Mesoscale All-Atom Influenza Virus Simulations Suggest New Substrate Binding Mechanism

Authors: Jacob Durrant[1,†], Sarah Kochanek[2,†], Lorenzo Casalino[2], Pek U Ieong[2], Abigail Dommer[2], Rommie E. Amaro[2*]

Affiliations:

[1]Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260

[2]Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, CA 92093-0340

† These authors contributed equally.

*Correspondence to: ramaro@ucsd.edu

Abstract

Influenza virus circulates in human, avian, and swine hosts, causing seasonal epidemic and occasional pandemic outbreaks. Influenza neuraminidase, a viral surface glycoprotein, has two sialic acid binding sites. The catalytic (primary) site, which also binds inhibitors such as oseltamivir carboxylate, is responsible for cleaving the sialic acid linkages that bind viral progeny to the host cell. In contrast, the functional annotation of the secondary site remains unclear. Here, we better characterize these two sites through the development of an all-atom, explicitly solvated, experimentally based integrative model of the pandemic influenza A H1N1 2009 viral envelope, containing ~160 million atoms and spanning ~115 nm in diameter. Molecular dynamics simulations of this crowded subcellular environment, coupled with Markov state model theory, provide a novel framework for studying realistic molecular systems at the mesoscale and allow us to quantify the kinetics of the 150-loop transition between the open and closed states. An analysis of chloride ion occupancy along the neuraminidase surface implies a potential new role for the neuraminidase secondary site, wherein the terminal sialic acid residues of the linkages may bind before transfer to the primary site where enzymatic cleavage occurs. Altogether, our work breaks

new ground for molecular simulation in terms of the size, complexity, and methodological analyses of the simulated components, as well as provides fundamental insights into the understanding of substrate recognition processes for this vital influenza drug target, suggesting a new strategy for the development of anti-influenza therapeutics.

Main

Influenza virus infection is responsible for millions of deaths worldwide each year. The Center for Disease Control estimates that pandemic influenza A H1N1 2009 (pH1N1) affected 60.8 million people, resulting in 12,468 casualties in the United States alone[1,2]. Along with others, this strain dramatically contributes to yearly epidemics, continuously fueling concerns about the emergence of a new pandemic strain. In addition, the increasingly widespread resistance to antiviral medications is compounding this threat[3], thus requiring the development of novel approaches for the prevention and treatment of influenza virus infection. One such strategy is to target the viral surface glycoprotein neuraminidase (NA), which promotes viral progeny release from the host cell by cleaving terminal sialic acid residues[4–6]. Previous work has identified the importance of characterizing the dynamics of the NA catalytic site for drug design[7–12], understanding mechanisms of antiviral resistance[13], and deciphering the mechanisms underlying substrate binding[14–18].

The catalytic (primary, 1°) site of NA is highly flexible, in part due to the adjacent 150- and 430-loops (residues 147–152 and 429–433, respectively, N2 numbering)[11,14,19]. The significance of this flexibility is highlighted by structural comparison of the phylogenetically distinct group-1 (N1, N4, N5, and N8) and group-2 (N2, N3, N6, N7, and N9) NAs, which illustrates that opening of the 150-loop in the group-1 structures leads to the formation of the so-called 150-cavity[12] that can bind compounds with increased specificity and potency[10]. However, crystal structures of pH1N1 NA (pN1) reveal that, unlike all other group-1 NAs, its 150-loop is closed and therefore no 150-cavity is present[20]. In contrast, previous investigations utilizing MD simulations have found that the 150-loop of pN1 is in the open state ~60-65% of the time[13,19,21].

NA also contains a secondary (2°) sialic acid binding site adjacent to the catalytic site. This site was first identified as a hemadsorption site in avian-origin influenza NAs[22–26] and was not initially believed to be present in swine-origin strains due to non-conservation of critical residues at this site[24,27]. However, more recent studies provide support for the presence of a 2° site in swine-origin influenza NAs, including pN1[16,17]. The precise mechanism by which this 2° site functions remains unclear, however a number of studies have demonstrated its role in receptor binding[28–32] and catalytic efficiency[28,29]. In addition, previous Brownian dynamics (BD) simulations of single glycoproteins and various ligands suggested that both endogenous substrates and the drug oseltamivir carboxylate bind faster to the 2° site than the 1° site (*i.e.*, the k_{on} rate is 2- and 7-fold higher for the N1 and N2 2° site, respectively, vs. the corresponding 1° site)[17]. Finally, the 2° site has recently been identified as a target for a novel influenza virus inhibitor[33,34], further highlighting the need to understand its role in viral infectivity.

To study the 1° and 2° sites in the context of the viral surface, we used integrative modeling to construct a fully atomistic model of the pH1N1 viral envelope (**Figure 1A**, and **Figure S1** of Supporting Information, SI). The model was built using high-resolution crystallographic structures of individual glycoproteins (~1.9 – 2.6 Å resolution)[20,35] that were spatially positioned according to a lower-resolution cryo-electron tomography (cryo-ET, ~16-20 Å resolution) map of a viral particle[36]. Our viral envelope construct includes 30 NA tetramers (120 monomers) and 236 hemagglutinin (HA) trimers (708 monomers) embedded in a phospholipid bilayer, with realistic density and patterning taken from the cryo-ET. The entire pH1N1 all-atom system modeled here amounts to ~160 million atoms (fully solvated) and is ~115 nm in diameter. As such, it is among the largest biophysical systems yet studied with all-atom molecular dynamics (MD)[37,38]. A complete description of the integrative modeling and computational approaches used to build the virion envelope is provided in the SI (section 1.3). Additional details can also be found in Amaro et al.[39]

Over the past decade, studies of viruses at the molecular and coarse-grained (CG) levels have given unique insights into these systems, complementing and extending available experimental data by providing highly detailed models at never-before-seen scales, as well as suggesting testable biological hypotheses (predictions)[40,41]. Work by Schulten and coworkers established the first explicitly solvated atomic MD simulation of an intact virus, the satellite tobacco mosaic virus (~17 nm diameter, ~1 million atoms, 50 ns dynamics), in 2006[42]. Zink and Grubmuller in 2009 used steered MD to explore the dynamics of the explicitly solvated icosahedral shell of the southern bean mosaic virus (~36 nm diameter, ~4.5 million atoms, 100 ns)[43]. In 2010, Ayton and Voth developed and simulated an implicitly solvated CG representation of the immature HIV-1 virion (~125 nm diameter, 280,000 CG particles)[44]. In 2012, Larsson and coworkers simulated with explicitly solvated all-atom MD the satellite tobacco necrosis virus (~17 nm diameter, ~1.2 million atoms, ~1 us)[44], and Roberts et al. developed a fully atomic poliovirus (~30 nm diameter, ~2.8-4 million atoms, 50 ns)[45]. In 2013, Schulten and coworkers built and simulated a fully atomic representation of the HIV capsid (~70 nm diameter, ~64 million atoms, ~100 ns)[46], and Andoh et al. in 2014 simulated an all-atom poliovirus capsid (~30 nm diameter, ~6.5 million atoms, ~200 ns)[47]. Sansom and colleagues in 2015 reported an explicitly solvated CG influenza virus simulation (~80 nm diameter, 5 million particles)[48]. In two separate studies in 2016, Reddy and Sansom[49], and Bond, Verma and coworkers[50], reported coarse grained simulations of the Dengue viral membrane (~50 nm diameter, ~1 million particles). Aside from the rich structural, dynamical, and biophysical insights that these studies each provided, the investigations have collectively pushed the capabilities of molecular simulation, often relying on the world's fastest and most advanced computing architectures.

As the first explicitly solvated atomic-scale simulation of a viral lipid envelope (~115 nm diameter, ~160 million particles, ~121 ns), the work reported here breaks new ground in molecular simulation. To further characterize the structural dynamics of the viral envelope and its glycoproteins, we combined our mesoscale all-atom MD simulations with Markov state model (MSM) theory [51–53], thus enabling the extraction of long-timescale (*e.g.*, microseconds) individual glycoprotein dynamics in a crowded environment from the short timescale MD (*e.g.*, nanoseconds) of the fully intact viral surface. The accuracy and utility of MSMs have been demonstrated by experimental validation for many use cases including protein-protein binding, small-molecule binding kinetics, and protein-folding rate prediction [54–56]. Correspondingly, the approach reported here, which relies on the many copies of single glycoproteins present within a biologically accurate environment, provides a novel methodological advance for extracting long time-scale dynamics from short simulations through the powerful MSM theoretical framework at subcellular and cellular scales.

Here, we quantitatively compare how calculated protein dynamics differ when simulating many proteins in a single subcellular environment vs. simulating single proteins in isolation. By exploiting the whole pH1N1 viral envelope treated entirely with atomic resolution, this study provides unique insights into the two sialic acid binding sites of NA (*e.g.*, 1° and 2°). Our mesoscale atomic simulations suggest that the NA 1° site is even more flexible than previously appreciated and provide the first rigorous kinetic characterization of the 150-loop dynamics.

Furthermore, our work suggests that the 2° site, which is more solvent exposed and, in some strains, has a higher k_{on} rate than the 1° site[17,39], may be responsible for initially capturing sialic acid residues, which are then electrostatically guided to the 1° site for enzymatic cleavage. Within this context, our mesoscale simulations unveil an unprecedented cooperative interplay between the two sites that further illuminates the process of sialic acid/oseltamivir carboxylate recognition and the 2° site functional annotation. This fundamental discovery may be used as a rationale for the development of novel anti-influenza small molecule therapeutics targeting NA.

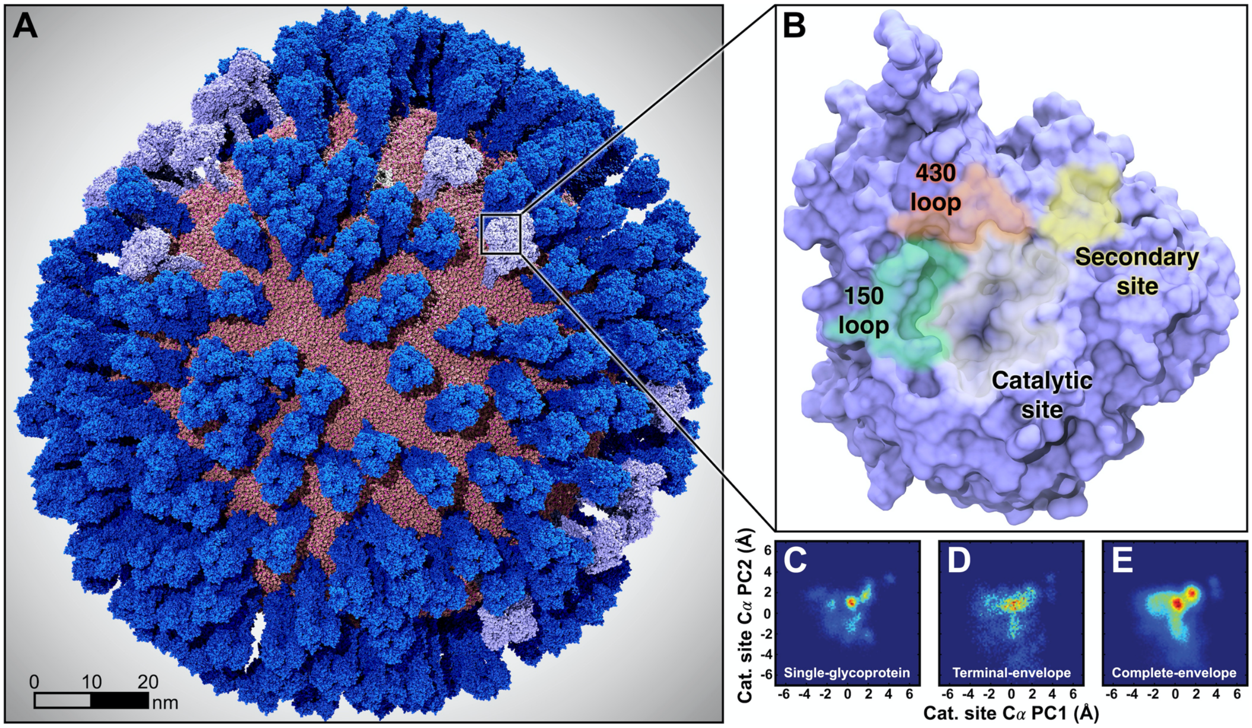

Figure 1. Mesoscale simulations enhance conformational sampling of the viral glycoproteins. **(A)** A fully intact all-atom model of the influenza A H1N1 2009 (pH1N1) virion envelope, containing over 160 million atoms, shown without explicit water molecules, was simulated with all-atom molecular dynamics simulations. Hemagglutinin (HA) glycoproteins shown in royal (dark) blue, neuraminidase (NA) glycoproteins shown in ice (light) blue. **(B)** Top view of a single NA monomer in surface representation with the catalytic site (white), secondary site (yellow), 150-loop (red) and 430-loop (green) highlighted. **(C-E)** Principal Component Analysis (PCA) was performed by considering the motions of the C_α atoms of 19 1°-pocket residues. PCA histograms were independently normalized so the bins containing the minimum and maximum number of points were blue and red, respectively. **(C)** PCA analysis of the four monomers sampled during a single-NA-tetramer simulation (“single-glycoprotein"). **(D)** PCA analysis of the 120 monomeric trajectories extracted during the last 8.33 ns of the viral envelope simulation (“terminal-envelope”). **(E)** PCA analysis of all 120 monomeric trajectories extracted from the full simulation of the viral envelope (“complete-envelope”).

Influenza Virus All-Atom Simulations

All-atom MD simulations of the pH1N1 viral envelope were performed using NAMD2.10[57] and CHARMM36 all-atom additive force fields[58]. The system was fully solvated with explicit water molecules (TIP3P force field[59]), while ions were described using Beglov and Roux force fields[60]. To broaden conformational sampling and more efficiently use supercomputer resources, the initial simulation was forked twice, generating two additional shorter daughter simulations (schematic representation in **Figure S2)**. Taken together, these simulations achieved a comprehensive simulation time of ~121 ns. The complete viral envelope simulation included 30 NA tetramers, yielding 14.5 μs of monomeric simulation (121 ns × 30 tetramers × 4 monomers/tetramer), and 236 HA tetramers, accounting for 85.6 μs of monomeric (121 ns × 236 tetramers × 3 monomers/trimer). Each glycoprotein structure used to build the initial viral envelope system was taken from fully equilibrated single-glycoprotein MD simulations (see SI, sections 1.1-1.2, for computational details relative to these sets of simulations). The viral envelope simulations were run on the Blue Waters petascale supercomputer, using 114,688 processors, equivalent to 16,384 Blue Waters nodes or 4,096 physical nodes. The simulation averaged 25.57 steps/sec. Frames were written every 10,000 steps (20 ps), ultimately occupying 11.66 terabytes of disk space. Data analysis drew upon conformations extracted at equally-spaced timepoints from these trajectories. The adopted MD protocol for the viral envelope simulations is fully described in the SI (section 1.4), including the NAMD input file. The physical properties of the virus (RMSD, RMSF analyses) and its lipid bilayer (curvature, motions) per the simulations are reported in the SI (sections 1.5-1.6, and **Figures S3** to **S9**).

To explore the flexibility of the 1° pocket (shown in **Figure S10**), we concatenated the MD trajectories of all 120 NA monomers and calculated the principal components of 19 catalytic-pocket-lining NA residues by considering their C_α atoms (heatmap, **Figures 1C-E**). We selected these 19 residues because they are homologous to those within 5 Å of the crystallographic oseltamivir carboxylate from the 2HU4 structure[12]. Principal component analysis (PCA) details, including indices of the catalytic and active site residues used in the analysis, are provided in the SI (section 1.7).

