

# Insights into the Human Glycan Receptor Conformation of 1918 Pandemic Hemagglutinin–Glycan Complexes Derived from Nuclear Magnetic Resonance and Molecular Dynamics Studies

Stefano Elli,<sup>†</sup> Eleonora Macchi,<sup>†</sup> Timothy R. Rudd,<sup>‡</sup> Rahul Raman,<sup>§</sup> Guillherme Sassaki,<sup>⊥</sup> Karthik Viswanathan,<sup>§</sup> Edwin A. Yates,<sup>||</sup> Zachary Shriver,<sup>§</sup> Annamaria Naggi,<sup>†</sup> Giangiacomo Torri,<sup>†</sup> Ram Sasisekharan,<sup>\*§</sup> and Marco Guerrini<sup>\*,†</sup>

<sup>†</sup>Istituto di Ricerche Chimiche e Biochimiche “G. Ronzoni”, Via Giuseppe Colombo 81, 20133 Milano, Italy

<sup>‡</sup>Diamond Light Source Ltd., Diamond House, Harwell Science and Innovation Campus, Didcot, Oxfordshire OX11 0DE, U.K.

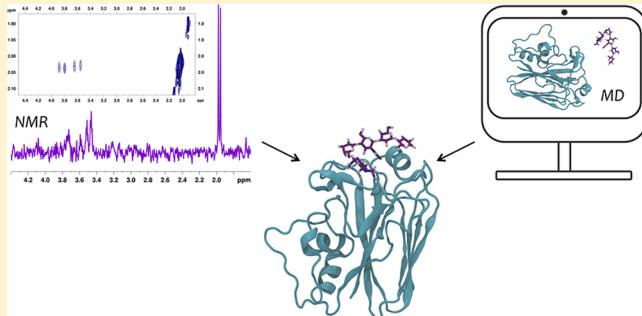
<sup>§</sup>Department of Biological Engineering, Koch Institute of Integrative Cancer Research, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, United States

<sup>⊥</sup>Departamento de Bioquímica e Biologia Molecular, Universidade Federal do Paraná, CEP 81.531-980, CP 19046, Curitiba, PR, Brazil

<sup>||</sup>Department of Structural and Chemical Biology, University of Liverpool, Liverpool L69 3BX, U.K.

## Supporting Information

**ABSTRACT:** The glycan receptor binding and specificity of influenza A viral hemagglutinin (HA) are critical for virus infection and transmission in humans. However, ambiguities in the interpretation of the receptor binding specificity of hemagglutinin from human- and avian-adapted viruses have prevented an understanding of its relationship with aerosol transmissibility, an exclusive property of human-adapted viruses. A previous conformational study, which we performed, indicated that human and avian receptors sample distinct conformations in solution. On the basis of detailed nuclear magnetic resonance (NMR) studies provided herein, we offer evidence of the distinct structural constraints imposed by hemagglutinin receptor binding sites on the glycan conformational space upon binding. The hemagglutinin from the SC18 virus, which has efficient aerosol transmissibility in humans (human-adapted), imposed the most stringent constraints on the conformational space of the human glycan receptor (LSTc), compared to single (NY18) or double (AV18) amino acid HA mutants, a property correlating to the ligand–HA binding strength. This relationship was also observed for the avian-adapted HA, where the high affinity binding partner, AV18, imposed the most stringent conformational constraints on the avian receptor, compared to those imposed by NY18. In particular, it is interesting to observe how different HAs when binding to human or avian glycosidic receptors impose significantly different conformational states, in terms of the states sampled by the glycosidic backbone and/or the entire molecule shape (linear or bent), when compared to the corresponding unbound glycans. Significantly, we delineate a “characteristic NMR signature” for the human adapted hemagglutinin (SC18) binding to human glycan receptors. Therefore, the conformational space constraints imposed by the hemagglutinin receptor binding site provide a characteristic signature that could be a useful tool for the surveillance of human adaptation of other (such as H7N9 and H5N1) deadly influenza viruses.



Influenza A viruses are among the most rapidly evolving pathogens with potential for new strains to adapt to human hosts and lead to a pandemic outbreak, with significant economic and public health impact.<sup>1,2</sup> The emergence of novel influenza strains such as 2009 H1N1 and 2010 H3N2 through genetic reassortment<sup>3–5</sup> poses a constant threat in terms of the evolution of various subtypes, including H5N1, H7N2, H7N7, H7N9, and H9N2, to generate a pandemic strain. The H5 and H7 subtypes, among others, are categorized as avian-adapted because they primarily circulate within birds, but through

contact with infected animals, they can jump species and infect humans. Some of these avian-adapted viruses, including H5N1 and H7N9, upon infection, can replicate efficiently in various human organs and lead to severe infection and mortality.<sup>6–11</sup> However, these avian-adapted subtypes are not capable of efficient human-to-human aerosol transmission,<sup>12,13</sup> a charac-

**Received:** March 19, 2014

**Revised:** May 30, 2014

**Published:** May 30, 2014

teristic feature of subtypes such as H1N1, H2N2, and H3N2, which are considered human-adapted.

A key factor governing human adaptation of the influenza A virus is the binding specificity of viral surface glycoprotein hemagglutinin (HA) to sialylated glycan receptors on the host cell surface [glycans terminated by  $\alpha$ -D-N-acetylneurameric acid (Neu5Ac)]. A canonical definition of this binding specificity based on the terminal sialic acid linkage has been used in the field in recent decades. HA from avian-adapted subtypes such as H5, H7, and H9 is known to bind specifically to glycans terminated by  $\alpha$ (2→3)-linked sialic acid [ $\alpha$ (2→3) glycans or avian receptors].<sup>14,15</sup> Meanwhile, HA from human-adapted subtypes such as H1, H2, and H3 is known to bind specifically to glycans terminated by  $\alpha$ (2→6)-linked sialic acid.<sup>16,17</sup> This definition based on sialic acid linkage alone, while useful for characterizing many influenza strains, has misclassified some notable strains in terms of their ability to effect efficient human-to-human respiratory droplet transmission,<sup>18–20</sup> in particular H5N1 and H7N9.

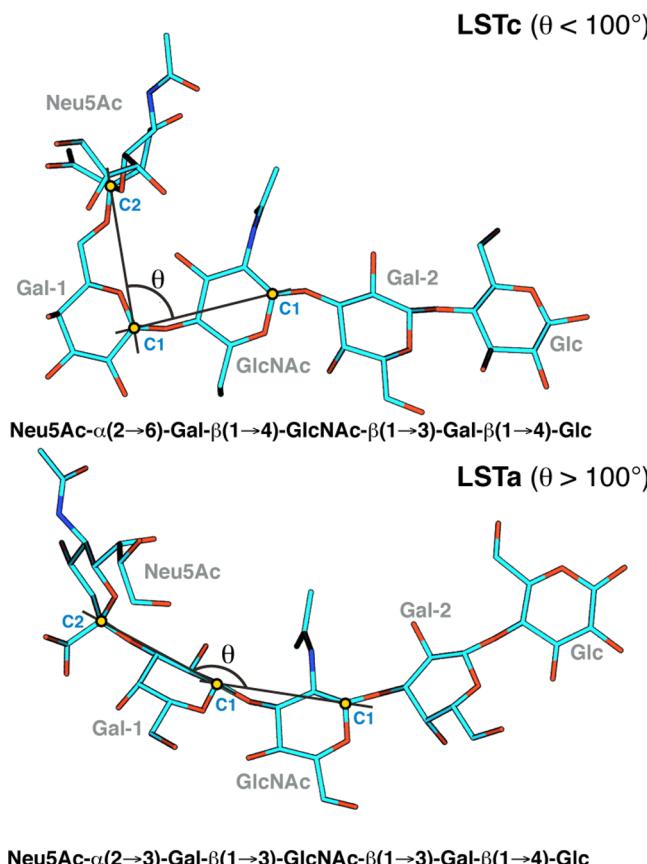
Glycan receptors have been defined according to their overall conformation, via a parameter called the  $\theta$  angle, which defines the form of the nonreducing end of the glycan receptors (Figure 1).<sup>18</sup> Specifically, in the case of influenza, we previously demonstrated that the apical surface of human upper respiratory epithelia, which is a primary target for human-adapted viruses, predominantly expresses glycan receptors with long oligosaccharide branches terminated by  $\alpha$ (2→6)-linked

sialic acid [ $\alpha$ (2→6) glycan or human receptors].<sup>18,21</sup> On the basis of analyses of glycan conformation and topology in HA-glycan X-ray cocrystal structures, we noted that the  $\alpha$ (2→6) glycans adopted an “umbrella-like” topology ( $\theta$  angle of  $<100^\circ$ ) in the receptor binding site (RBS) of H1 and H2 HAs, while  $\alpha$ (2→6) and  $\alpha$ (2→3) glycans adopted “cone-like” topologies ( $\theta$  angle of  $>100^\circ$ ) in the RBS of H5 HA.<sup>18</sup> This topology-based definition, in addition to the specific sialic acid linkage [i.e.,  $\alpha$ (2→3) vs  $\alpha$ (2→6)], distinguished HA from human-adapted subtypes binding to human receptors from the HA of avian-adapted viruses binding to avian and human receptors.<sup>22</sup>

The relationship between the glycan receptor specificity of HA and the aerosol transmissibility of the virus in ferrets (a well-established animal model for respiratory droplet transmission) was first demonstrated using the prototypic 1918 H1N1 pandemic HA (A/South Carolina/1/18 or SC18).<sup>23</sup> While SC18 showed efficient aerosol transmission in ferrets, a single amino acid mutation in HA (Asp225 → Gly; numbering based on H3 HA) resulted in a virus, NY18, that exhibited inefficient transmission, and a second further mutation, Asp190 → Glu, resulted in a virus, AV18, that could not be transmitted. We previously demonstrated that SC18 HA bound with high specificity and high affinity to human receptors, while AV18 HA bound with high affinity to avian receptors. NY18 bound to both avian and human receptors with a binding affinity substantially lower than that observed for AV18 and SC18. Despite dramatic differences in the aerosol transmissibility of these three viruses and their glycan receptor binding properties, they showed similar infectivity, replication efficiency, and tissue distribution in ferrets directly inoculated with virus. More recently, this approach was extended to the 2009 H1N1 system, revealing that the receptor specificity affects neither replication nor virulence of this pandemic virus in mice or ferrets, again after intranasal inoculation, but did affect animal-to-animal transmission by respiratory droplets.<sup>24</sup> These studies highlight the significance of HA mutations and glycan receptor binding specificity (given that all other genes among the three viruses are intact) in distinguishing aerosol transmissibility from other phenotypic properties of the virus such as infectivity and replication efficiency.

