



## OPEN ACCESS

## EDITED BY

Mohamad S. Hakim,  
Gadjah Mada University, Indonesia

## REVIEWED BY

Nongluk Sriwilajaroen,  
Thammasat University, Thailand  
Yanbing Li,  
Chinese Academy of Agricultural Sciences,  
China

## \*CORRESPONDENCE

Mikhail Matrosovich  
✉ m.matrosovich@gmail.com  
Stefano Elli  
✉ elli@ronzoni.it  
Sergei Kosakovsky Pond  
✉ spond@temple.edu

RECEIVED 07 October 2023

ACCEPTED 29 November 2023

PUBLISHED 19 December 2023

## CITATION

Elli S, Raffaini G, Guerrini M,  
Kosakovsky Pond S and Matrosovich M (2023)  
Molecular modeling and phylogenetic analyses  
highlight the role of amino acid 347 of the N1  
subtype neuraminidase in influenza virus host  
range and interspecies adaptation.  
*Front. Microbiol.* 14:1309156.  
doi: 10.3389/fmicb.2023.1309156

## COPYRIGHT

© 2023 Elli, Raffaini, Guerrini, Kosakovsky Pond  
and Matrosovich. This is an open-access article  
distributed under the terms of the [Creative  
Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The  
use, distribution or reproduction in other  
forums is permitted, provided the original  
author(s) and the copyright owner(s) are  
credited and that the original publication in this  
journal is cited, in accordance with accepted  
academic practice. No use, distribution or  
reproduction is permitted which does not  
comply with these terms.

# Molecular modeling and phylogenetic analyses highlight the role of amino acid 347 of the N1 subtype neuraminidase in influenza virus host range and interspecies adaptation

Stefano Elli<sup>1\*</sup>, Giuseppina Raffaini<sup>2</sup>, Marco Guerrini<sup>1</sup>,  
Sergei Kosakovsky Pond<sup>3\*</sup> and Mikhail Matrosovich<sup>4\*</sup>

<sup>1</sup>Istituto di Ricerche Chimiche e Biochimiche 'G. Ronzoni', Milan, Italy, <sup>2</sup>Department of Chemistry, Materials, and Chemical Engineering "Giulio Natta", Politecnico di Milano, Milan, Italy, <sup>3</sup>Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, PA, United States, <sup>4</sup>Institute of Virology, Philipps University, Marburg, Germany

The N1 neuraminidases (NAs) of avian and pandemic human influenza viruses contain tyrosine and asparagine, respectively, at position 347 on the rim of the catalytic site; the biological significance of this difference is not clear. Here, we used molecular dynamics simulation to model the effects of amino acid 347 on N1 NA interactions with sialyllacto-N-tetraoses 6'SLN-LC and 3'SLN-LC, which represent NA substrates in humans and birds, respectively. Our analysis predicted that Y347 plays an important role in the NA preference for the avian-type substrates. The Y347N substitution facilitates hydrolysis of human-type substrates by resolving steric conflicts of the Neu5Ac2-6Gal moiety with the bulky side chain of Y347, decreasing the free energy of substrate binding, and increasing the solvation of the Neu5Ac2-6Gal bond. Y347 was conserved in all N1 NA sequences of avian influenza viruses in the GISAID EpiFlu database with two exceptions. First, the Y347F substitution was present in the NA of a specific H6N1 poultry virus lineage and was associated with the substitutions G228S and/or E190V/L in the receptor-binding site (RBS) of the hemagglutinin (HA). Second, the highly pathogenic avian H5N1 viruses of the Gs/Gd lineage contained sporadic variants with the NA substitutions Y347H/D, which were frequently associated with substitutions in the HA RBS. The Y347N substitution occurred following the introductions of avian precursors into humans and pigs with N/D347 conserved during virus circulation in these hosts. Comparative evolutionary analysis of site 347 revealed episodic positive selection across the entire tree and negative selection within most host-specific groups of viruses, suggesting that substitutions at NA position 347 occurred during host switches and remained under pervasive purifying selection thereafter. Our results elucidate the role of amino acid 347 in NA recognition of sialoglycan substrates and emphasize the significance of substitutions at position 347 as a marker of host range and adaptive evolution of influenza viruses.

## KEYWORDS

influenza, neuraminidase, substrate specificity, MD simulation, natural selection, H5N1

## 1 Introduction

Wild aquatic birds are a primary natural reservoir of influenza A viruses (IVs). Transmission and occasional adaptation of aquatic bird viruses to other avian and mammalian species, with or without gene reassortment and/or additional host-switching events, led to the formation of the known variety of HxNy host-specific lineages of IVs, such as H5N1, H7N9 and H9N2 poultry IVs, H1N1 “classical” and “avian-like” swine IVs, H3N8 equine IVs, H3N8 and H3N2 canine IVs. Infrequent transmissions of animal IVs to humans result in isolated zoonotic infections; on very rare occasions, zoonotic IVs can adapt for efficient human-to-human transmission, initiate global pandemics, and continue to circulate in humans causing seasonal influenza disease [for recent reviews, see [Krammer et al. \(2018\)](#) and [Liu et al. \(2022\)](#)].

Interspecies transmission of IVs is usually accompanied by adaptive changes in their genomes required for improved viral fitness in the new host species. Adaptive changes in the receptor-binding properties of IVs and underlying amino acid substitutions in the attachment protein HA have been relatively well studied [reviewed by [Matrosovich et al. \(2006\)](#), [Shi et al. \(2014\)](#), [Thompson and Paulson \(2020\)](#), and [Liu et al. \(2023\)](#)]. Avian-to-human and avian-to-swine adaptation was shown to require a change in HA receptor-binding specificity from preferential recognition of Neu5Ac2–3Gal-terminated glycans (“avian-type” receptors) expressed on avian intestinal target cells to preferential recognition of Neu5Ac2–6Gal-terminated glycans (“human-type” receptors) expressed on target cells of airway epithelium in humans and pigs. Changes in receptor specificity were mediated by amino acid substitutions at the conserved positions of the avian HA, such as G228S and/or Q226L (the H3 numbering system is used throughout the text) in the case of H2N2/1957 and H3N2/1968 pandemic IVs and G225D/E and/or E190D (H1N1/1918 pandemic IV, H1N1 classical and avian-like swine IVs). Substitutions at some of these four “canonical” positions and/or several other HA positions of poultry-adapted IVs and sporadic mammalian isolates with H4, H5, H6, H7, H9, and H10 HAs have been found to facilitate binding to human-type receptors ([Paulson and de Vries, 2013](#); [Thompson and Paulson, 2020](#); [Liu et al., 2022, 2023](#)).

The primary function of IV neuraminidase is to attenuate HA interactions with decoy receptors [for reviews on structure and functions of IV NA, see [von Itzstein \(2007\)](#), [Shtyrya et al. \(2009\)](#), and [McAuley et al. \(2019\)](#)]. In the early stages of infection, NA removes sialic acid receptors from soluble sialoglycoproteins, mucous blanket, and the cellular glycocalyx, thereby facilitating virus motility and access to functional receptors on the cell membrane. In the late stages of infection, NA desialylates viral progeny and the surface of infected cells, preventing virus aggregation and promoting its release. As the receptor-destroying activity of NA counteracts the receptor-binding activity of HA, a balance of these activities with respect to the spectrum of sialoglycans present in the target host tissues is essential for IV fitness ([Wagner et al., 2002](#); [de Vries et al., 2020](#)). However, in contrast to HA, there is much less understanding of NA specificity for Neu5Ac2–3Gal- and Neu5Ac2–6Gal-containing sialoglycans and of alterations in its substrate specificity during avian-to-mammalian adaptation of IVs.

Studies on the desialylation of soluble monovalent sialyloligosaccharides showed that N1 and N2 NAs of avian IVs predominantly hydrolyze Neu5Ac2–3Gal-terminated substrates,

whereas NAs of human and porcine viruses have dual specificity with less efficient cleavage of Neu5Ac2–6Gal compared to Neu5Ac2–3Gal ([Baum and Paulson, 1991](#); [Couceiro and Baum, 1994](#); [Kobasa et al., 1999](#); [Mochalova et al., 2007](#); [Gerlach et al., 2012](#); [Garcia et al., 2013](#)). The catalytic activity correlated with the binding avidity of the substrates to the NA, suggesting that the NA specificity for the avian-type and human-type sialoglycans depends on differences in their binding to the catalytic site ([Mochalova et al., 2007](#); [Garcia et al., 2013](#)). Avian-origin N2 NA was introduced into the human population with the H2N2/1957 pandemic IV. Pandemic and early post-pandemic H2N2 virus strains had avian-type substrate specificity; the first detectable increase in their ability to cleave Neu5Ac2–6Gal was observed with A/England/12/1962 and correlated with the acquisition of the I275V amino acid substitution in the NA ([Kobasa et al., 1999](#)). The molecular mechanism behind the effect of this substitution on NA specificity remains undefined.

The NA of the H1N1/1918 pandemic IV differed from the closest avian N1 NAs by about 30 amino acid residues, including the Y347N substitution (the N2 numbering system is used throughout the text; position 347 corresponds to codon 344 of the N1 NA gene). This substitution was predicted to be a marker for mammalian adaptation ([Reid et al., 2000](#)) because it was also present in the N1 NA of the swine-adapted IV lineage that emerged from the independent introduction of an avian H1N1 virus into European swine in the late 1970s. In the crystal structures of N1 NA, the side chain of amino acid 347 protrudes into the solvent at the rim of the catalytic site; its main chain carbonyl oxygen forms a part of the Ca<sup>2+</sup>-binding site ([Russell et al., 2006](#); [Xu et al., 2008](#)). Several research groups have found that the identity of amino acid at position 347 affects N1 NA avidity for the substrate, catalytic activity, and sensitivity to the NA inhibitors oseltamivir and zanamivir ([Collins et al., 2009](#); [Rameix-Welti et al., 2011](#); [Yongkiettrakul et al., 2013](#); [Dai et al., 2016](#); [Su et al., 2021](#)). Although these results demonstrated a contribution of amino acid 347 to substrate binding and catalysis, they were obtained using the non-natural substrate MUNANA and did not provide information on the effect of amino acid 347 on NA specificity toward Neu5Ac2–3Gal- and Neu5Ac2–6Gal-terminated sialoglycans.

X-ray analysis of NA binding to sialoglycans could not be performed for technical reasons. Therefore, molecular dynamics (MD) simulation was used to study the interactions of Neu5Ac2–3Gal- and Neu5Ac2–6Gal-containing trisaccharides with N1 NAs from H5N1 avian and H1N1 human IVs ([Raab and Tvaroska, 2011](#); [Jongkon and Sangma, 2012](#); [Phanich et al., 2019](#)). Collectively, these studies predicted that the catalytic sites of avian-type NAs favor binding to Neu5Ac2–3Gal- over Neu5Ac2–6Gal-containing substrates, whereas human-type NAs bind both substrates. Binding specificity was related, in part, to the poor fit of the Neu5Ac2–6Gal-containing glycan, in its bended solution-dominant conformation, into the catalytic site of the avian-type NAs. In contrast, the nearly linear conformers of the Neu5Ac2–3Gal-containing glycan were accommodated by both avian-type and human-type NAs. The amino acid at position 347 was predicted to contribute to binding specificity by steric hindrance and by interactions with the Neu5Ac and Gal moieties of the substrate. Because these modeling experiments used wild-type N1 NAs separated by multiple amino acid substitutions, further studies are needed to confirm and specify the predictions regarding the role of amino acid 347 in substrate specificity of N1 NA.

Here, we used molecular modeling to characterize complexes of two N1 NAs differing by the single substitution N347Y with two pentasaccharides representing models of the human-type and avian-type NA substrates. The NA-ligand complexes were built by molecular docking, and their geometry was refined by hundred-nanosecond MD simulation in explicit solvent. Using these models, we determined the effects of amino acid 347 on the conformation and atomic interactions of the ligand in the catalytic site, the free energy of binding, and the solvation of the glycosidic oxygen. In addition, we performed analysis of the host- and lineage-specific variation of the amino acid 347 among N1 NAs in the GISAID EpiFlu sequence database and assessed selective pressures at this position using comparable phylogenetic dN/dS techniques.

## 2 Materials and methods

### 2.1 Molecular modeling

#### 2.1.1 Models of sialoglycans and NAs

Two pentasaccharides,  $\alpha$ -D-Neu5Ac(2 $\rightarrow$ 6) $\beta$ -D-Gal(1 $\rightarrow$ 4) $\beta$ -D-GlcNAc(1 $\rightarrow$ 3) $\beta$ -D-Gal(1 $\rightarrow$ 4) $\beta$ -D-Glc-OH, referred to in this study as 6S, and  $\alpha$ -D-Neu5Ac(2 $\rightarrow$ 3) $\beta$ -D-Gal(1 $\rightarrow$ 4) $\beta$ -D-GlcNAc(1 $\rightarrow$ 3) $\beta$ -D-Gal(1 $\rightarrow$ 4) $\beta$ -D-Glc-OH (referred to as 3S) were used as models of the human-type and avian-type glycan receptors, respectively (Supplementary Figure S1). The pyranose ring conformations of 6S and 3S were set as chair  ${}^2C_5$  for Neu5Ac and as  ${}^4C_1$  for the remaining saccharides. The glycosidic backbone conformation of these glycans was defined by the set of dihedral angles  $\phi_i/\psi_i$ ,  $i=1-4$ . The Neu5Ac2-3Gal linkage of 3S was defined by  $\phi_1$  (C1-C2-O2-C3) and  $\psi_1$  (C2-O3-C3-H3). The Neu5Ac2-6Gal linkage of 6S was defined by  $\phi_1$  (C1-C2-O6-C6),  $\psi_1$  (C2-O6-C6-C5), and  $\omega$  (O6-C6-C5-H5). The other pairs of glycosidic dihedral angles were defined by four consecutive atoms  $H_i-C_i-O_{i+1}-C_{i+1}$  and  $C_i-O_{i+1}-C_{i+1}-H_{i+1}$ . The conformations of the glycosidic bonds were set according to the previously determined dominant unbound state conformations of the glycans in solution (Sasaki et al., 2013; Elli et al., 2021). The coordinate files of the two glycans were generated using the tleap application included in AmberTools 14.0 (Case et al., 2014). The crystal structure of the NA of the H1N1/1918 pandemic influenza virus A/Brevig Mission/1/1918 (3B7E) was obtained from the RCSB Protein Databank and designated NA/N in accordance with the nature of the amino acid at position 347. The N347Y point mutant of NA/N (designated NA/Y) was generated using PyMOL 2.4.0 (Schrödinger, LLC) with the default settings for the side chain rotational states.

#### 2.1.2 Molecular docking

The ligands, 3S and 6S, were docked into the catalytic sites of NA/Y and NA/N using Autodock 4.2 (Morris et al., 2009) to generate four complexes, 3S-NA/Y, 6S-NA/Y, 3S-NA/N, and 6S-NA/N. To perform the docking, we determined the Gasteiger charges (Gasteiger and Marsili, 1978) for the atoms of the glycans (ligands) and NAs (acceptors). The conformational sampling of 6S and 3S during docking was set to allow a full conformational exploration of the Neu5Ac-Gal linkages. In contrast, the inter-glycosidic dihedrals  $\phi_2/\psi_2$ ,  $\phi_3/\psi_3$ , and the conformation of the whole disaccharide Gal1-4Glc were fixed in the docking simulations. Each docking simulation sampled 21 free rotational dihedral angles. The NAs were set as rigid macromolecules.

The grid box dimensions were  $L_x=L_y=80$  points,  $L_z=100$  points (spacing  $\Delta x=\Delta y=\Delta z=0.375$  Å); the center of the grid box was set at the hydroxyl oxygen of Y406, the key catalytic residue occupying the central position at the base of the NA active site (Taylor and von Itzstein, 1994). In each simulation, the Lamarckian genetic algorithm search was performed with the following parameters: number of runs, 100; population size, 4,000; maximum number of energy evaluations,  $3\cdot 10^7$ ; maximum number of generations, 300,000. The ligand-NA poses were selected for further geometry refinement by MD simulation based on the following criteria: (1) the position of the Neu5Ac residue in the center of the NA active site among the conserved residues R118, R292, R371, W178, D151, R152, and Y406; (2) the lowest predicted (Autodock 4.2) binding energy.

#### 2.1.3 MD simulation of the complexes 3S-NA/Y, 6S-NA/Y, 3S-NA/N, and 6S-NA/N

The four glycan-NA complexes obtained by docking were submitted to MD simulation in explicit solvent. Each complex was surrounded by a 15 Å-wide layer of TIP3P water molecules (Jorgensen et al., 1983) to form an orthogonal box (simulation cell) with edges of approximately 90 Å. The molecular mechanics force fields used, GLYCAM06 (Kirschner et al., 2008) and Amber (ff14SB) (Case et al., 2014), represented the “state of the art” force fields for glycans and proteins, respectively. Tleap (AmberTools 14.0) was used to build the topology and coordinate files of the complexes. The standard cut-off (12 Å) was applied to describe non-bonded electrostatic and dispersive interactions. Each simulation cell was minimized by running 100 K steps of the default minimization algorithm included in the NAMD 2.14 (Phillips et al., 2005). The software VMD 1.9.3 (Humphrey et al., 1996) was used for the MD simulation visualization and analysis. The number of particles (N), pressure (P), and temperature (T) were kept constant during the MD simulations. The constant simulation temperature (300 K) was maintained with a Lowe-Andersen thermostat; the Nosé-Hoover-Langevin piston algorithm controlled the pressure (1.01325 bar) applied to the cell walls. During the cell density equilibration steps (duration approximately 10–15 ns), a harmonic potential energy constraint (harmonic constant of  $2.0\text{ kcal mol}^{-1}$ ) was applied to all atoms of the complexes, while water molecules were allowed to move freely. No additional constraints were applied at the equilibration and production stages of the MD simulation. For each simulated complex the equilibration period was estimated following the time evolution of two different properties, (1) the glycosidic dihedral angles  $\phi_i(t)/\psi_i(t)$  ( $i=1,4$ ) of the glycan backbone; (2) the distance RMSD(t) of the glycan from its initial position ( $t=0$  ns); a stationary oscillatory behavior of these properties over time indicated the end of the equilibration and the beginning of the production MD simulation. During the MD simulation, all complexes were sampled every 10 ps for a period of 195 ns. The equilibration time was 110 ns for 6S-NA/N, 3S-NA/N and 3S-NA/Y, and 120 ns for 6S-NA/Y.

