JVI Journals.ASM.org

Structural Characterization of the Hemagglutinin Receptor Specificity from the 2009 H1N1 Influenza Pandemic

Rui Xu,[a] Ryan McBride,[b] Corwin M. Nycholat,[b] James C. Paulson,[a,b] and Ian A. Wilson[a,c]

Department of Molecular Biology[a] and Department of Chemical Physiology,[b] Skaggs Institute for Chemical Biology,[c] The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, California 92037, USA

Influenza virus hemagglutinin (HA) is the viral envelope protein that mediates viral attachment to host cells and elicits membrane fusion. The HA receptor-binding specificity is a key determinant for the host range and transmissibility of influenza viruses. In human pandemics of the 20th century, the HA normally has acquired specificity for human-like receptors before widespread infection. Crystal structures of the H1 HA from the 2009 human pandemic (A/California/04/2009 [CA04]) in complex with human and avian receptor analogs reveal conserved recognition of the terminal sialic acid of the glycan ligands. However, favorable interactions beyond the sialic acid are found only for α2-6-linked glycans and are mediated by Asp190 and Asp225, which hydrogen bond with Gal-2 and GlcNAc-3. For α2-3-linked glycan receptors, no specific interactions beyond the terminal sialic acid are observed. Our structural and glycan microarray analyses, in the context of other high-resolution HA structures with α2-6- and α2-3-linked glycans, now elucidate the structural basis of receptor-binding specificity for H1 HAs in human and avian viruses and provide a structural explanation for the preference for α2-6 siaylated glycan receptors for the 2009 pandemic swine flu virus.

The 2009 H1N1 influenza A virus pandemic was the first human pandemic in the last four decades. From the time that the first human cases were confirmed in April 2009 (2), the virus spread to 214 countries and caused more than 18,449 human deaths globally. The overall mortality rate for this pandemic virus was not much different than that for seasonal flu, but it did have much more adverse effects on children and young adults (http://www.who.int/csr/don/2010_08_06/en/index.html). The efficient human-to-human transmission of the 2009 pandemic virus (18, 33, 35, 36) suggested that the virus was well adapted for human infection.

Hemagglutinin (HA) is the major surface envelope protein of influenza virus and carries out the crucial viral functions of host recognition and membrane fusion (41). The receptor-binding specificity of HA is strongly linked to the host range of influenza viruses (30). Avian and human viruses preferentially bind to terminal sialosides that differ mainly in the linkages that sialic acid makes with the rest of the carbohydrate receptor. Otherwise, human and avian receptors have similar glycan compositions. The HA of human viruses recognizes sialic acid-containing sugars that have an α2-6 linkage to galactose (Gal). Such α2-6-linked glycans are found on the epithelial cell surface of the human upper respiratory tract, the primary site for viral infection. The HAs from avian viruses have specificity for α2-3-linked sialylated glycans, which are localized on human lung alveolar cells and on the digestive and respiratory tracts of birds.

In the three human pandemics of the last century, the HAs of avian origin likely had to switch their binding specificity toward human-like, α2-6-linked glycans before human infection became widespread. This shift in binding specificity was achieved by the mutation of as few as two key residues in the receptor-binding site: E190D/G225D for the H1 subtype and Q226L/G228S for the H2 and H3 subtypes (H3 numbering) (14, 31, 45). Structural studies of these HAs have revealed the molecular basis of this specificity switch in the H2 and H3 subtypes of influenza A viruses (16, 26, 48). In each case, the mutations eliminate favorable interactions between the HA and α2-3-linked glycans while enhancing interactions with glycans that contain a terminal sialic acid with an α2-6 linkage. As a result, most HAs from human viruses display highly specific binding toward α2-6-linked glycans (44).

The HA of the 2009 human pandemic is of the H1 subtype and a distant relative of previously studied human H1 HAs (20, 42). HAs from viral strains isolated during the pandemic generally contain Asp190 and Asp225, implying a binding preference for human-like α2-6 sialylated glycans (34). However, computational modeling of HA in complex with glycan receptor analogs (43) and neoglycolipid-based carbohydrate microarray analysis of whole viruses (11) suggested the dual specificity of 2009 pandemic viruses, with considerable binding reported to α2-3 sialic acids in addition to α2-6-linked glycans. The supposedly broader specificity of the 2009 pandemic virus seemed to correlate with its efficient replication in the lungs of experimental animals (22) but is in stark contrast to the glycan-binding results that we report here, as well as those from other groups (7, 10, 29, 49). Studies using isolated viruses (7, 10), as well as recombinant HAs (29, 49), have showed a strict preference for human-like α2-6-linked sialic acids on the glycan microarrays produced by the Consortium for Functional Glycomics (www.functionalglycomics.org). An understanding of such discrepancies is critical for assessing the role of HA receptor specificity in the emergence of new pandemic viruses and in viral pathogenesis. Adding to the complexity of the receptor specificity of the 2009 pandemic viruses is the significant presence of a D225G mutation. More than 1% of all HA sequences from the

Received 15 September 2011 **Accepted** 31 October 2011
Published ahead of print 9 November 2011
Address correspondence to Ian A. Wilson, wilson@scripps.edu.
This is publication 21056 from The Scripps Research Institute.

doi:10.1128/JVI.06322-11

pandemic had the avian-like Gly225 instead of the human signature Asp225 in H1 viruses (12). The D225G mutant was associated with a number of severe human infections and was transmitted locally in the human population (9, 23, 38). In the HA of the 1918 influenza virus, the D225G mutation significantly reduces avidity for α2-6-linked glycans and improves binding for α2-3-linked glycans (44), resulting in impaired transmission in ferret models (46). For the 2009 pandemic virus, the impact on glycan binding and transmission in ferrets and guinea pigs appears to be less dramatic (12, 49). Thus, a structural explanation is important to understand the effect of these HA mutations on receptor binding and transmission and to help predict viral behavior in humans.

To examine the HA receptor specificity of the 2009 pandemic at the atomic level, we determined crystal structures of HA from a 2009 pandemic strain (A/California/04/2009 [CA04]) in complex with receptor analogs at high resolution (2.0 to 2.5 Å). The crystal structures show the relatively conserved protein-glycan interactions around the terminal sialic acid for both α2-3- and α2-6-linked glycans. However, beyond the terminal sialic acid, the human-like receptor analogs are able to engage in multiple hydrogen bonds with the HA that stabilize HA-receptor association, whereas these interactions are absent from the avian-like receptor analog complexes. Comparison with previous HA complex structures thereby elucidates the structural basis of receptor specificity in the 2009 pandemic H1 HAs.

MATERIALS AND METHODS

The crystal structures were determined using a stabilizing mutant of CA04. The mutations G205C/R220C in the CA04 HA were introduced by the polymerase incomplete primer extension cloning method (24). The mutant was expressed and crystallized as described for wild-type HA (47). Crystals were soaked in reservoir solution plus 10% polyethylene glycol 400 (PEG400) and 10 mM glycan ligand for 10 min and flash-cooled in liquid nitrogen. Data sets were collected at the Advanced Photon Source and processed with HKL2000 (37). The structures were solved with molecular replacement by Phaser (32) using the coordinates from the wild-type CA04 HA trimer (Protein Data Bank [PDB] code 3LZG). The structures then were adjusted using COOT (17) and refined with PHENIX (1). Statistics for data collection and structure refinement are summarized in Table 1.