To judge whether mesoscale simulations enhance conformational sampling, we compared the viral envelope full simulation (referred to as "complete-envelope", **Figure 1E**) to five long time scale simulations of isolated NA tetramers embedded in small lipid-bilayer patches, described in a previous work ("single-glycoprotein", **Figure 1C**)[61]. A fair comparison requires that the sampling of the two systems be consistent in terms of the actual simulation length. The viral envelope full simulation sampled 14.5 μs of monomeric dynamics, but the five simulations of isolated NA tetramers sampled only 1.0 μs of monomeric dynamics (5 simulations × 50 ns/simulation × 1 NA tetramer × 4 monomers/tetramer). Thus, to improve comparison, we considered only the final 8.33 ns of the viral envelope simulation, which is equivalent to 1.0 μs of monomeric dynamics (1 simulation × 8.33 ns/simulation × 30 NA tetramer × 4 monomers/tetramer) (**Figure 1D**). We refer to this truncated segment of the full simulation as the "terminal-envelope" simulation. In all the cases, the motions of the C_α atoms of the same 19 residues were projected onto the first two principal components of the viral envelope NA trajectories, and the resulting heatmaps were compared (**Figures 1C-E**). Strikingly, PCA of the NA catalytic site residues indicates that the viral envelope simulation more thoroughly explored the conformational landscape, even after controlling for total simulation time.

To better study the 1° site conformations sampled by the viral envelope simulation, we applied *k*-means clustering to the PCA points of **Figure 1E**. Visual inspection of cluster centroids corresponding to four representative 1° site conformations (shown in **Figure S10**) reveals that R292 and R371, two key residues known to interact with the sialic acid carboxylate group, are the most flexible. In contrast, the carboxylate-stabilizing R152 residue moves outward in only one of the four representative conformations. Other pocket-lining residues such as R118 and D151, which previous works suggest may play a role in the molecular mechanisms of oseltamivir resistance[13], are also relatively flexible in the apo state.

The PCA analysis demonstrates that the viral envelope simulation more thoroughly sampled distinct 1° pocket states (**Figures 1E**) that are scarcely populated in the single-glycoprotein (**Figure 1C**). This holds true even when comparing the terminal-envelope (**Figure 1D**) simulation to the single-glycoprotein simulations (where the total sampling time is equal). The enhanced conformational sampling may simply be a product of the large number of NA copies blanketed across the viral surface; however, we do expect some effects from the viral surface environment including long-range electrostatic forces and glycoprotein-glycoprotein interactions that only the viral envelope simulation can capture.

pN1 Catalytic Site Structure and Dynamics

To explore the dynamics of the catalytic site, we analyzed the volumes of the 1° pocket and adjacent regions over the course of the entire viral envelope simulation (120 NA monomers, **Figure 2**)[62]. The volumes ranged from 450 to 4440 Å^3, with an average of 1536 Å^3 (**Table S1** and section 1.8 in the SI**)**. By comparison, the starting crystal structure pN1 (PDB ID: 3NSS[20]) with a closed 150-cavity has a volume of 800 Å^3, and a structure of a non-pandemic N1 (PDB ID: 2HTY[12]) with an open 150-cavity has a volume of 1088 Å^3. This indicates that the volume and depth of the catalytic pocket and adjacent regions can increase remarkably over what has been observed in crystal structures (**Figure 2C**). Contributing to this additional cavity volume and depth are two novel sub-pockets near residues G351 and E227, buried deep inside – but contiguous with – the 1° site (**Figure 2B**). Analysis with FTMap, a server for mapping ligand binding hot spots in macromolecules[63], suggests that the G351 sub-pocket can accommodate small-molecule ligands. Similar to the 150-cavity and 430-cavity, the G351- and E227-adjacent sub-pockets may provide new ligand-binding opportunities.

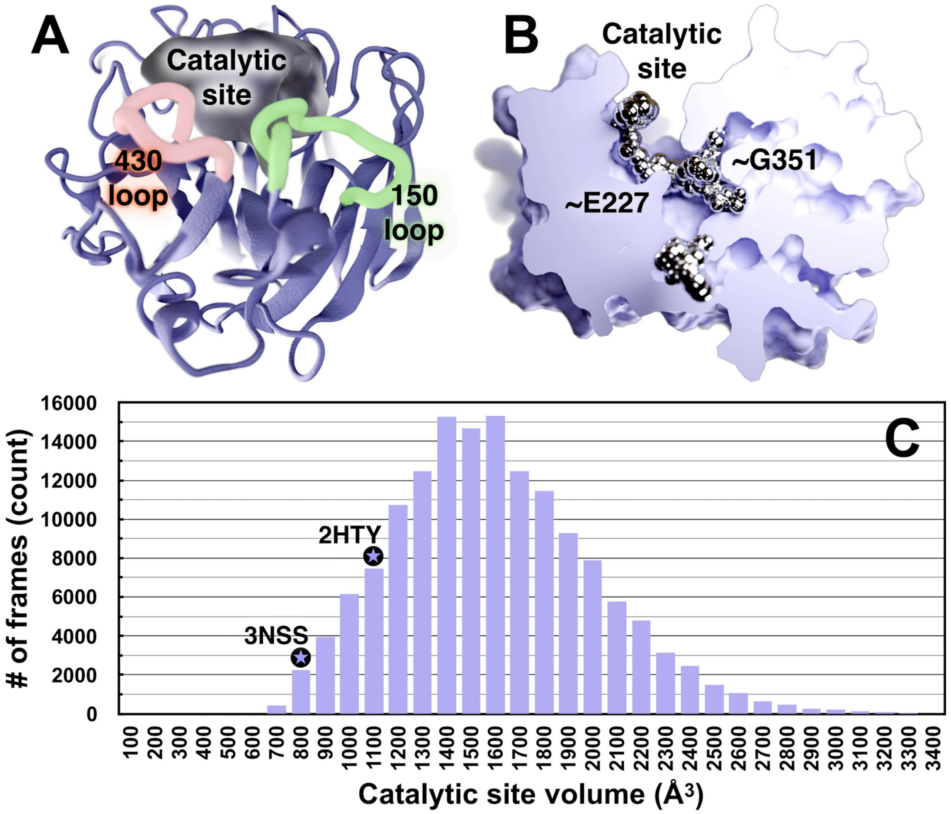

Figure 2. Volumetric and ligand binding "hot spot" analyses of the 1° catalytic site and adjacent regions. **(A)** NA is shown in blue ribbon, and the pocket volume is filled with semi-transparent gel. The 1° active site, 430-loop, and 150-loop are visible. **(B)** NA is shown as solid diffuse, and ligand-binding hotspots are metallic. A portion of the surface-rendered protein was removed to facilitate visualization of internal cavities. This NA conformation has a notably open G351 pocket, which has a high propensity to bind ligands. **(C)** A histogram of the NA catalytic-site volumes sampled during the MD simulations. As reference, the volumes of the same active-site cavity from two crystal structures are indicated with black circled stars. The 3NSS[20] structure (pH1N1 with a closed 150-cavity) has a volume of 800 Å^3, and the 2HTY[12] structure (H5N1 with an open 150-cavity) has a pocket volume of 1088 Å^3. The simulated-pocket volumes range from ~450 to 4440 Å^3 (intervals 3500-4500 not shown) as reported in **Table S1**; the average pocket volume is 1536 Å^3.

The volumetric and dynamical properties of the 1° site revealed in our simulations suggest that NA can bind many structurally distinct and complex sialoglycan receptors as part of the host-cell recognition process. Indeed, human glycans are vastly diverse in both their sugar composition and configuration (*e.g.*, long, short, biantennary, triantennary, etc.)[64]. As such, transient deepening and broadening of the 1° NA site may allow the glycoprotein to accommodate bulkier (*e.g.*, long, bi/triantennary) and longer glycan receptors. Given that our simulations model the entire viral envelope, it may be that full pocket opening only occurs in a crowded viral-surface environment. Alternatively, our simulations may capture full pocket opening because the viral coat includes many replicates of individual glycoproteins, enabling extensive conformational sampling.

Using the intramolecular distance between the 150- and 430-loops as a metric for 150-cavity formation, we constructed a two-state MSM from the conformations sampled by the viral envelope simulation to estimate the timescales of 150-loop opening and closing motions. Ultimately, we find the stationary distribution (equilibrium probabilities) of the open and closed states to be similar (0.53 and 0.47, respectively). Correspondingly, the time to transition between the two states (*i.e.*, the mean first-passage time, MFPT) is also roughly equal (39 ± 15 ns for open to closed, and 29 ± 11 ns from closed to open), indicating that loop opening and closing occur at similar rates (**Figure 3**). MSM calculations are detailed in the SI (section 1.9, and **Figures S11-S14)**.

To understand the impact of a crowded viral environment on loop sampling, we used the same protocol to construct an MSM from structures extracted from simulations of isolated NAs embedded in planar bilayers patches[61]. The 150-loop dynamics of both the viral envelope and single glycoprotein simulations are nearly equivalent, though the error associated with the viral-envelope MSM is much smaller, likely due to the increased simulation time. The single-glycoprotein equilibrium probabilities of the open and closed states were 0.61 and 0.39, respectively, and the MFPT ranges (open to closed and closed to open after 50 ± 96 ns and 72 ± 44 ns, respectively) overlap with those calculated using the viral envelope simulations. This comparison suggests that 150-cavity dynamics are not influenced by the crowded environment of the viral envelope, an expected result given that this pocket is oriented inwards (towards the neighboring three monomers of the same tetramer) rather than outwards (towards the environment).

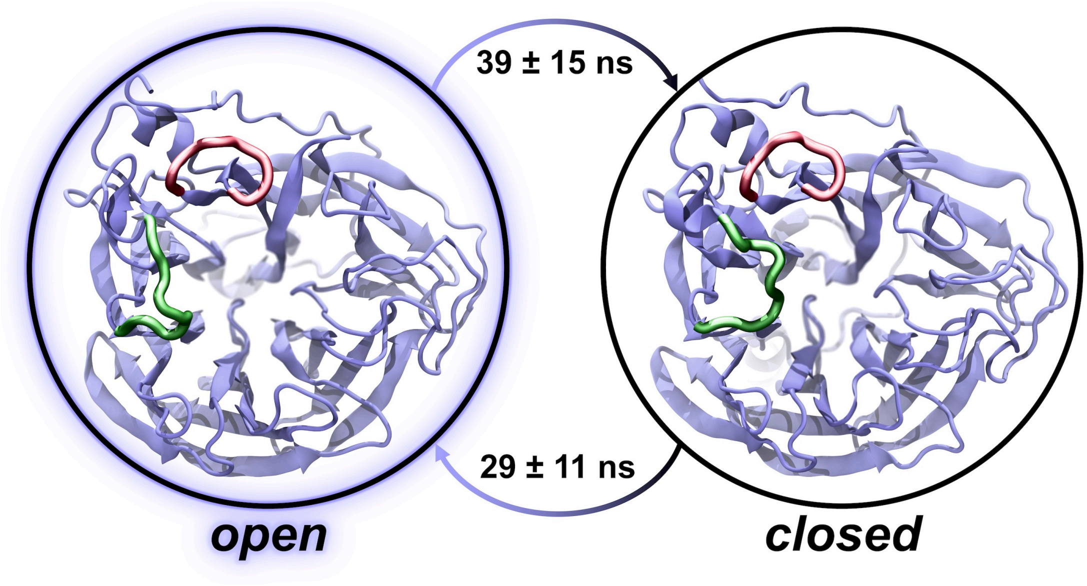

Figure 3. Two-state MSM with representative structures from the viral envelope simulation. The equilibrium populations of the open and closed states are approximately equal in both the viral envelope and single-glycoprotein simulations. Correspondingly, the mean first-passage times between the states are approximately equal. The 150-loop and 430-loop are represented as green and red ribbons, respectively.

Secondary binding site: functional annotation

We note that all FDA-approved NA inhibitors, as well as the endogenous ligand sialic acid, contain negatively charged carboxylate groups. Considering the hypothesis that the 2° site contributes to catalytic efficiency by recruiting and keeping substrates within close proximity to the catalytic site[28,29], and given that prior BD simulations indicate that substrates bind faster to the 2° site than the 1° site[17,39], we postulate that sialic acid first binds to the more solvent exposed 2° site. Subsequently, the electrostatics of the NA surface guides the substrates to the 1° enzymatic site.

Although sialic acid substrates were not included in the viral envelope simulation, we propose that the negatively charged chloride anions in the bulk solvent surrounding the NA monomers serve as a rough surrogate for negatively charged ligand moieties that may associate with the glycoprotein surface. To identify regions favorable to chloride occupancy, we concatenated the 120 monomeric NA simulations and aligned by the alpha-carbons of the 1° site. The chloride atoms were binned into 3,375,000 voxels (0.67 Å x 0.67 Å x 0.67 Å each). We focused on voxels containing chloride counts greater than three standard deviations above the mean. Notably, our simulations reveal that a volume of high chloride occupancy connects the 1° and 2° sialic acid binding sites (**Figure 4A**). This path is wide enough to allow negatively charged small molecules such as sialic acid or oseltamivir carboxylate to move from the 2° site to the 1° catalytic site. Additional regions of high chloride density are also depicted in **Figure 4A**. To further explore the role of electrostatics in the in this transfer mechanism, we calculated the electrostatic potential of the 120 NA monomers at the end of MD simulations using the adaptive Poisson-Boltzmann solver (APBS 1.4) software[65].

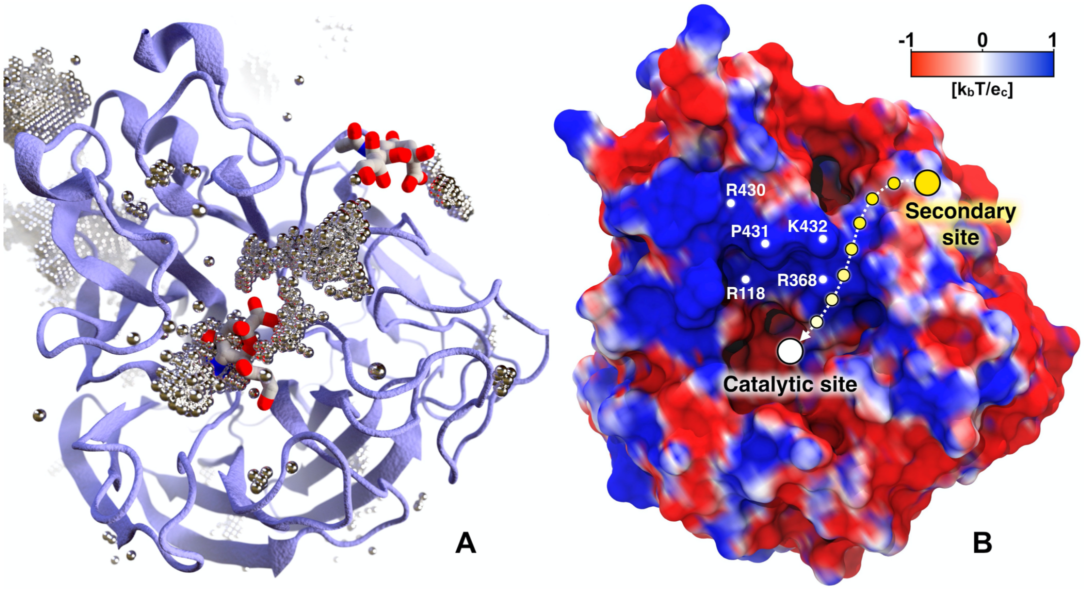

Figure 4. Chlorine anion distribution within the NA binding sites. The chlorine anion distribution **(A)** and the projection of the electrostatic potential onto the NA surface **(B)** shows the pathway between the 1° and 2° sites. In panel **A**, NA is drawn in ice blue cartoon. Regions of high chloride occupancy are illustrated as dotted silver bubbles. Two sialic acids (PDB ID: 1MWE[24]) are superimposed in the catalytic (center) and 2° (upper right) sites for reference[24]. In panel **B** the NA surface is colored with a palette varying from red (negative) to royal blue (positive), representing electrostatic potential values of $-1\ k_bT/e_c$ and $+1\ k_bT/e_c$, respectively. The path connecting the 2° site with the catalytic site is shown as a dashed arrow between circles fading from yellow (2° site) to white (catalytic site).

