A single mutation in dairy cow-associated H5N1 viruses increases receptor binding breadth

Marina R. Good[1*], Wei Ji[1*], Monica L. Fernández-Quintero[2], Andrew B. Ward[2], Jenna J. Guthmiller[1]

[1]Department of Immunology and Microbiology, University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA

[2]Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA 92037, USA

*Contributed equally

Correspondence: jenna.guthmiller@cuanschutz.edu

ABSTRACT

Clade 2.3.4.4b H5N1 is causing an unprecedented outbreak in dairy cows in the United States. To understand if recent H5N1 viruses are changing their receptor use, we screened recombinant hemagglutinin (HA) from historical and recent 2.3.4.4b H5N1 viruses for binding to distinct glycans bearing terminal sialic acids. We found that H5 from A/Texas/37/2024, an isolate from the dairy cow outbreak, has increased binding breadth to glycans bearing terminal α2,3 sialic acids, the avian receptor, compared to historical and recent 2.3.4.4b H5N1 viruses. We did not observe any binding to α2,6 sialic acids, the receptor used by human seasonal influenza viruses. We identified a single mutation outside of the receptor binding site, T199I, was responsible for increased binding breadth, as it increased receptor binding site flexibility. Together, these data show recent H5N1 viruses are evolving increased receptor binding breadth which could impact the host range and cell types infected with H5N1.

INTRODUCTION

Since 2021, clade 2.3.4.4b H5N1 viruses, a highly pathogenic avian influenza virus, have been causing a worldwide outbreak in wild bird populations, with reported cases on six continents. Numerous H5N1 spillover events in domestic animals, including poultry and minks, have led to massive culling events[1,2]. Moreover, H5N1 spillover into wild mammals, including aquatic and scavenger mammals, have been reported since 2022[3]. In March 2024, the United States Department of Agriculture reported an outbreak of H5N1 in domestic dairy cattle[4]. Since then, H5N1 has expanded to 12 states with over 100 farms affected[5]. H5N1 viruses from dairy cows have spilled over into domestic felines, alpacas, poultry, and house mice[6,7]. Importantly, H5N1 viruses from the ongoing outbreak in dairy cattle have led to three confirmed human infections, with two cases causing conjunctivitis and the third case cause mild respiratory symptoms[8,9].

H5N1 infection in dairy cows is largely restricted to the mammary tissue, leading to clinical manifestations of mastitis including reductions in milk production, milk discoloration, and increased milk thickness[10]. Analysis of infectious virus revealed titers ranging from 10^4-10^9 tissue culture infectious dose 50 ($TCID_{50}$)[11,12]. One in five retail milk samples within the United States (US) has detectable virus by PCR, although viable virus has not been recovered from these samples[13]. Moreover, pasteurization is an effective method to kill H5N1 viruses[11,14]. It remains unclear how H5N1 is being transmitted between cows and different hosts, although transmission is linked to raw milk consumption or exposure.

Avian influenza viruses, including H5N1, preferentially bind glycans bearing terminal α2,3 sialic acids[15]. In contrast, influenza viruses that cause seasonal influenza outbreaks in humans prefer glycans bearing terminal α2,6 sialic acids[16]. The influenza virus preference for α2,3 or α2,6 sialic acid linkages creates a major species barrier for avian influenza viruses to spill over into humans. Two recent studies show that dairy cow mammary tissue, and particularly the mammary alveoli, has abundant α2,3 sialic acid linked glycans[17,18]. Moreover, dairy cow mammary tissues also have α2,6 sialic

acid linked glycans[17,18], suggesting dairy cow mammary glands could be a site of viral evolution to adapt H5N1 to human-like receptors.

In this study, we investigated if recent H5N1 viruses are evolving their receptor binding specificities. We identified that H5 from the ongoing dairy cow outbreak has increased binding breadth to backbone glycans bearing α2,3 sialic acids relative to other H5N1 viruses, which was linked to a single mutation near, but not within, the receptor binding site (RBS). I199 emerged in late 2023, before the onset of the ongoing dairy cow outbreak, and is now the dominant amino acid at this residue in North American isolates. Our study indicates a single mutation near the RBS expands the types of backbone gylcans bound by H5, which could imply an increase in cell, tissue and host tropisms.

RESULTS

HA from circulating H5N1 in dairy cows is phylogenetically distinct

The phylogenetic analysis of HA genetic sequences from 94 H5Nx viruses representing various clades and ancestral and human H5N1 sequences revealed distinct branching patterns (**Fig. 1**; **Extended Data Table 1**). Human H5N1 sequences from Vietnam (2004) and Indonesia (2005) were closely related but formed separate branches from the 2.3.4.4 clusters, bridging the evolutionary paths between the ancestral A/Goose/Guangdong/1996 sequence and the 2.3.4.4 clades (**Fig. 1**). The 2.3.4.4 clade is further subdivided into distinct subclusters (2.3.4.4b, 2.3.4.4c, 2.3.4.4e, 2.3.4.4g, 2.3.4.4h), highlighting the diversity and evolutionary progression of H5Nx across various regions. The 2.3.4.4b clade represents the dominant H5N1 viruses globally since 2021[19-21]. 2.3.4.4.b H5N1 viruses segregated based on continent(s), with virus isolates from Eurasia and Africa more closely related to each other than to viruses from North and South America, which was independent of isolation date (**Fig. 1**). Moreover, viruses from North and South America are more closely related to each other than they are to viruses from Eurasia and Africa (**Fig. 1**). Within the Americas branches, the cattle-derived H5N1 viruses form a distinct group within the 2.3.4.4b clade (**Fig. 1**). The new group shows a cluster of H5N1 strains isolated from various hosts in the United States

in 2024, including dairy cows, domestic cats, raccoons, skunks, mountain lions, and several bird species. The clustering of sequences from the recent dairy cow outbreak indicates a common ancestor and suggests a potential transmission link between these species. Together, these data demonstrate that the recent outbreak of H5N1 in dairy cows is distinct from other circulating 2.3.4.4b H5N1 viruses.

