arise during egg passaging may alter the antigenic structure of the HA, which can have consequences for viruses used in vaccines (7, 9, 10).

The original MDCK cells, including the MDCK-NBL2 available from the American Type Culture Collection, have been shown to be highly heterogeneous, and many alternate MDCK cell lineages now exist (4). A number of studies have shown that different MDCK clones vary significantly in their properties and have been broadly grouped into "Type I" and "Type II" cells (5, 11–14). MDCK-Type I cells are generally sub-cloned from a low passage of the parental MDCK cell line and have a small, flat, spindle-like morphology with strong tight junctions (giving high electrical resistance) and higher density of NA-K proton pumps (11, 12, 15). In comparison, MDCK-Type II cells have been sub-cloned from a high passage of MDCK cells and are large, cuboidal cells that grow in clusters and are characterized as having "leaky" tight junctions (giving low electrical resistance), with fewer NA-K proton pumps, and a slower growth

rate (11, 12, 15). The MDCK-Type I and MDCK-Type II cells also differ in their lipid composition and their ability to develop hemicysts resulting from transport of liquid through the monolayer (12, 14). These different cell lineages provide interesting models for examining the effects of *in vitro* passaging on IAV variation. Previous research using a variety of different sub-clones of MDCK showed that Type-II-like cells supported a lower replication of IAV compared to Type-I-like clones, possibly due to differences in endogenous protease expression (16). However, it is not known whether the different MDCK cell lineages can select for different IAV variants.

One classic paradigm in studies of IAV host range is that IAV tropism is determined by the linkage of Sia to the underlying carbohydrate chain, or glycan, on the surface of cells (1). Human-adapted IAV strains preferentially bind to α2,6-linked Sia, the predominant linkage found in the upper respiratory tract of humans (17). In contrast most avian IAV strains prefer an α2,3-linked Sia, the dominant form found in the gastrointestinal tract of birds where the virus replicates (18). Pigs have been described as a "mixing vessel" due to the fact that they express both α2,6- and α2,3-linked Sia in their respiratory tract, allowing for the potential co-infection by both mammalian and avian IAV strains, which would permit reassortment (19). However, in addition to linkage type there is considerable variation in Sia structures, including the addition of a variety of chemical modifications (20, 21). The most basic Sia form is *N*-acetylneuraminic acid (Neu5Ac), and the hydroxylated form of Neu5Ac, *N*-glycolylneuraminic acid (Neu5Gc), is one of the most common Sia variants. This modification is synthesized by cytidine 5'-monophosphate-*N*-acetylneuraminic acid hydroxylase (CMAH) that is highly expressed in many mammals, but has been independently lost in some vertebrate lineages (22). Many natural hosts of IAV only have Neu5Ac due to loss of CMAH function, including humans, western breeds of

dogs, and birds, while other hosts such as horses and pigs express high levels of Neu5Gc in their respiratory tract and other tissues (22–25). While some effects of Neu5Gc have been characterized in functional assays on hemagglutinin (HA) and neuraminidase (NA) (26–28), it remains unclear how or if this modified Sia acts as a selective pressure on the virus.

MDCK cells may also be used as a model for modifying Sia properties by engineering the enzymes involved in Sia modification or linkages. Plasmid expression of β-galactoside α2,6-sialyltransferase 1 (ST6GAL1) has been used to produce MDCK-SiaT1 cells which express higher levels of α2,6-linked Sia compared to wild-type MDCK cells (29). Human viruses passed on standard MDCK cells may show mutations associated with selection by α2,3-linked Sia, while fewer mutations were observed after passage of viruses in MDCK-SiaT1 cells (6, 30, 31). MDCK-SiaT1 have therefore become the preferred cell line for culturing human IAV, particularly for clinical isolates (32). MDCK cells appear to express little or no detectable Neu5Gc, due to their origin in a western breed of dog, and therefore can act as a model for Neu5Gc presence through transfection with a CMAH expression plasmid.

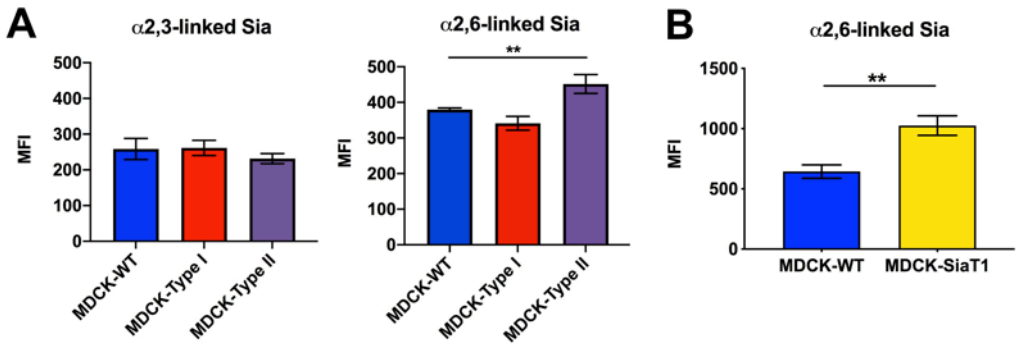
In this study, we passaged three different IAV strains (A/California/04/2009 pH1N1, A/Wyoming/2003 H3N2, and A/canine/Illinois/2015 H3N2) in different MDCK cell lines to determine the effects on sequence variation within virus populations. Each virus stock was derived from reverse genetic plasmids and serially passaged on MDCK-NBL2, MDCK-SiaT1, MDCK-CMAH, MDCK-Type I, or MDCK-Type II cells. The virus populations were analyzed in detail using deep sequencing to determine the roles of different cell types and Sia forms on IAV selection.

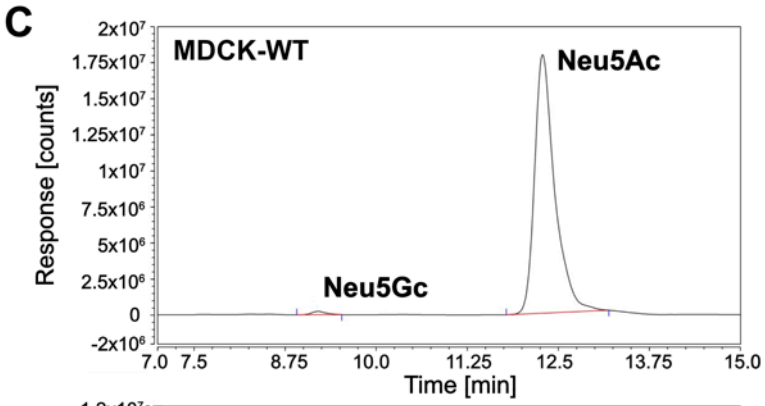
4.4 RESULTS

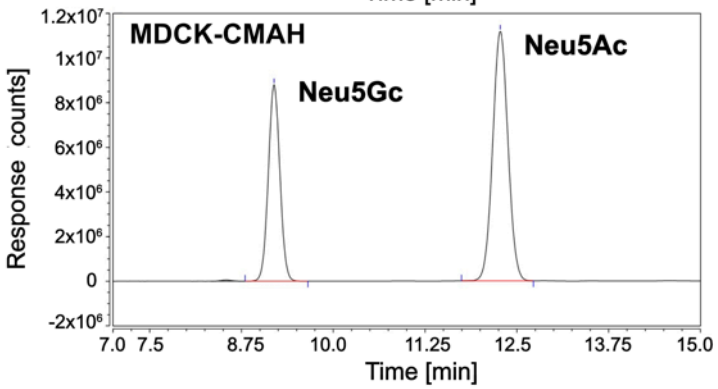
Development of MDCK cell lines. We used the standard ATCC line of MDCK-NBL2 (MDCK-WT) as the parental line to create the different MDCK cell lines for this study. MDCK-WT cells express both α2,3- and α2,6-linked Neu5Ac Sia, and only trace amounts of Neu5Gc (<1%), either due to uptake from the fetal bovine serum in growth media or possibly from residual CMAH activity (**Figs. 4.1A,C**) (33). To create MDCK cells with higher levels of the human-like α2,6-linked Sia, MDCK-WT cells were transfected with an expression plasmid carrying the human *ST6GAL1* gene to generate MDCK-SiaT1 cells (29). Compared to MDCK-WT, MDCK-SiaT1 cells showed increased staining in flow cytometry for the *Sambucus nigra* (SNA) lectin which recognizes α2,6-linked Sia (**Fig. 4.1B**). MDCK-WT cells were transfected with an expression plasmid carrying CMAH and sub-cloned to create a cell line that expressed Neu5Gc (MDCK-CMAH). MDCK-CMAH cells expressed ~40% of their total Sia as Neu5Gc (**Fig. 4.1C**) and maintained consistent Neu5Gc expression across cell passages. These levels of Neu5Gc are comparable to those seen in pig respiratory tissues (**Table 4.1**).

Other MDCK cell variants examined here include clones originally isolated and characterized as described in Nichols et al (11), which were a kind gift from Dr. William Young (University of Kentucky), that were defined as MDCK-Type I and MDCK-Type II cell lines. These MDCK-Type I and MDCK-Type II cells expressed the same amount of α2,3-linked Sia as MDCK-WT when stained with *Maakia amurensis type I* (MAA I) lectin, but MDCK-Type II had higher staining for α2,6-linked Sia when stained with SNA lectin (**Fig. 4.1A**).

Viral passaging and analysis. The experimental design used in this study is diagrammed in **Figure 4.2**. To compare results between different IAV strains, we tested three viruses:







	% Neu5Ac	% Neu5Gc	% Total Sia
MDCK-WT	97.9	0.9	100
MDCK-CMAH	62.7	37.3	100

Figure 4.1

A) Plant lectins MAA (α2,3-linked Sia) and SNA (α2,6-linked Sia) were used to compared linkage types between MDCK-WT, MDCK-Type I, and MDCK-Type II cells via flow cytometry. **B)** Plant lectin SNA was used to determine amount of α2,6-linked Sia on MDCK-WT cells compared to MDCK-SiaT1 cells via flow cytometry. **C)** HPLC analysis of total Sia collected from MDCK-WT and MDCK-CMAH cells.
Data analyzed by one-way ANOVA using PRISM software.
* = p-value ≤0.05; ** = p-value ≤0.01; *** = p-value ≤0.001.

Table 4.1.

HPLC analysis for total sialic acid content for tissue from pigs. Data given as percent of total Sia averaged across three runs

Tissue	%Neu5Gc	%Neu5Ac	%Neu5,9Ac2
Upper Lung	36.8	51.0	0.7
Lower Lung	37.9	52.0	0.5
Lower Trachea	14.4	81.4	0.3
Upper Trachea	14.1	83.9	0.3

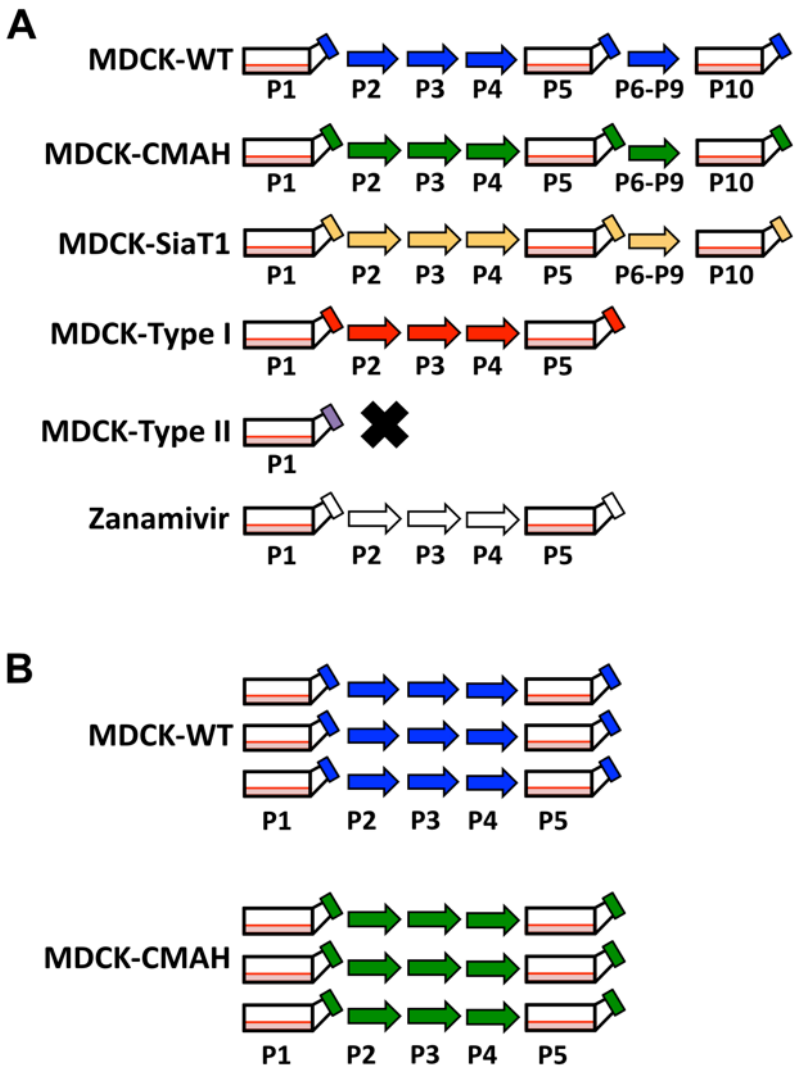


Figure 4.2.

A) Diagram of replicate one for virus passaging in MDCK cells. Passages were completed for human IAV strains wyoH3N2 and pH1N1, along with CIV H3N2 at a low MOI of 0.001. Virus was sequenced for passage 1, 5, and 10. MDCK-Type II cells did not produce enough progeny virus despite increasing MOI, so no virus was sequenced for this cell type. For Zanamivir passaged virus, an MOI of 0.005 was used and Zanamivir concentration was increased during each passage from 0.01 μM to 1 μM.
B) Replicate two was used to determine if there was variation between virus populations passaged in MDCK-WT and MDCK-CMAH, three individual populations of each virus were passaged 5 times. Virus in passage 1 and 5 were sequenced for each population.

pandemic H1N1 (pH1N1, A/California/04/2009), a seasonal H3N2 (wyoH3N2, A/Wyoming/2003)), and a canine H3N2 strain (CIV H3N2, A/canine/Illinois/2015). Viruses were recovered from reverse genetics plasmids as previously described (34), and stocks prepared after two passages in MDCK-WT cells. Each virus was titrated by $TCID_{50}$ assay in MDCK-WT cells, and passaged five times at low MOI of 0.001 in MDCK-WT, MDCK-CMAH, MDCK-Type I, or MDCK-SiaT1 cells. Viruses were then passaged for an additional five passages in MDCK-WT, MDCK-CMAH, and MDCK-SiaT1 cells, giving a total of ten passages. Passaging of virus in MDCK-Type II cells failed, which will be covered in more detail. The same passaging methods were repeated in MDCK-WT and MDCK-CMAH as a second replicate using three separate passage series per cell type for five passages to determine the reproducibility of the variants seen. As a positive control for virus selection, we also passaged each virus five times in the presence of increasing concentrations of 4-guanidino-2,4-dideoxy-2,3-dehydro-N-acetylneuraminic acid (Zanamivir), as previously described (35, 36).

Viral RNA was extracted from the stock virus, and from viruses recovered after passage 1, 5, and 10, and viral genomes was amplified using whole genome RT-PCR as previously described (37). Libraries were prepared using Illumina Nextera XT and sequenced using Illumina MiSeq with 2x250 base paired end reads. To control for error during sequencing, the plasmid stocks for each virus were also deep sequenced (**Fig. 4.3**), and to control for RT-PCR introduced error, a variant calling threshold of 2% was used. Average coverage for all gene segments for pH1N1 and wyoH3N2 (**Fig. 4.4A,B**) was 1000 to 10000 reads per base pair. The coverage for most gene segments for CIV H3N2 were comparably high, apart from the HA gene segment, which had consistently lower coverage (between 50-200 reads per base pair) for reasons that are unclear. CIV H3N2 HA-specific primers were therefore used to prepare

additional libraries with higher coverage (**Fig. 4.4C**). The second replicate of the wyoH3N2 passages could not be completed to passage 5 in MDCK-WT cells, as the virus titers dropped below detectable limits during passage 3 for all three flasks.

MDCK-Type II cells less permissive to IAV infection. All three viruses replicated well in the MDCK-WT and MDCK-Type I cells, while replication was less efficient in MDCK-Type II cells despite MDCK-Type II cells expressing similar amounts of Sia on their surface as both MDCK-WT and MDCK-Type I cells (**Fig. 4.1A**). Lower numbers of infected cells were seen at early time points of infection (**Fig. 4.5A,B**) and significantly less virus was collected in the supernatant after 48 hours (**Fig. 4.5C**). We also found that virus infections in these were very sensitive to the presence of trypsin in the media, as had previously been reported (16).

