Comparative Analysis of Mutational Constraints Across Influenza Hemagglutinin Subtypes

Jenny Ahn

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2026

Reading Committee:
Jesse D. Bloom, Chair
Neil King
Deborah Fuller

Program Authorized to Offer Degree:
Microbiology

University of Washington

Abstract

Comparative Analysis of Mutational Constraints Across Influenza Hemagglutinin Subtypes

Jenny Ahn

Chair of the Supervisory Committee:
Affiliate Professor Jesse D. Bloom
Fred Hutchinson Cancer Center

H7 avian influenza represents a significant pandemic threat due to its ability to cause severe disease in humans and its potential for cross-species transmission. To enhance our understanding of this virus and improve pandemic preparedness, we developed a comprehensive deep mutational scanning system to systematically assess the functional effects of mutations in H7 hemagglutinin (HA). This system utilizes a pseudotyped lentiviral platform to precisely determine how individual mutations affect critical viral functions including HA-mediated cell entry, pH stability, and receptor binding specificity. Through this approach, we generated new insights into H7 evolutionary constraints and identified novel entry mechanisms that challenge traditional models of influenza host range determination. The experimental framework we established provides a safe and generalizable method for mapping the functional landscape of viral entry proteins, offering valuable tools for assessing pandemic potential across diverse influenza subtypes and other emerging viruses.

In **chapter 2**, we examine the development of an improved pH stability assay capable of measuring how mutations affect HA protein stability under acidic conditions. This enhanced methodology addresses key limitations of traditional approaches by incorporating virus reconcentration steps, enabling high-throughput analysis of HA variants with varying pH activation thresholds. We demonstrate the assay's utility using H5N1 hemagglutinin, successfully identifying both stabilizing and destabilizing mutations that influence viral environmental persistence and transmission potential.

In **chapter 3**, we present the first comprehensive deep mutational scanning analysis of H7 hemagglutinin, systematically measuring the functional effects of nearly every possible amino acid mutation on viral cell entry. Our analysis reveals that H7 can efficiently bind both human-type α2-6 and avian-type α2-3 linked sialic acid receptors, confirming its dual receptor capability. We identify critical functional constraints across different HA domains and demonstrate how specific mutations present in circulating H7N9 strains enhance human receptor binding, providing insights into the virus's pandemic adaptation potential.

In **chapter 4,** we examine how mutational constraints differ among influenza HA subtypes by comparing deep mutational scanning datasets from H3, H5, and H7. Despite sharing highly conserved protein structures and identical functional mechanisms, these HA subtypes display dramatically different tolerance patterns for specific mutations. Our analysis reveals that approximately half of all HA sites show significantly diverged amino acid preferences across subtypes, demonstrating how evolutionary sequence changes can rewire protein constraints even when overall structure and function remain preserved.

Finally, in **chapter 5**, we investigate an alternative entry mechanism through which H7 can utilize Major Histocompatibility Complex class II (MHCII) molecules as cellular receptors. This discovery reveals that influenza entry pathways extend beyond traditional sialic acid-dependent mechanisms and may provide previously unrecognized routes for cross-species transmission. Through systematic mutational analysis, we identify putative MHCII binding interfaces and demonstrate that this alternative entry pathway operates through molecular determinants distinct from canonical receptor binding sites, with important implications for viral host range and pandemic risk assessment

Contents

Acknowledgements

I have been extremely fortunate throughout my PhD to have fantastic mentors, colleagues, and friends. I am certain that without these positive influences in my career, I would not have become the scientist or person that I am today.

Starting graduate school, I was given the advice: “pick the PI and not the project.” I was told that your mentor would have a very large role in your journey throughout your program, and so I endeavoured to take that advice to heart. With that being said, joining Jesse’s lab was one of the best things that I’ve ever done. Jesse is so talented in many ways, whether it’s writing pipelines, helping think of new experiment ideas, or managing to bike home uphill with his kids in the trailer behind him. He continues to amaze me with his ability to be so generous with his time and scientific ideas. As you can imagine, not everything in my graduate school career worked. Even in those times of failure, Jesse was so encouraging and ready to see what we could learn from things that didn’t work. Jesse has really taught me to see the value in all parts of science, the achievements and the setbacks, and for that I am a better scientist. Thank you so much, Jesse.

Being a part of the Bloom lab has also been one of the greatest privileges of my life. Being able to work side by side with some of the brightest scientists and kindest people has been a true highlight of my graduate experience. Thank you to everybody in the Bloom lab for always encouraging me to be a better scientist. I especially would like to thank Bernadeta Dadonaite, who truly taught me everything that I know about influenza and all the techniques that I learned in the Bloom lab (and book recommendations). You amaze me constantly with how efficiently and rigorously you do your work, and I hope to be half as great of a scientist as you one day. Caelan Radford, who also taught me everything I know about DMS, and who was always willing to walk with me to grab food, even if it was far. Andrea Loes for answering many of my questions and keeping the lab running, I wouldn't have been able to do any of this without you. Brendan Larsen provided the most insightful feedback during my talks. Cassie Simoch, for your fun stories of your family and for answering all my medical questions. Sara Sunshine, for always being there for great conversations about science, careers, and, most importantly, books. Sheri Harari was also so encouraging and warm when I was stressed about anything (which was a frequent occurrence). Xiaohui Ju, for being so encouraging about my presentations and experiments. Wenlin Ren, for our fun conversations about food and science. Teagan McMahon, for all of your help around the lab and your fun fishing stories. Gavin Juviler, for keeping the lab running and for the fun lunchtime conversations. Caleb Carr, for being so

generous to share your incubator space. Caroline Kikawa, for getting me into biking and always being there to make me laugh, thank you for being such a great labmate and friend. Andrew Butler, for being our computational wizard and also answering all my bike related questions. Tim Yu, for being the best deskmate and my lab bestie, thank you for being such a great friend and being willing to yap or go get food all the time, I Love Sushi hasn't been the same without you. Lizzie Plender, for being my new deskmate and always being willing to have fun and chaotic conversations. Catherine Jacob-Dolan, for always being my "bike to beers" buddy, outdoor guide, and amazing friend, thank you for being there to take me on new adventures. Lucas Kampman, for also being there to listen to my crazy ideas and also talk some sense into me, thank you for also being an amazing friend. Thank you to everybody for the fun conversations that we've had over the years at lunch. Thank you for entertaining my crazy questions and putting up with me in my stressful times.

Thank you to my mentor at UCLA, Jeffrey H. Miller, thank you for taking me in as an undergraduate researcher who knew nothing about the world of microbiology and research. You gave me the opportunity to run my own projects and were always so encouraging with all of the work that I did. Thank you to my other mentors, Nili Ostrov, Sabrina Tan, and Stephanie Jost who were also so welcoming when joining their labs and so helpful in learning how to be a better scientist.

I also would not have made it this far in my graduate career, much less in life, without the support of my friends. Even if my friends didn't quite understand my work, they were all incredibly supportive and encouraging. When I moved back to Seattle for graduate school, I didn't expect to make any new friends as I would be so busy studying, but I found truly incredible new friends both in and out of the lab. To my high school friends who have always been so supportive of me, thank you Aria, Ashley, Kristina, Jessica, and Kristi. My friends at UCLA, who were always there to study with me and hear me talk about my dreams of doing research and going to graduate school thank you Jane, Connie, and Rachel. The Seattlelite friend group (especially the girls Rachel, Nancy, Ellen, Allysia, Carissa, and Mai) for being so welcoming and supportive no matter what, I wouldn't have been able to do this without our dinners and hangouts. The Boston friend group, for always knowing how to make me laugh, feeding me, and for being so supportive, thank you Sean, Kiki, Pam, Kelia, Trevor, and Jason.

Thank you to my family who has always been extremely supportive of my dreams since I was young. My parents truly gave everything to move to the US when I was young, and they became successful business owners on their own in face of the challenges that come with living in a new country with a new language. They've taught me the importance of hard work and having a supportive community around you. Despite not fully understanding what I do, they've always been so supportive of me doing something I'm so passionate about. 엄마 아빠, 언제나 든든하게 지지해 주셔서 감사합니다. To my sister Julie Ahn who has always been there for me, whether it be a random call, errand day, or family events, thank you for being my best friend from day one. You inspire me with your kindness and generosity. The Leung-Perez

family, who have always had confidence in my abilities and supported my career without a doubt, thank you Anna, Joey, and Katie.

Finally, thank you to my partner Will Perez. I am so thankful that we met, even if it was outside in the sub-freezing temperatures of Boston's winter. Thank you for being my everything, my loudest cheerleader, my favorite travel partner, and my favorite person to talk to about everything and anything with. I can't put into words how grateful I am for you and your support. I truly would not have made it through this PhD without you and I am excited to continue our adventures together!

Chapter 1

1 Introduction

Influenza A viruses represent one of the most successful pathogens in nature, capable of infecting a remarkable diversity of host species while maintaining the constant threat of pandemic emergence. The ability of these viruses to jump between species, reassort genetic material, and rapidly evolve in response to immune pressure has made them a persistent challenge for global public health. Central to influenza's success is the surface glycoprotein hemagglutinin (HA), which mediates both viral entry into host cells and serves as the primary target for neutralizing antibodies. Understanding how HA functions, evolves, and adapts to new environments is therefore critical for predicting pandemic potential and developing effective countermeasures.

The hemagglutinin protein exemplifies the complex relationship between protein structure, function, and evolution. Despite sharing highly conserved structural architectures and functional mechanisms across subtypes, HA proteins from different influenza strains can differ by as much as 60% in their amino acid sequences. This remarkable sequence diversity occurs within the constraints of maintaining essential functions including receptor binding, membrane fusion, and viral entry. Recent technological advances in deep mutational scanning have made it possible to systematically map how thousands of individual mutations affect HA function, providing unprecedented insights into the molecular constraints that govern influenza evolution.

In this thesis, we will examine the molecular determinants of HA function and evolution through comprehensive mutational analysis, with particular focus on the H7 subtype that poses a significant pandemic threat due to its ability to cause severe disease in humans. The work presented here combines methodological innovation with fundamental discoveries about influenza biology, revealing both conserved principles of HA evolution and surprising diversity in entry mechanisms that challenge our current understanding of influenza host range.

The thesis will begin with the development of improved experimental methods for measuring HA stability, a critical property that influences viral transmission and environmental persistence. We will introduce a modified pH stability assay that overcomes limitations of traditional approaches and enables high-throughput characterization of HA variants. Next, we present the first comprehensive deep mutational scanning analysis of H7 HA, systematically measuring how nearly every possible amino acid substitution affects viral entry function. This

dataset reveals the functional landscape of a pandemic-relevant influenza subtype and provides insights into its evolutionary potential.

We then explore how sequence constraints differ across influenza subtypes by comparing mutational effects between H3, H5, and H7 hemagglutinins. Despite their structural similarity, these proteins show surprising differences in their tolerance for specific mutations, revealing how evolutionary divergence can rewire protein function even when overall structure remains conserved. The analysis demonstrates that proteins with nearly identical folds can be subject to dramatically different site-specific evolutionary pressures.

Finally, we investigate an entirely alternative entry mechanism through which H7 HA can utilize Major Histocompatibility Complex class II (MHCII) molecules as cellular receptors. This discovery challenges the traditional view that influenza entry depends solely on sialic acid receptors and reveals a potentially underappreciated pathway for cross-species transmission. Through systematic mutational analysis, we map the molecular requirements for MHCII-mediated entry and identify putative binding interfaces that differ dramatically from canonical receptor binding sites.

Together, these studies provide a comprehensive view of HA functional constraints and evolutionary dynamics, with important implications for pandemic preparedness, vaccine development, and our understanding of how viruses adapt to new hosts. The work demonstrates how systematic experimental approaches can reveal hidden complexity in well-studied biological systems and highlights the importance of considering alternative pathways when assessing viral threat potential.

1.1 Influenza Virus

Influenza viruses are single stranded RNA viruses that are members of the *Orthomyxoviridae* family. There are influenza A, B, C, and D viruses. While influenza A is most notable for causing human pandemics (Uyeki et al. 2022), influenza B and C are both capable of causing disease in humans. These viruses are additionally classified into antigenic subtypes based on thor hemagglutinin (HA) and neuraminidase (NA) envelope glycoproteins. For the purpose of this thesis, we will be focusing solely on influenza A viruses.

1.1.1 Genetics

Influenza viruses are categorized by segmented, negative-strand RNA genomes. Influenza A and B viruses are made up of 8 negative sense, single stranded viral RNA segments. These segments include; PB2, PB1, PA, HA, NP, NA, M, and NS (Bouvier and Palese 2008). This segmented genome enables antigenic shift, where an influenza strain can acquire an HA segment (or NA segment) from a different subtype of influenza virus(Uyeki et al. 2022). As previously mentioned, these viruses are further categorized into subtypes based on the genetic

divergence of the HA and NA surface glycoproteins. There are currently 19 HA subtypes identified, and 11 NA subtypes identified (Er 2025).

1.1.2 Virion Structure

Influenza virion particles are spherical or filamentous in shape (Bouvier and Palese 2008). The spherical forms are on the order of 100nm in diameter and the filamentous forms are typically around 300nm in length. The virions are studded with the primary surface glycoproteins HA (hemagglutinin) and NA (neuraminidase) at a ratio of ~4:1 (Samji 2009; Bouvier and Palese 2008). There are a smaller number of matrix (M2) ion channels that cross the lipid envelope, with the M2:HA ratio being around one M2 channel per 10^1-10^2 HA molecules (Zebedee and Lamb 1988).The envelope, including the 3 integral proteins HA, NA, and M2, overlay a matrix of M1 proteins which enclose the virion core. Inside, the M1 matrix is the nuclear export protein (NEP, also known as NS2, nonstructural protein) and the ribonucleoprotein (RNP) complex which is made up of the viral RNA segments covered with nucleoprotein (NP) and the RNA-dependent polymerase, made up of two polymerase basic (PB1 and PB2) and one polymerase acidic (PA).

The genome structure of influenza A and B viruses consists of eight negative-sense, single stranded viral (vRNA) segments (influenza C has seven segments). The eight segments of influenza A and B are numbered in ordering decreasing length. In these virus segments, 1, 3, 4, and 5 encode for the following proteins PB2, PA, HA, and NP. All viruses encode the polymerase subunit PB1 on segment 2, but in some strains of influenza A virus, this segment also codes for the accessory protein PB1-F2. Segment 6 of influenza A virus encodes for only the NA protein, while B viruses encode both the NA protein and, in a -1 alternate reading frame, the NB matrix protein. Segment 7 in both A and B viruses code for the M1 matrix protein. In the influenza A genome, the M2 ion channel is expressed from segment 7 by RNA splicing (Lamb et al. 1981). Both A and B viruses also possess segment 8, which is a single RNA segment which will express the interferon antagonist NS1 protein (Dauber et al. 2004; García-Sastre 2001; Kochs et al. 2007) and via mRNA splicing, the NEP/NS2 (Briedis and Lamb 1982; Lamb et al. 1981). THe NEP/NS2 is involved in viral RNP export from the host cell nucleus. The ends of each vRNA segment form a helical hairpin, bound by the heterotrimeric RNA polymerase complex. The rest of the segment is covered with NP and the positive charge binds the negatively charged phosphate backbone of the vRNA (Baudin et al. 1994; Compans et al. 1972; Murti et al. 1988). Each vRNA segment contains noncoding regions at both 3′ and 5′ ends

1.1.3 Replication cycle

Influenza viruses require entering the host cell to replicate. These viruses commonly enter cells via sialic acid receptors that are expressed on host cells. The viral life cycle is detailed below (G Bhanu Prakash Animated Medical Videos 2019). Recently, there has been evidence that some types of influenza are also able to enter via MHCII receptors.

Attachment and Entry

Influenza viruses require entering the host cell to replicate. The virus initiates infection by attaching to host cells via sialic acid receptors through its hemagglutinin (HA) protein. Following attachment, the virus is endocytosed and becomes encapsulated within the host cell cytoplasm. The endocytic vesicle containing the virus then fuses with the lysosome, an organelle that contains digestive enzymes and maintains an acidic interior environment. While lysosomes usually digest contents taken in from outside the cell, influenza has evolved to exploit this acidic environment for its own replication purposes.

Uncoating Process

Influenza uses the lysosome's acidic environment to its advantage during the uncoating process. The virus contains an ion channel that allows protons to enter the viral particle, creating acidic conditions within the virus itself. This internal acidity disrupts protein-protein interactions, causing matrix proteins to detach from the proteins covering the viral RNA genome. This detachment marks the beginning of the uncoating process. Simultaneously, the low pH environment triggers HA to undergo a dramatic structural change that allows it to insert into the vesicle membrane. In this new conformational state, HA stimulates membrane fusion between the viral envelope and the endocytic vesicle. The RNA genome and its associated proteins can then float freely away from the matrix proteins into the host cell cytoplasm, completing the uncoating process.

RNA and Protein Production

The viral genome consists of eight RNA segments that are covered in nucleoproteins (NP) and attached to several other proteins comprising the viral RNA polymerase complex. These RNA-protein units enter the host cell nucleus through nuclear pores to begin replication. The RNA segments exist as negative-sense RNA strands, which are non-coding and must be converted to complementary positive-sense strands by the viral RNA polymerase before they can be used for protein translation. To produce functional mRNA for translation by host ribosomes, the virus must add a 5' cap structure to the beginning of each mRNA transcript. The virus accomplishes this through a process called cap snatching, where the viral RNA polymerase cuts off the cap from one of the host cell's own mRNA molecules and uses this stolen cap to initiate transcription of viral RNA. The newly synthesized positive-strand viral RNA receives a poly-A tail that enhances mRNA stability. This viral mRNA then leaves the nucleus and enters the cytoplasm, where host cell ribosomes translate it into viral proteins. Envelope proteins are specifically synthesized by ribosomes located on the host cell's endoplasmic reticulum. Meanwhile, back in the nucleus, newly produced NP proteins prime and stabilize viral RNA as they are being synthesized into full-length positive strands that lack both cap and poly-A tail structures. Newly made RNA polymerase proteins enter the nucleus and participate in producing additional full-length negative strands that will serve as RNA genome segments for new viral particles.

Assembly and Release

Through these coordinated processes, ten different viral proteins are synthesized from the eight RNA segments, and new negative-strand RNA genome segments are produced for progeny viruses. The RNA genome segments, now complexed with RNA polymerase, NP, matrix proteins, and packaging proteins, are exported from the nucleus to the cytoplasm. Concurrently, envelope proteins that were synthesized on the endoplasmic reticulum move through transport vesicles from the ER to the Golgi apparatus and finally to the plasma membrane, where they become embedded. The RNA-protein complexes are then packaged into progeny viruses as they bud from the host cell membrane, completing the viral replication cycle and releasing new infectious particles that can infect additional cells.

1.1.4 Antigenic shift and drift

The segmented genome that influenza viruses possess enables antigenic shift, where influenza A virus strain can acquire the HA segment, and in addition possibly the NA segment, form an influenza virus of a different subtype. This reassortment can happen in cells that were infected with different human and animal viruses. The resulting virus could then encode completely different antigenic proteins to which there is limited preexisting immunity in the human population. Pandemic influenza strains can then arise when antigenic shift creates a virus in which humans have limited immunity and are susceptible. This shift was likely the cause of the 1918-1919 H1N1 influenza outbreak that caused millions of deaths worldwide.

Influenza viruses are also subject to antigenic drift; this consists of small changes (or mutations) in the influenza gene segments that can lead to changes in the surface proteins HA and NA. HA and NA, being the primary surface glycoproteins of influenza are also antigens. This means that they are recognized by the immune system and can trigger immune responses such as antibody production to fight off infections. The small changes associated with drift happen consistently over time as the virus replicates. Vaccines and antibody treatments are typically designed to target these surface proteins, however, the small changes in HA and NA can accumulate over time to result in viruses that are antigenically different from each other. This means that a person's antibodies will bind differently to this virus or not bind the virus at all meaning that there is limited protection against that specific influenza virus (CDC 2026).

1.1.5 Avian influenza virus prevalence

Avian influenza viruses represent one of the most significant pandemic threats facing global public health today. These viruses naturally circulate in wild bird populations worldwide, with periodic spillover events causing outbreaks in domestic poultry and sporadic infections in humans. Among the 16 known hemagglutinin subtypes, H5 and H7 viruses have emerged as particular concerns due to their ability to cause severe disease in both birds and humans, their demonstrated capacity for cross-species transmission, and their potential for further adaptation to mammalian hosts.

H5N1 viruses have caused widespread outbreaks in poultry since the late 1990s and have infected over 900 humans with a case fatality rate exceeding 50% (CDC 2025). More recently, H5N1 has expanded its host range to include marine mammals and has become endemic in North American dairy cattle (Lee et al. 2026), raising concerns about increased mammalian adaptation. H7 viruses pose a parallel threat, with H7N9 causing over 1,500 human infections since 2013 with a case fatality rate around 40% (Ke et al. 2017). Both subtypes have demonstrated concerning adaptive mutations that enhance binding to human-type receptors and increase their pandemic potential (Dadonaite et al. 2024; Ahn et al. 2026).

The continued circulation of these viruses in animal reservoirs, combined with their proven ability to infect humans and cause severe disease, makes H5 and H7 subtypes priority targets for pandemic preparedness efforts. Understanding their molecular constraints, evolutionary potential, and host adaptation mechanisms is critical for developing effective surveillance strategies, risk assessment frameworks, and countermeasures against future pandemic emergence.

1.2 Deep Mutational Scanning

Deep mutational scanning (DMS) is a high-throughput method for generating large libraries that contain genotype-to-phenotype links between the mutation and the outcome (Fowler et al. 2010). DMS involves creating a large library of different protein mutations, putting the library under a selective pressure, and then deep sequencing the library makeup prior to and post selection to quantify how much each mutational variant affects the phenotype.

Some of the earliest DMS studies on viral proteins were done on influenza HA (Thyagarajan and Bloom 2014; Wu et al. 2014). These initial studies used reverse genetics to produce libraries of viruses that contained the HA mutations in addition to the 7 other internal genes from a lab-adapted A/WSN/1933 (H1N1) strain. These saturated mutagenesis libraries created through reverse genetics were important in studying how mutations were tolerated by HA (Doud and Bloom 2016/6; Lee et al. 2018), and how they are able to escape monoclonal antibodies (Doud et al. 2017, 2018; Wu, Thompson, et al. 2020), polyclonal sera (Lee et al. 2019), and how the targets for antibodies against HA varied between age groups (Welsh et al. 2024). Combined mutagenesis libraries also created through reverse genetics were vital in studying the importance of epistasis in the receptor binding pocket (RBP) (Wu, Otwinowski, et al. 2020; Wu et al. 2018, 2017). Due to the replicative nature of these experiments, biosafety concerns were imminent, necessitating the need for a safer approach.

Another type of DMS that is safer uses yeast display. In these experiments a library of peptides with varying mutations are displayed on the surface of yeast. Most prominently, mutagenesis yeast display libraries of the SARS-Cov2 receptor binding domain (RBD) have been used to reveal mutational constraints on RBD folding and receptor binding (Starr, Greaney, Hilton, et al. 2020), and how mutations in the RBD can escape monoclonal antibodies (Cao et al. 2022; Starr, Greaney, Addetia, et al. 2020; Starr et al. 2021) and polyclonal sera (Greaney et al. 2021; Starr,

Greaney, Addetia, et al. 2020), and epistatic shifts (Starr, Greaney, Hannon, et al. 2022). Groups have also created yeast display libraries that contain combinations of different mutations, these libraries helped to reveal the role of epistasis in RBD receptor binding and antibody binding (Moulana et al. 2022, 2023). The limitations of this method arise in the size of the peptide that can be displayed. This is one of the reasons why the experiments are limited to RBD as the full SARS-CoV2 spike protein or the HA protein are too large. These experiments are also limited to measuring mutation effects on biochemical phenotypes, meaning that it's possible to measure antibody binding effects but not neutralization.

The final type of DMS experiments that we will describe is based on lentiviral pseudotyping (Dadonaite, Brown, et al. 2023). The technical innovation comes from conditionally replicative lentiviral particles that can only undergo a single cycle of infection and can be pseudotyped with mutant viral entry proteins. The advantages of this system are that they are safe, can accommodate any viral entry protein, and can be used to study how mutations affect infection rather than just biochemical phenotypes. So far, these libraries have been used to study the functional and antigenic effects of mutations to a diverse set of viral entry proteins including SARS-Cov2 Spike, HIV envelope, nipah glycoprotein, lassa glycoprotein, rabies glycoprotein, chikungunya virus entry protein, RSV F protein, and influenza HA.

1.4 Overview of the thesis

The goal of this thesis is to study how HA mutational effects can inform on viral protein evolution. In **chapter 2**, we talk about the development of a high-throughput pH assay to measure the stability effects of influenza HA. Using this system, we are able to validate previously known acid stabilizing mutations and identify new mutations.

In **chapter 3**, we perform pseudovirus deep mutational scanning on the H7 protein to discover how different mutations can affect cell entry via human receptors vs. avian receptors. Using this system, we are able to identify mutations that increase entry on cells expressing human sialic acid receptors. We analyze the presence of such mutations in current circulating strains and we provide evidence of adaptive mutations for human receptor binding.

In **chapter 4**, we compare pseudovirus deep mutational scanning datasets to study how mutations affect cell entry across different HA subtypes: human H3, avian H5, and avian H7. These HAs share ~40% amino acid identity yet their structures are quite similar. We find that despite sharing extremely similar sequences, these subtypes have dramatically different mutational constraints.

Finally in **chapter 5,** we perform pseudovirus deep mutational scanning to study the usage of the MHCII receptor in influenza H7N9 binding.

Chapter 2

A version of this chapter is published as:

Dadonaite, Bernadeta, Jenny J. Ahn, Jordan T. Ort, Jin Yu, Colleen Furey, Annie Dosey, William W. Hannon, et al. 2024. "**Deep Mutational Scanning of H5 Hemagglutinin to Inform Influenza Virus Surveillance**." *PLoS Biology* 22 (11): e3002916. https://doi.org/10.1371/journal.pbio.3002916.

2 Development of a pH Stability Assay for Measuring HA Protein Stability

2.1 Abstract

Influenza hemagglutinin (HA) undergoes pH-triggered conformational changes essential for viral membrane fusion and infectivity. Traditional pH stability assays require extensive virus dilution post-treatment, limiting their application to high-titer samples and hindering large-scale mutational studies. We developed an improved pH stability assay that incorporates virus reconcentration and neutralization steps, enabling measurement of HA stability in lower-titer samples suitable for deep mutational scanning (DMS) applications. Using H5N1 hemagglutinin as a model system, we validated the assay with citric acid buffers at pH 7.4, 6.0, and 5.5, measuring fractional infectivity through luciferase reporter activity. The H5N1 wild-type showed characteristic pH-dependent inactivation, with complete loss of infectivity at pH 5.5, consistent with reported activation thresholds for emerging H5N1 strains. SARS-CoV-2 Spike BA.2 served as a negative control, maintaining infectivity across all pH conditions and confirming HA-specific inactivation. The assay successfully distinguished known pH stability variants: the destabilizing mutant N198K showed premature inactivation at pH 6.0, while the stabilizing mutant L434I retained full infectivity even at pH 5.5 where wild-type virus was completely inactivated. These results demonstrate that our modified assay preserves biological relevance while improving practical applicability for high-throughput characterization of HA stability. This approach provides a robust platform for systematic investigation of HA structure-function relationships and has potential applications in vaccine development and pandemic preparedness efforts.

2.2 Introduction

Influenza A viruses (IAV) commonly circulate in a variety of wild animals (i.e. seals, minks, whales), aquatic birds, domestic animals (i.e. poultry, horses, swine), and humans. The ability of IAV to infect and spread amongst different species depends on a variety of different factors. During the influenza lifecycle, the virus is able to enter the host cells via the surface

glycoprotein hemagglutinin (HA). HA binds the sialic acid (SA) receptor expressed on the surface of host cells (Rogers and Paulson 1983; Rogers et al. 1983) and this causes membrane fusion in the endosome. Mature HA proteins, HA's cleaved into HA1/HA2 complexes, are trapped in a metastable conformation that is then triggered by low pH in endosomes to undergo an irreversible structural change that ultimately results in membrane fusion (Carr and Kim 1993; Carr et al. 1997). HA stability, meaning the ability of the HA to resist this structural change induced by low pH (or heating) has recently been shown to be an important adaptation of influenza viruses to humans and ferrets (Zaraket et al. 2013-9; Russell et al. 2018). The pH values that trigger this conformational change range from 4.8-6.2, with human-adapted strains having lower pH stability values (4.8-5.5) (Russell et al. 2018; Scholtissek 3/1985; Galloway et al. 2013). A stabilized HA protein, meaning one with a lower activation pH value, could have prolonged environmental stability and may be able to resist virion inactivation in mildly acidic extracellular environments (Poulson et al. 2016; Russier et al. 2020; Labadie et al. 2018). Additionally, the upper respiratory tract of humans are mildly acidic, meaning that upon infection a stabilized HA could enhance intra-host spread of infection by avoiding extracellular inactivation of virions (Russell et al. 2018). Within the cell, intermediate HA stability may also play a role (Singanayagam et al. 2019) in minimizing interferon responses triggered by dendritic cells by high-pH HA protein in the early endosome (Russier et al. 2020) or by interactions between low-pH HA proteins IFITM2 and IFITM3 proteins in the late endosomes (Gerlach et al. 2017).

With the increasing risks of pandemics, the need for better biosurveillance to prevent future pandemics has never been greater. The identification of high pathogenicity strains of avian influenza (HPAI) with the ability to sporadically infect humans has recently raised concerns regarding their potential to trigger a pandemic. When further evaluating the pandemic potential of influenza A viruses, risk assessment tools that analyze HA receptor-binding specificity are commonly used by the WHO and CDC (CDC 2024; "Tool for Influenza Pandemic Risk Assessment (TIPRA)," n.d.). The addition of considering HA stability in these existing algorithms could potentially enhance their predictive abilities (Long et al. 2019; Lipsitch et al. 2016; Russell et al. 2018). A challenge to adding HA stability to surveillance stems from the idea that the stability needs to be measured and cannot be inferred based on the HA gene sequence. There have been developments of several assays to measure HA stability, although each can be time-consuming, strain- dependent, and cumbersome which is not well suited for high-throughput measurements of this property. Here we discuss the development of a pH-based stability assay capable of measuring the stability effects of thousands of single mutations in a single experiment. We are then able to test and validate this system to use with our H5 DMS libraries and identify existing and new stability enhancing mutations.

2.3 Results

2.3.1 Development of pH stability assay

Membrane fusion of influenza is regulated by the conformational state of the influenza HA protein. The HA trimer natively exists in a prefusion structure, this structure represents a high-energy metastable configuration of the HA. Upon exposure to acidic pH's, typically in the cellular endosome, the HA protein undergoes an irreversible conformational change that ultimately results in membrane fusion (Carr and Kim 1993; Skehel and Wiley 2000; Ruigrok et al. 1988; Weber et al. 1994; Wharton et al. 1995). If the virion is exposed to sufficiently low pH outside of the host or host cell, the HA protein can be prematurely activated to refold irreversibly such that the virion is inactivated. This loss of viral infectivity at lower pH is widely used as a proxy to measure HA stability (W. Yang et al. 2021; Poulson et al. 2016; Russier et al. 2020). While the biological trigger for HA activation is a lower pH, HA refolding can also be triggered by other destabilizing agents such as urea and heat (Carr et al. 1997; Ruigrok et al. 1986)

Traditional pH stability assays required diluting virus samples after pH treatment to restore neutral conditions before adding onto the cells (G. Yang et al. 2021), ensuring cell viability during pH exposure. However, these methods demanded large amounts of starting virus, limiting their use to high-titer samples, and making it difficult to adapt to our large scale library system. To address these limitations, we developed an improved pH stability assay that would work with our DMS libraries. Our approach treats viruses with citric acid buffers across a range of pH values, reconcentrates and neutralizes the pH with PBS, and then the pH treated virus is put onto the cells. We initially validated this system using a luciferase-based reporter, where viruses encoded both the HA variant of interest and a luciferase gene for detection.

During the assay development, we used the HA from the A/American Wigeon/South Carolina/USDA-000345-001/2021 (H5N1)as our WT, an H5 pH destabilized variant N198K (N186K, (Timofeeva et al. 2020), an H5 pH stabilized variant L434I (L419, (Wang et al. 2018), and the SARS-CoV2 Spike BA.2 variant as control. HA inactivation values tend to fall within the range of pH 4.8-6.2 (Mair et al. 2014) but the average HA activation pH values was around 5.7 for emerging H5N1 strains (Zaraket et al. 2013-9). We first tested a range of pH buffers with the H5 WT to see when the WT would be inactivated (figure 1c). Based on the destabilizing pH values, we chose a range of pH values to test with our viruses, neutral media pH (~7.4), 6.0, and 5.5.