Protocols for the glycan microarray analysis of recombinant HA were as previously described (6, 48). Glycan arrays were custom printed on a MicroGridII (Digilab) contact microarray robot equipped with Stealth4B microarray pins (Telechem), and the glycans present on the array are listed in Table 2. Compound 1 was prepared as previously described (5). Glycans 2, 24, and 33 to 38 were gifts provided by Otsuka Chemical Co., Ltd. Glycans 3, 4, 6, 7, 20, and 26 were obtained from Lectinity, and glycans 8 to 19, 21 to 23, 25, and 27 to 31 were from the Consortium for Functional Glycomics (http://www.functionalglycomics.org/). The chemoenzymatic synthesis of compounds 5 and 32 will be published elsewhere. Briefly, for analyses using recombinant HA, HA-antibody complexes were prepared by mixing 1.5 μg of recombinant HA, mouse anti-penta-His-Alexa Fluor 488 (Qiagen), and anti-mouse-IgG-Alexa Fluor 488 (Invitrogen) in a molar ratio of 4:2:1, respectively. These prepared mixtures of complexes were incubated for 15 min on ice, diluted to 100 μl with PBS-T (phosphate-buffered saline containing 0.05% Tween 20), and incubated on the array surface in a humidified chamber for 1 h. Slides were subsequently washed by successive rinses with PBS-T, PBS, and deionized water. Washed arrays were dried by centrifugation and immediately scanned for Alexa Fluor 488 signal on a Perkin-Elmer Proscanarray Express. Fluorescent signal intensity was measured using Imagene (Biodiscovery), and raw signal data were calculated for mean intensity and graphed using Prism (GraphPad).

The enzyme-linked immunosorbent assay (ELISA)-based plate assays were carried out as previously described (8). A streptavidin-coated high-binding-capacity 384-well plate (Pierce) was loaded overnight at 4°C with 50 μl of 1.6 mM biotinylated glycans (NeuAcα2-6Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAc [6′-SLNLN] and NeuAcα2-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAc [3′-SLNLN]; obtained from the Consortium for Functional Glycomics). The plate subsequently was washed with PBS and incubated with preformed HA-antibody trimer complexes comprised of His-tagged HA protein and primary (mouse anti-Penta-His antibody; Qiagen) and secondary (horseradish peroxidase [HRP]-conjugated goat anti-mouse IgG; Pierce) antibodies mixed in a ratio of 4:2:1. After incubation and extensive washes, HRP activity was measured using an Amplex Red hydrogen peroxide/peroxidase assay kit (Invitrogen).

Protein structure accession numbers. The atomic coordinates and structure factors have been deposited in the Protein Data Bank (www.rcsb.org) under codes 3UBE, 3UBJ, 3UBN, and 3UBQ.

RESULTS

Complexes with human and avian receptors. Previously, we determined the crystal structure of the HA ectodomain of CA04 using a baculovirus expression system (47). Due to the weak association of the HA subunits in the wild-type trimer (47, 49), we introduced a stabilizing disulfide bridge across the subunit-subunit interface to facilitate the structural determination of receptor complexes. The mutant G205C/R220C was crystallized under the same conditions and in the same crystal lattice as the wild-type CA04 HA (47). The mutations themselves introduced little perturbation to the HA structure. The root square mean difference (RMSD) of Cα atoms between the mutant and wild type is 0.35 Å for the entire receptor-binding domain, suggesting that this stabilized HA trimer is a suitable surrogate for studying HA-receptor interaction.

Crystal structures of HA were determined in complex with the pentasaccharides LSTc and LSTa, which represent α2-6 and α2-3-linked glycan analogs of human and avian receptors, respectively. LSTc (NeuAcα2-6Galβ1-4GlcNAcβ1-3Galβ1-4Glc) is a close analog of the terminal sequence of extended N-linked glycans that are found on human and swine respiratory epithelial cells (4, 8), which have been proposed as the preferred ligands of human influenza viruses (8).

In the LSTc α2-6-linked glycan complex, strong electron density for the glycan ligand (Fig. 1A) is observed in all six of the potential binding sites (two HA trimers) in the crystal asymmetric unit and enabled interpretation and model buildup to the fourth sugar, Gal-4, of the five sugars. Ligand LSTc adopts an overall compact conformation, which is aided by the *cis* configuration of the glycosidic bond between Sia-1 and Gal-2 (Fig. 2A). The protein-ligand interactions involving the terminal sialic acid (Fig. 2A and B) generally are similar to those in H3 HA complexes (41); here, we focus on the differences between CA04 and H3 HA. Thr136 in CA04 H1 replaces Ser136 of H3 but makes the same hydrogen bond with the carboxylate of Sia-1. Gln226 forms hydrogen bonds with the carboxylate and 8-hydroxyl group of Sia-1 that cannot be made with Leu226 in H3. In addition to the highly conserved polar and van der Waals contacts between Sia-1 and HA, an extensive water-mediated, hydrogen bond network is observed in the CA04 structures. Ser186 and Asp190 together coordinate a water molecule (Wat1) that hydrogen bonds with the 9-hydroxyl group of Sia-1. In H2 and H3 HAs, this hydrogen bond to Sia-1 is made by the carboxyl group of Glu190 (16, 26). Two additional water molecules coordinate interactions between resi-

TABLE 1 Data collection and refinement statistics[a]

Parameter	Values for: CA04 + LSTc	CA04 + 6SLN	CA04 + LSTa	CA04 + 3SLN
Data collection statistics				
Wavelength (Å)	1.0332	1.0332	1.0332	1.0332
Space group	P1	P1	P1	P1
Unit cell dimensions				
a, b, c (Å)	66.8, 116.6, 118.7	67.0, 116.7, 119.7	66.9, 116.1, 118.4	66.9, 116.3, 118.4
α, β, γ (°)	60.9, 77.3, 80.7	60.6, 77.0, 80.4	61.0, 77.0, 80.4	60.9, 77.1, 80.4
Resolution range (Å)	50–2.15 (2.23–2.15)	50–2.50 (2.59–2.50)	50–2.25 (2.33–2.25)	50–2.00 (2.07–2.00)
No. of observations	310,510	207,943	274,313	348,771
No. of unique reflections	153,165 (12,262)	99,231 (7,960)	131,425 (9,713)	172,551 (10,978)
Completeness (%)	92.8 (74.3)	94.6 (75.7)	91.0 (67.5)	84.4 (53.7)
$I/\sigma(I)$	10.2 (1.2)	9.2 (1.3)	9.3 (1.4)	10.7 (1.4)
R_{sym}[b]	0.07 (0.53)	0.09 (0.47)	0.08 (0.43)	0.07 (0.46)
Refinement statistics				
Resolution range (Å)	49.5–2.15	49.7–2.50	49.3–2.25	49.2–2.00
No. of reflections (total)	153,068	99,186	130,790	172,491
No. of reflections (test)	7,685	4,908	6,522	8,648
R_{cryst}[c] (%)	19.5	20.5	20.1	19.9
R_{free}[d] (%)	25.2	25.3	25.2	25.0
Avg B value (Å^2)				
Protein	40.9	48.9	41.6	41.9
Ligand	49.2	56.1	57.0	57.9
Water	38.4	40.1	36.4	39.9
Wilson B value (Å^2)	33.4	42.3	31.3	29.5
Protein atoms	23,303	23,303	23,264	23,264
Ligand atoms	281	261	124	108
Water	881	485	783	1053
RMSD from ideal geometry				
Bond length (Å)	0.004	0.002	0.003	0.005
Bond angles (°)	0.82	0.61	0.75	0.89
Ramachandran statistics[e] (%)				
Favored	95.2	94.7	94.3	95.3
Outliers	0.3	0.5	0.6	0.4
PDB code	3UBE	3UBN	3UBJ	3UBQ

