INSIGHTS INTO THE PLASTICITY OF MOLECULAR DETERMINANTS ASSOCIATED WITH ANTIGENIC DRIFT IN THE INFLUENZA H3 HEMAGGLUTININ

by

JEFFERSON JOSE DA SILVA SANTOS

(Under the Direction of Daniel R. Perez)

ABSTRACT

Influenza A virus (IAV) remains a major threat to a wide range of animal hosts. The hemagglutinin (HA) of IAV initiates virus attachment and entry by binding to terminal sialic acid (SA) residues on host cells. HA gradually accumulates amino acid substitutions that allow IAV to escape population immunity through a mechanism known as antigenic drift. Recent work has identified seven amino acid positions (145, 155, 156, 158, 159, 189 and 193, H3 numbering) near the receptor-binding site (RBS) of H3N2 IAVs as the major determinants of antigenic drift. In the first part of the dissertation, we took an in-depth look at the amino acid plasticity at residue 145 and its impact on receptor binding and antibody recognition. We generated a panel of HA mutant viruses carrying substitution at residue 145 (H3 numbering) representing all 20 amino acids. Despite limited amino acid usage in nature, most substitutions at residue 145 were well tolerated and stably maintained *in vitro*. All substitutions retained receptor binding specificity, but frequently led to decreased receptor binding avidity. Glycan microarray analysis showed substitutions at residue 145 modulate binding to a broad range of

glycans. Furthermore, antigenic characterization identified specific substitutions at residue 145 that altered antibody recognition. In the second part of the dissertation, we combined deep mutational scanning with reverse genetics to examine the *in vitro* amino acid plasticity of these key antigenically relevant residues in the H3 HA. We envisioned a H3 HA antigenic virus library as an alternative IAV vaccine strategy that potentially induces broader protection. Our approach was efficient in producing H3 HA antigenic virus libraries in different donor virus strains. Despite limited diversity in nature, high-throughput targeted next-generation sequencing revealed that these residues exhibited extraordinary amino acid plasticity *in vitro*. Nonetheless, virus library diversity is severely reduced following virus rescue. To validate our experimental approach, a novel H3 HA variant virus was isolated and shown to possess distinct receptor binding and antigenic properties relative to wt H3 HA virus. The findings obtained in this dissertation have important implications for understanding virus evolution and aiding the development of novel vaccine design approaches.

INDEX WORDS: Influenza A virus, H3N2, hemagglutinin, antigenic drift, plasticity, receptor binding site, antigenicity, deep mutational scanning, sialic acid

INSIGHTS INTO THE PLASTICITY OF MOLECULAR DETERMINANTS ASSOCIATED WITH ANTIGENIC DRIFT IN THE INFLUENZA H3 HEMAGGLUTININ

by

JEFFERSON JOSE DA SILVA SANTOS

B.Sc., Universidade Federal de Pernambuco, Brazil, 2007

M.Sc., Universidade Federal de Pernambuco, Brazil, 2010

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

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2018

INSIGHTS INTO THE PLASTICITY OF MOLECULAR DETERMINANTS ASSOCIATED WITH ANTIGENIC DRIFT IN THE INFLUENZA H3 HEMAGGLUTININ

by

JEFFERSON JOSE DA SILVA SANTOS

Major Professor:	Daniel R. Perez
Committee:	Maricarmen Garcia
	Mark W. Jackwood
	Melinda Brindley
	S. Mark Tompkins

Electronic Version Approved:

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

DEDICATION

This dissertation is dedicated to my mother for her emotional resilience, unbeatable optimism and unconditional support.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Daniel R. Perez, for having my back. I will be forever grateful for the opportunity he gave me to join his laboratory. Dr. Perez has given me the guidance, space, and freedom to find my own compass on the project.

I also would like to thank Maricarmen Garcia, Mark Jackwood, Melinda Brindley and Mark Tompkins for their time and support to my graduate advisory committee.

I would like to express my gratitude to my former advisor, Laura Gil, for her continued support and guidance. Dr. Gil has been a champion for my scientific career. Many thanks also go to other important role models in my life for keeping me in line, former teachers Socorro Damascena and Cristina Valença.

My sincerest gratitude goes to former members of the Perez lab at the University of Maryland, College Park (UMD) as well as past and present members of the lab at the University of Georgia (UGA). Special thanks to Ade (a-day-BIM-pee, my "nemesis"), Ana Silvia, Andrea ("legacy" member), Brittany, Christina ("honorary" member), Dani, Ginger, Isabel, Joaquin, John, Jojo, Lucas, Maria Luisa, Silvia, Stivalis and Zhimin. You made the work fun. I am truly honored to have met you and worked by your side.

I would like to express my gratitude to Amy Vincent and Eugenio Abente at the USDA for the long-standing and fruitful collaboration.

I would like to thank fellow graduate students and other members of the Poultry Diagnostic and Research Center as well as the wider community in the College of

Veterinary Medicine at UGA for their support. Many thanks also go to the volleyball group at UGA to help keep me sane.

I would also like to express my gratitude to Alexsandra Câmara, Ana Lisa Gomes, Andrea Rangel, Fabiana Gomes, Jonathan Timothy, Karina Wacemberg, Natália Almeida, Otoniel Maranhão and Suzanne Brito for their friendship and support, even at a distance.

Finally, I would like to thank my family, in special, my mother (Maria Solange), brother (Roger), grandmother (Dona Maria), Edgar and Dila for their unconditional love and unwavering support.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

Influenza A virus (IAV) remains a major threat to a wide range of animal hosts, causing acute respiratory illness from mild to severe in humans, swine and avian species (1-4). Belonging to the *Orthomyxoviridae* family of segmented, negative-sense single-stranded RNA viruses, IAV is classified into subtypes based on the major viral surface glycoproteins hemagglutinin (HA) and neuraminidase (NA). The influenza virus replication is initiated by the attachment of HA on target cells through interaction with terminal sialic acid (SA) residues followed by the fusion of viral and cellular membranes, while NA mediates virus progeny release (5, 6).

HA has a pivotal role in receptor specificity and host range, antigenicity, and pathogenicity. To date, 16 HA subtypes (H1 to H16) have been found in nature, which are divided into two phylogenetic groups with marked antigenic differences: group 1 (H1, H2, H5, H6, H8, H9, H11, H12, H13 and H16) and group 2 (H3, H4, H7, H10, H14 and H15) (5, 7). Additionally, two influenza A-like HA subtypes (provisionally designated HL17 and HL18) have been recently discovered in bats (8). HA is a type I integral membrane glycoprotein with a N-terminal signal sequence, which is synthesized as a single polypeptide. Upon removal of the signal peptide, HA precursor (HA0) is cleaved by host proteases into HA1 (globular head domain) and HA2 (stalk domain) (9, 10).

Located in a small depression on the globular head of HA, the receptor-binding site (RBS) mediates virus binding to terminal SA residues on host cells (9, 10). The

binding site consists of a pocket of noncontiguous residues bordered by the 130-loop, the 150-loop, the 190-helix and the 220-loop. A set of well-conserved polar and nonpolar residues, including Tyr98, Trp153, His183, and Tyr195, forms the base of the binding site (9-12). SAs are regarded as the only known cellular receptors for IAV and occur as terminal monosaccharides in glycoproteins and glycolipids on cell surface or secreted glycoconjugates. Terminal SAs utilized as IAV receptors are bound to galactose (Gal) in α2,3 (SAα2,3Gal) or α2,6 (SAα2,6Gal) linkage configuration (13), with early observations demonstrating that IAV displayed differences in receptor specificity (14). Although receptor binding specificity has received considerable attention in recent years, particularly due to the emergence of IAV strains with pandemic potential, it remains poorly understood which receptors are relevant for IAV attachment, replication, and transmission *in vivo* (15).

HA is also the major target of the antibody response and antibodies elicited against HA by either natural infection or vaccination can block virus infection (16). The elicited antibody response is often strain-specific and directed towards the highly variable head domain of HA. The gradual accumulation of amino acid substitutions on the head of HA, known as antigenic drift, allows evasion of host immunity and remains a challenge for effective vaccination (9, 10, 16-18).

Early work in 1980s combining amino acid sequence information and protein structure data has defined five sites (designated A to E) on the globular head domain of HA H3 involved in antibody binding and recognition (19, 20). It was proposed that the emergence of antigenically drifted strains of epidemiologic importance would require amino acid substitutions located within these antigenic sites (20). More recently,

antigenic cartography, a computational method to quantify binding assay data, such as hemagglutination inhibition (HI) data, was used to characterize the antigenic evolution of human H3 IAV strains (16). Human H3 IAV strains circulating from 1968 to 2003 were grouped into 11 antigenic clusters, with each cluster eventually being replaced by an emerging cluster with distinct antigenic properties over time. Between 1 and 13 amino acid substitutions were associated with each of the antigenic cluster transitions (16). This study laid the groundwork to pinpoint the molecular determinants responsible for antigenic change.

Although more than 130 residues on the HA were previously characterized as putative antigenic residues (19, 20), only seven residues (145, 155, 156, 158, 159, 189 and 193, H3 numbering) near the receptor-binding site (RBS) of human H3N2 IAV strains have a major impact on the emergence of antigenically distinct clusters, as determined by HI data (21). Parallel studies characterizing the antigenic evolution of swine H3N2 IAV strains circulating in the United States have also defined antigenic clusters among swine viruses (18). The substitutions that resulted in marked antigenic differences were postulated to be located at six amino acid residues (145, 155, 156, 158, 159, and 189) in swine H3N2 IAV strains (18, 21). These residues have been validated experimentally, confirming them as major drivers of antigenicity (17).

Research objectives and outline

Despite their significance in antigenic drift, little is known about other functional implications of amino acid substitutions in these antigenically relevant residues. This is particularly important considering human, swine, avian, equine and canine H3 IAV strains exhibit restricted amino acid usage in these residues.

The main goal of this dissertation was to explore the amino acid plasticity in key residues near the RBS of the H3 HA that have been implicated in antigenic drift. In the first part of the dissertation, functional constraints associated with specific amino acid usage were investigated at the single amino acid level with regards to receptor binding and antigenic properties. We took an in-depth look at the plasticity of amino acid at HA residue 145 and its impact on receptor binding and antibody recognition. In the second part of the dissertation, a deep mutational scanning (DMS) strategy targeting the seven residues previously implicated as major determinants of antigenic drift was used to explore the dynamics of virus diversity *in vitro*. We performed a comprehensive analysis of the nucleotide and amino acid diversity of virus libraries, demonstrating the feasibility and robustness of the approach. The findings obtained in this dissertation have important implications for understanding virus evolution and aiding the development of novel vaccine design approaches. The research objectives of this dissertation are as follows:

1. Characterize the impact of single amino acid substitutions on receptor binding properties and antigenicity of influenza A viruses of the H3 subtype.
    a) Generate a panel of influenza H3 HA mutant viruses carrying single amino acid substitutions at residue 145, representing each of the 20 naturally occurring amino acids.
    b) Assess receptor binding preference and avidity of influenza H3 HA mutant viruses *in vitro*.
    c) Evaluate the antigenic profile of influenza H3 HA mutant viruses *in vitro*.

2. Determine the amino acid plasticity of an influenza virus H3 HA.
    a) Design a degenerate PCR library targeting antigenic residues in influenza H3 hemagglutinin and attempt virus rescue.
    b) Examine library amino acid diversity *in vitro*.
    c) Isolate influenza H3 HA mutant viruses with unique "antigenic motifs" and characterize their receptor binding and antigenic properties.

References

1. **Gordon A, Reingold A.** 2018. The Burden of Influenza: a Complex Problem. Curr Epidemiol Rep **5:**1-9.
2. **Paules C, Subbarao K.** 2017. Influenza. Lancet **390:**697-708.
3. **Petrova VN, Russell CA.** 2018. The evolution of seasonal influenza viruses. Nat Rev Microbiol **16:**47-60.
4. **Swayne DE.** 2017. Animal influenza, Second edition. ed. John Wiley and Sons, Inc., Ames, Iowa.
5. **Shaw M, Palese P.** 2013. Orthomyxoviridae, vol 1. Lippincott Williams & Wilkins, Philadelphia, PA.
6. **Taubenberger JK, Kash JC.** 2010. Influenza virus evolution, host adaptation, and pandemic formation. Cell Host Microbe **7:**440-451.
7. **Webster RG, Bean WJ, Gorman OT, Chambers TM, Kawaoka Y.** 1992. Evolution and ecology of influenza A viruses. Microbiol Rev **56:**152-179.
8. **Wu Y, Wu Y, Tefsen B, Shi Y, Gao GF.** 2014. Bat-derived influenza-like viruses H17N10 and H18N11. Trends Microbiol **22:**183-191.
9. **Wiley DC, Skehel JJ.** 1987. The structure and function of the hemagglutinin membrane glycoprotein of influenza virus. Annu Rev Biochem **56:**365-394.
10. **Wilson IA, Skehel JJ, Wiley DC.** 1981. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 A resolution. Nature **289:**366-373.
11. **Laursen NS, Wilson IA.** 2013. Broadly neutralizing antibodies against influenza viruses. Antiviral Res **98:**476-483.

12. **Xiong X, McCauley JW, Steinhauer DA.** 2014. Receptor binding properties of the influenza virus hemagglutinin as a determinant of host range. Curr Top Microbiol Immunol **385:**63-91.

13. **Nicholls JM, Chan RW, Russell RJ, Air GM, Peiris JS.** 2008. Evolving complexities of influenza virus and its receptors. Trends Microbiol **16:**149-157.

14. **Rogers GN, Paulson JC.** 1983. Receptor determinants of human and animal influenza virus isolates: differences in receptor specificity of the H3 hemagglutinin based on species of origin. Virology **127:**361-373.

15. **de Graaf M, Fouchier RA.** 2014. Role of receptor binding specificity in influenza A virus transmission and pathogenesis. EMBO J **33:**823-841.

16. **Smith DJ, Lapedes AS, de Jong JC, Bestebroer TM, Rimmelzwaan GF, Osterhaus AD, Fouchier RA.** 2004. Mapping the antigenic and genetic evolution of influenza virus. Science **305:**371-376.

17. **Abente EJ, Santos J, Lewis NS, Gauger PC, Stratton J, Skepner E, Anderson TK, Rajao DS, Perez DR, Vincent AL.** 2016. The Molecular Determinants of Antibody Recognition and Antigenic Drift in the H3 Hemagglutinin of Swine Influenza A Virus. J Virol **90:**8266-8280.

18. **Lewis NS, Anderson TK, Kitikoon P, Skepner E, Burke DF, Vincent AL.** 2014. Substitutions near the hemagglutinin receptor-binding site determine the antigenic evolution of influenza A H3N2 viruses in U.S. swine. J Virol **88:**4752-4763.

19. **Wiley DC, Wilson IA, Skehel JJ.** 1981. Structural identification of the antibody-binding sites of Hong Kong influenza haemagglutinin and their involvement in antigenic variation. Nature **289:**373-378.

20. **Wilson IA, Cox NJ.** 1990. Structural basis of immune recognition of influenza virus hemagglutinin. Annu Rev Immunol **8:**737-771.

21. **Koel BF, Burke DF, Bestebroer TM, van der Vliet S, Zondag GC, Vervaet G, Skepner E, Lewis NS, Spronken MI, Russell CA, Eropkin MY, Hurt AC, Barr IG, de Jong JC, Rimmelzwaan GF, Osterhaus AD, Fouchier RA, Smith DJ.** 2013. Substitutions near the receptor binding site determine major antigenic change during influenza virus evolution. Science **342:**976-979.

CHAPTER 2

LITERATURE REVIEW

Classification and nomenclature

Influenza A virus (IAV) belongs to the *Orthomyxoviridae* family of segmented, negative-sense single-stranded RNA viruses (1). According to the International Committee on Taxonomy of Viruses (ICTV), the family is divided into seven genera: *Influenzavirus A*, *Influenzavirus B*, *Influenzavirus C*, *Isavirus*, *Thogotovirus*, and the newly recognized *Quaranjavirus* and *Influenzavirus D* (2-5).

Of the seven genera within *Orthomyxoviridae*, only five have been associated with human disease. The sole representative of the *Isavirus* genus, Infectious Salmon Anemia virus (ISAV), is primarily implicated in causing serious disease to farmed Atlantic salmon (6). Influenza D virus (IDV) from the *Influenzavirus D* genus has been identified in swine, cattle, sheep, and goats. Among these hosts, cattle have been proposed as the natural reservoir (2). Members of the *Thogotovirus* and *Quaranjavirus* genera are transmitted by arthropods, usually ticks, and have occasionally been associated with disease in humans (5, 7, 8).

While influenza B virus (IBV, *Influenzavirus B* genus) and influenza C virus (ICV, *Influenzavirus C* genus) are mainly human pathogens, IAV (*Influenzavirus A* genus) is known to infect an expanded range of species. IAV is further classified into subtypes based on the antigenicity of the major viral surface glycoproteins hemagglutinin (HA) and neuraminidase (NA). To date, 16 HA subtypes (H1 to H16) and nine NA

subtypes (N1 to N9) have been found in nature. Wild aquatic waterfowl are believed to be the primary reservoir of all these influenza HA and NA subtypes (9). Additionally, two influenza A-like subtypes (provisionally designated HL17NL10 and HL18NL11) have been recently described in bats (10). These hemagglutinin- and neuraminidase-like proteins from bat-derived viruses lack key canonical functions and structures of influenza viruses (10).

Influenza viruses are named following a recommended nomenclature scheme stablished by the World Health Organization (11). According to the guidelines of this naming convention, influenza viruses are to be designated using the following components:

1. The antigenic type (e.g., A, B, C, D).
2. Host of origin if other than human (e.g., swine, canine, turkey, etc.). For viruses from non-human species, both the Latin binomial nomenclature and the common name of the host of origin should be recorded in the original publication describing the virus isolate [e.g., *Anas acuta* (pintail duck)]. Thereafter, the common name of the species should be used for the strain. When viruses are isolated from nonliving material, the nature of the material should be specified (e.g., lake water, environment, etc.).
3. Geographical origin (e.g., Texas, Brisbane, Brazil, etc.).
4. Strain number (e.g., 237, 9, etc.).
5. Year of isolation (e.g., 98, 2004, etc.).
6. HA and NA subtype designation in parentheses [for influenza A viruses only; e.g., (H1N1), (H3N2)].

As an example, the virus strain A/turkey/Ohio/313053/2004 (H3N2) designates an influenza **A** virus that came from a **turkey** in the state of **Ohio**, USA. The isolated virus strain was assigned the strain number **313053** during **2004** and classified as an **H3N2**.

While this nomenclature scheme has proven useful and is internationally accepted by the influenza research community, there are some shortcomings such as the lack of specificity in host designation due to the use of non-standardized animal names (e.g., duck, but which kind? Muscovy, Pekin, Mallard, etc.) and broad geographical origin specifications (e.g., Brazil, but which state?) (12). These limitations have important implications for the study of IAV ecology, epidemiology and evolution.

Influenza A virus structure and genome organization

IAVs are pleomorphic, ranging from spherical particles of about 100 nm in diameter to filamentous particles with elongated viral structures that exceed 1 μm in length. The filamentous particles show a uniform diameter of about 55 nm (1, 13). The filamentous forms are particularly predominant in human IAV clinical isolates before virus propagation in cell culture or chicken embryonated eggs (14). IAV lipid membrane is derived from the host cell and contains three virus proteins: the hemagglutinin (HA), the neuraminidase (NA), and the matrix protein 2 (M2). HA is the most abundant virus surface protein and forms homotrimers that appear as spikes while NA exists as mushroom-shaped homotetramers on the lipid membrane. The most abundant protein in virus particles, the matrix protein 1 (M1), lies just beneath the envelope and plays a major role in virus morphology (1, 15, 16).

The core of the virus particle is comprised of the eight virus ribonucleoprotein (vRNP) complexes arranged in a specific "1+7" pattern. Each vRNP complex consist of one negative-sense, genomic vRNA bound to multiple copies of the nucleocapsid protein (NP) and one virus polymerase complex [polymerase basic 1 (PB1), polymerase basic 2 (PB2) and polymerase acid (PA)] (17). The nonstructural protein 1 (NS1) and the nuclear export protein/nonstructural protein 2 (NEP/NS2) are also present in purified virus preparations (15, 18). Virus particles also incorporate abundant host cell proteins as part of the core virus architecture, including ubiquitin, annexins, cytoskeletal proteins and glycolytic enzymes. Interestingly, virus particles produced by mammalian and avian hosts have distinct host protein compositions (15).

With approximately 14 kb in length, the genome of IAV is divided in eight segments (1). Individual segments vary in length from 890 bp to 2,341 bp (**Table 2.1**). All segments possess 5′ and 3′ untranslated regions flaking an open reading frame (ORF). The terminal ends of these untranslated regions are conserved among all segments of all IAV strains. These conserved 13 nucleotides at the 5′ end, and 12 nucleotides at the 3′ end display partial and inverted complementarity (19-22). Once paired, the virus RNA is thought to form either a panhandle or corkscrew structure (1, 23). Each virus RNA (vRNA) segment exists as a vRNP complex in which the virus RNA is encapsidated by NP and forms a helical hairpin that is bound on one end by the virus polymerase complex (1, 17, 24).

Up to 15 proteins are encoded by the IAV genome. A summary of the genome organization, encoded virus proteins and their major function is presented for a representative IAV, A/turkey/Ohio/313053/2004 (H3N2), in **Table 2.1**.

Table 2.1. A/turkey/Ohio/313053/2004 (H3N2) virus segments, encoded virus proteins and their major function

Virus segment		Virus protein			
#	nt length	Protein	nt length	aa length	Function
1	2341	PB2	2280	759	Virus polymerase complex, cap binding
		PB1	2274	757	Virus polymerase complex, RNA polymerase
2	2341	PB1-F2	273	90	Strain dependent, modulates host innate response
		PB1-N40[1]	2157	718	Unclear, strain dependent
3	2233	PA	2151	716	Virus polymerase complex, endonuclease
		PA-N155[1]	1689	562	Unclear, likely participates in the virus replication cycle
		PA-N182[1]	1608	535	Unclear, likely participates in the virus replication cycle
		PA-X[1]	699	232	Modulates host gene expression (host cell shut off)
4	1762	HA	1701	566	Receptor binding, membrane fusion, antigenicity
5	1565	NP	1497	498	Virus RNA encapsidation, virus assembly
6	1466	NA	1410	469	Virus release, antigenicity
7	1027	M1	759	252	Virus assembly
		M2	294	97	Ion channel activity, uncoating
8	890	NS1	660	219	Multifunctional protein
		NEP/NS2	366	121	Virus assembly, nuclear export of vRNP complexes

[1]Computationally predicted and available in the Influenza Research Database (IRD) (25).

Surface glycoproteins

Hemagglutinin

The HA glycoprotein is the sole protein encoded by the IAV genome segment 4 and has a pivotal role in receptor specificity and host range, antigenicity, and pathogenicity (1, 9). HA is the major surface protein on the virus particle and is responsible to bind to the virus receptor on host cells and mediate fusion between the cellular endosome membrane and virus envelope (26). There is also a role for the HA in virus assembly (1, 27). HA is a type I integral protein, ranging from 562 to 566 amino acids long, that is synthesized as a single polypeptide precursor (HA0) containing a N-terminal signal peptide that is removed after HA0 insertion into the rough endoplasmic reticulum (RER). HA0 is cleaved by host proteases into HA1 and HA2 subunits that are covalently linked by a disulfide bond to form a monomer (28-30). Three monomers are then associated non-covalently to form a homotrimer (31). HA typically contain 5 to 11 glycosylation sites that affect receptor binding affinity and/or specificity, antigenicity, replication, fusion activity, virulence, and host range (1).

Cleavage of the HA0 precursor into HA1 and HA2 subunits is required for fusion activity and, consequently, infectivity (32, 33). For most IAV strains, this occurs at a single basic amino acid at the HA cleavage site, which is only recognized by tissue specific trypsin-like proteases (34, 35). For some IAV strains, commonly referred to as highly pathogenic avian influenza virus (HPAIV) due the increased pathogenicity in chickens (36), the tissue tropism is expanded by the accumulation of multiple basic residues within the HA cleavage site. This allows ubiquitous HA0 precursor cleavage by

the furin protease in the Golgi apparatus, leading to systemic infection in susceptible hosts (37, 38).

Structurally, each HA monomer consists of a globular head domain composed completely of the HA1 subunit and a stalk domain that is composed of the HA2 and parts of HA1 subunits. Located in a small depression on the globular head of HA, the receptor-binding site (RBS) mediates virus binding to the virus receptor on host cells (29, 30). The binding site consists of a pocket of noncontiguous residues bordered by the 130-loop, the 150-loop, the 190-helix and the 220-loop. A set of well-conserved polar and nonpolar residues, including Tyr98, Trp153, His183, and Tyr195, forms the base of the binding site (29, 30, 39).

There are 16 HA subtypes (H1 to H16) found in nature, which are divided into two phylogenetic groups with marked antigenic differences: group 1 (H1, H2, H5, H6, H8, H9, H11, H12, H13 and H16) and group 2 (H3, H4, H7, H10, H14 and H15) (1, 9). Even though the overall amino acid sequence identity can be less than 50%, the structure and functions of these HAs are highly conserved (1). Recently, two unique IAVs were identified in bats. The HAs of these bat-derived viruses have been provisionally designated HL17 and HL18 (10). Structural and functional analyses of HL17 revealed a HA fold similar to other IAV HAs (40, 41). However, HL17 possesses a distorted RBS and a distinct fusion mechanism. This is consistent with a unique mechanism of entry that is independent of binding to canonical IAV virus receptors (40-42).

Neuraminidase

The NA is the second major glycoprotein of IAV and the only protein encoded by the IAV genome segment 6. NA is a type II integral membrane protein of 469 amino

acids long that is responsible for the removal of terminal sialic acid residues from virus particle, host cell surface and the surrounding environment glycoconjugates (1, 43). This sialidase activity thus is essential at several stages of the virus replication cycle, such as in aiding virus release of newly formed virus particles from infected cells (44), promoting virus spread (45), facilitating virus entry (45) and preventing virus superinfection (46). The essential role of NA in the virus replication cycle is exemplified by the fact that commercially available influenza virus antiviral therapy is focused almost exclusively on neuraminidase inhibitors, such as zanamivir and oseltamivir, that act by directly interfering with the enzymatic activity of NA (47).

Although largely ignored in vaccine development, antibodies against NA were shown to restrict virus replication and prevent severe disease (48, 49). Recently, NA has re-emerged as a potential vaccine antigen candidate for broadly protective influenza virus vaccines (48, 50). There are 9 NA subtypes (N1 to N9) found in nature, which are divided into two major groups based on sequence comparisons: group 1 (N1, N4, N5 and N8) and group 2 (N3, N6, N7 and N9) (1, 51). Additionally, two unique IAVs were identified in bats and their NAs have been provisionally designated NL10 and NL11 due the lack of canonical NA activity (52, 53).

Other gene products of influenza A virus

Polymerase basic 2 (PB2)

The PB2 protein is one of the subunits of the virus polymerase complex (54). Encoded by the IAV genome segment 1 as a polypeptide of 759 amino acids long, PB2 plays an essential role in the initiation of the synthesis of virus messenger RNA (mRNA)

transcripts by binding to the 5'-methylated cap of host pre-mRNA molecules (55, 56). Consequently, all virus mRNA transcripts have host-derived sequence on their 5'-end. Two aromatic residues (F363 and F404) were identified as essential for cap-binding activity and proposed to function by a mechanism shared with other cap-binding proteins, such as eIF4E and CBP20. Interestingly, these residues are conserved in the PB2 protein of IBV and ICV strains (57).

PB2 is an important determinant of pathogenicity and host range restriction (58). The PB2 E627K substitution is a well characterized mammalian adaptation mutation shown to confer efficient polymerase activity of avian IAVs in mammalian cells (59). This substitution is associated with increased virulence of H5N1 HPAIV isolated from humans (60). More recently, it was demonstrated PB2 E627K substitution modulates vRNP inhibition by the retinoic acid-inducible gene I (RIG-I) pathogen sensor (61). Amino acid substitutions on PB2 were also shown to contribute to the temperature sensitivity (*ts*) phenotype of IAV master donor viruses (MDVs) used in live attenuated influenza vaccines (62, 63).

Polymerase basic 1 (PB1)

The PB1 protein is the primary gene product encoded by the IAV genome segment 2. Consisting of a polypeptide of 757 amino acids, PB1 is one of the subunits of the virus polymerase complex and possesses conserved motifs characteristic of RNA-dependent RNA polymerase (RdRp) enzymes (54, 64). PB1 catalyzes the addition of nucleotides during virus RNA synthesis (56). The active site for the polymerase catalytic activity is an S-D-D motif at residues 444 to 446. Mutational analysis revealed these amino acids are critical for PB1 activity and that mutation in any one of these residues

renders the protein nonfunctional (64). Despite the lack of canonical activity on HA and NA proteins, PB1 from bat-derived influenza viruses retain this highly conserved catalytic motif (54). Additionally, PB1 contains N-terminal and C-terminal binding domains for virus polymerase complex inter-subunits interaction with PA and PB2, respectively (65, 66). As for PB2, modifications on PB1 were shown to contribute to the temperature sensitivity (*ts*) and attenuated (*att*) phenotypes of IAV master donor viruses (MDVs) used in live attenuated influenza vaccines (62, 63).

Polymerase acid (PA)

The PA protein is the primary gene product encoded by the IAV genome segment 3 and consists of a polypeptide of 716 amino acids long. Although PA forms one of the subunit of the virus polymerase complex (54), its main function remained unknown until recently. It was shown that the endonuclease activity of the virus polymerase complex resides in the PA subunit (67, 68), resolving conflicting early reports that attributed endonuclease activity to either PB2 (69) or PB1 (70). The endonuclease activity of PA is critical for snatching capped primers from host mRNA molecules to initiate virus mRNA transcript synthesis (67, 68). The endonuclease active site of PA is highly conserved in all influenza viruses and consists of a histidine and a cluster of three acidic residues, which bind two manganese ions (67). Mutations in critical residues of PA severely impaired or abolished transcriptional activity while replication activity of the virus polymerase complex was unaffected, confirming the specific role of endonuclease activity in virus transcription (68).

Nucleocapsid protein (NP)

The NP protein (498 amino acids) is the only protein encoded by the IAV genome segment 5. It is an arginine-rich protein that functions as a single-stranded RNA binding protein. NP is primarily responsible for the encapsidation of the segmented vRNA genome (1). Each vRNA segment is associated with multiple NP molecules and a single virus polymerase complex, comprising of the PB2, PB1 and PA proteins, to form the vRNP complex (17). Contrary to the prevailing model for vRNA encapsidation, NP does not bind the vRNA uniformly and regions of low-NP binding are enriched for predicted RNA secondary structures (71, 72). Following uncoating, NP play an essential role in the transport of vRNP complexes from the cytoplasm to the nucleus, where virus replication and transcription occur (1). NP possesses nuclear localization signals (NLSs) that interact with members of the karyopherin α (also called importin α) family to mediate the nuclear import of vRNP complexes (73). While other proteins in the vRNP complex also contain NLSs (74), it is the signals on NP that have been shown to be necessary for the import of the vRNP complex (75).

Matrix protein 1 (M1) and matrix protein 2 (M2)

The M1 protein (252 amino acids) is encoded by the IAV genome segment 7. M1 lies on the inner surface of the lipid membrane and provides structural integrity to the virus envelope (1, 76). Electron microscopy of virus particles indicates M1 functions as bridge between the lipid membrane, where M1 makes contact with the cytoplasmic tails of the virus surface proteins, and the vRNP core (76). During the virus replication cycle, M1 interacts with vRNP and NEP/NS2 and mediates the export of newly synthesized vRNP complexes out of the nucleus and into the cytoplasm. M1 also play a crucial role in

virus assembly (77, 78). Association of M1 with the site of assembly and budding at the plasma membrane is dependent on the interaction of M1 with the cytoplasmic tails of HA and NA (79).

The M2 protein (97 amino acids) is expressed from a virus mRNA transcript produced by splicing of segment 7 (80). M2 is a tetrameric type III integral membrane protein consisting of a short ectodomain, a transmembrane domain, and an endodomain. M2 is integrated in the virus envelope as a homotetramer and possesses ion channel activity (81). Its major role is to conduct protons from endosomes into the virus particle to facilitate virus uncoating (1, 81). M2 is also involved in virus particle morphology and genome packaging (82, 83). The high sequence conservation of the M2 ectodomain has pushed its development as a potential target for universal influenza virus vaccines (84).

Nonstructural protein 1 (NS1) and nuclear export protein/nonstructural protein 2 (NEP/NS2)

The NS1 protein (219 amino acids) is a multifunctional protein encoded by the IAV genome segment 8 (85, 86). NS1 exists as a homodimer and is highly expressed in infected cells early on in the infection. While recognized as a nonstructural protein, a recent report suggests NS1 is present in substantial amounts as an internal component of purified virus particles (15). Whether NS1 is actively packed into virus particles or merely incorporated as a passive bystander remains to be determined. The primary function of NS1 is to suppress the virus-induced host type I interferon (IFN- α/ β) response (87). As a multifunctional protein, NS1 is involved in a plethora of activities during the virus replication cycle, such as temporal regulation of virus RNA synthesis, control of virus mRNA splicing, enhancement of virus mRNA translation, regulation of

virus particle morphology, and involvement in strain-dependent pathogenesis. All of these functions of NS1 rely on its ability to participate in a multitude of protein-protein and protein-RNA interactions (88).

The NEP/NS2 protein (121 amino acids) is produced from a virus mRNA transcript generated by splicing of segment 8 (85, 86). Amounts of spliced NEP/NS2 virus mRNA correspond to approximately 10% of unspliced NS1 virus mRNA (89). During the virus replication cycle, NEP/NS2 interacts with M1 and vRNP and recruits the nuclear export machinery to direct the export of newly synthesized vRNP complexes (1, 77, 78, 90). Consistent with its role in the nuclear export of vRNP complexes, NEP/NS2 was shown to interact with members of the nuclear export machinery, including Crm1 and nucleoporins (90, 91). NEP/NS2 has also been implicated in aiding the virus budding process (92) and regulating the switch between virus transcription and replication (93).

Accessory proteins

To overcome the compact nature of the virus genome, IAV employs several strategies to increase genomic coding capacity including gene splicing, downstream translation initiation in normal or alternative overlapping frames, and ribosomal frameshifting (94). Some of these proteins are present in all of the IAV strains and play important roles in the virus replication cycle, such as M2 and NEP/NS2 that are produced utilizing the host cell spliceosome machinery. However, most of them are dispensable for virus replication. Classified as accessory proteins, these proteins vary in prevalence among IAV strains and often exhibit protein length polymorphism. These accessory proteins include PB1-F2, PB1-N40, PA-X, PA-N155, PA-N182 and M42 (95-99).

Evolving complexity of glycan receptor recognition

The HA of IAV is responsible to mediate cell attachment and entry by binding to virus receptors on the surface of susceptible host cells (1). While George K. Hirst first noted the ability of IAV to agglutinate red blood cells in 1941 (100), it took more than a decade to identify N-acetylneuraminic acid (Neu5Ac), a sialic acid (SA), as a potential receptor for IAV. Alfred Gottschalk proposed that absorption of IAV particles to potential host cells and agglutination of erythrocytes by IAV are intimately associated with the Neu5Ac content (101). Although SAs are recognized as IAV receptors, whether specific sialic acid-containing proteins crucial for IAV infection exist it is unclear. A recent study showed that HA binds to the sialylated voltage-dependent Ca^{2+} channel Cav1.2 to trigger intracellular Ca^{2+} oscillations and subsequent IAV entry and replication (102).

SAs consist of a family of nine carbon monosaccharides that occur as terminal residues in glycoconjugates, including glycoproteins and glycolipids. Neu5Ac and N-glycolylneuraminic acid (Neu5Gc) are the most common SAs found in mammals, both of which serve as anchors for IAV. Neu5Gc is present in many animal species, but humans lack the ability to synthesize this particular SA due to a genetic defect on the enzyme gene responsible for its synthesis. Humans can incorporate Neu5Gc from the diet (103). SA are added to the penultimate sugar of glycoconjugates by sialyltransferases (STs). When added to galactose (Gal), SA is linked to the hydroxyl attached to carbon-3 of Gal to form an α2-3 glycosidic linkage (SAα2-3Gal) by ST3Gal transferases or to the hydroxyl group attached to carbon-6 of Gal to form an α2-6 glycosidic linkage (SAα2-6Gal) by ST6Gal transferases (103, 104).

Highlighting the importance of the SA glycosidic linkage stereochemistry for host range restriction, IAVs isolated from different hosts show specificity toward SAs with different linkages. Avian IAVs were shown to have a general preference for SAα2-3Gal (avian-type) while human IAVs exhibited marked specificity for SAα2-6Gal (human-type) (103, 105, 106). This simplistic view was consistent with early findings that human tracheal epithelial cells contained mostly SAα2-6Gal while gut epithelial cells from ducks possessed mostly SAα2-3Gal. Virus specificity is not absolute with some avian and human IAVs retaining the ability of binding to SA receptors with either linkages. Swine IAVs were shown to display preference to both SAα2-3Gal and SAα2-6Gal, or exclusively to SAα2-6Gal (1, 103, 105, 106).

Subsequent findings started to depict a more complex picture. *In vitro* differentiation studies of human respiratory tissues demonstrated nonciliated epithelial cells possessed SAα2-6Gal receptors and were predominantly infected by human IAVs. In contrast, ciliated cells contained SAα2-3Gal receptors and were preferentially infected by avian IAVs (45). Additional studies showed that while SAα2-6Gal receptors are dominant on epithelial cells in nasal mucosa, paranasal sinuses, pharynx, trachea, and bronchi, SAα2-3Gal receptors are found on nonciliated cuboidal bronchiolar cells at the junction between the respiratory bronchiole and alveolus, and on type II cells lining the alveolar wall. SAα2-3Gal receptors were also commonly found in corneal and conjunctival epithelial cells of the human eye (107, 108). These findings may explain, at least in part, the localization and severity of H5N1 IAV pneumonia as well as the conjunctivitis associated with H7 IAV infections in humans (107-109).