The dramatic changes in relative glycan receptor binding affinities and aerosol transmissibility resulting from single-amino acid changes to SC18 prompted an investigation into the structural nuances governing HA-glycan interactions. While HA-glycan X-ray cocrystal structures revealed differences in overall glycan topology (cone-like vs umbrella-like), when bound to different HAs, these “static” structures did not entirely capture the restrictions imposed on the conformational space of the glycan receptor by the RBS of different HAs, in moving from the free to HA-bound state.

For what is, to the best of our knowledge the first time, we present solution structures of LSTc [human receptor, Neu5Ac- $\alpha$ (2→6)-Gal- $\beta$ (1→4)-GlcNAc- $\beta$ (1→3)-Gal- $\beta$ (1→4)-Glc] and LSTA [avian receptor, Neu5Ac- $\alpha$ (2→3)-Gal- $\beta$ (1→3)-GlcNAc- $\beta$ (1→3)-Gal- $\beta$ (1→4)-Glc] bound to SC18, NY18, and AV18 HAs, based on comprehensive nuclear magnetic resonance (NMR) [saturation transfer difference spectroscopy (STD) and transferred nuclear Overhauser effect spectroscopy (tr-NOESY)] analyses and molecular dynamics (MD) simulations. The combination of both NMR and molecular dynamic simulations is particularly powerful when investigating protein-carbohydrate interactions, specifically when dealing with the highly mobile carbohydrate ligand.<sup>25</sup> We find that



**Figure 1.** Schematic of LSTc (top) and LSTA (bottom). The  $\theta$  angle parameter is defined by the angle across anomeric carbons of the three successive residues starting from the nonreducing end (Neu5Ac, Gal-1, and GlcNAc).

these “dynamic” structures are instrumental in delineating the conformational features of glycans when they are bound to HA. Our study captures the active changes in conformation,  $\theta$  angle, and glycosidic linkage torsional angles of LSTa and LSTc induced upon binding to HA and sheds light on distinct structural constraints imposed by the RBS that differ by one or two amino acids. We report the new finding of defining the restriction on the glycan conformational space and mobility of glycan bound to HA that, importantly, is not captured in the corresponding “snapshot” cocrystal structures. Importantly, these data are consistent with the observed differences in the biochemical binding affinities of these glycans for SC18, NY18, and AV18 HAs. We discuss the implications of these findings for our understanding of binding of HA to human receptors, specifically the significance of glycan conformation,  $\theta$  angle, and glycosidic torsional angles, and HA RBS interactions in determining the appropriate specificity and affinity. Our findings will greatly aid future studies aimed at delineating appropriate structural constraints on glycan topologies for other influenza subtypes (such as H5N1 and H7N9), including surveillance of adaption to humans by these deadly viruses.

## MATERIALS AND METHODS

**Cloning, Baculovirus Synthesis, Expression, and Purification of HA.** The soluble form of HA was expressed using the Baculovirus Expression Vector System (BEVS). SC18 (A/South Carolina/1/1918) baculovirus (generated from the pAcGP67-SC18-HA plasmid<sup>26,27</sup>) was a gift from J. Stevens. pAcGp67-NY18-HA and pAcGp67-AV18-HA plasmids were generated from pAcGP67-SC18-HA by Asp225 → Gly and Asp190 → Glu/Asp225 → Gly mutations, respectively. Mutagenesis was conducted using the QuikChange Multi Site-Directed Mutagenesis Kit (Stratagene). The primers used for mutagenesis were designed using the web-based program PrimerX (<http://bioinformatics.org/primerx/>) and synthesized by IDT DNA technologies (Coralville, IA). NY18 and AV18 baculoviruses were created from pAcGP67-NY18-HA and pAcGP67-AV18-HA constructs using a Baculogold system (BD Biosciences, San Jose, CA) according to the manufacturer’s instructions. The baculoviruses were used to infect 300 mL suspension cultures of Sf9 cells (BD Biosciences) cultured in BD BaculoGold Max-XP Insect Cell medium (BD Biosciences). These cultures were monitored for signs of infection and harvested 4–5 days postinfection. BEVS produces trimeric HA that provides multivalent binding to glycans. The soluble form of HA was purified from the supernatant of infected cells using the protocol described previously.<sup>27</sup> Briefly, the supernatant was concentrated using Centricon Plus-70 centrifugal filters (Millipore, Billerica, MA), and the trimeric HA was recovered from the concentrated cell supernatant using affinity chromatography with columns packed with Ni-NTA beads (Qiagen, Valencia, CA). Eluting fractions that contained HA were pooled and dialyzed overnight with 10 mM Tris-HCl, 50 mM NaCl buffer (pH 8.0). Ion exchange chromatography was then performed on the dialyzed samples using a Mono-Q HR10/10 column (GE healthcare, Piscataway, NJ). The fractions containing HA were pooled together and subjected to ultrafiltration using Amicon Ultra 100 K NMWL membrane filters (Millipore). The protein was then concentrated and reconstituted in PBS. The purified protein was quantified using Bio-Rad’s protein assay (Bio-Rad, Hercules, CA).

**NMR Analysis of SC18, NY18, and AV18 with LSTc and LSTa.** STD and tr-NOESY samples were prepared by washing

the proteins SC18, NY18, and AV18 (1 mg/mL) with a buffered solution [150 mM sodium chloride, 100 mM sodium phosphate, 0.3 mM *d*-EDTA, and D<sub>2</sub>O (pH 7.2)] using Amicon Ultra centrifugal filters, and a 10 kDa membrane (Millipore). Each ligand (LSTc or LSTa) was added to the corresponding protein sample, yielding final molar ratios of 100:1 (glycan receptor:HA) for the STD measurements and 25:1 (glycan receptor:HA) for the tr-NOESY measurements, and the protein concentration for the STD measurements was 0.01 mM and for the tr-NOESY experiments was 0.04 mM. NMR spectra were acquired using a Bruker 600 and 900 MHz AVANCE series NMR spectrometer, both equipped with a high-sensitivity 5 mm TCI cryoprobe. LSTc and LSTa resonances were previously assigned.<sup>28</sup> For the STD experiments, the on-resonance frequency was set at 7.3 ppm (6600 Hz) and the off-resonance frequency at 20.0 ppm (18000 Hz), a train of 40 Gaussian-shaped pulses of 50 ms each were applied to produce a selective saturation of 2 s, and D1 was 6 s. The number of scans was 1K, and the spectral width was 12626 Hz. The two-dimensional NOESY experiments were conducted using a mixing time of 300 ms; the data were recorded for quadrature detection in the indirect dimension and acquired using 16 scans per series of 2048 × 416 data points. The spectra were recorded at 295 K.

**Dose-Dependent Direct Binding of SC18, NY18, and AV18 to Glycan Receptors LSTa and LSTc.** LS-tetrasaccharide c [LSTc, Neu5Ac- $\alpha$ (2→6)-Gal- $\beta$ (1→4)-GlcNAc- $\beta$ (1→3)-Gal- $\beta$ (1→4)-Glc] and LS-tetrasaccharide a [LSTa, Neu5Ac- $\alpha$ (2→3)-Gal- $\beta$ (1→3)-GlcNAc- $\beta$ (1→3)-Gal- $\beta$ (1→4)-Glc] (Accurate Chemicals) were biotinylated with EZ-Link Biotin-LC-Hydrazide (Thermo) according to the manufacturer’s instructions. Streptavidin-coated High Binding Capacity 384-well plates (Pierce) were loaded to the full capacity of each well by incubating the well with 50  $\mu$ L of 2.4  $\mu$ M biotinylated LSTa or LSTc overnight at 4 °C. Excess glycans were removed through extensive washing with PBS. The trimeric HA unit comprises three HA monomers, and the spatial arrangement of the biotinylated glycans in the wells of the streptavidin plate array favors binding to only one of the three HA monomers in the trimeric HA unit. To specifically enhance the correct multivalency in the HA–glycan interactions, the recombinant HA proteins were precomplexed with the primary and secondary antibodies in a molar ratio of 4:2:1 (HA:primary:secondary). The identical arrangement of four trimeric HA units in the precomplex for all the HAs permitted comparison between their glycan binding affinities. A stock solution containing appropriate amounts of histidine-tagged HA protein, primary antibody (mouse anti-six-His tag IgG), and secondary antibody (HRP-conjugated goat anti-mouse IgG) (Santa Cruz Biotechnology, Santa Cruz, CA) was combined in a ratio 4:2:1 and incubated on ice for 20 min. Appropriate amounts of precomplexed stock HA were diluted to 250  $\mu$ L with 1% BSA in PBS; 50  $\mu$ L of this precomplexed HA was added to each of the glycan-coated wells and incubated at room temperature for 2 h followed by the wash steps described above. The binding signal was determined on the basis of HRP activity using the Amplex Red Peroxidase Assay (Invitrogen, CA) according to the manufacturer’s instructions. The experiments were conducted in triplicate. Minimal binding signals were observed in the negative controls, including binding of the precomplexed unit to wells without glycans and binding of the antibodies alone to the wells with glycans. The

data obtained from this analysis can be found in Figure 1 of the Supporting Information.

