Large-Scale Computational Modeling of H5 Influenza Variants Against HA1-Neutralizing Antibodies

Colby T. Ford[1,2,3,4], **Shirish Yasa**[1,2], **Khaled Obeid**[1,2], **Sayal Guirales-Medrano**[1,2], **Richard Allen White III**[1,2,5], **and Daniel Janies**[1,2,✉]

[1]University of North Carolina at Charlotte, Center for Computational Intelligence to Predict Health and Environmental Risks (CIPHER), Charlotte, NC, USA
[2]University of North Carolina at Charlotte, Department of Bioinformatics and Genomics, Charlotte, NC, USA
[3]University of North Carolina at Charlotte, School of Data Science, Charlotte, NC, USA
[4]Tuple LLC, Charlotte, NC, USA
[5]North Carolina Research Campus (NCRC), Kannapolis, NC, USA

In June 2024, the U.S. Department of Agriculture released a report that showed over 300 samples from 2022-2024 of highly pathogenic avian influenza have been detected in mammals (1). To date, the World Health Organization reports that there have been approximately 5 humans infected with H5N1 in 2024, but the broader potential impact to human health remains unclear. In this study, we computationally model 1,804 protein complexes consisting of various H5 isolates in 1959 to 2024 against 11 HA1-neutralizing antibodies. This study shows a trend of weakening binding affinity of existing antibodies against H5 isolates over time, indicating that the H5N1 virus is evolving immune escape of our medical defenses. This study also showcases the value of high-performance computing to rapidly model protein-protein interactions at-scale for functional insights into medical preparedness.

influenza | H5 | antibodies | protein modeling | docking

Correspondence: ***djanies@charlotte.edu***

Introduction

Wild aquatic birds (e.g., order Anseriformes) are fundamental hosts of influenza viruses, which are transmitted from domestic birds (e.g., Galliformes) and mammals (orders Artiodactyla, Carnivora, Cetacea, and Primates) (2, 3). In 1996, H5N1 influenza largely occurred in anseriforms and spread to humans and chickens in Hong Kong in 1997 (3, 4). As a response to H5N1 in chickens and occasional human infection chickens were culled in Hong Kong in the period 1997 to 2011 (5). In 2002, the most common hosts of H5N1 were anseriformes with occasional transmission to galliformes and humans throughout China and South East Asia (3, 6). Since 2003, various lineages of H5N1 have spread throughout China and Hong Kong, South East Asia, Russia, North Africa, the West Bank, Gaza strip and Israel, Pakistan, Bangladesh, India, Bhutan, Nepal, Europe, Japan and South Korea, using a wide variety of hosts (7, 8). Many avian taxa (Charadriiformes, Accipitriformes, Corvidae, Ardeidae, Columbidae, and Passeriformes) as well as primate, carnivore, artiodactyl, and arthropod hosts have been infected with H5N1 (3). H5N1 infections in humans have been reported by the World Health Organization (WHO) in: Hong Kong 1997–2003, in China and Hong Kong 2003–2014; in Thailand 2003–2007; in Indonesia 2005–2012; in Nigeria in 2007; in Bangladesh 2011-2013; in Azerbaijan, Turkey, Iraq, Myanmar, Pakistan, and Djibouti in 2006-2007; in Egypt 2003-2014; in Lao PDR 2007; in Vietnam 2003–2014; in Cambodia 2003–2014, and in Canada in 2014. The document produced by the WHO has not been updated since.

However, a comprehensive review from 2023 (9) illustrates the further spread of H5N1 in animals as follows:

- **Asia:** Hong Kong, India, Bhutan, Japan, Korea, Nepal, Philippines, Taiwan
- **Europe:** Austria, Belgium, Bulgaria, Czechia, Denmark, France, Germany, Hungary, Italy, Luxemburg, Moldova, Netherlands, Poland, Romania, Russia, Slovakia, Slovenia, Spain, Sweden, Croatia, Estonia, Finland, Latvia, Luxembourg, United Kingdom, Ireland, Lithuania, Norway, Serbia, Switzerland
- **The Middle East:** Israel, Turkey
- **Africa:** Gambia, Nigeria, Senegal, Bolivia
- **North and South America:** Canada, United States, Peru, Chile, Colombia, Panama, Brazil

And in 2024, H5N1 has been found in animals in Antarctica (10). Charostad et al. (2023) also report H5N1 human infections in 2022 and 2023 in Cambodia, Chile, China, Ecuador, Spain, United Kingdom, United States, and Vietnam. More specifically, In March 2024, an outbreak was reported among U.S dairy cattle which spread amongst herds, leading to fatal infections among cats, infection into poultry, and four reported infections for dairy workers (11). From 1997, through late April 2024, 909 human H5N1 cases were reported, with 52% of cases being fatal (12).

Continued transmission of avian strains to human populations and subsequent human-to-human transmission can have a devastating effect on public heath worldwide. As the human-animal interface increases due to shrinking of natural habitats, deforestation, and increased demand for animal products; animal-human disease transmission is becoming more common (13). Current human seasonal influenza vaccines do not confer immunity against H5N1 influenza or other animal influenza A viruses (14). Thus, it is of great public health interest to discover patterns and molecular insights into the effects of H5 mutations on existing human immunity (15). In this study, we present the results of a large computational corpus of molecular docking experiments between various H5 isolates against existing HA1-neutralizing antibodies and show their changes over time.

Results

From the study set of 18,693 H5 isolates, we show a breakdown of hosts similar to what has been reported in previous studies (2, 3). As shown in Table 1, approximately 88% of the isolates are from birds (class Aves). Also, note that all 666 isolates from Primates were collected from humans.

Class	Order	Isolates	% of Total
Aves	Galliformes	5,844	31.26%
	Anseriformes	4,469	23.91%
	Other Orders	934	5.00%
	Not Specified	5,287	28.28%
	Aves Total	*16,534*	*88.45%*
Mammalia	Primates	666	3.56%
	Carnivora	191	1.02%
	Artiodactyla	57	0.30%
	Other Orders	6	0.03%
	Not Specified	360	1.93%
	Mammalia Total	*1,280*	*6.85%*
Insecta	*Insecta Total*	*3*	*0.02%*
Other	Laboratory Derived	178	0.95%
	Other / Environmental	698	3.73%
	Other Total	*876*	*4.69%*
Grand Total		**18,693**	**100%**

Table 1. Taxonomic breakdown of the isolates used in this study.

As shown in Supplementary Figure 2, the proportion of H5 isolates collected from various continents has changed over time. To date, isolates collected in 2024 are predominately from Europe, driven by the increased sequencing of infected livestock.

Sequence Analyses. The clustering of the 18,693 HA1 sequences resulted in 250 distinct clusters at ≥97% identity. Further organization of the representative sequences of each cluster, shown in Supplementary Figure 3, indicates a continuous distribution of antibody binding performance.

Isolates EPI764389 and EPI597824 are outliers in that they have poor antibody binding performance across all of the antibodies tested (average van der Waals energies >-60. Conversely, isolates EPI2429052, EPI167665, and EPI275974 show the best antibody affinity, having average van der Waals energies <-100.

Of note, antibodies 12H5, 3C11, 65C6, AVFluIgG01, and H5M9 show the best performance overall, though there are exceptions of poor performance.

Phylogenetics. The alignment used in phylogenetics consisted of 18,508 taxa and 1,899 aligned positions. TNT searches at level 1 for 1000 replicates consistently produced 4 trees of 44,550 steps. Transmission network analyses of the best scoring tree of the H5 isolates in this study show a considerable pattern of transmission between various classes of organisms. Specifically, the most common being between avian-to-mammalian transmission, of which there are over 600 apomorphic transitions across the tree. See Figure 1a.

There is also a consistent transmission between many continents across the globe, especially between Europe and North America, as shown in Figure 1b.

Worsening Binding Affinity. With respect to collection date, there are significant correlations that show a worsening in antibody binding affinity in isolates collected from orders Anseriformes, Galliformes, and Primates. See Figure 2.

However, this worsening in binding affinity does not seem localized to specific antibodies. Though the antibodies' performances are from independent distributions, as shown by the Kruskal-Wallis test in Supplementary Figure 4, the overall trend suggests that more recent isolates have mutated to better evade existing antibodies in general.

Mutational Effects. As indicated in Shi et al. (2014), there are various mutations in the HA1 receptor binding domain that enable 'host jumping'. Our results show several statistically significant differences in binding affinity of antibodies given the presence of these mutations. Of note, N158S, T160A/S/V, E190N, and G225R all result in weakened antibody binding affinity across multiple metrics. Conversely, T160K and G228S are shown to increase binding affinity in some metrics. Significant changes based on van der Waals energies and HADDOCK Score are shown in Figure 4 and all other metrics are shown in Supplementary Figure 5.

In Figure 3, subfigures A and B represent the worst and best binding structures across the experimental results, isolates EPI168674 and EPI2429052, respectively. Though the epitope is different between isolates, note the variation in quantity of polar contacts within the respective complexes.

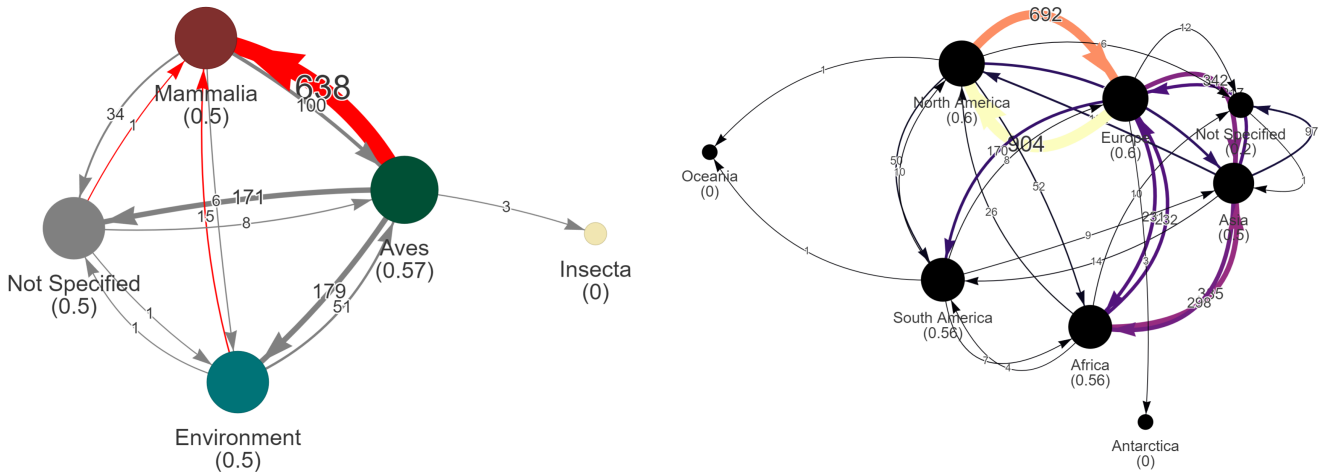
Figure 3D is an example of modest binding affinity between an isolate EPI658567 and antibody 12H5. However, this improved binding affinity compared to other isolates is not due to the G225R mutation as this residue is not in the epitope of the antigen.