When projected onto the NA surface, the electrostatic potential ranging from $-1\ k_bT/e_c$ to $+1\ k_bT/e_c$ shows a positive region connecting the two sites (**Figure 4B**). Positively charged residues such as R118, R368, R430, K432, and P431 (N2 numbering scheme) largely determine this profile. Interestingly, the same analysis performed on representative NA structures with open and closed 1° pockets (extracted with MSM and shown in **Figure 3**) reveal that these residues are less exposed in the closed state (**Figure S15**). These results provide evidence that the two sites may act cooperatively, supporting the work of Lai et al.[16], which confirmed that pN1 has a 2° site that can bind sialic acid. It also supports the work of Le et al.[66], which suggested electrostatic funneling as being the main driving force for oseltamivir carboxylate association to the active site. Chloride anion and electrostatic analyses are detailed in the SI, section 1.10.

Taken together, these results suggest a biophysical mechanism for the previously uncharacterized 2° site. Sialic acid receptors may first bind the 2° site before being transferred to the 1° sialidase site (**Figure 5**). We propose that our chloride distribution analysis is well suited for studying these possible mechanisms of molecular transfer. In contrast to a simple electrostatic map, our simulation-based analysis accounts for both electrostatic and steric factors, as well as for the conformational dynamics sampled over all 14.5 µs of monomeric simulation in the context of the whole viral-envelope environment. In addition, the proposed “bind and transfer” mechanism is in good agreement with prior experimental results and proposed mechanisms[28].

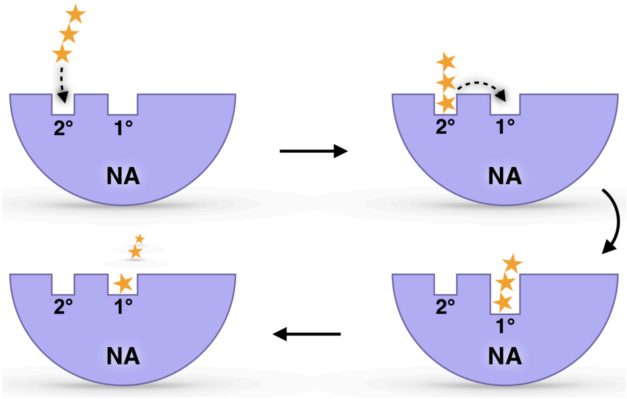

Figure 5. The predicted sialic acid "bind and transfer" mechanism. Yellow stars represent a sialic-acid-containing glycan receptor. Blue half circles represent NA. The 1° active site and 2° site are also labeled.

Conclusions

Our work suggests a novel NA binding mechanism wherein a sialic acid containing substrate (*e.g.*, a glycan receptor) first binds the 2° site, as predicted by earlier BD simulations[17]. After binding, the substrate is transferred to the catalytic site via electrostatic interactions. Finally, the catalytic site cleaves the terminal sialic acid substrate. In other words, a budding viral particle might use the 2° site to first attract the sialic-acid-tipped receptors before these are cleaved within the catalytic site, ultimately allowing viral escape from the infected host-cell surface. MSM and volumetric analyses also further expand the functional annotation of the 1° site and surrounding regions, disclosing exceptionally deep and broad catalytic-pocket conformations. These findings can be exploited to design novel multi-pronged inhibitors capable of reaching the additional NA cavities unveiled in our multiscale simulations.

Taken together, this information provides fundamental insights into our understanding of sialic-acid/oseltamivir-carboxylate recognition, suggesting new strategies for the development of NA inhibitors. Our work also provides a novel methodological advance for extracting long time-scale dynamics from short time-scale simulations by applying the powerful MSM theoretical framework at subcellular and cellular scales.

Data availability

A NAMD input script used for production run MD and the Jupyter notebook used to create the MSM are provided as Supporting Information. Additional analysis scripts are available upon request.

References

1. Centers for Disease Control and Prevention. Summary of the 2017-2018 Influenza Season. **11**, 7–10 (2018).
2. Wang, X. *et al.* Epidemiology of avian influenza A H7N9 virus in human beings across five epidemics in mainland China, 2013–17: an epidemiological study of laboratory-confirmed case series. *Lancet Infect. Dis.* **17**, 822–832 (2017).
3. De Clercq, E. Antiviral agents active against influenza A viruses. *Nat. Rev. Drug Discov.* **5**, 1015–1025 (2006).
4. Hamilton, B. S., Whittaker, G. R. & Daniel, S. Influenza virus-mediated membrane fusion: determinants of hemagglutinin fusogenic activity and experimental approaches for assessing virus fusion. *Viruses* **4**, 1144–1168 (2012).
5. von Itzstein, M. The war against influenza: discovery and development of sialidase inhibitors. *Nat Rev Drug Discov* **6**, 967–974 (2007).
6. Gamblin, S. J. & Skehel, J. J. Influenza hemagglutinin and neuraminidase membrane glycoproteins. *J Biol Chem* **285**, 28403–28409 (2010).

7. An, J. *et al.* A novel small-molecule inhibitor of the avian influenza H5N1 virus determined through computational screening against the neuraminidase. *J Med Chem* **52**, 2667–2672 (2009).

8. Feng, E. *et al.* Structure-based design and synthesis of C-1- and C-4-modified analogs of zanamivir as neuraminidase inhibitors. *J Med Chem* **56**, 671–684 (2013).

9. Wen, W. H. *et al.* Analogs of zanamivir with modified C4-substituents as the inhibitors against the group-1 neuraminidases of influenza viruses. *Bioorganic Med. Chem.* **18**, 4074–4084 (2010).

10. Rudrawar, S. *et al.* Novel sialic acid derivatives lock open the 150-loop of an influenza A virus group-1 sialidase. *Nat Commun* **1**, 113 (2010).

11. Amaro, R. E. *et al.* Remarkable Loop Flexibility in Avian Influenza N1 and Its Implications for Antiviral Drug Design. *J. Am. Chem. Soc.* **129**, 7764–7765 (2007).

12. Russell, R. J. *et al.* The structure of H5N1 avian influenza neuraminidase suggests new opportunities for drug design. *Nature* **443**, 45–49 (2006).

13. Woods, C. J. *et al.* Long time scale GPU dynamics reveal the mechanism of drug resistance of the dual mutant I223R/H275Y neuraminidase from H1N1-2009 influenza virus. *Biochemistry* **51**, 4364–4375 (2012).

14. Amaro, R. E., Cheng, X., Ivanov, I., Xu, D. & McCammon, J. A. A. Characterizing Loop Dynamics and Ligand Recognition in Human- and Avian-Type Influenza Neuraminidases via Generalized Born Molecular Dynamics and End-Point Free Energy Calculations. *J. Am. Chem. Soc.* **131**, 4702–4709 (2009).

15. Hausmann, J., Kretzschmar, E., Garten, W. & Klenk, H.-D. N1 neuraminidase of influenza virus A / FPV / Rostock / 34 has haemadsorbing activity. *J. Gen. Virol.* **76**,

1719–1728 (1995).

16. Lai, J. C. C. *et al.* A secondary sialic acid binding site on influenza virus neuraminidase: Fact or fiction? *Angew. Chemie - Int. Ed.* **51**, 2221–2224 (2012).

17. Sung, J. C., Wynsberghe, A. W. Van, Amaro, R. E., Li, W. W. & McCammon, J. A. Role of Secondary Sialic Acid Binding Sites in Influenza N1 Neuraminidase. *J. Am. Chem. Soc.* **132**, 2883–2885 (2010).

18. Wang, T. & Wade, R. C. Comparative Binding Energy (COMBINE) Analysis of Influenza Neuraminidase−Inhibitor Complexes. *J. Med. Chem.* **44**, 961–971 (2001).

19. Amaro, R. E. *et al.* Mechanism of 150-cavity formation in influenza neuraminidase. *Nat. Commun.* **2**, 388 (2011).

20. Li, Q. *et al.* The 2009 pandemic H1N1 neuraminidase N1 lacks the 150-cavity in its active site. *Nat. Struct. Mol. Biol.* **17**, 1266–1268 (2010).

21. Han, N. & Mu, Y. Plasticity of 150-Loop in Influenza Neuraminidase Explored by Hamiltonian Replica Exchange Molecular Dynamics Simulations. *PLoS One* **8**, e60995 (2013).

22. Laver, W. G., Colman, P. M., Webster, R. G., Hinshaw, V. S. & Air, G. M. Influenza virus neuraminidase with hemagglutinin activity. *Virology* **137**, 314–323 (1984).

23. Webster, R. G. *et al.* Antigenic structure and variation in an influenza virus N9 neuraminidase. *J. Virol.* **61**, 2910–6 (1987).

24. Varghese, J. N. *et al.* Structural evidence for a second sialic acid binding site in avian influenza virus neuraminidases. *Proc Natl Acad Sci U S A* **94**, 11808–11812 (1997).

25. Air, G. M. Influenza neuraminidase. *Influenza Other Respi. Viruses* **6**, 245–256 (2012).

26. Streltsov, V. A., Pilling, P., Barrett, S. & McKimm-Breschkin, J. L. Catalytic mechanism

and novel receptor binding sites of human parainfluenza virus type 3 hemagglutinin-neuraminidase (hPIV3 HN). *Antiviral Res.* **123**, 216–223 (2015).

27. Kobasa, D., Rodgers, M. E., Wells, K. & Kawaoka, Y. Neuraminidase hemadsorption activity, conserved in avian influenza A viruses, does not influence viral replication in ducks. *J. Virol.* **71**, 6706–6713 (1997).
28. Uhlendorff, J., Matrosovich, T., Klenk, H.-D. & Matrosovich, M. Functional significance of the hemadsorption activity of influenza virus neuraminidase and its alteration in pandemic viruses. *Arch. Virol.* **154**, 945–957 (2009).
29. Dai, M. *et al.* Mutation of the Second Sialic Acid-Binding Site, Resulting in Reduced Neuraminidase Activity, Preceded the Emergence of H7N9 Influenza A Virus. *J. Virol.* **91**, (2017).
30. Du, W. *et al.* Substrate Binding by the Second Sialic Acid-Binding Site of Influenza A Virus N1 Neuraminidase Contributes to Enzymatic Activity. *J. Virol.* **92**, e01243-18 (2018).
31. Du, W. *et al.* The 2nd sialic acid-binding site of influenza A virus neuraminidase is an important determinant of the hemagglutinin-neuraminidase-receptor balance. *PLOS Pathog.* **15**, e1007860 (2019).
32. Zeller, F., Luitz, M. P., Bomblies, R. & Zacharias, M. Multiscale Simulation of Receptor–Drug Association Kinetics: Application to Neuraminidase Inhibitors. *J. Chem. Theory Comput.* **13**, 5097–5105 (2017).
33. Ivinson, K. *et al.* Salivary Blockade Protects the Lower Respiratory Tract of Mice from Lethal Influenza Virus Infection. *J. Virol.* **91**, e00624-17 (2017).
34. Gilbertson, B., Ng, W. C., Crawford, S., McKimm-Breschkin, J. L. & Brown, L. E. Mouse

Saliva Inhibits Transit of Influenza Virus to the Lower Respiratory Tract by Efficiently Blocking Influenza Virus Neuraminidase Activity. *J. Virol.* **91**, e00145-17 (2017).

35. Xu, R. *et al.* Structural Basis of Preexisting Immunity to the 2009 H1N1 Pandemic Influenza Virus. *Science (80-.).* **328**, 357–360 (2010).
36. Harris, A. *et al.* Influenza virus pleiomorphy characterized by cryoelectron tomography. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 19123–19127 (2006).
37. Singharoy, A. *et al.* Atoms to Phenotypes: Molecular Design Principles of Cellular Energy Metabolism. *Cell* **179**, 1098-1111.e23 (2019).
38. Karplus, M. Development of Multiscale Models for Complex Chemical Systems: From H+H 2 to Biomolecules (Nobel Lecture). *Angew. Chemie Int. Ed.* **53**, 9992–10005 (2014).
39. Amaro, R. E. *et al.* A Computational Assay that Explores the Hemagglutinin/Neuraminidase Functional Balance Reveals the Neuraminidase Secondary Site as a Novel Anti-Influenza Target. *ACS Cent. Sci.* **4**, 1570–1577 (2018).
40. Hadden, J. A. & Perilla, J. R. All-atom virus simulations. *Curr. Opin. Virol.* **31**, 82–91 (2018).
41. Huber, R. G., Marzinek, J. K., Holdbrook, D. A. & Bond, P. J. Multiscale molecular dynamics simulation approaches to the structure and dynamics of viruses. *Prog. Biophys. Mol. Biol.* **128**, 121–132 (2017).
42. Freddolino, P. L., Arkhipov, A. S., Larson, S. B., McPherson, A. & Schulten, K. Molecular dynamics simulations of the complete satellite tobacco mosaic virus. *Structure* **14**, 437–449 (2006).
43. Zink, M. & Grubmüller, H. Mechanical properties of the icosahedral shell of southern bean mosaic virus: A molecular dynamics study. *Biophys. J.* **96**, 1350–1363 (2009).

44. Ayton, G. S. & Voth, G. A. Multiscale computer simulation of the immature HIV-1 virion. *Biophys. J.* **99**, 2757–2765 (2010).

45. Roberts, J. A., Kuiper, M. J., Thorley, B. R., Smooker, P. M. & Hung, A. Investigation of a predicted N-terminal amphipathic α-helix using atomistic molecular dynamics simulation of a complete prototype poliovirus virion. *J. Mol. Graph. Model.* **38**, 165–173 (2012).

46. Zhao, G. *et al.* Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature* **497**, 643–646 (2013).

47. Andoh, Y. *et al.* All-atom molecular dynamics calculation study of entire poliovirus empty capsids in solution. *J. Chem. Phys.* **141**, 165101 (2014).

48. Reddy, T. *et al.* Nothing to Sneeze At: A Dynamic and Integrative Computational Model of an Influenza A Virion. *Structure* **23**, 584–597 (2015).

49. Reddy, T. & Sansom, M. S. P. The Role of the Membrane in the Structure and Biophysical Robustness of the Dengue Virion Envelope. *Structure* **24**, 375–382 (2016).

50. Marzinek, J. K., Holdbrook, D. A., Huber, R. G., Verma, C. & Bond, P. J. Pushing the Envelope: Dengue Viral Membrane Coaxed into Shape by Molecular Simulations. *Structure* **24**, 1410–1420 (2016).

51. Pande, V. S., Beauchamp, K. & Bowman, G. R. Everything you wanted to know about Markov State Models but were afraid to ask. *Methods* **52**, 99–105 (2010).

52. Prinz, J. H. *et al.* Markov models of molecular kinetics: Generation and validation. *J. Chem. Phys.* **134**, 174105 (2011).

53. Chodera, J. D. & Noé, F. Markov state models of biomolecular conformational dynamics. *Curr. Opin. Struct. Biol.* **25**, 135–144 (2014).

