

UC San Diego

UC San Diego Previously Published Works

Title

Visualizing and analyzing 3D biomolecular structures using Mol* at RCSB.org: Influenza A H5N1 virus proteome case study.

Permalink

<https://escholarship.org/uc/item/05h74626>

Journal

Protein Science, 34(4)

Authors

Bittrich, Sebastian

Rose, Alexander

Sehnal, David

et al.

Publication Date

2025-04-01

DOI

10.1002/pro.70093

Peer reviewed

Visualizing and analyzing 3D biomolecular structures using Mol* at RCSB.org: Influenza A H5N1 virus proteome case study

Sebastian Bittrich¹  | Alexander S. Rose²  | David Sehnal³  |
 Jose M. Duarte¹  | Yana Rose¹  | Joan Segura¹  | Dennis W. Piehl⁴  |
 Brinda Vallat^{4,5}  | Chenghua Shao⁴  | Charmi Bhikadiya¹  |
 Jesse Liang¹ | Mark Ma¹ | David S. Goodsell^{4,5,6}  |
 Stephen K. Burley^{1,4,5,7,8}  | Shuchismita Dutta^{4,5} 

¹Research Collaboratory for Structural Bioinformatics Protein Data Bank, San Diego Supercomputer Center, University of California San Diego, La Jolla, California, USA

²Mol* Consortium, San Diego, California, USA

³National Centre for Biomolecular Research, Faculty of Science, Masaryk University, Brno, Czech Republic

⁴Research Collaboratory for Structural Bioinformatics Protein Data Bank, Institute for Quantitative Biomedicine, Rutgers, The State University of New Jersey, Piscataway, New Jersey, USA

⁵Rutgers Cancer Institute, Rutgers, The State University of New Jersey, New Brunswick, New Jersey, USA

⁶Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, California, USA

⁷Rutgers Artificial Intelligence and Data Science (RAD) Collaboratory, Rutgers, The State University of New Jersey, Piscataway, New Jersey, USA

⁸Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, Piscataway, New Jersey, USA

Correspondence

Sebastian Bittrich, Research Collaboratory for Structural Bioinformatics Protein Data Bank, San Diego Supercomputer Center, University of California San Diego, La Jolla, CA 92093, USA.

Email: sebastian.bittrich@rcsb.org

Shuchismita Dutta, Research Collaboratory for Structural Bioinformatics Protein Data Bank, Institute for Quantitative Biomedicine, Rutgers, The State University of New Jersey, Piscataway, NJ 08854, USA.

Email: shuchi.dutta@rcsb.org

Funding information

National Science Foundation, Grant/Award Numbers: DBI-2019297, DBI-2129634, DBI-2321666; Czech Science Foundation, Grant/Award Number: 22-30571M; U.S. Department of Energy, Grant/Award Number: DE-SC0019749; Biotechnology and Biological Sciences Research Council, Grant/Award Numbers: BB/V004247/1, BB/W017970/1; National Institutes of Health, Grant/Award Number: R01GM152779

Review Editor: Nir Ben-Tal

Abstract

The easiest and often most useful way to work with experimentally determined or computationally predicted structures of biomolecules is by viewing their three-dimensional (3D) shapes using a molecular visualization tool. Mol* was collaboratively developed by RCSB Protein Data Bank (RCSB PDB, [RCSB.org](https://www.rcsb.org)) and Protein Data Bank in Europe (PDBe, [PDBe.org](https://www.ebi.ac.uk/pdbe)) as an open-source, web-based, 3D visualization software suite for examination and analyses of biostructures. It is capable of displaying atomic coordinates and related experimental data of biomolecular structures together with a variety of annotations, facilitating basic and applied research, training, education, and information dissemination. Across [RCSB.org](https://www.rcsb.org), the RCSB PDB research-focused web portal, Mol* has been implemented to support single-mouse-click atomic-level visualization of biomolecules (e.g., proteins, nucleic acids, carbohydrates) with bound cofactors, small-molecule ligands, ions, water molecules, or other macromolecules. [RCSB.org](https://www.rcsb.org) Mol* can seamlessly display 3D structures from various sources, allowing structure interrogation, superimposition, and comparison. Using influenza A H5N1 virus as a topical case study of an important pathogen, we exemplify how Mol* has been embedded within various [RCSB.org](https://www.rcsb.org) tools—allowing users to view polymer sequence and structure-based annotations integrated from trusted

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *Protein Science* published by Wiley Periodicals LLC on behalf of The Protein Society.

bioinformatics data resources, assess patterns and trends in groups of structures, and view structures of any size and compositional complexity. In addition to being linked to every experimentally determined biostructure and Computed Structure Model made available at [RCSB.org](https://www.rcsb.org), Standalone Mol* is freely available for visualizing any atomic-level or multi-scale biostructure at [rcsb.org/3d-view](https://www.rcsb.org/3d-view).

KEYWORDS

3D biostructure, global health, influenza A H5N1 virus, molecular visualization, open-source, pandemic preparedness, Protein Data Bank, viral pathogen, virus life cycle, web-based

1 | INTRODUCTION

The volume and complexity of three-dimensional (3D) structure data for biological macromolecules (i.e., proteins, nucleic acids, carbohydrates, lipids, and their various complexes) is growing rapidly. At the time of writing, the Protein Data Bank or PDB (wwPDB consortium 2019) archives and makes available >230,000 3D biostructures, experimentally determined using macromolecular crystallography (MX), 3D electron microscopy (3DEM), or nuclear magnetic resonance (NMR) spectroscopy. In addition, hundreds of millions of computed structure models (CSMs) of proteins are freely available from the AlphaFold Protein Structure Database (Varadi et al. 2024) and the ModelArchive (Humphreys et al. 2021). The easiest way to interact with and explore these structures is to view them using molecular visualization tools. Since PDB was established as the first open-access digital data resource in biology (Protein Data Bank 1971), the quality of molecular rendering and the capabilities to visualize large and complex data sets have improved dramatically (Li and Wei 2024). Most modern 3D visualization tools offer options for changing representation and analyzing biomolecular structures at the atomic level. Importantly, newer web-based tools that operate entirely within a user's browser make macromolecular visualization accessible to anyone in the world using a desktop computer, a tablet, or a smartphone connected to the internet.

The Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB) (Berman et al. 2000; Burley et al. 2023; Burley et al. 2025) serves as the United States (US) data center of the Worldwide Protein Data Bank (wwPDB, [wwpdb.org](https://www.wwpdb.org)) partnership, which was established in 2003 to support joint management of the PDB archive as a global public good (Berman et al. 2003). The wwPDB was originally co-founded by RCSB PDB, Protein Data Bank in Europe (PDBe) (Armstrong et al. 2020), and Protein Data Bank Japan (PDBj) (Bekker et al. 2022). It has expanded over the years to encompass two specialist data resources: Electron Microscopy Data Bank (EMDB) (wwPDB Consortium 2023) and Biological

Magnetic Resonance Data Bank (BMRB) (Hoch et al. 2023). Together, RCSB PDB, PDBe, PDBj, EMDb, and BMRB now jointly manage complete deposition (Young et al. 2017), rigorous validation (Baskaran et al. 2024; Feng et al. 2021; Gore et al. 2017; Smart et al. 2018), and expert biocuration (Young et al. 2021) of data contributed to all three wwPDB Core Archives (PDB, EMDb, and BMRB). Recently, PDB China (PDBc) (Xu et al. 2023) joined the wwPDB partnership as an Associate Member.

The RCSB PDB research-focused [RCSB.org](https://www.rcsb.org) web portal provides open-access tools that support query, browsing, visualization, analysis, comparison, and mapping of annotations for >230,000 experimentally determined 3D biostructures plus more than 1 million CSMs of proteins. Herein, we describe the development of the Mol* web-based molecular visualization software suite and its deployment within various [RCSB.org](https://www.rcsb.org) tools. We exemplify the use of these powerful tools to study the proteome of the influenza A H5N1 virus, a globally important pathogen and a major public health concern recently in the news.

1.1 | Mol*: An international open-source software development project

The advent of the Web Graphics Library (WebGL) (icr.khronos.org/webgl/) marked a significant milestone in web application development, enabling robust, high-performance rendering of 3D visualizations within web browsers. This technical advance eliminated the need for dedicated software installations, such as desktop-based viewers like PyMOL (DeLano 2002) or ChimeraX (Meng et al. 2023). Two early adopters, the NGL Viewer (developed by RCSB PDB; Rose and Hildebrand 2015) and the LiteMol Suite (developed by PDBe; Sehnal et al. 2017), leveraged their respective capabilities to pioneer the development of more efficient data exchange formats (Bradley et al. 2017; Sehnal et al. 2020). To avoid duplication of effort, these parallel activities were merged with the launch of the open-source Mol* technology stack (Sehnal et al. 2018; Sehnal et al. 2021) by RCSB PDB, PDBe, and CEITEC

(Central European Institute of Technology, Brno, Czech Republic).

Mol* integrated elements of both NGL and LiteMol, establishing a new ecosystem of related data conversion and delivery tools to power 3D visualization of biomolecules (Sehnal et al. 2021). Initially, Mol* was implemented at [RCSB.org](https://www.rcsb.org) and [PDB.org](https://www.pdb.org), with subsequent adoption by other trusted biodata resources (e.g., UniProt (UniProt Consortium 2023) and AlphaFold Protein Structure Database (Varadi et al. 2024)). Mol* is an excellent choice for an in-browser biomolecule viewer, offering high-performance visualization and broad data format support. Other established tools, such as JSmol (Hanson et al. 2013), provide relevant features like extensive scripting capabilities for automating tasks and customizing visualizations—functionality not currently available in Mol*. Additionally, while Mol* excels in visualizing large and complex biomolecular systems, some users may find that it lacks advanced molecular dynamics simulation tools (e.g., VMD; Humphrey et al. 1996).

1.2 | Mol* capabilities

Mol* supports sophisticated visualization capabilities, rivaling those of traditional desktop applications, while providing a seamless, installation-free user experience. It can display various representations of atomic coordinates of biomolecular structures, MX electron density maps, 3DEM electric Coulomb potential maps (or density maps), molecular dynamics trajectories, and geometric shapes (e.g., polyhedra) in various formats (Sehnal et al. 2021). In addition, Mol* enables users to analyze and compare structures (e.g., measure interatomic distances, calculate angles, and superimpose structures). Finally, it can save scenes and export images, videos, and sessions for use in publications, presentations, and scientific collaborations. See the Mol* project page (molstar.org/viewer-docs) for detailed instructions. The versatility of Mol* makes it a powerful tool for visualizing biomolecules, enabling dozens of third-party applications to be built atop the viewer. It has exceptional performance across length scales—supporting visualization of biomolecules ranging from small chemicals to mesoscale atomic models of entire virus capsids or even whole cells (see the Mesoscale Explorer Tool, available at molstar.org/me; Rose et al. 2024).

1.3 | Integration into the [RCSB.org](https://www.rcsb.org) web portal

In 2020, Mol* became the default visualization tool and interaction interface at [RCSB.org](https://www.rcsb.org), supporting access and exploration of 3D biostructure data (hereafter

[RCSB.org](https://www.rcsb.org) Mol*). It has been linked to [RCSB.org](https://www.rcsb.org) Structure Summary Pages (SSPs) across the entire archive, providing single-mouse-click access for displaying and analyzing atomic coordinates, macromolecular assemblies, experimental density maps, etc. More recently, Mol* has been deployed within various [RCSB.org](https://www.rcsb.org) data display and analysis tools. It is used in the Sequence Annotations Tool (Segura et al. 2020; Segura et al. 2022) to display bidirectional mapping of sequence-based annotations from trusted bioinformatics data resources on 3D structures. It also powers our Pairwise Structure Alignment Tool (Bittrich et al. 2024b). Mol* can also be used to visualize small-molecule ligands and biologically interesting molecules (Biologically Interesting molecules Reference Dictionary or BIRD molecules) (Dutta et al. 2014; Sen et al. 2014) within their respective summary pages. Relevant implementations are summarized in Table 1, concerning examples presented below.

1.4 | Introducing the Influenza A H5N1 virus case study

Influenza infections (the “flu”) are caused by influenza viruses. Reports of flu epidemics affecting humans date as far back as the 5th century BCE (Potter 2001). One of the worst pandemics recorded in the 20th century was the 1918 Spanish flu, which killed an estimated 50 million people worldwide (Taubenberger and Morens 2009). All influenza viruses are enveloped, negative-sense, single-stranded RNA viruses with segmented genomes (Taubenberger and Morens 2010). Among the four known types of influenza virus (A, B, C, D), subtypes of A have caused pandemics in the modern era (Centers for Disease Control and Prevention and National Center for Immunization and Respiratory Diseases (NCIRD) 2023). Two viral envelope proteins, hemagglutinin (HA) and neuraminidase (NA), encoded by distinct genome segments, vary among influenza A virus subtypes. Combinations of 18 different HAs (HA1 through HA18) and 11 different NAs (NA1 through NA11) can yield many different viruses identified with H#N# designations (Centers for Disease Control and Prevention and National Center for Immunization and Respiratory Diseases (NCIRD) 2023). New combinations of HA and NA can arise when host cells become co-infected by two or more viruses with distinct HAs and/or NAs.

Some influenza A viruses can jump the species barrier from one host to another, for example, from birds or pigs to humans (World Health Organization 2023). While such zoonotic infections rarely spread from person to person, aquatic birds serve as a vast natural reservoir for influenza A viruses. Exposure to infected birds (dead or alive) can constitute a substantial public health risk. From 1997, when the first case of a human infected

TABLE 1 Exploring structure and function of influenza A H5N1 virus proteins using [RCSB.org](https://www.rcsb.org) Mol* and its deployments within various [RCSB.org](https://www.rcsb.org) tools.

