Pleiotropic mutational effects on function and stability constrain the antigenic evolution of influenza hemagglutinin

Timothy C. Yu[1,2], Caroline Kikawa[1,3,4], Bernadeta Dadonaite[1], Andrea N. Loes[1,5], Janet A. Englund[6,7], Jesse D. Bloom[1,3,5,*]

[1] Division of Basic Sciences and Computational Biology Program, Fred Hutchinson Cancer Center, Seattle, WA
[2] Molecular and Cellular Biology Graduate Program, University of Washington, Seattle, WA
[3] Department of Genome Sciences, University of Washington, Seattle, WA
[4] Medical Scientist Training Program, University of Washington, Seattle, WA
[5] Howard Hughes Medical Institute, Seattle, WA
[6] Seattle Children's Research Institute, Seattle, WA
[7] Department of Pediatrics, University of Washington, Seattle, WA

[*] Corresponding author. E-mail: jbloom@fredhutch.org

Abstract

The evolution of human influenza virus hemagglutinin (HA) involves simultaneous selection to acquire antigenic mutations that escape population immunity while preserving protein function and stability. Epistasis shapes this evolution, as an antigenic mutation that is deleterious in one genetic background may become tolerated in another. However, the extent to which epistasis can alleviate pleiotropic conflicts between immune escape and protein function/stability is unclear. Here, we measure how all amino acid mutations in the HA of a recent human H3N2 influenza strain affect its cell entry function, acid stability, and neutralization by human serum antibodies. We find that epistasis has enabled emergence of antigenic mutations that were detrimental to HA's cell entry function in earlier strains. However, epistasis appears insufficient to overcome the pleiotropic costs of antigenic mutations that impair HA's stability, explaining why some mutations that strongly escape human antibodies never fix in nature. Our results refine our understanding of the mutational constraints that shape influenza evolution: epistasis can enable antigenic change, but pleiotropic effects can restrict its trajectory.

Introduction

The evolution of human influenza viruses is shaped by selection from population immunity. This immune pressure is especially apparent in the evolution of the viral hemagglutinin (HA) protein, which is the major target of neutralizing antibodies[1,2]. Due largely to antibody-mediated immune pressure, the HA of human H3N2 influenza virus fixes an average of 3-4 amino acid substitutions per year[3–7].

However, HA's antigenic evolution is constrained by its essential role in viral fitness, as it binds to the host cell receptor (sialic acid) and mediates fusion of the viral and cell membranes. Many mutations impair these functions, yet human influenza viruses have demonstrated—both in the laboratory and in nature—the ability to rapidly adapt in the face of these constraints[8–15]. Prior work has revealed that epistatic interactions among mutations in HA play a role in facilitating these adaptations. For example, mutations that impair receptor binding can become tolerated in the presence of other mutations that help restore binding[8–14], while mutations that enhance receptor binding can help buffer the effects of deleterious mutations[15].

Although epistasis among mutations affecting receptor binding is an established mechanism for resolving constraints on HA evolution, it is less clear how epistasis involving other molecular phenotypes shapes evolution. Many studies have used deep mutational scanning to measure the effects of HA mutations at scale, but these measurements have been typically limited to a single phenotype: cell entry or viral replication in cell culture[9–11,16–20]. One phenotype with poorly understood evolutionary constraints is HA acid stability. As influenza virions are internalized into acidifying endosomes, HA undergoes a pH-triggered destabilization from its metastable pre-fusion form to a conformation that is primed for mediating membrane fusion[21]. While the structural transitions of this conformational change have been characterized in exquisite detail[22], we lack a complete understanding of how mutations to human H3N2 influenza HA affect acid stability, whether these effects impose pleiotropic costs on antigenic evolution, and whether such costs can be alleviated through epistasis.

Here, we used pseudovirus deep mutational scanning[23,24] to measure how all amino acid mutations in the HA of a recent human H3N2 influenza strain affect cell entry, acid stability, and neutralization by human serum antibodies. By comparing these data to amino acid frequencies at each HA site observed in natural viral evolution, we assessed the extent to which epistasis modulates mutation effects on cell entry and acid stability, and whether pleiotropic costs on these phenotypes could be alleviated to enable antigenic mutations to fix. The effects of many mutations on cell entry have changed over time—and epistatic interactions enabled an antigenic mutation that was highly deleterious to cell entry in 2009 to eventually fix in 2022. However, the effects of mutations on acid stability show little evidence of epistasis, and several antigenic mutations that pleiotropically reduce acid stability have never fixed. Our results indicate that

epistasis plays a central role in driving HA evolution, but its contribution to overcoming pleiotropic costs depends on the underlying molecular phenotype.

Results

Pseudovirus deep mutational scanning of HA from a recent human H3N2 strain

To measure the effects of mutations to HA on different key molecular phenotypes in high-throughput, we used a recently developed pseudovirus deep mutational scanning approach[23,24]. In brief, we generated genotype-phenotype linked pseudovirus libraries where each virion encodes a mutant HA gene in its genome that matches the HA protein expressed on its surface (**Fig. S1A-B**). Each mutant HA is coupled to a unique nucleotide barcode, forming variants whose phenotypic effects can be measured in a single multiplexed experiment via short-read sequencing of the barcodes (**Fig. 1A**). These pseudoviruses also express the matched H3N2 neuraminidase (NA) on their surfaces, but this NA is supplied from a separate plasmid and not encoded in the pseudovirus genome. Importantly, these pseudoviruses encode no viral genes other than HA and can only undergo a single round of cell entry. These pseudoviruses are therefore not infectious agents capable of causing disease and so provide a safe way to study HA mutations at biosafety-level 2.

We created duplicate libraries in the background of the HA from A/Massachusetts/18/2022 (MA22), which was the H3N2 strain included in the 2024-2025 seasonal influenza vaccine[25]. These libraries were designed to contain every possible amino acid mutation in the HA ectodomain (H3 numbering: 1 to 504) for a total of 504 x 19 = 9576 mutations. The final libraries contained 64,032 and 70,581 barcoded HA variants that covered 98.7% and 99.0% of all possible mutations, respectively (**Fig. S1C**). Most variants contained a single HA mutation (65%), while others contained zero (15%) or multiple mutations (20%) (**Fig. S1D**). To extract information from the multiply mutated variants, we used global epistasis models[26,27] to disentangle the effects of individual mutations on measured phenotypes (**Methods**).

Mutation effects on HA-mediated cell entry

We quantified the effects of HA mutations on pseudovirus entry into MDCK-SIAT1 cells[28], which express high levels of the a2-6-linked sialic acids preferred by human influenza HAs (**Fig. 1A-B, Fig. S3**, and interactive heatmaps at https://dms-vep.org/Flu_H3_Massachusetts2022_DMS/cell_entry.html). A negative effect indicates the mutation impairs cell entry, potentially via one or a combination of reasons such as impaired receptor binding, fusion competency, HA folding, or HA expression. The measurements of mutation effects were highly correlated between the two independent replicate libraries (r=0.95, **Fig. S2A**). We validated the deep mutational scanning measurements of mutation effects on cell entry for 15 mutations with a range of effects on cell entry using conditionally replicative influenza viruses that lack the PB1 gene (**Fig. S5A**)[29,30]. The titers of the conditionally replicative influenza virions were highly correlated with the cell entry effects obtained by deep mutational scanning (r=0.88, **Fig. S5C**).

The deep mutational scanning showed that the receptor binding pocket is heavily constrained overall (**Fig. 1C**), but that the distribution of cell entry effects varies across its different structural regions (**Table S1**). Some areas, like the base of the receptor binding pocket, are highly constrained, while others, like the 150-loop, are more tolerant of mutations (**Fig. 1D**, **Fig. S3**). These measurements of the tolerance of different regions to mutations with respect to their effects on cell entry correlate with the variability of different sites in HA during the natural evolution of human H3N2 since 1968 (compare mutation effects and entropy among natural sequences in **Fig. 1D**). For example, the 150-loop has seen substantial divergence during H3N2 HA's natural evolution, whereas the base of the receptor-binding pocket remains highly conserved, with only a single substitution at site 195 (Y to F) occurring in the early 2020s.

Many mutations within classically defined antigenic regions (epitopes A-E)[31,32] tend to be well tolerated for HA's cell entry function (**Fig. 1E**). However, epitopes that partially overlap with the receptor binding pocket (epitopes A, B, and D) are more constrained. Of these, epitope B is the most constrained yet exhibits the highest variability among natural human H3N2 influenza sequences (**Fig. 1E**). This discrepancy likely reflects the immunodominance of epitope B with respect to antibody neutralization in the human population, which imposes strong positive selection for mutations at sites within the region[33–35].

The deep mutational scanning also showed that nearly all mutations to the highly conserved fusion loop at sites 330-350 are strongly deleterious to cell entry (**Fig. 1F, Fig. S3**), consistent with the region's key role in mediating membrane fusion and also with prior deep mutational scanning on HAs from other viral subtypes[17–20,24,36]. Taken together, we have measured the effects of nearly all mutations to a recent H3N2 HA on cell entry; these measurements help explain the functional constraints on receptor binding, the general plasticity of antigenic regions, and the mutational constraint of the fusion loop.

Mutation effects on HA acid stability

We next used deep mutational scanning to measure the effects of mutations on HA's acid stability (**Fig. 2A-B, Fig. S4**, and interactive heatmaps at https://dms-vep.org/Flu_H3_Massachusetts2022_DMS/acid_stability.html). A negative effect indicates the mutation makes the HA less stable (more susceptible to inactivation at acidic pH). Again, we obtained highly correlated measurements between library replicates (r=0.9, **Fig. S2B**). Note we could only measure effects on stability of mutations that retained at least some minimal cell entry function (**Fig. S2C**). We validated the deep mutational scanning measurements of stability effects for a subset of mutations with varying effects using conditionally replicative influenza virions; the effects measured in the deep mutational scanning concord well with those measured in the validation assays using influenza virions (**Fig. S5D**). Importantly, the effects of mutations on acid stability are not strongly correlated with the effects of mutations on cell entry, indicating these assays capture distinct molecular phenotypes (**Fig. S2D**).

Most mutations that affect acid stability destabilize HA (**Fig. 2B, Fig. S4**). The mutations that most strongly affect stability tend to occur in structural regions that participate in the irreversible HA conformational changes that occur during membrane fusion (**Fig. 2C**)[22]. During this process, the HA1 protomers dilate and eventually dissociate[22,37–39]. Mutations at sites 165, 167, 205, and the 220-loop located in the trimer interface are often destabilizing, likely because they weaken interactions among neighboring trimers (**Fig. 2D**). Notably, site 165 carries a high-mannose N-linked glycan that packs against the 220-loop of a neighboring trimer[40] and all mutations to sites 165 or 167 destabilize except T167S, the one mutation that preserves the glycan (**Fig. 2E**). Positively charged arginine residues at sites 220 and 229 are thought to stabilize this region via electrostatic interactions[41], and indeed mutations at these sites destabilize HA (**Fig. S4**).

As HA1 dissociates, the HA1-HA2 interface undergoes conformational changes that are largely driven by intramonomer interactions. Consistent with prior work[39], a tetrad salt bridge involving sites 89, 109, 269 of HA1 and 396 of HA2 is important for HA stability (**Fig. 2F, Fig. S4**). Mutations at site 269 that preserved the net charge did not affect stability, but substituting an opposing charged residue was destabilizing (**Fig. 2G**, **Fig. S4**). Interestingly, charge-preserving mutations to K or H at site R109 were destabilizing, albeit less so than mutation to an oppositely charged E, suggesting the longer R109 side chain may also play a role in stability.

At acidic pH, the short α-helix, interhelical loop, and long α-helix of HA2 form an extended coil (**Fig. 2C**)[22,42]. In the prefusion conformation, these regions feature some of the largest structural differences between influenza subtypes[43]. The H3/H4/H14 clade-specific G at site 404 in the interhelical loop creates a unique sharp turn that is stabilized in part by a clade-specific S at site 107[43]. As expected, mutations at these sites and in the interacting periphery tend to destabilize HA, indicating the fragile yet essential balance governing this region (**Fig. 2H, Fig. S4**). In summary, we have produced a comprehensive high-throughput map of mutation effects on H3 acid stability; this map provides insight into the structural principles governing H3 HA stability and its conformational transitions.

Mutations exhibit phenotype-specific entrenchment

Having characterized the effects of mutations to HA on cell entry and acid stability in the background of a recent H3N2 HA, we next explored whether these effects have *changed* over the last few decades of evolution. To do this, we retraced mutations that swept to fixation in the past (**Fig. 3A**). During a sweep at a site, both the ancestral amino acid and the descendant amino acid are expected to be functionally tolerated. However, over time the descendant strains may lose tolerance for the ancestral amino acid due to other mutations that become contingent on the current amino acid, a form of epistasis called entrenchment[10,44,45].

The effects of many mutations on cell entry in the receptor binding pocket have become entrenched over time (**Fig. 3B**). For instance, a mutation at site 195 from Y to F emerged in 2020 and swept to fixation among human H3N2 strains, but the reversion F195Y in the MA22 background is highly deleterious to cell entry, indicating the mutation of site 195 from Y to F has become entrenched (**Fig. 3B**). This observation is consistent with recent work that found Y159N and T160I, which swept shortly after Y195F, are contingent on the F at site 195 to maintain receptor binding[14]. Similarly, the G186D mutation emerged in 2020 and subsequently swept to fixation; the reversion D186G in the MA22 background is highly deleterious to cell entry, indicating that G186D has become entrenched (**Fig. 3B**). This observation is also consistent with recent work that found G186D epistatically interacts with D190N, and the pair co-evolved together to preserve receptor binding[12]. We validated that conditionally replicative influenza virions encoding MA22 HA with reversions to each of these entrenched mutations (reversions D186G, S193F, F195Y, and D225N) had over 100-fold decreased titers relative to virions encoding the unmutated MA22 HA (**Fig. S5B**). Collectively, these results along with prior work[9–12,14], highlight the existence of extensive epistasis with respect to the effects of mutations within the HA receptor binding pocket on cell entry.

In contrast to the extensive epistatic entrenchment involving mutations in the receptor-binding pocket with respect to cell entry, we saw little evidence of entrenchment with respect to cell entry involving mutations in other regions of HA. Nearly all reversions to ancestral amino acids at sites outside of the receptor binding pocket are well tolerated with respect to cell entry with the single exception of T248N (**Fig. 3B**).

Strikingly, the effects of mutations on acid stability have not become entrenched in any region of HA. Reversions to ancestral amino acids at sites inside and outside of the receptor binding pocket remain well tolerated with respect to acid stability with the single exception of the mildly destabilizing reversion A163V (**Fig. 3B**). Sites where many mutations destabilize HA tend to have conserved amino-acid identities across all natural human H3N2 sequences, which further suggests constraints on acid stability may be constant across genetic backgrounds (**Fig. S6A**). At the destabilizing sites that do show variation among natural sequences, the natural mutations are exclusively the particular amino-acid changes that do not affect stability (**Fig. S6B**). For example, site 223 has exchanged hydrophobic residues valine and isoleucine during natural evolution of H3N2 HA, but (at least in MA22) such mutations have little effect on stability—and natural evolution has never sampled the many other amino-acid identities at this site that are destabilizing (**Fig. S6C**). Therefore, it appears that while epistasis commonly shifts the effects of HA mutations on cell entry to entrench mutations, such epistatic processes are much rarer with respect to the phenotype of HA acid stability.

Epistasis can alleviate the effects of antigenic mutations that impair cell entry but not stability

Phenotypes like cell entry and acid stability help determine which HA mutations are tolerated, but immune pressure largely drives positive selection for mutations in HA during the evolution of human influenza viruses. To define this immune pressure on the MA22 HA, we used deep mutational scanning to measure how HA mutations affected neutralization by human sera collected in 2023 from four children born between 2009 and 2021 (**Fig. 4A, Fig. S7A-B**, and interactive heatmaps at https://dms-vep.org/Flu_H3_Massachusetts2022_DMS/sera_neutralization.html). We used children sera since children may play an especially important role in driving influenza evolution[16,46]. The deep mutational scanning measurements are quantified such that a positive effect indicates the mutation confers viral escape from serum neutralization, whereas a negative effect indicates the mutation sensitizes the virus to neutralization by the serum. To complement these results, we also used a sequencing-based method[3,47] to measure neutralization titers for each sera against 78 H3N2 strains that either circulated in humans or were included in vaccines between 2012 and 2023.