54. Plattner, N. & Noé, F. Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models. *Nat. Commun.* **6**, 7653 (2015).

55. Plattner, N., Doerr, S., De Fabritiis, G. & Noé, F. Complete protein-protein association kinetics in atomic detail revealed by molecular dynamics simulations and Markov modelling. *Nat. Chem.* **9**, 1005–1011 (2017).

56. Voelz, V. A., Bowman, G. R., Beauchamp, K. & Pande, V. S. Molecular Simulation of ab Initio Protein Folding for a Millisecond Folder NTL9(1-39). *J. Am. Chem. Soc.* **132**, 1526–1528 (2010).

57. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **26**, 1781–1802 (2005).

58. Huang, J. & MacKerell, A. D. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* **34**, 2135–2145 (2013).

59. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926 (1983).

60. Beglov, D. & Roux, B. Finite representation of an infinite bulk system: Solvent boundary potential for computer simulations. *J. Chem. Phys.* **100**, 9050–9063 (1994).

61. Durrant, J. D., Bush, R. M. & Amaro, R. E. Microsecond Molecular Dynamics Simulations of Influenza Neuraminidase Suggest a Mechanism for the Increased Virulence of Stalk-Deletion Mutants. *J. Phys. Chem. B* **120**, 8590–8599 (2016).

62. Durrant, J. D., Votapka, L., Sørensen, J. & Amaro, R. E. POVME 2.0: An Enhanced Tool for Determining Pocket Shape and Volume Characteristics. *J. Chem. Theory Comput.* **10**,

5047–5056 (2014).

63. Ngan, C. H. *et al.* FTMAP: extended protein mapping with user-selected probe molecules. *Nucleic Acids Res* **40**, W271-5 (2012).

64. Marino, K., Bones, J., Kattla, J. J. & Rudd, P. M. A systematic approach to protein glycosylation analysis: a path through the maze. *Nat Chem Biol* **6**, 713–723 (2010).

65. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 10037–41 (2001).

66. Le, L., Lee, E. H., Hardy, D. J., Truong, T. N. & Schulten, K. Molecular Dynamics Simulations Suggest that Electrostatic Funnel Directs Binding of Tamiflu to Influenza N1 Neuraminidases. *PLoS Comput. Biol.* **6**, e1000939 (2010).

Supplementary Information

Supplementary Information available online.

Author Contributions

REA designed and oversaw the research project. JDD built the structures and performed the MD simulations. REA, JDD, SEK, LC, PUL, AD performed the simulation analyses. JDD, LC, PUL, AD created the figures. REA, JDD, SEK, LC, PUL, AD wrote the paper.

Acknowledgments

This work was funded in part by the Director’s New Innovator Award Program NIH DP2 OD007237 and the National Biomedical Computation Resource (NBCR) through NIH P41 GM103426 to REA. Computing support on the NSF Blue Waters Petascale Computer was provided through NSF OAC-1811685. We thank the NSF for access to the TACC Stampede supercomputer through CHE060073N to REA, and the Center for Research Computing at the University of Pittsburgh for a computing allocation to JDD. The authors also thank Dr. Robert Malmstrom for providing sample MSM-building scripts and useful discussions.

Supplementary Information

Mesoscale All-Atom Influenza Virus Simulations Suggest New Substrate Binding Mechanism

Authors: Jacob Durrant[1,†], Sarah Kochanek[2,†], Lorenzo Casalino[2], Pek U Ieong[2], Abigail Dommer[2], Rommie E. Amaro[2*]

Affiliations:

[1]Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260

[2]Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, CA 92093-0340

† These authors contributed equally.

*Correspondence to: ramaro@ucsd.edu

Table of contents:

1.1 Building single-glycoprotein hemagglutinin (HA) and neuraminidase (NA) structural models

To study the 2009 H1N1 pandemic strain, we built each component of the viral envelope. Our initial efforts focused on creating whole-construct glycoprotein models. Crystal structures of some glycoprotein components (e.g., the HA head and NA ectodomain regions) are present in the Protein Data Bank but other NA and HA motifs are incalcitrant to crystallography. We used protein/protein docking and homology modeling to model the structures of the NA stalk as well as the NA and HA transmembrane domains, which are typically absent from crystallographic constructs.

We have previously described how we modeled and simulated the NA stalk and transmembrane regions[1], but the HA modeling and simulations have not yet been published. To match the exact Shandong 2009 H1N1 HA sequence (accession number F2YI86), we created a homology model of the influenza A HA extra-virion domain using Schrödinger's Prime module. Prime recommended the 2WR0[2] structure as a template. We note that both F2YI86 and 2WR0 correspond to group-1 HA strains, which all share high structural and sequence homology. Protonation states were assigned using PDB2PQR[3] with PROPKA[4–7] at pH 7.0. The homology model was subsequently energy minimized and geometry-optimized using the default parameters in Schrödinger's Maestro suite.

No structures of the HA transmembrane domain have been deposited in the PDB. To create a homology model of this motif, we searched the Protein Data Bank for analogous extended trimeric alpha-helical bundles. The 2WPQ structure[8] was identified as a good template candidate. Modeled alpha helices with the appropriate HA sequences were aligned to the helices of the 2WPQ structure. Inter-virion domains are likely unstructured and so were added to the modeled transmembrane domain as an untemplated extension. The HA extra-virion domain with appropriate disulfide bonds was positioned relative to the transmembrane and inter-virion domains using VMD[9].

We separately used CHARMM-GUI[10,11] to construct a planar lipid bilayer containing an appropriate mixture of lipids per ref. 12, which describes the composition of the influenza viral envelope. We considered all membrane components of the inner and outer leaflets with mol % > 1.5. To utilize only lipid molecules present in the well-tested CHARMM 36 force field[13,14], we substituted some of the experimentally identified lipids with related, parameterized lipid molecules based on chemical similarity. 3-palmitoyl-2-oleoyl-d-glycero-1-phosphatidylethanolamine (POPE) was used for any phosphatidylethanolamine (PE) variant; 2,3-distearoyl-d-glycero-1-phosphatidylserine (DSPS) for phosphatidylserine (PS); 3-palmitoyl-2-oleoyl-d-glycero-1-phosphatidylglycerol (POPG) for the Forssman glycolipid hapten; and 3-palmitoyl-2-oleoyl-d-glycero-1-phosphatidylcholine (POPC) for sphingomyelin (SM). Cholesterol and phosphatidylcholine, also major components of the envelope membrane, are present in the CHARMM 36 force field and so were not substituted[13].

We next embedded the HA models within the generated bilayer, as described in ref[1]. To resolve steric clashes, any lipid or membrane component within 3.0 Å of the protein was removed using PyMolecule[15], an early version of the now-published Scoria Python package[16] for easily manipulating 3D molecular data. We further resolved steric clashes using serial iterations of minimization and geometry optimization in Schrödinger's Maestro, coupled with manual atomic and molecular manipulation in VMD[9] as needed.

We used the VMD plug-in cionize[17] to position sodium and chloride ions as required to bring the HA/membrane system to electrical neutrality and to simulate a 20 nM solution. We manually adjusted the positions of some ions that were placed far from any protein or lipid molecule. We then used Amber's XLEAP module[18] to generate a water box extending 15 Å beyond any protein or lipid atom. Excess water molecules were removed if they were not positioned directly above the bilayer, such that the 15 Å margin was ultimately retained only along the Z axis, perpendicular to the bilayer itself. The complete HA system contained 417,457 atoms.

1.2 Details of the single-glycoprotein HA and NA molecular dynamics simulations

We performed multiple independent all-atom molecular dynamics (MD) simulations of both the NA and HA complete-construct, single-glycoprotein, membrane-bound systems. Our goal in performing these simulations was to refine and fully equilibrate the modeled protein structures so that they could be subsequently used to populate the initial viral envelope model (described below).

The single-glycoprotein NA simulations are described in ref. 1. We followed a similar protocol when simulating the single-glycoprotein HA system. In brief, we used PSFGEN to parameterize the protein, lipid, and water residues according to the CHARMM22[13,19–21] and TIP3P force fields[22]. We then used NAMD 2.9[23] on the Stampede supercomputer at the Texas Advanced Computing Center to minimize, equilibrate, and simulate the system. Periodic boundary conditions used the particle mesh Ewald method to account for long-range electrostatics (smoothing cutoff: 14 Å).

To prepare the HA system for production simulation, we first performed 1000 1-fs steps of unrestrained conjugate gradient minimization, with NAMD’s velocity quenching option turned on and maximumMove parameter set to 0.0001. We then equilibrated the system using an NPT-ensemble protocol at 310 K. Langevin dynamics and a modified Langevin piston Nosé–Hoover thermostat maintained the temperature and pressure at 310 K and 1 atm, respectively. The gradual, constrained equilibration was performed in four phases consisting of 250,000 1-fs steps each. We applied harmonic-constraint force constants of 4, 3, 2, and 1 kcal/mol/Å^2 to the atoms of the protein backbone during each phase, respectively. Finally, we subjected the HA system to further unconstrained equilibration (10,000,000 1-fs steps).

Following equilibration, we performed five distinct HA productive runs of 100 ns each (~100 million 1-fs steps). We saved 8000 frames from the last 80% of each simulation for subsequent analysis.

1.3 Building the viral envelope model from a simplified cryoelectron-tomography “point model”

Here we summarize the computational approach that we adopted to model the influenza viral envelope, which we have already discussed in some detail in ref. 24. The final model is shown in **Figure 1** (main text) and **Figure S1** (cross section).

Point model of the influenza exterior. As a starting reference, experimental collaborators provided us with a simple point model of the influenza exterior derived from electron tomography[25]. Surface points represented the lipid-covered virion surface, and lines protruding from the viral surface represented the NA and HA “spikes.” This model provided both the morphological insights (glycoprotein distribution) and geometrical information (virion shape) required to model the lipid bilayer. More details can be found in ref. 24.

Positioning the glycoprotein models. To generate the initial viral envelope model, we selected an ensemble of representative structures from the fully equilibrated single-glycoprotein HA and NA simulations described above. We aligned the MD-sampled HA and NA conformations and clustered on their C$_\alpha$ atoms using RMSD clustering[26], as implemented in the GROMACS computer package[27]. We systematically and independently varied the RMSD cutoffs for HA and NA to identify values that grouped the respective glycoprotein conformations into five clusters each (see ref. 1 for details). The central member of each cluster was considered most representative. These centroid glycoprotein conformations were programmatically positioned at the appropriate locations on the cryoelectron-tomography-derived point model using PyMolecule[15,16].

The point model suggested that some glycoproteins were in very close proximity, making steric clashes unavoidable. We used a multi-step process to resolve these clashes. First, glycoproteins that were sufficiently distant from their neighbors were fixed. Using a Monte-Carlo method, the remaining glycoproteins were allowed to randomly "jump" to adjacent regions on the virion surface. With every random jump, the distances between neighboring glycoproteins were again evaluated, and any glycoprotein sufficiently distant from its new neighbors was also fixed. Additionally, with each jump, each glycoprotein had a 10% chance of being returned to its original location, assuming that location was not occupied by another glycoprotein. This ensured that glycoproteins did not wander too far from their initial, experimentally determined locations.

The ultimate goal of acceptably positioning every glycoprotein was not achievable using the above method alone, so we also gradually relaxed the distance cutoffs during the Monte Carlo procedure to keep the number of clashes to a minimum. Following the procedure, we performed a pairwise distance comparison between the heavy atoms of all neighboring glycoproteins to eliminate those occasional clashes that remained. Two glycoproteins were said to clash if the distance between any of their heavy atoms was less than 1.0 A. Those glycoproteins that clashed with the greatest number of their neighbors were removed first, followed by the less egregious offenders. Ultimately, it was only necessary to eliminate 10-15% of the glycoproteins to resolve all steric clashes.

Inserting M2 channels. For completeness sake, we randomly positioned eleven identical models of the M2 transmembrane domain, derived from the 2L0J structure[28], at viral-surface regions that were not occupied by glycoproteins. At the time of model construction, the published M2 structures and available structural templates based on sequence similarity did not allow us to create a full-sequence M2 homology model. Subsequent examination of the model (after extensive simulation) also revealed that the M2 channels had the incorrect orientation. However, these channels are sparsely distributed (i.e., only 11 throughout the entire viral envelope) and are so far from the glycoprotein ectodomains that their impact on the structure and dynamics of the NA and HA sialic-acid binding sites is expected to be negligible.

Generating the viral envelope lipid bilayer. We next used LipidWrapper[29], a PyMolecule-based program designed for creating large-scale lipid-bilayer models of arbitrary geometry, to carpet the entire asymmetrical surface of the virion point model with phosphatidylcholine (POPC) molecules. LipidWrapper used a large planar lipid-bilayer model as input, which we generated using the CHARMM-GUI server[10,11]. Lipid residues that came within 3 Å of any protein were deleted.

We modeled a pure POPC bilayer, rather than a mixed bilayer, for practical reasons. It is very difficult to create asymmetrical mixed-lipid models with leaflets that have identical densities. When modeling a small lipid patch, slight differences in leaflet densities are of little consequence. But these small differences are magnified many times over when modeling mesoscale structures such as the influenza viral envelope. Fortunately, we are primarily interested in the dynamics of the glycoprotein ectodomains, which are over 100 Å distant from the bilayer surface[24]. Regardless, lipid-relevant dynamics such as lateral diffusion through the bilayer occur on time scales that are much longer than those sampled by the viral envelope simulations described in this work (see below).

Ionization and solvation. The viral-surface model was solvated both internally and externally with explicit water molecules, leading to a total system that had dimensions 1,218 Å x 1,215 Å x 1,279 Å (~1,900,000 nm^3). We added Na+ and Cl- ions as required to first bring the system to electrical neutrality and then to model a counterion concentration of 0.15 M. The final system—including 236 HA trimers, 30 NA tetramers, 11 M2 ion channels, the quasi-spherical POPC lipid bilayer, waters, and ions—was comprised of 160,653,271 atoms.

1.4 Complete viral envelope, all-atom molecular dynamics simulations

We next used all-atom MD simulations to study the dynamics of the viral envelope model. The complete system was parameterized using PSFGEN according to the CHARMM36 all-atom additive force fields for proteins and lipids[13,14,21], the TIP3P force field for water molecules[22], and the Beglov and Roux force field[30] for the ions.

Minimizing and equilibrating so large a system was challenging. As mentioned above, the NA and HA glycoproteins used to populate the viral surface were themselves taken from single-glycoprotein MD simulations and so were effectively pre-equilibrated. To further equilibrate these glycoproteins in the context of the viral envelope bilayer, we used NAMD[23] to subject the viral envelope model to iterative rounds of minimization and simulation. Multiple cycles of minimization were performed using NAMD's default minimizer (conjugate gradient). Following minimization, the system was subjected to 3.2 ns (2 fs time step) of equilibration under an NPT ensemble. During the equilibration phases, Langevin dynamics with a damping coefficient of 1/ps[31] were used to gradually increase the temperature to 298 K, and a Nosé-Hoover Langevin piston[32,33] was used to maintain the pressure (1.01325 bar). Between these iterative rounds of minimization and equilibration, it was occasionally necessary to patch holes in the lipid bilayer. This process eventually produced a bilayer that was reasonably stable. We note, however, that full equilibration was not possible because our viral envelope model does not include the virion interior. To maintain the overall geometry of the model in the absence of structure-supporting interior components, we fixed the positions of every 10th inner-leaflet lipid head group. This same constraint was used during all subsequent viral envelope simulations.