#### 2.1.4 Analysis of conformations of 6S and 3S

The conformational analysis of the 6S and 3S in the unbound state was performed previously (Elli et al., 2021), and the results are presented in this study with permission from the Biochemical Journal via the Copyright Clearance Center. The conformations of 6S and 3S bound to the catalytic site of NA/N and NA/Y were sampled by MD simulation, and the Ramachandran plots  $\phi_i(t)/\psi_i(t)$  with



corresponding density color maps for the four glycosidic linkages connecting the five sugar residues of the ligands were calculated. In these plots, the color gradient from blue to red is proportional to the density of states that were sampled by the MD simulations (from 110 or 120 ns to 195 ns). The Ramachandran plots, the 2D binning procedure and the density color maps were generated using the statistical software R (R Core Team, 2013), as described previously (Elli et al., 2021). The most populated states of the  $\omega$  angle of 6S and of the internal dihedral angle C1-C2-C3-C4 of Neu5Ac (see Supplementary Figure S1) were determined at the production stage of the MD simulation using the 1D binning procedure implemented in the OriginPro 8.0 software (OriginLab Corp.). The internal dihedral angle C1-C2-C3-C4 was used to monitor the conformation of the Neu5Ac moiety. Values of this angle around  $60^\circ$  correspond to the chair  ${}^2C_5$  conformation typical for the terminal Neu5Ac residue in sialoglycans; values around  $\pm 180^\circ$  characterize the distorted boat conformation observed in the crystal structures of NA complexes with the free Neu5Ac molecule (Varghese et al., 1992).

### 2.1.5 Structure of the glycan-NA complexes

The structures of the NA complexes with 3S and 6S were determined by MD simulation as follows. First, the average structure of each glycan-NA complex was calculated for the production stage of the MD simulation using the Wordom 0.22-rc3 software (Seeber et al., 2007). Next, the MD snapshot having the smallest root mean square distance (RMSD) from the average structure of the complex was extracted from the MD simulation trajectory using VMD 1.9.3 software. Thus, the MD simulation snapshots taken at times 142.03, 150.44, 147.95, and 141.59 ns were selected to represent the complexes 6S-NA/N, 6S-NA/Y, 3S-NA/N, and 3S-NA/Y, respectively. These snapshots had RMSDs of 0.494, 0.511, 0.495, and 0.507 Å, respectively, from the corresponding average structures.

### 2.1.6 Free energy of binding determined using MMPBSA approximation

To estimate the Poisson Boltzmann free energy of binding ( $\Delta G_{PB}^{bind}$ ) from the MD simulation trajectories of the glycan-NA complexes, we used Molecular Mechanics Poisson Boltzmann Surface Area (MMPBSA) method described previously (Weis et al., 2006; Elli et al., 2021). For each glycan-NA complex, the  $\Delta G_{PB}^{bind}$  was determined as average on the production phase of MD simulation (between 130 to 195 ns) with a sampling frequency of 200 ps and a sample size of 325 poses covering a range of 65 ns. The standard error of the mean (SEM) of  $\Delta G_{PB}^{bind}$  was calculated as  $\sigma/(N)^{1/2}$ , where  $\sigma$  represents the estimated standard deviation, and  $N$  represents the number of samples. The contributions to  $\Delta G_{PB}^{bind}$  of individual binding counterparts (saccharide and amino acid residues) ( $\Delta G_{PB}^{bind}(i)$ ) were calculated using the MMPBSA.py application (Miller et al., 2012) included in Ambertools 14.0 package. The chosen set of residues included all glycan residues (Neu5Ac, Gal, GlcNAc, Gal2, and Glc), and all amino acids of NA characterized by an energy cut-off  $|\Delta G_{PB}^{bind}(i)| \geq 0.3$  Kcal mol $^{-1}$ . Negative and positive values of  $\Delta G_{PB}^{bind}(i)$  reflect, respectively, favorable and unfavorable contributions of the residue to the binding energy.

### 2.1.7 Distribution of the water molecules in proximity of the Neu5Ac-Gal glycosidic bond

The MD simulation in explicit solvent reproduces the distribution of water molecules at the interface between the complex (or the

macromolecular surface) and the surrounding water (Raffaini et al., 2006; Raffaini and Ganazzoli, 2006). Here, we focused on the glycosidic oxygen of the Neu5Ac-Gal linkage, which is the target of hydrolysis catalyzed by NA (Taylor and von Itzstein, 1994; von Itzstein, 2007). The profile of the local concentration of the water molecules surrounding the oxygen atom was investigated using the radial pair distribution function  $rdf(r)$  (Levine et al., 2011) as implemented in VMD 1.9.3:

$$rdf(r) = \lim_{dr \rightarrow 0} \left[ \frac{V}{4\pi N_{pair} r^2} \frac{p(r)}{dr} \right] \quad (1)$$

where the variable  $r$  represents the distance between the glycosidic oxygen of the Neu5Ac-Gal bond (reference atom) and the oxygen atom of each H $_2$ O molecule,  $N_{pair}$  denotes the number of potential atom pairs in the system, and  $V$  is the volume of the simulation cell.  $rdf(r) dr$  reflects the concentration of water molecules in a spherical layer of infinitesimal thickness (between  $r$  and  $r + dr$ ) centered on the glycosidic oxygen.  $rdf(r)$  profiles were calculated for each of the four simulated complexes sampled every 100 ps at the production stage (130 to 195 ns) of the MD simulation.

### 2.1.8 Hydrogen bonds

Tight H-bonds were defined based on the following criteria: a distance less than 3 Å between the atoms of donor X and acceptor Y (X-H---Y) and a corresponding angle (X-H---Y) greater than  $150^\circ$  (Almond et al., 2006). These conditions were checked for all possible H-bonds between the glycan and the protein in four studied complexes at the production stage of MD simulation. The percentage of time H-bonds persisted during the simulation was estimated using Origin 8 software.

## 2.2 Analyses of NA sequences

### 2.2.1 NA sequences, phylogenetic analyses, and amino acid prevalence at position 347

The sequences were obtained from the GISAID EpiFlu database (Shu and McCauley, 2017) accessed on June 19, 2023. Host- and lineage-specific nucleotide sequences were downloaded using the EpiFlu search filter “N1 NA, complete length” combined with each of the following filters: (1) “HA subtypes, H1-H4, H6-H16; Avian,” (2) “H5 HA; Avian,” (3) “H5 HA; Human,” (4) “H5 HA; Mammalian,” (5) “H1 HA; Swine,” (6) “Human; isolated before 2009,” and (7) “Human; isolated from 2009 to 2023.” Sequences were aligned using the MAFFT program implemented in the Unipro UGENE 47.0 (Okonechnikov et al., 2012). The datasets were processed using Bio-Edit 7.1.11 (Hall, 1999) and Jalview 2.11.2.7 (Waterhouse et al., 2009). Sequences with gaps, ambiguities, incomplete sequences, and sequences of laboratory-derived IVs were excluded; only one sequence from each cluster of identical sequences was retained. Due to the large number of H1N1pdm sequences (exceeding 30,000), a subset of 2,568 representative sequences was selected for the analyses. Maximum likelihood trees were generated using either the FastTree method (Price et al., 2010) implemented in Unipro UGENE or IQ-TREE 2 (Minh et al., 2020) with ModelFinder (Kalyanamoorthy et al.,

2017) and the ultrafast bootstrap approximation (Hoang et al., 2018). The trees were plotted using FigTree 1.4.4.<sup>1</sup> Based on the trees, the sequences of swine IVs were separated into the classical and the avian-like lineage; the sequences of human IVs were separated into the A/H1N1/1918-like seasonal lineage and the A/H1N1/2009-like H1N1pdm lineage. In some of the sequence datasets, a few sequences failed to cluster with the rest. Some of those sequences were discarded as laboratory artifacts; the remaining atypical sequences represented either swine-virus-like isolates from birds and humans or human-virus-like isolates from pigs and birds; they were combined and analyzed separately. To facilitate analyses of the trees, we annotated the sequences with the single-letter code for the amino acid at position 347 (corresponds to codon 344 of the N1 NA ORF). The number of sequences in each group is listed in Table 3. The group name, virus name, and accession number for each analyzed sequence are shown on the global tree (Supplementary material, Global NA tree.nex and Global NA tree.svg). Prevalence of amino acids at position 347 was determined using the positional numerical summary tool of Bio-Edit.

To annotate titles of NA and/or HA sequences with specific amino acids of both proteins, we downloaded the sequences using the search filter “HxN1; HA+NA; complete length.” Curated HA and NA sequences were concatenated with MEGA11 (Tamura et al., 2021). Using Bio-Edit, we selected specific columns in the alignments of translated concatenated sequences and copied them into sequence titles.

### 2.2.2 Selection pressure analyses

We applied several codon-based phylogenetic tests of natural selection to the global alignment of all N1 NA sequences [for a review see Spielman et al. (2019)]. These tests estimate the ratio of rates of non-synonymous and synonymous substitution (dN/dS) and compare it to the neutral expectation, dN/dS = 1 (FEL and MEME tests) (Kosakovsky Pond and Frost, 2005; Murrell et al., 2012). For analyses investigating within-group selection (e.g., within “Classical swine” group, or “H1N1pdm” group), we partitioned all branches in the combined NA tree using “conjunctive” labeling: each leaf of the tree has a label based on its annotation, and internal nodes (whose labels are not known) are assigned the same label as their descendants, but only when all of the descendants have the same label. The remaining internal nodes receive no label (“Unlabeled” group). This approach is conservative and will not attempt to infer virus group assignments for internal/ancestral nodes unless such assignment is unambiguous. Given G groups in the tree, we inferred G dN/dS ratios using maximum likelihood. For each group, we tested whether its dN/dS was different from 1, using the likelihood ratio test.

We also ran a multi-group MEME analysis, which allowed dN/dS to vary among branches within a group, and can be used to detect episodic selection, i.e., selection affecting only a proportion of branches in the tree (Murrell et al., 2012). For all dN/dS analyses, we used the HyPhy package (v 2.5.54).

## 3 Results

### 3.1 Molecular modeling

To study recognition by the catalytic sites of N1 NAs of human-type and avian-type sialoglycans and to dissect the role of amino acid at position 347 in recognition, we modeled binding of the sialopentasaccharides 6S and 3S to the NA of human-adapted pandemic IV A/Brevig Mission/1/1918 (H1N1) and to its “avianized” N347Y mutant. Four complexes, 3S-NA/Y, 6S-NA/Y, 3S-NA/N, and 6S-NA/N, were built by docking, and their geometry was analyzed and refined by MD simulation in explicit solvent (Figure 1). The complexes 6S-NA/N and 3S-NA/Y were used to model interactions between human-type and avian-type NAs, respectively, with corresponding host-specific sialoglycans (referred to as “homologous” interactions throughout the text). The complexes 6S-NA/Y and 3S-NA/N modeled “heterologous” interactions.

#### 3.1.1 Alteration of the conformations of 6S and 3S upon binding to NA

The Ramachandran plots and the most populated conformations of the glycosidic torsion angles  $\phi_i/\psi_i$  for both free and bound sialoglycans sampled during the productive stage of MD simulations are presented in Figure 2, Supplementary Figure S2, and Supplementary Table S1. The  $\phi_i/\psi_i$  torsions of free 6S demonstrated significant conformational flexibility of the Neu5Ac2–6Gal linkage and populated four states,  $-62^\circ/-163^\circ$  (41%),  $-73^\circ/155^\circ$  (26%),  $-66^\circ/110^\circ$  (20%) and  $-173^\circ/170^\circ$  (13%) (Elli et al., 2021) (see Figure 2A; Supplementary Table S1). In the homologous 6S-NA/N complex, the conformational space of the Neu5Ac2–6Gal linkage was reduced to one of these four states ( $-67^\circ/122^\circ$ ). In contrast, none of the four states were observed in the heterologous 6S-NA/Y complex, in which case the Neu5Ac2–6Gal linkage adopted a new conformation ( $\phi_1/\psi_1 = 52^\circ/-130^\circ$ ) not previously seen in free Neu5Ac2–6Gal-terminated glycans or in their complexes with IV HA (Xu et al., 2009; Elli et al., 2014; Macchi et al., 2016). Distinctions in the recognition of Neu5Ac2–6Gal linkage by NA/N and NA/Y were also evident at the level of the linkage dihedral angle  $\omega$  (O6-C6-C5-H5). More significant changes in  $\omega$  (from *gauche* (–) to *trans*) occurred in the case of 6S binding to avian-type NA/Y compared to the homologous NA/N (Figure 2C).

MD simulations with the avian-type glycan 3S revealed analogous pattern. The Neu5Ac2–3Gal linkage of free 3S populated three states,  $-67^\circ/-3^\circ$  (74%),  $-92^\circ/-58^\circ$  (15%) and  $-168^\circ/-23^\circ$  (11%) (Elli et al., 2021). The second conformation was selected when 3S bound to the homologous NA/Y ( $\phi_1/\psi_1 = -79^\circ/-42^\circ$ ), whereas binding to heterologous NA/N induced a new conformation ( $\phi_1/\psi_1 = -55^\circ/-100^\circ$ ) (Figure 2B; Supplementary Table S1).

In contrast to significant changes in the conformational space of the Neu5Ac-Gal linkage upon binding, much weaker effects were observed at the level of the Gal-GlcNAc linkage ( $\phi_2/\psi_2$ ) (Figure 2), and no effects were detected at the level of more distant saccharide residues ( $\phi_3/\psi_3$  of GlcNAc-Gal2 linkage and  $\phi_4/\psi_4$  of Gal2-Glc linkage, Supplementary Figure S2; Supplementary Table S1). This pattern suggests that the recognition of glycans by NA was primarily determined by the structure of the terminal Neu5Ac-Gal moieties.

<sup>1</sup> <https://tree.bio.ed.ac.uk>

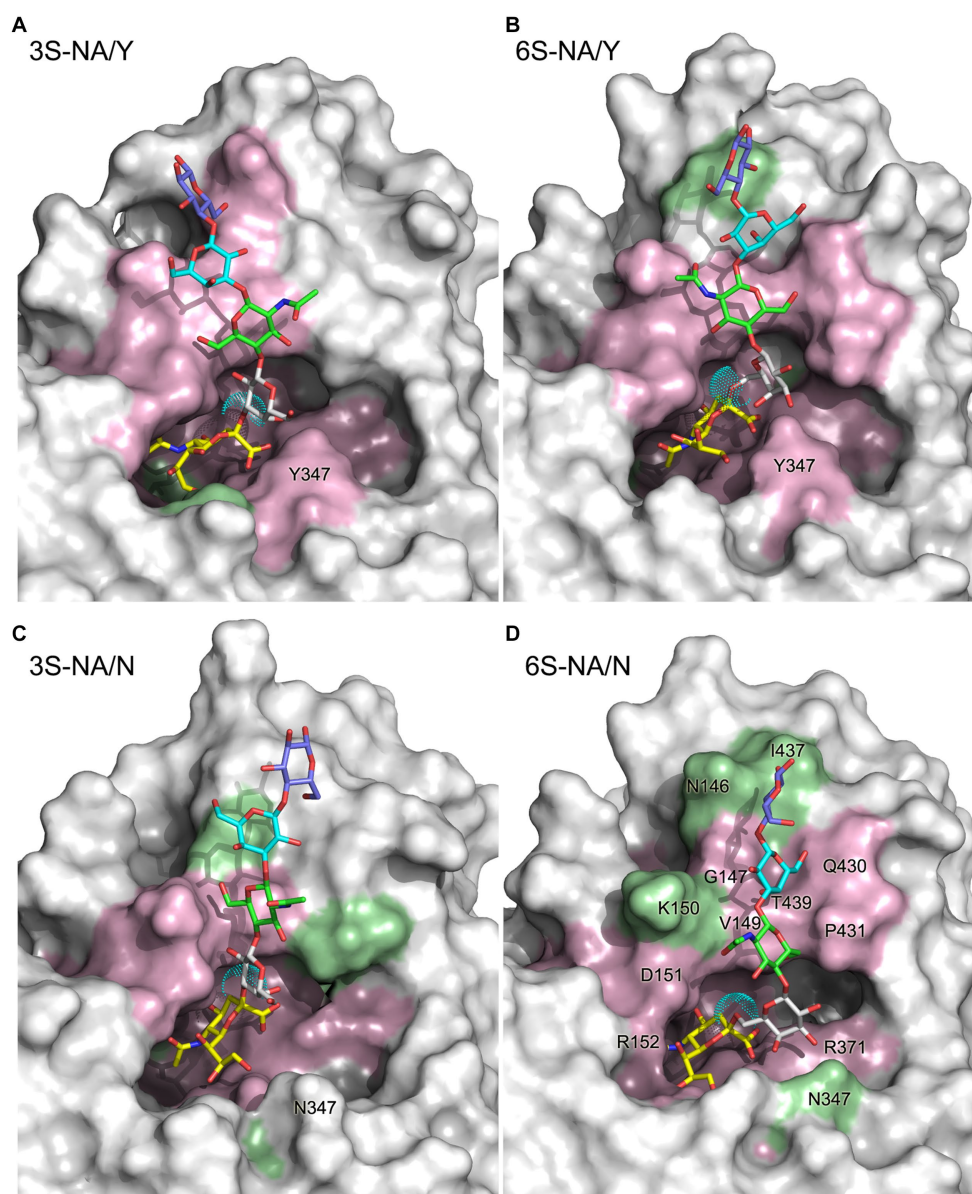


FIGURE 1

3D structure of the complexes 3S-NA/Y (A), 6S-NA/Y (B), 3S-NA/N (C), and 6S-NA/N (D) predicted by MD simulation. The sialoglycans are shown as stick models with carbon atoms colored in yellow (Neu5Ac), white (Gal), green (GlcNAc), cyan (Gal2) and slate blue (Glc), and oxygen and nitrogen atoms colored in red and blue, respectively. Cyan dots show the van der Waals surface of the oxygen atom of the Neu5Ac-Gal linkage. The protein is presented as a molecular surface with atoms colored according to their contact distance to the sialoglycan (pink, < 4 Å; green, between 4 Å and 5 Å; white, >5 Å). Selected residues are labeled on the panel (D).