Our deep mutational scan revealed mutations at site K140 in the MA22 HA increase sensitivity to neutralization (**Fig. 4A**, note the negative escape at site 140). Site 140 is in epitope A of HA, and changed from an I to K in 2022 (**Fig. 4B**). The fact that many mutations to site 140, including the reversion K140I, are sensitizing (**Fig. S7C**) suggests that sera contain neutralizing antibodies targeting the ancestral amino-acid identity of I140 that were escaped by the I140K mutation in 2022; the MA22 reversion K140I restores neutralization by these antibodies. We validated that conditionally replicative influenza virions with the MA22 HA carrying a K140I reversion increased sensitivity to neutralization in two of the three sera predicted to contain 140I-specific antibodies by the deep mutational scanning (**Fig. 4C**). For those sera, neutralization of historic and vaccine strains differing by up to 34 amino acids from MA22 could also be explained by site 140, though we cannot rule out the influence of other historical mutations (**Fig. S8C**). Taken together, these results indicate 140I-specific antibodies contribute substantially to the neutralizing antibody activity of sera from some individuals, and suggest that the I140K mutation that fixed in 2022 was likely selected by this immune pressure.

Given that changing site 140 from I to K causes such an appreciable reduction in antibody neutralization for some sera, why did this mutation not spread widely in human H3N2 influenza until 2022 (**Fig. 4B**)? At least part of the answer appears to be that the I140K mutation was highly deleterious for the cell entry function of HAs from older strains: for instance, I140K causes a nearly 40-fold drop in the titers of conditionally replicative influenza virions encoding the HA from the older A/Perth/16/2009 (H3N2) strain, whereas interchanging I and K at site 140 in the MA22 HA has little impact on the titers (**Fig. 4D**). These observations suggest I140K was evolutionarily inaccessible in 2009 due to its strongly negative effect on HA's cell entry function, and could not be positively selected until a permissive HA genetic background emerged. This finding parallels the recent demonstration by Lei et al.[12] that G186D and D190N,

two mutations that co-evolved together in 2020, are individually deleterious to receptor binding but together restore this function and lead to escape from human sera. Collectively, these results indicate that epistasis can alleviate pleiotropic constraints on cell entry function and facilitate antigenic evolution.

We next examined whether we could identify pleiotropic constraints on other mutations that affect serum neutralization of the MA22 HA. Mutations at a variety of HA sites reduce serum neutralization, including sites 145, 165, 189, 205, 220, and 229 (**Fig. 4A**, note positive escape at these sites). In the absence of other constraints, one might expect recent evolution to select for neutralization escape mutations at these sites. In some cases, this is occurring: S145N and K189R are single nucleotide accessible mutations from MA22 that cause sera escape in the deep mutational scanning (**Fig. 4E**) and both have been increasing in frequency among human H3N2 sequences since 2024 (**Fig. 4B**). However, some of the sites that strongly escape sera neutralization show no variation over decades of H3N2 HA evolution. For instance, mutations at sites 165, 205, 220, and 229 cause strong serum escape in the deep mutational scanning (**Fig. 4A,E)** and in validation assays with conditionally replicative influenza virions (**Fig. 4F**), but these sites have not changed during natural HA evolution (**Fig. 4B**). This disparity between the abundance of single nucleotide accessible escape mutations identified by deep mutational scanning and their absence in natural human H3N2 sequences suggests pleiotropic constraints could be constraining evolution at these sites. Notably, many of the escape mutations at these sites destabilize HA in our deep mutational scanning (**Fig. 4E**), even if they are not directly deleterious for HA-mediated cell entry. For instance, the serum escape mutations N165H, S205Y, R220T, and R229I are all roughly neutral with respect to HA-mediated cell entry (**Fig. S5B**), yet mutations at these sites strongly destabilize HA, making it more sensitive to acid inactivation (**Fig. S5D**) while conferring escape from sera in independent neutralization assays. Therefore, the effects of mutations on HA's acid stability may impose a strong pleiotropic constraint on its evolution, even when these mutations have no apparent effect on HA-mediated cell entry.

Discussion

Here, we measured the effects of all amino acid mutations to a recent H3 HA on cell entry, acid stability, and neutralization by serum antibodies. These genotype-to-phenotype maps are useful for understanding how individual mutations pleiotropically affect different HA molecular phenotypes. Interpreting these maps in the context of HA's natural evolution reveals that recent H3N2 evolution has alleviated constraints on cell entry through epistasis, but many antigenic mutations remain highly constrained because they pleiotropically decrease acid stability.

Our work shows that epistatic entrenchment is common with respect to the impact of mutations on HA's cell entry function, but not on its acid stability. One possible explanation is that acid stability reflects a single underlying biophysical property that is less influenced by epistasis, whereas cell entry is a higher order phenotype that involves several underlying properties (e.g., protein stability, receptor-binding, and membrane fusion). There is evidence for other proteins

that mutations often have additive effects on underlying biophysical properties, and epistasis in higher-order phenotypes often arises simply from their non-linear dependence on underlying molecular properties[26,48–50]. Additionally, receptor binding involves the coordinated action of many amino-acid side chains in a manner that may be especially conducive to epistasis[9–14].

Our deep mutational scanning shows that there are sites outside classically defined antigenic regions where mutations strongly escape serum antibody neutralization but also destabilize HA (e.g. sites 165, 205, 220, and 229). Interestingly, these sites are highly conserved and located in the trimer interface. While non-neutralizing anti-H3 and neutralizing anti-H7 antibodies targeting this region have been described[51–54], to our knowledge, humans have not been shown to possess neutralizing anti-H3 trimer interface antibodies. While it is unclear how these mutations are escaping serum antibody neutralization, we speculate there are two possible mechanisms. These mutations could abrogate antibody binding, either by directly affecting the binding site or altering nearby epitopes, or the destabilizing effect could accelerate membrane fusion, a previously reported strategy for antibody escape[55]. In any case, our work suggests antibodies targeting these sites may be particularly difficult to escape due to strong pleiotropic constraints on destabilizing mutations during HA evolution.

Overall, our study highlights how mutations to HA often have pleiotropic effects on cell entry, acid stability, and antibody neutralization. The extent to which pleiotropic constraints can be alleviated by epistasis differs across phenotypes: for example, constraints on cell entry appear more readily alleviated than constraints on acid stability. A deeper understanding of how pleiotropy and epistasis shape HA evolution will be useful for forecasting viral evolution and designing therapeutics that are resistant to viral escape.

Acknowledgements

This work was supported in part by the NIH/NIGMS CMB Training Grant (T32 GM007270) to TCY, NSF Graduate Research Fellowship (DGE-2140004) to TCY, NIH/NIAID under award R01AI165821 to JDB, and NIH/NIAID under contract 75N93021C00015 to JDB. JDB is an investigator of the Howard Hughes Medical Institute. This research was also supported by the Genomics & Bioinformatics Shared Resource (RRID: SCR_022606), the Flow Cytometry Shared Resource (RRID: SCR_022613) of the Fred Hutch/University of Washington/Seattle Children's Cancer Consortium (P30 CA015704), and by Fred Hutch Scientific Computing, NIH grants S10-OD-020069 and S10-OD-028685. We thank the Bedford lab at Fred Hutch for maintaining Nextstrain builds, John Huddleston for technical assistance, and Brendan Larsen for help with data visualization.

Competing interests

JDB consults on topics related to viral evolution for Apriori Bio, Invivyd, the Vaccine Company, Pfizer, and Moderna. JDB, BD, and ANL are inventors on Fred Hutch licensed patents related to viral deep mutational scanning.

Author contributions

TCY and JDB conceived the study. TCY and CK collected the data. TCY and JDB analyzed the data. TCY and JDB wrote the paper. All authors edited and approved the paper.

Main figures

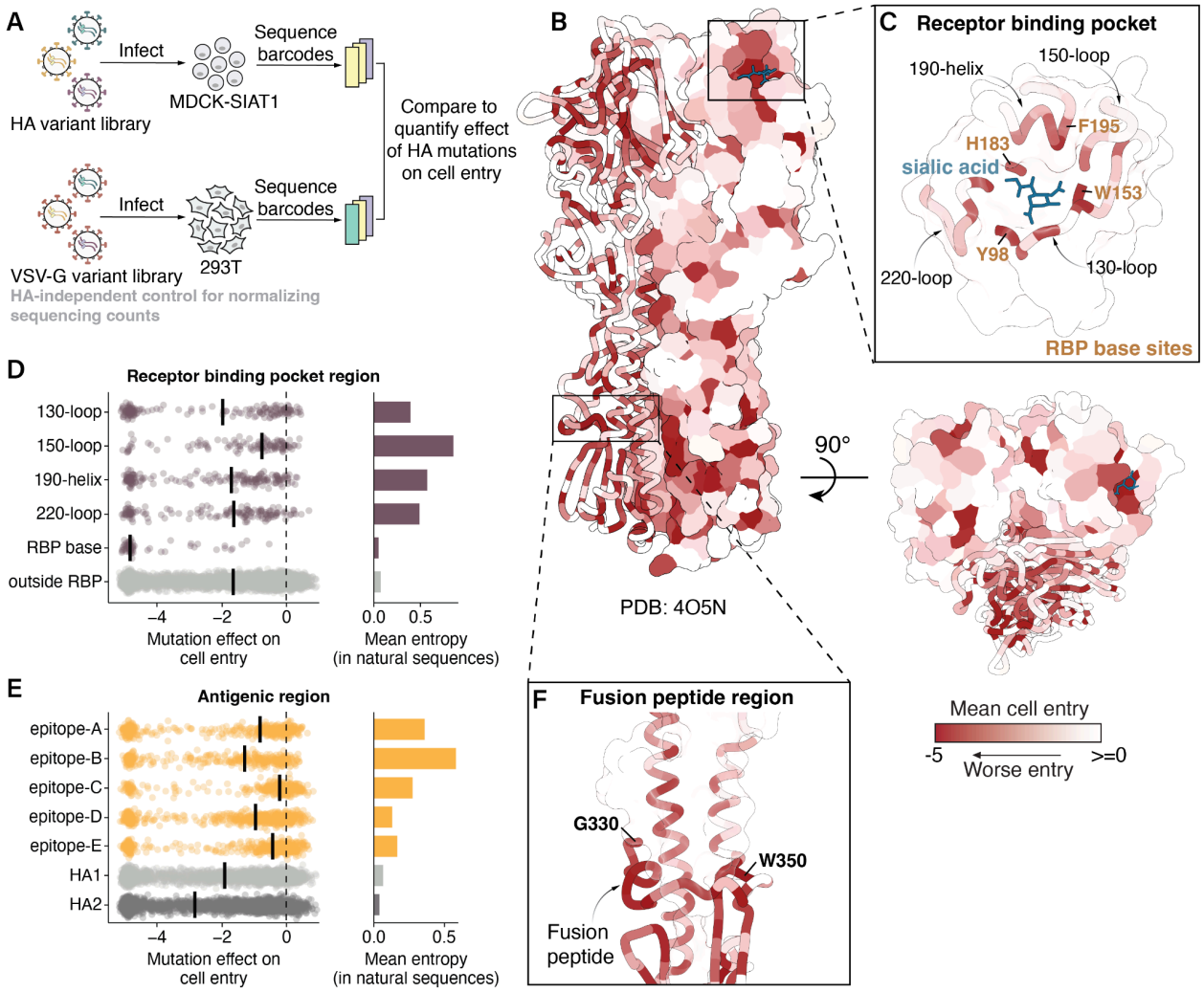


Figure 1 | Mutation effects on HA-mediated cell entry.

A) To measure how all HA mutations affect cell entry, we create libraries of barcoded pseudoviruses expressing different HA mutants on their surface (**Fig. S1**). We use deep sequencing to quantify the ability of each HA mutant to enter cells, normalizing the sequencing counts to a copy of the pseudovirus library where all virions express VSV-G and so do not rely on HA for cell entry. The effect of each mutation is quantified as the $\log_2$ of its frequency relative to unmutated variants in the HA condition relative to the VSV-G condition, so negative values indicate impaired cell entry. **B)** Mean effect of mutations at each site on cell entry mapped onto the HA structure (Protein Data Bank 4O5N) viewed from the side or top, with darker red indicating worse cell entry. See **Fig. S3** for a heatmap of all mutation effects. **C)** Zoomed in version of the structure showing the receptor binding pocket. Distribution of mutation effects on cell entry in **D)** receptor binding pocket regions and **E)** antigenic regions, with the median effect

in each region indicated with a solid black line. The mean Shannon entropy of all sites in each region across a subsampled tree of natural human H3N2 evolution since 1968 is shown on the right. **F)** Zoomed in version of structure showing sites that make up the fusion peptide and periphery.

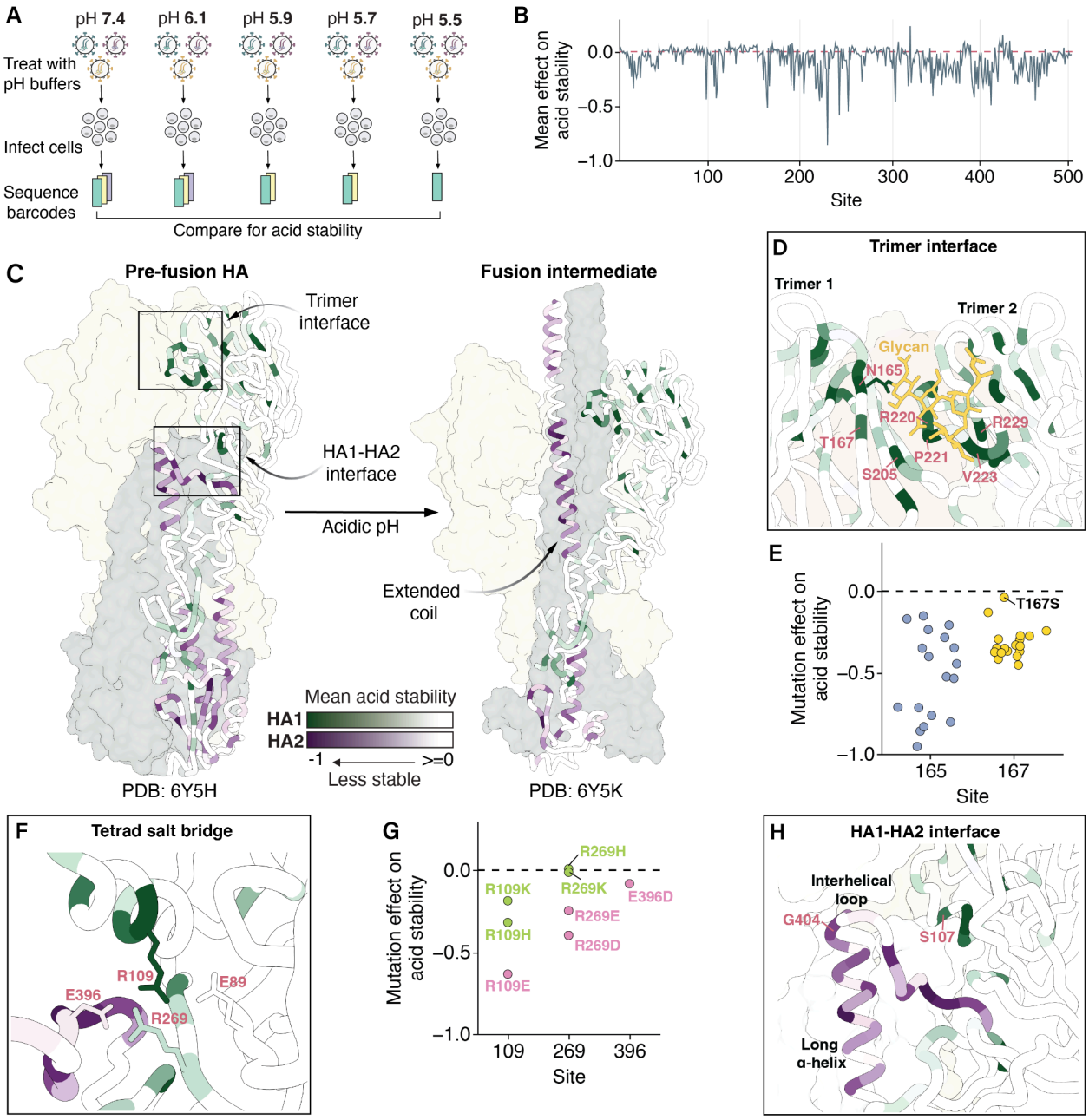


Figure 2 | Mutation effects on HA acid stability.