**Molecular Dynamics Simulations.** The dynamic and conformational properties of the interaction between LSTc and HA were studied by comparing the MD simulation trajectories of three complexes: LSTc–SC18, LSTc–NY18, and LSTc–AV18. The X-ray cocrystal structures of SC18 and NY18 were recently determined with LSTA and LSTc.<sup>29</sup> However, during the preparation of this paper, the available cocrystal structures were those of SC18 with LSTc [Protein Data Bank (PDB) entry 2WRG], a swine H1N1 HA (A/swine/Iowa/30) with LSTA and LSTc, and another human H1N1 (A/Puerto Rico/8/34, PDB entry 1RVX<sup>30</sup>) with LSTA. These X-ray cocrystal structures constituted reasonable starting models of HA–glycan complexes for MD simulations. The LSTc–SC18 complex was built starting with LSTc cocrystallized with SC18 HA [PDB entry 2WRG; coordinates were available for tetrasaccharide Neu5Ac- $\alpha$ (2→6)-Gal- $\beta$ (1→4)-GlcNAc- $\beta$ (1→3)-Gal-].<sup>31</sup> The protein, HA portion, of the complex was taken directly from the PDB coordinates, selecting a sequence of 60–260 amino acids (2WRG numbering) that includes the HA RBS. The solution conformation of LSTc, determined previously,<sup>28</sup> was superposed onto the cocrystallized glycan structure, with the nonreducing residues Neu5Ac and Gal-1 giving a root-mean-square deviation (rmsd) of 6.5 Å. The remaining complexes were obtained from the previously built LSTc–SC18 complex by applying the single mutation Asp225 → Gly *in silico* to generate the LSTc–NY18 complex and the double mutation Asp225 → Gly/Asp190 → Glu to generate the LSTc–AV18 complex. It is important to note that at the start of the MD simulation, the three complexes are characterized by exactly the same conformation of LSTc and HA with the exception of mutated residues.

The LSTA–AV18 complex was built by superimposing a previously selected conformation of LSTA<sup>28</sup> on the LSTA-like trisaccharides cocrystallized with H1 (A/Puerto Rico/8/34 H1N1) with amino acids within the binding site typical of an avian HA (PDB entry 1RVX). The rmsd between the superposed glycans is 6.2 Å, calculated on residues Neu5Ac and Gal-1. The previously built AV18 protein was superposed on the HA in PDB entry 1RVX, matching the protein Cα backbone (rmsd = 0.38 Å); then the complex was built by taking LSTA and AV18. The LSTA–NY18 complex was built from the latter complex (LSTA–AV18) by substituting Glu190 with Asp. Even in that case, the LSTA–AV18 and LSTA–NY18 model complexes have the same geometry at the beginning of the MD simulations.

Ambertools 1.4 was used to build the force field. GLYCAM06/Amber was used to describe the glycan and protein part of the complexes. The simulation cell was built by enveloping each macromolecule by a water layer (TIP3P water model) 15 Å wide in three directions, resulting in an orthogonal cell with an edge of approximately 100 Å. The nonbonded potential energy was described using the standard cutoff (12 Å) technique for both electrostatic and dispersive interactions. Each cell was minimized using 100 K steps of the default minimization algorithm included in NAMD. Then 1 ns of MD simulation sampling the NPT ensemble was used to equilibrate the cell density. The simulation temperature was set at 295 K and maintained by a Langevin thermostat as implemented in NAMD, while the Nosé-Hoover Langevin piston algorithm controlled the pressure (1.01325 bar) applied to the cell walls. During the minimization and cell density

equilibration steps, a harmonic potential energy restraint (harmonic constant of 50 kcal mol<sup>-1</sup>) on all the atoms of the complex was applied, while the water molecules were allowed to move freely. The MD simulation for all the modeled complexes was ~120 ns and was completed by applying a soft harmonic restraint on the HA backbone atoms (Cα, N, and carbonyl carbon) with a harmonic constant of 2.0 kcal mol<sup>-1</sup>. This allows the ligand and the side chain residues to be adjusted, while the secondary structure elements are maintained. The MD simulation trajectory was sampled every 10 ps, and the comparisons between the different complexes were conducted by monitoring selected distances between the ligand and the HA active site residues (Figure 4).

**MD Simulation Trajectory Principal Component Analysis (PCA).** PCA has been a powerful tool for the analysis of protein<sup>32,33</sup> and glycan<sup>28,34</sup> molecular dynamics. Although in these examples the PCA was used with the aim of extracting distinct and independent motional modes, it has been used in our study to identify the final bound states in the LSTx–HA complexes. Each frame of the glycan–HA MD trajectories was converted from Cartesian coordinates to a distance matrix, measured between the non-carbon and hydrogen atoms of the glycan (excluding the glycosidic linkage oxygens and including the carbon of the N-acetyl groups) and the amino acid side chains of HA, the last carbon in the amino acid side chain. A 6 Å cutoff was applied to the distance matrices; this means that only glycan–HA interactions were observed and not the glycan rearranging in solution, away from the RBS. These matrices were then converted into a vector, and all of them, for a single MD trajectory, were placed into a matrix; this data set was then mean-centered before principal component analysis could be performed. Density-based cluster analysis was performed on the first two component loadings (the most significant); the time each cluster appeared in the MD trajectory could then be compared. The first 20 ns of each trajectory was discarded; this left the time interval from 20 to 120 ns to be investigated, which was decomposed into 10000 distance matrices that were examined by PCA. This approach allows the evolution of the glycan–HA complex to be observed, from the initial state to the final state. For the final bound states, the  $\phi$ ,  $\psi$ , and  $\theta$  angles were determined as well as the average glycan–protein contacts for that subset of conformers. The glycan–protein contacts are represented as networks, with the edge thickness being inversely proportional to distance (the thicker the edge, the closer the vertices are); these can be found in Figure 5 and Figures 8–11 of the Supporting Information.

**Parameters That Define the Glycan Conformation and Topology.** The torsional angles ( $\phi$  and  $\psi$ ) are defined as the following pairs of dihedral angles:  $\phi_1$  and  $\psi_1$ ,  $\phi_2$  and  $\psi_2$ ,  $\phi_3$  and  $\psi_3$ , and  $\phi_4$  and  $\psi_4$  (starting from the nonreducing termini). For LSTA, the first pair is defined as C1–C2–O3–C3 ( $\phi_1$ ) and C2–O3–C3–H3 ( $\psi_1$ ), while for LSTc,  $\phi_1$  is the C1–C2–O6–C6 angle and  $\psi_1$  the C2–O6–C6–C5 angle, as previously defined by Xu et al.<sup>35</sup> Thereafter, successive pairs are defined as H1–C1–O4'–C4' ( $\phi_i$ ) and C1–O4'–C4'–H4' ( $\psi_i$ ) for the 1→4 linkage or H1–C1–O3'–C3' ( $\phi_i$ ) and C1–O3'–C3'–H3' ( $\psi_i$ ) for the 1→3 linkage. Atoms labeled with a prime belong to the monosaccharide on the reducing side of the glycosidic linkage, while atoms without a prime are on the nonreducing side of the glycosidic linkage. To be consistent with Chandrasekaran et al.,<sup>18</sup> Xu et al.,<sup>35</sup> and Sasaki et al.,<sup>28</sup> torsional angles  $\phi$  and  $\psi$  were illustrated in the range of -120°

**Table 1.** Glycosidic Torsional Angles for LSTc, Free and Bound to SC18, NY18, and AV18<sup>a</sup>

	linkage	cluster	size	percentage	$\phi$ (deg)	standard deviation (SD) of $\phi$ (deg)	$\psi$ (deg)	SD of $\psi$ (deg)
LSTc	1	1	9858	98.6	-57	11	190	17
		2	48	0.5	-71	12	-114	5
		3	73	0.7	-60	9	111	13
	2	1	5430	54.3	42	11	-3	11
		2	2419	24.2	-38	13	-29	11
	3	1	2151	21.5	24	13	-32	10
		2	4691	46.9	20	14	-47	13
		3	3981	39.8	39	15	34	18
	4	1	1328	13.3	-25	11	-43	11
		2	5934	59.3	39	13	-8	18
LSTc-SC18	1	1	2763	100.0	-58	7	189	7
		2	2763	100.0	50	9	-5	9
	3	1	2724	98.6	-2	21	-42	11
		2	39	1.4	40	7	25	8
	4	1	1192	43.1	45	10	0	11
		2	1069	38.7	19	13	-30	11
		3	502	18.2	-33	18	-28	11
LSTc-NY18	1	1	1826	100.0	-46	10	194	7
		2	1826	100.0	41	9	-7	11
	3	1	1826	100.0	17	14	-43	12
		4	1826	100.0	-31	13	-27	9
LSTc-AV18	1	1	2698	100.0	-51	8	192	7
		2	2698	100.0	-24	14	-32	11
	3	1	2698	100.0	24	13	34	11
		4	2698	100.0	-28	12	-24	9

<sup>a</sup>These angles for the ligand-receptor complexes were determined for those conformers that were extracted by PCA: g3, LSTc-SC18; g2, LSTc-NY18; and g2, LSTc-AV18 (Figure 6 of the Supporting Information). The angles were determined by cluster analysis of the data illustrated in Figure 13 of the Supporting Information, and the analysis used nonparametric density estimation<sup>36</sup> to determine the members of each cluster. These angles are average values for each  $\phi$  and  $\psi$  cluster.

to 240°. The torsional angles for the LSTc-HA and LSTA-HA complexes are listed in Tables 1 and 2, respectively. The angles were determined by cluster analysis of the data illustrated in Figures 13 and 14 of the Supporting Information, and the analysis used nonparametric density estimation<sup>36</sup> to determine the members of each cluster. The angles represented in the tables are the average values for each  $\phi$  and  $\psi$  cluster.