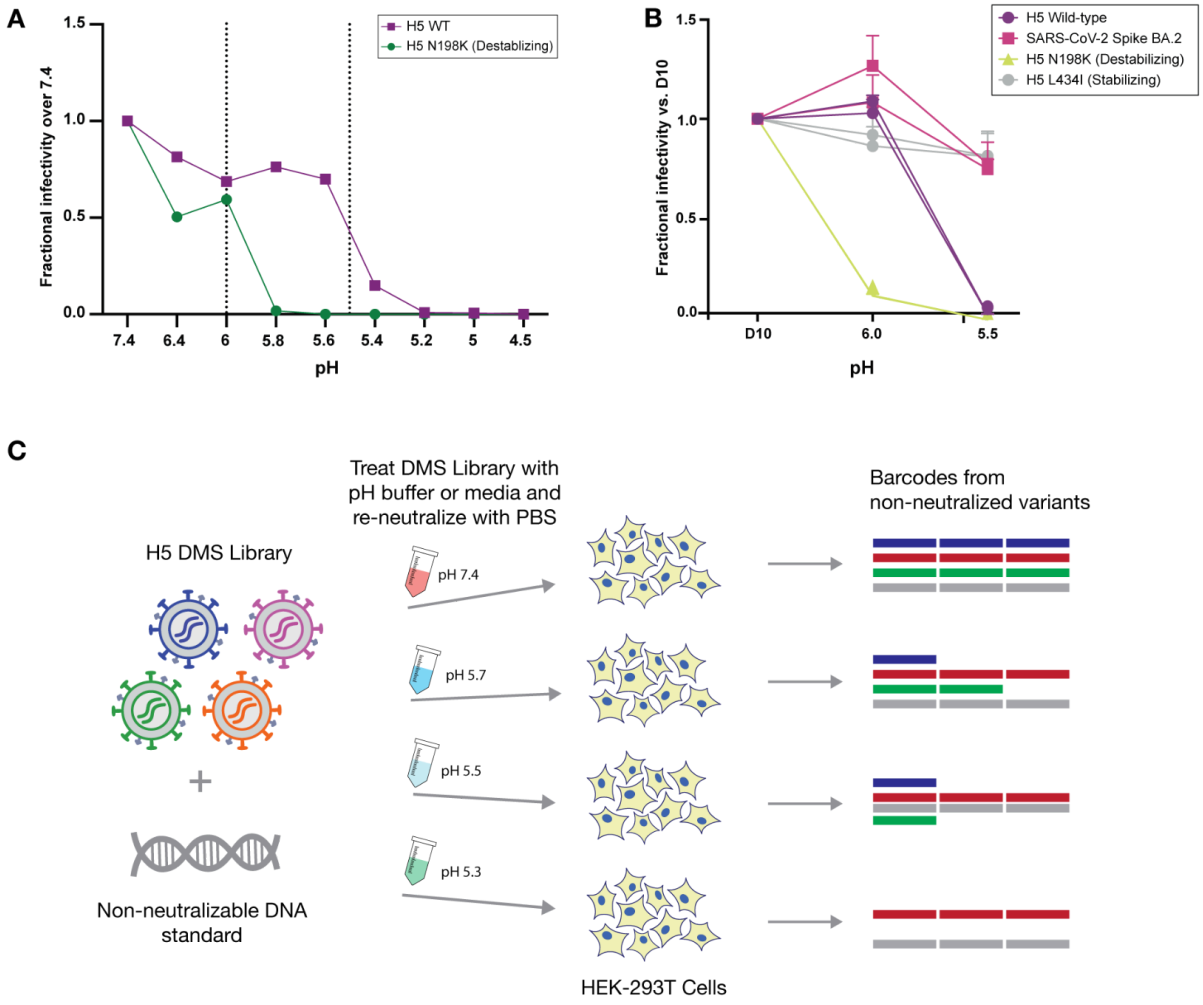


Figure 2.1 | pH Stability pipeline development and validation.
A) Graph showing the pH buffer range that was tested using H5 WT (purple) and N198K destabilizing mutation (green) to determine the range of buffer to use for the final H5 pH DMS selections. **B)** Graph showing results from treating H5 WT (purple), SARS-CoV-2 Spike BA.2 (pink), H5 N198K destabilizing mutation (yellow), and H5 L434I stabilizing mutation (grey) with a range of pH buffers and re-concentrating the virus prior to putting them on cells. **C)** Diagram detailing the pH stability pipeline. First the libraries are treated with a range of citrate buffers of the media control. After treatment viruses are re-concentrated and neutralized with PBS. After neutralization, viruses are concentrated again and then put onto the target cells. The barcodes from the non-neutralized variants are then read out. Stability was quantified as the resistance of each mutant to inactivation by pH.

To measure the pH of virion inactivation, we calculated the target RLU readouts (~5M) for each virus and brought up the volume of the virus to 0.5ml D10. We then added a larger volume of either D10 or each of the pH adjusted citric acid buffers. The viruses were treated with the

buffers or D10 for 1 hour at 37˚. After incubation, all the contents were spun down through a protein concentrator. After the spin down, the spin through liquid was discarded and the virus concentrated in the filter was topped off with a larger volume of PBS to re-neutralize the viruses. Viruses were spun down again at the same speed and time and the virus was then collected. Each virus was brought up to a final volume of ~2ml in D10 and 100µl of pH treated virus was then added onto the cells.

In this assay we measured the fractional infectivity (FI), meaning that this was the proportion of virus that was able to infect the cells over the non-pH treated control conditions. For the Spike BA.2 variant, we saw that despite the lower pH treatments, the FI remained consistent around 1. This meant that infectivity seen at the lower pH treatments remained similar to the un-treated (**Fig 1B**). However, when looking at the H5 WT, we see that the FI starts off around 1 at the D10 pH and 6.0, but as you get to pH 5.5 the FI drops to 0. We also saw the expected results with our stabilizing and destabilizing mutants. The N198K (destabilizing) mutant had a FI of around 1 at the neutral D10 pH, but as the pH dropped, the FI dropped at pH 6.0, indicating a loss of infectivity at this pH. For the L434I (stabilizing) mutant, we saw that despite the lower pH treatments, the FI remained consistent around 1, similar to the BA.2 variant. This stabilizing mutant, retained its fractional infectivity even at pH 5.5 when the WT FI dropped to 0 a pH 5.5.

2.3.2 Effects of mutations on H5 pH Stability

Mutations that increase HA stability are often associated with airborne transmissibility of influenza viruses (Linster et al. 2014; Sun et al. 2024; Mair et al. 2014; Zaraket et al. 2013-9; Poulson et al. 2016). To identify mutations that affect HA's stability, we incubated the pseudovirus libraries in progressively more acidic pH buffers and quantified the retention of infectivity as a measure of HA stability.

Mutations at a variety of sites increase HA stability (**Fig 2.2**) and an interactive plot at (https://dms-vep.org/Flu_H5_American-Wigeon_South-Carolina_2021-H5N1_DMS/stability.html). Although the stabilizing mutations are widely distributed among the primary sequence of the HA protein, they are concentrated in 2 regions of HA's three-dimensional structure: the stem's alpha helices the base of the head domain. The spatial location of these mutations is consistent with the fact that the acidic conditions or high temperatures can trigger that irreversible conformational change that involves the displacement of the head domain (Carr and Kim 1993; Xu et al. 2010). Overall the deep mutational scanning both identified previously known stabilizing mutations (ie. Y17H, A19T, H24Q, E31K, H110Y, and T318I) (Imai et al. 2012; Reed et al. 2010; Hanson et al. 2015; Russier et al. 2016; Singanayagam et al. 2019) as well as many additional mutations that increase stability.

We were also able to validate the deep mutational scanning stability measurements by generating conditionally replicative influenza virions (**Fig S1A**) carrying key mutations and quantifying their infectivity after acid treatment. The deep mutational scanning measurements correlate well with the stability of the influenza virions (**Fig S1A**), further validating the approach. We note that the map of the stabilizing mutations could inform on the design of stabilized HA immunogens (Milder et al. 2022) for vaccines as well as inform on viral surveillance.

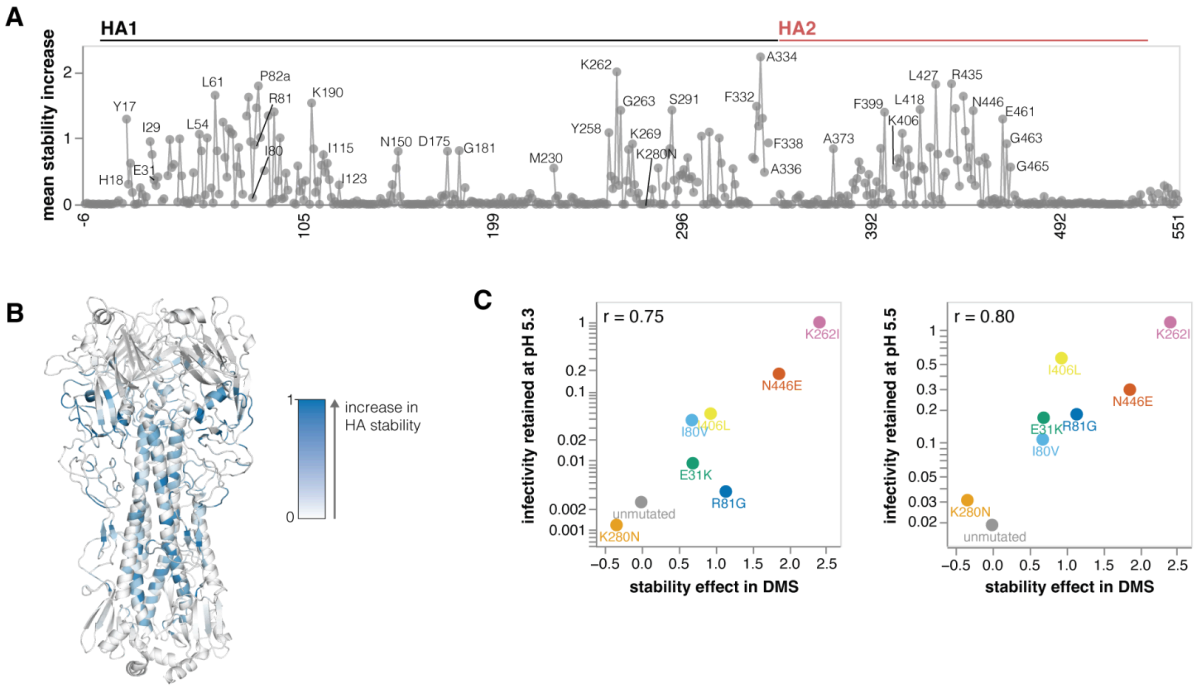


Figure 2.2 | H5 pH stability DMS Data

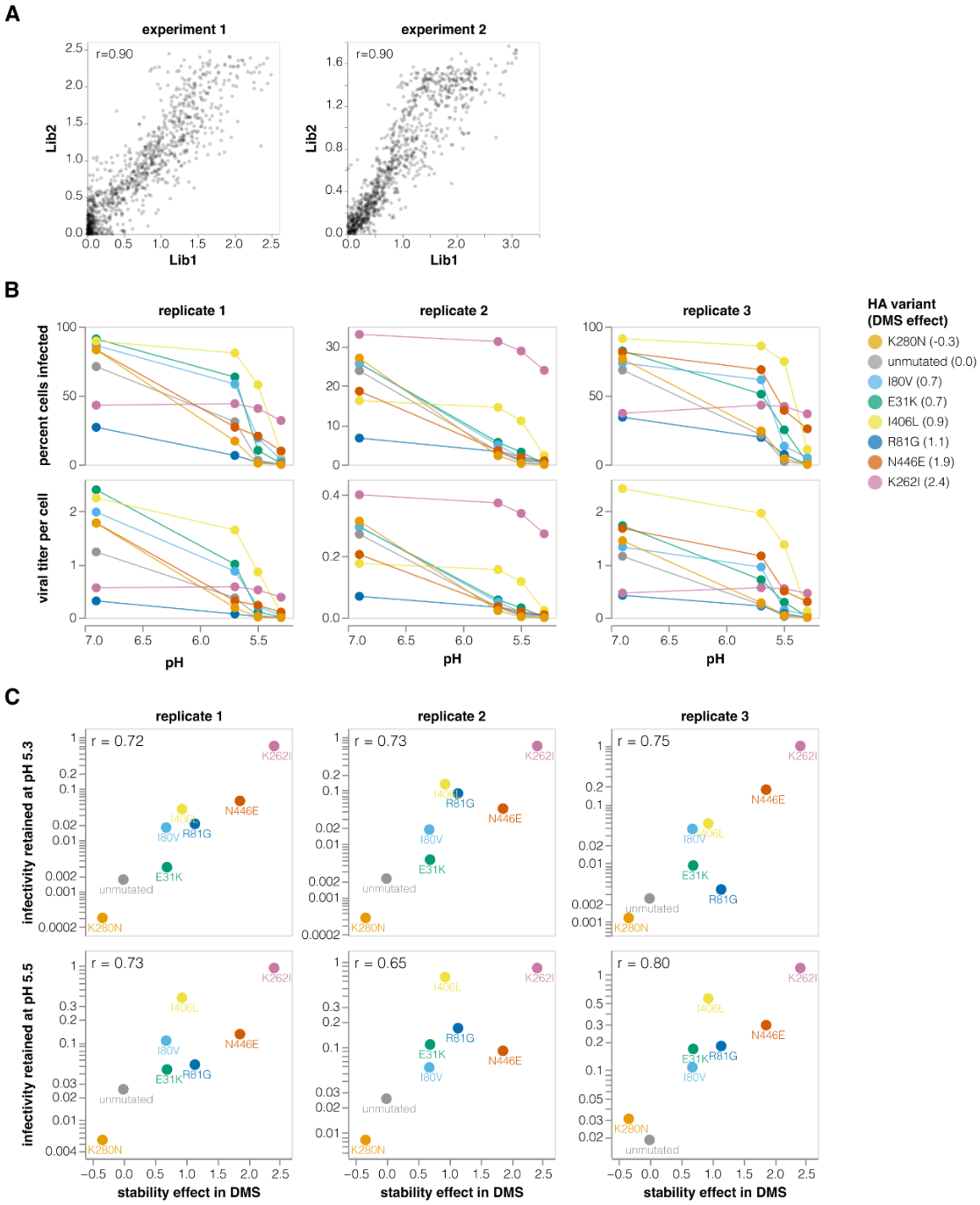
A) The average increase in HA stability caused by mutations at each site. See https://dms-vep.org/Flu_H5_American-Wigeon_South-Carolina_2021-H5N1_DMS/stability.html for an interactive heatmap of the effects of all mutations that include options to show destabilizing as well a stabilizing mutations **B)** HA structure colored by the average increase in HA stability at each site. Sites where mutations increase HA stability are shown in darker blue and sites where there was little to no effect in HA stability increased are shown in light blue or white. **C)** Validation of deep mutational scanning measurements by comparison to direct measurements of acid-induced inactivation of conditionally replicative influenza virions carrying the indicated mutations. The mutations validated here were chosen to span a wide range of effects on stability. The r value in the upper-left of each panel indicates the Pearson correlation.

2.4 Discussion

We successfully developed and validated a modified pH stability assay that addresses key limitations of traditional methods while maintaining biological relevance for measuring HA conformational stability. Our approach eliminates the need for extensive virus dilution by incorporating a reconcentration step, making it suitable for lower-titer samples and high-throughput applications like DMS libraries. The assay effectively distinguished between viruses with different pH stability profiles. As expected, the H5N1 WT showed characteristic pH-dependent inactivation, with complete loss of infectivity at pH 5.5, consistent with reported activation pH values of ~5.7 for emerging H5N1 strains (Zaraket et al. 2013-9). The SARS-CoV-2 Spike BA.2 variant served as an appropriate negative control, maintaining full infectivity across all pH conditions tested, confirming that the observed inactivation was specific to influenza HA-mediated fusion rather than non-specific pH effects on viral particles or cell viability.

Importantly, our assay successfully detected both stabilizing and destabilizing mutations previously identified in literature, in addition to new ones. The N198K destabilizing mutant showed premature inactivation at pH 6.0, demonstrating increased pH sensitivity as predicted from previous studies (Timofeeva et al. 2020). Conversely, the L434I stabilizing mutant retained infectivity even at pH 5.5, where the WT was completely inactivated, confirming enhanced stability consistent with prior characterization (Wang et al. 2018). These results validate that our modified assay preserves the biological signal of HA conformational changes while improving practical applicability. The successful discrimination between known stabilizing and destabilizing variants demonstrates that this assay can reliably measure the functional consequences of HA mutations on pH stability. This capability is particularly valuable for DMS applications, where large libraries of variants need to be systematically characterized. The improved efficiency and reduced virus requirements make this approach well-suited for high-throughput screening of HA stability across diverse mutation backgrounds, providing a robust platform for understanding structure-function relationships in influenza HA and potentially informing vaccine strain selection and pandemic preparedness efforts.

2.5 Supplementary Materials



S1 Fig. Validation of stability enhancing HA mutations
(A) Correlations between the effects of mutations on HA stability as measured by the deep mutational scanning for each of the 2 independent library replicates. Experiments 1 and 2 were performed on different dates. Only stabilizing mutations are shown. The Pearson correlation is shown in the upper left. Throughout the rest of this paper, we report the average effect across

libraries and experiments. The data underlying this panel can be found in S3 Data. (B and C) Stability-enhancing mutations identified in deep mutational scanning were validated using the conditionally replicative influenza virions. **(B)** Conditionally replicative influenza virions carrying each of the indicated mutations were incubated at a buffer of the indicated pH for 30 minutes, diluted into a neutral pH buffer, and then used to infect target cells. The top row shows the percent of cells infected by each virus after treatment at each pH, and the bottom row shows the titer per cell (MOI) calculated using the Poisson distribution. **(C)** The infectivity remaining for each mutant at the indicated pHs relative to pH 6.9 (as calculated from the bottom row of panel (B) versus the stability effects of the mutation measured in the deep mutational scanning. The Pearson correlation is indicated in the upper-left of each plot.

2.6 Materials and methods

Cell lines and media

293T cells were maintained D10 media (Dulbecco's Modified Eagle Medium with 10% heat-inactivated fetal bovine serum, 2 mM l-glutamine, 100 U/mL penicillin, and 100 µg/mL streptomycin). D10 media was used for all experiments involving pseudovirus. MDCK-SIAT1-PB1 and MDCK-SIAT1-PB1-TMPRSS2 cells were passaged in D10 media and for assays described above plated in either Influenza growth media (Opti-MEM supplemented with 0.1% heat-inactivated FBS, 0.3% bovine serum albumin, 100 µg/mL of calcium chloride, 100 U/mL penicillin, and 100 µg/mL streptomycin), or Neutralization assay media (Medium-199 supplemented with 0.01% heat-inactivated FBS, 0.3% BSA, 100 U/mL penicillin, 100 µg/mL streptomycin, 100 µg/mL calcium chloride, and 25mM HEPES). Influenza growth media was used for all conditionally replicative virus rescue experiments, while Neutralization assay media was used for all experiments involving fluorescence readings because riboflavin present in Opti-MEM has high background fluorescence.

Measuring effects of mutations on HA Stability

To measure the effects of mutations on HA stability, we followed the workflow outlined in **(Fig 2.1C**). Specifically, approximately 5 million transcription units of the pseudovirus library were incubated with D10 media or citrate-based acidic buffers. Buffers were prepared at 0.1 M by mixing together sodium citrate dihydrate and citric acid solutions and adjusting to pHs 5.7, 5.5, or 5.3 using hydrochloric acid or sodium hydroxide. Pseudovirus libraries were incubated for 1 hour at 37°C at 1:12 library to buffer volume ratio. After incubation, libraries were then concentrated using 100k Amicon spin columns (Milipore, UFC910008) by spinning for 15 minutes at 1,500g to approximately 1 ml volume and resuspended in 11 ml of PBS to exchange and neutralize acidic buffers. Viruses were spun down again for 20 minutes at 1,500g to reduce sample volume and exchange PBS for 2 ml of D10 media. Oseltamivir was added to the libraries at 1 µM and they were then used to infect 293T cells that were plated a day before in the presence of 2.5 µg/ml of amphotericin B. At 12 to 15 hours postinfection unintegrated, lentiviral genomes were recovered from cells using Qiagen Miniprep Kit. During the miniprep lysis step (P2 buffer), 0.003 ng of DNA standard was added to a final molar concentration of approximately 1% of recovered lentiviral DNA (this was calculated based on the number of expected nonintegrated lentiviral genomes recovered given the amount of virus used for infections). The DNA standard consisted of a plasmid encoding lentiviral backbone with a barcoded mCherry gene (plasmid map for nonneutralizable standard is at https://github.com/dms-vep/Flu_H5_American-Wigeon_South-Carolina_2021-H5N1_DMS/blob/main/library_design/3068_ForInd_mC_BCs_pool1.gb, and barcodes in this standards can be found at https://github.com/dms-vep/Flu_H5_American-Wigeon_South-Carolina_2021-H5N1_DMS/blob/main/data/neutralization_standard_barcodes.csv), and the purpose of this standard is to enable conversion of relative sequencing counts to absolute amount of each barcoded pseudovirus variant (relative to the standard) at each pH condition. Note that unlike for sera selections (described below) where we use VSV-G or RdPro pseudotyped viruses as

nonneutralizable standard, here we choose to use DNA-based standard because VSV-G or RdPro may be affected by the pH treatment. Lentiviral genomes were then used to make amplicon libraries for Illumina sequencing as described previously (Dadonaite, Crawford, et al. 2023).

The sequencing counts were converted to the fraction infectivity retained at each pH condition relative to D10 by using the counts of each variant relative to the DNA standard at each condition. The polyclonal (Yu et al. 2022) software (v6.11) was used to analyze mutation-level retained infectivity at each pH and to fit a pH-dependent neutralization curve while estimating the effect of each mutation on stability. An example notebook of this analysis can be found at https://dms-vep.org/Flu_H5_American-Wigeon_South-Carolina_2021-H5N1_DMS/notebooks/fit_escape_stability_Lib2-231002-pH.html. HA stability data can be viewed interactively in the context of the H5 HA protein structure using dms-viz (Hannon and Bloom 2023) at https://dms-viz.github.io/v0/?data=https%3A%2F%2Fraw.githubusercontent.com%2Fdms-vep%2FFlu_H5_American-Wigeon_South-Carolina_2021-H5N1_DMS%2Fmain%2Fresults%2Fdms-viz%2Fdms-viz.json.

Validation of stability enhancing mutations using conditionally replicative influenza virus
MDCK-SIAT1-PB1 cells were plated in NAM at 100 k per well in a 6-well dish, 4 hours prior to infection. Wild-type and mutant viruses were diluted to the target final MOI of 0.75 to 1 and incubated with citrate buffers at pHs 6.9, 5.7, 5.5, or 5.3 for 30 minutes at 37℃ at a dilution of 4µl virus to 96 µl of buffer. Citrate buffers were prepared at 0.1 M by mixing together sodium citrate dihydrate and citric acid solutions and adjusted to the required pH using hydrochloric acid or sodium hydroxide. pH-treated viruses were then brought up to a neutral pH by 3.9mL of NAM media. About 2mL of the NAM-diluted virus was then added to MDCK-SIAT1-PB1 cells. At 16 to 20 hours postinfection, cells were collected for flow cytometry analysis and the percent positive eGFP cells were measured for each mutant. We conducted single replicates of each mutant and repeated these across 3 different days. The fractional infectivity was then calculated by taking the proportion of the percent positive green cells at pH 5.3
or pH 5.5 over the percent positive green cells at pH 6.9 (S5 Fig).

Production of conditionally replicative influenza viruses
Conditionally replicative influenza viruses with wild-type, mutant, or selected strain HA genes were generated as described previously (Hooper and Bloom 2013; Doud et al. 2018) with several modifications explained here. The native HA sequences (not codon optimized) lacking the polybasic cleavage site for all strains and variants were cloned into the bidirectional pHW2000 influenza reverse genetics plasmid (Hannon and Bloom 2023) (plasmid maps for all strains can be found at
https://github.com/dms-vep/Flu_H5_American-Wigeon_South-Carolina_2021-H5N1_DMS/tree/main/library_design/conditionally_replicative_virus_plasmids). The polybasic cleavage site was removed from the HA sequence to further enhance the safety of this system. The native NA sequence from A/turkey/Indiana/22-003707-003/2022 strain was also cloned into a pHW2000

plasmid (plasmid map https://github.com/dms-vep/Flu_H5_American-Wigeon_South-Carolina_2021-H5N1_DMS/blob/main/library_design/conditionally_replicative_virus_plasmids/3972_pHW_N1_TurkeyH5.gb). A mixture of 5e5 293T-PB1 cells (Bloom et al. 2010) and 50,000 MDCK-SIAT1-PB1-TMPRSS2 cells (Lee et al. 2018) per well was plated into a 6-well plate. The next day, cells were transfected with plasmid mix consisting of A/WSN/1933 PB2, PA, NP, M, and NS reverse genetics plasmids (Hoffmann et al. 2000), pHH-PB1flank-eGFP plasmid that consists of noncoding terminal regions from A/WSN/1933 PB1 segment and eGFP gene in place of PB1 coding sequence (Bloom et al. 2010), pHW_HA_American-Wigeon_USDA-000345-001_H5N1_delPBCS plasmid and pHW_N1_TurkeyH5 plasmid. Next day, cell media was changed to Influenza growth media (see Cell lines methods subsection below) and virus supernatant was collected 48 hours after media change. Virus titers were determined by serially diluting virus stocks and infecting MDCK-SIAT1-PB1 cells in Neutralization Assay Medium (NAM) media. Flow cytometry was used to determine eGFP positive cell numbers, 16 hours postinfection, and Poisson distribution was used to determine virus titers. In cases where high virus titers were needed, rescued viruses were further passaged at low MOI (0.01) in MDCK-SIAT1-PB1-TMPRSS2 cells in NAM. For stability mutation validations, passaged viruses were further concentrated using 100k Amicon spin columns. To avoid any possibility of reassortment between the conditionally replicative influenza viruses with the H5N1 HA and NA and fully replication competent influenza, work with these conditionally replicative viruses was strictly segregated away from tissue-culture incubators and biosafety cabinets where fully replicative seasonal or lab-adapted influenza viruses were being studied.

Neutralization assays using conditionally replicative influenza virus

To validate sera escape mutations, we performed neutralization assays using conditionally replicative virus as described previously (Hooper and Bloom 2013). In brief, 50 k MDCK-SIAT1-PB1 were plated into 96-well plates in NAM media 4 hours before infection. Sera were serially diluted in NAM media and mixed with wild-type or mutant conditionally replicative influenza viruses. The amount of virus to use was chosen such that the green fluorescence signal would change linearly with degree of sera neutralization. Virus and sera were incubated for 45 minutes at 37°C and preplated MDCK-SIAT1-PB1 were infected. Fluorescence signal was determined using Tecan M1000 plate reader, 16 hours postinfection. Fraction infectivity was determined relative to no serum controls and neutralization curves were plotted using neutcurve software (Loes et al. 2024).

Chapter 3

A version of this chapter is published as:

Jenny J Ahn, Timothy C Yu, Bernadeta Dadonaite, Caelan E Radford, Jesse D Bloom, **Influenza hemagglutinin subtypes have different sequence constraints despite sharing extremely similar structures,** *Virus Evolution*, Volume 12, Issue 1, 2026, veag018, https://doi.org/10.1093/ve/veag018

3 Mapping the Functional Landscape of H7 Avian Influenza for Pandemic Preparedness

3.1 Abstract

We performed deep mutational scanning (DMS) on H7 hemagglutinin from the A/Anhui/1/2013 strain, a Eurasian H7N9 virus that serves as the backbone for WHO candidate vaccine viruses. Using pseudoviral libraries containing 95-96% of all possible amino acid mutations in the HA ectodomain, we systematically measured the functional effects of mutations on cell entry into engineered 293 cells expressing α2-3 linked, α2-6 linked, or mixed sialic acid receptors. Our comprehensive mutational analysis revealed distinct patterns of functional constraint across the H7 HA protein. HA2 showed greater sensitivity to mutations than HA1, with particularly strong constraint in the fusion loop region (sites 330-350) critical for membrane fusion. In contrast, many areas of the HA1 globular head demonstrated high mutational tolerance, including regions commonly targeted by neutralizing antibodies. These results provide the first comprehensive mutational landscape for H7 HA and extend previous findings from H3 and H5 subtypes, revealing conserved patterns of functional constraint across influenza A hemagglutinins. This dataset offers critical insights into H7 HA evolutionary potential and identifies both constrained regions suitable for universal vaccine targets and highly mutable areas prone to immune escape.

3.2 Introduction

Influenza has caused several past pandemics due to the reassortment of seasonal and avian viruses, making future occurrences likely. Moreover, the high fatality rates of current avian flu outbreaks are especially concerning, as the virus could potentially spill over into the human population, either directly or through reassortment. Particularly concerning are H5 and H7 which have caused numerous zoonotic outbreaks. H7 viruses have been first reported in humans in 2013 and since then, there have been multiple additional infections with case fatality rates around 40% (Liu et al. 08/2021; Xiang 2016; Ning et al. 2019). HA plays a major role in virus entry, transmission and neutralizing antibody development and therefore a better understanding of the effects of mutations to H7 HA can help us to understand the pandemic potential of this virus and be better prepared in case of its pandemic emergence. Additionally, because antibodies against HA are a primary correlate of protection, they are a key target for

influenza vaccines. Previous studies have shown that mutations in HA not only impact the virus's antigenicity, potentially rendering vaccines and antibody treatments less effective, but can also impact the ability of the virus to infect and spread efficiently between humans (Tanner et al. 2015; Watanabe et al. 2014; Si et al. 2014). With continued sporadic outbreaks of H7N9 globally, and a lack of immunity in the population, the potential for a pandemic remains high. In the event of a pandemic, we currently lack much knowledge necessary to respond effectively in the case of such a scenario.

3.3 Results

3.3.1 H7 HA deep mutational scanning libraries

We have previously described a pseudovirus deep mutational scanning approach that can be used to measure the effects of all individual amino-acid mutations to HA on its ability to mediate the entry of pseudotyped lentiviral particles into cells (Dadonaite, Crawford, et al. 2023) (**Fig. S3.1A,B**), and applied this approach to H5 and H3 HAs (Dadonaite et al. 2024; Yu et al. 2025). Note that this experimental approach uses lentiviral particles that encode no viral protein other than HA, and are therefore only able to undergo a single round of cell entry, this means that they are not fully infectious pathogens capable of causing disease, making them a safe tool to study the effects of mutations to HA. Here we extend this previously described approach to measure the effects of mutations to a H7 HA on cell entry in order to create an expanded dataset to facilitate comparison of sequence constraints across HA subtypes.

For these new experiments, we selected the HA of one of the World Health Organization's candidate H7 vaccine viruses (WHO, n.d.), the Eurasian lineage A/Anhui/1/2013 strain. This strain is from the Eurasian H7N9 lineage that first caused human infections in 2013 in China (Gao et al. 2013; Wen and Klenk 2013). The human infections were traced to closely related viruses from this same lineage circulating in poultry and live bird markets (J. Shi et al. 2013; Lam et al. 2013), with poultry infections identified concurrently with the human cases. The H7 HA from the A/Anhui/1/2013 strain is the background for three current candidate vaccine viruses, NIBRG-268, NIIDRG-10.1, and IDCDC-RG33A (WHO 2025). The HA from this A/Anhui/1/2013 strain has been shown to bind both the α2-3 linked sialic acids that are the typical receptor for avian-adapted influenza viruses and the α2-6 linked sialic acids that are the typical receptor for human-adapted influenza viruses(Y. Shi et al. 2013). The ability of the A/Anhui/1/2013 H7 HA to bind α2-6 linked sialic acids is due primarily to the fact that it has the Q226L mutation (Y. Shi et al. 2013), which promotes binding to α2-6 linked sialic acids in multiple HA subtypes (Stevens et al. 2006; Sun et al. 2024; Bateman et al. 2008).

We created duplicate mutant libraries of this H7 HA in the context of pseudoviruses capable of undergoing only a single round of cell entry. The two libraries contained 95% and 96% of all possible amino-acid mutations to the HA ectodomain, respectively. Most of the barcoded HA

variants in the libraries contained single amino acid mutations, although a small fraction contained no mutations or multiple mutations (**Fig. S3.1C-E**). We measured the ability of each barcoded HA variant to mediate pseudovirus entry into cells using previously described approaches (Dadonaite, Crawford, et al. 2023; Dadonaite et al. 2024; Yu et al. 2025) (**Fig S3.1B**), and used global epistasis models (Haddox et al. 2023; Otwinowski et al. 2018) to determine the effect of each mutation on cell entry. The effect of each mutation is quantified as the $\log_2$ of the cell entry of pseudovirus with that mutant HA relative to the unmutated HA, so mutations that do not impact cell entry are assigned effects of zero while mutations that impair entry are assigned negative effects.

3.3.2 Effects of HA mutations on entry into cells expressing α2-6- versus α2-3-linked sialic acids

We measured entry into previously described 293 cells (Narimatsu et al. 2019; Büll et al. 2021) that have been engineered to express only α2-3 linked sialic acids, only α2-6 linked sialic acids, or an equal mix of cells expressing each of the two types of sialic acids. We performed two experimental replicates for each of the two independent pseudovirus mutant libraries, and the measured effects of the mutations on cell entry were highly correlated across both replicates of the same library and between the two independent libraries (**Fig. S3.2**); throughout this paper we report the median of the measurements across all replicates.

The effects of nearly all mutations to the H7 HA ectodomain on entry into a mix of 293 cells expressing α2-3 linked or α2-6 linked sialic acids are shown in **Fig. 3.1A** (see also the interactive heatmap at https://dms-vep.org/Flu_H7_Anhui13_DMS/cell_entry.html). As can be seen from this figure, there is wide variation in mutational tolerance across the HA protein, with some sites being quite tolerant of mutations whereas at other sites nearly any mutation strongly impairs HA's cell-entry function. Functional constraint was generally higher in HA2 than HA1, especially in regions of HA2 like the fusion loop (sites 330-350) that are critical for mediating HA's cell entry function (**Fig. 3.1A**). However, many regions of HA1, especially near the top of the globular head of the protein, are quite tolerant of mutations (**Fig. 3.1A,B**). Overall, these results are in line with prior deep mutational scanning studies of HA from other influenza A subtypes, which have found the most constraint in regions of HA2 involved in membrane fusion, and higher mutational tolerance in many portions of the HA1 globular head including the epitopes most commonly targeted by neutralizing antibodies (Dadonaite et al. 2024; Wu, Thompson, et al. 2020; Lee et al. 2018; Wu et al. 2014; Thyagarajan and Bloom 2014).

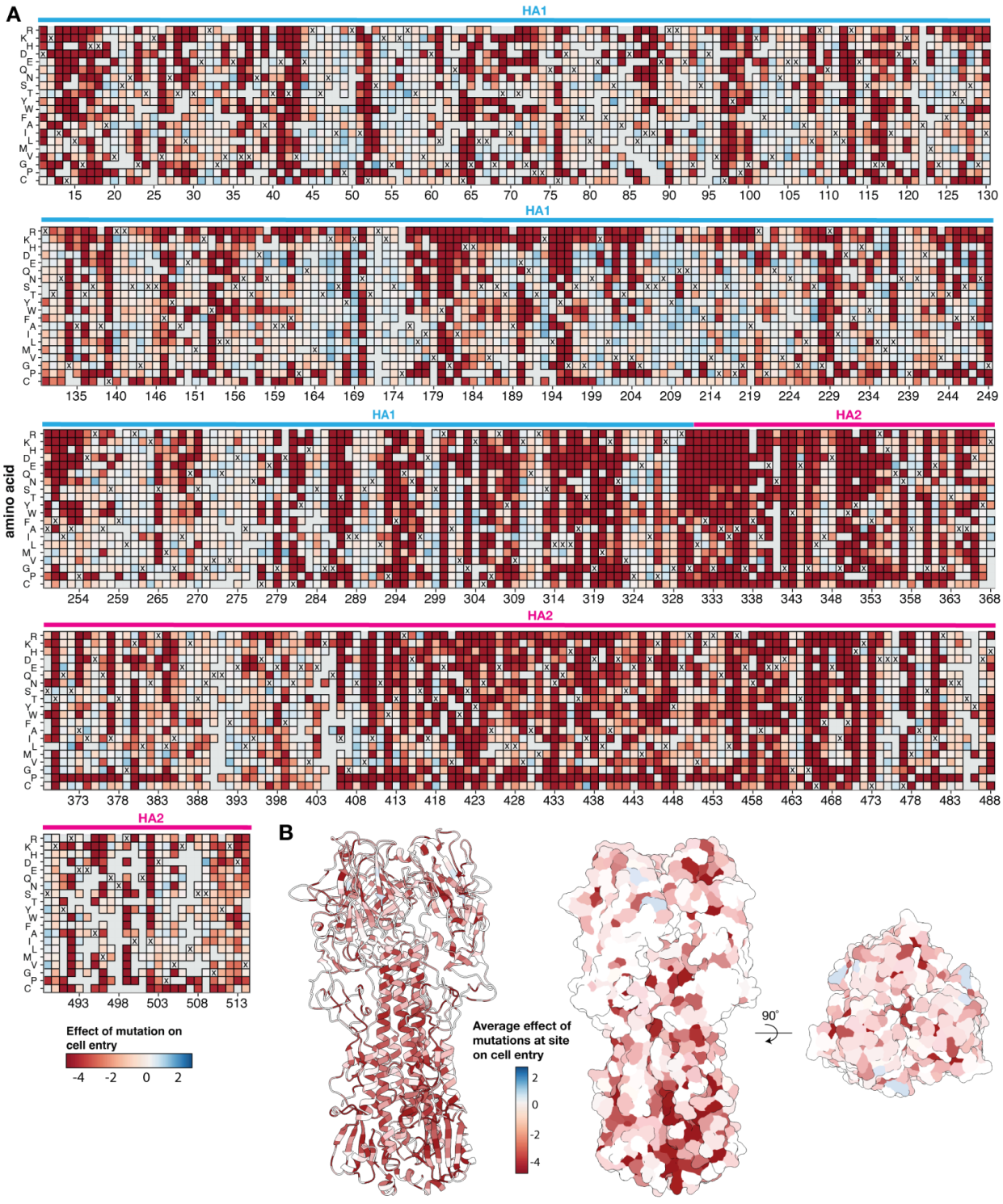


Figure 3.1 | Effects of mutations to the ectodomain of H7 HA on its ability to mediate entry into a mix of cells expressing α2-3 and α2-6 linked sialic acids.