Dairy cow-related H5 has increased glycan binding breadth

To understand if recent H5 has changed its receptor binding specificity, we tested recombinant H5 (rH5) from an ancestral H5N1 virus, 2.3.4.4b H5N1 viruses from 2022, and a recent H5 isolated from dairy farm worker (A/Texas/37/2024) on a N-acetylneuraminic (Neu5Ac) and N-glyconeuraminic acid (Neu5Gc) glycan microarray (**Extended Data Table 2**). This microarray includes an array of distinct glycans with terminal sialic acids of both the α2,3 and α2,6 Neu5Ac linkages, which correspond to the receptors for avian and human influenza viruses, respectively. This microarray includes glycans with distinct branches, with most glycans incorporating a single branch with α2,3 or α2,6 Neu5Ac linkage and a second branch of varying lengths and compositions (**Extended Data Table 2**). We observed ancestral rH5 from A/Vietnam/1204/2004 exhibited a dominant preference for α2,3 linked lactosamine glycans (**Fig. 2A**). In contrast, rH5 from A/Colorado/18/2022, which was isolated from a human involved in culling 2.3.4.4b H5N1 infected poultry, exhibited restricted binding to 3' sialyl Lewis X glycans. Other isolates from 2022 2.3.4.4b H5N1 viruses revealed expanded binding breadth to 3' sialyl Lewis X and α2,3 linked lactosamine glycans (**Fig. 2A**). Compared to other 2.3.4.4b rH5s, A/Texas/37/2024 has gained further binding breadth to nearly all α2,3 sialic linked lactosamine glycans, including those with asymmetrical branches (**Fig. 2A**). Moreover, A/Texas/37/2024 had augmented binding signal for 3' sialyl Lewis X glycans relative other 2.3.4.4b H5 viruses and to α2,3 sialic acid-linked lactosamine glycans relative to A/Vietnam/1204/2004 (**Fig. 2A; Extended Data Fig. 1A-E**). Importantly, we did not observe any binding to glycans bearing only terminal α2,6 sialic acids, indicating recent H5N1 viruses have not gained binding affinity to receptors used by human seasonal influenza virus subtypes (**Extended Data Fig. 1A-E**).

To understand the mechanism of how A/Texas/37/2024 rH5 has gained increased binding breadth, we performed molecular dynamics (MD) simulations. Sequences of rH5 from A/Texas/37/2024 and A/Colorado/18/2022 were used to model binding to LSTa, an α2,3 sialic acid avian analog receptor (**Fig. 2B**). We identified similar residues of both A/Colorado/18/2022 and A/Texas/37/2024 involved in LSTa binding (**Fig. 2B**). Additionally, we observed a strong water-mediated, hydrogen bond network of LSTa with both A/Texas/37/2024 and A/Colorado/18/2022. In particular, we found a long-lasting water-mediated hydrogen bond of LSTa with residue E190 (**Fig. 2B**), an important residue for mediating α2,3 sialic acid receptor specificity[22]. A/Texas/37/2024 exhibited more variability in the receptor binding site (RBS), adopting 9 different conformations (**Fig. 2C**). In contrast A/Colorado/18/2022 reveals a more stable RBS with only 3 conformational states (**Fig. 2C**). To quantify the differences in flexibility, we mapped the B-factor obtained from the MD simulations onto the structure of A/Colorado/18/2022 and A/Texas/372/2024 in complex with LSTa. A/Texas/37/2024 exhibited overall a dramatically higher B-factor. This is particularly pronounced in the 190-helix, whereas A/Colorado/18/2022 revealed an overall rather low B-factor (**Fig. 2D**). Thus, our findings suggest that the expanded binding breadth to glycans even bearing terminal α2,3 sialic acids, is mediated by an increased flexibility within the RBS of A/Texas/37/2024.

H5N1 viruses circulating in the Americas have acquired mutations near the RBS

Several mutations have arisen in 2.3.4.4b viruses since 2022, particularly at L111M, T199I, and V214A (**Fig. 3A**). Notably, these three mutations lie outside of the traditional RBS, which is comprised of the 130-loop, 190-helix, and 220-loop[23]. Analysis of these mutations based on location and outbreak revealed that all three mutations are specific to H5N1 viruses in the Americas (**Fig. 3B**). Importantly, we observed T199I was found in all H5N1 viruses from the dairy cow outbreak, as well as some H5N1 viruses circulating in the Americas not related to the ongoing dairy cow outbreak (**Fig. 3B**). T199I is the only amino acid difference between A/Texas/37/2024 and A/pelican/Chile/7087-1/2022 (**Fig. 3C-D**). Importantly, HAs from the ongoing dairy cow outbreak are highly conserved, as the amino acid sequence of A/Texas/37/2024 and A/Michigan/90/2015 are identical (**Fig. 3C**). Structurally, position 199 is located on the backside of the 190-

helix (**Fig. 3A**). We observed that the T199I mutation arose in the second half of 2023, with I199 becoming dominate by November 2023 (**Fig. 3D-E**). Notably, A/Texas/37/2024 HA NT sequence clusters more closely with an HA sequence collected from a mountain lion, also known as a cougar (*Puma concolor*) in Montana than sequences related to the ongoing dairy outbreak (**Fig. 3E**). Interestingly, the mountain lion isolate was collected in January 2024 and it was recently proposed that the dairy cow outbreak has been ongoing since late 2023[24]. These data would suggest a closely related common ancestor from the mountain lion case to the ongoing dairy cow outbreak. Together, these data show that H5N1 viruses in the Americas have accumulated mutations within the RBS, with T199I being the only mutation specific to the ongoing dairy cow outbreak.

T199I is responsible for increased α2,3 sialic acid binding breadth

T199I resides in a loop on the backside of the 190-helix that leads into the 220-loop of the RBS. To determine if T199I augments binding breadth, we reverted A/Texas/37/2024 from I199 to T199 (I199T) and tested glycan binding breadth. A/Texas/37/2024 I199T demonstrated identical binding breadth to A/pelican/Chile/7087-1/2022 (**Fig. 4A-B; Extended Data Figure 1F**), indicating a mutation outside of the RBS massively affects receptor binding specificity. MD simulations predicted A/Texas/37/2024 T199 only has four different conformational states (data not shown). Furthermore, we find that T199 hydroxyl group hydrogen bonds with the amide of N248 on the same protomer, stabilizing the 190-helix and 220-loop (**Fig. 4C**). Thus, the T199I mutation would lose this stabilizing hydrogen bond, leading to more flexibility within the RBS. This additional stabilization of T199I is further emphasized by a more favorable interaction energy of T199 compared to I199 (~-60 kcal/mol to ~-45 kcal/mol). Analysis of the B-factor of A/Texas/37/2024 with T199 shows a decreased flexibility globally and in the 190-helix relative to A/Texas/37/2024 with the naturally occurring I199 (**Fig. 2D** and **Fig. 4D**). These data demonstrate that a single mutation outside the RBS can improve binding breadth binding breadth to distinct backbone glycans bearing terminal α2,3 sialic acids. Mechanistically, we propose that T199 stabilizes the RBS, leading to more restricted receptor binding.

DISCUSSION

Our study shows that H5N1 viruses from the ongoing dairy cow outbreak have increased their receptor binding breadth to bind more glycans bearing α2,3 sialic acids. We observed that A/Texas/37/2024 could bind 3' sialyl Lewis X glycans, which has a fucosylated sialoside, and α2,3 sialic acid-linked lactosamine glycans. We observed a historical H5N1 virus, A/Vietnam/1204/2004, preferentially bound to α2,3 sialic acid-linked lactosamine glycans, whereas A/Colorado/18/2022, the first human case of 2.3.4.4b virus in the US, was highly specific to glycan with a 3' sialyl Lewis X structure. A prior study found A/Vietnam/1194/2004, an isolate closely related to A/Vietnam/1204/2004, binds both 2,3 sialic acid-linked lactosamine glycans and 3' sialyl Lewis X, albeit the former with 3-times stronger affinity[25]. Moreover, an analysis of Asian 2003-2004 H5 isolates from chickens and humans preferred sulfated α2,3 sialic acid-linked glycans, including binding to a sulfated sialyl Lewis X glycan[15]. Avian influenza viruses are known to have restricted sialic acid binding breadth as a mechanism to have specific and limited host tropism[26]. As it stands, our understanding of the glycan structures bearing α2,3 sialic acids on distinct cell types, tissues, and hosts remains poorly understood. Moreover, how host glycosylation patterns impact influenza virus evolution to augment receptor binding affinity and breadth is not well characterized. Thus, a deeper understanding of how glycan binding specificity and breadth across diverse hosts is needed to perform risk assessment of potential pandemic influenza viruses, such as H5Nx.