Passage with Zanamivir. Passaging of virus with the NA inhibitor Zanamivir has previously been shown to select for mutations in both NA and HA (35, 38, 39). As a positive control for selection, we therefore passaged the three IAVs in increasing concentrations of Zanamivir starting at 0.01 μM in passage one up to 1.0 μM in passage 5. For pH1N1 virus, no single nucleotide variants (SNVs) arose in NA or in the HA1 domain of HA where the receptor-binding site is located (**Fig. 4.6A**). Some SNVs did arise in the stalk of pH1N1 HA, however these same SNVs also arose in all other pH1N1 virus populations passaged in the different MDCK cell lines, as will be discussed in more detail. For wyoH3N2, no SNVs arose in NA, but some did arise in HA (**Fig. 4.6B**), including A163T (62%) and L244V (28%) in the HA1 domain. In CIV H3N2, the variant S231G in HA reached near fixation and is near to a R229I variant that was previously found to confer NA inhibitor resistance in cell-based assays (40). CIV H3N2 also saw fixation of the A27T variant in NA that was present in the stock virus and

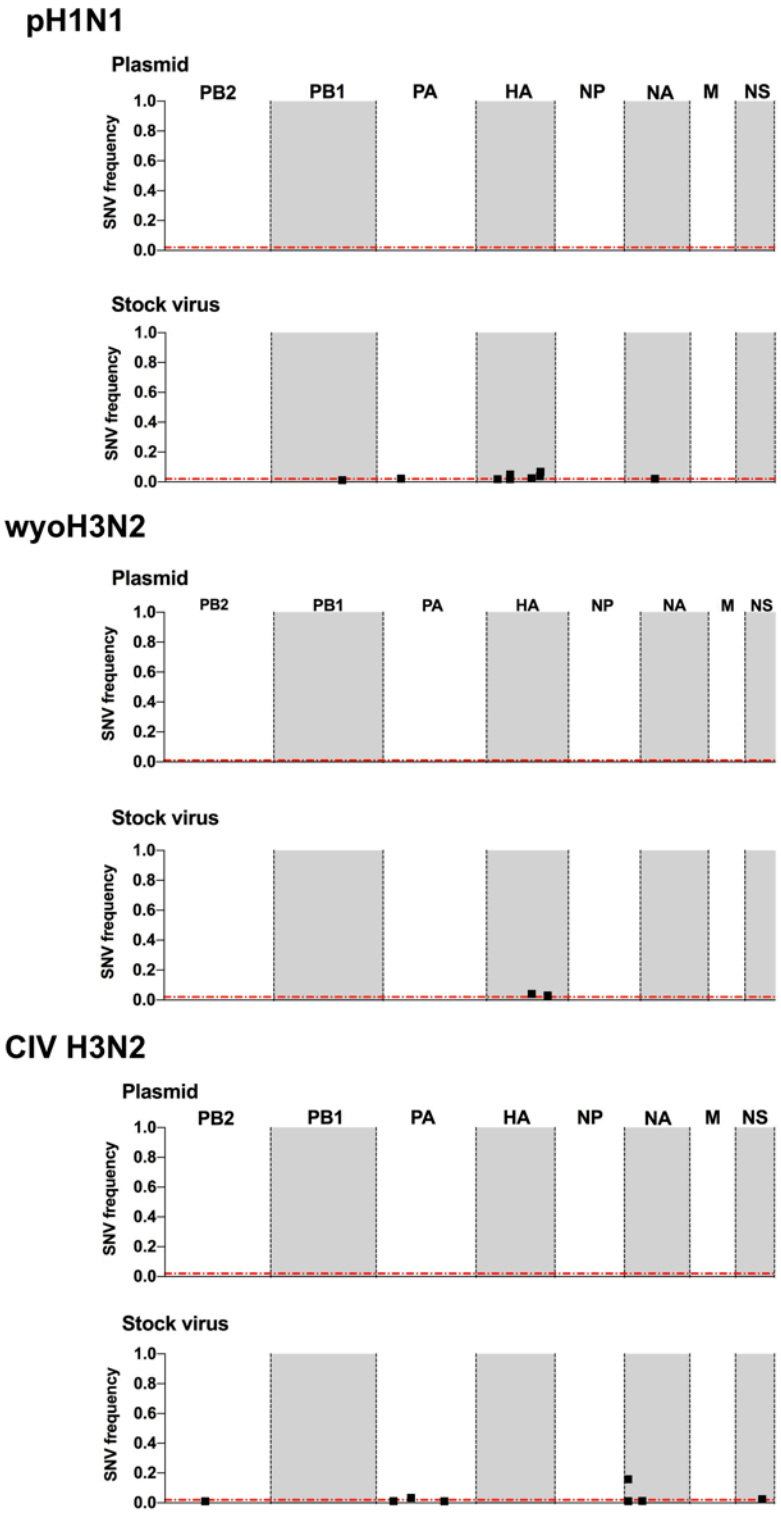


Figure 4.3

Single nucleotide variants (SNV) in plasmids and stock virus were analyzed for pH1N1, wyoH3N2, and CIV H3N2. Data is shown for the whole genome. Non-synonymous mutations are filled in shapes while synonymous mutations are open shapes.

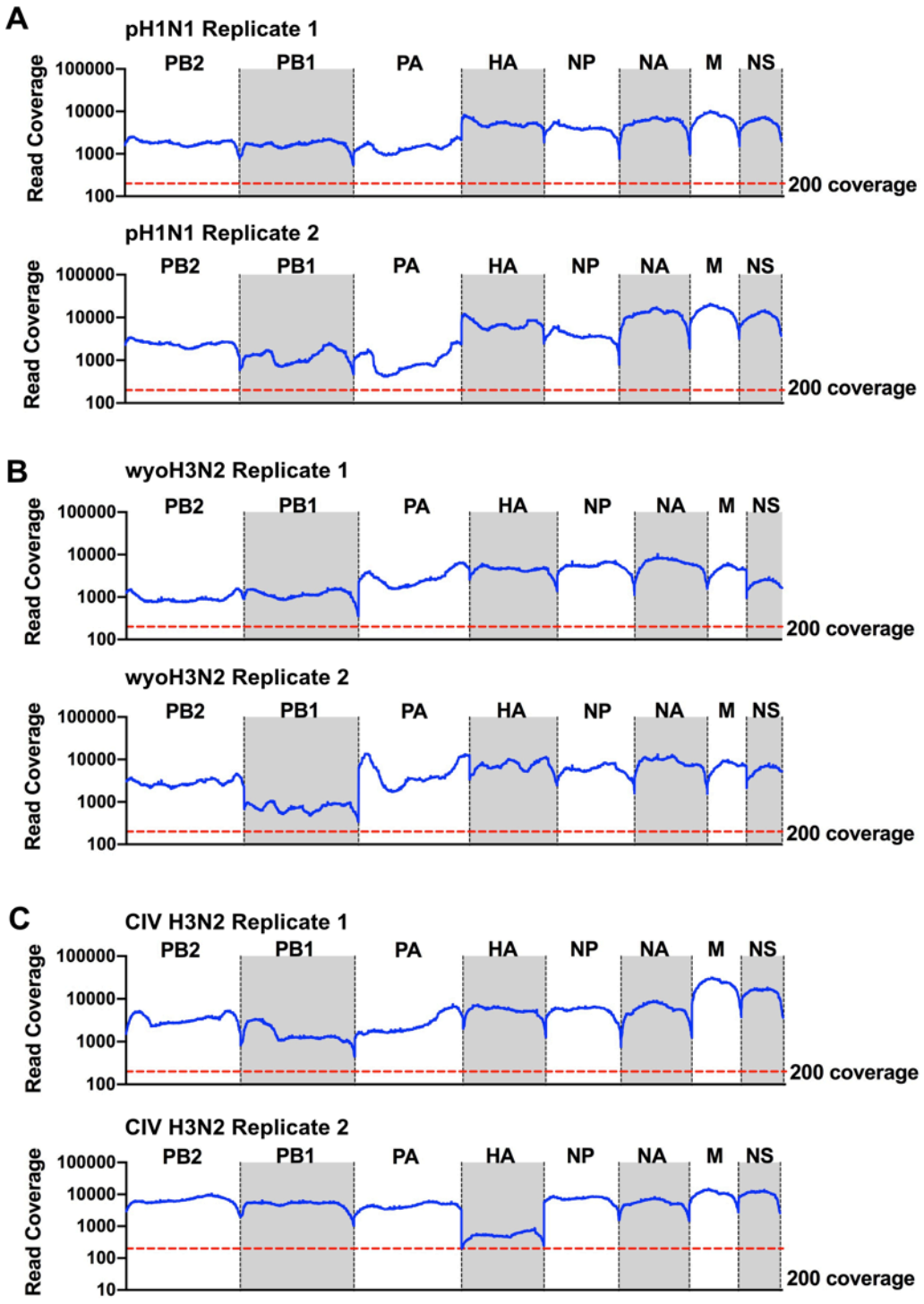


Figure 4.4.

Average coverage for each genome segment for replicate one and replicate two were determined for **A)** pH1N1, **B)** wyoH3N2, and **C)** CIV H3N2. For analysis, a cutoff of 200 reads per base pair was set.

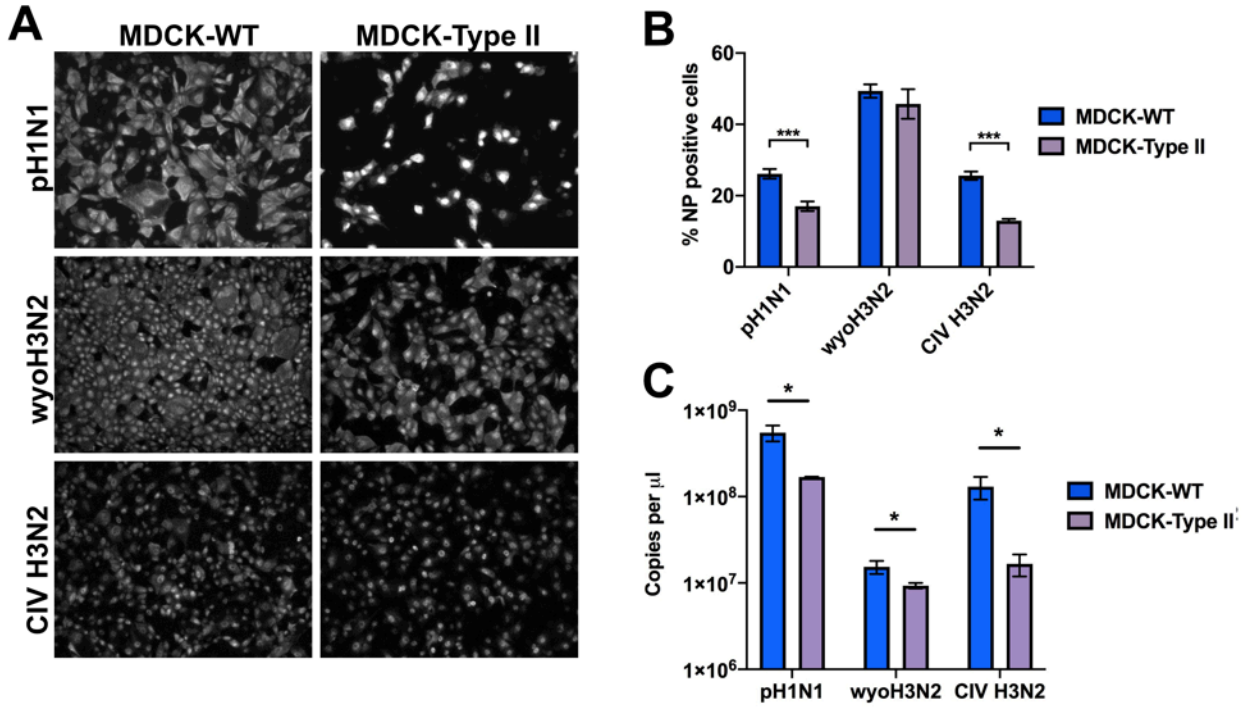


Figure 4.5.

MDCK-Type II cells had low number of infected cells and less progeny virus than MDCK-WT cells. **A)** MDCK-WT and MDCK-Type II cells were infected at an MOI of 0.25 for 12 hours, fixed and stained for NP protein. Cell were imaged at 10x. **B)** Flow cytometry of MDCK-WT and MDCK-Type II cells infected at an MOI of 0.25 for 6 hours, fixed and stained for NP protein. **C)** Supernatant from MDCK-WT and MDCK-Type II cells infected at an MOI of 0.01 for 48 hours was collected and virus titers determined by RT-qPCR for the M gene segment. Data analyzed by t-test using PRISM software.
* = p-value ≤0.05; ** = p-value ≤0.01; *** = p-value ≤0.001.

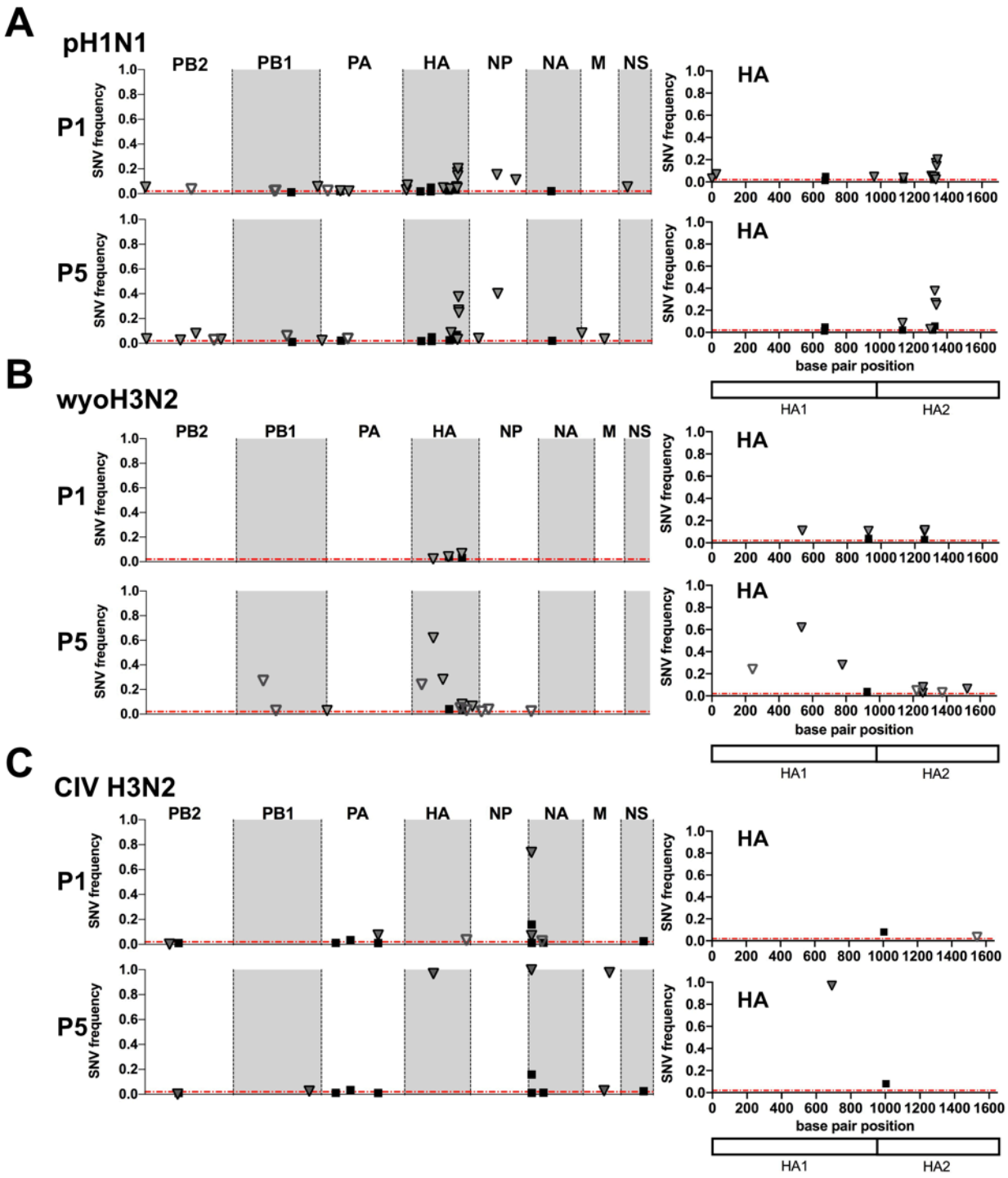


Figure 4.6.

Single nucleotide variants (SNV) were analyzed for **A)** pH1N1, **B)** wyoH3N2, and **C)** CIV H3N2 passaged in Zanamivir. Data is shown for the whole genome for passage one and passage five on the left panel, while the nucleotide position of SNVs for HA are mapped on the right side. Non-synonymous mutations are filled in shapes while synonymous mutations are open shapes.

fixation of a G228R variant in M1 in the M segment, both of which also reached high frequency in other CIV H3N2 passaging populations.