The topology of LSTA and LSTc is defined using a  $\theta$  angle parameter. The  $\theta$  angle is defined by the C2, C1, and C1 atoms of residues Neu5Ac, Gal, and GlcNAc (*N*-acetylgalactosamine), going from the nonreducing end to the reducing end (Figure 1).

## RESULTS

### NMR Structural Analyses of HA-Glycan Interactions.

NMR analyses were performed using a 900 MHz spectrometer, which permitted unprecedented resolution for observation of HA-glycan interactions. NMR STD experiments were employed to obtain a qualitative description of the glycan residues that interact with HA. NMR analysis indicated that for both LSTA and LSTc (Figure 1) in complex with the different HAs, the main sugar residue involved in the contact is the terminal sialic acid (Neu5Ac). However, there are differences in the mode of binding for these glycans to the various HAs.

To aid the assignment of the one-dimensional STD signals, spectra of LSTc bound to SC18 and NY18 were superimposed on the HSQC spectrum of LSTc (Figures 2A and 3A). While the interaction between LSTc and SC18 occurs primarily through the nonreducing end Neu5Ac moiety, involving

protons H3ax, H3eq, H4, H5, H7, and H9 (Figure 2A), signals belonging to Gal-1 (H6), GlcNAc (methyl group), Gal-2 (H4 and H1), and Glc (H5 and H6) are also present in the STD spectrum (Figure 2A). Similarly, LSTc interacts with NY18 principally through Neu5Ac, (H4, H5, H7, and H9), whereas signals belonging to Gal-2 and Glc appear to be absent or, at best, weaker than those found in the STD spectrum of the LSTc-SC18 complex. Notably, both methyl groups of GlcNAc and Neu5Ac still interact with NY18, but these signals are weaker than those observed in the LSTc-SC18 complex (Figure 3A).

The greater number of proton signals corresponding to the monosaccharides of LSTc that interact with SC18 compared to NY18 is consistent with the dose-dependent direct biochemical binding of these HAs to this glycan in an array platform (Figure 1 of the Supporting Information) and also with previously reported human receptor affinity of SC18 HA that is higher than that of NY18 HA.<sup>20</sup> The absence of STD signals observed for the LSTc-AV18 complex (Figure 3 of the Supporting Information) indicates that LSTc does not interact with AV18, consistent with minimal to no binding of AV18 to LSTc that is observed in the dose-dependent binding assay (Figure 1 of the Supporting Information).

The avian receptor, LSTA, on the other hand, interacts with AV18 and NY18 almost exclusively through Neu5Ac (Figure 2B). Because of the overlapping signals of the methyl groups belonging to Neu5Ac and GlcNAc, it was not possible to establish definitively which group interacts with the corresponding HA (Figure 3B). Among the few STD signals that

Table 2. Glycosidic Torsional Angles for LSTa, Free and Bound to NY18 and AV18<sup>a</sup>

	linkage	cluster	size	percentage	$\phi$ (deg)	SD of $\phi$ (deg)	$\psi$ (deg)	SD of $\psi$ (deg)
LSTa	1	1	8771	87.7	-62	10	-7	13
		2	1191	11.9	-81	9	-55	9
		3	21	0.2	230	6	47	8
		4	17	0.2	209	12	-26	8
	2	1	7317	73.2	41	13	-1	17
		2	1917	19.2	-37	13	-20	13
		3	766	7.7	61	12	57	10
	3	1	5939	59.4	39	16	32	19
		2	3290	32.9	20	13	-48	14
		3	771	7.7	-28	11	-44	13
	4	1	7861	78.6	40	13	-6	17
		2	2139	21.4	-34	14	-30	12
LSTa–AV18	1	1	2426	100	206	15	-20	8
		2	2426	100	51	9	19	15
		3	1118	46.1	49	9	53	13
		2	1308	53.9	25	12	42	12
	4	1	2426	100	54	22	5	21
		2	101	2.9	-73	10	-50	8
	2	1	2674	75.9	-40	11	-16	14
		2	484	13.7	45	11	1	12
		3	364	10.3	23	13	-26	10
LSTa–NY18	3	1	1939	55.1	22	15	-44	14
		2	1234	35	-24	17	-42	14
		3	349	9.9	47	14	14	19
		4	1760	50	-36	14	-30	12
	4	1	1762	50	35	15	-11	19
		2	1762	50	35	15	-11	19

<sup>a</sup>These angles for the ligand–receptor complexes were determined for those conformers that were extracted by PCA: g2–g5, LSTa–NY18; g1 and g2, LSTa–AV18 (Figure 7 of the Supporting Information). Unlike when LSTc is bound to HA, multiple bound states were evident when LSTa was bound to HA. These angles were determined by cluster analysis of the data illustrated in Figure 14 of the Supporting Information, and the analysis used nonparametric density estimation<sup>36</sup> to determine the members of each cluster. These angles are average values for each  $\phi$  and  $\psi$  cluster.

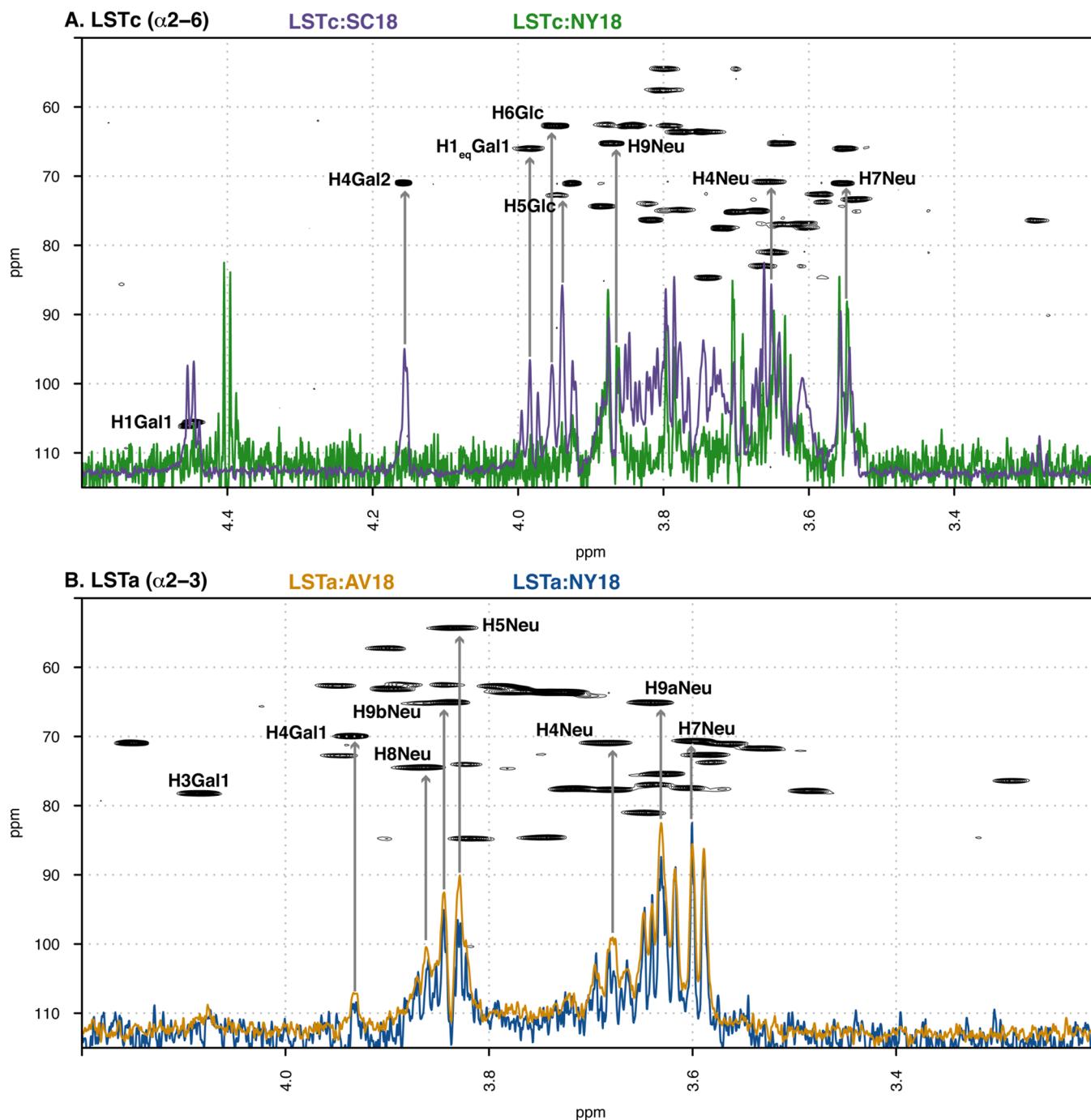
were observed for the interaction between LSTa and AV18 or NY18 that did not belong to NeuSAc, weak signals were observed for H3 and H4 of Gal-1, consistent with the partial involvement of the Gal-1 residue in binding.