[a] Numbers in parentheses refer to the highest-resolution shell.
[b] $R_{sym} = \Sigma_{hkl}|<I_i>|/\Sigma_{hkl} I_i$, where I_i is the scaled intensity of the ith measurement and $<I_i>$ is the average intensity for that reflection.
[c] $R_{cryst} = \Sigma_{hkl}|F_o - F_c|/\Sigma_{hkl}|F_o| \times 100$.
[d] R_{free} was calculated as described for R_{cryst} but on a test set comprising 5% of the data excluded from refinement.
[e] Calculated using Molprobity.

dues from the 220 loop and the hydroxyl groups of Sia-1. Interestingly, the HA-ligand interaction also is stabilized by a well-ordered water (Wat4) that bridges the 4-hydroxyl of Sia-1 and the carbonyl oxygen of Lys133A. This water-mediated interaction is unique to H1 HAs, as Lys133A corresponds to a 1-residue insertion in the 130 loop, unlike H2 and H3 HAs. Beyond Sia-1, protein-glycan interactions between CA04 and LSTc are dominated by two residues, Asp190 and Asp225, that function as the specificity switches in H1 subtypes. Gal-2 forms extensive hydrogen bonds with Asp225 and the neighboring Lys222. The side-chain conformation of Lys222 is stabilized by ionic interactions with Asp225 and Glu227. After Gal-2, LSTc makes a sharp turn toward the 190 helix, where Asp190 is the only residue that contacts the glycan. The Asp190 carboxyl hydrogen bonds with the 2-acetamido nitrogen of GlcNAc-3 and the 2-hydroxyl of Gal-4.

For the LSTa α2-3-linked glycan complex, relatively weak density was observed for the glycan in 4 of the 6 receptor-binding sites in the two copies of the HA trimer in the crystal asymmetric unit, and no interpretable density was found in the other two sites (Fig. 1B). Only Sia-1 and the connecting Gal-2 could be modeled for the glycan ligand (Fig. 2C and D). As in the LSTc complex, the protein-glycan interactions around Sia-1, including the bridging water molecules, are very well conserved, but very few contacts are found between the HA and the rest of the glycan. The 4-hydroxyl of Gal-2 forms one hydrogen bond with a water molecule (Wat3) that is associated with the 220 loop. This lack of sugar-protein interactions explains the disorder in the rest of the glycan and the weak binding to the HA for LSTa (see below).

Complexes with trisaccharides. Crystal structures also were determined in complex with the trisaccharides 3′-SLN (3′-sialyl-*N*-acetyllactosamine) and 6′-SLN (6′-sialyl-*N*-acetyllactosamine). The two trisaccharides contain only three sugar moieties: Sia-1, Gal-2, and GlcNAc-3. Similarly to the structural complexes of the two LST pentasaccharides, the α2-6-linked glycan 6′-SLN is well ordered in all six receptor-binding sites in the asymmetric unit, whereas only half of the sites are occupied for 3′-SLN, and with

TABLE 2 List of glycans imprinted on the microarray[a]

Glycan	Glycan name
1	Galβ(1-4)-GlcNAcβ-ethyl-NH_2
2	Galβ(1-4)-GlcNAcβ(1-2)-Manα(1-3)-[Galβ(1-4)-GlcNAcβ(1-2)-Manα(1-6)]-Manβ(1-4)-GlcNAcβ(1-4)-GlcNAcβ-Asn-NH_2
3	NeuAcα(2-3)-Galβ(1-4)-6-O-sulfo-GlcNAcβ-propyl-NH_2
4	NeuAcα(2-3)-Galβ(1-4)-[Fucα(1-3)]-6-O-sulfo-GlcNAcβ-propyl-NH_2
5	NeuAcα(2-3)-6-O-sulfo-Galβ(1-4)-GlcNAcβ-ethyl-NH_2
6	NeuAcα(2-3)-6-O-sulfo-Galβ(1-4)-[Fucα(1-3)]-GlcNAcβ-propyl-NH_2
7	NeuAcα(2-3)-Galβ(1-3)-6-O-sulfo-GlcNAcβ-propyl-NH_2
8	NeuAcα(2-3)-Galβ(1-4)-Glcβ-ethyl-NH_2
9	NeuAcα(2-3)-Galβ(1-4)-GlcNAcβ-ethyl-NH_2
10	NeuAcα(2-3)-Galβ(1-4)-GlcNAcβ(1-3)-Galβ(1-4)-GlcNAcβ-ethyl-NH_2
11	NeuAcα(2-3)-Galβ(1-4)-GlcNAcβ(1-3)-Galβ(1-4)-GlcNAcβ(1-3)-Galβ(1-4)-GlcNAcβ-ethyl-NH_2
12	NeuAcα(2-3)-GalNAcβ(1-4)-GlcNAcβ-ethyl-NH_2
13	NeuAcα(2-3)-Galβ(1-3)-GlcNAcβ-ethyl-NH_2
14	NeuAcα(2-3)-Galβ(1-3)-GlcNAcβ(1-3)-Galβ(1-4)-GlcNAcβ-ethyl-NH_2
15	NeuAcα(2-3)-Galβ(1-3)-GlcNAcβ(1-3)-Galβ(1-3)-GlcNAcβ-ethyl-NH_2
16	NeuAcα(2-3)-Galβ(1-3)-GalNAcβ(1-3)-Gala(1-4)-Galβ(1-4)-Glcβ-ethyl-NH_2
17	NeuAcα(2-3)-[GalNAcβ(1-4)]-Galβ(1-4)-GlcNAcβ-ethyl-NH_2
18	NeuAcα(2-3)-[GalNAcβ(1-4)]-Galβ(1-4)-Glcβ-ethyl-NH_2
19	Galβ(1-3)-GalNAcβ(1-4)-[NeuAcα(2-3)]-Galβ(1-4)-Glcβ-ethyl-NH_2
20	NeuAcα(2-3)-Galβ(1-4)-[Fucα(1-3)]-GlcNAcβ-propyl-NH_2
21	NeuAcα(2-3)-Galβ(1-3)-[Fucα(1-4)]-GlcNAcβ(1-3)-Galβ(1-4)-[Fucα(1-3)]-GlcNAcβ-ethyl-NH_2
22	NeuAcα(2-3)-Galβ(1-4)-[Fucα(1-3)]-GlcNAcβ(1-3)-Galβ(1-4)-[Fucα(1-3)]-GlcNAcβ-ethyl-NH_2
23	NeuAcα(2-3)-Galβ(1-4)-[Fucα(1-3)]-GlcNAcβ(1-3)-Galβ(1-4)-[Fucα(1-3)]-GlcNAcβ(1-3)-Galβ(1-4)-[Fucα(1-3)]-GlcNAcβ-ethyl-NH_2
24	NeuAcα(2-3)-Galβ(1-4)-GlcNAcβ(1-2)-Manα(1-3)-[NeuAcα(2-3)-Galβ(1-4)-GlcNAcβ(1-2)-Manα(1-6)]-Manβ(1-4)-GlcNAcβ(1-4)-GlcNAcβ-Asn-NH_2
25	NeuGcα(2-3)-Galβ(1-4)-GlcNAcβ-ethyl-NH_2
26	NeuAcα(2-6)-Galβ(1-4)-6-O-sulfo-GlcNAcβ-propyl-NH_2
27	NeuAcα(2-6)-Galβ(1-4)-Glcβ-ethyl-NH_2
28	NeuAcα(2-6)-Galβ(1-4)-GlcNAcβ-ethyl-NH_2
29	NeuAcα(2-6)-Galβ(1-4)-GlcNAcβ(1-3)-Galβ(1-4)-GlcNAcβ-ethyl-NH_2
30	NeuAcα(2-6)-Galβ(1-4)-GlcNAcβ(1-3)-Galβ(1-4)-GlcNAcβ(1-3)-Galβ(1-4)-GlcNAcβ-ethyl-NH_2
31	NeuAcα(2-6)-GalNAcβ(1-4)-GlcNAcβ-ethyl-NH_2
32	NeuAcα(2-6)-Galβ(1-4)-GlcNAcβ(1-3)-[NeuAcα(2-6)]-Galβ(1-4)-GlcNAcβ-ethyl-NH_2
33	Galβ(1-4)-GlcNAcβ(1-2)-Manα(1-3)-[NeuAcα(2-6)-Galβ(1-4)-GlcNAcβ(1-2)-Manα(1-6)]-Manβ(1-4)-GlcNAcβ(1-4)-GlcNAcβ-Asn-NH_2
34	NeuAcα(2-6)-Galβ(1-4)-GlcNAcβ(1-2)-Manα(1-3)-[Galβ(1-4)-GlcNAcβ(1-2)-Manα(1-6)]-Manβ(1-4)-GlcNAcβ(1-4)-GlcNAcβ-Asn-NH_2
35	GlcNAcβ(1-2)-Manα(1-3)-[NeuAcα(2-6)-Galβ(1-4)-GlcNAcβ(1-2)-Manα(1-6)]-Manβ(1-4)-GlcNAcβ(1-4)-GlcNAcβ-Asn-NH_2
36	NeuAcα(2-6)-Galβ(1-4)-GlcNAcβ(1-2)-Manα(1-3)-[NeuAcα(2-6)-Galβ(1-4)-GlcNAcβ(1-2)-Manα(1-6)]-Manβ(1-4)-GlcNAcβ(1-4)-GlcNAcβ-Asn-NH_2
37	NeuAcα(2-3)-Galβ(1-4)-GlcNAcβ(1-2)-Manα(1-3)-[NeuAcα(2-6)-Galβ(1-4)-GlcNAcβ(1-2)-Manα(1-6)]-Manβ(1-4)-GlcNAcβ(1-4)-GlcNAcβ-Asn-NH_2
38	NeuAcα(2-6)-Galβ(1-4)-GlcNAcβ(1-2)-Manα(1-3)-[NeuAcα(2-3)-Galβ(1-4)-GlcNAcβ(1-2)-Manα(1-6)]-Manβ(1-4)-GlcNAcβ(1-4)-GlcNAcβ-Asn-NH_2