A) Heatmap showing the effects of mutations on HA-mediated cell entry. Negative cell entry effects (red) indicate impaired cell entry relative to the unmutated HA. Gray indicates mutations

that were not reliably measured in our experiments, and for each site the X indicates the amino-acid identity in the unmutated A/Anhui/1/2013 HA. See https://dms-vep.org/Flu_H7_Anhui13_DMS/cell_entry.html for an interactive version of this heatmap that also shows effects of mutations on entry into 293 cells expressing only α2-3 or only α2-6 linked sialic acids. In this figure and throughout the rest of this paper we use H3 sequence numbering(Burke and Smith 2014). **B)** The structure of the A/Anhui/1/2013 H7 HA(Huang et al. 2019) (PDB 6ii9) colored by the mean effects of all mutations at that site on cell entry.

The effects of H7 HA mutations on entry into 293 cells expressing only α2-3 linked or only α2-6 linked sialic acids were very similar to the effects of mutations on entry into the mix of cells expressing α2-3 linked and α2-6 linked sialic acids (**Fig. S3.3A**).

There were only a few sites where the effects of mutations were appreciably different in the α2-3 linked only or α2-6 linked only sialic acid cells; these were mostly sites in the receptor-binding loops known to be important for receptor binding specificity such as 193, 220, and 226 (**Fig. S3.3A**). The reason that the effects of mutations are so similar on entry in both types of cells is likely because most HA mutations affect entry by modulating HA folding and membrane fusion rather than receptor-binding, and because the A/Anhui/1/2013 H7 HA used in our experiments is already capable of binding both α2-3 linked and α2-6 linked sialic acids due to the fact that it contains the Q226L mutation(Y. Shi et al. 2013). To confirm this latter point, we created pseudovirus with the L226Q reversion and measured its entry into the 293 cells expressing only α2-3 linked or only α2-6 linked sialic acids (**Fig S3.3B**). Pseudovirus with the unmutated A/Anhui/1/2013 H7 HA entered both cells similarly, but pseudovirus with the Q226L revertant of this protein were better at entering the α2-3 linked sialic acid expressing cells. As additional controls, we confirmed pseudovirus expressing an avian H5N1 HA was better at entering the α2-3 linked sialic acid expressing cells, while pseudovirus expressing a human H3N2 HA was better at entering the α2-6 linked sialic acid expressing cells (**Fig S3.3B**). Overall, these results show that the effects of most mutations to the H7 HA on its cell entry function are the same regardless of the type of sialic acid expressed on the target cells, although a small number of mutations in the receptor binding loops affect entry in a way that depends on the sialic acid expressed on the target cells.

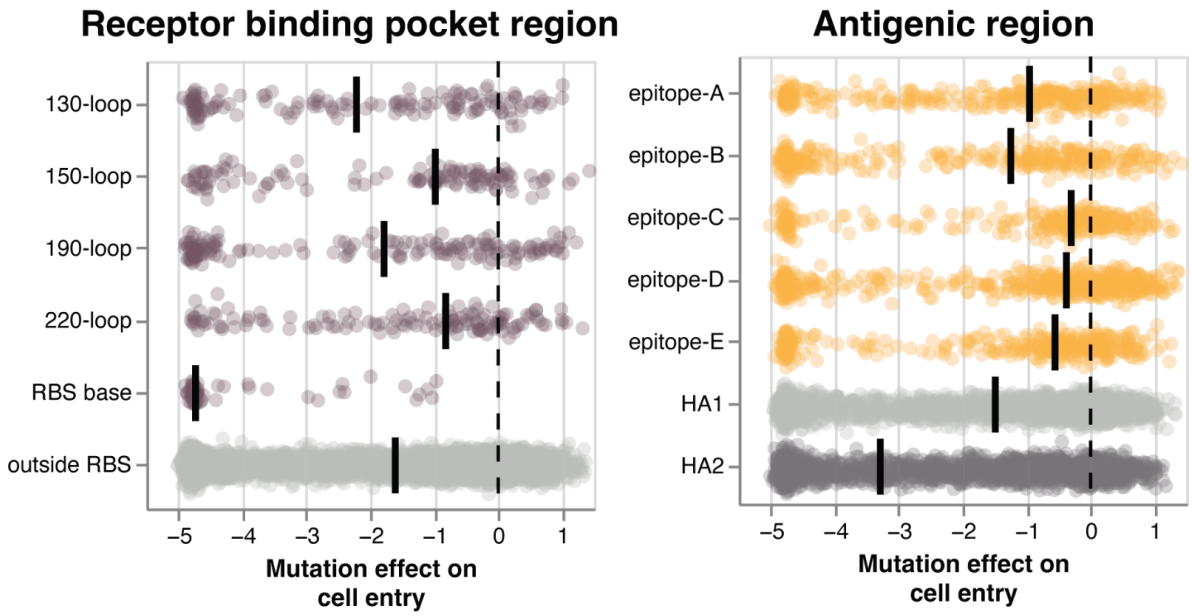


Figure 3.2 | Mutation effect on cell entry group by receptor binding pocket and antigenic regions.

A) Scatter plot comparing mutation effects on cell entry across different functional regions of H7 hemagglutinin. Left panel depicts mutations within receptor binding pocket regions, showing the distinct constraint patterns across the RBS base (dark purple), structural loops (130-loop, 150-loop, 190-loop, 220-loop in varying shades of purple), and regions outside the RBS (grey). Right panel depicts mutations within antigenic regions, showing the five major epitopes (A-E in orange) compared to the broader HA1 and HA2 domains (grey). X-axis measures mutation effect on cell entry as log2 fold-change relative to wildtype, where negative values indicate impaired function and positive values indicate enhanced function. Y-axis shows the different functional regions or epitopes where mutations occur. The data reveal that the receptor binding site base shows the highest functional constraint with most mutations causing severe defects (effects ranging from -4 to -5), while surrounding structural loops display intermediate constraint. In contrast, all five major antigenic epitopes demonstrate remarkable mutational tolerance with most mutations having minimal impact on cell entry function, highlighting the evolutionary trade-off between maintaining essential binding functions and allowing immune escape through antigenic variation.

Additionally, the mutations show distinct patterns of functional constraint when analyzed by receptor binding regions and antigenic sites. In the receptor binding pocket region **(Fig 3.2**), the RBS base shows the highest constraint, with most mutations causing severe functional defects (effects ranging from -4 to -5), reflecting the critical importance of direct receptor contact residues for cell entry. The structural loops surrounding the receptor binding site (130-loop, 150-loop, 190-loop, and 220-loop) display intermediate constraint, with mutation effects typically ranging from -1 to -3, indicating their supportive role in maintaining proper receptor binding geometry. Regions outside the RBS show much greater mutational tolerance, with effects clustered near zero.

When examining antigenic regions (**right panel**), all five major epitopes (A through E) demonstrate remarkably high mutational tolerance, with most mutations having minimal impact on cell entry function (effects near zero). This contrasts sharply with the highly constrained HA2 region, where mutations consistently impair function, and the moderately constrained HA1 backbone regions. These findings highlight the evolutionary trade-off in influenza HA, where antigenic sites maintain high mutational flexibility to facilitate immune escape while core functional regions like the receptor binding site and fusion machinery remain under strong purifying selection.

3.3.3 Presence of mutations that confer increased entry in alpha 2,6 cells in current circulating strains

We wanted to determine whether mutations enhancing entry into cells expressing human-type 2,6-linked sialic acid receptors had been naturally selected in H7N9 populations. To address this, we pulled 1,451 H7N9 sequences from GISAID (isolate IDs hidden for privacy) and systematically compared them against our DMS-identified beneficial mutations.

The analysis yielded a striking pattern. Among the ~65 sites where we found beneficial mutations that enhance 2,6 receptor-mediated entry (effects ranging from +0.50 to +1.80), several showed remarkable convergence between our experimental results and natural evolution. Five sites emerged as critical adaptive hotspots, each showing near-universal prevalence in circulating strains.

Most notably, the strongest entry enhancement mutations were already nearly fixed in natural H7N9 populations. Q210Y and K193T both conferred substantial entry effects (>1.2) and appeared in ~98% of analyzed sequences (**Fig 3.3A**). This high prevalence suggests these represent essential adaptive mutations for human receptor binding. Similarly prevalent were K169Y (effect +0.80, found in 97.8% of sequences) and P175S (effect +0.91, present in 97.2%). The K166T mutation showed an effect of +0.82 and appeared in 96% of sequences—clear evidence of strong positive selection for enhanced 2,6 binding. Interestingly, site 166 displayed functional redundancy. While K166T dominated, we identified four other beneficial mutations at this position (K→D, I, L, V), though these occurred at much lower frequencies (<1%). This pattern indicates multiple evolutionary solutions exist for enhanced entry, but one has become strongly

favored. We also found nine additional mutations present in more than 10% of sequences, suggesting their ongoing establishment within the H7N9 population.

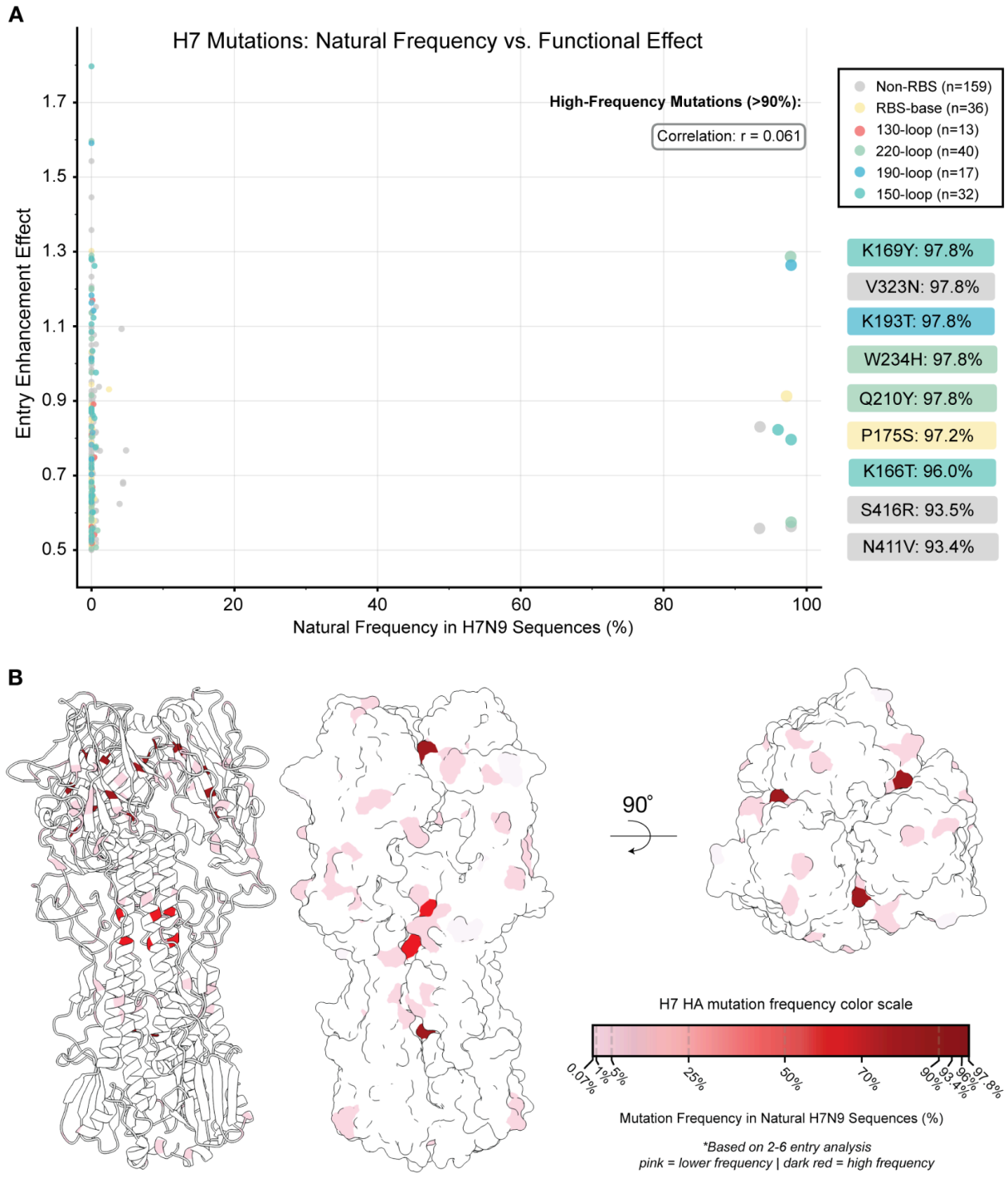


Figure 3.3 | Presence of 2-6 Entry Enhancing Mutations in H7N9 Natural Sequences.
A) Scatter plot comparing the cell entry effects of mutations on 2-6 cells (y-axis) and the natural frequency (x-axis) of the respective mutations in H7N9 sequences. Each dot represents an individual mutation and the dots are colored by their respective RBS region. Non-RBS sites (gre), RBS-base (yellow), 130-loop (red), 220-loop (green), 190-loop (blue), and 150-loop (cyan). Higher frequency mutations are also noted. **B)** The structure of the A/Anhui/1/2013 H7 HA (Huang et al. 2019) (PDB 6ii9) colored by the mutation frequency in natural H7N9 sequences. Mutations at a site that are more frequent in natural sequences are shown in red - dark red. Mutations that are less frequent in natural sequences are shown in light pink to dark pink, with the darker the shade conveying higher frequency.

Many high-effect mutations (+1.0 to +1.8) were rare or completely absent in circulating strains. For example, N158bD showed a +1.80 effect but was not detected in any sequences. R482T conferred a +1.45 effect but appeared in only 0.14% of isolates. E81N (+1.54 effect) and K200V (+1.59 effect) were entirely absent from natural populations. Even some moderately beneficial mutations like G23A remained rare, appearing in <1% of sequences and suggesting ongoing evolutionary dynamics where advantageous changes haven't reached fixation.

These results illuminate H7N9's evolutionary trajectory. Substantial adaptive evolution toward human receptor binding has already occurred, with the most beneficial mutations becoming nearly universal. However, the discovery of numerous high-effect mutations that remain rare indicates significant untapped evolutionary potential. This could be realized under appropriate selective pressures, suggesting that while H7N9 has undergone key adaptive changes, additional functional enhancement remains possible through mutations not yet naturally selected.

The convergence between our experimentally identified beneficial mutations and those prevalent in nature validates our DMS approach. More importantly, it demonstrates how systematic mutational analysis can uncover both realized and potential evolutionary pathways in influenza adaptation.

3.3 Discussion

We successfully extended our deep mutational scanning approach to H7 hemagglutinin, providing the first comprehensive mutational landscape for this pandemic-relevant subtype. Our analysis of the A/Anhui/1/2013 strain serves as the backbone for current WHO candidate vaccine viruses and reveals important insights into H7 HA functional constraints and evolutionary potential. The overall pattern of mutational tolerance in H7 HA closely mirrors findings from previous DMS studies of H3 and H5 subtypes. This suggests conserved structural and functional constraints across influenza A hemagglutinins. HA2 showed greater functional constraint than HA1, as expected, particularly in critical regions such as the fusion loop (sites 330-350) that mediates membrane fusion. Conservation of these constraint patterns indicates that fundamental aspects of HA structure-function relationships are maintained across subtypes, despite significant sequence divergence.

Our experimental system benefits from the dual receptor binding capability of the A/Anhui/1/2013 H7 HA. This strain contains the Q226L mutation enabling binding to both avian-type α2-3 and human-type α2-6 linked sialic acids. We found that most mutations had nearly identical effects regardless of the sialic acid type expressed on target cells. Only a few sites in receptor-binding loops (193, 220, 226) showed differential effects. This similarity likely reflects the fact that most HA mutations affect protein folding and membrane fusion rather than receptor specificity. It also confirms that the Q226L-containing H7 HA can efficiently utilize both receptor types. This dual functionality carries epidemiological significance, as it may facilitate viral transmission between avian and mammalian hosts.

Comparing our DMS-identified beneficial mutations with circulating H7N9 sequences from GISAID revealed striking evidence of natural selection for enhanced human receptor binding. Among the ~65 sites where we identified mutations that enhance 2,6 receptor-mediated entry, five emerged as critical adaptive hotspots. These showed near-universal prevalence (>95%) in natural populations: K193T, Q210Y, K169Y, P175S, and K166T. These mutations displayed some of the strongest entry enhancement effects in our screen (+0.8 to +1.3) and are nearly fixed in circulating H7N9 strains. This demonstrates that substantial adaptive evolution toward human receptor binding has already occurred. The convergence between experimental and evolutionary data validates our DMS approach while highlighting how natural selection has favored mutations that optimize viral entry into human-type receptor-expressing cells.

Our DMS also revealed something unexpected: substantial "cryptic" adaptive potential that remains largely untapped in nature. Many mutations with strong beneficial effects (+1.0 to +1.8) were rare or entirely absent in circulating strains. Examples include N158bD (+1.80), E81N (+1.54), and K200V (+1.59). While H7N9 has undergone key adaptive changes, this pattern suggests significant evolutionary potential exists. This could be realized under appropriate selective pressures, potentially enabling further enhancement of human adaptation.

Different functional regions showed distinct constraint patterns that highlight evolutionary trade-offs inherent in HA evolution. The receptor binding site base showed the most severe constraint. Mutations here typically caused 4-5 log2 reductions in cell entry, reflecting the critical importance of direct receptor contacts. Surrounding structural loops displayed intermediate constraint, supporting their role in maintaining proper receptor binding geometry. All five major antigenic epitopes (A-E) demonstrated remarkable mutational tolerance in stark contrast. Most mutations in these regions had minimal impact on cell entry function. This pattern suggests that antigenic sites can readily accommodate mutations that facilitate immune escape without compromising viral fitness. This explains the rapid antigenic drift observed in circulating influenza viruses.

These findings carry important implications for pandemic preparedness and vaccine development. The high mutational tolerance of antigenic regions, combined with evidence that beneficial receptor binding mutations have already been naturally selected, suggests H7 viruses possess substantial capacity for both immune escape and further human adaptation. The identification of cryptic beneficial mutations not yet present in nature provides a roadmap for potential evolutionary pathways that could enhance pandemic risk. Conversely, strong functional constraints in receptor binding and fusion regions identify potential targets for universal vaccine approaches. These would be less susceptible to viral evolution. Understanding these constraint patterns and the realized versus potential adaptive landscape will be crucial for predicting H7 evolutionary trajectories. This knowledge will also prove essential for developing more robust countermeasures against this pandemic threat.

3.4 Supplementary Materials

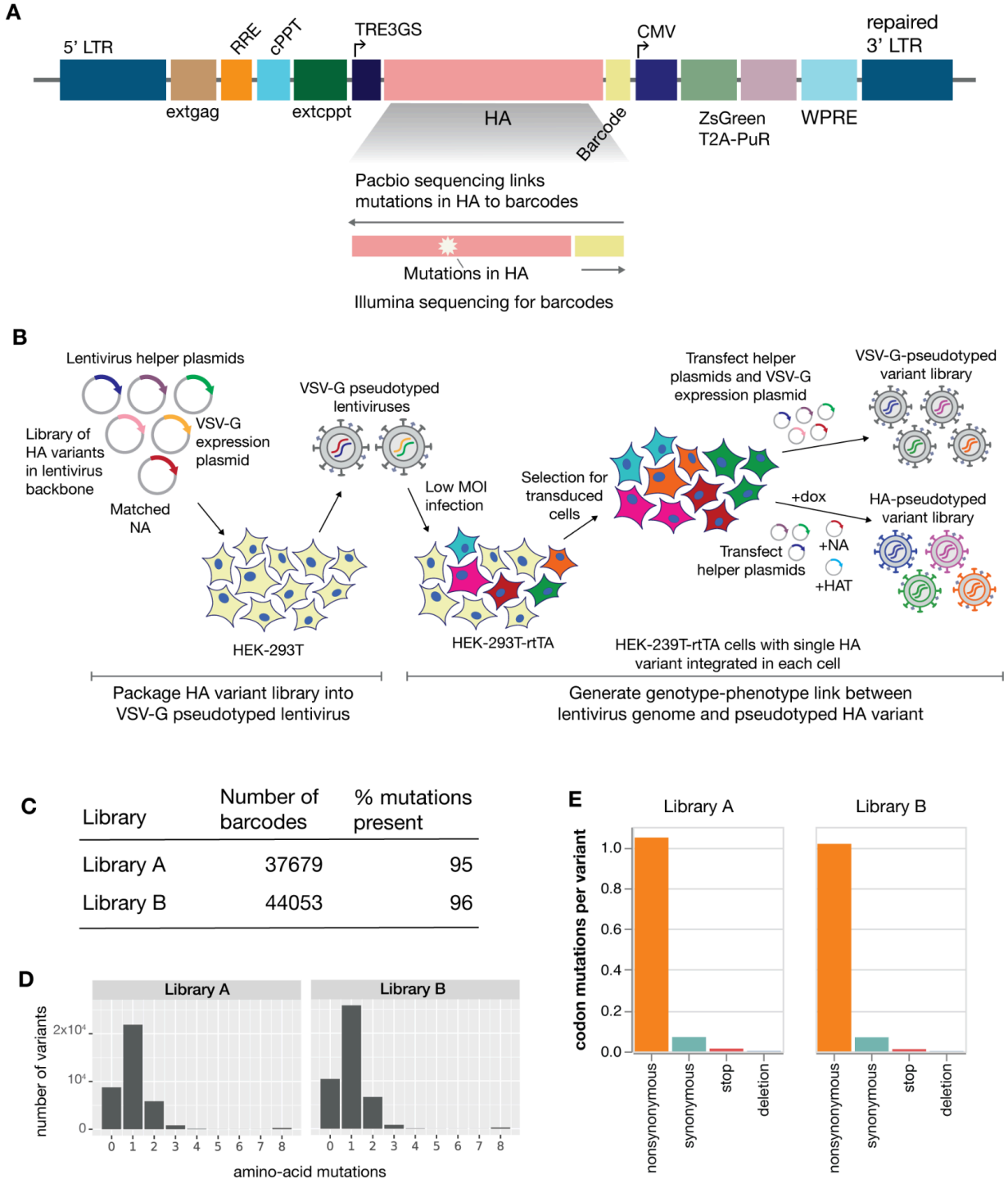


Figure S3.1 | Overview of pseudovirus deep mutational scanning of H7 HA

A) The pseudovirus deep mutational scanning uses a lentiviral backbone encoding the HA protein followed by an identifying 16-nucleotide barcode under the control of an inducible promoter. The lentiviral backbone has a full 3′-LTR so that integrated proviruses can be reactivated by transfection of appropriate helper plasmids. **B)** Schematic of pseudovirus deep mutational scanning experiments. We create a library of mutants of the HA in the context of the barcoded-HA expressing lentiviral backbone plasmid. Cells are transfected with this HA-encoding lentiviral backbone along with helper plasmids encoding the other proteins needed to produce lentiviral particles (Gag-Pol, Tat, Rev), a plasmid encoding the matched N9 NA, and a plasmid encoding VSV-G. The resulting lentiviral particles, which lack a genotype-phenotype link and express VSV-G on their surface, are the infected into rtTA expressing 293T cells at low multiplicity of infection, and transduced cells are infected so that each cell encode a unique HA mutant as a provirus in its genome. These cells are then transfected with the helper plasmids and NA-encoding plasmids to produce pseudoviruses that express unique HA mutants. Cells are also transfected with VSV-G to produce control pseudoviruses that are not dependent on HA to infect cells. To measure cell entry of each HA variant, we use sequencing to compare the efficiency of that barcoded lentiviral particle at entering cells in the presence or absence of VSV-G (particles always enter cells when VSV-G is expressed, but otherwise require a functional HA for cell entry). Note that these lentiviral particles encode no proteins other than HA, and so are not capable of undergoing multicycle replication and so provide a safe way to study the effects of mutations to HA. **C)** Number of uniquely barcoded HA variants and the fraction of all possible HA ectodomain amino-acid mutations that are present in at least one barcoded variant in each of the two independent H7 HA pseudovirus libraries. **D)** Distribution of amino-acid mutations per barcoded HA variant in each library. **E)** Average number of nonsynonymous, synonymous, stop-codon, and deletion mutations per barcoded HA variant.

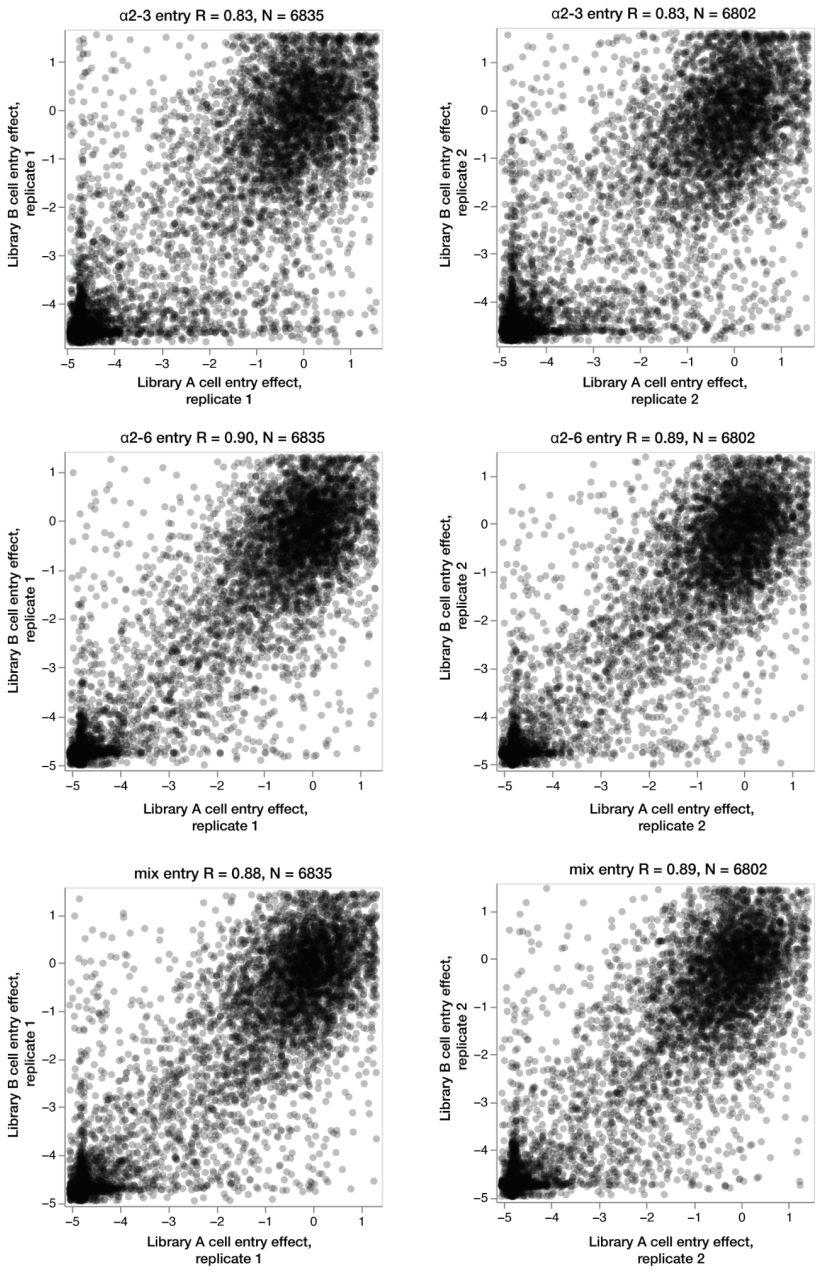


Figure S3.2 | H7 HA cell entry effects measured using independent deep mutational scanning libraries are highly correlated

Correlation of the effects of H7 HA mutations on cell entry as measured with independent pseudovirus libraries on 293 cells expressing α2-3 linked sialic acids (top row), α2-6 linked sialic acids (middle row), or a mix of both cells (bottom row). Each point represents the effect of a mutation measured in each replicate. For each of the two libraries, two technical replicates were performed. Each panel shows the correlations between two independent libraries (Library A and B) for each technical replicate (replicate 1 or 2). R is the Pearson correlation and N is the number of mutations measured in both pairs of measurements.

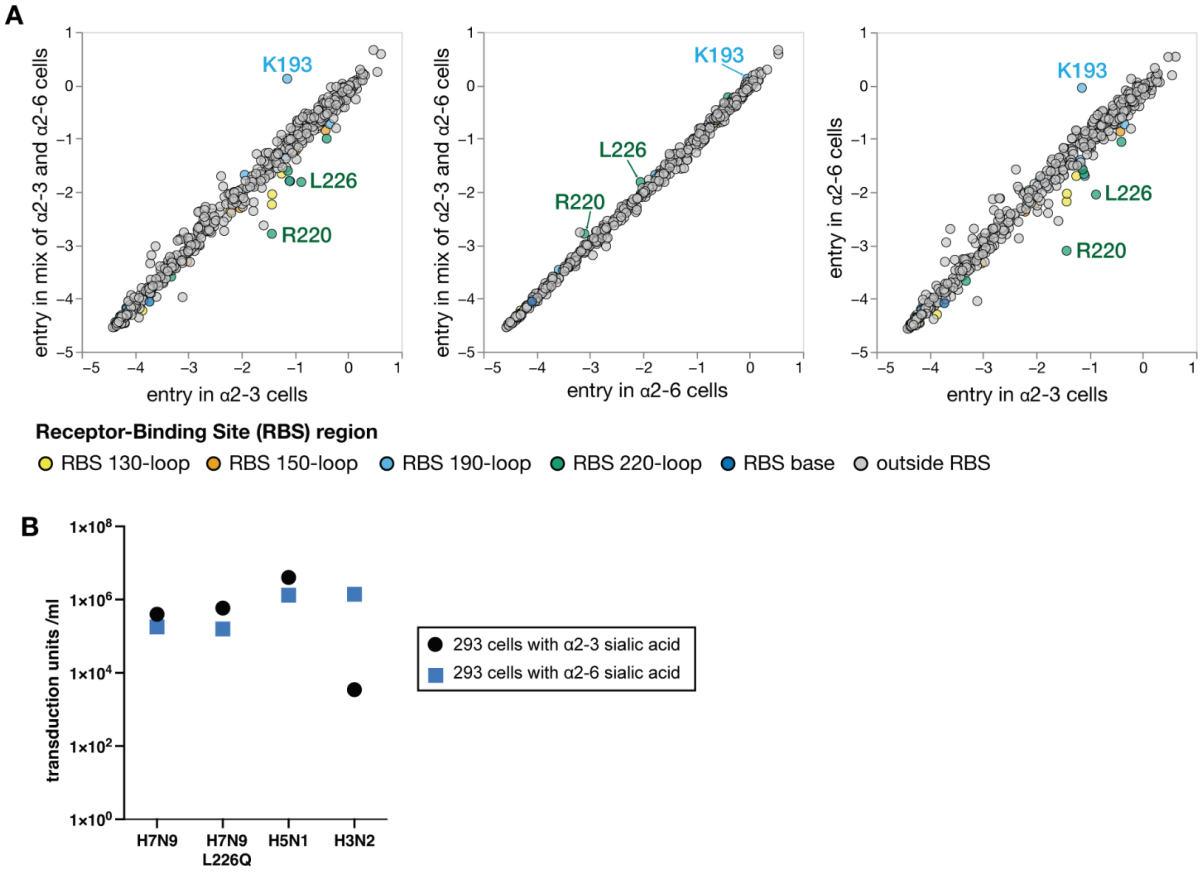


Figure S3.3 | Effects of mutations to the H7 HA on entry in 293 cells expressing only α2-6 or only α2-3 linked sialic acids

A) Scatter plots showing the average effects of mutations at each site on entry into 293 cells expressing only α2-3 linked sialic acids, only α2-6 linked sialic acids, or an equal mix of the two cells. Sites are colored by whether they are part of the receptor-binding site (RBS). See https://dms-vep.org/Flu_H7_Anhui13_DMS/cell_entry.html for interactive versions of these plots. **B)** Titers on 293 cells expressing only α2-3 linked sialic acids or only α2-6 linked sialic acids of lentiviral particles pseudotyped with the HA and NA of A/Anhui/1/2013 (H7N9) strain, the same H7N9 HA and NA but with the L226Q mutation in HA, the HA and NA of the avian A/American Wigeon/South Carolina/USDA-000345-001/2021 (H5N1) strain, or the HA and NA of the human A/Massachusetts/18/2022 (H3N2) strain. text

3.5 Materials and methods

Code and data availability

The computer code for the deep mutational scanning of the H7 HA described in this manuscript is available on GitHub at https://github.com/dms-vep/Flu_H7_Anhui13_DMS and can be

viewed interactively at https://dms-vep.org/Flu_H7_Anhui13_DMS/ via GitHub Pages. The quality filtered measurements of mutation effects on cell entry for the H7 HA are available in tabular format at https://github.com/dms-vep/Flu_H7_Anhui13_DMS/blob/main/results/summaries/entry_all_cells.csv.