These results demonstrated that the binding of sialoglycans to NA significantly reduced their conformational space. The extent of this effect (and the associated energy cost) depended on the interplay between the type of Neu5Ac-Gal linkage and the amino acid residue at position 347; the effect was less pronounced in the case of homologous glycan-NA interactions.

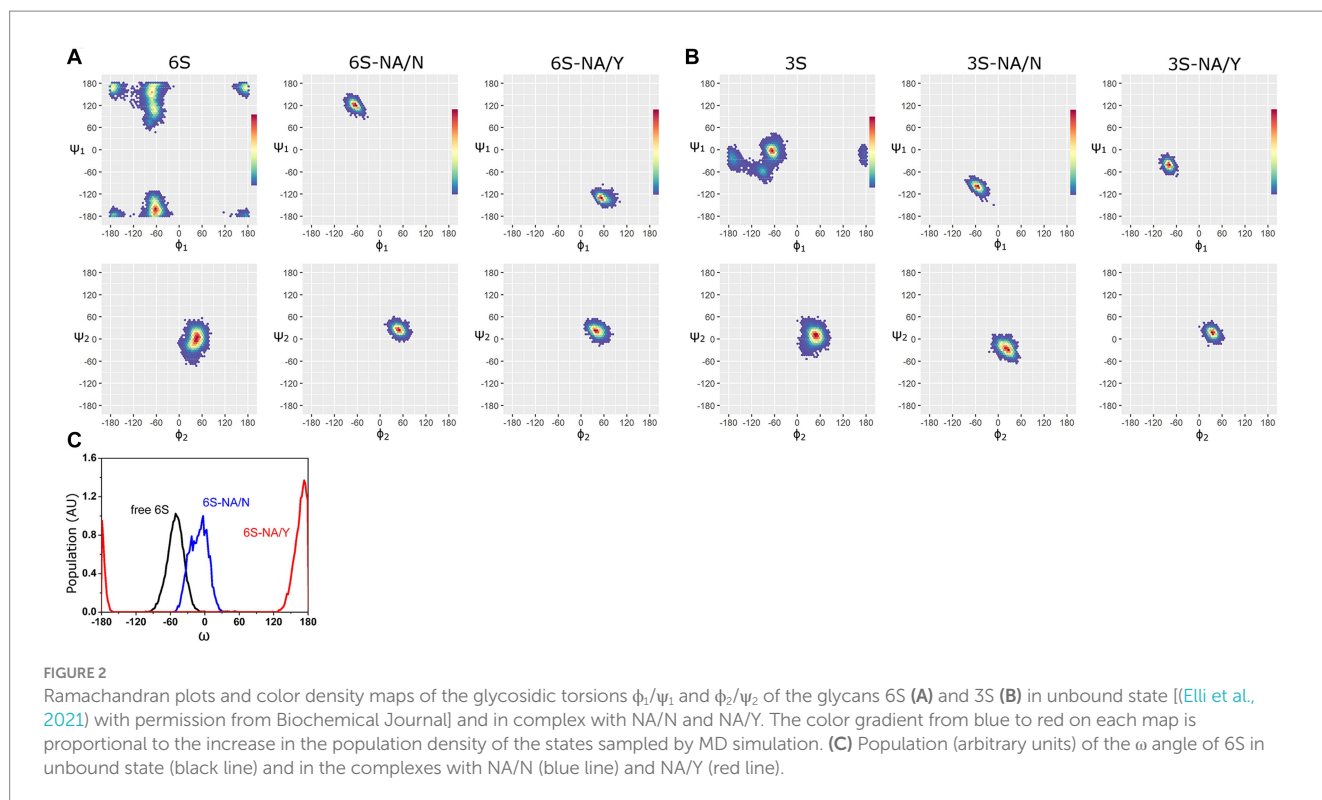
### 3.1.2 Atomic contacts in the NA complexes with sialoglycans

The contact distances between sialoglycans and the protein in the complexes are color coded in Figure 1. Figure 3 illustrates the interactions between the terminal Neu5Ac-Gal moiety and

the catalytic site of the NA. Selected contact distances, H-bonds and their persistence during the production phase of the MD simulation are presented in Table 1. In all four complexes, the Neu5Ac residue was located between the conserved amino acid residues R118, E119, D151, R152, R156, W178, R292, E276, R371, and Y406. However, the position of Neu5Ac with respect to the arginine triad (R118, R292, R371), residues R152, R156, and Y406, as well as loops 150 and 430, and contact distances between Neu5Ac and the protein differed between homologous and heterologous complexes.

The homologous complexes 3S-NA/Y and 6S-NA/N showed considerable parallelism in the NA interactions with the terminal





Neu5Ac moiety of the glycan (Figures 3A,D; Table 1). Neu5Ac was co-planar with the arginine triad residues in both complexes. Tight H-bonds were formed by Neu5Ac with R292, R371, E119, E276 and either R152 (3S-NA/Y) or R156 (6S-NA/N). The distances between the hydroxyl group of the key nucleophilic residue Y406 and its target C2 atom of Neu5Ac (4.0 Å, 3S-NA/Y; 3.9 Å, 6S-NA/N) were compatible with the proposed role of Y406 in NA-mediated hydrolysis (von Itzstein, 2007; Vavricka et al., 2013). Comparison of 3S-NA/Y and 6S-NA/N with the published crystal structures of NA complexes with free Neu5Ac revealed a marked similarity in the contact distances of the Neu5Ac residue in these complexes (Table 1). This finding suggests that, in the homologous complexes, the asialic moieties of the bound glycan do not interfere with the optimal position of the terminal Neu5Ac moiety in the NA catalytic site. Importantly, the Gal residues of the glycan in the complexes 3S-NA/Y and 6S-NA/N facilitated binding through the formation of polar contacts and weak transient H-bonds with the side chain of amino acid 347. Thus, during the MD simulation, the 4-hydroxyl of Gal of 3S oscillated within 3.0–3.5 Å from the hydroxyl group of Y347, whereas the 4-hydroxyl of Gal of 6S oscillated within 4.7–5.9 Å from the side chain of the carboxamido group of N347. In addition to interacting with amino acid 347, the asialic residues of 3S and 6S also interacted with V149. Gal made polar contacts with V149 in the 3S-NA/Y complex, whereas GlcNAc was involved in hydrophobic interactions with this amino in the 6S-NA/N complex. Interestingly, in 6S-NA/N, the GlcNAc and Gal2 residues of the glycan approached loop 430 and were engaged in polar and hydrophobic interactions with Q430 and P431, respectively (Figure 1D; Table 1). This finding supports the previous hypothesis about potential role of amino acid 430 in the substrate specificity of human-type NAs (Jongkon and Sangma, 2012).

The heterologous complexes 6S-NA/Y and 3S-NA/N shared a few structural features that differentiated them from both the homologous complexes and the published Neu5Ac-NA co-crystal structures. Thus, 6S bound to NA/Y in a tilted orientation of the Neu5Ac residue with respect to the arginine triad. Compared to the 3S-NA/Y complex, the Neu5Ac moiety of 6S moved away from loop 340 toward loop 150, lost H-bonds with R292 (distance, 5.1 Å), R152 (6.1 Å) and E276 (4.4 Å), and acquired tight H-bonds with R118 (2.8 Å) and R156 (3.2 Å) (Figures 1B, 3B; Table 1). The distance between the hydroxyl group of Y406 and the C2 carbon atom of Neu5Ac has increased to 4.8 Å in the 6S-NA/Y complex, compared to 4.0 Å in the 3S-NA/Y complex. Observed differences in the binding of 3S and 6S to NA/Y were likely caused by a larger footprint of the Neu5Ac2–6Gal moiety, a steric conflict of the 6-linked Gal residue with Y347 (see the pink surface in Figure 1B), and resolution of this conflict by the shift of the sialoglycan toward loop 150.

Similar to the binding of 6S to NA/Y, 3S bound to NA/N in a tilted orientation characterized by the absence of H-bonds between Neu5Ac and R292, R152 and E276 and by the formation of H-bonds with R118, R371 and R156 (Table 1). Due to this orientation, the Neu5Ac-Gal moiety shifted toward loop 150, and the distance between Y406 and the target C2 atom of Neu5Ac increased to 5.6 Å. The tilted orientation of the Neu5Ac-Gal moiety in 3S-NA/N was stabilized by the H-bond and hydrophobic interactions of Gal with V149 (Figure 3C; Table 1). The close contact between the ligand and loop 150 in 3S-NA/N and 6S-NA/Y is illustrated in the Figures 1B,C by the dotted van der Waals spheres of glycosidic oxygen atoms approaching this loop.

In summary, these results suggested that N347 and Y347 bind the homologous sialoglycans 6S and 3S, respectively, with an almost

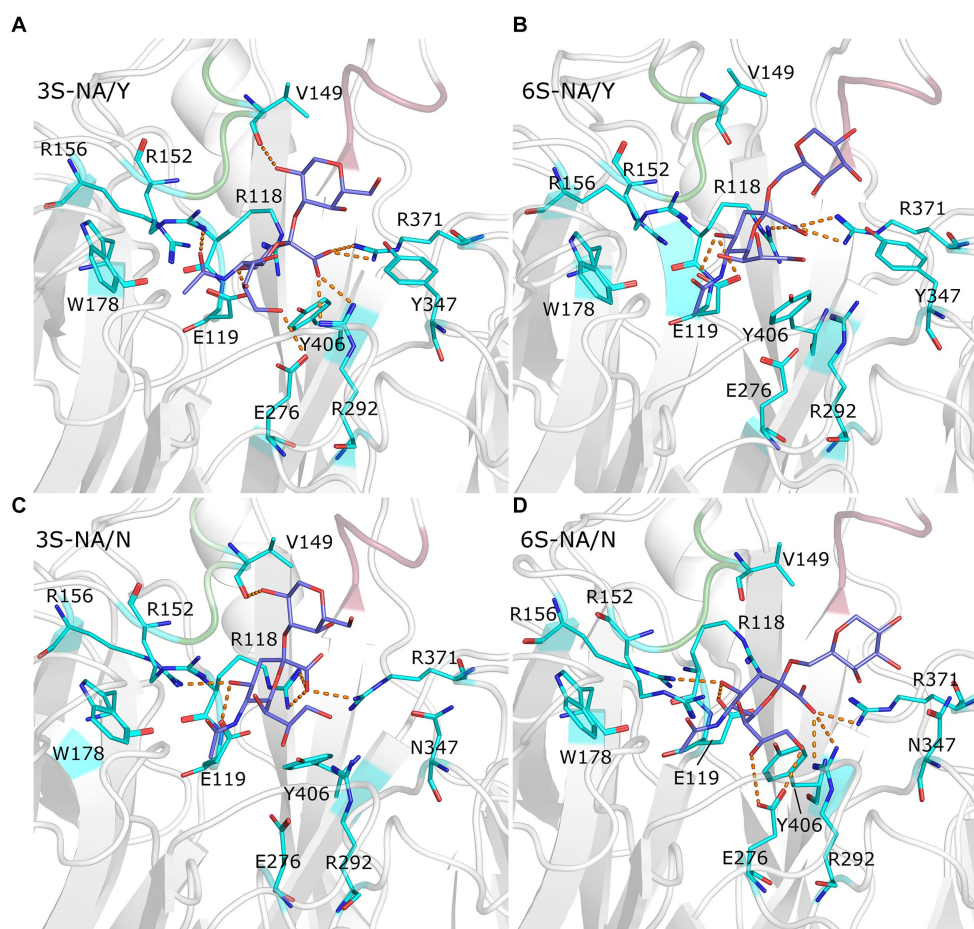


FIGURE 3

Close up view of the complexes 3S-NA/Y (A), 6S-NA/Y (B), 3S-NA/N (C), and 6S-NA/N (D). The white ribbon depicts the protein backbone, with loop 150 and loop 430 colored in green and pink, respectively. Stick models show terminal Neu5Ac-Gal moiety of the ligand (carbon atoms in slate blue) and selected contact residues (carbon atoms in cyan); oxygen and nitrogen atoms are colored in red and blue, respectively. The hydrogen bonds with the highest population are indicated by orange dashed lines.

identical optimal orientation of the sialic acid moiety with respect to the contact residues of the catalytic site. This orientation depends, at least in part, on van der Waals and polar interactions of amino acid 347 with the Gal residue of the glycan.

### 3.1.3 Poisson-Boltzmann free energy of binding

The Poisson Boltzmann free energy of binding  $\Delta G_{PB}^{bind}$  (Weis et al., 2006) was calculated from the MD trajectories of glycan-NA complexes and decomposed into the contributions of individual saccharide residues of the glycan (Table 2) and amino acid residues of the protein (Figure 4). It should be noted that the MMPBSA procedure used to calculate  $\Delta G_{PB}^{bind}$  does not account for the conformational changes (and associated energy costs) that the ligand and receptor undergo upon binding.

Based on the values of total  $\Delta G_{PB}^{bind}$ , the human-type NA/N bound 6S with a higher avidity in comparison to 3S, whereas avianized NA/Y bound 3S more strongly than 6S. The N347Y substitution markedly reduced NA binding to 6S ( $\Delta G_{PB}^{bind}$  difference, 19.2 Kcal mol<sup>-1</sup>) and had weak unfavorable effect on binding to 3S (3.9 Kcal mol<sup>-1</sup>). This pattern generally agreed with the results of MD simulation performed

using sialotrisaccharides and NAs from wild type human and avian IVs (Phanich et al., 2019). Neu5Ac provided the major favorable contribution to the binding energy among the saccharide residues of the ligand, with the highest contribution in the case of 6S-NA/N complex (−16.8 Kcal mol<sup>-1</sup>). The other residues, including penultimate Gal, made insignificant and frequently unfavorable contributions in all complexes (Table 2).

The amino acid residues with charged side chains made the highest favorable and unfavorable contributions to  $\Delta G_{PB}^{bind}$  (Figure 4), highlighting the key role of electrostatic forces in the binding. The arginine triad R118, R292 and R371 together with R152, R156, and R224 had the most significant favorable effect on binding. A few amino acids without charged side chains, including S179, Y/N347 and Y406, also favorably contributed to binding avidity. In contrast, residues with the side chain carboxylic groups (E119, E227, E276, E277) destabilized the glycan-NA complexes. Interestingly, loop 150 containing positively charged R152 and R156 had a much higher favorable energy share compared to loop 430 (Q430, P431, K432). In most cases, both favorable and unfavorable effects were more pronounced in the homologous complex than in the heterologous



TABLE 1 Selected contact distances (Å) in NA complexes determined in this study and in published crystal structures 2BAT, 1MWE, and 1W21.