[a] The array contains two neutral glycans (1 and 2), 23 α2-3-linked glycans (3 to 25), 11 α2-6-linked sialylated glycans (26 to 36), and two glycans of mixed linkages (37 and 38).

considerably weaker electron density (Fig. 1C and D). Structural superposition with the corresponding pentasaccharide complexes shows nearly identical glycan conformation for the α2-6-linked glycans (Fig. 3A). Protein-glycan interactions, including those mediated by water molecules, are very well conserved between the LSTc and 6′-SLN complexes. In the 3′-SLN complex structure, the glycan is bound slightly deeper in the receptor-binding site than in the LSTa complex. For one of the sites, electron density is present for GlcNAc-3, implying slightly better binding for the trisaccharide. Taking these results together, the discrimination of these glycans with different glycosidic linkages is achieved mainly by specific HA recognition of the terminal three sugars, as previously observed in other HA subtypes (16, 21, 26).

Glycan binding and receptor specificity. The structural analysis of the glycan complexes supports the earlier reports that indicated α2-6-linked glycans are preferentially recognized by the receptor-binding site of CA04 H1 HA (7, 10, 29, 49). To examine the specificity more quantitatively, we performed glycan array analysis and the titration of recombinant HA binding to sialosides in plate assays. Analysis with the recombinant HA precludes any possible impact of the neuraminidase (NA) on the results, which occurs when whole virus is used. The glycan array (Table 2) was specifically customized for the study of influenza viruses. It includes 36 unique natural sialosides (α2,3 linkage, 3 to 25; α2,6 linkage, 26 to 36; mixed linkage, 37 and 38) that are relevant to influenza biology and 2 neutral glycans (1 and 2) as controls. On the glycan microarray, the HAs from three 2009 pandemic strains (A/California/04/2009, A/Netherlands/602/2009, and A/New York/06/2009) show significant binding only to α2-6-linked glycans and not to α2-3-linked receptor analogs (Fig. 4A to C). All tested HAs bind more favorably to long linear α2-6 glycans, including α2-6 sialylated tri-*N*-acetyllactosamine 30 and α2-6

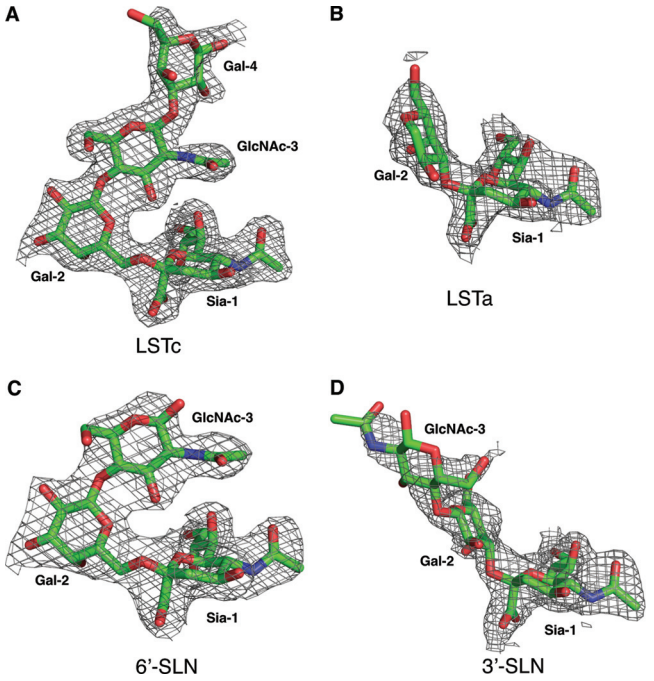

FIG 1 Electron density maps (2Fo-Fc) of glycan ligands in the CA04 HA crystal structures. (A) LSTc; (B) LSTa; (C) 6′-SLN; (D) 3′-SLN. The maps are contoured at a 1σ level. For clarity, only density for the glycan receptors is shown.

sialylated di-*N*-acetyllactosamine 29, which are analogous to the LSTc pentasaccharide used for the crystal structures. HAs from A/Netherlands/602/2009 and A/New York/06/2009 also display binding to the shorter α2-6 sialylated *N*-acetyllactosamine 28 and to sulfated α2-6 sialylated *N*-acetyllactosamine 26. However, the HA of CA04 displays considerably lower avidity than those of the other two 2009 pandemic strains tested. The difference in avidity likely resulted from a single-residue difference near the 190 helix. CA04 contains Thr200, whereas Ala200 is present in A/Netherlands/602/2009 and A/New York/06/2009. Although residue 200 is not among the HA residues directly interacting with glycan receptors, the T200A substitution in CA04 was shown to improve the receptor binding of recombinant HA on the glycan array (15).