Design of H7 HA deep mutational scanning libraries

Deep mutational scanning libraries were designed in the background of the HA of the A/Anhui/1/2013 (H7N9) strain. The HA gene was codon optimized for use in the lentiviral backbone, see https://github.com/dms-vep/Flu_H7_Anhui13_DMS/blob/main/library_design/pH2rU3_d1GTSS_extgag_extcppt_ForInd_Anhui13_A-Anhui-1-2013_genscript2_CMV_ZsGT2APurR.gb for a map of the lentiviral backbone plasmid encoding the HA gene. The codon optimization tool used was the Gensmart codon optimization offered by Genscript. This lentiviral backbone contains several modifications relative to commonly used lentiviral vectors. We used an extended gag sequence, as has been shown to improve genome packaging efficiency (Humes et al. 2021; Liu et al. 2017). Under the same rationale, we used an extended cppt sequence with more of the native HIV sequence following the cppt. We also mutated the transcription start site of HIV to "TTG" from "GGG" to generate TTG-initiated transcripts more likely to be packaged than other transcription start initiation site forms(Brown et al. 2020; Nikolaitchik et al. 2023).

The library was designed to have all possible amino acid mutations to each site in the HA ectodomain and a stop codon at every other position for the first 20 positions of HA. Single mutant libraries were ordered from Twist Biosciences. Quality control data for library production by Twist is at (https://github.com/dms-vep/Flu_H7_Anhui13_DMS/blob/main/library_design/Final_QC_Report_Q-367972_for_twist_libraries.xlsx)

H7 HA deep mutational scanning plasmid library production

The H7 library ordered from Twist was designed to have mutations to all amino acids in the 506 residue ectodomain of HA. Of the 506 sites, 12 sites failed to be properly mutagenized in the Twist library production; of the sites that failed, 2 sites were in antigenic regions and within the receptor-binding pocket. To add mutations at the missing sites, forward and reverse primers with NNG/NNC codons were designed and ordered from IDT, sequences can be found at (https://github.com/dms-vep/Flu_H7_Anhui13_DMS/blob/main/library_design/Anhui13_gen2_ha_missing_mut_primers.csv and https://github.com/dms-vep/Flu_H7_Anhui13_DMS/blob/main/library_design/Anhui13_gen2_ha_missing_site_NNS_primers.csv). Forward and reverse primer pools were made by pooling the primers at equal molar ratios for a final concentration of 5µM. The linear HA template was produced by digesting the lentivirus backbone encoding the HA (plasmid map linked above) with XbaI and NotI-HF and then gel purifying the reaction. PCR mutagenesis was

then performed as previously described (Yu et al. 2025; Dadonaite, Crawford, et al. 2023) with a few changes, 9 PCR cycles were used for the mutagenesis and after the joining PCR a DpnI digest was performed for 20 minutes at 37˚ to remove any wildtype template. This PCR mutagenesis reaction was performed in duplicate, one for library A and one for library B, resulting in two mutagenesis pools. We barcoded the library pools generated from Twist Biosciences independently in 2 separate reactions to generate library A and library B pools to serve as biological replicates, each of which were then handled independently throughout deep mutational scanning. The two mutagenized pools, A and B, were also barcoded separately, resulting in 4 separate barcoding PCR reactions. The reactions were barcoded with primers containing a random 16-nucleotide sequence downstream of the stop codon. The lentiviral backbone was digested from the same plasmid that was used to generate the linear HA fragment (https://github.com/dms-vep/Flu_H7_Anhui13_DMS/blob/main/library_design/pH2rU3_d1GTSS_extgag_extcppt_ForInd_Anhui13_A-Anhui-1-2013_genscript2_CMV_ZsGT2APurR.gb) by incubating with XbaI and NotI-HF for 2 hours at 37˚ followed by 65˚ for 20 minutes to heat inactivate XbaI. The barcoded libraries and the digested lentiviral backbone were then run on a 0.8% agarose gel at 90V for 1 hour. The bands of the appropriate size were excised using the Nucleospin Gel and PCR Clean-up Kit (Macherey-Nagel, Cat. No. 740609.5) and then purified using Ampure XP beads (Beckman Coulter, Cat. No. A63881). All products were eluted in molecular grade water.

The barcoded libraries were then cloned into the lentiviral backbone at a 1:2 insert to vector ratio in a HiFi assembly for 1 hour at 50˚C. This was done using the NEBuilder HiFi DNA Assembly kit (NEB E5520S). The reactions were then purified with Ampure XP beads and eluted in molecular grade water. The purified products were then transformed into 10-beta electrocompetent cells (NEB, Cat. No. C3020K) using a BioRad MicroPulser Electroporator (Cat. No. 1652100), shocking at 2 kV for 5 milliseconds. For the Twist library pools A and B, 5 reactions were electroporated per library. For the mutagenesis library pools, 3 reactions were electroporated per library A and B. The resulting tubes of electroporated cells were placed on a 37˚ shaker for 1 hour to recover. After recovery, the Twist library pools were pooled per condition (5x tubes for Library A and 5x tubes for Library B) were pooled and transferred into ~300ml of LB + Ampicillin (final concentration at 10µg/ml). The mutagenesis pools were pooled per condition (3x library A and 3x library B) and transferred into ~150ml of LB + Ampicillin at the same final concentration as mentioned previously. We then plated 50µl of a 1:1000 dilution of the Twist library pools and mutagenized HA pools to calculate the CFU/ml. The total number of colonies from the electroporations were as follows: Twist A 7.2e7 CFU/ml, Twist B 4.8e7 CFU/ml, Mutagenesis A 3.6e6 CFU/ml, Mutagenesis B 4.2e6 CFU/ml. The high number of colonies is necessary to ensure high diversity of the library and to prevent bottlenecking of the mutants. After shaking overnight, the plasmids were then extracted using the QIAGEN HiSpeed Plasmid Maxi Kit (Cat. No. 12662). The mutagenized HA fragments were then pooled together with the library pool generated by Twist Biosciences at a molar ratio proportional to the number of mutants in each pool so that each variant was represented equally in the final

library. For every 1µg of Twist pools, 0.0285µg of mutagenized HA pools were used because long-read PacBio sequencing of the plasmid libraries showed this ratio resulted in the most even distribution of mutations across the HA for the combined libraries.

Production of cell-stored deep mutational scanning libraries

We generated cell-stored deep mutational scanning libraries where each cell is integrated with a single copy of a barcoded HA mutant to enable rescue of the genotype-phenotype linked pseudoviruses(Dadonaite, Crawford, et al. 2023; Dadonaite et al. 2024; Yu et al. 2025). VSV-G pseudotyped viruses were produced by transfecting 2x10cm plates seeded with 8 million 293T cells with the lentiviral backbone containing barcode HA deep mutational scanning libraries (10µg per plate). Lentiviral Gag/pol, Tat, and Rev helper plasmids (AddGene numbers 204152, 204154, 204153; 1.25 µg per plate), a VSV-G expression plasmid (1µg per plate, AddGene number 204156), and a plasmid expressing the matched N9 NA from A/Anhui/1/2013 (1µg per plate). BioT transfection reagent (Bioland Scientific, Cat. No. B01-02) was used according to the manufacturer's instructions. The NA gene came from the A/Anhui/1/2013 strain. The sequence for this NA can be found at https://github.com/dms-vep/Flu_H7_Anhui13_DMS/blob/main/library_design/HDM_N9_genscript_H7N9_A_Anhui_1_2013.gb. This NA was used because it provided high pseudovirus titers. It is important to note that VSV-G pseudoviruses do not require NA or virus release or entry, the produced virions will also have HA expressed from the lentiviral backbone on their surface, so NA is necessary to prevent HA from binding the virions back to the producer cell surface. This is important for the HA library, because different HA mutants will have different expression and sialic-acid binding properties. This could lead to a different tendency of different virions to be bound on the cell surface by HA if NA is not added at a high excess. 48 hours post transfection, the supernatant was filtered through a 0.45µm syringe filter (Corning, Cat. No. 431220) and then stored at -80˚C. These viruses were then used to infect 293T cells to determine the titer in transcription units (TU) per mL, which was determined by measuring the percentage of zsGreen positive cells via flow cytometry.

The VSV-G pseudoviruses were then used to infect 293T-rtTA cells (using a specific clone that had shown to yield good pseudovirus titer(Dadonaite, Crawford, et al. 2023; Dadonaite et al. 2024; Yu et al. 2025) at an infection rate of 0.75% so that most transduced cells receive only a single integrated proviral genome. The MOI was then confirmed by measuring the percentage of zsGreen positive cells via flow cytometry 48 hours post transduction. After measuring the positive cells, Library A contained 37,679 variants and, and library B contained 44,053 (**Fig. S2C**).. This allowed for coverage of each barcode ~4x times. Transduced cells were then selected using 0.75µg/mL puromycin for 1 week (fresh media with puromycin was replenished every 48 hours) leading to a population of cells where each cell contains an integrated proviral genome encoding a single barcoded variant of HA. These cells were then frozen in 20 million cells per aliquot and stored in liquid nitrogen until further use.

Rescue of HA and VSV-G expressing pseudovirus libraries

To rescue HA expressing pseudoviruses from the integrated cells, 150 million cells were plated in 5-layer flasks in phenol-free tetracycline-free D10 supplemented with 1µg/mL of doxycycline to induce HA expression from the integrated genomes. The next day, each flask was transfected with 40µg of each helper plasmids encoding Gag/Pol, Tat, and Rev (AddGene numbers 204152, 204154, 204153), 15µg of plasmid expressing human airway trypsin-like protease (https://github.com/dms-vep/Flu_H7_Anhui13_DMS/blob/main/library_design/HDM_HAT.gb) to activate the HA for membrane fusion, and 15µg of the plasmid expressing NA. BioT transfection reagent was used according to the manufacturer's instructions. 16 hours post transfection, the phenol-free tetracycline-free media was aspirated off, and 150mL of fresh Influenza Growth Media with 1µg/mL of doxycycline was added. This low serum media is necessary due to the non-specific inhibitors in FBS that can inactivate HA and interfere with pseudovirus infection. 32 hours after the media swap, the supernatant was then filtered through 0.45µm SFCA Nalgene 500 mL Rapid-Flow filter unit (Cat. No. 09-740-44B). Filtered supernatant was then concentrated using 100K MWCO Pierce protein concentrators (Thermo Fisher Cat. No. 88537) that were spun down at 1500xg at 4˚ for 1 hour. After centrifugation, supernatant was discarded and the concentrated virus was pooled and aliquoted. The estimated titers were ~3.2e6TU/mL. 1mL aliquots of the concentrated library viruses were frozen and stored at -80˚C until further use.

To rescue VSV-G expressing pseudoviruses from integrated cells, 20 million cells were plated in 15cm plates in phenol-free tetracycline-free D10. The next day, each plate was transfected with 6.75µg of each helper plasmids expressing Gag/Pol, Rev, and Tat, 3µg of plasmid expressing NA, and 6.75µg of plasmid expressing VSV-G. 48 hours post transfection, the supernatant was filtered through a 0.45µM SFCA 500mL Rapid-Flow filter unit and concentrated using 100K MWCO Pierce protein concentrators. Aliquots of the concentrated VSV-G expressing pseudoviruses were frozen at -80˚C for use in linking mutations to barcodes and cell entry experiments.

Long-read PacBio sequencing for variant-barcode linkage

1e6 293T cells were plated per well in a 6-well plate coated with Poly-L-Lysine to help with cell adhesion. The next day, 15 million TU's of VSV-G expressing pseudoviruses that were rescued from cell-stored deep mutational scanning libraries were used to infect the cells across 3 wells. 12 hours post infection, the non-integrated reverse-transcribed lentiviral genomes were recovered by minipreping the 293T cells using a QIAprep Spin Miniprep Kit as previously described(Dadonaite, Crawford, et al. 2023; Dadonaite et al. 2024; Yu et al. 2025). Eluted genomes were split into 2 separate PCR reactions, where each reaction added a unique tag for detecting strand exchange, which could prevent correct mutation-barcode linkage. The resulting set of PCR reactions were then pooled together, and another round of PCR was performed. In each PCR round, we minimized the number of reaction cycles to limit strand exchange. The reactions were then purified with Ampure XP beads and eluted in elution buffer.Long read sequencing was then performed using a PacBio Sequel IIe machine. The PacBio circular consensus sequences (CCSs) were then aligned to the target amplicon

(https://github.com/dms-vep/Flu_H7_Anhui13_DMS/blob/main/data/PacBio_amplicon.gb) using alignparse(Crawford and Bloom 2019), and consensus sequences for HA variants associated with each barcode were determined by requiring at least 2 CCSs per barcode. The final barcode-variant lookup table can be found at https://raw.githubusercontent.com/dms-vep/Flu_H7_Anhui13_DMS/refs/heads/main/results/variants/codon_variants.csv. Library A had 37,679 variants and library B had 44,053 variants (**Fig. S2C**).

Mutation effects on cell entry in α2-3 and α2-6 linked sialic acids expressing cells

Cell entry effects of each mutation in the library were determined as previously described (Dadonaite et al. 2024). To measure effects of mutations on cell entry, 293 cells expressing primarily α2-3 or α2-6 linked sialic acids were used. Additionally, we created an equal pool of the α2-3 or α2-6 linked sialic acid expressing cells, and measured cell entry on this equal pool as well. These cells were then either infected with libraries pseudotyped with VSV-G envelope (produced as described above) or HA pseudotyped libraries. For HA pseudotyped libraries, the α2-3, α2-6, or an α2-3/α2-6 cell pool were plated in the presence of 2.5µg/ml of amphotericin B a day before (amphotericin B was shown to increase library virus titers), and infections were performed in the presence of 1µM oseltamivir to inhibit NA and prevent cleavage of receptors on target cell surface. For VSV-G pseudotyped library infections approximately 8 million transcription units were used and for the H7 pseudotyped library infections approximately 3.2 million transcription units were used. Because the initial titers on α2-3 vs. α2-6-linked sialic acid cells were similar, we were able to infect both cell lines with the same amount of virus. After infecting the cells with the viruses, the plates were spun in a 30˚ centrifuge for 1 hour at 300g to maximize infection rate. At 12 to 15 hours postinfection, cells were collected, nonintegrated lentiviral genomes were recovered using QIAprep Spin Miniprep Kit (Qiagen, cat. no. 27104), and amplicon libraries for barcode sequencing were prepared as previously described (Dadonaite, Crawford, et al. 2023; Dadonaite et al. 2024; Yu et al. 2025).

Cell entry scores for each variant were calculated using log enrichment ratio: $\log_2$ ([n^{v}_{post} / n^{wt}_{post}]/[n^{v}_{pre} / n^{wt}_{pre}]), where n^{v}_{post} is the count of variant v in the H7-pseudotyped infection (postselection condition), n^{v}_{pre} is the count of variant v in the VSV-G-pseudotyped infection (preselection condition, and n^{wt}_{post} and n^{v}_{pre} are the counts for wildtype variants. Positive cell entry scores indicate that a variant is better at entering the cells compared to the unmutated parental HA, and negative scores indicate entry worse than the unmutated HA.

To calculate the mutational-level cell entry effects, a sigmoid global-epistasis (Otwinowski et al. 2018) function was fitted to variant entry scores after truncating the values at a lower bound of -5, using the *multi-dms* software package(Haddox et al. 2023). The rationale for fitting the global epistasis models is that some of the variants have multiple HA mutations, and so the model enables deconvolution of the mutation effects from these multiple mutated variants as well as just the single mutants. An interactive heatmap showing the effect of each mutation on cell entry on α2-3, α2-6, and an equal pool of both cells is at (https://dms-vep.org/Flu_H7_

Anhui13_DMS/cell_entry.html). For the mutation effects, values of zero mean no effect on cell entry, negative values mean impaired cell entry, and positive values mean improved cell entry. For plots in the paper, we show only mutations observed in average of at least 2 barcoded variants per library and measured in both biological library replicates.
Text

Chapter 4

A version of this chapter is published as:

Jenny J Ahn, Timothy C Yu, Bernadeta Dadonaite, Caelan E Radford, Jesse D Bloom, **Influenza hemagglutinin subtypes have different sequence constraints despite sharing extremely similar structures,** *Virus Evolution*, Volume 12, Issue 1, 2026, veag018, https://doi.org/10.1093/ve/veag018

4 Influenza hemagglutinin subtypes have different sequence constraints despite sharing extremely similar structures

4.1 Abstract

Hemagglutinins (HA) from different influenza A virus subtypes share as little as ~40% amino acid identity, yet their protein structure and cell entry function are highly conserved. Here we examine the extent that sequence constraints on HA differ across three subtypes. To do this, we first use pseudovirus deep mutational scanning to measure how all amino-acid mutations to an H7 HA affect its cell entry function. We then compare these new measurements to previously described measurements of how all mutations to H3 and H5 HAs affect cell entry function. We find that ~50% of HA sites display substantially diverged preferences for different amino acids across the HA subtypes. The sites with the most divergent amino-acid preferences tend to be buried and have biochemically distinct wildtype amino acids in the different HA subtypes. We provide an example of how rewiring the interactions among contacting residues has dramatically shifted which amino acids are tolerated at specific sites. Overall, our results show how proteins with the same structure and function can become subject to very different site-specific evolutionary constraints as their sequences diverge.

4.2 Introduction

Proteins often evolve at the sequence level while largely retaining their structures and functions (Lesk and Chothia 1980; Zuckerkandl and Pauling 1965). A particularly striking example of this evolutionary dynamic is the hemagglutinin (HA) of influenza virus. There are at least 19 subtypes of HA (numbered H1 to H19), most of which are found in avian influenza viruses with some also found in viruses that have jumped to humans or other mammals (Parrish et al. 2015), and a handful found only in bat viruses (W. Yang et al. 2021). These HA subtypes have highly conserved structures and share the ability to mediate cell entry via a pH-dependent conformational conversion after binding to receptor (Bouvier and Palese 2008; Wu and Wilson 2020; Gamblin et al. 2021), which is usually sialic acid although sometimes

MHC class II (Karakus et al. 2024, 2019). Yet despite this structural and functional conservation, HA subtypes share as little as ~40% amino-acid sequence identity. This high sequence divergence in the context of a conserved structure and function is driven in part by host immune pressure for antigenic diversification of HA to evade antibodies elicited by prior infections (Hill et al. 2022; Smith et al. 2004). The remarkable extent to which HA maintains its functional properties despite its sequence divergence is illustrated by the fact that two influenza pandemics in the last century were caused by human strains acquiring HAs of different subtypes while retaining most other viral genes: in 1957 a human H1N1 strain acquired a H2 HA with only 67% identity while retaining 5 of the other 7 viral genes, and in 1968 a human H2N2 strain acquired a H3 HA with only 41% identity while retaining 6 of the other 7 genes (Palese 2004).

A fundamental question in protein evolution, which is also of public-health relevance in the case of influenza HA, is how this sequence divergence on a nearly fixed structural backbone affects tolerance to further mutations. Amino acids have different propensities for certain secondary structures (Muñoz and Serrano 1994; Chou and Fasman 1978) or hydrophobic versus polar environments (Tien et al. 2013; Miller et al. 1987), and so when the backbone structure is preserved there is expected to be some conservation of the preferences of specific sites for different amino acids. But even on a nearly fixed backbone, sequence divergence can result in epistatic shifts in a site's amino-acid preferences due to changes in side-chain interactions or subtle shifts in the backbone (Starr and Thornton 2016; Pollock et al. 2012). Experiments have shown that mutations often have similar effects in closely related homologs albeit with occasional epistasis (Haddox et al. 2018; Moulana et al. 2025; Starr, Greaney, Hannon, et al. 2022), but that the consistency of effects of mutations decays as the homologs become more diverged (Park et al. 2022). Over long evolutionary times, these shifts can lead to coevolution at structurally interacting sites (Ovchinnikov et al. 2014; Marks et al. 2012, 2011; Weigt et al. 2009).

Here we compare the effects of mutations across subtypes of influenza HA, which as mentioned above exemplifies a protein with high sequence divergence but exceptionally conserved structure and function. To do this, we first use pseudovirus deep mutational scanning (Dadonaite, Crawford, et al. 2023) to measure how all mutations to the ectodomain of a H7 HA affect its ability to mediate virion entry into cells. We then compare these measurements to previously generated experimental data for H5 and H3 HAs (Dadonaite et al. 2024; Yu et al. 2025). Overall, we find that the evolutionary divergence of HA across subtypes has led to dramatic changes in mutational tolerance via rewiring of residue interactions. In addition, the new H7 HA deep mutational scanning data reported here provides sequence-function information that can help inform vaccine immunogen design and viral surveillance (Dadonaite et al. 2024; Yu et al. 2025) for an influenza subtype of potential relevance to public health.

4.3 Results

4.3.1 Phylogenetic and structural comparison of HAs from H3, H5, and H7 subtypes

The HA proteins of influenza A viruses of different subtypes are highly diverged at the sequence level, with amino-acid identities among different subtypes ranging from 36% to 80% (**Fig. 4.1A,B**). Here we focus on HAs from three subtypes: H3, H5, and H7 (**Fig. 4.1A**). H3 and H7 HAs have 47% amino-acid identity, while the H5 HA is more diverged from these other two, with identities of 40% and 43% to H3 and H7, respectively (**Fig. 4.1B**).

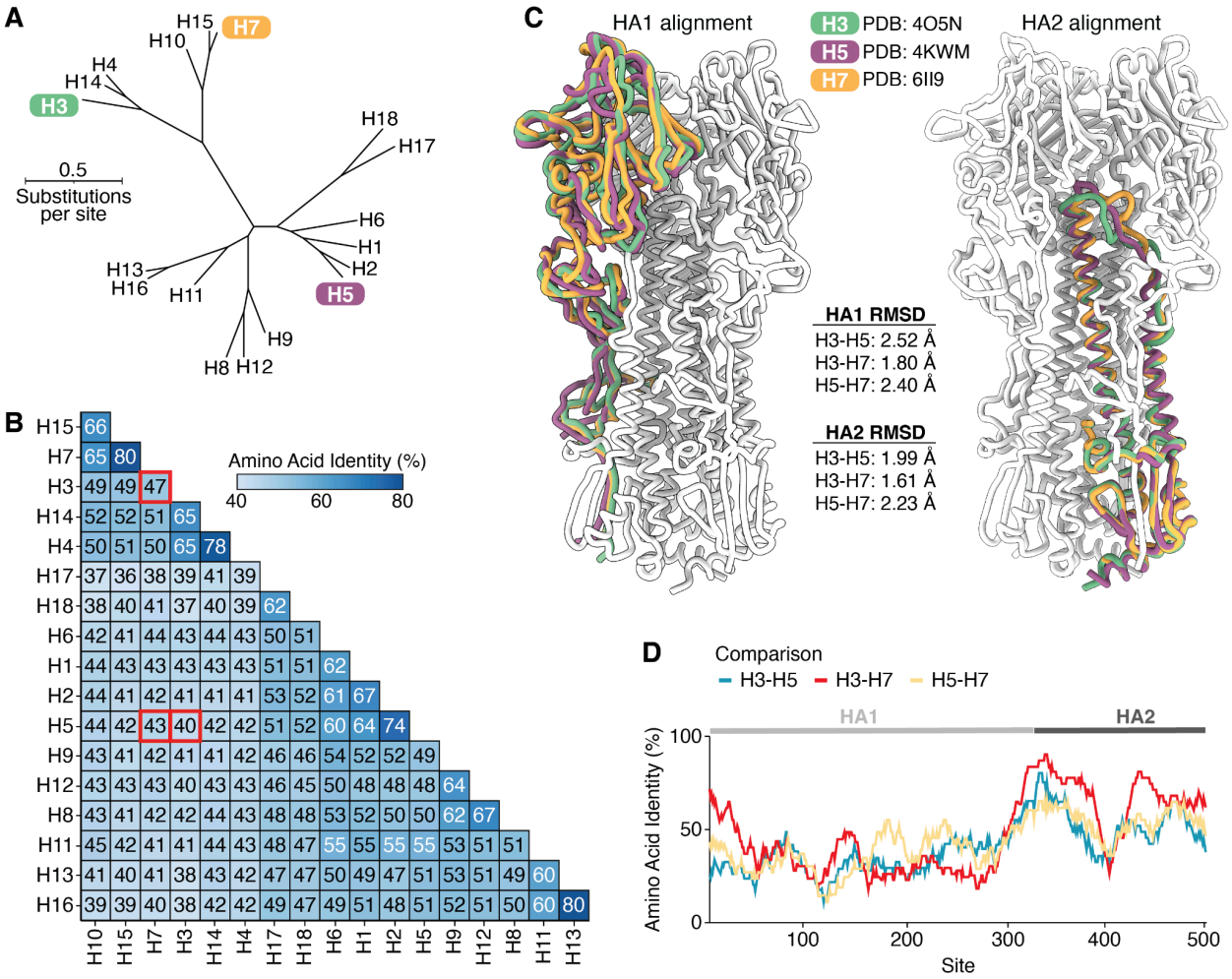


Figure 4.1 | Phylogenetic and structural comparison of H3, H5, and H7 HAs.
A) Maximum-likelihood phylogenetic tree of HA protein sequences inferred using IQ-TREE(Nguyen et al. 2015) with the Jones-Taylor-Thornton amino-acid substitution model with rate variation. **B)** Pairwise amino acid identity between the HAs shown in **A**, with red boxes indicating comparisons between the three subtypes that are the focus of the experimental work in this study. **C)** Structural alignments of single HA1 or HA2 monomers of H3 (green), H5 (purple), and H7 (orange) HAs. The rest of the H3 HA trimer (white) is included to provide structural context. HA1 and HA2 were aligned separately due to a shift in the relative orientation of these domains (see **Fig. S4.1**). The root mean square deviation (RMSD) of carbons between the aligned HA1 and HA2 monomers are also reported. **D)** Pairwise amino acid identities between the different HA subtypes computed along the length of the primary sequence with a sliding window of 30 residues. The HA structures used in this figure are PDB 6II9 (H7)(Huang et al. 2019), PDB 4O5N (H3)(Lee et al. 2014), and PDB 4KWM (H5)(Shore et al. 2013).

Despite their extensive sequence divergence, H3, H5, and H7 HAs have highly conserved structures (**Fig. 4.1C**). The root mean square deviations of the structurally aligned backbones of the two HA polypeptide chains (HA1 and HA2) are within 2.5 angstroms across experimentally determined structures of HAs from the three subtypes (**Fig. 4.1C**), although there is some modest shift in the relative orientation of HA1 and HA2 (**Fig. S4.1A,B**). All three HAs also have conserved biological functions of binding to sialic acid receptors and then mediating membrane fusion at acidic pH (Bouvier and Palese 2008; Wu and Wilson 2020; Gamblin et al. 2021). Viruses with H3 HAs have circulated in a variety of avian and mammalian species, including humans, dogs, and horses(Parrish et al. 2015). While avian H3 HAs preferentially bind α2-3 linked sialic acids, some mammalian-adapted H3 strains (including the H3N2 strains that are currently endemic in humans) preferentially bind to α2-6 linked sialic acids (Matrosovich et al. 2000). Viruses with H5 and H7 HAs mostly circulate in avian species, but they have caused sporadic infections of humans and other mammalian species (Shi et al. 2023), and a H5N1 strain recently became endemic in dairy cattle (Baker et al. 2025). Avian H5 and H7 strains preferentially bind α2-3 linked sialic acids, although there are known mutations that can shift this binding preference from α2-3 to α2-6 linked sialic acids (Tharakaraman et al. 2013; Vries et al. 2017; Linster et al. 2014; Ayora-Talavera et al. 2009; Lin et al. 2024).

The sequence divergence among H3, H5, and H7 HAs is asymmetrically distributed across the protein (**Fig. 4.1D**). The fusion peptide spanning sites 330-350 (throughout this paper, we use mature H3 numbering) is the most conserved region, consistent with its key function in mediating membrane fusion. Sites in HA1 tend to be more variable among the HA subtypes than sites in HA2, reflecting HA1's higher mutational tolerance (Thyagarajan and Bloom 2014; Heaton et al. 2013) and positive selection for HA1 mutations that reduce antibody recognition(Broecker et al. 2018; Popova et al. 2012).

4.3.2 Comparing amino-acid preferences across H3, H5, and H7 HAs

We next compared the H7 HA deep mutational scanning dataset described above to previously published H3 and H5 datasets (Dadonaite et al. 2024; Yu et al. 2025) to assess how the effects of amino-acid mutations on HA's cell entry function differ across these subtypes. To make these comparisons, we converted the measured effects of mutations on cell entry to normalized values quantifying the preference of each site for each amino acid. Quantitatively, the preference $\pi_{r,a}$ of site r for amino acid a is defined as proportional to the exponential of the effect $x_{r,a}$ of that mutation on cell entry, namely as $\pi_{r,a} = exp(x_{r,a}) / \left(\sum_{a'} exp(x_{r,a'})\right)$ where the sum over a' is taken over all amino-acid identities. If all amino-acid mutations except the wildtype identity are highly deleterious at a site then only that wildtype amino acid is preferred; when many mutations are well tolerated at a site then many amino acids have similar preferences (**Fig. 4.2A**). There are two advantages of converting the measured effects of

mutations to amino-acid preferences for comparing across subtypes. First, unlike the mutation effects, the amino-acid preferences at a site are not defined relative to its wildtype amino-acid identity and so facilitate direct comparison across H3, H5, and H7, which often have different amino acids at the same site (**Fig. 4.1B**). Second, quantitative

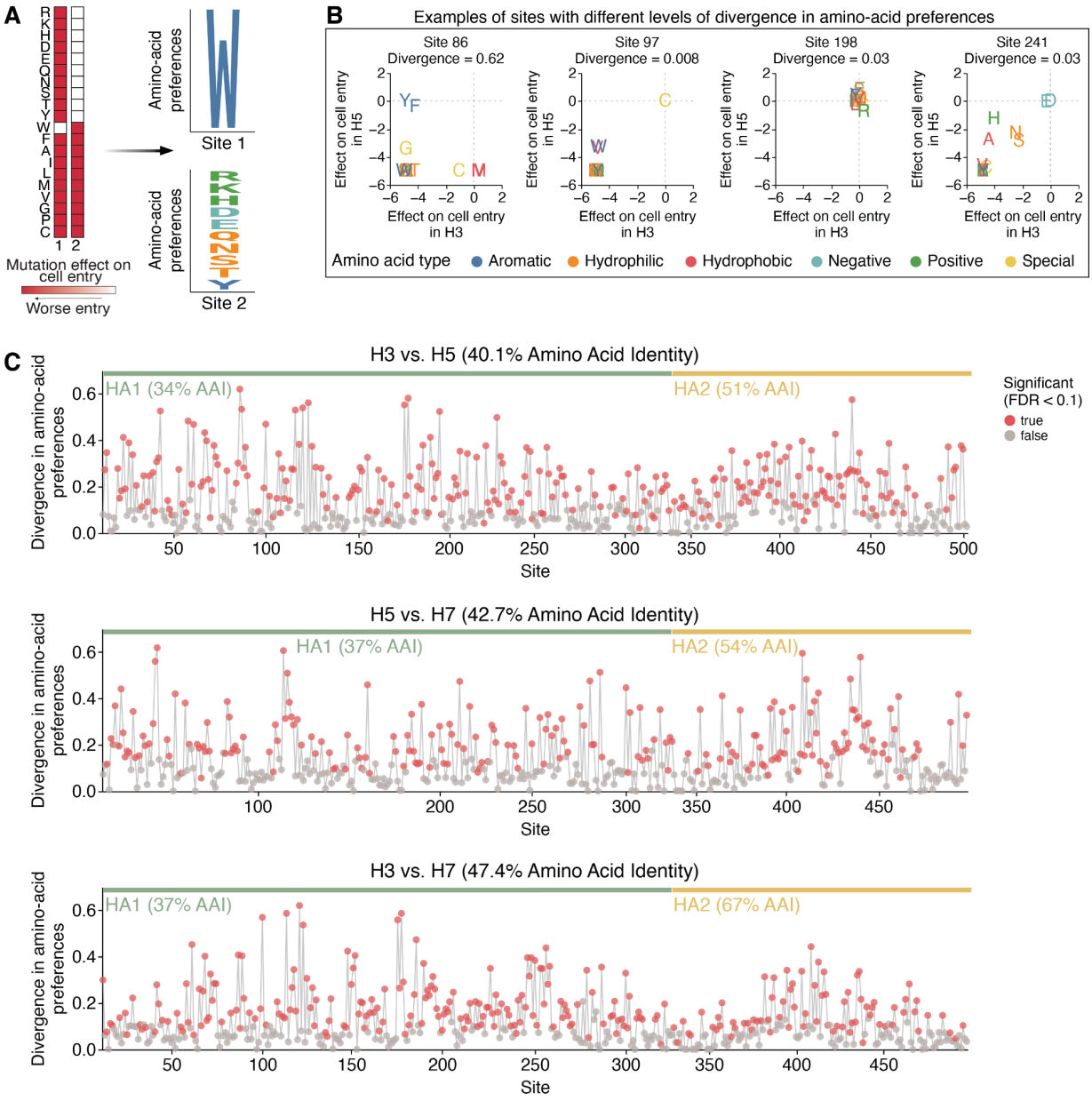


Figure 4.2 | Sequence constraints differ at many sites among H3, H5, and H7 HAs

A) Mutation effects on cell entry at a site are converted into amino-acid preferences, and the amino-acid preferences, which can be plotted as logo plots where the height of the letter is proportional to the preference of that site for that amino acid. **B)** Examples of sites with different Jensen-Shannon divergence in amino-acid preferences between HA subtypes. The scatter plot shows the effect of mutating to each amino acid in each of the two subtypes, with the wildtype identity assigned an effect of zero. **C)** Divergence in amino-acid preferences at each site in HA between H3-H5, H5-H7, and H3-H7. Red points indicate sites where preferences have significantly diverged (false discovery rate < 0.1) between HA subtypes. See https://jbloomlab.github.io/ha-preference-shifts/htmls/combined_interactive_lineplots.html for an interactive version of this plot.

comparisons using amino-acid preferences emphasize differences in which amino acids at a site are well tolerated, whereas quantitative comparisons using directly measured mutation effects also emphasize less relevant differences between moderately and strongly deleterious amino acid identities at a site. Given the experimentally measured set of amino-acid preferences at a site in two different HA subtypes, we quantified the difference in evolutionary constraint at the site by the Jensen-Shannon divergence. This divergence is high when mutations at a site have dramatically different effects between HA subtypes, and low when mutations have similar effects in both HA subtypes (**Fig. 4.2B**).

Many HA sites display high divergence in amino-acid preferences among subtypes (**Fig. 4.2C**). Overall, 58% of sites exhibited significant divergence in amino-acid preferences between H3-H5, 49% of sites between H5-H7, and 54% of sites between H3-H7 (FDR < 0.1, see red points in **Fig. 4.2C**). At those significantly diverged sites, HAs of different subtypes often preferred distinct amino acid types (**Fig. 4.2B**, left panel), demonstrating how the biochemical determinants of mutational tolerance differs across HA subtypes. Although the experimental measurements of the effects of mutations on cell entry for the different HA subtypes were performed in different target cells (a mix of 293 cells expressing α2-3- and α2-6-linked sialic acids for H7, MDCK-SIAT1 for H3 (Yu et al. 2025), and 293T for H5 (Dadonaite et al. 2024)) that express varying ratios of α2-3- and α2-6-linked sialic acids, the use of different target cells is not a major contributor to the divergence in amino-acid preferences across subtypes. This fact is demonstrated by observing that there is very little divergence between the amino-acid preferences of the H7 HA between measurements made using 293-α2-3 versus 293-α2-6 cells (**Fig. S4.2A,B**), with only a few sites in the receptor-binding pocket exhibiting modest differences in amino-acid preferences between cell types. Collectively, these results suggest the evolutionary constraints on cell entry are dramatically different across H3, H5, and H7 HAs due to large changes in sequence that alter the preferences for specific amino acids at different sites due to their impacts on fundamental aspects of HA protein folding and function.