2.3.4.4 viruses in the mid-2010s gained mutations at positions K222Q and S227R, which increase binding to fucosylated sialosides, such as 3' sialyl Lewis X[27]. K222 sterically clashes with the fucose group on 3' sialyl Lewis X[27], which could explain the selection of mutations at this site that improves binding to fucosylated glycans. Importantly, 2.3.4.4b clade H5N1 viruses have retained K222Q and S227R, which could help explain their preferential binding to 3' sialyl Lewis X. Our data adds T199I to the list of mutations that change receptor binding, by expanding the H5 RBS binding to α2,3 sialic acid-linked lactosamine glycans. Our MD data shows that T199 stabilizes the RBS through hydrogen bonds formed with N248 on the same protomer. Since T199 is within a loop directly following the 190-helix and leads into the 220-loop, these hydrogen

bonds likely stabilize the RBS, limiting the number of conformations it can adopt while binding α2,3 sialic acid-linked glycans.

Our glycan binding data shows that 2.3.4.4b H5N1 viruses, including those related to the ongoing dairy cow outbreak, have not gained binding to α2,6 sialic acids, the most abundant human receptor for influenza viruses, despite the presence of α2,6 sialic acids within the cow mammary glands[17,18]. Two mutations, E190D and G225D, are defined mutations for receptor switches between α2,3 and α2,6 sialic acid-linked glycans[22]. E190 and D190 function as direct α2,3 and α2,6 sialic acid contacts, respectively[22,28], whereas G225D mutation introduces a bulky amino acid within the 220-loop, making binding specific to α2,6 sialic acids[29,30]. While mutations at these two sites are not observed within the circulating 2.3.4.4b, our data support that a mutation near the RBS is impacting receptor binding specificity and breadth. Notably, our data supports that mutations not directly within the RBS can dramatically change receptor binding properties. Deep mutational scanning tools for emerging influenza viruses can provide insight into mutations permissible for increased binding to α2,3 and α2,6 sialic acids[31]. A proactive analysis of H5 sequences and their potential to increase binding breadth and specificity can alert to their potential to cause a pandemic.

Study Limitations

A limitation of our study is that we only used recombinant HA to study HA-glycan interactions, which will be limited to low avidity interactions. As a result, our approach is only detecting high affinity interactions but not low affinity high avidity interactions that would be detected using a viral particle. The glycan microarray did not include sulfated glycans, which are known to be recognized by H5Nx viruses[15]. Sulfated glycans with Neu5Ac may be an important glycan specificity of recent H5N1 viruses. Lastly, the MD data presented depend on predicted structures and serve as models of what may be occurring. Analysis of HA structures with LTSa would confirm structural confirmations taken by emerging H5N1 viruses.

METHODS

Sequence analyses

We downloaded 94 H5Nx sequences (Extended Data Table 2) from different clades (2.3.4.4b, 2.3.4.4c, 2.3.4.4e, 2.3.4.4g, 2.3.4.4h) along with the ancestral H5N1 sequence (A/Goose/Guangdong/1996-01-01), and two human H5N1 sequences from Asia (A/Vietnam/2004 and A/Indonesia/2005) from the GISAID database[32-34]. H5N1 avian influenza virus sequences from the 2.3.4.4b clade in the Americas were aligned using MEGA11[35,36]. The best-fit nucleotide substitution model was identified using MEGA11. A Maximum Clade Credibility (MCC) tree was constructed with BEAST v2.6.3, using the TN93+Gamma5 substitution model and partitioning by positions 1, 2, and 3[37]. The analysis employed an uncorrelated relaxed clock with a chain length of 10,000,000 generations, sampling every 1,000 generations, and discarding 10% of the samples as burn-in. The resulting file was analyzed and annotated using Tracer v1.7.1 and TreeAnnotator v1.10.4, and the annotated MCC tree was visualized using FigTree v1.4.4[38,39]. The mammal symbol on the tree originates from the Pixabay website and BioRender. Weblogo plots were generated as previously described[40].

Cloning and protein purification

HA sequences were downloaded from GISAID. HA ectodomains were synthesized from Integrated DNA Technologies (IDT) or Twist Biosciences and cloned into a vector with a Fibritin Foldon Domain and a his-tag. PCR-based site-directed mutagenesis was used to introduce I199T into the A/Texas/37/2024 construct. PCR reactions with mutagenesis primers were performed PrimeSTAR Max DNA Polymerase (Takara). The PCR product was treated with DpnI (New England Biolabs). All plasmids were transformed into *E. coli* New England Biolabs), miniprepped, correct clones selected, and maxiprepped. Sequence verified maxipreps were used for transfections in HEK293T cells (ATCC) or Expi293F Cells (Thermofisher). Expi293F suspension cells were maintained at 125 rpm at 37°C with 8% CO_2 in FreeStyle™293 expression medium (Gibco). HEK293T cells were grown in 37°C with 5% CO_2. HAs were produced in-house via transfection of HEK293T cells or Expi293F cells. HAs were purified from the supernatant using nickel-NTA agarose (Qiagen) and disposable 10mL polypropylene columns (Thermofisher). HA

concentrations were determined using a Pierce BCA Protein Assay (Thermofisher). HA was aliquoted and maintained at -80°C.

Neu5Ac and Neu5Gc glycan microarray

We used a Neu5Ac and Neu5Gc glycan microarray (Zbiotech; Lot # 04242301); structures can be found here (https://www.zbiotech.com/product/neu5gc-neu5ac-n-glycan-microarray/). HA proteins were diluted into 200µL of glycan microarray assay buffer (GAAB) supplemented with 1% BSA, for a final concentration 40, 20 or 10 µg/mL of HA. Subsequently, an anti-6x His tag antibody and anti-rabbit immunoglobulin (H+L) (Cy3) antibody were added to the HA + GAAB mixture at a final concentration of 3.2 µg/mL each. The entire mixture was then incubated at room temperature for 60 minutes with gentle vortexing as part of the precomplexing process. To prepare the microarray slide for analysis, the slide was pretreated with glycan microarray blocking buffer (GABB) supplemented with 1% BSA at room temperature for 60 minutes. Following this, the precomplexed HA protein samples were added to the microarray, with 100µL added to each submicroarray. The slide was incubated for 60 minutes at room temperature to facilitate binding interactions. After this incubation period, the slide was thoroughly washed to remove any unbound components. The slide was then scanned at 532 nm using high intensity (1 PMT) to detect and visualize any interactions. Innopsys' Mapix software was used to analyze the microarray scans. Positive binding signals were determined by subtracting the background and negative control signals from all experimental sample signals.