Most of these studies made use of lectins as an indirect probe for the presence and distribution of IAV glycan receptors in tissue sections (45, 107, 108, 110). Lectins form a diverse group of proteins that have in common the ability to bind defined carbohydrate structures, such as the SAα2-3Gal linkage-specific lectin *Maackia amurensis* agglutinin-I (MAL-I) isolated from seeds of *M. amurensis* and the SAα2-6Gal linkage-specific lectin *Sambucus nigra* agglutinin (SNA) isolated from the elderberry plant *S. nigra* (103, 111-114). Interestingly, a recent report (115), evaluating glycan recognition in the porcine lung, showed that IAV binding profile did not directly correlate with the recognition profile determined for the lectins SNA and MAL-I (45, 103, 107, 108). Whereas some viruses bound to glycans that reflected the profile observed with specific lectins, other viruses bound to highly specific subsets of glycans. For a number of virus strains, significant binding was observed to glycans recognized by neither SNA or MAL-I (115). These results question the utility of these lectins, which are often used to probe the abundance of glycan receptors with defined specificity (45, 103, 107, 108, 115).

It is clear that not only sialic acid linkage but the underlying glycan composition and structure, chain length, presentation of receptors and conformational entropy also play a role in the complex interactions that lead to productive attachment of IAV to host cells (115-117). A pioneering study showed that a characteristic structural topology, and not the linkage itself, enabled specific binding to sialylated glycans and that the recognition of this topology may be critical for adaptation of IAV to bind glycans in the human upper respiratory tract (116). The topology of SAα2-3Gal and SAα2-6Gal is governed by the glycosidic torsion angle with the SAα2-3Gal linkage adopting a cone-like topology while the SAα2-6Gal linkage, due to conformational flexibility, occupies a

cone-like structure as well as a wider umbrella-like topology (116). Whereas the cone-like topology is characteristic of SAα2-3Gal as wells as short SAα2-6Gal glycans such as single lactosamine branches, the umbrella-like topology is unique to SAα2-6Gal and is typically adopted by long glycans with multiples repeating lactosamine units (116).

Advances in carbohydrate chemistry has led to the development of defined glycan microarrays that have been successfully applied to examine the binding specificity of IAV (118-120). These glycan platforms consist of hundreds of synthetic glycan motifs displayed on the surface of the array. These arrays not only revealed clear differences in receptor preference for SAα2-3Gal or SAα2-6Gal linkage, but also detected fine differences in HA specificity, such as preferences for fucosylation, sulfation and sialylation in the underlying glycan structure (119). However, a limitation of most of these arrays is that they are based on chemically or enzymatically synthesized glycan structures and represent only a subset of the endogenous glycan diversity found in natural hosts (115). A newly reported glycan array has addressed this limitation by incorporating a series of glycans with extended lactosamine branches, previously shown to be essential for adaptation of IAV in the human upper respiratory tract (116, 121).

Though these arrays provide valuable information on virus binding, which receptors are relevant for IAV attachment, replication, and transmission *in vivo* remain poorly understood (122). For example, the restrictions imposed by the synthetic nature of such arrays was evidenced by a recent report in which isolated mutant IAVs showed no detectable binding to the glycans in the tested array but are capable of replicating both in cell culture and in the mouse respiratory tract (115, 123). The need for more relevant arrays has led to the adaptation of glycan platforms to the characterization and profiling

of natural glycans receptors derived from pertinent tissues in natural hosts of IAV, including human airway epithelial cell line (116), human respiratory tract tissues (124), ferret respiratory tract tissues (125), porcine airway epithelial cells (126) and porcine lung (115). These studies demonstrated a diversity of glycan structures that IAVs recognize but are not comprehensively represented on synthesized glycan arrays. Nevertheless, details on the density, distribution and organization of glycans in host tissues are lost during the preparation of these natural glycan arrays.

Although HA has a canonical role in mediating binding to virus receptors, recent reports have suggested that NA can also acquire receptor-binding activity, adding another layer of complexity to IAV glycan receptor recognition (127, 128). Human H3N2 IAVs were shown to acquire the D151G substitution on the NA that resulted in a change in the specificity of NA such that it gained the capacity to bind receptors, which were refractory to enzymatic cleavage, without altering its ability to remove receptors for HA (128). The residue 151 is located in the 150-loop at the edge of the active site of NA. Additional changes in other residues in the 150-loop associated with receptor-binding activity in NA from different subtypes suggests a more general role for the conformation of the 150-loop in determining the specificity to bind SA receptors and mediate virus attachment (127-130).

Molecular determinants of glycan receptor specificity

The receptor-binding specificity of HA is responsible for the host range restriction of IAV. Consequently, efficient human-to-human transmission requires preferential recognition of "human-type" receptors, a premise supported by the finding that the

earliest isolates in the 1918, 1957, and 1968 pandemics preferentially recognized SAα2-6Gal, rather than SAα2-3Gal, receptors. The differential recognition of SAα2-3Gal or SAα2-6Gal by IAVs is determined by the nature of the amino acids that form the receptor-binding pocket (103, 131), with the changes and positions required to switch receptor specificity varying among HA subtypes. Early work showed a single change in residue 226 (H3 numbering) from leucine to glutamine in an H3 IAV led to a change in binding preference from SAα2-6Gal to SAα2-3Gal (132). Later studies undoubtedly demonstrated that residue 226 as well as 228 play an important role in the receptor specificity of H2 and H3 IAVs during the human adaptation of avian IAVs. Glutamine at residue 226 (226Q) and glycine at residue 228 (228G) determine specificity for SAα2-3Gal, whereas leucine (226L) and serine (228S) confer SAα2-6Gal specificity (132, 133).

For H1 IAVs, residues 190 and 225 determine receptor specificity with glutamic acid at residue 190 (190E) and glycine at residue 225 (225G) conferring SAα2-3Gal specificity, while aspartic acid at both positions (190D/225D) is required for SAα2-6Gal specificity (134, 135). For H7N9 IAVs, several alternative combinations of 3 amino acid substitutions (V186G/K-K193T-G228S or V186N-N224K-G228S) conferred a switch to SAα2-6Gal specificity (136). In case of H9N2 and H6N1 IAVs, a single amino acid change (Q226L for H9N2 and G225D for H6N1) led to a switch to SAα2-6Gal preference (110, 137). Interestingly, a H5N1 IAV carrying substitutions associated with SAα2-6Gal specificity in H3 and H2 IAVs (226L/228S) retained SAα2-3Gal binding while gaining SAα2-6Gal specificity. However, when substitutions associated with SAα2-6Gal specificity in H1 IAV (190D/225D) were introduced, the resulting H5N1 IAV lost the ability to binding to both SAα2-3Gal or SAα2-6Gal glycans (138).

Given the relatively plastic nature of glycans, binding to HA incurs an entropic penalty proportional to the degree of conformational constriction (117). Considering the structural topology of sialylated glycan receptors and the conformational flexibility of SAα2-6Gal, (116), it has been proposed that preferential binding of SAα2-6Gal glycans require amino acid substitutions in the RBS that overcome the larger entropic penalty associated with binding to more flexible SAα2-6Gal receptors (117). It is important to emphasize that while these residues are critical to receptor binding specificity, other residues on the HA have the potential to affect receptor binding functions, including receptor binding avidity (123, 139-141).

The evolution of influenza A viruses

IAV strains evolve via a complex process that involves the gradual accumulation of mutations and genome reassortment, which are known as antigenic drift and antigenic shift, respectively (131, 142). Antigenic drift occurs as a result of the introduction of point mutations in the virus genome during the error-prone activity of the virus polymerase complex. While such point mutations occur continuously, antigenic drift refers specifically to the process by which they become fixed in variant strains. For human and swine IAVs, the major selective pressure driving antigenic drift is imposed by antibody responses against the HA and, to a lesser extent, NA. When these variant strains are no longer neutralized by antibodies to the "parental" strain, its genotype predominates and the mutations it carries become fixed (131, 143). Nonsynonymous mutations on the HA and NA of human IAVs occur at a frequency of less than 1% per year. Nonetheless,

antigenic drift variant strains can circulate for 1 to 5 years before being completely replaced by a newly variant strain (131).

Antigenic shift involves the introduction of a novel, antigenically distinct HA or NA segment by reassortment to which the human population lack preexisting immunity. Reassortment refers to the exchange of intact gene segments between viruses that coinfect the same cell. If the novel IAV can sustain efficient human to human transmission, it will lead to high infection rates in the immunologically naive population and cause a global pandemics (131, 143). While antigenic shift represent sporadic occurrences, 5 antigenic shift events have occurred over the last 100 years: (i) in 1918, with the emergence of the H1N1 virus associated with the "Spanish influenza"; (ii) in 1957, with the replacement of H1N1 virus by the H2N2 virus, causing the "Asian influenza"; (iii) in 1968, when the H3N2 virus replaced H2N2 virus, leading to the "Hong Kong influenza"; (iv) in 1977, when the H1N1 virus reappeared, causing the "Russian influenza"; and (v) in 2009, when a novel, antigenically distinct H1N1 virus (A(H1N1)pdm2009) replaced the H1N1 virus (seasonal H1N1), leading to the "2009 H1N1 pandemic" (131, 144, 145). The antigenic shift that caused the 1957 and 1968 pandemics resulted from reassortment between human and avian IAVs (131). By contrast, the A(H1N1)pdm2009 virus resulted from reassortment events with swine, avian and human IAVs (144, 145). Phylogenetic evidence indicates that the H1N1 virus associated with the Spanish influenza pandemic was directly introduced from an avian reservoir into the human population (146, 147).

Molecular determinants of antigenic drift in the H3 hemagglutinin

As previously mentioned, the HA protein is the major virus surface glycoprotein and the primary target of the antibody response following natural IAV exposure. The majority of the antibody responses to HA are directed against the HA globular head (1). Due to a mechanism known as antigenic drift, HA can escape the host preexisting immunity. Early work in 1980s combining x-ray crystallography, comparative sequence analysis, and the characterization of escape mutants defined 5 antigenic sites (designated A to E) on the globular head of HA H3 involved in antibody binding and recognition. Site A is formed by a protruding loop (residues 140-146); site B is formed by a loop (residues 155-160) and an α-helix (residues 188-198) and is situated at the membrane distal end of HA; site C is located at the base of the globular domain in the antiparallel sheet of HA1; site D is situated near the trimeric interface of the globular head domains; and site E lies near the bottom of the globular distal domain between sites C and A (148, 149).

It was proposed that the emergence of antigenically drifted strains of epidemiologic importance would require four or more amino acid substitutions located in two or more of these antigenic sites (149). Because antigenic drift can result in reduced vaccine effectiveness (131), IAV vaccines must be periodically updated. The World Health Organization coordinates a global influenza surveillance network that evaluates the antigenic properties of human IAVs using the hemagglutination inhibition (HI) assay. Additional surveillance information is provided by sequencing a subset of these strains. The combined antigenic, epidemiological, and genetic data are used to select strains for use in the vaccine (150). Computational approaches have become an invaluable tool for

studying and predicting antigenic drift using data from global influenza surveillance network (131, 150).

Using antigenic cartography, a computational method to quantify binding assay data such as hemagglutination inhibition (HI) data, Smith *et al* characterized the antigenic evolution of human H3N2 IAV strains. The punctuated nature of antigenic evolution led to the clustering of antigenically related IAV strains rather than to form a continuous antigenic lineage. Strikingly, the order of clusters in the map was mostly chronological. Smith *et al* showed that human H3N2 IAVs circulating from 1968 to 2003 could be grouped into 11 antigenic clusters, with each cluster eventually being replaced by an emerging cluster with distinct antigenic properties over time (150). Between 1 and 13 amino acid substitutions were associated with each of the antigenic cluster transitions (150). Defining the precise molecular determinants of antigenic drift has important implications for understanding IAV evolution and this study laid the groundwork to pinpoint the molecular basis responsible for antigenic change.

Although more than 130 residues on the HA were previously characterized as putative antigenic sites (148, 149), Koel *et al* showed that only seven residues (145, 155, 156, 158, 159, 189 and 193) have major impact on the emergence of antigenically distinct variant strains of human H3N2 IAVs. All seven residues were located adjacent to the RBS, indicating these residues may also influence receptor binding (151). Parallel studies characterizing the antigenic evolution of swine H3N2 IAV strains circulating in the United States defined distinct antigenic clusters among swine H3N2 IAVs (152). Surprisingly, the substitutions that resulted in marked antigenic differences were located at six amino acid positions (145, 155, 156, 158, 159, and 189) near the RBS (151, 152).

These positions have been validated experimentally, confirming them as major drivers of antigenicity (153). These results suggest greater antigenic diversity of swine H3N2 viruses circulating in United States considering the short timespan since these viruses have been introduced, further hampering the development of effective vaccination strategies.

Several reports have demonstrated the significance of amino acid substitutions near the RBS in modulating antigenic drift and immune evasion for avian H5N1 (154), equine H3N8 (155), and human H1N1 IAVs (151, 156). Considering that only a small number of amino acid substitutions are required to the emergence of new antigenically variant strains, it is surprising that new variants are observed only every few years (142). The close proximity of these substitutions to the RBS has led to the hypothesis that emerging substitutions in these residues must drive antigenic change and immune escape without disrupting receptor binding properties (142, 151). In agreement with this hypothesis, IAVs appear to utilize a small subset of amino acid substitutions at these residues that are recycled sporadically (151-153, 157, 158). Little is known about the plasticity and functional consequences of amino acid substitutions in residues identified as major determinants of antigenic drift.

References

1. **Shaw M, Palese P.** 2013. Orthomyxoviridae, vol 1. Lippincott Williams & Wilkins, Philadelphia, PA.
2. **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.
3. **Adams MJ, Lefkowitz EJ, King AMQ, Harrach B, Harrison RL, Knowles NJ, Kropinski AM, Krupovic M, Kuhn JH, Mushegian AR, Nibert M, Sabanadzovic S, Sanfacon H, Siddell SG, Simmonds P, Varsani A, Zerbini FM, Gorbalenya AE, Davison AJ.** 2017. Changes to taxonomy and the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2017). Arch Virol **162:**2505-2538.
4. **Lefkowitz EJ, Dempsey DM, Hendrickson RC, Orton RJ, Siddell SG, Smith DB.** 2018. Virus taxonomy: the database of the International Committee on Taxonomy of Viruses (ICTV). Nucleic Acids Res **46:**D708-D717.
5. **Presti RM, Zhao G, Beatty WL, Mihindukulasuriya KA, da Rosa AP, Popov VL, Tesh RB, Virgin HW, Wang D.** 2009. Quaranfil, Johnston Atoll, and Lake Chad viruses are novel members of the family Orthomyxoviridae. J Virol **83:**11599-11606.
6. **Aamelfot M, Dale OB, Falk K.** 2014. Infectious salmon anaemia - pathogenesis and tropism. Journal of Fish Diseases **37:**291-307.

7. **Kosoy OI, Lambert AJ, Hawkinson DJ, Pastula DM, Goldsmith CS, Hunt DC, Staples JE.** 2015. Novel thogotovirus associated with febrile illness and death, United States, 2014. Emerg Infect Dis **21:**760-764.

8. **Taylor RM, Hurlbut HS, Work TH, Kingston JR, Hoogstraal H.** 1966. Arboviruses isolated from ARGAS TICKS IN Egypt: Quaranfil, Chenuda, and Nyamanini. Am J Trop Med Hyg **15:**76-86.

9. **Webster RG, Bean WJ, Gorman OT, Chambers TM, Kawaoka Y.** 1992. Evolution and ecology of influenza A viruses. Microbiol Rev **56:**152-179.

10. **Wu Y, Wu Y, Tefsen B, Shi Y, Gao GF.** 2014. Bat-derived influenza-like viruses H17N10 and H18N11. Trends Microbiol **22:**183-191.

11. **Anonymous.** 1980. A revision of the system of nomenclature for influenza viruses: a WHO memorandum. Bull World Health Organ **58:**585-591.

12. **Kuhn JH, Bao Y, Bavari S, Becker S, Bradfute S, Brister JR, Bukreyev AA, Chandran K, Davey RA, Dolnik O, Dye JM, Enterlein S, Hensley LE, Honko AN, Jahrling PB, Johnson KM, Kobinger G, Leroy EM, Lever MS, Muhlberger E, Netesov SV, Olinger GG, Palacios G, Patterson JL, Paweska JT, Pitt L, Radoshitzky SR, Saphire EO, Smither SJ, Swanepoel R, Towner JS, van der Groen G, Volchkov VE, Wahl-Jensen V, Warren TK, Weidmann M, Nichol ST.** 2013. Virus nomenclature below the species level: a standardized nomenclature for natural variants of viruses assigned to the family Filoviridae. Arch Virol **158:**301-311.

13. **Calder LJ, Wasilewski S, Berriman JA, Rosenthal PB.** 2010. Structural organization of a filamentous influenza A virus. Proc Natl Acad Sci U S A **107:**10685-10690.

14. **Chu C.** 1949. Filamentous Forms Associated with Newly Isolated Influenza Virus. The Lancet **253**.

15. **Hutchinson EC, Charles PD, Hester SS, Thomas B, Trudgian D, Martinez-Alonso M, Fodor E.** 2014. Conserved and host-specific features of influenza virion architecture. Nat Commun **5:**4816.

16. **Bourmakina SV, Garcia-Sastre A.** 2003. Reverse genetics studies on the filamentous morphology of influenza A virus. J Gen Virol **84:**517-527.

17. **Eisfeld AJ, Neumann G, Kawaoka Y.** 2015. At the centre: influenza A virus ribonucleoproteins. Nat Rev Microbiol **13:**28-41.

18. **Richardson JC, Akkina RK.** 1991. NS2 protein of influenza virus is found in purified virus and phosphorylated in infected cells. Arch Virol **116:**69-80.

19. **Desselberger U, Racaniello VR, Zazra JJ, Palese P.** 1980. The 3' and 5'-terminal sequences of influenza A, B and C virus RNA segments are highly conserved and show partial inverted complementarity. Gene **8:**315-328.

20. **Skehel JJ, Hay AJ.** 1978. Nucleotide sequences at the 5' termini of influenza virus RNAs and their transcripts. Nucleic Acids Res **5:**1207-1219.

21. **Moss B, Keith JM, Gershowitz A, Ritchey MB, Palese P.** 1978. Common sequence at the 5' ends of the segmented RNA genomes of influenza A and B viruses. J Virol **25:**312-318.

22. **Robertson JS.** 1979. 5' and 3' terminal nucleotide sequences of the RNA genome segments of influenza virus. Nucleic Acids Res **6:**3745-3757.

23. **Hsu MT, Parvin JD, Gupta S, Krystal M, Palese P.** 1987. Genomic RNAs of influenza viruses are held in a circular conformation in virions and in infected cells by a terminal panhandle. Proc Natl Acad Sci U S A **84:**8140-8144.

24. **Brownlee GG, Sharps JL.** 2002. The RNA polymerase of influenza a virus is stabilized by interaction with its viral RNA promoter. J Virol **76:**7103-7113.

25. **Zhang Y, Aevermann BD, Anderson TK, Burke DF, Dauphin G, Gu Z, He S, Kumar S, Larsen CN, Lee AJ, Li X, Macken C, Mahaffey C, Pickett BE, Reardon B, Smith T, Stewart L, Suloway C, Sun G, Tong L, Vincent AL, Walters B, Zaremba S, Zhao H, Zhou L, Zmasek C, Klem EB, Scheuermann RH.** 2017. Influenza Research Database: An integrated bioinformatics resource for influenza virus research. Nucleic Acids Res **45:**D466-D474.

26. **Skehel JJ, Wiley DC.** 2000. Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. Annu Rev Biochem **69:**531-569.

27. **Rossman JS, Lamb RA.** 2011. Influenza virus assembly and budding. Virology **411:**229-236.

28. **Segal MS, Bye JM, Sambrook JF, Gething MJ.** 1992. Disulfide bond formation during the folding of influenza virus hemagglutinin. J Cell Biol **118:**227-244.

29. **Wilson IA, Skehel JJ, Wiley DC.** 1981. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 A resolution. Nature **289:**366-373.

30. **Wiley DC, Skehel JJ.** 1987. The structure and function of the hemagglutinin membrane glycoprotein of influenza virus. Annu Rev Biochem **56:**365-394.

31. **Wiley DC, Skehel JJ, Waterfield M.** 1977. Evidence from studies with a cross-linking reagent that the haemagglutinin of influenza virus is a trimer. Virology **79:**446-448.

32. **Lazarowitz SG, Choppin PW.** 1975. Enhancement of the infectivity of influenza A and B viruses by proteolytic cleavage of the hemagglutinin polypeptide. Virology **68:**440-454.

33. **Klenk HD, Rott R, Orlich M, Blodorn J.** 1975. Activation of influenza A viruses by trypsin treatment. Virology **68:**426-439.

34. **Bottcher-Friebertshauser E, Freuer C, Sielaff F, Schmidt S, Eickmann M, Uhlendorff J, Steinmetzer T, Klenk HD, Garten W.** 2010. Cleavage of influenza virus hemagglutinin by airway proteases TMPRSS2 and HAT differs in subcellular localization and susceptibility to protease inhibitors. J Virol **84:**5605-5614.

35. **Baron J, Tarnow C, Mayoli-Nussle D, Schilling E, Meyer D, Hammami M, Schwalm F, Steinmetzer T, Guan Y, Garten W, Klenk HD, Bottcher-Friebertshauser E.** 2013. Matriptase, HAT, and TMPRSS2 activate the hemagglutinin of H9N2 influenza A viruses. J Virol **87:**1811-1820.

36. **Swayne DE, Suarez DL.** 2000. Highly pathogenic avian influenza. Rev Sci Tech **19:**463-482.

37. **Stieneke-Grober A, Vey M, Angliker H, Shaw E, Thomas G, Roberts C, Klenk HD, Garten W.** 1992. Influenza virus hemagglutinin with multibasic cleavage site is activated by furin, a subtilisin-like endoprotease. EMBO J **11:**2407-2414.

38. **Veits J, Weber S, Stech O, Breithaupt A, Graber M, Gohrbandt S, Bogs J, Hundt J, Teifke JP, Mettenleiter TC, Stech J.** 2012. Avian influenza virus hemagglutinins H2, H4, H8, and H14 support a highly pathogenic phenotype. Proc Natl Acad Sci U S A **109:**2579-2584.

39. **Xiong X, McCauley JW, Steinhauer DA.** 2014. Receptor binding properties of the influenza virus hemagglutinin as a determinant of host range. Curr Top Microbiol Immunol **385:**63-91.

40. **Zhu X, Yu W, McBride R, Li Y, Chen LM, Donis RO, Tong S, Paulson JC, Wilson IA.** 2013. Hemagglutinin homologue from H17N10 bat influenza virus exhibits divergent receptor-binding and pH-dependent fusion activities. Proc Natl Acad Sci U S A **110:**1458-1463.

41. **Sun X, Shi Y, Lu X, He J, Gao F, Yan J, Qi J, Gao GF.** 2013. Bat-derived influenza hemagglutinin H17 does not bind canonical avian or human receptors and most likely uses a unique entry mechanism. Cell Rep **3:**769-778.

42. **Moreira EA, Locher S, Kolesnikova L, Bolte H, Aydillo T, Garcia-Sastre A, Schwemmle M, Zimmer G.** 2016. Synthetically derived bat influenza A-like viruses reveal a cell type- but not species-specific tropism. Proc Natl Acad Sci U S A doi:10.1073/pnas.1608821113.

43. **Hirst GK.** 1942. The Quantitative Determination of Influenza Virus and Antibodies by Means of Red Cell Agglutination. J Exp Med **75:**49-64.

44. **Palese P, Tobita K, Ueda M, Compans RW.** 1974. Characterization of temperature sensitive influenza virus mutants defective in neuraminidase. Virology **61:**397-410.

45. **Matrosovich MN, Matrosovich TY, Gray T, Roberts NA, Klenk HD.** 2004. Human and avian influenza viruses target different cell types in cultures of human airway epithelium. Proc Natl Acad Sci U S A **101:**4620-4624.

46. **Huang IC, Li W, Sui J, Marasco W, Choe H, Farzan M.** 2008. Influenza A virus neuraminidase limits viral superinfection. J Virol **82:**4834-4843.

47. **Moscona A.** 2005. Neuraminidase inhibitors for influenza. N Engl J Med **353:**1363-1373.

48. **Krammer F, Fouchier RAM, Eichelberger MC, Webby RJ, Shaw-Saliba K, Wan H, Wilson PC, Compans RW, Skountzou I, Monto AS.** 2018. NAction! How Can Neuraminidase-Based Immunity Contribute to Better Influenza Virus Vaccines? MBio **9**.

49. **Eichelberger MC, Wan H.** 2015. Influenza neuraminidase as a vaccine antigen. Curr Top Microbiol Immunol **386:**275-299.

50. **Chen YQ, Wohlbold TJ, Zheng NY, Huang M, Huang Y, Neu KE, Lee J, Wan H, Rojas KT, Kirkpatrick E, Henry C, Palm AE, Stamper CT, Lan LY, Topham DJ, Treanor J, Wrammert J, Ahmed R, Eichelberger MC, Georgiou G, Krammer F, Wilson PC.** 2018. Influenza Infection in Humans Induces Broadly Cross-Reactive and Protective Neuraminidase-Reactive Antibodies. Cell **173:**417-429 e410.

51. **Russell RJ, Haire LF, Stevens DJ, Collins PJ, Lin YP, Blackburn GM, Hay AJ, Gamblin SJ, Skehel JJ.** 2006. The structure of H5N1 avian influenza neuraminidase suggests new opportunities for drug design. Nature **443:**45-49.

52. **Zhu X, Yang H, Guo Z, Yu W, Carney PJ, Li Y, Chen LM, Paulson JC, Donis RO, Tong S, Stevens J, Wilson IA.** 2012. Crystal structures of two subtype N10 neuraminidase-like proteins from bat influenza A viruses reveal a diverged putative active site. Proc Natl Acad Sci U S A **109:**18903-18908.

53. **Li Q, Sun X, Li Z, Liu Y, Vavricka CJ, Qi J, Gao GF.** 2012. Structural and functional characterization of neuraminidase-like molecule N10 derived from bat influenza A virus. Proc Natl Acad Sci U S A **109:**18897-18902.

54. **Pflug A, Guilligay D, Reich S, Cusack S.** 2014. Structure of influenza A polymerase bound to the viral RNA promoter. Nature **516:**355-360.

55. **Bouloy M, Plotch SJ, Krug RM.** 1980. Both the 7-methyl and the 2'-O-methyl groups in the cap of mRNA strongly influence its ability to act as primer for influenza virus RNA transcription. Proc Natl Acad Sci U S A **77:**3952-3956.

56. **Braam J, Ulmanen I, Krug RM.** 1983. Molecular model of a eucaryotic transcription complex: functions and movements of influenza P proteins during capped RNA-primed transcription. Cell **34:**609-618.

57. **Fechter P, Mingay L, Sharps J, Chambers A, Fodor E, Brownlee GG.** 2003. Two aromatic residues in the PB2 subunit of influenza A RNA polymerase are crucial for cap binding. J Biol Chem **278:**20381-20388.

58. **Almond JW.** 1977. A single gene determines the host range of influenza virus. Nature **270:**617-618.

59. **Subbarao EK, London W, Murphy BR.** 1993. A single amino acid in the PB2 gene of influenza A virus is a determinant of host range. J Virol **67:**1761-1764.

60. **Hatta M, Gao P, Halfmann P, Kawaoka Y.** 2001. Molecular basis for high virulence of Hong Kong H5N1 influenza A viruses. Science **293:**1840-1842.

61. **Weber M, Sediri H, Felgenhauer U, Binzen I, Banfer S, Jacob R, Brunotte L, Garcia-Sastre A, Schmid-Burgk JL, Schmidt T, Hornung V, Kochs G, Schwemmle M, Klenk HD, Weber F.** 2015. Influenza virus adaptation PB2-627K modulates nucleocapsid inhibition by the pathogen sensor RIG-I. Cell Host Microbe **17:**309-319.

62. **Jin H, Zhou H, Lu B, Kemble G.** 2004. Imparting temperature sensitivity and attenuation in ferrets to A/Puerto Rico/8/34 influenza virus by transferring the genetic signature for temperature sensitivity from cold-adapted A/Ann Arbor/6/60. J Virol **78:**995-998.

63. **Pena L, Vincent AL, Ye J, Ciacci-Zanella JR, Angel M, Lorusso A, Gauger PC, Janke BH, Loving CL, Perez DR.** 2011. Modifications in the polymerase genes of a swine-like triple-reassortant influenza virus to generate live attenuated vaccines against 2009 pandemic H1N1 viruses. J Virol **85:**456-469.

64. **Biswas SK, Nayak DP.** 1994. Mutational analysis of the conserved motifs of influenza A virus polymerase basic protein 1. J Virol **68:**1819-1826.

65. **Sugiyama K, Obayashi E, Kawaguchi A, Suzuki Y, Tame JR, Nagata K, Park SY.** 2009. Structural insight into the essential PB1-PB2 subunit contact of the influenza virus RNA polymerase. EMBO J **28:**1803-1811.

66. **Perez DR, Donis RO.** 2001. Functional analysis of PA binding by influenza a virus PB1: effects on polymerase activity and viral infectivity. J Virol **75:**8127-8136.

67. **Dias A, Bouvier D, Crepin T, McCarthy AA, Hart DJ, Baudin F, Cusack S, Ruigrok RW.** 2009. The cap-snatching endonuclease of influenza virus polymerase resides in the PA subunit. Nature **458:**914-918.

68. **Yuan P, Bartlam M, Lou Z, Chen S, Zhou J, He X, Lv Z, Ge R, Li X, Deng T, Fodor E, Rao Z, Liu Y.** 2009. Crystal structure of an avian influenza polymerase PA(N) reveals an endonuclease active site. Nature **458:**909-913.

69. **Shi L, Summers DF, Peng Q, Galarz JM.** 1995. Influenza A virus RNA polymerase subunit PB2 is the endonuclease which cleaves host cell mRNA and functions only as the trimeric enzyme. Virology **208:**38-47.

70. **Li ML, Rao P, Krug RM.** 2001. The active sites of the influenza cap-dependent endonuclease are on different polymerase subunits. EMBO J **20:**2078-2086.

71. **Williams GD, Townsend D, Wylie KM, Kim PJ, Amarasinghe GK, Kutluay SB, Boon ACM.** 2018. Nucleotide resolution mapping of influenza A virus nucleoprotein-RNA interactions reveals RNA features required for replication. Nat Commun **9:**465.

72. **Lee N, Le Sage V, Nanni AV, Snyder DJ, Cooper VS, Lakdawala SS.** 2017. Genome-wide analysis of influenza viral RNA and nucleoprotein association. Nucleic Acids Res **45:**8968-8977.

73. **Wang P, Palese P, O'Neill RE.** 1997. The NPI-1/NPI-3 (karyopherin alpha) binding site on the influenza a virus nucleoprotein NP is a nonconventional nuclear localization signal. J Virol **71:**1850-1856.

74. **Jones IM, Reay PA, Philpott KL.** 1986. Nuclear location of all three influenza polymerase proteins and a nuclear signal in polymerase PB2. EMBO J **5:**2371-2376.

75. **Cros JF, Garcia-Sastre A, Palese P.** 2005. An unconventional NLS is critical for the nuclear import of the influenza A virus nucleoprotein and ribonucleoprotein. Traffic **6:**205-213.

76. **Ruigrok RW, Barge A, Durrer P, Brunner J, Ma K, Whittaker GR.** 2000. Membrane interaction of influenza virus M1 protein. Virology **267:**289-298.

77. **Martin K, Helenius A.** 1991. Nuclear transport of influenza virus ribonucleoproteins: the viral matrix protein (M1) promotes export and inhibits import. Cell **67:**117-130.

78. **Yasuda J, Nakada S, Kato A, Toyoda T, Ishihama A.** 1993. Molecular assembly of influenza virus: association of the NS2 protein with virion matrix. Virology **196:**249-255.

79. **Ali A, Avalos RT, Ponimaskin E, Nayak DP.** 2000. Influenza virus assembly: effect of influenza virus glycoproteins on the membrane association of M1 protein. J Virol **74:**8709-8719.

80. **Lamb RA, Choppin PW.** 1981. Identification of a second protein (M2) encoded by RNA segment 7 of influenza virus. Virology **112:**729-737.

81. **Pinto LH, Holsinger LJ, Lamb RA.** 1992. Influenza virus M2 protein has ion channel activity. Cell **69:**517-528.

82. **Roberts PC, Lamb RA, Compans RW.** 1998. The M1 and M2 proteins of influenza A virus are important determinants in filamentous particle formation. Virology **240:**127-137.

83. **Grantham ML, Stewart SM, Lalime EN, Pekosz A.** 2010. Tyrosines in the influenza A virus M2 protein cytoplasmic tail are critical for production of infectious virus particles. J Virol **84:**8765-8776.

84. **Neirynck S, Deroo T, Saelens X, Vanlandschoot P, Jou WM, Fiers W.** 1999. A universal influenza A vaccine based on the extracellular domain of the M2 protein. Nat Med **5:**1157-1163.

85. **Lamb RA, Choppin PW.** 1979. Segment 8 of the influenza virus genome is unique in coding for two polypeptides. Proc Natl Acad Sci U S A **76:**4908-4912.

86. **Inglis SC, Barrett T, Brown CM, Almond JW.** 1979. The smallest genome RNA segment of influenza virus contains two genes that may overlap. Proc Natl Acad Sci U S A **76:**3790-3794.

87. **Garcia-Sastre A, Egorov A, Matassov D, Brandt S, Levy DE, Durbin JE, Palese P, Muster T.** 1998. Influenza A virus lacking the NS1 gene replicates in interferon-deficient systems. Virology **252:**324-330.

88. **Hale BG, Randall RE, Ortin J, Jackson D.** 2008. The multifunctional NS1 protein of influenza A viruses. J Gen Virol **89:**2359-2376.

89. **Lamb RA, Choppin PW, Chanock RM, Lai CJ.** 1980. Mapping of the two overlapping genes for polypeptides NS1 and NS2 on RNA segment 8 of influenza virus genome. Proc Natl Acad Sci U S A **77:**1857-1861.

90. **O'Neill RE, Talon J, Palese P.** 1998. The influenza virus NEP (NS2 protein) mediates the nuclear export of viral ribonucleoproteins. EMBO J **17:**288-296.

91. **Neumann G, Hughes MT, Kawaoka Y.** 2000. Influenza A virus NS2 protein mediates vRNP nuclear export through NES-independent interaction with hCRM1. EMBO J **19:**6751-6758.

92. **Gorai T, Goto H, Noda T, Watanabe T, Kozuka-Hata H, Oyama M, Takano R, Neumann G, Watanabe S, Kawaoka Y.** 2012. F1Fo-ATPase, F-type proton-translocating ATPase, at the plasma membrane is critical for efficient influenza virus budding. Proc Natl Acad Sci U S A **109:**4615-4620.

93. **Robb NC, Smith M, Vreede FT, Fodor E.** 2009. NS2/NEP protein regulates transcription and replication of the influenza virus RNA genome. J Gen Virol **90:**1398-1407.

94. **Yewdell JW, Ince WL.** 2012. Virology. Frameshifting to PA-X influenza. Science **337:**164-165.

95. **Wise HM, Hutchinson EC, Jagger BW, Stuart AD, Kang ZH, Robb N, Schwartzman LM, Kash JC, Fodor E, Firth AE, Gog JR, Taubenberger JK, Digard P.** 2012. Identification of a novel splice variant form of the influenza A virus M2 ion channel with an antigenically distinct ectodomain. PLoS Pathog **8:**e1002998.

96. **Wise HM, Foeglein A, Sun J, Dalton RM, Patel S, Howard W, Anderson EC, Barclay WS, Digard P.** 2009. A complicated message: Identification of a novel PB1-related protein translated from influenza A virus segment 2 mRNA. J Virol **83:**8021-8031.

97. **Chen W, Calvo PA, Malide D, Gibbs J, Schubert U, Bacik I, Basta S, O'Neill R, Schickli J, Palese P, Henklein P, Bennink JR, Yewdell JW.** 2001. A novel influenza A virus mitochondrial protein that induces cell death. Nat Med **7:**1306-1312.

98. **Jagger BW, Wise HM, Kash JC, Walters KA, Wills NM, Xiao YL, Dunfee RL, Schwartzman LM, Ozinsky A, Bell GL, Dalton RM, Lo A, Efstathiou S, Atkins JF, Firth AE, Taubenberger JK, Digard P.** 2012. An overlapping protein-coding region in influenza A virus segment 3 modulates the host response. Science **337:**199-204.

99. **Muramoto Y, Noda T, Kawakami E, Akkina R, Kawaoka Y.** 2013. Identification of novel influenza A virus proteins translated from PA mRNA. J Virol **87:**2455-2462.

100. **Hirst GK.** 1941. The Agglutination of Red Cells by Allantoic Fluid of Chick Embryos Infected with Influenza Virus. Science **94:**22-23.

101. **Gottschalk A.** 1957. Neuraminidase: the specific enzyme of influenza virus and Vibrio cholerae. Biochim Biophys Acta **23:**645-646.

102. **Fujioka Y, Nishide S, Ose T, Suzuki T, Kato I, Fukuhara H, Fujioka M, Horiuchi K, Satoh AO, Nepal P, Kashiwagi S, Wang J, Horiguchi M, Sato Y, Paudel S, Nanbo A, Miyazaki T, Hasegawa H, Maenaka K, Ohba Y.** 2018. A Sialylated Voltage-Dependent Ca(2+) Channel Binds Hemagglutinin and Mediates Influenza A Virus Entry into Mammalian Cells. Cell Host Microbe **23:**809-818 e805.

103. **Nicholls JM, Chan RW, Russell RJ, Air GM, Peiris JS.** 2008. Evolving complexities of influenza virus and its receptors. Trends Microbiol **16:**149-157.

104. **Harduin-Lepers A, Mollicone R, Delannoy P, Oriol R.** 2005. The animal sialyltransferases and sialyltransferase-related genes: a phylogenetic approach. Glycobiology **15:**805-817.

105. **Rogers GN, Paulson JC.** 1983. Receptor determinants of human and animal influenza virus isolates: differences in receptor specificity of the H3 hemagglutinin based on species of origin. Virology **127:**361-373.

106. **Palese P, Shaw M.** 2007. Orthomyxoviridae: the viruses and their replication, vol 2. Lippincott Williams & Wilkins, Philadelphia, PA.

107. **van Riel D, Munster VJ, de Wit E, Rimmelzwaan GF, Fouchier RA, Osterhaus AD, Kuiken T.** 2006. H5N1 Virus Attachment to Lower Respiratory Tract. Science **312:**399.

108. **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.