Contacting atoms (bold)		Complex						
NA	Glycan	3S-NA/Y	6S-NA/Y	3S-NA/N	6S-NA/N	2BAT	1MWE	1W21
<b>H-bonds and polar contacts</b>								
R118 NCNH <sub>2</sub>	Neu5Ac COO-	5.7	<b>2.8</b> (57)	<b>2.6</b> (86, 55)	4.8	<b>3.1</b>	<b>2.8</b>	<b>2.9</b>
R292 NCNH <sub>2</sub>	Neu5Ac COO-	<b>2.6</b> (26, 53)	5.1	5.8	<b>3.1</b> (62, 28)	<b>3.2</b>	<b>3.3</b>	<b>3.3</b>
R371 NCNH <sub>2</sub>	Neu5Ac COO-	<b>2.8</b> (21, 56)	<b>3.0</b> (67, 43)	<b>2.9</b> (86, 62)	<b>2.7</b> (94, 6)	<b>2.8</b>	<b>2.7</b>	<b>2.9</b>
E119 COO-	Neu5Ac 4OH	<b>2.6</b> (67)	<b>3.0</b> (48, 15)	<b>2.6</b> (87)	<b>2.6</b> (92)	<b>3.3</b>	<b>3.3</b>	<b>3.1</b>
R152 NCNH <sub>2</sub>	Neu5Ac NHCOMe	<b>3.1</b> (49)	6.1	5.6	4.0 (6)	4.0	<b>2.6</b>	<b>2.9</b>
R156 NCNH <sub>2</sub>	Neu5Ac 4OH	3.9	<b>3.2</b> (10)	<b>2.9</b> (38)	3.8 (37)	4.8	4.7	4.7
E276 COO-	Neu5Ac 8OH	4.7	5.5	4.9	<b>3.4</b> (95)	<b>2.6</b>	<b>2.7</b>	3.6
E276 COO-	Neu5Ac 9OH	<b>3.0</b> (60)	4.4	7.0	<b>2.6</b> (89)	<b>3.1</b>	<b>3.4</b>	<b>2.7</b>
Y347 OH	Neu5Ac 9OH	7.4	4.2 (8)	-	-	-	-	-
N347 CONH <sub>2</sub>	Neu5Ac 9OH	-	-	6.1	7.5	10.8	8.5	-
V149 CO	Gal 2OH	<b>2.9</b> (69)	-	<b>2.8</b> (58)	-	-	-	-
Q430 CONH <sub>2</sub>	Gal2 6OH	6.3	6.1	7.6	3.8	-	-	-
K150 NH <sub>3</sub> <sup>+</sup>	GlcNAc NHCOMe	12.4	3.6 (5)	7.8	6.2	-	-	-
Y347 OH	Gal 4OH	<b>3.0</b>	3.7	-	-	-	-	-
Y347 OH	Gal 3OH	-	<b>2.8</b>	-	-	-	-	-
N347 CONH <sub>2</sub>	Gal 4OH	-	-	9.4	5.5	-	-	-
N347 CONH <sub>2</sub>	Gal 3OH	-	-	10.9	5.9	-	-	-
<b>van der Waals contacts</b>								
Y406 OH	Neu5Ac C2	4.0	4.8	5.6	3.9	<b>3.2</b>	<b>3.2</b>	<b>3.1</b>
W178 Cε3	Neu5Ac NHCOMe	4.0	6.9	6.4	6.1	3.9	3.7	3.9
V149 CHMe <sub>2</sub>	GlcNAc C3,C5,C6	4.4 (C3)	5.1 (C3)	3.9 (C5)	4.4 (C6)	-	-	-
P431 Cδ	GlcNAc C6	7.3	5.3	10.5	4.0	-	-	-

The closest contacts (<3.5 Å) are shown in bold. Values in brackets depict percent of H-bond persistence (if higher than 5%). 2BAT, 1MWE, and 1W21, complexes of free Neu5Ac with N2, N9, and N6 NAs, respectively (Varghese et al., 1992, 1997; Rudiño-Piñera et al., 2006).

TABLE 2 Total  $\Delta G_{PB}^{bind}$  ( $\pm$ SEM) of the complexes and contributions of saccharide residues (Kcal mol<sup>-1</sup>).

$\Delta G_{PB}^{bind}$ component	Complex			
	3S-NA/Y	6S-NA/Y	3S-NA/N	6S-NA/N
Total	-26.8 ( $\pm$ 0.3)	-18.7 ( $\pm$ 0.3)	-30.7 ( $\pm$ 0.4)	-37.9 ( $\pm$ 0.4)
Neu5Ac	-11.5	-10.5	-9.5	-16.8
Gal	1.1	0.9	-0.5	1.7
GlcNAc	0.3	1.5	0.0	2.1
Gal2	-0.4	0.0	-0.4	2.1
Glc	0.9	-0.1	-0.3	0.2

complex of the same sialoglycan (as shown by the data in the same panels of Figure 4). However, R118 deviated from this pattern, suggesting a potential unique function for this residue in recognition. Of note, V149 contributed favorably to the binding in both 3S complexes as a result of its hydrophobic interactions with Gal (Table 1).

In summary, estimation of the free energy of binding reveals that avian-type NA/Y binds human-type 6S significantly weaker than the homologous 3S and that substitution Y347N markedly improves binding of 6S and slightly increases binding of 3S. Decomposition data

show that this pattern does not correlate with the direct contribution of amino acid 347 to  $\Delta G_{PB}^{bind}$  but primarily depends on the effect of this amino acid on favorable electrostatic interactions of 3S and 6S with the arginine triad, E119, E276, and other conserved residues.

### 3.1.4 Distortion of the Neu5Ac moiety in the complexes with NA

The initial stages of the proposed mechanism of NA-mediated hydrolysis involve the binding of sialoglycan with its Neu5Ac moiety in the chair <sup>2</sup>C<sub>5</sub> conformation (with the carboxylate in the axial position),

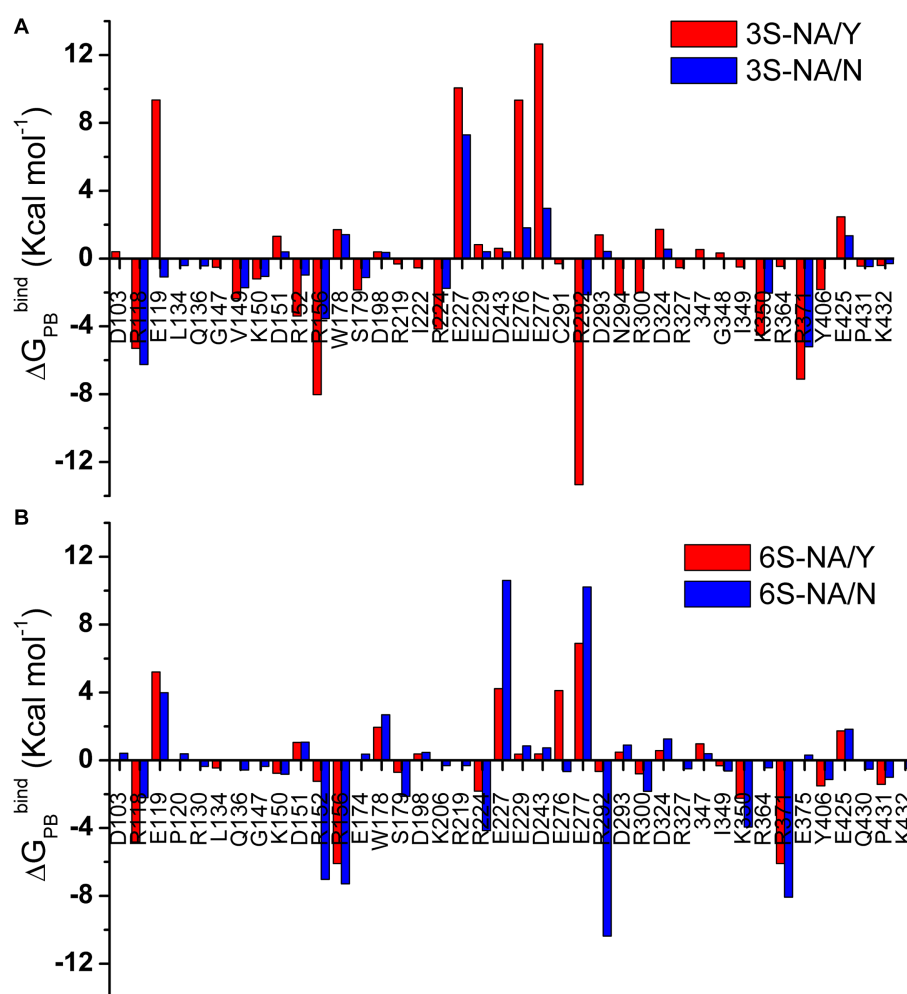


FIGURE 4

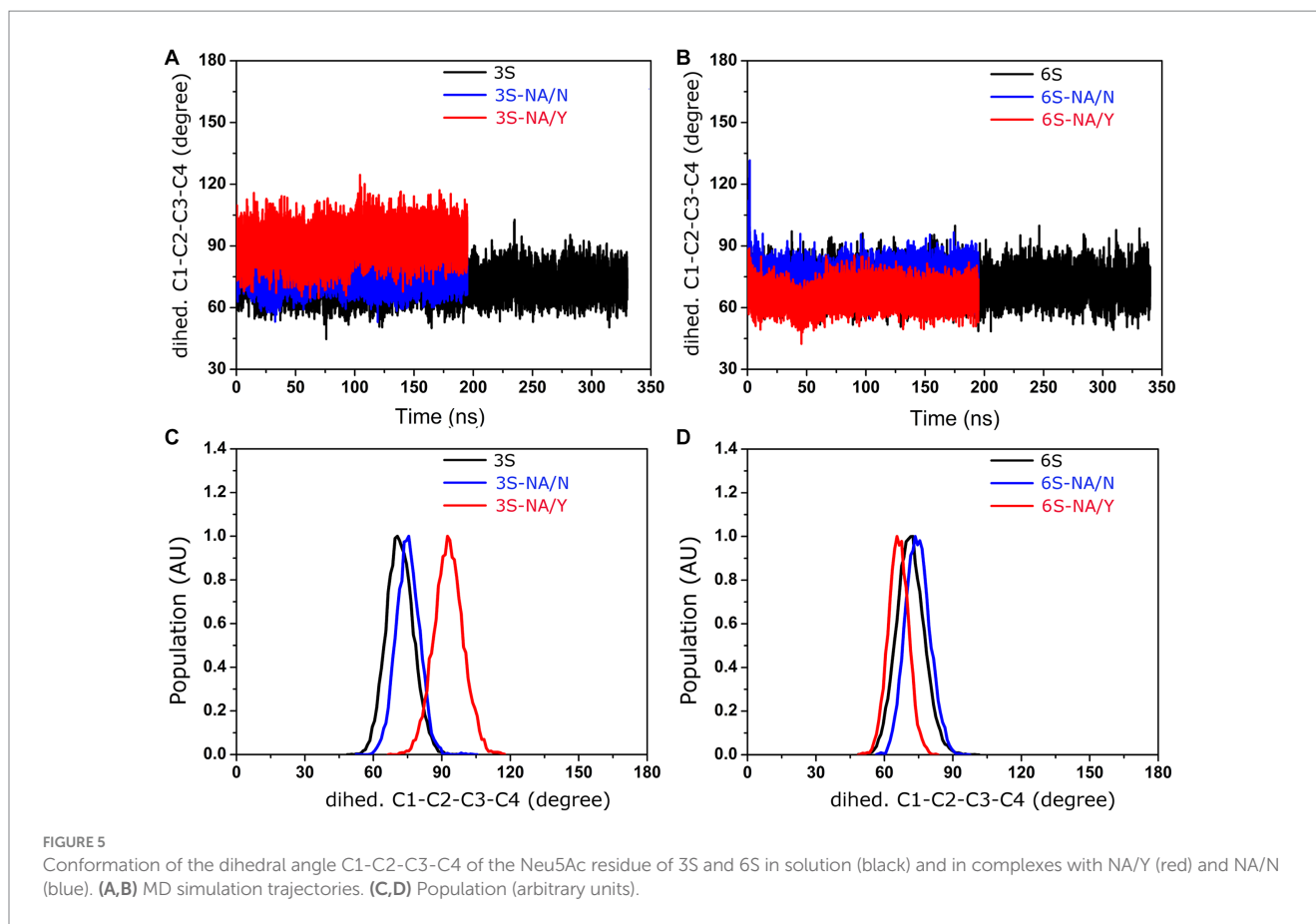
Per amino acid residue decomposition of  $\Delta G_{PB}^{bind}$  for the complexes of 3S (A) and 6S (B) with NA/Y (red) and NA/N (blue). Only residues with absolute values of  $\Delta G_{PB}^{bind}$  greater than 0.3 Kcal mol<sup>-1</sup> are shown.

followed by the distortion of the Neu5Ac ring to a pseudo-boat conformation (with the carboxylate in the pseudo equatorial position). The distortion is mediated by the ionic, hydrogen bond, and steric interactions of Neu5Ac with active site residues; it induces formation of a strained oxocarbenium ion, ultimately resulting in the cleavage of the glycosidic bond (Janakiraman et al., 1994; Taylor and von Itzstein, 1994). Taking this mechanism into account, we decided to examine evolution of the Neu5Ac ring conformation during interaction of 3S and 6S with the NAs (Figure 5). MD simulation of the 3S-NA/Y complex revealed substantial alteration of the internal dihedral angle C1-C2-C3-C4 of Neu5Ac (mean value 93°) with respect to its conformation in the free ligand (mean value 72°) indicative of ring distortion toward pseudo-boat. Minimal distortion (if any) was observed in the case of the 3S-NA/N complex (Figures 5A,C). In the case of 6S, there was a slightly greater Neu5Ac distortion in the 6S-NA/N complex compared to 6S-NA/Y complex (Figures 5B,D), although the effect was smaller than that observed in 3S-NA/Y.

### 3.1.5 Solvation of the glycosidic oxygen of Neu5Ac-Gal linkage

According to the proposed solvent-assisted mechanism of hydrolysis catalyzed by NAs, water must be present in the catalytic

pocket. In particular, the water molecule located near D151 and R152 may participate in the NA-mediated proton transfer to the glycosidic oxygen atom of the substrate (Chong et al., 1992; Taylor and von Itzstein, 1994). Thus, the efficiency of hydrolysis may be affected by the concentration of water molecules surrounding the Neu5Ac-Gal linkage. As the MD simulation in explicit solvent allows the analysis of water molecules at the interface between the complex and the solvent, we calculated the distribution profile of water around the glycosidic oxygen in both free 3S and 6S, as well as their complexes with NA/Y and NA/N as described in the Methods section (Figure 6). In the case of free glycans, local maxima of water concentration were observed at 3.2, 5.5, and 7.4 Å (3S) and 2.8, 5.3, and 7.5 Å (6S). These maxima represented the first, second and third hydration shells, respectively, surrounding the target glycosidic oxygen. The concentrations of water molecules in the second and the third shells decreased in all complexes as compared to free 3S and 6S. Moreover, the first three peaks were shifted in the 3S-NA/N complex toward longer distances indicating a significant depletion of water in this complex in comparison to the homologous 3S-NA/Y complex (Figure 6A). Similarly, a greater depletion of water was observed in the complex 6S-NA/Y compared to its homologous counterpart 6S-NA/N (Figure 6B). These effects correlate with the observed shift of the



Neu5Ac-Gal moieties in heterologous complexes 6S-NA/Y and 3S-NA/N toward loop 150 and the accompanying shielding of the glycosidic oxygen from the solvent (see models in the [Figure 1](#)).

## 3.2 Analysis of NA sequences

### 3.2.1 Prevalence of amino acid 347 in avian and mammalian IV lineages

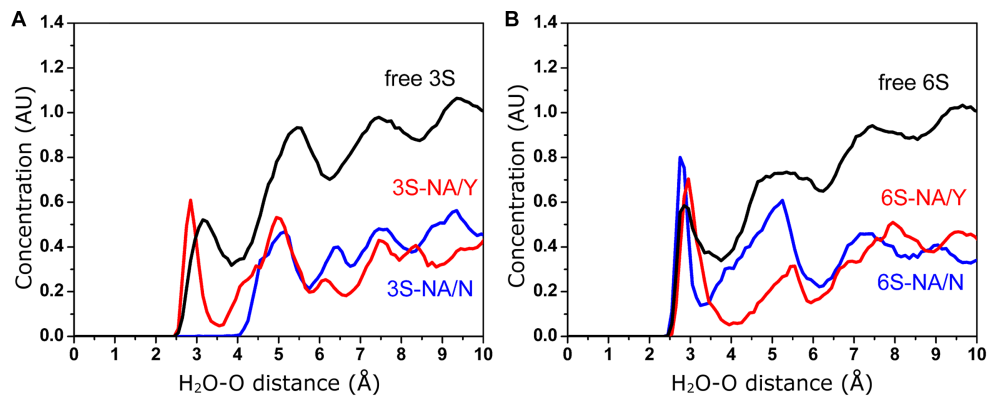
To investigate the prevalence of amino acid 347 in N1 NAs in different host species, we analyzed sequences retrieved from the GISAID EpiFlu database ([Figure 7](#)). In agreement with previous reports ([Xu et al., 2012](#); [Hufnagel et al., 2023](#); [Yang et al., 2023](#)), the N1 NAs were represented by the following major groups and lineages. No HxN1 sequences with HA subtypes H13-H16 were found in the database. The NAs of avian IVs with HA subtypes H1-H4 and H6-H12 were located near the root of the global phylogenetic tree and shared a common ancestor with two H1N1 mammalian lineages. Both these lineages, the “seasonal” human IVs and the “classical” swine IVs, were derived from the 1918 influenza pandemic. The second circulating swine IV lineage, commonly called “avian-like swine” (ALS) lineage, originated from an H1N1 avian ancestor; the first ALS viruses were isolated in Belgium and Germany in 1979. The 2009 influenza pandemic was initiated by a reassortant swine virus (H1N1pdm) that contained the NA gene from the ALS lineage. The novel H1N1pdm lineage replaced the previous seasonal human IVs. The H1N1pdm viruses were repeatedly reintroduced from humans to pigs, and this phenomenon significantly influenced the evolution of the ALS lineage ([Hufnagel et al., 2023](#)). Therefore, we divided the ALS sequences into two groups as shown in [Figure 7](#). The sequences of highly pathogenic

and panzootic H5N1 avian IVs comprised the last major N1 NA group. Due to the importance of H5N1 viruses for animal and human health, their sequences were significantly overrepresented compared to other avian sequences. We divided the NAs of all H5N1 IVs into four groups containing sequences from human isolates (H5hum), sequences from other mammalian isolates (H5mam), and two groups of avian sequences ([Figure 7](#)). The H5av-1 group included NAs of the A/goose/Guangdong/1996-like viruses from 1997–2023 (Gs/Gd-like lineage). The H5av-2 group included NAs of clade 2.3.4.4b H5N1 IVs that emerged in 2020, spread globally, and have been causing large outbreaks of infection in wild birds, poultry and mammals in 2020–2023 ([Kandeil et al., 2023](#); [Tian et al., 2023](#)). Both H5av groups also included a small number of non-zoonotic H5N1 IVs from wild birds and poultry.