MD simulations

The starting structures of A/Colorado/18/2022, A/Texas/37/2024, and A/Texas/37/2024 I199T, were predicted using *ColabFold*, which increased the accessibility of protein structure prediction tools by combining *AF2* with the rapid homology search capability of MMseqs2, making it an easy to use and fast software (~90-fold speed up in prediction) to predict homo-and heteromeric complexes, matching the prediction quality of AF2 and AF-multimer[41,42]. As template to model the sialic acid complex (LSTa) we used the available X-ray structure of A/California/04/2009 in complex with LSTa (PDB ID: 3UBJ).

In addition, we used the GlycoShape tool to ensure the known glycosylation sites are glycosylated for the simulations[43]. For our simulations we capped the C-terminal and N-terminal parts of each domain with acetylamide and N-methylamide to avoid perturbations by free charged functional groups. For each H5 variant, we performed three repetitions of 500 ns of classical molecular dynamics simulations using the AMBER 22 simulation software package which contains the pmemd.cuda module[44]. The structures were prepared using CHARMM-GUI[45,46]. The structure models were placed into cubic water boxes of TIP3P water molecules[47] with a minimum wall distance to the protein of 12 Å[48,49]. Parameters for all simulations were derived from the AMBER force field 14SB[50,51]. To neutralize the charges, we used uniform background charges[52-54]. Each system was carefully equilibrated using a multistep equilibration protocol[55]. Bonds involving hydrogen atoms were restrained using the SHAKE algorithm, allowing a timestep of 2.0 femtoseconds[56]. The systems' pressure was maintained at 1 bar by applying weak coupling to an external bath using the Berendsen algorithm[57]. The Langevin Thermostat was utilized to keep the temperature at 300K during the simulations[58].

MD analysis

For all investigated H5 variants, we calculated the respective contacts of the HA protomers with LSTa in solution using the GetContacts software (Stanford University; https://getcontacts.github.io/). This tool can compute interactions within one protein structure, but also between different protein interfaces and allows to monitor the evolution of contacts during the simulation. Apart from visualizing and quantifying the contacts of the different poses, we calculated the residue-wise B-factor, as measure of global flexibility implemented in cpptraj[59] to identify differences in the conformational diversity between the HA variants. We used PyMOL to visualize protein structures (PyMOL - The PyMOL Molecular Graphics System, Version 3.0 Schrödinger, LLC).

HA modeling

The protein structure 7DEA, from A/duck Northern China/22/2017 (H5N6), was retrieved from the Protein Data Bank (PDB) and visualized using PyMOL (Version 2.6,

Schrödinger, LLC). All numbering in this manuscript is H3-numbering, based on Burke and Smith[60].

ACKNOWLEDGEMENTS

We gratefully acknowledge all data contributors, including the authors and their originating laboratories responsible for obtaining the specimens, and their submitting laboratories for generating the genetic sequence and metadata and sharing via the GISAID Initiative, on which this research is based. We also thank Jian Zheng, Xi Chen, and Nick Berning of Z Biotech for their discussion and suggestions on glycan analyses. We thank Julianna Han for her critical feedback on data interpretation.

FUNDING

This project was funded in part by National Institute of Allergies and Infectious Diseases (NIAID) Centers of Excellence in Influenza Research and Response (CEIRR) grant #75N93021C00045 (JJG) and the Collaborative Influenza Vaccine Innovation Centers (CIVIC) Grant #75N93019C00051 (JJG and ABW). This work was supported in part by The Howard Hughes Medical Institute (HHMI) Emerging Pathogens Initiative (JJG) and the American Heart Association Grant #24PRE1189305.

AUTHOR CONTRIBUTIONS

Conceptualization: MRG, WJ, JJG; Methodology: MRG, WJ, MLF-Q, JJG; Investigation: MRG, WJ, MLF-Q, JJG; Visualization: MRG, WJ, MLF-Q, JJG; Funding acquisition: JJG and ABW; Project administration: JJG; Supervision: JJG and ABW; Writing – original draft: MRG, WJ, JJG; Writing – review & editing: MRG, WJ, MLF-Q, ABW.

COMPETING INTERESTS

The authors have no competing interests to declare.

Fig. 1: Phylogenetic tree of highly pathogenic avian H5N1. The Neighbor-Joining (NJ) phylogenetic analysis of 94 hemagglutinin gene sequences. Clade 2.3.4.4 is subdivided into distinct subclades, including 2.3.4.4e, 2.3.4.4h, 2.3.4.4g, 2.3.4.4c, and the currently dominant 2.3.4.4b clade. The new group belonging to clade 2.3.4.4b includes strains isolated from domestic dairy cows and humans and animals linked to H5N1-positive dairy farms (highlighted with a blue region and animal symbols) in the United States in 2024. Distinct clades of the virus are labeled on the right side of the figure. Tips are labeled with H5Nx strain names, host species, and isolation dates. Nodes represent inferred common ancestors of the grouped tips. Branch lengths are proportional to the number of nucleotide substitutions per site, indicating the divergence between nodes.

A

	SA1	SA5	SA11	SA17	SA23	SA37	SA41	Sialyl Lewis X: SA9	SA15	SA21	SA25	SA29	SA35
Vietnam 2004	✓ 3636.0 ± 209.9	✓ 286.0 ± 68.7	✓ 854.8 ± 270.2	✓ 1282.4 ± 337.2	✓ 269.1 ± 53.1								
Colorado 2022								✓ 1222.6 ± 458.9	✓ 1440.3 ± 73.2	✓ 6993.5 ± 1572.6	✓ 6710.5 ± 629.4	✓ 1158.4 ± 160.3	✓ 490.8 ± 85.6
Mink/Spain 2022	✓ 1235.4 ± 82.9			✓ 998.4 ± 126.9	✓ 181.3 ± 82.2			✓ 5283.1 ± 403.9	✓ 3765.8 ± 832.5	✓ 9320.9 ± 235.1	✓ 8601 ± 141.9	✓ 4801.8 ± 422.5	✓ 1041.4 ± 311.2
Pelican/Chile 2022	✓ 708.0 ± 226.0							✓ 3632.8 ± 2411.4	✓ 5545.4 ± 529.4	✓ 8402.1 ± 532.5	✓ 8121.3 ± 213.6	✓ 5164.6 ± 1179.1	✓ 2624.9 ± 258.7
Texas 2024	✓ 7675.1 ± 174.6		✓ 1664.4 ± 115.3	✓ 5922.9 ± 952.0	✓ 3668 ± 631.6	✓ 2130 ± 199.6	✓ 2046.6 ± 185.8	✓ 15998.6 ± 378.0	✓ 12085.3 ± 2049.1	✓ 19928.8 ± 582.9	✓ 21530.8 ± 650.6	✓ 13545.1 ± 1171.9	✓ 9007.6 ± 625.3