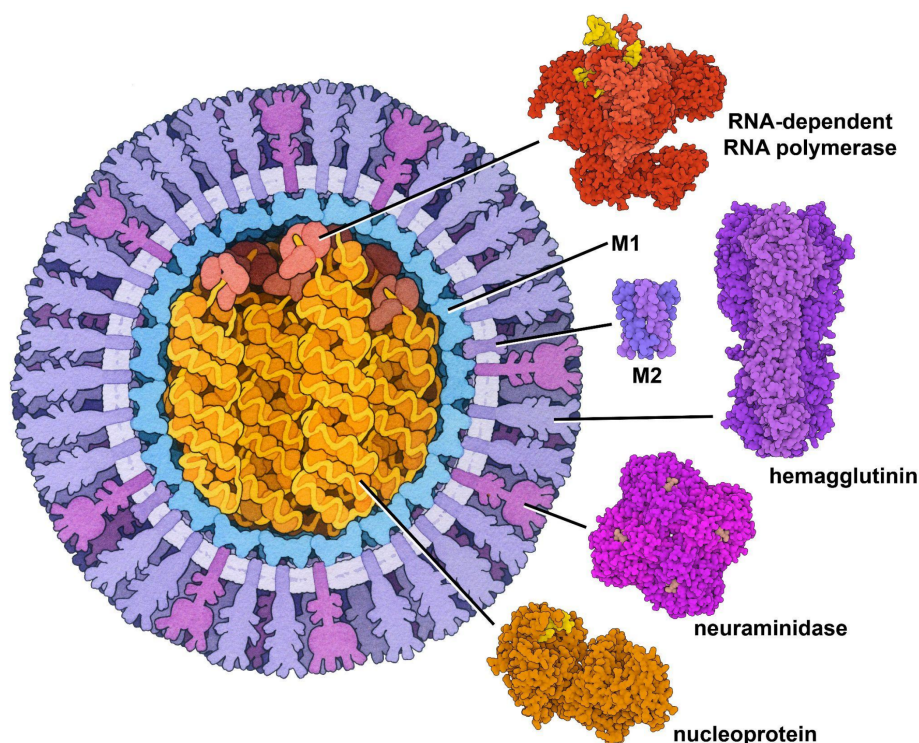
H5N1 biology	Protein(s) and/or complex(es)	Focus	Mol* capability	Number
Structures of H5N1 proteins	HA: PDB ID 2fk0 (Stevens et al. 2006)	Overall shape and assembly	Basic Mol* features (Sehna et al. 2021)	2A, 2B, 2D
		Protein glycosylation patterns	Glycan visualization (Neelamegham et al. 2019; Shao et al. 2021)	2C
		Side-by-side, linked display of functional annotations of 1D sequence and 3D structure	Sequence annotations in 1D/3D (Segura et al. 2020; Segura et al. 2022)	3B
		Goodness-of-fit of atomic coordinates to experimental density maps	Electron density map visualization	3C
		Structure quality	wwPDB validation reports	3D
	NA, bound to an inhibitor (zanamivir): PDB ID 3ckz (Collins et al. 2008)		Real-space correlation coefficient (Shao et al. 2022)	3E
		Atomic-level structures of small-molecule ligands or BIRD molecules	Ligand and BIRD Summary Pages (Dutta et al. 2014; Sen et al. 2014)	4B, 4E
		Interactions in the vicinity of an amino acid residue or a bound small molecule	Non-covalent interactions	4C
		Local and global symmetry and pseudosymmetry in biomolecular assemblies	Symmetry visualization	4D
		Mapping of phospholipid bilayer location onto membrane protein structures	Predict membrane position (Bittrich et al. 2022)	5B
	M2 bound to an inhibitor: PDB ID 6nv1 (Thomaston et al. 2020)		Custom representations	5C–5E
		Selectively show/hide polypeptide chains, amino acid residues, ligands, and their interactions to explain and illustrate function		
		Compare distinct structures of the same protein to learn about conformational changes (e.g., upon RNA binding)	Standalone Mol* and built-in alignment tools	6A–6D
Key steps in virus life cycle	Transmissible and non-transmissible HAs bound to human and avian sialic acid receptors: PDB IDs 4bh0, 4bh1, 4bh3, 4bh4 (Xiong et al. 2013)	Compare structures of evolutionarily related proteins interacting with human versus avian sialic acid receptors to understand zoonotic H5N1 infection at the atomic level	Pairwise structure alignment (Bittrich et al. 2024b). Close up and examination of interactions in the neighborhood of specific amino acids in these structures	7A–7F
	H5N1 RNA-dependent RNA polymerase (RdRp) bound to human ANP32B: PDB ID 8r1j (Staller et al. 2024)	Overview of structure and visualization of location of key components of the complex	Transparent surface	8A
		Launch views of all structures including any part of ANP32B annotated with the same UniProt ID to compare the PDB structure of ANP32B with that of its AlphaFold2 CSM	Sequence Alignment Viewer (Segura et al. 2020; Segura et al. 2022)	8C
	AlphaFold2 CSM of human ANP32B: RCSB.org ID AF_AFQ92688F1 (Jumper et al. 2021)	View CSM color coded by prediction confidence (dark blue denotes highest confidence)	pLDDT coloring	8D
		Side-by-side display of structural and functional annotations on both 1D sequence and 3D structure	Sequence Annotations Viewer	8E

TABLE 1 (Continued)

H5N1 biology	Protein(s) and/or complex(es)	Focus	Mol* capability	Number
Structure exploration facilitating antiviral drug discovery	Comparison of NA co-crystal structures with zanamivir and oseltamivir carboxylate: PDB IDs 3ckz, 3cl0 (Collins et al. 2008)	Identify common interactions for drug binding	Sequence Alignment View (Segura et al. 2020; Segura et al. 2022)	9A, 9B
	Select NA amino acid residues interacting with zanamivir for Structure Motif Search	Identify 3D biostructures (experimentally determined or computationally predicted) at RCSB.org with the defined structure motif (for possible zanamivir binding sites in human sialidases, etc.)	Structure Motif Search (Bittrich et al. 2020)	9C–9F

Note: Core Mol* features have been described previously (Sehna et al. 2021).

FIGURE 1 David Goodsell's rendition of the H5N1 virion in cross-section (left) and experimentally determined structures of select H5N1 proteins (right) drawn using Mol*: RdRp (PDB ID 8h69; Li et al. 2023), M2 (PDB ID 6nv1; Thomaston et al. 2020), HA (PDB ID 2fk0; Stevens et al. 2006), NA (PDB ID 2hu4; Russell et al. 2006), and NP (PDB ID 7dpx; Tang et al. 2021). The shape of M1 in the artistic rendition is based on structures from other influenza A viruses (e.g., PDB ID 7jm3; Selzer et al. 2020).



with an influenza A H5N1 virus was detected, to April 2024, 912 sporadic cases of human influenza A H5N1 virus infections have been reported in 24 countries with disturbingly high case-fatality rates exceeding 50% (Centers for Disease Control and Prevention 2024). More recently, so-called “bird flu” viruses have been shown to infect dairy cows (Abbasi 2024). Hence, dairy workers are now thought to be at higher risk of infection than the general population (Garg et al. 2024), raising concerns about the possibility of a serious epidemic or even a global pandemic that could kill millions.

To exemplify the utility of [RCSB.org](https://www.rcsb.org) Mol* and [RCSB.org](https://www.rcsb.org) tools reliant on Mol*, we present influenza A H5N1 virus (hereafter H5N1) as a topical case study

(Figure 1). We use data from the PDB archive and trusted bioinformatics resources to explore the H5N1 proteome in 3D at the atomic level, to examine key steps in the influenza A virus life cycle, and discuss the potential impact of structural knowledge of the H5N1 proteome for epidemic/pandemic preparedness (i.e., discovering/developing drug therapies or designing vaccines to counter H5N1). Distinct deployments of Mol* used in these explorations are summarized in Table 1, with reference to examples presented herein. Instructions for recreating Figure 1 using Mol* are provided in Figure S1, Supporting Information, with similar details for subsequent figures in their corresponding figures.

2 | RESULTS

2.1 | Architecture of the H5N1 virion

Figure 1 (left) presents an artist's depiction of a cross-sectional view of the H5N1 virion (Goodsell 2024) that integrates 3D structure information from MX and 3DEM. The RNA genome (yellow) is shown packaged/protected by many copies of the nucleoprotein (NP, orange), surrounded by a shell of matrix proteins (M1, light blue), and encased within a lipid envelope (mauve). Three distinct populations of surface proteins embedded within the lipid bilayer viral envelope include hemagglutinin (HA, purple), neuraminidase (NA, magenta), and a proton-conducting channel (M2, purple). Once the virus enters a host cell nucleus, the RNA-dependent RNA polymerase (RdRp, light red) transcribes viral mRNAs and replicates the viral genome (Du et al. 2023). Experimentally determined 3D structures of functionally important parts of most H5N1 influenza virus proteins are archived in PDB and illustrated using RCSB.org Mol* (Figure 1, right). Hereafter, we exemplify the use of RCSB.org Mol* based tools to view H5N1 protein structures and explore their function.

2.2 | Exploring the H5N1 proteome in 3D

2.2.1 | Hemagglutinin

Hemagglutinin (HA) is the major viral surface protein. It is a homotrimeric glycoprotein in the influenza virus

envelope that recognizes sialic acid-containing glycans displayed on host-cell surfaces to gain viral entry. The protein is synthesized as a long single polypeptide chain (HA0). Late in the infection, host cell proteases cleave HA0 into two segments (HA1, the sialic acid receptor binding subunit; HA2, the membrane fusion subunit), which are covalently linked to one another by disulfide bridges (Antanasijevic et al. 2020). At the time of writing, PDB holdings included >500 experimentally determined structures of influenza A virus HAs.

The first H5N1 HA structure contributed to PDB was determined using MX at ~2.95 Å resolution (PDB ID 2fk0) (Stevens et al. 2006). A snapshot of the homotrimeric structure is displayed in the top left corner of its SSP (Figure 2a). The image consists of three copies of each of the HA1 (green, orange, and purple) and HA2 (yellow, light green, and pink) protein chains. Overall, the stalk-like hemagglutinin trimer has a bulbous head with three receptor binding sites for binding sialic acid moieties occurring within host-cell surface glycoprotein glycans. All HAs are glycosylated. Blue cubes and green spheres linked to the polypeptide chains shown throughout Figure 2 denote experimentally observed carbohydrate post-translational modifications.

Single-mouse-click of the Structure tab or the hyperlinked Structure button at the bottom of the thumbnail image shown in the top left corner of the SSP (Figure 2a) launches RCSB.org Mol*, revealing the 3D structure at the atomic level (Figure 2b). It utilizes (i) a Sequence panel that lists polymer sequences and small molecule entities present in the structure; (ii) a 3D

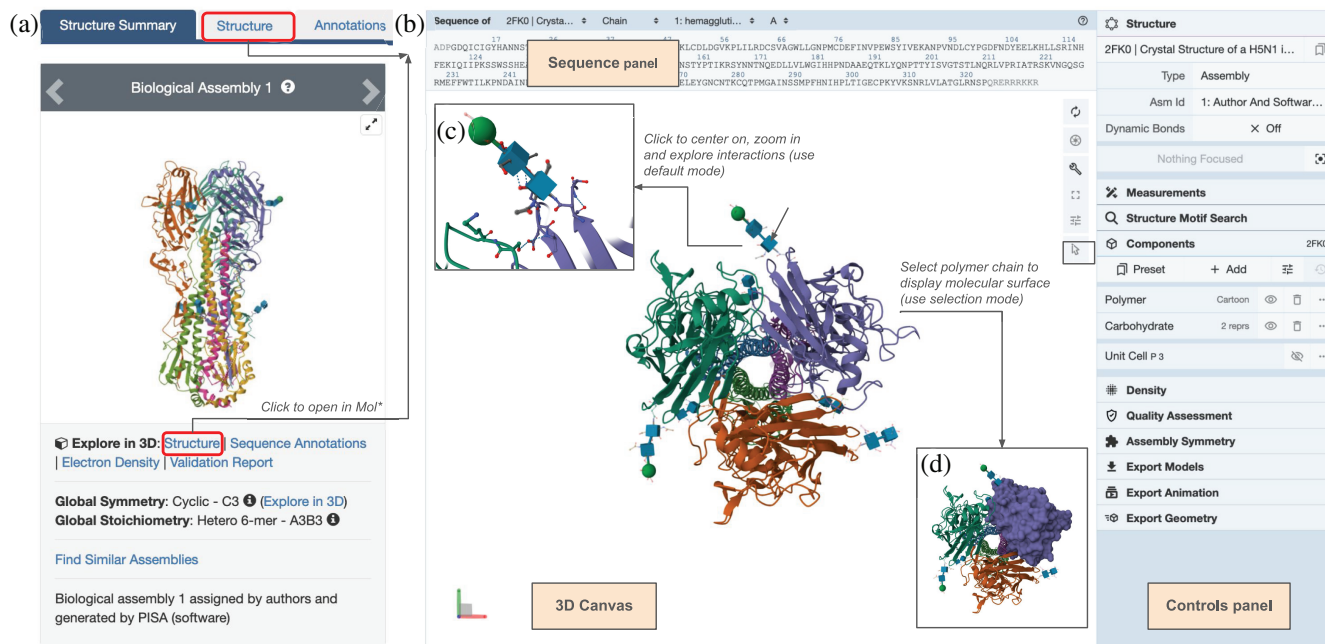


FIGURE 2 Structure of H5N1 HA trimer (PDB ID 2fk0; Stevens et al. 2006). (a) Screenshot of the top left corner of the SSP. (b) Single-mouse-click of the Structure tab or the hyperlinked word Structure (red boxes in (a)) launches RCSB.org Mol*, revealing Sequence panel, 3D Canvas, and Controls panel. (c) Close-up view of a covalently-bound glycan attached to Asn169, illustrated using Symbol Nomenclature for Glycans representation (Neelamegham et al. 2019; Shao et al. 2021), showing nearby covalent and non-covalent interactions. (d) View of the symmetric homotrimer with one HA depicted with molecular surface representation (invoked in Mol* Selection mode).

Canvas that displays the biomolecule structure interactively; and (iii) a Controls panel that provides options for managing the graphical display. [RCSB.org Mol*](https://www.rcsb.org/mol) operates in two different modes. In default mode, clicking on any part of the structure positions it at the center of the 3D Canvas, zooms in, and displays both covalent and non-covalent interactions within a 5 Å radius. The outcome of a single-mouse-click on the glycan marked with a black arrow is shown in Figure 2c. The Mol* Selection mode is activated by clicking on the open arrow button in the vertical menu on the right side of the 3D Canvas. Doing so presents a horizontal menu at the top of the 3D Canvas with options to select and change representations and colors of the selected portion of the 3D biostructure. Figure 2d shows the HA protein representation changed to a molecular surface representation for one protomer within the symmetric homotrimer (see Figure S2 for details).

Every SSP offers various single-mouse-click features providing immediate access to different representations and aspects of structure. The Sequence Annotations button visible in Figure 3a presents the 1D polymer sequence and 3D biostructure in side-by-side linked panels (Figure 3b). Clicking on the Electron Density button (Figure 3a) shows available experimental density maps (Figure 3c). Clicking on the Validation Report button (Figure 3a) displays the entire structure, colored based on structure quality—parts of the

structure with no or few wwPDB validation-reported issues are shown in blue, while regions with more issues are shown in orange and yellow (Figure 3d). Although not directly available from the SSP, an additional validation color scheme for MX structures identifies per-residue confidence based on the real-space correlation coefficient (RSCC) of experimentally determined electron density versus atomic coordinates (Shao et al. 2022) (Figure 3e, available from the Quality Assessment section of the controls panel in [RCSB.org Mol*](https://www.rcsb.org/mol)). See Figure S3 for details.