We next subjected the system to productive MD simulations using a Langevin thermostat (298 K) and a Nosé-Hoover Langevin barostat (1.01325 bar) to achieve temperature and pressure control, respectively. We used a 12 Å cutoff to evaluate non-bonded interactions at each time step. A full electrostatic evaluation was performed every 3 time steps using the particle mesh Ewald method[34] to account for long-range electrostatic interactions. The simulations were run with a 2 fs integration time step. Atomic bonds involving hydrogen atoms were fixed with SHAKE[35] using the option "rigidbonds all".

Simulations were run using a memory-optimized version of NAMD 2.10[23] on the Blue Waters supercomputer (114,688 processors, equivalent to 16,384 Blue Waters nodes or 4,096 physical nodes). Over the course of the simulations, NAMD performed 252 benchmark calculations, which averaged 25.57 steps/second (i.e., 51.14 fs of simulation per second of real time). Frames were saved to disk every 10,000 steps (20 ps), ultimately occupying 11.66 terabytes of disk space. More details describing how to simulate mesoscale systems with NAMD can be found at https://www.ks.uiuc.edu/Research/namd/wiki/index.cgi?NamdMemoryReduction.

After 40.14 ns, the productive simulation was forked into two daughter simulations that continued for 30.60 and 15.16 ns, respectively (**Figure S2**). The second daughter simulation was then also forked into simulations that lasted 20.70 and 14.44 ns, respectively. Prior to forking, a single hole developed in the virion bilayer after roughly 20 ns. This hole comprised roughly 0.2%

of the entire virion surface area and so had a minimal effect on overall viral envelope dynamics. It was repaired prior to launching the 14.44-ns simulation. Taken together, these simulations constitute 121.04 ns total. The NAMD configuration file used for the production run is provided as additional material in the Supporting Information.

1.5 RMSD and RMSF analyses of the viral envelope MD simulations

As mentioned above, the NA and HA glycoproteins used to populate the viral surface were themselves taken from multicopy MD simulations and so were effectively pre-equilibrated. To verify that the NA ectodomains remained equilibrated over the course of the viral envelope simulation, for each simulation frame we calculated the average root mean square distance (RMSD) between the 3NSS crystal structure[36] and each of the 120 viral envelope NA monomers. We similarly calculated the per-frame average RMSD between a 3LZG-based HA model and the 708 viral envelope HA monomers. These average RMSD values did not drift substantially over the course of the simulation (**Figure S3**).

In addition to RMSD, we provide a root mean square fluctuation (RMSF) analysis of the C_α of the NA (**Figure S4**) and HA (**Figure S5**) glycoproteins over the course of the viral envelope simulations. To calculate the RMSF values of each residue, all monomeric trajectories were aligned by the head-domain C_α (i.e., those residues resolved in the 3NSS structure). The atoms of each residue were assigned the RMSF values calculated from the corresponding C_α movements.

To ascertain the structural variation overall, we computed the RMSD for the first frame of the simulation versus the last frame, over all the glycoprotein C_α. The results of this calculation are shown in **Figure S6**.

1.6 Viral envelope lipid and curvature analysis

Although this investigation focuses primarily on the dynamics of the two glycoproteins HA and NA, we here provide some analysis of the lipid bilayer and the overall curvature of the simulated envelope. These analyses serve as basic system checks and inform on the other aspects contained within the simulated data.

Phospholipid curvature values were calculated by extracting the coordinates of the headgroup phosphorous atoms and fitting a sphere to surrounding lipids within a 200 Å radius (**Figure S7**). The inverse radius of the fitted sphere gives the curvature value. The sphere fit was calculated using the least squares algorithm available in NumPy[21].

The area per lipid was estimated at each lipid coordinate by dividing the number of lipids within the 200 Å radius by the surface area of the spherical cap as calculated from the radius of curvature.

To investigate the dynamics and movement of lipids in the bilayer, we computed lipid streamlines for the lipids at the end of a 70 ns trajectory over a 3.4 ns timestep (**Figure S8**). Streamlines were calculated using the MDAnalysis[37,38] streamplot algorithm developed and described by Chavent and Reddy et al.[39] The calculation uses a vector field to describe the motion of lipid groups. A three-dimensional grid first subdivides the influenza model, followed by a calculation of the lipid center-of-mass coordinates corresponding to each grid box. The same calculation is performed for that center of mass after a timestep *t*, producing a vector describing the displacement in the x, y and z directions. We used a grid resolution of 20 Å. The vector field datasets were visualized and plotted using the MayaVi visualization package[40].

Others have observed that concerted lipid motion in both vesicular and planar bilayer simulations is often highly transient, lasting on the order of nanoseconds[39,41,42]. Our streamline visualization of the outer leaflet revealed instances of lipid vortices around some, but not all, embedded glycoproteins. Though beyond the scope of this work, further investigation of lipid diffusion coefficients, protein diffusion, and rigorous correlation plots may reveal how vesicular membrane composition is likely to modulate observed nanometer-scale motions. However, the short duration of the simulation prohibits a thorough analysis of longer-timescale events such as lipid mixing and the impact of protein crowding on lateral lipid diffusion[41–43].

Finally, we plotted the glycoprotein distribution with respect to local surface curvature to ascertain whether the presence of glycoproteins caused a significant deformation in the viral envelope (**Figure S9**). The protein distribution was calculated by mapping the C_α, onto the curvature value associated with the nearest lipid. Rather than mapping a single representative protein to a point, the C_α were selected and mapped to their nearest lipid neighbors to account for the total diversity of lipid-protein interactions and curvature values. The resulting histogram, normalized by curvature/surface area, indicates that the protein content skews slightly toward higher curvature values, indicating that the glycoproteins may induce local bilayer deformation. It is important to note, however, that the simulation length does not account for large-scale protein diffusion[39] and that the membrane, which contains only POPC molecules, does not accurately capture the full chemical complexity of the bilayer.

1.7 NA catalytic site principal component analysis

To perform the principal component analysis (PCA) shown in **Figure 1E**, we considered the C_α of the 19 NA active-site residues homologous to those that come within 5 Å of the crystallographic oseltamivir molecule solved in ref. [44] (PDB ID: 2HU4). In 2HU4, these residues are ARG:118, ARG:152, ARG:156, ARG:224, ARG:292, ARG:371, ASN:294, ASP:151, GLU:119, GLU:227, GLU:276, GLU:277, GLY:348, ILE:222, SER:179, SER:246, TRP:178, TYR:347, and TYR:406. The homologous residues from the F2YI87 sequence used to create our NA models are ARG:118, ARG:152, ARG:156, ARG:225, ARG:293, ARG:368, ASN:295, ASP:151, GLU:119, GLU:228, GLU:277, GLU:278, GLY:345, ILE:223, SER:180, SER:247, TRP:179, ASN:344, and TYR:402. Eighteen of these 19 residues are conserved between the two sequences. We used the PCA function from the sklearn.decomposition Python library[45]. Each PCA histogram in **Figure 1E** was independently normalized so that the bins containing the minimum and maximum number of 2D PCA points were blue and red, respectively. A k-means clustering performed PCA on 155,280 regularly spaced, active-site-aligned monomeric conformations extracted from the viral envelope simulation. To approximately represent each conformation, we assigned each to a vector comprised of the first and second principal components. We normalized this set of vectors by dividing each principal component by its standard deviation across all conformations (using scipy.cluster.vq.whiten). We then used K-means clustering (scipy.cluster.vq.kmeans2) to group the vectors into 15 clusters. To focus on the most representative conformations, we discarded all but the five largest clusters. For each of the remaining five clusters, we identified the conformation with the PCA vector closest the cluster centroid. These conformations are shown in **Figure S10**.

1.8 Volumetric and FTMap analysis

After aligning all 120 NA monomeric trajectories by the active-site C_α, we used POVME 2.0[46] to measure the pocket shapes and volumes across all trajectories (grid spacing 2.0 Å). A volumetric analysis of the NA catalytic site was performed over ~150,000 frames, which were subsequently binned into the histogram shown in **Figure 2** (main text). **Table S1** reports all calculated volume intervals with a resolution of 100 Å^3 (starting from 0-100 Å^3) as well as the number of frames populating each interval. Note that in **Figure 2** we omitted the volume intervals larger than 3400 Å^3 for the sake of clarity.

Volume (Å^3)	**# FRAMES**
100	0
200	0
300	0
400	0
500	1
600	32
700	447
800	2267
900	3947
1000	6152
1100	7477
1200	10730
1300	12481
1400	15251
1500	14677
1600	15292
1700	12456
1800	11432
1900	9275
2000	7904
2100	5773
2200	4801
2300	3142
2400	2473
2500	1509
2600	1049
2700	643
2800	479
2900	263
3000	204
3100	137
3200	76
3300	52
3400	42
3500	24
3600	21
3700	6
3800	8
3900	3
4000	0
4100	2
4200	0
4300	0
4400	0
4500	1

Table S1. The NA catalytic-site volumes calculated using POVME are divided into intervals with a resolution of 100 Å^3 (left column). The number of frames populating each volume interval is reported in the right column.

The frame with the largest G351-adjacent pocket was fed into FTMap to predict druggable hotspots[47], as described in ref. 1. As a control, we performed the same FTMap analysis using an N1 crystal structure that was crystallized in complex with oseltamivir carboxylate (PDB ID: 2HU4)[44], but with the oseltamivir carboxylate molecule removed. This control showed that FTMap is able to identify the oseltamivir druggable hotspot (data not shown).

1.9 Markov state models

Markov state models (MSM) of the 150-cavity were constructed from both the viral envelope simulations and the isolated NA simulations with trajectory frames taken every 0.02 ns and 0.05 ns, respectively. To describe the motion of the 150-loop, the minimum distance between I149 and P431 was selected as the input feature for model construction. Trajectory frames were clustered into 300 microstates by k-means clustering, as implemented in the PyEMMA software package[48]. Based on the implied timescale plots (**Figure S11**), a lag time of 10 ns was selected for model construction, and the resulting models were validated by the Chapman-Kolmogorov (CK) test (**Figure S12**). Both the implied timescale plots as well as the CK tests give a sense of the 'correctness' and convergence of the MSM. Subsequently, the resulting MSMs were coarse grained using PCCA++ in conjunction with hidden Markov state models (HMMs). This procedure is a standard approach in the MSM field. For these HMMs, a lag time of 2 ns was selected based on the implied timescale plots (**Figure S13**). The final models were again validated by the CK test (**Figure S14**). Confidence intervals were calculated using Bayesian hidden Markov state models (BHMMs) corresponding to the described HMMs. The Jupyter notebook used for model construction is provided as Supporting Information.

1.10 Chloride anion and electrostatic analyses

To calculate the chloride anion distributions shown in **Figure 4**, we concatenated all catalytic-site-aligned NA monomers, together with nearby chloride ions, into one trajectory. A box with dimensions 100 Å x 100 Å x 100 Å, centered on the catalytic sialic-acid binding site, was divided into 3,375,000 voxels (0.67 Å x 0.67 Å x 0.67 Å each). We then tallied the number of chloride anions falling within each voxel, regardless of the associated monomer or frame. To identify regions that favor chloride occupancy, we applied a high-pass filter, retaining only voxels with chloride counts greater than three standard deviations above the mean (**Figure 4)**.

To further investigate the role of electrostatics in our proposed "bind and transfer" mechanism, we performed electrostatic calculations using the Adaptive Poisson-Boltzmann Solver (APBS1.4) software[49]. APBS evaluates the electrostatic properties of biomolecules by solving the Poisson-Boltzmann electrostatic equation (PBE)[49]. We used APBS to evaluate the conformations of all NA monomers present at the end of the third daughter simulation, as well as representative "open" and "closed" structures identified from the MSM (**Figure S15**). We first used PDB2PQR[3] with unvaried protonation states (CHARMM36 all-atom additive force field) to prepare the monomer structures for APBS. APBS calculations were then carried out using the Linearized Poisson-Boltzmann Equation (LPBE) with cubic B-spline discretization to map atomic partial charges to grid points. Grid spacing was automatically detected according to the system and set to ~0.7 Å. We fixed the temperature at 298.15 K and used 150 mM ionic strength for monovalent ions. The external dielectric constant was set to 78.54 to mimic the aqueous medium, and the solute dielectric constant was tuned to 2.0.

Among the extracted NA monomers, one with the 150-loop in a semi-open state was selected to showcase the projection of the electrostatic potential (-1 k_bT/e_c to +1 k_bT/e_c) onto the protein surface (**Figure 4B**). At 300 K, 1 k_bT/e_c corresponds to ~25.8 mV. The projection of the electrostatic potential onto the "open" and "closed" NA structures is shown in **Figure S15**. Interestingly, when the 150-loop is in a closed state, fewer positively charged regions are exposed, possibly limiting the transferring of sialic acid moieties from the secondary site to the catalytic site.

NOTE: As this study was entirely computational in nature, no unexpected or unusually high safety hazards were encountered.

2. References

1. Durrant, J. D., Bush, R. M. & Amaro, R. E. Microsecond Molecular Dynamics Simulations of Influenza Neuraminidase Suggest a Mechanism for the Increased Virulence of Stalk-Deletion Mutants. *J. Phys. Chem. B* **120**, 8590–8599 (2016).
2. Liu, J. *et al.* Structures of receptor complexes formed by hemagglutinins from the Asian Influenza pandemic of 1957. *Proc Natl Acad Sci U S A* **106**, 17175–17180 (2009).
3. Dolinsky, T. J. *et al.* PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* **35**, W522–W525 (2007).
4. Olsson, M. H., Sondergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J Chem Theory Comput* **7**, 525–537 (2011).
5. Sondergaard, C. R., Olsson, M. H., Rostkowski, M. & Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *J Chem Theory Comput* **7**, 2284–2295 (2011).
6. Li, H., Robertson, A. D. & Jensen, J. H. Very fast empirical prediction and rationalization of protein pKa values. *Proteins Struct. Funct. Bioinforma.* **61**, 704–721 (2005).
7. Bas, D. C., Rogers, D. M. & Jensen, J. H. Very fast prediction and rationalization of pKa values for protein–ligand complexes. *Proteins Struct. Funct. Bioinforma.* **73**, 765–783 (2008).
8. Hartmann, M. D. *et al.* A coiled-coil motif that sequesters ions to the hydrophobic core. *Proc Natl Acad Sci U S A* **106**, 16950–16955 (2009).
9. Humphrey, W., Dalke, A. & Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).
10. Jo, S., Kim, T. & Im, W. Automated Builder and Database of Protein/Membrane Complexes for Molecular Dynamics Simulations. *PLoS One* **2**, e880 (2007).
11. Jo, S., Lim, J. B., Klauda, J. B. & Im, W. CHARMM-GUI membrane builder for mixed bilayers and its application to yeast membranes. *Biophys. J.* **97**, 50–58 (2009).
12. Gerl, M. J. *et al.* Quantitative analysis of the lipidomes of the influenza virus envelope and MDCK cell apical membrane. *J. Cell Biol.* **196**, 213–221 (2012).
13. Klauda, J. B. *et al.* Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *J. Phys. Chem. B* **114**, 7830–7843 (2010).
14. Klauda, J. B., Monje, V., Kim, T. & Im, W. Improving the CHARMM Force Field for

Polyunsaturated Fatty Acid Chains. *J. Phys. Chem. B* **116**, 9424–9431 (2012).