Symbols: Man, GlcNAc, Gal, L-Fuc, Neu5Ac, Neu5Gc

B

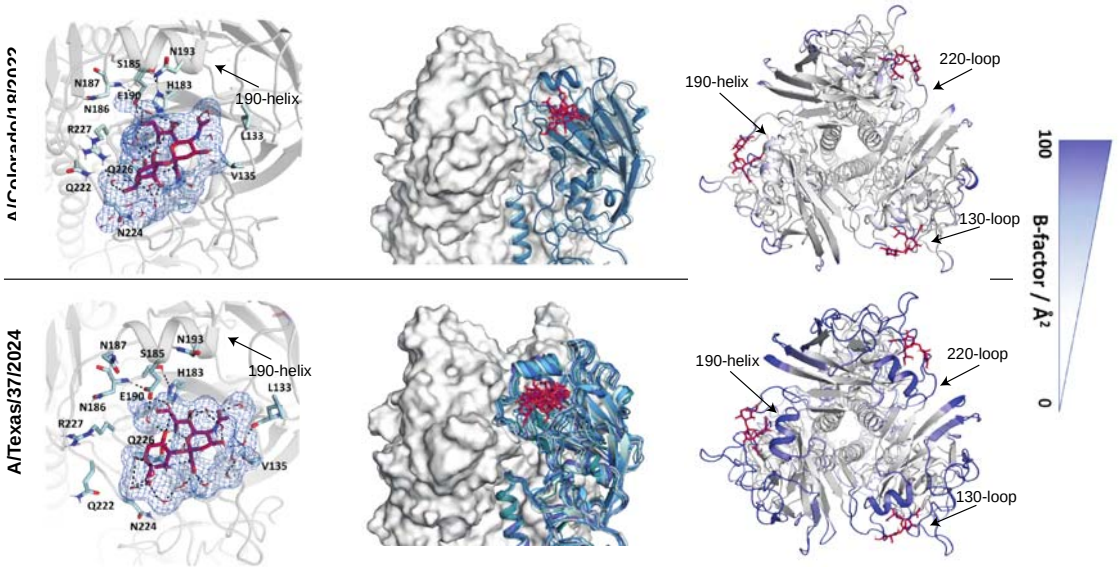


Fig. 2: A dairy cow associated H5N1 virus exhibits increased glycan binding breadth. (A) rH5 binding to distinct Neu5Ac glycans. Green checkmarks indicate a positive binding result for the corresponding glycan. Normalized relative fluorescence unit (RFU) values ± standard deviation are indicated below checkmarks. Value above each glycan indicates the glycan number in the array. (B-D) MD simulations of A/Colorado/18/2022 and A/Texas/37/2024 to characterize the LSTa binding site properties. (B) Representative structure obtained from MD simulations showing interactions of LSTa (burgundy) with A/Colorado/18/2022 and A/Texas/37/2024 in solution. (C) Conformational states of A/Colorado/18/2022 and A/Texas/37/2024 binding to LSTa. Each shade of blue represents a distinct confirmation. (D) Residue-wise B-factor, as a measure of flexibility, mapped on the respective A/Colorado/18/2022 and A/Texas/37/2024 structure.