Passage of pH1N1 in MDCK-WT, MDCK-SiaT1, and MDCK-Type I cells. When pH1N1 was passed in MDCK-WT, MDCK-SiaT1, and MDCK-Type I cells, SNVs were detected in most gene segments; however, the majority of these SNVs were at low frequency, synonymous, and not consistently carried between passages. The exception to this was in the HA segment encoding hemagglutinin, which had the greatest number of non-synonymous SNVs with very few synonymous variants (**Fig. 4.7; Table 4.2**). Most SNVs introduced substitutions into the HA2 stalk domain, while few SNVs were seen in HA1 corresponding to the receptor binding site domain. None of the SNVs in HA1 were selected or carried through passage 5 for MDCK-Type I or passage 10 for MDCK-WT and MDCK-SiaT1. Minority SNVs in the stalk domain increased in frequency between passages p1 and p5 from 25%-45% up to 45%-85%, with the highest frequency being at amino acid position 445, a variant that was also present at low frequency in the stock virus (**Table 4.2**). By passage 10, minority variants at position 445 in both MDCK-WT and MDCK-SiaT1 were fixed, with K445E fixed in MDCK-SiaT1 cells and K445M fixed in MDCK-WT (**Fig. 4.7A,B, Table 4.2**). Interesting, neither variant was carried to passage 5 in MDCK-Type I, but an SNV near by, T436N, reached a frequency of 58% (**Fig. 4.7C, Table 4.2**). MDCK-WT passaged virus also saw fixation of another stalk variant, R509G, by passage 10 that arose between passage 5 and 10 (**Fig. 4.7A**). Other SNVs clustered in this same region of HA2, including at residues 436, 439, 441, 442, 443, 446, and 448, with some reaching high frequencies of 10-50%. Glycosylation sites on HA were maintained across all passaged virus except in one population of MDCK-WT-passaged virus in replicate two, where loss of the

proposed glycosylation site at site 21 occurred in passage 1 at a frequency of 13% and was maintained through passage 5 at a frequency of 12% (**Fig. 4.8C; Table 4.2**).

Passaging of pH1N1 in MDCK-CMAH cells. Initial passage of virus in MDCK-CMAH cells showed more SNVs that introduced coding changes into the HA1 domain, so that a second replicate of five passages in three separate virus populations was performed to determine if there was variation in SNVs (**Fig. 4.3**). MDCK-CMAH passaged virus populations showed most SNVs in the HA gene segment, along with some changes in PA that were primarily synonymous and not consistent between virus populations (**Fig. 4.8B**). MDCK-CMAH passaged pH1N1 showed emergence of SNVs in the HA2 stalk domain at positions 441, 442, and 445 as seen in virus passed in other MDCK cell lines. Several apparently MDCK-CMAH specific SNVs arose near the receptor-binding site of HA during the later passages in these cells (**Fig. 4.8, Table 4.2**). These include polymorphisms of residues 138 and 228 of HA1 in all MDCK-CMAH-passaged virus populations in both replicate one and two, although these were not maintained through passage 10. Residues 128 and 228 fall within the well-characterized outer loops of the receptor-binding site and affect α2,3- or α2,6-Sia binding preference (41). Other SNVs near the receptor-binding site also arose in different MDCK-CMAH passaged virus populations, although there was variability between virus populations (**Fig. 4.8C**). However, none of the SNVs near the receptor-binding site were present at higher frequency than 15% by passage 5, except one passage 5 population in replicate two (A138S, ~30%). After 10 passages in MDCK-CMAH cells, only a low frequency variant, I269N (4%), was present in HA1. Some pH1N1 virus passed in MDCK-CMAH cells also showed loss of glycosylation sites at 21 and 33 at low frequency (<10%) (**Fig. 4.8B,C**), similar to that seen in some populations of MDCK-WT passed virus.

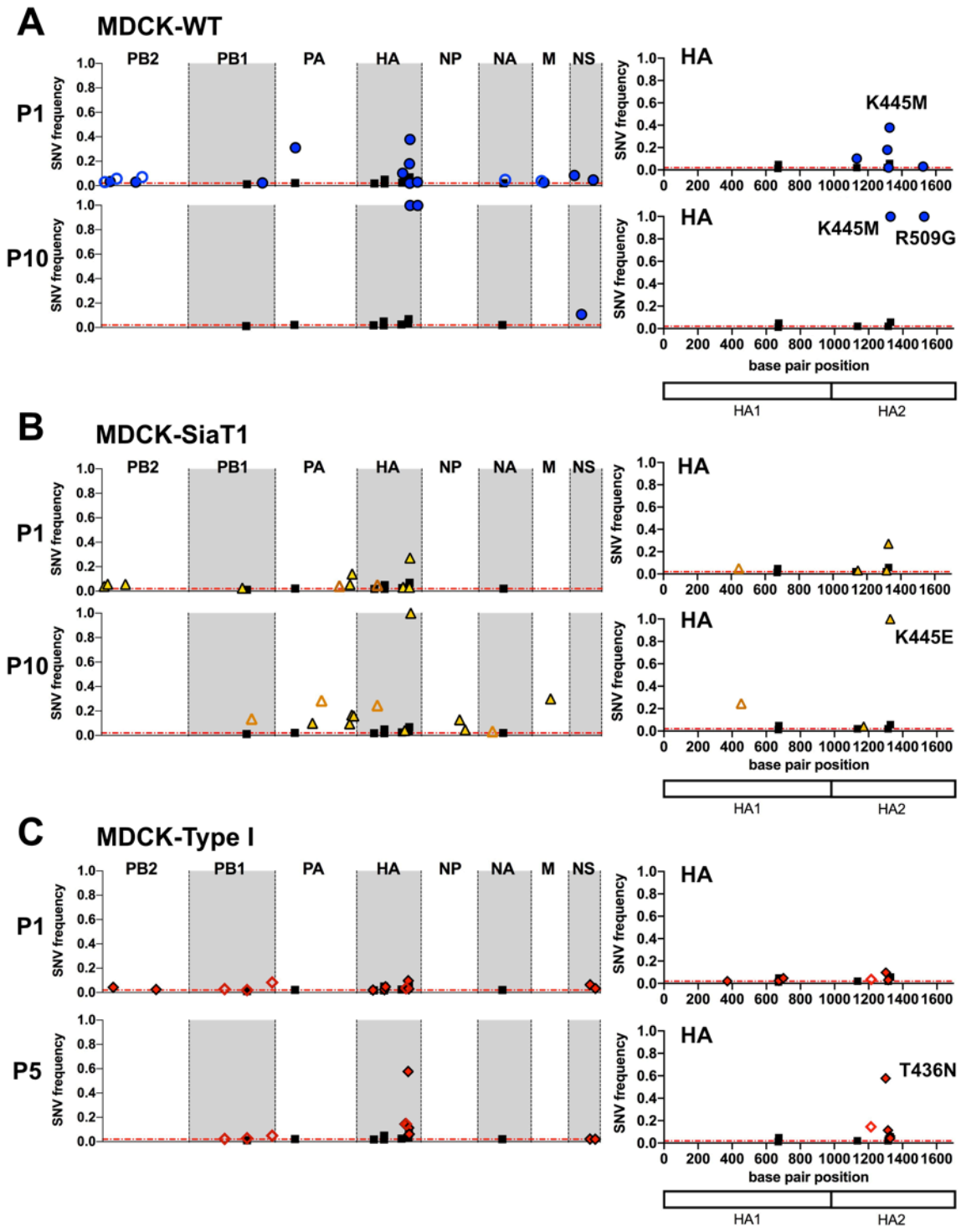


Figure 4.7

Single nucleotide variants (SNV) were analyzed for pH1N1 passaged in **A)** MDCK-WT, **B)** MDCK-SiaT1, and **C)** MDCK-Type I cells. Data is shown for the whole genome for passage one and passage 10 (**A,B**) or passage 5 (**C**) on the left panel, while the nucleotide position of SNVs for HA are mapped on the right side with particular amino acid changes noted. Non-synonymous mutations are filled in shapes while synonymous mutations are open shapes.

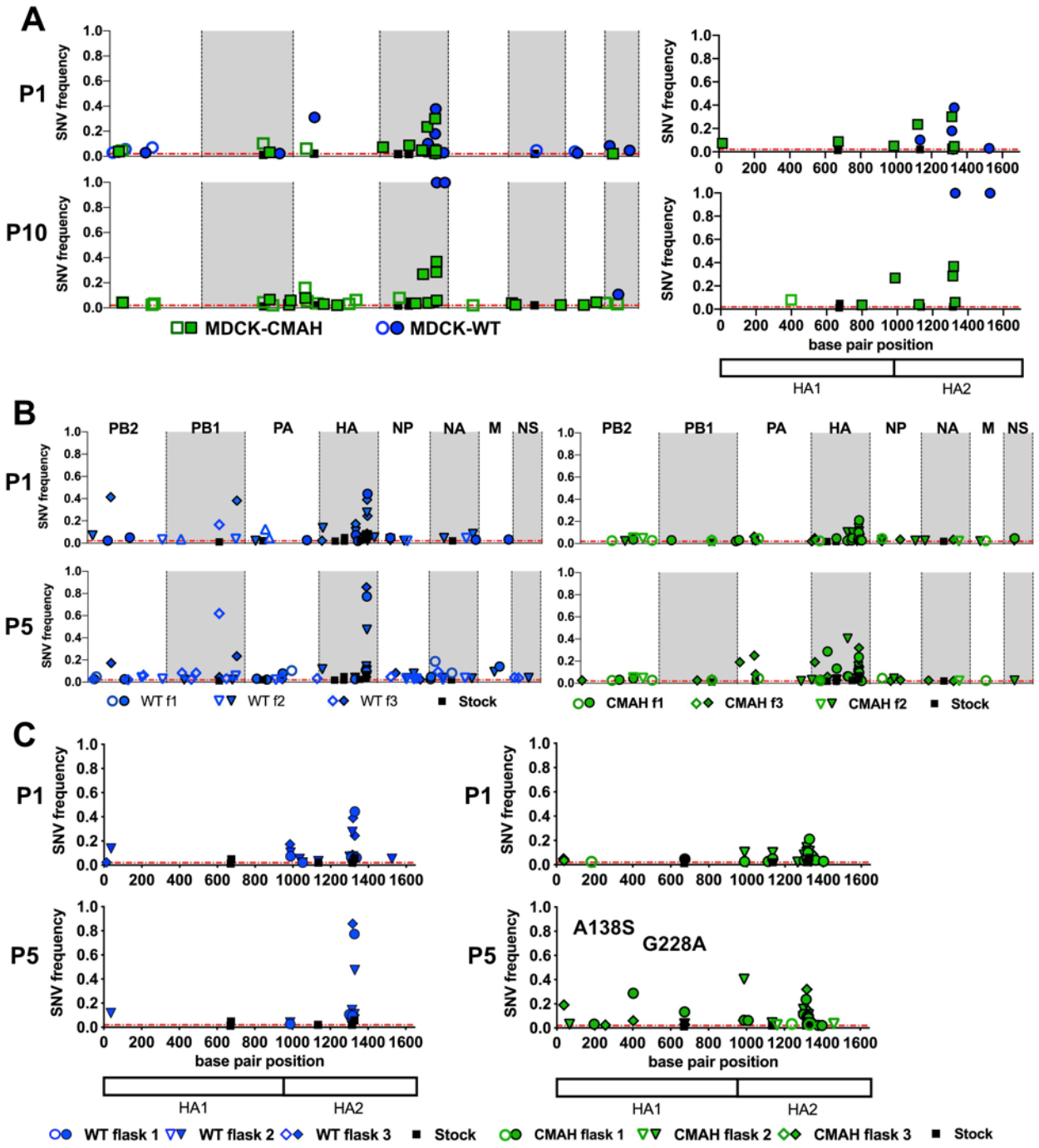


Figure 4.8

Single nucleotide variants (SNV) are mapped for pH1N1 passaged in MDCK-WT and MDCK-CMAH. **A)** SNVs from replicate one are mapped for the whole genome (left panel) and for HA alone (right panel) for passage 1 and passage 10. **B)** Whole genome map for SNVs from the three independent virus populations of pH1N1 passaged in MDCK-WT or MDCK-CMAH. **C)** The nucleotide position of SNVs mapped for HA only from the three virus populations of pH1N1 in MDCK-WT or MDCK-CMAH. Particular amino acid changes as result of SNVs are also noted. Non-synonymous mutations are filled in shapes while synonymous mutations are open shapes.

Table 4.2. Single Nucleotide Variants in pH1N1 segments HA and NA

HA Segment					
	Replicate	**Flask**	**Passage 1**	**Passage 5**	**Passage 10**
	Stock		G228A (5%), V379I (2%), K445M (6%),		
MDCK-WT	Rep 1	F1	V381I (10%), D441N (18%), N443S (2%), K445M (38%), R509G (3%)	*L59 (3%),* V381I (6%), D441N (28%), K445M (57%), N446D (10%)	K445M (100%), R509G (100%)
	Rep 2	F1	F332L (7%), Y353F (2%), Y439D (6%), K445E/M (6%, 44%), Y448C (6%)	F332L (3%), T436N (11%), S442L (10%), K445M (77%)	
		F2	T23P (13%,) F332L (10%), D348Y (5%), V381I (3%), T436N (7%), D441Y/N/E (27%, 5%, 3%), S442L (7%), K445M (2%), R509G (5%)	T23P (12%,) F332L (4%), T436N (6%), D441Y/N (13%, 14%,), K445M (11%), N446D (47%)	
		F3	I15L (2%), L331I (17%), D441N/G (3%, 9%), S442L (39%), K445M (24%)	D441N (5%), S442L (86%), K445M (10%)	
MDCK-CMAH	Rep 1	F1	Y17H (7%), G228A (9%), F332L (5%), E376D (24%), D441N (30%), S442L (3%), K445M (5%)	G228A (3%), F332L (11%), Y363C (2%), E376D (16%), D441N (49%), S442L (8%), K460E/M (3%, 8%)	*T136 (8%),* I269N (4%), F332L (27%), E376D (4%), D441N (29%), S442L (37%), K445E (6%)
	Rep 2	F1	*L71 (2%),* G228A (5%), F332L (3%), A373V (3%), V381I (5%), T436N (3%), D441N/Y (5%, 2%), N443S (10%), K445E/M (3%, 21%), K456R (4%), Y470N (3%)	E75V (3%), A138S (29%), G228A (13%), F332L (6%), I339L (6%), *D430 (3%)*, T436N (11%), D441Y/N (24%, 9%), S442L (9%), N443Y/S (4%, 10%), K445E/M (10%, 8%), *L462 (3%)*, I462T (2%), F469L (2%)	
		F2	F332L (10%), V381I (10%), A425T (2%), T436N (8%), D441Y/N (14%, 9%), K445M (11%)	N33D (3%), G228A (4%), F332L (40%), V381I (4%), *N404 (3%)*, T436N (15%), D441Y/N (21%, 16%), K445M (11%), *P504 (3%)*	
		F3	D11N (2%), T23P (5%), D24N (4%), V381I (4%), T436N (5%), D441N/Y (14%, 7%), S442L (7%), N443Y (4%), K445M (9%), Y448C (3%), V451I (9%)	T23P (19%), D93N (2%), A138S (6%), G228A (3%), L331I (6%), T436N (13%), D441N/Y (16%, 9%), S442L (32%), K460M (6%), V451I (6%)	
MDCK-SiaT1	Rep 1	F1	*L151 (5%)*, I385V (3%), D441G (3%), K445E (27%)	D441G (2%), K445E (94%)	*L154 (25%),* T393I (4%), K445E (100%)
MDCK-Type I	Rep 1	F1	P128L (2%), G228A (2%), V237L (5%), *E407 (4%*), T436N (10%), D441N (3%)	*E407 (15%)*, T436N (58%), D441N (11%), K460E/M (6%, 4%)	
MDCK-Type II	Rep 1	F1			
MDC+ Zanamivir	Rep 1	F1	D11N (3%), A19G (7%), I323F (5%), V381I (4%), T436N (5%), D441N (5%), K445Q/M (2%, 17%), N446S (14%), Y448C (20%)	V381I (9%), T436N (3%), K445Q (38%), N446S (27%), Y448C (25%)	

Table 4.2 continued

NA Segment					
	Replicate	**Flask**	**Passage 1**	**Passage 5**	**Passage 10**
MDCK-WT	Rep 1	F1	*V234 (5%)*		
	Rep 2	F1	V453M (3%)	N28K (3%), S31P (5%), *N71 (19%), C233 (8%)*	
		F2	G147R (5%), *R361 (5%)*, I427V (8%)	*N146 (2%),* G147R (4%), *R361 (4%)*, I427V (3%)	
		F3		*Y100 (9%)*	
MDCK-CMAH	Rep 1	F1			N42D (4%), T72I (3%), C446W (2%)
	Rep 2	F1			
		F2	C49Y (3%), *P377 (2%)*		
		F3	F322L (4%)	V81I (3%), F322L (2%)	
MDCK-SiaT1	Rep 1	F1		I263V (17%), I389L (3%)	*Q136 (3%)*, I263V (13%)
MDCK-Type I	Rep 1	F1			
MDCK-Type II	Rep 1	F1			
MDCK+ Zanamivir	Rep 1	F1			

Italics give synonymous mutations, non-italics give non-synonymous mutations

Passage of wyoH3N2 in MDCK-WT, MDCK-SiaT1, and MDCK-Type I cells. The wyoH3N2 results varied, and the virus did not propagate past passage 3 in the MDCK-WT cells in the second replicate, likely due to the poor cell culture growth of recent H3N2 human viruses (42, 43). In other passage series in all MDCK cell lines, most SNVs were seen in the HA gene segment (**Fig. 4.9A, B; Table. 4.3**). Changes of residues in HA1 near the receptor-binding site of HA included residue N216K, which arose after five passages in both MDCK-WT and MDCK-Type I at 16% and 2% respectively (**Fig. 4.9A,D**). In MDCK-WT, N216K reached over 80% by passage 10. However, few SNVs in HA1 were seen in the MDCK-SiaT1 passaged virus and none were carried to passage 10, which is consistent with previous reports (**Fig. 4.9B**) (31).