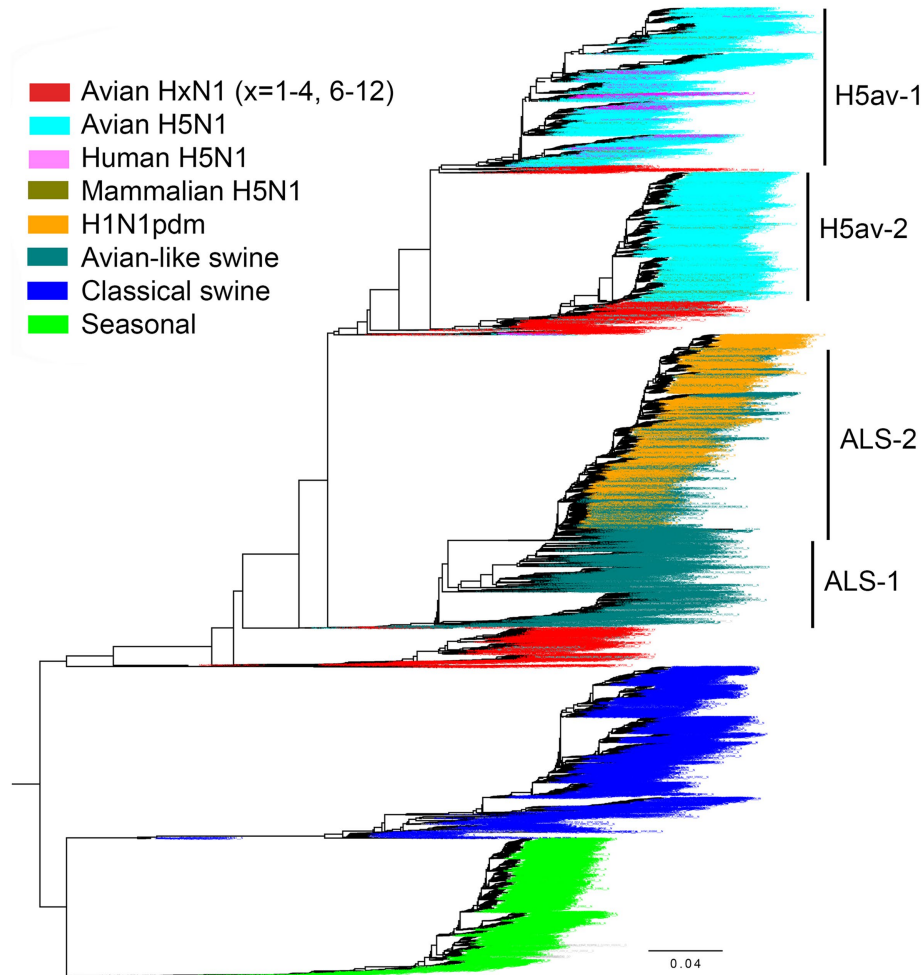
The number of NA sequences in each group and the prevalence of amino acid 347 are shown in [Table 3](#). Our initial analysis of avian NAs revealed an unusually high frequency of substitutions at position 347 in H6N1 samples. Therefore, we decided to analyze H6N1 viruses separately. Y347 was perfectly conserved in the NAs of avian viruses with HA subtypes H1-H4 and H7-H12. The H6N1 group contained 11.6% of NA variants with the Y347F substitution. The H5av-1 group contained 1.33% of mutant NA sequences, ten times more than the H5av-2 group. It remains unclear whether this difference reflects a shorter evolutionary time of the H5av-2 lineage, unique virus properties, or a combination both.

Among the human and porcine IV NAs, only one H1N1pdm sequence (0.09%) and 14 ALS-1 sequences (1.1%) contained the avian-type 347Y. NAs from classical swine, ALS, and H1N1pdm lineages contained almost exclusively N347, whereas NAs from seasonal human IVs contained either N347 or D347.





**FIGURE 6**  
 Concentration (arbitrary units) of water molecules in the vicinity of the glycosidic oxygen of the Neu5Ac-Gal linkage determined by MD simulation. **(A)** Free and bound 3S. **(B)** Free and bound 6S.



**FIGURE 7**  
 Phylogenetic relationship of 18,097 N1 NA gene sequences from the GISAID EpiFlu database used in this study. For the H1N1pdm lineage, only 2,568 representative sequences were included in the tree. The colors of the taxa names correspond to the virus group as indicated in the legend. Group separation of avian H5N1 IVs and avian-like swine IVs is illustrated. The tree was generated with FastTree. The fully annotated tree in nexus and svg formats can be found in the [Supplementary information](#).

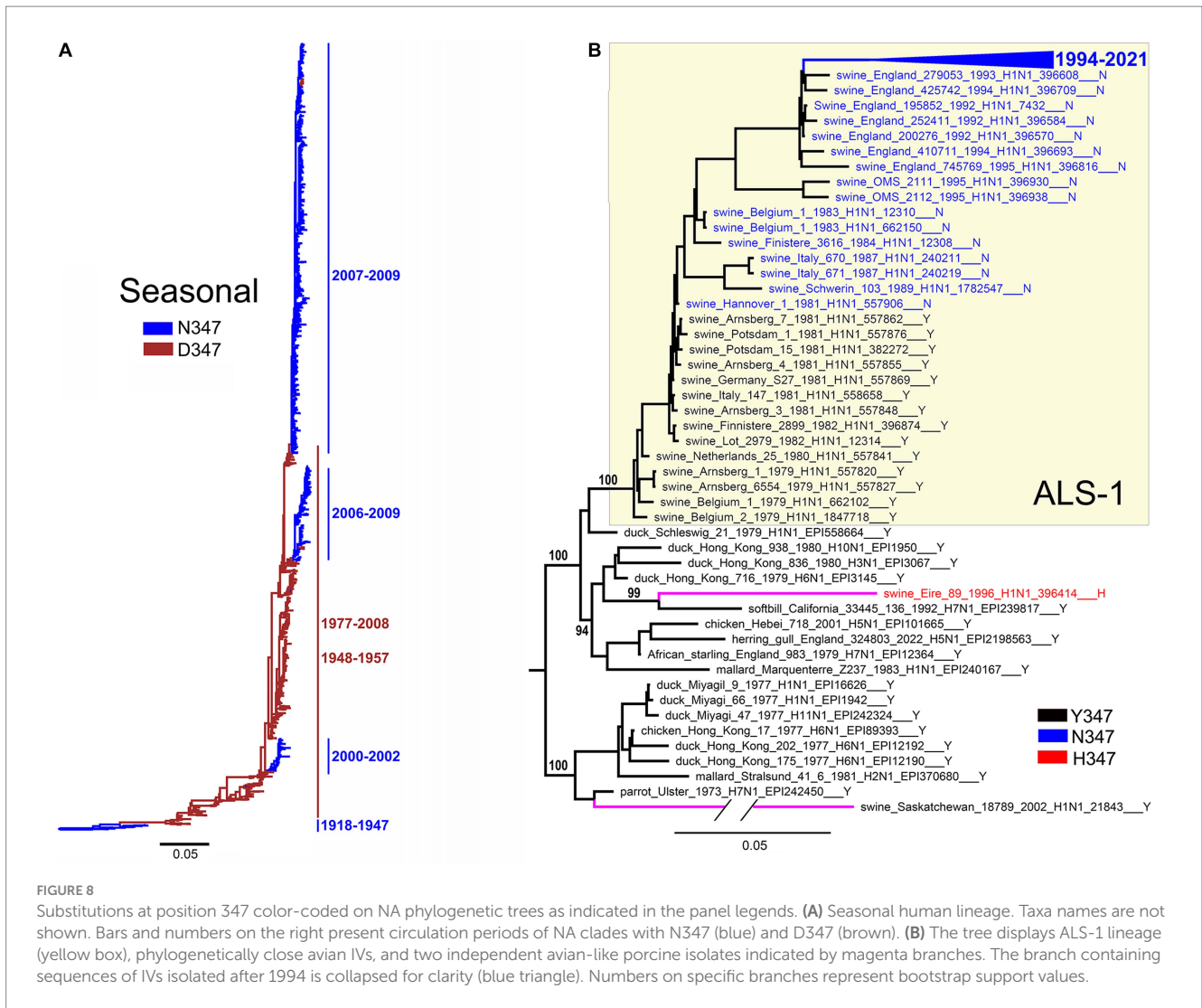


FIGURE 8

Substitutions at position 347 color-coded on NA phylogenetic trees as indicated in the panel legends. (A) Seasonal human lineage. Taxa names are not shown. Bars and numbers on the right present circulation periods of NA clades with N347 (blue) and D347 (brown). (B) The tree displays ALS-1 lineage (yellow box), phylogenetically close avian IVs, and two independent avian-like porcine isolates indicated by magenta branches. The branch containing sequences of IVs isolated after 1994 is collapsed for clarity (blue triangle). Numbers on specific branches represent bootstrap support values.

The NAs of 101 atypical sporadic swine-virus-like isolates from humans and birds and human-virus-like isolates from pigs as a rule contained the same residue 347 as their closely related original host species (see sequences labeled “Atypical” in the tree in [Supplementary material](#)).

### 3.2.2 Evolution of position 347 in seasonal and ALS-1 lineages

For the NA groups containing more than 1% of position-347 variants, we studied the location of these variants on the phylogenetic tree ([Figures 8–10](#) and [Global tree.nex/Global tree.svg](#) in [Supplementary material](#)).

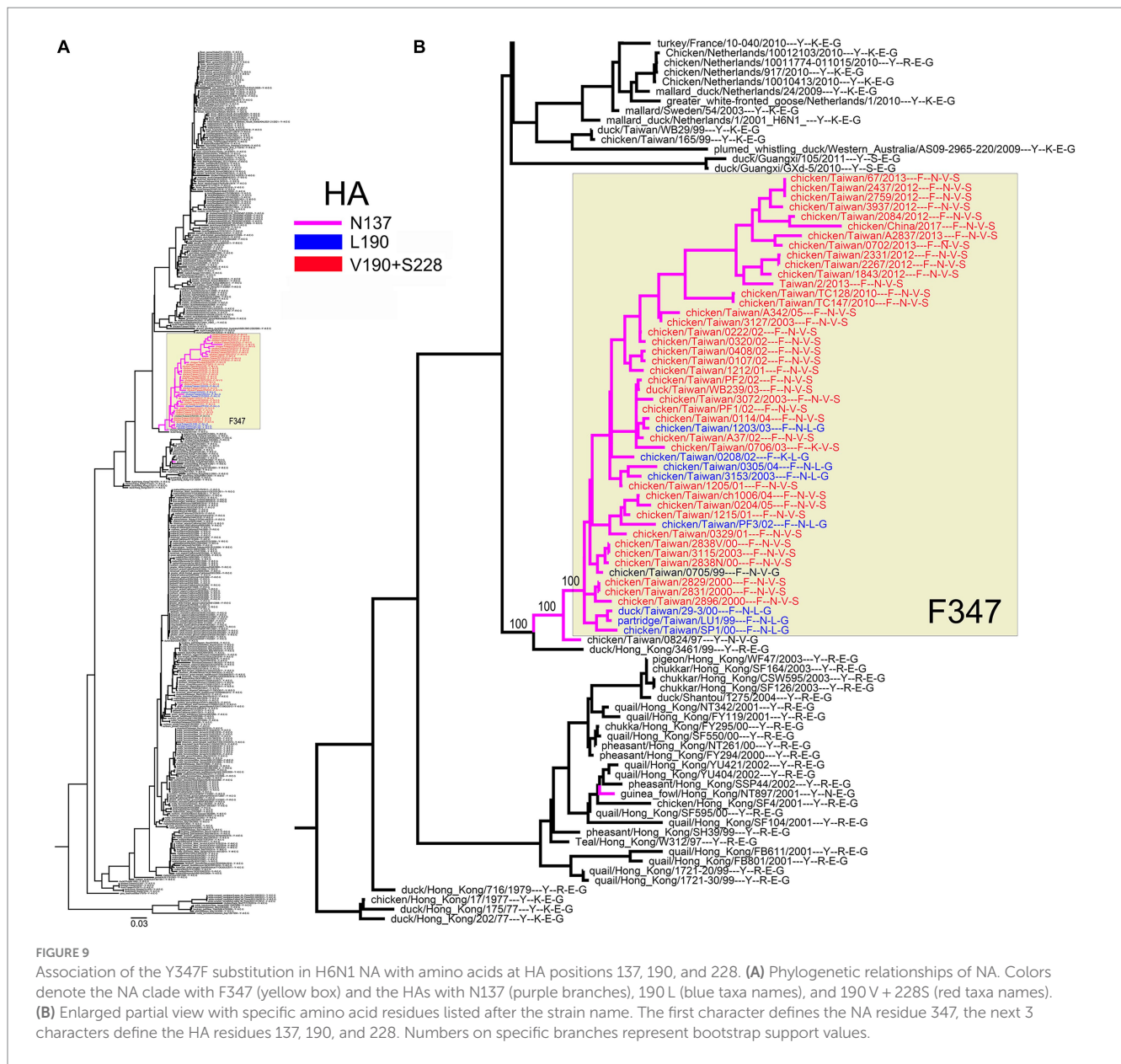
The H1N1/1918-derived seasonal human IVs circulated in 1918–1957 and 1977–2009 ([Krammer et al., 2018](#)). The NAs of the 1918 virus and its successors isolated between 1933 and 1947 contained N347 ([Figure 8A](#)). The N347D substitution occurred around 1947, and viruses with this substitution circulated in 1948–1957 and 1977–2008. During this time, the IVs with the D347N reversion emerged on three separate occasions and circulated together with the D347 variants in 2000–2002 and 2006–2009. These observations suggest that both N347 and D347 support the replication of seasonal viruses in humans and

that occasional swapping of N and D occurs, conferring epidemiological advantage to the novel variant and leading to its expansion.

All 14 NAs of the ALS-1 lineage containing the avian-type Y347 were located at the root of the avian-like lineage ([Figure 8B](#)). The Y347N substitution occurred in 1981 and was strictly maintained thereafter. The only other substitution in the ALS-1 NAs, N347D ([Table 3](#)), was present in a cluster of two viruses isolated in 2013 (see [Global tree](#) in [Supplementary material](#)). In addition to the stable ALS lineage, two atypical H1N1 IVs, A/swine/Eire/1996 and A/swine/Saskatchewan/2002, represented independent introductions of avian IVs into swine ([Figure 8B](#); [Karasin et al., 2004](#); [Lycett et al., 2012](#)). Interestingly, these IVs contained canonical markers of HA adaptation to Neu5Ac2-6-Gal-terminated receptors, the E190D substitution (both viruses) and the Q226L substitution (A/swine/Eire/1996). The latter virus also had the Y347H substitution in NA.

### 3.2.3 Substitutions in the NA of H6N1 and H5N1 viruses are associated with substitutions in HA RBS

All H6N1 viruses with the Y347F NA substitution belonged to the specific clade of poultry IVs from Taiwan ([Figure 9A](#)). A case of human



**FIGURE 9**  
 Association of the Y347F substitution in H6N1 NA with amino acids at HA positions 137, 190, and 228. **(A)** Phylogenetic relationships of NA. Colors denote the NA clade with F347 (yellow box) and the HAs with N137 (purple branches), 190 L (blue taxa names), and 190 V + 228S (red taxa names). **(B)** Enlarged partial view with specific amino acid residues listed after the strain name. The first character defines the NA residue 347, the next 3 characters define the HA residues 137, 190, and 228. Numbers on specific branches represent bootstrap support values.

infection with the virus from this clade in 2013 prompted research into the structure and receptor-binding properties of HA [for a review, see Everest et al. (2020)]. One of the reports predicted that the substitutions E190V and G228S in the HA RBS, in combination with N137, could allow the H6N1 viruses to bind to human-type receptors (Ni et al., 2015). In this view, we generated a phylogenetic tree of the H6N1 NA annotated with amino acids 137, 190, and 228 of the corresponding HAs (Figure 9B). According to the phylogeny, the Y347F substitution emerged and was fixed in parallel with unique substitutions in HA, R137N and either E190L or E190V + G228S.