A) We incubated pseudovirus HA variants in different acidic pH buffers prior to infection and sequenced the pseudovirus barcodes within cells after infection. To quantify mutation effects on acid stability, we compared these barcode counts to those of pseudoviruses treated with neutral pH media, which served as an infection baseline. **B)** Mean effect of mutations at each site on acid stability. Negative values indicate sites where mutations decrease stability (e.g., lead to viral inactivation at a higher pH). See **Fig. S4** for a heatmap of all mutation effects. **C)** Mean effect of mutations at each site on acid stability mapped onto the structures of pre-fusion

(Protein Data Bank 6Y5H) and fusion intermediate (Protein Data Bank 6Y5K) HAs, with darker shades of green and purple indicating greater destabilizing mutation effects in HA1 and HA2, respectively. **D)** Zoomed in view of the trimer interface, with the same color scale used in C. **E)** Mutation effects on acid stability at sites 165 and 167. **F)** Zoomed in view of sites that participate in a tetrad salt bridge, with the same color scale used in C. **G)** Mutation effects of charged amino acids at sites that participate in the tetrad salt bridge. **H)** Zoomed in view of the HA1-HA2 interface, with the same color scale used in C.

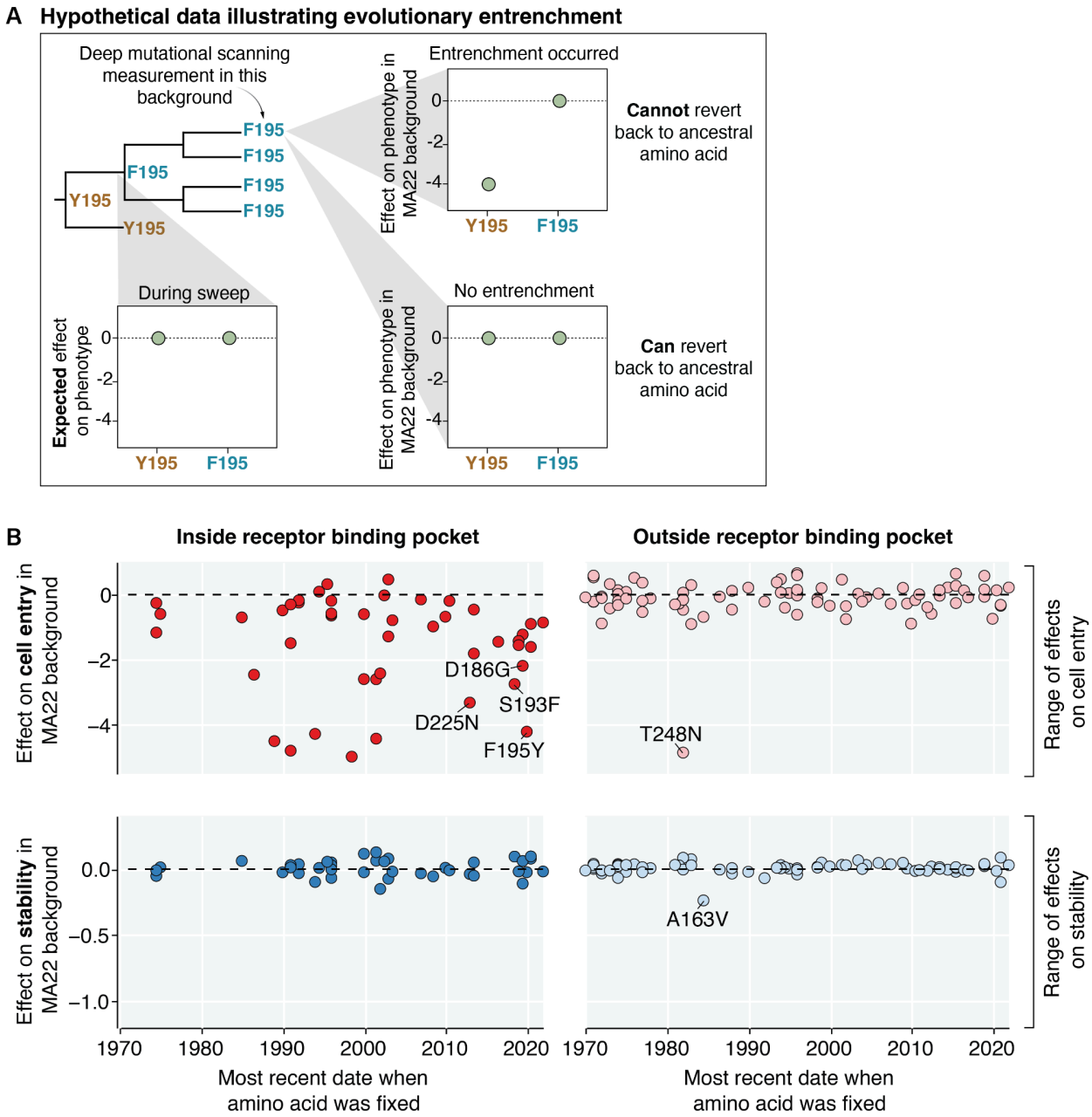


Figure 3 | Mutations exhibit phenotype-specific epistatic entrenchment.
A) Hypothetical data illustrating entrenchment of a mutation via epistasis. The phylogenetic tree shows a sweep of F replacing Y at site 195. During the sweep, both F and Y are expected to be tolerated. However, the effect of reverting to the ancestral amino acid (Y) may change over time due to entrenchment. The HA used for deep mutational scanning contains the current amino acid (F) at site 195. If entrenchment occurred, then reverting to the ancestral amino-acid Y195 will be deleterious in the recent genetic background used for deep mutational scanning. In the absence of entrenchment, Y195 remains tolerated in newer genetic backgrounds. **B**) Actual experimental data showing the effects of reversions to all ancestral amino acids that were

previously fixed in human H3N2 strains since 1968 on HA-mediated cell entry (top row) or HA acid stability (bottom row) as measured in the deep mutational scanning on the MA22 HA. Mutations are placed on the x-axis by the most recent date that the ancestral amino acid was fixed, and the panel columns indicate whether the mutation is at a site inside (left column) or outside (right column) of the receptor binding pocket. The range of the y-axis for each phenotype is set to span the range of effects of all mutations to HA (not just those that fixed during natural H3N2 evolution) in the deep mutational scanning. There is extensive epistatic entrenchment of mutations in the receptor binding pocket with respect to cell entry, but no substantial entrenchment of mutations with respect to stability. To mouseover the individual mutations, see https://dms-vep.org/Flu_H3_Massachusetts2022_DMS/entrenchment.html for an interactive version of this plot.

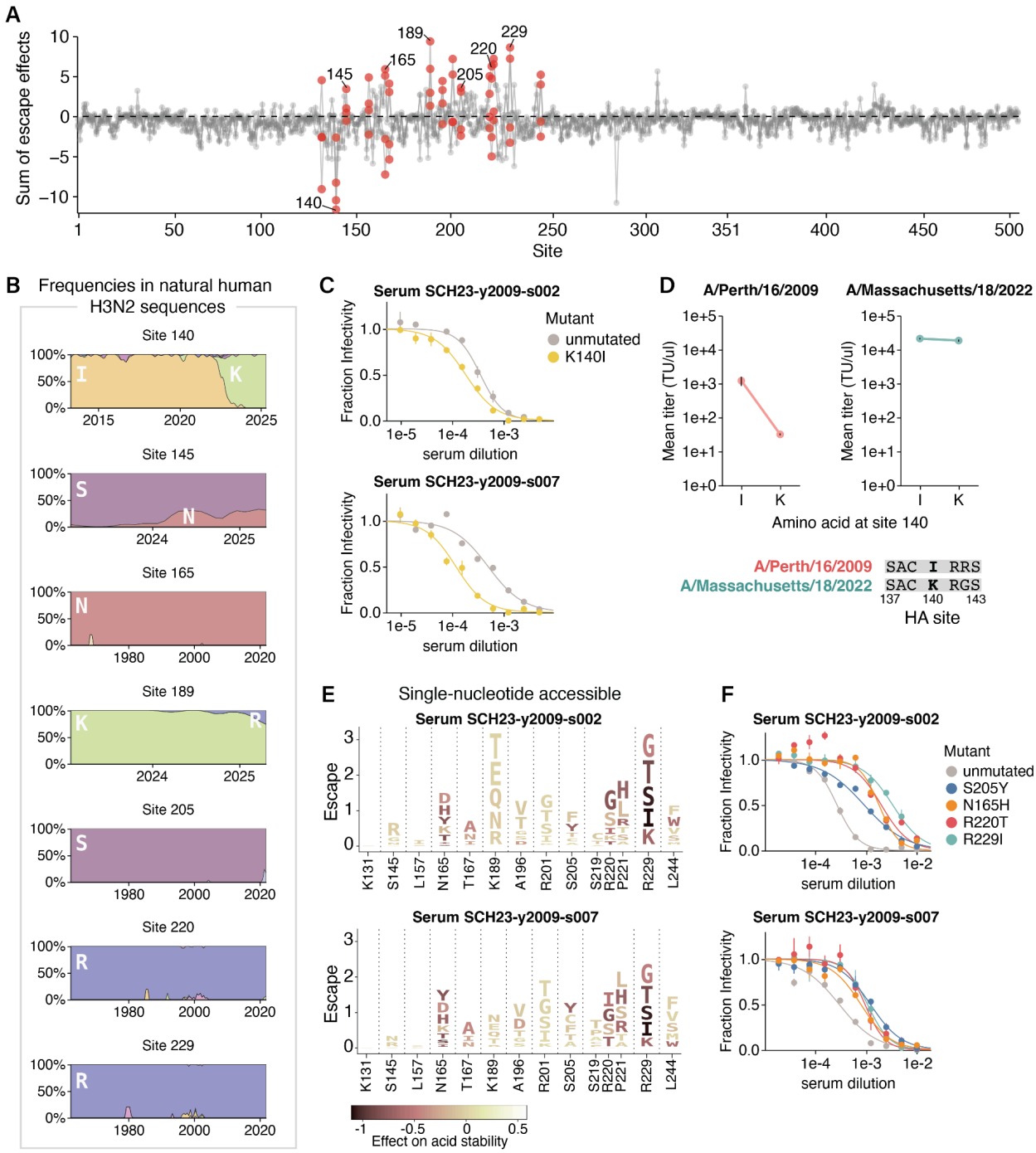
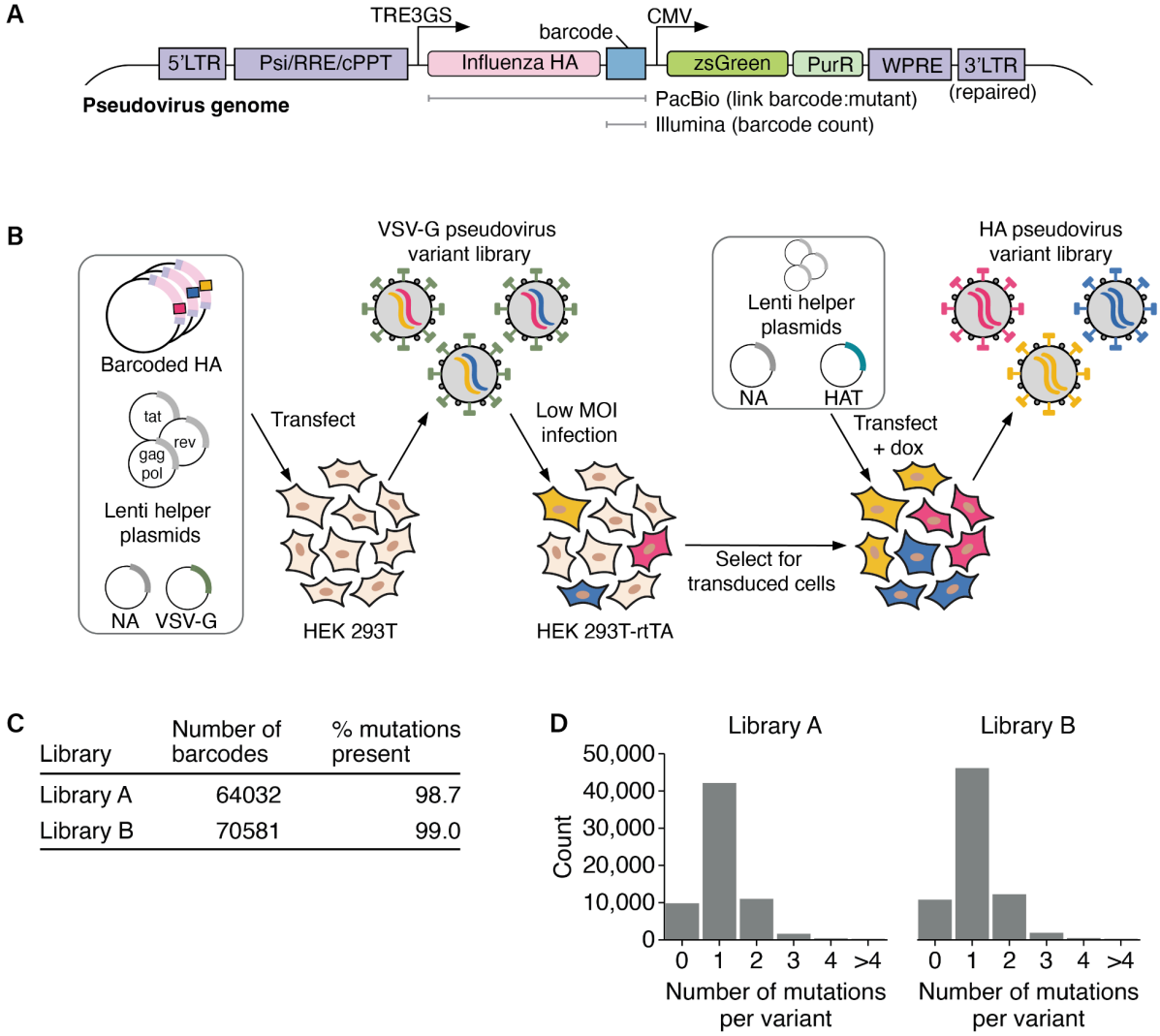


Figure 4 | During H3N2 evolution, epistasis alleviates the effects of antigenic mutations that impair cell entry but not stability.

A) Sum of mutation effects at each site in the MA22 HA on escape from four human sera collected in 2023. Key sites of escape or sensitization included in logoplots are colored red, and sites where escape mutants were validated in independent neutralization assays are labeled with text. See **Fig. S7** for data for individual sera. **B)** Frequencies of amino acids observed in

natural human H3N2 HA sequences over time (x-axis indicates year) for several key sites. Note the x-axis range varies as some subplots show changes in recent evolution, while others show conservation across decades of evolution. **C)** Neutralization by two human sera of conditionally replicative influenza virions with MA22 HA with or without the mutation K140I. **D)** Titers of conditionally replicative influenza virions with HA from either A/Perth/16/2009 or MA22 with either an I or a K amino acid at site 140. **E)** Logoplots displaying amino-acid mutations to the MA22 HA that are accessible by single nucleotide changes. The height of each letter is proportional to the escape from the indicated serum as measured by deep mutational scanning. Each mutation is colored by its effect on HA acid stability as measured in the deep mutational scanning, with darker colors indicating decreased stability. **F)** Neutralization of conditionally replicative MA22 virus with destabilizing mutations by the two sera in E.