109. **Fouchier RA, Schneeberger PM, Rozendaal FW, Broekman JM, Kemink SA, Munster V, Kuiken T, Rimmelzwaan GF, Schutten M, Van Doornum GJ, Koch G, Bosman A, Koopmans M, Osterhaus AD.** 2004. Avian influenza A virus (H7N7) associated with human conjunctivitis and a fatal case of acute respiratory distress syndrome. Proc Natl Acad Sci U S A **101:**1356-1361.

110. **Wan H, Perez DR.** 2007. Amino acid 226 in the hemagglutinin of H9N2 influenza viruses determines cell tropism and replication in human airway epithelial cells. J Virol **81:**5181-5191.

111. **Sharon N.** 2007. Lectins: carbohydrate-specific reagents and biological recognition molecules. J Biol Chem **282:**2753-2764.

112. **Loris R.** 2002. Principles of structures of animal and plant lectins. Biochim Biophys Acta **1572:**198-208.

113. **Shibuya N, Goldstein IJ, Broekaert WF, Nsimba-Lubaki M, Peeters B, Peumans WJ.** 1987. The elderberry (Sambucus nigra L.) bark lectin recognizes the Neu5Ac(alpha 2-6)Gal/GalNAc sequence. J Biol Chem **262:**1596-1601.

114. **Wang WC, Cummings RD.** 1988. The immobilized leukoagglutinin from the seeds of Maackia amurensis binds with high affinity to complex-type Asn-linked oligosaccharides containing terminal sialic acid-linked alpha-2,3 to penultimate galactose residues. J Biol Chem **263:**4576-4585.

115. **Byrd-Leotis L, Liu R, Bradley KC, Lasanajak Y, Cummings SF, Song X, Heimburg-Molinaro J, Galloway SE, Culhane MR, Smith DF, Steinhauer DA, Cummings RD.** 2014. Shotgun glycomics of pig lung identifies natural endogenous receptors for influenza viruses. Proc Natl Acad Sci U S A **111:**E2241-2250.

116. **Chandrasekaran A, Srinivasan A, Raman R, Viswanathan K, Raguram S, Tumpey TM, Sasisekharan V, Sasisekharan R.** 2008. Glycan topology determines human adaptation of avian H5N1 virus hemagglutinin. Nat Biotechnol **26:**107-113.

117. **Ji Y, White YJ, Hadden JA, Grant OC, Woods RJ.** 2017. New insights into influenza A specificity: an evolution of paradigms. Curr Opin Struct Biol **44:**219-231.

118. **Smith DF, Song X, Cummings RD.** 2010. Use of glycan microarrays to explore specificity of glycan-binding proteins. Methods Enzymol **480:**417-444.

119. **Stevens J, Blixt O, Paulson JC, Wilson IA.** 2006. Glycan microarray technologies: tools to survey host specificity of influenza viruses. Nat Rev Microbiol **4:**857-864.

120. **Blixt O, Head S, Mondala T, Scanlan C, Huflejt ME, Alvarez R, Bryan MC, Fazio F, Calarese D, Stevens J, Razi N, Stevens DJ, Skehel JJ, van Die I, Burton DR, Wilson IA, Cummings R, Bovin N, Wong CH, Paulson JC.** 2004. Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. Proc Natl Acad Sci U S A **101:**17033-17038.

121. **Peng W, de Vries RP, Grant OC, Thompson AJ, McBride R, Tsogtbaatar B, Lee PS, Razi N, Wilson IA, Woods RJ, Paulson JC.** 2017. Recent H3N2 Viruses Have Evolved Specificity for Extended, Branched Human-type Receptors, Conferring Potential for Increased Avidity. Cell Host Microbe **21:**23-34.

122. **de Graaf M, Fouchier RA.** 2014. Role of receptor binding specificity in influenza A virus transmission and pathogenesis. EMBO J **33:**823-841.

123. **Bradley KC, Galloway SE, Lasanajak Y, Song X, Heimburg-Molinaro J, Yu H, Chen X, Talekar GR, Smith DF, Cummings RD, Steinhauer DA.** 2011. Analysis of influenza virus hemagglutinin receptor binding mutants with limited receptor recognition properties and conditional replication characteristics. J Virol **85:**12387-12398.

124. **Walther T, Karamanska R, Chan RW, Chan MC, Jia N, Air G, Hopton C, Wong MP, Dell A, Malik Peiris JS, Haslam SM, Nicholls JM.** 2013. Glycomic analysis of human respiratory tract tissues and correlation with influenza virus infection. PLoS Pathog **9:**e1003223.

125. **Jia N, Barclay WS, Roberts K, Yen HL, Chan RW, Lam AK, Air G, Peiris JS, Dell A, Nicholls JM, Haslam SM.** 2014. Glycomic characterization of respiratory tract tissues of ferrets: implications for its use in influenza virus infection studies. J Biol Chem **289:**28489-28504.

126. **Bateman AC, Karamanska R, Busch MG, Dell A, Olsen CW, Haslam SM.** 2010. Glycan analysis and influenza A virus infection of primary swine respiratory epithelial cells: the importance of NeuAc{alpha}2-6 glycans. J Biol Chem **285:**34016-34026.

127. **Hooper KA, Bloom JD.** 2013. A mutant influenza virus that uses an N1 neuraminidase as the receptor-binding protein. J Virol **87:**12531-12540.

128. **Lin YP, Gregory V, Collins P, Kloess J, Wharton S, Cattle N, Lackenby A, Daniels R, Hay A.** 2010. Neuraminidase receptor binding variants of human influenza A(H3N2) viruses resulting from substitution of aspartic acid 151 in the catalytic site: a role in virus attachment? J Virol **84:**6769-6781.

129. **Mohr PG, Deng YM, McKimm-Breschkin JL.** 2015. The neuraminidases of MDCK grown human influenza A(H3N2) viruses isolated since 1994 can demonstrate receptor binding. Virol J **12:**67.

130. **Tamura D, Nguyen HT, Sleeman K, Levine M, Mishin VP, Yang H, Guo Z, Okomo-Adhiambo M, Xu X, Stevens J, Gubareva LV.** 2013. Cell culture-

selected substitutions in influenza A(H3N2) neuraminidase affect drug susceptibility assessment. Antimicrob Agents Chemother **57:**6141-6146.

131. **Wright PF, Neumann G, Kawaoka Y.** 2013. Orthomyxoviruses, vol 1. Lippincott Williams & Wilkins, Philadelphia, PA.

132. **Rogers GN, Paulson JC, Daniels RS, Skehel JJ, Wilson IA, Wiley DC.** 1983. Single amino acid substitutions in influenza haemagglutinin change receptor binding specificity. Nature **304:**76-78.

133. **Connor RJ, Kawaoka Y, Webster RG, Paulson JC.** 1994. Receptor specificity in human, avian, and equine H2 and H3 influenza virus isolates. Virology **205:**17-23.

134. **Gamblin SJ, Haire LF, Russell RJ, Stevens DJ, Xiao B, Ha Y, Vasisht N, Steinhauer DA, Daniels RS, Elliot A, Wiley DC, Skehel JJ.** 2004. The structure and receptor binding properties of the 1918 influenza hemagglutinin. Science **303:**1838-1842.

135. **Matrosovich M, Tuzikov A, Bovin N, Gambaryan A, Klimov A, Castrucci MR, Donatelli I, Kawaoka Y.** 2000. Early alterations of the receptor-binding properties of H1, H2, and H3 avian influenza virus hemagglutinins after their introduction into mammals. J Virol **74:**8502-8512.

136. **de Vries RP, Peng W, Grant OC, Thompson AJ, Zhu X, Bouwman KM, de la Pena ATT, van Breemen MJ, Ambepitiya Wickramasinghe IN, de Haan CAM, Yu W, McBride R, Sanders RW, Woods RJ, Verheije MH, Wilson IA, Paulson JC.** 2017. Three mutations switch H7N9 influenza to human-type receptor specificity. PLoS Pathog **13:**e1006390.

137. **de Vries RP, Tzarum N, Peng W, Thompson AJ, Ambepitiya Wickramasinghe IN, de la Pena ATT, van Breemen MJ, Bouwman KM, Zhu X, McBride R, Yu W, Sanders RW, Verheije MH, Wilson IA, Paulson JC.** 2017. A single mutation in Taiwanese H6N1 influenza hemagglutinin switches binding to human-type receptors. EMBO Mol Med **9:**1314-1325.

138. **Stevens J, Blixt O, Tumpey TM, Taubenberger JK, Paulson JC, Wilson IA.** 2006. Structure and receptor specificity of the hemagglutinin from an H5N1 influenza virus. Science **312:**404-410.

139. **Hensley SE, Das SR, Bailey AL, Schmidt LM, Hickman HD, Jayaraman A, Viswanathan K, Raman R, Sasisekharan R, Bennink JR, Yewdell JW.** 2009. Hemagglutinin receptor binding avidity drives influenza A virus antigenic drift. Science **326:**734-736.

140. **Li Y, Bostick DL, Sullivan CB, Myers JL, Griesemer SB, Stgeorge K, Plotkin JB, Hensley SE.** 2013. Single hemagglutinin mutations that alter both antigenicity and receptor binding avidity influence influenza virus antigenic clustering. J Virol **87:**9904-9910.

141. **Wu NC, Thompson AJ, Xie J, Lin CW, Nycholat CM, Zhu X, Lerner RA, Paulson JC, Wilson IA.** 2018. A complex epistatic network limits the mutational reversibility in the influenza hemagglutinin receptor-binding site. Nat Commun **9:**1264.

142. **Petrova VN, Russell CA.** 2018. The evolution of seasonal influenza viruses. Nat Rev Microbiol **16:**47-60.

143. **Lowen AC.** 2017. Constraints, Drivers, and Implications of Influenza A Virus Reassortment. Annu Rev Virol **4:**105-121.

144. **Novel Swine-Origin Influenza AVIT, Dawood FS, Jain S, Finelli L, Shaw MW, Lindstrom S, Garten RJ, Gubareva LV, Xu X, Bridges CB, Uyeki TM.** 2009. Emergence of a novel swine-origin influenza A (H1N1) virus in humans. N Engl J Med **360:**2605-2615.

145. **Mena I, Nelson MI, Quezada-Monroy F, Dutta J, Cortes-Fernandez R, Lara-Puente JH, Castro-Peralta F, Cunha LF, Trovao NS, Lozano-Dubernard B, Rambaut A, van Bakel H, Garcia-Sastre A.** 2016. Origins of the 2009 H1N1 influenza pandemic in swine in Mexico. Elife **5**.

146. **Reid AH, Fanning TG, Hultin JV, Taubenberger JK.** 1999. Origin and evolution of the 1918 "Spanish" influenza virus hemagglutinin gene. Proc Natl Acad Sci U S A **96:**1651-1656.

147. **Taubenberger JK, Reid AH, Krafft AE, Bijwaard KE, Fanning TG.** 1997. Initial genetic characterization of the 1918 "Spanish" influenza virus. Science **275:**1793-1796.

148. **Wiley DC, Wilson IA, Skehel JJ.** 1981. Structural identification of the antibody-binding sites of Hong Kong influenza haemagglutinin and their involvement in antigenic variation. Nature **289:**373-378.

149. **Wilson IA, Cox NJ.** 1990. Structural basis of immune recognition of influenza virus hemagglutinin. Annu Rev Immunol **8:**737-771.

150. **Smith DJ, Lapedes AS, de Jong JC, Bestebroer TM, Rimmelzwaan GF, Osterhaus AD, Fouchier RA.** 2004. Mapping the antigenic and genetic evolution of influenza virus. Science **305:**371-376.

151. **Koel BF, Burke DF, Bestebroer TM, van der Vliet S, Zondag GC, Vervaet G, Skepner E, Lewis NS, Spronken MI, Russell CA, Eropkin MY, Hurt AC, Barr IG, de Jong JC, Rimmelzwaan GF, Osterhaus AD, Fouchier RA, Smith DJ.** 2013. Substitutions near the receptor binding site determine major antigenic change during influenza virus evolution. Science **342:**976-979.

152. **Lewis NS, Anderson TK, Kitikoon P, Skepner E, Burke DF, Vincent AL.** 2014. Substitutions near the hemagglutinin receptor-binding site determine the antigenic evolution of influenza A H3N2 viruses in U.S. swine. J Virol **88:**4752-4763.

153. **Abente EJ, Santos J, Lewis NS, Gauger PC, Stratton J, Skepner E, Anderson TK, Rajao DS, Perez DR, Vincent AL.** 2016. The Molecular Determinants of Antibody Recognition and Antigenic Drift in the H3 Hemagglutinin of Swine Influenza A Virus. J Virol **90:**8266-8280.

154. **Koel BF, van der Vliet S, Burke DF, Bestebroer TM, Bharoto EE, Yasa IW, Herliana I, Laksono BM, Xu K, Skepner E, Russell CA, Rimmelzwaan GF, Perez DR, Osterhaus AD, Smith DJ, Prajitno TY, Fouchier RA.** 2014. Antigenic variation of clade 2.1 H5N1 virus is determined by a few amino acid substitutions immediately adjacent to the receptor binding site. MBio **5:**e01070-01014.

155. **Lewis NS, Daly JM, Russell CA, Horton DL, Skepner E, Bryant NA, Burke DF, Rash AS, Wood JL, Chambers TM, Fouchier RA, Mumford JA, Elton DM, Smith DJ.** 2011. Antigenic and genetic evolution of equine influenza A (H3N8) virus from 1968 to 2007. J Virol **85:**12742-12749.

156. **Koel BF, Mogling R, Chutinimitkul S, Fraaij PL, Burke DF, van der Vliet S, de Wit E, Bestebroer TM, Rimmelzwaan GF, Osterhaus AD, Smith DJ, Fouchier RA, de Graaf M.** 2015. Identification of amino acid substitutions supporting antigenic change of influenza A(H1N1)pdm09 viruses. J Virol **89:**3763-3775.

157. **Barr IG, Russell C, Besselaar TG, Cox NJ, Daniels RS, Donis R, Engelhardt OG, Grohmann G, Itamura S, Kelso A, McCauley J, Odagiri T, Schultz-Cherry S, Shu Y, Smith D, Tashiro M, Wang D, Webby R, Xu X, Ye Z, Zhang W, Writing Committee of the World Health Organization Consultation on Northern Hemisphere Influenza Vaccine Composition f.** 2014. WHO recommendations for the viruses used in the 2013-2014 Northern Hemisphere influenza vaccine: Epidemiology, antigenic and genetic characteristics of influenza A(H1N1)pdm09, A(H3N2) and B influenza viruses collected from October 2012 to January 2013. Vaccine **32:**4713-4725.

158. **Klimov AI, Garten R, Russell C, Barr IG, Besselaar TG, Daniels R, Engelhardt OG, Grohmann G, Itamura S, Kelso A, McCauley J, Odagiri T, Smith D, Tashiro M, Xu X, Webby R, Wang D, Ye Z, Yuelong S, Zhang W, Cox N, Writing Committee of the World Health Organization Consultation on Southern Hemisphere Influenza Vaccine Composition f.** 2012. WHO

recommendations for the viruses to be used in the 2012 Southern Hemisphere Influenza Vaccine: epidemiology, antigenic and genetic characteristics of influenza A(H1N1)pdm09, A(H3N2) and B influenza viruses collected from February to September 2011. Vaccine **30:**6461-6471.

CHAPTER 3

PLASTICITY OF AMINO ACID SUBSTITUTIONS NEAR THE RECEPTOR BINDING SITE (RBS) OF H3 INFLUENZA A VIRUSES AND ITS IMPACT ON RECEPTOR BINDING AND ANTIBODY RECOGNITION[1]

[1]Jefferson Santos, Eugenio Abente, Adebimpe Obadan, Andrew J. Thompson, Lucas Ferreri, Ginger Geiger, Ana Silvia Gonzalez-Reiche, David F. Burke, Daniela S. Rajão, James C. Paulson, Amy L. Vincent and Daniel R. Perez. To be submitted to PLOS Pathogens.

Abstract

The influenza A virus (IAV) hemagglutinin (HA) initiates the virus life cycle by binding to terminal sialic acid (SA) residues on host cells. HA gradually accumulates amino acid substitutions that allow these viruses to escape population immunity through a mechanism known as antigenic drift. We recently confirmed that a small set of amino acid residues are largely responsible for driving antigenic drift in H3 influenza A viruses in swine. All identified residues are located adjacent to the HA receptor binding site (RBS), suggesting that substitutions associated with antigenic drift may also influence receptor binding. To determine whether there are functional constraints to substitutions near the RBS and their impact on receptor binding and antigenic properties, we carried out site-directed mutagenesis experiments at the single amino acid level using a swine-origin H3N2 virus backbone. We generated a panel of HA mutant viruses carrying substitution at residue 145 (H3 numbering) representing all 20 amino acids. Despite limited amino acid usage in nature, most substitutions at residue 145 were well tolerated and stably maintained without major impact on virus replication *in vitro*. All substitutions retained receptor binding specificity, but frequently led to decreased receptor binding avidity. Glycan microarray analysis showed that substitutions at residue 145 modulate binding to a broad range of glycans. Furthermore, antigenic characterization identified specific substitutions at residue 145 that altered antibody recognition. This work provides a better understanding of the functional effects of amino acid substitutions near the RBS and the interplay between receptor binding and antigenic drift.

Introduction

The surface hemagglutinin (HA) glycoprotein of influenza A virus (IAV) has a pivotal role in initiating the virus life cycle by binding to the virus receptor on target cells. Sialic acid (SA) residues are the receptors for IAV and occur as terminal monosaccharides in glycoproteins and glycolipids on the cell surface. SA receptors engaged by IAV are bound to galactose (Gal) in a α2-3 (SAα2-3Gal) or α2-6 (SAα2-6Gal) linkage configuration (1). Located in a small depression on the globular head of HA, the receptor-binding site (RBS) is composed of the 130-loop, the 150-loop, the 190-helix and the 220-loop. A set of conserved residues, including Tyr98, Trp153, His183, and Tyr195 (H3 numbering), forms the base of the RBS and is important for SA interaction (2, 3). Although some HA residues on the RBS are critical for receptor specificity (4-6), other residues may influence binding by modulating virus receptor binding avidity (7, 8).

IAV remains an important pathogen for humans and swine (9). While influenza vaccines are commercially available, the relative effectiveness of these vaccines are heavily dependent on the antigenic match of vaccine strains to circulating virus strains (10, 11). Most of the humoral immune response elicited by influenza vaccination or natural exposure is directed against the HA to block virus infection. Through a mechanism known as antigenic drift, IAV can circumvent the preexisting antibody response by rapidly selecting amino acid substitutions in key HA epitopes on the globular head, leading to the emergence of escape mutant viruses (12). Antigenic drift is a frequent cause of reduced vaccine effectiveness, especially for H3N2 IAVs (13).

Defining the molecular basis of antigenic drift has important implications for understanding IAV evolution and has been facilitated by methodological advances, such as antigenic cartography (14). A recent study identified seven residues (145, 155, 156, 158, 159, 189 and 193) on the globular head of the HA as the major determinants of antigenic drift during the evolution of human H3N2 IAVs (15). Interestingly, all seven residues were located adjacent to the RBS (15), and therefore these residues may also influence receptor binding. The importance of this small set of HA residues as major drivers of antigenic evolution has been demonstrated for IAVs circulating in other hosts, including swine H3N2 IAVs (16).

Antigenic changes in human H3N2 IAVs over the course of time were shown to be frequently caused by a single amino acid substitution in one of the seven residues, with specific substitutions involved in antigenic change more than once (15). One of those substitutions, N145K, was shown in two separate instances to be the sole major determinant of antigenic drift during the evolution of human H3N2 IAVs (14, 15). While there is evidence that N145K caused large antigenic changes (7, 14-16) and also altered receptor binding avidity (7), it remains unclear whether there are functional constraints limitations to substitutions at residue 145 or how alternative amino acid substitutions at this residue may affect receptor binding and antigenic properties.

In the present study, we carried out site-directed mutagenesis experiments to better understand functional constraints to substitutions near the RBS. A panel of H3 HA mutant viruses carrying a single amino acid substitution at residue 145, representing each of the 20 possible amino acids, was generated to evaluate the impact of substitution at this key HA residue on receptor binding and antibody recognition. Our data indicates that

residue 145 displayed remarkable amino acid plasticity *in vitro* tolerating multiple amino acid substitutions, many of which have not yet been observed in nature. Mutant viruses carrying substitutions at residue 145 showed no major impairment on virus replication. While all substitutions retained binding to SAα2-6Gal glycans, mutant viruses with substitutions not commonly found in nature displayed diminished receptor binding avidity. Antigenic characterization confirmed the impact of HA residue 145K in antibody immunodominance. These findings have important implications for understanding virus evolution and aiding the development of novel vaccine design approaches.

Materials and Methods

Cells

Madin-Darby canine kidney (MDCK) and human embryonic kidney 293T cells were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS). Cells were propagated at 37˚C in a humidified incubator under 5% CO_2 atmosphere.

Molecular cloning and virus rescue

A/turkey/Ohio/313053/2004 (H3N2), herein referred to as OH/04 wt, is a prototypic swine origin virus amenable to genetic manipulation by a stablished reverse genetics (RG) system (16-18). Single amino acid substitutions representing each of the 20 naturally occurring amino acids (**Table 3.1**) were inserted by site-directed mutagenesis at the codon corresponding to HA residue 145 (H3 numbering). To prevent the carryover of OH/04 wt HA plasmid DNA during PCR amplification, the cloned OH/04 wt HA segment was split into two overlapping plasmids: pDP-SD1 and pDP-SD2. The pDP-SD1

plasmid carries the mouse RNA polymerase I terminator (t1) followed by nucleotides 1-522 of OH/04 wt HA segment. The pDP-SD2 plasmid contains nucleotides 500-1762 of OH/04 wt HA segment followed by the human RNA polymerase I promoter (polI). To introduce the desired mutations into OH/04 wt HA, a fragment containing t1 followed by the first portion of the HA was amplified with forward primer 5'-ACC GGA GTA CTG GTC GAC CTC CGA AGT TGG GGG GGA GCA AAA GCA GG-3' and the respective reverse primer (**Table 3.1**) using pDP-SD1 as template. Similarly, a fragment comprising of the second portion of HA followed by polI was amplified from pDP-SD2 using reverse primer 5'-ATG CTG ACA ACG TCC CCG GCC CGG CGC TGC T-3' and the respective forward primer (**Table 3.1**). All PCR products were purified by gel extraction using QIAquick Gel Extraction Kit (Qiagen, Valencia, CA) and combined to produce RG-ready PCR-based HA segments for individual OH/04 HA 145 single amino acid mutants by overlapping PCR as previously described (19). All PCR reactions were performed with Phusion High Fidelity DNA polymerase (New England Biolabs, Ipswich, MA) and confirmed to be free of unwanted mutations by sequencing. Viruses were rescued by PCR-based RG using a co-culture of 293T/MDCK cells as previously described (19, 20). To generate OH/04 HA 145 single amino acid mutant viruses, the respective RG PCR-based HA segment was paired with the seven plasmids representing the remaining OH/04 wt gene segments. Following transfection, cells were incubated at 35˚C. After 24 h incubation, media was replaced with Opti-MEM I (Life Technologies, Carlsbad, CA) containing 1 μg/mL TPCK-trypsin (Worthington Biochemicals, Lakewood, NJ) and 1% antibiotics/antimycotic solution (Sigma-Aldrich, St. Louis, MO). Following virus rescue, virus stocks were amplified in MDCK cells. Virus stocks were

titrated by tissue culture infectious dose 50 ($TCID_{50}$) and virus titers were determined by the Reed and Muench method (21).

Whole-genome sequencing

Virus RNA from tissue culture supernatant virus stocks were purified using the RNeasy mini kit (Qiagen, Valencia, CA) or MagNA Pure LC RNA Isolation Kit (Roche Life Science, Mannheim, Germany). Isolated virus RNA served as template in a one-step reverse transcriptase PCR (RT-PCR) reaction for multi-segment, whole genome amplification (22). Amplicon sequence libraries were prepared as previously described (22) or using Nextera XT DNA Library Prep Kit (Illumina, San Diego, CA) according to the manufacturer's protocol. Barcoded libraries were multiplexed and sequenced on the high-throughput Illumina MiSeq sequencing platform in a paired-end 150 nt run format. De novo genome assembly was performed as described previously (22) and HA and NA-specific reads were mapped to OH/04 reference sequences using Geneious 10.1.3 (23).

In vitro growth kinetics

Confluent monolayers of MDCK cells were inoculated at 0.01 multiplicity of infection (MOI) for each virus. After 1 h incubation at 37˚C, virus inoculum was removed, and cells washed twice with 1X phosphate-buffered saline (PBS). Then, Opti-MEM I (Life Technologies, Carlsbad, CA) containing TPCK-trypsin (Worthington Biochemicals, Lakewood, NJ) and antibiotics/antimycotic solution (Sigma-Aldrich, St. Louis, MO) was added to the cells. At indicated time points, tissue culture supernatant from inoculated cells was collected for virus titer quantification. Virus RNA from tissue culture supernatant was isolated using the MagMAX-96 AI/ND Viral RNA Isolation Kit (Thermo Fisher Scientific, Waltham, MA). Virus titers were determined using a real-time

reverse transcriptase PCR (rRT-PCR) assay based on the influenza A matrix gene (24). The rRT-PCR was performed in a LightCycler 480 Real Time PCR instrument (Roche Diagnostics, Rotkreuz, Switzerland) using the LightCycler 480 RNA Master Hydrolysis Probes kit (Roche Life Science, Mannheim, Germany). A standard curve was generated using 10-fold serial dilutions from an OH/04 wt virus stock of known titer to correlate qPCR crossing point (Cp) values with virus titers, as previously described (25). Virus titers were expressed as $\log_{10}$ $TCID_{50}$/ml equivalents.

Hemagglutination (HA) assay

Chicken (Poultry Diagnostic and Research Center, Athens, GA), turkey (Poultry Diagnostic and Research Center, Athens, GA) and horse (Lampire Biologicals, Pipersville, PA) red blood cells (RBCs) were prepared from whole blood preparations using standard techniques (26). Virus HA assays were carried out using 0.5% (vol/vol in PBS) chicken RBCs, 0.5% (vol/vol in PBS) turkey RBCs or 1% (vol/vol in PBS) horse RBCs. Briefly, 50 µl of RBC preparations were added to 50 µl of two-fold serial dilutions of virus stocks allowed to incubate at room temperature for 45 min. After incubation agglutination was measured, and data expressed as the inverse of the highest dilution that allowed full agglutination.

Receptor cell binding assay

The receptor cell binding assay was performed as previously described (8, 16, 27). Briefly, 10% (vol/vol in PBS) turkey RBCs were pretreated for 1 h at 37°C with two-fold serial dilutions of bacterial neuraminidase from *Arthrobacter ureafaciens* (New England BioLabs, Ipswich, MA) or *Clostridium perfringens* (New England BioLabs, Ipswich, MA). Following neuraminidase treatment, treated RBCs were then washed twice with

cold PBS, and then resuspended to 1% (vol/vol in PBS). 50 μl of 1% RBCs treated with the different neuraminidase concentrations were added to 50 μl of each virus (8 HAU, as determined on untreated RBCs) and allowed to incubate at room temperature for 45 min. After incubation agglutination was measured, and data expressed as the maximal concentration of neuraminidase that allowed full agglutination.

Solid-phase assay of receptor binding specificity

The receptor-binding specificity was determined in a solid phase direct binding assay using monospecific preparations of peroxidase (HRP)-conjugated fetuin (fet-HRP). Monospecific preparations of fet-HRP were synthesized using α2-3-sialyltransferase from *Pasteurella multocida* (Sigma, St. Louis, MO) for 3-modified fetuin (3-fet-HRP) or α2-6-sialyltransferase from *Photobacterium damsela* (Sigma, St. Louis, MO) for 6-modified fetuin (6-fet-HRP), essentially as described previously (28). 96-well native fetuin-coated flat-bottom plates (Greiner Bio-One, Monroe, NC) were incubated overnight at 4ºC with 128 HAU of each virus in 0.02 M tris-buffered saline (TBS), pH 7.2–7.4. Virus samples were run in duplicate. Plates were washed three times with PBS and blocked with blocking solution [BS, PBS containing 0.1% neuraminidase-treated bovine serum albumin (BSA-NA)] for 2 h at room temperature. After blocking, plates were washed twice with ice-cold washing solution (WS, PBS containing 0.02% Tween 80) and incubated with two-fold serial dilution of 3-Fet-HRP (SAα2-3Gal) or 6-Fet-HRP (SAα2-6Gal) in reaction solution (RS, PBS containing 0.02% Tween-80, 0.1% BSA-NA and 2 μM oseltamivir carboxylate) for 1 h at 4ºC. After incubation, plates were washed five times with ice-cold WS before adding freshly prepared substrate solution (SS, 0.01% 3,3',5,5'-tetramethylbenzidine in 0.05 M sodium acetate with 0.03% H_20_2). Reactions

were allowed to proceed at room temperature for 30 min unless otherwise stated. Reactions were stopped with 3% (vol/vol in ddH_2O) H_2SO_4. Absorbance readings were obtained at 450 nm using a Victor x3 Multilabel Plate Reader (PerkinElmer, Waltham, MA).

Glycan array analysis

Virus stocks were grown in MDCK cells, clarified by low-speed centrifugation, and inactivated by treatment with 0.1% β-propiolactone (BPL) for 1 day at 4°C. Inactivated virus stocks were concentrated as described previously (29). Concentrated virus stocks were resuspended in PBS with 5% glycerol, aliquoted and stored at −80°C. HA titers of concentrated virus stocks were determined by HA assay using 0.5% (vol/vol in PBS) turkey RBCs. Glycan array analysis was performed using an NHS ester-coated glass microarray slide containing six replicates of 128 synthetic sialic acid-containing glycans, including terminal sequences as well as intact N-linked and O-linked glycans found on mammalian and avian glycoproteins and glycolipids (30). Whole influenza virus samples were diluted to 256 HAU in PBS containing 3% BSA (PBS-BSA) and incubated on the array surface for 1 h at room temperature in a humidity-controlled chamber. After incubation, slides were washed in PBS and incubated with OH/04 wt swine antisera (16) diluted 1:200 in PBS-BSA for 1 h. Slides were washed in PBS and incubated for 1 h in goat anti-pig IgG conjugated to fluorescein isothiocyanate (FITC; Thermo Fisher Scientific, Waltham, MA) diluted in PBS-BSA (20 μg/ml final concentration). Slides were washed twice in PBS, and in dH_2O, then dried prior to detection. Slide scanning to detect bound virus was conducted using an InnoScan 1100AL (Innopsys, Carbonne, France) fluorescent microarray scanner. Fluorescent signal intensity was measured using

Mapix (Innopsys, Carbonne, France) and mean intensity minus mean background of 4 replicate spots was calculated. A complete list of the glycans present in the array is presented in Appendix A. The array is comprised of non-sialoside control (1-10; Grey), SAα2-3Gal (11-76; Yellow) and SAα2-6Gal (77-128; Green) glycans. Glycans are grouped by structure type: L, linear; O, O-linked; N, N-linked and L^x, sialyl Le^x.

Antisera

Swine antisera against A/swine/New York/A01104005/2011 (H3N2) [NY/11], A/swine/Iowa/A01480656/2014 (H3N2) [IA/14] and A/turkey/Ohio/313053/2004 (H3N2) [OH/04] were generated in previous studies (16, 31). For each virus, two pigs were primed and boosted intramuscularly with UV-inactivated whole virus vaccine combined with commercial adjuvant. The pigs were humanely euthanized for blood collection. Goat anti-swine IgG-HRP (Seracare, Milford, MA) polyclonal antibody was used as secondary antibodies.

Hemagglutination inhibition (HI) assay

HI assays were performed as previously described (26). Prior to HI testing, swine antisera were treated overnight with receptor-destroying enzyme (Denka Seiken, Tokyo, Japan) and heat inactivated at 56°C for 30 min. Serial two-fold dilutions starting at 1:10 were tested for the ability to inhibit the agglutination of 0.5% turkey RBCs with 4 HAU of each virus. HI titers were recorded as the inverse of the highest dilution that inhibited hemagglutination.

Enzyme-linked immunosorbent assay (ELISA)

For ELISA, 96-well flat-bottom plates (Greiner Bio-One, Monroe, NC) were incubated overnight at 4ºC with 16 HAU of each virus in 1X coating solution (Seracare, Milford,

MA). Virus samples were run in duplicate. Plates were blocked with StartingBlock (PBS) blocking buffer (Thermo Fisher Scientific, Waltham, MA) for 1 h at room temperature. After blocking, plates were washed three times with PBS containing 0.05% tween 20 (PBS-T). Two-fold serial dilutions of swine antisera was added and allowed to incubate for 1 h at room temperature. After incubation, plates were washed three times with PBS-T before adding the secondary, HRP-conjugated polyclonal antibody. Plates were incubated at room temperature for 1 h. After incubation, plates were washed three times with PBS-T before adding freshly prepared SS buffer or ABTS 1-component microwell peroxidase substrate (Seracare, Milford, MA) for 1 h at room temperature. The reaction was stopped by adding 3% (vol/vol in ddH_2O) H_2SO_4 or ABTS peroxidase stop solution (Seracare, Milford, MA), respectively. Absorbance readings were obtained at 405 nm using a Victor x3 Multilabel Plate Reader (PerkinElmer, Waltham, MA).

Statistical analysis

All statistical analyses were performed using the GraphPad Prism Software Version 7 (GraphPad Software Inc., San Diego, CA). For multiple comparisons, two-way ANOVA was performed followed by a post-hoc test. When indicated, a *P* value below 0.05 ($P<0.05$) was considered significant.

Structure modeling

A model of the structure of the HA of A/turkey/Ohio/313053/2004 (H3N2) was built by homology modeling using Modeller v9.16 (32) based upon the crystal structure of multiple H3 HA proteins [Protein Data Bank (PDB) codes 2YP7, 1HA0, 2YP2, 4WE8 and 4WE5). The generated model was subsequently rendered with PyMOL v2.1 (33).

Computational analysis of HA sequences

The frequency distribution of amino acid identities at residue 145 (H3 numbering) was computed using the sequence variation analysis tool in the Influenza Research Database (IRD) (34). HA amino acid sequences from IAVs of the H3 subtype isolated from swine, human, avian, canine and equine hosts and publicly available in the IRD as of September 06, 2017 were analyzed. For humans, the frequency was calculated from precomputed data in the IRD database. Only amino acids that reached a frequency of at least 1% are labeled in the plot legend.

Results

HA residue 145 displayed remarkable amino acid plasticity *in vitro*.

The HA N145K substitution results in significant antigenic changes in both human and swine H3N2 IAVs (7, 14-16). To examine the amino acid frequency distribution at residue 145 in natural isolates, HA amino acid sequences from IAVs of the H3 subtype isolated in swine, human, avian, canine and equine hosts were obtained from the IRD (34) and analyzed (**Fig. 3.1A**). In swine-origin IAVs, around 64% of isolates possessed asparagine (Asn or N) at residue 145, followed by lysine (Lys or K) at ~33% and serine (Ser or S) at ~2.5%. In human-origin IAVs, about 42% of isolates carried Asn, followed by Ser at ~39% and Lys at ~18%. In avian-origin IAVs, the frequency was ~49% for Asn, ~45% for Ser and ~4.6% for glycine (Gly or G). In canine-origin IAVs, around ~73% of isolates possessed Asn, followed by aspartic acid (Asp or D) at ~24%. In equine-origin IAVs, 99% of isolates carried Asn (~99%). Remaining isolates possessing arginine (Arg or R), histidine (His or H), isoleucine (Ile or I), cysteine (Cys or C), methionine (Met or M), glutamine (Gln or Q), or threonine (Thr or T) at residue 145 were

found at a frequency below 1%. Interestingly, alanine (Ala or A), glutamic acid (Glu or E), leucine (Leu or L), phenylalanine (Phe or F), proline (Pro or P), tryptophan (Trp or W), tyrosine (Tyr or Y) and valine (Val or V) were not observed in any of the isolates from analyzed hosts.

To evaluate the impact of substitutions at residue 145, a panel of viruses carrying single substitutions representing each of the 20 amino acids was generated by reverse genetics (19, 20) (**Fig. 3.1B**). Since these viruses were generated in the context of a swine-origin IAV (**Fig. 3.1A**), 145N (wt), 145K and 145S are referred to as naturally occurring substitutions whereas the remaining substitutions are termed alternative or non-naturally occurring. Substitutions were introduced into the HA of OH/04 wt virus that naturally carries an Asn at residue 145 (**Fig. 3.1C**). Of the 19 OH/04 HA 145 single amino acid mutant viruses rescued, next generation sequencing analysis revealed that 12 substitutions (145A, 145C, 145G, 145H, 145K, 145L, 145M, 145P, 145Q, 145R, 145S and 145T) were well tolerated with 98-100% of sequenced reads possessed the expected codon at residue 145. For substitutions 145F, 145I, 145V and 145Y, the percentage of reads bearing the mutated codon was around 90% while the remaining 10% of reads showed a partial reversion to the wt codon (I145N, V145N and Y145N) or a partial transition for a codon specifying Ser (F145S). None of these HA substitutions introduced compensatory substitutions on the neuraminidase (NA) segment. Three substitutions (145D, 145E and 145W) were not well tolerated, leading to partial reversion to the wt codon (D145N) or partial transition to a codon specifying either Gly (E145G) or Leu (W145L). Additionally, compensatory substitutions emerged on the HA for 145E and

145W (T128A) or on the NA for 145E (T148I). For these reasons 145D, 145E and 145W viruses were excluded in further analysis.

As a first step toward characterizing the impact of substitutions at residue 145, viruses were assayed for their ability to agglutinate red blood cells (RBCs) from different species (turkey, chicken and horse) by standard hemagglutination (HA) assay. Turkey and chicken RBCs are known to carry both SAα2-3Gal and SAα2-6Gal receptors on their cell surface while horse RBCs mainly display SAα2-3Gal (35-37). In comparison to 145N (wt) virus, nearly all mutant viruses were found to agglutinate turkey RBCs efficiently, with the exception of 145C virus that displayed low HA titers (**Table 3.2**). In contrast, agglutination of chicken RBCs was less consistent with HA titers of some mutant viruses comparable to that of the 145N (wt) virus while other mutant viruses exhibited an 8 to 16-fold decrease in HA titers (145F, 145G, 145V and 145Y viruses) or no agglutination (145C virus) (**Table 3.2**). Only the 145F virus showed detectable, albeit low, HA titers when using horse RBCs (**Table 3.2**). Due the decreased ability to agglutinate RBCs, the 145C virus was not tested in further assays.