15. Durrant, J. D. & McCammon, J. A. Autoclickchem: Click chemistry in silico. *PLoS Comput. Biol.* **8**, e1002397–e1002397 (2012).
16. Ropp, P., Friedman, A. & Durrant, J. D. Scoria: a Python module for manipulating 3D molecular data. *J. Cheminform.* **9**, 52 (2017).
17. Stone, J. E. *et al.* Accelerating molecular modeling applications with graphics processors. *J. Comput. Chem.* **28**, 2618–2640 (2007).
18. Case, D. A. *et al.* The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688 (2005).
19. MacKerell, A. & Bashford, D. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. ...* **5647**, 3586–3616 (1998).
20. Mackerell, A. D., Feig, M. & Brooks, C. L. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* **25**, 1400–1415 (2004).
21. Huang, J. & MacKerell, A. D. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* **34**, 2135–2145 (2013).
22. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926 (1983).
23. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **26**, 1781–1802 (2005).
24. Amaro, R. E. *et al.* A Computational Assay that Explores the Hemagglutinin/Neuraminidase Functional Balance Reveals the Neuraminidase Secondary Site as a Novel Anti-Influenza Target. *ACS Cent. Sci.* **4**, 1570–1577 (2018).
25. Harris, A. *et al.* Influenza virus pleiomorphy characterized by cryoelectron tomography. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 19123–19127 (2006).
26. Daura, X. *et al.* Peptide folding: When simulation meets experiment. *Angew. Chemie-International Ed.* **38**, 236–240 (1999).
27. Christen, M. *et al.* The GROMOS software for biomolecular simulation: GROMOS05. *J. Comput. Chem.* **26**, 1719–1751 (2005).
28. Sharma, M. *et al.* Insight into the Mechanism of the Influenza A Proton Channel from a Structure in a Lipid Bilayer. *Science (80-.).* **330**, 509–512 (2010).
29. Durrant, J. D. & Amaro, R. E. LipidWrapper: an algorithm for generating large-scale membrane models of arbitrary geometry. *PLoS Comput. Biol.* **10**, e1003720 (2014).
30. Beglov, D. & Roux, B. Finite representation of an infinite bulk system: Solvent boundary potential for computer simulations. *J. Chem. Phys.* **100**, 9050–9063 (1994).
31. Turq, P., Lantelme, F. & Friedman, H. L. Brownian dynamics: Its application to ionic solutions. *J. Chem. Phys.* **66**, 3039–3044 (1977).
32. Martyna, G. J., Tobias, D. J. & Klein, M. L. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **101**, 4177–4189 (1994).
33. Feller, S. E., Zhang, Y., Pastor, R. W. & Brooks, B. R. Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.* **103**, 4613–4621 (1995).
34. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N ·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).

35. Ryckaert, J.-P., Ciccotti, G. & Berendsen, H. J. . Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **23**, 327–341 (1977).
36. Li, Q. *et al.* The 2009 pandemic H1N1 neuraminidase N1 lacks the 150-cavity in its active site. *Nat. Struct. Mol. Biol.* **17**, 1266–1268 (2010).
37. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **32**, 2319–2327 (2011).
38. Gowers, R. *et al.* MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. in 98–105 (2016). doi:10.25080/Majora-629e541a-00e
39. Chavent, M. *et al.* Methodologies for the analysis of instantaneous lipid diffusion in md simulations of large membrane systems. *Faraday Discuss.* **169**, 455–475 (2014).
40. Ramachandran, P. & Varoquaux, G. Mayavi: 3D Visualization of Scientific Data. *Comput. Sci. Eng.* **13**, 40–51 (2011).
41. Metzler, R., Jeon, J.-H. & Cherstvy, A. G. Non-Brownian diffusion in lipid membranes: Experiments and simulations. *Biochim. Biophys. Acta - Biomembr.* **1858**, 2451–2467 (2016).
42. Moradi, S., Nowroozi, A. & Shahlaei, M. Shedding light on the structural properties of lipid bilayers using molecular dynamics simulation: a review study. *RSC Adv.* **9**, 4644–4658 (2019).
43. Hong, C., Tieleman, D. P. & Wang, Y. Microsecond Molecular Dynamics Simulations of Lipid Mixing. *Langmuir* **30**, 11993–12001 (2014).
44. Russell, R. J. *et al.* The structure of H5N1 avian influenza neuraminidase suggests new opportunities for drug design. *Nature* **443**, 45–49 (2006).
45. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* 2825–2830 (2012).
46. Durrant, J. D., Votapka, L., Sørensen, J. & Amaro, R. E. POVME 2.0: An Enhanced Tool for Determining Pocket Shape and Volume Characteristics. *J. Chem. Theory Comput.* **10**, 5047–5056 (2014).
47. Ngan, C. H. *et al.* FTMAP: extended protein mapping with user-selected probe molecules. *Nucleic Acids Res* **40**, W271-5 (2012).
48. Scherer, M. K. *et al.* PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J. Chem. Theory Comput.* **11**, 5525–5542 (2015).
49. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 10037–41 (2001).
50. Xu, R. *et al.* Structural Basis of Preexisting Immunity to the 2009 H1N1 Pandemic Influenza Virus. *Science (80-.).* **328**, 357–360 (2010).
51. Corti, D. *et al.* A Neutralizing Antibody Selected from Plasma Cells That Binds to Group 1 and Group 2 Influenza A Hemagglutinins. *Science (80-.).* **333**, 850–856 (2011).
52. Xu, R. *et al.* A recurring motif for antibody recognition of the receptor- binding site of influenza hemagglutinin. *Nat. Struct. Mol. Biol.* **20**, 363–370 (2013).
53. Dreyfus, C. *et al.* Highly Conserved Protective Epitopes on Influenza B Viruses. *Science (80-.).* **337**, 1343–1348 (2012).
54. Ekiert, D. C. *et al.* Antibody Recognition of a Highly Conserved Influenza Virus Epitope. *Science (80-.).* **324**, 246–251 (2009).

55. Kashyap, A. K. *et al.* Protection from the 2009 H1N1 Pandemic Influenza by an Antibody from Combinatorial Survivor-Based Libraries. *PLoS Pathog.* **6**, e1000990 (2010).
56. Sui, J. *et al.* Structural and functional bases for broad-spectrum neutralization of avian and human influenza A viruses. *Nat. Struct. Mol. Biol.* **16**, 265–273 (2009).
57. Ekiert, D. C. *et al.* A Highly Conserved Neutralizing Epitope on Group 2 Influenza A Viruses. *Science (80-.).* **333**, 843–850 (2011).

3. Supplementary Figures

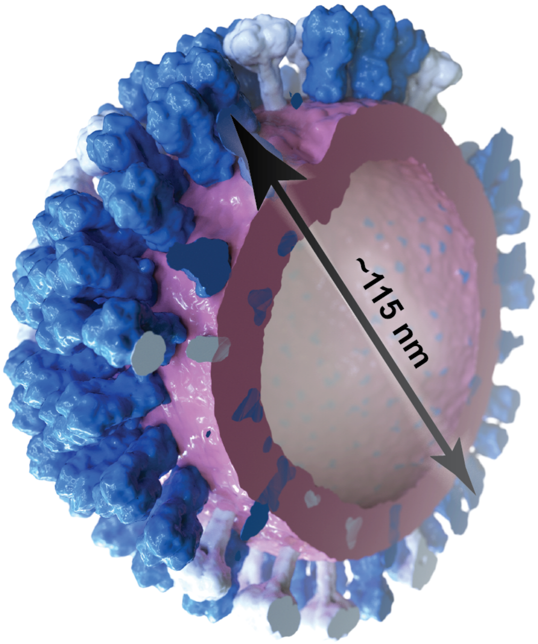

Figure S1. Cross section of the influenza viral envelope model. The phospholipid bilayer is colored in pink, and the two glycoproteins HA and NA are colored in blue and light gray, respectively.

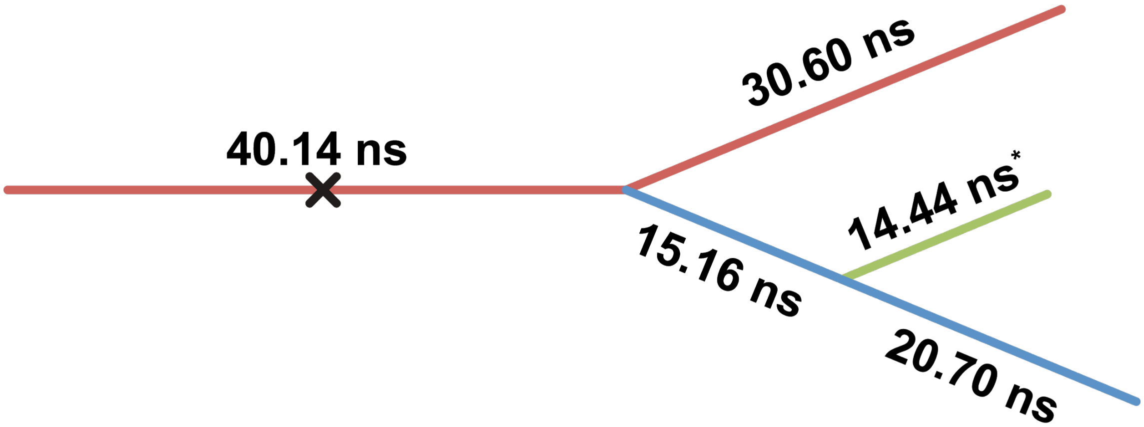

Figure S2. Simulation workflow. After 40.14 ns, the productive simulation was forked into two daughter simulations that continued for 30.60 and 15.16 ns, respectively. The second daughter simulation was also forked into simulations that lasted 20.70 and 14.44 ns, respectively. A single hole developed in the virion bilayer after roughly 20 ns, at the approximate time marked with a black X. This hole comprised roughly 0.2% of the entire viral envelope surface area and is thus expected to have a minimal impact on overall envelope dynamics. The hole was repaired prior to launching the 14.44 ns simulation (marked with an asterisk). Taken together, these simulations constitute 121.04 ns total.

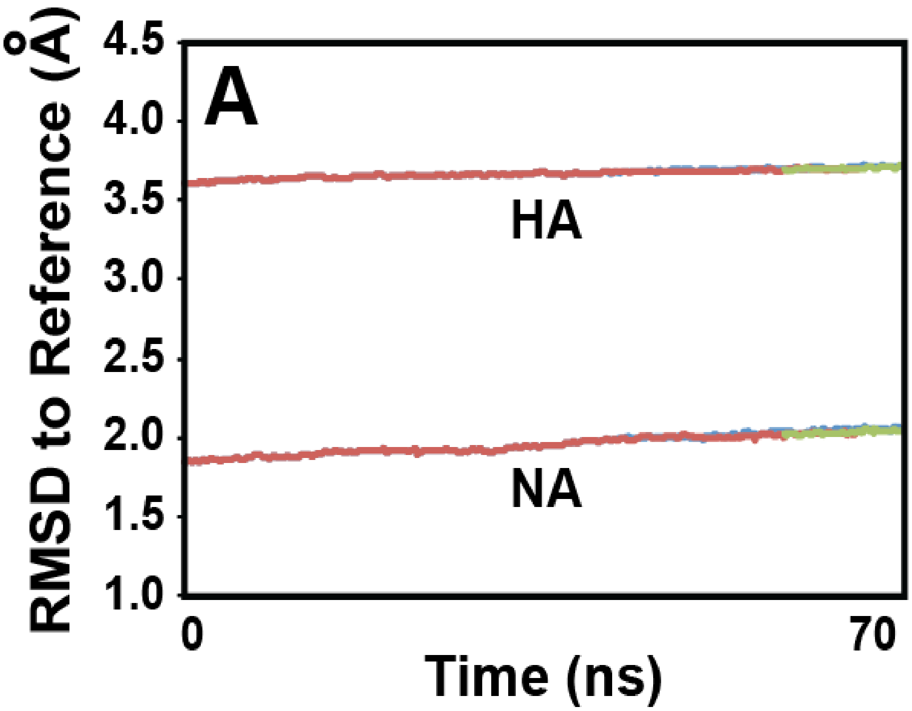

Figure S3. Average RMSD of HA and NA glycoproteins (whole viral envelope simulation). The average C$_\alpha$ RMSD from a reference structure to each glycoprotein monomer, per frame. The model included 120 and 708 monomers of NA and HA, respectively. The NA and HA reference structures were the 3NSS crystal structure[36] and a 3LZG-based model[50], respectively. The colors of the (sometimes overlapping) lines correspond to the branches of the simulations described in Figure S2 (red, blue, and green, respectively).

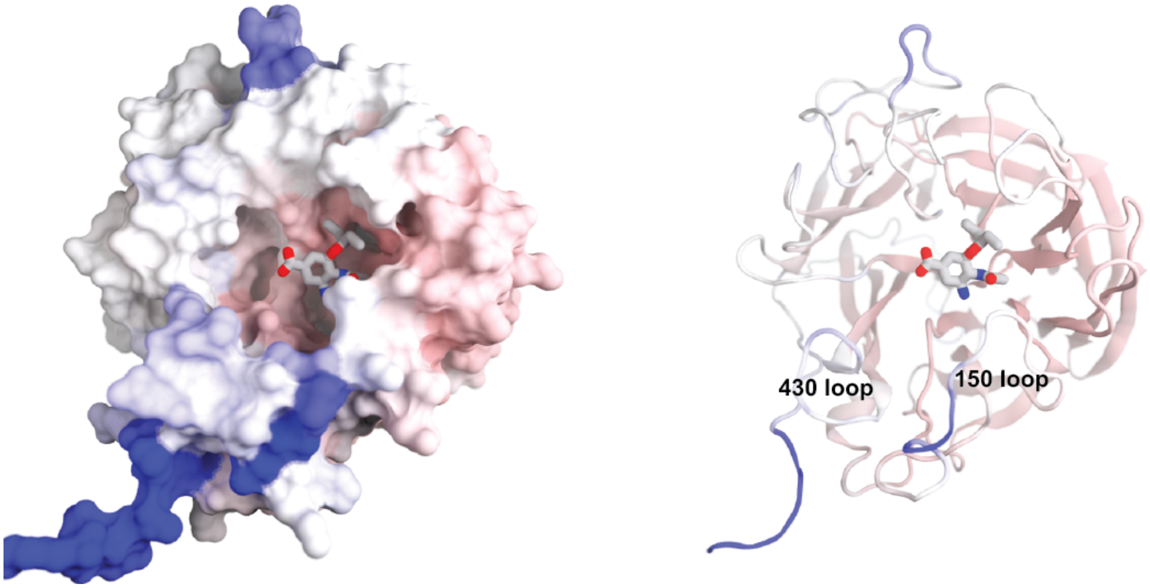

Figure S4. RMSF analysis of NA (whole viral envelope simulation). RMSF values ranging from 0 (red) to ≥3 (blue) are visualized on a structure selected from the simulation. An oseltamivir molecule taken from 2HU4[44] was superimposed for reference, though no such molecule was present in the simulations.