2.2.2 | Neuraminidase

Neuraminidase (NA) is the second major influenza virus surface glycoprotein. It is an enzyme that acts on glycans to cleave terminal sialic acid residues (from both viral and host cell glycoproteins). The enzymatic activity of NA is essential for the efficient release of newly assembled virions from host cells because it prevents them from binding and being endocytosed. At the time of writing, PDB holdings included >200 experimentally determined structures of influenza A virus NAs.

Many of the NA structures in PDB provide atomic-level insight into binding of inhibitors within the enzyme active site. One such inhibitor is a commonly used small-molecule drug treatment for influenza virus

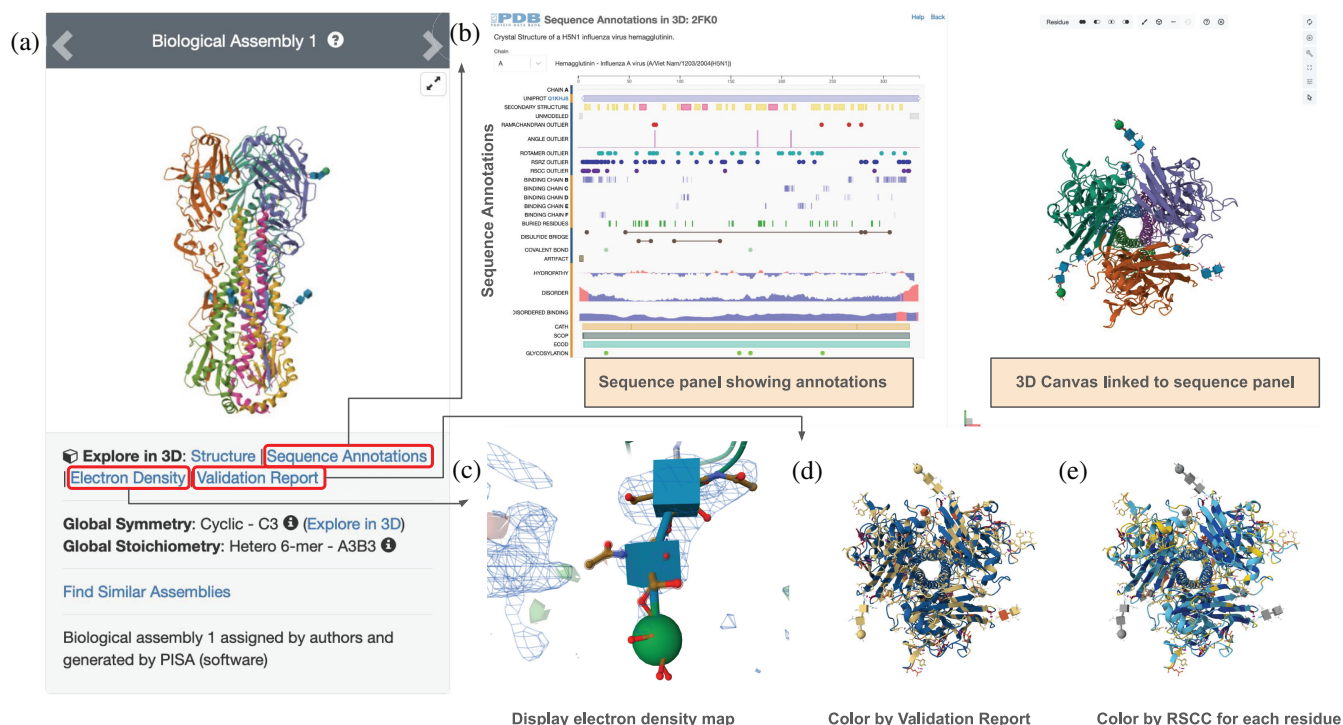


FIGURE 3 Exploring the H5N1 HA homotrimer (PDB ID 2fk0; Stevens et al. 2006). (a) Screenshot of the top left corner of the SSP. (b) Interface showing Sequence Annotations mapped onto the 3D structure in the linked side-by-side view. (c) Electron density map for a selected portion of a monomer structure. (d) Ribbon representation of the HA homotrimer colored by wwPDB Validation Report issues. (e) Same HA homotrimer colored by individual residue RSCC values.

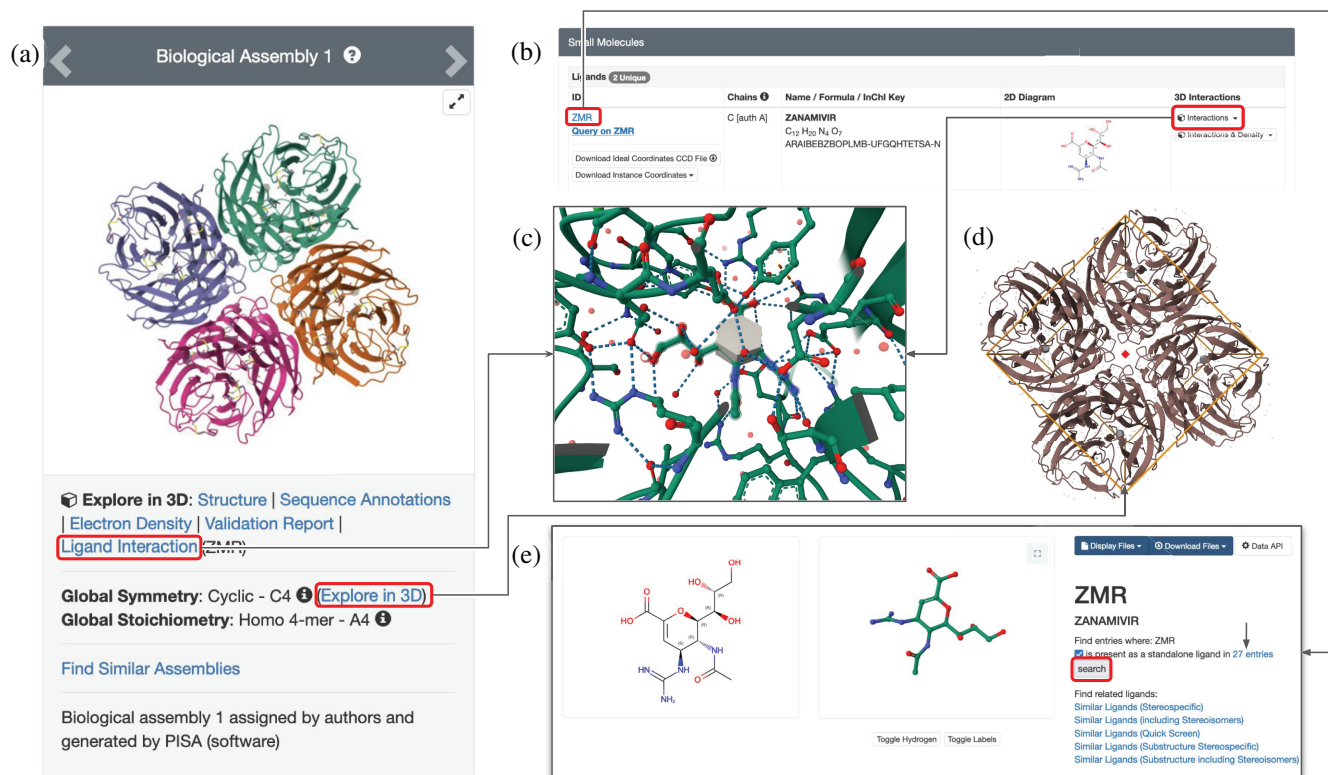


FIGURE 4 Co-crystal structure of H5N1 NA inhibited by zanamivir (PDB ID 3ckz; Collins et al. 2008). (a) Screenshot of the top left corner of the SSP. (b) Exploring covalent and non-covalent interactions in the vicinity of the bound drug. (c) Screenshot of the Small Molecules section of the SSP. (d) Screenshot of the Ligand Summary Page for zanamivir (ZMR), showing 2D and 3D images of the drug. (e) Exploring the global symmetry of the structure with the C4 symmetry axis denoted by a red square.

infections—zanamivir (brand name Relenza, wwPDB Chemical Component Dictionary, or CCD ID ZMR). The first PDB structure of an H5N1 NA-zanamivir complex was determined using MX at 1.9 Å resolution (PDB ID 3ckz) (Collins et al. 2008). The enzyme is a symmetric homotetramer (Figure 4a) of a largely β -sheet protein stabilized by several disulfide bridges. The active site of each protomer in PDB ID 3ckz is occupied by zanamivir. Drug–protein interactions can be viewed with [RCSB.org](https://www.rcsb.org/mol) Mol* by single-mouse-click of the Ligand Interaction button at the top left corner of the SSP (Figure 4a), or the Interactions button in the Small Molecules section of the SSP (Figure 4b), or by single-mouse-click of the bound ligand displayed in the 3D Canvas (in default mode). The resulting zoomed-in view is centered on zanamivir. All amino acid residues located within 5 Å of the ligand are shown in ball-and-stick representation together with nearby non-covalent intra- and inter-molecular interactions (Figure 4c, blue dashed lines denote hydrogen bonds or H-bonds, orange dashed lines represent π -cation interactions).

Single-mouse-click of a button in the top left corner of the SSP (Figure 4a) supports exploration of Global Symmetry in a structure (Duarte et al. 2022). In Figure 4d, the symmetry of the tetrameric NA assembly

is denoted by the small red square positioned at the tip of the symmetry axis and a cube-shaped cage denoting Cyclic Global C4 Symmetry. Clicking the CCD ID in the Small Molecules section of the SSP (Figure 4b) opens the corresponding Ligand Summary Page (Figure 4e), wherein Mol* is used to display the 3D structure of the ligand interactively. Single-mouse-click of the Search button in this panel executes a search for all PDB structures containing this ligand. At the time of writing, this search returned 27 hits, including structures of 24 influenza NAs, one human sialidase (PDB ID 2f0z) (Chavas et al. 2005), and two bacterial proteins. See Figure S4 for details.

2.2.3 | M2 ion channel

The M2 protein forms an ion channel embedded in the viral envelope. It is necessary for the release of the viral genome into the cytoplasm of the host cell at the beginning of the infectious process and later during virion packaging (Pinto and Lamb 2006). Once the virion is internalized by host cell endocytosis, the resulting endosome is acidified, and the viral membrane fuses with the endosome lipid bilayer. In addition to stabilizing the fusion-promoting conformation of HA (Skehel and

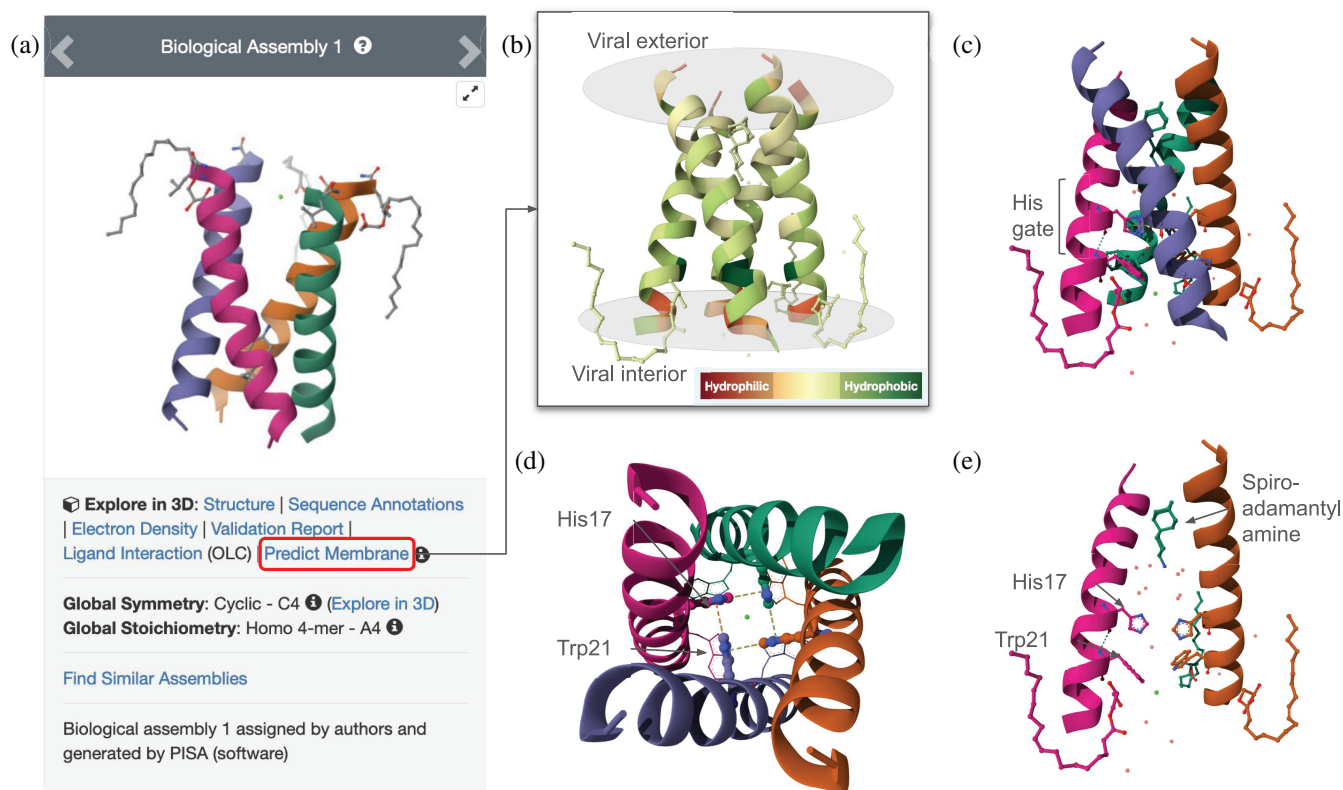


FIGURE 5 Structure of the H5N1 M2 ion channel pore. (a) Screenshot of the top left corner of the SSP for PDB ID 6nv1 (Thomaston et al. 2020) showing the Predict Membrane button within a red rectangle. (b) RCSB.org Mol* image showing the predicted membrane orientation for PDB ID 6nv1 as circular gray planes corresponding to outer and inner membrane leaflets. Polypeptide chain segments of M2 are shown in ribbon representation, color-coded for hydrophobicity versus hydrophilicity (see panel inset). (c) Side-view of the M2 ion channel shown in ribbon representation with different colored polymer chains (orientation as in (b), location of His17 [auth 37] ion gate is marked). (d) Outside-to-inside view of the M2 ion focusing on key residues forming the ion gate (histidine residues in ball-and-stick representation, tryptophan residues shown as thin lines). (e) Identical view of the M2 ion channel as in panel (c), with two of the polymer chains hidden for clarity showing His17 [auth 37] and Trp21 [auth 41] and the channel blocking ligand.