Figure 3E shows the poor binding affinity of isolate EPI3178330 with antibody FLD194 due to the E190N mutation, though this residue is still forming a polar contact with the antibody CDR loop. The mutation from glutamic acid (E), a negatively charged side chain, to asparagine (N), a polar uncharged side chain.

Isolate EPI242227 is the oldest isolate in the dataset, collected in 1959. Note that this isolate contains the N158D mutation and its interaction with various antibodies results in a wide range of van der Waals energies from -52.81 (worst) to -92.84 (best, with antibody 65C3 as shown in Figure 3F).

Interfacing Residue Prevalence. An analysis of the interfacing residues in the best complexes from all 1,804 experiments shows patterns in particular residues forming polar contacts with the antibodies tested. Residues 156, 193, 222 are often interfaced (≥25% prevalence) in the antigen epitope. These are shown in red in Figure 5.

Many of the frequently interfaced residues shown in this study closely neighbor antigenic and receptor binding sites reported in Sriwilaijaroen and Suzuki (2012) Figure 2. Residues 130, 132, 158, and 159 are part of two glycosylation sites that, in 2012, were only found in seasonal H1 viruses (17).



(a) Host transmission network. **(b)** Geographic transmission network.

Fig. 1. Transmission networks, generated by StrainHub, showing apomorphic changes between (a) hosts and (b) continents. Node values in parentheses represent the source-hub ratio of that class/location. A source-hub ratio of 1 indicates that the state is always the source of the transmission. The numerical labels annotated on the edges of the graphs represent the number of apomorphic changes seen across the phylogenetic tree between metadata states.

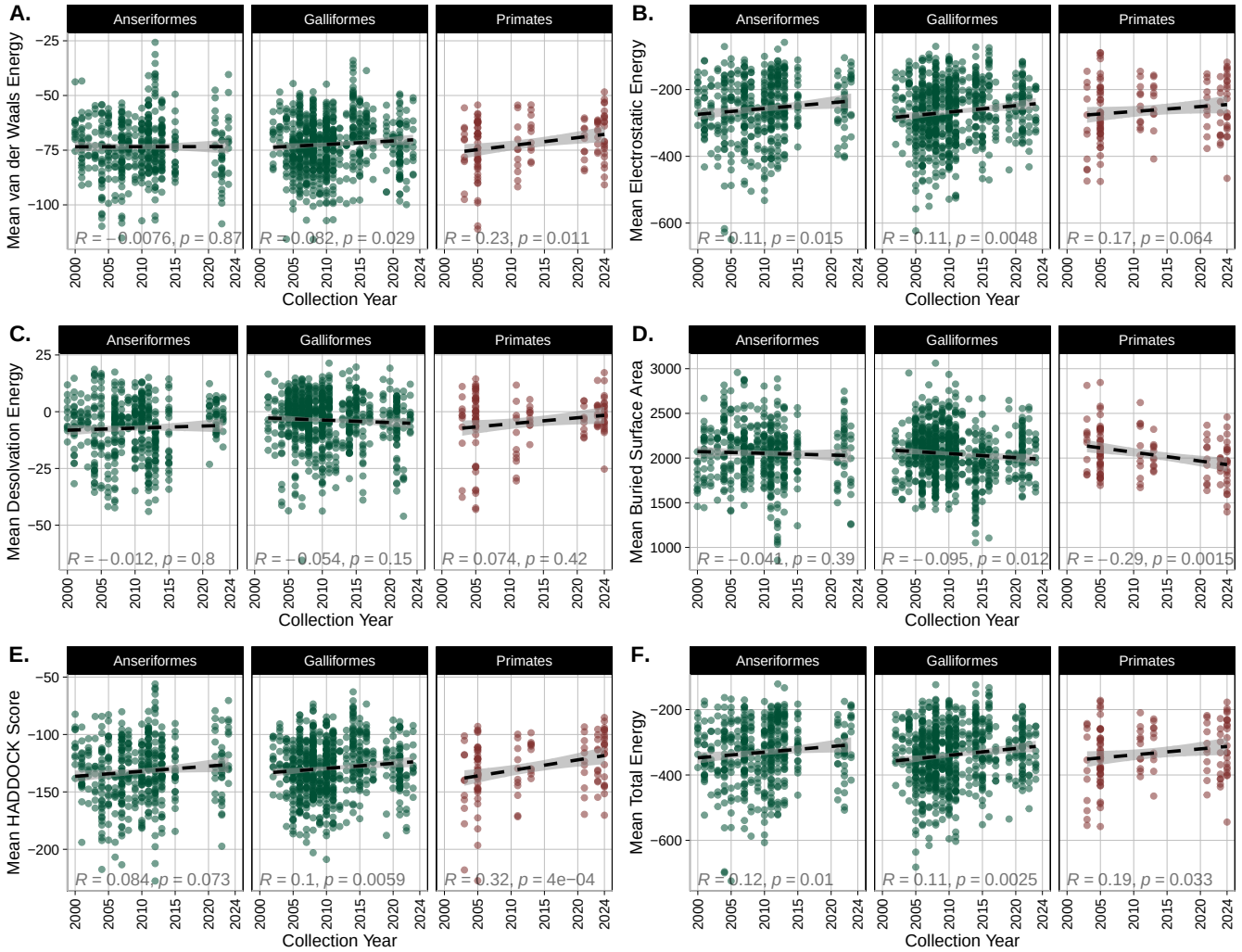


Fig. 2. Antibody binding performance metrics over time for isolates collected from Anseriformes (waterfowl), Galliformes (chickens, etc.), and Primates (humans). Note: Spearman correlations are shown based on all isolates in the data (from 1959-2024), but the plots are truncated to show collection years from 2000-2024.

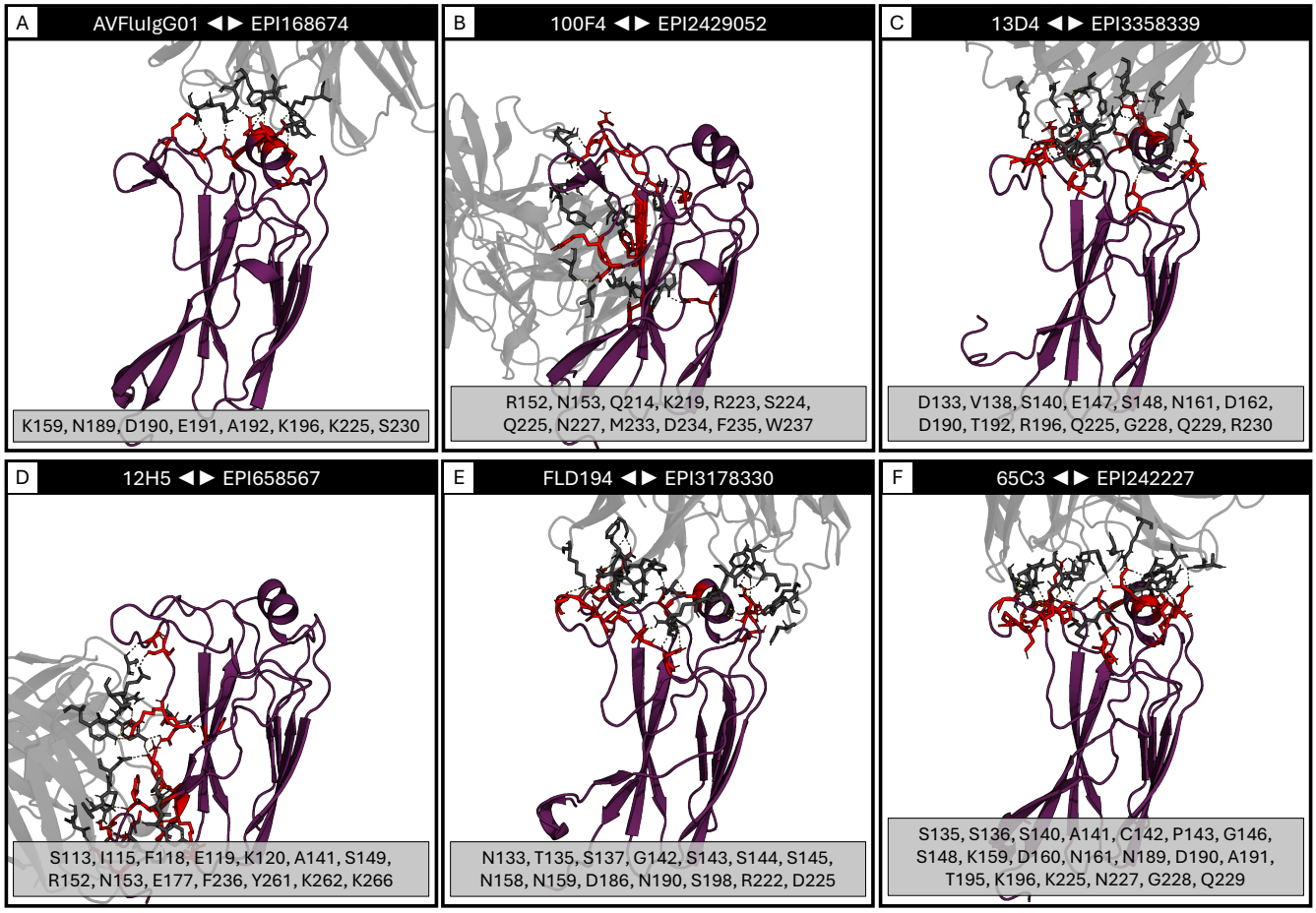


Fig. 3. Example interface renderings showing the diversity in epitopes, residues, and binding affinity. The grey cartoon structure is the Fab portion of the docked antibody and the purple structure is the HA1 antigen with sticks designating the polar contacts between them (shown as sticks). The list below each subfigure contains the interfacing residues on the antigen chain.