Supplemental figures



C

Library	Number of barcodes	% mutations present
Library A	64032	98.7
Library B	70581	99.0

Figure S1 | Pseudovirus deep mutational scanning of influenza hemagglutinin.

A) Diagram of the lentiviral genome used to produce genotype-phenotype-linked pseudovirus libraries for deep mutational scanning. The genome is flanked by long terminal repeat (LTR) sequences, with the typical 3′ LTR deletion repaired so the lentiviral genome can be transcribed after integration. A zsGreen reporter and a puromycin resistance marker are constitutively expressed from a CMV promoter. Expression of the HA gene is regulated by a doxycycline inducible TRE3GS promoter. PacBio sequencing is performed to map each barcode to an HA mutant. Then, effects of HA mutations can be quantified by Illumina sequencing the barcodes. **B)** Schematic of the "two-step" method for generating genotype-phenotype-linked pseudovirus libraries described in Dadonaite et al.[23,24]. In the first step, a plasmid library encoding the lentiviral genomes with the HA mutants is co-transfected into HEK293T cells alongside three lentiviral helper plasmids (tat, rev, gagpol), a plasmid expressing a strain-matched neuraminidase (NA), and a plasmid expressing the glycoprotein from vesicular stomatitis virus

(VSV-G). This results in pseudoviruses that encode HA mutants within their genomes but express VSV-G and NA on their surfaces. The NA ensures HA expression does not prevent virions from detaching from producing cells. These VSV-G pseudotyped viruses are transduced into a HEK293T-rtTA cell line at low MOI to ensure most infected cells integrate a single lentiviral genome, and puromycin is used to select for integrated cells. In the second step, helper plasmids, a plasmid expressing NA, and a plasmid expressing the HA-activating human airway trypsin-like (HAT) protease are co-transfected into the integrated cells. Doxycycline is added at this step to induce HA expression. This results in genotype-phenotype-linked HA-pseudotyped pseudoviruses. These pseudoviruses can undergo a single round of cell entry, but are not fully infectious agents as they do not encode the genes needed to undergo multiple rounds of replication. **C)** Number of barcodes and mutation coverage in the two pseudovirus library replicates. In this table, "% mutations present" indicates the percentage of all HA ectodomain amino-acid mutations found in at least one of the barcoded variants. **D)** Distribution of the number of HA amino-acid mutations per variant in the two pseudovirus library replicates. Most variants contain a single mutation.

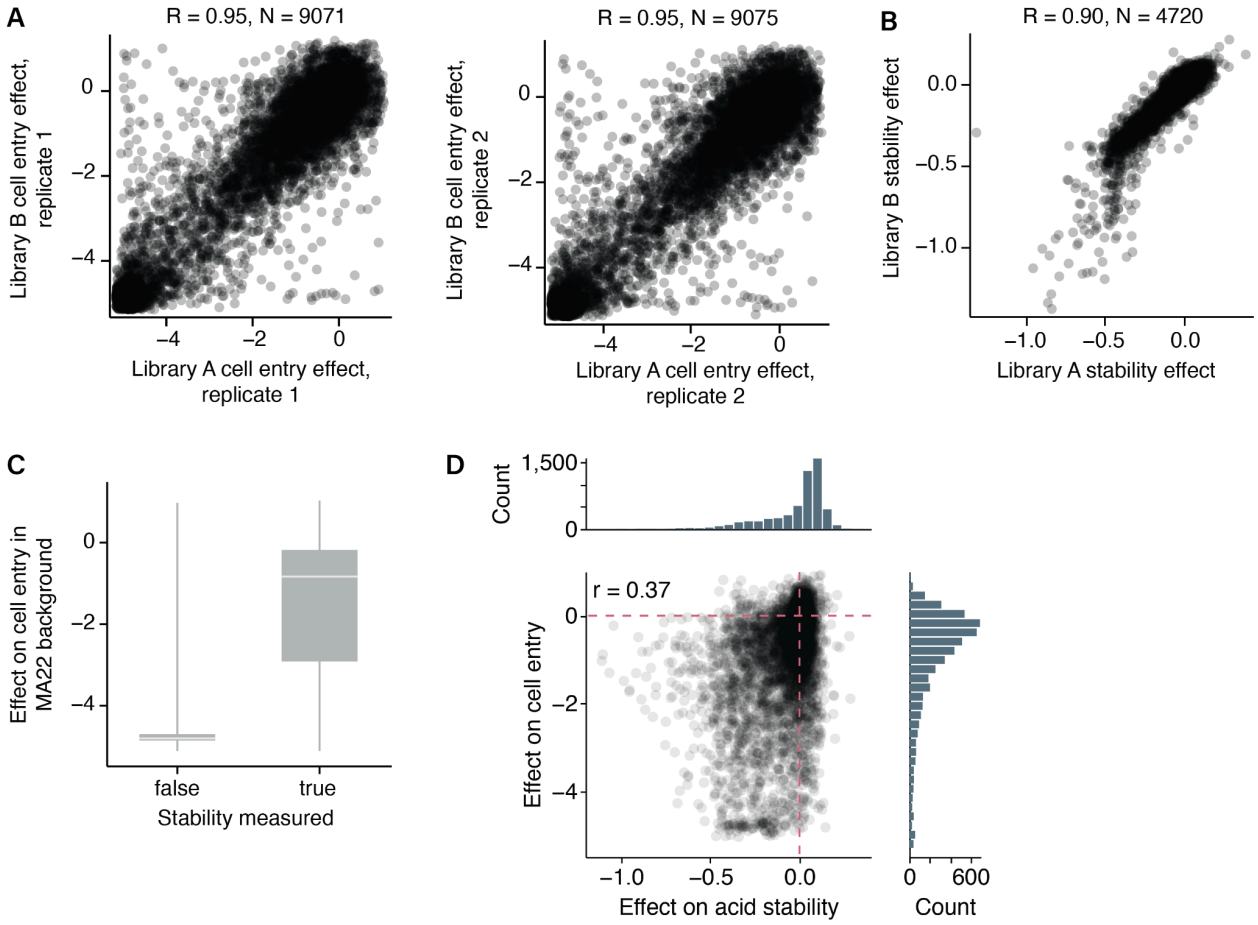


Figure S2 | Correlations of mutation effects measured by deep mutational scanning.

A) Correlation of the effects of HA mutations on cell entry between the two pseudovirus library replicates. Each point represents the effect of a different mutation as measured in each replicate. Two technical replicates were performed for each of the two libraries, and the two panels show correlations between the two independent libraries for each technical replicate. Throughout this paper, we report the median effect of mutations across the four replicates. **B)** Correlation of the effects of HA mutations on acid stability between the two pseudovirus library replicates. Note that there are fewer mutations with measured effects on stability because we can only measure stability for mutations with at least some cell entry. **C)** Most mutations for which it was not possible to make measurements of acid stability correspondingly have very poor cell entry. Shown is the distribution of effects on cell entry of mutations for which it was or was not possible to measure the effect on stability. **D)** Correlation between effects of mutations on cell entry and effects of mutations on acid stability. These effects of mutations on these two phenotypes are only weakly correlated.

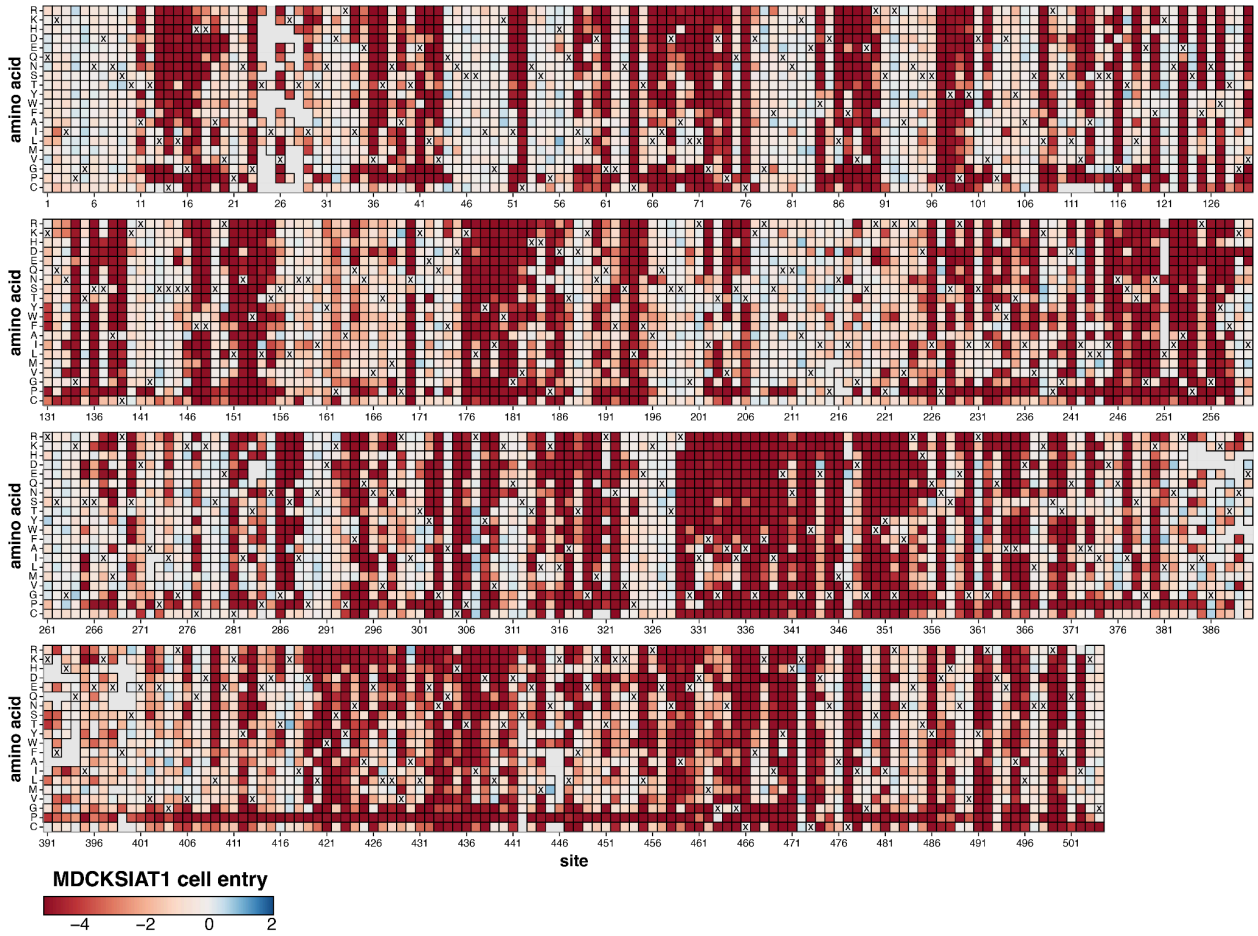


Figure S3 | Mutation effects on HA-mediated cell entry.

Each tile represents a mutation at an HA site, colored by the effect of that mutation on entry into MDCK-SIAT1 cells. Red indicates impaired entry, white indicates no effect, and blue indicates improved entry. To visualize these mutation effects in the context of the HA structure, see **Fig. 1B**. Tiles with an 'X' denote the amino acid identity in the unmutated MA22 strain. Empty gray tiles indicate mutations that were either missing from the library or lacked a reliable cell entry measurement. See https://dms-vep.org/Flu_H3_Massachusetts2022_DMS/cell_entry.html for an interactive version of this heatmap.

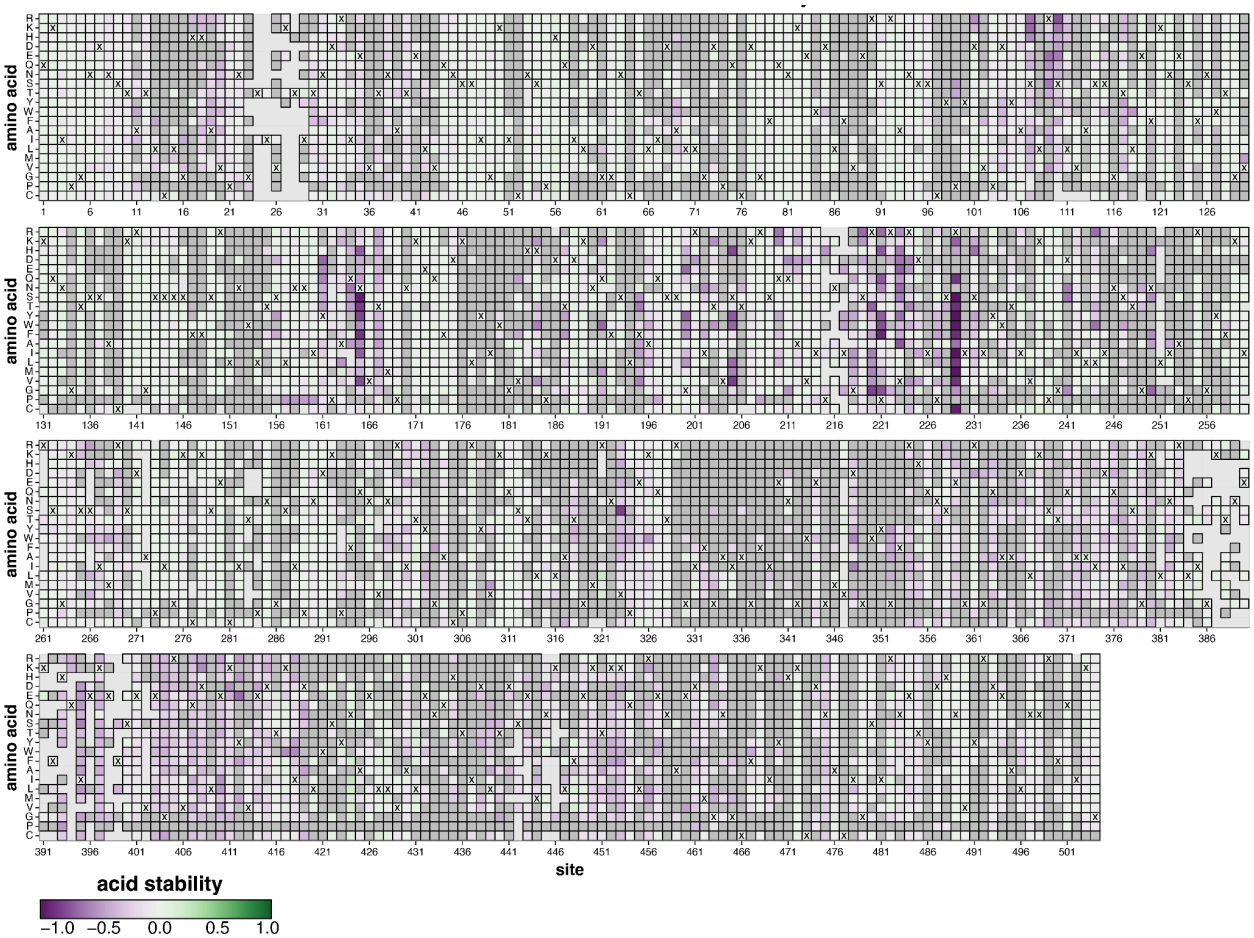


Figure S4 | Mutation effects on HA acid stability.

Each tile represents a mutation at an HA site, colored by the effect of that mutation on HA acid stability. Purple indicates decrease stability, white indicates no effect, and green indicates increased stability. To visualize these mutation effects in the context of the HA structure, see **Fig. 2C**. Tiles with an 'X' denote the amino acid identity in the unmutated MA22 strain. Dark gray tiles indicate mutations that are too deleterious for cell entry to reliably measure their effect on acid stability, while light gray tiles indicate mutations that were missing (not measured) in the library. See https://dms-vep.org/Flu_H3_Massachusetts2022_DMS/acid_stability.html for an interactive version of this heatmap.