Wiley 2000), the pH drop activates the M2 ion channel, allowing protons to enter the virion nucleocapsid. Doing so disrupts interactions between the viral ribonucleoprotein (RNP), the matrix protein 1 (M1), and the viral membrane, thereby uncoating and ejecting the viral genome into the host cell cytoplasm (Pielak and Chou 2011). During virus maturation in the normally acidic trans-Golgi network, the M2 ion channel acts locally to increase pH, thereby preventing newly translated HA proteins from assuming the fusion-promoting conformation.

The M2 ion channel is a 27-amino acid protein with a transmembrane α -helix that forms a symmetric homotetramer (PDB ID 6nv1) (Thomaston et al. 2020). At the time of writing, PDB archival holdings included >20 experimentally determined structures of influenza A virus M2 proteins. Much of our understanding of M2 function comes from MX studies of hydrophobic synthetic peptides, corresponding to the transmembrane portion of the ion channel (Figure 5a,b), reconstituted in the presence of membrane phospholipid-mimicking detergents plus spiro-adamantyl amine (a small-

molecule channel blocker). Single-mouse-click of the Predict Membrane button in the top left corner of the SSP (Figure 5a) displays the structure depicted in Figure 5b with two circular planes in the RCSB.org Mol* 3D Canvas, representing the predicted extent of the phospholipid bilayer.

Two amino acid sidechains (His17 [auth 37] and Trp21 [auth 41]), originate from each of the four α -helices pointing towards the center of the ion channel (Note: In some PDB structures, an amino acid residue may possess two identifying numbers: the PDB-assigned ordinal number (label_seq_id) and an author-provided number (auth_seq_id)). The sidechains of His17 [auth 37] interact with each other (via π - π interactions) and water molecules (via H-bonds) to stabilize the tetrameric assembly and form a gate preventing H^+ ion flow across the viral membrane. RCSB.org Mol* provides various options to select or change representations, hide or show parts of a structure, and display specific interactions to illustrate a molecular story. Key elements of the M2 gate are shown in side- (Figure 5c), top- (Figure 5d), and cut-away side-views (Figure 5e). Within the acidified endosome (pH $\sim$ 5.5), the

imidazole moiety of His17 [auth 37] is protonated, destabilizing interactions within the M2 gate and allowing passage of protons into the nucleocapsid. Single-mouse-click of the His residue (in default mode) displays interactions within 5 Å. See Figure S5 for details.

2.2.4 | Nucleoprotein

The H5N1 nucleoprotein (NP) is responsible for packaging the segmented RNA genome within the nucleocapsid. In addition to protecting against nuclease digestion, NP interacts with the viral RNA-dependent RNA polymerase (RdRp) and plays a role in determining whether the viral genome is transcribed into mRNAs encoding viral proteins or used as a template for genome replication (Ng et al. 2008). Multiple NPs bound to the RNA genome (viral ribonucleoprotein or vRNP) become localized within the host cell nucleus early in the infection process. A bipartite nuclear localization signal (NLS) found within the NP facilitates active nuclear import of vRNPs via binding to karyopherins (Ozawa et al. 2007). Later in the infectious process, this polypeptide chain segment becomes occluded by the matrix protein M1 to prevent inappropriate reimportation of newly assembled vRNPs exported from the nucleus.

To examine the consequences of conformational changes due to RNA binding, NP structures with and without bound RNA can be superimposed using the RCSB PDB Standalone Mol* ([rcsb.org/3d-view](https://www.rcsb.org/3d-view); also accessible from the [RCSB.org](https://www.rcsb.org) homepage via the “Visualize” dropdown menu). This version of Mol* is not connected to any specific SSP. It supports visualization of multiple experimentally determined PDB structures and/or predicted CSMs and/or uploaded atomic coordinates. Figure 6a shows a ~ 3.3 Å resolution MX structure of the NP alone (PDB ID 2q06) (Ng et al. 2008) superimposed on a ~ 2.3 Å resolution MX structure of NP bound to RNA (PDB ID 7dyp) (Tang et al. 2021) generated by Standalone Mol*. Viewing either structure reveals that the NP is an α/β -protein (Figure 6b) with a solvent-accessible bipartite NLS (yellow in Figure 6b). Binding of an RNA oligonucleotide causes loops near the RNA binding site to become ordered (Figure 6c), while loops near the NLS become disordered (compare Figure 6b,d). See sect. 6 of Data S1 for details.

2.2.5 | M1 matrix protein

The M1 matrix protein is ~ 250 amino acids in length. It interacts with both the viral envelope and the segmented RNA genome to form the protein coat of the nucleocapsid.

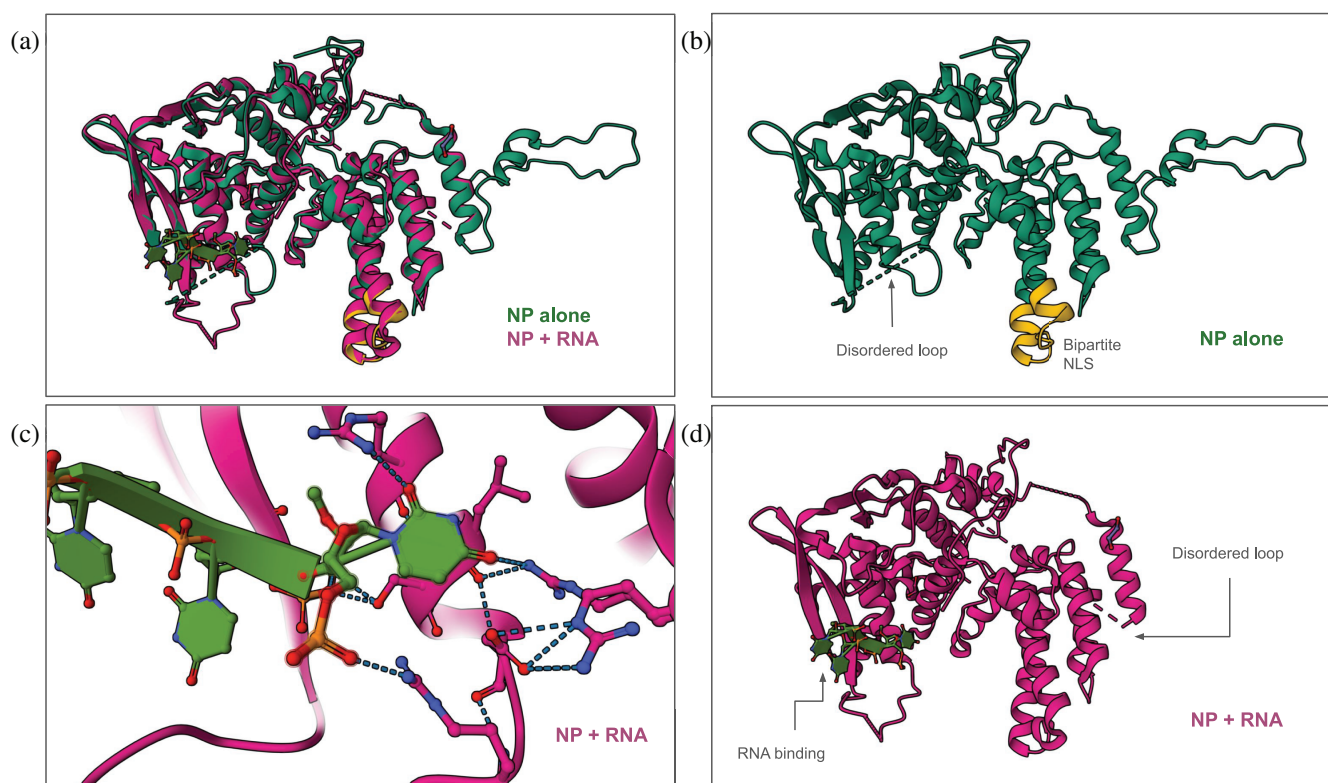


FIGURE 6 NP conformational changes upon RNA binding. (a) Superimposition of the structures of NP alone (PDB ID 2q06; green; Ng et al. 2008) and NP bound to RNA (PDB ID 7dyp; magenta; Tang et al. 2021). (b) Structure of NP alone, showing the yellow-colored bipartite NLS (residues 204 [auth 198] to 222 [auth 216]). (c) Zoomed in view of an RNA oligonucleotide bound to the NP showing intermolecular interactions in atomic detail. (d) Structure of NP bound to RNA.

At the time of writing, PDB holdings included ~20 experimentally determined structures of influenza A virus M1 proteins, although none were derived from an H5N1 virus. A ~3.4 Å resolution 3DEM structure of the M1 protein (PDB ID 7jm3; EMDB ID EMD-22384) (Selzer et al. 2020) revealed that the N- and C-terminal domains are largely α -helical and connected by a flexible linker. [RCSB.org](https://www.rcsb.org) Mol* tools exemplified in Figures 2 and 3 can be used to visualize and analyze this protein.

2.2.6 | RNA-dependent RNA polymerase

The viral RdRp is tripartite, consisting of polymerase acidic protein (PA), RdRp catalytic subunit (PB1), and polymerase basic protein 2 (PB2). At the time of writing, PDB holdings included >50 experimentally determined structures of influenza A virus RdRp heterotrimers. The first structure of an H5N1 RdRp heterotrimer was determined via 3DEM at ~3.2 Å resolution (PDB ID 8r1j; EMDB ID EMD-18818) (Staller et al. 2024). Again, [RCSB.org](https://www.rcsb.org) Mol* tools exemplified in Figures 2 and 3 can be used to visualize and analyze this complex assembly.

2.2.7 | Influenza A virus life cycle

The influenza A virus life cycle spans binding to sialylated cell-surface proteins, endocytosis, acidification of

the endosome, fusion of the viral lipid bilayer with the endosomal membrane, ejection of the viral nucleocapsid into the host cell cytoplasm, transport of the segmented viral genome into the nucleus, viral RNA transcription and translation, genome replication, and assembly of new virions. Viral transmembrane proteins are translocated into the endoplasmic reticulum (ER) and trafficked to the Golgi, where they induce budding. Cytoplasmic M1 packages the viral genome into budding virions. Finally, sialic acid cleavage by NAs facilitates the release of nascent virions (Carter and Iqbal 2024) ready to infect nearby cells or other animal or human hosts. To further exemplify how [RCSB.org](https://www.rcsb.org) Mol* has been deployed within various RCSB PDB structure display and analysis tools, two key steps of the viral life cycle are examined in 3D: (i) species-specific HA binding and (ii) RdRp switching between mRNA transcription and genome replication.

2.2.8 | HA binding to avian versus human sialic acid receptors

HA homotrimers displayed on the surface of the viral envelope recognize and bind to sialic acid receptors decorating host cell surface glycoproteins and glycolipids (Figure 7). Bird (or avian) host cells display glycans with sialic acid moieties α 2,3-linked to galactose, whereas glycans in human cells utilize α 2,6-linkages

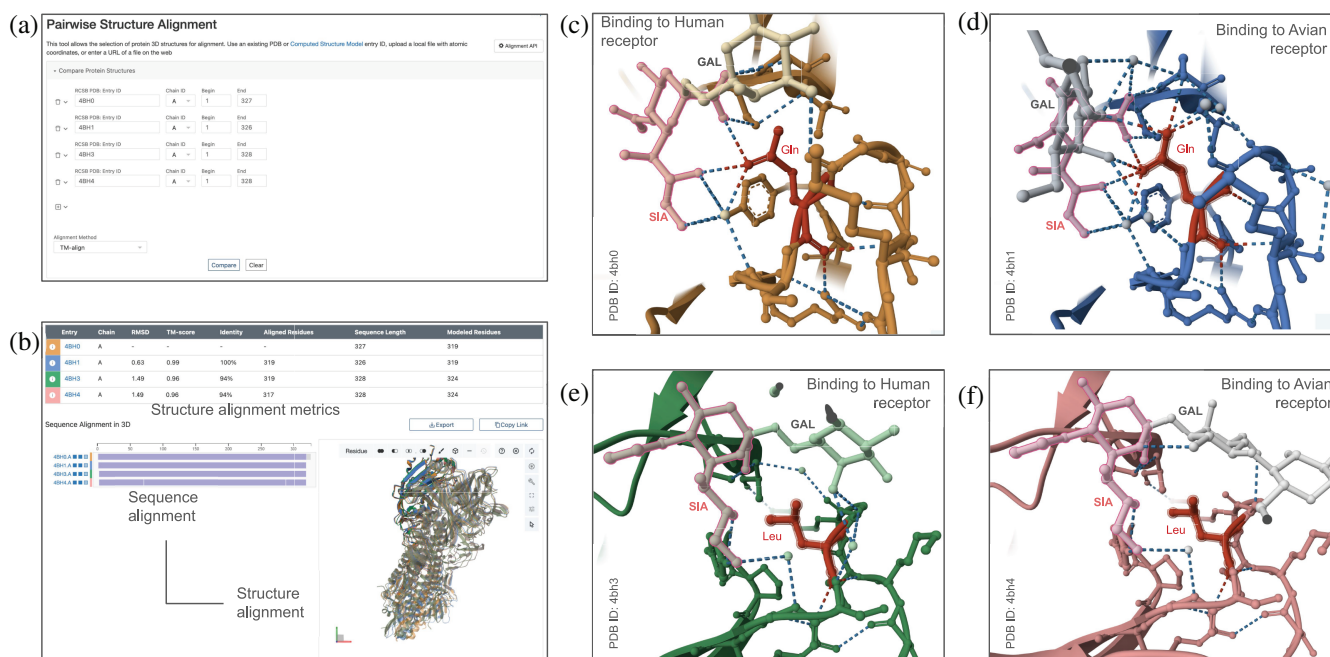


FIGURE 7 Comparing co-crystal structures of non-transmissible (Asn/Gln) and transmissible-mutant (Lys/Leu) HAs (Xiong et al. 2013). (a) Pairwise Structure Alignment Tool displaying the list of structures to be compared. (b) Results of the above structure comparison (structures are distinctly colored). (c) Non-transmissible HA (Asn/Gln) bound to a human sialic acid receptor (PDB ID 4bh0). (d) Non-transmissible HA (Asn/Gln) bound to an avian sialic acid receptor (PDB ID 4bh1). (e) Transmissible-mutant HA (Lys/Leu) bound to a human sialic acid receptor (PDB ID 4bh3). (f) Transmissible-mutant HA (Lys/Leu) bound to an avian sialic acid receptor (PDB ID 4bh4).

(Shinya and Kawaoka 2006). Thus, for H5N1 to cross the species barrier from birds to humans, changes in HA sialic acid receptor binding preferences must play a role.