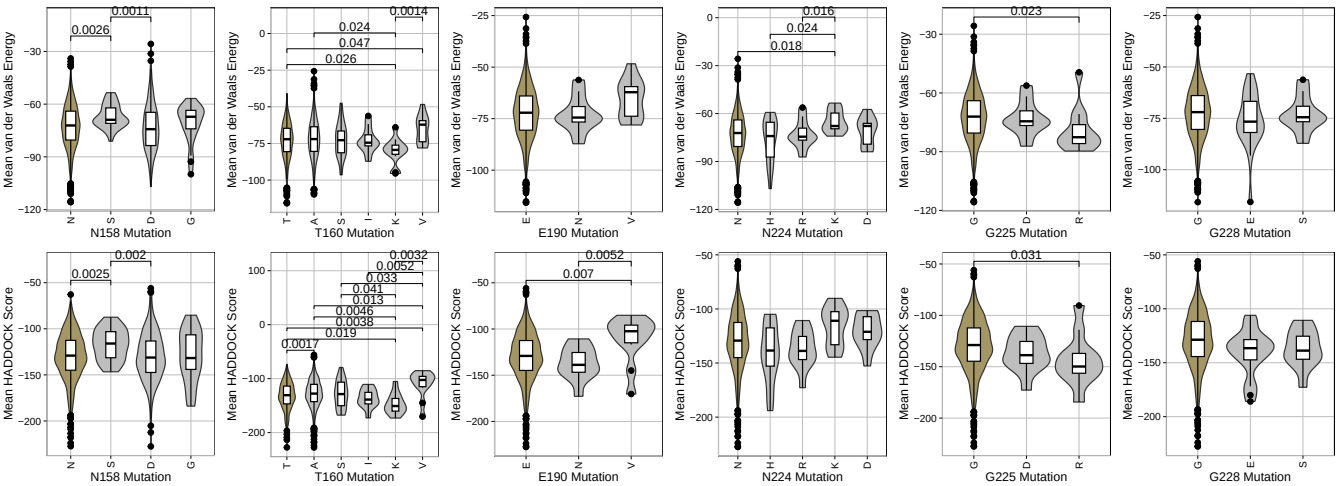


Fig. 4. The distribution of van der Waals energies and HADDOCK score docking metrics broken out by antigen mutations. The first amino acid shown on the left of each plot in gold represents the reference residue at that position as described in Shi et al. (2014). Statistical comparisons shown are significant Wilcoxon Rank Sum test p-values at the α<0.05 level.

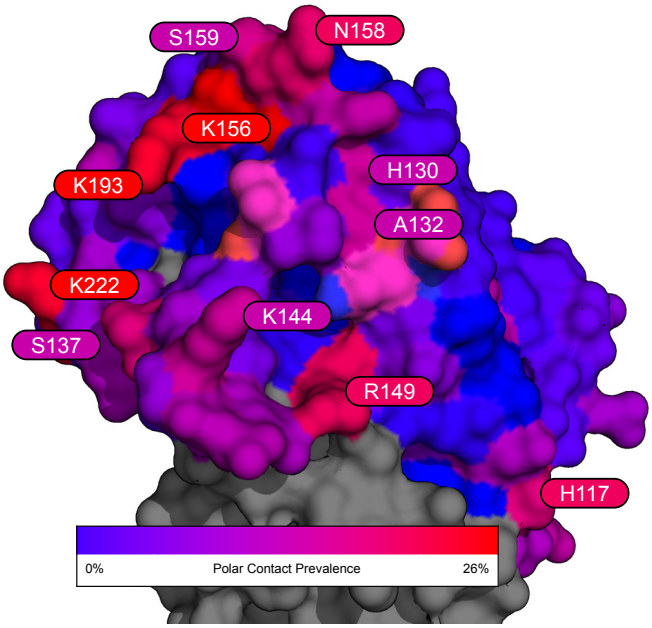


Fig. 5. Surface rendering of the HA1 globular head domain (reference PBD 2FK0) showing the prevalence of each residue to form polar contacts (within 3Å of antibody residues) across the experiments in this study. Annotated residues are those with $\geq$16% prevalence.

Discussion

Relation to Prior Studies. This *in silico* analysis yields the trend of a reduction in binding affinity for neutralizing antibodies against H5N1-designated influenza variants. This reduction in binding affinity reinforces previous studies that evolution has occurred to yield HA proteins that are elusive of antibody neutralization (18, 19). The trend observed is consistent with empirical studies and strengthens the relatively novel *in silico* approach taken within this study.

Diversity of H5N1 Interactions. As shown in Supplementary Figure 3, high disparities in antibody binding affinity exist between different sequences designated as H5N1. This disparity leads to concerns over the nomenclature of H5N1. For future clinical practice to effectively treat H5N1, clinicians must be able to more precisely identify sub-variants of H5N1 to identify antibodies that have specific neutralizing capabilities against the identified variant. For example, in Supplementary Figure 3, sequence EPI597824 has a large disparity in van der Waals energy between antibody 3C11 and AvFluIgG03. In addition, Supplementary Figure 5 demonstrates the effect of a single amino acid mutation on docking metrics. One change in an amino acid residue can lead to statistically significant differences in antibody binding affinity. Broadly categorizing a viral diversity as simply "H5N1" will lead to a lack of precision in the design of therapeutics and treatments, which may will have adverse health outcomes.

Zoonosis Analysis. In Figure 1a, a notable transition pattern is observed from avian species (class Aves) to mammals (class Mammalia), likely attributable to a mutagenic drift. Recent empirical studies have investigated the mutational dynamics of H5N1, revealing changes in the hemagglutinin (HA) protein. While Influenza prefers binding to the α2-3 sialic receptors in birds, this study demonstrated a binding affinity of H5N1 to α2-6 sialic acid receptors, predominant in mammals, at almost equal proportion (20). Concurrently, mutations in the HA protein also show reduced susceptibility to neutralization by antibodies.

Our *in silico* study aligns with these findings, indicating a progressive decrease in H5N1's binding affinity to antibodies in our isolates over time. As illustrated in Figure 2, while the binding affinity to antibodies remains relatively stable for avian isolates, there is a marked decline in affinity for primate isolates. This suggests a potential evolutionary trajectory for H5N1 towards increased infection in primate/mammal hosts, facilitated by decreased susceptibility to neutralizing antibodies.

Isolate EPI3358339. EPI3358339, a H5N2 subtype isolate, was added to this study as it is from the first ever human infection of H5N2 avian influenza seen in Mexico. Unfortunately, this strain was found in a person from Mexico who died of complications due to the infection. However, it is not yet known if the zoonotic "jump" of the isolate is cause for concern or if the individual had other comorbidities that played a role is his death. It is also not known if this case is related to recent poultry outbreaks in the area.

The experimental docking conformation between antibodies and this antigen (antibody 65C3 shown in Figure 3F) are predicted to have a relatively strong binding affinity (e.g., van der Waals energy: [-51.96, -95.45]) across all the tested antibodies.

Thus, our experiments using this isolate do not indicate that this isolate is highly mutated, though some mutations may have reduced the van der Waals and electrostatic energies of the interaction with this individual's existing antibodies, if any.

Structural Analysis. When comparing the predicted docking outputs to empirical structures, such as those listed in Table 2, we see similar binding epitopes in the the predicted docking complexes versus the empirical structures (21–28).

Furthermore, the binding conformations seen in the empirical structures often mimic the predicted complexes in the experiments in this study, though various mutations effect the binding angle, polar contacts, electrostatics, and overall affinity. In general, this supports the confidence in the predictive accuracy of the *in silico* experiments given their similarity to empirically derived structures. However, empirical studies are still needed to validate specific complexes.

Motifs of Interest and Future Research. H5N1 exists as an endemic in avian populations. This epidemic creates a pool containing vast numbers of host species in which the RNA virus evolves rapidly. It is an endemic that is difficult to diminish due to the nature of the HA protein's host receptors lacking homogeneity. Serological immunity from vaccination or prior infection in avian hosts may have yielded selective pressures in the evolution of specific mechanisms of

entry for H5N1 (29). Subsequently high infection rates lead to new mechanism of entry. Over time, serological immunity from original vaccination and/or infection is reduced, and the cycle of influenza transmission continues.

Highly conserved portions of HA are of high interest. The three primary conserved elements of the receptor-binding site (RBS) on the HA1 subunit are the 130 and 220-loops and the 190-helix (30–32). As shown in Figures 3 and 5, the antibodies docked to conserved motifs on the studied H5N1 strains, further supporting the empirical literature that initially identified their neutralizing capability. The inhibition of HA1-Host Receptor binding by RBS targeting antibodies may not however be the only mechanism of action of antibodies targeting HA to inhibit viral replication for future research. Inhibition of Neuraminidase activity via the steric hindrance provided by the antibody-HA1 complex has been demonstrated in the past (33). In the future, targeting of conserved portions of the HA1 protein whilst also inhibiting neuraminidase is a strategy that may yield substantial neutralization capacity to future therapeutics.

More recent development of multivalent mRNA-based vaccines has been successful in H5 influenza A clade 2.3.4.4b (from which there are 15 sequences used in this study) (34). The selection of high-quality sequences that elicit strong antibody responses is a complex process in mRNA vaccine development. However, the utility of *in silico* modeling as presented in this study will help to reduce the future lab workload by positing predicted candidate sequences from which vaccines can be developed.

Computational modeling of immunoprotein interactions as shown in this study and previous works (35–40) have proven to be highly effective in the understanding of the health impacts of pathogen variants. For H5 influenza, this study, along with recent preprints (41, 42), show an overall trend of worsening antibody binding and depicts the recent increase in avian-to-mammalian transmissions due to various mutations. This suggests that there is am impending danger to human health for highly pathogenic strains of H5 influenza that can infect avian and mammalian livestock and jump to humans.

Methods

This study closely follows the published methods in Tomezsko and Ford et al. (2024) and other protein modeling studies by this team (36–40). The specific workflow for this study is depicted in Figure 6.