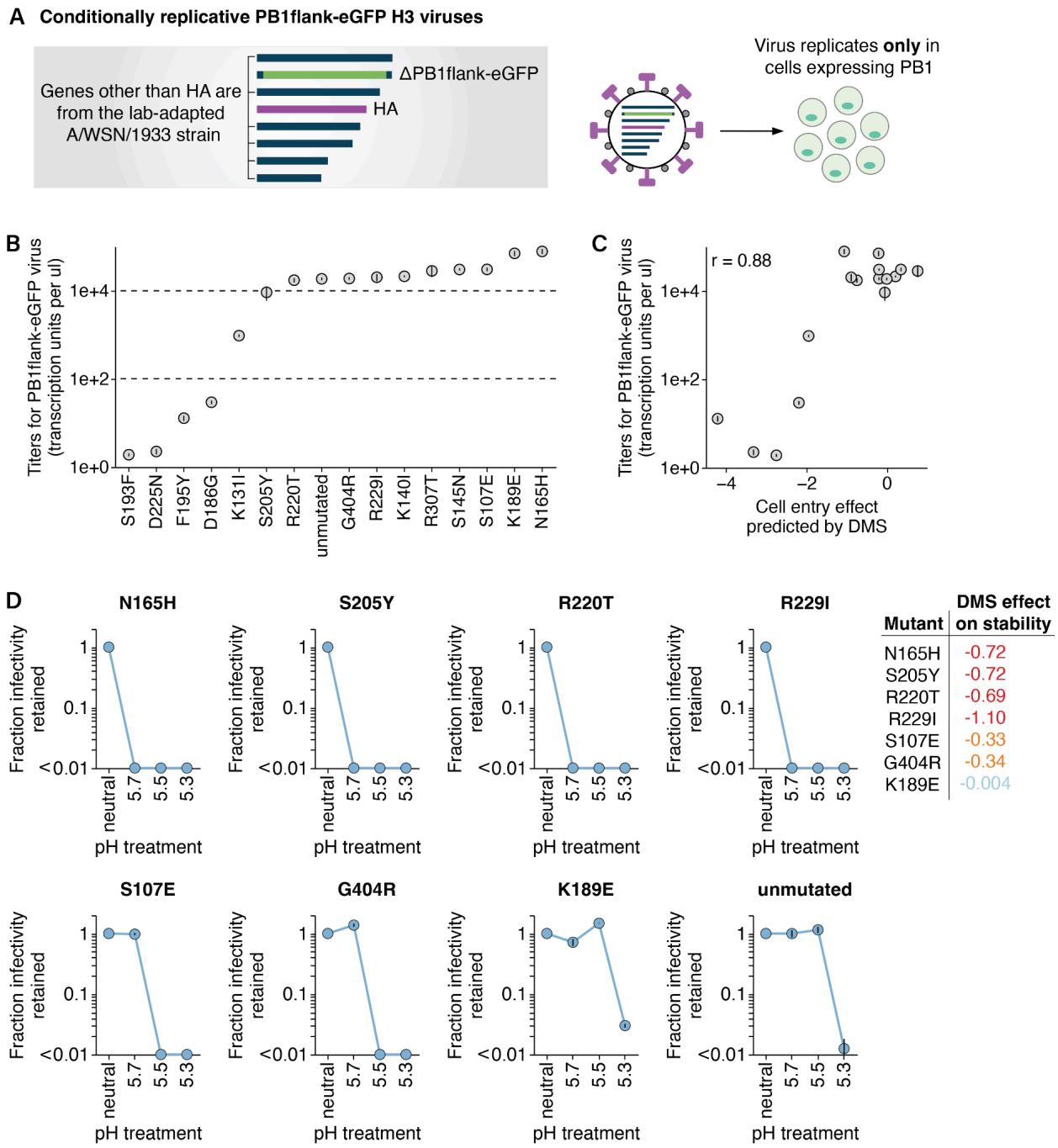


Figure S5 | Validation of mutation effects on cell entry and acid stability with conditionally replicative influenza virions.

A) Diagram of a conditionally replicative PB1flank-eGFP influenza virus genome. The PB1 gene is replaced with an eGFP and the remaining segments besides HA are derived from a lab-adapted A/WSN/1933 strain. These viruses can be rescued by reverse genetics, can only replicate in cells that express PB1, and are safe to use at biosafety-level 2. **B)** Titers of

conditionally replicative virions carrying single amino acid mutations in the MA22 HA that were present in the supernatant after virus production by reverse genetics. Each point in the plot is the mean of four titer measurements, two technical replicates from the same virion rescue stock and two biological replicate stocks rescued from independent plasmid preparations. The error bars indicate mean ± 1 standard error. **C)** Correlation between the titers of conditionally replicative virions carrying single amino acid mutations to the MA22 HA and the effects of those mutations on cell entry measured by deep mutational scanning. **D)** The fraction infectivity retained after treating conditionally replicative virions with the indicated MA22 HA mutations with either neutral media or acidic pH buffers. Single replicates were performed across two different days and averaged. The error bars indicate mean ± 1 standard error. The fractions are normalized to the infectivity in the neutral condition. The effect of each mutant on acid stability measured by deep mutational scanning is included on the right, and generally tracks with the pH sensitivity measured in the validation assay.

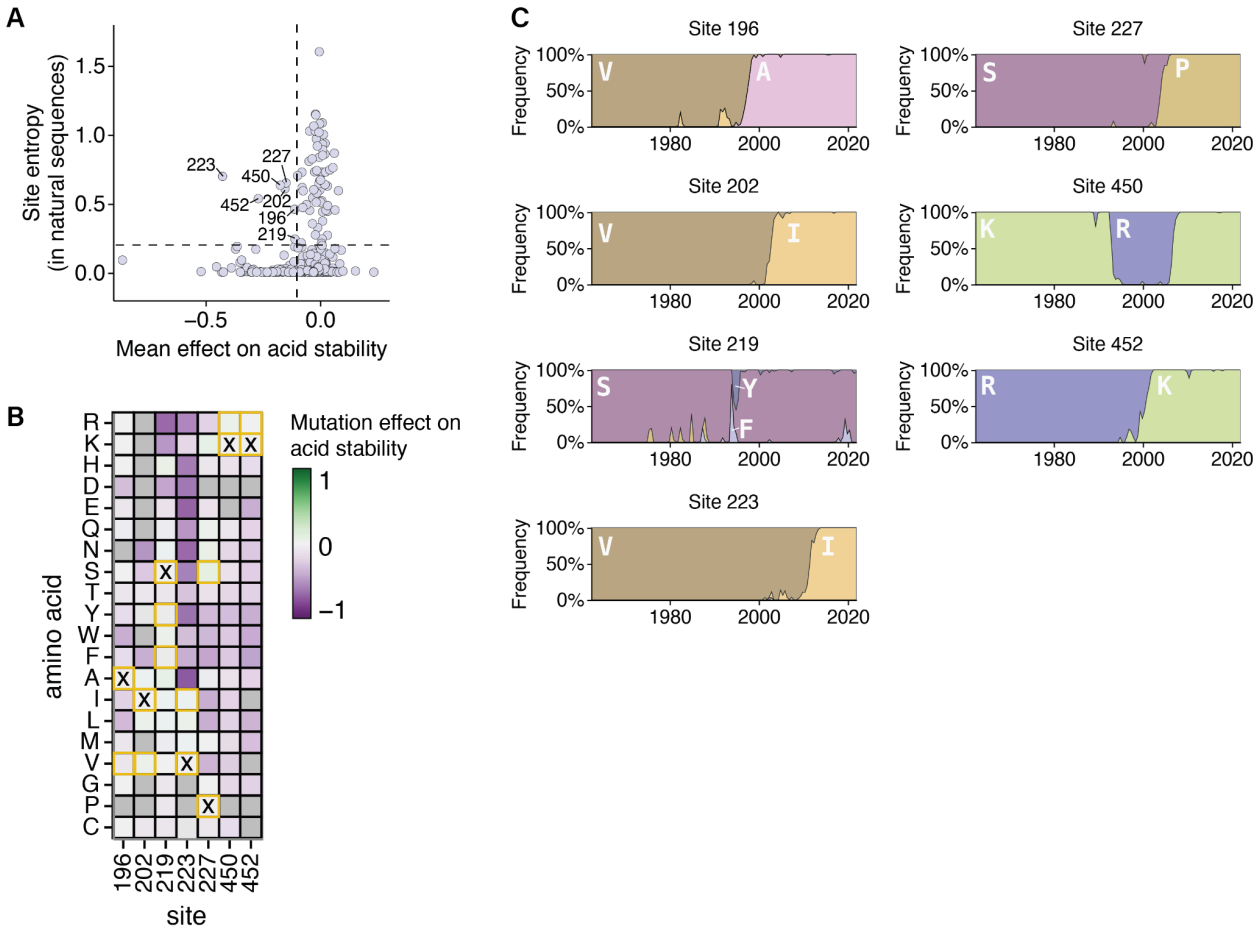


Figure S6 | Conservation at HA sites with destabilizing mutations.

A) Correlation between the Shannon entropy at each HA site in an alignment from a subsampled tree of natural human H3N2 evolution since 1968 and the mean effect on acid stability of all mutations at each HA site as measured by deep mutational scanning. Most sites with many mutations that destabilize HA are strongly conserved, with exceptions labeled. **B)** Heatmap of all mutation effects on acid stability at sites labeled in A. The 'X' indicates the amino acid in the unmutated MA22 HA, and squares boxed in yellow indicate amino acids that increased in frequency during natural evolution. At sites with many destabilizing mutations, natural evolution exclusively samples only the (relatively rare) amino-acid identities that do not destabilize HA. **C)** Frequencies over time (the x-axis indicates year) of amino acids observed in natural human H3N2 sequences at the sites labeled in A.

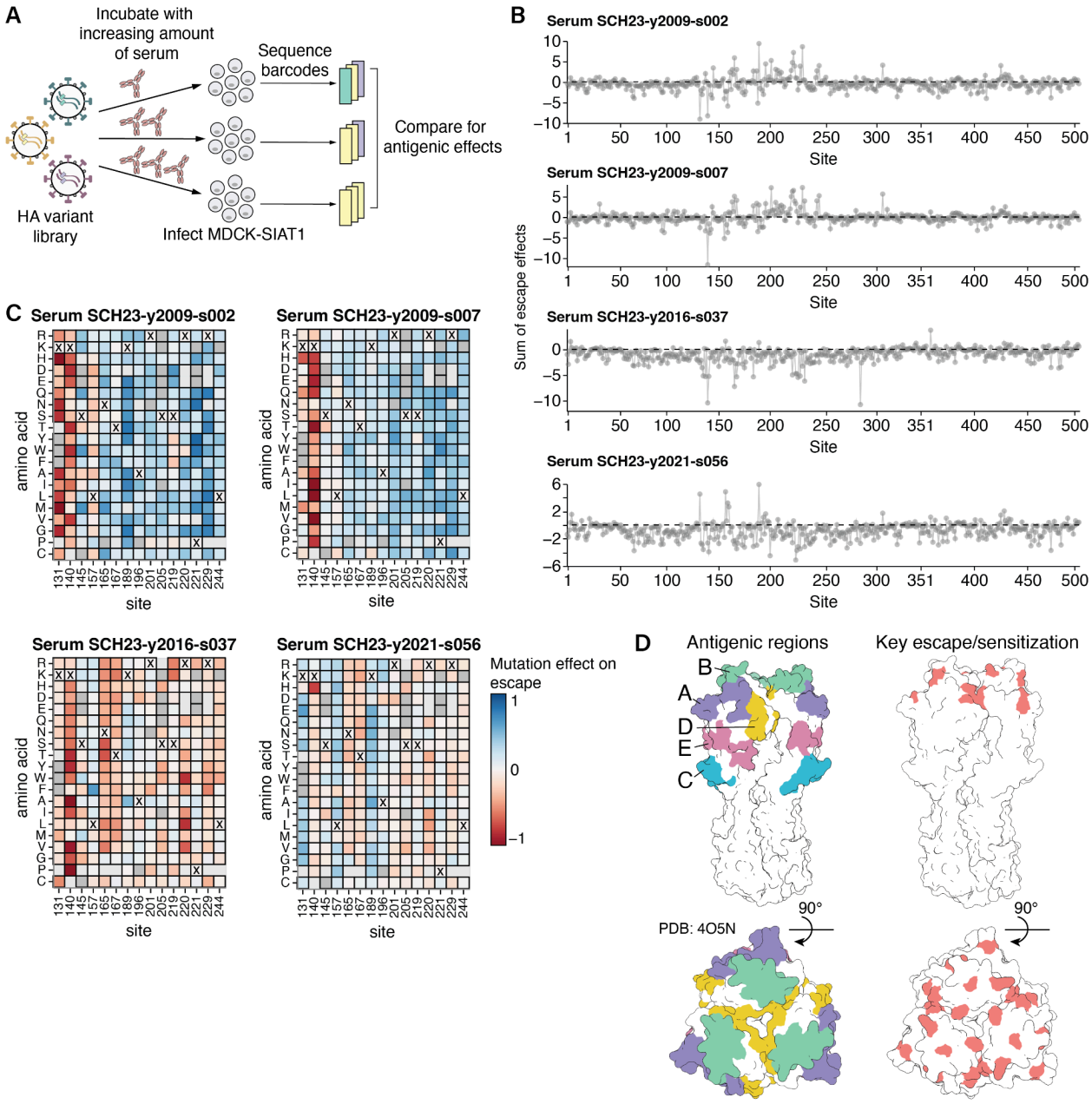


Figure S7 | Mapping the effects of HA mutations on neutralization by human sera using deep mutational scanning.

A) We incubated the pseudovirus HA library with increasing concentrations of serum, then infected MDCK-SIAT1 cells and sequenced the barcodes of pseudoviruses that were still able to enter cells after serum treatment. To quantify mutation effects on antigenicity, we compared these barcode counts to those of pseudoviruses that were not incubated with serum, which served as an infection baseline; a neutralization standard is used to convert these counts into fraction infectivity at each serum concentration (**Methods**)[23,24]. **B)** The sum of mutation effects on escape at each site in HA for four human sera. **Fig. 4A** shows these four plots overlaid. See

https://dms-vep.org/Flu_H3_Massachusetts2022_DMS/sera_neutralization.html for a version of these lineplots that is interactive, along with interactive heatmaps that show how individual mutations affect sera neutralization. **C)** Mutation-level escape and sensitization at key sites that are highlighted in **Fig. 4A** for the four sera. The X's indicate the amino acid in the unmutated MA22 strain. Tiles that are more blue indicate mutations that escape sera, while tiles that are more red indicate mutations that have a sensitizing effect. **D)** HA structures (Protein Data Bank 4O5N) showing antigenic regions and locations of key sites of escape or sensitization that are highlighted in **Fig. 4A**.