The RCSB PDB Pairwise Structure Alignment Tool (Figure 7a; [rcsb.org/alignment](https://www.rcsb.org/alignment); also accessible from the [RCSB.org](https://www.rcsb.org) homepage via the “Analyze” dropdown menu; Bittrich et al. 2024b) was used to superimpose co-crystal structures of HAs from two H5N1 virions: non-transmissible (turkey/Turkey/1/2/005(H5N1): PDB IDs 4bh0—human-type sialic acid receptor; 4bh1—avian-type sialic acid receptor) and transmissible/mutant (A/Vietnam/1203/2004(H5N1): PDB IDs 4bh3—human-type sialic acid receptor; 4bh4—avian-type sialic acid receptor) (Xiong et al. 2013). The default TM-align option was used, which leverages secondary structure information to perform a pairwise structure alignment (Zhang and Skolnick 2005). Structure superimposition confirmed that the two proteins are highly similar in 3D structure with root-mean-square-deviation or RMSD ~ 1.5 Å (PDB IDs 4bh3 vs. 4bh0, sequence identity $\sim 94\%$, 319 equivalent α -carbon atomic pairs) (Figure 7a,b). Within the sialic acid receptor binding site of transmissible/mutant A/Vietnam/1203/2004(H5N1) HA, two amino acid differences are thought to be important for transmissibility—i.e., Asn 220 is changed to Lys and Gln 222 is changed to Leu (Note: The Xiong et al. 2013 manuscript refers to these amino acids as Asn 224 and Gln 226, respectively. The Pairwise Structure Alignment Tool refers to these amino acids by their assigned ordinal numbers (Asn 222 and Gln 224). However, the paragraph below uses the author-specified numbering (Asn 220 and Gln 222)).

Figure 7c–f depicts the binding of non-transmissible HA (Asn/Gln) and transmissible HA (Lys/Leu) to human- and avian-type sialic acid receptors. HA (Asn/Gln) makes several H-bonds involving Gln 222 and Tyr 91 sidechains with the human-type sialic acid receptor (PDB ID 4bh0; Figure 7c). When binding to the avian-type sialic acid receptor, Gln 222 in HA (Asn/Gln) makes similar interactions to those depicted in Figure 7c, plus it forms H-bonds to a nearby galactose moiety and Ser 133 (PDB ID 4bh1; Figure 7d). These additional interactions enable HA (Asn/Gln) to bind more tightly and selectively to the avian sialic acid receptor than to its human counterpart and may explain non-transmissibility. Because the Asn 220 sidechain is oriented away from the sialic acid, it has a limited impact on receptor binding.

In contrast, transmissible-mutant HA (Lys/Leu) makes very few H-bonds with either type of bound glycan (PDB ID 4bh3—human-type sialic acid receptor, Figure 7e; PDB ID 4bh4—avian-type sialic acid receptor, Figure 7f). In both cases, the hairpin conformation of the bound glycan wrapping around the hydrophobic Leu side chain is markedly different from that observed for either of the HA (Asn/Gln) co-crystal structures

shown in Figure 7c,d. Lys 220 forms H-bonds with backbone atoms of Ala 134 and Cys 90, in both human-type and avian-type receptor-bound structures. Additional water-mediated H-bonds are seen when HA (Lys/Leu) binds to the human-type sialic acid receptor. Biophysical characterization of HA–receptor interactions (Xiong et al. 2013, fig. 1a) documented that HA (Lys/Leu) binds human-type sialic acid receptors (Figure 7e) more tightly than avian-type sialic acid receptors (Figure 7f), thereby facilitating infection of both human and avian hosts. See Figure S7 for details.

More recently, Wilson and co-workers reported the experimental structure determination of an HA from the first-reported human-infecting bovine H5N1 virus (A/Texas/37/2024, Texas) using MX at ~ 2.7 Å resolution (PDB ID 9diq) (Lin et al. 2024). Additional structural studies from Wilson’s research group revealed (i) avian-type sialic acid receptor binding preference (PDB ID 9dip), and (ii) human-type sialic acid receptor binding preference for the Gln226Leu modified form of the “Texas HA” (PDB ID 9dio). Moreover, binding to human-type sialic acid receptors was enhanced for the Gln226Leu/Asn224Lys doubly modified form of the “Texas HA.”

2.2.9 | Viral transcription and replication

Once inside the host cell, the segmented viral genome remains bound to one or more NPs, forming viral ribonucleoproteins (vRNPs). The NP NLS mediates the transport of vRNPs and associated RdRps into the nucleus, wherein they both synthesize viral protein mRNAs (transcription) and replicate the segmented viral genome. Primers for transcription initiation are obtained by the RdRp using a process known as “cap snatching” (Plotch et al. 1981). Resulting viral mRNAs are translated by host cell ribosomes, yielding the accumulation of the RdRp, NP, M1, and NP proteins in the cytoplasm, and the HA, NA, and M2 proteins embedded in the lipid bilayer of the ER.

Replication of the influenza A viral genome in human cells requires one or the other of two related acidic nuclear phosphoproteins (ANP32 A or B) (Staller et al. 2019). The N-terminal segments of ANP32 A or B contain several leucine-rich repeats, while their C-terminal portions include many negatively charged amino acids thought to interact with positively charged viral NPs. A ~ 3.2 Å resolution 3DEM structure of the H5N1 RdRp heterotrimer (PA, PB1, and PB2) bound to human ANP32B (PDB ID 8r1j) (Staller et al. 2024) is depicted in Figure 8a. The PA subunit interacts with the globular N-terminal domain of ANP32B (Figure 8a) (Note: Atomic coordinates for the C-terminal domain of ANP32B are absent in PDB ID 8r1j).

To learn more about the missing C-terminal portion of ANP32B, [RCSB.org](https://www.rcsb.org) can be queried for other

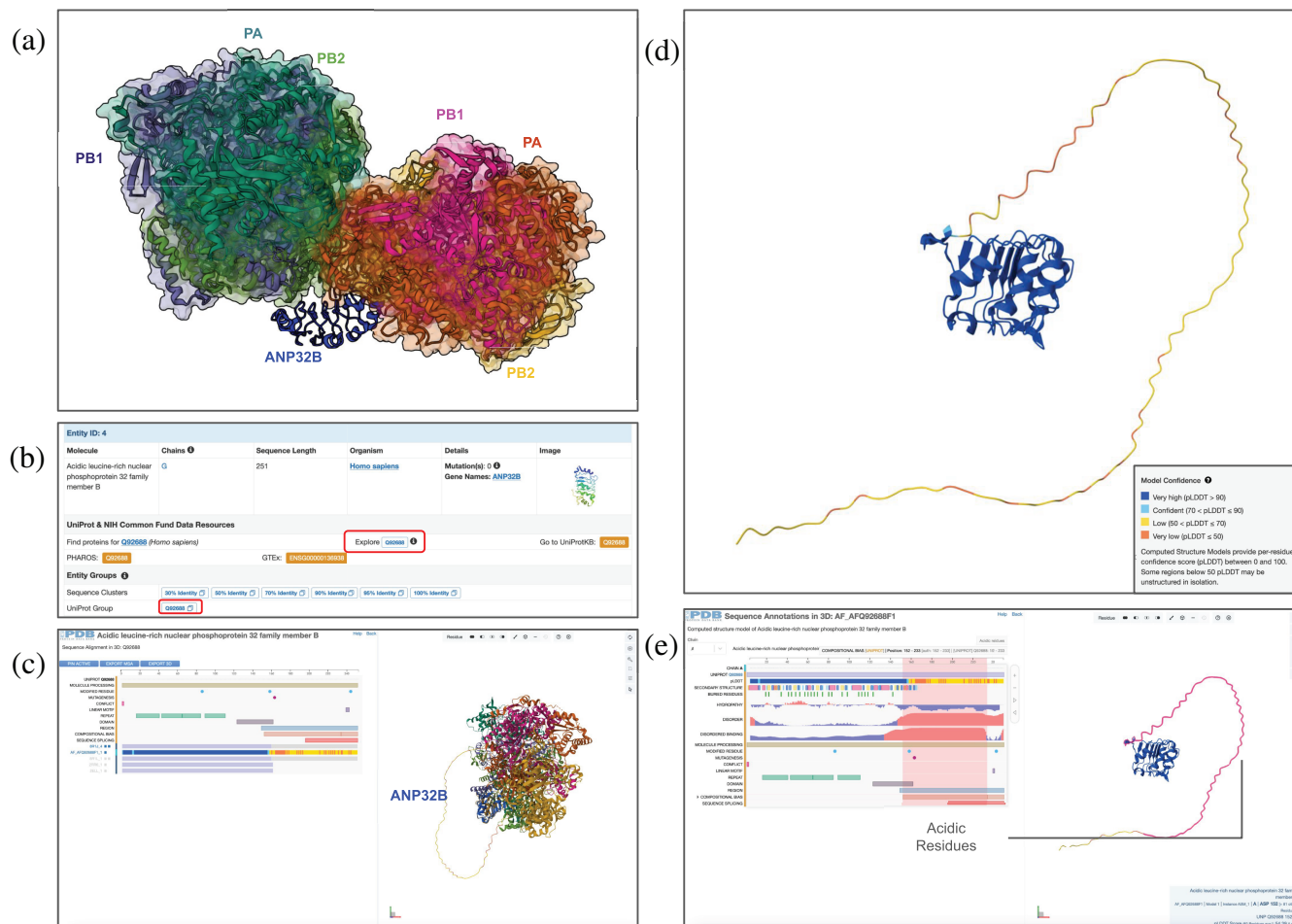


FIGURE 8 3DEM structure of two copies of H5N1 RdRp bound to one copy of human ANP32B (PDB ID 8r1j; Staller et al. 2024). (a) Component proteins (PA, PB1, and PB2) are shown with semi-transparent molecular surfaces, bound to ANP32B (dark blue, ribbon representation). (b) Macromolecules section of the SSP with information and links for the human ANP32B protein. (c) Sequence Alignment in 3D page for human ANP32B with the amino acid sequence of UniProt Accession Number Q92688. (d) Mol* view of CSM AF_AFQ92688F1, shown in ribbon representation, colored by pLDDT values. (e) Sequence Annotations in 3D showing linked 1D sequence panel (left) and 3D structure display (right).

structures that may include this region. A single-mouse-click of the UniProt Group button from the SSP for PDB ID 8r1j (Figure 8b) opens a Group Summary Page (Segura et al. 2023) displaying information about 3D biostructures, including UniProt Accession Number Q92688 available at [RCSB.org](https://www.rcsb.org). Visualization of the CSM for the full-length protein (AF_AFQ92688F1, predicted using AlphaFold2) (Jumper et al. 2021) reveals that the C-terminal portion of the human protein (Figure 8d,e) is likely to be intrinsically disordered (at least in the absence of a binding partner). Inspection of the Sequence Alignment in 3D page (Figure 8c) allows selective viewing of the CSM pre-aligned with each experimentally determined PDB structure of ANP32B. The polypeptide chain of the CSM is colored (by default) based on per-residue prediction confidence (predicted local distance difference test or pLDDT), ranging from 0 (lowest confidence, orange) to 100 (highest confidence, dark blue) (Tunyasuvunakool

et al. 2021). Viewing Sequence Annotations for ANP32B reveals that its C-terminal region is rich in acidic residues (Figure 8e), which explains why it is predicted to be unstructured in isolation. See Figure S8 for details.

2.3 | Preparing for a possible H5N1 epidemic/pandemic

Seasonal influenza viruses cause acute respiratory infections, some of which may require hospitalization and, in rare cases, prove fatal. Zoonotic influenza viruses (causing bird flu, swine flu, etc.) can spread to humans, potentially causing a global public health crisis. Enhanced surveillance, vaccination programs, non-pharmacologic precautions (e.g., frequent hand washing, N95/KN95 mask usage), and timely clinical management with antiviral agents are proven means of supporting public health and preventing fatalities.

PDB data and structure-guided approaches are widely used across the biopharmaceutical industry to support the discovery and development of new small-molecule drugs (Burley 2021; Burley et al. 2024; Westbrook et al. 2020; Westbrook and Burley 2019). Below, we describe how visualizing and analyzing 3D biostructure information freely available from [RCSB.org](https://www.rcsb.org) can help us understand how drugs act and how they can bind to other human proteins (potentially giving rise to unwanted off-target-related side effects). This information can help guide the discovery and development of next-generation antiviral agents to reduce case-fatality rates in the face of an H5N1 epidemic/pandemic.

The three classes of drugs currently used to treat influenza virus infections include NA inhibitors, adamantanes, and inhibitors of the RdRp cap-dependent endonuclease activity. The US Food and Drug Administration (FDA) has approved two chemically similar NA inhibitors (zanamivir and oseltamivir—wwPDB CCD IDs ZMR and G39, respectively) that prevent the spread of infection with efficacy against a broad range of NA subtypes (Note: Oseltamivir is an oral pro-drug, which is converted during passage through the liver into its active metabolite oseltamivir carboxylate). The US FDA-approved drug amantadine (wwPDB CCD ID 308) blocks the M2 ion channel and interferes with viral

entry into the host cell cytoplasm. Liabilities limiting its usage include toxic side effects and the emergence of resistance. The most recently approved influenza drug baloxavir marboxil is a prodrug rapidly converted to baloxavir acid (wwPDB CCD ID EwZ) by arylacetamide deacetylases occurring in cells of the blood, liver, and lumen of the small intestine.

Comparing MX co-crystal structures of NA bound to zanamivir (PDB ID 3ckz) (Collins et al. 2008) and oseltamivir carboxylate (PDB ID 3cl0) (Collins et al. 2008) can reveal amino acids critical for the binding of both drugs. The Pairwise Structure Alignment Tool, described above, can be used to accomplish this objective. Visual inspection of the binding sites of both drugs reveals a triad of Arg residues (Arg 36 [auth 118], Arg 211 [auth 292], Arg 286 [auth 371]) that form analogous ionic interactions with zanamivir and with oseltamivir carboxylate (Figure 9a,b). Binding mode similarity suggests that it is worth enquiring whether a similar Arg-triad exists in human and/or other viral proteins. Answering this question could help identify potential off-target interactions with human proteins and/or identify additional viral targets for which one or both drugs may be effective.