WyoH3N2 virus passaged in MDCK-WT also showed a low frequency loss of a glycosylation site at residue 165 in HA1 (N165T), which was maintained at a frequency of 5-10% through passage 10 in MDCK-WT passaged virus. SNVs in the stalk domain of HA2 occurred after passage in MDCK-WT, MDCK-Type I, and MDCK-SiaT1 cells (**Fig. 4.9; Table 4.3**). HA residue 404 had two variants, G404R and G404E, present in most populations at high frequency in passage 5 (16-75%) and G404E was maintained through passage 10 in MDCK-WT passaged virus at 12% frequency (**Fig. 4.9A**). Both variants were present in the stock virus at low frequencies (**Fig. 4.3**). Other stalk mutations include G465S, which was completely fixed in MDCK-SiaT1 cells in passage 5 and maintained through passage 10 (**Fig. 4.9B**), and G463D, which reached over 40% in MDCK-Type I cells (**Fig. 4.9C**). Neither of these mutations was detected in the stock virus or the MDCK-WT grown virus. For the other gene segments, only a few low frequency (<3%) SNVs occurred, but most were synonymous mutations and were not retained across passages (**Fig. 4.9**).

Passage of wyoH3N2 in MDCK-CMAH cells. WyoH3N2 passaged in MDCK-CMAH showed more low frequency SNVs near the receptor-binding site than were seen in other MDCK cell lines, although results varied between virus populations (**Fig 4.10; Table 4.3**). SNVs included substitutions of HA residues V186I, P221S, and Y233H at low frequencies (<5%) in both replicate 1 and replicate 2. WyoH3N2 passaged in MDCK-CMAH saw a high frequency SNV at residue N216K by passage 10 in replicate 1 (92%) similar to MDCK-WT passaged virus, as well as SNVs G404R/E (20-50% by passage 5), G465S (~10% by passage 5), and the loss of the glycosylation site at position 165 (2-16%) that were also seen in other cell passaged populations. However, only the D487N variant was maintained through all 10 passages in replicate 1. Few non-synonymous SNVs arose in the NA gene segment across wyoH3N2 passages; however, SNVs did arise in passage 10 that led to changes in the proposed secondary Sia binding site of NA in MDCK-WT (S367G, 11%) and MDCK-CMAH (R428K, 37%). As this secondary binding site has only recently been described (44), the importance of this site for infection in cell culture and the effects of these mutations will require additional research. Additionally, one SNV appeared repeatedly in the PA gene segment in MDCK-WT (42% in passage 5, lost by passage 10) and in MDCK-CMAH (10% in passage 5, 5% in passage 10) in replicate one, F600S. This variant is located in the C-domain of PA that interacts with PB1, however position 600 is not directly involved in this interaction and it is unclear if this variant has any function.

CIV H3N2 in MDCK-WT, MDCK-SiaT1, and MDCK-Type I cells. The CIV H3N2 showed little consistent selection for HA stalk mutations, with only a few SNVs arising at very low frequency including a SNV at position I335K that was also present in the stock virus at a frequency of 8% and reached 10-30% by passage 5 but was not maintained through passage 10

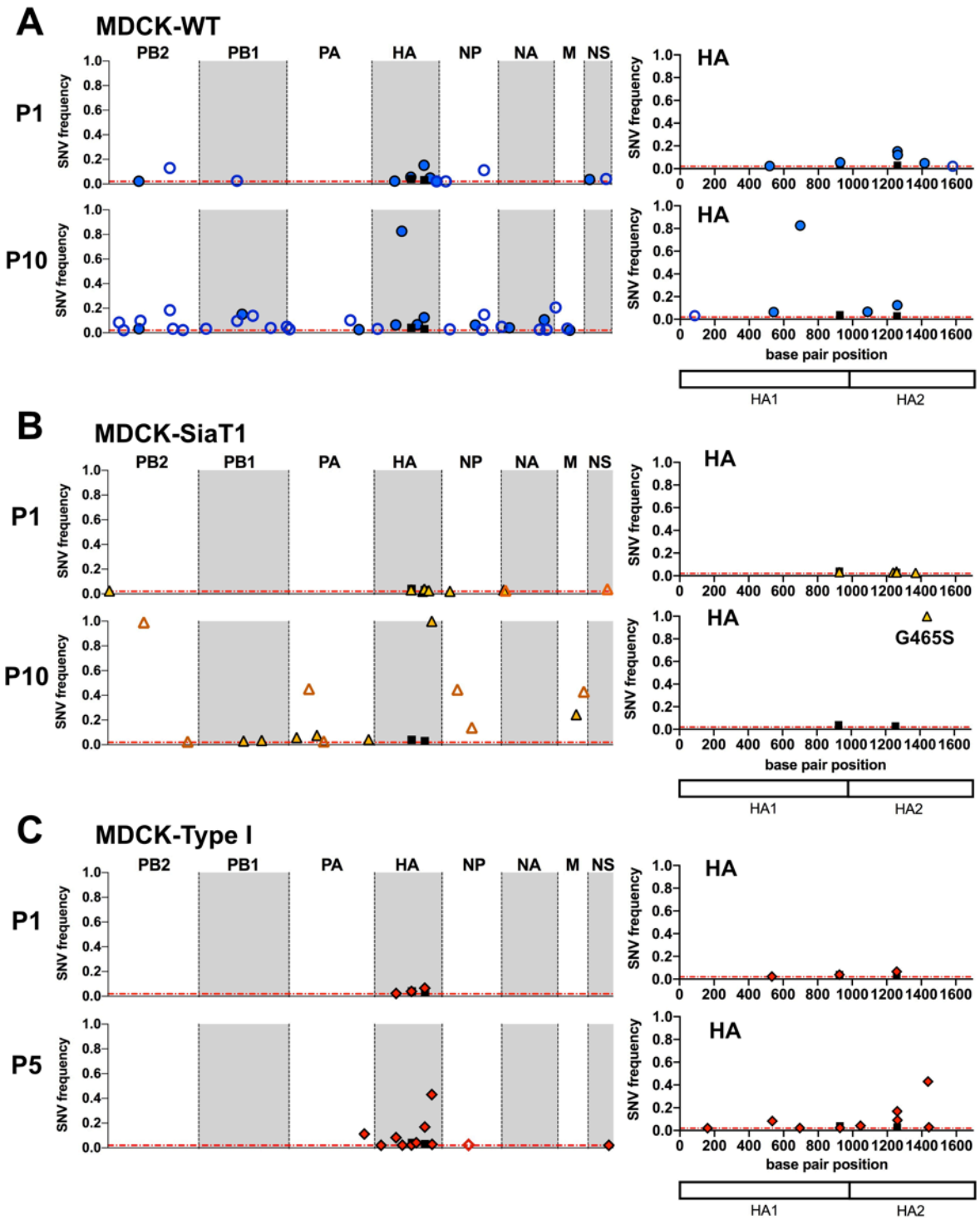


Figure 4.9

Single nucleotide variants (SNV) were analyzed for wyoH3N2 passaged in **A)** MDCK-WT, **B)** MDCK-SiaT1, and **C)** MDCK-Type I cells. Data is shown for the whole genome for passage one and passage ten (**A,B**) or passage 5 (**C**) on the left panel, while the nucleotide position of SNVs for HA are mapped on the right side with particular amino acid changes noted. Non-synonymous mutations are filled in shapes while synonymous mutations are open shapes.

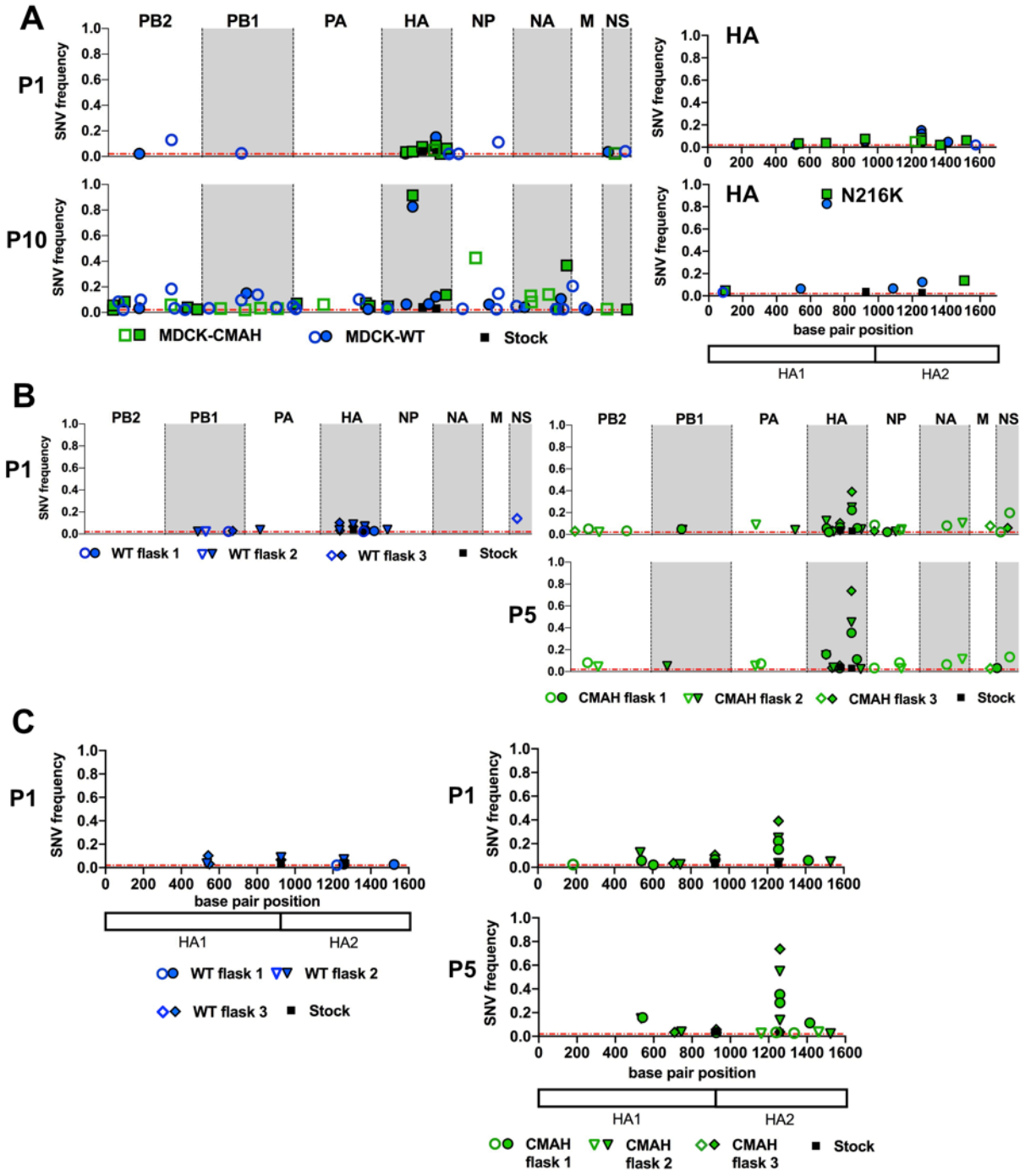


Figure 4.10

Single nucleotide variants (SNV) are mapped for wyoH3N2 passaged in MDCK-WT and MDCK-CMAH. **A)** SNVs from replicate one are mapped for the whole genome (left panel) and for HA alone (right panel) for passage one and passage ten. **B)** Whole genome map for SNVs from the three independent virus populations of wyoH3N2 passaged in MDCK-WT or MDCK-CMAH. **C)** The nucleotide position of SNVs mapped for HA only from the three virus populations of wyoeH3N2 in MDCK-WT or MDCK-CMAH. Non-synonymous mutations are filled in shapes while synonymous mutations are open shapes.

Table 4.3. Single Nucleotide Variants in wyoH3N2 segments HA, PA, and NA

HA Segment					
	Replicate	**Flask**	**Passage 1**	**Passage 5**	**Passage 10**
Stock			P293L (4%), G404R/E (3%, 1%)		
MDCK-WT	Rep 1	F1	L157S (2%), P293L (6%), G404R/E (15%, 12%), R456K (5%), *G510 (2%)*	N165T (9%), N216K (16%), K397R (7%), G404R/E (19%, 47%)	*T12 (3%),* N165S (6%), N216K (83%), V347M (7%), G404E (12%)
	Rep 2	F1	*K391 (2%),* G404R (3%), R492K (3%)	virus failed to replicate to p5	
		F2	A163T (4%), P293L (9%), G404R/E (7%, 3%)	virus failed to replicate to p5	
		F3	N165T (10%), T167N (3%), P293L (7%), G404R/E (6%, 4%)	virus failed to replicate to p5	
MDCK-CMAH	Rep 1	F1	A163T (4%), N216K (4%), P293L (7%), *K391 (5%*), G404R/E (9%, 6%), D441Y (2%), R492K (6%)	H17R (12%), *T37 (8%),* A163T (3%), N165T (2%), T167N (2%), N216K (55%), *S247 (3%),* P293L (10%), G333S (7%), K397R (4%), G404R/E (3%, 4%), G463D (3%), G465S (9%), D487N (8%)	H17R (5%), N216K (92%), D487N (14%)
	Rep 2	F1	N165T (5%), V186I (2%), P293L (7%), G404R/E (22%, 15%), R456K (6%)	N165T (16%), P293L (3%), G404R/E (35%, 28%), R456K (11%)	
		F2	A163T (12%), Y233H (3%), P293L (4%), G404R/E (25%, 3%), A495T (5%)	A163T (15%), Y233H (4%), G404R/E (45%, 14%), R492K (2%)	
		F3	P221S (3%), P293L (10%), G404E (39%)	P221S (3%), P293L (5%), G404R/E (4%, 74%)	
MDCK-SiaT1	Rep 1	F1	P293L (3%), K391R (2%), K397R (3%), G404R/E (4%, 3%), D441N/Y (2%, 3%), G465S (1%)	G465S (99%)	G465S (99%)
MDCK-Type I	Rep 1	F1	A163T (2%), P293L (4%), G404R (7%)	N38D (2%), A163T (8%), N216K (2%), P293L (2%), G333S (4%), G404R/E (17%, 9%), G463D (43%), G465S (3%)	
MDCK-Type II	Rep 1	F1			
MDCK+ Zanamivir	Rep 1	F1	A163T (11%), P293L (11%), G404R/E (10%, 11%)	*T65 (24%),* A163T (62%), L244V (28%), *K391 (5%),* G404R/E (3%, 8%), *S442 (3%),* R492K (6%)	

Table 4.3 continued

PA Segment					
	Replicate	**Flask**	**Passage 1**	**Passage 5**	**Passage 10**
MDCK-WT	Rep 1	F1		F600S (42%)	*D3 (3%), D516 (10%),* S588P (3%)
	Rep 2	F1		virus failed to replicate to p5	
		F2	I145T (4%), F600S (4%)	virus failed to replicate to p5	
		F3		virus failed to replicate to p5	
MDCK-CMAH	Rep 1	F1		*N228 (6%), K340 (4%),* V541M (3%), C561Y (2%), *C562 (2%*), F600S (10%)	*N228 (6%),* S571* (7%), F600S (5%)
	Rep 2	F1	*K281 (5%)*	*K281 (7%)*	
		F2	*N228 (9%)*	*N228 (5%)*	
		F3			
MDCK-SiaT1	Rep 1	F1		*R168 (11%), G555 (4%)*	L65I (6%), *R168 (45%)*, G235A (8%), *I292 (3%*), V668I (4%)
MDCK-Type I	Rep 1	F1			
MDCK-Type II	Rep 1	F1			
MDCK+ Zanamivir	Rep 1	F1		D3G (3%)	

NA Segment					
	Replicate	**Flask**	**Passage 1**	**Passage 5**	**Passage 10**
MDCK-WT	Rep 1	F1		M51I (3%), *L255 (5%), D463 (9%)*	*I8 (5%),* I73T (4%), *N329 (3%),* S367G (11%), *N387 (3%), D463 (21%)*
	Rep 2	F1		virus failed to replicate to p5	
		F2		virus failed to replicate to p5	
		F3		virus failed to replicate to p5	
MDCK-CMAH	Rep 1	F1		*P282 (2%)*	*G137 (13%), V143 (8%), P282 (14%),* N358K (3%), R428K (37%)
	Rep 2	F1	*L255 (8%)*	*L255 (7%)*	
		F2	*R403 (10%)*	*R403 (11%)*	
		F3			
MDCK-SiaT1	Rep 1	F1	I20T (3%), *L35 (3%)*		
MDCK-Type I	Rep 1	F1			
MDCK-Type II	Rep 1	F1			
MDCK+ Zanamivir	Rep 1	F1			

Italics give synonymous mutations, non-italics give non-synonymous mutations

in any virus populations (**Fig. 4.3**; **Fig. 4.11**). This I335K variant was positioned just after the cleavage site between HA1 and HA2 and may be associated with protease susceptibility. Other SNVs in HA varied between the replicates in MDCK-WT cells. Residue S186I reached over 50% in passage 5 and over 90% frequency by passage 10 in replicate one, while in the three populations in replicate two had only low frequency SNVs in HA by passage 5 (**Fig. 4.11A**, **Fig. 4.12B**). Passage in both MDCK-Type I and MDCK-WT cells had a variant at position 186 reach ~50% frequency (**Fig. 4.11C**). In MDCK-SiaT1 cells, the SNV at position S219P in the receptor-binding site reached a frequency of 98% by passage 10 (**Fig. 4.11B**). Previous reports have shown mutations at position 219 associated with changes in receptor binding preference in human H3N2 virus passaged in eggs (45).