HA1 Sequence Procurement. 18,693 influenza A H5 sequences were procured from the GISAID EpiFlu database (43, 44) on June 17, 2024. Isolate metadata, including collection date, country of origin, and host information, were also collected. We then derived taxonomic classifications from the provided host metadata along with continent of origin from the country information.

These amino acid sequences were clustered using CD-HIT v4.8.1 (45, 46). Ranging in size from 11 to 576 amino acids, the clustering resulted in 250 clusters, based on 97% identity. A representative sequence from each of these clusters was selected by the clustering program, which was then output in a FASTA file.

These representative sequences were then aligned using MUSCLE v3.8.425 (47) using default settings. Then, sequences that were of low quality, incomplete, or did not contain the desired HA1 domain were removed. This resulted in 164 amino acid sequences that were then used in the subsequent steps. These sequences were then trimmed to the HA1 receptor binding domain (approximately residues 111-269, depending on the presence of indels) (17).

Also, clusters consisting of only lab derived isolates ($n = 3$) were analyzed, but have been left out of the reported statistics and visualizations.

Structure Prediction. Structures for each of the 164 HA1 sequences were predicted using ColabFold v1.5.5 (48), a protein folding framework that uses AlphaFold2 (49) accelerated with MMseqs2 (50), with default settings. The side chains of these predicted structures were relaxed using the OpenMM/Amber method (51). The .PDB file output of the top structure (i.e., the one with the highest pLDDT[1] confidence) was selected and used for subsequent analyses.

Neutralizing Antibody Collection. Existing structures for 11 HA1-neutralizing antibodies were collected by a database search through Thera-SAbDab (52) and the Protein Data Bank (53). Each of these 11 antibody structures have an epitope on the target HA1 domain of the hemagglutinin protein, though not all share the same epitope. These antibodies and their respective PDB IDs are listed in Table 2.

Antibody ID	PDB ID	Reference	H/L Chain Subgroups	Source Information
100F4	5dur	(21)	II/I	Human Memory B-Cell, Recovered from H5N1 Infection
12H5	7fah	(22)	I/IV	Mouse, Immunized with three H1N1 strains, Humanized
13D4	6a0z	(23)	I/I	Mouse, Immunized with five H5N1 strains, Humanized
3C11	6iuv	(24)	I/II	Human Memory B-cell, Infected by H5N1 viruses
65C6	5dum	(21)	I/III	Human, Infected by H5N1 viruses
AVFluIgG01	6iut	(24)	II/I	Human, Infected by H5N1 viruses
AVFluIgG03	5dup	(21)	III/I	Human, Infected by H5N1 viruses
FLD194	5a3i	(25)	II/I	Human Memory B-cell, Recovered from H5N1 infection
FLD21.140	6a67	(26)	?/I	Human, Recovered from H5N1 Infection
H5M9	4mhh	(27)	I/IV	Mouse, Immunized with H5N1, Humanized
H5.3	4xrc	(28)	II/?	Human, Immunized with one H5N1 strain

Table 2. HA1-neutralizing antibodies.

Docking Analyses. Using HADDOCK3, a computational framework for the integrative modeling of biomolecular complexes, each antibody was docked to each antigen across the dataset. Given the 11 antibody structures and 164 HA1 structures, this resulted in 1,804 docking experiments to be performed.

Experiment Generation. HADDOCK3 inputs for each experiment were generated programmatically, defining the antibody and antigen .PDB file inputs on which to dock. Other experiment files were also copied or created programmatically that include the scripts to run the docking process and togenerate other configuration files.

[1]Predicted local distance difference test (pLDDT): An estimate of local confidence, scaled from [0, 100], where higher scores indicate higher confidence in the protein conformation.

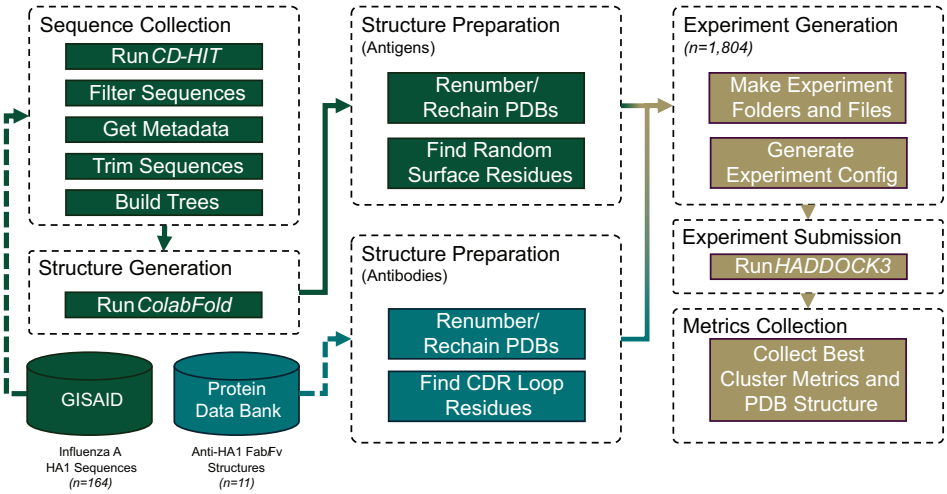


Fig. 6. Workflow diagram of the data procurement, data preparation, and analysis steps.

HADDOCK3 requires the definition of active/inactive residue restraints (AIRs) to help guide the protein docking process. To avoid biasing the docking placement of the antibody on the HA1 antigens, a random subset of surface residues were selected as "active" and were then included in the AIR file on which to dock. For the antibody structures, residues in the CDR loops were detected using ANARCI (54), a Python package for numbering antibody sequences.

Lastly, HADDOCK3 configuration files were generated programmatically, which define the input .PDB files, the output directory, and the steps of the docking process. The logic for this programmatic generation of HADDOCK3 configuration files is available in the supplementary GitHub repository.

Docking Process. HADDOCK3 provides a configurable interface for defining the individual steps of the docking process, including the rigid-body docking, flexible refinement, and solvent-based refinement along with any desired clustering and filtering steps.

For this study, we customized the published HADDOCK3 protocol for antibody-antigen modeling (55) to focus on generating the singular best cluster of docking results for each experiment and reducing excess work by the docking process. The specific steps of our HADDOCK3 configuration that was used for the experiments are shown in Table 3.

The HADDOCK3 system outputs multiple metrics for the predicted binding affinities and an output set of PDB files containing the antibody docked against the HA1 antigen. Some main metrics include:

- van der Waals intermolecular energy (*vdw*)
- Electrostatic intermolecular energy (*elec*)
- Desolvation energy (*desolv*)
- Restraints violation energy (*air*)
- Buried surface area (*bsa*)
- Total energy (*total*): $1.0vdw + 1.0elec$
- HADDOCK score: $1.0vdw + 0.2elec + 1.0desolv + 0.1air$

Note that the HADDOCK Score is a conglomerate metric used for assessing the best complexes (or best cluster of complexes) that get promoted through the various refinement iterations in the pipeline.

Computational Scalability. For this study, we utilized a Docker containerized version of HADDOCK3[2,3], which contains all of the software dependencies to allow HADDOCK3 to run more readily in a high-performance computing (HPC) environment.

HADDOCK3 was run in a Singularity container on the UNC Charlotte Orion HPC cluster on 14 nodes, each with dual 18-core Intel Xeon Gold 6154 3.00GHz CPUs (36 cores per node). The average walltime of the experiments was $14.6\bar{3}$ minutes, with the entire set of 1,804 experiments completing in under 2 days.

Post Processing. Once all experiments were completed, the metrics for each experiment were either retrieved from the CAPRI evaluation outputs (if the FCC clustering algorithm (56) reached convergence) or from the REMARK entries of the best cluster's .PDB files. These metrics were organized in a single aggregate table, representing each experiment's best cluster metrics, for subsequent visualization and statistical analyses. The full table of experiment results is available in the Supplementary Data.

Phylogenetic, Statistical, and Protein Structure Analyses. For phylogenetic analyses, laboratory derived isolates were filtered out ($n = 178$, shown in Table 1), resulting in a set of 18,515 isolates. An alignment of the input 18,515 sequences was generated using MAFFT v7.471 (57) and then a phylogenetic tree search was performed using this alignment with TNT v1.6 (58). One of the best scoring trees was used for downstream analyses. StrainHub v2.0.0 (59) was used to generate transmission networks of the phylogenetic tree by host class and continent in Figure 1.

Statistical analyses were performed using R v4.3.4 (60) and plots were generated using ggplot2 (61) and ggpubr (62). Any statistical significance reported in this study is based on a p-value threshold of $\alpha < 0.05$.

Visualizations and analyses of the protein complexes were generated using PyMOL v2.4.1 (63) and BioPandas v0.4.1 (64).

[2]Container GitHub Repository: https://github.com/colbyford/HADDOCKer
[3]Docker Hub Images: https://hub.docker.com/r/cford38/haddock

Step		Description	# Models
1	Topology Modeling	Creates CNS all-atom topology	-
2	Rigid Body Modeling $n = 200$	Rigid body energy minimization with CNS (*it0*). Samples n models.	200
3	Cluster with FCC	Calculates the contacts between chains and then calculates the FCC matrix for each model. Clusters the models based on the calculated contacts.	200
4	Select Top Clusters $n = 5, m = 10$	Selects top n clusters of m models based on HADDOCK score	50
5	Flexible Refinement	Semi-flexible refinement using a simulated annealing protocol. Uses molecular dynamics simulations in torsion angle space (*it1*)	50
6	Cluster with FCC	Calculates the contacts between chains and then calculates the FCC matrix for each model. Clusters the models based on the calculated contacts.	50
7	Select Top Cluster $n = 1, m = 10$	Selects top n clusters of m models based on HADDOCK score	10
8	Solvent Refinement	Refinement by a short molecular dynamics simulation in explicit solvent (*itw*)	10
9	Molecular Dynamics Scoring	Performs a short molecular dynamics simulation on the input models and scores them.	10
10	Cluster with FCC	Calculates the contacts between chains and then calculates the FCC matrix for each model. Clusters the models based on the calculated contacts.	$\sim$10
11	Calculate CAPRI Metrics	Calculates i-RMSD, I-RMSD, Fnat, DockQ of the top scoring models.	$\sim$10

Table 3. Descriptions of the steps of the HADDOCK3 docking configuration.