Given the conformational flexibility of glycans, arising from the numerous glycosidic torsion angles, the parameter  $\theta$  has been defined as quantifying the form of the nonreducing end of the receptors (Figure 1 and Materials and Methods). To compare the conformational space sampled by the free and bound ligand, tr-NOESY experiments were performed for LSTc interacting with SC18 and NY18 (Figure 5 of the Supporting Information). Notably, via comparison of the results from bound and unbound LSTc, the NOE signals of the GlcNAc methyl group are substantially different. Whereas only the NOE between the methyl group of GlcNAc and H5 of NeuSAc was observed for unbound LSTc, additional NOEs are observed in the bound state with SC18, including signals associated with protons H8 and H9. These findings agree with a decrease in the distance between the GlcNAc ( $\text{CH}_3$  protons) and NeuSAc (H5 and H8/H9 protons), which is probably correlated with a reduction in the  $\theta$  angle between the free and bound state. This analysis indicates that, upon binding to SC18 and to a lesser extent NY18, LSTc undergoes a conformational change, reducing the  $\theta$  angle and subsequently the level of conformational freedom of the NeuSAc residue, particularly in the region between atoms C6 and C9.

**Molecular Dynamics Simulation of HA–Glycan Interactions.** The impact of glycan conformational differences and

points of glycan–HA interactions observed in the NMR experiments was further investigated using MD simulations, allowing a structural and dynamic comparison between the various complexes (LSTc–SC18, –NY18, and –AV18 and LSTa–AV18 and –NY18) to be made.

During the MD simulations, the conformational and dynamic properties of the HA–glycan complexes progressively change, especially those of the glycan conformation, with differences arising in the complexes due to the amino acid mutations within the RBS. It should be noted that the glycan starting geometries are the same in the two sets of MD simulations, with the LSTa and LSTc starting geometries taken from the solution forms identified by Sasaki et al.<sup>28</sup> The mobility observed for the glycan in the HA RBS was monitored using PCA, as described in Materials and Methods. Briefly, the analysis was performed on the distance matrix between the HA and receptor; unlike the conventional PCA of protein MD simulations, which uses the position of the protein backbone or the dihedral angles of the protein, this focuses the analysis on the interaction between the receptor and ligand. The temporal changes in the conformational state of the LSTc–SC18 complex are reported in Figure 6C of the Supporting Information; the conformer subset identified by cluster g3 represents the final bound state of the complex, and a similar representation for the LSTc–NY18 and LSTc–AV18 complexes is shown in Figure 6 of the Supporting Information. When LSTc is bound to any of the HAs, it appears to find a final singular state; this is not the case for the LSTa–AV18 or



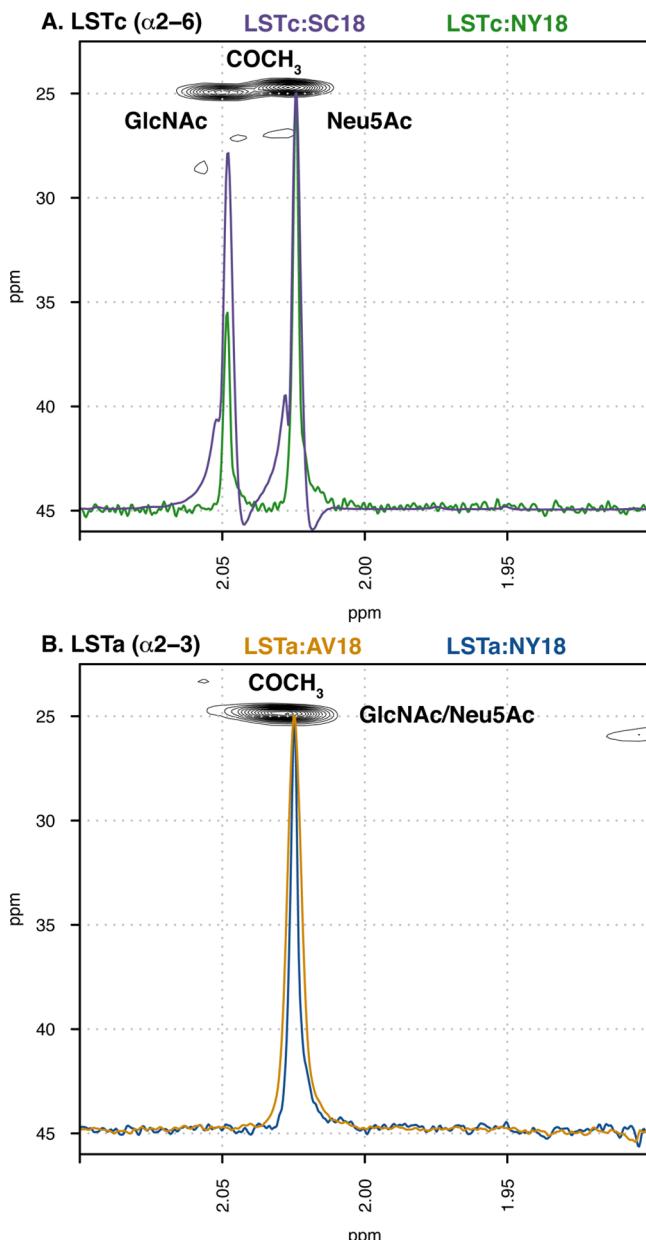
**Figure 2.** Main regions of the overlaid STD HSQC spectra of LSTc– and LSTa–receptor complexes. (A) STD spectra of LSTc–SC18 (purple) and LSTc–NY18 (green) complexes overlaid upon the HSQC spectrum of LSTc. (B) STD spectra of LSTa–AV18 (orange) and LSTa–NY18 (blue) complexes overlaid upon the HSQC spectrum of LSTa. The HSQC spectra and chemical shift assignments of LSTc and LSTa can be found in ref 28. Individual figures, with each STD spectrum plotted over the relevant glycan HSQC spectrum, can be found in Figures 2 and 4 of the Supporting Information.

LSTa–NY18 complex, where the final state of the latter MD simulation is represented by at least four conformational subsets (Figure 7C of the Supporting Information, clusters g2–g5) and the former has two final conformational subsets (Figure 7F of the Supporting Information, clusters g1 and g2), indicating a greater level of conformational freedom compared to that for the case in which LSTc interacts with HA.

Inspection of the MD trajectory indicates that, in the case of the LSTc–SC18 complex, all the monosaccharides of LSTc are positioned to interact with the RBS (Figure 5A and Table 1 of

the Supporting Information). It is also apparent that both hydrogen bonds and dispersive forces are important components in the interaction of LSTc with SC18; specifically, the methyl groups of GlcNAc show persistent contacts with Leu194 and Asp190, while the methyl group of Neu5Ac interacts with Gly134 and Trp153 (Figure 5A). Other noteworthy interactions within the LSTc–SC18 complex are between Gal-1 and Gln226, Lys222, and Asp225.

The presence of Gly instead of Asp at position 225, in going from wild-type SC18 to NY18, removes the hydrogen bond



**Figure 3.** N-Acetyl regions of the overlaid STD HSQC spectra of LSTc- and LSTA-receptor complexes. (A) STD spectra of LSTc-SC18 (purple) and LSTc-NY18 (green) complexes overlaid upon the HSQC spectrum of LSTc. (B) STD spectra of LSTA-AV18 (orange) and LSTA-NY18 (blue) complexes overlaid upon the HSQC spectrum of LSTA. The HSQC spectra and chemical shift assignments of LSTc and LSTA can be found in ref 28. Individual figures, with each STD spectrum plotted over the relevant glycan HSQC spectrum, can be found in Figures 2 and 4 of the Supporting Information.

interaction between Gal-1 (OH3 and OH4) and the RBS (specifically involving residues Gln226, Lys222, and Asp225), thereby promoting a greater distance between Gal-1 and HA (loop220) compared to that in SC18 (Figures 3B and 5B and Table 1 of the Supporting Information). This result is clearly supported by the NMR STD data (Figures 2 and 3), where Gal-1, GlcNAc, Gal-2, and Glc resonances are absent from the spectrum of the LSTc-NY18 complex or weaker than those in the spectrum of the wild-type complex, corresponding to a weaker interaction.