The [RCSB.org](https://www.rcsb.org) Structure Motif Search Tool (Bittrich et al. 2020) is ideally suited to this task. By visualizing the structure of NA bound to zanamivir (PDB ID 3ckz)

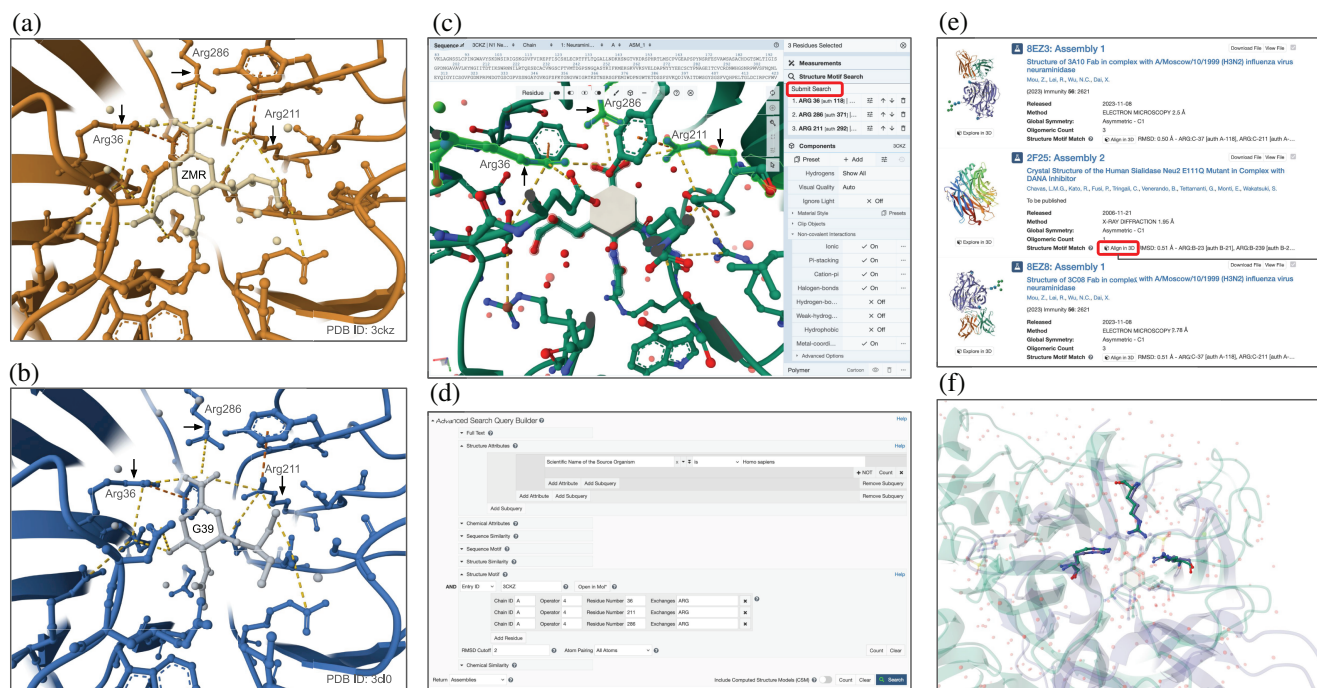


FIGURE 9 Launching a Structure Motif Search using Mol*. (a) Panel showing zanamivir binding to H5N1 NA (PDB ID 3ckz; Collins et al. 2008). Interacting amino acids within 5 Å are shown in ball-and-stick representation, with ionic interactions rendered as orange dashed lines. (b) Panel showing ionic interactions stabilizing oseltamivir binding (PDB ID 3cl0; Collins et al. 2008). (c) Selection of the Arg-triad in PDB ID 3ckz for Structure Motif Search. (d) Advanced Search Query Builder set up to run the Structure Motif Search (specified in (c)) and refined to list only result structures that include at least one human protein. (e) Structure Motif Search results page listing human sialidase (NEU2). (f) Structure alignment of the Arg-triad in both the H5N1 NA (green) and human NEU2 (violet).

(Collins et al. 2008) in Mol* (Figure 9c; hydrogen bonds have been hidden for the sake of clarity), each member of the Arg-triad can be selected (in Selection mode) to define the search. At the time of writing, single-mouse-click of the Search button (Figure 9c) returned more than 775 experimentally determined PDB structures. Selecting the subset including at least one human protein narrows the result set to ~115 cases. Scrolling through the query results reveals Sialidase Neu2 (PDB ID 2f25) (Chavas et al. 2005). A representation of this refined query in the Advanced Search Query Builder (Figure 9d) can be saved for re-running at a later date. See Figure S9 for details.

Single-mouse-click of the Align in 3D button on the results page (Figure 9e) delivers a comparison of the query structure (PDB ID 3ckz) and the Sialidase Neu2 search result (PDB ID 2f25) in a new Mol* window (Figure 9f). The RMSD for non-hydrogen atom positions of the Arg-triad is ~0.5 Å for PDB ID 3ckz versus PDB ID 2f25. In contrast, comparing the structures of the viral NA (PDB ID 3ckz) and the human sialidase (PDB ID 2f25) yields RMSD ~3.9 Å (sequence identity ~11%, 253 equivalent α -carbon atomic pairs). While these two proteins appear somewhat distantly related in evolutionary terms, the Arg-triad is common to both structures suggesting that zanamivir may bind to the active site of human NEU2.

In the clinic, neuropsychiatric adverse events and fatalities due to oseltamivir therapy have been reported (Fuyuno 2007; Maxwell 2007). A follow-up study examining in vitro inhibitory effects of both oseltamivir carboxylate and zanamivir on select human sialidases (NEU1, NEU2, NEU3, and NEU4) documented that oseltamivir carboxylate does not bind to any of the four human enzymes (albeit the studies use recombinant preparations of enzyme) at 1 mM concentration, whereas zanamivir binds to NEU1, NEU2, NEU3, and NEU4 in the micromolar range (Hata et al. 2008). At the time of writing, repeating the Arg-triad Structure Motif Search to include more than 1 million CSMs available at [RCSB.org](https://www.rcsb.org) (by activating the CSM toggle switch visible in Figure 9d) returned 10 human protein CSMs—7 are sialidases (wherein the Arg-triad matches with RMSD <1 Å) plus Transmembrane protein 132A, Breast cancer type 1 susceptibility protein, and Ubiquitin carboxyl-terminal hydrolase 6 (all with higher RMSDs ~1.5–2 Å).

Oseltamivir treatment adverse events may be due to the binding of its active metabolite to one or more human sialidases not previously tested. Not surprisingly, the Arg-triad Structure Motif Search also returned experimentally determined structures of other influenza virus NAs. As the genomes of new influenza viruses are isolated and sequenced and relevant NA structures are deposited into the PDB, [RCSB.org](https://www.rcsb.org) tools described herein could be used to understand resistance to zanamivir and/or oseltamivir therapy (if/when it becomes clinically apparent) and provide structural guidance for

the discovery/development of novel antiviral drugs based on chemically distinct chemotypes.

The PDB archive also houses co-crystal structures of amantadine and baloxavir acid bound to their respective target proteins. PDB ID 3c9j (Stouffer et al. 2008) revealed at the atomic level the molecular mechanism of action of amantadine blocking an influenza A H5N1 M2 ion channel. Similarly, PDB ID 6fs6 (Omoto et al. 2018) revealed at the atomic level the molecular mechanism of action of baloxavir acid inhibiting influenza A H1N1 RdRp cap-dependent endonuclease. Both of these structures can be explored using [RCSB.org](https://www.rcsb.org) web portal tools as described for the NA inhibitors above.

Successful management of any zoonotic influenza A H5N1 virus pandemic/epidemic will require large-scale vaccination. Viral subtypes known to circulate routinely in humans targeted by seasonal flu vaccines include A(H1N1), A(H3N2), and B(Victoria) (Centers for Disease Control and Prevention and National Center for Immunization and Respiratory Diseases (NCIRD) 2024). Their main immunogens are HA proteins (Nichol and Treanor 2006). Because influenza virus RdRps are error-prone, viral genomes undergo continuous evolution potentially rendering seasonal vaccines ineffective against new subtypes (including H5N1). At the time of writing, the globally predominant H5N1 clade is 2.3.4.4b (Li et al. 2020). Several inactivated or attenuated vaccines against influenza A H5N1 viruses have been tested in animal models (Cargnin Faccin and Perez 2024).

3D structures of H5N1 HA proteins and those bound to antibody Fab fragments (e.g., PDB IDs 3fku (Sui et al. 2009), 4fqi (Dreyfus et al. 2012), and 8txt (McIntire et al. 2024)) archived in the PDB are freely available to help guide vaccine designers (Goodsell and Burley 2022). Recently, a nucleoside-modified mRNA vaccine encoding clade 2.3.4.4b HA was tested in animal models (Furey et al. 2024). Experience with structure-guided design, development, and regulatory approval of mRNA vaccines against SARS-CoV-2 (Burley 2025; Corbett et al. 2020) bodes well for the successful management of possible zoonotic H5N1 influenza virus epidemics (or global pandemic).

3 | SUMMARY OF EXPLORATIONS USING RCSB.org Mol*

Exploration of influenza A H5N1 virus proteins (Figure 1) and their complexes with one another and small-molecule drugs has demonstrated various ways to use [RCSB.org](https://www.rcsb.org) Mol* and its deployments across RCSB PDB software tools.

Our explorations fell into three groups:

1. **Exploration of individual structures** to learn about overall shape, oligomeric assemblies and

their symmetry, detailed structural features, and intermolecular interactions. PDB structures and CSMs can be visualized in various formats. wwPDB validation report information, experimental density maps, and structure–function annotations can be mapped onto 3D biostructures to learn more about structure quality and the biochemical and biological role(s) played by each protein (Figures 2–5).

2. **Comparing structures** to understand conformational changes due to RNA binding and the impact of mutations/sequence variations on 3D structure. [RCSB.org](https://www.rcsb.org) structure comparison features and our Pairwise Structure Alignment Tool support analyses of diverse structures (Figures 6, 7, and 9). In addition, groups of structures related by UniProt Accession Number (100% identity) or polymer sequence (lower sequence identity) can be viewed together and compared to learn about ligand binding and conformational flexibility (Figure 8).
3. **Identifying related structures** using the [RCSB.org](https://www.rcsb.org) Structure Motif Search Tool can reveal common structural features of interest to both evolutionary biologists and drug hunters (Figure 9).

The [RCSB.org](https://www.rcsb.org) Mol*-based tools can support identification, exploration, examination, and comparison of structures relevant to almost any biological theme or topic of interest.

4 | COMMUNICATING KNOWLEDGE AND SUPPORTING COLLABORATIONS

[RCSB.org](https://www.rcsb.org) supports several ways to export and store Mol* visualizations. Users can bookmark or modify links from SSPs by changing relevant parameters. For preparation of publication-ready figures, Mol* offers an image export option via the Screenshot/State menu (camera lens aperture icon) located in the top-right corner of the Mol* 3D Canvas. This menu also supports the export of the current state or session file saving (bundling all source data and required when working with local files).

Three additional data export options are available at the bottom of the right-hand Mol* panel:

1. **Export models:** Downloads atomic coordinates for PDB structures in PDBx/mmCIF or BinaryCIF formats.
2. **Export animation:** Creates short movies, such as a 360° rotation of a 3D structure in .mov file format.
3. **Export geometry:** Converts atomic coordinates into 3D shapes ready for 3D printing in .glb, .stl, .obj, or .usdz file formats.

5 | CONCLUSION

The [RCSB.org](https://www.rcsb.org) Mol* is a powerful, versatile, and easily accessible, no-cost, web-based molecular visualization tool. It can perform complex calculations and mappings onto 3D biostructure data and create publication-quality images and animations. A wide range of features, use cases, and related tools are showcased herein with the proteome of the influenza A H5N1 virus, a potential threat to global public health. Other biologically and medically important topics and themes may be similarly explored using [RCSB.org](https://www.rcsb.org) Mol*.

5.1 | Opportunities for future enhancement for Mol*

Displaying sophisticated, customized scenes using Mol* can be challenging, particularly for more complex visualizations. PDBe and RCSB PDB are co-developing the MolViewSpec standard, which allows users to define molecular visualizations and scenes in a declarative manner (Bittrich et al. 2024a). MolViewSpec provides a step-wise process for composing human-readable and human-editable JSON files that combine small, intuitive building blocks. These files can be read and displayed by [RCSB.org](https://www.rcsb.org) Mol*, consistently recreating previously defined scenes. Looking ahead, we aim to migrate aspects of [RCSB.org](https://www.rcsb.org) Mol* to MolViewSpec, which will make scene definition and customization more transparent and more easily reusable, enhancing accessibility and collaboration across research and educational communities. Additional ongoing improvements to the core Mol* library include support for mesoscale visualizations encompassing billions of atoms or entire cells (molstar.org/me) (Rose et al. 2024).

6 | MATERIALS AND METHODS

6.1 | [RCSB.org](https://www.rcsb.org) Mol* wrapper

A dedicated wrapper implements common Mol* usage patterns within the [RCSB.org](https://www.rcsb.org) web portal. This custom wrapper enables branding and supports a range of additional features relevant to our research-focused web portal. Related open-source software projects, including the rcsb-molstar wrapper, are enumerated in the “Accessibility Statement” section.

6.2 | Standalone Mol*

All Mol* deployments across [RCSB.org](https://www.rcsb.org), except Standalone Mol*, are linked to individual PDB IDs or CSMs,

or specific [RCSB.org](https://www.rcsb.org) tools. Standalone Mol* is accessible at [rcsb.org/3d-view](https://www.rcsb.org/3d-view), allowing the loading of multiple structures, including those from external sources. After loading the desired data, structure alignments can be computed directly in Mol* provided the biopolymer chains are similar in sequence.

6.3 | Predicting the location of the phospholipid bilayer

The predict membrane tool (Bittrich et al. 2022) supports the prediction of the membrane bilayer location. It is available for any PDB structure annotated as a membrane protein by UniProt and represented in one or more trusted external membrane protein data resources (Dobson et al. 2024; Lomize et al. 2012; Newport et al. 2019; White 2009). Membrane placement (Postic et al. 2016) is based on amino acid residue solvent exposure and optimized for the orientation and thickness of the membrane bilayer such that superficial hydrophobic amino acid sidechains fall within the phospholipid bilayer while hydrophilic residues are solvent accessible.

6.4 | Symmetry-related visualization

Visual display of symmetry axes detected in a macromolecular assembly can be accessed from SSPs or requested in Mol* directly using the Assembly Symmetry panel. Pre-calculated symmetry-related data is fetched by the [RCSB.org](https://www.rcsb.org) Data API at data.rcsb.org.

6.5 | Static images and data preparation for the [RCSB.org](https://www.rcsb.org) web portal

The Molrender Tool is used to generate static images in JPEG format. These images are displayed on SSPs for individual PDB structures or CSMs and serve as previews in search results. The core Mol* technology stack plays a crucial role in data preparation for weekly updates of the [RCSB.org](https://www.rcsb.org) web portal. Command-line tools are used to create optimized binary representations of PDBx/mmCIF source files in BinaryCIF format (Sehna et al. 2020). These tools also preprocess experimental density map data from MX and 3DEM studies, facilitating the efficient delivery of volumetric data.

6.6 | Viewing sequence annotations in 3D

The RCSB PDB Sequence Annotations in 3D Viewer (Segura et al. 2022) connect information about the 1D

sequence to locations of key residues in the 3D view and vice versa. Users can hover over or single-mouse-click on individual residues in the 3D view (right) to highlight the corresponding position in the 1D sequence view (left). Conversely, doing the same with a position or feature in the 1D sequence view highlights its location in the 3D view (Figure 3b).