To further assess the impact of substitutions at residue 145, viruses were then compared in a multiple-step infection cycle *in vitro*. MDCK cells were infected at a low multiplicity of infection (MOI of 0.01). All viruses grew to high titers and displayed similar growth kinetics. There was no discernable difference in peak titers (~6.5 to 7.0 $\log_{10}$ $TCID_{50}$/ml equivalents) at 72 hpi (**Fig. 3.2**), with the exception of 145P and 145H viruses that showed a slightly lower peak titers (~5.5 $\log_{10}$ $TCID_{50}$/ml equivalents) (**Fig. 3.2**). Taken together, these results indicate that residue 145 demonstrated increased plasticity *in vitro* despite of the limited amino acid usage in nature. Furthermore, single

amino acid substitutions did not have a major impact in growth kinetics *in vitro* but modulated the ability of mutant viruses to agglutinate RBCs from different hosts.

Viruses with alternative substitutions at residue 145 retained SAα2-6Gal binding, but frequently displayed decreased receptor binding avidity

To further expand on the receptor binding characterization of the 145 mutants, we analyzed whether these substitutions were involved in modulating receptor avidity or receptor specificity. We tested virus receptor avidity by measuring agglutination of turkey RBCs previously treated with different concentrations of bacterial neuraminidase (**Fig. 3.3A-B**). All of the mutant viruses bound to desialylated RBCs to a varying degree. Naturally occurring substitutions showed the highest avidity to the receptor. Binding of 145N (wt) virus was indistinguishable from either the 145K (both neuraminidases) or 145S (*C. perfringens* neuraminidase) viruses (**Fig. 3.3A**). Nearly all mutant viruses carrying alternative substitutions at residue 145 displayed decreased receptor binding avidity, with the notable exception of the 145M virus that showed no discernable difference in binding compared to 145N (wt) virus (**Fig. 3.3A**). Overall, the results were consistent using different bacterial neuraminidases regardless of the neuraminidase specificity (**Fig. 3.3A-B**).

To evaluate the impact of substitutions at residue 145 on receptor binding specificity, we performed an ELISA-based assay using monospecific preparations of Fet-HRP as surrogates of binding to SAα2-3Gal (3-Fet-HRP) or SAα2-6Gal (6-Fet-HRP) (28) glycans (**Fig. 3.4A-R**). The assay reliably discriminated receptor binding specificity as evidenced by the viruses used as controls, with human pH1N1 showing a preference to SAα2-6Gal (**Fig. 3.4Q**) while avian ΔH5N1 displayed restricted binding to SAα2-3Gal

(**Fig. 3.4R**). Consistent with the location of HA residue 145 on the HA structure (**Fig. 3.1C**), all of the mutant viruses retained binding to SAα2-6Gal with no residual binding to SAα2-3Gal (**Fig. 3.4A-P**). In agreement with previous results, naturally occurring substitutions [145N (wt), 145K and 145S viruses] showed the greatest binding to SAα2-6Gal (**Fig. 3.4A, G and M**). As expected, mutant viruses carrying alternative substitutions at residue 145 displayed weaker binding to SAα2-6Gal (**Fig. 3.4B-F, H-L and N-P**) compared to viruses possessing naturally occurring substitutions (**Fig. 3.4A, G and M**). Among mutant viruses carrying alternative substitutions, 145M and 145P viruses demonstrated the highest binding (**Fig. 3.4I and J**) while 145F and 145G showed substantial decrease in binding to SAα2-6Gal (**Fig. 3.4C and D**). Overall, these results suggest substitutions in HA at residue 145 do not affect receptor specificity but modulate receptor avidity. Nearly all alternative substitutions led to decreased receptor binding avidity.

Substitutions at residue 145 modulated binding to a broad range of SAα2-6Gal glycans

To further compare the impact of substitutions at residue 145 on receptor binding specificity, glycan array analysis was performed for all viruses using a glycan microarray containing linear, O-linked, and N-linked glycans with extended poly-N-acetyl-lactosamine (poly-LacNAc) repeats found in human, swine and ferret airway tissues (30). The array provides a broad qualitative view of the binding preference for specific viruses. All of the viruses retained specificity to SAα2-6Gal glycans (**Fig. 3.5A-P**), corroborating the results from the simple linear glycan ELISA-based assay(**Fig. 3.4A-P**). Viruses carrying naturally occurring substitutions showed expanded binding to nearly all SAα2-6Gal glycan types in the array. OH/04 wt (145N) virus had slightly preferred binding to

O-glycans (**Fig. 3.5A**) while 145K virus appeared to favor binding to N-glycans (**Fig. 3.5G**). 145S virus binding was comparable to OH/04 wt (145N) virus with slight reduction in binding to linear and some of the smaller O-glycans (**Fig. 3.5M**). There were distinct patterns of binding for viruses carrying alternative substitutions. The 145F, 145L, 145M, 145P 145Q, 145T and 145V mutant viruses demonstrated binding to a broader range of glycans that was comparable to OH/04 wt (145N) virus or slight better (**Fig. 3.5C, H, I-K, N and O**). The 145A, 145G, 145H, 145I, 145R and 145Y mutant viruses displayed reduced binding to most receptors across the entire array, although they showed slightly preferred binding to O-linked glycans (**Fig. 3.5B, D, E, F, L and P**). Taken together, these results indicated that substitutions in HA at residue 145 modulate binding to a broader range of SAα2-6Gal glycans.

Substitutions at residue 145 modulated sera reactivity

For human H3N2 IAVs, it has been proposed that the emergence of the HA N145K substitution lead to a change in the antibody immunodominance (7). To assess whether a similar phenomenon occurs in swine-origin IAVs and how alternative substitutions at residue 145 affect antigenicity, the sera reactivity of each mutant virus in the panel was tested by whole virus ELISA (**Fig. 3.6A-F**) and HI assays (**Fig. 3.7A and B**). For these experiments, swine antisera against NY/11 or IA/14 wild type viruses were tested. The HA1 domain of these viruses differs by only two amino acids: NY/11 virus possesses 145N/289P while IA/14 carries 145K/289S. HA residue 289 is not antigenically relevant. Additionally, swine antisera against OH/04 wt virus was tested. To prevent confounding effects, the HA gene segment of the NY/11 and IA/14 viruses were rescued by reverse genetics in the background of 7 gene segments from the OH/04 wt

virus. The 1+7 NY/11 and IA/14 reverse genetics viruses were used as controls in ELISA and HI assays.

Reactivity of sera generated against the OH/04 wt virus was similar across the entire panel of OH/04 145 single amino acid mutant viruses (**Fig. 3.6A and D**), indicating substitutions at residue 145 in the context of OH/04 HA do not change antigenicity against wt homologous sera. Consistent with amino acid divergence between OH/04 wt and control viruses, OH/04 wt virus antisera had reduced reactivity to both NY/11 and IA/14 control viruses (**Fig. 3.6A and D**). Surprisingly, sera raised against the NY/11 virus displayed no discernible difference in reactivity among all of the virus evaluated regardless of the 145 amino acid substitution, including the NY/11 and IA/14 viruses (**Fig. 3.6B and E**). As observed for human H3N2 IAVs, the presence of 145K in a swine-origin IAV led to a refocus of the antibody immunodominance to target an epitope bearing HA residue 145K (**Fig. 3.6C and F**). Sera generated against the IA/14 virus showed decreased reactivity to the NY/11 virus (possessing HA residue 145N) compared to the IA/14 virus (carrying HA residue 145K) (**Fig. 3.6C and F**). Nearly all of the OH/04 145 single amino acid mutant viruses, including 145N (wt) virus, exhibited diminished reactivity to the IA/14 virus antisera (**Fig. 3.6C and F**). Interestingly, the 145K virus was the only mutant virus to consistently show increased reactivity to sera generated against the IA/14 virus (**Fig. 3.6C and F**), confirming the impact of HA residue 145K in the antibody immunodominance.

Antigenic characterization by HI assay was mostly consistent with the whole virus ELISA data. The OH/04 wt virus antisera reacted to all OH/04 145 single amino acid mutant viruses. Relative to the 145N (wt) virus, changes in HI titers were almost

indistinguishable (**Fig. 3.7A and B**). When tested against the OH/04 wt virus antisera, the IA/14 virus showed higher decrease in sera reactivity compared to the NY/11 virus (**Fig. 3.7A and B**). It is unclear whether reduced HI titers of the IA/14 control virus to the OH/04 wt virus antisera reflected increased receptor avidity (7, 8). In contrast to the ELISA data, most substitutions at HA residue 145 had subtle impact to NY/11 virus antisera reactivity, with the 145K virus exhibiting the most profound decrease in HI titers (**Fig. 3.7A and B**). Reactivity of the IA/14 virus antisera was drastically decreased to nearly all of the OH/04 145 single amino acid mutant viruses. In agreement with previous results, the 145K virus was the only mutant virus to show almost no discernible difference in sera reactivity compared to IA/14 virus (**Fig. 3.7A and B**). Collectively, these results indicate that residue 145K has a profound impact in the antibody immunodominance. Sera raised against a 145K bearing virus can recognize the epitope bearing such substitution in the context of a distinct H3 HA.

Discussion

HA engagement with terminal SA residues on host cells is an essential step in the IAV replication cycle. Not surprisingly, receptor binding specificity is a major host range restriction factor. In a simplistic view, IAVs of avian origin prefer binding to SAα2-3Gal glycans whereas those that circulate in humans favor binding to SAα2-6Gal receptors (5, 38-40). During human adaptation of avian-origin IAVs, receptor specificity switch is accompanied by substitutions on the HA (E190D/G225D in the H1 subtype and Q226L/G228S in the H2 and H3 subtypes) (5, 38, 41, 42). While these residues are critical to receptor binding specificity, other residues on the HA have the potential to

affect receptor binding functions, including receptor binding avidity (7, 8, 37, 43). The identification of up to 7 residues (145, 155, 156, 158, 159, 189 and 193) near the RBS as the major determinants of antigenic drift during the evolution of human and swine H3N2 IAVs (15, 16, 31) has led to the hypothesis that emerging substitutions in these residues must drive antigenic change and immune escape without disrupting receptor binding properties, potentially limiting the flexibility of aa residues in these positions (15, 44). In agreement with this hypothesis, analysis of HA sequences showed the occurrence of only a small subset of amino acid substitutions at these residues that are recycled sporadically (15, 16, 31, 45, 46).

Here we examined the amino acid plasticity at residue 145, one of the identified key determinants of antigenic drift, by mutating this residue to create a panel of H3 HA mutant viruses carrying a single amino acid substitution representing every possible amino acid in a swine H3N2 backbone. HA residue 145 showed extraordinary plasticity with 16 out of 19 substitutions (145A, 145C, 145F, 145G, 145H, 145I, 145K, 145L, 145M, 145P, 145Q, 145R, 145S, 145T, 145V and 145Y) being well tolerated *in vitro* (90-100% of sequenced reads possessing the mutated codon at residue 145). Whether these substitutions are tolerated *in vivo* is beyond the scope of the present report and remain to be determined. Consistent with the intricate balance between immune escape and receptor binding, three substitutions (145D, 145E and 145W) were not tolerated and led to the emergence of compensatory substitutions in either HA (T128A) and/or NA (T148I). HA T128A disrupted a potential glycosylation site (47, 48) and was shown to emerge along with a substitution at residue 145 in human H3N2 IAVs during the 2013-2014 season

(45). NA T148I has been associated with reduced NA activity, decreased susceptibility to neuraminidase inhibitors, and NA-mediated binding and agglutination (49, 50).

With the exception of 145C virus, there was no discernible difference on the ability of tested mutant viruses to agglutinate turkey RBCs. While turkey and chicken RBCs carry both SAα2-3Gal and SAα2-6Gal glycans on their cell surface (35-37), some of tested mutant viruses exhibited impaired ability to agglutinate chicken RBCs. In our glycan array analysis, some of these mutant viruses showed a preference to poly-LacNAc extended O-linked glycans terminated by α2-6-linked SA. Analysis of cell surface glycans of chicken RBCs revealed an absence of LacNAc repeats (51). Additionally, most of the sialylated O-linked glycans attached to surface glycoproteins of chicken RBCs terminated are by α2-3-linked SA (52). Considering these pieces of evidence, it is possible to speculate that these mutant viruses lost their ability to recognize particular SAα2-6Gal glycans present in chicken RBCs. Loss of the ability to agglutinate RBCs has plagued the antigenic characterization of recent human H3N2 IAVs (53). As previously reported for chicken RBCs (51), an in-depth analysis of major glycan structures present on RBCs from other hosts will help inform and address current issues on the antigenic characterization of IAVs.

Using independent assays, we demonstrated that all tested mutant viruses retained binding to SAα2-6Gal glycans and these observations are consistent with the location of residue 145 on the HA structure. Viruses possessing naturally occurring substitutions [145N (wt), 145K and 145S] showed the highest avidity to the receptor. With the exception of 145M virus, all viruses carrying alternative substitutions (145A, 145C, 145F, 145G, 145H, 145I, 145L, 145P, 145Q, 145R, 145T, 145V and 145Y) displayed

decreased receptor binding avidity. In contrast to other reports, no compensatory substitutions were identified on the HA or NA by next-generation sequencing (54, 55). This indicates changes in receptor avidity that are directly related to single amino acid substitutions at residue 145. It is important to emphasize that although viruses carrying alternative substitutions exhibited decreased receptor binding avidity, the lower threshold for biologically relevant avidity is unknown. Glycan array analysis revealed a broad range of SAα2-6Gal glycan interactions that were modulated by substitutions at residue 145. Interestingly, one of the viruses with the lowest avidity in the panel (145F virus) displayed extensive binding to SAα2-6Gal glycans, including extended glycans (30). While the expanded glycan array provides important insights on the binding preference, which receptors are relevant for IAV attachment, replication, and transmission *in vivo* remain to be elucidated. Furthermore, the density, distribution and organization of glycans on host tissues are poorly defined.

As observed in human H3N2 IAVs, the emergence of the HA N145K substitution in swine H3N2 IAVs led to a change in the antibody immunodominance. It is astonishing that reactivity to sera raised against a virus possessing HA 145N was indistinguishable among all of the viruses carrying substitutions at residue 145, but reactivity to sera raised against a virus possessing HA 145K was highly skewed to recognize an epitope bearing HA residue 145K even in the context of a distinctly related H3 HA. Antibody immunodominance may be key for understanding antigenic drift and refers to the immunological phenomenon in which the immune system preferentially mounts a response to complex antigens in a dynamic hierarchical order (12). Immunodominance hierarchy can occur at the level of viruses within multivalent immunogens, proteins

within viruses, antigenic sites within proteins, epitopes within antigenic sites and, as showed in the present report and by others, single amino acid substitutions within epitopes (7, 12, 14, 15, 56, 57).

The complex antigenic composition of IAV remains a challenge for vaccine selection and effective vaccination. We provided a better understanding of the functional effects of amino acid substitutions near the RBS implicated in antigenic drift and the consequences to receptor binding and antigenicity. The impact of these substitutions at residue 145 on the immunogenicity and antibody immunodominance remains to be fully characterized. In light of a recent report revealing carbohydrate dependent activation of B cells (58), the ability to modulate binding to a broad range of SAα2-6Gal glycans by substitutions near the RBS offers a possibility to augment B cell activation by combining non-cognate (carbohydrate dependent) with cognate (antigen specific) interactions that warrants further investigation. While there may be a fitness cost associated with the emergence of such substitutions in nature, the ability to manipulate the amino acid plasticity near the RBS may offer an alternative approach to induce broader protection and, perhaps, potentially elicit receptor mimicry, broad-spectrum neutralizing antibodies (59).

Acknowledgements

This study was supported by a contract (HHSN272201400008C) from the National Institute of Allergy and Infectious Diseases (NIAID) Centers for Influenza Research and Surveillance (CEIRS). JJSS received a short-term training award from the NIAID CEIRS Training Program (HHSN272201400008C). This study was supported in part by

resources and technical expertise from the Georgia Advanced Computing Resource Center, a partnership between the University of Georgia's Office of the Vice President for Research and Office of the Vice President for Information Technology.

References

1. **Nicholls JM, Chan RW, Russell RJ, Air GM, Peiris JS.** 2008. Evolving complexities of influenza virus and its receptors. Trends Microbiol **16:**149-157.
2. **Wilson IA, Skehel JJ, Wiley DC.** 1981. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 A resolution. Nature **289:**366-373.
3. **Skehel JJ, Wiley DC.** 2000. Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. Annu Rev Biochem **69:**531-569.
4. **Matrosovich M, Tuzikov A, Bovin N, Gambaryan A, Klimov A, Castrucci MR, Donatelli I, Kawaoka Y.** 2000. Early alterations of the receptor-binding properties of H1, H2, and H3 avian influenza virus hemagglutinins after their introduction into mammals. J Virol **74:**8502-8512.
5. **Connor RJ, Kawaoka Y, Webster RG, Paulson JC.** 1994. Receptor specificity in human, avian, and equine H2 and H3 influenza virus isolates. Virology **205:**17-23.
6. **Wan H, Perez DR.** 2007. Amino acid 226 in the hemagglutinin of H9N2 influenza viruses determines cell tropism and replication in human airway epithelial cells. J Virol **81:**5181-5191.
7. **Li Y, Bostick DL, Sullivan CB, Myers JL, Griesemer SB, Stgeorge K, Plotkin JB, Hensley SE.** 2013. Single hemagglutinin mutations that alter both antigenicity and receptor binding avidity influence influenza virus antigenic clustering. J Virol **87:**9904-9910.
8. **Hensley SE, Das SR, Bailey AL, Schmidt LM, Hickman HD, Jayaraman A, Viswanathan K, Raman R, Sasisekharan R, Bennink JR, Yewdell JW.** 2009.

Hemagglutinin receptor binding avidity drives influenza A virus antigenic drift. Science **326:**734-736.

9. **Nelson MI, Vincent AL.** 2015. Reverse zoonosis of influenza to swine: new perspectives on the human-animal interface. Trends Microbiol **23:**142-153.

10. **Belongia EA, Simpson MD, King JP, Sundaram ME, Kelley NS, Osterholm MT, McLean HQ.** 2016. Variable influenza vaccine effectiveness by subtype: a systematic review and meta-analysis of test-negative design studies. Lancet Infect Dis **16:**942-951.

11. **Rajao DS, Perez DR.** 2018. Universal Vaccines and Vaccine Platforms to Protect against Influenza Viruses in Humans and Agriculture. Front Microbiol **9:**123.

12. **Altman MO, Angeletti D, Yewdell JW.** 2018. Antibody Immunodominance: The Key to Understanding Influenza Virus Antigenic Drift. Viral Immunol **31:**142-149.

13. **Chambers BS, Parkhouse K, Ross TM, Alby K, Hensley SE.** 2015. Identification of Hemagglutinin Residues Responsible for H3N2 Antigenic Drift during the 2014-2015 Influenza Season. Cell Rep **12:**1-6.

14. **Smith DJ, Lapedes AS, de Jong JC, Bestebroer TM, Rimmelzwaan GF, Osterhaus AD, Fouchier RA.** 2004. Mapping the antigenic and genetic evolution of influenza virus. Science **305:**371-376.

15. **Koel BF, Burke DF, Bestebroer TM, van der Vliet S, Zondag GC, Vervaet G, Skepner E, Lewis NS, Spronken MI, Russell CA, Eropkin MY, Hurt AC, Barr IG, de Jong JC, Rimmelzwaan GF, Osterhaus AD, Fouchier RA, Smith**

DJ. 2013. Substitutions near the receptor binding site determine major antigenic change during influenza virus evolution. Science **342:**976-979.

16. **Abente EJ, Santos J, Lewis NS, Gauger PC, Stratton J, Skepner E, Anderson TK, Rajao DS, Perez DR, Vincent AL.** 2016. The Molecular Determinants of Antibody Recognition and Antigenic Drift in the H3 Hemagglutinin of Swine Influenza A Virus. J Virol **90:**8266-8280.

17. **Pena L, Vincent AL, Ye J, Ciacci-Zanella JR, Angel M, Lorusso A, Gauger PC, Janke BH, Loving CL, Perez DR.** 2011. Modifications in the polymerase genes of a swine-like triple-reassortant influenza virus to generate live attenuated vaccines against 2009 pandemic H1N1 viruses. J Virol **85:**456-469.

18. **Tang Y, Lee CW, Zhang Y, Senne DA, Dearth R, Byrum B, Perez DR, Suarez DL, Saif YM.** 2005. Isolation and characterization of H3N2 influenza A virus from turkeys. Avian Dis **49:**207-213.

19. **Chen H, Ye J, Xu K, Angel M, Shao H, Ferrero A, Sutton T, Perez DR.** 2012. Partial and full PCR-based reverse genetics strategy for influenza viruses. PLoS One **7:**e46378.

20. **Perez DR, Angel M, Gonzalez-Reiche AS, Santos J, Obadan A, Martinez-Sobrido L.** 2017. Plasmid-Based Reverse Genetics of Influenza A Virus. Methods Mol Biol **1602:**251-273.

21. **Reed LJ, Muench H.** 1938. A simple method for estimating fifty percent endpoints. Am J Hyg **27:**493-497.

22. **Mena I, Nelson MI, Quezada-Monroy F, Dutta J, Cortes-Fernandez R, Lara-Puente JH, Castro-Peralta F, Cunha LF, Trovao NS, Lozano-Dubernard B,**

Rambaut A, van Bakel H, Garcia-Sastre A. 2016. Origins of the 2009 H1N1 influenza pandemic in swine in Mexico. Elife **5**.

23. **Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A.** 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics **28:**1647-1649.

24. **Spackman E, Senne DA, Myers TJ, Bulaga LL, Garber LP, Perdue ML, Lohman K, Daum LT, Suarez DL.** 2002. Development of a real-time reverse transcriptase PCR assay for type A influenza virus and the avian H5 and H7 hemagglutinin subtypes. J Clin Microbiol **40:**3256-3260.

25. **Santos JJS, Obadan AO, Garcia SC, Carnaccini S, Kapczynski DR, Pantin-Jackwood M, Suarez DL, Perez DR.** 2017. Short- and long-term protective efficacy against clade 2.3.4.4 H5N2 highly pathogenic avian influenza virus following prime-boost vaccination in turkeys. Vaccine **35:**5637-5643.

26. **World Health Organization.** 2011. Manual for the laboratory diagnosis and virological surveillance of influenza. World Health Organization, Geneva.

27. **Lakdawala SS, Lamirande EW, Suguitan AL, Jr., Wang W, Santos CP, Vogel L, Matsuoka Y, Lindsley WG, Jin H, Subbarao K.** 2011. Eurasian-origin gene segments contribute to the transmissibility, aerosol release, and morphology of the 2009 pandemic H1N1 influenza virus. PLoS Pathog **7:**e1002443.

28. **Matrosovich MN, Gambaryan AS.** 2012. Solid-phase assays of receptor-binding specificity. Methods Mol Biol **865:**71-94.

29. **Stevens J, Chen LM, Carney PJ, Garten R, Foust A, Le J, Pokorny BA, Manojkumar R, Silverman J, Devis R, Rhea K, Xu X, Bucher DJ, Paulson JC, Cox NJ, Klimov A, Donis RO.** 2010. Receptor specificity of influenza A H3N2 viruses isolated in mammalian cells and embryonated chicken eggs. J Virol **84:**8287-8299.

30. **Peng W, de Vries RP, Grant OC, Thompson AJ, McBride R, Tsogtbaatar B, Lee PS, Razi N, Wilson IA, Woods RJ, Paulson JC.** 2017. Recent H3N2 Viruses Have Evolved Specificity for Extended, Branched Human-type Receptors, Conferring Potential for Increased Avidity. Cell Host Microbe **21:**23-34.

31. **Lewis NS, Anderson TK, Kitikoon P, Skepner E, Burke DF, Vincent AL.** 2014. Substitutions near the hemagglutinin receptor-binding site determine the antigenic evolution of influenza A H3N2 viruses in U.S. swine. J Virol **88:**4752-4763.

32. **Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, Pieper U, Sali A.** 2006. Comparative protein structure modeling using Modeller. Curr Protoc Bioinformatics **Chapter 5:**Unit-5 6.

33. **Schrodinger, LLC.** 2015. The PyMOL Molecular Graphics System, Version 2.1.

34. **Zhang Y, Aevermann BD, Anderson TK, Burke DF, Dauphin G, Gu Z, He S, Kumar S, Larsen CN, Lee AJ, Li X, Macken C, Mahaffey C, Pickett BE, Reardon B, Smith T, Stewart L, Suloway C, Sun G, Tong L, Vincent AL,**

Walters B, Zaremba S, Zhao H, Zhou L, Zmasek C, Klem EB, Scheuermann RH. 2017. Influenza Research Database: An integrated bioinformatics resource for influenza virus research. Nucleic Acids Res **45:**D466-D474.

35. **Ito T, Suzuki Y, Mitnaul L, Vines A, Kida H, Kawaoka Y.** 1997. Receptor specificity of influenza A viruses correlates with the agglutination of erythrocytes from different animal species. Virology **227:**493-499.

36. **Medeiros R, Escriou N, Naffakh N, Manuguerra JC, van der Werf S.** 2001. Hemagglutinin residues of recent human A(H3N2) influenza viruses that contribute to the inability to agglutinate chicken erythrocytes. Virology **289:**74-85.

37. **Bradley KC, Galloway SE, Lasanajak Y, Song X, Heimburg-Molinaro J, Yu H, Chen X, Talekar GR, Smith DF, Cummings RD, Steinhauer DA.** 2011. Analysis of influenza virus hemagglutinin receptor binding mutants with limited receptor recognition properties and conditional replication characteristics. J Virol **85:**12387-12398.

38. **Matrosovich MN, Gambaryan AS, Teneberg S, Piskarev VE, Yamnikova SS, Lvov DK, Robertson JS, Karlsson KA.** 1997. Avian influenza A viruses differ from human viruses by recognition of sialyloligosaccharides and gangliosides and by a higher conservation of the HA receptor-binding site. Virology **233:**224-234.

39. **Rogers GN, Paulson JC.** 1983. Receptor determinants of human and animal influenza virus isolates: differences in receptor specificity of the H3 hemagglutinin based on species of origin. Virology **127:**361-373.

40. **Yamada S, Suzuki Y, Suzuki T, Le MQ, Nidom CA, Sakai-Tagawa Y, Muramoto Y, Ito M, Kiso M, Horimoto T, Shinya K, Sawada T, Kiso M, Usui T, Murata T, Lin Y, Hay A, Haire LF, Stevens DJ, Russell RJ, Gamblin SJ, Skehel JJ, Kawaoka Y.** 2006. Haemagglutinin mutations responsible for the binding of H5N1 influenza A viruses to human-type receptors. Nature **444:**378-382.

41. **Gamblin SJ, Haire LF, Russell RJ, Stevens DJ, Xiao B, Ha Y, Vasisht N, Steinhauer DA, Daniels RS, Elliot A, Wiley DC, Skehel JJ.** 2004. The structure and receptor binding properties of the 1918 influenza hemagglutinin. Science **303:**1838-1842.

42. **Rogers GN, Paulson JC, Daniels RS, Skehel JJ, Wilson IA, Wiley DC.** 1983. Single amino acid substitutions in influenza haemagglutinin change receptor binding specificity. Nature **304:**76-78.

43. **Wu NC, Thompson AJ, Xie J, Lin CW, Nycholat CM, Zhu X, Lerner RA, Paulson JC, Wilson IA.** 2018. A complex epistatic network limits the mutational reversibility in the influenza hemagglutinin receptor-binding site. Nat Commun **9:**1264.

44. **Petrova VN, Russell CA.** 2018. The evolution of seasonal influenza viruses. Nat Rev Microbiol **16:**47-60.

45. **Barr IG, Russell C, Besselaar TG, Cox NJ, Daniels RS, Donis R, Engelhardt OG, Grohmann G, Itamura S, Kelso A, McCauley J, Odagiri T, Schultz-Cherry S, Shu Y, Smith D, Tashiro M, Wang D, Webby R, Xu X, Ye Z, Zhang W, Writing Committee of the World Health Organization**

Consultation on Northern Hemisphere Influenza Vaccine Composition f. 2014. WHO recommendations for the viruses used in the 2013-2014 Northern Hemisphere influenza vaccine: Epidemiology, antigenic and genetic characteristics of influenza A(H1N1)pdm09, A(H3N2) and B influenza viruses collected from October 2012 to January 2013. Vaccine **32:**4713-4725.

46. **Klimov AI, Garten R, Russell C, Barr IG, Besselaar TG, Daniels R, Engelhardt OG, Grohmann G, Itamura S, Kelso A, McCauley J, Odagiri T, Smith D, Tashiro M, Xu X, Webby R, Wang D, Ye Z, Yuelong S, Zhang W, Cox N, Writing Committee of the World Health Organization Consultation on Southern Hemisphere Influenza Vaccine Composition f.** 2012. WHO recommendations for the viruses to be used in the 2012 Southern Hemisphere Influenza Vaccine: epidemiology, antigenic and genetic characteristics of influenza A(H1N1)pdm09, A(H3N2) and B influenza viruses collected from February to September 2011. Vaccine **30:**6461-6471.

47. **Govorkova EA, Matrosovich MN, Tuzikov AB, Bovin NV, Gerdil C, Fanget B, Webster RG.** 1999. Selection of receptor-binding variants of human influenza A and B viruses in baby hamster kidney cells. Virology **262:**31-38.

48. **Lin Y, Wharton SA, Whittaker L, Dai M, Ermetal B, Lo J, Pontoriero A, Baumeister E, Daniels RS, McCauley JW.** 2017. The characteristics and antigenic properties of recently emerged subclade 3C.3a and 3C.2a human influenza A(H3N2) viruses passaged in MDCK cells. Influenza Other Respir Viruses **11:**263-274.

49. **Tamura D, Nguyen HT, Sleeman K, Levine M, Mishin VP, Yang H, Guo Z, Okomo-Adhiambo M, Xu X, Stevens J, Gubareva LV.** 2013. Cell culture-selected substitutions in influenza A(H3N2) neuraminidase affect drug susceptibility assessment. Antimicrob Agents Chemother **57:**6141-6146.

50. **Mohr PG, Deng YM, McKimm-Breschkin JL.** 2015. The neuraminidases of MDCK grown human influenza A(H3N2) viruses isolated since 1994 can demonstrate receptor binding. Virol J **12:**67.

51. **Aich U, Beckley N, Shriver Z, Raman R, Viswanathan K, Hobbie S, Sasisekharan R.** 2011. Glycomics-based analysis of chicken red blood cells provides insight into the selectivity of the viral agglutination assay. FEBS J **278:**1699-1712.

52. **Duk M, Krotkiewski H, Stasyk TV, Lutsik-Kordovsky M, Syper D, Lisowska E.** 2000. Isolation and characterization of glycophorin from nucleated (chicken) erythrocytes. Arch Biochem Biophys **375:**111-118.

53. **Zost SJ, Parkhouse K, Gumina ME, Kim K, Diaz Perez S, Wilson PC, Treanor JJ, Sant AJ, Cobey S, Hensley SE.** 2017. Contemporary H3N2 influenza viruses have a glycosylation site that alters binding of antibodies elicited by egg-adapted vaccine strains. Proc Natl Acad Sci U S A **114:**12578-12583.

54. **Das SR, Hensley SE, David A, Schmidt L, Gibbs JS, Puigbo P, Ince WL, Bennink JR, Yewdell JW.** 2011. Fitness costs limit influenza A virus hemagglutinin glycosylation as an immune evasion strategy. Proc Natl Acad Sci U S A **108:**E1417-1422.

55. **Hensley SE, Das SR, Gibbs JS, Bailey AL, Schmidt LM, Bennink JR, Yewdell JW.** 2011. Influenza A virus hemagglutinin antibody escape promotes neuraminidase antigenic variation and drug resistance. PLoS One **6:**e15190.

56. **Huang KY, Rijal P, Schimanski L, Powell TJ, Lin TY, McCauley JW, Daniels RS, Townsend AR.** 2015. Focused antibody response to influenza linked to antigenic drift. J Clin Invest **125:**2631-2645.

57. **Stark SE, Caton AJ.** 1991. Antibodies that are specific for a single amino acid interchange in a protein epitope use structurally distinct variable regions. J Exp Med **174:**613-624.

58. **Villar RF, Patel J, Weaver GC, Kanekiyo M, Wheatley AK, Yassine HM, Costello CE, Chandler KB, McTamney PM, Nabel GJ, McDermott AB, Mascola JR, Carr SA, Lingwood D.** 2016. Reconstituted B cell receptor signaling reveals carbohydrate-dependent mode of activation. Sci Rep **6:**36298.

59. **Ekiert DC, Kashyap AK, Steel J, Rubrum A, Bhabha G, Khayat R, Lee JH, Dillon MA, O'Neil RE, Faynboym AM, Horowitz M, Horowitz L, Ward AB, Palese P, Webby R, Lerner RA, Bhatt RR, Wilson IA.** 2012. Cross-neutralization of influenza A viruses mediated by a single antibody loop. Nature **489:**526-532.

Figure legends

Figure 3.1. Amino acid plasticity at residue 145. (A) Publicly available H3 HA sequences were retrieved and the relative frequency of identified amino acids at residue 145 were calculated for swine, human, avian, canine and equine AIVs. Amino acids present at frequency below 1% are not labeled in the figure. (B) Schematic representation of the HA gene segment of A/turkey/Ohio/313053/2004 (H3N2) depicting the codon corresponding to residue 145 and respective amino acid substitutions introduced by site-directed mutagenesis. (C) HA monomeric structure of A/turkey/Ohio/313053/2004 (H3N2) indicating the location of HA residue 145 (colored in cyan).

Figure 3.2. Substitutions at residue 145 showed no major impact on virus growth. Confluent monolayers of MDCK cells were inoculated with H3 viruses carrying amino acid substitutions at residue 145 at a MOI of 0.01 and incubated at 37°C. At 6, 12, 24, 48, and 72 hpi, tissue culture supernatants from inoculated cells were collected for virus RNA quantification by rRT-PCR and expressed as $\log_{10}$ $TCID_{50}$/ml equivalents. Plotted data represent means ± standard errors (SD).

Figure 3.3. Substitutions at residue 145 led to decreased receptor binding avidity. Turkey red blood cells pretreated with different amounts of neuraminidase from either (A) *Clostridium perfringens* or (B) *Arthrobacter ureafaciens* were mixed with H3 viruses carrying amino acid substitutions at residue 145 to quantify virus agglutination as measure of virus binding avidity. Data are expressed as the maximal amount of

neuraminidase that allowed full agglutination. Plotted data represent means ± standard errors (SD).

Figure 3.4. Substitutions at residue 145 retained binding to SAα2-6Gal. H3 viruses carrying amino acid substitutions at residue 145 were tested for receptor binding specificity with varying concentrations of SAα2-3Gal (α2-3-linked SA) or SAα2-6Gal (α2-6-linked SA). (A) 145N (wt), (B) 145A, (C) 145F, (D) 145G, (E) 145H, (F) 145I, (G) 145K, (H) 145L, (I) 145M, (J) 145P, (K) 145Q, (L) 145R, (M) 145S, (N) 145T, (O) 145V and (P) 145Y viruses. (Q) Human pH1N1 and (R) avian ΔH5N1 were used as binding control to SAα2-6Gal and SAα2-3Gal, respectively. Glycan concentration is expressed as arbitrary units (AU). Plotted data represent means ± standard errors (SD).

Figure 3.5. Substitutions at residue 145 modulated binding to a broad range of SAα2-6Gal glycans. Glycan microarray analysis of H3 viruses carrying amino acid substitutions at residue 145. The array is comprised of non-sialoside control (1-10; Grey), SAα2-3Gal (11-76; Yellow) and SAα2-6Gal (77-128; Green) glycans. Glycans are grouped by structure type: L, linear; O, O-linked; N, N-linked and L^x, sialyl Le^x. (A) 145N (wt), (B) 145A, (C) 145F, (D) 145G, (E) 145H, (F) 145I, (G) 145K, (H) 145L, (I) 145M, (J) 145P, (K) 145Q, (L) 145R, (M) 145S, (N) 145T, (O) 145V and (P) 145Y viruses. Plotted data represent means ± standard errors (SD). RFU, relative fluorescent units.

Figure 3.6. Substitutions at residue 145 modulated sera reactivity. Antibody responses to H3 viruses carrying amino acid substitutions at residue 145 was determined by ELISA using swine antisera generated against (A and D) OH/04, (B and E) NY/11 possessing HA residue 145N or (C and F) IA/14 possessing HA residue 145K. Two sets of sera were tested independently. Plotted data represent means ± standard errors (SD). O.D., optical density.

Figure 3.7. Substitutions at residue 145 impacted HI titers. HI titers were measured against H3 viruses carrying amino acid substitutions at residue 145 using swine antisera generated against OH/04 (cyan), NY/11 possessing HA residue 145N (red) or IA/14 possessing HA residue 145K (light green). (A and B) two sets of sera were tested independently. Sera reactivity of all H3 mutant viruses to the respective swine antisera are depicted as the mean fold change [(log2 HI titer mutant virus)/(log2 HI titer homologous virus)]. Plotted data represent means ± standard errors (SD). Colors are based on the antigenic cluster designation for swine H3N2 IAVs.