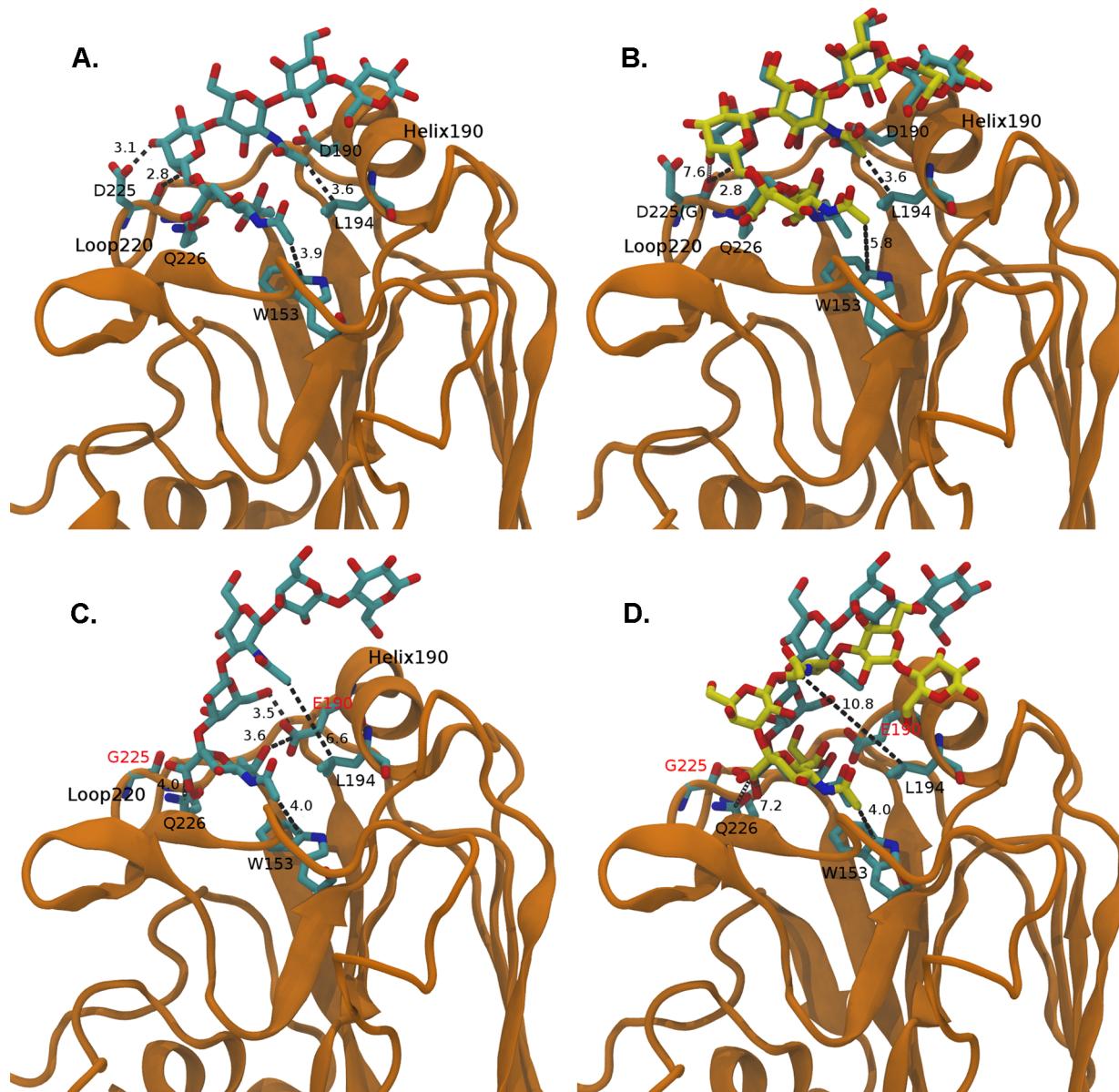
In the LSTc-AV18 complex, the further mutation of Asp190 to Glu introduces greater steric hindrance by preventing the optimal interaction between the reducing end of LSTc and helix190 of AV18, as can be seen by comparing the corresponding distances in the MD simulation trajectories of the LSTc-AV18 complex with those of the LSTc-SC18 and LSTc-NY18 reference complexes (Figure 5 and Table 1 of the Supporting Information). The consequence of this is that the interaction between Gal-1 and the HA RBS is re-established and Neu5Ac is drawn closer to the RBS, while the interaction of GlcNAc is weaker than that in the LSTc-SC18 and LSTc-NY18 complexes (specifically, Asp190 and Leu194 with GlcNAc-CMe), with the whole residue moving away from the RBS.

The analyses here reaffirm that it is interaction of Neu5Ac, Gal-1, and GlcNAc with HA that is important for human adaptation of influenza. These results provide a structural description of the effect caused by the single- and double-amino acid mutations in the RBS of HA, which correlates with a progressive weakening of the interaction between LSTc and HA (SC18 > NY18 > AV18), a result also supported by NMR STD experiments and the measured binding affinities.

In the case of the LSTA-AV18 complex, the primary contacts involve Neu5Ac of LSTA, with the carboxyl, acetyl, and sialyl groups of Neu5Ac interacting with the HA RBS (Figure 4C and Figure 8 and Table 2 of the Supporting Information). Importantly, contacts are also observed between AV18 and Gal-1 in LSTA. This involves the interaction of O6-Gal-2 with Glu190 and Pro186, which is afforded by the presence of the  $\alpha$ (2 $\rightarrow$ 3) linkage in LSTA. PCA of the LSTA-AV18 MD simulation trajectory also indicates the mobility of the reducing end region of LSTA is greater than that of its nonreducing terminus (Neu5Ac), which remains strongly attached to the RBS on the MD simulation time scale (Figures 7 and Figure 8 and Table 2 of the Supporting Information). In the case of the LSTA-NY18 complex, the mutation of Glu190 to Asp reduces the extent of interaction between Gal-1 and the RBS, with only the nonreducing end Neu5Ac residue interacting with the RBS. A consequence of this is that the number of interactions between the sialyl group of Neu5Ac and the RBS is decreased and the interaction between Gln226 and the carboxyl group of Neu5Ac is no longer observed (Figure 4D and Figure 9 and Table 3 of the Supporting Information). This corresponds to a weaker interaction between LSTA and NY18 than between LSTA and AV18, in agreement with the NMR STD experiments and biochemical assay. The LSTA-NY18 complex samples two conformational states during the MD simulation, which differ on the basis of contacts between Neu5Ac and the HA RBS (Figure 9 of the Supporting Information). The features of binding of LSTA to AV18 and NY18 obtained from the MD simulations are consistent with the corresponding NMR STD signals, where data from both complexes indicate that Neu5Ac is the main interacting residue with HA, while the methyl group of GlcNAc is no longer a key point of interaction with either NY18 or AV18.

We have compared the available X-ray crystallographic structures with our MD and NMR structures (Figures 10 and 11 of the Supporting Information). The comparisons have been limited to glycan-protein contact networks as the forms of the glycans within the crystal structures are distorted, precluding the determination of conformational angles.

**Dynamics of LSTA and LSTc Conformations upon Binding to HA.** The  $\theta$  angle parameter is a key conformational



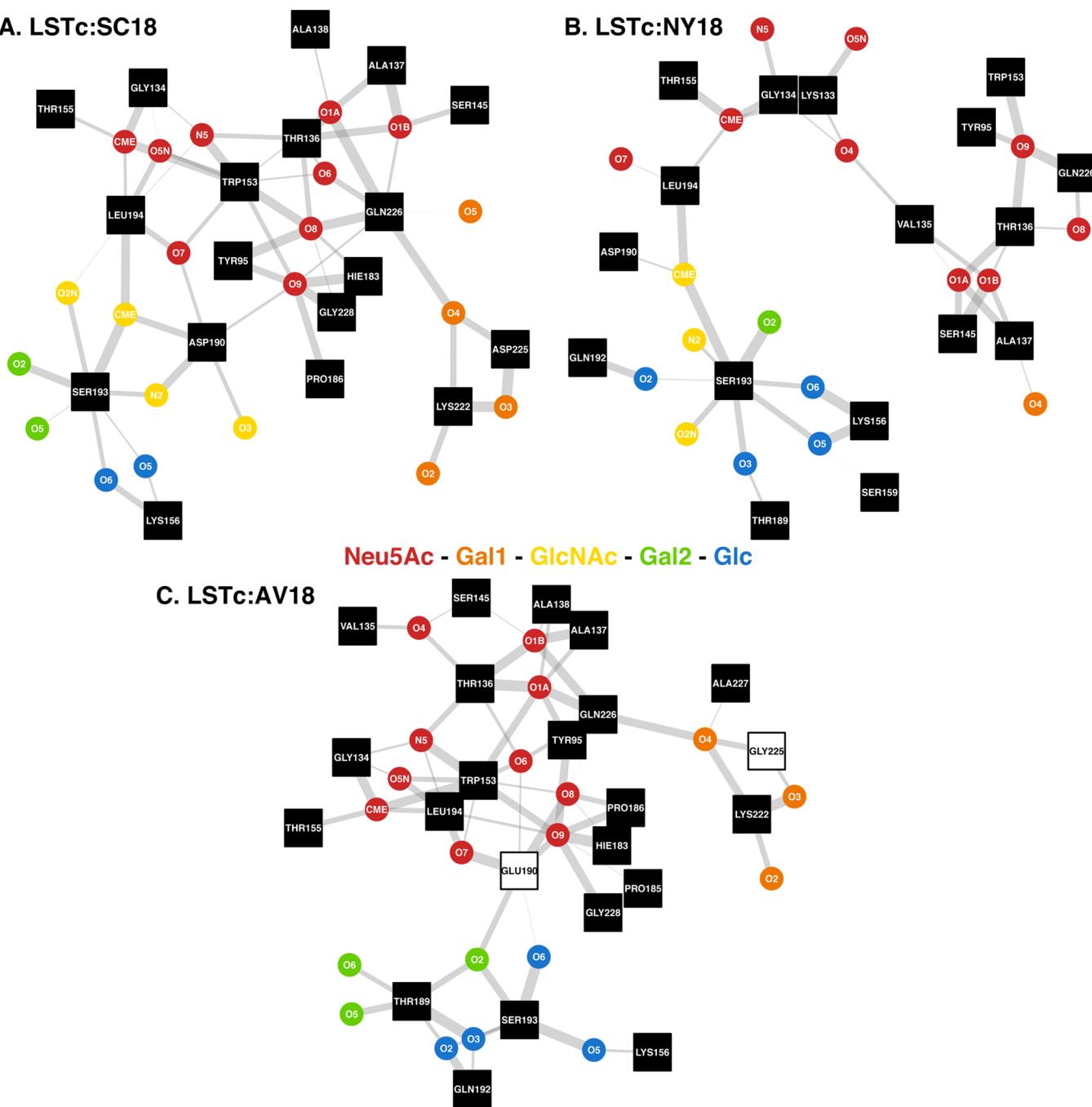
**Figure 4.** (A) Structure of the LSTc–SC18 complex in the g3 cluster conformation subset (Figure 6C of the Supporting Information). (B) Superposition of the RBSs of the LSTc–SC18 complex with that of the LSTc–NY18 complex ( $\text{rmsd} = 0.78 \text{ \AA}$ ). The reported complex structures belong to g3 and g2 conformations, respectively (Figure 6C,F of the Supporting Information). The carbon skeleton of LSTc bound to SC18 is colored cyan, while LSTc interacting with NY18 is colored yellow. (C) Structure of the LSTa–AV18 complex corresponding to conformation subset g2 of the PCA conformational characterization (Figure 7C of the Supporting Information). (D) Superimposed structures of the LSTa–AV18 and LSTa–NY18 complexes ( $\text{rmsd} = 0.74 \text{ \AA}$ ). The reported structure of the LSTa–NY18 complex corresponds to g2 and g1 conformation subsets as obtained from PCA (Figure 7F of the Supporting Information). LSTa linked to AV18 is colored cyan, while LSTa interacting with NY18 is colored yellow. The relevant amino acid residues of the shown HA active sites are underlined by a tube representation, with the name and numbering relative to PDB entry 2WRG. The reported distances are in angstroms.