6.7 | Pairwise structure alignments

3D structure alignments can be performed using the [RCSB.org](https://www.rcsb.org) Pairwise Structure Alignment Tool ([rcsb.org/alignment](https://www.rcsb.org/alignment)) (Bittrich et al. 2024b). This user-friendly feature supports alignment of up to nine chains versus a single reference chain with well-established alignment algorithms (i.e., TM-align (Zhang and Skolnick 2005) or jFATCAT, jCE, jCE-CP, Smith-Waterman 3D from the BioJava library (Lafita et al. 2019)). Results are presented interactively, allowing users to explore alignments at both 1D sequence and 3D structure levels. Sequence alignments are visualized using the Sequence Annotations in 3D Viewer (Segura et al. 2022), which summarizes structurally aligned regions, sequence-only matches, and unaligned regions or gaps. In parallel, Mol* provides a dynamic 3D structure view of aligned polypeptide chains. Alignments can be readily interrogated as selections made in the 1D sequence view are automatically reflected in the 3D structure view and vice versa.

6.8 | Comparing 1D sequence and 3D structure alignments

RCSB PDB Group Summary Pages (Segura et al. 2023) allow users to explore related entities (e.g., those with a common UniProt Accession Number). This web page presents an interactive side-by-side view, similar to the Pairwise Structure Alignment Tool (described below). Panels on the left of Figure 8c,e display the 1D sequence alignment of all corresponding entities, together with shared key UniProt annotations. Entities are sorted in descending order based on the extent of sequence coverage concerning the full-length polypeptide chain. Unresolved portions of polypeptide chains for experimentally determined structures in the PDB are indicated by gray sections of the horizontal track. CSM horizontal tracks are colored according to pLDDT confidence scores (Tunyasuvunakool et al. 2021). Initially, the first entity appearing in the alignment list is displayed in 3D (right). Additional entities in the 1D sequence alignment can be displayed in the 3D view by clicking on their desaturated identifier labels in the left-hand panel. For finer control over displayed entities, the three square boxes can be used to gain an overview of entity composition

and toggle 3D representations of the aligned entity, other polymeric chains, or bound ligands (as applicable). This approach to displaying available structure information allows users to quickly assess the relevance of individual PDB IDs or CSMs to the scientific question(s) at hand.

6.9 | Structure motif searching of 3D structures

The [RCSB.org](https://www.rcsb.org) web portal offers a dedicated Structure Motif Search Tool (Bittrich et al. 2020), which can rapidly identify similar 3D arrangements of selected amino acids in PDB structures or CSMs. This tool harnesses Mol* to streamline the construction of queries and interrogation of results as 3D structure alignments. Users can also define Structure Motif queries using the [RCSB.org](https://www.rcsb.org) Advanced Query Builder by specifying a reference structure and selecting from 2 to 10 residues by specifying chain identifiers and sequence positions. Because this process can be tedious and requires knowledge of exact sequence positions for residues of interest, [RCSB.org](https://www.rcsb.org) Mol* provides a dedicated user interface allowing users to define structure motifs interactively in 3D.

6.10 | Accessibility statement

The [RCSB.org](https://www.rcsb.org) web portal is freely accessible and does not require user registration. All software dependencies and the entire Mol* technology stack are shared via GitHub (github.com/molstar). General-purpose functionality is part of the core Mol* project (github.com/molstar/molstar, landing page: molstar.org, available as npm download: npmjs.com/package/molstar). Custom functionality specific to [RCSB.org](https://www.rcsb.org) is implemented in a dedicated [RCSB.org](https://www.rcsb.org) Mol* wrapper (github.com/molstar/rcsb-molstar, available as npm download: npmjs.com/package/@rcsb/rcsb-molstar). The Sequence Annotations Viewer software stack is available at: github.com/rcsb/rcsb-saguaro, github.com/rcsb/rcsb-saguaro-app, and github.com/rcsb/rcsb-saguaro-3d. The user interface of the Pairwise Structure Alignment Tool is shared at: github.com/rcsb/rcsb-pecos-app. Static images can be generated at scale using the Mol* Molrender Tool (github.com/molstar/molrender). Ongoing joint progress by PDBe and RCSB PDB on the MolViewSpec standard is also freely available (github.com/molstar/mol-view-spec, landing page: molstar.org/mol-view-spec, PyPI download: pypi.org/project/molviewspec). Mol* makes direct use of the RCSB PDB Data API (data.rcsb.org) (e.g., to retrieve data for the visualization of protein symmetry). Note: PDBe maintains its custom wrapper at github.com/molstar/pdbe-molstar.

AUTHOR CONTRIBUTIONS

Sebastian Bittrich: Writing – review and editing; writing – original draft; visualization; software. **Alexander S. Rose:** Conceptualization; project administration; visualization; software; supervision. **David Sehnal:** Supervision; project administration; conceptualization; funding acquisition; software; visualization. **Jose M. Duarte:** Conceptualization; project administration; software; visualization; supervision; writing – review and editing. **Yana Rose:** Conceptualization; visualization; project administration; supervision; software. **Joan Segura:** Supervision; software; visualization. **Dennis W. Piehl:** Visualization; writing – review and editing; software. **Brinda Vallat:** Software; visualization; writing – review and editing. **Chenghua Shao:** Software; validation; visualization; writing – review and editing. **Charmi Bhikadiya:** Software; visualization. **Jesse Liang:** Software; visualization. **Mark Ma:** Software; visualization. **David S. Goodsell:** Visualization; writing – original draft; writing – review and editing. **Stephen K. Burley:** Conceptualization; writing – original draft; writing – review and editing; funding acquisition; supervision. **Shuchismita Dutta:** Conceptualization; writing – original draft; writing – review and editing; supervision.

ACKNOWLEDGMENTS

RCSB PDB Core Operations are jointly funded by the U.S. National Science Foundation (NSF) (DBI-2321666, PI: S.K. Burley), the U.S. Department of Energy (DE-SC0019749, PI: S.K. Burley), and the National Cancer Institute, the National Institute of Allergy and Infectious Diseases, and the National Institute of General Medical Sciences of the National Institutes of Health (R01GM157729, PI: S.K. Burley). The US NSF and UK Biotechnology and Biological Sciences Research Council jointly supported the development of the Next Generation PDB archive (DBI-2019297, PI: S.K. Burley, BB/V004247/1, PI: S. Velankar) and new Mol* features (DBI-2129634, PI: S.K. Burley, BB/W017970/1, PI: S. Velankar). D.S. was supported by the Czech Science Foundation (22-30571M). Funding for open access charge: NIH (R01GM152779). The authors gratefully acknowledge contributions from all members of the RCSB PDB past and present and our wwPDB partners. The authors thank Dr. Sameer Velankar for helpful discussions and the acquisition of the co-funded NSF/BBSRC awards. Finally, we acknowledge and thank all existing and past contributors to the open-source Mol* project and welcome participation by new contributors.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Sebastian Bittrich  <https://orcid.org/0000-0003-3576-0387>

Alexander S. Rose  <https://orcid.org/0000-0002-0893-5551>

David Sehna  <https://orcid.org/0000-0002-0682-3089>

Jose M. Duarte  <https://orcid.org/0000-0002-9544-5621>

Yana Rose  <https://orcid.org/0000-0003-1018-5718>

Joan Segura  <https://orcid.org/0000-0001-5593-735X>

Dennis W. Piehl  <https://orcid.org/0000-0002-8315-2899>

Brinda Vallat  <https://orcid.org/0000-0002-4076-8984>

Chenghua Shao  <https://orcid.org/0000-0001-6817-7476>

Charmi Bhikadiya  <https://orcid.org/0000-0002-1587-5139>

David S. Goodsell  <https://orcid.org/0000-0002-5932-2130>

Stephen K. Burley  <https://orcid.org/0000-0002-2487-9713>

Shuchismita Dutta  <https://orcid.org/0000-0002-2109-5411>

REFERENCES

- Abbasi J. Bird flu outbreak in dairy cows is widespread, raising public health concerns. *JAMA*. 2024;331(21):1789–91.
- Antanasijevic A, Durst MA, Lavie A, Caffrey M. Identification of a pH sensor in Influenza hemagglutinin using X-ray crystallography. *J Struct Biol*. 2020;209(1):107412.
- Armstrong DR, Berrisford JM, Conroy MJ, Gutmanas A, Anyango S, Choudhary P, et al. PDBe: improved findability of macromolecular structure data in the PDB. *Nucleic Acids Res*. 2020;48(D1):D335–43.
- Baskaran K, Ploskon E, Tejero R, Yokochi M, Harrus D, Liang Y, et al. Restraint validation of biomolecular structures determined by NMR in the Protein Data Bank. *Structure*. 2024;32(6):824–837.e1.
- Bekker GJ, Yokochi M, Suzuki H, Ikegawa Y, Iwata T, Kudou T, et al. Protein Data Bank Japan: celebrating our 20th anniversary during a global pandemic as the Asian hub of three dimensional macromolecular structural data. *Protein Sci*. 2022;31(1):173–86.
- Berman HM, Henrick K, Nakamura H. Announcing the worldwide Protein Data Bank. *Nat Struct Mol Biol*. 2003;10(12):980.
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. *Nucleic Acids Res*. 2000;28(1):235–42.
- Bittrich S, Burley SK, Rose AS. Real-time structural motif searching in proteins using an inverted index strategy. *PLoS Comput Biol*. 2020;16(12):e1008502.
- Bittrich S, Midlik A, Varadi M, Velankar S, Burley SK, Young JY, et al. Describing and sharing molecular visualizations using the MolViewSpec Toolkit. *Curr Protoc*. 2024a;4(7):e1099.
- Bittrich S, Rose Y, Segura J, Lowe R, Westbrook J, Duarte JM, et al. RCSB Protein Data Bank: improved annotation, search, and visualization of membrane protein structures archived in the PDB. *Bioinformatics*. 2022;38:1452–4.
- Bittrich S, Segura J, Duarte JM, Burley SK, Rose Y. RCSB Protein Data Bank: exploring protein 3D similarities via comprehensive structural alignments. *Bioinformatics*. 2024b;40(6):btac370.
- Bradley AR, Rose AS, Pavelka A, Valasatava Y, Duarte JM, Pricl A, et al. MMTF—an efficient file format for the transmission, visualization, and analysis of macromolecular structures. *PLoS Comput Biol*. 2017;13(6):e1005575.
- Burley SK. Impact of structural biologists and the Protein Data Bank on small-molecule drug discovery and development. *J Biol Chem*. 2021;296:100559.
- Burley SK. Protein Data Bank: from two epidemics to the global pandemic to mRNA vaccines and Paxlovid. *Curr Opin Struct Biol*. 2025;90:102954.
- Burley SK, Bhikadiya C, Bi C, Biester A, Biswas P, et al. Updated resources for exploring experimental PDB structures and computed structure models at the RCSB Protein Data Bank. *Nucleic Acids Res*. 2025;53:D564–74.
- Burley SK, Bhikadiya C, Bi C, Bittrich S, Chao H, Chen L, et al. RCSB Protein Data Bank (RCSB.org): delivery of experimentally-determined PDB structures alongside one million computed structure models of proteins from artificial intelligence/machine learning. *Nucleic Acids Res*. 2023;51:D488–508.
- Burley SK, Wu-Wu A, Dutta S, Ganesan S, Zheng SXF. Impact of structural biology and the Protein Data Bank on US FDA new drug approvals of low molecular weight antineoplastic agents 2019–2023. *Oncogene*. 2024;43(29):2229–43.
- Cargnin Faccin F, Perez DR. Pandemic preparedness through vaccine development for avian influenza viruses. *Hum Vaccin Immunother*. 2024;20(1):2347019.
- Carter T, Iqbal M. The influenza A virus replication cycle: a comprehensive review. *Viruses*. 2024;16(2):316.
- Centers for Disease Control and Prevention. Technical report: June 2024 Highly Pathogenic Avian Influenza A(H5N1) Viruses. 2024. Available from: <https://www.cdc.gov/bird-flu/php/technical-report/h5n1-06052024.html>
- Centers for Disease Control and Prevention and National Center for Immunization and Respiratory Diseases (NCIRD). Types of Influenza viruses. 2023. Available from: <https://www.cdc.gov/flu/about/viruses/types.htm>
- Centers for Disease Control and Prevention and National Center for Immunization and Respiratory Diseases (NCIRD). Trivalent Influenza vaccines. 2024. Available from: <https://www.cdc.gov/flu/prevent/trivalent.htm>
- Chavas LMG, Kato R, McKimm-Breschkin J, Colman PM, Fusi P, Tringali C, et al. Crystal structure of the human Sialidase Neu2 in complex with Zanamivir inhibitor. 2005. Available from: <https://doi.org/10.2210/pdb2212f2210z/pdb>
- Collins PJ, Haire LF, Lin YP, Liu J, Russell RJ, Walker PA, et al. Crystal structures of oseltamivir-resistant influenza virus neuraminidase mutants. *Nature*. 2008;453(7199):1258–61.
- Corbett KS, Edwards DK, Leist SR, Abiona OM, Boyoglu-Barnum S, Gillespie RA, et al. SARS-CoV-2 mRNA vaccine design enabled by prototype pathogen preparedness. *Nature*. 2020;586(7830):567–71.
- DeLano WL. The PyMOL molecular graphics system. 2002. Available from: <https://pymol.org/>
- Dobson L, Gerdan C, Tusnady S, Szekeres L, Kuffa K, Lango T, et al. UniTmP: unified resources for transmembrane proteins. *Nucleic Acids Res*. 2024;52(D1):D572–8.
- Dreyfus C, Laursen NS, Kwaks T, Zuijdgheest D, Khayat R, Ekiert DC, et al. Highly conserved protective epitopes on influenza B viruses. *Science*. 2012;337(6100):1343–8.
- Du R, Cui Q, Chen Z, Zhao X, Lin X, Rong L. Revisiting influenza A virus life cycle from a perspective of genome balance. *Virol Sin*. 2023;38(1):1–8.
- Duarte JM, Dutta S, Goodsell DS, Burley SK. Exploring protein symmetry at the RCSB Protein Data Bank. *Emerg Top Life Sci*. 2022;6(3):231–43.
- Dutta S, Dimitropoulos D, Feng Z, Persikova I, Sen S, Shao C, et al. Improving the representation of peptide-like inhibitor and antibiotic molecules in the Protein Data Bank. *Biopolymers*. 2014;101(6):659–68.