4.3.3 Divergence in amino-acid preferences across structural and functional regions of HA

We next examined if divergence in amino-acid preferences differs across the two polypeptide chains of HA: HA1 and HA2. HA1 contains the antigenic regions that are under positive selection from antibodies, while HA2 comprises the stem region that is under purifying selection to maintain HA stability. For the H3-H7 comparison, sites in the HA2 domain showed lower divergence than those in the HA1 domain ($p < 0.001$, two-sided Mann-Whitney-U test, **Fig. S4.3A**). In contrast, no significant differences in divergence were observed between HA1 and HA2 for the H3-H5 and H5-H7 comparisons. This pattern reflects the higher amino acid conservation (67%) of HA2 between H3-H7 versus H3-H5 and H5-H7 (**Fig. 4.2C**), which was likely maintained due to strong purifying selection against mutations at sites in the HA stem. In general, sites where amino acids are conserved between HA subtypes are typically mutationally intolerant in both subtypes (**Fig. S4.3B**). However, the finding that there is no significant difference in divergence across HA1 and HA2 between H3-H5 and H5-H7 indicates that while

the networks of interacting residues in HA2 are more preserved than those for HA1 between H3-H7, they have diverged similarly in HA1 and HA2 over the broader evolutionary split between group 1 and group 2 HAs.

Finally, to explore how divergence differs across functional domains in HA, we compared divergence across the receptor-binding pocket and antigenic regions (**Fig. S4.3C**). In contrast with the comparison across HA1 and HA2 regions, phylogenetic relatedness among HAs does not always underlie divergence at these functional domains. In the 190-helix of the receptor-binding pocket, which overlaps with antigenic epitope B (H3 classification), sites showed the highest median divergence in amino-acid preferences in the H3–H7 comparison relative to the H3–H5 and H5–H7 comparisons (**Fig. S4.3C**). In this case, divergence may instead reflect distinct selective pressures in the human and avian reservoirs that drove adaptations in receptor binding and antigenic escape. We suggest that a common pressure to maintain receptor specificity in birds convergently shaped the mutational tolerance of the 190-helix and antigenic epitope B for H5 and H7, whereas pressure for changes in receptor specificity and antibody escape in humans drove the divergence in these regions for H3.

4.3.4 Factors associated with divergence in amino-acid preferences among HA subtypes

We sought to determine what specific factors were associated with high divergence in the amino-acid preferences of the same site across HA subtypes. Sites where the wildtype amino acid identity changed between HAs tended to show significantly higher divergence compared to sites where the wildtype amino acids are conserved ($p = 0.07$ for H3-H5; $p < 0.001$ for H3-H7; $p < 0.001$ for H5-H7; two-sided Mann-Whitney-U test, **Fig. 4.3A**). This difference likely reflects the fact that conserved sites usually cannot tolerate most mutations across HAs (**Fig. S4.3B**). However, even sites with conserved wildtype amino-acid identities sometimes have different preferences between HA subtypes (**Fig. 4.3A**). We therefore examined whether additional structural or biochemical features beyond conservation could help predict a site's divergence.

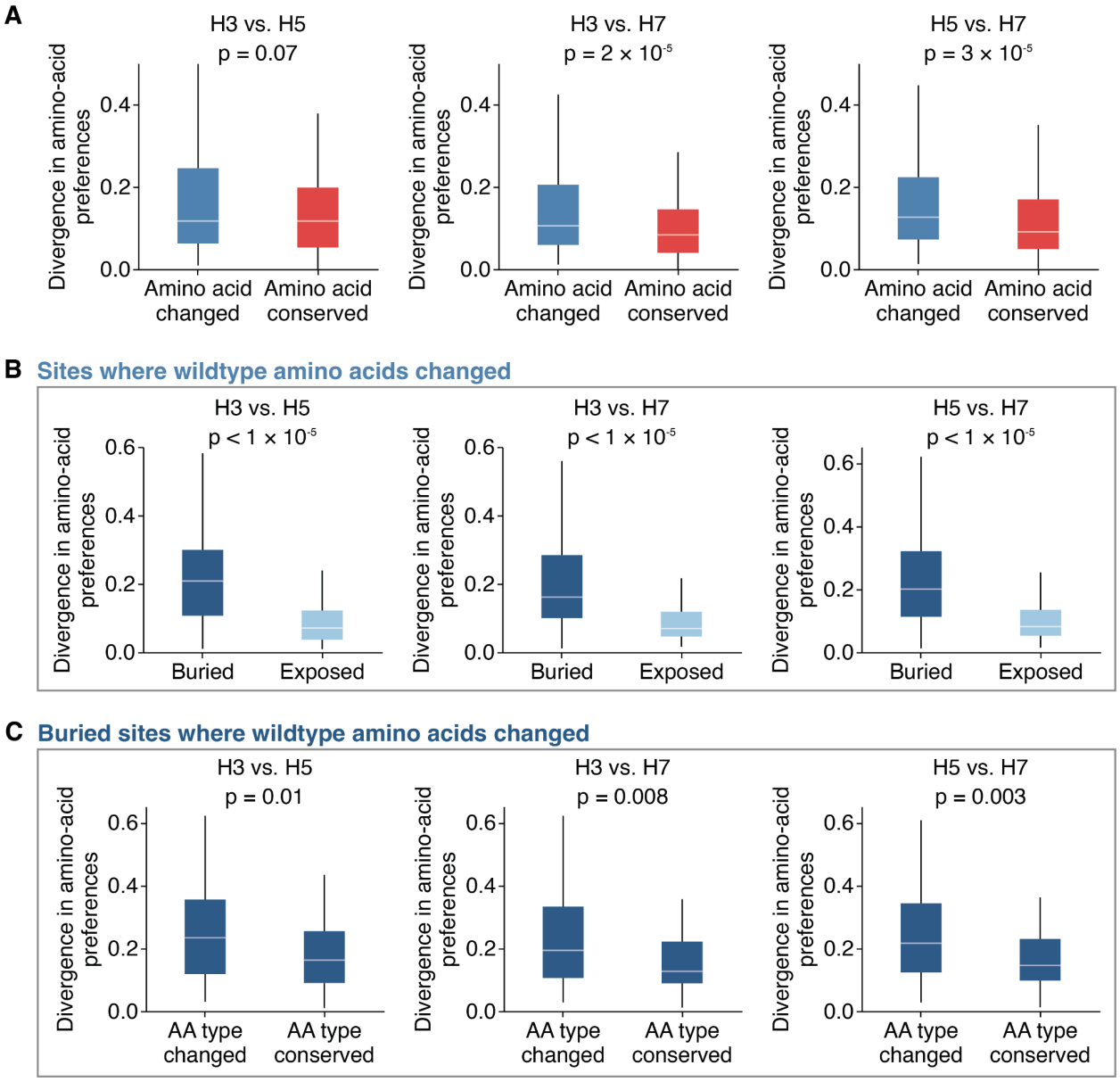


Figure 4.3 | Sites with divergent amino-acid preferences across HAs tend to be buried and have different wildtype amino-acid types.

A) Divergence in amino-acid preferences at sites where the wildtype amino-acid identity is conserved or changed between HA subtypes. **B)** Among sites where the wildtype identity changed, buried sites show significantly higher divergence in amino-acid preferences compared to exposed sites. Buried sites are defined as having a relative solvent accessibility < 0.2. **C)** Among buried sites where the wildtype amino-acid identity changed, sites where the wildtype amino-acid type is different (e.g., hydrophobic versus hydrophilic) show significantly higher divergence compared to sites where the amino acid type is conserved. Two-sided Mann-Whitney-U test was used for significance testing.

There is no correlation between the distance of Cα backbone atoms in structurally aligned HAs and divergence (**Fig. S4.4**), indicating structural shifts in the backbone residues between HA subtypes is not a major driver of divergence in amino-acid preferences. However, among sites where the wildtype amino-acid identity changed between HAs, sites that are buried (relative solvent accessibility ≤ 0.2) showed significantly higher divergence in amino-acid preferences compared to sites that are surface exposed ($p < 0.001$ for all comparisons, two-sided Mann-Whitney-U test, **Fig. 4.3B**). In contrast, among sites where the wildtype identity is conserved between HAs, solvent accessibility does not significantly explain divergence ($p > 0.1$ for all comparisons, two-sided Mann-Whitney-U test, **Fig. S4.5B**). These patterns arise because conserved sites tend to be more mutationally intolerant regardless of solvent accessibility (**Fig. S4.3B**), whereas variable sites that are exposed tend to be more mutationally tolerant in both HA backgrounds compared to variable sites that are buried (**Fig. S4.5C**). For example, the wildtype amino acids at the exposed site 173 (relative solvent accessibility = 0.85) are Q, R, and K in H3, H5, and H7, respectively, and all measured mutations are tolerated in all backgrounds, resulting in low divergence in amino-acid preferences (**Fig. S4.5D**). Together, these data suggest that among variable sites where the wildtype amino acid changed, the ones with highly diverged amino-acid preferences are typically buried within the protein.

Yet while the most divergence in amino-acid preferences tends to occur at buried sites with different wildtype amino acids between subtypes, divergence still ranges widely at such sites (**Fig. 4.3B**). What accounts for this variation? Again, structural deviation of the protein backbone at these sites is not strongly associated with divergence amino-acid preferences ($r < 0.1$ for all comparisons, **Fig. S4.4**). However, buried sites where the wildtype amino-acid identity differs in biochemical type across subtypes show significantly higher divergence compared to sites where the wildtype amino acids are of the same type ($p \leq 0.01$ across all comparisons, two-sided Mann-Whitney-U test, **Fig. 4.3C**). Therefore, the type of amino acid at a buried site where the wildtype identity changed across subtypes is indicative of divergence in amino-acid preferences.

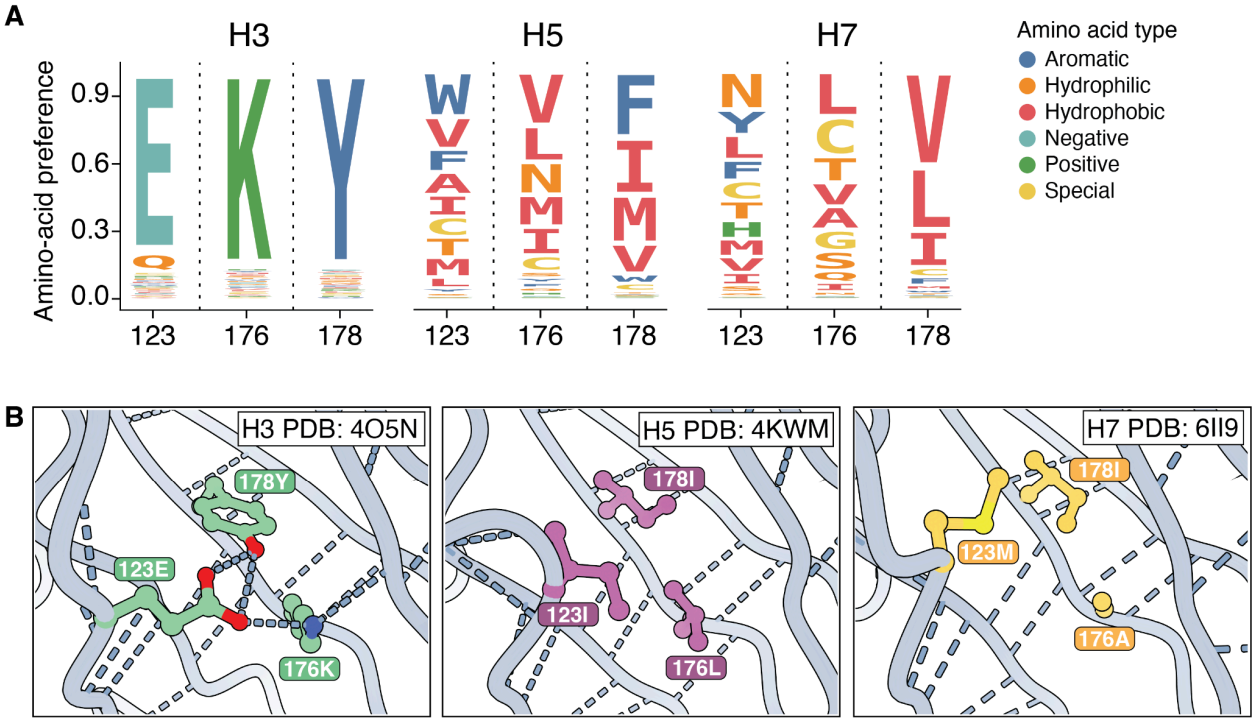


Figure 4.4 | Example of contacting sites where interactions have been rewired from a hydrogen bond network in H3 HA to a hydrophobic environment in H5 and H7 HAs.
A) Logoplots showing preferences for amino acids at buried sites 123, 176, and 178 in H3, H5, and H7. The height of each letter is proportional to the amino-acid preference measured by deep mutational scanning. Each amino acid is colored by its biochemical type. The amino-acid preferences at these sites are relatively similar between H5 and H7, but both are highly diverged from H3 HA. **B)** Sites 123, 176, and 178 form a hydrogen bond network in H3, but are part of a hydrophobic environment in H5 and H7 HAs.

The high divergence of amino-acid preferences at buried sites with different wildtype identities across HA subtypes can arise from fundamental rewiring of how sets of residues interact, even when the backbone structure has not changed. For example, the wildtype amino acids at the buried sites 123/176/178 are E/K/Y in H3, I/L/I in H5, and M/A/I in H7. The amino-acid preferences at these three sites are fairly similar between H5 and H7 HAs, but both are sharply diverged from H3 HA (**Fig. 4.4A, Fig S4.6**). This high divergence of amino-acid preferences is likely because sites 123/176/178 form a hydrogen bond network in H3 HAs, but this network is absent in H5 and H7 HAs, which instead have a more hydrophobic environment in this region (**Fig. 4.4B**). Collectively, the amino-acid preferences and protein structure show how evolution has rewired the constraints in this region: H3 relies on a highly constrained hydrogen bond network to maintain cell entry function, while H5 and H7 rely on a hydrophobic environment that does not require hydrogen bonds at all.

4.4 Discussion

Here we have measured how nearly all amino-acid mutations to the ectodomain of a H7 HA affect its cell entry function, and compared the results to previously generated data for H3 and H5 HAs. We find that the effects of mutations at many sites have substantially diverged among the HA subtypes despite the proteins having virtually superimposable backbone structures and mediating the same biochemical function. This divergence in site-specific amino-acid preferences is distributed across HA's sequence and structure, with the greatest divergence being at buried sites where the wildtype amino-acid type differs between subtypes. In some cases, the highly diverged amino-acid preferences are due to clear rewiring of how the neighboring residues interact even though the protein backbone structure remains unchanged.

It is interesting to contrast the highly diverged amino-acid preferences of different HA subtypes reported here to recent studies that have extensively examined the effects of mutations to the SARS-CoV-2 spike for different viral variants (Starr, Greaney, Hannon, et al. 2022; Starr, Greaney, Stewart, et al. 2022; Haddox et al. 2023; Moulana et al. 2022; Taylor and Starr 2024; Moulana et al. 2025). The spikes of even the most diverged SARS-CoV-2 variants are far more similar at the sequence level than the different HA subtypes studied here (current SARS-CoV-2 variants diverged over ~6 years, whereas HA subtypes diverged over many millenia), and most sites in SARS-CoV-2 spike have similar amino-acid preferences albeit with some examples of epistatic shifts in mutational effects (Moulana et al. 2025; Taylor and Starr 2024; Moulana et al. 2022; Haddox et al. 2023; Starr, Greaney, Stewart, et al. 2022; Starr, Greaney, Hannon, et al. 2022). Furthermore, most of the strong epistastic shifts in amino-acid preferences that have occurred during the short-term evolution of the SARS-CoV-2 spike involve sites that affect receptor binding (Moulana et al. 2025; Taylor and Starr 2024; Moulana et al. 2022; Starr, Greaney, Stewart, et al. 2022; Starr, Greaney, Hannon, et al. 2022), whereas the most striking changes in amino-acid preferences across the highly diverged HA subtypes involve residues buried in the protein structure that are more involved in folding and stability than receptor binding. This contrast is consistent with recent work on other proteins suggesting that over long evolutionary time

frames, pervasive shifts in site-specific amino-acid preferences across a protein accumulate due to many small epistatic interactions among fixed mutations (Park et al. 2022).

It is important to note that our experiments only measure how mutations affect HA-mediated cell entry, which may not fully capture all the ways that HA contributes to viral fitness. In particular, pseudotyped lentiviral particles are not identical in morphology to actual influenza virions, and HA can contribute to transmissibility and immune recognition in ways beyond its ability to mediate cell entry. Therefore, our findings should be interpreted specifically in terms of how mutations affect HA's cell entry function, which is related to but not identical to its contributions to viral fitness.

These results have implications for how large-scale sequence-function measurements can inform viral surveillance and vaccine immunogen design. One of the rationales for characterizing the effects of mutations to viral proteins is to provide information that can help interpret the consequences of changes that are observed during genomic surveillance of ongoing viral evolution. Indeed, this rationale is why we focused the new experimental work in this study on H7 HA, since that subtype is thought to pose a potential risk to humans (Sutton 2018). But of course viruses themselves are always evolving, and so the protein sequence that is the subject of any particular experimental study is unlikely to be identical to that of future viruses of public-health interest. Our results underscore the limits to the extent that experimental measurements made in one genetic background can predict the impacts of mutations in another genetic background. Therefore, reliable prediction of mutation effects across an entire family of a highly diverged but structurally and functionally conserved protein such as HA cannot be achieved simply by large-scale experimental studies—although such studies can be highly predictive for closely related variants of a protein as might be found in the viral strains circulating in humans at any given time.

4.5 Supplementary Materials

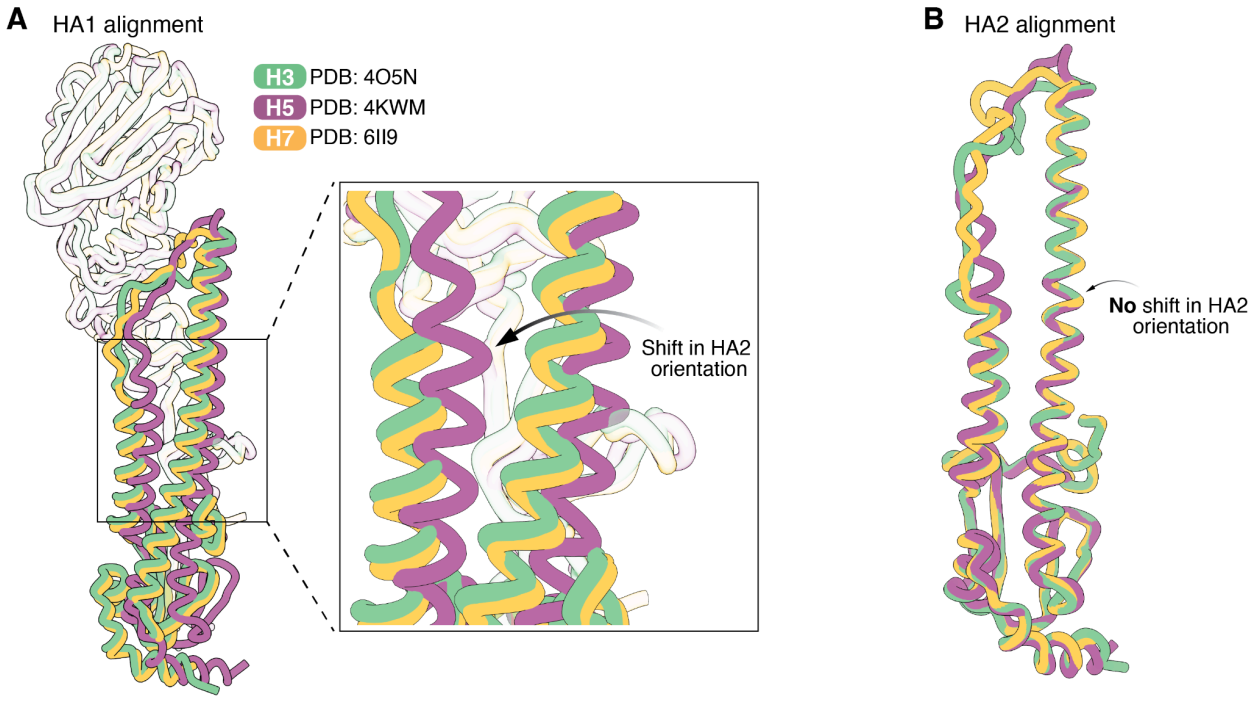


Figure S4.1 | The relative orientations of HA1 and HA2 are shifted in H5 HA relative to H3 and H7 HAs.

A) Structural alignment of the HA1 domain (same as left structure in **Fig. 4.1C**) with the HA2 domains from H3 (green), H5 (purple), and H7 (orange) HAs. Due to a tilt in the orientation of the HA1 and HA2 domains, the HA2 domain of H5 becomes shifted relative to where the HA2 domains of H3 and H7 are located. **B)** Aligning the HA2 domain alone (same as right structure in **Fig. 4.1C**) reveals the folds in HA2 are actually highly conserved across H3, H5, and H7.

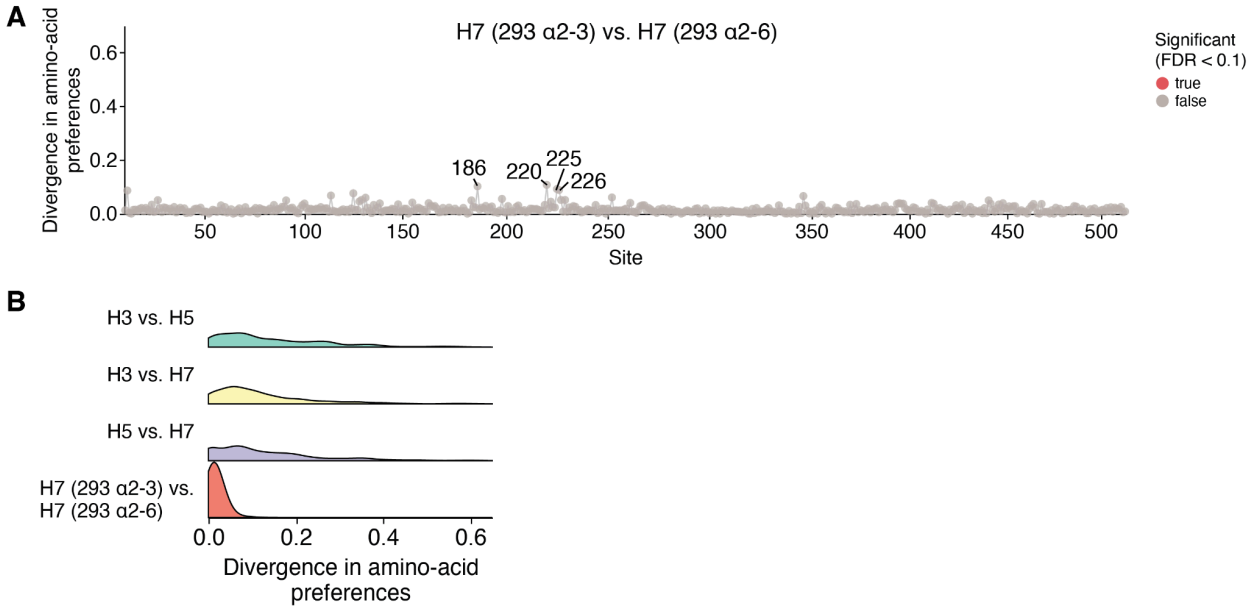


Figure S4.2 | Divergence in H7 HA amino-acid preferences for entry into 293 cells expressing α2-3 versus α2-6 linked sialic acids is minimal compared to the divergence between HA subtypes.

A) Divergence in amino-acid preferences at each site of the H7 HA as measured in 293 cells expressing only α2-3 versus only α2-6 linked sialic acids. Note that the divergence at all sites is low compared to the cross HA-subtype comparisons in **Fig. 4.2C**, and no sites are significantly diverged (false discovery rate < 0.1). The largest differences are at sites like 186, 220, 225, and 226, which are important for receptor binding. **B)** Distributions of divergence in amino-acid preferences for all sites between subtypes compared to the distribution of divergence in H7 HA for measurement made using 293 cells expressing entry only α2-3 versus only α2-6 linked sialic acids.

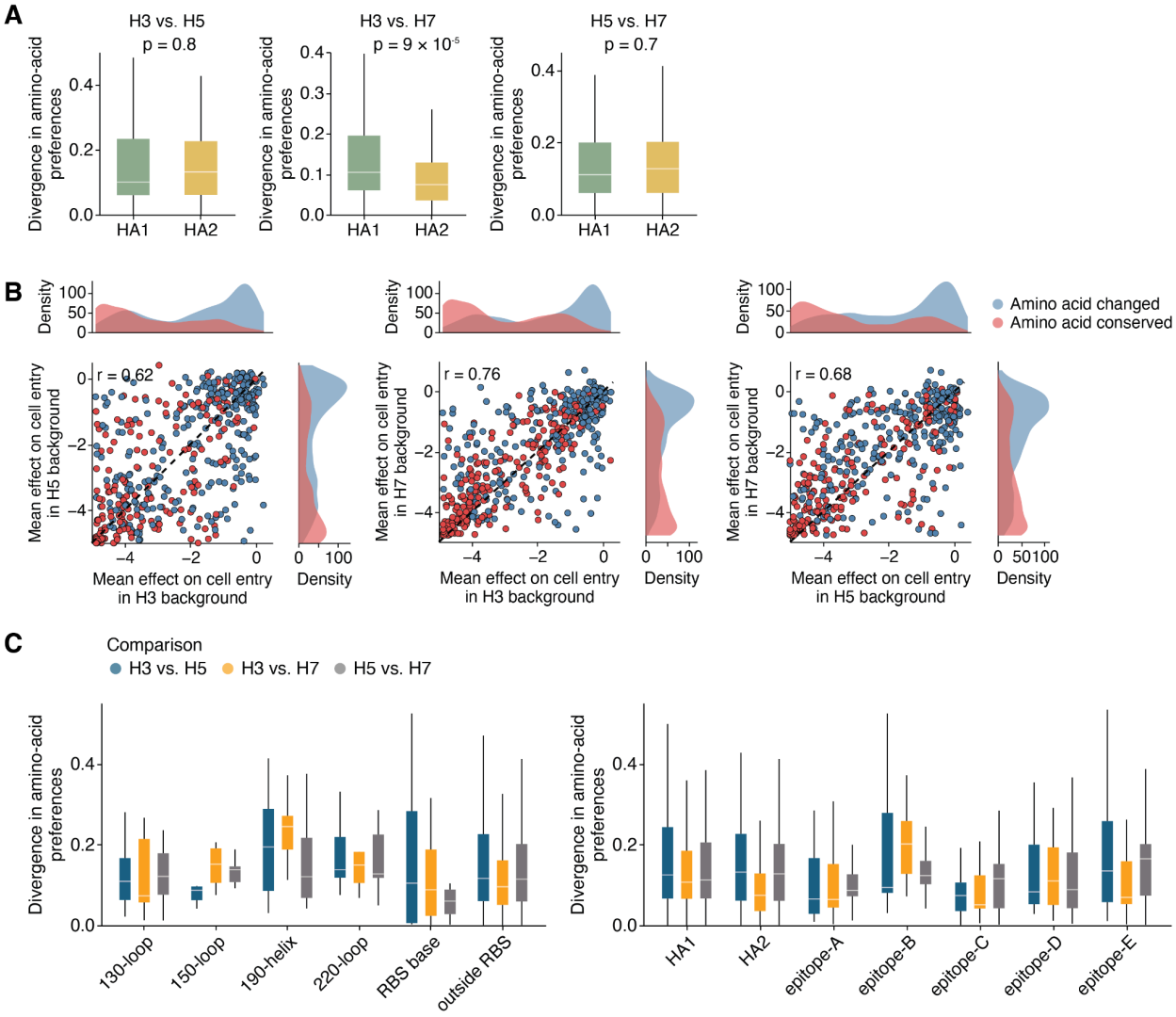


Figure S4.3 | Divergence in amino-acid preferences for different domains or regions of HA.

A) Distributions of divergence in amino-acid preferences between HA subtypes for sites in the HA1 or HA2 domains of HA. In H3-H7, HA2 sites show significantly lower divergence compared to HA1. **B)** Correlation of the mean effect of mutations at each site between each pair of HA subtypes. Sites where the wildtype amino-acid identity differs between subtypes are colored blue, while sites with conserved wildtype amino-acid identities are red. **C)** Distributions of divergence in the amino-acid preferences by receptor binding pocket region (left) and antigenic region (right) across each HA comparison.

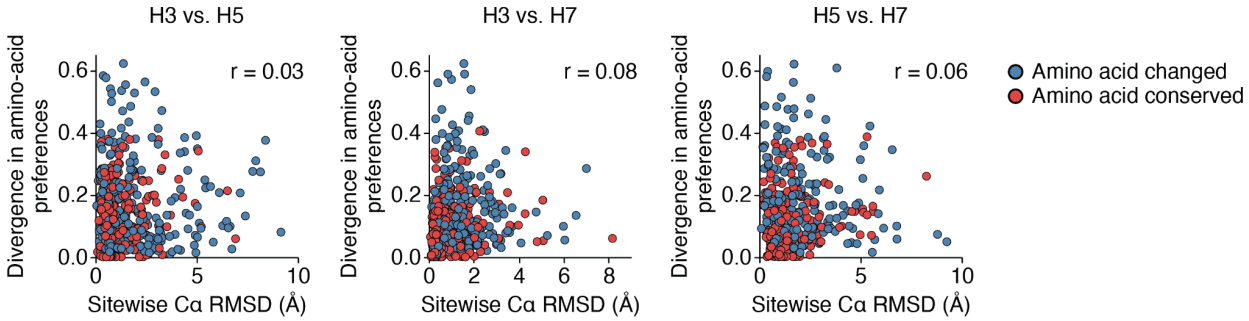


Figure S4.4 | Structural deviation is a poor predictor of divergence in amino-acid preferences at a site.

Correlation between divergence in amino-acid preferences and the Cα root mean square deviation at each site in the structurally aligned backbones of HA1 and HA2 across pairwise HA comparisons. Sites where the wildtype amino acids changed between subtypes are colored blue, while conserved sites are red.

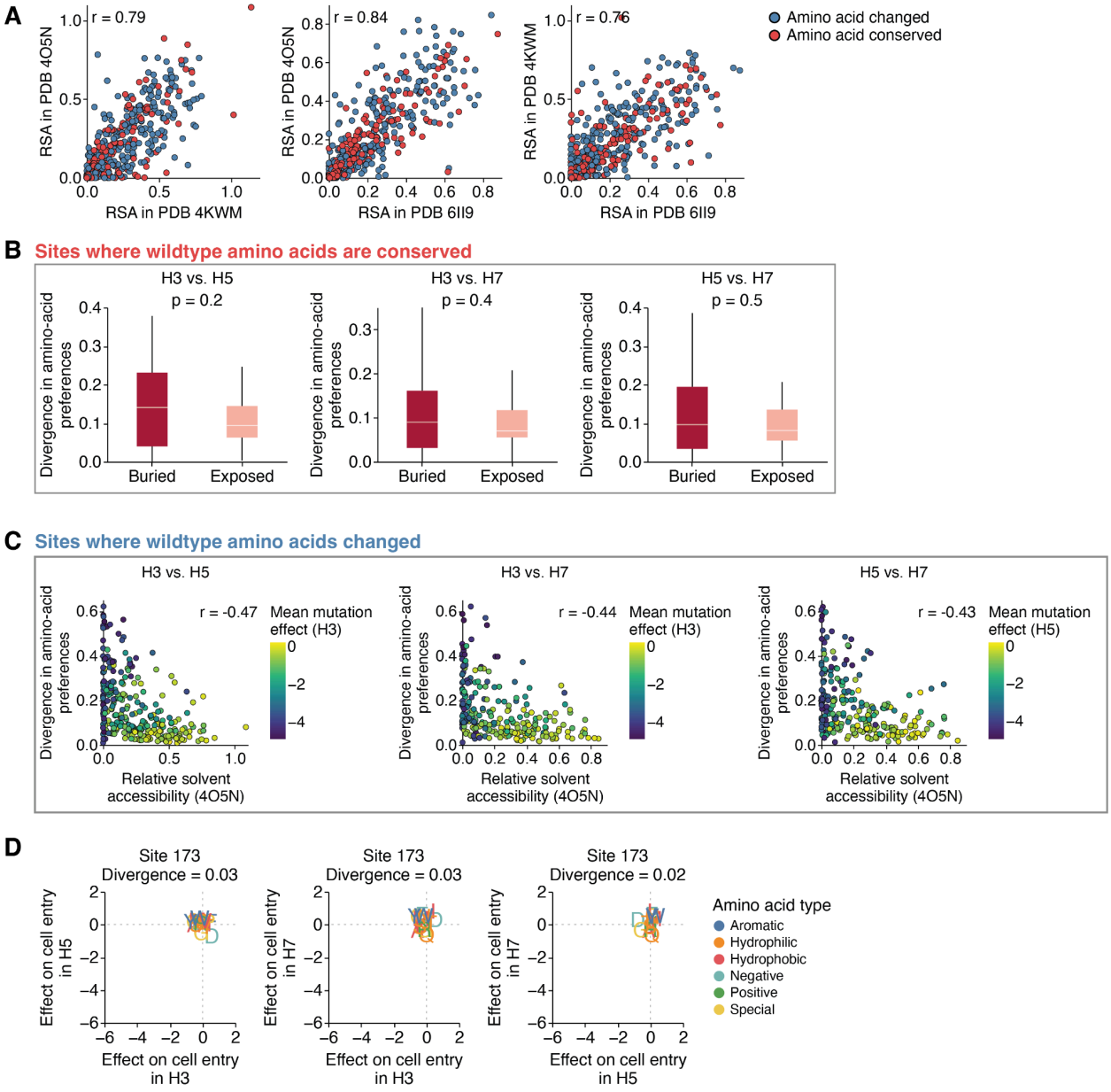


Figure S4.5 | Surface exposed sites are more mutationally tolerant and display less divergent amino acid preferences across HAs.