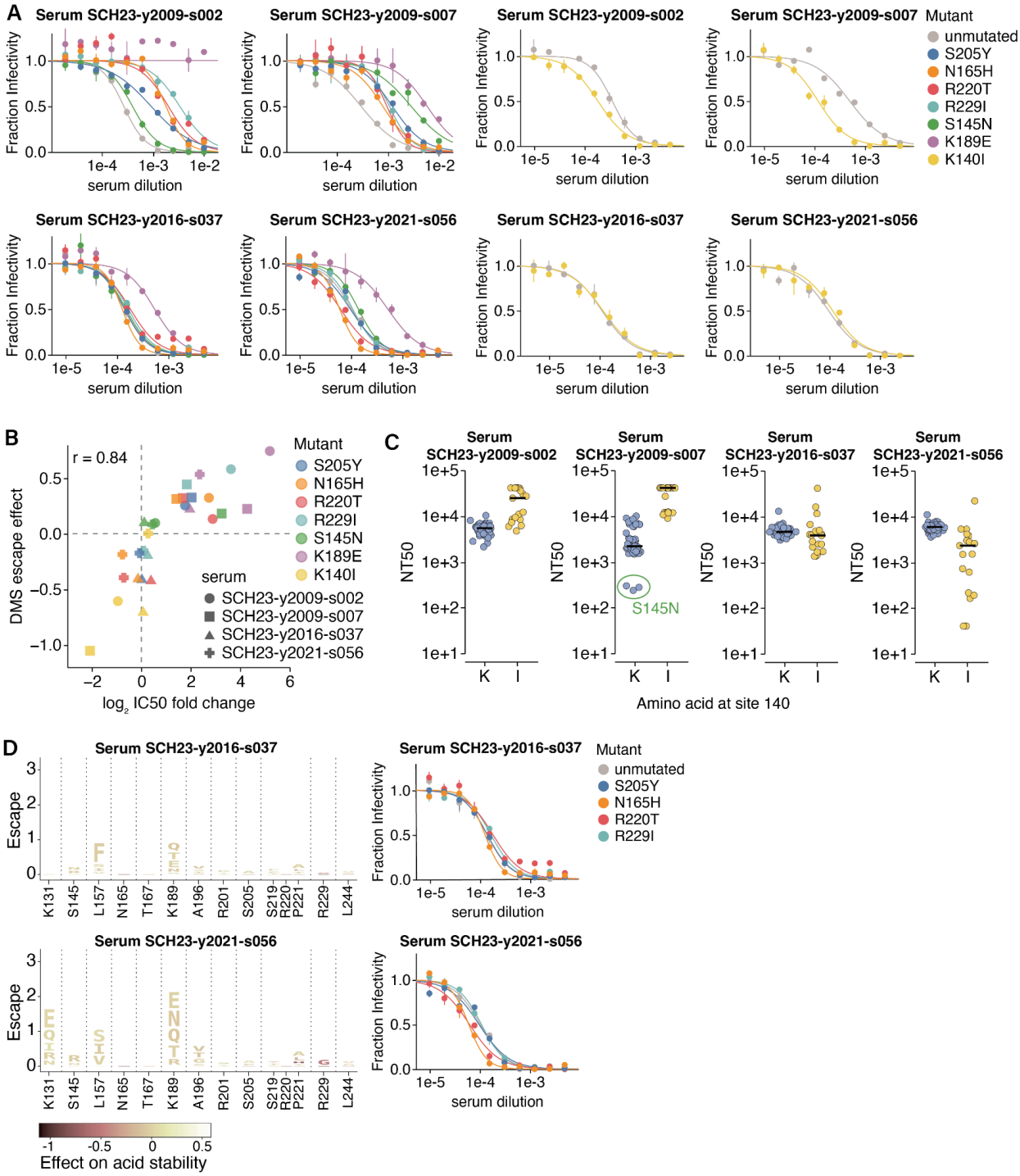


Figure S8 | Validation of mutation effects on serum neutralization with conditionally replicative influenza virions.

A) Neutralization of conditionally replicative influenza virions with the MA22 HA carrying the indicated mutations (S205Y, N165H, R220T, R229I, S145N, K189E, and K140I) by four human sera collected in 2023. A subset of these curves are shown in **Fig. 4C**, **Fig. 4F**, and **Fig. S8D**. **B)** Correlation between the antigenic effect of mutations measured by deep mutational scanning

and the change in IC50 measured by the independent neutralization assay in A. **C)** Neutralization for 78 vaccine or circulating H3N2 strains between 2012 and 2023 by the four sera, as measured in Kikawa et al.[3]. The points are stratified by whether or not the strain includes a K or I at site 140. The black line indicates the median NT50. Two of the sera have higher titers against strains with 140I, consistent with K140I being a sensitizing mutation for these sera. There are three recent strains that contain S145N that escape SCH23-y2009-s007, consistent with this being an escape mutation from this sera. **D)** Logoplots displaying single nucleotide accessible mutations from MA22 HA with positive escape at the key sites highlighted red in **Fig. 4A** for two sera. The height of each letter is proportional to the escape from the indicated serum as measured by deep mutational scanning. Each mutation is colored by its effect on HA acid stability as measured in the deep mutational scanning, with darker colors indicating decreased stability. The logoplots and neutralization curves for the other two sera mapped by deep mutational scanning are shown in **Fig. 4E** and **Fig. 4F**.

Supplemental tables

Table S1 | Sites within receptor binding pocket and antigenic regions.

Receptor binding pocket region	Sites	Reference
130-loop	128, 130, 131, 133-138	[10,12,14,56,57]
150-loop	155-160	[10,12,14,56,57]
190-helix	186, 189-194, 196-198	[10,12,14,56,57]
220-loop	221-228	[10,12,14,56,57]
Base	98, 153, 183, 195	[10,12,14,56,57]
Other	145	[58]
Antigenic region		
Epitope A	122-146	[33]
Epitope B	155-160, 186-198	[33]
Epitope C	44-54, 273-280	[33]
Epitope D	166-181, 201-219	[33]
Epitope E	62-65, 78-94, 260-265	[33]

Methods

Code and data availability

See https://dms-vep.org/Flu_H3_Massachusetts2022_DMS for a collection of interactive visualizations. Code for reproducing the analysis is at https://github.com/dms-vep/Flu_H3_Massachusetts2022_DMS and output of the analysis is at https://dms-vep.org/Flu_H3_Massachusetts2022_DMS/appendix.html.

Data that have been pre-filtered for quality control criteria are available in CSV format at these links:

- Mutation effects on cell entry and acid stability: https://github.com/dms-vep/Flu_H3_Massachusetts2022_DMS/blob/main/results/summaries/Phenotypes.csv
- Mutation effects on sera neutralization: https://github.com/dms-vep/Flu_H3_Massachusetts2022_DMS/blob/main/results/summaries/Phenotypes_per_antibody_escape.csv

Cell lines and media

The following cell lines were used: 293T (ATCC, CRL-3216), 293T-rtTA (from Dadonaite et al.[23]), 293T-CMV-PB1 (from Bloom et al.[59]), MDCK-SIAT1 (HPA Cultures, 05071502), MDCK-SIAT1-CMV-PB1 (from Bloom et al.[59]), and MDCK-SIAT1-CMV-PB1-TMPRSS2 (from Lee et al.[19]).

All cell lines were maintained in D10 media (Dulbecco's Modified Eagle Medium supplemented 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. For virus rescue and infection with HA expressing pseudovirus libraries and conditionally replicative influenza viruses, we used Influenza Growth Media (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) or Neutralization Assay Media (NAM, Medium-199 supplemented with 0.01% heat-inactivated fetal bovine serum, 0.3% bovine serum albumin, 100 µg/mL calcium chloride, 100 U/mL penicillin, 100 µg/mL streptomycin, and 25 mM HEPES).

Plasmids and primers

Plasmid maps can be found at: https://github.com/dms-vep/Flu_H3_Massachusetts2022_DMS/tree/main/data/supplemental_data/plasmids
Primer sequences can be found at: https://github.com/dms-vep/Flu_H3_Massachusetts2022_DMS/tree/main/data/supplemental_data/primers

Human sera

Human sera samples were obtained from Seattle Children's Hospital during routine blood draws from children receiving medical care in December 2023. This was approved by the Seattle Children's Hospital Institutional Review Board with a waiver of consent. Sera samples were treated with receptor-destroying enzyme (RDE) and heat-inactivated to remove non-specific inhibitors prior to use in deep mutational scanning library selections and neutralization assays. RDE was prepared by resuspending one vial of lyophilized RDE II (Seikan) in 20 mL PBS and filtering through a 0.22 µM filter. Sera and RDE were combined at a 1:3 sera to RDE ratio, incubated at 37℃ for 2.5 hours, and then incubated at 55℃ for 30 minutes as previously described[60]. RDE-treated sera were stored at -80C until further use.

Design of deep mutational scanning libraries

We used a previously described lentiviral backbone that is schematized in **Fig. S1A**[23]. The libraries were designed in the background of the A/Massachusetts/18/2022 HA, which was the 2024-2025 cell-based vaccine strain. The HA gene was codon-optimized via the GenSmart codon optimization tool offered by GenScript, as we found this codon optimization increases viral titers. The plasmid map for the lentiviral backbone with codon-optimized HA sequence is at: https://github.com/dms-vep/Flu_H3_Massachusett

s2022_DMS/blob/main/data/supplemental_data/plasmids/4570_pH2rU3_ForInd_Massachusetts2022HA_GenscriptV1_T7_CMV_ZsGT2APurR.gb. We aimed to include all single amino acid mutations in the HA ectodomain (H3 numbering: 1-504). 20 stop codons located at alternating positions from the start of the ectodomain were also included as negative controls for cell entry measurements. We ordered a site-saturation variant library with these specifications from Twist Biosciences. The final Twist quality control report for the library is at: https://github.com/dms-vep/Flu_H3_Massachusetts2022_DMS/blob/main/data/supplemental_data/Final_QC_Twist_VariantProportion.csv

Cloning of deep mutational scanning plasmid library

We cloned the deep mutational scanning plasmid libraries following an approach first described in Dadonaite et al.[23]. Barcoding PCR was performed using the Twist library as template to append random 16 nucleotide barcodes downstream of the HA gene stop codon. 5 ng of Twist library (1 µL) was combined with 1.5 µL of ForInd_AddBC_2 primer (10 µM), 1.5 µL of 5'for_lib_bcing primer (10 µM), 21 µL of molecular biology grade water, and 25 µL of KOD Hot Start Master Mix (ThermoFisher, Ref. No. 71842-4). The PCR cycling conditions were:

1. 95℃, 2 min
2. 95℃, 20 sec
3. 55.5℃, 20 sec, cooling at 0.5℃/sec
4. 70℃, 1 min
5. Return to Step 2, 9 cycles
6. 12℃ hold

The barcoding was performed in two independent reactions, yielding two barcoded PCR products to serve as biological library replicates (libraries A and B). Therefore, the two libraries contain unique barcodes, and all subsequent cloning and virus generation steps were carried out separately for each library.

The lentiviral backbone was digested from a plasmid containing mCherry in place of the HA insert (3137_pH2rU3_ForInd_mCherry_CMV_ZsGT2APurR) by incubating with XbaI and MluI for 2 hours at 37℃, followed by 20 minutes at 65℃ to inactivate XbaI.

Both the barcoded HA libraries and digested lentiviral backbone were run on a 0.8% agarose gel, and bands of the expected size were excised and purified using the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel, Cat. No. 740609.5) followed by additional purification with Ampure XP beads (Beckman Coulter, Cat. No. A63881) to ensure high purity. All were eluted in molecular biology grade water.

Barcoded HA libraries were cloned into the lentiviral backbone at a 1:2 insert to vector ratio in a 1 hour Hifi assembly reaction using the NEBuilder HiFi DNA Assembly kit (NEB E5520S). The Hifi reactions were purified with Ampure XP beads and eluted in molecular biology grade water, 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. 15 electroporation reactions were performed for each library and bacteria were plated on 15 cm LB+ampicillin plates and grown overnight at 37℃. The next day, colonies were scraped with LB+ampicillin and plasmids were extracted using the QIAGEN HiSpeed Plasmid Maxi Kit (Cat. No. 12662). The total number of colonies for Library A and B were 8.4e6 and 7.5e6 CFU's respectively. Large numbers of colonies at this stage are necessary to ensure library diversity does not become bottlenecked.

Based on the Twist quality control report, 35 sites were missing >75% of mutations and 19 mutations at other sites were missing in the library. Therefore, we aimed to clone a "spike-in" plasmid library that contains these missing mutations using a previously described mutagenesis PCR protocol[24,61]. We designed NNS primers for missing sites with https://github.com/jbloomlab/CodonTilingPrimers and primers for missing mutations with https://github.com/jbloomlab/TargetedTilingPrimers. Forward and reverse primer pools were created by combining either forward or reverse NNS and targeted mutation primers at an equal molar ratio per codon. To prepare a linear HA template for mutagenesis PCR, the lentiviral backbone plasmid encoding the codon-optimized HA was incubated with NotI and NdeI for 2 hours at 37℃, followed by 20 minutes at 65℃ to inactivate both enzymes. The digest product

was run on a 0.8% agarose gel, and the band corresponding to the linear HA fragment was purified as described previously.

The protocol for mutagenesis involves two reactions: a mutagenesis PCR and a joining PCR. Two separate replicates were performed to form biological library replicates (spike-in libraries A and B) as was done for the Twist libraries. The first mutagenesis PCR was divided into forward and reverse reactions. Both forward and reverse reactions shared the following PCR conditions: 4 µL of linear HA template (3 ng/µL), 8 µL of molecular biology grade water, and 15 µL of KOD Hot Start Master Mix. 1.5 µL of the forward primer pool (5 µM) and 1.5 µL of 3'rev_linjoin_KHDC primer (5 µM) were added to the forward reaction. 1.5 µL of the reverse primer pool (5 µM) and 1.5 µL of VEP_Amp_For primer (5 µM) were added to the reverse reaction. The PCR cycling conditions were:

1. 95℃, 2 min
2. 95℃, 20 sec
3. 70℃, 1 sec
4. 54℃, 20 sec, cooling at 0.5℃/sec
7. 70℃, 50 sec
8. Return to Step 2, 9 cycles
9. 4℃ hold

The forward and reverse mutagenesis PCR products were diluted 1:4 in molecular biology grade water and 4 µL of each were added to the joining PCR reaction along with the following: 1.5 µL of 3'rev_linjoin_KHDC primer (5 µM), 1.5 µL of VEP_Amp_For primer (5 µM), 4 µL of molecular biology grade water, and 15 µL of KOD Hot Start Master Mix. The PCR cycling conditions were:

1. 95℃, 2 min
2. 95℃, 20 sec
3. 70℃, 1 sec
4. 54℃, 20 sec, cooling at 0.5℃/sec
5. 70℃, 65 sec
6. Return to Step 2, 19 cycles
7. 4℃ hold

A DpnI digest was performed afterwards to remove any potential unmutated HA template (which would be methylated) by incubating the joining PCR products with DpnI for 20 minutes at 37℃. DpnI-digested joining PCR products were run on a 0.8% gel and the expected bands were excised and purified as described above. The mutagenized HA fragments were barcoded and cloned into the lentiviral backbone following the same approach described above for the Twist library. However, spike-in plasmids were extracted using the QIAprep Spin Miniprep Kit (Qiagen, cat. no. 27104) instead of via maxipreps.

Corresponding replicates of the Twist plasmid libraries and spike-in plasmid libraries were combined at a 1:4 Twist to spike-in molar ratio per codon, with the spike-in library intentionally added at four times the amount required for an equal per-codon molar ratio because long-read PacBio sequencing of the plasmid libraries revealed this ratio results in the most even distribution of mutants in the combined libraries.

Production of cell-stored deep mutational scanning libraries

Deep mutational scanning requires genotype-phenotype linked pseudoviruses. The rationale for this is described in detail in the caption of **Fig. S1B**. We followed a previously described protocol to generate cell-stored deep mutational scanning libraries where each cell is integrated with a single copy of a barcoded HA mutant to enable rescue of genotype-phenotype linked pseudoviruses, with a few changes[23,24]. 15cm plates were plated with ~20 million 293T cells. On the next day, each plate was transfected with 12.5 µg of plasmid library encoding the lentiviral backbone with barcoded HA mutants, 3.125 µg of each lentiviral helper plasmid (26_HDM_Hgpm2, 27_HDM_tat1b, and 28_pRC_CMV_Rev1b), 3.125 µg of plasmid expressing VSV-G (29_HDM_VSV_G), and 3.125 µg of plasmid expressing a strain-matched codon-optimized neuraminidase gene (4576_HDM_Massachusetts2022NA_Genscript). BioT transfection reagent (Bioland Scientific, Cat. No. B01-02) was used according to manufacturer's instructions. Note the NA is important because the virions produced here will also have HA expressed on

their surface from the lentiviral backbone. The NA expression prevents HA from binding to producing cells, as this binding could bias the library since the HA mutants will have different abilities to bind to sialic acids on the producing 293T cells. 48 hours post transfection, the supernatant was filtered through a 0.45 µm syringe filter (Corning, Cat. No. 431220) and stored at -80°C. 4 x 15cm plates were transfected with each library replicate, resulting in ~100 ml of VSV-G pseudotyped library viruses. An aliquot of these viruses was used to infect 293T cells, and the titer in transcription units (TU) per mL was determined by measuring the percentage of zsGreen positive cells via flow cytometry.