Author Contributions

Authors SY, KO, and SGM retrieved the H5 antigen sequences from GISAID and antibody structures from Protein Data Bank. SY performed the structure prediction of the antigens. Author RAW performed the clustering analysis. Authors KO and CTF curated the metadata of the H5 sequences. Author DJ performed the phylogenetic analyses. Authors SY and CTF performed the protein structure analyses. Author CTF performed the docking experiments and statistical analyses and generated all figures. All authors wrote and reviewed the manuscript.

Acknowledgements

We gratefully acknowledge all GISAID data contributors (i.e., the authors and their originating laboratories) responsible for obtaining the specimens, and their submitting laboratories for generating the genetic sequence and metadata and sharing via the GISAID Initiative, on which this research is based.

We acknowledge the following entities at the University of North Carolina at Charlotte: Academic Affairs, The Office of Research, The Center for Computational Intelligence to Predict Health and Environmental Risks (CIPHER), The Department of Bioinformatics and Genomics, The College of Computing and Informatics, and the University Research Computing group. We acknowledge the support of the Belk Family.

Data Availability Statement

All code, data, results, and additional analyses are openly available on GitHub at:
https://github.com/colbyford/Influenza_H5-Antibody_Predictions.
These data include all sequences and folded structures for the isolates and antibodies used in this study, analysis scripts, and docking metrics.

Conflict of Interest Disclosure

Author CTF is the owner of Tuple, LLC, a biotechnology consulting firm. All other authors declare no conflicts of interest.

Bibliography

1. USDA Animal and Plant Health Inspection Service. USDA APHIS Livestock and Poultry Health: Avian Influenza High Pathogenicity Avian Influenza Detections in Mammals. https://www.aphis.usda.gov/livestock-poultry-disease/avian/avian-influenza/hpai-detections/mammals, 2024. Accessed: 2024-06-20.
2. R G Webster, W J Bean, O T Gorman, T M Chambers, and Y Kawaoka. Evolution and ecology of influenza A viruses. *Microbiological Reviews*, 56(1):152–179, 1992. doi: 10.1128/mr.56.1.152-179.1992.
3. Daniel Janies, Andrew W Hill, Robert Guralnick, Farhat Habib, Eric Waltari, and Ward C Wheeler. Genomic Analysis and Geographic Visualization of the Spread of Avian Influenza (H5N1). *Systematic Biology*, 56(2):321–329, 04 2007. ISSN 1063-5157. doi: 10.1080/10635150701266848.
4. Kennedy F. Shortridge. Influenza - a continuing detective story. *The Lancet*, 354:SIV29, Dec 1999. ISSN 0140-6736. doi: 10.1016/S0140-6736(99)90372-0.
5. The New York Times. Hong kong culls chickens after bird flu is found. *The New York Times*, 2011. Accessed: 2024-06-25.
6. Stephanie Sonnberg, Richard J. Webby, and Robert G. Webster. Natural history of highly pathogenic avian influenza h5n1. *Virus Research*, 178(1):63–77, 2013. ISSN 0168-1702. doi: https://doi.org/10.1016/j.virusres.2013.05.009. H5N1.
7. Steven L. Salzberg, Carl Kingsford, Giovanni Cattoli, David J. Spiro, Daniel A. Janies, Magdi M. Aly, and Ilaria Capua. Genome Analysis Linking Recent European and African Influenza (H5N1) Viruses. *Emerging Infectious Diseases*, 13(5):713, 2007. doi: 10.3201/eid1305.070013.
8. World Health Organization. Influenza Virus Infections in Humans. https://www.who.int/publications/m/item/influenza-virus-infections-in-humans, February 2014. Accessed: 2024-06-25.
9. Javad Charostad, Mohammad Rezaei Zadeh Rukerd, Shahab Mahmoudvand, Davood Bashash, Seyed Mohammad Ali Hashemi, Mohsen Nakhaie, and Keivan Zandi. A comprehensive review of highly pathogenic avian influenza (HPAI) H5N1: An imminent threat at doorstep. *Travel Medicine and Infectious Disease*, 55:102638, 2023. ISSN 1477-8939. doi: https://doi.org/10.1016/j.tmaid.2023.102638.
10. ABC News. Bird Flu Mutation Reaches Antarctica, Sparks Australia Outbreak. https://www.abc.net.au/news/2024-05-21/bird-flu-mutation-reaches-antarctica-australia-outbreak/103844240, May 2024. Accessed: 2024-06-25.
11. Features of H5N1 influenza viruses in dairy cows may facilitate infection, transmission in mammals, Jul 2024.
12. Shikha Garg. Outbreak of highly pathogenic avian influenza A (H5N1) viruses in US dairy cattle and detection of two human cases—United States, 2024. *MMWR. Morbidity and Mortality Weekly Report*, 73, 2024.
13. Falguni Debnath, Debjit Chakraborty, Alok Kumar Deb, Malay Kumar Saha, and Shanta Dutta. Increased human-animal interface & emerging zoonotic diseases: An enigma requiring multi-sectoral efforts to address. *Indian Journal of Medical Research*, 153(5-6):577–584, 2021.
14. Centers for Disease Control and Prevention (CDC). CDC Media Release: CDC to Provide Vaccine Candidate Viruses to Better Protect Against Avian Influenza. https://www.cdc.gov/media/releases/2024/p0401-avian-flu.html. Accessed: 2024-07-13.
15. Smriti Mallapaty. Bird flu could become a human pandemic. How are countries preparing? *Nature*, July 2024. doi: 10.1038/d41586-024-02237-4.
16. Yi Shi, Ying Wu, Wei Zhang, Jianxun Qi, and George F. Gao. Enabling the 'host jump': structural determinants of receptor-binding specificity in influenza A viruses. *Nature Reviews Microbiology*, 12(12):822–831, Dec 2014. ISSN 1740-1534. doi: 10.1038/nrmicro3362.
17. Nongluk Sriwilaijaroen and Yasuo Suzuki. Molecular basis of the structure and function of H1 hemagglutinin of influenza virus. *Proceedings of the Japan Academy, Series B*, 88(6): 226–249, 2012. doi: 10.2183/pjab.88.226.
18. Giovanni Cattoli, Adelaide Milani, Nigel Temperton, Bianca Zecchin, Alessandra Buratin, Eleonora Molesti, Mona Meherez Aly, Abdel Arafa, and Ilaria Capua. Antigenic drift in H5N1 avian influenza virus in poultry is driven by mutations in major antigenic sites of the hemagglutinin molecule analogous to those for human influenza virus. *Journal of virology*, 85(17):8718–8724, 2011.
19. Christian Grund, El-Sayed M Abdelwhab, Abdel-Satar Arafa, Mario Ziller, Mohamed K Hassan, Mona M Aly, Hafez M Hafez, Timm C Harder, and Martin Beer. Highly pathogenic avian influenza virus H5N1 from Egypt escapes vaccine-induced immunity but confers clinical protection against a heterologous clade 2.2. 1 Egyptian isolate. *Vaccine*, 29(33):5567–5573, 2011.
20. Bernadeta Dadonaite, Jenny J Ahn, Jordan T Ort, Jin Yu, Colleen Furey, Annie Dosey, William W Hannon, Amy Vincent Baker, Richard J Webby, Neil P King, and et al. Deep mutational scanning of h5 hemagglutinin to inform influenza virus surveillance. *BioRxiv*, May 2024. doi: 10.1101/2024.05.23.595634.
21. Teng Zuo, Jianfeng Sun, Guiqin Wang, Liwei Jiang, Yanan Zuo, Danyang Li, Xuanling Shi, Xi Liu, Shilong Fan, Huanhuan Ren, Hongxing Hu, Lina Sun, Boping Zhou, Mifang Liang, Paul Zhou, Xinquan Wang, and Linqi Zhang. Comprehensive analysis of antibody recognition in convalescent humans from highly pathogenic avian influenza h5n1 infection. *Nature Communications*, 6(1):8855, Dec 2015. ISSN 2041-1723. doi: 10.1038/ncomms9855.
22. Tingting Li, Junyu Chen, Qingbing Zheng, Wenhui Xue, Limin Zhang, Rui Rong, Sibo Zhang, Qian Wang, Minqing Hong, Yuyun Zhang, Lingyan Cui, Maozhou He, Zhen Lu, Zhenyong Zhang, Xin Chi, Jinjin Li, Yang Huang, Hong Wang, Jixian Tang, Dong Ying, Lizhi Zhou, Yingbin Wang, Hai Yu, Jun Zhang, Ying Gu, Yixin Chen, Shaowei Li, and Ningshao Xia. Identification of a cross-neutralizing antibody that targets the receptor binding site of h1n1 and h5n1 influenza viruses. *Nature Communications*, 13(1):5182, Sep 2022. ISSN 2041-1723. doi: 10.1038/s41467-022-32926-5.
23. Qingshan Lin, Tingting Li, Yixin Chen, Siu-Ying Lau, Minxi Wei, Yuyun Zhang, Zhenyong Zhang, Qiaobin Yao, Jinjin Li, Zhihai Li, Daning Wang, Qingbing Zheng, Hai Yu, Ying Gu, Jun Zhang, Honglin Chen, Shaowei Li, and Ningshao Xia. Structural basis for the broad, antibody-mediated neutralization of h5n1 influenza virus. *Journal of Virology*, 92 (17):10.1128/jvi.00547–18, 2018. doi: 10.1128/jvi.00547-18.
24. Pengfei Wang, Yanan Zuo, Jianfeng Sun, Teng Zuo, Senyan Zhang, Shichun Guo, Xuanling Shi, Mifang Liang, Paul Zhou, Linqi Zhang, and Xinquan Wang. Structural and functional definition of a vulnerable site on the hemagglutinin of highly pathogenic avian influenza a virus h5n1. *Journal of Biological Chemistry*, 294(12):4290–4303, Mar 2019. ISSN 0021-9258. doi: 10.1074/jbc.RA118.007008.
25. Xiaoli Xiong, Davide Corti, Junfeng Liu, Debora Pinna, Mathilde Foglierini, Lesley J. Calder, Stephen R. Martin, Yi Pu Lin, Philip A. Walker, Patrick J. Collins, Isabella Monne, Amorsolo L. Suguitan, Celia Santos, Nigel J. Temperton, Kanta Subbarao, Antonio Lanzavecchia, Steven J. Gamblin, and John J. Skehel. Structures of complexes formed by h5 influenza hemagglutinin with a potent broadly neutralizing human monoclonal antibody. *Proceedings of the National Academy of Sciences*, 112(30):9430–9435, 2015. doi: 10.1073/pnas.1510816112.
26. Yanan Zuo, Pengfei Wang, Jianfeng Sun, Shichun Guo, Guiqin Wang, Teng Zuo, Shilong Fan, Paul Zhou, Mifang Liang, Xuanling Shi, Xinquan Wang, and Linqi Zhang. Complementary recognition of the receptor-binding site of highly pathogenic h5n1 influenza viruses by two human neutralizing antibodies. *Journal of Biological Chemistry*, 293(42):16503–16517, Oct 2018. ISSN 0021-9258. doi: 10.1074/jbc.RA118.004604.
27. Xueyong Zhu, Yong-Hui Guo, Tao Jiang, Ya-Di Wang, Kwok-Hung Chan, Xiao-Feng Li, Wenli Yu, Ryan McBride, James C. Paulson, Kwok-Yung Yuen, Cheng-Feng Qin, Xiao-Yan Che, and Ian A. Wilson. A unique and conserved neutralization epitope in h5n1 influenza viruses identified by an antibody against the a/goose/guangdong/1/96 hemagglutinin. *Journal of Virology*, 87(23):12619–12635, 2013. doi: 10.1128/jvi.01577-13.
28. Katie L. Winarski, Natalie J. Thornburg, Yingchun Yu, Gopal Sapparapu, James. E. Crowe, and Benjamin W. Spiller. Vaccine-elicited antibody that neutralizes h5n1 influenza and variants binds the receptor site and polymorphic sites. *Proceedings of the National Academy of Sciences*, 112(30):9346–9351, 2015. doi: 10.1073/pnas.1502762112.
29. Huaiyu Tian, Yujun Cui, Lu Dong, Sen Zhou, Xiaowen Li, Shanqian Huang, Ruifu Yang, and Bing Xu. Spatial, temporal and genetic dynamics of highly pathogenic avian influenza A (H5N1) virus in China. *BMC Infectious Diseases*, 15(1):54, Feb 2015. ISSN 1471-2334. doi: 10.1186/s12879-015-0770-x.
30. John J Skehel and Don C Wiley. Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. *Annual review of biochemistry*, 69(1):531–569, 2000.
31. Ya Ha, David J Stevens, John J Skehel, and Don C Wiley. X-ray structures of H5 avian and H9 swine influenza virus hemagglutinins bound to avian and human receptor analogs. *Proceedings of the National Academy of Sciences*, 98(20):11181–11186, 2001.
32. Netanel Tzarum, Robert P De Vries, Xueyong Zhu, Wenli Yu, Ryan McBride, James C Paulson, and Ian A Wilson. Structure and receptor binding of the hemagglutinin from a human H6N1 influenza virus. *Cell host & microbe*, 17(3):369–376, 2015.
33. Ivan Kosik, Davide Angeletti, James S Gibbs, Matthew Angel, Kazuyo Takeda, Martina Kosikova, Vinod Nair, Heather D Hickman, Hang Xie, Christopher B Brooke, et al. Neuraminidase inhibition contributes to influenza a virus neutralization by anti-hemagglutinin stem antibodies. *Journal of Experimental Medicine*, 216(2):304–316, 2019.
34. Colleen Furey, Gabrielle Scher, Naiqing Ye, Lisa Kercher, Jennifer DeBeauchamp, Jeri Carol Crumpton, Trushar Jeevan, Christopher Patton, John Franks, Adam Rubrum, Mohamad-Gabriel Alameh, Steven H. Y. Fan, Anthony T. Phan, Christopher A. Hunter, Richard J. Webby, Drew Weissman, and Scott E. Hensley. Development of a nucleoside-modified mRNA vaccine against clade 2.3.4.4b H5 highly pathogenic avian influenza virus. *Nature Communications*, 15(1):4350, May 2024. ISSN 2041-1723. doi: 10.1038/s41467-024-48555-z.
35. Phillip J. Tomezsko, Colby T. Ford, Avery E. Meyer, Adam M. Michaleas, and Rafael Jaimes. Human cytokine and coronavirus nucleocapsid protein interactivity using large-scale virtual screens. *Frontiers in Bioinformatics*, 4, 2024. ISSN 2673-7647. doi: 10.3389/fbinf.2024.1397968.
36. Colby T. Ford. PD-1 Targeted Antibody Discovery Using AI Protein Diffusion. *bioRxiv*, 2024. doi: 10.1101/2024.01.18.576323.
37. Shirish Yasa, Sayal Guirales-Medrano, Denis Jacob Machado, Colby T. Ford, and Daniel A. Janies. Predicting Antibody and ACE2 Affinity for SARS-CoV-2 BA.2.86 and JN.1 with *In Silico* Protein Modeling and Docking. *Frontiers in Virology*, 4, 2024. ISSN 2673-818X. doi: 10.3389/fviro.2024.1419276.
38. Colby T. Ford, Shirish Yasa, Denis Jacob Machado, Richard Allen White III, and Daniel A. Janies. Predicting changes in neutralizing antibody activity for SARS-CoV-2 XBB.1.5 using *in silico* protein modeling. *Frontiers in Virology*, 3, 2023. ISSN 2673-818X. doi: 10.3389/fviro.2023.1172027.
39. Colby T. Ford, Denis Jacob Machado, and Daniel A Janies. Predictions of the SARS-CoV-2 Omicron Variant (B.1.1.529) Spike Protein Receptor-Binding Domain Structure and Neutralizing Antibody Interactions. *Frontiers in Virology*, 2022. doi: 10.3389/fviro.2022.830202.
40. Cheikh Cambel Dieng, Colby T. Ford, Anita Lerch, Dickson Doniou, Kovidh Vegesna, Daniel Janies, Liwang Cui, Linda Amoah, Yaw Afrane, and Eugenia Lo. Genetic variations of *Plasmodium falciparum* circumsporozoite protein and the impact on interactions with human immunoproteins and malaria vaccine efficacy. *Infection, Genetics and Evolution*, 2023. ISSN 1567-1348. doi: https://doi.org/10.1016/j.meegid.2023.105418.
41. Leonardo C. Caserta, Elisha A. Frye, Salman L. Butt, Melissa Laverack, Mohammed Nooruzzaman, Lina M. Covaleda, Alexis C. Thompson, Melanie Prarat Koscielny, Brittany Cronk, Ashley Johnson, Katie Kleinhenz, Erin E. Edwards, Gabriel Gomez, Gavin Hitchener, Mathias Martins, Darrell R. Kapczynski, David L. Suarez, Ellen Ruth Alexander Morris, Terry Hensley, John S. Beeby, Manigandan Lejeune, Amy K. Swinford, François Elvinger, Kiril M. Dimitrov, and Diego G. Diel. From birds to mammals: spillover of highly pathogenic avian influenza H5N1 virus to dairy cattle led to efficient intra- and interspecies transmission. *bioRxiv*, 2024. doi: 10.1101/2024.05.22.595317.
42. Bernadeta Dadonaite, Jenny J. Ahn, Jordan T. Ort, Jin Yu, Colleen Furey, Annie Dosey, William W. Hannon, Amy L. Vincent Baker, Richard Webby, Neil P. King, Yan Liu, Scott E. Hensley, Thomas P. Peacock, Louise H. Moncla, and Jesse D. Bloom. Deep mutational scanning of H5 hemagglutinin to inform influenza virus surveillance. *bioRxiv*, 2024. doi: 10.1101/2024.05.23.595634.
43. Yuelong Shu and John McCauley. GISAID: Global initiative on sharing all influenza data – from vision to reality. *Eurosurveillance*, 22(13):30494, 2017. doi: https://doi.org/10.2807/