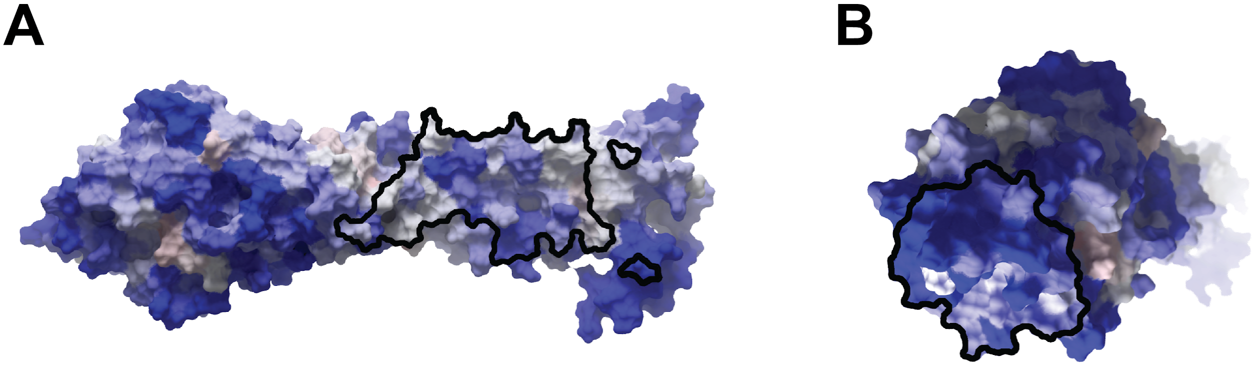

Figure S5. RMSF analysis of HA (whole viral envelope simulation). RMSF values ranging from 0 (red) to ≥5 (blue) are visualized on an HA structure selected from the simulation. The "footprint" of selected HA-binding antibodies is shown as a black outline[51–57]. A) The binding site of broadly neutralizing, stem-binding antibodies. B) The binding site of head-binding antibodies.

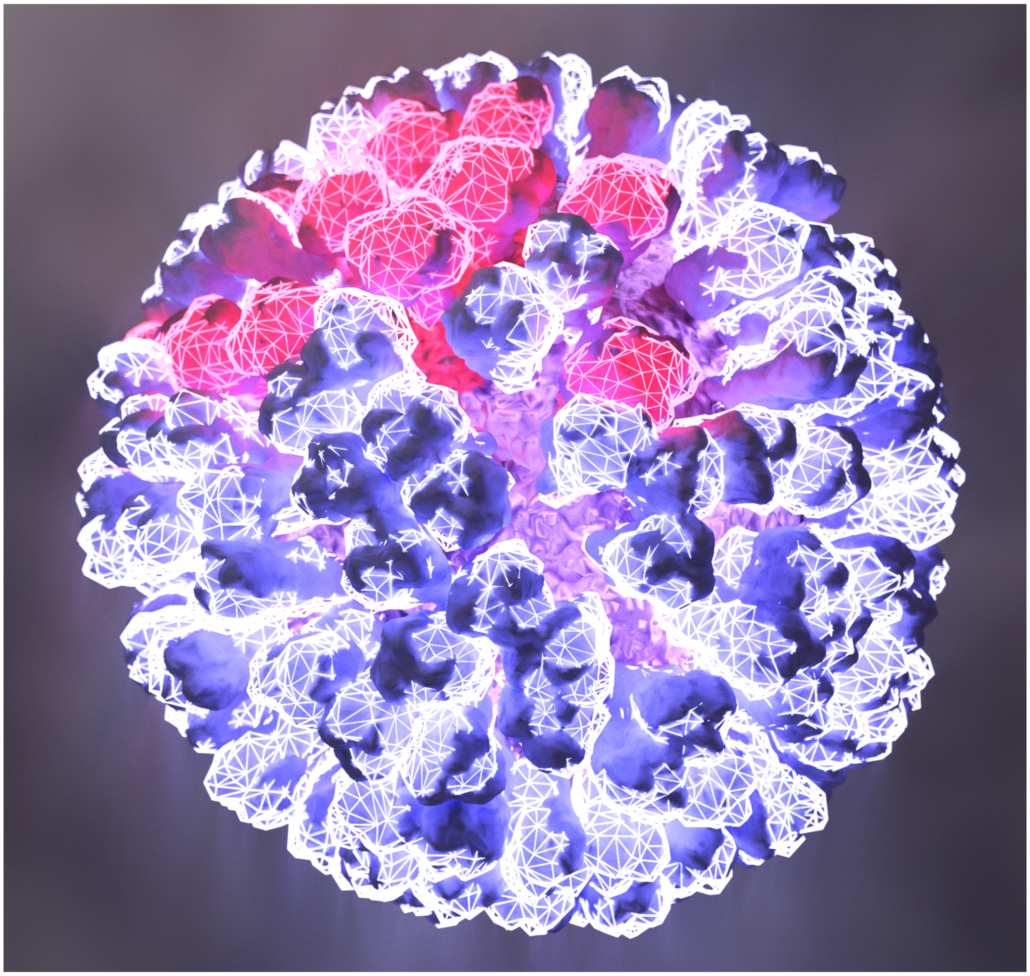

Figure S6. The first and last frames of the viral envelope simulation. NA and HA are shown in red and blue, respectively. The glycoprotein conformations of the first and last frames are shown in solid and glowing-mesh representations, respectively. The lipid bilayer was taken from the first frame. The first-frame/last-frame RMSD between all glycoprotein C_α was 13.18 Å.

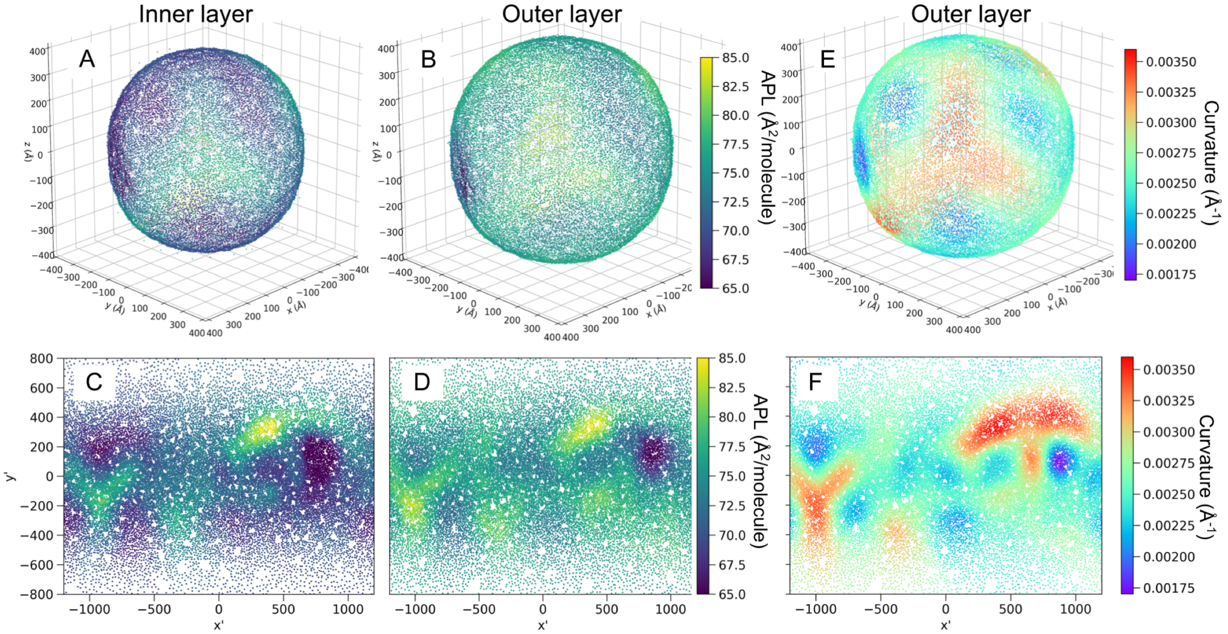

Figure S7. 3D and Mercator projection plots of area-per-lipid and lipid-curvature values. Plots of area-per-lipid (A-D) and curvature (E-F) values associated with the phospholipids at time *t* = 70 ns. The phospholipid headgroups are represented via 3D plot (A-B) or Mercator projection (C-D). The left and right columns show the lipids from the inner and outer layers of the phospholipid bilayer, respectively. E-F) Curvature values associated with the outer layer are shown via 3D plot and Mercator projection, respectively.

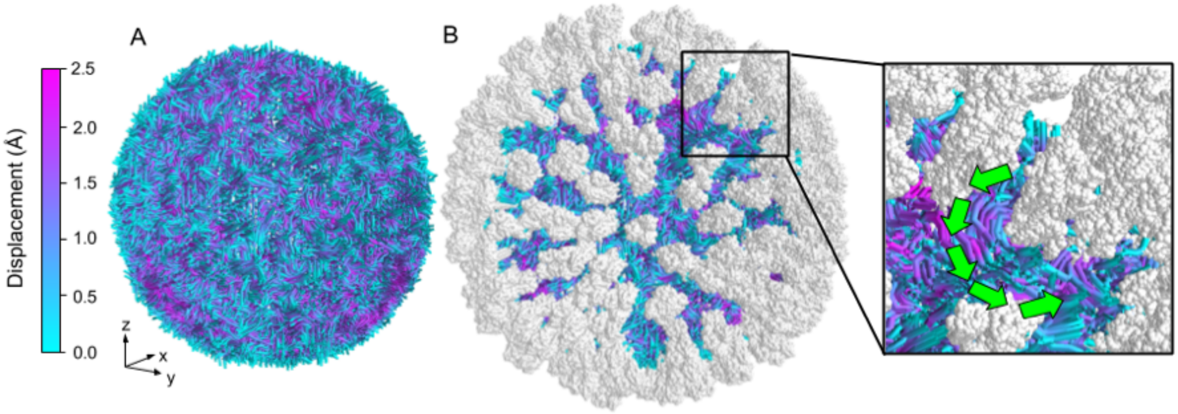

Figure S8. Streamlines of surface lipid motions after 70 ns of simulation. Lipid streamlines from the outer leaflet are plotted with (B) and without (A) embedded surface proteins, colored by displacement (Å). The magnification in B shows a lipid vortex around a surface protein. The streamlines were calculated using a 4 ns timestep after 70 ns of simulation.

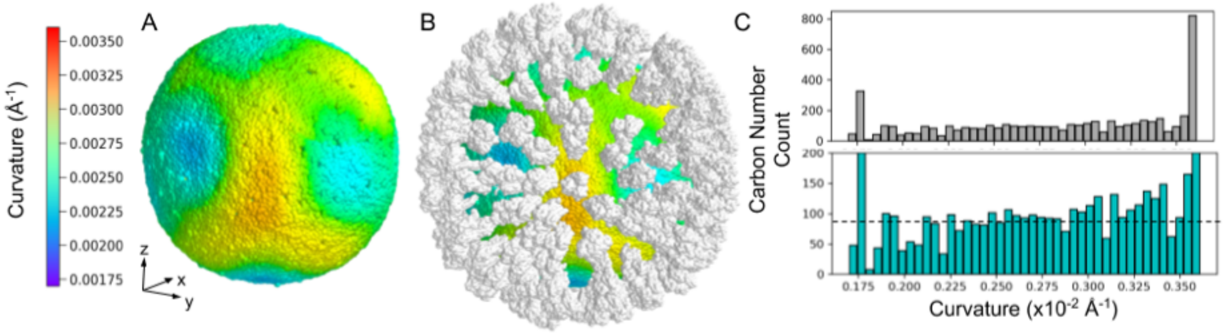

Figure S9. **Curvature and glycoprotein distribution analyses.** A) Lipid headgroups of the outer leaflet colored by associated curvature values, after 70 ns of simulation. B) Added hemagglutinin and neuraminidase proteins (white). C) Grey: distribution of protein C_α across curvature values, normalized by the curvature/surface area. Cyan: protein distribution slightly skews towards areas of higher curvature (note scale difference). Dashed line indicates carbon number counts for a theoretically even protein distribution across curvature values.

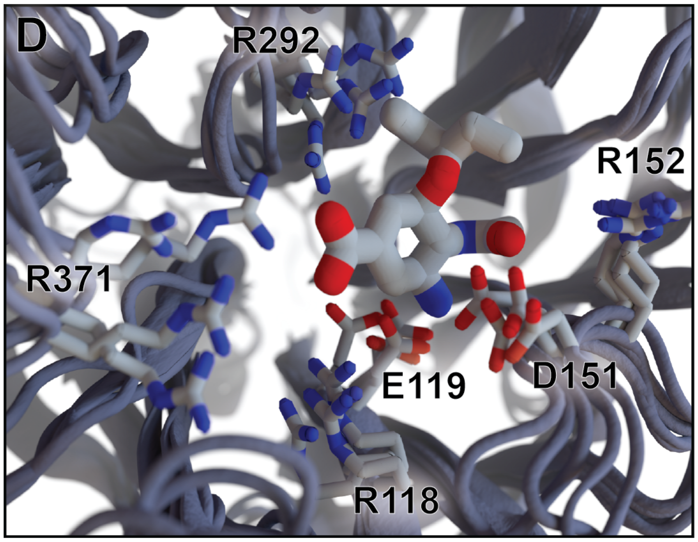

Figure S10. An illustration of four representative catalytic-pocket conformations taken from the viral envelope simulation. Key residues are shown as licorice. An oseltamivir molecule taken from 2HU4 has been superimposed on the structure for reference but was not included in the simulations.

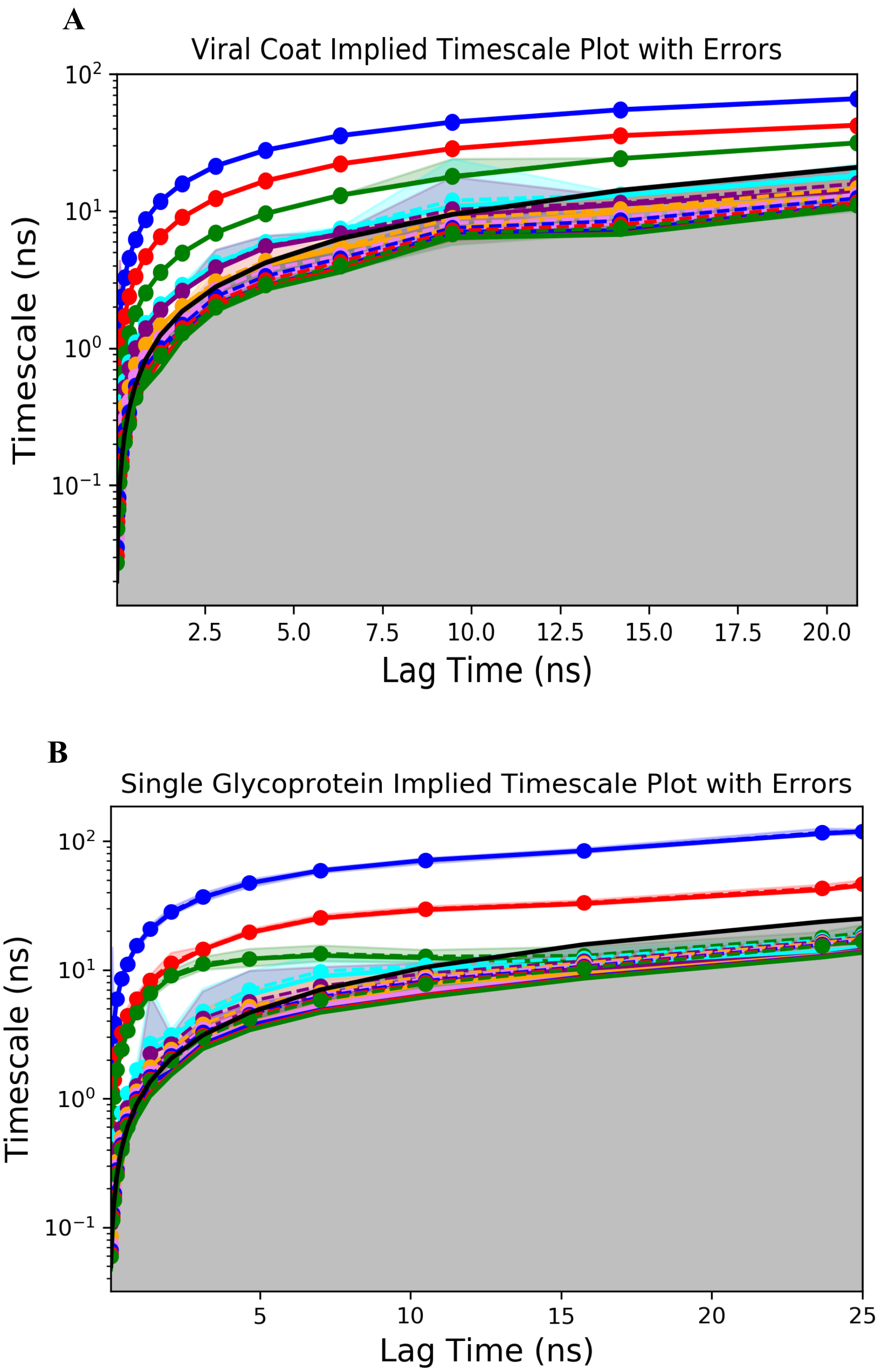

Figure S11. Implied time scale plots with errors, for the (A) neuraminidase glycoprotein 150-loop MSM taken from the viral envelope simulations and (B) single-glycoprotein neuraminidase 150-loop MSMs. Each colored line represents the timescales of different dynamical processes (motions) identified by the decomposition of the transition matrix (eigenvalues). If the model is Markovian (at the chosen lag time) then the timescales will be constant for all longer lag times that are also short enough to resolve the process. Bayesian errors are indicated as similarly colored shaded areas.