An A27T variant in the NA was present in the stock virus (14%) and reached 60-95% by passage 5 in all cells (**Fig. 4.11D**). In the passage 10 virus population in MDCK-WT cells, A27T reached 97% frequency while in passage 10 virus from MDCK-SiaT1, the population was split between A27T (48%) and L35Q (48%). One MDCK-WT passage 5 virus population in replicate two also had the L35Q variant (80%) (**Fig. 4.12D; Table 4.4**). Several additional SNVs occurred in the NA transmembrane and stalk domain by passage 5 and passage 10, including residues 28, 30, 35, 36, and 37 at frequencies of 5-70%. Similar mutations were also seen in canine H3N2 viruses when they were passaged on feline cells, so could be a general cell culture adaptation for increased stability or budding efficiency (46).

The other gene segments of CIV H3N2 virus showed few consistent minority SNVs between passages and cell types. C489S variant in PA was present in the stock virus and was maintained through passage 5 and passage 10 of all virus populations at <10% frequency. An NS segment variation within the NEP protein at position 35 was seen in all virus populations at low

(<5%) frequency in passage 1 and was maintained in MDCK-WT passaged virus through passage 10 at low frequency. Also, a G228R variant in M1 reach high frequency in MDCK-WT passaged virus by passage 10. This variant also reached fixation in virus passaged in Zanamivir (**Table 4.4**).

CIV H3N2 in MDCK-CMAH cells. These viruses showed a similar inconsistency between replicate one and replicate two virus populations (**Fig 12. A,B**), with more SNVs in HA1 in replicate two MDCK-CMAH passed virus compared to MDCK-WT passed virus, with variants near the receptor-binding site at residues 183, 186, 188, 219, and 221, with many reaching frequencies of 20-70% (**Fig. 4.12B,C; Table 4.4**). However, in replicate one only the S186I variant reached high frequency (96% by passage 10) in MDCK-CMAH passed virus, with other minority variants at positions 129, 183, 221, and 252 present at 10% or lower frequencies. The S186I is also nearly fixed in MDCK-WT passaged CIV H3N2 in replicate one by passage 10 (**Fig. 4.12A**). However, the S186I appeared in all MDCK-CMAH passaged viruses at a frequency of 17% to 67% by passage 5 across both replicates, but was detected only in replicate one in MDCK-WT. The variants S219P and P221L were seen at low frequency across several of the MDCK-CMAH populations by passage 5 and are within the 220-loop of the receptor-binding site. P221L only occured in MDCK-CMAH passaged virus (43% frequency in passage 5, replicate two) while S219P also occurred in MDCK-SiaT1 passaged virus (98% by passage 10). Similar to passage in other MDCK cell lines, viruses passaged in MDCK-CMAH also showed variation of HA residue 335 in the stalk at low frequency, as well as SNVs in NA at positions 20, 27, 28, 35, and 36, with A27T having high frequency in all passage 5 virus populations. For passage 10 virus in MDCK-CMAH, the frequency of A27T decreased due to the rise in frequency of V20A (**Table 4.4**). The low frequency C489S variant in PA, F35L in NEP sequence

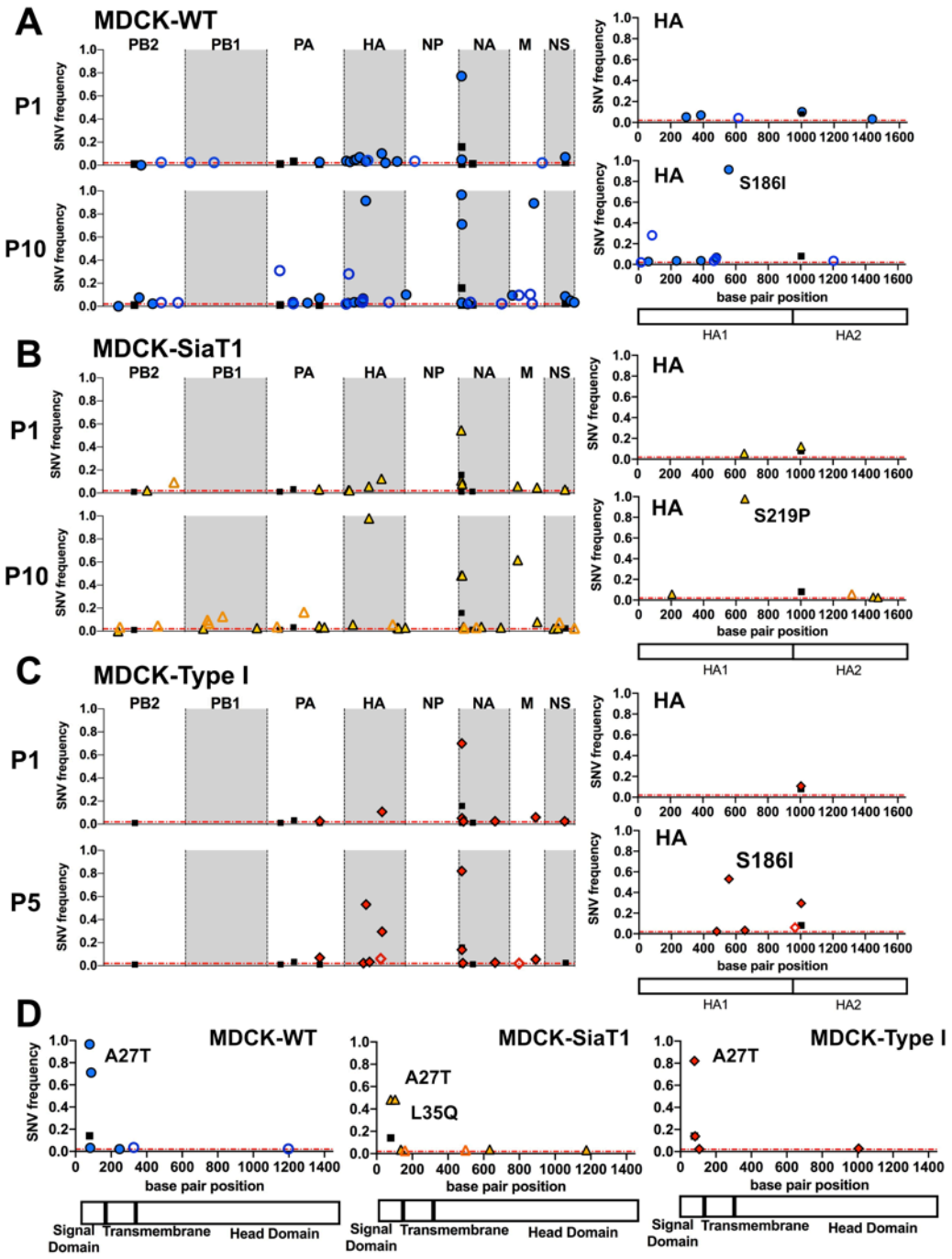


Figure 4.11

Single nucleotide variants (SNV) were analyzed for CIV H3N2 passaged in **A)** MDCK-WT, **B)** MDCK-SiaT1, and **C)** MDCK-Type I cells. Data is shown for the whole genome for passage one and passage ten (**A,B**) or passage 5 (**C**) on the left panel, while the nucleotide position of SNVs for HA are mapped on the right side with particular amino acid changes noted. **D**) SNVs mapped for NA for passage 10 for MDCK-WT and MDCK-SiaT1 or passage 5 for MDCK-Type I. Non-synonymous mutations are filled in shapes while synonymous mutations are open shapes.

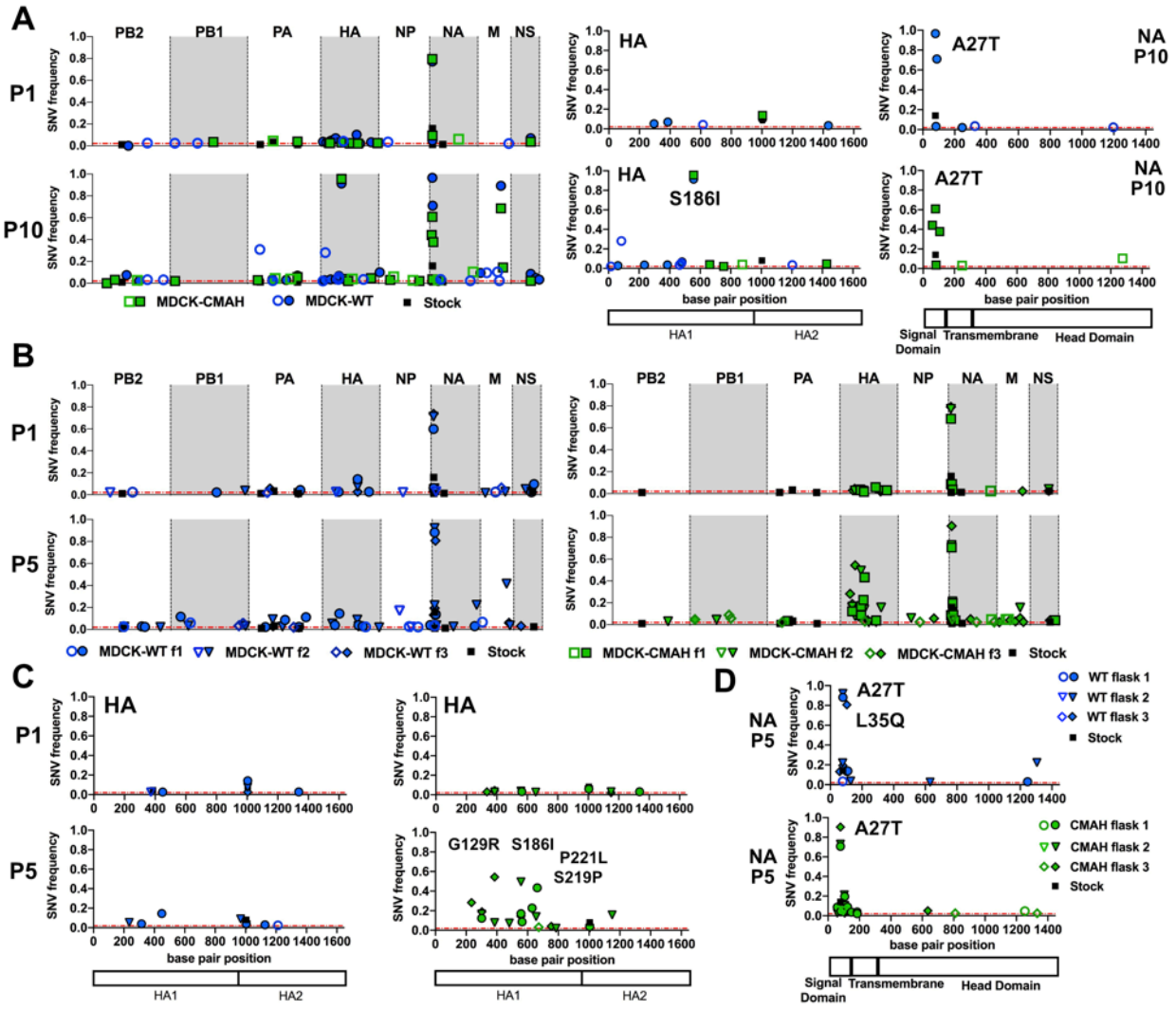


Figure 4.12

Single nucleotide variants (SNV) are mapped for CIV H3N2 passaged in MDCK-WT and MDCK-CMAH. **A)** SNVs from replicate one are mapped for the whole genome (left panels) and for HA alone (middle panels) and NA alone (right panels) for passage one and passage ten. **B)** Whole genome map for SNVs from the three independent virus populations of CIV H3N2 passaged in MDCK-WT or MDCK-CMAH. **C)** Nucleotide position of SNVs mapped for HA only from the three virus populations of CIV H3N2 in MDCK-WT or MDCK-CMAH. **D)** Nucleotide position of SNVs mapped for NA only for the three virus populations of CIV H3N2 in MDCK-WT or MDCK-CMAH. Non-synonymous mutations are filled in shapes while synonymous mutations are open shapes.

Table 4.4. Single Nucleotide Variants in CIV H3N2 segments HA, PA, NA, NS, and M

HA Segment					
	Replicate	**Flask**	**Passage 1**	**Passage 5**	**Passage 10**
Stock			I335K (8%)		
MDCK-WT	Rep 1	F1	P99S (5%), G129R (7%), *S205 (4%)*, I335K (10%), I478M (3%)	*T28 (4%),* F79L (9%), P99S (5%), G129R (10%), *T155 (14%), N159 (18%),* Y161H (19%), S186I (53%), Q189K (7%), I335K (17%), *E401 (11%)*	*G5 (2%),* P21T (3%), *T28 (28%),* F79L (4%), G129R (4%), *T155 (4%), N159 (6%),* Y161H (7%), S186I (91%), *E401 (4%)*
	Rep 2	F1	R150I (3%), I335K (14%), K446R (3%)	A106T (4%), R150I (15%), I335K (4%), Q376L (3%), *G404 (2%)*	
		F2	*G124 (3%),* I335K (8%)	F79L (5%), V323I (9%), I335K (3%)	
		F3	I335K (3%)	I335K (3%)	
MDCK-CMAH	Rep 1	F1		G129R (10%), H183Q (4%), S186I (67%), P221L (5%), I335K (14%), K508T (2%)	S186I (96%), P221L (4%), I252V (2%), *K292 (4%), A476 (5%)*
	Rep 2	F1	N188K (3%), I335K (6%), K446R (3%)	D101N (12%), S186I (17%), N188K (9%), Q210R (23%), P221L (43%), I335K (4%)	
		F2	G129R (3%), S186I (4%), S219P (3%), I335K (5%), V384M (3%)	D101N (17%), G129R (8%), Y161H (8%), S186I (50%), S219P (14%), I262S (2%), I335K (2%), V384M (16%)	
		F3	V112A (3%), G129R (4%), V384M (2%)	F79L (28%), D101N (19%), G129R (54%), S186I (14%), *R240 (3%)*, I252V (4%), I335K (4%)	
MDCK-SiaT1	Rep 1	F1	S219P (6%), I335K (12%)	S186I (3%), S219P (82%), I335K (7%)	A69S (6%), S219P (98%), *L439 (6%),* N483D (3%), D493N (3%)
MDCK-Type I	Rep 1	F1	I335K (11%)	Y161H (3%), S186I (53%), S219P (3%), *N322 (6%)*, I335K (30%)	
MDCK-Type II	Rep 1	F1			
MDCK+ Zanamivir	Rep 1	F1	*L516 (3%)*	S231G (97%)	

PA Segment					
	Replicate	**Flask**	**Passage 1**	**Passage 5**	**Passage 10**
Stock, p2			C489S (1%)		
MDCK-WT	Rep 1	F1	C489S (3%)	C489S (4%)	*I120 (31%),* D243N (4%), *L246 (2%),* V379M (3%), C489S (7%)
	Rep 2	F1	C489S (2%), R508K (4%)	T162N (2%), I354T (9%), C489S (2%), V557M (11%)	
		F2		N234D (9%, truncates PA-X), K328E (3%)	
		F3	R212S (6%), C489S (3%)	E237G (3%), *H427 (2%)*	
MDCK-CMAH	Rep 1	F1	*S250 (4%),* C489S (4%)	*I423 (3%)*, C489S (5%)	V100A (3%), *L270 (4%), I423 (4%),* C489S (6%)
	Rep 2	F1		I201T (3%)	
		F2			
		F3		*I145 (2%)*	