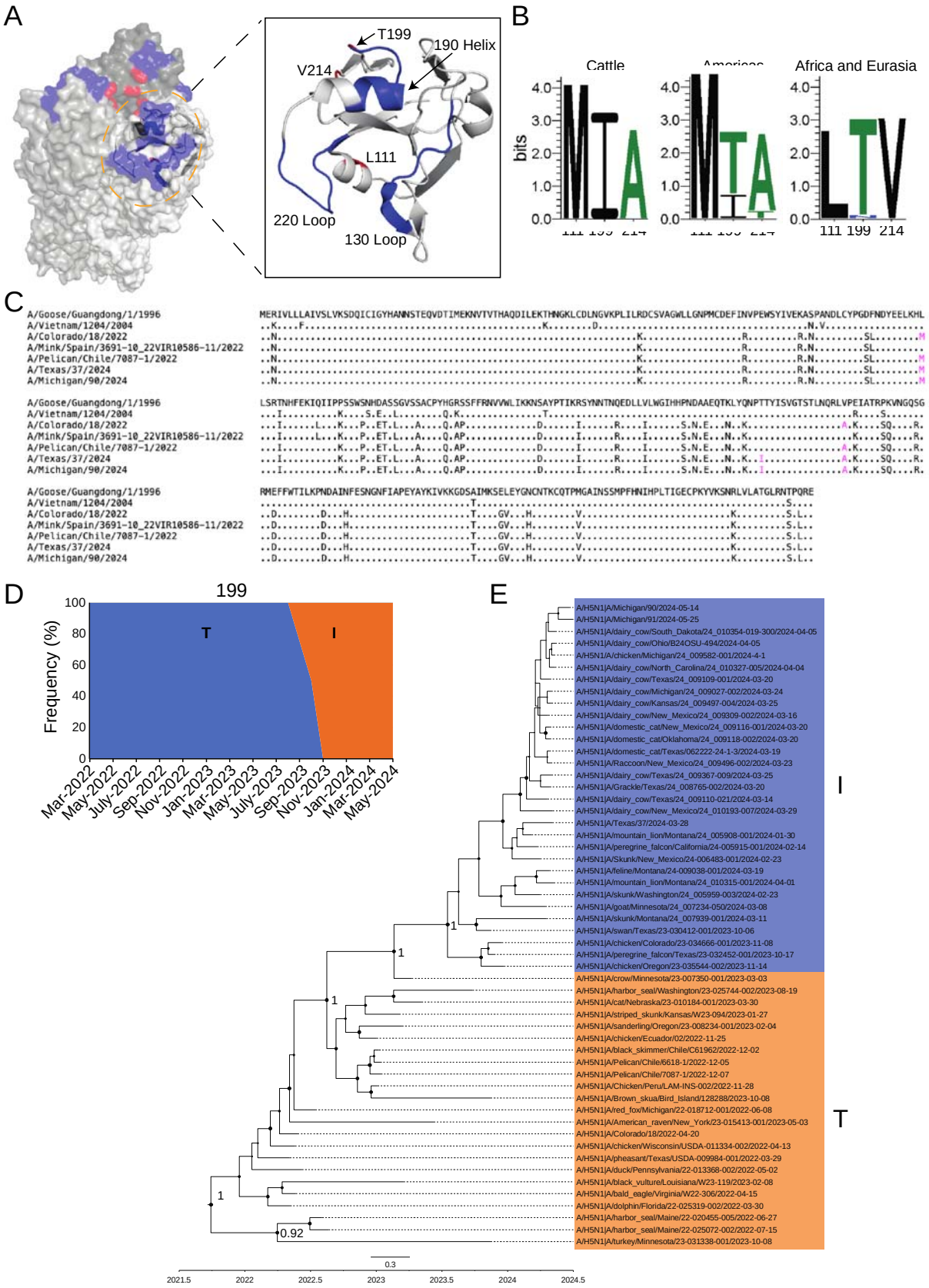


Fig. 3: 2.3.4.4b H5N1 viruses in the Americas recently acquired T199I. (A) Structural depiction on A/duck/Northern China/22/2017 H5 (PDB: 7DEA) of the RBS and recent mutations. Blue residues indicate those found within the 130-loop, 190-helix, or 220-loop. Red residues indicate mutations of interest. (B) Logo plots of positions 111, 199, and 214 based on geographical location. “Americas” logo plot does not include sequences from the dairy cow outbreak. (C) Amino acid alignment of HA1 from H5N1 viruses in this study. Residues in magenta are positions, 111, 199, and 214. (D) Frequency of T199 (blue) and I199 (orange) in circulating 2.3.4.4b H5N1 viruses in the Americas, including the dairy cow outbreak, between March 2022 and May 2024. (E) Maximum Clade Credibility tree of 2.3.4.4b clade H5N1 viruses in the Americas with T199 or I199 from 2022 to 2024. Posterior probabilities were marked with black dots at the nodes, with the main clades labeled by number. The size of the dots corresponds to the posterior probability values. The larger the black dot, the higher the value it represents. The scale bar at the bottom represents 0.3 substitutions per site.

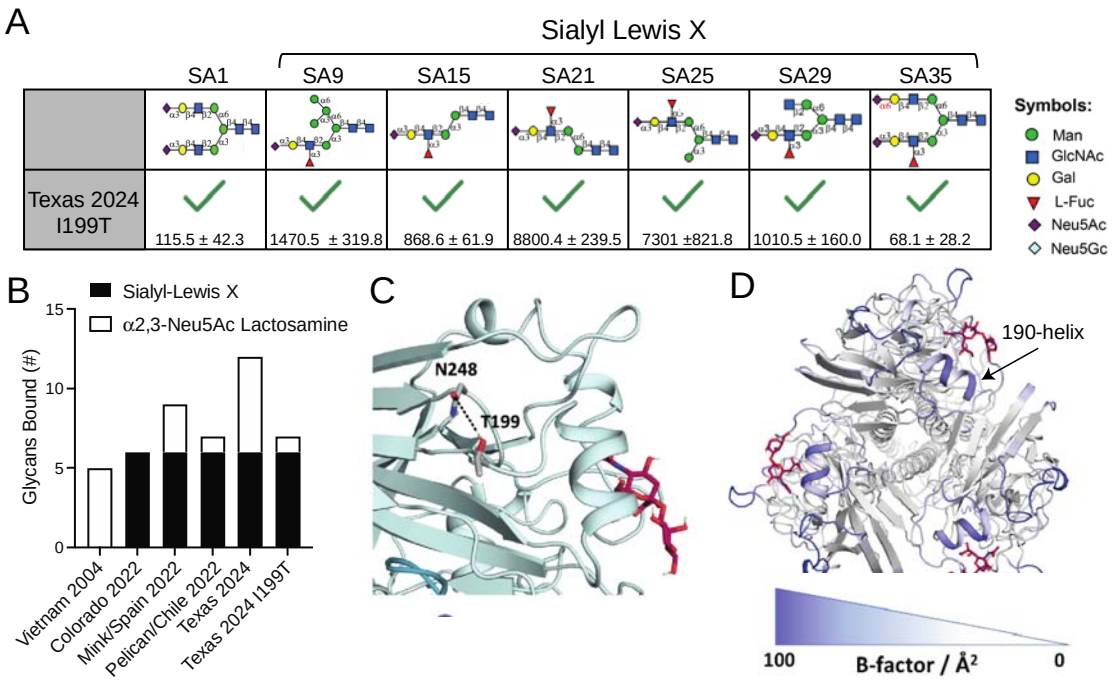
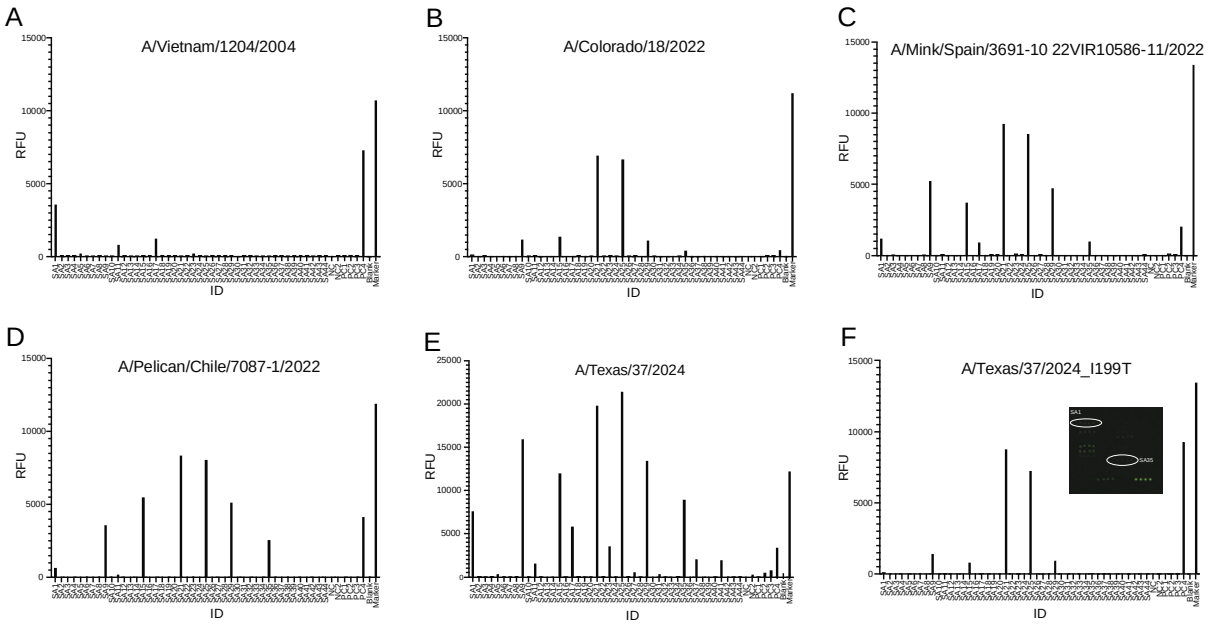


Fig. 4: T199I is responsible for increased glycan binding breadth in A/Texas/37/2024. (A) A/Texas/37/2024 with an I199T mutation binding to distinct Neu5Ac glycans. Green checkmarks indicate a positive binding result for the corresponding glycan. Normalized relative fluorescence unit (RFU) values ± standard deviation are indicated below checkmarks. Only glycans above the background signal are shown. (B) Number and type of glycans bound by each H5. (C) Hydrogen bond analysis showing the stabilizing role of I199T with N248. (D) B-factor analysis of A/Texas/37/2024 I199T.