The lower avidity of the CA04 HA relative to those of the other pandemic strains also was evident in a plate-based ELISA. While all three recombinant HAs bind to α2-6-sialylated di-*N*-acetyllactosamine, CA04 HA exhibited more than 10-fold weaker avidity (Fig. 4D to F). In contrast, only minuscule binding to α2-3-sialylated di-*N*-acetyllactosamine could be detected, even at high protein concentrations, in the ELISA.

DISCUSSION

Recognition of human-like α2-6 receptors. The α2-6-linked human receptor analogs adopt compact, folded configurations in the receptor-binding site of HA. The ligand conformation and its positioning in the binding site are similar to previously reported receptor complexes of H1 HAs from A/South Carolina/1/18 (SC1918) (26) (Fig. 5A). Some minor differences do occur in the receptor-binding site, most notably in the relative position of the 190 helix. In CA04, the 190 helix is located slightly further from the 220 loop, possibly due to the bulky Glu227 in CA04 instead of Ala227 in SC1918. Because of the structural variations in the receptor-binding site, Sia-1 and Gal-2 of the α2-6-linked receptor analog are positioned closer to the 130 loop in the CA04 complex. Nevertheless, Asp190 and Asp225 continue to play critical roles in glycan recognition through hydrogen bonds with GlcNAc-3 and Gal-2, respectively. Asp190 and Asp225 thus serve as reliable indicators for the specific recognition of α2-6-linked sialylated glycans in H1 HAs.

The HA from an early human H1 isolate (A/Puerto Rico/8/34 [PR8/34]) carries a D190E substitution and displays dual specificity for both avian and human receptors (19, 39). In the complex structure with LSTc, Glu190 of PR8/34 does not directly contact GlcNAc-3 and instead participates in a water-mediated hydrogen bond network involving residues from the 220 loop (19). This paucity of interaction between PR8/34 and GlcNAc-3 likely relaxes the specificity for α2-6-linked receptor analogs that is seen by the lack of interpretable electron density beyond GlcNAc-3 in the PR8/34 crystal structure.

Recognition of avian-like α2-3 receptors. A signature motif in the H1 HA receptor-binding site for specific avian receptor binding has not been quite so clear from previous studies. Before this study, the only human H1 HA structure in complex with an avian-like receptor was for the PR8/34 HA (19), which has dual specificity for glycan receptors (39). In the complex structure with α2-3-linked receptor analog LSTa, PR8/34 displays HA-glycan interactions similar to those observed in the crystal structure of avian H1 HA (A/wdk/JX/12416/2005) complex (19, 25) and leads to ambiguity as to what makes human HAs less effective in binding to avian receptors. Comparison between the CA04 H1 HA and an avian H1 HA, however, elucidates the structural determinants that are specific for the recognition of α2-3-linked glycans. LSTa maintains a conformation and location in the binding site of CA04 that are similar to those in avian HA. In the avian structure, Gln226 makes close contacts with Gal-2, forming hydrogen bonds with O-3 and O-4 of Gal-2 (Fig. 5B) (25). This pair of hydrogen bonds are the only interactions between avian HA and the avian-like receptor analog LSTa besides those that involve Sia-1. Similar HA-receptor interactions also are observed in the crystal structures of avian H2 and H3 HAs (21, 26), implying a common mechanism for avian receptor recognition. These favorable interactions for α2-3-linked glycans are absent from the structure of human CA04 (Fig. 5B). Although Gln226 moves about 0.5 Å toward Gal-2 in structures complexed with α2-3-linked sialylated glycans compared to those in complex with α2-6-linked glycans, the side chain of Gln226 still is more than 4 Å from Gal-2, which eliminates effective hydrogen bonding. In comparison, in PR8/34 HA, Gln226 moves close enough to hydrogen bond with O-4, but not O-3, of Gal-2 in LSTa. This interaction likely is facilitated by the hydrogen bond network established between the 220 loop and Glu190, a residue normally conserved in avian but not human H1 HAs (19). Thus, favorable binding for avian receptors in avian H1 HA is achieved by close contacts between Gln226 of H1 HA and Gal-2 of the glycan receptor. This difference in avian receptor recognition between human and avian HAs is in large part determined by the overall conformation of the 220 loop, which contains Gly225 as well as Gln226 (Fig. 5B). The 220 loop of avian HAs, with its conserved Gly225, displays a conformation distinct from that of human HAs (Fig. 5B) (25). Such an avian-like 220 loop conformation is energetically prohibitive for most human HAs, as the dihedral angles of Gly225 ($\phi = 139.1°$; $\psi = -69.7°$) in