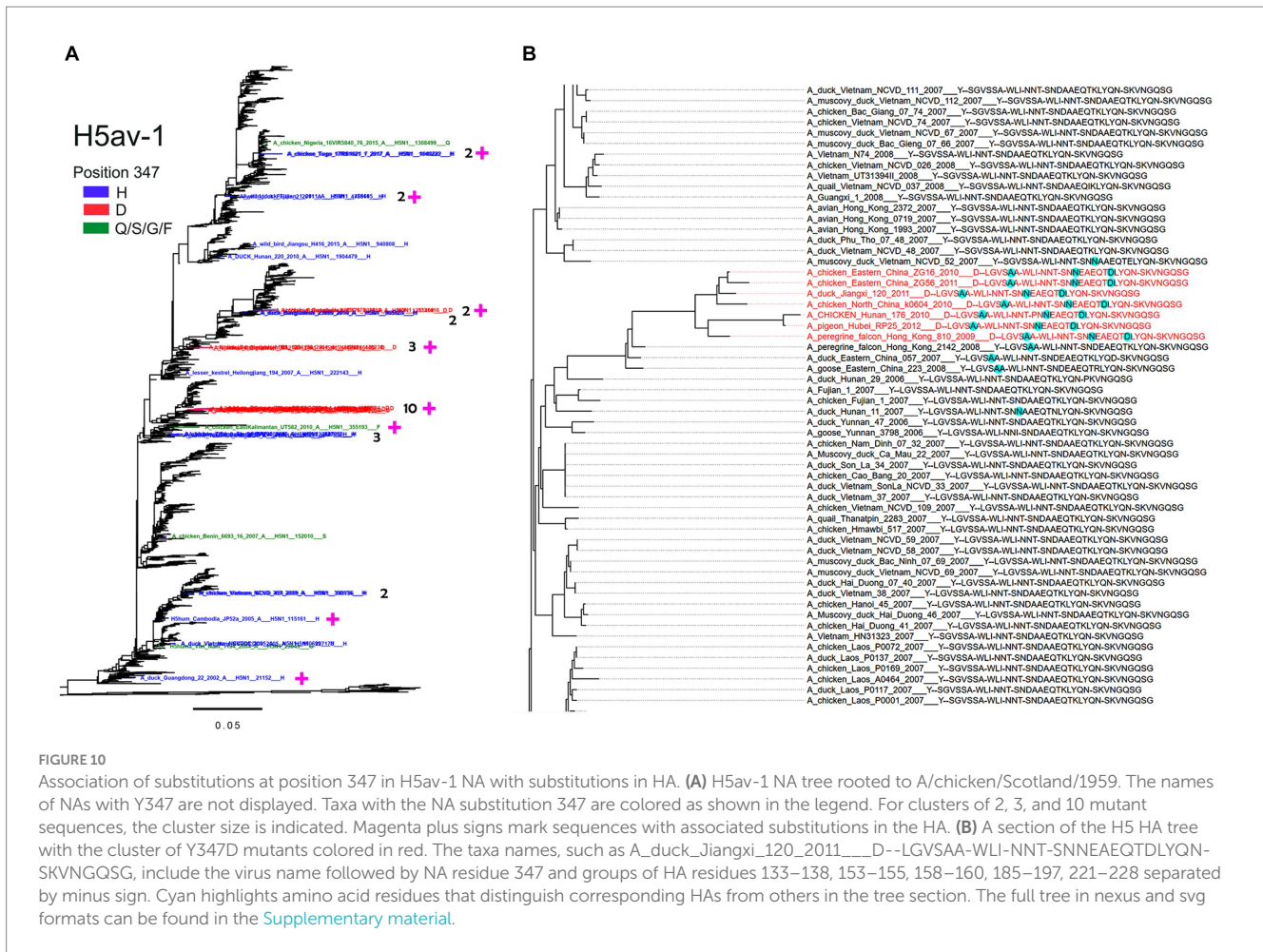
In the H5av-1 group, the non-Y NA variants were scattered across different clades of Gs/Gd-like IVs (Figure 10A). Occasional clusters of the Y347H and Y347D mutants containing from 2 to 10 sequences were observed, suggesting limited expansion of these mutants. Variants having Q, E, and S were represented by a single sequence each. Two H5hum mutants (Y347H, Y347G) and one H5mam mutant (Y347H) clustered with avian-derived NAs that had Y347 (see the trees in

Supplementary material). We next analyzed the phylogeny of H5 HA sequences that included annotations of NA residue 347 and groups of HA residues, that form the RBS. As an example, the part of the phylogeny shown in Figure 10B demonstrates that the cluster of viruses with D347 in the NA differs from the close Y347-containing IVs by 3 amino acid substitutions in HA, namely S137A, D187N, and K193D. Using this analysis, we found association between NA substitutions and HA substitutions in roughly half of the H5N1 virus clusters and single sequences examined (Figure 10A; Table 4). Remarkably, substitutions at most of these positions have been previously shown to affect antigenic and receptor-binding properties of H5 HA (Paulson and de Vries, 2013; Koel et al., 2014; Suttie et al., 2019).

### 3.2.4 Selection pressure analyses

These analyses were performed to determine, whether site 347 evolved under natural selection pressure and whether pressures varied depending on the virus host species and/or phylogenetic lineage.





**FIGURE 10**  
 Association of substitutions at position 347 in H5av-1 NA with substitutions in HA. **(A)** H5av-1 NA tree rooted to A/chicken/Scotland/1959. The names of NAs with Y347 are not displayed. Taxa with the NA substitution 347 are colored as shown in the legend. For clusters of 2, 3, and 10 mutant sequences, the cluster size is indicated. Magenta plus signs mark sequences with associated substitutions in the HA. **(B)** A section of the H5 HA tree with the cluster of Y347D mutants colored in red. The taxa names, such as A\_duck\_Jiangxi\_120\_2011\_D-LGVSA-WLI-NNT-SNNEAETDLYQN-SKVNGQSG, include the virus name followed by NA residue 347 and groups of HA residues 133–158, 153–155, 158–160, 185–197, 221–228 separated by minus sign. Cyan highlights amino acid residues that distinguish corresponding HAs from others in the tree section. The full tree in nexus and svg formats can be found in the [Supplementary material](#).

**TABLE 3** Prevalence of amino acids at position 347 of N1 NA.

Virus group	SeqN	N	D	Y	F	H	K	Q	S	T	G	%
Avian HxN1 (x = 1–4, 7–12)	918			918								0.00
Avian H6N1	464			410	54							<b>12.6</b>
H5av-1	2,627		15	2,592	1	17		1	1			<b>1.33</b>
H5av-2	2,315			2,312		3						0.13
H5hum	334			332		1					1	0.60
H5mam	186			185		1						0.54
ALS-1	1,588	1,572	2	14								<b>1.07</b>
ALS-2	1,269	1,262	1			1	5					0.55
Classical swine	3,174	3,167	7									0.22
Seasonal	2,551	1,728	823									<b>32.3</b>
H1N1pdm, all	32,752	32,723	12	1		2	11		2	1		0.089
H1N1pdm, representative	2,568	2,566	1				1					0.078

SeqN, number of sequences analyzed; %, percentage of amino acids other than Y in avian IVs and other than N in human and swine IVs (percentages greater than 1 are shown in bold).

Using the combined alignment of all sequences ( $N = 17,856$ ) we found that site 347 is evolving under overall (pervasive) purifying selection (FEL  $dN/dS = 0.20$ , value of  $p$  for non-neutral evolution  $< 10^{-20}$ ). However, a more sensitive method (MEME) identified that

this site also experienced episodic diversifying selection (MEME  $dN/dS = 8.7$ , proportion assigned to  $dN/dS > 1 = 10\%$ , value of  $p$  for episodic positive selection = 0.025). Next, using the combined tree, we estimated  $dN/dS$  values for subsets of branches partitioned by

TABLE 4 Substitutions in H5 HA associated with NA substitutions at position 347.

Virus	347	N	Substitutions
A/duck/Guangdong/22/2002	H	1	S133L
A/Cambodia/JP52a/2005	H	1	T192I
A/peregrine falcon/Hong Kong/810/2009	D	6	S137A, D187N, K193D*
A/chicken/East Kalimantan/UT581/2010	F	2	L154I, S159N
A/wild duck/Fujian/1/2011	H	2	Q196K + R227S
A/chicken/Bangladesh/3012/2011 <sup>†</sup>	D	1	K193Q
A/chicken/Bangladesh/11RS-1984-30/2011 <sup>†</sup>	D	1	S185P, D187N, K193Q
A/chicken/Bangladesh/08C178/2016	D	2	D158N (G+), D187N, K193T
A/chicken/Togo/17RS1021-7/2017	H	2	S221P

N, number of HA sequences in the cluster; G+, substitution creates glycosylation sequon; \*, D193 was not found in any other H5 HA sequence analyzed; <sup>†</sup>, NAs belong to the same cluster of 3 NAs in Figure 10.

TABLE 5 Evolutionary analysis of site 347.

Virus group	BN	dN/dS (CI)	p	Inferred substitutions*
Avian HxN1 (x = 1–4, 7–12)	1,211	0.00 (0.00–0.29)	4×10 <sup>-6</sup>	Y(5)
Avian H6N1	643	0.14 (0.00–0.89)	0.03	Y(2), Y-F(1)
H5av-1	3,685	0.72 (0.36–1.27)	0.66	Y(8), Y-H(10), Y-D(3), Y-F(1), Y-Q(1), Y-S(1)
H5av-2	2,993	0.21 (0.02–0.86)	0.03	Y(1), Y-H(2)
H5hum + H5mam	618	1.25 (0.24–3.62)	0.68	Y-G(1), Y-H(2)
Classical swine	4,866	0.11 (0.03–0.29)	9×10 <sup>-10</sup>	N(8), N-D(5)
ALS-1	2,603	0.02 (0.00–0.10)	4×10 <sup>-6</sup>	N(16), N-D(1)
ALS-2	1,787	0.23 (0.06–0.61)	7×10 <sup>-4</sup>	N (5), N-D(1), N-K(4)
Seasonal	3,383	0.42 (0.14–0.93)	0.03	N(9), D(2), D-N(5), N-D(3)
H1N1pdm, representative	3,295	0.09 (0.01–0.36)	2×10 <sup>-6</sup>	N(5), N-D(1), N-K(1)
Unlabeled	1,570	0.24 (0.05–0.70)	0.035	Y(7), N(1), D-N(1), N-D(1), N-Y(1), Y-N(1)

dN/dS is estimated using FEL, and profile likelihood confidence intervals (CI, 99%) are shown. The value of p for non-neutral evolution is obtained under a likelihood ratio test, where only the dN/dS ratio for this group is constrained to be 1 in the combined tree with all sequences; the Holm-Bonferroni correction for multiple testing is applied. Substitutions are inferred based on the joint maximum likelihood ancestral codon reconstruction. BN, number of branches tested; P, value of p for dN/dS ≠ 1; \*, X(n), n synonymous substitutions for amino acid X; X-Z(n), n non-synonymous substitutions from amino acid X to amino acid Z.

groups as described above (Figure 7; Table 3) (Note, that we placed H5N1 sequences isolated from mammals or humans in one group to improve statistical power, given the relatively small sizes of each group). We found that 8 out of 10 studied groups of branches evolved subject to purifying selection ( $p < 0.05$ , LRT with Holm-Bonferroni correction for multiple testing) (Table 5). Overall, unlabeled branches were also subject to purifying selection. As only three substitutions have been accumulated on the combined H5hum + H5mam lineages, there was insufficient statistical power to discern selection. H5av-1 branches encompassed considerably more variation, with 8 synonymous and 16 non-synonymous substitutions, consistent with neutrality (Table 5). However, there was a significant difference ( $p = 0.05$ , LR test) between the evolutionary regimes along internal branches ( $dN/dS = 1.3$ ), which, by necessity, represent some host-to-host transmissions, and terminal branches ( $dN/dS = 0.4$ ), which often reflect within-host evolution. Thus, while we cannot reject the hypothesis of neutrality on all H5av-1 branches, an observation that  $dN/dS$  is significantly higher along internal H5av-1 branches, enriched for evolution between hosts (and host species), is consistent with a functional role of this site. We next ran MEME, which is capable of

detecting episodic diversifying selection (EDS) affecting only a proportion of branches. Although we detected EDS for the whole NA alignment, no individual group provided statistical support for EDS, likely due to loss of power and relatively few substitutions compared to those accumulated along the entire tree.

We also interrogated the entire NA alignment for evidence of sites that may be co-evolving with site 347 using a Bayesian Graphical Model phylogenetic method (Poon et al., 2007). No evidence of interactions among site 347 and other sites in NA was found.

In summary, negative selection within most groups of viruses indicates that the amino acids at position 347 specifically adapted to particular hosts are maintained, whereas episodic positive selection across the entire tree is consistent with the hypothesis that adaptation occurs during or following some host switches.

## 4 Discussion

The N1 NAs of avian IVs hydrolyze the Neu5Ac2–6Gal linkage much less efficiently than the Neu5Ac2–3Gal linkage. In contrast, N1

NAs of swine and human IVs show less restricted linkage specificity due to increased activity against the 2–6 linkage and diminished activity against the 2–3 linkage (Mochalova et al., 2007; Gerlach et al., 2012). Previous molecular dynamics simulation studies on N1 NAs from phylogenetically distant avian and human IV isolates suggested that amino acid 347 contributed to the specificity of the enzyme for the Neu5Ac-Gal linkage type (Raab and Tvaroska, 2011; Jongkon and Sangma, 2012; Phanich et al., 2019). Here, we explored this hypothesis by modeling the interaction between sialoglycan substrates and two N1 NAs that differed only by amino acid 347.

Comparison of the complexes of avian-type NA/Y with 3S and 6S revealed the following differences. First, 3S bound to NA/Y in one of its solution-populated conformations, while 6S was forced to adopt a novel conformation that was not populated in its unbound state (Figure 2). Second, the Neu5Ac moiety of the bound 3S was located in the center of the NA catalytic pocket, with optimal contact distances of Neu5Ac to the arginine triad and E276 [residues known to make the major contributions to the binding energy (Taylor and von Itzstein, 1994)], and to the key catalytic residues, such as Y406. In contrast, the Neu5Ac moiety of 6S was tilted in the complex, resulting in a network of contacts that differed from those in both 3S-NA/Y and crystal structures of NA complexes with free Neu5Ac (Table 1; Figures 3A,B). Third, the estimated binding free energy was less favorable in the case of 6S (Table 2). Fourth, catalysis-promoting distortion of the Neu5Ac ring from chair to pseudo-boat was observed in the 3S-NA/Y complex, with little, if any, distortion evident in the 6S-NA/Y complex (Figure 5). Finally, the solvation of the glycosidic bond was less efficient in the 6S-NA/Y complex (Figure 6). All of these differences are expected to decrease the binding avidity and hydrolysis efficiency of 6S in comparison to 3S. Remarkably, the Y347N substitution alleviated most of the mentioned adverse effects by allowing the binding of 6S to NA/N in a solution-populated conformation, improving the orientation of the Neu5Ac moiety and the solvation of the glycosidic linkage, and considerably enhancing the free energy of binding. At the same time, replacement of Y with N negatively impacted binding characteristics of 3S. These findings advance and substantiate previous notions about the significance of amino acid 347 for the NA recognition of the Neu5Ac-Gal linkage. However, substrate binding represents only the first step of the NA-mediated catalysis (Taylor and von Itzstein, 1994). Although catalytic activity correlates with binding affinity (Mochalova et al., 2007; Garcia et al., 2013), other steps, such as donation of the proton from the solvent, formation of the endocyclic sialoside cation intermediate, and release of the asialic part of the substrate may also depend on the type of the Neu5Ac-Gal linkage. Therefore, further analyses of the catalytic activity of the 347 mutants are required to refine the predictions made by modeling.

Our MD simulation data suggest the following molecular mechanisms for the effect of residue 347 on substrate specificity of the NA. We found that optimal binding of 3S to NA/Y depends on favorable polar interactions between the 4-OH group of Gal and the OH group of Y347. In this conformation of bound 3S, the 2-hydroxyl of Gal forms an H-bond with the main chain carbonyl group of V149, while GlcNAc makes a van der Waals contact with the side chain of V149 (Figures 1A, 3A; Table 1). In the case of 6S, the Neu5Ac2–6Gal moiety approaches the side chain of Y347 with the apolar side of the Gal residue. To avoid a steric conflict and unfavorable polar-apolar

interactions between Gal and the OH group of Y347, the glycan is shifted toward loop 150; this shift negatively affects interactions with key catalytic residues and reduces binding avidity. In contrast to NA/Y, the less bulky side chain of N347 in the NA/N variant allows accommodation of the Neu5Ac2–6Gal moiety in the catalytic pocket in one of its solution-populated conformations. Furthermore, the 6S-NA/N complex is stabilized by polar contacts of the 3-OH and 4-OH groups of Gal with the side chain of N347 and by interactions of GlcNAc and Gal2 with P431 and Q430, respectively (Table 1). The unfavorable effect of the Y347N substitution on NA complex with 3S can be explained by the loss of stabilizing polar interactions between Gal and residue 347. To compensate for this loss, the glycan tilts and shifts toward loop 150, increasing interactions of the Gal and GlcNAc residues with V149 (Table 1).

The initial notion about the association of amino acid 347 with the IV host range (Fanning et al., 2000) was based on the analysis of a small number of N1 NAs and, to the best of our knowledge, has not been followed by systematic analysis. To fill this gap, we assessed the identity of residue 347 in all currently available N1 NA sequences of IVs in different host species. Remarkably, with two exceptions, Y347 was perfectly conserved in avian IVs regardless of their HA subtype, while N347 was highly conserved in classical swine, avian-like swine, and H1N1pdm virus lineages (Table 3). Natural selection analyses based on dN/dS ratios suggested that position 347 was under pervasive purifying selective pressure in these avian and mammalian viruses (Table 5). The NAs of the avian-origin 1918 pandemic virus and its successors isolated in 1933–1947 also contained N347 (Figure 8A). Combined with the modeling data, these results indicate that efficient cleavage of 2–3-linked sialoglycans associated with Y347 is essential for avian IV fitness, and that the Y347N substitution and resulting changes in substrate specificity of NA are essential for avian virus adaptation to humans and pigs. This pattern correlates perfectly with changes in HA receptor-binding specificity during avian-to-human and avian-to-swine transmission (Matrosovich et al., 2006; Shi et al., 2014; Thompson and Paulson, 2020; Liu et al., 2023).