1560-7917.ES.2017.22.13.30494.
44. Stefan Elbe and Gemma Buckland-Merrett. Data, disease and diplomacy: Gisaid's innovative contribution to global health. *Global Challenges*, 1(1):33–46, 2017. doi: https://doi.org/10.1002/gch2.1018.
45. Weizhong Li and Adam Godzik. CD-HIT: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22(13):1658–1659, 05 2006. ISSN 1367-4803. doi: 10.1093/bioinformatics/btl158.
46. Limin Fu, Beifang Niu, Zhengwei Zhu, Sitao Wu, and Weizhong Li. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28(23):3150–3152, 10 2012. ISSN 1367-4803. doi: 10.1093/bioinformatics/bts565.
47. Robert C. Edgar. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5):1792–1797, 03 2004. ISSN 0305-1048. doi: 10.1093/nar/gkh340.
48. Milot Mirdita, Konstantin Schütze, Yoshitaka Moriwaki, Lim Heo, Sergey Ovchinnikov, and Martin Steinegger. ColabFold: making protein folding accessible to all. *Nature Methods*, 19 (6):679–682, Jun 2022. ISSN 1548-7105. doi: 10.1038/s41592-022-01488-1.
49. John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, Alex Bridgland, Clemens Meyer, Simon A. A. Kohl, Andrew J. Ballard, Andrew Cowie, Bernardino Romera-Paredes, Stanislav Nikolov, Rishub Jain, Jonas Adler, Trevor Back, Stig Petersen, David Reiman, Ellen Clancy, Michal Zielinski, Martin Steinegger, Michalina Pacholska, Tamas Berghammer, Sebastian Bodenstein, David Silver, Oriol Vinyals, Andrew W. Senior, Koray Kavukcuoglu, Pushmeet Kohli, and Demis Hassabis. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, Aug 2021. ISSN 1476-4687. doi: 10.1038/s41586-021-03819-2.
50. Martin Steinegger and Johannes Söding. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nature Biotechnology*, 35(11):1026–1028, Nov 2017. ISSN 1546-1696. doi: 10.1038/nbt.3988.
51. Peter Eastman, Jason Swails, John D. Chodera, Robert T. McGibbon, Yutong Zhao, Kyle A. Beauchamp, Lee-Ping Wang, Andrew C. Simmonett, Matthew P. Harrigan, Chaya D. Stern, Rafal P. Wiewiora, Bernard R. Brooks, and Vijay S. Pande. Openmm 7: Rapid development of high performance algorithms for molecular dynamics. *PLOS Computational Biology*, 13 (7):1–17, 07 2017. doi: 10.1371/journal.pcbi.1005659.
52. Matthew I J Raybould, Claire Marks, Alan P Lewis, Jiye Shi, Alexander Bujotzek, Bruck Taddese, and Charlotte M Deane. Thera-SAbDab: the Therapeutic Structural Antibody Database. *Nucleic Acids Research*, 48(D1):D383–D388, 09 2019. ISSN 0305-1048. doi: 10.1093/nar/gkz827.
53. Helen M. Berman, John Westbrook, Zukang Feng, Gary Gilliland, T. N. Bhat, Helge Weissig, Ilya N. Shindyalov, and Philip E. Bourne. The Protein Data Bank. *Nucleic Acids Research*, 28(1):235–242, 01 2000. ISSN 0305-1048. doi: 10.1093/nar/28.1.235.
54. James Dunbar and Charlotte M. Deane. ANARCI: antigen receptor numbering and receptor classification. *Bioinformatics*, 32(2):298–300, 09 2015. ISSN 1367-4803. doi: 10.1093/bioinformatics/btv552.
55. Bonvin Lab. HADDOCK3 Antibody-Antigen Tutorial. https://www.bonvinlab.org/education/HADDOCK3/HADDOCK3-antibody-antigen/, 2024. Accessed: 2024-06-21.
56. João P. G. L. M. Rodrigues, Mikaël Trellet, Christophe Schmitz, Panagiotis Kastritis, Ezgi Karaca, Adrien S. J. Melquiond, and Alexandre M. J. J. Bonvin. Clustering biomolecular complexes by residue contacts similarity. *Proteins: Structure, Function, and Bioinformatics*, 80(7):1810–1817, 2012. doi: https://doi.org/10.1002/prot.24078.
57. Kazutaka Katoh and Daron M. Standley. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, 30(4):772–780, 01 2013. ISSN 0737-4038. doi: 10.1093/molbev/mst010.
58. Pablo A. Goloboff and Martín E. Morales. TNT version 1.6, with a graphical interface for MacOS and Linux, including new routines in parallel. *Cladistics*, 39(2):144–153, 2023. doi: https://doi.org/10.1111/cla.12524.
59. Adriano de Bernardi Schneider, Colby T Ford, Reilly Hostager, John Williams, Michael Cioce, Ümit V Çatalyürek, Joel O Wertheim, and Daniel Janies. StrainHub: A phylogenetic tool to construct pathogen transmission networks. *Bioinformatics*, 08 2019. ISSN 1367-4803. doi: 10.1093/bioinformatics/btz646.
60. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021.
61. Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016. ISBN 978-3-319-24277-4.
62. Alboukadel Kassambara. *ggpubr: 'ggplot2' Based Publication Ready Plots*, 2023. R package version 0.6.0.
63. Schrödinger, LLC. The PyMOL molecular graphics system. version 2.4.1, November 2015.
64. Sebastian Raschka. BioPandas: Working with molecular structures in pandas DataFrames. *The Journal of Open Source Software*, 2(14), jun 2017. doi: 10.21105/joss.00279.