A) Correlation between relative solvent accessibility values at structurally aligned sites between HAs. These values quantify how exposed or buried a site is, and the correlation is strong ($r >$ 0.75) because the HA structures are conserved. PDB accessions are 4O5N (H3), 4KWM (H5), 6II9 and (H7). **B)** Among sites where the wildtype amino-acid identity is conserved (red boxplots in **Fig. 4.3A**), there is no significant difference in divergence in amino-acid preferences at buried versus exposed sites. **C)** Among sites where the wildtype amino-acid identity changed (blue boxplots in **Fig. 4.3A**), there is a negative correlation between divergence in amino-acid preferences and relative solvent accessibility. Sites are colored by mean mutation effect on entry in the indicated HA background, showing that sites that are more exposed tend to tolerate

mutations better. **D)** Correlations of effects of mutations on cell entry between HAs at the exposed site 173, showing how all measured mutations are tolerated across HAs at this site. Two-sided Mann-Whitney-U test was used for all significance testing.

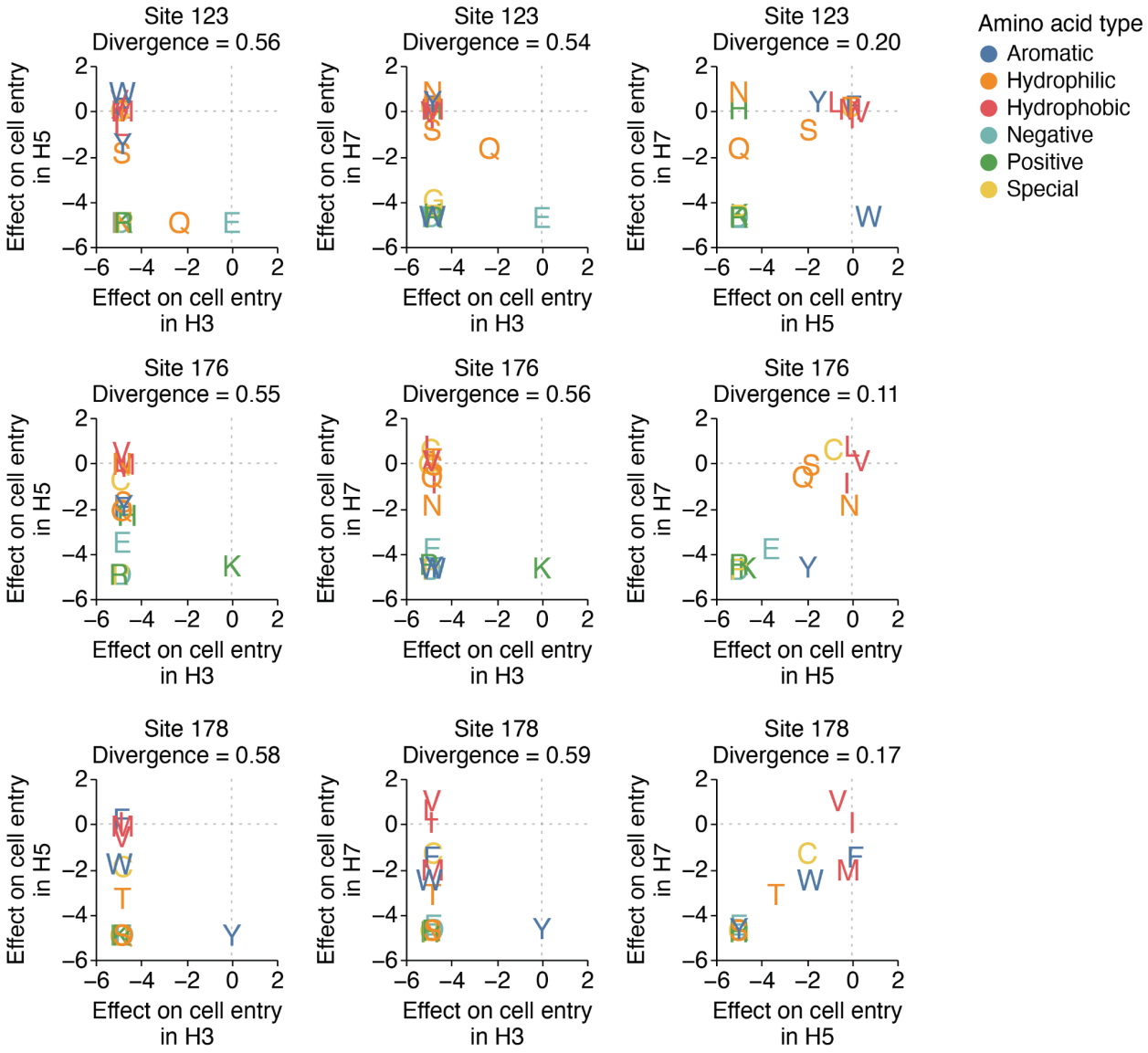


Figure S4.6 | Comparison of mutation effects across H3, H5, and H7 at buried sites 123, 176, and 178.

Correlation of mutation effects on cell entry as measured by deep mutational scanning between pairs of HAs. The letters show the effect of mutation to each amino acid in each of the two indicated HA homologs. The mutation effects at these sites are relatively similar between H5 and H7, but both are highly diverged from H3 HA.

4.6 Materials and methods

Code and data availability

The computer code for the deep mutational scanning of the H7 HA described in this manuscript is available on GitHub at https://github.com/dms-vep/Flu_H7_Anhui13_DMS and can be viewed interactively at https://dms-vep.org/Flu_H7_Anhui13_DMS/ via GitHub Pages. The quality filtered measurements of mutation effects on cell entry for the H7 HA are available in tabular format at https://github.com/dms-vep/Flu_H7_Anhui13_DMS/blob/main/results/summaries/entry_all_cells.csv.

A single CSV with the effects of mutations at each site on entry for the H7 HA measured in this study as well as the previously measured values for the H5 and H3 HAs(Dadonaite et al. 2024; Yu et al. 2025) is at https://github.com/jbloomlab/ha-preference-shifts/blob/main/results/combined_effects/combined_mutation_effects.csv.

- The computer code for comparing mutation effects across H3, H5, and H7 HAs is on GitHub at https://github.com/jbloomlab/ha-preference-shifts and an interactive plot showing divergence in amino-acid preferences across the three HAs is at https://jbloomlab.github.io/ha-preference-shifts/htmls/combined_interactive_lineplots.html.

Biosafety

All experiments reported in this paper used lentiviral particles pseudotyped with HA and NA. These particles encode no viral genes other than HA and so are only able to undergo a single round of cell entry, meaning that they are not fully infectious viruses and so are not human pathogens.

Cell lines and media

HEK293 isogenic α2-3 and α2-6 cells were stably engineered as previously described (Büll et al. 2021; Narimatsu et al. 2019). The α2-3 cells are 293 cells with combinatorial knockout of ST3GAL3, ST3GAL4, ST3GAL6, ST6GAL1, and ST6GAL2 genes and knock-in of the human ST3GAL4 gene. The α2-6 cells are 293 cells with knockout of ST3GAL3, ST3GAL4, ST3GAL6, ST6GAL1, and ST6GAL2 genes and knock-in of the human ST6GAL1 gene. It is important to note that both cell lines express α2-3 and α2-6-linked sialic acids on O-linked glycans.

These cells were maintained in D10 media (Dulbecco's Modified Eagle Medium with 10% heat-inactivated fetal bovine serum, 2 mM l-glutamine, 100 U/mL penicillin, and 100 µg/mL streptomycin). To suppress rtTA activation, 293T-rtTA cells were grown in tet-free D10, which is made with tetracycline-negative fetal bovine serum (Gemini Bio, Ref. No. 100-800) instead. Additionally, a phenol free version of D10 was used for library production and rescue. Influenza growth medium was also used during library production (IGM, Opti-MEM supplemented with

0.01% heat-inactivated fetal bovine serum, 0.3% bovine serum albumin, 100 µg/mL of calcium chloride, 100 U/mL penicillin, and 100 µg/mL streptomycin).

Structural alignment of H3, H5, and H7 HAs

Experimentally determined HA protein structures were obtained from the Protein Data Bank (accession IDs 4O5N for H3, 4KWM for H5, and 6II9 for H7). Foldmason (Gilchrist et al. 2024) was used to structurally align the HA1 and HA2 domains of the three HAs. Separate alignment of the domains was necessary due to the relative shift in HA1/HA2 orientation shown in **Fig. S1**. The alignment results are at:

- HA1: https://github.com/jbloomlab/ha-epistasis/blob/main/results/foldmason_alignment/chain_A/result_aa.fa
- HA2: https://github.com/jbloomlab/ha-epistasis/blob/main/results/foldmason_alignment/chain_B/result_aa.fa

Jensen-Shannon divergence in amino-acid preferences

For reasons described in the main text, amino-acid preferences are more suitable for comparing effects of mutations across HAs. Therefore, for each site r and amino-acid identity a, we converted the mutation effect on cell entry $x_{r,a}$ to an amino-acid preference by:

$$\pi_{r,a} = exp(x_{r,a}) / \left(\sum_{a'} exp(x_{r,a'})\right)$$

where $\pi_{r,a}$ is the preference of site r for amino-acid identity a, and the sum in the denominator is over all amino-acid identities a' at that site. These amino-acid preferences can be interpreted as the probability of observing the amino-acid identity a at a site r in an evolutionarily equilibrated alignment evolving according to the single-mutant effects measured in the deep mutational scanning(Bloom 2014; Hilton et al. 2017).

To compare the amino-acid preferences between HAs, we quantified the Jensen-Shannon divergence (JSD) between their amino-acid preference probability distributions at each site. Let P be the amino-acid preferences at a site in one HA and Q be the amino-acid preferences at the same site in another HA. Then, the JSD at the site is the average Kullback-Leibler divergence (KLD) of P and Q from their per-index-mean vector M:

$$JSD(P||Q) = \frac{KLD(P||M) + KLD(Q||M)}{2}$$

where the KLD is calculated as

$$KLD(P||m) = \sum_{a} P_a log\left(\frac{P_a}{M_a}\right)$$

We calculated the Jensen-Shannon divergence between HAs at every site where there were at least 10 amino acids measured in both HA backgrounds.

Significance testing for divergent amino-acid preferences between HAs

We implemented a simulation approach to test for significant divergence in amino-acid preferences between HAs. Under the null hypothesis that the true mutation effects (and therefore preferences) are identical across HAs and that any divergence is due only to measurement uncertainty, we generated a simulated null distribution for the JSD at each site.

For a given background X (e.g., H3), we drew two independent replicate measurements of each mutation effect by $\tilde{x}_{r,a} \sim N\left(x_{r,a}, \sigma_{r,a}^{2}\right)$, where $x_{r,a}$ is the measured mutation effect and $\sigma_{r,a}$ is the population standard deviation across replicate measurements made for that HA (note that all cell entry effects for all three HAs were made in replicate with this paper typically just reporting the median value of the replicate measurements for each mutation effect). For each site, the sampled mutation effects across amino acids were converted into two amino-acid preference vectors, and the JSD between these vectors was computed. This procedure was repeated 1000 times to yield a null distribution $\left\{JSD_{null,X}\right\}$. The same procedure was performed for background Y (e.g., H5) to obtain $\left\{JSD_{null,Y}\right\}$. To account for potentially different levels of noise in the two backgrounds, the final null distribution was defined as the elementwise average of the two distributions, $JSD_{null} = \left(JSD_{null,X} + JSD_{null,Y}\right) / 2$.

For each site, we computed an empirical one-sided p-value as the fraction of null JSD values greater than or equal to the observed JSD. We controlled for multiple testing across sites by the Benjamini-Hochberg false discovery rate (FDR) procedure and considered sites as significant at FDR < 0.1.

Note this approach makes two assumptions about the null distribution. First, it assumes Gaussian noise in log-space. Second, it assumes the standard deviations can be faithfully derived from 2-4 replicate measurements. Therefore, a limitation is that these standard deviation estimates could be noisy, leading to false positives and negatives at sites with modest divergence values. However, clear separation between significant and non-significant sites (when divergence > 0.2) demonstrates that sites with larger divergence are robustly called significant despite potential uncertainty in the standard deviation estimates (**Fig. 3C**).

Structural analysis

UCSF ChimeraX v.1.10.1 (Meng et al. 2023) was used for structural analysis and visualizations. In **Fig. S7**, we analyzed structural deviation between aligned HAs. The structures were aligned by Foldmason as described above and the distances between corresponding Cα atoms were computed by the rmsd function in ChimeraX. In **Fig. S3.8**, we analyzed the surface accessibility

of sites in HA. The biological assemblies of each HA were used as input for the Define Secondary Structure of Proteins (DSSP) program (Hekkelman et al. 2025) to calculate the absolute solvent accessibility values of each HA site. These values were converted to relative solvent accessibility values based on the maximum allowed solvent accessibilities of residues reported in Tien et al. 2013 (Tien et al. 2013). In this work, we define a site as buried if its relative solvent accessibility < 0.2.

Chapter 5

A version of this chapter is in progress:

> Dadonaite, Bernadeta, Jenny J. Ahn, et al. 2026. "**Structural Basis of MHC Class II Binding by Influenza Hemagglutinin**."

5 Deep Mutational Analysis of H7 MHCII-Dependent Entry Mechanisms

5.1 Abstract

Influenza A viruses traditionally enter host cells through sialic acid receptors, but recent discoveries have revealed that certain subtypes can utilize Major Histocompatibility Complex class II (MHCII) molecules as alternative entry receptors. While bat-origin H17/H18 and duck-origin H19 subtypes exclusively use MHCII for cellular entry, the molecular determinants governing this interaction remain poorly understood. Here, we employed deep mutational scanning to systematically map the functional requirements for MHCII-mediated entry using H7 hemagglutinin as a model system. Our comprehensive mutational analysis revealed distinct constraint patterns compared to traditional sialic acid receptor binding, with the canonical receptor binding pocket showing remarkably relaxed functional constraints in MHCII-expressing cells. Instead, we identified alternative sites in the HA head domain (positions 131, 67, 247, 138, and 139) that displayed strong functional constraint, suggesting these residues may comprise a novel MHCII binding interface. Additionally, specific regions in the HA stalk (sites 330-337 and 408-470) showed critical functional importance for MHCII-mediated entry. These findings demonstrate that influenza-MHCII interactions utilize binding determinants distinct from the traditional sialic acid pathway and provide the first systematic map of molecular constraints governing this alternative entry mechanism. Understanding these requirements has important implications for predicting viral host range and pandemic potential, as MHCII-mediated entry may represent an underappreciated pathway for cross-species transmission that operates independently of well-characterized sialic acid receptor barriers.

5.2 Introduction

Influenza A viruses have traditionally been understood to enter host cells primarily through interactions with sialic acid receptors displayed on the cell surface. The specific linkage of these sialic acid structures plays a crucial role in determining influenza host range, as

avian-adapted influenza viruses demonstrate a preference for α2-3 linked sialic acids while human-adapted strains preferentially bind α2-6 linked variants (Stevens et al. 2006; Rajao et al. 2025; Cardenas et al. 2026). This receptor specificity means that mutations occurring in or near the hemagglutinin (HA) receptor binding pocket can significantly modulate the binding preferences of the virus, potentially altering host tropism and transmission patterns (Glaser et al. 2005; Nobusawa et al. 2000; Obadan et al. 2019; Dadonaite, Crawford, et al. 2023; Ahn et al. 2026). However, recent discoveries have revealed that influenza receptor usage is more diverse than previously appreciated, with several subtypes demonstrating the ability to utilize alternative entry pathways.

A paradigm shift in our understanding of influenza entry mechanisms emerged with the discovery that certain influenza subtypes can utilize Major Histocompatibility Complex class II (MHCII) molecules as their primary entry receptor. Bat-origin subtypes H17 and H18, along with duck-origin H19, have been shown to exclusively use MHCII for cellular entry, completely bypassing the traditional sialic acid receptor pathway (Giotis et al. 2019; Karakus et al. 2019, 2024). This discovery has opened new avenues for understanding influenza host range determination and has important implications for pandemic preparedness, as it suggests that influenza viruses may have access to entry mechanisms that are independent of the well-characterized sialic acid pathway.

MHCII molecules represent sophisticated antigen presentation machinery that plays fundamental roles in adaptive immunity. These heterodimeric transmembrane proteins are essential components of the immune system, responsible for presenting antigenic peptides derived from extracellularly acquired proteins on the surface of antigen-presenting cells to CD4+ T lymphocytes (Roche and Furuta 2015; Rock et al. 2016; Wieczorek et al. 2017). The structural architecture of MHCII molecules consists of distinct alpha (α) and beta (β) polypeptide chains, each containing membrane-proximal, barrel-shaped immunoglobulin-like domains (α2 and β2) and membrane-distal domains (α1 and β1) that contribute approximately equally to the formation of the peptide-binding groove (Rock et al. 2016; Brown et al. 1993; Unanue et al. 2016). This peptide-binding groove represents the functional core of the MHCII molecule and is critical for its role in antigen presentation.

The cellular trafficking and maturation of MHCII molecules involves a highly regulated pathway that ensures proper peptide loading and surface expression. MHCII molecules initially fold within the endoplasmic reticulum (ER), where they associate with the invariant chain, a chaperone protein that facilitates their transport to late endosomal compartments (Rock et al. 2016; Wieczorek et al. 2017). Within these specialized compartments, the invariant chain undergoes proteolytic degradation, allowing MHCII molecules to be loaded with antigenic peptides. This peptide loading process is facilitated by HLA-DM, a specialized chaperone molecule that represents a nonclassical MHCII variant (Wieczorek et al. 2017; Sanderson et al. 1994). HLA-DM resides within late endosomal compartments and exhibits high structural similarity to classical MHCII molecules, though it notably lacks a functional peptide-binding groove (Wieczorek et al. 2017; Mosyak et al. 1998). Following successful peptide loading, the

mature peptide-MHCII complexes are trafficked to the plasma membrane, where they can fulfill their antigen presentation function (Bosch et al. 2013).

The utilization of MHCII as an entry receptor by H17-H19 subtypes represents a complete departure from sialic acid-dependent entry mechanisms. Comprehensive glycan array studies have definitively demonstrated that H17-H19 hemagglutinins fail to bind any of the extensive panel of sialic acid-containing glycans tested, confirming that their entry pathway is entirely independent of traditional sialic acid receptors (Karakus et al. 2024; Sun et al. 2013). Furthermore, H17 viruses exhibit pronounced species-specific preferences for different MHCII variants, with human HLA-DR (human leukocyte antigen isotype DR) molecules enabling efficient infection of otherwise non-permissive cells, while the swine homologue (SLA-DR) provides significantly reduced support for viral entry(Karakus et al. 2019). This species specificity suggests that subtle structural differences between MHCII molecules from different species can dramatically influence viral tropism.

Detailed molecular analysis has begun to reveal the specific determinants of MHCII-mediated influenza entry. Studies of H18N11 infection have identified conserved residues within the α1, α2, and β1 domains of HLA-DR molecules as critical for supporting efficient viral infection of human cells (Olajide et al. 2023). These findings suggest that MHCII molecules may play previously unrecognized roles in determining influenza host range specificity, potentially providing an additional layer of species barrier that operates independently of sialic acid receptor distribution.

While the precise binding site for H18 hemagglutinin on MHCII molecules remains to be determined, functional studies have revealed that classical MHCII molecules from all vertebrate species tested to date can support H18-mediated viral entry (Karakus et al. 2019). This broad cross-species compatibility suggests that highly conserved structural features of MHCII molecules, rather than species-specific variations, are responsible for facilitating viral entry. Understanding which specific residues and structural elements are required for this interaction represents a critical gap in our knowledge of MHCII-mediated influenza entry. To address this question and systematically identify the key molecular determinants of MHCII-influenza interactions, we employed deep mutational scanning approaches to comprehensively map the functional requirements for this alternative entry pathway.

5.3 Results

5.3.1 Effects of HA mutations on entry into cells expressing tufted duck MHCII receptors

To investigate whether H7 is able to use MHCII receptors to infect cells, we used our H7 pseudovirus DMS library (produced as described in chapter 3). We measured entry into cells with no sialic acid receptors that were engineered to express the tufted duck MHCII receptor. We performed two experimental replicates for each of the two independent pseudovirus

mutants libraries, and the measured effects of the mutations on cell entry were highly correlated across both replicates of the same library and between the two independent libraries (**Fig. S5.1**); throughout this paper we report the median of the measurements across all replicates. The effects of nearly all mutations to the H7 HA ectodomain on entry into cells expressing the tufted duck MHCII are shown in (**Fig 5.1A**) (see also the interactive heatmap at https://dms-vep.org/Flu-H7-Anhui13-MHCII-binding/htmls/293_noSA_tufted_duck_MHCII_entry_func_effects.html).

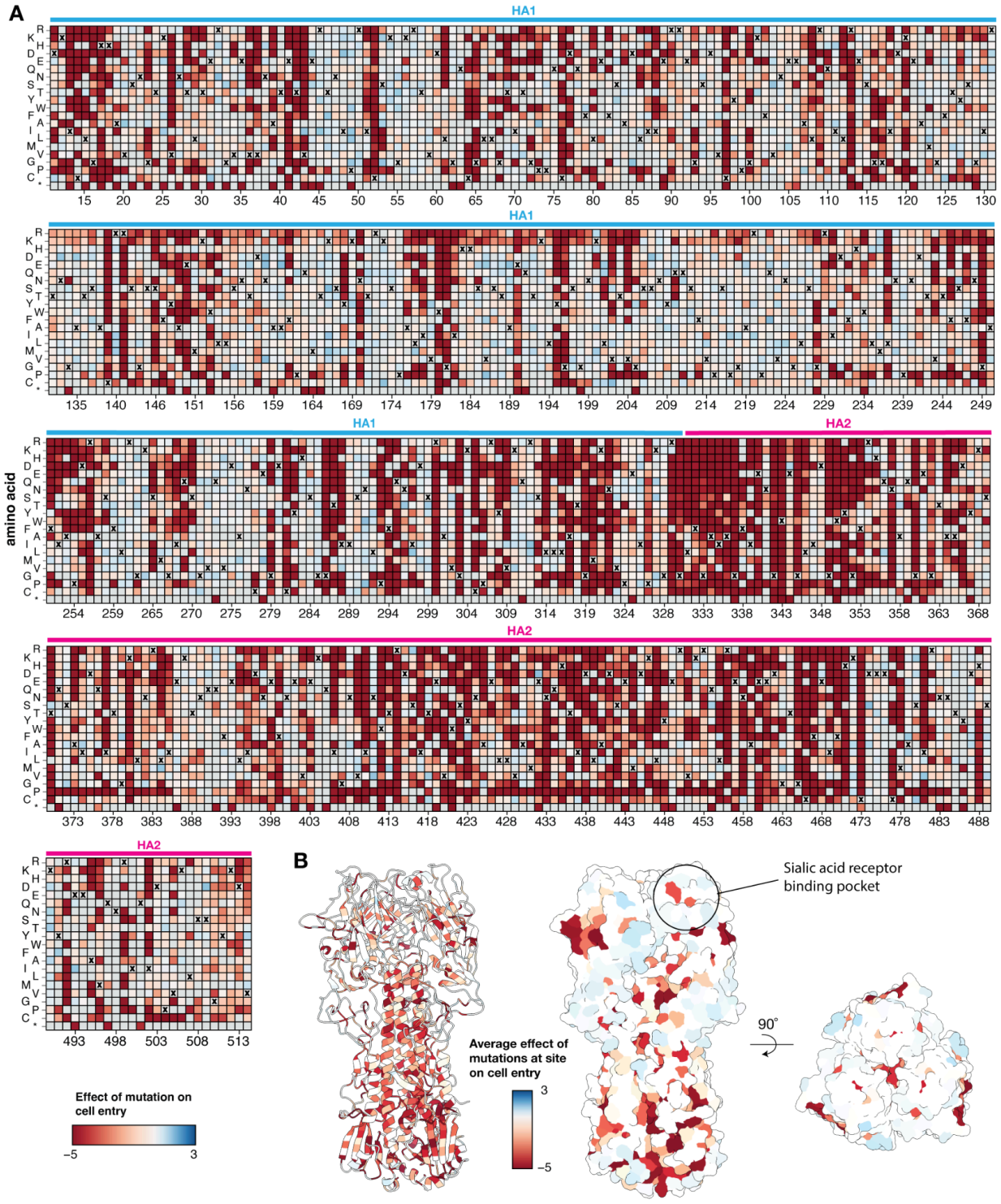


Figure 5.1 | Effects of mutations to the ectodomain of H7 HA on its ability to mediate entry into 293 cells expressing tufted duck MHCII receptors.

A) Heatmap showing the effects of mutations on HA-mediated cell entry. Negative cell entry effects (red) indicate impaired cell entry relative to the unmutated HA. Gray indicates mutations that were not reliably measured in our experiments, and for each site the X indicates the

amino-acid identity in the unmutated A/Anhui/1/2013 HA. See https://dms-vep.org/Flu-H7-Anhui13-MHCII-binding/htmls/293_noSA_tufted_duck_MHCII_entry_func_effects.html for an interactive version of this heatmap. In this figure and throughout the rest of this paper we use H3 sequence numbering (Burke and Smith 2014). **B)** The structure of the A/Anhui/1/2013 H7 HA (Huang et al. 2019) (PDB 6ii9) colored by the mean effects of all mutations at that site on tufted duck MHCII expressing cell entry. Traditional sialic acid receptor binding pocket is indicated by a circle.

As illustrated in this figure, there is wide variation in mutational tolerance across the HA protein, with some sites accommodating mutations readily while others show severe functional impairment from nearly any amino acid change when tested on tufted duck MHCII-expressing cells. The overall pattern of functional constraint mirrors what we observed on α2-3 and α2-6 receptor-expressing cells, with HA2 generally displaying higher constraint than HA1. This is particularly evident in critical HA2 regions such as the fusion loop (sites 330-350), which mediates the essential membrane fusion step of viral entry (**Fig. 5.1A**). In contrast, many regions of HA1, especially those near the top of the globular head domain, remain quite tolerant of mutations (**Fig. 5.1B**). These results align well with previous deep mutational scanning studies of HA from other influenza A subtypes. Those studies consistently identified the strongest functional constraints in HA2 regions involved in membrane fusion, while finding higher mutational tolerance in portions of the HA1 globular head, including epitopes commonly targeted by neutralizing antibodies (Dadonaite et al., 2024; J. M. Lee et al., 2018; Thyagarajan & Bloom, 2014; Wu et al., 2014; Wu, Thompson, et al., 2020). However, our MHCII entry data revealed some notable differences from the canonical sialic acid receptor binding patterns.

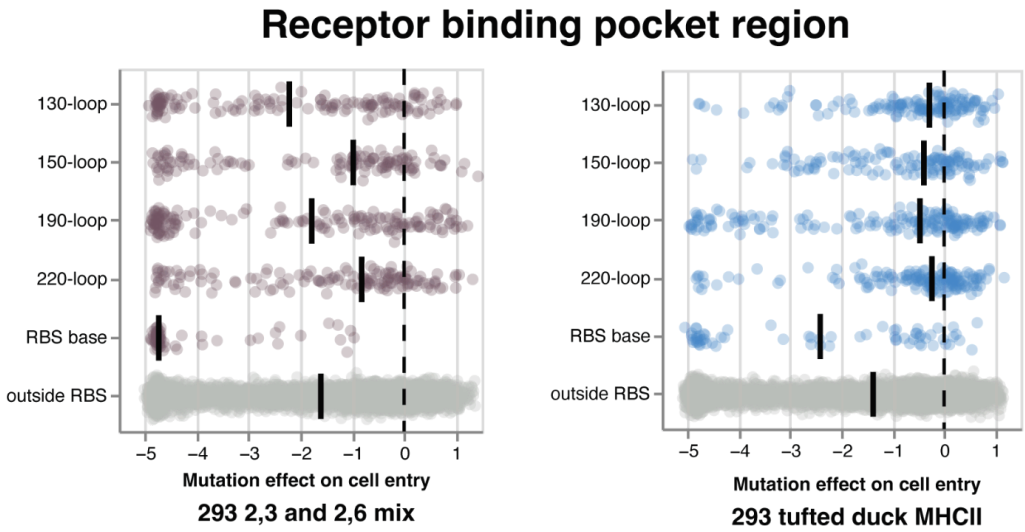


Figure 5.2 | Comparison of mutation effects to the receptor binding pocket of H7 HA on its ability to mediate entry into 293 cells expressing tufted duck MHCII receptors vs. 293 mix cells.

A) Scatter plot comparing mutation effects on cell entry across different receptor binding pocket regions when tested on two distinct cell types. The left panel shows results from 293 cells expressing a mixture of α2-3 and α2-6 linked sialic acid receptors, while the right panel displays data from 293 cells expressing tufted duck MHCII receptors. Each point represents an individual mutation, with the y-axis indicating the specific receptor binding pocket loop where the mutation occurs (130-loop, 150-loop, 190-loop, 220-loop, and RBS base), and the x-axis showing the measured mutation effect on cell entry from the deep mutational scanning experiment. Mutations within the traditional receptor binding site regions are colored dark purple (left panel) and blue (right panel), while mutations outside the RBS are shown in grey. The data reveal that mutations in the receptor binding pocket loops are generally better tolerated in the MHCII-expressing cells compared to the sialic acid receptor-expressing cells, with most mutations showing less severe negative effects (values closer to zero) in the MHCII context. This pattern suggests that the traditional sialic acid receptor binding pocket may not be the primary binding site for MHCII-mediated entry, as mutations that significantly impair sialic acid receptor binding have minimal impact on MHCII-mediated cell entry.

Most strikingly, the traditional receptor binding site for sialic acid receptors showed relatively low constraint in our MHCII entry assays. The majority of sites within this pocket tolerated mutations well, a pattern that extended to the various loops that comprise the receptor binding pocket. When we directly compared mutation effects on cell entry across these different loops, mutations were consistently better tolerated in 293 cells expressing tufted duck MHCII compared to the mixed 293 cells expressing a2,3 and a2,6 receptors (**Fig 5.2**). Even the receptor binding site base appeared less constrained in the MHCII context than in standard receptor binding assays.

This relaxed constraint pattern supports the hypothesis that HA binds to MHCII receptors at a location distinct from the traditional sialic acid receptor binding pocket. Instead, our constraint mapping identified alternative sites that may be critical for MHCII engagement. Sites 131, 67, 247, 138, and 139 in the head domain emerged as among the most functionally constrained, suggesting these positions could comprise a potential MHCII receptor binding interface. The strong constraint at these sites indicates that maintaining proper amino acid identity is crucial for MHCII-mediated entry, consistent with their involvement in receptor contact or binding site architecture. In the stalk region, sites 330-337 and 408-470 showed the strongest functional constraint. While the 330-337 region corresponds to the well-characterized fusion loop, the constraint pattern at sites 408-470 suggests additional critical functions in this domain that may be particularly important for MHCII-mediated entry or subsequent fusion events.

To validate our deep mutational scanning measurements, we generated single mutant viruses containing specific mutations that either increased or decreased entry efficiency on 293 cells expressing tufted duck MHCII. These validation experiments confirmed the accuracy of our high-throughput screening approach and provided additional confidence in the constraint patterns we identified (**Fig S5.2**).

5.4 Discussion

Our deep mutational scanning analysis of H7 hemagglutinin entry via MHCII receptors provides the first comprehensive map of functional constraints governing this alternative influenza entry pathway. The striking differences in mutational tolerance patterns between MHCII-mediated and sialic acid receptor-mediated entry demonstrate that these represent fundamentally distinct molecular interactions with different structural requirements.

The most significant finding from our analysis is the relaxed functional constraint observed in the traditional sialic acid receptor binding pocket when viral entry occurs via MHCII receptors. This pattern strongly supports previous biochemical evidence that H17-H19 subtypes utilize binding sites distinct from the canonical receptor binding pocket(Karakus et al. 2024; Sun et al. 2013). The tolerance of mutations that would severely impair sialic acid binding suggests that the MHCII interaction surface is spatially separated from this well-characterized binding site. This finding has important evolutionary implications, as it suggests that viruses could

potentially acquire MHCII binding capability without losing sialic acid receptor function, potentially creating dual-receptor viruses with expanded cellular tropism.

Our identification of highly constrained sites in the HA head domain (positions 131, 67, 247, 138, and 139) provides the first evidence-based prediction of a putative MHCII binding interface. The strong functional constraint at these positions suggests they play critical roles in either direct receptor contact or maintaining the proper structural architecture required for MHCII recognition. Notably, these sites are distributed across different regions of the HA1 head domain, suggesting that MHCII binding may involve a more extended or conformationally complex interface compared to the relatively compact sialic acid binding pocket. This prediction aligns with the larger size and more complex topology of MHCII molecules compared to sialic acid receptors, which may require more extensive protein-protein contacts for stable binding.

The functional constraints we identified in the HA stalk region provide additional insights into the mechanistic requirements for MHCII-mediated entry. While constraint in the fusion loop region (sites 330-337) likely reflects universal requirements for membrane fusion that are independent of receptor type, the strong constraint observed at sites 408-470 may indicate MHCII-specific functional requirements. This could suggest that MHCII-mediated entry involves distinct conformational changes or membrane fusion kinetics compared to sialic acid receptor-mediated entry. Alternatively, these stalk region constraints might reflect requirements for proper HA folding or stability that become more critical when the receptor binding interaction occurs at an alternative site.

The broader implications of MHCII-mediated influenza entry extend beyond mechanistic understanding to questions of viral evolution and host range determination. MHCII molecules are highly conserved across vertebrate species and are expressed on professional antigen-presenting cells throughout the body(Rock et al. 2016; Wieczorek et al. 2017). This conservation could explain why H18 can utilize MHCII from diverse vertebrate species(Karakus et al. 2019), but it also raises concerning possibilities for broad host range and zoonotic potential. Unlike sialic acid receptor distribution, which varies significantly between species and tissue types, MHCII expression patterns are more uniform across vertebrates, potentially providing a pathway for cross-species transmission that bypasses traditional receptor-based species barriers.

Our findings also highlight important gaps in our understanding of influenza receptor diversity. While H7 showed the capability to utilize MHCII for entry in our experimental system, it remains unclear whether other conventional influenza subtypes possess latent MHCII binding potential or whether this represents an evolutionary innovation specific to certain lineages. The identification of specific constraint patterns associated with MHCII usage provides a framework for evaluating other subtypes and assessing their potential for alternative receptor utilization.

From a pandemic preparedness perspective, these results underscore the importance of surveillance strategies that account for alternative entry pathways. Current monitoring efforts

focus heavily on sialic acid receptor binding properties as predictors of zoonotic and pandemic potential(Cardenas et al. 2026; Rajao et al. 2025). However, viruses capable of MHCII-mediated entry might circumvent these predictions entirely, as their host range determinants would be governed by different molecular rules. Understanding the structural requirements we have identified here will be crucial for developing more comprehensive risk assessment frameworks that account for the full spectrum of potential influenza entry mechanisms.