Interestingly, a functional similarity can be observed between Y347 of N1 NA and conserved Q226 of the avian HA. First, both Q226 of HA and Y347 of NA play important roles in protein binding to one of the solution-dominant conformations of Neu5Ac2–3Gal-terminated receptors via interactions with the 4-hydroxyl group of Gal (Supplementary Figure S3). Second, both residues hinder the accommodation of the Gal residue in solution-dominant conformers of Neu5Ac2–6Gal-terminated receptors. Third, the Q226L substitution in avian influenza viruses of diverse HA subtypes occurs during the virus adaptation to humans and swine and alters its preference for the type of Neu5Ac-Gal linkage.

The first viruses of the ALS lineage isolated in 1979 carried substitutions E190D and/or G225E in the HA and bound to both Neu5Ac2–3Gal- and Neu5Ac2–6Gal-terminated receptors. The binding preference of the virus for 6-linked receptors and transmissibility in pigs under experimental conditions increased over time (Ito et al., 1998; Matrosovich et al., 2000; Su et al., 2021). Because the Y347N substitution appeared and became fixed after 1981 (Figure 8B), we conclude that it was not essential for the initial avian-to-swine transmission and was selected at a later stage, likely, as a compensation for the increased HA avidity for Neu5Ac2–6Gal.

Seasonal human IVs exhibited a distinct pattern of NA evolution with an initial circulation of N347-containing IVs, their replacement



by the D347 variants, and the periodic emergence of new N347-containing lineages (Figure 8A). The D347N substitution was previously found to increase NA avidity and catalytic activity with respect to the synthetic substrate MUNANA (Collins et al., 2009; Rameix-Welti et al., 2011; Duan et al., 2014). Collins et al. (2009) observed this effect in IVs from three independent clades of D-to-N mutants. As suggested by the authors, the substitution increased substrate binding by eliminating unfavorable electrostatic interactions between the negatively charged carboxylic groups of Neu5Ac and D347. It has been speculated that the exchange of N and D at position 347 and corresponding changes in NA catalytic activity reflect epistatic interactions of amino acid 347 with other amino acids of NA and/or with the HA (Collins et al., 2009; Duan et al., 2014). For example, either D or N may be selected to compensate alterations of NA catalytic activity and/or HA receptor-binding activity of the antigenic drift variants of these proteins. Although we found no evidence of interactions between residue 347 and other sites in NA, further studies are needed to test this hypothesis and to characterize underlying mechanisms of N/D switching in at the NA position 347 in seasonal IVs.

Most avian IVs bind to Neu5Ac2–3Gal-terminated receptors, however, avian viruses may differ in the ability to recognize the sub-terminal saccharide parts of the sialoglycans. Thus, mallards and other dabbling ducks (*Anatinae*) carry all IV subtypes except H13 and H16 and are thought to play a key role in the maintenance of IVs in nature (Fouchier and Munster, 2009; Verhagen et al., 2021). Duck-adapted IVs exhibit similar receptor-binding traits, in particular, share a preference for binding to Neu5Ac2–3Gal1–3/4GlcNAc-containing receptors, with low tolerance for fucosylation of the GlcNAc moiety [(Gambaryan et al., 2012, 2018) and references therein]. Observed conservation of the NA residue Y347 in avian IVs with HA subtypes H1–H4 and H7–H12 (Tables 3, 5) is consistent with the high conservation of the HA RBS and receptor specificity of IVs in ducks.

Adaptation of aquatic bird viruses to land-based gallinaceous birds, such as chicken and quail, is often accompanied by substitutions in the HA and alteration of the receptor-binding specificity. In contrast to duck IVs, poultry-adapted viruses typically show high-avidity binding to Neu5Ac $\alpha$ 2–3Gal $\beta$ 1–4GlcNAc-terminated receptors containing fucose and/or sulfate at the GlcNAc residue. Moreover, some of these viruses, for instance, certain H9N2 and H7N9 lineages, acquire substitutions at the conserved positions of the RBS, such as 190, 225, and 226, which facilitate HA binding to Neu5Ac2–6Gal-terminated receptors [for reviews, see Matrosovich et al. (2008), Sriwilajaroen and Suzuki (2020), Thompson and Paulson (2020), and Zhao and Pu (2022)]. Remarkably, two groups of poultry-adapted IVs studied here differed from the other avian IVs by the presence of substitutions at NA position 347.

IVs with the H6 subtype HA have a distinctive receptor-binding specificity (Gambaryan et al., 2018) and the broadest host range in birds compared to other IV subtypes (Everest et al., 2020). The Y347F substitution emerged on a single occasion in the chicken H6N1 viruses in parallel with the substitutions at HA positions 137, 228 and/or 190 (Figure 9; Table 5). There is no consensus between different research groups about the effect of these substitutions on receptor-binding specificity. Two studies concluded that the substitutions increase HA binding to Neu5Ac2–6Gal-terminated receptors (Ni et al., 2015; Wang et al., 2015), and one study showed that the substitution E190V increased HA binding to sulfated species of Neu5Ac2–3Gal-containing receptors (Kikutani et al., 2020). In any case, the non-conservative

substitutions in the canonical conserved positions of the avian RBS should significantly alter receptor specificity of these H6N1 IVs.

Highly pathogenic H5N1 avian IVs with the Gs/Gd-like HA emerged in 1996 in Southern China, spread to other countries and have been causing continuous outbreaks in poultry and wild birds with occasional infections of mammals including humans. These IVs diverged into multiple H5 HA clades with different NA subtypes (Lycett et al., 2019; Demirev et al., 2023; Yang et al., 2023). Due to the wide geographical spread of H5N1 IVs, their broad host range and tissue tropism, as well as immune pressure imposed by vaccination of poultry, a large number of substitutions in the vicinity of the HA RBS were selected that affected both antigenicity and receptor-binding properties of the HA (Paulson and de Vries, 2013; Koel et al., 2014; Suttie et al., 2019). Most of these substitutions likely reflect antigenic drift and/or adaptation of the HA to different species- and tissue-specific Neu5Ac2–3Gal-containing receptors in wild and domestic avian hosts (Kuchipudi et al., 2021; Zhao and Pu, 2022); some of the substitutions marginally increase HA avidity for human-type receptors. We found that the Gs/Gd-like lineage contained a number of independently emerging NA mutants (typically with H347 and D347), and that the substitutions in the NA were associated with substitutions in or near the HA RBS (Table 4). For example, all three different NA clusters with D347 shown in Figure 10A carried, among other substitutions, the HA substitution at position 193 from K to D, Q or T. A positively charged amino acid (K/R) at this site is responsible for the high-avidity binding of H5, H7, H13 and H16 subtype IVs to sulfated sialoglycans, such as Su-3SLN and Su-SLex (Gambaryan et al., 2012, 2018, and references therein). Therefore, the substitution of K193 with either a negatively charged or uncharged amino acid should have a significant effect on the receptor-binding properties of H5 HA. It is noteworthy that the Y347H NA mutants had different substitutions in the HA compared to the Y347D mutants, indicating the probable functional interplay between the HA and NA substitutions in these H5N1 IVs. We conclude that the substitutions at position 347 in the NA of both H5N1 and H6N1 were selected to alter the NA catalytic activity and restore the HA/NA balance that was disrupted by the substitutions in the HA. However, it cannot be ruled out that the NA substitutions emerged first, followed by selection of corresponding compensatory substitutions in HA.

In summary, our molecular modeling data combined with the analyses of NA and HA sequences strengthen previous observations on the role of amino acid 347 in the N1 NA recognition of the Neu5Ac-Gal linkage, predict the underlying molecular mechanisms, and suggest that substitutions at NA position 347 represent a novel marker of viral host range, interspecies transmission, and adaptive evolution. These findings call for further studies, including in-depth analyses of the catalytic activity and epistatic interactions of position-347 N1 NA mutants as well as identification of analogous host-specific substitutions within the catalytic domain of other NA subtypes.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding authors.

## Author contributions

MM conceived and planned the structure of the article and supervised the whole project. MM and MG acquired funding. SE, GR, and MG generated and analyzed MD simulation data, MM prepared the sequence datasets and performed phylogenetic analyses, SKP performed selection analyses. SE, SKP and MM analyzed the results, prepared the figures and tables and wrote the manuscript. All authors contributed to the article and approved the final version of the manuscript.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was funded by the Deutsche Forschungsgemeinschaft (German Research Foundation), project number 197785619-SFB 1021 (MM) and by G Ronzoni Foundation (SE and MG). SK was supported in part by a grant from NIH (GM144468). Open Access funding was provided by the Open Access Publishing Fund of Philipps-Universität Marburg.

## Acknowledgments

We gratefully acknowledge the Authors from the Originating laboratories responsible for obtaining the specimens and the Submitting laboratories where genetic sequence data were generated

## References

- Almond, A., Deangelis, P. L., and Blundell, C. D. (2006). Hyaluronan: the local solution conformation determined by NMR and computer modeling is close to a contracted left-handed 4-fold helix. *J. Mol. Biol.* 358, 1256–1269. doi: 10.1016/j.jmb.2006.02.077
- Baum, L. G., and Paulson, J. C. (1991). The N2 neuraminidase of human influenza virus has acquired a substrate specificity complementary to the hemagglutinin receptor specificity. *Virology* 180, 10–15. doi: 10.1016/0042-6822(91)90003-t
- Case, D., Babin, V., Berryman, J., Betz, R., Cai, Q., Cerutti, D., et al. (2014). *AMBER 14*. University of California: San Francisco.
- Chong, A. K., Pegg, M. S., Taylor, N. R., and von Itzstein, M. (1992). Evidence for a sialosyl cation transition-state complex in the reaction of sialidase from influenza virus. *Eur. J. Biochem.* 207, 335–343. doi: 10.1111/j.1432-1033.1992.tb17055.x
- Collins, P. J., Haire, L. F., Lin, Y. P., Liu, J., Russell, R. J., Walker, P. A., et al. (2009). Structural basis for oseltamivir resistance of influenza viruses. *Vaccine* 27, 6317–6323. doi: 10.1016/j.vaccine.2009.07.017
- Couceiro, J. N., and Baum, L. G. (1994). Characterization of the hemagglutinin receptor specificity and neuraminidase substrate specificity of clinical isolates of human influenza A viruses. *Mem. Inst. Oswaldo Cruz* 89, 587–591. doi: 10.1590/s0074-02761994000400015
- Dai, M., Guo, H., Dortmans, J. C., Dekkers, J., Nordholm, J., Daniels, R., et al. (2016). Identification of residues that affect oligomerization and/or enzymatic activity of influenza virus H5N1 neuraminidase proteins. *J. Virol.* 90, 9457–9470. doi: 10.1128/JVI.01346-16
- de Vries, E., Du, W., Guo, H., and de Haan, C. A. M. (2020). Influenza A virus hemagglutinin-neuraminidase-receptor balance: preserving virus motility. *Trends Microbiol.* 28, 57–67. doi: 10.1016/j.tim.2019.08.010
- Demirev, A. V., Park, H., Lee, K., Park, S., Bae, J.-Y., Park, M.-S., et al. (2023). Phylodynamics and molecular mutations of the hemagglutinin affecting global transmission and host adaptation of H5Nx viruses. *Transbound. Emerg. Dis.* 2023, 1–14. doi: 10.1155/2023/8855164
- Duan, S., Govorkova, E. A., Bahl, J., Zaraket, H., Baranovich, T., Seiler, P., et al. (2014). Epistatic interactions between neuraminidase mutations facilitated the emergence of the oseltamivir-resistant H1N1 influenza viruses. *Nat. Commun.* 5:5029. doi: 10.1038/ncomms6029
- Elli, S., Gambacorta, N., Rudd, T. R., Matrosovich, M., and Guerrini, M. (2021). MD simulation of the interaction between sialoglycans and the second sialic acid binding site of influenza A virus N1 neuraminidase. *Biochem. J.* 478, 423–441. doi: 10.1042/BCJ20200670
- Elli, S., Macchi, E., Rudd, T. R., Raman, R., Sasaki, G., Viswanathan, K., et al. (2014). Insights into the human glycan receptor conformation of 1918 pandemic hemagglutinin-glycan complexes derived from nuclear magnetic resonance and molecular dynamics studies. *Biochemistry* 53, 4122–4135. doi: 10.1021/bi500338r
- Everest, H., Hill, S. C., Daines, R., Sealy, J. E., James, J., Hansen, R., et al. (2020). The evolution, spread and global threat of H6Nx avian influenza viruses. *Viruses* 12:673. doi: 10.3390/v12060673
- Fanning, T. G., Reid, A. H., and Taubenberger, J. K. (2000). Influenza A virus neuraminidase: regions of the protein potentially involved in virus-host interactions. *Virology* 276, 417–423. doi: 10.1006/viro.2000.0578
- Fouchier, R. A., and Munster, V. J. (2009). Epidemiology of low pathogenic avian influenza viruses in wild birds. *Rev. Sci. Tech.* 28, 49–58. doi: 10.20506/rst.28.1.1863
- Gambaryan, A. S., Matrosovich, T. Y., Boravleva, E. Y., Lomakina, N. F., Yamnikova, S. S., Tuzikov, A. B., et al. (2018). Receptor-binding properties of influenza viruses isolated from gulls. *Virology* 522, 37–45. doi: 10.1016/j.viro.2018.07.004
- Gambaryan, A. S., Matrosovich, T. Y., Philipp, J., Munster, V. J., Fouchier, R. A., Cattoli, G., et al. (2012). Receptor-binding profiles of H7 subtype influenza viruses in different host species. *J. Virol.* 86, 4370–4379. doi: 10.1128/JVI.06959-11
- Garcia, J. M., Lai, J. C., Haselhorst, T., Choy, K. T., Yen, H. L., Peiris, J. S., et al. (2013). Investigation of the binding and cleavage characteristics of N1 neuraminidases from avian, seasonal, and pandemic influenza viruses using saturation transfer difference nuclear magnetic resonance. *Influenza Other Respir. Viruses* 8, 235–242. doi: 10.1111/irv.12184
- Gasteiger, J., and Marsili, M. (1978). A new model for calculating atomic charges in molecules. *Tetrahedron Lett.* 19, 3181–3184. doi: 10.1016/s0040-4039(01)94977-9
- Gerlach, T., Kuhling, L., Uhlendorff, J., Laukemper, V., Matrosovich, T., Czudai-Matwisch, V., et al. (2012). Characterization of the neuraminidase of the

and shared via the GISAID Initiative, on which part of this research is based.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of *Frontiers*, at the time of submission. This had no impact on the peer review process and the final decision.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1309156/full#supplementary-material>