Tables and Figures

Table 3.1. Primers used to introduce amino acid substitutions at residue 145

aa at position 145	Forward primer sequence (5'-3')*	Reverse primer sequence (5'-3')*
A	GGAATCTGTTGCTAGTTTCTTTAGTAGATT	TAAAGAAACTAGCAACAGATTCCCTTCT
C	GGAATCTGTTTGTAGTTTCTTTAGTAGATT	TAAAGAAACTACAAACAGATTCCCTTCT
D	GGAATCTGTTGATAGTTTCTTTAGTAGATT	TAAAGAAACTATCAACAGATTCCCTTCT
E	GGAATCTGTTGAAAGTTTCTTTAGTAGATT	TAAAGAAACTTTCAACAGATTCCCTTCT
F	GGAATCTGTTTTTAGTTTCTTTAGTAGATT	TAAAGAAACTAAAAACAGATTCCCTTCT
G	GGAATCTGTTGGAAGTTTCTTTAGTAGATT	TAAAGAAACTTCCAACAGATTCCCTTCT
H	GGAATCTGTTCATAGTTTCTTTAGTAGATT	TAAAGAAACTATGAACAGATTCCCTTCT
I	GGAATCTGTTATAAGTTTCTTTAGTAGATT	TAAAGAAACTTATAACAGATTCCCTTCT
K	GGAATCTGTTAAAAGTTTCTTTAGTAGATT	TAAAGAAACTTTTAACAGATTCCCTTCT
L	GGAATCTGTTCTCAGTTTCTTTAGTAGATT	TAAAGAAACTGAGAACAGATTCCCTTCT
M	GGAATCTGTTATGAGTTTCTTTAGTAGATT	TAAAGAAACTCATAACAGATTCCCTTCT
P	GGAATCTGTTCCAAGTTTCTTTAGTAGATT	TAAAGAAACTTGGAACAGATTCCCTTCT
Q	GGAATCTGTTCAAAGTTTCTTTAGTAGATT	TAAAGAAACTTTGAACAGATTCCCTTCT
R	GGAATCTGTTCGGAGTTTCTTTAGTAGATT	TAAAGAAACTCCGAACAGATTCCCTTCT
S	GGAATCTGTTAGCAGTTTCTTTAGTAGATT	TAAAGAAACTGCTAACAGATTCCCTTCT
T	GGAATCTGTTACAAGTTTCTTTAGTAGATT	TAAAGAAACTTGTAACAGATTCCCTTCT
V	GGAATCTGTTGTCAGTTTCTTTAGTAGATT	TAAAGAAACTGACAACAGATTCCCTTCT
W	GGAATCTGTTTGGAGTTTCTTTAGTAGATT	TAAAGAAACTCCAAACAGATTCCCTTCT
Y	GGAATCTGTTTATAGTTTCTTTAGTAGATT	TAAAGAAACTATAAACAGATTCCCTTCT

*Underlined nucleotides indicate the mutated codon associated to HA residue 145.

Table 3.2. Agglutination of erythrocytes by OH/04 145 single amino acid mutant viruses

aa at position 145	HA titer (HAU)		
	0.5% turkey RBCs	0.5% chicken RBCs	1% horse RBCs
N (wt)	128	256	<2
A	128	64	<2
C	4	<2	<2
F	128	16	8
G	128	32	<2
H	256	128	<2
I	128	64	<2
K	128	128	<2
L	128	128	<2
M	128	128	<2
P	128	64	<2
Q	256	128	<2
R	128	64	<2
S	128	128	<2
T	128	64	<2
V	128	32	<2
Y	128	32	<2

Figure 3.1

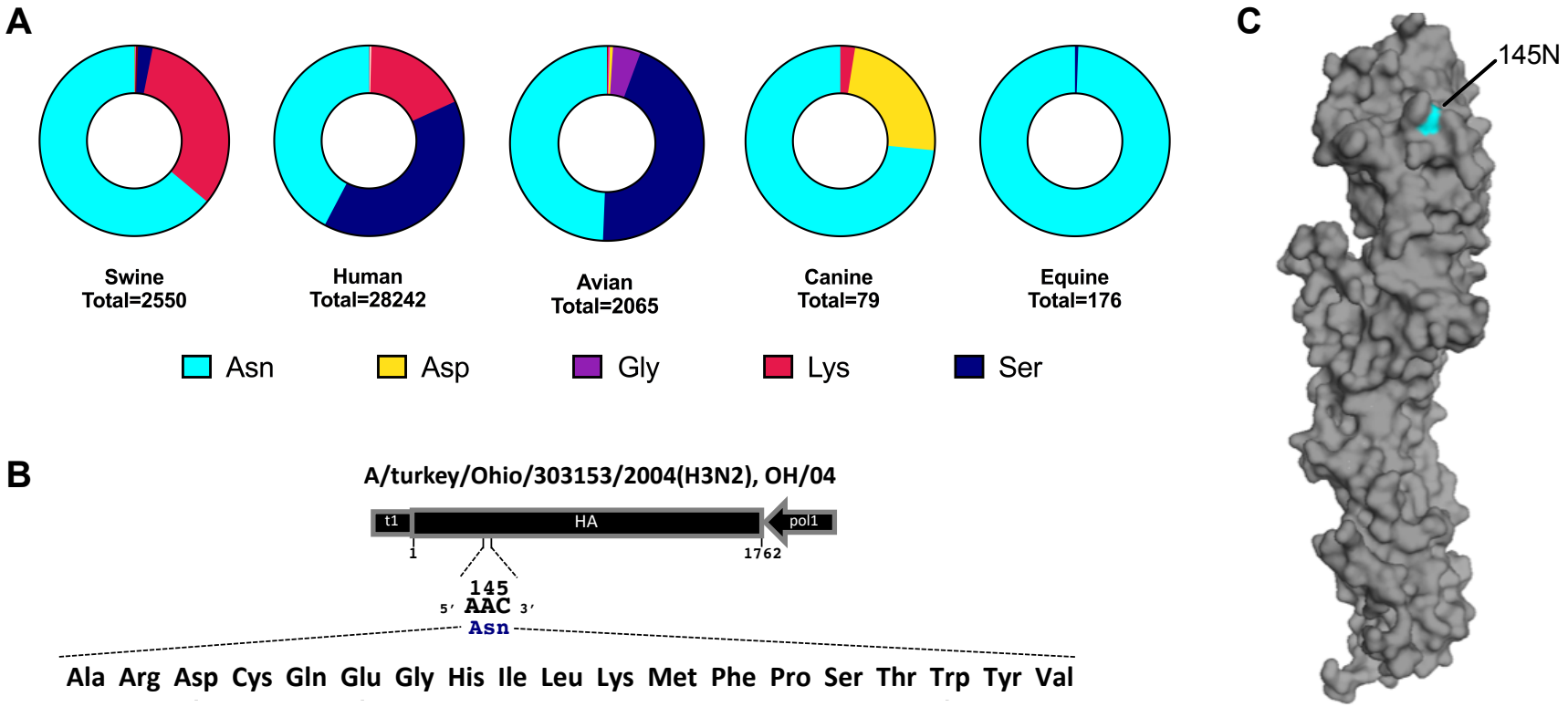

A
Swine
Total=2550
Human
Total=28242
Avian
Total=2065
Canine
Total=79
Equine
Total=176
Asn
Asp
Gly
Lys
Ser
B
A/turkey/Ohio/303153/2004(H3N2), OH/04
t1
HA
pol1
1
1762
145
5′ AAC 3′
Asn
Ala Arg Asp Cys Gln Glu Gly His Ile Leu Lys Met Phe Pro Ser Thr Trp Tyr Val
C
145N

Figure 3.2

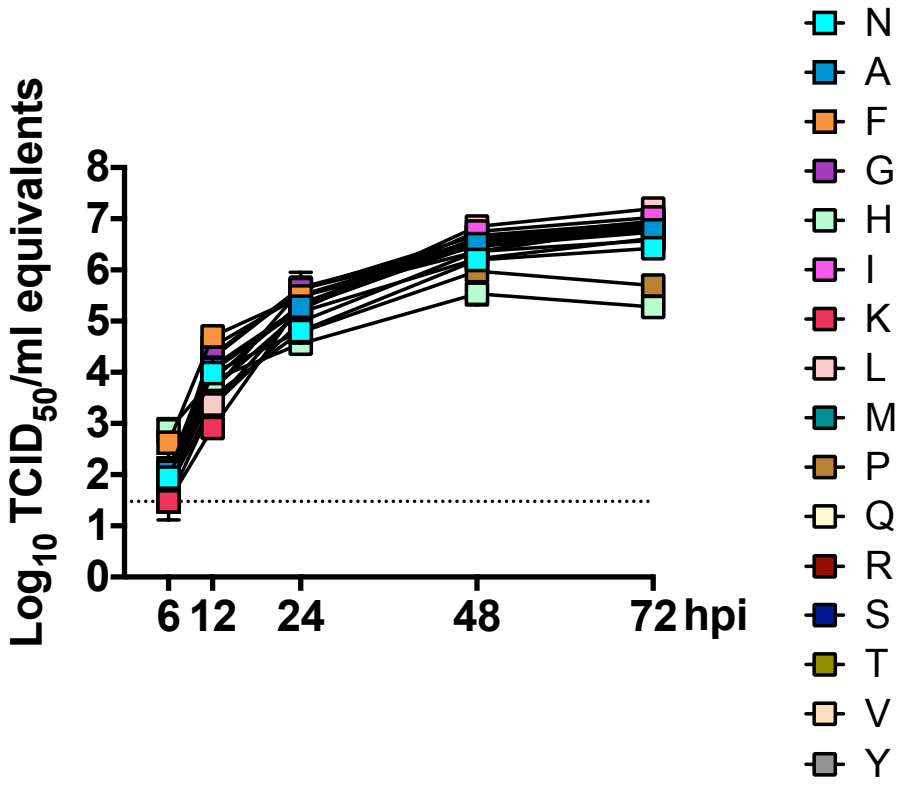

Log10 TCID50/ml equivalents
0
1
2
3
4
5
6
7
8
6
12
24
48
72
hpi
N
A
F
G
H
I
K
L
M
P
Q
R
S
T
V
Y

Figure 3.3

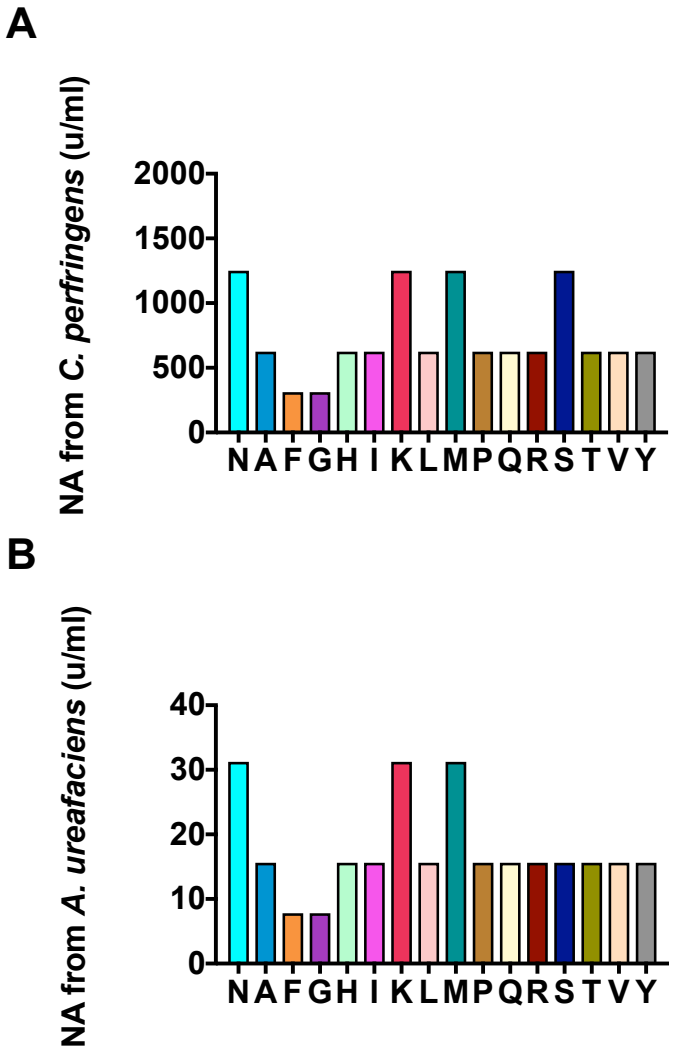

A
NA from C. perfringens (u/ml)
2000
1500
1000
500
0
N A F G H I K L M P Q R S T V Y
B
NA from A. ureafaciens (u/ml)
40
30
20
10
0
N A F G H I K L M P Q R S T V Y

Figure 3.4

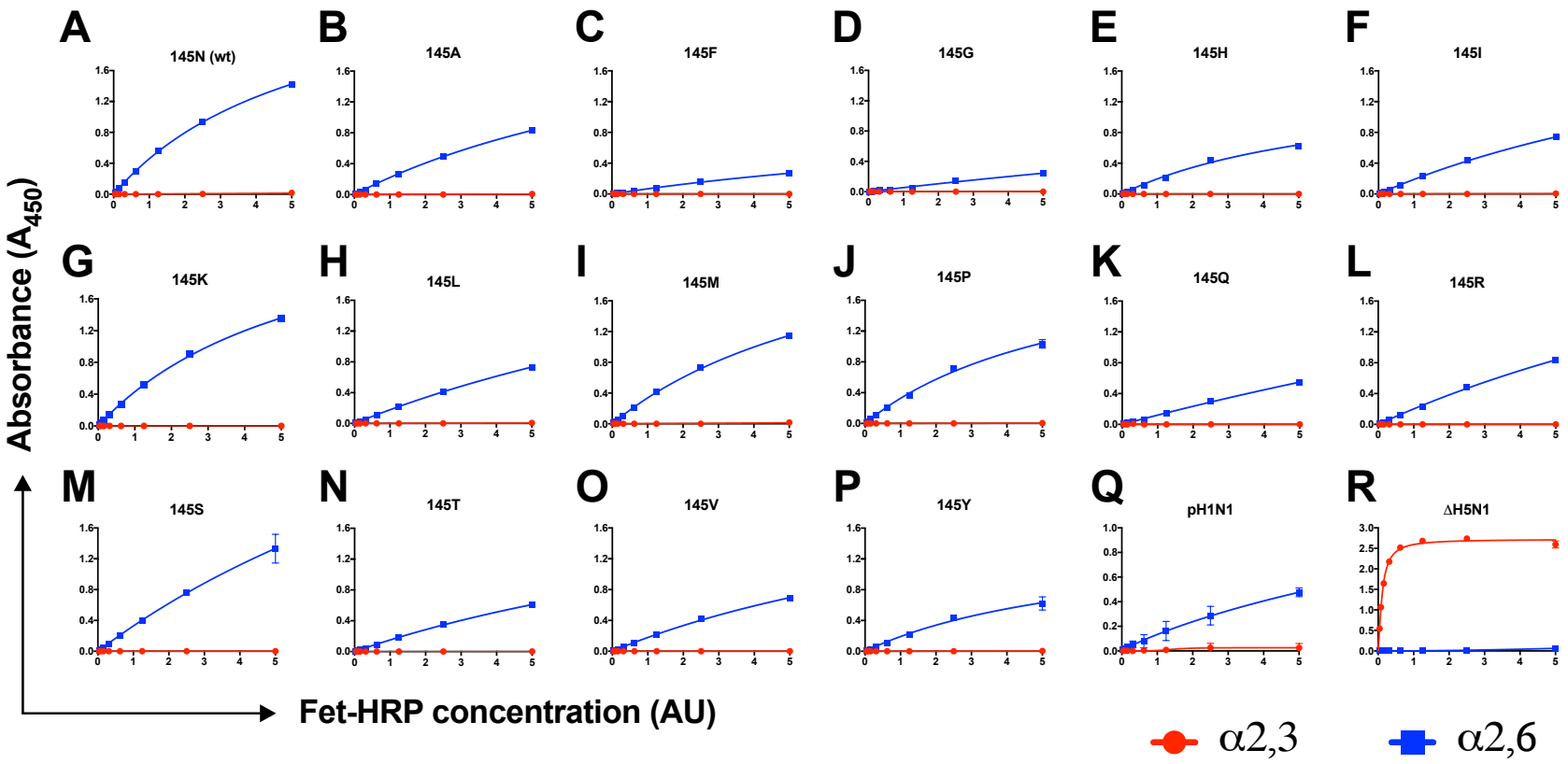

A
145N (wt)
B
145A
C
145F
D
145G
E
145H
F
145I
G
145K
H
145L
I
145M
J
145P
K
145Q
L
145R
M
145S
N
145T
O
145V
P
145Y
Q
pH1N1
R
ΔH5N1
Absorbance (A450)
Fet-HRP concentration (AU)
α2,3
α2,6

Figure 3.5

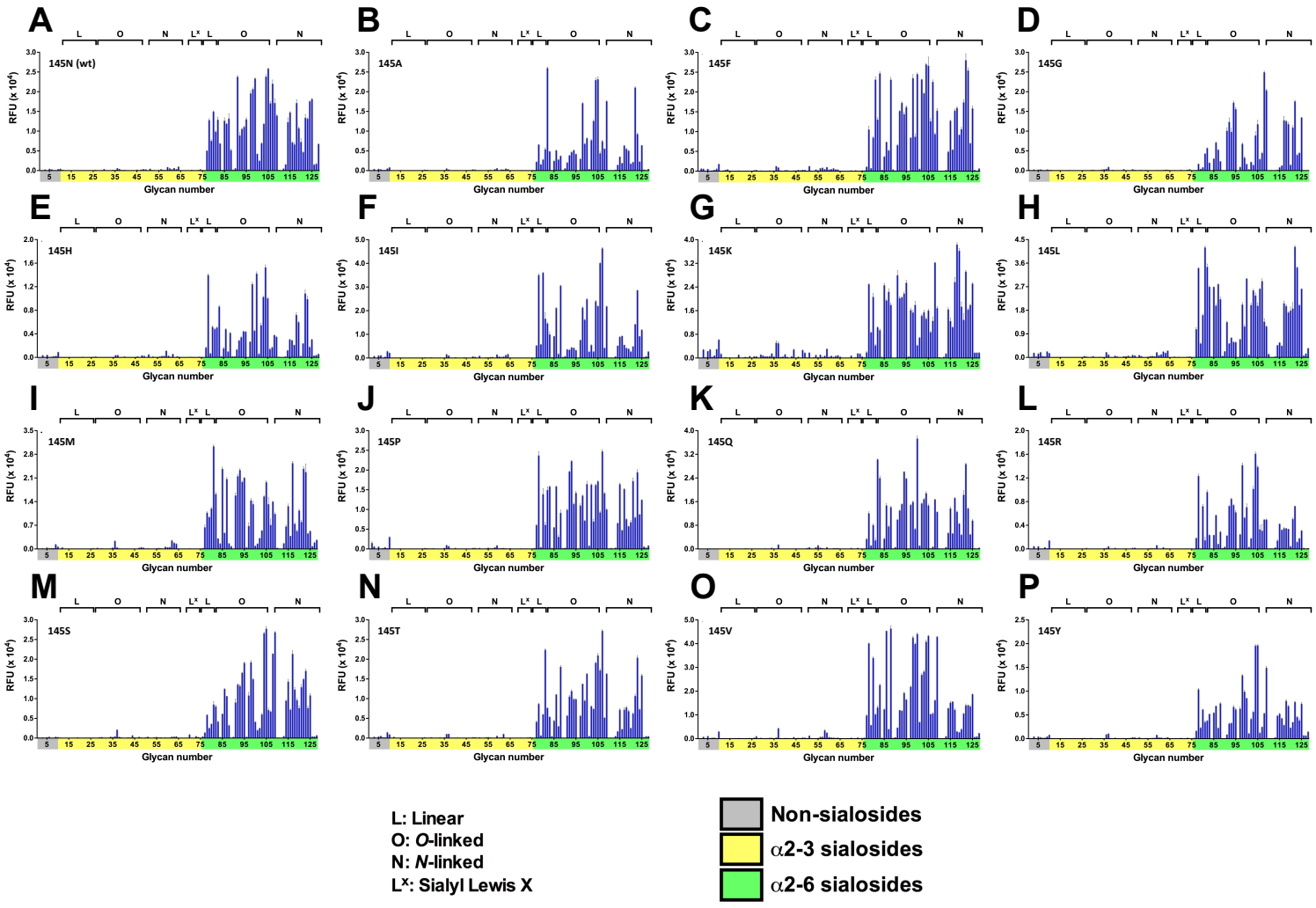

A
145N (wt)
B
145A
C
145F
D
145G
E
145H
F
145I
G
145K
H
145L
I
145M
J
145P
K
145Q
L
145R
M
145S
N
145T
O
145V
P
145Y
L O N Lˣ L O N
RFU (x 10⁴)
Glycan number
L: Linear
O: O-linked
N: N-linked
Lˣ: Sialyl Lewis X
Non-sialosides
α2-3 sialosides
α2-6 sialosides

Figure 3.6

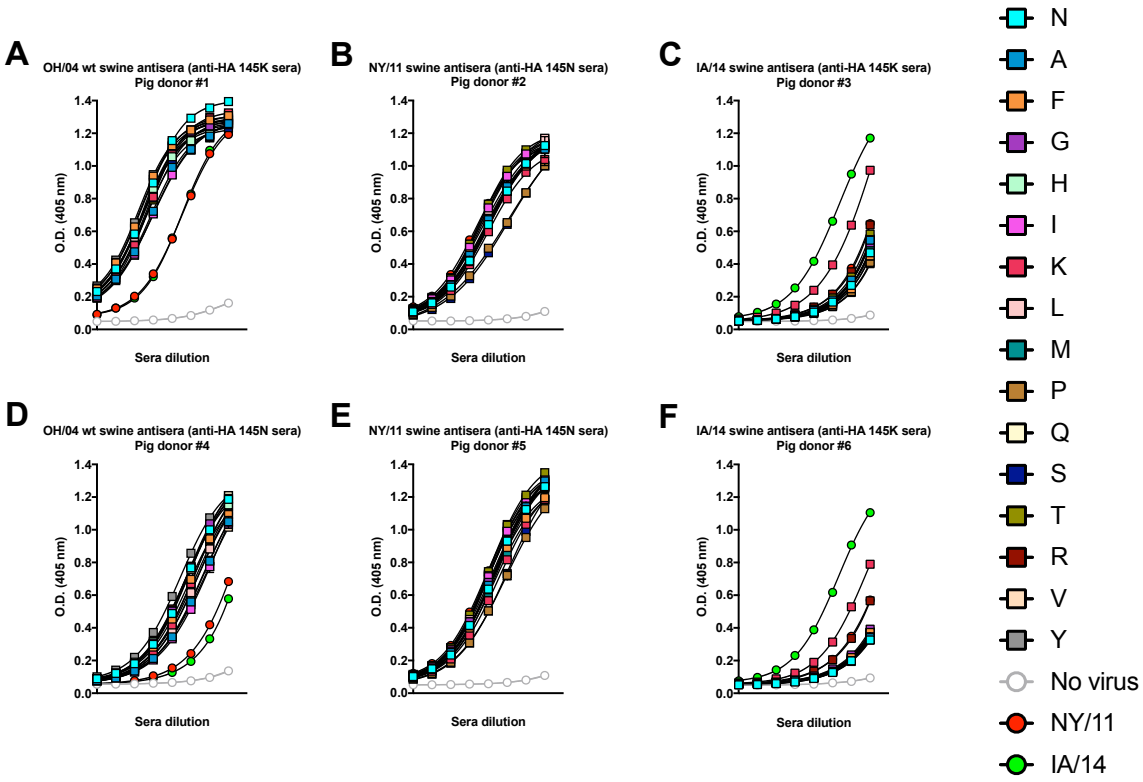
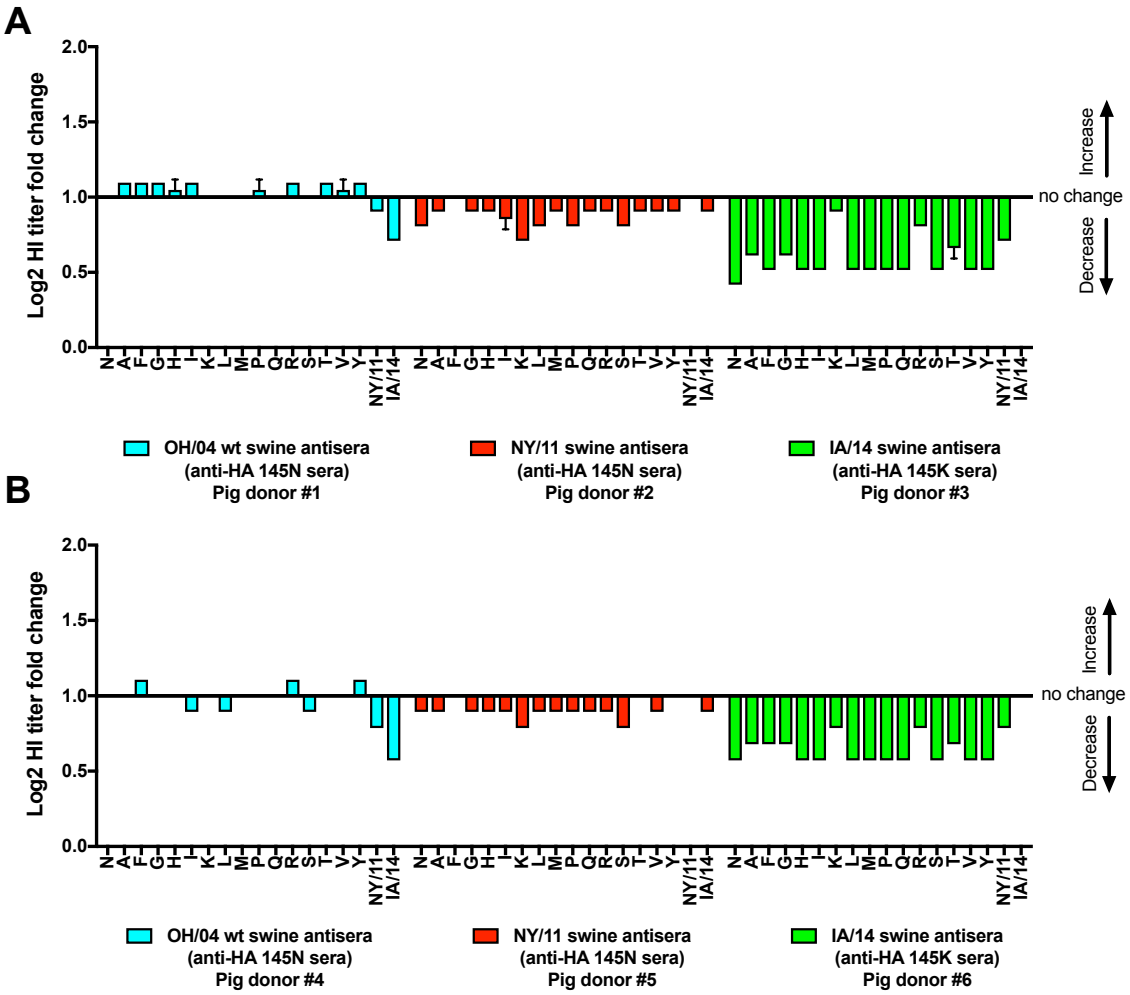

A
OH/04 wt swine antisera (anti-HA 145K sera)
Pig donor #1
B
NY/11 swine antisera (anti-HA 145N sera)
Pig donor #2
C
IA/14 swine antisera (anti-HA 145K sera)
Pig donor #3
D
OH/04 wt swine antisera (anti-HA 145N sera)
Pig donor #4
E
NY/11 swine antisera (anti-HA 145N sera)
Pig donor #5
F
IA/14 swine antisera (anti-HA 145K sera)
Pig donor #6
O.D. (405 nm)
Sera dilution
N
A
F
G
H
I
K
L
M
P
Q
S
T
R
V
Y
No virus
NY/11
IA/14

Figure 3.7


A
Log2 HI titer fold change
2.0
1.5
1.0
0.5
0.0
Increase
no change
Decrease
N A L G H I K L M P Q R S T V Y NY/11 IA/14 N A F G H I K L M P Q R S T V Y NY/11 IA/14 N A F G H I K L M P Q R S T V Y NY/11 IA/14
OH/04 wt swine antisera
(anti-HA 145N sera)
Pig donor #1
NY/11 swine antisera
(anti-HA 145N sera)
Pig donor #2
IA/14 swine antisera
(anti-HA 145K sera)
Pig donor #3
B
Log2 HI titer fold change
2.0
1.5
1.0
0.5
0.0
Increase
no change
Decrease
N A F G H I K L M P Q R S T V Y NY/11 IA/14 N A F G H I K L M P Q R S T V Y NY/11 IA/14 N A F G H I K L M P Q R S T V Y NY/11 IA/14
OH/04 wt swine antisera
(anti-HA 145N sera)
Pig donor #4
NY/11 swine antisera
(anti-HA 145N sera)
Pig donor #5
IA/14 swine antisera
(anti-HA 145K sera)
Pig donor #6

CHAPTER 4

CONSTRUCTION AND CHARACTERIZATION OF AN INFLUENZA A VIRUS H3 HEMAGGLUTININ ANTIGENIC LIBRARY [1]

[1]Jefferson Santos, Adebimpe Obadan, Eugenio Abente, Daniela S. Rajão, Amy L. Vincent and Daniel R. Perez. To be submitted to the Journal of Virology.

Abstract

Vaccination is considered the first line of defense to prevent influenza A virus (IAV) infections. However, IAV antigenic variation remains a challenge for effective vaccination. IAV vaccines primarily work by eliciting neutralizing antibodies against the hemagglutinin (HA) surface protein. Through a mechanism known as antigenic drift, IAV can circumvent the preexisting host immunity by acquiring amino acid substitutions in the HA globular head. Therefore, universal or broadly protective IAV vaccines are greatly needed. Recent work has identified seven amino acid positions (145, 155, 156, 158, 159, 189 and 193, H3 numbering) near the receptor-binding site (RBS) of H3N2 IAVs as the major determinants of antigenic drift. Here we combined deep mutational scanning with reverse genetics to examine the *in vitro* amino acid plasticity of these key antigenically relevant residues in the influenza H3 hemagglutinin. We envisioned a H3 HA antigenic virus library as an alternative IAV vaccine strategy that potentially induces broader protection. Our approach was efficient in producing H3 HA antigenic virus libraries in different donor virus strains. Despite limited diversity in nature, high-throughput targeted next-generation sequencing revealed that these residues exhibited extraordinary amino acid plasticity *in vitro*. Nonetheless, virus library diversity is severely reduced following virus rescue. To validate our experimental approach, a novel H3 HA variant virus was isolated and shown to possess distinct receptor binding and antigenic properties relative to wt H3 HA virus. These findings have important implications for understanding IAV antigenic diversity and aiding the development of an alternative vaccine approach.

Introduction

Antigenic variation represents a critical challenge facing the development of broadly protective vaccines against many pathogens, including influenza A virus (IAV) and human immunodeficiency virus-1 (HIV-1). Commercially available IAV vaccines rely largely on eliciting neutralizing antibodies against the hemagglutinin (HA), the major surface glycoprotein and the primary target of the antibody response following natural IAV exposure. Due to antigenic drift, the natural accumulation of amino acid substitutions in the globular head domain of HA to escape population immunity, IAV vaccines must be periodically updated (1-4). Therefore, IAV vaccine development has shifted in the past few years towards universal vaccine approaches that will require fewer updates and provide longer lasting immunity (5, 6). For the most part, these efforts have largely focused on eliciting immune responses to more conserved components of IAV, such as the HA stem, the extracellular domain of M2 (M2e) or internal proteins (3).

Located in a small depression on the apex of the HA globular head, the receptor binding site (RBS) mediates virus binding to sialylated glycan receptors on the host cell surface (7). The RBS is relatively conserved on an otherwise hypervariable HA globular head, but it has been overlooked as a possible target for developing broadly protective vaccines. Despite its small footprint, a number of receptor mimicking, broadly neutralizing antibodies targeting the RBS have been recently identified (8, 9). Nevertheless, IAV have evolved mechanisms of evasion to protect this vulnerable site. Consistently, several reports have demonstrated the significance of a small set of amino acid substitutions near the RBS in modulating antigenic drift and immune evasion for avian H5N1 (10), human H3N2 (11), swine H3N2 (12, 13), equine H3N8 (14), and

human H1N1 IAVs (11, 15). In the case of human and swine H3N2 IAVs, amino acid substitutions in up to seven residues (145, 155, 156, 158, 159, 189 and 193, H3 numbering) adjacent to the RBS were responsible for the major antigenic changes observed in these viruses (11, 12).

Elucidating the relationship between amino acid substitutions and antigenic drift has important implications for understanding IAV antigenic evolution and informing vaccine selection (10-12, 15, 16). Nevertheless, these substitutions only represent a small fraction of the sequence space sampled by the natural evolutionary process. For example, we previously generated a panel of HA mutant viruses carrying single amino acid substitutions, representing all 20 possible amino acids, in a key residue on the HA globular head. We showed that most amino acid substitutions not commonly found in nature were well tolerated *in vitro* but possessed important functional consequences to receptor binding and antigenicity (Chapter 3). Deep mutational scanning is a new high-throughput experimental methodology that combines saturation mutagenesis with next-generation sequencing to probe tolerance or functional consequences of a large number of amino acid substitutions in variants of a protein on a massive scale (17). Deep mutational scanning has many important applications (17), including understanding protein evolution (18), and has been extensively used in IAV research in recent years (19-21). Disrupting antigenic variation (22) near the RBS by applying deep mutational scanning to target antigenically relevant residues may be useful in generating more effective and broadly protective vaccines.

In the present report, we aimed to systematically explore the *in vitro* amino acid plasticity and functional sequence space of key antigenically relevant residues (145, 155,

156, 158, 159, 189 and 193) in the influenza H3 hemagglutinin by coupling saturation mutagenesis with next-generation sequencing. Despite limited diversity in nature, our results revealed that these residues exhibited extraordinary amino acid plasticity *in vitro*. Our approach was robust and efficient in producing H3 HA antigenic virus libraries in different donor virus strains. Yet, virus library diversity is drastically reduced following virus rescue, indicating potential bottlenecks. Interestingly, an isolated virus variant not previously found in nature was further characterized and shown to possess distinct receptor binding and antigenic properties relative to wt HA. These findings have important implications for understanding IAV antigenic diversity and developing an alternative IAV vaccine approach that induces broader protection.

Materials and Methods

Cells

Madin-Darby canine kidney (MDCK) and human embryonic kidney 293T cells were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal bovine serum (FBS). Cells were propagated at 37˚C in a humidified incubator under 5% CO_2 atmosphere.

Construction of a hemagglutinin antigenic library and virus rescue

A/turkey/Ohio/313053/2004 (H3N2) or OH/04 wt is a prototypic swine origin virus amenable to genetic manipulation by a stablished reverse genetics (RG) system (12, 23, 24). To generate an HA library targeting antigenically relevant positions in the influenza H3 hemagglutinin, a DNA fragment (~450 nt) was synthesized spanning a portion of the OH/04 wt HA coding sequences, including the HA residues 145 through 193 (Genscript,

Piscataway, NJ). Antigenically relevant amino acid positions 145, 155, 156, 158, 159, 189 and 193 encoded a degenerate (NNK) codon while the other positions retained the corresponding wild-type nucleotide sequence. A series of overlapping PCR reactions were carried out in order to reconstitute a full-length RG-ready PCR-based HA segment. Briefly, the fragment library described above was amplified with primers 5'-AAC TGT TAC CCT TAT TAT GT-3' and 5'-TAT GTC TCC CGG TTT TAC TAT T-3'. The remaining portions of the HA were amplified from OH/04 wt HA plasmid DNA. To prevent the carryover of OH/04 wt HA plasmid DNA during PCR amplification, the cloned OH/04 wt HA segment was split into two overlapping plasmids: pDP-SD1 and pDP-SD2. The pDP-SD1 plasmid carries the mouse RNA polymerase I terminator (t1) followed by nucleotides 1-522 of OH/04 wt HA segment. The pDP-SD2 plasmid contains nucleotides 500-1762 of OH/04 wt HA segment followed by the human RNA polymerase I promoter (polI). A second fragment containing t1 followed by the initial portion of the HA was amplified with primers 5'-ACC GGA GTA CTG GTC GAC CTC CGA AGT TGG GGG GGA GCA AAA GCA GG-3' and 5'-ACA TAA TAA GGG TAA CAG TT-3'. A final fragment comprising of the last portion of HA followed by polI was amplified from pDP-SD2 using primers 5'-AAT AGT AAA ACC GGG AGA CAT A-3' and 5'-ATG CTG ACA ACG TCC CCG GCC CGG CGC TGC T-3'. All three resulting PCR fragments were purified by gel extraction using QIAquick Gel Extraction Kit (Qiagen, Valencia, CA) and combined to produce a full-length RG-ready PCR-based HA segment carrying degenerate codons at antigenically relevant amino acid positions by overlapping PCR (25). All PCR reactions were performed with Phusion High Fidelity DNA polymerase (New England Biolabs, Ipswich, MA) and confirmed to be devoid of

unwanted mutations by sequencing. Virus libraries were rescued by PCR-based reverse genetics using a co-culture of 293T/MDCK cells as previously described (25, 26). The RG-ready PCR-based HA segment was paired with the six plasmids representing the internal gene segments of OH/04 wt, the live attenuated OH/04 att (23) or A/Puerto Rico/8/34 (H1N1), herein referred as PR/8. All virus libraries were rescued using the OH/04 wt NA plasmid. Following transfection, cells were incubated at 35˚C. After 24 h incubation, media was replaced with Opti-MEM I (Life Technologies, Carlsbad, CA) containing 1 μg/mL TPCK-trypsin (Worthington Biochemicals, Lakewood, NJ) and 1% antibiotics/antimycotic solution (Sigma-Aldrich, St. Louis, MO). Following rescue of virus libraries, virus stocks were amplified once in MDCK cells before sequencing.

Virus library sequencing preparation

Virus RNA from tissue culture supernatant virus stocks was extracted using the RNeasy mini kit (Qiagen, Valencia, CA). Isolated virus RNA was treated with DNase using TURBO DNA-free kit (Thermo Fisher Scientific, Waltham, MA) and then used as template in a one-step reverse transcriptase PCR (RT-PCR) reaction with SuperScript III and Platinum Taq High Fidelity DNA Polymerase system (Thermo Fisher Scientific, Waltham, MA). Synthesized fragment and RG-ready PCR-based HA segment were amplified with NEBNext High-Fidelity 2X PCR Master Mix (New England Biolabs, Ipswich, MA). All libraries were amplified with primers 5'-ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TCT GGA ACA AGC TAT GCT TGC AGA-3' and 5'-AGA CGT GTG CTC TTC CGA TCT CTG TAA CTC TCC CTG ATG CTT-3'. These primers contain part of the adaptor sequence required for Illumina sequencing. A second PCR reaction was carried out to add the rest of the adaptor and index (for multiplex

sequencing) sequences to the amplicons using primers 5'-AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG AT-3' and 5'-CAA GCA GAA GAC GGC ATA CGA GAT XXX XXX GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T-3'. Positions annotated by an "X" indicate the nucleotides for the index sequences. The second PCR was performed using NEBNext High-Fidelity 2X PCR Master Mix (New England Biolabs, Ipswich, MA). For virus RNA template, the cycling conditions were set as follows: 2 min at 55˚C followed by 60 min at 42˚C for reverse transcription; then 2 min at 94˚C followed by 25 three-step cycles of 30 s at 94˚C for denaturation, 30 s at 56˚C for primer annealing, and 30 s at 68˚C for extension, and a 10 min final extension at 68˚C. For DNA template, the thermocycler was set as follows: 30 s at 98˚C, then 15 three-step cycles of 10 s at 98˚C for denaturation, 30 s at 56˚C for primer annealing, and 30 s at 72˚C for extension, and a 2 min final extension at 72˚C. In the second PCR reaction, the conditions were the same as above (for DNA template), except the cycle number was 8. All PCR products were purified by gel extraction using QIAquick Gel Extraction Kit (Qiagen, Valencia, CA). Barcoded libraries were multiplexed and sequenced on the high-throughput Illumina MiSeq sequencing platform in a paired-end 250 nt run format.