Future studies should focus on validating our predicted MHCII binding interface through structural and biochemical approaches, determining whether the constraint patterns we identified are conserved across other subtypes capable of MHCII binding, and assessing the prevalence of MHCII-binding capability among circulating influenza strains. Additionally, investigating the tissue-specific implications of MHCII-mediated entry will be important for understanding the pathogenetic consequences of this alternative pathway and its potential role in influenza disease progression.

5.5 Supplementary Materials

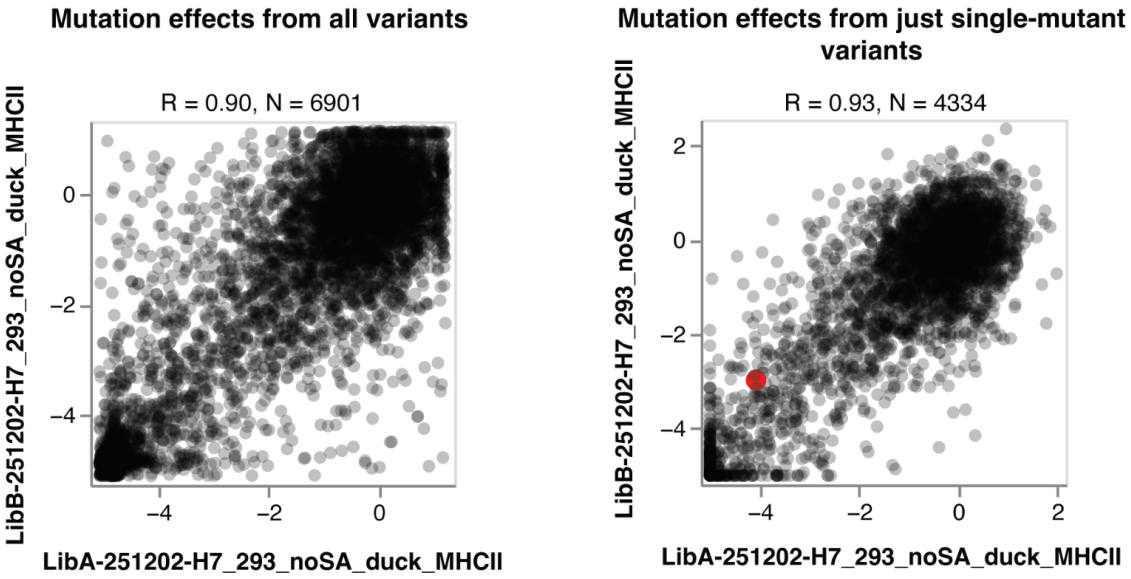


Figure S5.1 | H7 HA cell entry effects measured using independent deep mutational scanning libraries are highly correlated

Correlation of the effects of H7 HA mutations on cell entry as measured with independent pseudovirus libraries on 293 cells with no sialic acid receptors and expressing tufted duck MHCII receptors. Each point represents the effect of a mutation measured in each replicate. For each of the two libraries, two technical replicates were performed. The left panel shows the correlations between two independent libraries (Library A and B) across all variants. The panel on the right shows the correlations between two independent libraries (A and B) across just single-mutant variants. R is the Pearson correlation and N is the number of mutations measured in both pairs of measurements.

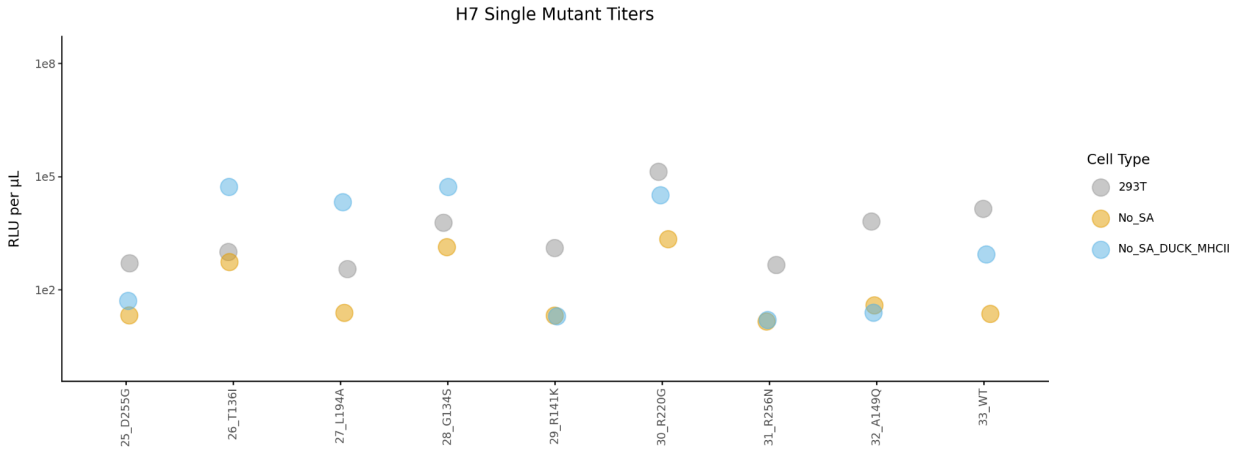


Figure S5.2 | Single mutant validations of H7 HA cell entry effects measured on 293T cells, 293 no sialic acid expressing cells, and tufted duck MHCII expressing cells

This titer plot validates deep mutational scanning predictions using single-mutant H7 luciferase viruses tested across three cell types: 293T cells with natural sialic acid receptors (grey), sialic acid-depleted 293T cells (No SA, yellow), and sialic acid-depleted cells expressing tufted duck MHCII receptors (No SA DUCK MHCII, blue). The x-axis shows nine selected mutations (named as original amino acid, position, and substitution), while the y-axis displays viral titers in relative light units (RLU) per microliter. This experimental design distinguishes between sialic acid-dependent and MHCII-dependent entry pathways. Mutations showing enhanced titers in MHCII-expressing cells compared to sialic acid-depleted cells indicate MHCII-specific entry improvements, validating the functional significance of constraint patterns identified through deep mutational scanning and confirming distinct molecular requirements for alternative influenza entry mechanisms.

5.6 Materials and methods

Cell lines and media

293 cells were maintained D10 media (Dulbecco's Modified Eagle Medium with 10% tetracycline-free heat-inactivated fetal bovine serum, 2 mM l-glutamine, 100 U/mL penicillin, and 100 µg/mL streptomycin). D10 media was used for all experiments involving pseudovirus.

To make the 293 cells expressing the tufted duck MHCII that have no sialic acid, on day 1 293T cells were transfected with 0.25µg of each helper plasmid each helper plasmids encoding Gag/Pol, Tat, and Rev (AddGene numbers 204152, 204154, 204153), 1µg of the lentivirus backbone coding for the tufted duck MHCII and purmycin resistance gene (plasmid 5490 https://github.com/jbloomlab/plasmids/blob/master/genbank_maps/5490_pHAGE2_EF1a_Duck_MHCII-DR_IRES_Puro.gb). On day 2, 1M 293 cells with no sialic acid were plated in a 6-well plate. On day 3, supernatant was collected from the transfected cells using a 0.45µM SFCA filter. 1ml of the collected virus was then used to infect the previously plated 293-no-sialic-acid cells, since these viruses cannot be titrated without the presence of a fluorescent protein, an estimated high MOI was used to infect. On day 5, 48 hours post infection, cells were then trypsinized and then replated in a 10cm dish in the presence of 0.75µg/ml puromycin to select for successfully transduced cells. Cell media was changed with media with fresh puromycin every 2 days until no more cell death was observed. The cells were then expanded in D10 media with no puromycin and then frozen down.

Generation of single mutant luciferase pseudoviruses

Single mutant luciferase pseudoviruses were made by plating 293T cells in 6-well dishes. Each well was seeded with ~800,000 cells/ml in a total volume of 2ml. The next day cells were then transfected with the following plasmids; 1µg of the 2727 luciferase backbone, 340ng Gag/Pol (AddGene numbers 204152), 150ng of plasmid expressing human airway trypsin-like protease

(https://github.com/dms-vep/Flu_H7_Anhui13_DMS/blob/main/library_design/HDM_HAT.gb) to activate the HA for membrane fusion, 50ng of the plasmid expressing NA, and 500ng of plasmid expressing the H7 HA with the individual mutations. BioT transfection reagent was used according to the manufacturer's instructions. 48 hours after transfection, the viral supernatant was collected and filtered through 0.45µM syringe filters (Corning® 28 mm Diameter Syringe Filters, 0.45 µm Pore SFCA Membrane). Viruses were then titrated by plating 30,000 cells per well of 293T cells in 96-well black wall plates coated in Poly-L-lysine (Grenier, Item No.: 655936). The cells were also plated in the presence of 2.5µg/ml of amphotericin B (amphotericin B was shown to increase virus titers). The next day, the viruses were then diluted at a starting dilution of 30µl virus into 210µl of D10 media (total 240µl), 120µl of the previous dilution was then diluted into 120µl for a 1:2 dilution going down the plate. 48 hours after adding the virus 100µl of the supernatant was removed and 30µl of brightglo was added and used according to the manufacturer's instructions (Promega, Cat. E2620). The plate was then read-out for luciferase activity after 5 minutes and the titers were calculated.

Mutation effects on cell entry in tufted duck MHCII expressing cells

Cell entry effects of each mutation in the library were determined as previously described (Dadonaite et al. 2024). To measure effects of mutations on cell entry, 293 cells expressing no sialic acid but did express the tufted duck MHCII were used. These cells were then either infected with libraries pseudotyped with VSV-G envelope (produced as described above) or HA pseudotyped libraries. For HA pseudotyped libraries, the tufted duck MHCII cells were plated in the presence of 2.5µg/ml of amphotericin B a day before (amphotericin B was shown to increase library virus titers), and infections were performed in the presence of 1µM oseltamivir to inhibit NA and prevent cleavage of receptors on target cell surface. For VSV-G pseudotyped library infections approximately 8 million transcription units were used and for the H7 pseudotyped library infections approximately 3.2 million transcription units were used. At 12 to 15 hours postinfection, cells were collected, nonintegrated lentiviral genomes were recovered using QIAprep Spin Miniprep Kit (Qiagen, cat. no. 27104), and amplicon libraries for barcode sequencing were prepared as previously described (Dadonaite, Crawford, et al. 2023; Dadonaite et al. 2024; Yu et al. 2025).

Cell entry scores for each variant were calculated using log enrichment ratio: $\log_2$ ([n^{v}_{post} / n^{wt}_{post}]/[n^{v}_{pre} / n^{wt}_{pre}]), where n^{v}_{post} is the count of variant v in the H7-pseudotyped infection (postselection condition), n^{v}_{pre} is the count of variant v in the VSV-G-pseudotyped infection (preselection condition, and n^{wt}_{post} and n^{v}_{pre} are the counts for wildtype variants. Positive cell entry scores indicate that a variant is better at entering the cells compared to the unmutated parental HA, and negative scores indicate entry worse than the unmutated HA.

To calculate the mutational-level cell entry effects, a sigmoid global-epistasis(Otwinowski et al. 2018) function was fitted to variant entry scores after truncating the values at a lower bound of -5, using the *multi-dms* software package(Haddox et al. 2023). The rationale for fitting the global

epistasis models is that some of the variants have multiple HA mutations, and so the model enables deconvolution of the mutation effects from these multiple mutated variants as well as just the single mutants. An interactive heatmap showing the effect of each mutation on cell entry on 293 cells expressing tufted duck MHCII is at (https://dms-vep.org/Flu-H7-Anhui13-MHCII-binding/htmls/293_noSA_tufted_duck_MHCII_entry_func_effects.html). For the mutation effects, values of zero mean no effect on cell entry, negative values mean impaired cell entry, and positive values mean improved cell entry. For plots in the paper, we show only mutations observed in average of at least 2 barcoded variants per library and measured in both biological library replicates.

Chapter 6

6 Conclusions and Future Directions

In **chapter 2**, we developed an improved pH stability assay that enables high-throughput identification of HA stabilizing and destabilizing mutations, overcoming traditional limitations through virus reconcentration methods. In **chapter 3**, we mapped the complete functional landscape of H7 hemagglutinin, revealing its dual receptor binding capability and identifying key mutations that enhance human adaptation, many of which have already been naturally selected in circulating H7N9 strains. In **chapter 4**, we discovered that amino acid preferences have diverged dramatically across HA subtypes despite nearly identical structures, demonstrating how evolutionary changes can rewire protein constraints even when overall architecture remains conserved. In **chapter 5**, we characterized an alternative MHCII-mediated entry pathway that operates through entirely different molecular determinants than traditional sialic acid binding, challenging conventional models of influenza host range determination.

Coming into the lab, I was under the impression that my thesis work needed to yield some sort of groundbreaking, field-changing discovery otherwise it wouldn't be important. What I've learned is that science is constantly evolving, and it's often the accumulation of smaller findings that ultimately leads to major breakthroughs. More often than not, my most significant progress came from the parts of experiments that didn't work as expected and digging into why they failed. These apparent setbacks frequently led the project in unexpected directions, but those pivots often resulted in the most interesting discoveries. A perfect example of this came when Jesse suggested we pivot from our planned serum and antibody selections to instead compare mutation effects across all the HA subtypes we already had data for. This turned out to be one of the most compelling directions of the project, revealing just how dramatically mutational tolerance differs across HA subtypes despite their nearly identical structures. What started as a natural next step in our experimental pipeline uncovered something fundamental about protein evolution that we hadn't anticipated.

The MHCII work followed a similar trajectory. What began as a side experiment to test alternative entry mechanisms revealed an entirely new dimension of influenza biology that challenges our basic assumptions about viral host range. These discoveries remind me that some of the most important contributions come not from confirming what we expect, but from revealing gaps in our understanding.

With all the information presented in this thesis, this conclusion wraps up the applications of my findings while considering what directions might come next. The work demonstrates how

methodological advances, unexpected experimental results, and willingness to follow interesting leads can collectively reshape our understanding of complex biological systems.

6.1 Applications of high-throughput acid stability assay

The development of our improved pH stability assay represents a significant methodological advance with broad applications for influenza research and public health preparedness. By incorporating reconcentration and neutralization steps, we have overcome the traditional limitations that required high-titer virus samples, making systematic stability measurements feasible for large-scale mutational studies and surveillance applications. Our validation using H5N1 hemagglutinin demonstrated the assay's ability to accurately distinguish between stabilizing and destabilizing mutations, with results that correlate well with direct measurements using conditionally replicative viruses.

The practical implications of this methodology extend far beyond academic research. HA stability represents a critical parameter for assessing pandemic potential, as viruses with enhanced environmental persistence may have increased transmission efficiency and broader geographic spread. Current WHO and CDC risk assessment frameworks could be significantly enhanced by incorporating stability measurements alongside existing evaluations of receptor binding specificity and antigenicity. Our high-throughput approach makes it feasible to rapidly characterize emerging viral isolates for this important phenotype, providing public health officials with additional data for risk stratification.

From a vaccine development perspective, the ability to systematically map stability-enhancing mutations offers valuable insights for immunogen design. Stabilized HA variants may serve as improved vaccine antigens with enhanced shelf life and immunogenicity, while understanding the molecular determinants of stability provides guidance for rational vaccine optimization. The broad applicability of our assay across different HA subtypes makes it a valuable tool for both seasonal vaccine strain selection and pandemic preparedness efforts, enabling rapid characterization of novel variants as they emerge in surveillance systems.

6.2 Can deep mutational scanning of H7 help inform on pandemic preparedness?

Our comprehensive deep mutational scanning analysis of H7 hemagglutinin provides the first systematic map of functional constraints for this pandemic-relevant subtype, with important implications for risk assessment and preparedness strategies. The identification of nearly 65 sites where mutations enhance entry into human-type receptor-expressing cells reveals the substantial adaptive potential that exists within H7 populations. Perhaps most significantly, our comparison with circulating H7N9 sequences demonstrates that many of the most beneficial mutations have already been naturally selected, indicating that substantial human adaptation has already occurred in this subtype.

The discovery that H7N9 viruses have achieved near-universal fixation of key adaptive mutations like K193T and Q210Y suggests these viruses have progressed further along the pathway toward human adaptation than previously appreciated. However, our identification of numerous high-effect beneficial mutations that remain rare in nature reveals significant "cryptic"

evolutionary potential that could be realized under appropriate selective pressures. This finding has profound implications for pandemic risk assessment, as it suggests that H7 viruses retain the capacity for further functional enhancement beyond their current state.

The systematic constraint mapping we have provided offers a framework for interpreting future H7 evolution and assessing the significance of novel mutations as they arise in surveillance. Mutations occurring in highly constrained regions can be flagged as potentially significant for viral fitness, while changes in regions we have identified as mutationally tolerant may be of less immediate concern. This functional annotation of the H7 genome provides public health officials with evidence-based tools for prioritizing surveillance findings and making informed decisions about vaccine strain updates and pandemic preparedness measures.

6.2 MHCII-Mediated Influenza Entry: Implications for Host Range and Alternative Transmission Pathways

Our discovery that H7 hemagglutinin can utilize MHCII molecules for cellular entry fundamentally challenges traditional models of influenza host range determination and reveals a potentially underappreciated pathway for cross-species transmission. The systematic constraint mapping we performed provides the first evidence-based prediction of molecular determinants governing this alternative entry mechanism, identifying putative binding interfaces that operate independently of canonical sialic acid receptor interactions. This finding has profound implications for our understanding of influenza ecology and pandemic emergence.

The biological significance of MHCII-mediated entry extends beyond mechanistic interest to practical concerns about viral surveillance and risk assessment. Unlike sialic acid receptor distribution, which varies significantly between species and provides well-characterized barriers to cross-species transmission, MHCII molecules are highly conserved across vertebrates and expressed on professional antigen-presenting cells throughout diverse tissues. This conservation suggests that MHCII-mediated entry could provide influenza viruses with access to transmission pathways that bypass traditional receptor-based species barriers, potentially enabling broader host ranges and more efficient zoonotic transmission.

Our findings highlight critical gaps in current pandemic risk assessment frameworks, which focus heavily on sialic acid receptor binding properties as predictors of zoonotic potential. Viruses capable of MHCII-mediated entry might circumvent these predictions entirely, operating through molecular determinants governed by different evolutionary rules. The identification of specific constraint patterns associated with MHCII usage provides a foundation for developing more comprehensive surveillance strategies that account for alternative entry mechanisms. Future pandemic preparedness efforts must consider the possibility that influenza

viruses may possess multiple, independent pathways for host cell entry, each with distinct implications for cross-species transmission and public health risk.

Bibliography

Ahn, Jenny J., Timothy C. Yu, Bernadeta Dadonaite, Caelan E. Radford, and Jesse D. Bloom. 2026. "Influenza Hemagglutinin Subtypes Have Different Sequence Constraints despite Sharing Extremely Similar Structures." In *bioRxiv*. BioRxiv, January 6. https://doi.org/10.64898/2026.01.05.697808.

Ayora-Talavera, Guadalupe, Holly Shelton, Margaret A. Scull, et al. 2009. "Mutations in H5N1 Influenza Virus Hemagglutinin That Confer Binding to Human Tracheal Airway Epithelium." *PloS One* 4 (11): e7836.

Baker, Amy L., Bailey Arruda, Mitchell V. Palmer, et al. 2025. "Dairy Cows Inoculated with Highly Pathogenic Avian Influenza Virus H5N1." *Nature* 637 (8047): 913–920. https://doi.org/10.1038/s41586-024-08166-6.

Bateman, Allen C., Marc G. Busch, Alexander I. Karasin, Nicolai Bovin, and Christopher W. Olsen. 2008. "Amino Acid 226 in the Hemagglutinin of H4N6 Influenza Virus Determines Binding Affinity for alpha2,6-Linked Sialic Acid and Infectivity Levels in Primary Swine and Human Respiratory Epithelial Cells." *Journal of Virology* 82 (16): 8204–8209. https://doi.org/10.1128/JVI.00718-08.

Baudin, F., C. Bach, S. Cusack, and R. W. Ruigrok. 1994. "Structure of Influenza Virus RNP. I. Influenza Virus Nucleoprotein Melts Secondary Structure in Panhandle RNA and Exposes the Bases to the Solvent." *The EMBO Journal* 13 (13): 3158–3165. https://doi.org/10.1002/j.1460-2075.1994.tb06614.x.

Bloom, Jesse D. 2014. "An Experimentally Determined Evolutionary Model Dramatically Improves Phylogenetic Fit." *Molecular Biology and Evolution* 31 (8): 1956–1978. https://doi.org/10.1093/molbev/msu173.

Bloom, Jesse D., Lizhi Ian Gong, and David Baltimore. 2010. "Permissive Secondary Mutations Enable the Evolution of Influenza Oseltamivir Resistance." *Science (New York, N.Y.)* 328 (5983): 1272–1275. https://doi.org/10.1126/science.1187816.

Bosch, Berta, Erica L. Heipertz, James R. Drake, and Paul A. Roche. 2013. "Major Histocompatibility Complex (MHC) Class II-Peptide Complexes Arrive at the Plasma

Membrane in Cholesterol-Rich Microclusters." *The Journal of Biological Chemistry* 288 (19): 13236–13242. https://doi.org/10.1074/jbc.M112.442640.

Bouvier, Nicole M., and Peter Palese. 2008. "The Biology of Influenza Viruses." *Vaccine* 26 Suppl 4 (Suppl 4): D49–53. https://doi.org/10.1016/j.vaccine.2008.07.039.

Briedis, D. J., and R. A. Lamb. 1982. "Influenza B Virus Genome: Sequences and Structural Organization of RNA Segment 8 and the mRNAs Coding for the NS1 and NS2 Proteins." *Journal of Virology* 42 (1): 186–193. https://doi.org/10.1128/JVI.42.1.186-193.1982.

Broecker, Felix, Sean T. H. Liu, Weina Sun, Florian Krammer, Viviana Simon, and Peter Palese. 2018. "Immunodominance of Antigenic Site B in the Hemagglutinin of the Current H3N2 Influenza Virus in Humans and Mice." *Journal of Virology* 92 (20). https://doi.org/10.1128/JVI.01100-18.

Brown, J. H., T. S. Jardetzky, J. C. Gorga, et al. 1993. "Three-Dimensional Structure of the Human Class II Histocompatibility Antigen HLA-DR1." *Nature* 364 (6432): 33–39. https://doi.org/10.1038/364033a0.

Brown, Joshua D., Siarhei Kharytonchyk, Issac Chaudry, et al. 2020. "Structural Basis for Transcriptional Start Site Control of HIV-1 RNA Fate." *Science (New York, N.Y.)* 368 (6489): 413–417. https://doi.org/10.1126/science.aaz7959.

Büll, Christian, Rebecca Nason, Lingbo Sun, et al. 2021. "Probing the Binding Specificities of Human Siglecs by Cell-Based Glycan Arrays." *Proceedings of the National Academy of Sciences of the United States of America* 118 (17). https://doi.org/10.1073/pnas.2026102118.

Burke, David F., and Derek J. Smith. 2014. "A Recommended Numbering Scheme for Influenza A HA Subtypes." *PloS One* 9 (11): e112302. https://doi.org/10.1371/journal.pone.0112302.

Cao, Yunlong, Jing Wang, Fanchong Jian, et al. 2022. "Omicron Escapes the Majority of Existing SARS-CoV-2 Neutralizing Antibodies." *Nature* 602 (7898): 657–663. https://doi.org/10.1038/s41586-021-04385-3.

Cardenas, Matias, Sasha Compton, C. Joaquin Caceres, Adolfo García-Sastre, Daniel R. Perez, and Daniela S. Rajao. 2026. "MHC Class II Functions as a Host-Specific Entry Receptor for Representative Human and Swine H3N2 Influenza A Viruses." *Nature Communications* 17

(1): 2560. https://doi.org/10.1038/s41467-026-69267-6.

Carr, Chavela M., Charu Chaudhry, and Peter S. Kim. 1997. "Influenza Hemagglutinin Is Spring-Loaded by a Metastable Native Conformation." *Proceedings of the National Academy of Sciences of the United States of America* 94 (26): 14306–14313. https://doi.org/10.1073/pnas.94.26.14306.

Carr, Chavela M., and Peter S. Kim. 1993. "A Spring-Loaded Mechanism for the Conformational Change of Influenza Hemagglutinin." *Cell* 73 (4): 823–832. https://doi.org/10.1016/0092-8674(93)90260-w.

CDC. 2024. "Influenza Risk Assessment Tool (IRAT)." Pandemic Flu, October 25. https://www.cdc.gov/pandemic-flu/php/national-strategy/influenza-risk-assessment-tool.html.

CDC. 2025. "CDC A(H5N1) Bird Flu Response Update March 19, 2025." Avian Influenza (Bird Flu), July 24. https://www.cdc.gov/bird-flu/spotlights/h5n1-response-03192025.html.

CDC. 2026. "How Flu Viruses Can Change: 'Drift' and 'Shift.'" Influenza (Flu), March 25. https://www.cdc.gov/flu/php/viruses/change.html.

Chou, P. Y., and G. D. Fasman. 1978. "Prediction of the Secondary Structure of Proteins from Their Amino Acid Sequence." *Advances in Enzymology and Related Areas of Molecular Biology* 47: 45–148. https://doi.org/10.1002/9780470122921.ch2.

Compans, R. W., J. Content, and P. H. Duesberg. 1972. "Structure of the Ribonucleoprotein of Influenza Virus." *Journal of Virology* 10 (4): 795–800. https://doi.org/10.1128/JVI.10.4.795-800.1972.

Crawford, Katharine H. D., and Jesse D. Bloom. 2019. "Alignparse: A Python Package for Parsing Complex Features from High-Throughput Long-Read Sequencing." *Journal of Open Source Software* 4 (44): 1915. https://doi.org/10.21105/joss.01915.

Dadonaite, Bernadeta, Jenny J. Ahn, Jordan T. Ort, et al. 2024. "Deep Mutational Scanning of H5 Hemagglutinin to Inform Influenza Virus Surveillance." *PLoS Biology* 22 (11): e3002916. https://doi.org/10.1371/journal.pbio.3002916.

Dadonaite, Bernadeta, Jack Brown, Teagan E. McMahon, et al. 2023. "Full-Spike Deep Mutational Scanning Helps Predict the Evolutionary Success of SARS-CoV-2 Clades." In *bioRxivorg*. BioRxiv, November 14. https://doi.org/10.1101/2023.11.13.566961.

Dadonaite, Bernadeta, Katharine H. D. Crawford, Caelan E. Radford, et al. 2023. "A Pseudovirus System Enables Deep Mutational Scanning of the Full SARS-CoV-2 Spike." *Cell* 186 (6): 1263–1278.e20. https://doi.org/10.1016/j.cell.2023.02.001.

Dauber, Bianca, Gudrun Heins, and Thorsten Wolff. 2004. "The Influenza B Virus Nonstructural NS1 Protein Is Essential for Efficient Viral Growth and Antagonizes Beta Interferon Induction." *Journal of Virology* 78 (4): 1865–1872. https://doi.org/10.1128/jvi.78.4.1865-1872.2004.

Doud, Michael B., and Jesse D. Bloom. 2016/6. "Accurate Measurement of the Effects of All Amino-Acid Mutations on Influenza Hemagglutinin." *Viruses* 8 (6): 155. https://doi.org/10.3390/v8060155.

Doud, Michael B., Scott E. Hensley, and Jesse D. Bloom. 2017. "Complete Mapping of Viral Escape from Neutralizing Antibodies." *PLoS Pathogens* 13 (3): e1006271. https://doi.org/10.1371/journal.ppat.1006271.

Doud, Michael B., Juhye M. Lee, and Jesse D. Bloom. 2018. "How Single Mutations Affect Viral Escape from Broad and Narrow Antibodies to H1 Influenza Hemagglutinin." *Nature Communications* 9 (1): 1386. https://doi.org/10.1038/s41467-018-03665-3.

Er, Jwee Chiek. 2025. "Improving Influenza Nomenclature Based on Transmission Dynamics." *Viruses* 17 (5): 633. https://doi.org/10.3390/v17050633.

Fowler, Douglas M., Carlos L. Araya, Sarel J. Fleishman, et al. 2010. "High-Resolution Mapping of Protein Sequence-Function Relationships." *Nature Methods* 7 (9): 741–746. https://doi.org/10.1038/nmeth.1492.

Galloway, Summer E., Mark L. Reed, Charles J. Russell, and David A. Steinhauer. 2013. "Influenza HA Subtypes Demonstrate Divergent Phenotypes for Cleavage Activation and pH of Fusion: Implications for Host Range and Adaptation." *PLoS Pathogens* 9 (2): e1003151. https://doi.org/10.1371/journal.ppat.1003151.

Gamblin, Steven J., Sébastien G. Vachieri, Xiaoli Xiong, Jie Zhang, Stephen R. Martin, and John J. Skehel. 2021. "Hemagglutinin Structure and Activities." *Cold Spring Harbor Perspectives in Medicine* 11 (10): a038638. https://doi.org/10.1101/cshperspect.a038638.

Gao, Rongbao, Bin Cao, Yunwen Hu, et al. 2013. "Human Infection with a Novel Avian-Origin Influenza A (H7N9) Virus." *The New England Journal of Medicine* 368 (20): 1888–1897. https://doi.org/10.1056/NEJMoa1304459.

García-Sastre, A. 2001. "Inhibition of Interferon-Mediated Antiviral Responses by Influenza A Viruses and Other Negative-Strand RNA Viruses." *Virology* 279 (2): 375–384. https://doi.org/10.1006/viro.2000.0756.

G Bhanu Prakash Animated Medical Videos. 2019. "Influenza Virus Replication Cycle Animation - Medical Microbiology USMLE Step 1." Youtube, November 12. 296 sec. https://www.youtube.com/watch?v=tB5FQZi4HKY.

Gerlach, Thomas, Luca Hensen, Tatyana Matrosovich, et al. 2017. "PH Optimum of Hemagglutinin-Mediated Membrane Fusion Determines Sensitivity of Influenza A Viruses to the Interferon-Induced Antiviral State and IFITMs." *Journal of Virology* 91 (11): e00246–17. https://doi.org/10.1128/JVI.00246-17.

Gilchrist, Cameron L. M., Milot Mirdita, and Martin Steinegger. 2024. "Multiple Protein Structure Alignment at Scale with FoldMason." In *bioRxiv*. August 1. https://doi.org/10.1101/2024.08.01.606130.

Giotis, Efstathios S., George Carnell, Erik F. Young, et al. 2019. "Entry of the Bat Influenza H17N10 Virus into Mammalian Cells Is Enabled by the MHC Class II HLA-DR Receptor." *Nature Microbiology* 4 (12): 2035–2038. https://doi.org/10.1038/s41564-019-0517-3.

Glaser, Laurel, James Stevens, Dmitriy Zamarin, et al. 2005. "A Single Amino Acid Substitution in 1918 Influenza Virus Hemagglutinin Changes Receptor Binding Specificity." *Journal of Virology* 79 (17): 11533–11536. https://doi.org/10.1128/JVI.79.17.11533-11536.2005.

Greaney, Allison J., Andrea N. Loes, Katharine H. D. Crawford, et al. 2021. "Comprehensive Mapping of Mutations in the SARS-CoV-2 Receptor-Binding Domain That Affect Recognition by Polyclonal Human Plasma Antibodies." *Cell Host & Microbe* 29 (3):

463–476.e6. https://doi.org/10.1016/j.chom.2021.02.003.

Haddox, Hugh K., Adam S. Dingens, Sarah K. Hilton, Julie Overbaugh, and Jesse D. Bloom. 2018. "Mapping Mutational Effects along the Evolutionary Landscape of HIV Envelope." *eLife* 7 (March): e34420. https://doi.org/10.7554/eLife.34420.

Haddox, Hugh K., Jared G. Galloway, Bernadeta Dadonaite, Jesse D. Bloom, Frederick A. Matsen, and William S. DeWitt. 2023. *Jointly Modeling Deep Mutational Scans Identifies Shifted Mutational Effects among SARS-CoV-2 Spike Homologs*. Molecular Biology. https://doi.org/10.1101/2023.07.31.551037.

Hannon, William W., and Jesse D. Bloom. 2023. "Dms-Viz: Structure-Informed Visualizations for Deep Mutational Scanning and Other Mutation-Based Datasets." In *bioRxivorg*. BioRxiv, November 1. https://doi.org/10.1101/2023.10.29.564578.

Hanson, Anthony, Masaki Imai, Masato Hatta, et al. 2015. "Identification of Stabilizing Mutations in an H5 Hemagglutinin Influenza Virus Protein." *Journal of Virology* 90 (6): 2981–2992. https://doi.org/10.1128/JVI.02790-15.

Heaton, Nicholas S., David Sachs, Chi-Jene Chen, Rong Hai, and Peter Palese. 2013. "Genome-Wide Mutagenesis of Influenza Virus Reveals Unique Plasticity of the Hemagglutinin and NS1 Proteins." *Proceedings of the National Academy of Sciences of the United States of America* 110 (50): 20248–20253. https://doi.org/10.1073/pnas.1320524110.

Hekkelman, Maarten L., Daniel Álvarez Salmoral, Anastassis Perrakis, and Robbie P. Joosten. 2025. "DSSP 4: FAIR Annotation of Protein Secondary Structure." *Protein Science* 34 (8): e70208. https://doi.org/10.1002/pro.70208.

Hill, Nichola J., Mary Anne Bishop, Nídia S. Trovão, et al. 2022. "Ecological Divergence of Wild Birds Drives Avian Influenza Spillover and Global Spread." *PLoS Pathogens* 18 (5): e1010062. https://doi.org/10.1371/journal.ppat.1010062.

Hilton, Sarah K., Michael B. Doud, and Jesse D. Bloom. 2017. "Phydms: Software for Phylogenetic Analyses Informed by Deep Mutational Scanning." *PeerJ* 5 (July): e3657. https://doi.org/10.7717/peerj.3657.

Hoffmann, E., G. Neumann, Y. Kawaoka, G. Hobom, and R. G. Webster. 2000. "A DNA

Transfection System for Generation of Influenza A Virus from Eight Plasmids." *Proceedings of the National Academy of Sciences of the United States of America* 97 (11): 6108–6113. https://doi.org/10.1073/pnas.100133697.

Hooper, Kathryn A., and Jesse D. Bloom. 2013. "A Mutant Influenza Virus That Uses an N1 Neuraminidase as the Receptor-Binding Protein." *Journal of Virology* 87 (23): 12531–12540. https://doi.org/10.1128/jvi.01889-13.

Huang, Kuan-Ying A., Pramila Rijal, Haihai Jiang, et al. 2019. "Structure-Function Analysis of Neutralizing Antibodies to H7N9 Influenza from Naturally Infected Humans." *Nature Microbiology* 4 (2): 306–315. https://doi.org/10.1038/s41564-018-0303-7.

Humes, Daryl, Stephanie Rainwater, and Julie Overbaugh. 2021. "The TOP Vector: A New High-Titer Lentiviral Construct for Delivery of sgRNAs and Transgenes to Primary T Cells." *Molecular Therapy. Methods & Clinical Development* 20 (March): 30–38. https://doi.org/10.1016/j.omtm.2020.10.020.