- Feng Z, Westbrook JD, Sala R, Smart OS, Bricogne G, Matsubara M, et al. Enhanced validation of small-molecule ligands and carbohydrates in the protein databank. *Structure*. 2021;29:393–400.e1.
- Furey C, Scher G, Ye N, Kercher L, DeBeauchamp J, Crumpton JC, et al. Development of a nucleoside-modified mRNA vaccine against clade 2.3.4.4b H5 highly pathogenic avian influenza virus. *Nat Commun*. 2024;15(1):4350.
- Fuyuno I. Tamiflu side effects come under scrutiny. *Nature*. 2007;446(7134):358–9.
- Garg S, Reed C, Davis CT, Uyeki TM, Behraves CB, Kniss K, et al. Outbreak of highly pathogenic avian influenza A (H5N1) viruses in U.S. dairy cattle and detection of two human cases - United States, 2024. *MMWR Morb Mortal Wkly Rep*. 2024;73(21):501–5.
- Goodsell DS. Influenza virus. 2024. Available from: https://doi.org/10.2210/rcsb_pdb/goodsell-gallery-049
- Goodsell DS, Burley SK. RCSB Protein Data Bank resources for structure-facilitated design of mRNA vaccines for existing and emerging viral pathogens. *Structure*. 2022;30:252–262.e4.
- Gore S, Sanz Garcia E, Hendrickx PMS, Gutmanas A, Westbrook JD, Yang H, et al. Validation of structures in the Protein Data Bank. *Structure*. 2017;25(12):1916–27.
- Hanson RM, Prilusky J, Renjian Z, Nakane T, Sussman JL. JSmol and the next-generation web-based representation of 3D molecular structure as applied to Proteopedia. *Isr J Chem*. 2013;53(3–4):207–16.
- Hata K, Koseki K, Yamaguchi K, Moriya S, Suzuki Y, Yingsakmongkon S, et al. Limited inhibitory effects of oseltamivir and zanamivir on human sialidases. *Antimicrob Agents Chemother*. 2008;52(10):3484–91.
- Hoch JC, Baskaran K, Burr H, Chin J, Eghbalnia HR, Fujiwara T, et al. Biological magnetic resonance data bank. *Nucleic Acids Res*. 2023;51(D1):D368–76.
- Humphrey W, Dalke A, Schulten K. VMD: visual molecular dynamics. *J Mol Graph*. 1996;14(1):33–8.
- Humphreys IR, Pei J, Baek M, Krishnakumar A, Anishchenko I, Ovchinnikov S, et al. Computed structures of core eukaryotic protein complexes. *Science*. 2021;374(6573):eabm4805.
- Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596(7873):583–9.
- Lafita A, Bliven S, Pric A, Guzenko D, Rose PW, Bradley A, et al. BioJava 5: a community driven open-source bioinformatics library. *PLoS Comput Biol*. 2019;15(2):e1006791.
- Li H, Li Q, Li B, Guo Y, Xing J, Xu Q, et al. Continuous reassortment of clade 2.3.4.4 H5N6 highly pathogenic avian influenza viruses demonstrating high risk to public health. *Pathogens*. 2020;9(8):670.
- Li H, Wei X. A concise review of biomolecule visualization. *Curr Issues Mol Biol*. 2024;46(2):1318–34.
- Li H, Wu Y, Li M, Guo L, Gao Y, Wang Q, et al. An intermediate state allows influenza polymerase to switch smoothly between transcription and replication cycles. *Nat Struct Mol Biol*. 2023;30(8):1183–92.
- Lin T-H, Zhu X, Wang S, Zhang D, McBride R, Yu W, et al. A single mutation in bovine influenza H5N1 hemagglutinin switches specificity to human receptors. *Science*. 2024;386(6726):1128–34.
- Lomize MA, Pogozheva ID, Joo H, Mosberg HI, Lomize AL. OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res*. 2012;40:D370–6.
- Maxwell SR. Tamiflu and neuropsychiatric disturbance in adolescents. *BMJ*. 2007;334(7606):1232–3.
- McIntire KM, Meng H, Lin TH, Kim W, Moore NE, Han J, et al. Maturation of germinal center B cells after influenza virus vaccination in humans. *J Exp Med*. 2024;221(8):e20240668. <https://doi.org/10.1084/jem.20240668>
- Meng EC, Goddard TD, Pettersen EF, Couch GS, Pearson ZJ, Morris JH, et al. UCSF ChimeraX: tools for structure building and analysis. *Protein Sci*. 2023;32(11):e4792.
- Neelamegham S, Aoki-Kinoshita K, Bolton E, Frank M, Lisacek F, Lutteke T, et al. Updates to the symbol nomenclature for Glycans guidelines. *Glycobiology*. 2019;29(9):620–4.
- Newport TD, Sansom MSP, Stansfeld PJ. The MemProtMD database: a resource for membrane-embedded protein structures and their lipid interactions. *Nucleic Acids Res*. 2019;47(D1):D390–7.
- Ng AK, Zhang H, Tan K, Li Z, Liu JH, Chan PK, et al. Structure of the influenza virus A H5N1 nucleoprotein: implications for RNA binding, oligomerization, and vaccine design. *FASEB J*. 2008;22(10):3638–47.
- Nichol KL, Treanor JJ. Vaccines for seasonal and pandemic influenza. *J Infect Dis*. 2006;194(Suppl 2):S111–8.
- Omoto S, Speranzini V, Hashimoto T, Noshi T, Yamaguchi H, Kawai M, et al. Characterization of influenza virus variants induced by treatment with the endonuclease inhibitor baloxavir marboxil. *Sci Rep*. 2018;8(1):9633.
- Ozawa M, Fujii K, Muramoto Y, Yamada S, Yamayoshi S, Takada A, et al. Contributions of two nuclear localization signals of influenza A virus nucleoprotein to viral replication. *J Virol*. 2007;81(1):30–41.
- Pielak RM, Chou JJ. Influenza M2 proton channels. *Biochim Biophys Acta*. 2011;1808(2):522–9.
- Pinto LH, Lamb RA. The M2 proton channels of influenza A and B viruses. *J Biol Chem*. 2006;281(14):8997–9000.
- Plotch SJ, Bouloy M, Ulmanen I, Krug RM. A unique cap(m7GpppXm)-dependent influenza virion endonuclease cleaves capped RNAs to generate the primers that initiate viral RNA transcription. *Cell*. 1981;23(3):847–58.
- Postic G, Ghouzam Y, Guiraud V, Gelly JC. Membrane positioning for high- and low-resolution protein structures through a binary classification approach. *Protein Eng Des Sel*. 2016;29(3):87–91.
- Potter C. A history of influenza. *J Appl Microbiol*. 2001;91:572–9.
- Protein Data Bank. Crystallography: Protein Data Bank. *Nature New Biol*. 1971;233(42):223.
- Rose AS, Hildebrand PW. NGL viewer: a web application for molecular visualization. *Nucleic Acids Res*. 2015;43(W1):W576–9.
- Rose AS, Sehnal D, Goodsell DS, Autin L. Mesoscale explorer—visual exploration of large-scale molecular models. *Protein Sci*. 2024;33(10):e5177.
- Russell RJ, Haire LF, Stevens DJ, Collins PJ, Lin YP, Blackburn GM, et al. The structure of H5N1 avian influenza neuraminidase suggests new opportunities for drug design. *Nature*. 2006;443(7107):45–9.
- Segura J, Rose Y, Bi C, Duarte J, Burley SK, Bittrich S. RCSB Protein Data Bank: visualizing groups of experimentally determined PDB structures alongside computed structure models of proteins. *Front Bioinform*. 2023;3:1311287.
- Segura J, Rose Y, Bittrich S, Burley SK, Duarte JM. RCSB Protein Data Bank 1D3D module: displaying positional features on macromolecular assemblies. *Bioinformatics*. 2022;38:3304–5.
- Segura J, Rose Y, Westbrook J, Burley SK, Duarte JM. RCSB Protein Data Bank 1D tools and services. *Bioinformatics*. 2020;36(22–23):5526–7.
- Sehnal D, Bittrich S, Deshpande M, Svobodova R, Berka K, Bazgier V, et al. Mol* Viewer: modern web app for 3D visualization and analysis of large biomolecular structures. *Nucleic Acids Res*. 2021;49:W431–7.
- Sehnal D, Bittrich S, Velankar S, Koca J, Svobodova R, Burley SK, et al. BinaryCIF and CIFTTools-Lightweight, efficient and extensible macromolecular data management. *PLoS Comput Biol*. 2020;16(10):e1008247.
- Sehnal D, Deshpande M, Varekova RS, Mir S, Berka K, Midlik A, et al. LiteMol suite: interactive web-based visualization of large-

- scale macromolecular structure data. *Nat Methods*. 2017; 14(12):1121–2.
- Sehna D, Rose A, Koca J, Burley S, Velankar S. Mol*: towards a common library and tools for web molecular graphics. *Proceedings of the workshop on molecular graphics and visual analysis of molecular data*. Germany: ACM; 2018. p. 29–33.
- Selzer L, Su Z, Pintilie GD, Chiu W, Kirkegaard K. Full-length three-dimensional structure of the influenza A virus M1 protein and its organization into a matrix layer. *PLoS Biol*. 2020;18(9):e3000827.
- Sen S, Young J, Berrisford JM, Chen M, Conroy MJ, Dutta S, et al. Small molecule annotation for the Protein Data Bank. *Database (Oxford)*. 2014;2014:bau116.
- Shao C, Bittrich S, Wang S, Burley SK. Assessing PDB macromolecular crystal structure confidence at the individual amino acid residue level. *Structure*. 2022;30:1385–94.
- Shao C, Feng Z, Westbrook JD, Peisach E, Berrisford J, Ikegawa Y, et al. Modernized uniform representation of carbohydrate molecules in the Protein Data Bank. *Glycobiology*. 2021;31:1204–18.
- Shinya K, Kawaoka Y. Influenza virus receptors in the human airway. *Virus*. 2006;56(1):85–9.
- Skehel JJ, Wiley DC. Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. *Annu Rev Biochem*. 2000;69:531–69.
- Smart OS, Horsky V, Gore S, Svobodova Varekova R, Bendova V, Kleywegt GJ, et al. Validation of ligands in macromolecular structures determined by X-ray crystallography. *Acta Crystallogr D Biol Crystallogr*. 2018;74(Pt 3):228–36.
- Staller E, Carrique L, Swann OC, Fan H, Keown JR, Sheppard CM, et al. Structures of H5N1 influenza polymerase with ANP32B reveal mechanisms of genome replication and host adaptation. *Nat Commun*. 2024;15(1):4123.
- Staller E, Sheppard CM, Neasham PJ, Mistry B, Peacock TP, Goldhill DH, et al. ANP32 proteins are essential for influenza virus replication in human cells. *J Virol*. 2019;93(17):e00217–00219.
- Stevens J, Blixt O, Tumpey TM, Taubenberger JK, Paulson JC, Wilson IA. Structure and receptor specificity of the hemagglutinin from an H5N1 influenza virus. *Science*. 2006;312(5772):404–10.
- Stouffer AL, Acharya R, Salom D, Levine AS, Di Costanzo L, Soto CS, et al. Structural basis for the function and inhibition of an influenza virus proton channel. *Nature*. 2008;451(7178):596–9.
- Sui J, Hwang WC, Perez S, Wei G, Aird D, Chen LM, et al. Structural and functional bases for broad-spectrum neutralization of avian and human influenza A viruses. *Nat Struct Mol Biol*. 2009;16(3):265–73.
- Tang YS, Xu S, Chen YW, Wang JH, Shaw PC. Crystal structures of influenza nucleoprotein complexed with nucleic acid provide insights into the mechanism of RNA interaction. *Nucleic Acids Res*. 2021;49(7):4144–54.
- Taubenberger JK, Morens DM. Pandemic influenza—including a risk assessment of H5N1. *Rev Sci Tech*. 2009;28(1):187–202.
- Taubenberger JK, Morens DM. Influenza: the once and future pandemic. *Public Health Rep*. 2010;125(Suppl 3):16–26.
- Thomaston JL, Konstantinidi A, Liu L, Lambrinidis G, Tan J, Caffrey M, et al. X-ray crystal structures of the influenza M2 proton channel drug-resistant V27A mutant bound to a spiroadamantyl amine inhibitor reveal the mechanism of adamantane resistance. *Biochemistry*. 2020;59(4):627–34.
- Tunyasuvunakool K, Adler J, Wu Z, Green T, Zielinski M, Zidek A, et al. Highly accurate protein structure prediction for the human proteome. *Nature*. 2021;596(7873):590–6.
- UniProt Consortium. UniProt: the universal protein knowledgebase in 2023. *Nucleic Acids Res*. 2023;51:D523–31.
- Varadi M, Bertoni D, Magana P, Paramval U, Piduchna I, Radhakrishnan M, et al. AlphaFold protein structure database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res*. 2024;52(D1):D368–75.
- Westbrook JD, Burley SK. How structural biologists and the Protein Data Bank contributed to recent FDA new drug approvals. *Structure*. 2019;27:211–7.
- Westbrook JD, Soskind R, Hudson BP, Burley SK. Impact of Protein Data Bank on antineoplastic approvals. *Drug Discov Today*. 2020;25:837–50.
- White SH. Biophysical dissection of membrane proteins. *Nature*. 2009;459(7245):344–6.
- World Health Organization. Influenza (Avian and other zoonotic). 2023. Available from: [https://www.who.int/news-room/fact-sheets/detail/influenza-\(avian-and-other-zoonotic\)](https://www.who.int/news-room/fact-sheets/detail/influenza-(avian-and-other-zoonotic))
- wwPDB consortium. Protein Data Bank: the single global archive for 3D macromolecular structure data. *Nucleic Acids Res*. 2019; 47(D1):D520–8.
- wwPDB Consortium. EMDB: the electron microscopy data bank. *Nucleic Acids Res*. 2023;52:D456–65.
- Xiong X, Coombs PJ, Martin SR, Liu J, Xiao H, McCauley JW, et al. Receptor binding by a ferret-transmissible H5 avian influenza virus. *Nature*. 2013;497(7449):392–6.
- Xu W, Velankar S, Patwardhan A, Hoch JC, Burley SK, Kurisu G. Announcing the launch of Protein Data Bank China as an associate member of the worldwide protein data Bank Partnership. *Acta Crystallogr D Struct Biol*. 2023;79:792–5.
- Young JY, Berrisford J, Chen M. wwPDB biocuration: on the front line of structural biology. *Nat Methods*. 2021;18(5):431–2.
- Young JY, Westbrook JD, Feng Z, Sala R, Peisach E, Oldfield TJ, et al. OneDep: unified wwPDB system for deposition, biocuration, and validation of macromolecular structures in the PDB archive. *Structure*. 2017;25(3):536–45.
- Zhang Y, Skolnick J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res*. 2005;33(7):2302–9.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Bittrich S, Rose AS, Sehna D, Duarte JM, Rose Y, Segura J, et al. Visualizing and analyzing 3D biomolecular structures using Mol* at RCSB.org: Influenza A H5N1 virus proteome case study. *Protein Science*. 2025;34(4):e70093. <https://doi.org/10.1002/pro.70093>