Extended Data Figure 1: rHA binding to glycan microarray. (A-F) rH5 binding RFUs for each glycan on the microarray. rH5s tested are A/Vietnam/1204/2024 (A), A/Colorado/18/2022 (B), A/Mink/Spain/3691-1022VIR10586-11/2022 (C), A/Pelican/Chile/7087-1/2022 (D), A/Texas/37/2024 (E), and A/Texas/37/2024 I199T (F). Microarray fluorescence is shown in F, with SA1 and SA35 circled to indicate detection of fluorescence despite low RFU values.

418 **Extended Data Table 1**: The HA sequences used for the phylogenetic tree analysis.

419 **Extended Data Table S2:** Glycans in the Neu5Ac and Neu5Gc microarray.

420
421

REFERENCES

1 Aguero, M. *et al.* Highly pathogenic avian influenza A(H5N1) virus infection in farmed minks, Spain, October 2022. *Euro Surveill* **28**, doi:10.2807/1560-7917.ES.2023.28.3.2300001 (2023).

2 Lindh, E. *et al.* Highly pathogenic avian influenza A(H5N1) virus infection on multiple fur farms in the South and Central Ostrobothnia regions of Finland, July 2023. *Euro Surveill* **28**, doi:10.2807/1560-7917.ES.2023.28.31.2300400 (2023).

3 European Food Safety, A. *et al.* Avian influenza overview December 2022 - March 2023. *EFSA J* **21**, e07917, doi:10.2903/j.efsa.2023.7917 (2023).

4 Agriculture, U. S. D. o. (United States Department of Agriculture, 2024).

5 Agriculture, U. S. D. o. *Highly Pathogenic Avian Influenza (HPAI) Detections in Livestock*, <https://www.aphis.usda.gov/livestock-poultry-disease/avian/avian-influenza/hpai-detections/livestock> (2024).

6 Burrough, E. R. *et al.* Highly Pathogenic Avian Influenza A(H5N1) Clade 2.3.4.4b Virus Infection in Domestic Dairy Cattle and Cats, United States, 2024. *Emerg Infect Dis* **30**, doi:10.3201/eid3007.240508 (2024).

7 Agriculture, U. S. D. o. (United States Department of Agriculture, 2024).

8 Uyeki, T. M. *et al.* Highly Pathogenic Avian Influenza A(H5N1) Virus Infection in a Dairy Farm Worker. *N Engl J Med* **390**, 2028-2029, doi:10.1056/NEJMc2405371 (2024).

9 Prevention, C. f. D. C. a. *Technical Report: June 2024 Highly Pathogenic Avian Influenza A(H5N1) Viruses*, <https://www.cdc.gov/bird-flu/php/technical-report/h5n1-06052024.html> (2024).

10 Garg, S. *et al.* Outbreak of Highly Pathogenic Avian Influenza A(H5N1) Viruses in U.S. Dairy Cattle and Detection of Two Human Cases - United States, 2024. *MMWR Morb Mortal Wkly Rep* **73**, 501-505, doi:10.15585/mmwr.mm7321e1 (2024).

11 Guan, L. *et al.* Cow's Milk Containing Avian Influenza A(H5N1) Virus - Heat Inactivation and Infectivity in Mice. *N Engl J Med*, doi:10.1056/NEJMc2405495 (2024).

12 Caserta, L. C. *et al.* From birds to mammals: spillover of highly pathogenic avian influenza H5N1 virus to dairy cattle led to efficient intra- and interspecies transmission. *bioRxiv*, 2024.2005.2022.595317, doi:10.1101/2024.05.22.595317 (2024).

13 Spackman, E. *et al.* Characterization of highly pathogenic avian influenza virus in retail dairy products in the US. *medRxiv*, 2024.2005.2021.24307706, doi:10.1101/2024.05.21.24307706 (2024).

14 Schafers, J. *et al.* Pasteurisation temperatures effectively inactivate influenza A viruses in milk. *medRxiv*, 2024.2005.2030.24308212, doi:10.1101/2024.05.30.24308212 (2024).

15 Gambaryan, A. *et al.* Evolution of the receptor binding phenotype of influenza A (H5) viruses. *Virology* **344**, 432-438, doi:10.1016/j.virol.2005.08.035 (2006).

16 Connor, R. J., Kawaoka, Y., Webster, R. G. & Paulson, J. C. Receptor specificity in human, avian, and equine H2 and H3 influenza virus isolates. *Virology* **205**, 17-23, doi:10.1006/viro.1994.1615 (1994).

17 Kristensen, C., Jensen, H. E., Trebbien, R., Webby, R. J. & Larsen, L. E. The avian and human influenza A virus receptors sialic acid (SA)-α2,3 and SA-α2,6 are widely expressed in the bovine mammary gland. *bioRxiv*, 2024.2005.2003.592326, doi:10.1101/2024.05.03.592326 (2024).

18 Ríos Carrasco, M., Gröne, A., van den Brand, J. M. A. & de Vries, R. P. The mammary glands of cows abundantly display receptors for circulating avian H5 viruses. *bioRxiv*, 2024.2005.2024.595667, doi:10.1101/2024.05.24.595667 (2024).

19 Bevins, S. N. *et al.* Intercontinental Movement of Highly Pathogenic Avian Influenza A(H5N1) Clade 2.3.4.4 Virus to the United States, 2021. *Emerg Infect Dis* **28**, 1006-1011, doi:10.3201/eid2805.220318 (2022).

20 Nagy, A., Cernikova, L. & Stara, M. A new clade 2.3.4.4b H5N1 highly pathogenic avian influenza genotype detected in Europe in 2021. *Arch Virol* **167**, 1455-1459, doi:10.1007/s00705-022-05442-6 (2022).

21 Sagong, M. *et al.* Emergence of clade 2.3.4.4b novel reassortant H5N1 high pathogenicity avian influenza virus in South Korea during late 2021. *Transbound Emerg Dis* **69**, e3255-e3260, doi:10.1111/tbed.14551 (2022).

22 Stevens, J. *et al.* Structure and receptor specificity of the hemagglutinin from an H5N1 influenza virus. *Science* **312**, 404-410, doi:10.1126/science.1124513 (2006).

23 de Graaf, M. & Fouchier, R. A. Role of receptor binding specificity in influenza A virus transmission and pathogenesis. *The EMBO journal* **33**, 823-841 (2014).

24 Nguyen, T.-Q. *et al.* Emergence and interstate spread of highly pathogenic avian influenza A(H5N1) in dairy cattle. *bioRxiv*, 2024.2005.2001.591751, doi:10.1101/2024.05.01.591751 (2024).

25 Xiong, X. *et al.* Recognition of sulphated and fucosylated receptor sialosides by A/Vietnam/1194/2004 (H5N1) influenza virus. *Virus Res* **178**, 12-14, doi:10.1016/j.virusres.2013.08.007 (2013).