Imai, Masaki, Tokiko Watanabe, Masato Hatta, et al. 2012. "Experimental Adaptation of an Influenza H5 Haemagglutinin (HA) Confers Respiratory Droplet Transmission to a Reassortant H5 HA/H1N1 Virus in Ferrets." *Nature* 486 (7403): 420–428. https://doi.org/10.1038/nature10831.

Karakus, Umut, Ignacio Mena, Jithesh Kottur, et al. 2024. "H19 Influenza A Virus Exhibits Species-Specific MHC Class II Receptor Usage." *Cell Host & Microbe* 32 (7): 1089–1102.e10. https://doi.org/10.1016/j.chom.2024.05.018.

Karakus, Umut, Thiprampai Thamamongood, Kevin Ciminski, et al. 2019. "MHC Class II Proteins Mediate Cross-Species Entry of Bat Influenza Viruses." *Nature* 567 (7746): 109–112. https://doi.org/10.1038/s41586-019-0955-3.

Ke, Changwen, Chris Ka Pun Mok, Wenfei Zhu, et al. 2017. "Human Infection with Highly Pathogenic Avian Influenza A(H7N9) Virus, China." *Emerging Infectious Diseases* 23 (8): 1332–1340. https://doi.org/10.3201/eid2308.170600.

Kochs, Georg, Adolfo García-Sastre, and Luis Martínez-Sobrido. 2007. "Multiple Anti-Interferon Actions of the Influenza A Virus NS1 Protein." *Journal of Virology* 81 (13): 7011–7021.

https://doi.org/10.1128/JVI.02581-06.

Labadie, Thomas, Christophe Batéjat, Jean-Claude Manuguerra, and India Leclercq. 2018. "Influenza Virus Segment Composition Influences Viral Stability in the Environment." *Frontiers in Microbiology* 9 (1496): 1496. https://doi.org/10.3389/fmicb.2018.01496.

Lamb, R. A., C. J. Lai, and P. W. Choppin. 1981. "Sequences of mRNAs Derived from Genome RNA Segment 7 of Influenza Virus: Colinear and Interrupted mRNAs Code for Overlapping Proteins." *Proceedings of the National Academy of Sciences of the United States of America* 78 (7): 4170–4174. https://doi.org/10.1073/pnas.78.7.4170.

Lam, Tommy Tsan-Yuk, Jia Wang, Yongyi Shen, et al. 2013. "The Genesis and Source of the H7N9 Influenza Viruses Causing Human Infections in China." *Nature* 502 (7470): 241–244. https://doi.org/10.1038/nature12515.

Lee, Carolyn, Natalie N. Tarbuck, Hannah J. Cochran, et al. 2026. "Dairy Cows Infected with Influenza A(H5N1) Reveals Low Infectious Dose and Transmission Barriers." *Nature Communications*, May 24, 1–10. https://doi.org/10.1038/s41467-026-73490-6.

Lee, Juhye M., Rachel Eguia, Seth J. Zost, et al. 2019. "Mapping Person-to-Person Variation in Viral Mutations That Escape Polyclonal Serum Targeting Influenza Hemagglutinin." *eLife* 8 (e49324). https://doi.org/10.7554/eLife.49324.

Lee, Juhye M., John Huddleston, Michael B. Doud, et al. 2018. "Deep Mutational Scanning of Hemagglutinin Helps Predict Evolutionary Fates of Human H3N2 Influenza Variants." *Proceedings of the National Academy of Sciences of the United States of America* 115 (35): E8276–E8285. https://doi.org/10.1073/pnas.1806133115.

Lee, Peter S., Nobuko Ohshima, Robyn L. Stanfield, et al. 2014. "Receptor Mimicry by Antibody F045-092 Facilitates Universal Binding to the H3 Subtype of Influenza Virus." *Nature Communications* 5 (1): 3614. https://doi.org/10.1038/ncomms4614.

Lesk, A. M., and C. Chothia. 1980. "How Different Amino Acid Sequences Determine Similar Protein Structures: The Structure and Evolutionary Dynamics of the Globins." *Journal of Molecular Biology* 136 (3): 225–270. https://doi.org/10.1016/0022-2836(80)90373-3.

Linster, Martin, Sander van Boheemen, Miranda de Graaf, et al. 2014. "Identification,

Characterization, and Natural Selection of Mutations Driving Airborne Transmission of A/H5N1 Virus." *Cell* 157 (2): 329–339.

Lin, Ting-Hui, Xueyong Zhu, Shengyang Wang, et al. 2024. "A Single Mutation in Bovine Influenza H5N1 Hemagglutinin Switches Specificity to Human Receptors." *Science (New York, N.Y.)* 386 (6726): 1128–1134. https://doi.org/10.1126/science.adt0180.

Lipsitch, Marc, Wendy Barclay, Rahul Raman, et al. 2016. "Viral Factors in Influenza Pandemic Risk Assessment." *eLife* 5 (November): e18491. https://doi.org/10.7554/eLife.18491.

Liu, William J., Haixia Xiao, Lianpan Dai, et al. 08/2021. "Avian Influenza A (H7N9) Virus: From Low Pathogenic to Highly Pathogenic." *Frontiers of Medicine* 15 (4): 507–527. https://doi.org/10.1007/s11684-020-0814-5.

Liu, Yang, Olga A. Nikolaitchik, Sheikh Abdul Rahman, Jianbo Chen, Vinay K. Pathak, and Wei-Shau Hu. 2017. "HIV-1 Sequence Necessary and Sufficient to Package Non-Viral RNAs into HIV-1 Particles." *Journal of Molecular Biology* 429 (16): 2542–2555. https://doi.org/10.1016/j.jmb.2017.06.018.

Loes, Andrea N., Rosario Araceli L. Tarabi, John Huddleston, et al. 2024. "High-Throughput Sequencing-Based Neutralization Assay Reveals How Repeated Vaccinations Impact Titers to Recent Human H1N1 Influenza Strains." *Journal of Virology* 98 (10): e0068924. https://doi.org/10.1128/jvi.00689-24.

Long, Jason S., Bhakti Mistry, Stuart M. Haslam, and Wendy S. Barclay. 2019. "Host and Viral Determinants of Influenza A Virus Species Specificity." *Nature Reviews. Microbiology* 17 (2): 67–81. https://doi.org/10.1038/s41579-018-0115-z.

Mair, Caroline M., Kai Ludwig, Andreas Herrmann, and Christian Sieben. 2014. "Receptor Binding and pH Stability - How Influenza A Virus Hemagglutinin Affects Host-Specific Virus Infection." *Biochimica et Biophysica Acta* 1838 (4): 1153–1168. https://doi.org/10.1016/j.bbamem.2013.10.004.

Marks, Debora S., Lucy J. Colwell, Robert Sheridan, et al. 2011. "Protein 3D Structure Computed from Evolutionary Sequence Variation." *PloS One* 6 (12): e28766. https://doi.org/10.1371/journal.pone.0028766.

Marks, Debora S., Thomas A. Hopf, and Chris Sander. 2012. "Protein Structure Prediction from Sequence Variation." *Nature Biotechnology* 30 (11): 1072–1080. https://doi.org/10.1038/nbt.2419.

Matrosovich, M., A. Tuzikov, N. Bovin, et al. 2000. "Early Alterations of the Receptor-Binding Properties of H1, H2, and H3 Avian Influenza Virus Hemagglutinins after Their Introduction into Mammals." *Journal of Virology* 74 (18): 8502–8512. https://doi.org/10.1128/jvi.74.18.8502-8512.2000.

Meng, Elaine C., Thomas D. Goddard, Eric F. Pettersen, et al. 2023. "UCSF ChimeraX: Tools for Structure Building and Analysis." *Protein Science* 32 (11): e4792. https://doi.org/10.1002/pro.4792.

Milder, Fin J., Mandy Jongeneelen, Tina Ritschel, et al. 2022. "Universal Stabilization of the Influenza Hemagglutinin by Structure-Based Redesign of the pH Switch Regions." *Proceedings of the National Academy of Sciences of the United States of America* 119 (6): e2115379119. https://doi.org/10.1073/pnas.2115379119.

Miller, S., J. Janin, A. M. Lesk, and C. Chothia. 1987. "Interior and Surface of Monomeric Proteins." *Journal of Molecular Biology* 196 (3): 641–656. https://doi.org/10.1016/0022-2836(87)90038-6.

Mosyak, L., D. M. Zaller, and D. C. Wiley. 1998. "The Structure of HLA-DM, the Peptide Exchange Catalyst That Loads Antigen onto Class II MHC Molecules during Antigen Presentation." *Immunity* 9 (3): 377–383. https://doi.org/10.1016/s1074-7613(00)80620-2.

Moulana, Alief, Thomas Dupic, and Michael M. Desai. 2025. "Epistasis and Background Dependence in the Evolution of Omicron Variants of the SARS-CoV-2 Spike Protein." In *bioRxivorg*. BioRxiv, November 19. https://doi.org/10.1101/2025.11.18.689118.

Moulana, Alief, Thomas Dupic, Angela M. Phillips, et al. 2022. "Compensatory Epistasis Maintains ACE2 Affinity in SARS-CoV-2 Omicron BA.1." *Nature Communications* 13 (1): 7011. https://doi.org/10.1038/s41467-022-34506-z.

Moulana, Alief, Thomas Dupic, Angela M. Phillips, et al. 2023. "The Landscape of Antibody Binding Affinity in SARS-CoV-2 Omicron BA.1 Evolution." *eLife* 12 (e83442): e83442.

https://doi.org/10.7554/eLife.83442.

Muñoz, V., and L. Serrano. 1994. "Intrinsic Secondary Structure Propensities of the Amino Acids, Using Statistical Phi-Psi Matrices: Comparison with Experimental Scales." *Proteins* 20 (4): 301–311. https://doi.org/10.1002/prot.340200403.

Murti, K. G., R. G. Webster, and I. M. Jones. 1988. "Localization of RNA Polymerases on Influenza Viral Ribonucleoproteins by Immunogold Labeling." *Virology* 164 (2): 562–566. https://doi.org/10.1016/0042-6822(88)90574-0.

Narimatsu, Yoshiki, Hiren J. Joshi, Rebecca Nason, et al. 2019. "An Atlas of Human Glycosylation Pathways Enables Display of the Human Glycome by Gene Engineered Cells." *Molecular Cell* 75 (2): 394–407.e5. https://doi.org/10.1016/j.molcel.2019.05.017.

Nguyen, Lam-Tung, Heiko A. Schmidt, Arndt von Haeseler, and Bui Quang Minh. 2015. "IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies." *Molecular Biology and Evolution* 32 (1): 268–274. https://doi.org/10.1093/molbev/msu300.

Nikolaitchik, Olga A., Saiful Islam, Jonathan P. Kitzrow, et al. 2023. "HIV-1 Usurps Transcription Start Site Heterogeneity of Host RNA Polymerase II to Maximize Replication Fitness." *Proceedings of the National Academy of Sciences of the United States of America* 120 (23): e2305103120. https://doi.org/10.1073/pnas.2305103120.

Ning, Tingting, Jianhui Nie, Weijin Huang, et al. 2019. "Antigenic Drift of Influenza A(H7N9) Virus Hemagglutinin." *The Journal of Infectious Diseases* 219 (1): 19–25. https://doi.org/10.1093/infdis/jiy408.

Nobusawa, E., H. Ishihara, T. Morishita, K. Sato, and K. Nakajima. 2000. "Change in Receptor-Binding Specificity of Recent Human Influenza A Viruses (H3N2): A Single Amino Acid Change in Hemagglutinin Altered Its Recognition of Sialyloligosaccharides." *Virology* 278 (2): 587–596. https://doi.org/10.1006/viro.2000.0679.

Obadan, Adebimpe O., Jefferson Santos, Lucas Ferreri, et al. 2019. "Flexibility in Vitro of Amino Acid 226 in the Receptor-Binding Site of an H9 Subtype Influenza A Virus and Its Effect in Vivo on Virus Replication, Tropism, and Transmission." *Journal of Virology* 93 (6): e02011–18.

https://doi.org/10.1128/JVI.02011-18.

Olajide, Okikiola M., Maria Kaukab Osman, Jonathan Robert, et al. 2023. "Evolutionarily Conserved Amino Acids in MHC-II Mediate Bat Influenza A Virus Entry into Human Cells." *PLoS Biology* 21 (7): e3002182. https://doi.org/10.1371/journal.pbio.3002182.

Otwinowski, Jakub, David M. McCandlish, and Joshua B. Plotkin. 2018. "Inferring the Shape of Global Epistasis." *Proceedings of the National Academy of Sciences of the United States of America* 115 (32). https://doi.org/10.1073/pnas.1804015115.

Ovchinnikov, Sergey, Hetunandan Kamisetty, and David Baker. 2014. "Robust and Accurate Prediction of Residue-Residue Interactions across Protein Interfaces Using Evolutionary Information." *eLife* 3 (May): e02030. https://doi.org/10.7554/eLife.02030.

Palese, Peter. 2004. "Influenza: Old and New Threats." *Nature Medicine* 10 (12 Suppl): S82–7. https://doi.org/10.1038/nm1141.

Park, Yeonwoo, Brian P. H. Metzger, and Joseph W. Thornton. 2022. "Epistatic Drift Causes Gradual Decay of Predictability in Protein Evolution." *Science (New York, N.Y.)* 376 (6595): 823–830. https://doi.org/10.1126/science.abn6895.

Parrish, Colin R., Pablo R. Murcia, and Edward C. Holmes. 2015. "Influenza Virus Reservoirs and Intermediate Hosts: Dogs, Horses, and New Possibilities for Influenza Virus Exposure of Humans." *Journal of Virology* 89 (6): 2990–2994. https://doi.org/10.1128/JVI.03146-14.

Pollock, David D., Grant Thiltgen, and Richard A. Goldstein. 2012. "Amino Acid Coevolution Induces an Evolutionary Stokes Shift." *Proceedings of the National Academy of Sciences of the United States of America* 109 (21): E1352–9. https://doi.org/10.1073/pnas.1120084109.

Popova, Lyubov, Kenneth Smith, Ann H. West, et al. 2012. "Immunodominance of Antigenic Site B over Site A of Hemagglutinin of Recent H3N2 Influenza Viruses." *PloS One* 7 (7): e41895. https://doi.org/10.1371/journal.pone.0041895.

Poulson, R. L., S. M. Tompkins, R. D. Berghaus, J. D. Brown, and D. E. Stallknecht. 2016. "Environmental Stability of Swine and Human Pandemic Influenza Viruses in Water under Variable Conditions of Temperature, Salinity, and pH." *Applied and Environmental Microbiology* 82 (13): 3721–3726. https://doi.org/10.1128/AEM.00133-16.

Rajao, Daniela, Matias Cardenas, Sasha Compton, C. Joaquin Caceres, Adolfo Garcia-Sastre, and Daniel Perez. 2025. "MHC Class II Is a Functional Receptor for H3N2 Influenza A Viruses and Mediates Host-Specificity." In *Research Square*. May 2. https://doi.org/10.21203/rs.3.rs-6278351/v1.

Reed, Mark L., Olga A. Bridges, Patrick Seiler, et al. 2010. "The pH of Activation of the Hemagglutinin Protein Regulates H5N1 Influenza Virus Pathogenicity and Transmissibility in Ducks." *Journal of Virology* 84 (3): 1527–1535. https://doi.org/10.1128/JVI.02069-09.

Roche, Paul A., and Kazuyuki Furuta. 2015. "The Ins and Outs of MHC Class II-Mediated Antigen Processing and Presentation." *Nature Reviews. Immunology* 15 (4): 203–216. https://doi.org/10.1038/nri3818.

Rock, Kenneth L., Eric Reits, and Jacques Neefjes. 2016. "Present Yourself! By MHC Class I and MHC Class II Molecules." *Trends in Immunology* 37 (11): 724–737. https://doi.org/10.1016/j.it.2016.08.010.

Rogers, G. N., and J. C. Paulson. 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 (2): 361–373. https://doi.org/10.1016/0042-6822(83)90150-2.

Rogers, G. N., J. C. Paulson, R. S. Daniels, J. J. Skehel, I. A. Wilson, and D. C. Wiley. 1983. "Single Amino Acid Substitutions in Influenza Haemagglutinin Change Receptor Binding Specificity." *Nature* 304 (5921): 76–78. https://doi.org/10.1038/304076a0.

Ruigrok, R. W., A. Aitken, L. J. Calder, et al. 1988. "Studies on the Structure of the Influenza Virus Haemagglutinin at the pH of Membrane Fusion." *The Journal of General Virology* 69 (Pt 11) (November): 2785–2795. https://doi.org/10.1099/0022-1317-69-11-2785.

Ruigrok, R. W., S. R. Martin, S. A. Wharton, J. J. Skehel, P. M. Bayley, and D. C. Wiley. 1986. "Conformational Changes in the Hemagglutinin of Influenza Virus Which Accompany Heat-Induced Fusion of Virus with Liposomes." *Virology* 155 (2): 484–497. https://doi.org/10.1016/0042-6822(86)90210-2.

Russell, Charles J., Meng Hu, and Faten A. Okda. 2018. "Influenza Hemagglutinin Protein Stability, Activation, and Pandemic Risk." *Trends in Microbiology* 26 (10): 841–853.

https://doi.org/10.1016/j.tim.2018.03.005.

Russier, Marion, Guohua Yang, Benoit Briard, et al. 2020. "Hemagglutinin Stability Regulates H1N1 Influenza Virus Replication and Pathogenicity in Mice by Modulating Type I Interferon Responses in Dendritic Cells." *Journal of Virology* 94 (3). https://doi.org/10.1128/JVI.01423-19.

Russier, Marion, Guohua Yang, Jerold E. Rehg, et al. 2016. "Molecular Requirements for a Pandemic Influenza Virus: An Acid-Stable Hemagglutinin Protein." *Proceedings of the National Academy of Sciences of the United States of America* 113 (6): 1636–1641. https://doi.org/10.1073/pnas.1524384113.

Samji, Tasleem. 2009. "Influenza A: Understanding the Viral Life Cycle." *The Yale Journal of Biology and Medicine* 82 (4): 153–159. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2794490/.

Sanderson, F., M. J. Kleijmeer, A. Kelly, et al. 1994. "Accumulation of HLA-DM, a Regulator of Antigen Presentation, in MHC Class II Compartments." *Science (New York, N.Y.)* 266 (5190): 1566–1569. https://doi.org/10.1126/science.7985027.

Scholtissek, Ch. 3/1985. "Stability of Infectious Influenza A Viruses to Treatment at Low pH and Heating." *Archives of Virology* 85 (1-2): 1–11. https://doi.org/10.1007/BF01317001.

Shi, Jianzhong, Guohua Deng, Peihong Liu, et al. 2013. "Isolation and Characterization of H7N9 Viruses from Live Poultry Markets — Implication of the Source of Current H7N9 Infection in Humans." *Kexue Tongbao [Chinese Science Bulletin]* 58 (16): 1857–1863. https://doi.org/10.1007/s11434-013-5873-4.

Shi, Jianzhong, Xianying Zeng, Pengfei Cui, Cheng Yan, and Hualan Chen. 2023. "Alarming Situation of Emerging H5 and H7 Avian Influenza and Effective Control Strategies." *Emerging Microbes & Infections* 12 (1): 2155072. https://doi.org/10.1080/22221751.2022.2155072.

Shi, Yi, Wei Zhang, Fei Wang, et al. 2013. "Structures and Receptor Binding of Hemagglutinins from Human-Infecting H7N9 Influenza Viruses." *Science (New York, N.Y.)* 342 (6155): 243–247. https://doi.org/10.1126/science.1242917.

Shore, David A., Hua Yang, Amanda L. Balish, et al. 2013. "Structural and Antigenic Variation among Diverse Clade 2 H5N1 Viruses." *PloS One* 8 (9): e75209. https://doi.org/10.1371/journal.pone.0075209.

Singanayagam, Anika, Maria Zambon, and Wendy S. Barclay. 2019. "Influenza Virus with Increased pH of Hemagglutinin Activation Has Improved Replication in Cell Culture but at the Cost of Infectivity in Human Airway Epithelium." *Journal of Virology* 93 (17): e00058–19. https://doi.org/10.1128/JVI.00058-19.

Si, Youhui, Jianguo Li, Yuqiang Niu, et al. 2014. "Entry Properties and Entry Inhibitors of a Human H7N9 Influenza Virus." *PloS One* 9 (9): e107235. https://doi.org/10.1371/journal.pone.0107235.

Skehel, John J., and Don C. Wiley. 2000. "Receptor Binding and Membrane Fusion in Virus Entry: The Influenza Hemagglutinin." *Annual Review of Biochemistry* 69 (1): 531–569. https://doi.org/10.1146/annurev.biochem.69.1.531.

Smith, Derek J., Alan S. Lapedes, Jan C. de Jong, et al. 2004. "Mapping the Antigenic and Genetic Evolution of Influenza Virus." *Science (New York, N.Y.)* 305 (5682): 371–376. https://doi.org/10.1126/science.1097211.

Starr, Tyler N., Allison J. Greaney, Amin Addetia, et al. 2020. "Prospective Mapping of Viral Mutations That Escape Antibodies Used to Treat COVID-19." In *bioRxivorg*. BioRxiv, December 1. https://doi.org/10.1101/2020.11.30.405472.

Starr, Tyler N., Allison J. Greaney, Adam S. Dingens, and Jesse D. Bloom. 2021. "Complete Map of SARS-CoV-2 RBD Mutations That Escape the Monoclonal Antibody LY-CoV555 and Its Cocktail with LY-CoV016." *Cell Reports. Medicine* 2 (4): 100255. https://doi.org/10.1016/j.xcrm.2021.100255.

Starr, Tyler N., Allison J. Greaney, William W. Hannon, et al. 2022. "Shifting Mutational Constraints in the SARS-CoV-2 Receptor-Binding Domain during Viral Evolution." *Science (New York, N.Y.)* 377 (6604): 420–424. https://doi.org/10.1126/science.abo7896.

Starr, Tyler N., Allison J. Greaney, Sarah K. Hilton, et al. 2020. "Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2

Binding." *Cell* 182 (5): 1295–1310.e20. https://doi.org/10.1016/j.cell.2020.08.012.

Starr, Tyler N., Allison J. Greaney, Cameron M. Stewart, et al. 2022. "Deep Mutational Scans for ACE2 Binding, RBD Expression, and Antibody Escape in the SARS-CoV-2 Omicron BA.1 and BA.2 Receptor-Binding Domains." *PLoS Pathogens* 18 (11): e1010951. https://doi.org/10.1371/journal.ppat.1010951.

Starr, Tyler N., and Joseph W. Thornton. 2016. "Epistasis in Protein Evolution: Epistasis in Protein Evolution." *Protein Science: A Publication of the Protein Society* 25 (7): 1204–1218. https://doi.org/10.1002/pro.2897.

Stevens, James, Ola Blixt, Terrence M. Tumpey, Jeffery K. Taubenberger, James C. Paulson, and Ian A. Wilson. 2006. "Structure and Receptor Specificity of the Hemagglutinin from an H5N1 Influenza Virus." *Science (New York, N.Y.)* 312 (5772): 404–410. https://doi.org/10.1126/science.1124513.

Sun, Xiangjie, Jessica A. Belser, Joanna A. Pulit-Penaloza, et al. 2024. "Dissecting the Role of the HA1-226 Leucine Residue in the Fitness and Airborne Transmission of an A(H9N2) Avian Influenza Virus." *Journal of Virology* 98 (12): e0092824. https://doi.org/10.1128/jvi.00928-24.

Sun, Xiaoman, Yi Shi, Xishan Lu, et al. 2013. "Bat-Derived Influenza Hemagglutinin H17 Does Not Bind Canonical Avian or Human Receptors and Most Likely Uses a Unique Entry Mechanism." *Cell Reports* 3 (3): 769–778. https://doi.org/10.1016/j.celrep.2013.01.025.

Sutton, Troy C. 2018. "The Pandemic Threat of Emerging H5 and H7 Avian Influenza Viruses." *Viruses* 10 (9): 461. https://doi.org/10.3390/v10090461.

Tanner, W. D., D. J. A. Toth, and A. V. Gundlapalli. 2015. "The Pandemic Potential of Avian Influenza A(H7N9) Virus: A Review." *Epidemiology and Infection* 143 (16): 3359–3374. https://doi.org/10.1017/S0950268815001570.

Taylor, Ashley L., and Tyler N. Starr. 2024. "Deep Mutational Scanning of SARS-CoV-2 Omicron BA.2.86 and Epistatic Emergence of the KP.3 Variant." *Virus Evolution* 10 (1): veae067. https://doi.org/10.1093/ve/veae067.

Tharakaraman, Kannan, Akila Jayaraman, Rahul Raman, et al. 2013. "Glycan-Receptor Binding of the Influenza A Virus H7N9 Hemagglutinin." *Cell* 153 (7): 1486–1493.

https://doi.org/10.1016/j.cell.2013.05.034.

Thyagarajan, Bargavi, and Jesse D. Bloom. 2014. "The Inherent Mutational Tolerance and Antigenic Evolvability of Influenza Hemagglutinin." *eLife* 3 (July): e03300. https://doi.org/10.7554/eLife.03300.

Tien, Matthew Z., Austin G. Meyer, Dariya K. Sydykova, Stephanie J. Spielman, and Claus O. Wilke. 2013. "Maximum Allowed Solvent Accessibilites of Residues in Proteins." *PloS One* 8 (11): e80635. https://doi.org/10.1371/journal.pone.0080635.

Timofeeva, T. A., G. K. Sadykova, N. F. Lomakina, et al. 2020. "The Effect of I155T, K156Q, K156E and N186K Mutations in Hemagglutinin on the Virulence and Reproduction of Influenza A/H5N1 Viruses." *Molecular Biology* 54 (6): 861–869. https://doi.org/10.1134/S0026893320060126.

"Tool for Influenza Pandemic Risk Assessment (TIPRA)." n.d. Accessed April 9, 2026. https://www.who.int/teams/global-influenza-programme/avian-influenza/tool-for-influenza-pandemic-risk-assessment-(tipra).

Unanue, Emil R., Vito Turk, and Jacques Neefjes. 2016. "Variations in MHC Class II Antigen Processing and Presentation in Health and Disease." *Annual Review of Immunology* 34 (1): 265–297. https://doi.org/10.1146/annurev-immunol-041015-055420.

Uyeki, Timothy M., David S. Hui, Maria Zambon, David E. Wentworth, and Arnold S. Monto. 2022. "Influenza." *Lancet* 400 (10353): 693–706. https://doi.org/10.1016/S0140-6736(22)00982-5.

Vries, Robert P. de, Wenjie Peng, Oliver C. Grant, et al. 2017. "Three Mutations Switch H7N9 Influenza to Human-Type Receptor Specificity." *PLoS Pathogens* 13 (6): e1006390. https://doi.org/10.1371/journal.ppat.1006390.

Wang, Wei, Hyo Sook Song, Paul W. Keller, Esmeralda Alvarado-Facundo, Russell Vassell, and Carol D. Weiss. 2018. "Conformational Stability of the Hemagglutinin of H5N1 Influenza A Viruses Influences Susceptibility to Broadly Neutralizing Stem Antibodies." *Journal of Virology* 92 (12): e00247–18. https://doi.org/10.1128/JVI.00247-18.

Watanabe, Tokiko, Shinji Watanabe, Eileen A. Maher, Gabriele Neumann, and Yoshihiro

Kawaoka. 2014. “Pandemic Potential of Avian Influenza A (H7N9) Viruses.” *Trends in Microbiology* 22 (11): 623–631. https://doi.org/10.1016/j.tim.2014.08.008.

Weber, T., G. Paesold, C. Galli, R. Mischler, G. Semenza, and J. Brunner. 1994. “Evidence for H(+)-Induced Insertion of Influenza Hemagglutinin HA2 N-Terminal Segment into Viral Membrane.” *The Journal of Biological Chemistry* 269 (28): 18353–18358. https://doi.org/10.1016/s0021-9258(17)32314-1.

Weigt, Martin, Robert A. White, Hendrik Szurmant, James A. Hoch, and Terence Hwa. 2009. “Identification of Direct Residue Contacts in Protein-Protein Interaction by Message Passing.” *Proceedings of the National Academy of Sciences of the United States of America* 106 (1): 67–72. https://doi.org/10.1073/pnas.0805923106.

Welsh, Frances C., Rachel T. Eguia, Juhye M. Lee, et al. 2024. “Age-Dependent Heterogeneity in the Antigenic Effects of Mutations to Influenza Hemagglutinin.” *Cell Host & Microbe* 32 (8): 1397–1411.e11. https://doi.org/10.1016/j.chom.2024.06.015.

Wen, Yu-Mei, and Hans-Dieter Klenk. 2013. “H7N9 Avian Influenza Virus - Search and Re-Search: H7N9 Avian Influenza Virus - Search and Re-Search.” *Emerging Microbes & Infections* 2 (4): e18. https://doi.org/10.1038/emi.2013.18.

Wharton, S. A., L. J. Calder, R. W. Ruigrok, J. J. Skehel, D. A. Steinhauer, and D. C. Wiley. 1995. “Electron Microscopy of Antibody Complexes of Influenza Virus Haemagglutinin in the Fusion pH Conformation.” *The EMBO Journal* 14 (2): 240–246. https://doi.org/10.1002/j.1460-2075.1995.tb06997.x.

WHO. 2025. “Summary of Status of Development and Availability of Avian Influenza A(H7N9) Candidate Vaccine Viruses and Potency Testing Reagents.” Preprint. https://cdn.who.int/media/docs/default-source/influenza/cvvs/cvv-zoonotic-northern-hemisphere-2025-2026/h7n9_summary_a_h7n9_cvv_20250228.pdf?sfvrsn=b116cfda_5.

WHO. n.d. “Influenza A(H7N9) Candidate Vaccine Viruses and Potency Testing Reagents.” Preprint. https://cdn.who.int/media/docs/default-source/influenza/cvvs/cvv-zoonotic-northern-hemisphere-2024-2025/h7n9_summary_a_h7n9_cvv_-20240223.pdf?sfvrsn=5371c12d_3.

Wieczorek, Marek, Esam T. Abualrous, Jana Sticht, et al. 2017. "Major Histocompatibility Complex (MHC) Class I and MHC Class II Proteins: Conformational Plasticity in Antigen Presentation." *Frontiers in Immunology* 8 (March): 292. https://doi.org/10.3389/fimmu.2017.00292.

Wu, Nicholas C., Jakub Otwinowski, Andrew J. Thompson, Corwin M. Nycholat, Armita Nourmohammad, and Ian A. Wilson. 2020. "Major Antigenic Site B of Human Influenza H3N2 Viruses Has an Evolving Local Fitness Landscape." *Nature Communications* 11 (1): 1233. https://doi.org/10.1038/s41467-020-15102-5.

Wu, Nicholas C., Andrew J. Thompson, Juhye M. Lee, et al. 2020. "Different Genetic Barriers for Resistance to HA Stem Antibodies in Influenza H3 and H1 Viruses." *Science (New York, N.Y.)* 368 (6497): 1335–1340. https://doi.org/10.1126/science.aaz5143.

Wu, Nicholas C., Andrew J. Thompson, Jia Xie, et al. 2018. "A Complex Epistatic Network Limits the Mutational Reversibility in the Influenza Hemagglutinin Receptor-Binding Site." *Nature Communications* 9 (1): 1264. https://doi.org/10.1038/s41467-018-03663-5.

Wu, Nicholas C., and Ian A. Wilson. 2020. "Structural Biology of Influenza Hemagglutinin: An Amaranthine Adventure." *Viruses* 12 (9): 1053. https://doi.org/10.3390/v12091053.

Wu, Nicholas C., Jia Xie, Tianqing Zheng, et al. 2017. "Diversity of Functionally Permissive Sequences in the Receptor-Binding Site of Influenza Hemagglutinin." *Cell Host & Microbe* 21 (6): 742–753.e8. https://doi.org/10.1016/j.chom.2017.05.011.

Wu, Nicholas C., Arthur P. Young, Laith Q. Al-Mawsawi, et al. 2014. "High-Throughput Profiling of Influenza A Virus Hemagglutinin Gene at Single-Nucleotide Resolution." *Scientific Reports* 4 (1): 4942. https://doi.org/10.1038/srep04942.

Xiang, Nijuan. 2016. "Assessing Change in Avian Influenza A(H7N9) Virus Infections During the Fourth Epidemic — China, September 2015–August 2016." *MMWR. Morbidity and Mortality Weekly Report* 65. https://doi.org/10.15585/mmwr.mm6549a2.

Xu, Rui, Damian C. Ekiert, Jens C. Krause, Rong Hai, James E. Crowe Jr, and Ian A. Wilson. 2010. "Structural Basis of Preexisting Immunity to the 2009 H1N1 Pandemic Influenza Virus." *Science (New York, N.Y.)* 328 (5976): 357–360.

https://doi.org/10.1126/science.1186430.

Yang, Guohua, Chet R. Ojha, and Charles J. Russell. 2021. "Relationship between Hemagglutinin Stability and Influenza Virus Persistence after Exposure to Low pH or Supraphysiological Heating." *PLoS Pathogens* 17 (9): e1009910. https://doi.org/10.1371/journal.ppat.1009910.

Yang, Wenyu, Tony Schountz, and Wenjun Ma. 2021. "Bat Influenza Viruses: Current Status and Perspective." *Viruses* 13 (4): 547. https://doi.org/10.3390/v13040547.

Yu, Timothy C., Caroline Kikawa, Bernadeta Dadonaite, Andrea N. Loes, Janet A. Englund, and Jesse D. Bloom. 2025. "Pleiotropic Mutational Effects on Function and Stability Constrain the Antigenic Evolution of Influenza Haemagglutinin." *Nature Ecology & Evolution*, December 1, 1–15. https://doi.org/10.1038/s41559-025-02895-1.

Yu, Timothy C., Zorian T. Thornton, William W. Hannon, et al. 2022. "A Biophysical Model of Viral Escape from Polyclonal Antibodies." *Virus Evolution* 8 (2): veac110. https://doi.org/10.1093/ve/veac110.

Zaraket, Hassan, Olga A. Bridges, Susu Duan, et al. 2013-9. "Increased Acid Stability of the Hemagglutinin Protein Enhances H5N1 Influenza Virus Growth in the Upper Respiratory Tract but Is Insufficient for Transmission in Ferrets." *Journal of Virology* 87 (17): 9911–9922. https://doi.org/10.1128/JVI.01175-13.

Zebedee, S. L., and R. A. Lamb. 1988. "Influenza A Virus M2 Protein: Monoclonal Antibody Restriction of Virus Growth and Detection of M2 in Virions." *Journal of Virology* 62 (8): 2762–2772. https://doi.org/10.1128/JVI.62.8.2762-2772.1988.

Zuckerkandl, Emile, and Linus Pauling. 1965. "Evolutionary Divergence and Convergence in Proteins." In *Evolving Genes and Proteins*. Elsevier. https://doi.org/10.1016/b978-1-4832-2734-4.50017-6.