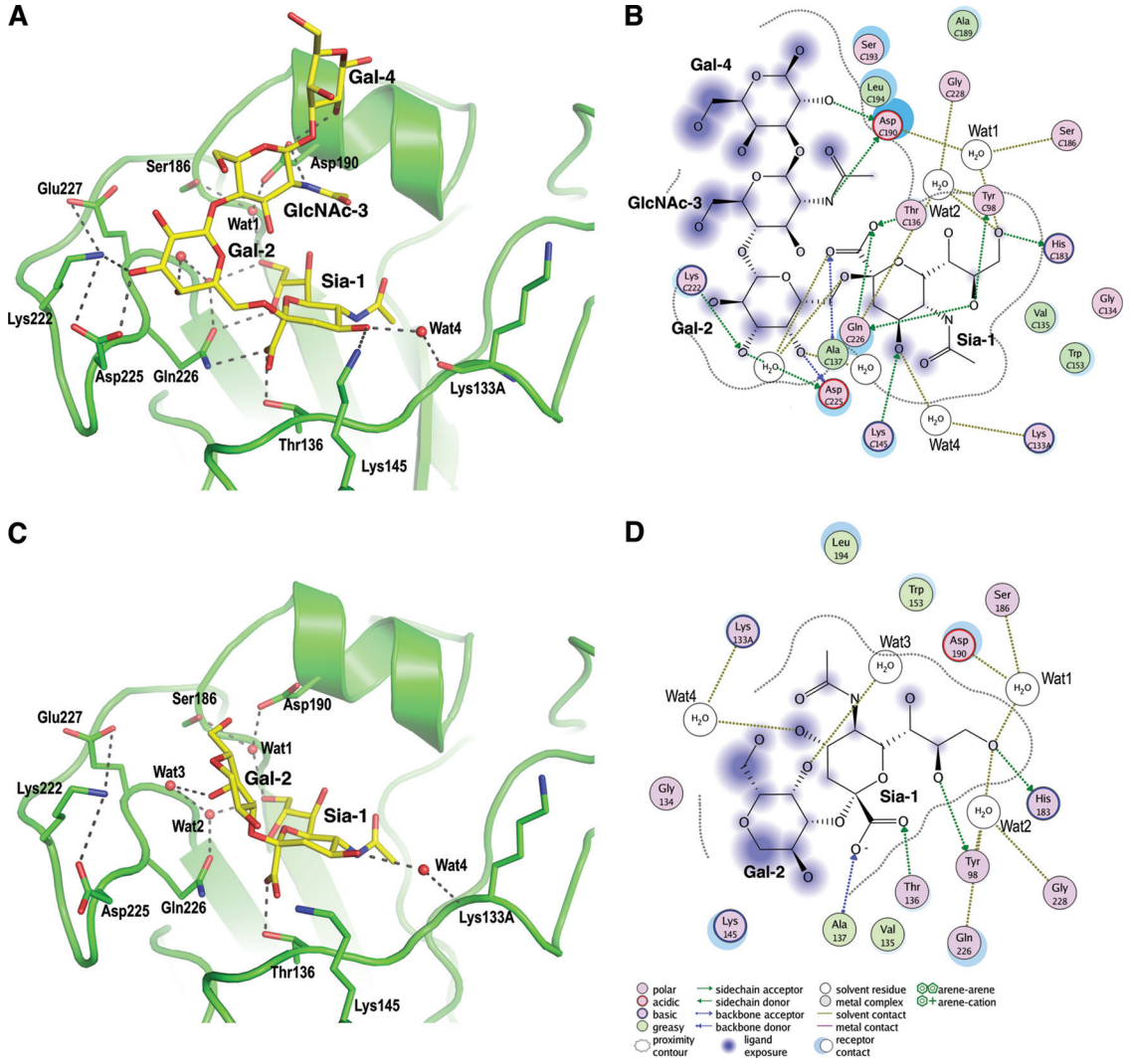

FIG 2 Molecular interactions in the crystal structure of the CA04 HA from the 2009 influenza pandemic with human and avian receptor analogs. The human analog LSTc (A and B) and avian receptor analog LSTa (C and D) are shown in a close view of the receptor-binding site (A and C) in the crystal structures and in molecular interaction diagrams (B and D). The glycans are represented with yellow carbons, red oxygens, and blue nitrogens, and the receptor-binding site backbone is in green. HA-glycan interactions are conserved in the Sia-1 region. LSTc binding is further stabilized by hydrogen bonds (dashes) to Asp190 and Asp225. Because of the difference in the overall sugar configuration due to the α2-3 versus α2-6 linkage, LSTa lacks specific HA-glycan contacts beyond its Sia-1 sugar ring. The two-dimensional interaction diagrams in B and D were drawn using MOE (Molecular Operating Environment) (13).

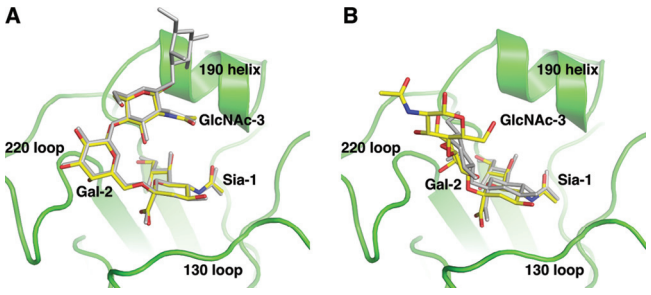

FIG 3 Receptor-binding specificity of CA04 HA is determined mainly by interactions between HA and the three terminal glycans of the receptors. Shown is a comparison of bound trisaccharide and pentasaccharide receptor analogs in the CA04 receptor binding site (green). The trisaccharide receptor analogs (in yellow) show binding modes similar to those of the pentasaccharide analogs (in gray) that contain the same glycosidic linkage between Sia-1 and Gal-2. (A) α2-6-linked glycans; (B) α2-3-linked glycans.

avian HA are energetically allowed only for glycine residues (28). The optimal positioning of Gln226 for avian receptor binding (Fig. 5B) thus is aided by the dual substitutions of D190E and D225G, which reshape the 220 loop and enable Gln226 to move closer to Gal-2, and this is mediated and aided by the formation of a water-mediated hydrogen bond network between Glu190 and Gln226. Specific changes in these two amino acids (residues 190 and 225) not only establish favorable interactions with avian receptors beyond Sia-1 but also eliminate potential interactions that are specific for human receptor recognition. Single substitutions of these residues generate HAs of dual specificities, whereas decreased avidity for α2-6-linked glycans and increased affinity for α2-3-linked glycans is seen for the Glu190/Gly225 combination (39, 44).

Residue 225 in 2009 pandemic viruses. Among the thousands of available H1 HA protein sequences from 2009 pandemic vi-

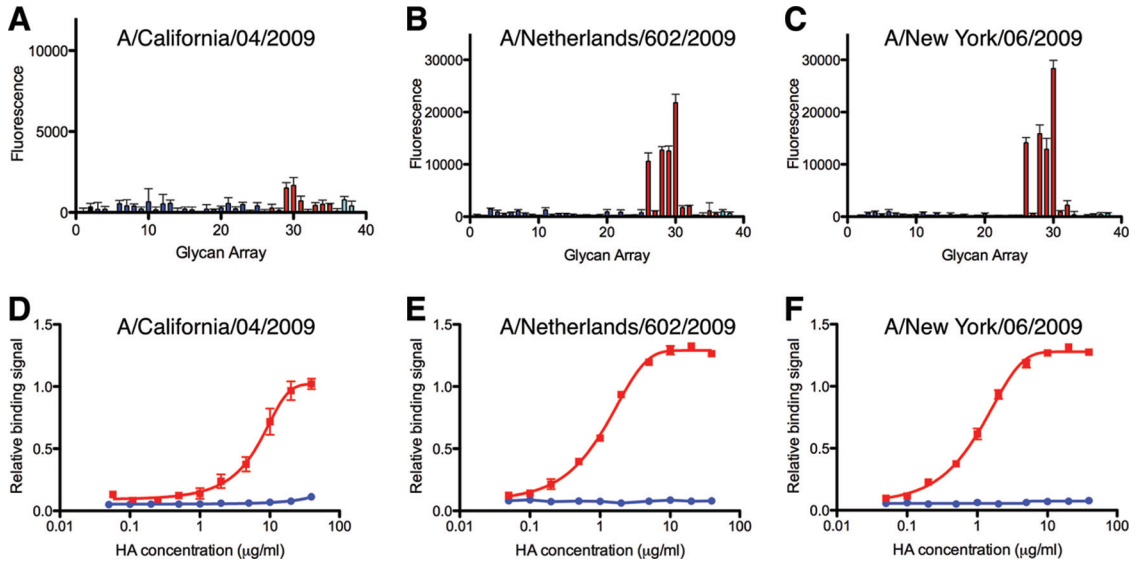

FIG 4 Glycan binding analysis of recombinant HAs of the 2009 pandemic. (A to C) In the glycan microarray studies, the HAs show specific binding toward certain α2-6-linked sialylated glycans (red bars; 26 to 36) but not to any α2-3-linked glycans (blue; 3 to 25), neutral glycans (black; 1 and 2), or glycans of mixed linkages (cyan; 37 and 38). The list of glycans on the array is provided in Table 2. Note that panel A is on a different scale than B and C due to the relatively low avidity of CA04. (D to F) In an ELISA-based plate assay, the tested HAs display significant binding to biotinylated 6′-SLNLN (Neu5Acα2-6Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAc) (in red) but not to 3′-SLNLN (Neu5Acα2-3Galβ1-4GlcNAcβ1-3Galβ1-4GlcNAc) (in blue). All error bars in the figure are indicative of standard deviations from three experiments.