Supplementary Information

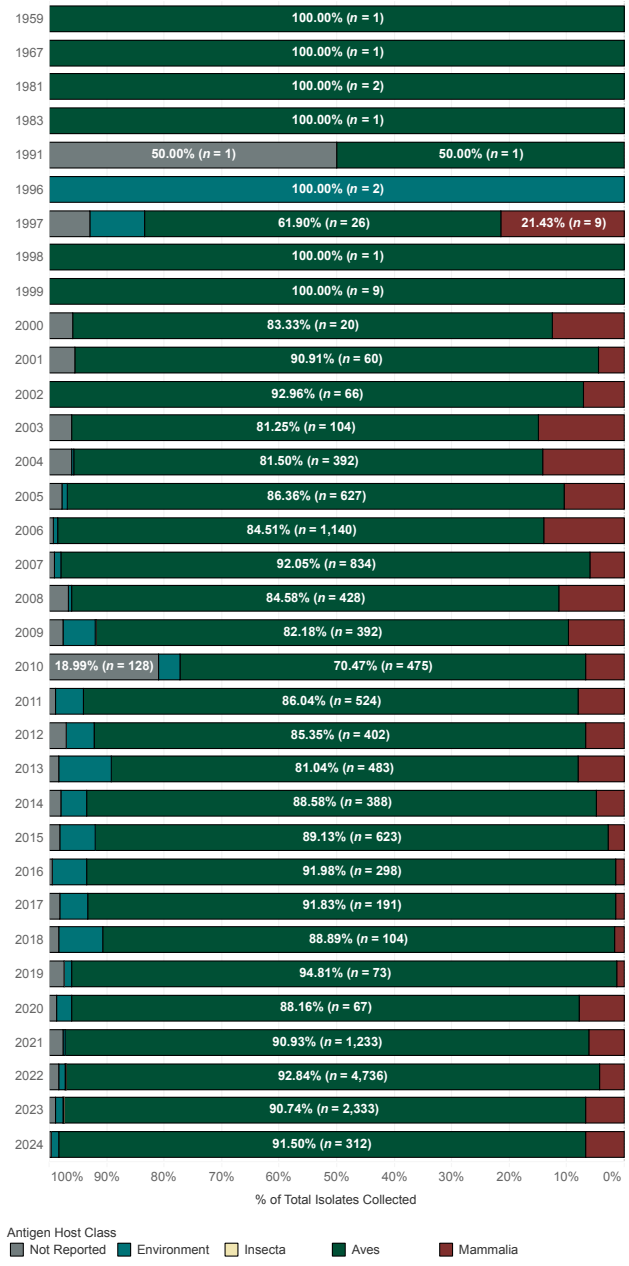


Fig. 1. Bar chart of the proportion of isolates collected from hosts of various taxonomic classes by year.

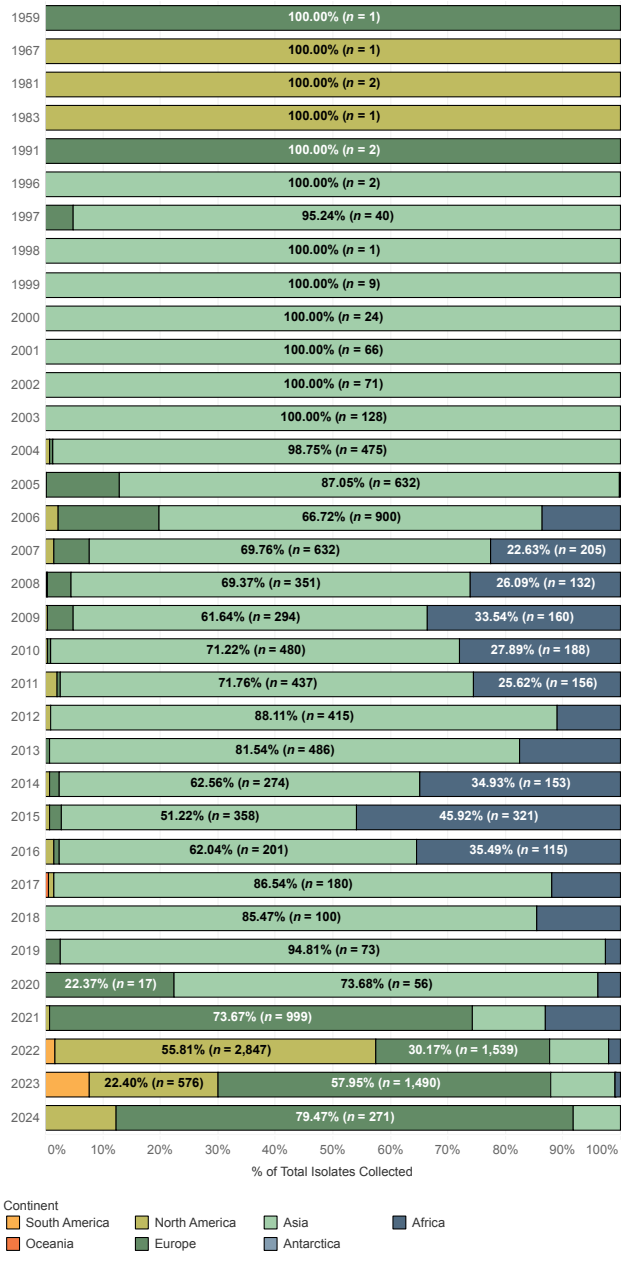


Fig. 2. Bar chart of the proportion of isolates collected from each continent by year.