descriptor of the nonreducing end of the glycan receptor, indicating the different forms of the glycans. As shown previously, the  $\theta$  angle of unbound LSTc had a predominant distribution at  $86^\circ$  and a smaller population located at  $119^\circ$  (Figure 6A).<sup>28</sup> Binding to SC18 substantially restricts the conformational population of LSTc, as reflected by the narrow distribution of the  $\theta$  parameter, with a single distribution centered at  $82^\circ$ . Binding to NY18 also restricts the conformational space sampled by LSTc, with a  $\theta$  angle distribution being located at approximately  $90^\circ$ . Interestingly, the small population of  $\theta$  angle values around  $119^\circ$  in the unbound LSTc is absent in the bound state, consistent with an earlier

study in which it was postulated that long  $\alpha(2\rightarrow6)$  glycans would predominantly adopt an umbrella-like topology (characterized by  $\theta < 100^\circ$ ) when bound to the RBS of human-adapted HAs.<sup>18</sup>

This restriction of the  $\theta$  angle when LSTc binds to SC18 and NY18 is supported experimentally by the new NOE signal appearing between GlcNAc and Neu5Ac H8/H9 in the tr-NOESY spectra of the LSTc–SC18 complex and by the stronger NOE signal between GlcNAc and Neu5Ac H5 in the LSTc–NY18 complex (Figure 5 of the Supporting Information). In contrast, low-affinity interaction of LSTc with AV18 predominantly samples the cone-like topology as indicated by

## A. LSTc:SC18



**Figure 5.** Contact network between LSTc and HA. If two vertices are linked, they are  $<6\text{ \AA}$  apart; the thicknesses of the graph edges are inversely proportional to the distance between the glycan atom (circular vertex) and the protein amino acid (square vertex) (the thicker the edge, the closer the two are together). The square vertices that are white with a black boundary indicate that this amino acid has been mutated, for example, Asp190 → Glu. The networks represent the average distances found in the final, “bound”, conformer subsets identified by PCA (Figure 6 of the Supporting Information): cluster g3, LSTc–SC18; cluster g2, LSTc–AV18; and cluster g2, LSTc–NY18. The distances illustrated here are listed in Table 1 of the Supporting Information. In this figure, the distinct modes of the interaction between LSTc and HA (SC18, NY18, and AV18) can be observed. When binding to its natural ligand, SC18, LSTc interacts with HA along its entire length. The single-point mutation forming NY18, Asp225 → Gly, drastically alters the interaction between the nonreducing end of the receptor and HA, and the major interactions with Gal-1 are abolished (panel B compared to panel A). The additional modification forming AV18, Asp190 → Glu, allows Gal-1 to re-engage with HA, which leads to the interaction between GlcNAc and amino acids 190 and Ser193 being lost (compare panel C to panels B and A).

the  $\theta$  distribution around  $113^\circ$ , which is consistent with the constraints imposed by avian-adapted HAs<sup>18</sup> and is due to the lack of interaction between GlcNAc and the RBS. These results re-emphasize the ability of human receptors to sample a distinct set of topologies (both umbrella-like and cone-like) in the RBS of SC18, NY18, and AV18 HA.

In the unbound state, the  $\theta$  angle of LSTA samples two distinct populations centered around  $118^\circ$  and  $154^\circ$  (Figure 6B), both of which correspond to cone-like topologies ( $\theta > 100^\circ$ ). The binding of LSTA to NY18 does not significantly alter the position of the  $\theta$  angle distribution compared to that of free glycan, distributions located at  $117^\circ$  and  $158^\circ$  in the

bound state, but does affect strongly their relative populations, with the population located at the smaller  $\theta$  angle being the largest. The binding of LSTa to AV18 HA imposes restrictions on the  $\theta$  angle, removing the population distributed around  $117^\circ$  (Figure 6B). Consistent with our previous studies, the avian receptor LSTa adopts exclusively a cone-like topology regardless of whether it is present in the unbound form or is bound to NY18 or AV18 HA. For the sake of completeness, the  $\theta$  angle density plots for the entire MD trajectory (20–100 ns) can be found in Figure 12 of the Supporting Information.

To complete our conformational studies, we also considered the glycosidic torsion angles of the free and bound glycan receptor sampled during the MD simulation; in the case of the bound receptors, the conformer subset extracted by PCA is considered (Tables 1 and 2).

First, consistent with the  $\theta$  angle distributions described above, comparison of the glycosidic torsional angles in the unbound and various HA-bound glycans indicates there are distinct structural constraints imposed by the RBS of SC18, NY18, and AV18 HA on LSTa and LSTc. When LSTc binds to SC18 and NY18, the distributions of states seen for  $\phi_1$  and  $\psi_1$ ,  $\phi_2$  and  $\psi_2$ , and  $\phi_3$  and  $\psi_3$  are very similar; only by using density cluster analysis is it possible to observe the different central locations of the states (Figure 13 and Table 1 of the Supporting Information). Indicating the modification of amino acid 190 (Asp  $\rightarrow$  Gly) and the subsequent change in the interaction between LSTc and HA, the loss of the interaction of Gal-1 with Gln226, Asp225, and Lys222 does not have a strong effect on the glycosidic torsional angles, whereas binding of LSTc to AV18 modifies amino acids 190 and 225, which eradicates the interaction between GlcNAc and Leu194, Asp190, and Ser193, affecting the positions of  $\phi_2$  and  $\psi_2$  and those of  $\phi_3$  and  $\psi_3$ , which are distinct from those observed in the LSTc–SC18 and LSTc–NY18 complexes (Figure 13 and Table 1 of the Supporting Information).

When LSTa binds to HA, the interactions with AV18 and NY18 are very different. The interaction between LSTa and NY18 is solely through the nonreducing end Neu5Ac (Figure 9 of the Supporting Information), and this is evident in the diversity of states observed for  $\phi_2$  and  $\psi_2$ ,  $\phi_3$  and  $\psi_3$ , and  $\phi_4$  and  $\psi_4$  (Figure 14 and Table 2 of the Supporting Information). The opposite is seen for the LSTa–AV18 complex; even though the glycan is dynamic with four conformational states being extracted by PCA of the MD trajectory of the complexes (Figure 7 of the Supporting Information), the glycosidic torsional angle states observed in the glycan are restricted for all four linkages.

## ■ DISCUSSION AND CONCLUSION

The NMR and MD simulation analyses described in this study offer new insights into the interaction between hemagglutinin and its glycan receptors, providing a detailed description of the contacts observed in the interactions between LSTc and LSTa with human- and avian-adapted HA and the consequent change in glycan conformation. The principal consequence of modifying SC18 to form NY18, Asp225  $\rightarrow$  Gly, is that Gal-1 of LSTc can no longer interact with the HA RBS (Gln226, Lys222, and amino acid 225), which allows Gal-1 to move away from the protein surface, also affecting the interaction between the Neu5Ac residue of LSTc and the RBS. A further second modification forming the avian-adapted AV18, Asp190  $\rightarrow$  Glu, permits Gal-1 of LSTc to re-engage with the HA RBS, interacting with Gly225 and Lys222. While GlcNAc can no

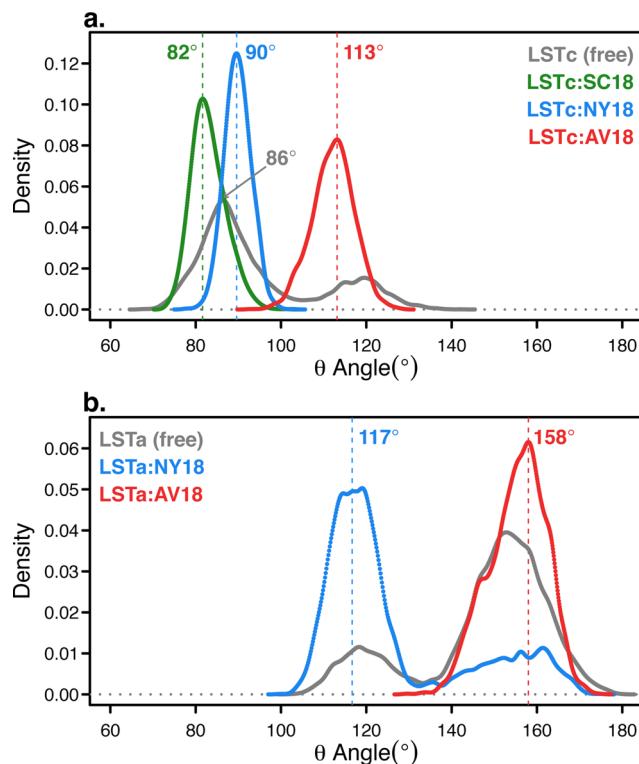
longer interact with the RBS in the LSTc–AV18 complex, it is this interaction that is a key decider for human adaption (Figure 5). The interaction between LSTa and AV18, avian-adapted HA, is principally via Neu5Ac and Gal-1 of the receptor. The nonreducing end  $\alpha(2\rightarrow3)$  linkage between Neu5Ac and Gal-1 in LSTa permits Glu190 and Pro186 to interact with the sialyl group of Neu5Ac and O6-Gal-1 of LSTa.