The VSV-G pseudotyped library viruses were used to infect 293T-rtTA cells at a multiplicity of infection (MOI) of 0.7% to ensure each cell integrates at most one copy of provirus. The MOI was confirmed by measuring the percentage of zsGreen positive cells via flow cytometry at 48 hours post transduction. Based on the measured MOI and number of cells present during infection, cells were pooled such that each library would contain an estimated 60,000 infected cells. This number was chosen to be high enough to ensure each mutant is associated with multiple barcodes to increase measurement accuracy (for ~10,000 mutants, each mutant would have ~6 barcodes), while also being low enough that it would be possible to measure all variants given our pseudovirus titers in selection experiments. Note the final numbers ended up being close to this target: 64,032 for library A and 70,581 for library B (**Fig. S1C**). Integrated cells were selected by growing in the presence of 0.75 µg/mL of puromycin for 1 week (fresh media with puromycin was replenished every 48 hours). After selection was complete, integrated cells were expanded in tet-free D10 for 24 hours and then frozen down in liquid nitrogen in 2e7 cell aliquots for long-term storage.

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 tet-free D10 supplemented with 1 µg/mL of doxycycline to induce HA expression from the integrated genomes. On the next day, each flask was transfected with 43.75 µg of each helper plasmid (26_HDM_Hgpm2, 27_HDM_tat1b, and 28_pRC_CMV_Rev1b), 15 µg of plasmid expressing human airway trypsin-like protease (3781_HDM_HAT) to activate HA for membrane fusion, and 3.75 µg of plasmid expressing NA (4576_HDM_Massachusetts2022NA_Genscript). BioT transfection reagent was used according to manufacturer's instructions. At 16 hours post transfection, the tet-free D10 in each flask was aspirated and 150 mL of IGM supplemented with 1 µg/mL of doxycycline was added. This swap to low-serum media is absolutely necessary because non-specific inhibitors in FBS can inactivate HA and interfere with pseudovirus infection. At 32 hours post media swap, the supernatant was filtered through a 0.45µm SFCA Nalgene 500 mL Rapid-Flow filter unit (Cat. No. 09-740-44B). Filtered supernatant was then concentrated by adding LentiX Concentrator (Takara, Cat. No. 631232) at a 1:3 virus to concentrator ratio, incubating at 4°C overnight, and spinning at 1500 x g and 4°C for 45 minutes. Following centrifugation, supernatant was discarded and viral pellets resuspended in NAM to an estimated titer of ~2e6 TU/mL. 1 mL aliquots of concentrated HA expressing pseudoviruses were frozen at -80°C for use in downstream selection experiments.

To rescue VSV-G expressing pseudoviruses from integrated cells, 30 million cells were plated in 10 cm plates in tet-free D10. On the next day, each plate was transfected with 7.3125 µg of each helper plasmid (26_HDM_Hgpm2, 27_HDM_tat1b, and 28_pRC_CMV_Rev1b), 0.75 µg of plasmid expressing NA (4576_HDM_Massachusetts2022NA_Genscript), and 7.3125 µg of plasmid expressing VSV-G (29_HDM_VSV_G). At 48 hours post transfection, the supernatant was filtered through a 0.45µm SFCA Nalgene 500 mL Rapid-Flow filter unit and concentrated using LentiX Concentrator but viral pellets were resuspended in D10. Aliquots of concentrated VSV-G expressing pseudoviruses were frozen at -80°C for use in linking mutations to barcodes and cell entry selection experiments.

Long-read sequencing to link mutations to barcodes

1e6 293T cells were plated in each well of 6-well plates coated with poly-L-lysine to help with cell adhesion. On 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. At 12 hours post infection, the

non-integrated reverse-transcribed lentiviral genomes were recovered by miniprepping the 293T cells using the QIAprep Spin Miniprep Kit.

Amplicons for long-read sequencing of the miniprepped genomes were prepared by following an approach described in Dadonaite et al.[23]. Briefly, the eluted minipreps were split into two separate reactions so each could be uniquely tagged for detecting strand exchange events from the PCR. The number of PCR cycles was chosen intentionally to limit the possibility of strand exchange. Both reactions shared the following PCR conditions: 20 µL of KOD Hot Start Master Mix and 18 µL of miniprepped DNA. 1 µL of 5_PacBio_G primer (10 µM) and 1 µL of 3_PacBio_C primer (10 µM) were added to the first reaction. 1 µL of 5_PacBio_C primer (10 µM) and 1 µL of 3_PacBio_G primer (10 µM) were added to the second reaction. The PCR cycling conditions were:

1. 95℃, 2 min
2. 95℃, 20 sec
3. 70℃, 1 sec
4. 60℃, 10 sec, cooling at 0.5℃/sec
5. 70℃, 60 sec
6. Return to Step 2, 7 cycles
7. 70℃, 1 min
8. 4℃ hold

The round 1 PCR products were purified with 50 µL of Ampure XP beads and eluted in 35 µL of elution buffer. For each library, equal volumes of the two separate round 1 PCR reactions were pooled. The round 2 PCR reactions contained: 25 µL of KOD Hot Start Master Mix, 21 µL of pooled round 1 product, 2 µL of 5_PacBio_Rnd2 primer (10 µM) and 2 µL of 3_PacBio_Rnd2 primer (10 µM). The PCR cycling conditions were:

1. 95℃, 2 min
2. 95℃, 20 sec
3. 70℃, 1 sec
4. 60℃, 10 sec, cooling at 0.5℃/sec
5. 70℃, 1 min
6. Return to Step 2, 10 cycles
7. 70℃, 1 min
8. 4℃ hold

The round 2 PCR products were purified with 50 µL of Ampure XP beads and eluted in 40 µL of elution buffer. PCR reactions for each library were combined and amplicon length was verified by TapeStation prior to sequencing. Libraries were sequenced on a single SMRT cell with a movie length of 30 hours on a PacBio Sequel IIe sequencer. For details on computational analysis, see the '**PacBio sequencing analysis**' section.

Mutation effects on cell entry

To measure effects of HA mutations on cell entry, we followed the approach described in Dadonaite et al.[23]. Briefly, we infected MDCK-SIAT1 cells with the HA expressing pseudovirus library and infected 293T cells with the VSV-G expressing pseudovirus library. The VSV-G expressing library is necessary to provide a baseline for infection as VSV-G can mediate cell entry without relying on HA. We used 293T cells for VSV-G infection because titers of VSV-G expressing pseudoviruses are higher when infecting 293T cells compared to MDCK-SIAT1 cells.

1e6 293T cells in D10 or 7e5 MDCK-SIAT1 cells in NAM were plated in each well of 6-well plates. 2.5 µg/mL of amphotericin B was added to the MDCK-SIAT1 cells when plating as this improves HA pseudovirus titers. On the next day, we infected the MDCK-SIAT1 cells with ~1.2e6 TU's of HA pseudovirus library and the 293T cells with ~8e6 TU's of VSV-G pseudovirus library. Prior to infection, the HA pseudovirus library was treated with 500 nM of oseltamivir for 20 minutes on ice to inhibit NA from interfering with cell entry. Note infections with the HA pseudovirus library must be done in NAM, as the serum in D10 contains non-specific inhibitors that inhibit H3 infection.

At 12 hours post infection, the non-integrated reverse-transcribed lentiviral genomes were recovered by miniprepping the 293T and MDCK-SIAT1 cells. To prepare the amplicons for Illumina sequencing, two rounds of PCR were performed: the first round appends the Illumina Truseq Read 1 and Read 2 sequences, and the second round attaches indices for multiplexing. The Round 1 PCR reactions contained: 22 µL of miniprepped DNA, 25 µL of KOD Hot Start Master Mix, 1.5 µL of Illumina_Rnd1_For primer (10 µM), and 1.5 µL of Illumina_Rnd1_Rev3 primer (10 µM). The PCR cycling conditions were:

1. 95℃, 2 min
2. 95℃, 20 sec
3. 70℃, 1 sec
4. 58℃, 10 sec, cooling at 0.5℃/sec 5. 70℃, 20 sec
5. Return to Step 2, 27 cycles
6. 70℃, 1 min
7. 4℃ hold

Round 1 PCR products were purified with 150 µL of Ampure XP beads and eluted in 50 µL of elution buffer. Concentrations of each PCR product were determined by Qubit 4 Fluorometer (ThermoFisher, Ref. No. Q33238). The Round 2 PCR reactions contained: 20 ng of Round 1 PCR product, 25 µL of KOD Hot Start Master Mix, 2 µL of each of the Round 2 indexing primers (10 µM each), and up to 25 µL of molecular biology grade water. The same PCR cycling conditions as Round 1 were used, except only 20 cycles were performed. Concentrations of each round 2 PCR product were determined by Qubit 4 Fluorometer. The samples were then pooled in equal DNA amounts and run on a 1% agarose gel. The correct size band (283 bp) was excised, purified with Ampure XP beads, diluted to a concentration of 4 nM, and sequenced on a Illumina NextSeq 2000 (with P2 reagent kit) or NovaSeq X Plus system. For details on how sequencing counts were converted to mutation effects on cell entry, see the '**Illumina sequencing barcode analysis**' section.

Mutation effects on acid stability

To measure effects of HA mutations on acid stability, we followed the approach in Dadonaite et al.[24]. Briefly, we incubated the HA pseudovirus library in different acidic pH buffers prior to infecting MDCK-SIAT1 cells. We also included a condition where the HA pseudovirus library was incubated with neutral pH media prior to infection.

7e5 MDCK-SIAT1 cells in NAM were plated in each well of 6-well plates. 2.5 µg/mL of amphotericin B was added to the MDCK-SIAT1 cells when plating. On the next day, aliquots containing ~2.4e6 TU/mL of HA pseudovirus library were incubated with citrate-based acidic buffers at pH 6.1, 5.9, 5.7, 5.5, 5.3 or NAM (neutral pH condition) for 60 minutes at 37℃. After incubation, libraries were concentrated with 100k Amicon spin columns (Milipore, UFC910008) by spinning for 15 minutes at 1500 x g, resuspending in 11 mL of PBS to neutralize the acidic buffers, and spun down again for 20 minutes at 1500 x g. The libraries were then resuspended in 2 mL of NAM, treated with 500 nM of oseltamivir for 20 minutes on ice to inhibit NA, and used to infect the plated MDCK-SIAT1 cells.

At 12 hours post infection, the non-integrated reverse-transcribed lentiviral genomes were recovered by miniprepping the MDCK-SIAT1 cells as previously described, with one important modification. In the miniprep lysis step where P2 buffer is added, we spiked in a DNA standard at an amount calculated to be approximately 3% of the recovered lentiviral DNA (based on the estimated number of non-integrated lentiviral genomes) under normal infection conditions with no acidic buffer treatment. The rationale for including the DNA spike-in standard is to enable relative sequencing counts to be converted into absolute quantities of each barcoded pseudovirus variant, normalized to the standard, across different acidic buffer conditions. This DNA standard is a plasmid that encodes the lentiviral backbone with a barcoded mCherry gene; the plasmid map is 3068_ForInd_mC_BCs_pool1 and the barcodes are at: https://github.com/dms-vep/Flu_H3_Massachusetts2022_DMS/blob/main/data/neutralization_standard_barcodes.csv. Afterwards, amplicons for Illumina sequencing were prepared as described previously. For details on how sequencing counts were converted to mutation effects on acid stability, see the '**Illumina sequencing barcode analysis**' section.

Mutation effects on sera neutralization

To measure effects of HA mutations on sera neutralization, we followed the approach in Dadonaite et al.[23]. 7e5 MDCK-SIAT1 cells in NAM were plated in each well of 6-well plates. 2.5 µg/mL of amphotericin B was added to the MDCK-SIAT1 cells when plating. On the next day, aliquots containing ~1.2e6 TU/mL of HA pseudovirus library were treated with 500 nM oseltamivir and incubated with three concentrations of sera estimated to span between IC98 and IC98*16. These IC values were determined by a luciferase-based pseudovirus neutralization assay. Multiple dilutions of sera are necessary for improving estimation of mutation effects on sera neutralization. Libraries and sera were incubated for 60 minutes at 37℃. After incubation, libraries were used to infect the plated MDCK-SIAT1 cells. At 12 hours post infection, the non-integrated reverse-transcribed lentiviral genomes were recovered by miniprepping with the spike-in DNA standard and amplicons for Illumina sequencing were prepared as described previously. For details on how sequencing counts were converted to mutation effects on sera neutralization, see the '**Illumina sequencing barcode analysis**' section.

Production of conditionally replicative influenza viruses

Conditionally replicative influenza viruses that lack the PB1 gene were produced by reverse genetics as previously described[29,62]. The native HA sequence was cloned into a bidirectional pHW2000 influenza reverse genetics plasmid. The plasmid map for A/Massachusetts/18/2022 HA is 5012_pHW_MA22_HA and the plasmid map for A/Perth/16/2009 HA is 1442_pHWPerth09_HA. Mutant HA plasmids were cloned by PCR with partially overlapping primers that contain the mutation of interest, followed by HiFi assembly. All plasmids were sequence confirmed by Plasmidsaurus. To perform the virus rescue, 5e5 293T-CMV-PB1 cells and 4e5 MDCK-SIAT1-CMV-PB1-TMPRSS2 cells were plated in D10 in each well of 6-well plates. On the next day, each well was transfected with 2 µg total of plasmids including: 0.25 µg each of six reverse genetics plasmids expressing genes from A/WSN/1933 (30_pHW181_PB2, 32_pHW183_PA, 34_pHW185_NP, 35_pHW186_NA, 36_pHW187_M, and 37_pHW188_NS), 0.25 µg of plasmid that expresses eGFP in place of PB1 (208_pHH_PB1flank_eGFP), and 0.25 µg of HA reverse genetics plasmid. At 24 hours post transfection, D10 was aspirated and each well was replenished with 2 mL of IGM. At 48 hours post media swap, viral supernatant was spun down for 4 minutes at 845 x g, and aliquots of clarified supernatant were collected and frozen down at -80°C.

Validation of cell entry effects

Conditionally replicative influenza viruses were serially diluted in NAM in 96-well plates. 5e4 MDCK-SIAT1-CMV-PB1 cells in NAM were added to each well. Note these cells do not express TMPRSS2, so the influenza viruses can only undergo a single cycle of infection. At 16 hours post infection, wells with 1-10% percent eGFP-positive cells were selected. Precise measurements of the percent of eGFP-positive cells in these wells were obtained by flow cytometry and viral titers were calculated using a Poisson distribution.

Validation of acid stability effects

1.5e5 MDCK-SIAT1-CMV-PB1 cells in D10 were plated in each well of a 12-well plate. Conditionally replicative influenza viruses were diluted to a target MOI of 0.5 to 1 (~2 to 10 µL of virus) in 100 µL of citrate-based acidic buffers at pH 5.7, 5.5, and 5.3 or NAM (neutral pH condition) and incubated for 60 minutes at 37°C. The pH-treated viruses were then brought back to neutral pH by diluting the 100 µL into 2 mL of NAM. 4 hours after plating the MDCK-SIAT1-CMV-PB1 cells, D10 was aspirated and cells were washed with 1 mL of 1X PBS before 2 mL of the NAM-diluted viruses were added. At 16 hours post infection, the percent of eGFP-positive cells in each well was determined by flow cytometry. The fraction infectivity retained was calculated as the ratio of percent eGFP-positive cells when virus was treated with acidic pH over the percent eGFP-positive cells when virus was treated with NAM.

Neutralization assays

Sera were serially diluted in NAM in 96-well plates. Conditionally replicative influenza viruses were diluted to a target MOI that falls within a range where the fluorescence signal would change linearly with respect to neutralization. The virus and sera dilutions were incubated for 60 minutes at 37°C. Afterwards, 4e4 MDCK-SIAT1-CMV-PB1 cells were added to each well. At 16 hours post infection, the fluorescence signal was read on a Tecan M1000 plate reader and the fraction infectivity was determined relative to no serum controls.

PacBio sequencing analysis

PacBio circular consensus sequences (CCSs) were aligned to the HA reference sequence using *alignparse*[63]. Consensus sequences for each barcode were determined by requiring at least 3 CCSs per barcode. See https://github.com/dms-vep/Flu_H3_Massachusetts2022_DMS/blob/main/results/variants/codon_variants.csv for the final barcode-variant table.