- H1N1/09 pandemic influenza virus. *Vaccine* 30, 7348–7352. doi: 10.1016/j.vaccine.2012.09.078
- Hall, T. A. (1999). “BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT” in *Nucleic acids symposium series* (Oxford: Oxford University Press)
- Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 35, 518–522. doi: 10.1093/molbev/msx281
- Hufnagel, D. E., Young, K. M., Arendsee, Z. W., Gay, L. C., Caceres, C. J., Rajão, D. S., et al. (2023). Characterizing a century of genetic diversity and contemporary antigenic diversity of N1 neuraminidase in influenza A virus from north American swine. *Virus Evol.* 9:15. doi: 10.1093/ve/vead015
- Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graph.* 14, 33–38. doi: 10.1016/0263-7855(96)00018-5
- Ito, T., Couceiro, J. N., Kelm, S., Baum, L. G., Krauss, S., Castrucci, M. R., et al. (1998). Molecular basis for the generation in pigs of influenza A viruses with pandemic potential. *J. Virol.* 72, 7367–7373. doi: 10.1128/JVI.72.9.7367-7373.1998
- Janakiraman, M. N., White, C. L., Laver, W. G., Air, G. M., and Luo, M. (1994). Structure of influenza virus neuraminidase B/Lee/40 complexed with sialic acid and a dehydro analog at 1.8-Å resolution: implications for the catalytic mechanism. *Biochemistry* 33, 8172–8179. doi: 10.1021/bi00193a002
- Jongkon, N., and Sangma, C. (2012). Receptor recognition mechanism of human influenza A H1N1 (1918), avian influenza A H5N1 (2004), and pandemic H1N1 (2009) neuraminidase. *J. Mol. Model.* 18, 285–293. doi: 10.1007/s00894-011-1071-y
- Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79, 926–935. doi: 10.1063/1.445869
- Kalyanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., and Jermini, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* 14, 587–589. doi: 10.1038/nmeth.4285
- Kandeil, A., Patton, C., Jones, J. C., Jeevan, T., Harrington, W. N., Trifkovic, S., et al. (2023). Rapid evolution of a(H5N1) influenza viruses after intercontinental spread to North America. *Nat. Commun.* 14:3082. doi: 10.1038/s41467-023-38415-7
- Karasin, A. I., West, K., Carman, S., and Olsen, C. W. (2004). Characterization of avian H3N3 and H1N1 influenza A viruses isolated from pigs in Canada. *J. Clin. Microbiol.* 42, 4349–4354. doi: 10.1128/JCM.42.9.4349-4354.2004
- Kikutani, Y., Okamoto, M., Nishihara, S., Takase-Yoden, S., Hiono, T., de Vries, R. P., et al. (2020). E190V substitution of H6 hemagglutinin is one of key factors for binding to sulfated sialylated glycan receptor and infection to chickens. *Microbiol. Immunol.* 64, 304–312. doi: 10.1111/1348-0421.12773
- Kirschner, K. N., Yongye, A. B., Tschampel, S. M., Gonzalez-Outeirino, J., Daniels, C. R., Foley, B. L., et al. (2008). GLYCAM06: a generalizable biomolecular force field, Carbohydrates. *J. Comput. Chem.* 29, 622–655. doi: 10.1002/jcc.20820
- Kobasa, D., Kodihalli, S., Luo, M., Castrucci, M. R., Donatelli, I., Suzuki, Y., et al. (1999). Amino acid residues contributing to the substrate specificity of the influenza A virus neuraminidase. *J. Virol.* 73, 6743–6751. doi: 10.1128/JVI.73.8.6743-6751.1999
- Koel, B. F., van der Vliet, S., Burke, D. F., Bestebroer, T. M., Bharoto, E. E., Yasa, I. W., et al. (2014). Antigenic variation of clade 2.1 H5N1 virus is determined by a few amino acid substitutions immediately adjacent to the receptor binding site. *MBio* 5, e01070–e01014. doi: 10.1128/mBio.01070-14
- Kosakovsky Pond, S. L., and Frost, S. D. W. (2005). Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol. Biol. Evol.* 22, 1208–1222. doi: 10.1093/molbev/msi105
- Krammer, F., Smith, G. J. D., Fouchier, R. A. M., Peiris, M., Kedzierska, K., Doherty, P. C., et al. (2018). Influenza. *Nat. Rev. Dis. Primers.* 4:3. doi: 10.1038/s41572-018-0002-y
- Kuchipudi, S. V., Nelli, R. K., Gontu, A., Satyakumar, R., Surendran Nair, M., and Subbiah, M. (2021). Sialic acid receptors: the key to solving the enigma of zoonotic virus spillover. *Viruses* 13:262. doi: 10.3390/v13020262
- Levine, B. G., Stone, J. E., and Kohlmeyer, A. (2011). Fast analysis of molecular dynamics trajectories with graphics processing units-radial distribution function Histogramming. *J. Comput. Phys.* 230, 3556–3569. doi: 10.1016/j.jcp.2011.01.048
- Liu, M., van Kuppeveld, F. J., de Haan, C. A., and de Vries, E. (2023). Gradual adaptation of animal influenza A viruses to human-type sialic acid receptors. *Curr. Opin. Virol.* 60:101314. doi: 10.1016/j.coviro.2023.101314
- Liu, W. J., Wu, Y., Bi, Y., Shi, W., Wang, D., Shi, Y., et al. (2022). Emerging HxNy influenza A viruses. *Cold Spring Harb. Perspect. Med.* 12:8406. doi: 10.1101/cshperspect.a038406
- Lycett, S. J., Baillie, G., Coulter, E., Bhatt, S., Kellam, P., McCauley, J. W., et al. (2012). Estimating reassortment rates in co-circulating Eurasian swine influenza viruses. *J. Gen. Virol.* 93, 2326–2336. doi: 10.1099/vir.0.044503-0
- Lycett, S. J., Duchatel, F., and Digard, P. (2019). A brief history of bird flu. *Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci.* 374:20180257. doi: 10.1098/rstb.2018.0257
- Macchi, E., Rudd, T. R., Raman, R., Sasisekharan, R., Yates, E. A., Naggi, A., et al. (2016). Nuclear magnetic resonance and molecular dynamics simulation of the interaction between recognition protein H7 of the novel influenza virus H7N9 and glycan cell surface receptors. *Biochemistry.* 55, 6605–6616. doi: 10.1021/acs.biochem.6b00693
- Matrosovich, M. N., Gambaryan, A. S., and Klenk, H.-D. (2008). “Receptor specificity of influenza viruses and its alteration during interspecies transmission” in *Avian Influenza*. eds. H.-D. Klenk, M. N. Matrosovich and J. Stech (Basel: Karger), 134–155.
- Matrosovich, M. N., Klenk, H.-D., and Kawaoka, Y. (2006). “Receptor specificity, host range and pathogenicity of influenza viruses” in *Influenza virology: Current topics*. ed. Y. Kawaoka (Wymondham: Caister Academic Press), 95–137.
- Matrosovich, M., Tuzikov, A., Bovin, N., Gambaryan, A., Klimov, A., Castrucci, M. R., et al. (2000). Early alterations of the receptor-binding properties of H1, H2, and H3 avian influenza virus hemagglutinins after their introduction into mammals. *J. Virol.* 74, 8502–8512. doi: 10.1128/JVI.74.18.8502-8512.2000
- McAuley, J. L., Gilbertson, B. P., Trifkovic, S., Brown, L. E., and McKimm-Breschkin, J. L. (2019). Influenza virus neuraminidase structure and functions. *Front. Microbiol.* 10:39. doi: 10.3389/fmicb.2019.00039
- Miller, B. R. 3rd, McGee, T. D. Jr., Swails, J. M., Homeyer, N., Gohlke, H., and Roitberg, A. E. (2012). MMPBSA.py: an efficient program for end-state free energy calculations. *J. Chem. Theory Comput.* 8, 3314–3321. doi: 10.1021/ct300418h
- Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., et al. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* 37, 1530–1534. doi: 10.1093/molbev/msaa015
- Mochalova, L., Kurova, V., Shtyrya, Y., Korzhagina, E., Gambaryan, A., Belyanchikov, I., et al. (2007). Oligosaccharide specificity of influenza H1N1 virus neuraminidases. *Arch. Virol.* 152, 2047–2057. doi: 10.1007/s00705-007-1024-z
- Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., et al. (2009). AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* 30, 2785–2791. doi: 10.1002/jcc.21256
- Murrell, B., Wertheim, J. O., Moola, S., Weighill, T., Scheffler, K., and Kosakovsky Pond, S. L. (2012). Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* 8:e1002764. doi: 10.1371/journal.pgen.1002764
- Ni, F., Kondrashkina, E., and Wang, Q. (2015). Structural and functional studies of influenza virus A/H6 hemagglutinin. *PLoS One* 10:e0134576. doi: 10.1371/journal.pone.0134576
- Okonechnikov, K., Golosova, O., and Fursov, M. (2012). Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* 28, 1166–1167. doi: 10.1093/bioinformatics/bts091
- Paulson, J. C., and de Vries, R. P. (2013). H5N1 receptor specificity as a factor in pandemic risk. *Virus Res.* 178, 99–113. doi: 10.1016/j.virusres.2013.02.015
- Phanich, J., Threerachep, S., Kungwan, N., Rungrotmongkol, T., and Hannongbua, S. (2019). Glycan binding and specificity of viral influenza neuraminidases by classical molecular dynamics and replica exchange molecular dynamics simulations. *J. Biomol. Struct. Dyn.* 37, 3354–3365. doi: 10.1080/07391102.2018.1514326
- Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., et al. (2005). Scalable molecular dynamics with NAMM. *J. Comput. Chem.* 26, 1781–1802. doi: 10.1002/jcc.20289
- Poon, A. F. Y., Lewis, F. I., Pond, S. L. K., and Frost, S. D. W. (2007). An evolutionary-network model reveals stratified interactions in the V3 loop of the HIV-1 envelope. *PLoS Comput. Biol.* 3:e231. doi: 10.1371/journal.pcbi.0030231
- Price, M. N., Dehal, P. S., and Arkin, A. P. (2010). FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS One* 5:e9490. doi: 10.1371/journal.pone.0009490
- R Core Team (2013). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Available at: <http://www.R-project.org/>
- Raab, M., and Tvaroska, I. (2011). The binding properties of the H5N1 influenza virus neuraminidase as inferred from molecular modeling. *J. Mol. Model.* 17, 1445–1456. doi: 10.1007/s00894-010-0852-z
- Raffaini, G., Elli, S., and Ganazzoli, F. (2006). Computer simulation of bulk mechanical properties and surface hydration of biomaterials. *J. Biomed. Mater. Res. A* 77A, 618–626. doi: 10.1002/jbm.a.30670
- Raffaini, G., and Ganazzoli, F. (2006). Adsorption of charged albumin subdomains on a graphite surface. *J. Biomed. Mater. Res. A* 76A, 638–645. doi: 10.1002/jbm.a.30546
- Rameix-Welti, M. A., Munier, S., Le Gal, S., Cuvelier, F., Agou, F., Enouf, V., et al. (2011). Neuraminidase of 2007–2008 influenza A(H1N1) viruses shows increased affinity for sialic acids due to the D344N substitution. *Antivir. Ther.* 16, 597–603. doi: 10.3851/IMP1804
- Reid, A. H., Fanning, T. G., Janczewski, T. A., and Taubenberger, J. K. (2000). Characterization of the 1918 “Spanish” influenza virus neuraminidase gene. *Proc. Natl. Acad. Sci. U. S. A.* 97, 6785–6790. doi: 10.1073/pnas.100140097
- Rudiño-Piñera, E., Tunnah, P., and Lukacik, P. (2006). The crystal structure of type A influenza virus neuraminidase subtype N6 reveals the existence of two separate Neu5Ac



- binding sites. PDB ID: 1W21. Available at: <https://www.rcsb.org/structure/1W21> (Accessed October 24, 2022).
- Russell, R. J., Haire, L. F., Stevens, D. J., Collins, P. J., Lin, Y. P., Blackburn, G. M., et al. (2006). The structure of H5N1 avian influenza neuraminidase suggests new opportunities for drug design. *Nature* 443, 45–49. doi: 10.1038/nature05114
- Sassaki, G. L., Elli, S., Rudd, T. R., Macchi, E., Yates, E. A., Naggi, A., et al. (2013). Human ( $\alpha 2 \rightarrow 6$ ) and avian ( $\alpha 2 \rightarrow 3$ ) Sialylated receptors of influenza a virus show distinct conformations and dynamics in solution. *Biochemistry* 52, 7217–7230. doi: 10.1021/bi400677n
- Seeber, M., Cecchini, M., Rao, F., Settanni, G., and Caffisch, A. (2007). Wordom: a program for efficient analysis of molecular dynamics simulations. *Bioinformatics* 23, 2625–2627. doi: 10.1093/bioinformatics/btm378
- Shi, Y., Wu, Y., Zhang, W., Qi, J., and Gao, G. F. (2014). Enabling the 'host jump': structural determinants of receptor-binding specificity in influenza a viruses. *Nat. Rev. Microbiol.* 12, 822–831. doi: 10.1038/nrmicro3362
- Shtyrya, Y. A., Mochalova, L. V., and Bovin, N. V. (2009). Influenza virus neuraminidase: structure and function. *Acta Nat.* 1, 26–32. doi: 10.32607/20758251-2009-1-2-26-32
- Shu, Y., and McCauley, J. (2017). GISAI: global initiative on sharing all influenza data – from vision to reality. *Euro Surveill.* 22:494. doi: 10.2807/1560-7917.ES.2017.22.13.30494
- Spielman, S. J., Weaver, S., Shank, S. D., Magalis, B. R., Li, M., and Kosakovsky Pond, S. L. (2019). Evolution of viral genomes: interplay between selection, recombination, and other forces. *Methods Mol. Biol.* 1910, 427–468. doi: 10.1007/978-1-4939-9074-0\_14
- Sriwilajaroen, N., and Suzuki, Y. (2020). Host receptors of influenza viruses and coronaviruses-molecular mechanisms of recognition. *Vaccines (Basel)* 8:587. doi: 10.3390/vaccines8040587
- Su, W., Harfoot, R., Su, Y. C. F., DeBeauchamp, J., Joseph, U., Jayakumar, J., et al. (2021). Ancestral sequence reconstruction pinpoints adaptations that enable avian influenza virus transmission in pigs. *Nat. Microbiol.* 6, 1455–1465. doi: 10.1038/s41564-021-00976-y
- Suttie, A., Deng, Y. M., Greenhill, A. R., Dussart, P., Horwood, P. F., and Karlsson, E. A. (2019). Inventory of molecular markers affecting biological characteristics of avian influenza a viruses. *Virus Genes* 55, 739–768. doi: 10.1007/s11262-019-01700-z
- Tamura, K., Stecher, G., and Kumar, S. (2021). MEGA11: molecular evolutionary genetics analysis version 11. *Mol. Biol. Evol.* 38, 3022–3027. doi: 10.1093/molbev/msab120
- Taylor, N. R., and von Itzstein, M. (1994). Molecular modeling studies on ligand binding to sialidase from influenza virus and the mechanism of catalysis. *J. Med. Chem.* 37, 616–624. doi: 10.1021/jm00031a011
- Thompson, A. J., and Paulson, J. C. (2020). Adaptation of influenza viruses to human airway receptors. *J. Biol. Chem.* 296:100017. doi: 10.1074/jbc.REV120.013309
- Tian, J., Bai, X., Li, M., Zeng, X., Xu, J., Li, P., et al. (2023). Highly pathogenic avian influenza virus (H5N1) clade 2.3.4.4b introduced by wild birds, China, 2021. *Emerg. Infect. Dis.* 29, 1367–1375. doi: 10.3201/eid2907.221149
- Varghese, J. N., Colman, P. M., van Donkelaar, A., Blick, T. J., Sahasrabudhe, A., and McKimm-Breschkin, J. L. (1997). Structural evidence for a second sialic acid binding site in avian influenza virus neuraminidases. *Proc. Natl. Acad. Sci. U. S. A.* 94, 11808–11812. doi: 10.1073/pnas.94.22.11808
- Varghese, J. N., McKimm-Breschkin, J. L., Caldwell, J. B., Kortt, A. A., and Colman, P. M. (1992). The structure of the complex between influenza virus neuraminidase and sialic acid, the viral receptor. *Proteins* 14, 327–332. doi: 10.1002/prot.340140302
- Vavricka, C. J., Liu, Y., Kiyota, H., Sriwilajaroen, N., Qi, J., Tanaka, K., et al. (2013). Influenza neuraminidase operates via a nucleophilic mechanism and can be targeted by covalent inhibitors. *Nat. Commun.* 4:1491. doi: 10.1038/ncomms2487
- Verhagen, J. H., Eriksson, P., Leijten, L., Blixt, O., Olsen, B., Waldenstrom, J., et al. (2021). Host range of influenza a VIRUS H1 to H16 in eurasian ducks based on tissue and receptor binding studies. *J. Virol.* 95:873. doi: 10.1128/JVI.01873-20
- von Itzstein, M. (2007). The war against influenza: discovery and development of sialidase inhibitors. *Nat. Rev. Drug Discov.* 6, 967–974. doi: 10.1038/nrd2400
- Wagner, R., Matrosovich, M., and Klenk, H. D. (2002). Functional balance between haemagglutinin and neuraminidase in influenza virus infections. *Rev. Med. Virol.* 12, 159–166. doi: 10.1002/rmv.352
- Wang, F., Qi, J., Bi, Y., Zhang, W., Wang, M., Zhang, B., et al. (2015). Adaptation of avian influenza a (H6N1) virus from avian to human receptor-binding preference. *EMBO J.* 34, 1661–1673. doi: 10.15252/embj.201509060
- Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009). Jalview version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191. doi: 10.1093/bioinformatics/btp033
- Weis, A., Katebzadeh, K., Soderhjelm, P., Nilsson, L., and Ryde, U. (2006). Ligand affinities predicted with the MM/PBSA method: dependence on the simulation method and the force field. *J. Med. Chem.* 49, 6596–6606. doi: 10.1021/jm0608210
- Xu, D., Newhouse, E. I., Amaro, R. E., Pao, H. C., Cheng, L. S., Markwick, P. R., et al. (2009). Distinct glycan topology for avian and human sialopentasaccharide receptor analogues upon binding different hemagglutinins: a molecular dynamics perspective. *J. Mol. Biol.* 387, 465–491. doi: 10.1016/j.jmb.2009.01.040
- Xu, J., Davis, C. T., Christman, M. C., Rivailler, P., Zhong, H., Donis, R. O., et al. (2012). Evolutionary history and phylodynamics of influenza a and B neuraminidase (NA) genes inferred from large-scale sequence analyses. *PLoS One* 7:e38665. doi: 10.1371/journal.pone.0038665
- Xu, X., Zhu, X., Dwek, R. A., Stevens, J., and Wilson, I. A. (2008). Structural characterization of the 1918 influenza virus H1N1 neuraminidase. *J. Virol.* 82, 10493–10501. doi: 10.1128/JVI.00959-08
- Yang, J., Zhang, C., Yuan, Y., Sun, J., Lu, L., Sun, H., et al. (2023). Novel avian influenza virus (H5N1) clade 2.3.4.4b reassortants in migratory birds, China. *Emerg. Infect. Dis.* 29, 1244–1249. doi: 10.3201/eid2906.221723
- Yongkiettrakul, S., Nivitchanyong, T., Pannengpetch, S., Wanitchang, A., Jongkaewwattana, A., and Srimanote, P. (2013). Neuraminidase amino acids 149 and 347 determine the infectivity and oseltamivir sensitivity of pandemic influenza a/H1N1 (2009) and avian influenza a/H5N1. *Virus Res.* 175, 128–133. doi: 10.1016/j.virusres.2013.04.011
- Zhao, C., and Pu, J. (2022). Influence of host sialic acid receptors structure on the host specificity of influenza viruses. *Viruses* 14:2141. doi: 10.3390/v14102141