Table 4.4 continued

PA Segment, continued					
	Replicate	**Flask**	**Passage 1**	**Passage 5**	**Passage 10**
MDCK-SiaT1	Rep 1	F1	C489S (3%)	*I348 (26%),* C489S (5%)	*T98 (4%), I348 (17%),* C489S (4%), E538D (3%)
MDCK-Type I	Rep 1	F1	C489S (3%)	C489S (7%)	
MDCK-Type II	Rep 1	F1			
MDCK+ Zanamivir	Rep 1	F1	C489S (7%)		

NS Segment					
	Replicate	**Flask**	**Passage 1**	**Passage 5**	**Passage 10**
MDCK-WT	Rep 1	F1	F35L (7%, NEP)	F35L (4%, NEP)	F35L (9%, NEP), L87V (5%), L120F (3%)
	Rep 2	F1	F35L (2%, NEP), R211I (10%, NEP), D160Y (10%, NS1)		
		F2	T129I (5%), F35L (3%, NEP)		
		F3	F35L (3%, NEP)	D74N (3%)	
MDCK-CMAH	Rep 1	F1	F35L (4%), NEP)	F35L (4%, NEP)	F35L (2%, NEP)
	Rep 2	F1		K88R (4%)	
		F2	F35L (4%, NEP)		
		F3	F35L (3%, NEP)	F35L (4%, NEP)	
MDCK-SiaT1	Rep 1	F1	F35L (3%, NEP)	*T58 (2%)*	S87P (2%), N127T (3%), *R140 (7%), Stop (3%)*
MDCK-Type I	Rep 1	F1	F35L (3%, NEP)		
MDCK-Type II	Rep 1	F1			
MDCK+ Zanamivir	Rep 1	F1			

NA Segment					
	Replicate	**Flask**	**Passage 1**	**Passage 5**	**Passage 10**
Stock, p2			A27T (14%)		
MDCK-WT	Rep 1	F1	A27T (77%), I28L (5%)	A27T (95%), I28L (5%), A30T (4%)	A27T (97%), I28L (3%), A30T (71%), E83K (2%), *S109 (4%), D399 (2%)*
	Rep 2	F1	A27T (60%), I28L (6%), Y36H (3%), F37S (3%)	*I26* (4%), A27T (88%), Y36H (13%), F37S (14%), S416G (3%)	
		F2	A27T (71%), I28L (5%), Y36H (6%), *V50 (3%)*	I26T (22%), A27T (92%), N43S (4%), M210I (3%), R435S (22%)	
		F3	A27T (74%), I28L (3%), L35Q (3%)	V20A (13%), A27T (19%), L35Q (81%)	
MDCK-CMAH	Rep 1	F1	A27T (80%), I28L (9%), *Y284 (6%)*	V20A (9%), A27T (84%), I28L (8%), Y36H (8%), *L426 (17%)*	V20A (44%), A27T (61%), I28L (4%), Y36H (38%), *E83 (3%), L426 (10%)*
	Rep 2	F1	A27T (68%), I28L (9%), Y36H (9%), C42R (5%), *V418 (2%)*	V20A (9%), A27T (71%), I28L (5%), Y36H (20%), C42R (9%), V50A (4%), I62T/M (4%, 2%), *V418 (5%)*	

Table 4.4 continued

NA Segment, continued					
	Replicate	**Flask**	**Passage 1**	**Passage 5**	**Passage 10**
MDCK-CMAH	Rep 2	F2	A27T (77%), I28L (3%), Y36H (2%)	V20A (3%), A27T (73%), *T32 (5%)*, L35Q (21%), Y36H (11%), F37S (4%)	
		F3	A27T (79%), I28L (5%)	A27T (90%), I28L/F (4%, 3%), Y36H (3%), V212A (5%), *G270 (3%), V444 (3%)*	
MDCK-SiaT1	Rep 1	F1	A27T (55%), I28L (11%), L35Q (8%)	A27T (60%), I28L (3%), L35Q (35%)	A27T (48%), L35Q (48%), S46A (4%), *C53 (3%), P166 (3%),* V212A (4%)
MDCK-Type I	Rep 1	F1	A27T (70%), I28L (5%), C42R (2%), N336D (3%)	A27T (82%), I28L (14%), Y36H (3%), N336 (3%)	
MDCK-Type II	Rep 1	F1			
MDCK+ Zanamivir	Rep 1	F1	A27T (74%), I28L (7%), *R118 (2%)*	A27T (100%)	

M Segments					
	Replicate	**Flask**	**Passage 1**	**Passage 5**	**Passage 10**
MDCK-WT	Rep 1	F1	*E225 (2%)*	*D89 (12%),* G228R (29%)	L28F (10%), *D89 (10%), S195 (10%), V213 (2%),* G228R (89%)
	Rep 2	F1	*H162 (3%)*	*K21 (7%)*	
		F2	I27T (3%, M2)	I27T (42%, M2), R61G (4%, M2)	
		F3	*I219 (6%)*	I51V (6%, M2)	
MDCK-CMAH	Rep 1	F1		M248V (4%)	G228R (68%), M248V (14%)
	Rep 2	F1		*D89 (5%), K113 (5%)*	
		F2		G228R (16%)	
		F3	S31G (2%, M2)	E23K (2%), G129D (4%), A182T (3%), G228R (6%), S31G (3%, M2)	
MDCK-SiaT1	Rep 1	F1	V80A (6%), S260N (5%)	V80A (50%), S31N (7%, M2)	V80A (62%), S31N (8%, M2)
MDCK-Type I	Rep 1	F1	N13K (6%, M2)	*D89 (2%),* N13K(6%, M2)	
MDCK-Type II	Rep 1	F1			
MDCK+ Zanamivir	Rep 1	F1		A182T (3%), G228R (98%)	

Italics give synonymous mutations, non-italics give non-synonymous mutations

of the NS segment, and G228R variant in M were also seen in the MDCK-CMAH passaged virus.

3.5 DISCUSSION

MDCK cell types varied in virus infection efficiency. Many MDCK cell stocks, including the standard ATCC lineage, are well known to be heterogeneous and to give rise to phenotypically distinct cells when passaged only 20 or more times, and some lineages differ in susceptibility to IAV strains (6, 16). While MDCK-WT cells and MDCK-Type I cells examined here were quite susceptible and grew the virus to high titers, the MDCK Type II clone cells were infected at much lower levels by all three IAVs tested. A previous study with similar cells showed that some clones lacked necessary proteases to activate HA for infection (16). Whether there are further differences in the cells, including activation of interferon responses or trafficking of viral proteins during infection remains to be seen. However, the marked differences between the MDCK-Type II cells and the MDCK-WT and MDCK-Type I cells in terms of glycolipid composition, metabolism, and polarization could contribute to this decrease in infection efficiency (13–15).

Only low levels of variation were detected in viruses passaged in different MDCK lineages. The three viruses used here all derived from reverse genetics plasmids, so only a small number of variants were detected in the starting virus populations grown in MDCK-NBL2 cells, primarily in the HA segment. After repeated passages of the viruses in the different MDCK cell lines, only a few SNVs were present within most of the viral gene segments, and most variants were in the HA gene segment (**Fig. 4.13**), and for CIV H3N2 also the NA gene segment. It could be that there was greater variability present in the population beneath our variant frequency cutoff of 2%; however, it is likely that if any of those variants had increased fitness or

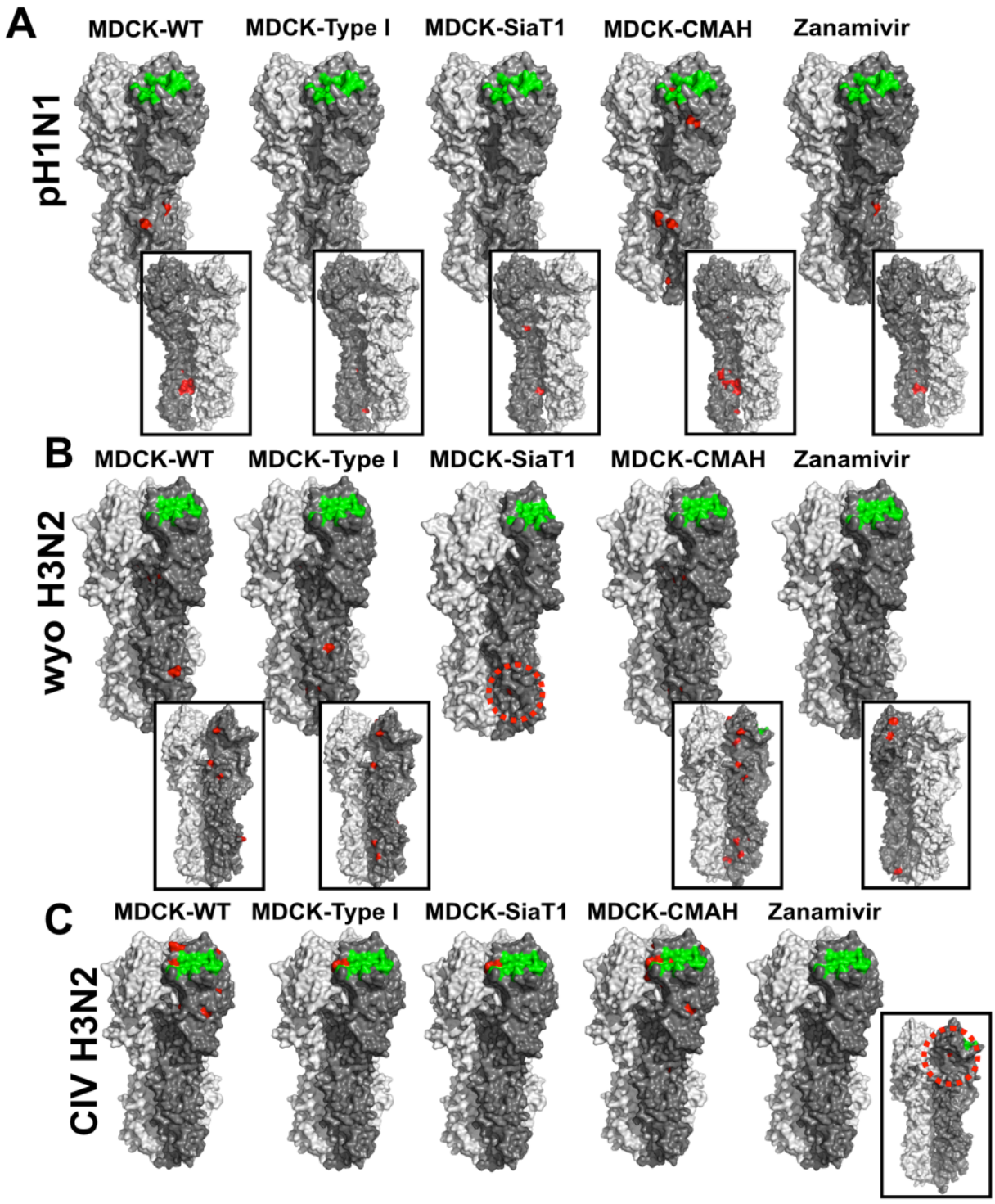


Figure 4.13

Models of IAV HA with receptor binding (green) and amino acid changes (red) for each passage virus population. Inset images show the interior of the HA trimer with one monomer removed to highlight amino acid variants that are not solvent exposed. HA models were created using PYMOL with structures from A/Aichi/2/1968 H3 (2YPG) and A/California/04/2009 H1 (3LZG).

functionality within the virus population, they would have risen to an appreciable level during the time course of this study. While replication in cell culture lacks many of the selective pressures found in natural infection, low levels of diversity within IAV populations have also been reported in human, equine, and canine infection (47–50). Very similar results were seen in a recent publication from our lab, which showed an equally low level of population variation in IAV passaged in both wild type and $CMAH^{-/-}$ mice (37). However, we did see different minority variants arise in individual virus populations in our second replicate experiments, where three virus populations were passaged five times in MDCK-WT or MDCK-CMAH cells. Thus, our data seems to indicate little variation within populations, but possibly higher variation between populations. This would be in-line with current thinking that intra-host variation is much lower than global variation in human IAVs (47).

For the pH1N1 and wyoH3N2 human IAV strains, most coding changes altered residues within the stalk region of HA gene in regions of the protein associated with pH stability, with some increasing to high frequency between 40% and 99%. Similar mutations effecting pH stability have been suggested to influence host tropism and transmission stability (51, 52). These mutations may therefore result from selection for HA stability at different pHs, or for more efficient protease cleavage or fusion in MDCK cells under the culture conditions of our studies. Comparing the sequences to those in databases showed that many of the stalk mutations that reached fixation in this study are present in other IAV isolates. In pH1N1, the K445E/M mutations were also present in some swine viruses and a few human isolates. Where culture information was available, most of these isolates had been passed in MDCK cells, so it is possible that these represent culture adaptation. The G404R/E mutations that arose in several wyoH3N2 populations, however, were primarily seen in equine and canine isolates, although

passaging information for these were not always available. The lack of information about passaging history of many isolates in the database makes it difficult to draw any solid conclusions. SNVs in the HA stalk were not as prevalent in CIV H3N2 virus, so are likely associated with the specific sequences of the human viruses, or their replication in canine cells.

Sialic acid linkages and effects on viruses. Human IAV favor cell binding and infection using the α2,6-linked Sia, which is common in the human upper respiratory tract, and the MDCK-SiaT1 cells have higher expression of this form of Sia linkage. Human viruses passaged in wild-type MDCK cells have previously been reported to have mutations around their Sia binding sites (30, 53). We saw no significant difference in the SNVs near the HA receptor-binding site in pH1N1 when passed in MDCK-WT compared to MDCK-SiaT1 cells. In wyoH3N2, one mutation at position 216 near the receptor-binding site reached a high frequency when passed in MDCK-WT while no mutations in HA1 arose in MDCK-SiaT1 cells. This confirms that MDCK-SiaT1 maintain virus receptor-binding preference, although this appears to be more important for human H3N2 than H1N1 viruses in our study. CIV H3N2 virus specifically binds to α2,3-linked Sia (54), and when passaged in MDCK-SiaT1 cells, CIV H3N2 showed near fixation of residue S219P near the receptor-binding site, suggesting selection by the Sia linkage type. Across the different MDCK cell lines, CIV H3N2 virus also showed a number of SNVs arising in the NA gene close to or within the trans-membrane and stalk region of NA, centered around position 27. Some of these mutations rose to near fixation in all CIV H3N2 passages, suggesting that these mutations increase stability or budding efficiency in cell culture. Indeed, similar mutations were seen in canine IAV strains passaged in feline cells (46). While some of these mutations did not correspond to any IAV isolates in the NCBI database (A27T and L35Q), others were seen in natural canine and avian isolates.

Cell expression of Neu5Gc shows varying effects. Most IAV that have been tested show a preference for binding Neu5Ac, which is the Sia form found in humans, birds, western breeds of dogs, and on MDCK cells (55, 56). But, pigs and horses do express high levels of Neu5Gc in their respiratory tract (24, 25). Some viruses from horses, particularly the H7N7 strain that circulated from 1956 until the mid-1970s, appear to bind Neu5Gc preferentially (26). Additionally, laboratory-generated HA proteins that bind Neu5Gc were made through a T155Y mutation, with residues 143 and 158 also being related to this shift in Sia binding preference (57). Passaging of all three IAVs in MDCK-CMAH cells, which express ~40% Neu5Gc, resulted in more SNVs near the receptor-binding site compared to viruses passaged in MDCK-WT cells, although these SNVs were consistently low frequency and none of them matched the previously described mutations for changing binding preference from Neu5Ac to Neu5Gc (26, 57). Few of these SNVs in HA1 arose consistently across virus populations, with instead a variety of low frequency SNVs in the HA1 region arising in different MDCK-CMAH passaged virus populations. The variants identified have not been previously reported as effecting Neu5Gc binding, but were similar to mutations effecting α2,3- or α2,6-linked Sia binding (41). Not all mutations that arose in MDCK-CMAH passaged virus were within the receptor-binding site, and many fell within important antigenic regions nearby (58). The low frequency of these variants may result from the fact that ~40% of Sia on cells was Neu5Gc, so viruses were able to bind the remaining ~60% of Neu5Ac on the cell surface to bypass the Neu5Gc. However, a smaller pool of available receptor may result in selection for mutations in or near the receptor-binding site that optimize Sia binding to the remaining Sia receptors or to relax Sia binding preference to improve binding to Neu5Gc receptors (26). The levels of Neu5Gc in our MDCK-CMAH cells were

similar to those found in pig respiratory tract, where they may affect receptor-binding preferences and alter antigenic sites during natural infection in pigs.

As a control to test for the ability to detect the emergence of mutations under direct selection, we also passaged each virus in the presence of 2-deoxy-2,3-didehydro-*N*-acetylneuraminic acid (DANA or Zanamavir). By passage 5, pH1N1 did not see any SNVs arise in NA or HA besides the stalk mutations that consistently arose in all MDCK passaged virus. WyoH3N2 had two SNVs arise in HA near the receptor-binding site at position 163 and 244, while CIV H3N2 had a S231G mutation reach fixation in HA. Previous studies have found that IAV strains can avoid NA inhibition through mutations in HA that change Sia binding affinity, so that the mutations in HA we saw in wyoH3N2 and CIV H3N2 could be tied to Zanamivir resistance (38, 40, 59). Mutations in NA that give resistance often require longer passaging (60).