Virus library sequencing data analysis

Raw sequencing data was processed using different bioinformatic tools from BBTools package v35.82. FASTAQ files containing forward and reverse reads were combined using Reformat and paired-end reads merged with BBMerge. Paired-end reads were filtered out if corresponding forward and reverse reads did not perfectly overlap. Passing filter reads were quality-trimmed using BBDuk and then trimmed with cutadapt v1.14

(27) to produce read of same length. Trimmed reads were mapped to OH/04 wt HA reference sequences using BBMap. Mapped reads were imported into R for data analysis. Amino acid identities at positions corresponding to HA residues 145, 155, 156, 158, 159, 189 and 193 were retrieved from individual reads. The frequencies of different amino acids at residues of interest or antigenic motifs (the combination of HA residues 145, 155, 156, 158, 159, 189 and 193) were computed.

Isolation of H3 HA variant viruses

Individual variants from virus libraries were isolated by limiting dilution as described elsewhere (28). Briefly, confluent monolayers of MDCK cells in 96-well plates were infected with ten-fold serial dilutions of virus stocks from virus libraries. Following 72 h incubation at 35˚C, supernatant from wells inoculated with the highest dilution that produced extensive signs of cytopathic effect (CPE) were harvested for a second and final round of limiting dilution. After isolation, virus stocks for individual variants were amplified in MDCK cells. Virus stocks were titrated by tissue culture infectious dose 50 ($TCID_{50}$) and virus titers were determined by the Reed and Muench method (29).

Whole-genome sequencing of H3 HA variant viruses

Following virus isolation by limiting dilution, virus RNA from tissue culture supernatant virus stocks was purified using the RNeasy mini kit (Qiagen, Valencia, CA) or MagNA Pure LC RNA Isolation Kit (Roche Life Science, Mannheim, Germany). Isolated virus RNA served as template in a one-step reverse transcriptase PCR (RT-PCR) reaction for multi-segment, whole genome amplification (30). Amplicon sequence libraries were prepared as previously described (30) or using Nextera XT DNA Library Prep Kit (Illumina, San Diego, CA) according to the manufacturer's protocol. Barcoded libraries

were multiplexed and sequenced on the high-throughput Illumina MiSeq sequencing platform in a paired-end 150 nt run format. De novo genome assembly was performed as described previously (30) and HA and NA-specific reads were mapped to OH/04 reference sequences using Geneious 10.1.3 (31).

Solid-phase assay of receptor binding specificity

The receptor-binding specificity was determined in a solid phase direct binding assay using either monospecific preparations of peroxidase (HRP)-conjugated fetuin (fet-HRP) or sialyglycopolymers (32-35). Monospecific preparations of fet-HRP were synthesized using α2-3-sialyltransferase from *Pasteurella multocida* (Sigma, St. Louis, MO) for 3-modified fetuin (3-fet-HRP) or α2-6-sialyltransferase from *Photobacterium damsela* (Sigma, St. Louis, MO) for 6-modified fetuin (6-fet-HRP), essentially as described previously (32). For the assay based on monospecific preparations of fet-HRP, 96-well native fetuin-coated flat-bottom plates (Greiner Bio-One, Monroe, NC) were incubated overnight at 4ºC with 128 HAU of each virus in 0.02 M tris-buffered saline (TBS), pH 7.2–7.4. Virus samples were run in duplicate. Plates were washed three times with PBS and blocked with blocking solution [BS, PBS containing 0.1% neuraminidase-treated bovine serum albumin (BSA-NA)] for 2 h at room temperature. After blocking, plates were washed twice with ice-cold washing solution (WS, PBS containing 0.02% Tween 80) and incubated with two-fold serial dilution of 3-Fet-HRP (SAα2-3Gal) or 6-Fet-HRP (SAα2-6Gal) in reaction solution (RS, PBS containing 0.02% Tween-80, 0.1% BSA-NA and 2 μM oseltamivir carboxylate) for 1 h at 4ºC. For the assay using sialyglycopolymers, 96-well fetuin-coated flat-bottom plates (Greiner Bio-One, Monroe, NC) were incubated overnight at 4ºC with 32 HAU of each virus in TBS buffer. Plates

were blocked with BS buffer for 2 h at room temperature and then washed twice with ice-cold WS buffer. Plates were incubated with two-fold serial dilution of 3'SLN (SAα2-3Gal) or 6'SLN (SAα2-6Gal) high molecular weight biotinylated sialyglycopolymer (Glycotek, Gaithersburg MD) in RS at 4ºC for 1 h. Subsequently, plates were washed five times with ice-cold WS buffer and incubated with a 1:1000 dilution of Streptavidin-HRP (Thermo Fisher Scientific, Waltham, MA) for 1 h at 4ºC. After the last incubation, plates from both assays were washed five times with ice-cold WS buffer before adding freshly prepared substrate solution (SS, 0.01% 3,3',5,5'-tetramethylbenzidine in 0.05 M sodium acetate with 0.03% H_20_2). Reactions were allowed to proceed at room temperature for 30 min unless otherwise stated. Reactions were stopped with 3% (vol/vol in ddH_2O) H_2SO_4. Absorbance readings were obtained at 450 nm using a Victor x3 Multilabel Plate Reader (PerkinElmer, Waltham, MA).

Swine antisera and antibodies

A/turkey/Ohio/313053/2004 (H3N2) was generated in a previous study (12). Two pigs were primed and boosted intramuscularly with UV-inactivated whole virus vaccine combined with commercial adjuvant. The pigs were humanely euthanized for blood collection. F49 (Takara Bio USA, Mountain View, CA) monoclonal antibody (mAb) binds to a conserved region of the H3 stalk and was used as a control. Goat anti-swine IgG-HRP (Seracare, Milford, MA) or goat anti-mouse IgG-HRP (Jackson ImmunoResearch, West Grove, PA) polyclonal antibodies were used as secondary antibodies.

Enzyme-linked immunosorbent assay (ELISA)

For ELISA, 96-well flat-bottom plates (Greiner Bio-One, Monroe, NC) were incubated overnight at 4ºC with 16 HAU of each virus in 1X coating solution (Seracare, Milford, MA). Virus samples were run in duplicate. Plates were blocked with StartingBlock (PBS) blocking buffer (Thermo Fisher Scientific, Waltham, MA) for 1 h at room temperature. After blocking, plates were washed three times with PBS containing 0.05% tween 20 (PBS-T). Two-fold serial dilutions of swine antisera or a monoclonal antibody were added and allowed to incubate for 1 h at room temperature. After incubation, plates were washed three times with PBS-T before adding a secondary, HRP-conjugated polyclonal antibody. Plates were incubated at room temperature for 1 h. After incubation, plates were washed three times with PBS-T before adding freshly prepared SS buffer or ABTS 1-component microwell peroxidase substrate (Seracare, Milford, MA) for 1 h at room temperature. The reaction was stopped by adding 3% (vol/vol in ddH_2O) H_2SO_4 or ABTS peroxidase stop solution (Seracare, Milford, MA), respectively. Absorbance readings were obtained at 405 nm using a Victor x3 Multilabel Plate Reader (PerkinElmer, Waltham, MA).

Structure modeling

A model of the structure of the HA of A/turkey/Ohio/313053/2004 (H3N2) was built by homology modeling using Modeller v9.16 (36) based upon the crystal structure of multiple H3 HA proteins [Protein Data Bank (PDB) codes 2YP7, 1HA0, 2YP2, 4WE8 and 4WE5). Generated model was subsequently rendered with PyMOL (37).

Computational analysis of HA sequences

The frequency distribution of antigenic motifs (the amino acid pattern of HA residues 145, 155, 156, 158, 159, 189 and 193) was analyzed by downloading all HA amino acid sequences from swine influenza A viruses of the H3 subtype publicly available in the Influenza Research Database (IRD) (38) as of June 10, 2017. After manual inspection, sequences that were less than full length or were otherwise anomalous were removed from the dataset. Sequences were aligned with MUSCLE (39) as integrated in the IRD database (38), converted from FASTA to tabular format using Galaxy (40) and imported into R for data analysis. Amino acid identities at HA residues 145, 155, 156, 158, 159, 189 and 193 were retrieved from individual sequences and combined together. The frequencies of different antigenic motifs were computed and plotted by year (from 2007 to 2017). Antigenic motifs represented by ≤ 5 sequences were categorized as "Other" in the plot.

Results

Amino acid diversity in antigenically relevant residues near the RBS of H3 HA

The recent identification of up to 7 residues (145, 155, 156, 158, 159, 189 and 193) near the RBS as major drivers of antigenic evolution in human and swine H3N2 IAVs (11-13) offers the possibility to explore antigenic diversity by looking at the amino acid pattern at these residues, previously referred to as antigenic motifs, in natural virus isolates (12). Publicly available HA amino acid sequences from swine IAVs of the H3 subtype were analyzed and the amino acid identities at HA residues 145, 155, 156, 158, 159, 189 and 193 were retrieved from individual viruses. While there are limitations on

using genetic information (HA sequence data) alone to infer antigenic diversity (41), our simplified approach provided a glimpse of the antigenic complexity in swine H3 IAVs over time (**Fig. 4.1A**). The emergence, expansion, decline and extinction of natural isolates with particular antigenic motifs were readily discernable (**Fig. 4.1A**). For example, the KTHNFKS antigenic motif represents the human-like H3 viruses that emerged in the swine population in 2012 and have been expanding thereafter (42). As not all substitutions are created equal, some motifs may not represent true antigenic motifs. Instead, they may be antigenically related and reflect amino acid sequence variation of particular antigenic groups that cannot be resolved by HA sequence data alone. Although there is an unquestionable level of antigenic motif diversification, amino acid diversity is rather limited at the level of individual residues (**Fig. 4.1A**).

Designing an H3 HA antigenic library targeting key residues near the RBS

To explore the amino acid plasticity at antigenically relevant residues in the influenza H3 hemagglutinin, a DNA fragment (~450 nt), herein referred to as "synthesized library", was chemically synthesized. Positions corresponding to residues 145, 155, 156, 158, 159, 189 and 193 were engineered to encode a degenerate (NNK) codon while the remaining positions on the HA retained the expected wild-type nucleotide sequence (**Fig. 4.1B**). A series of overlapping PCR reactions were carried out to reconstitute independently full-length RG-ready PCR-based HA segments. Theoretically, our deep mutational scanning strategy is expected to yield up to ~20^7 unique putative antigenic motifs, of which the large majority does not exist in nature. Mutations were introduced into the HA of OH/04 wt virus that naturally possesses the NHNNYRN antigenic motif (**Fig. 4.1C**). Diversity of the synthesized library and the

derived full-length RG-ready PCR-based HA segment was examined by targeted next generation sequencing. Each library was sequenced at ~$1x10^5$ reads. At both the nucleotide and amino acid levels, libraries demonstrated high diversity.

Relative to OH/04 wt HA reference sequence, the percentage of nucleotide variation for the synthesized library was 64.9-84.0% at codon position 1, 67.6-83.6% at codon position 2 and 41.7-99.9% at codon position 3 across the 7 degenerate codons (**Fig. 4.2A**). The percentage of nucleotide variation after PCR amplification of full-length RG-ready PCR-based HA segments was 65.1-84.1% at codon position 1, 67.4-83.8% at codon position 2 and 42.4-99.9% at codon position 3 (**Fig. 4.2A**). Analysis of wt HA by targeted next generation sequencing revealed no significant nucleotide variation (**Fig. 4.2A**). While all 4 nucleotides were represented across the 7 degenerate codons, the nucleotide frequency distribution per site diverged in the synthesized library (**Fig. 4.2B**). Nonetheless, PCR amplification did not alter the nucleotide frequency compared to the synthesized library (**Fig. 4.2B**). Based on the lack of nucleotide variation, analysis of wt HA revealed no changes in the nucleotide frequency distribution (**Fig. 4.2B**).

In agreement with previous results, the codon frequency distribution at antigenically relevant residues was indistinguishable between the synthesized library (**Fig. 4.3**) and after PCR amplification (**Fig. 4.4**). Codons specifying all possible amino acids as well as a stop codon were represented. Overall, codon frequency distribution was consistent with the number of codons determining a specific amino acid. However, there were sites in which some amino acids were overrepresented and/or underrepresented (**Fig. 4.3 and 4.4**). Taken together, these results suggest our deep mutational scanning approach generated high diversity at both the nucleotide and amino acid levels. While it

is not possible to control nucleotide bias in the synthesized library, sequence diversity was maintained after PCR amplification of full-length RG-ready PCR-based HA segments.

Efficient rescue of virus libraries from an H3 HA antigenic library targeting key residues near the RBS

To evaluate whether virus libraries can be recovered by reverse genetics using an H3 HA antigenic library, we attempted virus rescue of the full-length RG-ready PCR-based HA segment in the backbone of OH/04 wt (twice independently, $_{Ex01}$OH/04 wt and $_{Ex02}$OH/04 wt), the live attenuated OH/04 att ($_{Ex02}$OH/04 att) or PR/8 (twice independently, $_{Ex01}$PR/8 and $_{Ex02}$PR/8). Diversity of antigenic motifs was examined by targeted next generation sequencing (**Fig. 4.5A and B**). Each virus library was sequenced at ~$1x10^5$ reads and compared with synthesized library and input full-length RG-ready PCR-based HA segment (FL-PCR1 or FL-PCR2). Relative to wt HA, a high level of diversity in the number of unique antigenic motifs was observed in synthesized library (~ 67,709), FL-PCR1 (~ 68,969), and FL-PCR2 (~ 65,655) (**Fig. 4.5A and B**). However, virus library diversity is drastically reduced following virus rescue regardless of the virus backbone utilized (**Fig. 4.5A and B**), suggesting a potential bottleneck.

Despite the fact that the wt HA antigenic motif (NHNNYRN) is the most prevalent motif in either synthesized library or full-length RG-ready PCR-based HA segment, viruses with unique and very distinct antigenic motifs were rapidly selected *in vitro* (**Fig. 4.5B, Tables 4.1 and 4.2**). Although not surprisingly considering the massive number of possible antigenic motifs (~20^7), completely distinct antigenic motifs emerged from independent experiments in the same backbone (**Table 4.1 and 4.2**). The most

prevalent antigenic motifs were RHAESWG at 37.64% of ($_{Ex01}$OH/04 wt), QHKSGLY at 10.17% ($_{Ex02}$OH/04 wt), RFSSGGG at 94.75% ($_{Ex02}$OH/04 att), NVREFAS at 32.94% ($_{Ex01}$PR/8) and STKSLTK at 36.40% ($_{Ex02}$PR/8). Within the most representative antigenic motifs for each virus library, many antigenic motifs shared similarities in amino acid sequences at multiple sites. For instance, in the $_{Ex01}$OH/04 wt, RHAESWG, RHAESVS, RHAESKN, RHAESAG and RHAESFR all shared the RHAES amino acid motif for the first 5 residues (**Table 4.1**). Similarly, in the $_{Ex02}$OH/04wt, QHKSGLY, QHKSGVR, QHKSGIR and QHKSGRR all shared the QHKSG amino acid motif for the first 5 residues (**Table 4.2**). Interestingly, viruses with the most prevalent antigenic motifs emerged after rescue from rare, low-frequency variants (**Table 4.3**).

Collectively, these results indicate an H3 HA antigenic library lead to efficient virus rescue in different virus backbones. Although diversity of antigenic motifs is severely reduced following virus rescue, viruses with unique antigenic motifs were rapidly selected *in vitro*.

Substitutions in key residues near the RBS impact receptor binding and immune escape

To validate our experimental approach, we attempted to isolate individual virus variants with unique antigenic motifs from one of the virus libraries ($_{Ex01}$OH/04 wt) by limiting dilution. Out of 24 clones isolated and sequenced, all of them possessed the antigenic motif RHAESWG variant. These results are not completely unexpected considering the high prevalence of this specific variant in the $_{Ex01}$OH/04 wt virus library (**Table 4.1 and Fig. 4.5A**). The antigenic motif RHAESWG variant differed from the wt HA antigenic motif (NHNNYRN) in all, but one residue (155H). Next generation sequencing analysis revealed these substitutions are remarkably stable with 99.5-100% of

sequenced reads possessing the selected codon at designated HA residues (145, 156, 158, 159, 189 and 193) even after further amplification in tissue culture. No compensatory substitutions were detected on HA or the neuraminidase (NA).

As these substitutions are located near the RBS, it is important to evaluate their impact on receptor binding specificity and/or avidity. Initially, receptor binding was examined on an ELISA-based assay using monospecific preparation of Fet-HRP for SAα2-3Gal or SAα2-6Gal glycan (**Fig. 4.6A-D, top**) (32). Viruses used as controls displayed the expected receptor binding specificity. Human pH1N1 demonstrated preference to SAα2-6Gal while avian ΔH5N1 displayed exclusive binding to SAα2-3Gal (**Fig. 4.6A and B, top**). As expected, the binding profile of OH/04 wt virus (NHNNYRN) showed restricted binding to SAα2-6Gal (**Chapter 3 and Fig. 4.6C, top**). Surprisingly, RHAESWG variant showed no binding to either SAα2-6Gal or SAα2-3Gal (**Fig. 4.6D, top**).

In order to determine whether RHAESWG lack of binding in the Fet-HRP assay reflects decreased receptor avidity, we carried out a modified version of the ELISA-based assay using biotinylated sialyglycopolymers (**Fig. 4.6A-D, bottom**). These compounds are synthetic high molecular weight biotinylated probes with higher density of monospecific sialylated glycans capable of amplifying the signal for low avidity viruses (32, 34, 35). Consistent with the Fet-HRP assay, human pH1N1 and avian ΔH5N1 showed the expected binding preference (**Fig. 4.6A and B, bottom**). However, human pH1N1 and avian ΔH5N1 also displayed weak binding to SAα2-3Gal and SAα2-6Gal, respectively (**Fig. 4.6A and B, bottom**). Whether these observations are biologically relevant or reflect background noise of the assay remain to be determined. OH/04 wt

virus (NHNNYRN) showed strong binding to SAα2-6Gal although residual binding to SAα2-3Gal was evident (**Fig. 4.6C, bottom**). In this assay, we demonstrated RHAESWG variant binding exclusively to SAα2-6Gal but with diminished avidity compared to OH/04 wt virus (NHNNYRN) (**Fig. 4.6D, bottom**).

The accumulation of amino acid substitutions near the RBS can lead to significant antigenic change (11-13, 41). To examine the impact of substitutions in RHAESWG variant on immune escape, antibody binding was evaluated by whole virus ELISA (**Fig. 4.7A and B**). Binding to a monoclonal antibody (F49), targeting a conserved epitope on the HA stalk domain, was almost indistinguishable between RHAESWG variant and OH/04 wt virus (NHNNYRN) (**Fig. 4.7A**). This indicates the amount of virus used in these assays was equivalent, as was determined by HA assay. RHAESWG variant consistently showed decreased reactivity to swine sera raised against OH/04 wt virus (NHNNYRN) (**Fig. 4.7B**), indicating immune escape. Collectively, these results suggest that virus variants with unique antigenic motifs isolated from virus libraries, such as RHAESWG variant, are highly genetic stable in vitro. RHAESWG variant retained binding to SAα2-6Gal but exhibited reduced receptor binding avidity. Additionally, substitutions present in RHAESWG variant led to immune escape.

Discussion

Here we applied deep mutational scanning and reverse genetics to examine the *in vitro* amino acid plasticity in the H3 hemagglutinin. Instead of carrying out saturation mutagenesis in specific protein domains or in a whole protein (18, 20), our approach is unique in the way that we employ deep mutational scanning to target key residues (145,

155, 156, 158, 159, 189 and 193) on the HA globular head, previously identified as major molecular determinants of antigenic drift in H3N2 IAVs. Moreover, our next-generation sequencing strategy enabled us to explore not only the amino acid plasticity at the individual amino acid level (19, 20), but also the combined amino acid pattern of mutated residues (antigenic motifs) (12). Recognizing the importance of these 7 residues does not imply that amino acid changes outside the antigenic motifs are irrelevant. As we and others have pointed out, some of these substitutions may represent neutral hitchhikers or compensatory mutations necessary to retain protein function, such as receptor binding avidity and thermodynamic stability (11, 12, 41, 43). With new advances in high-throughput sequencing technologies and synthetic biology, our approach can be tailored to reduce mutation potential of specific residues to encode only amino acid substitutions previously found in nature, incorporate additional relevant residues or target entire antigenic sites on the HA globular head. While we used a swine origin HA segment, our strategy can be readily customized to HA segments from different host origins.

Targeted next-generation sequencing analysis of the synthesized library revealed high diversity at the nucleotide and amino acid levels, and consequently, at the antigenic motif level. More importantly, PCR amplification of the synthesized library to generate full-length RG-ready PCR-based HA segments did not reduce or alter library diversity. Considering the size of the synthesized library (~20^7 antigenic motifs), it is impractical to sequence every possible antigenic motif. With individual libraries sequenced at ~$1x10^5$ reads before virus rescue, around 67,000 unique antigenic motifs that appeared at least once were identified. This indicates most antigenic motifs are represented by a single read. While a stringent workflow was used for data cleanup and analysis, future

experiments should aim at greatly increasing sequencing depth to improve representative sampling of antigenic motifs in the library.

Although our goal was to generate highly diverse virus libraries of antigenic motifs, there was compelling evidence of bottleneck following virus rescue: (i) severe reduction in library diversity, as previously reported (20, 44, 45), (ii) emergence of viruses with strikingly distinct antigenic motifs from replicate experiments, and (iii) top antigenic motifs, within individual libraries, with similarities in sequence at multiple residues. This genetic bottleneck is in part explained by the segmented nature of the IAV genome, which consists of 8 negative-sense RNA segments (23, 25, 26, 44). When generating IAV by reverse genetics, only a fraction of cells likely produce most of the initial virus variants. In addition, many of the H3 HA variants generated by mutagenesis in the synthesized library could be nonfunctional. The establishment of cells lines expressing surface HA and NA proteins may mitigate the impact of bottleneck. This should also improve the replicate-to-replicate consistency (44). Furthermore, future experiment should focus on evaluating the temporal dynamics of virus libraries following rescue to identify optimal time-points to harvest virus libraries at their peak of antigenic motif diversity. Validating our experimental approach, we successfully isolated a variant virus possessing a very distinctive antigenic motif (RHAESWG) relative to the wt HA virus (NHNNYRN). These changes led to reduced receptor binding avidity and immune escape. Moreover, the RHAESWG variant showed exceptional genetic stability. The isolation of additional variant viruses may offer important insights on the impact of unnatural amino acid substitutions near RBS to receptor binding and antigenic properties.

While beyond the scope of the present report, we envisioned this strategy as an alternative IAV vaccine approach that potentially induces broader protection. A previous report showed that the disruption of antigenic variation in an antigenically variable parasite was critical to generate a broadly protective vaccine (22). A conceptually similar approach has emerged as a promising strategy for a global HIV-1 vaccine (46-49). It is fundamental to determine whether a H3 HA antigenic virus library can augment both the breadth and depth without compromising the magnitude of antigen-specific antibody responses as compared with wt HA virus antigens in animal models. This will require the evaluation of the best vaccine platform (live or inactivated immunization), route of administration, prime-boost regime, vaccine dose spacing, impact of pre-existing immunity and so on. As our approach targets key residues near the RBS, the ability of inducing receptor-mimicry broadly neutralizing antibodies (bnAbs) need to be investigated. This new class of antibodies recognizes the RBS and showed increased breadth of recognition against antigenic drifted virus strains. The maturation pathway to the development of receptor-mimicry bnAbs or vaccination strategies for inducing receptor-mimicry bnAbs remain poorly understood.

Acknowledgements

This study was supported by a contract (HHSN272201400008C) from the National Institute of Allergy and Infectious Diseases (NIAID) Centers for Influenza Research and Surveillance (CEIRS). JJSS received a short-term training award from the NIAID CEIRS Training Program (HHSN272201400008C). This study was supported in part by resources and technical expertise from the Georgia Advanced Computing Resource

Center, a partnership between the University of Georgia's Office of the Vice President for Research and Office of the Vice President for Information Technology.

References

1. **Altman MO, Angeletti D, Yewdell JW.** 2018. Antibody Immunodominance: The Key to Understanding Influenza Virus Antigenic Drift. Viral Immunol **31:**142-149.
2. **Belongia EA, Simpson MD, King JP, Sundaram ME, Kelley NS, Osterholm MT, McLean HQ.** 2016. Variable influenza vaccine effectiveness by subtype: a systematic review and meta-analysis of test-negative design studies. Lancet Infect Dis **16:**942-951.
3. **Rajao DS, Perez DR.** 2018. Universal Vaccines and Vaccine Platforms to Protect against Influenza Viruses in Humans and Agriculture. Front Microbiol **9:**123.
4. **Grohskopf LA, Sokolow LZ, Broder KR, Walter EB, Bresee JS, Fry AM, Jernigan DB.** 2017. Prevention and Control of Seasonal Influenza with Vaccines: Recommendations of the Advisory Committee on Immunization Practices - United States, 2017-18 Influenza Season. MMWR Recomm Rep **66:**1-20.
5. **Erbelding EJ, Post D, Stemmy E, Roberts PC, Augustine AD, Ferguson S, Paules CI, Graham BS, Fauci AS.** 2018. A Universal Influenza Vaccine: The Strategic Plan for the National Institute of Allergy and Infectious Diseases. J Infect Dis doi:10.1093/infdis/jiy103.
6. **Paules CI, Marston HD, Eisinger RW, Baltimore D, Fauci AS.** 2017. The Pathway to a Universal Influenza Vaccine. Immunity **47:**599-603.
7. **Wiley DC, Skehel JJ.** 1987. The structure and function of the hemagglutinin membrane glycoprotein of influenza virus. Annu Rev Biochem **56:**365-394.

8. **Schmidt AG, Therkelsen MD, Stewart S, Kepler TB, Liao HX, Moody MA, Haynes BF, Harrison SC.** 2015. Viral receptor-binding site antibodies with diverse germline origins. Cell **161:**1026-1034.

9. **Laursen NS, Wilson IA.** 2013. Broadly neutralizing antibodies against influenza viruses. Antiviral Res **98:**476-483.

10. **Koel BF, van der Vliet S, Burke DF, Bestebroer TM, Bharoto EE, Yasa IW, Herliana I, Laksono BM, Xu K, Skepner E, Russell CA, Rimmelzwaan GF, Perez DR, Osterhaus AD, Smith DJ, Prajitno TY, Fouchier RA.** 2014. Antigenic variation of clade 2.1 H5N1 virus is determined by a few amino acid substitutions immediately adjacent to the receptor binding site. MBio **5:**e01070-01014.

11. **Koel BF, Burke DF, Bestebroer TM, van der Vliet S, Zondag GC, Vervaet G, Skepner E, Lewis NS, Spronken MI, Russell CA, Eropkin MY, Hurt AC, Barr IG, de Jong JC, Rimmelzwaan GF, Osterhaus AD, Fouchier RA, Smith DJ.** 2013. Substitutions near the receptor binding site determine major antigenic change during influenza virus evolution. Science **342:**976-979.

12. **Abente EJ, Santos J, Lewis NS, Gauger PC, Stratton J, Skepner E, Anderson TK, Rajao DS, Perez DR, Vincent AL.** 2016. The Molecular Determinants of Antibody Recognition and Antigenic Drift in the H3 Hemagglutinin of Swine Influenza A Virus. J Virol **90:**8266-8280.

13. **Lewis NS, Anderson TK, Kitikoon P, Skepner E, Burke DF, Vincent AL.** 2014. Substitutions near the hemagglutinin receptor-binding site determine the

antigenic evolution of influenza A H3N2 viruses in U.S. swine. J Virol **88:**4752-4763.

14. **Lewis NS, Daly JM, Russell CA, Horton DL, Skepner E, Bryant NA, Burke DF, Rash AS, Wood JL, Chambers TM, Fouchier RA, Mumford JA, Elton DM, Smith DJ.** 2011. Antigenic and genetic evolution of equine influenza A (H3N8) virus from 1968 to 2007. J Virol **85:**12742-12749.

15. **Koel BF, Mogling R, Chutinimitkul S, Fraaij PL, Burke DF, van der Vliet S, de Wit E, Bestebroer TM, Rimmelzwaan GF, Osterhaus AD, Smith DJ, Fouchier RA, de Graaf M.** 2015. Identification of amino acid substitutions supporting antigenic change of influenza A(H1N1)pdm09 viruses. J Virol **89:**3763-3775.

16. **Li C, Hatta M, Burke DF, Ping J, Zhang Y, Ozawa M, Taft AS, Das SC, Hanson AP, Song J, Imai M, Wilker PR, Watanabe T, Watanabe S, Ito M, Iwatsuki-Horimoto K, Russell CA, James SL, Skepner E, Maher EA, Neumann G, Klimov AI, Kelso A, McCauley J, Wang D, Shu Y, Odagiri T, Tashiro M, Xu X, Wentworth DE, Katz JM, Cox NJ, Smith DJ, Kawaoka Y.** 2016. Selection of antigenically advanced variants of seasonal influenza viruses. Nat Microbiol **1:**16058.

17. **Fowler DM, Fields S.** 2014. Deep mutational scanning: a new style of protein science. Nat Methods **11:**801-807.

18. **Starr TN, Picton LK, Thornton JW.** 2017. Alternative evolutionary histories in the sequence space of an ancient protein. Nature **549:**409-413.

19. **Doud MB, Hensley SE, Bloom JD.** 2017. Complete mapping of viral escape from neutralizing antibodies. PLoS Pathog **13:**e1006271.

20. **Thyagarajan B, Bloom JD.** 2014. The inherent mutational tolerance and antigenic evolvability of influenza hemagglutinin. Elife **3**.

21. **Wu NC, Xie J, Zheng T, Nycholat CM, Grande G, Paulson JC, Lerner RA, Wilson IA.** 2017. Diversity of Functionally Permissive Sequences in the Receptor-Binding Site of Influenza Hemagglutinin. Cell Host Microbe **21:**742-753 e748.

22. **Rivero FD, Saura A, Prucca CG, Carranza PG, Torri A, Lujan HD.** 2010. Disruption of antigenic variation is crucial for effective parasite vaccine. Nat Med **16:**551-557, 551p following 557.

23. **Pena L, Vincent AL, Ye J, Ciacci-Zanella JR, Angel M, Lorusso A, Gauger PC, Janke BH, Loving CL, Perez DR.** 2011. Modifications in the polymerase genes of a swine-like triple-reassortant influenza virus to generate live attenuated vaccines against 2009 pandemic H1N1 viruses. J Virol **85:**456-469.

24. **Tang Y, Lee CW, Zhang Y, Senne DA, Dearth R, Byrum B, Perez DR, Suarez DL, Saif YM.** 2005. Isolation and characterization of H3N2 influenza A virus from turkeys. Avian Dis **49:**207-213.

25. **Chen H, Ye J, Xu K, Angel M, Shao H, Ferrero A, Sutton T, Perez DR.** 2012. Partial and full PCR-based reverse genetics strategy for influenza viruses. PLoS One **7:**e46378.

26. **Perez DR, Angel M, Gonzalez-Reiche AS, Santos J, Obadan A, Martinez-Sobrido L.** 2017. Plasmid-Based Reverse Genetics of Influenza A Virus. Methods Mol Biol **1602:**251-273.

27. **Martin M.** 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal **17**.

28. **Kimble JB, Angel M, Wan H, Sutton TC, Finch C, Perez DR.** 2014. Alternative reassortment events leading to transmissible H9N1 influenza viruses in the ferret model. J Virol **88:**66-71.

29. **Reed LJ, Muench H.** 1938. A simple method for estimating fifty percent endpoints. Am J Hyg **27:**493-497.

30. **Mena I, Nelson MI, Quezada-Monroy F, Dutta J, Cortes-Fernandez R, Lara-Puente JH, Castro-Peralta F, Cunha LF, Trovao NS, Lozano-Dubernard B, Rambaut A, van Bakel H, Garcia-Sastre A.** 2016. Origins of the 2009 H1N1 influenza pandemic in swine in Mexico. Elife **5**.

31. **Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A.** 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics **28:**1647-1649.

32. **Matrosovich MN, Gambaryan AS.** 2012. Solid-phase assays of receptor-binding specificity. Methods Mol Biol **865:**71-94.

33. **Matrosovich M, Matrosovich T, Carr J, Roberts NA, Klenk HD.** 2003. Overexpression of the alpha-2,6-sialyltransferase in MDCK cells increases influenza virus sensitivity to neuraminidase inhibitors. J Virol **77:**8418-8425.

34. **Matrosovich M, Tuzikov A, Bovin N, Gambaryan A, Klimov A, Castrucci MR, Donatelli I, Kawaoka Y.** 2000. Early alterations of the receptor-binding properties of H1, H2, and H3 avian influenza virus hemagglutinins after their introduction into mammals. J Virol **74:**8502-8512.

35. **Gambaryan AS, Karasin AI, Tuzikov AB, Chinarev AA, Pazynina GV, Bovin NV, Matrosovich MN, Olsen CW, Klimov AI.** 2005. Receptor-binding properties of swine influenza viruses isolated and propagated in MDCK cells. Virus Res **114:**15-22.

36. **Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen MY, Pieper U, Sali A.** 2006. Comparative protein structure modeling using Modeller. Curr Protoc Bioinformatics **Chapter 5:**Unit-5 6.

37. **Schrodinger, LLC.** 2015. The PyMOL Molecular Graphics System, Version 2.1.

38. **Zhang Y, Aevermann BD, Anderson TK, Burke DF, Dauphin G, Gu Z, He S, Kumar S, Larsen CN, Lee AJ, Li X, Macken C, Mahaffey C, Pickett BE, Reardon B, Smith T, Stewart L, Suloway C, Sun G, Tong L, Vincent AL, Walters B, Zaremba S, Zhao H, Zhou L, Zmasek C, Klem EB, Scheuermann RH.** 2017. Influenza Research Database: An integrated bioinformatics resource for influenza virus research. Nucleic Acids Res **45:**D466-D474.

39. **Edgar RC.** 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res **32:**1792-1797.

40. **Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Cech M, Chilton J, Clements D, Coraor N, Eberhard C, Gruning B, Guerler A, Hillman-Jackson J, Von Kuster G, Rasche E, Soranzo N, Turaga N, Taylor J, Nekrutenko A, Goecks J.** 2016. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res **44:**W3-W10.

41. **Smith DJ, Lapedes AS, de Jong JC, Bestebroer TM, Rimmelzwaan GF, Osterhaus AD, Fouchier RA.** 2004. Mapping the antigenic and genetic evolution of influenza virus. Science **305:**371-376.

42. **Rajao DS, Gauger PC, Anderson TK, Lewis NS, Abente EJ, Killian ML, Perez DR, Sutton TC, Zhang J, Vincent AL.** 2015. Novel Reassortant Human-Like H3N2 and H3N1 Influenza A Viruses Detected in Pigs Are Virulent and Antigenically Distinct from Swine Viruses Endemic to the United States. J Virol **89:**11213-11222.

43. **Hensley SE, Das SR, Bailey AL, Schmidt LM, Hickman HD, Jayaraman A, Viswanathan K, Raman R, Sasisekharan R, Bennink JR, Yewdell JW.** 2009. Hemagglutinin receptor binding avidity drives influenza A virus antigenic drift. Science **326:**734-736.

44. **Doud MB, Bloom JD.** 2016. Accurate Measurement of the Effects of All Amino-Acid Mutations on Influenza Hemagglutinin. Viruses **8**.

45. **Varble A, Albrecht RA, Backes S, Crumiller M, Bouvier NM, Sachs D, Garcia-Sastre A, tenOever BR.** 2014. Influenza A virus transmission

bottlenecks are defined by infection route and recipient host. Cell Host Microbe **16:**691-700.

46. **Seaman MS, Xu L, Beaudry K, Martin KL, Beddall MH, Miura A, Sambor A, Chakrabarti BK, Huang Y, Bailer R, Koup RA, Mascola JR, Nabel GJ, Letvin NL.** 2005. Multiclade human immunodeficiency virus type 1 envelope immunogens elicit broad cellular and humoral immunity in rhesus monkeys. J Virol **79:**2956-2963.

47. **Fischer W, Perkins S, Theiler J, Bhattacharya T, Yusim K, Funkhouser R, Kuiken C, Haynes B, Letvin NL, Walker BD, Hahn BH, Korber BT.** 2007. Polyvalent vaccines for optimal coverage of potential T-cell epitopes in global HIV-1 variants. Nat Med **13:**100-106.

48. **Barouch DH, Stephenson KE, Borducchi EN, Smith K, Stanley K, McNally AG, Liu J, Abbink P, Maxfield LF, Seaman MS, Dugast AS, Alter G, Ferguson M, Li W, Earl PL, Moss B, Giorgi EE, Szinger JJ, Eller LA, Billings EA, Rao M, Tovanabutra S, Sanders-Buell E, Weijtens M, Pau MG, Schuitemaker H, Robb ML, Kim JH, Korber BT, Michael NL.** 2013. Protective efficacy of a global HIV-1 mosaic vaccine against heterologous SHIV challenges in rhesus monkeys. Cell **155:**531-539.

49. **Barouch DH, O'Brien KL, Simmons NL, King SL, Abbink P, Maxfield LF, Sun YH, La Porte A, Riggs AM, Lynch DM, Clark SL, Backus K, Perry JR, Seaman MS, Carville A, Mansfield KG, Szinger JJ, Fischer W, Muldoon M, Korber B.** 2010. Mosaic HIV-1 vaccines expand the breadth and depth of cellular immune responses in rhesus monkeys. Nat Med **16:**319-323.