ruses, the majority have Asp190 and Asp225 (3, 12). These viruses, thus, have receptor specificity for α2-6-linked glycans, confirming the correlation between human receptor-binding specificity and efficient transmission in humans (30). Nevertheless, sporadic identifications of HA D225G mutants have been reported and linked to cases of severe disease and fatality (12, 27). Structural analysis suggests that the D225G substitution removes a favorable interaction with the Gal-2 of α2-6 sialylated glycans and increases conformational flexibility for adaptation to α2-3 sialylated glycans, as noted above. Indeed, glycan-binding studies suggest that the D225G mutation in H1 improves binding for α2-3-linked glycans, while it decreases avidity for α2-6-linked glycans (12, 27, 44, 49). The slight shift in receptor specificity then would affect cell tropism and possibly the virulence of these viruses (12, 27), but it also may reduce its ability to transmit (46), as reflected in its vastly reduced prevalence. Interestingly, the D225G substitution in 2009 pandemic viruses has a much more subtle effect on receptor binding than the same mutation in the HA of 1918 influenza virus, as deduced from glycan microarray analyses (49). A plausible explanation is the presence of Glu227 in 2009 pandemic viruses com-

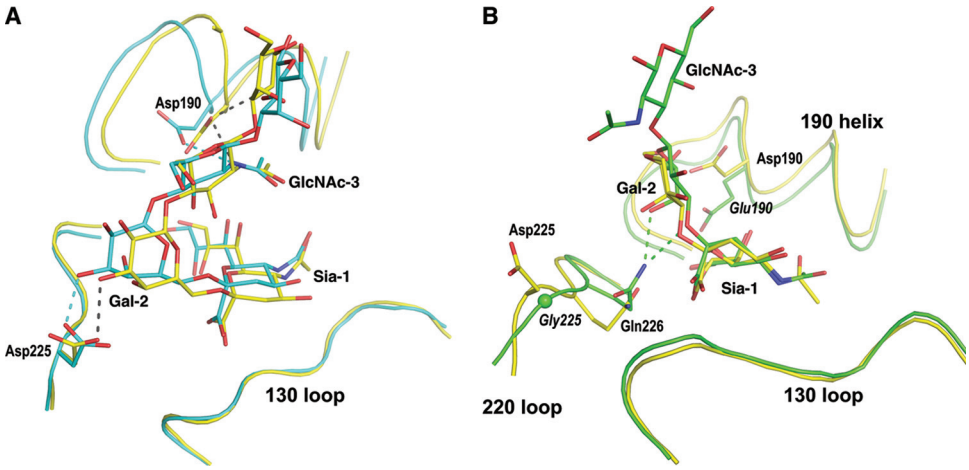

FIG 5 Structural comparison of receptor complexes of human (A) and avian (B) H1 HAs. (A) Similar binding interactions are observed for the LSTc complex of two human viruses: CA04 2009 (in yellow) and SC1918 (in blue; PDB code 2WRG [26]). Asp190 and Asp225 interact with the glycan receptor (hydrogen bonds in dashes; gray for CA04 and blue for SC1918) and are key determinants for the binding of human-like receptors. (B) Avian H1 HA (in green; PDB code 3HTP [25]) achieves preferential binding for α2-3-linked glycans through close contacts between Gln226 and the Gal-2 sugar (hydrogen bonds between Gln226 and Gal-2 are shown as green dashes). A similar arrangement in CA04 HA (in yellow) is unfavorable because of the stereochemical restraints imposed by the replacement of Gly225 by Asp225. Glu190 and Gly225 are specific for avian HAs and are labeled in italics.

pared to Ala227 in 1918 HA. In CA04, Glu227 is part of the charge network, including Asp225 and Lys222. In the absence of Asp225, Glu227 may compensate for the loss of Asp 225 and help maintain the overall conformation of the 220 loop and stabilize the interactions between Lys222 and human receptors.

In conclusion, we conclusively demonstrate that the HA from the 2009 H1N1 pandemic viruses preferentially recognizes α2-6 sialylated receptors (7, 10, 11, 29, 43, 49). Preferential binding of α2-6 sialylated glycans is achieved by hydrogen bonds between the glycan and HA mediated mainly by Asp190 and Asp225. In the crystal structures with human-like receptor analogs, Asp190 interacts with GlcNAc-3 while Asp225 contacts Gal-2. Due to differences in the overall configurations of α2-6- versus α2-3-linked sialylated glycans, these favorable interactions are absent from the complex structures with avian-like glycan receptors. The lack of any significant HA interactions with α2-3-linked glycans beyond sialic acid (Sia-1) explains the considerably weaker HA avidity to avian-like receptor analogs. Although other groups have reported binding to α2-3-linked glycans (11, 43), we conclude that this must be a result of the sensitive detection of very weak α2-3 binders relative to α2-6 receptors. Our crystal structures confirm the relatively low affinity for α2-3-linked receptor analogs, as evidenced by the relatively disordered electron densities and low occupancy of ligands in these structures. Such small differences in individual HA/glycan affinity are amplified by multivalent binding into distinguishable binding properties (40, 48), as also suggested by our glycan-binding experiments as well as by others (7, 10, 15, 29, 49).

ACKNOWLEDGMENTS

The work was supported in part by NIAID grant AI058113 (I.A.W. and J.C.P.), the Skaggs Institute for Chemical Biology, and the Scripps Microarray Core Facility.

Some glycans used for the plate binding assay were provided by the Consortium for Functional Glycomics, which is funded by NIGMS grant GM62116 (J.C.P.). X-ray diffraction data sets were collected at the Advanced Photon Source beamline 23ID-D.