For full details on the analysis, see these notebooks for:

- Analyzing the PacBio CCS's: https://dms-vep.org/Flu_H3_Massachusetts2022_DMS/notebooks/analyze_pacbio_ccs.html
- Building PacBio consensus sequences: https://dms-vep.org/Flu_H3_Massachusetts2022_DMS/notebooks/build_pacbio_consensus.html
- Building the final barcode-variant table: https://dms-vep.org/Flu_H3_Massachusetts2022_DMS/notebooks/build_codon_variants.html

Illumina sequencing barcode analysis

From the Illumina short-read sequencing data, barcodes were counted by https://jbloomlab.github.io/dms_variants/dms_variants.illuminabarcodeparser.html and then mutation effects were calculated using approaches described previously[23,24].

To convert barcode counts into mutation effects on cell entry, we first calculated functional scores. Briefly, a functional score for a variant v was calculated as $\log_2[(n^v_{post}/n^{wt}_{post})/(n^v_{pre}/n^{wt}_{pre})]$, where each n is a count of barcodes that entered cells. Specifically, n^v_{post} is the count of each variant in the HA pseudovirus library, n^v_{pre} is the count of each variant in the VSV-G pseudovirus library, n^{wt}_{post} and n^{wt}_{pre} are counts of the unmutated (wildtype) variants in these libraries. Positive functional scores indicate the variant is better at entering cells relative to the unmutated HA, while negative functional scores indicate the variant is worse at entering cells relative to the unmutated HA. Since some variants contain multiple mutations, we used *multidms*[64] (https://matsengrp.github.io/multidms) to fit a global epistasis model with a sigmoid function using the functional scores to obtain individual mutation effects on cell entry. For more details on fitting, see the notebooks under 'Functional effects of mutations' at https://dms-vep.org/Flu_H3_Massachusetts2022_DMS/appendix.html. We report the median mutation effect across library replicates and filter for mutations that are seen in at least two different barcoded variants (averaged across libraries). See https://dms-vep.org/Flu_H3_Massachusetts2022_DMS/cell_entry.html for interactive visualizations of mutation effects on cell entry.

To convert barcode counts into mutation effects on acid stability and sera neutralization, we calculated the fraction infectivity of each variant retained at each acidic pH buffer treatment or serum concentration, normalizing to the counts of spike-in standard barcodes in each condition. We then fit a biophysical model to these fractional infectivity data using *polyclonal*[27] (https://jbloomlab.github.io/polyclonal) to obtain individual mutation effects on acid stability and sera neutralization. For more details on fitting, see the notebooks under 'Antibody/serum escape' and 'Stability' at https://dms-vep.org/Flu_H3_Massachusetts2022_DMS/appendix.html. We report the average mutation effect across library replicates, and filter for mutations that are seen in at least two different barcoded variants (averaged across libraries) and have a cell entry score > -3. See https://dms-vep.org/Flu_H3_Massachusetts2022_D

MS/acid_stability.html for interactive visualizations of mutation effects on acid stability, and https://dms-vep.org/Flu_H3_Massachusetts2022_DMS/sera_neutralization.html for interactive visualizations of mutation effects on sera neutralization.

Evolutionary entrenchment analysis

The frequencies of amino acids at different timepoints were obtained from a Nextstrain H3N2/HA/60y build at https://nextstrain.org/groups/blab/flu/seasonal/h3n2/ha/60y. An amino acid was considered fixed if at any timepoint its frequency at a given site was >95%. Sites were considered inside the receptor binding pocket if they were within 4Å of sialic acid or previously reported to affect receptor binding. See **Table S1** for the full definition of receptor binding pocket sites. See https://dms-vep.org/Flu_H3_Massachusetts2022_DMS/entrenchment.html for an interactive plot of the analysis.

Structural analysis

UCSF ChimeraX v1.8[65] was used for structural visualizations. All Protein Data Bank accession IDs used are included in figure legends.

References

1. Couch, R. B. & Kasel, J. A. Immunity to influenza in man. *Annu. Rev. Microbiol.* **37**, 529–549 (1983).
2. Krammer, F. The human antibody response to influenza A virus infection and vaccination. *Nat. Rev. Immunol.* **19**, 383–397 (2019).
3. Kikawa, C. *et al.* High-throughput neutralization measurements correlate strongly with evolutionary success of human influenza strains. *bioRxiv* 2025.03.04.641544 (2025) doi:10.1101/2025.03.04.641544.
4. Smith, D. J. *et al.* Mapping the antigenic and genetic evolution of influenza virus. *Science* **305**, 371–376 (2004).
5. Bedford, T. *et al.* Integrating influenza antigenic dynamics with molecular evolution. *Elife* **3**, e01914 (2014).
6. Petrova, V. N. & Russell, C. A. The evolution of seasonal influenza viruses. *Nat. Rev. Microbiol.* **16**, 47–60 (2018).
7. Fitch, W. M., Bush, R. M., Bender, C. A. & Cox, N. J. Long term trends in the evolution of H(3) HA1 human influenza type A. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 7712–7718 (1997).
8. Myers, J. L. *et al.* Compensatory hemagglutinin mutations alter antigenic properties of influenza viruses. *J. Virol.* **87**, 11168–11172 (2013).
9. Wu, N. C. *et al.* Diversity of functionally permissive sequences in the receptor-binding site of influenza hemagglutinin. *Cell Host Microbe* **21**, 742–753.e8 (2017).
10. Wu, N. C. *et al.* A complex epistatic network limits the mutational reversibility in the influenza hemagglutinin receptor-binding site. *Nat. Commun.* **9**, 1264 (2018).
11. Wu, N. C. *et al.* Major antigenic site B of human influenza H3N2 viruses has an evolving local fitness landscape. *Nat. Commun.* **11**, 1233 (2020).
12. Lei, R. *et al.* Epistasis mediates the evolution of the receptor binding mode in recent human H3N2 hemagglutinin. *Nat. Commun.* **15**, 5175 (2024).
13. Thompson, A. J. *et al.* Evolution of human H3N2 influenza virus receptor specificity has substantially expanded the receptor-binding domain site. *Cell Host Microbe* **32**, 261–275.e4 (2024).
14. Liang, R. *et al.* Epistasis in the receptor binding domain of contemporary H3N2 viruses that reverted to bind sialylated diLacNAc repeats. *bioRxiv* (2024).
15. Lee, C.-Y. *et al.* Epistasis reduces fitness costs of influenza A virus escape from stem-binding antibodies. *Proc. Natl. Acad. Sci. U. S. A.* **120**, e2208718120 (2023).
16. Welsh, F. C. *et al.* Age-dependent heterogeneity in the antigenic effects of mutations to influenza hemagglutinin. *Cell Host Microbe* **32**, 1397–1411.e11 (2024).
17. Thyagarajan, B. & Bloom, J. D. The inherent mutational tolerance and antigenic evolvability of influenza hemagglutinin. *Elife* **3**, (2014).
18. Doud, M. B. & Bloom, J. Accurate measurement of the effects of all amino-acid mutations on influenza hemagglutinin. *Viruses* **8**, 155 (2016).
19. Lee, J. M. *et al.* Deep mutational scanning of hemagglutinin helps predict evolutionary fates of human H3N2 influenza variants. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E8276–E8285 (2018).
20. Wu, N. C. *et al.* High-throughput profiling of influenza A virus hemagglutinin gene at single-nucleotide resolution. *Sci. Rep.* **4**, 4942 (2014).
21. Skehel, J. J. & Wiley, D. C. Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. *Annu. Rev. Biochem.* **69**, 531–569 (2000).
22. Benton, D. J., Gamblin, S. J., Rosenthal, P. B. & Skehel, J. J. Structural transitions in influenza haemagglutinin at membrane fusion pH. *Nature* **583**, 150–153 (2020).
23. Dadonaite, B. *et al.* A pseudovirus system enables deep mutational scanning of the full SARS-CoV-2

spike. *Cell* **186**, 1263–1278.e20 (2023).
24. Dadonaite, B. *et al.* Deep mutational scanning of H5 hemagglutinin to inform influenza virus surveillance. *PLoS Biol.* **22**, e3002916 (2024).
25. Frutos, A. M. *et al.* Interim estimates of 2024-2025 seasonal influenza vaccine effectiveness - four vaccine effectiveness networks, United States, October 2024-February 2025. *MMWR Morb. Mortal. Wkly. Rep.* **74**, 83–90 (2025).
26. Otwinowski, J., McCandlish, D. M. & Plotkin, J. B. Inferring the shape of global epistasis. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E7550–E7558 (2018).
27. Yu, T. C. *et al.* A biophysical model of viral escape from polyclonal antibodies. *Virus Evolution* **8**, veac110 (2022).
28. Matrosovich, M., Matrosovich, T., Carr, J., Roberts, N. A. & Klenk, H.-D. Overexpression of the alpha-2,6-sialyltransferase in MDCK cells increases influenza virus sensitivity to neuraminidase inhibitors. *J. Virol.* **77**, 8418–8425 (2003).
29. Hooper, K. A. & Bloom, J. D. A mutant influenza virus that uses an N1 neuraminidase as the receptor-binding protein. *J. Virol.* **87**, 12531–12540 (2013).
30. Doud, M. B., Hensley, S. E. & Bloom, J. D. Complete mapping of viral escape from neutralizing antibodies. *PLoS Pathog.* **13**, e1006271 (2017).
31. Webster, R. G. & Laver, W. G. Determination of the number of nonoverlapping antigenic areas on Hong Kong (H3N2) influenza virus hemagglutinin with monoclonal antibodies and the selection of variants with potential epidemiological significance. *Virology* **104**, 139–148 (1980).
32. Skehel, J. J. *et al.* A carbohydrate side chain on hemagglutinins of Hong Kong influenza viruses inhibits recognition by a monoclonal antibody. *Proc. Natl. Acad. Sci. U. S. A.* **81**, 1779–1783 (1984).
33. Broecker, F. *et al.* Immunodominance of antigenic site B in the hemagglutinin of the current H3N2 influenza virus in humans and mice. *J. Virol.* **92**, (2018).
34. Popova, L. *et al.* Immunodominance of antigenic site B over site A of hemagglutinin of recent H3N2 influenza viruses. *PLoS One* **7**, e41895 (2012).
35. Chambers, B. S., Parkhouse, K., Ross, T. M., Alby, K. & Hensley, S. E. Identification of hemagglutinin residues responsible for H3N2 antigenic drift during the 2014-2015 influenza season. *Cell Rep.* **12**, 1–6 (2015).
36. Wu, N. C. *et al.* Different genetic barriers for resistance to HA stem antibodies in influenza H3 and H1 viruses. *Science* **368**, 1335–1340 (2020).
37. Godley, L. *et al.* Introduction of intersubunit disulfide bonds in the membrane-distal region of the influenza hemagglutinin abolishes membrane fusion activity. *Cell* **68**, 635–645 (1992).
38. Kemble, G. W., Bodian, D. L., Rosé, J., Wilson, I. A. & White, J. M. Intermonomer disulfide bonds impair the fusion activity of influenza virus hemagglutinin. *J. Virol.* **66**, 4940–4950 (1992).
39. Rachakonda, P. S. *et al.* The relevance of salt bridges for the stability of the influenza virus hemagglutinin. *FASEB J.* **21**, 995–1002 (2007).
40. Parsons, L. M., An, Y., de Vries, R. P., de Haan, C. A. M. & Cipollo, J. F. Glycosylation characterization of an influenza H5N7 hemagglutinin series with engineered glycosylation patterns: Implications for structure-function relationships. *J. Proteome Res.* **16**, 398–412 (2017).
41. Mair, C. M., Ludwig, K., Herrmann, A. & Sieben, C. Receptor binding and pH stability - how influenza A virus hemagglutinin affects host-specific virus infection. *Biochim. Biophys. Acta* **1838**, 1153–1168 (2014).
42. Carr, C. M. & Kim, P. S. A spring-loaded mechanism for the conformational change of influenza hemagglutinin. *Cell* **73**, 823–832 (1993).
43. Gamblin, S. J. *et al.* Hemagglutinin structure and activities. *Cold Spring Harb. Perspect. Med.* **11**,

a038638 (2021).
44. Shah, P., McCandlish, D. M. & Plotkin, J. B. Contingency and entrenchment in protein evolution under purifying selection. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E3226–35 (2015).
45. Pollock, D. D., Thiltgen, G. & Goldstein, R. A. Amino acid coevolution induces an evolutionary Stokes shift. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E1352–9 (2012).
46. Ranjeva, S. *et al.* Age-specific differences in the dynamics of protective immunity to influenza. *Nat. Commun.* **10**, 1660 (2019).
47. Loes, A. N. *et al.* High-throughput sequencing-based neutralization assay reveals how repeated vaccinations impact titers to recent human H1N1 influenza strains. *J. Virol.* **98**, e0068924 (2024).
48. Sailer, Z. R. & Harms, M. J. Detecting high-order epistasis in nonlinear genotype-phenotype maps. *Genetics* **205**, 1079–1088 (2017).
49. Sailer, Z. R. & Harms, M. J. High-order epistasis shapes evolutionary trajectories. *PLoS Comput. Biol.* **13**, e1005541 (2017).
50. Morrison, A. J. & Harms, M. J. An experimental demonstration of ensemble epistasis in the lac repressor. *bioRxiv* 2022.10.14.512271 (2022) doi:10.1101/2022.10.14.512271.
51. Zost, S. J. *et al.* Canonical features of human antibodies recognizing the influenza hemagglutinin trimer interface. *J. Clin. Invest.* **131**, (2021).
52. Turner, H. L. *et al.* Potent anti-influenza H7 human monoclonal antibody induces separation of hemagglutinin receptor-binding head domains. *PLoS Biol.* **17**, e3000139 (2019).
53. Watanabe, A. *et al.* Antibodies to a conserved influenza head interface Epitope protect by an IgG subtype-dependent mechanism. *Cell* **177**, 1124–1135.e16 (2019).
54. Bangaru, S. *et al.* A site of vulnerability on the influenza virus hemagglutinin head domain trimer interface. *Cell* **177**, 1136–1152.e18 (2019).
55. Chai, N. *et al.* Two escape mechanisms of influenza A virus to a broadly neutralizing stalk-binding antibody. *PLoS Pathog.* **12**, e1005702 (2016).
56. Shi, Y., Wu, Y., Zhang, W., Qi, J. & Gao, G. F. Enabling the 'host jump': structural determinants of receptor-binding specificity in influenza A viruses. *Nat. Rev. Microbiol.* **12**, 822–831 (2014).
57. Kong, H. *et al.* H3N2 influenza viruses with 12- or 16-amino acid deletions in the receptor-binding region of their hemagglutinin protein. *MBio* **12**, e0151221 (2021).
58. Li, Y. *et al.* Single hemagglutinin mutations that alter both antigenicity and receptor binding avidity influence influenza virus antigenic clustering. *J. Virol.* **87**, 9904–9910 (2013).
59. Bloom, J. D., Gong, L. I. & Baltimore, D. Permissive secondary mutations enable the evolution of influenza oseltamivir resistance. *Science* **328**, 1272–1275 (2010).
60. Lee, J. M. *et al.* Mapping person-to-person variation in viral mutations that escape polyclonal serum targeting influenza hemagglutinin. *Elife* **8**, (2019).
61. Bloom, J. D. An experimentally determined evolutionary model dramatically improves phylogenetic fit. *Mol. Biol. Evol.* **31**, 1956–1978 (2014).
62. Doud, M. B., Lee, J. M. & Bloom, J. How single mutations affect viral escape from broad and narrow antibodies to H1 influenza hemagglutinin. *Nat. Commun.* **9**, 1386 (2018).
63. Crawford, K. H. D. & Bloom, J. D. alignparse: A Python package for parsing complex features from high-throughput long-read sequencing. *J. Open Source Softw.* **4**, (2019).
64. Haddox, H. K. *et al.* Jointly modeling deep mutational scans identifies shifted mutational effects among SARS-CoV-2 spike homologs. *bioRxivorg* 2023.07.31.551037 (2023) doi:10.1101/2023.07.31.551037.
65. Meng, E. C. *et al.* UCSF ChimeraX: Tools for structure building and analysis. *Protein Sci.* **32**, e4792 (2023).