Figure legends

Figure 4.1. Amino acid plasticity in key antigenically relevant residues near the receptor-binding site (RBS) of H3 hemagglutinin. (A) Amino acid identities at residues 145, 155, 156, 158, 159, 189 and 193 were retrieved from individual swine H3 HA sequences and combined to create an antigenic motif. The frequencies of different antigenic motifs were computed and plotted by year. Antigenic motifs are listed by frequency in descending order. Antigenic motifs represented by ≤ 5 sequences were categorized as "Other" in the plot. (B) Schematic representation of the HA gene segment of A/turkey/Ohio/313053/2004 (H3N2) depicting the codons corresponding to residues 145, 155, 156, 158, 159, 189 and 193. These positions were submitted to saturation mutagenesis, producing degenerate (NNK) codons that encode all possible amino acids at these residues. (C) HA monomeric structure of A/turkey/Ohio/313053/2004 (H3N2) indicating the location of HA residue 145, 155, 156, 158, 159, 189, and 193 (colored in cyan).

Figure 4.2. Nucleotide diversity in an H3 HA antigenic library. An HA library targeting antigenically relevant residues (145, 155, 156, 158, 159, 189 and 193) in the influenza H3 hemagglutinin was chemically synthesized. Positions corresponding to these key residues encoded a degenerate (NNK) codon while the remaining positions on the HA retained the corresponding wild-type nucleotide sequence. (A) Nucleotide (nt) variation and (B) frequency distribution in the synthesized library and after library PCR amplification were examined by targeted next generation sequencing. Sequencing

analysis for wt HA was included as control for comparison. A, adenine; C, cytosine; G, guanine and T, thymine.

Figure 4.3. Codon frequency distribution in an H3 HA antigenic library. An HA library targeting antigenically relevant residues (145, 155, 156, 158, 159, 189 and 193) in the influenza H3 hemagglutinin was chemically synthesized. Positions corresponding to these key residues encoded a degenerate (NNK) codon while the remaining positions on the HA retained the corresponding wild-type nucleotide sequence. Codon frequency distribution at the specified residues was analyzed by targeted next generation sequencing for the synthesized library. *, stop codon; A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan and Y, tyrosine.

Figure 4.4. Maintenance of codon frequency distribution in an H3 HA antigenic library after PCR amplification. An HA library targeting antigenically relevant residues (145, 155, 156, 158, 159, 189 and 193) in the influenza H3 hemagglutinin was chemically synthesized. Positions corresponding to these key residues encoded a degenerate (NNK) codon while the remaining positions on the HA retained the corresponding wild-type nucleotide sequence. Codon frequency distribution at the specified residues was analyzed by targeted next generation sequencing after library PCR amplification. *, stop codon; A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M,

methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan and Y, tyrosine.

Figure 4.5. Diversity and enrichment of antigenic motifs in an H3 HA antigenic library. An HA library targeting antigenically relevant residues (145, 155, 156, 158, 159, 189 and 193) in the influenza H3 hemagglutinin was chemically synthesized. Positions corresponding to these key residues encoded a degenerate (NNK) codon while the remaining positions on the HA retained the corresponding wild-type nucleotide sequence. The synthesized library was used as template to produce full-length reverse genetics (RG)-ready PCR-based HA segments (FL-PCR 1 and FL-PCR 2, independent amplifications). FL-PCR 1 and FL-PCR 2 were used in independent virus rescue in the background of the OH/04 wt, the laboratory-adapted strain PR/8 and in the live attenuated strain OH/04 att. The synthesized library, derived full-length RG-ready PCR-based HA segments (FL-PCR 1 and FL-PCR 2), and resulting viruses were analyzed by targeted next generation sequencing and antigenic motifs retrieved from sequencing data. Sequencing analysis for wt HA was included as control for comparison. (A) Relative frequency of antigenic motifs for each analyzed condition. Each bubble depicts a unique antigenic motif. The size of the bubble reflects the antigenic motif frequency in percent. (B) Frequency distribution of antigenic motifs for each analyzed condition. The total number of unique antigenic motifs reflects the number of individual antigenic motifs that appeared at least once.

Figure 4.6. A H3 HA variant virus with a unique antigenic motif retained binding to SAα2-6Gal. H3 viruses carrying specific antigenic motifs were tested for receptor binding specificity with varying concentrations of either peroxidase (HRP)-conjugated fetuins (fet-HRP) or sialyglycopolymers (SGP). These compounds are monospecific preparations of either SAα2-3Gal (α2-3) or SAα2-6Gal (α2-6) glycans. (A) Human pH1N1 and (B) avian ΔH5N1 viruses were used as binding control to SAα2-6Gal and SAα2-3Gal, respectively. (C) OH/04 wt virus (NHNNYRN). (D) RHAESWG variant. Glycan concentration is expressed in arbitrary units (AU) for fet-HRP or micromolar (μM) . Plotted data represent means ± standard errors (SD).

Figure 4.7. A H3 HA variant virus with a unique antigenic motif escaped antibody recognition. (A) F49 antibody binding or (B) OH/04 antisera reactivity to H3 viruses carrying specific antigenic motifs were determined by ELISA. F49 mAb binds to a conserved epitope on the H3 HA stalk domain. Plotted data represent means ± standard errors (SD). O.D., optical density.

Tables and Figures

Table 4.1. Relative frequency of the most representative antigenic motifs from experiment 01

Synthesized library		Experiment 01						wt HA[#]	
		FL-PCR 1		Rescue in OH/04 wt		Rescue in PR/8			
aa motif	freq. in %	aa motif	freq. in %	aa motif	freq. in %	aa motif	freq. in %	aa motif	freq. in %
NHNNYRN	0.018	NHNNYRN	0.017	RHAESWG	37.64	NVREFAS	32.94	NHNNYRN	98.50
AAAAAWP	0.008	RHAESWG	0.011	KYPVNVS	3.96	HVSRFVR	16.44	NHNNYGN	0.14
RFSSGGG	0.006	RFSSGGG	0.010	QYSYAKN	3.42	NVREFVR	7.36	NHNNYRD	0.12
AAAAALF	0.005	LAAAAGS	0.006	RHAESVS	3.28	HVSRFAS	6.60	NRNNYRN	0.11
ATHPAW*	0.005	AVYNVTL	0.005	RHAESKN	2.52	PVSAVYR	6.08	SHNNYRN	0.09
LPTWSVG	0.005	HAAAALL	0.005	RHAESAG	2.40	NVREFYR	3.39	NHNDYRN	0.09
LRGKFSS	0.005	LAAAALC	0.005	RHAESFR	2.22	PVSAVAS	3.31	NHNNYRS	0.07
LVMNLND	0.005	PFI*Q*V	0.005	LFSGYFR	2.01	HVREFAS	2.67	DHNNYRN	0.05
NVREFAS	0.005	PGWFIVI	0.005	QYSYAWG	2.00	NVSRFVR	2.61	NHNNCRN	0.05
PAAAASR	0.005	PPV*PLG	0.005	KYPVNWG	1.97	PVSAVVR	2.48	NHNSYRN	0.04
Others	99.933	Others	99.926	Others	38.58	Others	16.12	Others	0.74

* specifies a stop codon

[#] included as control. For comparison purpose only

Table 4.2. Relative frequency of the most representative antigenic motifs from experiment 02

Synthesized library[1]		Experiment 02							
		FL-PCR 2		Rescue in OH/04 wt		Rescue in OH/04 att		Rescue in PR/8	
aa motif	freq. in %	aa motif	freq. in %	aa motif	freq. in %	aa motif	freq. in %	aa motif	freq. in %
NHNNYRN	0.018	NHNNYRN	0.015	QHKSGLY	10.17	RFSSGGG	94.75	STKSLTK	36.40
AAAAAWP	0.008	NVREFAS	0.008	NAAAARR	8.57	RFSSGTR	0.89	ATLQERR	25.73
RFSSGGG	0.006	RFSSGGG	0.007	NTGLNVR	6.81	NFSLGGG	0.42	ATLQETK	10.67
AAAAALF	0.005	LYRSWWT	0.005	QFAHFIR	5.75	NFSLGTR	0.42	STKSLRR	8.69
ATHPAW*	0.005	RHAESWG	0.005	QHKSGVR	2.29	RFSSGHT	0.35	ATKSLTK	3.24
LPTWSVG	0.005	TAYVL*G	0.005	AVGGYVR	2.13	RFSLGGG	0.29	STLQERR	2.72
LRGKFSS	0.005	*AAAALD	0.004	NAAAAVR	2.03	RFSLGTR	0.27	PHGADGR	1.21
LVMNLND	0.005	*DKPYVL	0.004	NHESYLD	1.94	NFSSGGG	0.21	STLQETK	1.12
NVREFAS	0.005	*EQDWCE	0.004	QHKSGIR	1.79	RLSSGGG	0.17	ATLQEGR	0.91
PAAAASR	0.005	*GDRRVA	0.004	QHKSGRR	1.70	RFSSGGR	0.14	STKSLGR	0.90
Others	99.933	Others	99.939	Others	56.82	Others	2.09	Others	8.41

* specifies a stop codon

[1]Same as shown in table 4.1. For comparison purpose only.

Table 4.3. Enriched antigenic motifs emerge from rare, low-frequency variants

aa motif	Synthesized library		After PCR		After rescue		Origin
	read count	freq. in %	read count	freq. in %	read count	freq. in %	
RHAESWG	4	0.005	9	0.010	31,770	37.64	$_{Ex01}$OH/04 wt
NVREFAS	4	0.005	3	0.003	26,701	32.94	$_{Ex01}$PR/8
QHKSGLY	2	0.002	ND	ND	7,851	10.17	$_{Ex02}$OH/04 wt
RFSSGGG	5	0.006	5	0.007	74,179	94.75	$_{Ex02}$OH/04 att
STKSLTK	4	0.005	2	0.002	32,669	36.40	$_{Ex02}$PR/8

ND, not detected

Figure 4.1

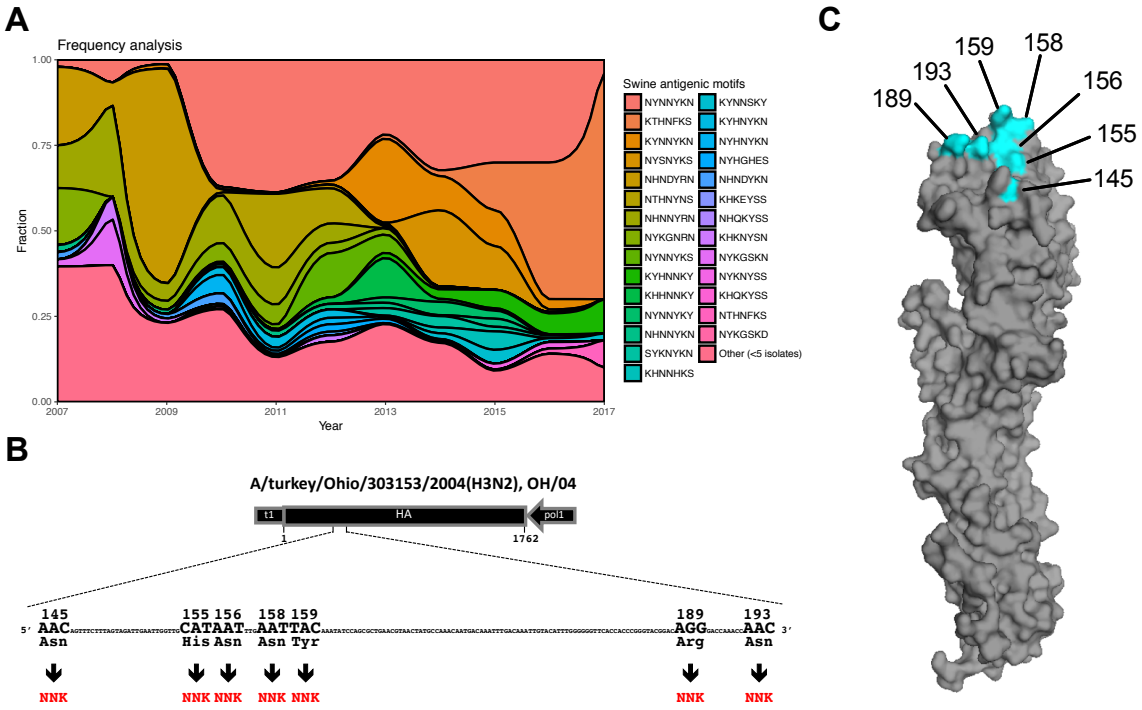

A
Frequency analysis
Fraction
Year
Swine antigenic motifs
NYNNYKN
KTHNFKS
KYNNYKN
NYSNYKS
NHNDYRN
NTHNYNS
NHNNYRN
NYKGNRN
NYNNYKS
KYHNNKY
KHHNNKY
NYNNYKY
NHNNYKN
SYKNYKN
KHNNHKS
KYNNSKY
KYHNYKN
NYHNYKN
NYHGHES
NHNDYKN
KHKEYSS
NHQKYSS
KHKNYSN
NYKGSKN
NYKNYSS
KHQKYSS
NTHNFKS
NYKGSKD
Other (<5 isolates)
B
A/turkey/Ohio/303153/2004(H3N2), OH/04
t1
HA
pol1
1
1762
145
AAC
Asn
155 156
CATAAT
His Asn
158 159
AATTAC
Asn Tyr
189
AGG
Arg
193
AAC
Asn
NNK
NNK NNK
NNK NNK
NNK
NNK
C
159
158
193
156
189
155
145

Figure 4.2

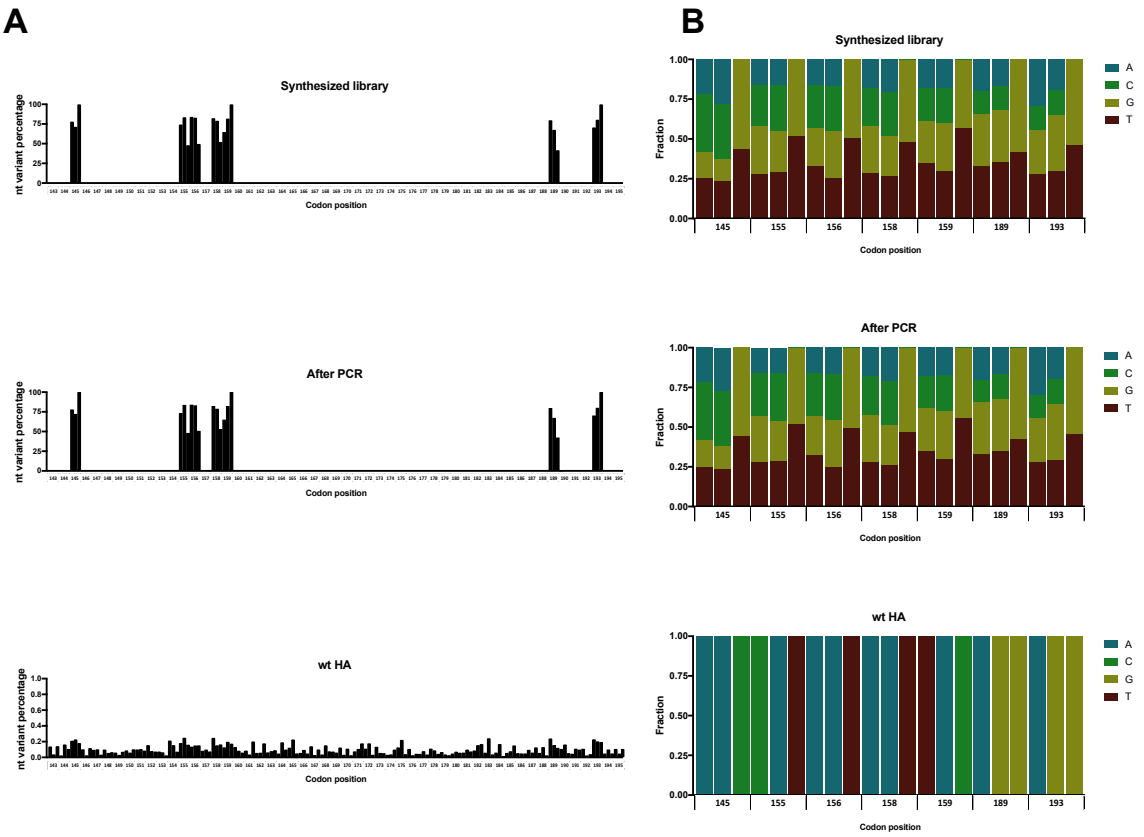

A
Synthesized library
nt variant percentage
Codon position
After PCR
nt variant percentage
Codon position
wt HA
nt variant percentage
Codon position
B
Synthesized library
Fraction
A
C
G
T
145
155
156
158
159
189
193
Codon position
After PCR
Fraction
Codon position
wt HA
Fraction
Codon position

Figure 4.3

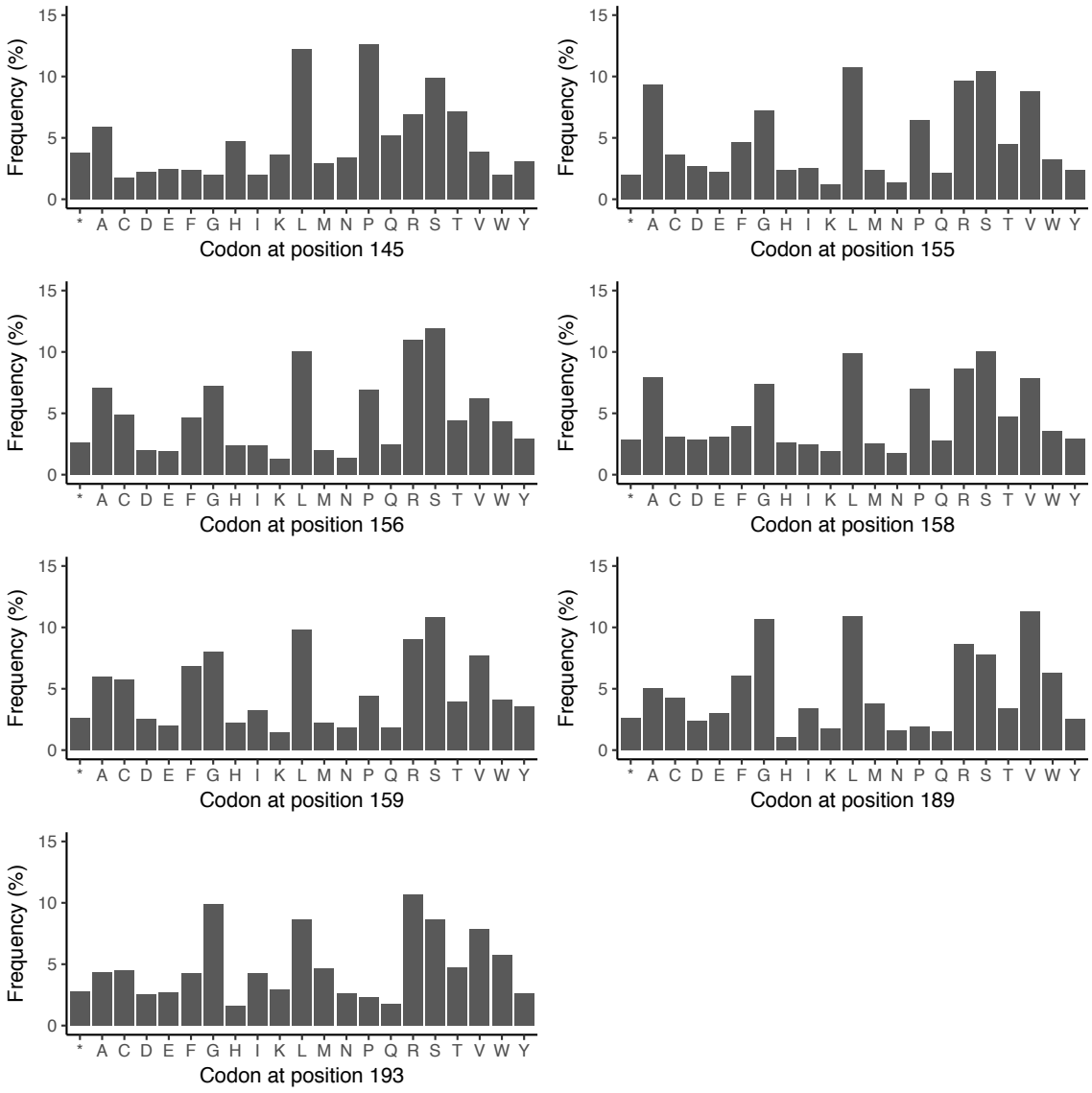

Frequency (%)
0
5
10
15
* A C D E F G H I K L M N P Q R S T V W Y
Codon at position 145
Codon at position 155
Codon at position 156
Codon at position 158
Codon at position 159
Codon at position 189
Codon at position 193

Figure 4.4

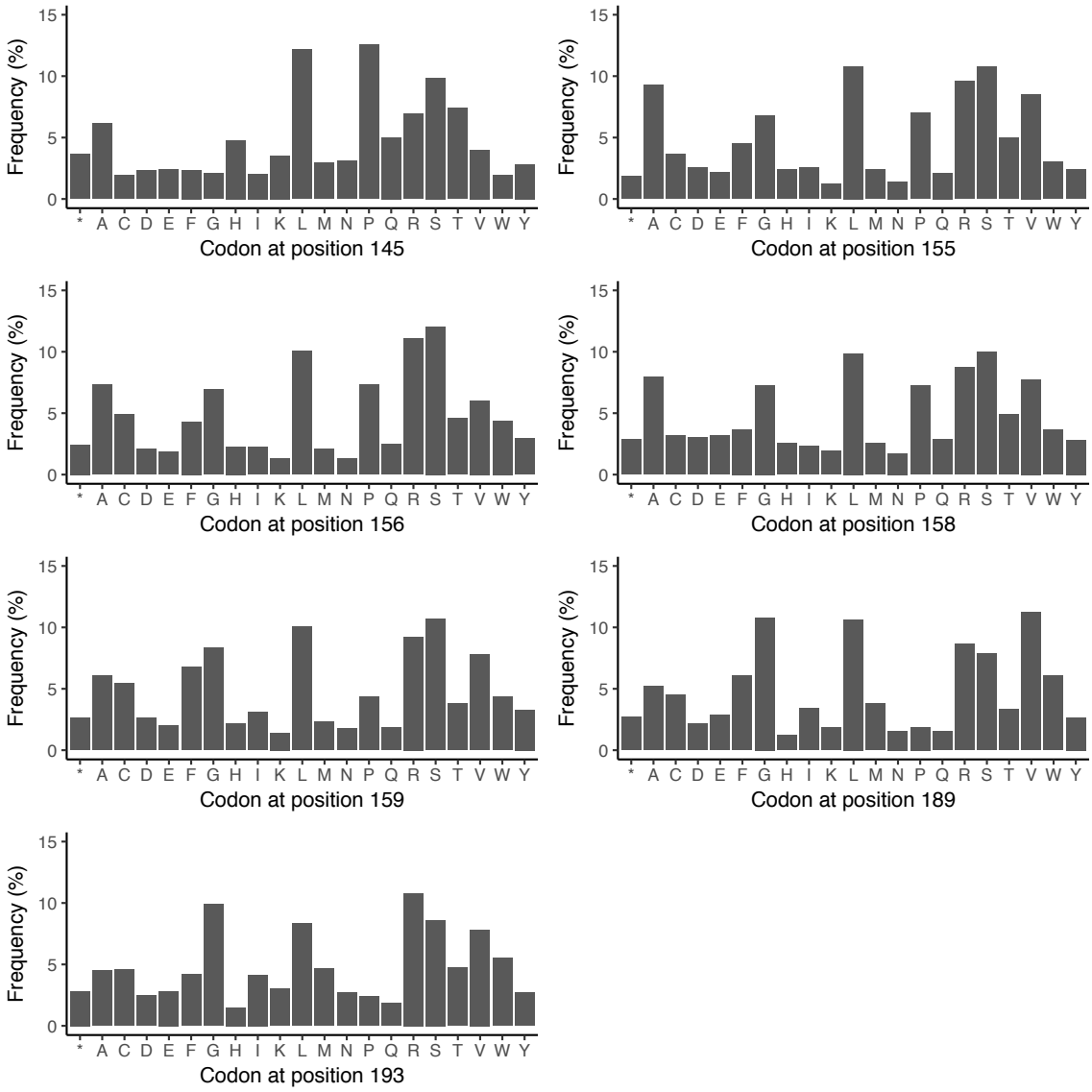

Frequency (%)
0
5
10
15
* A C D E F G H I K L M N P Q R S T V W Y
Codon at position 145
Codon at position 155
Codon at position 156
Codon at position 158
Codon at position 159
Codon at position 189
Codon at position 193

Figure 4.5

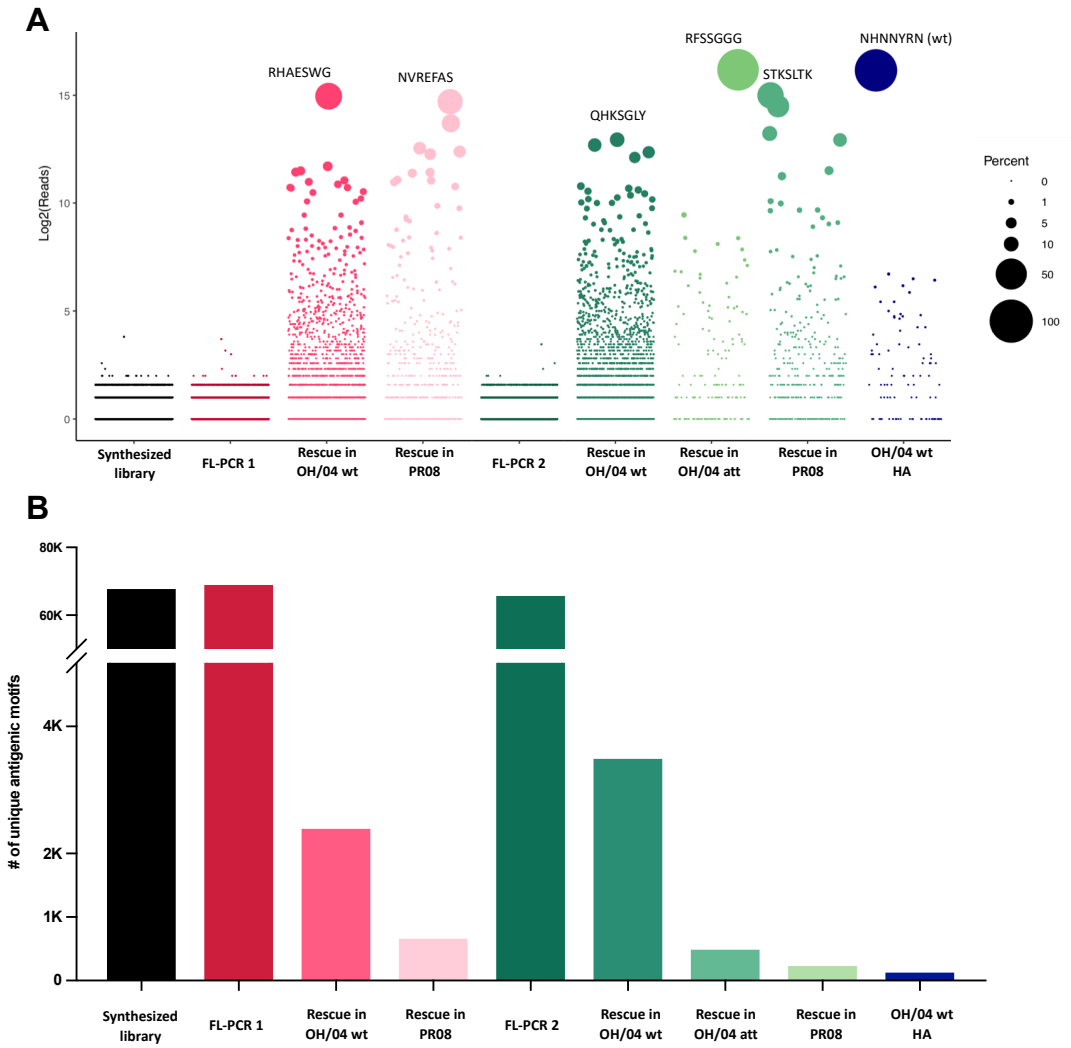

A
RHAESWG
NVREFAS
QHKSGLY
RFSSGGG
STKSLTK
NHNNYRN (wt)
Log2(Reads)
15
10
5
0
Percent
0
1
5
10
50
100
Synthesized library
FL-PCR 1
Rescue in OH/04 wt
Rescue in PR08
FL-PCR 2
Rescue in OH/04 wt
Rescue in OH/04 att
Rescue in PR08
OH/04 wt HA
B
# of unique antigenic motifs
80K
60K
4K
2K
1K
0
Synthesized library
FL-PCR 1
Rescue in OH/04 wt
Rescue in PR08
FL-PCR 2
Rescue in OH/04 wt
Rescue in OH/04 att
Rescue in PR08
OH/04 wt HA

Figure 4.6

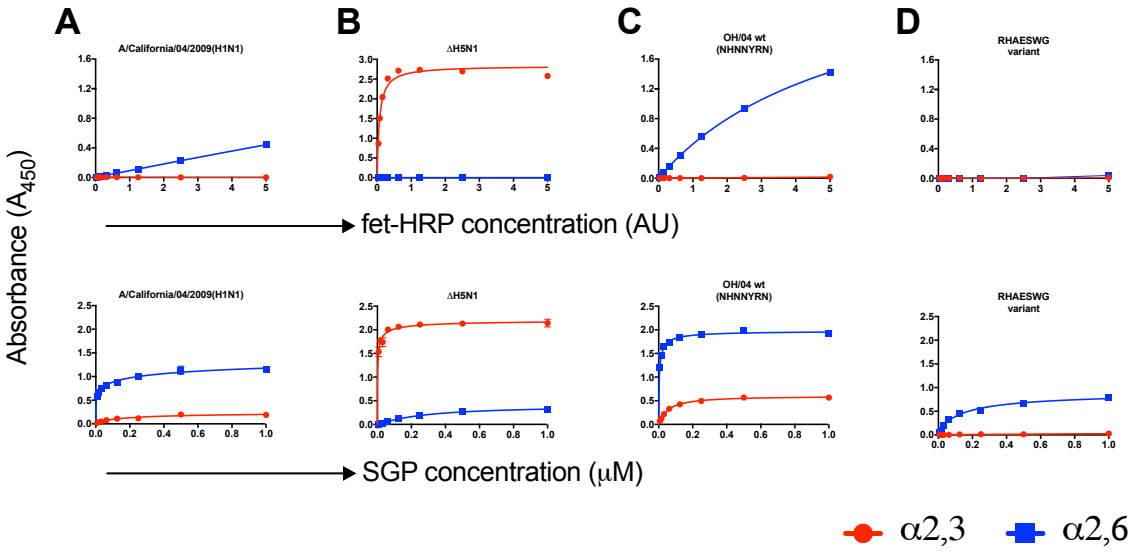
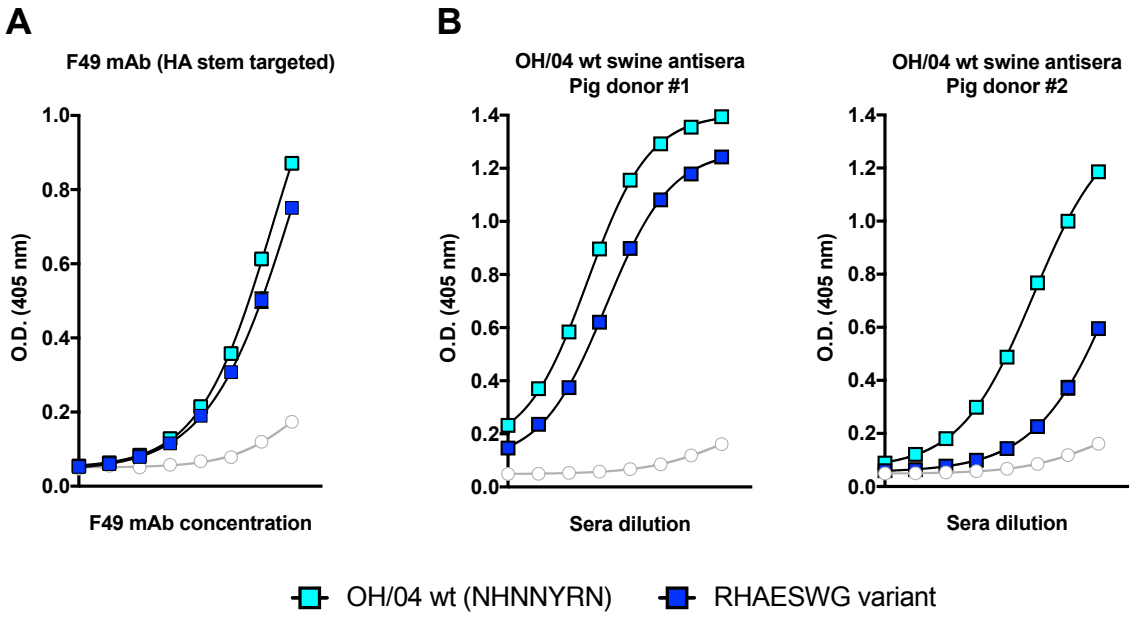

A
A/California/04/2009(H1N1)
B
ΔH5N1
C
OH/04 wt
(NHNNYRN)
D
RHAESWG
variant
Absorbance (A450)
fet-HRP concentration (AU)
SGP concentration (μM)
α2,3
α2,6

Figure 4.7


A
F49 mAb (HA stem targeted)
O.D. (405 nm)
1.0
0.8
0.6
0.4
0.2
0.0
F49 mAb concentration
B
OH/04 wt swine antisera
Pig donor #1
O.D. (405 nm)
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Sera dilution
OH/04 wt swine antisera
Pig donor #2
O.D. (405 nm)
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
Sera dilution
OH/04 wt (NHNNYRN)
RHAESWG variant

CHAPTER 5

SUMMARY AND CONCLUSIONS

Influenza A virus (IAV) remains an important pathogen for humans and swine. While vaccines are commercially available, the relative effectiveness of vaccines against IAV depends on the antigenic match of vaccine strains to circulating virus strains. Most of the antibody response stimulated following natural infection is directed against the hemagglutinin (HA), the major surface glycoprotein, to prevent virus infection and reduce clinical disease. For these reasons, HA is the main component of current IAV vaccines. Through a mechanism known as antigenic drift, IAV can circumvent preexisting antibodies by rapidly selecting amino acid substitutions in key HA epitopes, leading to immune escape. Antigenic drift is a frequent cause of reduced vaccine effectiveness, especially for IAV of the H3N2 subtype.

HA has an essential role in mediating virus attachment and entry by binding to terminal sialic acid (SA) receptors on the surface of susceptible host cells. The receptor binding site (RBS) resides in the globular head domain of HA and is relatively conserved on an otherwise hypervariable globular head domain. The identification of up to seven residues (145, 155, 156, 158, 159, 189 and 193) near the RBS as the major determinants of antigenic drift during the evolution of human and swine H3N2 IAVs has led to the hypothesis that emerging substitutions in these residues must drive antigenic change and immune escape without disrupting receptor binding properties. The main goal of this

dissertation was to explore the amino acid plasticity and the functional implications of amino acid substitutions in these key residues.

In the first part of this dissertation, we carried out site-directed mutagenesis experiments to better understand functional constraints to substitutions near the RBS, with a focus on residue 145. We generated a panel of H3 HA mutant viruses carrying a single amino acid substitution at this residue, representing each of the 20 possible amino acids. The results indicated that residue 145 displayed remarkable amino acid plasticity *in vitro* tolerating multiple amino acid substitutions, of which many have not been yet observed in nature. Mutant viruses carrying substitutions at residue 145 showed no major impact on virus replication. While substitutions at this residue did not alter receptor binding specificity, mutant viruses with substitutions not commonly found in nature displayed diminished receptor binding avidity. Interestingly, changes at this residue also modulated binding to particular glycan structures in a glycan array. In human H3N2 IAVs, it has been proposed that the emergence of the HA N145K substitution leads to a change in the antibody immunodominance. Our work on the antigenic characterization of the panel of mutant viruses confirmed the impact of residue 145K in shifting the hierarchy of the antibody immunodominance in swine H3N2 IAVs.

In the second part of this dissertation, we systematically explored the *in vitro* amino acid plasticity and functional sequence space of these 7 antigenically relevant residues in the H3 HA. By combining saturation mutagenesis with next-generation sequencing, we envisioned the development of a highly diverse H3 HA antigenic library targeting these key residues for potential vaccine use and testing. Our deep mutational scanning approach generated high diversity at both the nucleotide and amino acid levels.

While our strategy is efficient in producing H3 HA antigenic virus libraries in different donor virus strains, virus library diversity was drastically reduced following virus rescue, indicating evidence of a bottleneck. Nonetheless, our results revealed that despite limited diversity in nature, residues 145, 155, 156, 158, 159, 189 and 193 exhibited extraordinary amino acid plasticity *in vitro*, and viruses with very distinct antigenic motifs were rapidly selected *in vitro*. To validate our approach, we attempted to isolate virus variants with unique antigenic motifs. An isolated virus variant not previously found in nature was further characterized and shown to exhibit exceptional stability at the genetic level. *In vitro* characterization revealed this isolated virus variant possess distinct receptor binding and antigenic properties relative to the wt HA virus.

The complex antigenic composition of IAV remains a challenge for vaccine selection and effective vaccination. This work provided a better understanding of the functional effects of amino acid substitutions near the RBS implicated in antigenic drift and the consequences to receptor binding and antigenicity. These findings have important implications for understanding IAV antigenic diversity and aiding the development of an alternative IAV vaccine approach that potentially induces broader protection against antigenically distinct H3 virus strains. Nonetheless, many important questions still remain unanswered. Considering the findings presented in this dissertation, future directions of the work should aim at:

1. **Characterize the impact of single amino acid substitutions on receptor binding properties and antigenicity of influenza A viruses of the H3 subtype.**

a. While residue 145 displayed remarkable plasticity *in vitro*, the impact and stability of substitutions at this residue to IAV replication and transmission *in vivo* remains to be fully characterized.
b. Antigenic characterization indicated most substitutions at residue 145 had little antigenic effects against the selected panel of sera tested. Yet, it will be critical to assess the individual impact of these substitutions on the immunogenicity (elicited antibody response) and antibody immunodominance in animal models.
c. It is unclear how the emergence of the HA N145K substitution alters the hierarchy of antibody immunodominance, but likely reflects differential or preferential activation of specific B cell lineages. In light of a recent report revealing carbohydrate dependent activation of B cells, the ability to modulate binding to a broad range of glycans by substitutions near the RBS offers a possibility to augment B cell activation by combining non-cognate (carbohydrate dependent) with cognate (antigen specific) interactions that warrants further investigation.
d. In addition to residue 145, there are 6 other residues as major determinants of antigenic drift. The individual effect of amino acid substitutions in these key relevant residues are largely unknown and need to be addressed.