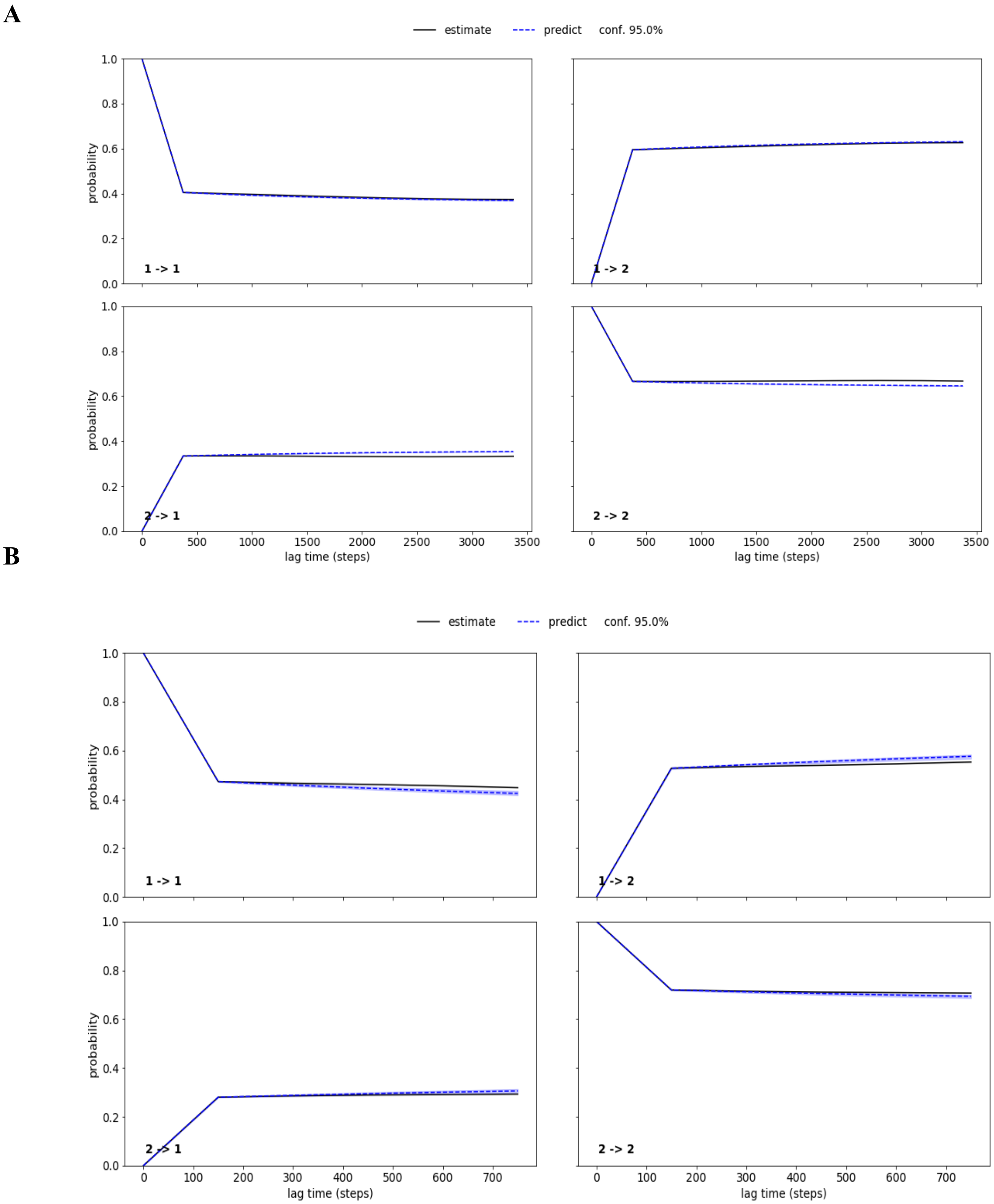

Figure S12. Chapman-Kolmogorov test, for the (**A**) viral envelope and (**B**) isolated-NA MSMs. Plots indicate convergence of the presented models.

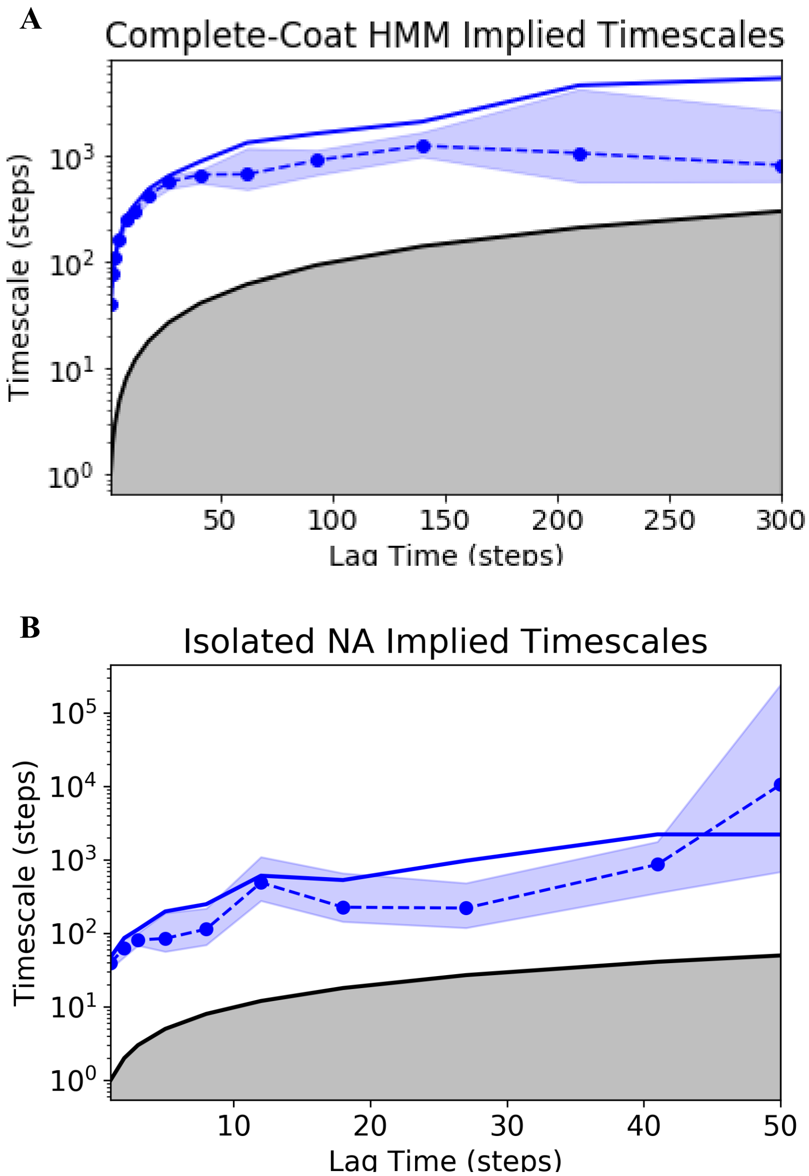

Figure S13. Implied time scale plots with errors, for the HMMs from the (**A**) viral envelope and (**B**) single-glycoprotein simulations. Bayesian errors are indicated as shaded areas. Each blue line represents the timescales of dynamical processes (motions) identified by the decomposition of the transition matrix (eigenvalues). Bayesian errors are indicated as similarly colored shaded areas.

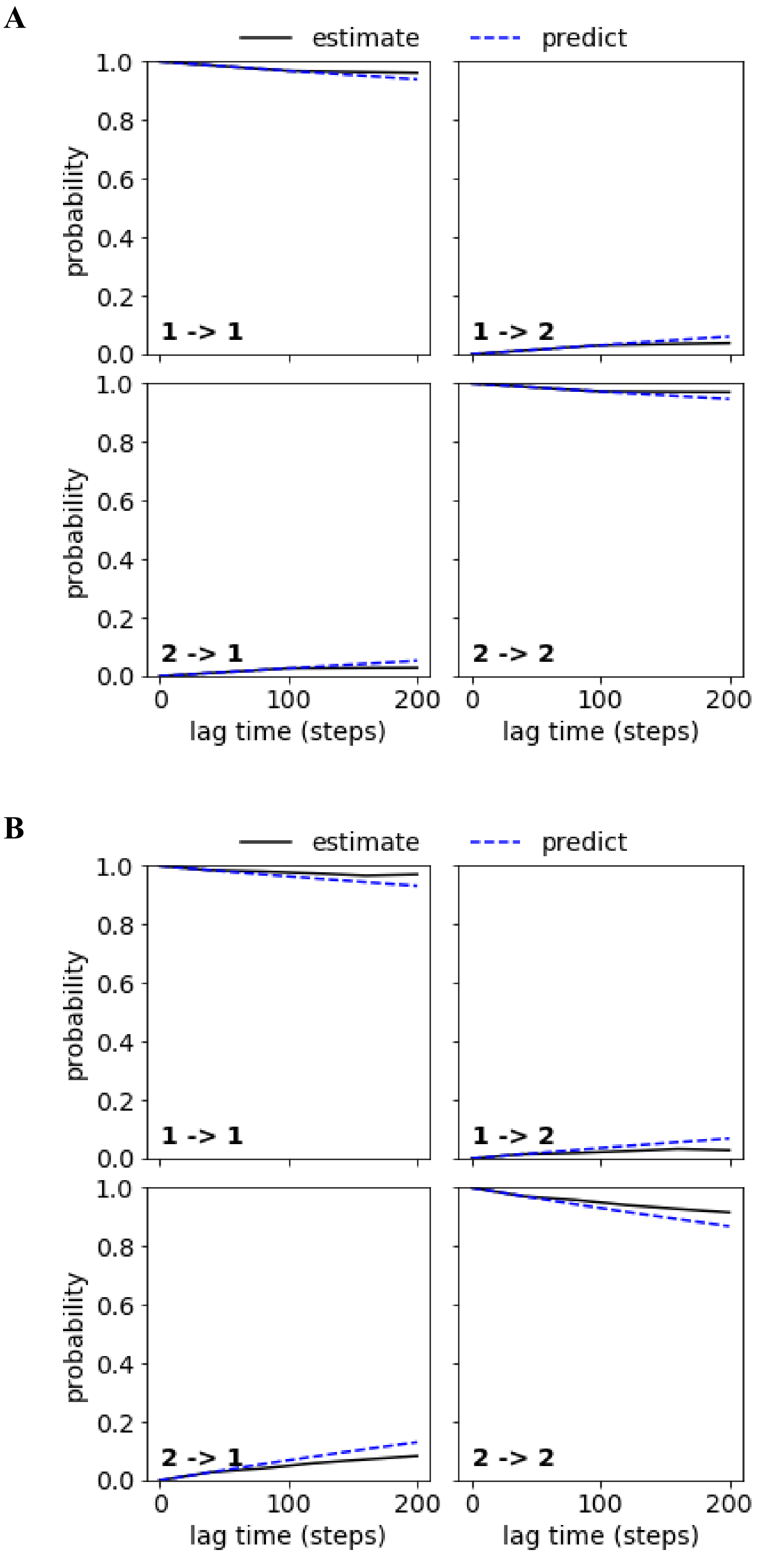

Figure S14. Chapman-Kolmogorov test, for the (A) viral envelope and (B) isolated-NA HMMs. Plots indicate convergence of the presented models.

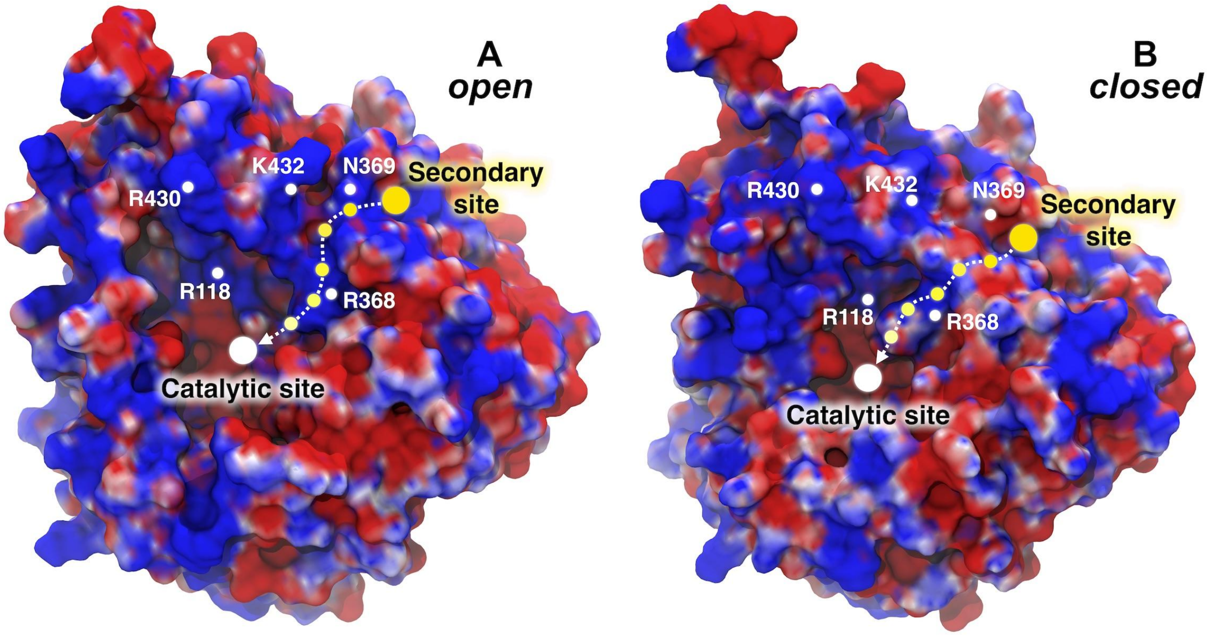

Figure S15. Electrostatic potential of NA monomer with 150-loop in "open" and "closed" states. The electrostatic potential projected onto the NA surface with the 150-loop in an open (A) and closed (B) state. The surface is colored from red (negative) to blue (positive), representing electrostatic potential values of -1 k_bT/e_c to +1 k_bT/e_c. Residues responsible of positively charged regions connecting the two binding sites are highlighted with white dots and numbered using the N2 numbering scheme. The path connecting the secondary site with the catalytic site is shown with a dashed arrow. The sequence of circles fades from yellow (secondary site) to white (catalytic site).