REFERENCES

1. **Adams PD, et al.** 2002. PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr. D Biol. Crystallogr. **58**:1948–1954.
2. **Anonymous.** 2009. Swine influenza A (H1N1) infection in two children–Southern California, March-April 2009. MMWR Morb. Mortal. Wkly. Rep. **58**:400–402.
3. **Bao Y, et al.** 2008. The influenza virus resource at the National Center for Biotechnology Information. J. Virol. **82**:596–601.
4. **Bateman AC, et al.** 2010. Glycan analysis and influenza A virus infection of primary swine respiratory epithelial cells: the importance of NeuAcα2-6 glycans. J. Biol. Chem. **285**:34016–34026.
5. **Blixt O, Brown J, Schur MJ, Wakarchuk W, Paulson JC.** 2001. Efficient preparation of natural and synthetic galactosides with a recombinant β-1,4-galactosyltransferase-/UDP-4′-gal epimerase fusion protein. J. Org. Chem. **66**:2442–2448.
6. **Blixt O, et al.** 2004. Printed covalent glycan array for ligand profiling of diverse glycan binding proteins. Proc. Natl. Acad. Sci. U. S. A. **101**: 17033–17038.
7. **Bradley KC, et al.** 2011. Comparison of the receptor binding properties of contemporary swine isolates and early human pandemic H1N1 isolates (novel 2009 H1N1). Virology **413**:169–182.
8. **Chandrasekaran A, et al.** 2008. Glycan topology determines human adaptation of avian H5N1 virus hemagglutinin. Nat. Biotechnol. **26**: 107–113.
9. **Chen GW, Shih SR.** 2009. Genomic signatures of influenza A pandemic (H1N1) 2009 virus. Emerg. Infect. Dis. **15**:1897–1903.
10. **Chen LM, et al.** 2011. Receptor specificity of subtype H1 influenza A viruses isolated from swine and humans in the United States. Virology **412**:401–410.
11. **Childs RA, et al.** 2009. Receptor-binding specificity of pandemic influenza A (H1N1) 2009 virus determined by carbohydrate microarray. Nat. Biotechnol. **27**:797–799.
12. **Chutinimitkul S, et al.** 2010. Virulence-associated substitution D222G in the hemagglutinin of 2009 pandemic influenza A(H1N1) virus affects receptor binding. J. Virol. **84**:11802–11813.
13. **Clark AM, Labute P, Santavy M.** 2006. 2D structure depiction. J. Chem. Infect. Model. **46**:1107–1123.
14. **Connor R, Kawaoka Y, Webster R, Paulson J.** 1994. Receptor specificity in human, avian, and equine H2 and H3 influenza virus isolates. Virology **205**:17–23.
15. **de Vries RP, et al.** 2011. Only two residues are responsible for the dramatic difference in receptor binding between swine and new pandemic H1 hemagglutinin. J. Biol. Chem. **286**:5868–5875.
16. **Eisen M, Sabesan S, Skehel J, Wiley D.** 1997. Binding of the influenza A virus to cell-surface receptors: structures of five hemagglutinin-sialyloligosaccharide complexes determined by X-ray crystallography. Virology **232**:19–31.
17. **Emsley P, Cowtan K.** 2004. Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. **60**:2126–2132.
18. **Fraser C, et al.** 2009. Pandemic potential of a strain of influenza A (H1N1): early findings. Science **324**:1557–1561.
19. **Gamblin S, et al.** 2004. The structure and receptor binding properties of the 1918 influenza hemagglutinin. Science **303**:1838–1842.
20. **Garten RJ, et al.** 2009. Antigenic and genetic characteristics of swine-origin 2009 A(H1N1) influenza viruses circulating in humans. Science **325**:197–201.
21. **Ha Y, Stevens D, Skehel J, Wiley D.** 2003. X-ray structure of the hemagglutinin of a potential H3 avian progenitor of the 1968 Hong Kong pandemic influenza virus. Virology **309**:209–218.
22. **Itoh Y, et al.** 2009. In vitro and in vivo characterization of new swine-origin H1N1 influenza viruses. Nature **460**:1021–1025.
23. **Kilander A, Rykkvin R, Dudman SG, Hungnes O.** 2010. Observed association between the HA1 mutation D222G in the 2009 pandemic influenza A(H1N1) virus and severe clinical outcome, Norway 2009–2010. Euro. Surveill. **15**:19498.
24. **Klock HE, Lesley SA.** 2009. The polymerase incomplete primer extension (PIPE) method applied to high-throughput cloning and site-directed mutagenesis. Methods Mol. Biol. **498**:91–103.
25. **Lin T, et al.** 2009. The hemagglutinin structure of an avian H1N1 influenza A virus. Virology **392**:73–81.
26. **Liu J, et al.** 2009. Structures of receptor complexes formed by hemagglutinins from the Asian influenza pandemic of 1957. Proc. Natl. Acad. Sci. U. S. A. **106**:17175–17180.
27. **Liu Y, et al.** 2010. Altered receptor specificity and cell tropism of D222G hemagglutinin mutants isolated from fatal cases of pandemic A(H1N1) 2009 influenza virus. J. Virol. **84**:12069–12074.
28. **Lovell SC, et al.** 2003. Structure validation by Cα geometry: Φ, Ψ and Cβ deviation. Proteins **50**:437–450.
29. **Maines TR, et al.** 2009. Transmission and pathogenesis of swine-origin 2009 A(H1N1) influenza viruses in ferrets and mice. Science **325**:484–487.
30. **Matrosovich M, Stech J, Klenk HD.** 2009. Influenza receptors, polymerase and host range. Rev. Sci. Tech. **28**:203–217.
31. **Matrosovich M, et al.** 2000. Early alterations of the receptor-binding properties of H1, H2, and H3 avian influenza virus hemagglutinins after their introduction into mammals. J. Virol. **74**:8502–8512.
32. **McCoy AJ, Grosse-Kunstleve RW, Storoni LC, Read RJ.** 2005. Likelihood-enhanced fast translation functions. Acta Crystallogr. D Biol. Crystallogr. **61**:458–464.
33. **Munayco CV, et al.** 2009. Epidemiological and transmissibility analysis of influenza A(H1N1)v in a southern hemisphere setting: Peru. Euro Surveill. **14**:19299.
34. **Neumann G, Noda T, Kawaoka Y.** 2009. Emergence and pandemic potential of swine-origin H1N1 influenza virus. Nature **459**:931–939.
35. **Nishiura H, Castillo-Chavez C, Safan M, Chowell G.** 2009. Transmission potential of the new influenza A(H1N1) virus and its age-specificity in Japan. Euro Surveill. **14**:19227.
36. **Nishiura H, Wilson N, Baker MG.** 2009. Estimating the reproduction number of the novel influenza A virus (H1N1) in a Southern Hemisphere setting: preliminary estimate in New Zealand. N. Z. Med. J. **122**:73–77.

37. **Otwinowski Z, Minor W.** 1997. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol. **276**:307–326.
38. **Puzelli S, et al.** 2010. Transmission of hemagglutinin D222G mutant strain of pandemic (H1N1) 2009 virus. Emerg. Infect. Dis. **16**:863–865.
39. **Rogers GN, D'Souza BL.** 1989. Receptor binding properties of human and animal H1 influenza virus isolates. Virology **173**:317–322.
40. **Sauter NK, et al.** 1989. Hemagglutinins from two influenza virus variants bind to sialic acid derivatives with millimolar dissociation constants: a 500-MHz proton nuclear magnetic resonance study. Biochemistry **28**:8388–8396.
41. **Skehel J, Wiley D.** 2000. Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. Annu. Rev. Biochem. **69**:531–569.
42. **Smith GJ, et al.** 2009. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. Nature **459**:1122–1125.
43. **Soundararajan V, et al.** 2009. Extrapolating from sequence–the 2009 H1N1 "swine" influenza virus. Nat. Biotechnol. **27**:510–513.
44. **Stevens J, et al.** 2006. Glycan microarray analysis of the hemagglutinins from modern and pandemic influenza viruses reveals different receptor specificities. J. Mol. Biol. **355**:1143–1155.
45. **Stevens J, et al.** 2006. Structure and receptor specificity of the hemagglutinin from an H5N1 influenza virus. Science **312**:404–410.
46. **Tumpey T, et al.** 2007. A two-amino acid change in the hemagglutinin of the 1918 influenza virus abolishes transmission. Science **315**:655–659.
47. **Xu R, et al.** 2010. Structural basis of preexisting immunity to the 2009 H1N1 pandemic influenza virus. Science **328**:357–360.
48. **Xu R, McBride R, Paulson JC, Basler CF, Wilson IA.** 2010. Structure, receptor binding, and antigenicity of influenza virus hemagglutinins from the 1957 H2N2 pandemic. J. Virol. **84**:1715–1721.
49. **Yang H, Carney P, Stevens J.** 2010. Structure and receptor binding properties of a pandemic H1N1 virus hemagglutinin. PLoS Curr. **2**:RRN1152.