2. Determine the amino acid plasticity of an influenza virus H3 HA.

a. Our deep mutational scanning approach generated an immense diversity of antigenic motifs, but virus libraries diversity was greatly reduced

following virus rescue. These results suggest strong selective pressure and potential bottlenecks. However, our sequencing analysis was focused on a single, late time-point. Future experiment should focus on evaluating the temporal dynamics of antigenic virus libraries following rescue to identify optimal time-points to harvest virus libraries at their peak of diversity.

b. Part of the bottleneck is intimately related to the segmented nature of the IAV genome. During transfection, only a fraction of cells likely produce most of the initial virus variants. Furthermore, many of the H3 HA variants in the library could be nonfunctional. The development of cells lines expressing surface HA and NA proteins may ameliorate the impact of bottlenecks. This should also improve the replicate-to-replicate consistency.

c. Our results also showed that reduced diversity of antigenic motifs in virus libraries varied with the virus donor strain used. It is essential to identify and reduce the impact of backbone-dependent constraints to virus library diversity.

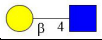
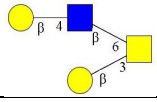
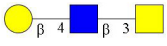
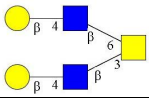
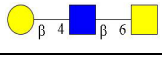
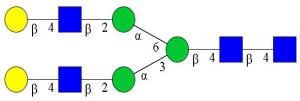
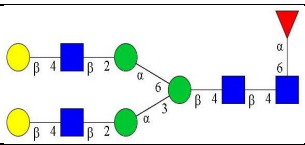
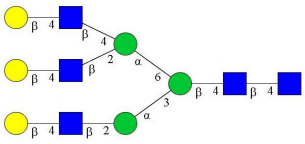
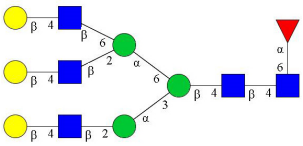
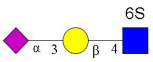
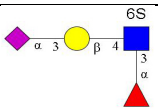
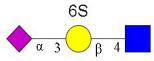
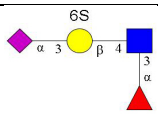
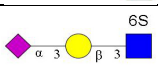
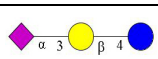
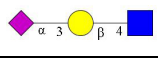
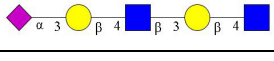
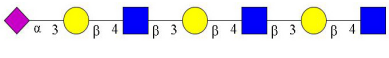
d. While we focused on the amino acid plasticity *in vitro*, our platform can be useful as an experimental system to model antigenic drift by tracking the dynamics of antigenic virus libraries in naïve and immunized hosts.

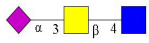
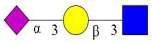
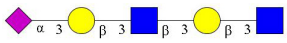
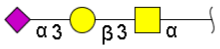
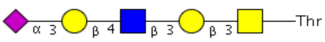
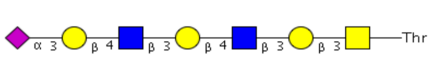
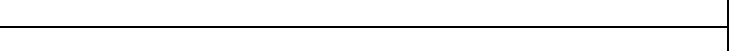
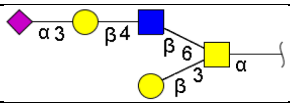
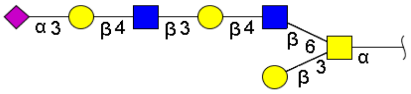
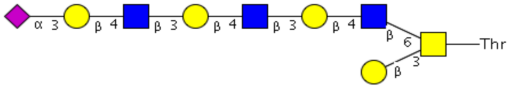
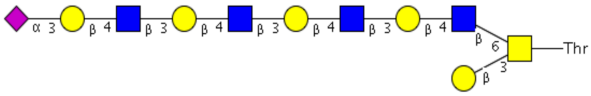
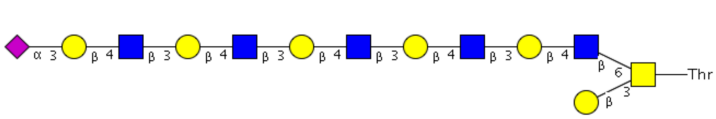
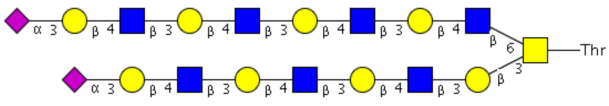
e. We envisioned this strategy as an alternative IAV vaccine approach that induces broader protection. This hypothesis need to be tested to determine whether a H3 HA antigenic virus library can augment both the breadth and depth without compromising the magnitude of antigen-specific antibody

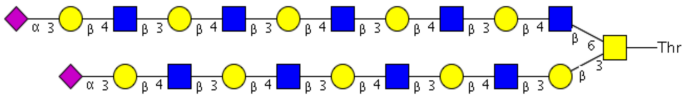
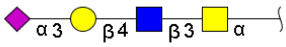
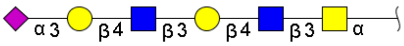
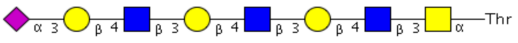
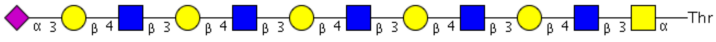
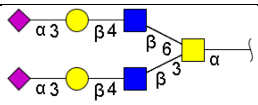
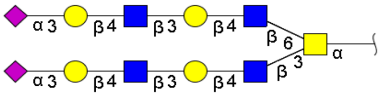
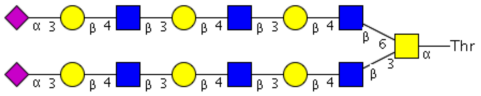
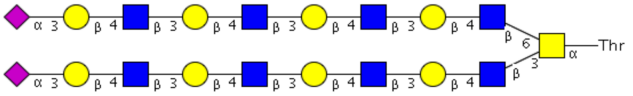
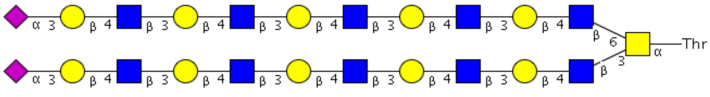
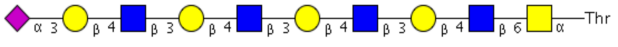
responses as compared with wt HA virus antigens in animal models. This will require the evaluation of the best vaccine platform (live or inactivated immunization), route of administration, prime-boost regime, vaccine dose spacing, impact of pre-existing immunity and so on.

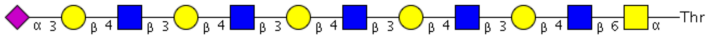
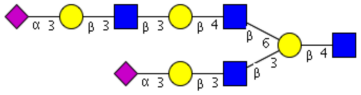
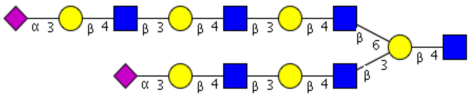
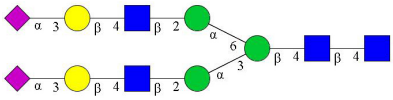
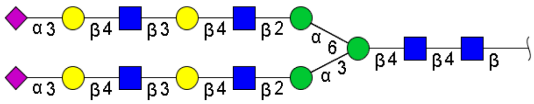
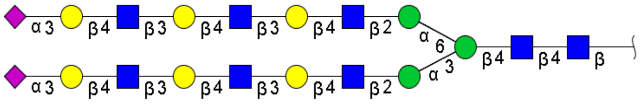
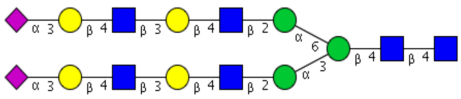
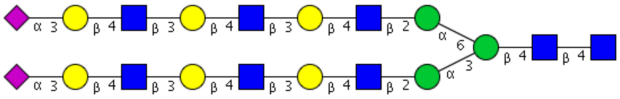
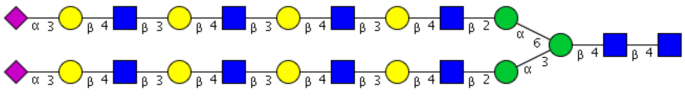
f. IAV receptor-mimicry broadly neutralizing antibodies (bnAbs) is a new class of antibodies with increased breadth of recognition against antigenic drifted virus strains. The maturation pathway to the development of receptor-mimicry bnAbs or vaccination strategies for inducing receptor-mimicry bnAbs remain poorly understood. As our approach targets key residues near the RBS, the ability of inducing such class of antibodies by our strategy need to be investigated.

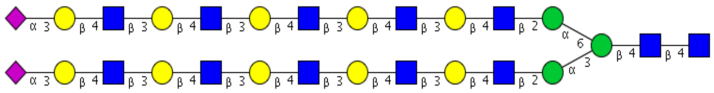
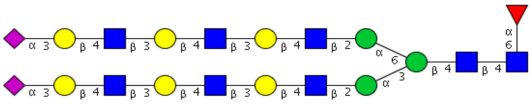
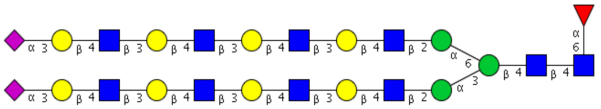
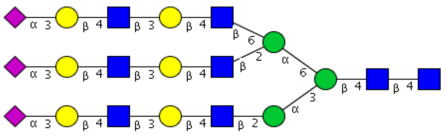
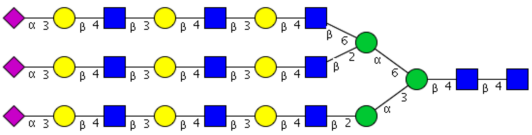
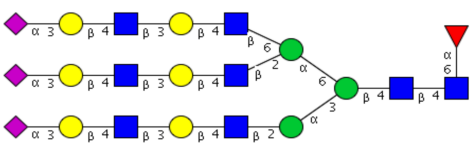
Appendix A. List of glycans present in the glycan array

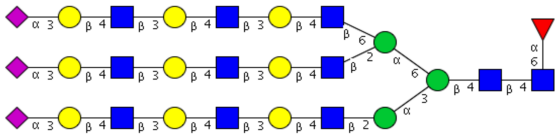
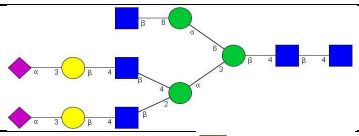
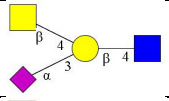
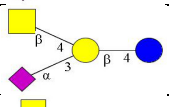
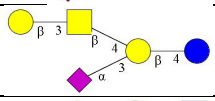
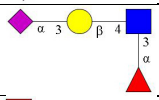
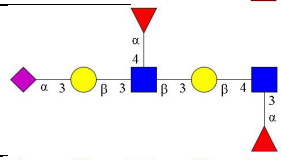
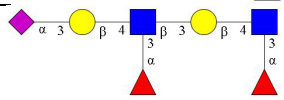
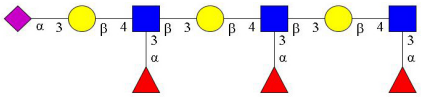
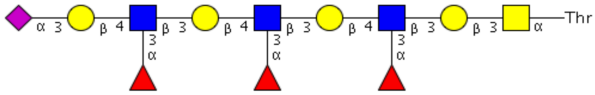
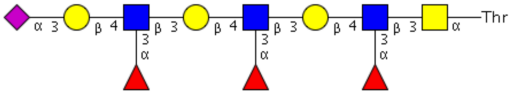
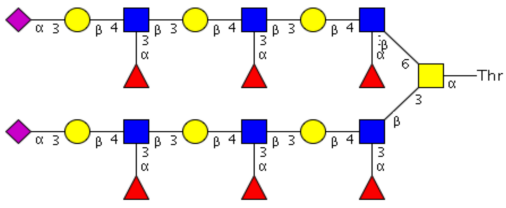
Glycan #	Common Name	Structure
1	Galβ(1-4)GlcNAcβ-ethyl-NH2	
2	Galβ(1-4)GlcNAcβ(1-3)Galβ(1-3)GalNAcα-Thr-NH2	
3	Galβ(1-4)GlcNAcβ(1-6)[Galβ(1-3)]-GalNAcα-Thr-NH2	
4	Galβ(1-4)GlcNAcβ(1-3)GalNAcα-Thr-NH2	
5	Galβ(1-4)GlcNAcβ(1-3)[Galβ(1-4)GlcNAcβ(1-6)]-GalNAcα-Thr-NH2	
6	Galβ(1-4)GlcNAcβ(1-6)GalNAcα-Thr-NH2	
7	Galβ(1-4)GlcNAcβ(1-2)Manα (1-3)[Galβ(1-4)GlcNAcβ(1-2)Manα(1-6)]-Manβ(1-4)GlcNAcβ(1-4)GlcNAcβ-Asn-NH2	
8	Galβ(1-4)GlcNAcβ(1-2)Manα(1-3)[Galβ(1-4)GlcNAcβ(1-2)Manα(1-6)]-Manβ(1-4)GlcNAcβ(1-4)[Fucα(1-6)]-GlcNAcβ-Asn-Ser-Thr-NH2	
9	Galβ(1-4)GlcNAcβ(1-2)Manα(1-3){Galβ(1-4)GlcNAcβ(1-2)[Galβ(1-4)GlcNAcβ(1-2)]-Manα(1-6)}-Manβ(1-4)GlcNAcβ(1-4)GlcNAcβ-Asn-Lys-NH2	
10	Galβ(1-4)GlcNAcβ(1-2)Manα(1-3){Galβ(1-4)GlcNAcβ(1-2)[Galβ(1-4)GlcNAcβ(1-2)]-Manα(1-6)}-Manβ(1-4)GlcNAcβ(1-4)[Fucα(1-6)]-GlcNAcβ-(Lys-Val-Ala)Asn-Lys-Thr-NH2	
11	NeuAcα(2-3)Galβ(1-4)6-O-sulfo-GlcNAcβ-propyl-NH2	
12	NeuAcα(2-3)Galβ(1-4)[Fucα(1-3)]-6-O-sulfo-GlcNAcβ-propyl-NH2	
13	NeuAcα(2-3)6-O-sulfo-Galβ(1-4)GlcNAcβ-ethyl-NH2	
14	NeuAcα(2-3)6-O-sulfo-Galβ(1-4)[Fucα(1-3)]-GlcNAcβ-propyl-NH2	
15	NeuAcα(2-3)Galβ(1-3)6-O-sulfo-GlcNAcβ-propyl-NH2	
16	NeuAcα(2-3)Galβ(1-4)Glcβ-ethyl-NH2	
17	NeuAcα(2-3)Galβ(1-4)GlcNAcβ-ethyl-NH2	
18	NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ-ethyl-NH2	
19	NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ-ethyl-NH2	

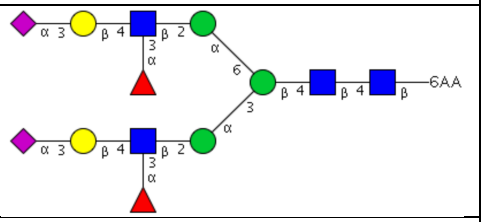
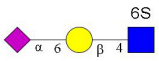
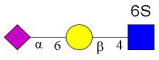
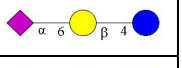
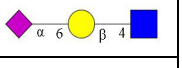
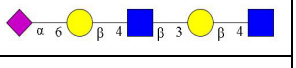
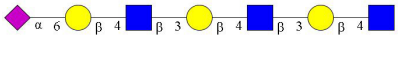
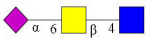
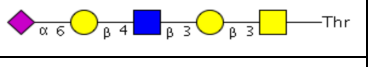
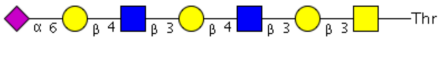
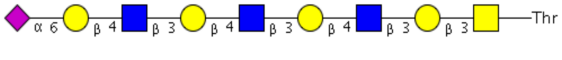
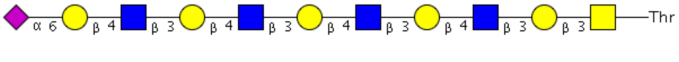
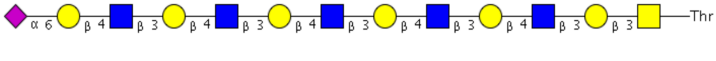
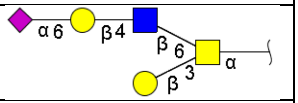
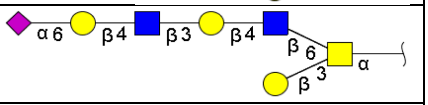
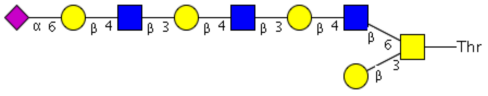
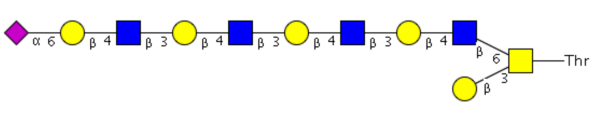
Glycan #	Common Name	Structure
20	NeuAcα(2-3)GalNAcβ(1-4)GlcNAcβ-ethyl-NH2	
21	NeuAcα(2-3)Galβ(1-3)GlcNAcβ-ethyl-NH2	
22	NeuAcα(2-3)Galβ(1-3)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ-ethyl-NH2	
23	NeuAcα(2-3)Galβ(1-3)GlcNAcβ(1-3)Galβ(1-3)GlcNAcβ-ethyl-NH2	
24	NeuAcα(2-3)Galβ(1-3)GalNAcβ(1-3)Galα(1-4)Galβ(1-4)Glcβ-ethyl-NH2	
25	NeuAcα(2-3)Galβ(1-3)GalNAcα-Thr-NH2	
26	NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-3)GalNAcα-Thr-NH2	
27	NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-3)GalNAcα-Thr-NH2	
28	NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-3)GalNAcα-Thr-NH2	
29	NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-3)GalNAcα-Thr-NH2	
30	NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-3)GalNAcα-Thr-NH2	
31	NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-6)[Galβ(1-3)]-GalNAcα-Thr-NH2	
32	NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-6)[Galβ(1-3)]-GalNAcα-Thr-NH2	
33	NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-6)[Galβ(1-3)]-GalNAcα-Thr-NH2	
34	NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-6)[Galβ(1-3)]-GalNAcα-Thr-NH2	
35	NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-6)[Galβ(1-3)]-GalNAcα-Thr-NH2	
36	NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-6)[NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-3)]-GalNAcα-Thr-NH2	

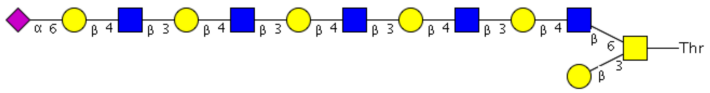
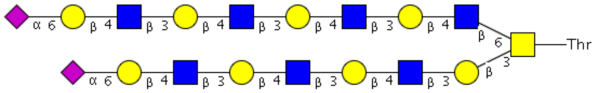
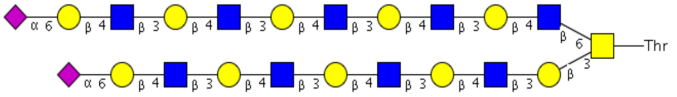
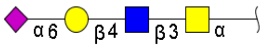
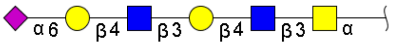
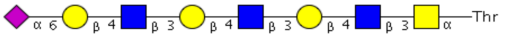
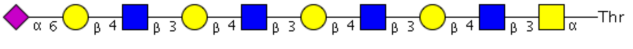
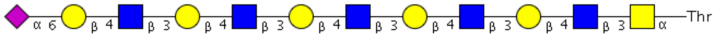
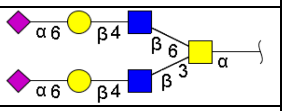
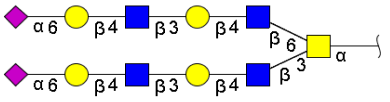
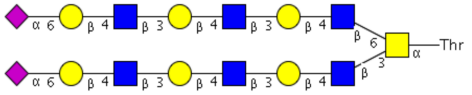
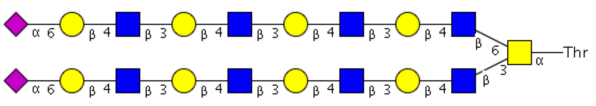
Glycan #	Common Name	Structure
37	NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-6)[NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-3)]-GalNAcα-Thr-NH2	
38	NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)GalNAcα-Thr-NH2	
39	NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)GalNAcα-Thr-NH2	
40	NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)GalNAcα-Thr-NH2	
41	NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)GalNAcα-Thr-NH2	
42	NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)GalNAcα-Thr-NH2	
43	NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)[NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-6)]-GalNAcα-Thr-NH2	
44	NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)[NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-6)]-GalNAcα-Thr-NH2	
45	NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)[NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-6)]-GalNAcα-Thr-NH2	
46	NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)[NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-6)]-GalNAcα-Thr-NH2	
47	NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)[NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-6)]-GalNAcα-Thr-NH2	
48	NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-6)GalNAcα-Thr-NH2	

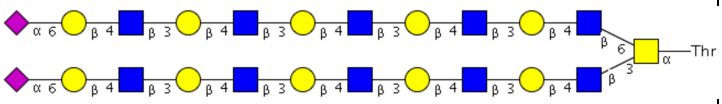
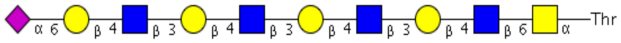
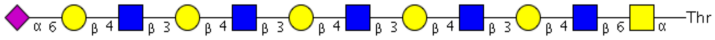
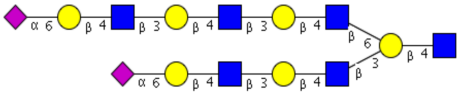
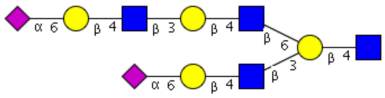
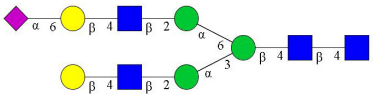
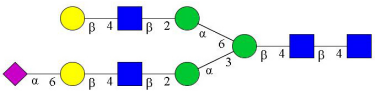
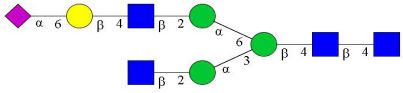
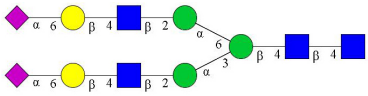
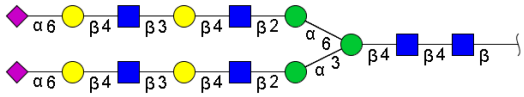
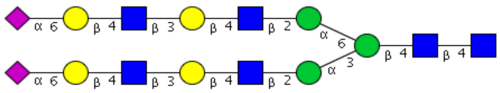
Glycan #	Common Name	Structure
49	NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-6)GalNAcα-Thr-NH2	
50	NeuAcα(2-3)Galβ(1-3)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-6)[NeuAcα(2-3)Galβ(1-3)GlcNAcβ(1-3)] Galβ(1-4)GlcNAcβ-ethyl-NH2	
51	NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-6)[NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)] Galβ(1-4)GlcNAcβ-ethyl-NH2	
52	NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-2)Manα(1-3)[NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-2)Manα(1-6)]-Manβ(1-4)GlcNAcβ(1-4)GlcNAcβ-Asn-NH2	
53	NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)Manα(1-3)[NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)Manα(1-6)]-Manβ(1-4)GlcNAcβ(1-4)GlcNAcβ-Asn-NH2	
54	NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)Manα(1-3)[NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)Manα(1-6)]-Manβ(1-4)GlcNAcβ(1-4)GlcNAcβ-Asn-NH2	
55	NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)Manα(1-3)[NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)Manα(1-6)]-Manβ(1-4)GlcNAcβ(1-4)GlcNAcβ-(Lys-Val-Ala)Asn-Lys-Thr-NH2	
56	NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)Manα(1-3)[NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)Manα(1-6)]-Manβ(1-4)GlcNAcβ(1-4)GlcNAcβ-(Lys-Val-Ala)Asn-Lys-Thr-NH2	
57	NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)Manα(1-3)[NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)Manα(1-6)]-Manβ(1-4)GlcNAcβ(1-4)GlcNAcβ-(Lys-Val-Ala)Asn-Lys-Thr-NH2	

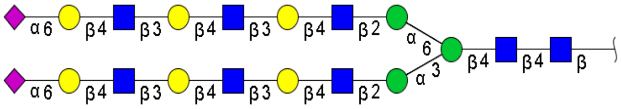
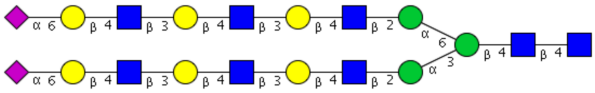
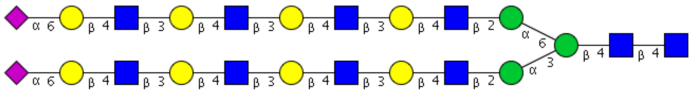
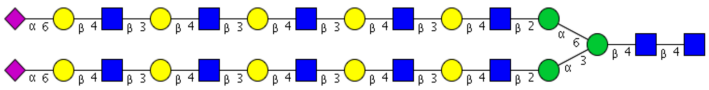
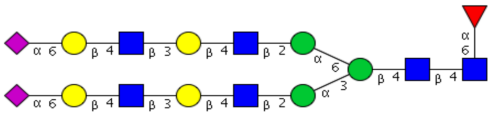
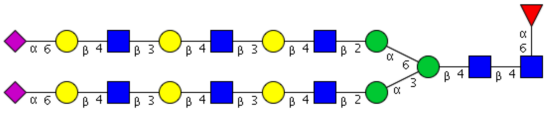
Glycan #	Common Name	Structure
58	NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)Manα(1-3)[NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)Manα(1-6)]-Manβ(1-4)GlcNAcβ(1-4)GlcNAcβ-(Lys-Val-Ala)Asn-Lys-Thr-NH2	
59	NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)Manα(1-3)[NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)Manα(1-6)]-Manβ(1-4)GlcNAcβ(1-4)[Fucα(1-6)]-GlcNAcβ-(Lys-Val-Ala)Asn-Lys-Thr-NH2	
60	NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)Manα(1-3)[NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)Manα(1-6)]-Manβ(1-4)GlcNAcβ(1-4)[Fucα(1-6)]-GlcNAcβ-(Lys-Val-Ala)Asn-Lys-Thr-NH2	
61	NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)Manα(1-3){NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)[NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-6)Manα(1-6)]}-Manβ(1-4)GlcNAcβ(1-4)GlcNAcβ-(Lys-Val-Ala)Asn-Lys-Thr-NH2	
62	NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)Manα(1-3){NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)[NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-6)Manα(1-6)]}-Manβ(1-4)GlcNAcβ(1-4)GlcNAcβ-(Lys-Val-Ala)Asn-Lys-Thr-NH2	
63	NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)Manα(1-3){NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)[NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-6)Manα(1-6)]}-Manβ(1-4)GlcNAcβ(1-4)[Fucα(1-6)]-GlcNAcβ-(Lys-Val-Ala)Asn-Lys-Thr-NH2	

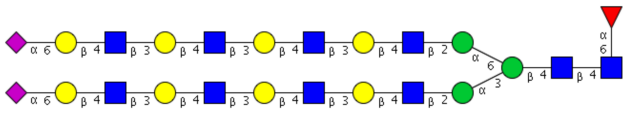
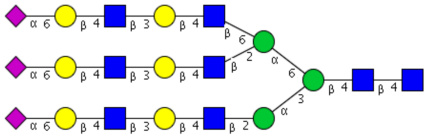
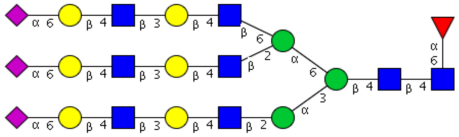
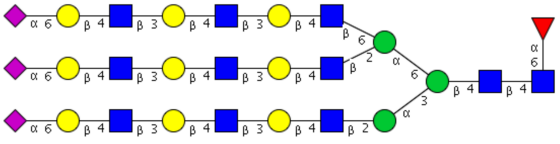
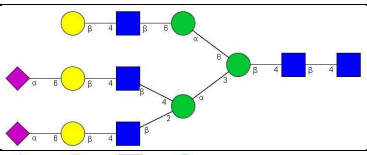
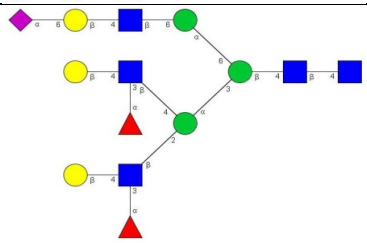
Glycan #	Common Name	Structure
64	NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)Manα(1-3){NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)[NeuAcα(2-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-6)Manα(1-6)]}-Manβ(1-4)GlcNAcβ(1-4)[Fucα(1-6)]-GlcNAcβ-(Lys-Val-Ala)Asn-Lys-Thr-NH2	
65	Gn/3'SLN/3'SLN-TriN	
66	NeuAcα(2-3)[GalNAcβ(1-4)]-Galβ(1-4)GlcNAcβ-ethyl-NH2	
67	NeuAcα(2-3)[GalNAcβ(1-4)]-Galβ(1-4)Glcβ-ethyl-NH2	
68	Galβ(1-3)GalNAcβ(1-4)[NeuAcα(2-3)]-Galβ(1-4)Glcβ-ethyl-NH2	
69	NeuAcα(2-3)Galβ(1-4)[Fucα(1-3)]-GlcNAcβ-propyl-NH2	
70	NeuAcα(2-3)Galβ(1-3)[Fucα(1-4)]-GlcNAcβ(1-3)Galβ(1-4)[Fucα(1-3)]-GlcNAcβ-ethyl-NH2	
71	NeuAcα(2-3)Galβ(1-4)[Fucα(1-3)]-GlcNAcβ(1-3)Galβ(1-4)[Fucα(1-3)]-GlcNAcβ-ethyl-NH2	
72	NeuAcα(2-3)Galβ(1-4)[Fucα(1-3)]-GlcNAcβ(1-3)Galβ(1-4)[Fucα(1-3)]-GlcNAcβ(1-3)Galβ(1-4)[Fucα(1-3)]-GlcNAcβ-ethyl-NH2	
73	NeuAcα(2-3)Galβ(1-4)[Fucα(1-3)]-GlcNAcβ(1-3)Galβ(1-4)[Fucα(1-3)]-GlcNAcβ(1-3)Galβ(1-4)[Fucα(1-3)]-GlcNAcβ(1-3)Galβ(1-3)GalNAcα-Thr-NH2	
74	NeuAcα(2-3)Galβ(1-4)[Fucα(1-3)]-GlcNAcβ(1-3)Galβ(1-4)[Fucα(1-3)]-GlcNAcβ(1-3)Galβ(1-4)[Fucα(1-3)]-GlcNAcβ(1-3)GalNAcα-Thr-NH2	
75	NeuAcα(2-3)Galβ(1-4)[Fucα(1-3)]-GlcNAcβ(1-3)Galβ(1-4)[Fucα(1-3)]-GlcNAcβ(1-3)Galβ(1-4)[Fucα(1-3)]-GlcNAcβ(1-3)[NeuAcα(2-3)Galβ(1-4)[Fucα(1-3)]-GlcNAcβ(1-3)Galβ(1-4)[Fucα(1-3)]-GlcNAcβ(1-3)Galβ(1-4)[Fucα(1-3)]-GlcNAcβ(1-6)]-GalNAcα-Thr-NH2	

Glycan #	Common Name	Structure
76	NeuAcα(2-3)Galβ(1-4)[Fucα(1-3)]-GlcNAcβ(1-2)Manα(1-3)[NeuAcα(2-3)Galβ(1-4)[Fucα(1-3)]-GlcNAcβ(1-2)Manα(1-6)]-Manβ(1-4)GlcNAcβ(1-4)GlcNAcβ-(Lys-Val-Ala)Asn-(Lys-Thr)NH2	
77	NeuAcα(2-6)Galb(1-4)(6S)GlcNacb-ethyl-NH2	
78	NeuAcα(2-6)Galβ(1-4)6-O-sulfo-GlcNAcβ-propyl-NH2	
79	NeuAcα(2-6)Galβ(1-4)Glcβ-ethyl-NH2	
80	NeuAcα(2-6)Galβ(1-4)GlcNAcβ-ethyl-NH2	
81	NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ-ethyl-NH2	
82	NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ-ethyl-NH2	
83	NeuAcα(2-6)GalNAcβ(1-4)GlcNAcβ-ethyl-NH2	
84	NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-3)GalNAcα-Thr-NH2	
85	NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-3)GalNAcα-Thr-NH2	
86	NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-3)GalNAcα-Thr-NH2	
87	NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-3)GalNAcα-Thr-NH2	
88	NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-3)GalNAcα-Thr-NH2	
89	NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-6)[Galβ(1-3)]-GalNAcα-Thr-NH2	
90	NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-6)[Galβ(1-3)]-GalNAcα-Thr-NH2	
91	NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-6)[Galβ(1-3)]-GalNAcα-Thr-NH2	
92	NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-6)[Galβ(1-3)]-GalNAcα-Thr-NH2	

Glycan #	Common Name	Structure
93	NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-6)[Galβ(1-3)]-GalNAcα-Thr-NH2	
94	NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-6)[NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-3)]-GalNAcα-Thr-NH2	
95	NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-6)[NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-3)]-GalNAcα-Thr-NH2	
96	NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)GalNAcα-Thr-NH2	
97	NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)GalNAcα-Thr-NH2	
98	NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)GalNAcα-Thr-NH2	
99	NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)GalNAcα-Thr-NH2	
100	NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)GalNAcα-Thr-NH2	
101	NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)[NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-6)]-GalNAcα-Thr-NH2	
102	NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)[NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-6)]-GalNAcα-Thr-NH2	
103	NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)[NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-6)]-GalNAcα-Thr-NH2	
104	NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)[NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-6)]-GalNAcα-Thr-NH2	

Glycan #	Common Name	Structure
105	NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)[NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-6)]-GalNAcα-Thr-NH2	
106	NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-6)GalNAcα-Thr-NH2	
107	NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-6)GalNAcα-Thr-NH2	
108	NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-6)[NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)] Galβ(1-4)GlcNAcβ-ethyl-NH2	
109	NeuAcα(2-6)Galβ(1-3)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-6)[NeuAcα(2-6)Galβ(1-3)GlcNAcβ(1-3)] Galβ(1-4)GlcNAcβ-ethyl-NH2	
110	Galβ(1-4)GlcNAcβ(1-2)Manα(1-3)[NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-2)Manα(1-6)]-Manβ(1-4)GlcNAcβ(1-4)GlcNAcβ-Asn-NH2	
111	NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-2)Manα(1-3)[Galβ(1-4)GlcNAcβ(1-2)Manα(1-6)]-Manβ(1-4)GlcNAcβ(1-4)GlcNAcβ-Asn-NH2	
112	GlcNAcβ(1-2)Manα(1-3)[NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-2)Manα(1-6)]-Manβ(1-4)GlcNAcβ(1-4)GlcNAcβ-Asn-NH2	
113	NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-2)Manα(1-3)[NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-2)Manα(1-6)]-Manβ(1-4)GlcNAcβ(1-4)GlcNAcβ-Asn-NH2	
114	NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)Manα(1-3)[NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)Manα(1-6)]-Manβ(1-4)GlcNAcβ(1-4)GlcNAcβ-Asn-NH2	
115	NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)Manα(1-3)[NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)Manα(1-6)]-Manβ(1-4)GlcNAcβ(1-4)GlcNAcβ-(Lys-Val-Ala)Asn-Lys-Thr-NH2	

Glycan #	Common Name	Structure
116	NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)Manα(1-3)[NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)Manα(1-6)]-Manβ(1-4)GlcNAcβ(1-4)GlcNAcβ-Asn-NH2	
117	NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)Manα(1-3)[NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)Manα(1-6)]-Manβ(1-4)GlcNAcβ(1-4)GlcNAcβ-(Lys-Val-Ala)Asn-Lys-Thr-NH2	
118	NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)Manα(1-3)[NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)Manα(1-6)]-Manβ(1-4)GlcNAcβ(1-4)GlcNAcβ-(Lys-Val-Ala)Asn-Lys-Thr-NH2	
119	NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)Manα(1-3)[NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)Manα(1-6)]-Manβ(1-4)GlcNAcβ(1-4)GlcNAcβ-(Lys-Val-Ala)Asn-Lys-Thr-NH2	
120	NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)Manα(1-3)[NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)Manα(1-6)]-Manβ(1-4)GlcNAcβ(1-4)[Fucα(1-6)]-GlcNAcβ-(Lys-Val-Ala)Asn-Lys-Thr-NH2	
121	NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)Manα(1-3)[NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)Manα(1-6)]-Manβ(1-4)GlcNAcβ(1-4)[Fucα(1-6)]-GlcNAcβ-(Lys-Val-Ala)Asn-Lys-Thr-NH2	

Glycan #	Common Name	Structure
122	NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)Manα(1-3)[NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)Manα(1-6)]-Manβ(1-4)GlcNAcβ(1-4)[Fucα(1-6)]-GlcNAcβ-(Lys-Val-Ala)Asn-Lys-Thr-NH2	
123	NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)Manα(1-3){NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)[NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-6)Manα(1-6)]}-Manβ(1-4)GlcNAcβ(1-4)GlcNAcβ-(Lys-Val-Ala)Asn-Lys-Thr-NH2	
124	NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)Manα(1-3){NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)[NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-6)Manα(1-6)]}-Manβ(1-4)GlcNAcβ(1-4)[Fucα(1-6)]-GlcNAcβ-(Lys-Val-Ala)Asn-Lys-Thr-NH2	
125	NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)Manα(1-3){NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-2)[NeuAcα(2-6)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-3)Galβ(1-4)GlcNAcβ(1-6)Manα(1-6)]}-Manβ(1-4)GlcNAcβ(1-4)[Fucα(1-6)]-GlcNAcβ-(Lys-Val-Ala)Asn-Lys-Thr-NH2	
126	LN/6'SLN/6'SLN-TriN	
127	6'SLN/LeX/LeX-TriN	
128	6'SLNLN/LeX/LeX-TriN