In summary, passage of the three viruses up to 10 passages in the variant MDCK cells or those expressing different Sia forms resulted in only a small number of changes in the HA gene segment, or a small number of variants in the NA gene segment of CIV H3N2, with little variation across all other gene segments.

3.6 MATERIALS AND METHODS.

Cells and viruses. MDCK-NBL2 and HEK293T cells were obtained from ATCC. MDCK-SiaT1 cells were prepared by transfection of the *ST6Gal1* gene in a plasmid under the control of the CMV promoter (pcDNA3.1, Invitrogen). Cell clones with increased levels of α2,6-linked Sia were identified by staining with the *Sambucus nigra* (SNA) lectin (Vector laboratories). MDCK-CMAH cells were prepared by transfection of the human CMAH gene in a plasmid under the control of the CMV promoter (pcDNA3.1, Invitrogen). Clones with Neu5Gc expression were determined by HPLC analysis as previous described (61). MDCK-Type I cells

(clone AA7) and MDCK-Type II cells (clone BG12) were a gift from Dr. William Young (University of Kentucky) and were originally cloned and characterized by Dr. Guy E. Nichols (11). All cells were grown in DMEM with 10% fetal calf serum, and 50µg/ml gentamycin.

Three IAV strains were derived from reverse genetics plasmids, comprising (i) human H1N1 pandemic IAV (A/California/04/2009, pH1N1) in plasmid pDP2002, (ii) human H3N2 seasonal IAV (A/Wyoming/3/2003, wyoH3N2) in plasmid pDZ, and (iii) a canine H3N2 IAV (A/Canine/IL/11613/2015, CIV H3N2) in pDZ. The plasmid encoding each viral segment was prepared from a single bacterial colony, and an 8-plasmid mixture for each virus was prepared and used for transfection of a 3:1 co-culture of HEK293T cells and MDCK cells (MDCK-SiaT1 cells for wyoH3N2). Each virus was passaged two additional times in the same MDCK, or MDCK-SiaT1 cells, to generate a passage-3 stock, which was tested for infectivity by $TCID_{50}$ assay. Each plasmid mixture (as DNA) and the virus stocks were then used to generate libraries for Illumina sequencing, as described below, revealing the original sequences and any baseline variation of the virus populations used for cell passaging.

Lectin staining by flow cytometry. Cells were seeded at sub-confluency and incubated overnight at 37°C and 5% CO_2. Cells were collected using Accutase (Sigma) to retain surface glycans, then fixed in 4% PFA for 15 min. Cells were blocked using Carbo-Free Blocking Solution (Vector Laboratories). Fluorescein-conjugated plant lectins from *Sambucus nigra* (SNA) and *Maackia amurensis* lectin I (MAA I) (Vector Laboratories) were diluted 1:1200 in blocking buffer and incubated with cells for one hour . A Millipore Guava EasyCyte Plus flow cyotometer (EMD Millipore, Billerica, MA) was used to collect data, analysis using FlowJo software (TreeStar, Ashland, OR). Statistical analyses were performed in PRISM software (GraphPad, version 8).

Passaging of virus in cells. For virus passaged in MDCK-WT, MDCK-SiaT1, MDCK-CMAH, and MDCK-Type I cells, virus was passaged at low MOI of 0.001 for passage one and two, then blind passaged for passage three through five or ten, depending on cell type (see **Figure 4.2**). For the Zanamivir passaged virus, Zanamivir concentration was 0.01μM for passage one, 0.1 μM for passage two, and kept at 1.0 μM for passage three through five. Virus was inoculated at a constant MOI of 0.005 following titering after each passage. For infections, cells were seeded at sub-confluency in T12.5 flasks and allowed to settle for 4 to 6 hrs. Once settled, cells were washed and inoculated with virus diluted in infection media (DMEM with 0.3% BSA and 1μg/ml TPCK-treated trypsin (Sigma Aldrich)) and allowed to adsorb for one hour. Inoculum was removed and infection media replaced. Virus was collected at 48 hrs.

Infection of MDCK Type II cells. For analysis of infected by fluorescent microscopy, cells were seeded onto Thermanox plastic coverslips (Thermo Fisher Scientific) at near confluency and allowed to settle for 8 hrs before infection at an MOI of 0.25 for 12 hrs. Coverslips were then fixed with 4% paraformaldehyde and stained with a mouse anti-NP antibody and fluorescence-conjugated goat anti-mouse secondary. Infected cells were imaged using a Nikon TE300 fluorescent microscope. For analysis of infected cells via flow cytometry, cells were seeded into 12-well plates to be near confluency and allow to settle for 8 hrs before infection at an MOI of 0.25. At 6 hours post infection, cells were trypsinized and fixed with 4% paraformaldehyde and stained with a mouse anti-NP antibody and fluorescence-conjugated goat anti-mouse secondary. A Millipore Guava EasyCyte Plus flow cyotometer (EMD Millipore, Billerica, MA) was used to collect data, analysis using FlowJo software (TreeStar). Statistical analyses were performed in PRISM software (GraphPad, version 8). For titering of virus, cells were seeded in 12-well plates as described and infected at an MOI of 0.01 and supernatant was

collected at 48 hrs post infection. Viral RNA was isolated using the QIAamp Viral RNA Mini Kit (QIAGEN). Influenza genome copies were quantified from RNA isolations by reverse transcription and quantitative PCR (RT-qPCR) for the M segment modified from CDC protocol (62).Products were amplified using Path-ID (Applied Biosystems) with M-specific primers (5' to 3', F: GACCRATCCTGTCACCTCTGAC, R: AGGGCATTYTGGACAAAKCGTCTA), probed with 5'-TGCAGTCCTCGCTCACTGGGCACG-3' and run on a 7500 Fast Real-Time platform against a standard curve.

Viral RNA extraction and virus titering. Viral RNA (vRNA) was isolated from cell culture supernatant using the QIAamp Viral RNA Mini Kit (QIAGEN). Influenza virus titers were determined by $TCID_{50}$ on MDCK-NBL2 cells.

Library generation and NGS sequencing. Library generation was performed as previously described (37). In brief, total vRNA was incubated with Superscript III and Platinum Taq-HiFi (Invitrogen) in the presence of universal influenza amplifying primers (5' to 3', uni12a: GTTACGCGCCAGCAAAAGCAGG; uni12b: GTTACGCGCCAGCGAAAGCAGG; uni13: GTTACGCGCCAGTAGAAACAAGG). For CIV H3N2, HA-specific primers (forward: AGCGAAAGCAGGGGATACT; reverse: CCAGTAGAAACAAGGGTGTTTT) were used in addition to the universal primer set to increase amplification of this segment. Viral cDNA products were purified by Agencourt AMPureXP magnetic beads (Beckman Coulter) and quantified by QuBit (Invitrogen). NGS sequencing libraries were prepared using Nextera XT DNA Library Prep Kit (Illumina). Pooled libraries were run on an Illumina MiSeq v2 for 250bp paired reads in the Cornell Genomics Facility of the Biotechnology Research Center.

Sequence analysis and variant calling. Analysis was performed in Geneious v.11.1.5. Read trimming was performed by the BBDuk script (https://jgi.doe.gov/data-and-

tools/bbtools/bb-tools-user-guide/bbduk-guide/), followed by read merging, and alignment to the reference genomic plasmid sequence for each virus. For variant calling, we considered those at >2% frequency with minimum 200 read coverage.

Graphing and analysis. All figure graphs were generated in GraphPad Prism v.8.1.2.

3.7 ACKNOWLEDGMENTS

We thank Wendy Weichert for expert technical help and coordinating sequencing of canine cells. We also would like to thank Dr. Edward Holmes (University of Sydney) for his invaluable feedback and suggestions, and Dr. Henry Wan (University of Missouri) for the swine tissue samples.

SUPPORT

Supported by CRIP (Center of Research in Influenza Pathogenesis), an NIAID funded Center of Excellence in Influenza Research and Surveillance (CEIRS) contract HHSN272201400008C to Colin Parrish, NIH grants R01 GM080533 to Colin Parrish.

REFERENCES

1. de Graaf M, Fouchier R a M. 2014. Role of receptor binding specificity in influenza A virus transmission and pathogenesis. The EMBO journal 33:823–41.

2. Dou D, Revol R, Östbye H, Wang H, Daniels R. 2018. Influenza A Virus Cell Entry, Replication, Virion Assembly and Movement. Frontiers in Immunology 9.

3. Tobita K, Sugiura A, Enomote C, Furuyama M. 1975. Plaque assay and primary isolation of influenza A viruses in an established line of canine kidney cells (MDCK) in the presence of trypsin. Med Microbiol Immunol 162:9–14.

4. Dukes JD, Whitley P, Chalmers AD. 2011. The MDCK variety pack: choosing the right strain. BMC Cell Biology 12:43.

5. Barker G, Simmons NL. 1981. Identification of two strains of cultured canine renal epithelial cells (MDCK cells) which display entirely different physiological properties. Quarterly Journal of Experimental Physiology 66:61–72.

6. Tsai H-C, Lehman CW, Lin C-C, Tsai S-W, Chen C-M. 2019. Functional evaluation for adequacy of MDCK-lineage cells in influenza research. BMC Research Notes 12.

7. Parker L, Wharton SA, Martin SR, Cross K, Lin Y, Liu Y, Feizi T, Daniels RS, McCauley JW. 2016. Effects of egg-adaptation on receptor-binding and antigenic properties of recent influenza A (H3N2) vaccine viruses. Journal of General Virology 97:1333–1344.

8. Stevens J, Chen L-M, Carney PJ, Garten R, Foust A, Le J, Pokorny BA, Manojkumar R, Silverman J, Devis R, Rhea K, Xu X, Bucher DJ, Paulson J, Cox NJ, Klimov A, Donis RO. 2010. Receptor Specificity of Influenza A H3N2 Viruses Isolated in Mammalian Cells and Embryonated Chicken Eggs. Journal of Virology 84:8287–8299.

9. Wu NC, Lv H, Thompson AJ, Wu DC, Ng WWS, Kadam RU, Lin C-W, Nycholat CM, McBride R, Liang W, Paulson JC, Mok CKP, Wilson IA. 2019. Preventing an Antigenically Disruptive Mutation in Egg-Based H3N2 Seasonal Influenza Vaccines by Mutational Incompatibility. Cell Host & Microbe 25:836-844.e5.

10. Subbarao K, Barr I. 2019. A Tale of Two Mutations: Beginning to Understand the Problems with Egg-Based Influenza Vaccines? Cell Host & Microbe 25:773–775.

11. Nichols GE, Lovejoy JC, Borgman CA, Sanders JM, Young WW. 1986. Isolation and characterization of two types of MDCK epithelial cell clones based on glycosphingolipid pattern. Biochim Biophys Acta 887:1–12.

12. Hansson GC, Simons K, van Meer G. 1986. Two strains of the Madin-Darby canine kidney (MDCK) cell line have distinct glycosphingolipid compositions. EMBO J 5:483–489.

13. Nichols GE, Shiraishi T, Allietta M, Tillack TW, Young WW. 1987. Polarity of the Forssman glycolipid in MDCK epithelial cells. Biochim Biophys Acta 930:154–166.

14. Nichols GE, Shiraishi T, Young WW. 1988. Polarity of neutral glycolipids, gangliosides, and sulfated lipids in MDCK epithelial cells. J Lipid Res 29:1205–1213.

15. Gekle M, Wünsch S, Oberleithner H, Silbernagl S. 1994. Characterization of two MDCK-cell subtypes as a model system to study principal cell and intercalated cell properties. Pflügers Archiv European Journal of Physiology 428:157–162.

16. Lugovtsev VY, Melnyk D, Weir JP. 2013. Heterogeneity of the MDCK Cell Line and Its Applicability for Influenza Virus Research. PLoS ONE 8:e75014.

17. Shinya K, Ebina M, Yamada S, Ono M, Kasai N, Kawaoka Y. 2006. Avian flu: influenza virus receptors in the human airway. Nature 440:435–436.

18. Causey D, Edwards SV. 2008. Ecology of Avian Influenza Virus in Birds. The Journal of Infectious Diseases 197:S29–S33.

19. Kahn RE, Ma W, Richt JA. 2014. Swine and influenza: a challenge to one health research. Curr Top Microbiol Immunol 385:205–218.

20. Varki NM, Varki A. 2007. Diversity in cell surface sialic acid presentations: implications for biology and disease. Laboratory investigation; a journal of technical methods and pathology 87:851–857.

21. Varki A, Schauer R. 2009. Chapter 14 Sialic AcidsEssentials of Glycobiology - second edition, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Springs Harbor, NY.

22. Peri S, Kulkarni A, Feyertag F, Berninsone PM, Alvarez-Ponce D. 2018. Phylogenetic Distribution of CMP-Neu5Ac Hydroxylase (CMAH), the Enzyme Synthetizing the Proinflammatory Human Xenoantigen Neu5Gc. Genome Biol Evol 10:207–219.

23. Hashimoto Y, Yamakawa T, Tanabe Y. 1984. Further studies on the red cell glycolipids of various breeds of dogs. A possible assumption about the origin of Japanese dogs. Journal of biochemistry 96:1777–1782.

24. Scocco P, Pedini V. 2008. Localization of influenza virus sialoreceptors in equine respiratory tract. Histology and Histopathology 23:973–978.

25. Wang R-G, Ruan M, Zhang R-J, Chen L, Li X-X, Fang B, Li C, Ren X-Y, Liu J-Y, Xiong Q, Zhang L-N, Jin Y, Li L, Li R, Wang Y, Yang H-Y, Dai Y-F. 2018. Antigenicity of tissues and organs from GGTA1/CMAH/β4GalNT2 triple gene knockout pigs. J Biomed Res.

26. Broszeit F, Tzarum N, Zhu X, Nemanichvili N, Eggink D, Leenders T, Li Z, Liu L, Wolfert MA, Papanikolaou A, Martínez-Romero C, Gagarinov IA, Yu W, García-Sastre A, Wennekes T, Okamatsu M, Verheije MH, Wilson IA, Boons G-J, de Vries RP. 2019. N-

Glycolylneuraminic Acid as a Receptor for Influenza A Viruses. Cell Reports 27:3284-3294.e6.

27. Higa H, Rogers G, Paulson C. 1985. Influenza virus hemagglutinins differentiate between receptor determinants bearing N-acetyl-, N-glycollyl-, and N,O-diacetylneuraminic acids. Virology 144:279–282.

28. Xu G, Suzuki T, Maejima Y, Mizoguchi T, Tsuchiya M, Kiso M, Hasegawa A, Suzuki Y. 1995. Sialidase of swine influenza A viruses: variation of the recognition specificities for sialyl linkages and for the molecular species of sialic acid with the year of isolation. Glycoconjugate journal 12:156–161.

29. Matrosovich M, Matrosovich T, Carr J, Roberts NA, Klenk H-D. 2003. Overexpression of the -2,6-Sialyltransferase in MDCK Cells Increases Influenza Virus Sensitivity to Neuraminidase Inhibitors. Journal of Virology 77:8418–8425.

30. McWhite CD, Meyer AG, Wilke CO. 2016. Sequence amplification via cell passaging creates spurious signals of positive adaptation in influenza virus H3N2 hemagglutinin. Virus Evolution 2:vew026.

31. Matsumoto S, Chong Y, Kang D, Ikematsu H. 2019. High genetic stability in MDCK-SIAT1 passaged human influenza viruses. Journal of Infection and Chemotherapy 25:222–224.

32. Oh DY, Barr IG, Mosse JA, Laurie KL. 2008. MDCK-SIAT1 cells show improved isolation rates for recent human influenza viruses compared to conventional MDCK cells. J Clin Microbiol 46:2189–2194.

33. Uno Y, Kawakami S, Ochiai K, Omi T. 2019. Molecular characterization of cytidine monophospho-N-acetylneuraminic acid hydroxylase (CMAH) associated with the erythrocyte antigens in dogs. Canine Genetics and Epidemiology 6.

34. Hoffmann E, Krauss S, Perez D, Webby R, Webster RG. 2002. Eight-plasmid system for rapid generation of influenza virus vaccines. Vaccine 20:3165–3170.

35. Gubareva LV, Bethell R, Hart GJ, Murti KG, Penn CR, Webster RG. 1996. Characterization of mutants of influenza A virus selected with the neuraminidase inhibitor 4-guanidino-Neu5Ac2en. Journal of virology 70:1818–1827.

36. Blick TJ, Tiong T, Sahasrabudhe a, Varghese JN, Colman PM, Hart GJ, Bethell RC, McKimm-Breschkin JL. 1995. Generation and characterization of an influenza virus neuraminidase variant with decreased sensitivity to the neuraminidase-specific inhibitor 4-guanidino-Neu5Ac2en. Virology 214:475–484.

37. Wasik BR, Voorhees IEH, Barnard KN, Alford-Lawrence BK, Weichert WS, Hood G, Nogales A, Martínez-Sobrido L, Holmes EC, Parrish CR. 2019. Influenza Viruses in

Mice: Deep Sequencing Analysis of Serial Passage and Effects of Sialic Acid Structural Variation. Journal of Virology 93.