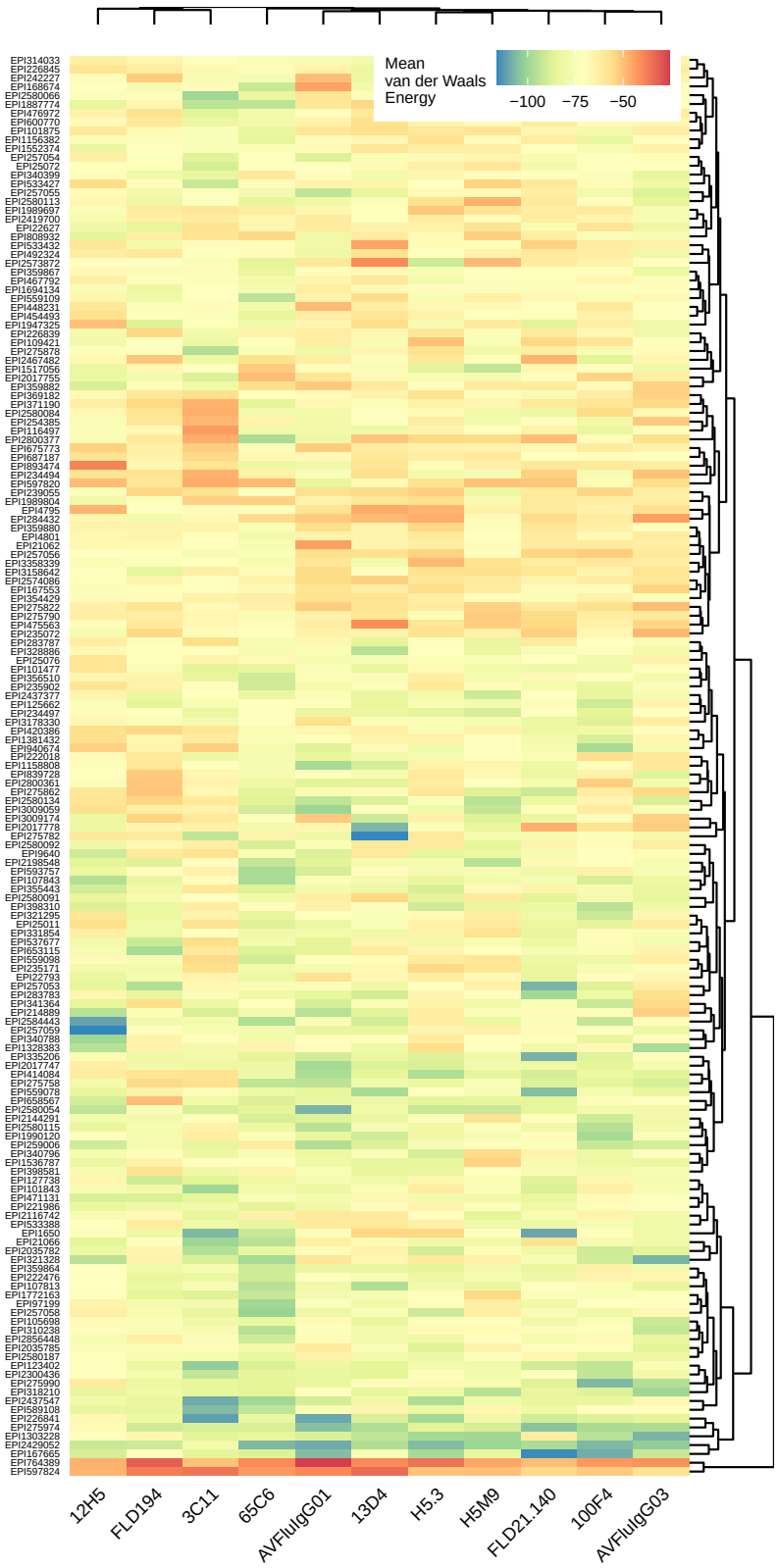


Fig. 3. Heatmap of van der Waals energies for each antibody-antigen experiment, shown with cluster dendrograms.

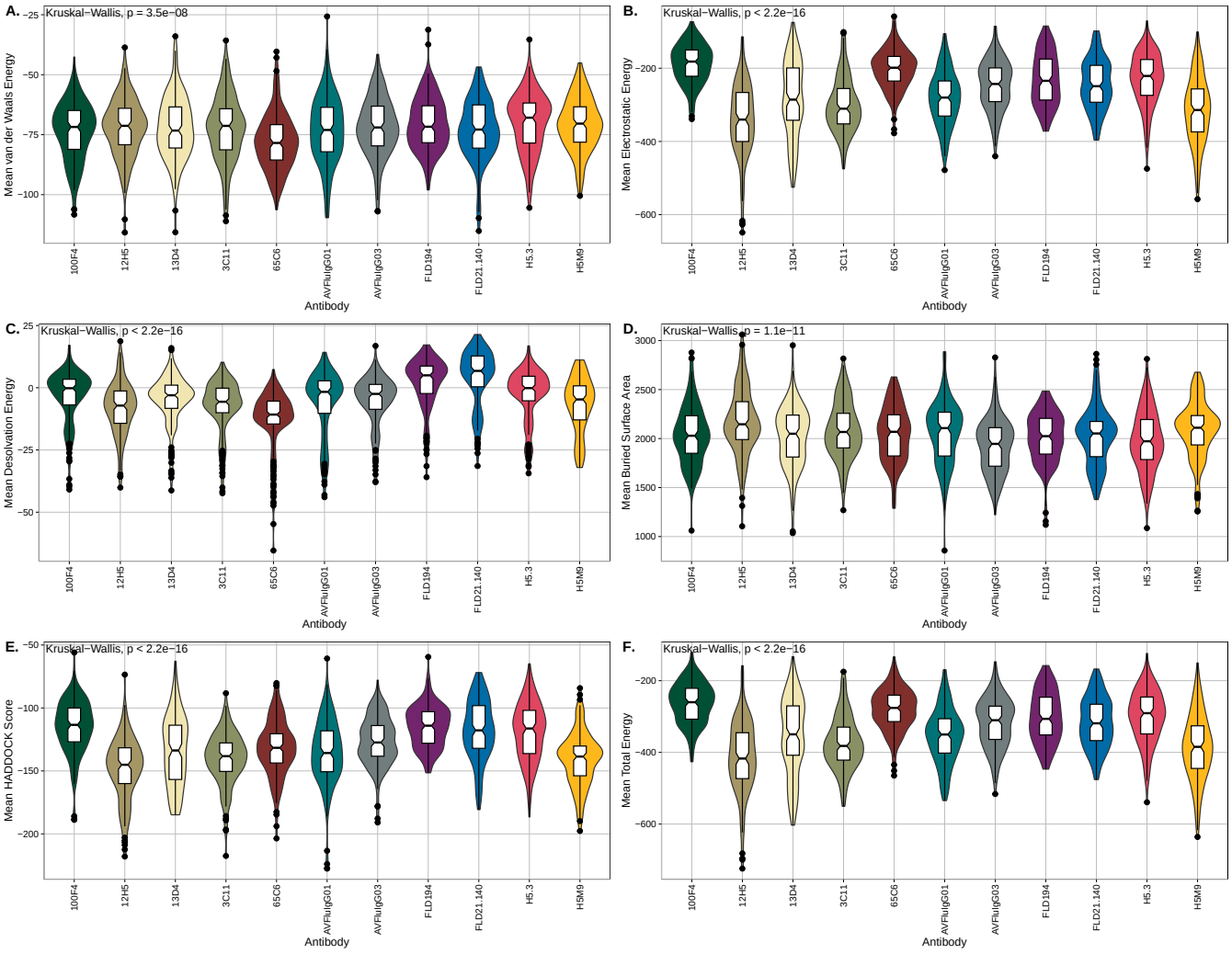


Fig. 4. The distribution of various docking metrics broken out by antibody. Statistics shown are Kruskal-Wallis test p-values.

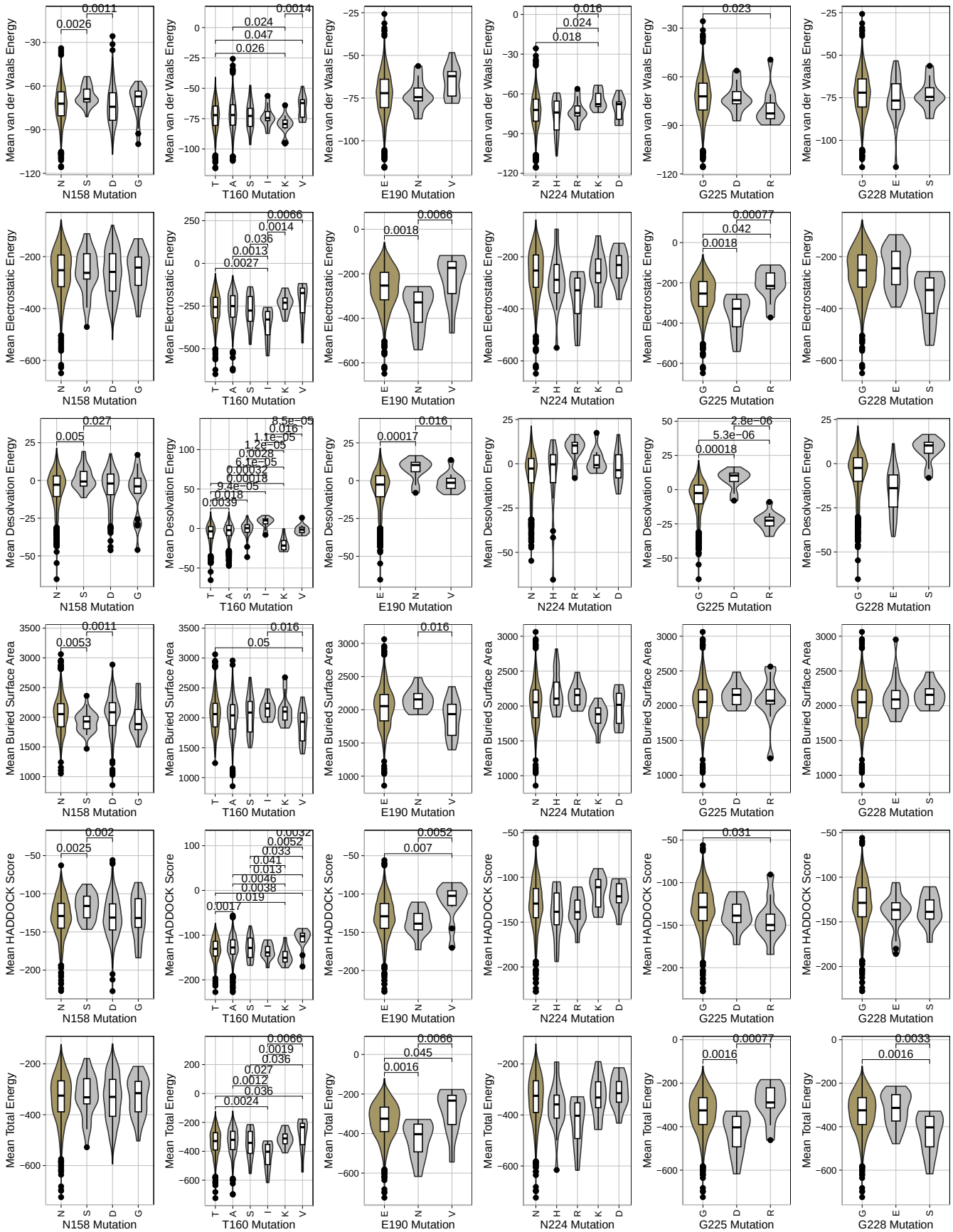


Fig. 5. The distribution of various docking metrics broken out by antigen mutations. The first amino acid shown on the left of each plot in gold represents the reference residue at that position as described in Shi et al. (2014). Statistical comparisons shown are significant Wilcoxon Rank Sum test p-values at the α<0.05 level.