26 Gambaryan, A. *et al.* Receptor specificity of influenza viruses from birds and mammals: new data on involvement of the inner fragments of the carbohydrate chain. *Virology* **334**, 276-283, doi:10.1016/j.virol.2005.02.003 (2005).

27 Guo, H. *et al.* Highly Pathogenic Influenza A(H5Nx) Viruses with Altered H5 Receptor-Binding Specificity. *Emerg Infect Dis* **23**, 220-231, doi:10.3201/eid2302.161072 (2017).

28 Xu, R., McBride, R., Nycholat, C. M., Paulson, J. C. & Wilson, I. A. Structural characterization of the hemagglutinin receptor specificity from the 2009 H1N1 influenza pandemic. *J Virol* **86**, 982-990, doi:10.1128/JVI.06322-11 (2012).

29 Stevens, J. *et al.* Glycan microarray analysis of the hemagglutinins from modern and pandemic influenza viruses reveals different receptor specificities. *J Mol Biol* **355**, 1143-1155, doi:10.1016/j.jmb.2005.11.002 (2006).

30 Gamblin, S. J. *et al.* The structure and receptor binding properties of the 1918 influenza hemagglutinin. *Science* **303**, 1838-1842, doi:10.1126/science.1093155 (2004).

31 Dadonaite, B. *et al.* Deep mutational scanning of H5 hemagglutinin to inform influenza virus surveillance. *bioRxiv*, 2024.2005.2023.595634, doi:10.1101/2024.05.23.595634 (2024).

32 Khare, S. *et al.* GISAID's role in pandemic response. *China CDC weekly* **3**, 1049 (2021).
33 Elbe, S. & Buckland-Merrett, G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Global challenges* **1**, 33-46 (2017).
34 Shu, Y. & McCauley, J. GISAID: Global initiative on sharing all influenza data–from vision to reality. *Eurosurveillance* **22**, 30494 (2017).
35 Nei, M. & Kumar, S. Molecular Evolution and Phylogenetics Oxford Univ. *Press, New York* (2000).
36 Tamura, K., Stecher, G. & Kumar, S. MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* **38**, 3022-3027 (2021).
37 Drummond, A. J., Suchard, M. A., Xie, D. & Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol. Biol. Evol.* **29**, 1969-1973 (2012).
38 Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Syst. Biol.* **67**, 901-904 (2018).
39 Rambaut, A. FigTree–Tree Figure Drawing Tool Version v. 1.4. 4. *Institute of Evolutionary Biology, University of Edinburgh: Edinburgh* (2018).
40 Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res* **14**, 1188-1190, doi:10.1101/gr.849004 (2004).
41 Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat Methods* **19**, 679-682, doi:10.1038/s41592-022-01488-1 (2022).
42 Tunyasuvunakool, K. *et al.* Highly accurate protein structure prediction for the human proteome. *Nature* **596**, 590-596, doi:10.1038/s41586-021-03828-1 (2021).
43 Ives, C. M. *et al.* Restoring Protein Glycosylation with GlycoShape. *bioRxiv*, 2023.2012.2011.571101, doi:10.1101/2023.12.11.571101 (2023).
44 D.A. Case, H. M. A., K. Belfon, I.Y. Ben-Shalom, J.T. Berryman, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, G.A. Cisneros, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G. Giambasu, M.K. Gilson, H. Gohlke, A.W. Goetz, R. Harris, S. Izadi, S.A. Izmailov, K. Kasavajhala, M.C. Kaymak, E. King, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, M. Machado, V. Man, M. Manathunga, K.M. Merz, Y. Miao, O. Mikhailovskii, G. Monard, H. Nguyen, K.A. O'Hearn, A. Onufriev, F. Pan, S. Pantano, R. Qi, A. Rahnamoun, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, A. Shajan, J. Shen, C.L. Simmerling, N.R. Skrynnikov, J. Smith, J. Swails, R.C. Walker, J. Wang, J. Wang, H. Wei, R.M. Wolf, X. Wu, Y. Xiong, Y. Xue, D.M. York, S. Zhao, and P.A. Kollman. Amber 2022. *University of California San Francisco* (2022).
45 Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J Comput Chem* **29**, 1859-1865, doi:10.1002/jcc.20945 (2008).
46 Lee, J. *et al.* CHARMM-GUI supports the Amber force fields. *J Chem Phys* **153**, 035103, doi:10.1063/5.0012280 (2020).
47 Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **79**, 926-935, doi:10.1063/1.445869 (1983).

48 El Hage, K., Hedin, F., Gupta, P. K., Meuwly, M. & Karplus, M. Valid molecular dynamics simulations of human hemoglobin require a surprisingly large box size. *Elife* **7**, doi:10.7554/eLife.35560 (2018).

49 Gapsys, V. & de Groot, B. L. Comment on 'Valid molecular dynamics simulations of human hemoglobin require a surprisingly large box size'. *Elife* **8**, doi:10.7554/eLife.44718 (2019).

50 Cornell, W. D. *et al.* A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *Journal of the American Chemical Society* **117**, 5179-5197, doi:10.1021/ja00124a002 (1995).

51 Maier, J. A. *et al.* ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput* **11**, 3696-3713, doi:10.1021/acs.jctc.5b00255 (2015).

52 Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **98**, 10089-10092, doi:10.1063/1.464397 (1993).

53 Salomon-Ferrer, R., Case, D. A. & Walker, R. C. An overview of the Amber biomolecular simulation package. *WIREs Computational Molecular Science* **3**, 198-210, doi:https://doi.org/10.1002/wcms.1121 (2013).

54 Hub, J. S., de Groot, B. L., Grubmuller, H. & Groenhof, G. Quantifying Artifacts in Ewald Simulations of Inhomogeneous Systems with a Net Charge. *J Chem Theory Comput* **10**, 381-390, doi:10.1021/ct400626b (2014).

55 Wallnoefer, H. G., Handschuh, S., Liedl, K. R. & Fox, T. Stabilizing of a globular protein by a highly complex water network: a molecular dynamics simulation study on factor Xa. *J Phys Chem B* **114**, 7405-7412, doi:10.1021/jp101654g (2010).

56 Andersen, H. C. Rattle: A “velocity” version of the shake algorithm for molecular dynamics calculations. *Journal of Computational Physics* **52**, 24-34, doi:https://doi.org/10.1016/0021-9991(83)90014-1 (1983).

57 Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **81**, 3684-3690, doi:10.1063/1.448118 (1984).

58 Adelman, S. A. & Doll, J. D. Generalized Langevin equation approach for atom/solid-surface scattering: General formulation for classical scattering off harmonic solids. *The Journal of Chemical Physics* **64**, 2375-2388, doi:10.1063/1.432526 (1976).

59 Roe, D. R. & Cheatham, T. E., 3rd. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J Chem Theory Comput* **9**, 3084-3095, doi:10.1021/ct400341p (2013).

60 Burke, D. F. & Smith, D. J. A recommended numbering scheme for influenza A HA subtypes. *PLoS One* **9**, e112302, doi:10.1371/journal.pone.0112302 (2014).