38. Wagner R, Matrosovich M, Klenk HD. 2002. Functional balance between haemagglutinin and neuraminidase in influenza virus infections. Reviews in Medical Virology 12:159–166.

39. Gaymard A, Le Briand N, Frobert E, Lina B, Escuret V. 2016. Functional balance between neuraminidase and haemagglutinin in influenza viruses. Clinical Microbiology and Infection 22:975–983.

40. Abed Y, Bourgault A, Fenton RJ, Morley PJ, Gower D, Owens IJ, Tisdale M, Boivin G. 2002. Characterization of 2 Influenza A(H3N2) Clinical Isolates with Reduced Susceptibility to Neuraminidase Inhibitors Due to Mutations in the Hemagglutinin Gene. The Journal of Infectious Diseases 186:1074–1080.

41. Shi Y, Wu Y, Zhang W, Qi J, Gao GF. 2014. Enabling the "host jump": structural determinants of receptor-binding specificity in influenza A viruses. Nature reviews Microbiology 12:822–31.

42. Peng W, de Vries RP, Grant OC, Thompson AJ, McBride R, Tsogtbaatar B, Lee PS, Razi N, Wilson IA, Woods RJ, Paulson JC. 2016. Recent H3N2 Viruses Have Evolved Specificity for Extended, Branched Human-type Receptors, Conferring Potential for Increased Avidity. Cell Host & Microbe 1–12.

43. Kumari K, Gulati S, Smith DF, Gulati U, Cummings RD, Air GM. 2007. Receptor binding specificity of recent human H3N2 influenza viruses. Virology Journal 4:42.

44. Du W, Guo H, Nijman VS, Doedt J, van der Vries E, van der Lee J, Li Z, Boons G-J, van Kuppeveld FJM, de Vries E, Matrosovich M, de Haan CAM. 2019. The 2nd sialic acid-binding site of influenza A virus neuraminidase is an important determinant of the hemagglutinin-neuraminidase-receptor balance. PLOS Pathogens 15:e1007860.

45. Lu B, Zhou H, Ye D, Kemble G, Jin H. 2005. Improvement of Influenza A/Fujian/411/02 (H3N2) Virus Growth in Embryonated Chicken Eggs by Balancing the Hemagglutinin and Neuraminidase Activities, Using Reverse Genetics. Journal of Virology 79:6763–6771.

46. Kamiki H, Matsugo H, Ishida H, Kobayashi-Kitamura T, Sekine W, Takenaka-Uema A, Murakami S, Horimoto T. 2019. Adaptation of H3N2 canine influenza virus to feline cell culture. PLOS ONE 14:e0223507.

47. Xue KS, Moncla LH, Bedford T, Bloom JD. 2018. Within-Host Evolution of Human Influenza Virus. Trends in Microbiology 26:781–793.

48. Murcia PR, Baillie GJ, Daly J, Elton D, Jervis C, Mumford JA, Newton R, Parrish CR, Hoelzer K, Dougan G, Parkhill J, Lennard N, Ormond D, Moule S, Whitwham A,

McCauley JW, McKinley TJ, Holmes EC, Grenfell BT, Wood JLN. 2010. Intra- and interhost evolutionary dynamics of equine influenza virus. Journal of virology 84:6943–54.

49. Hoelzer K, Murcia PR, Baillie GJ, Wood JLN, Metzger SM, Osterrieder N, Dubovi EJ, Holmes EC, Parrish CR. 2010. Intrahost evolutionary dynamics of canine influenza virus in naive and partially immune dogs. Journal of virology 84:5329–35.

50. Debbink K, McCrone JT, Petrie JG, Truscon R, Johnson E, Mantlo EK, Monto AS, Lauring AS. 2017. Vaccination has minimal impact on the intrahost diversity of H3N2 influenza viruses. PLOS Pathogens 13:e1006194.

51. Byrd-Leotis L, Galloway SE, Agbogu E, Steinhauer DA. 2015. Influenza Hemagglutinin (HA) Stem Region Mutations That Stabilize or Destabilize the Structure of Multiple HA Subtypes. Journal of Virology 89:4504–4516.

52. Russell CJ, Hu M, Okda FA. 2018. Influenza Hemagglutinin Protein Stability, Activation, and Pandemic Risk. Trends in Microbiology 26:841–853.

53. DuPai CD, McWhite CD, Smith CB, Garten R, Maurer-Stroh S, Wilke CO. 2019. Influenza passaging annotations: what they tell us and why we should listen. Virus Evolution 5.

54. Yang G, Li S, Blackmon S, Ye J, Bradley KC, Cooley J, Smith D, Hanson L, Cardona C, Steinhauer DA, Webby R, Liao M, Wan X-F. 2013. Mutation tryptophan to leucine at position 222 of haemagglutinin could facilitate H3N2 influenza A virus infection in dogs. Journal of General Virology 94:2599–2608.

55. Gulati S, Smith DF, Cummings RD, Couch RB, Griesemer SB, St. George K, Webster RG, Air GM. 2013. Human H3N2 Influenza Viruses Isolated from 1968 To 2012 Show Varying Preference for Receptor Substructures with No Apparent Consequences for Disease or Spread. PLoS ONE 8.

56. Gulati S, Lasanajak Y, Smith DF, Cummings RD, Air GM. 2014. Glycan array analysis of influenza H1N1 binding and release. Cancer Biomarkers 14:43–53.

57. Masuda H, Suzuki T, Sugiyama Y, Horiike G, Murakami K, Miyamoto D, Jwa Hidari KI-P, Ito T, Kida H, Kiso M, Fukunaga K, Ohuchi M, Toyoda T, Ishihama A, Kawaoka Y, Suzuki Y. 1999. Substitution of amino acid residue in influenza A virus hemagglutinin affects recognition of sialyl-oligosaccharides containing *N* -glycolylneuraminic acid. FEBS Letters 464:71–74.

58. Wu NC, Wilson IA. 2017. A Perspective on the Structural and Functional Constraints for Immune Evasion: Insights from Influenza Virus. Journal of Molecular Biology 429:2694–2709.

59. McKimm-Breschkin JL. 2000. Resistance of influenza viruses to neuraminidase inhibitors–a review. Antiviral research 47:1–17.

60. Hurt AC, Ho H-T, Barr I. 2006. Resistance to anti-influenza drugs: adamantanes and neuraminidase inhibitors. Expert Review of Anti-infective Therapy 4:795–805.

61. Varki A, Diaz S. 1984. The release and purification of sialic acids from glycoconjugates: Methods to minimize the loss and migration of O-acetyl groups. Analytical Biochemistry 137:236–247.

62. Shu B, Wu K-H, Emery S, Villanueva J, Johnson R, Guthrie E, Berman L, Warnes C, Barnes N, Klimov A, Lindstrom S. 2011. Design and Performance of the CDC Real-Time Reverse Transcriptase PCR Swine Flu Panel for Detection of 2009 A (H1N1) Pandemic Influenza Virus. Journal of Clinical Microbiology 49:2614–2619.

CHAPTER FIVE

Summary and Conclusions

The goal of this thesis was to determine the role of modified sialic acids (Sia) on influenza A (IAV) infection efficiency and tropism. The focus was on 7,9-*O*- and 9-*O*-acetyl Sia, as well as *N*-glycolyl Sia (Neu5Gc), and how the expression of these modified Sia varied between individual cells, in tissues, and between IAV host species. Sia linkage variation has been previously shown to be an important host tropism-determining factor for IAV; however, little research has been done on how other structural variants of Sia, such as *O*-acetyl and *N*-glycoyl modifications, affect the virus. As Sia modifications are also an understudied area of glycobiology, understanding the regulation and expression of these modified Sia in animals is key to understanding how these modifications act in the ecological landscape of IAV evolution.

This thesis sought to improve our understanding of modified Sia expression and their interaction with IAV from three approaches. The first focused on the expression of 7,9-*O*- and 9-*O*-Ac on cells in culture and their effect on IAV, influenza B (IBV), influenza C (ICV), and influenza D (IDV) viruses. The second determined differences in modified Sia expression across the tissues in mice, an important model species, as well as between secreted mucus from different IAV host species. It also looked functionally at the effect of these modifications on IAV glycoproteins, hemagglutinin and neuraminidase. Finally, the third was an in depth look at changes in IAV populations during serial passages on different MDCK cell lineages, including cells expressing different Sia modifications. All three approaches have built a more complete picture of the complexity of interactions between IAV and its Sia receptor.

5.1 7,9-*O*- and 9-*O*-acetyl Sia on cells and effects on influenza viruses

This paper was the first examination of how 7,9-*O*- and 9-*O*-Ac are regulated by the sialate *O*-acetyltransferase, CasD1, and the sialate *O*-acetylesterase, SIAE. It showed how the localization of these *O*-acetyl Sia modifications varies between different cell types and

confirmed that CasD1 was responsible for expression of both 9-*O*-Ac and 7,9-*O*-Ac modifications. This was also the first quantification of *O*-Ac modifications in different cell lines using HPLC analysis, which revealed that these modifications are a minority of total Sia in a cell. However, despite making up 1-2% of the total Sia, it seems that regulation of these modifications could be quite important for cell homeostasis. Dysregulation of these modifications through knocking-out SIAE and over-expressing CasD1 led to changes in cell growth, reducing growth in HEK-293 cells and increasing growth in A549 cells. It also may influence glycoprotein trafficking through the ER/Golgi pathway, as increasing *O*-Ac levels through over-expression of CasD1 led to most glycoproteins with these modifications being retained in the Golgi.

Functional assays showed that high levels of 7,9-*O*- and 9-*O*-Ac can inhibit IAV hemagglutinin (HA) binding and neuraminidase (NA) cleavage. However, while 9-*O*- and 7,9-*O*-Ac are necessary for infection of ICV and IDV, which utilize 9-*O*-Ac as their primary receptor, the low levels on cell surfaces do not effect IAV and IBV, as plenty of unmodified Sia remain to act as receptors for these viruses.

There are many questions that remain to be addressed from both a cell biology aspect, as well as for virus:host interactions. First, it is not clear what role these *O*-acetyl modifications might be playing in cell-cell signaling and homeostasis. Dysregulation of CasD1 and SIAE did seem to effect cell growth, which may play a role in some types of cancer as they often show changes to glycosylation patterns, including *O*-acetyl modifications. Similarly, how these modifications regulate trafficking of glycoproteins in the ER/Golgi pathway will require more research. It also appears from this work and others that 7,9-*O*-Ac and 9-*O*-Ac are regulated differently, as over-expressing CasD1 only increased 9-*O*-Ac and not 7,9-*O*-Ac, although the mechanism for this differential expression is currently unclear.

While 7,9-*O*- and 9-*O*-Ac did inhibit IAV HA binding and NA cleavage, the low levels present on cell did not effect virus infection for both IAV and IBV viruses. However, for ICV and IDV which utilize 9-*O*-Ac as their primary receptor, the 1-2% present on the cell surface is adequate for infection. Despite rescuing expression of 9-*O*-Ac in our MDCK CasD1 over-expression cells, the surface levels of 9-*O*-Ac for these cells are lower than on wild-type MDCK cells. Not all ICV strains regained infection in MDCK CasD1 over-expression cells, and IDV virus had lower infection rates. This raises the question of how much receptor is necessary to trigger uptake of virus for infection? While the small minority of 9-*O*-Ac on MDCK wild-type cells can support infection, some necessary threshold appears to exist below which infection does not occur.

5.2 Modified Sia on mucins and effects on influenza A virus HA and NA

This paper sought to determine variations in modified Sia expression across different tissues in an important animal model, the mouse, as well as across secreted mucus in different IAV host species. Additionally, the effect of Sia modifications on IAV HA and NA were examined in functional assays and during infection. I found that across mouse tissues, particularly mucosal tissues, there was variation in both *O*-acetyl modifications and Neu5Gc expression, confirming previous reports. Most *O*-acetyl modifications were associated with mucus in the respiratory and gastrointestinal tracts. In mice knocked out for the *CMAH* gene, the lack of Neu5Gc led to higher expression of *O*-acetyl modified Sia across most tissues. Secreted mucus in the saliva of some IAV host species also showed enrichment for *O*-acetyl modifications while pigs showed extremely high levels of Neu5Gc. However, secreted mucus in saliva from humans, as well as from primary and immortalized cells from human respiratory tissues, had very little Sia modifications. Similarly, the Sia modifications present on erythrocytes also varied

between different species, which is important as erythrocytes are often used to test virus binding in hemagglutination assays.

In functional assays, IAV HA showed decreased binding to modified Sia across human H1 and H3 glycoproteins. NA also had much lower cleavage against Neu5Gc and 7,9-*O*-acetyl modified Sia, although with some strain differences possibly related to the host tropism of the virus. Despite being inhibitory of both HA and NA, mucin with these modifications was actually less inhibitory than mucin without them in infection assays. This is likely because mucin with unmodified Sia has a higher density of decoy receptors, leading to more efficient trapping of virus.

The variation seen between IAV host species on modified Sia in secreted mucin raises interesting questions about the effect of these differences on host-specificity of different IAV strains. In this study, I focused on human strains of IAV; however, virus that is adapted to pigs or horses, which have higher levels of Neu5Gc, might have different sensitivities towards inhibition by Sia modifications. Previous reports have shown that the equine H7N7 virus can bind Neu5Gc as its receptor. Similarly, how other viruses that infect in the respiratory or gastrointestinal interact with these modified Sia in mucus will be a fascinating area of future research. Viruses are not the only pathogens that interact with Sia, as many bacteria and parasites also either utilize Sia during infection, or would need to penetrate mucus as part of the infection process. The effect of these modifications on commensal bacteria and how these might vary between different species also remains to be seen.

5.3 Sequence variation of influenza A viruses grown in MDCK cell lineages

In this chapter, the changes in IAV populations when passaged in different MDCK lineages were examined. MDCK cells are heterogeneous and many different clonal cell lines

have been developed from them, including MDCK-Type I and MDCK-Type II cells. Additionally, MDCK cells have been used as a model for expressing different Sia forms including increased α2,6-linked Sia in MDCK-SiaT1 cells and Neu5Gc expression in MDCK-CMAH cells. While previous reports have shown mutations in different IAV strains when serially passed in standard MDCK cells, in depth population changes in viruses serially passaged on different MDCK cell lineages have not been examine.

While all of the MDCK cell lineages were developed from the same standard MDCK cell lineage from ATCC (MDCK-NBL2, denoted WT here), it was surprising to find that one of the lineages was less permissive to virus infection. MDCK-Type II cells showed lower infection efficiency and virus release than the other MDCK cell lines. The cause of this is likely due differences in host factors, such as protease production which had previously been reported for MDCK-Type II clones. Infection in these cells appears to be highly dependent on trypsin presence in infection media. For the other cell lineages, all of them were able to grow virus for the complete five or ten passages in the experiment.

We found that for the three virus strains examined, A/California/04/2009 (pH1N1), A/Wyoming/2003 (wyoH3N2), and A/canine/Illinois/2015 (CIV H3N2), very few single nucleotide variants (SNV) arose across the genome in all the cell lineages except for in HA. Both pH1N1 and wyoH3N2 primarily showed SNV in the stalk region of HA, likely related to pH stability. In MDCK-CMAH cells that expressed Neu5Gc, pH1N1 and wyoH3N2 also showed more low frequency SNVs near the receptor-binding site than in MDCK-WT cells although these rarely rose above a frequency of 10% and were inconsistent between virus populations. For CIV H3N2, most SNVs were in HA1 with more SNVs near the receptor-binding site in MDCK-CMAH compared to MDCK-WT, although again maintained at low frequency. CIV H3N2 also

showed high frequency SNVs in the transmembrane and stalk domain of NA in all cell lineages, possibly related to budding efficiency or stability in cell culture.

The results show few SNVs across the genomes of all three IAV, except in HA. This is consistent with previous research performed in our lab, passaging virus in mice, as well as studies showing low diversity within individual infected humans, horses, and dogs. It also appears that Neu5Gc acts as a weak selective pressure, confirming a previous study from our lab showing no selected mutations in virus passaged in mice expressing Neu5Gc compared to those that lacked functional CMAH activity.

5.4 Final comments

Overall, the papers presented in this thesis represent an in depth analysis of the expression of modified Sia on cells, in tissues, and in secreted mucus across different influenza host species and animal models. How these modifications functionally effect IAV HA binding and NA cleavage was also determined, as well as how variation in Sia expression can affect virus evolution over time in the important *in vitro* model of MDCK cells. Variation in Sia expression is an important aspect of IAV evolution and host adaptation, but previous research had mainly been focused on the effects of Sia linkage type. The research here on *O*-acetyl and Neu5Gc forms of modified Sia is an expansion of this research and showed the importance of studying how other chemical and structural changes to Sia can effect virus infection and tropism. Sia exist in nature as over 50 different distinct forms that vary between different species, tissues, and cell types. The complexity of all these Sia variations to host tropism and infection will continue to be a valuable area of research not just for IAV, but many other human and animal pathogens.