The interaction between the glycan receptors and HA imposes conformational constraints upon the glycan, which is characterized in terms of the  $\theta$  angle and the glycosidic torsional angles. We also correlated this analysis with biochemical HA–glycan binding specificity and affinity to ensure that this analysis is consistent with available crystal structure information. Importantly, using the  $\theta$  angle as a parameter to characterize the overall shape of the glycan, our study demonstrates key differences in the form of LSTa and LSTc when they are free versus their bound states (LSTc–SC18, –NY18, and –AV18 and LSTa–AV18 and –NY18), which have not been possible to “capture” through X-ray crystal structures.

Previously, we have noted that human receptors in their unbound state sample a conformational space that resembles both a cone-like topology ( $\theta \geq 110^\circ$ ) and an umbrella-like topology ( $\theta < 100^\circ$ ), whereas avian receptors exclusively sample a cone-like topology.<sup>28</sup> By analyzing HA–glycan cocrystal structures, we postulated that glycans binding to “avian-adapted” HA RBS (such as AV18) would impose constraints on the human receptor (for example, LSTc) to preferentially sample a cone-like topology, whereas glycans binding to “human-adapted” HA (such as SC18) would constrain the glycan to preferentially sample an umbrella-like topology.<sup>18</sup> On the basis of the NMR analyses and MD simulations presented here, we demonstrate the former hypothesis to be correct. The  $\theta$  angle of LSTc (Figure 6A) samples two distinct populations in the unbound state corresponding to umbrella-like and cone-like topologies. Upon binding to AV18 HA, LSTc predominantly samples a cone-like topology ( $\theta \sim 110^\circ$ ); this is due to GlcNAc of LSTc not being able to interact with the HA RBS. On the other hand, binding to the RBS of SC18 or NY18 imposes constraints on LSTc such that the glycan exclusively samples an umbrella-like topology.

Furthermore, compared to free LSTc, the bound glycan has restricted glycosidic torsional angles, with the torsional states observed for the LSTc–SC18 and LSTc–NY18 complexes being similar, whereas the mutation of Asp190 to Gly, seen in AV18, which abolishes the interaction between GlcNAc of LSTc and HA, produces unique glycosidic angles for  $\phi_2$  and  $\psi_2$  and for  $\phi_3$  and  $\psi_3$ , which are distinct from those seen in the LSTc–SC18 and LSTc–NY18 complexes. This restriction in the glycosidic torsional angles is also observed in the LSTa–AV18 complex, even though the receptor is dynamic in the RBS, as indicated by the multiple states observed at the end of the MD simulation (Figures 7 and 8 of the Supporting Information), whereas in the LSTa–NY18 complex, where the nonreducing end Neu5Ac is the sole point of interaction with HA, linkages  $\phi_2$  and  $\psi_2$ ,  $\phi_3$  and  $\psi_3$ , and  $\phi_4$  and  $\psi_4$  all have glycosidic torsional angles that are similar to those seen in the free glycan (Figure 14 of the Supporting Information).

This stronger restriction correlates with a greater number of LSTc–SC18 HA contacts, which can be observed in the NMR STD measurements and higher measured binding affinity compared to those of the LSTc–NY18 HA complex. The



**Figure 6.** Density distribution plots of the topological  $\theta$  angles of LSTc (A) and LSTa (B) free and bound to HA. The interaction of LSTc or LSTa with HA alters the topological  $\theta$  angle assumed by the glycan receptor. The  $\theta$  angles shown here are for the PCA-extracted conformers; a comparison of these with the  $\theta$  angles for all of the conformers can be found in Figure 12 of the Supporting Information.

constraints on the avian receptor, LSTa, are consistent with the predominant contacts made by HA RBS with the Neu5Ac- $\alpha$ (2 $\rightarrow$ 3)-Gal motif as measured by NMR STD signals, which permit a higher degree of flexibility to the sugars on the reducing end of this terminal motif (as seen in the conformational map of  $\phi_3$  and  $\psi_3$  and of  $\phi_4$  and  $\psi_4$  in Figure 14 of the Supporting Information).

The results presented here clearly indicate that even one or two amino acid changes in the HA RBS impose different constraints on the conformation and topology of bound glycan receptors, which in turn governs the biochemical binding specificity and affinity. On the basis of this evidence, it is important to carefully assess the effects of transferring amino acid changes that lead to specific receptor binding properties for a given HA to, and from, a completely different strain or subtype. As stated earlier, the receptor specificity of avian- and human-adapted HAs has been broadly classified solely on the basis of a preference for  $\alpha$ (2 $\rightarrow$ 3)- and  $\alpha$ (2 $\rightarrow$ 6)-linked sialic acid. On the basis of the data presented here, we find that the amino acid composition of the RBS of various avian-adapted HA subtypes would critically govern structural constraints imposed on diverse and distinct sets of glycans expressed in different tissues, consistent with the ability of viruses from HS, H7, and H9 subtypes to infect distinct tissue types. On the other hand, human-adapted HAs share the characteristic binding to glycans expressed in human upper respiratory epithelia (particularly nonciliated goblet cells).<sup>20,21,37</sup> This characteristic binding can be explained on the basis of the structural constraints imposed by SC18 and NY18 on LSTc, which, in turn, is reflected by the  $\theta$  parameter distribution.

Given that goblet cells secrete mucins, it is possible that the characteristic binding of human-adapted HAs to these cell types would increase their propensity for aerosol formation and transmission.

Finally, the methods and framework presented in this study to measure the restriction imposed by the RBS of different HAs on the conformational space and topology sampled by glycan receptors can serve as a very useful tool for allowing more exact surveillance of emerging influenza viruses such as H7N9 and H5N1, to closely monitor their ability to bind to human receptors and acquire the capability for human-to-human transmission.

## ASSOCIATED CONTENT

### S Supporting Information

Additional MD methods, supporting NMR spectra, and the multivariate analysis of the MD simulation data. This material is available free of charge via the Internet at <http://pubs.acs.org>.

## AUTHOR INFORMATION

### Corresponding Authors

\*E-mail: rams@mit.edu.

\*E-mail: guerrini@ronzoni.it.

### Author Contributions

S.E., E.M., and T.R.R. made complementary and equal contributions to this work.

### Funding

This work was funded in part by the National Institutes of Health (Grant R37 GM057073-13) and the National Research Foundation supported Interdisciplinary Research group in Infectious Diseases of SMART (Singapore MIT Alliance for Research and Technology). The 900 MHz spectra were recorded at the SONNMR Large Scale Facility in Utrecht, which was made possible by the financial support of the Access to Research Infrastructures activity in the seventh Framework Programme of the EC (Contract 261863, EU-NMR).

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We are very grateful for access to the NMR spectrometer (900 MHz) at the Utrecht NMR Facility and the assistance of Dr. Hans Wienk.

## REFERENCES

- 1) Ahmed, R., Oldstone, M. B., and Palese, P. (2007) Protective immunity and susceptibility to infectious diseases: Lessons from the 1918 influenza pandemic. *Nat. Immunol.* 8, 1188–1193.
- 2) Perez Velasco, R., Praditsithikorn, N., Wichmann, K., Mohara, A., Kotirum, S., Tantivess, S., Vallenas, C., Harmanci, H., and Teerawattananon, Y. (2012) Systematic review of economic evaluations of preparedness strategies and interventions against influenza pandemics. *PLoS One* 7, e30333.
- 3) Fraser, C., Donnelly, C. A., Cauchemez, S., Hanage, W. P., Van Kerkhove, M. D., Hollingsworth, T. D., Griffin, J., Baggaley, R. F., Jenkins, H. E., Lyons, E. J., Jombart, T., Hinsley, W. R., Grassly, N. C., Balloux, F., Ghani, A. C., Ferguson, N. M., Rambaut, A., Pybus, O. G., Lopez-Gatell, H., Alpuache-Aranda, C. M., Chapela, I. B., Zavala, E. P., Guevara, D. M., Checchi, F., Garcia, E., Hugonnet, S., and Roth, C. (2009) Pandemic potential of a strain of influenza A (H1N1): Early findings. *Science* 324, 1557–1561.
- 4) Itoh, Y., Shinya, K., Kiso, M., Watanabe, T., Sakoda, Y., Hatta, M., Muramoto, Y., Tamura, D., Sakai-Tagawa, Y., Noda, T., Sakabe, S.,

- Imai, M., Hatta, Y., Watanabe, S., Li, C., Yamada, S., Fujii, K., Murakami, S., Imai, H., Kakugawa, S., Ito, M., Takano, R., Iwatsuki-Horimoto, K., Shimojima, M., Horimoto, T., Goto, H., Takahashi, K., Makino, A., Ishigaki, H., Nakayama, M., Okamatsu, M., Warshauer, D., Shult, P. A., Saito, R., Suzuki, H., Furuta, Y., Yamashita, M., Mitamura, K., Nakano, K., Nakamura, M., Brockman-Schneider, R., Mitamura, H., Yamazaki, M., Sugaya, N., Suresh, M., Ozawa, M., Neumann, G., Gern, J., Kida, H., Ogasawara, K., and Kawaoka, Y. (2009) In vitro and in vivo characterization of new swine-origin H1N1 influenza viruses. *Nature* 460, 1021–1025.
- (5) Pearce, M. B., Jayaraman, A., Pappas, C., Belser, J. A., Zeng, H., Gustin, K. M., Maines, T. R., Sun, X., Raman, R., Cox, N. J., Sasisekharan, R., Katz, J. M., and Tumpey, T. M. (2012) Pathogenesis and transmission of swine origin A(H3N2)v influenza viruses in ferrets. *Proc. Natl. Acad. Sci. U.S.A.* 109, 3944–3949.
- (6) Chen, Y., Liang, W., Yang, S., Wu, N., Gao, H., Sheng, J., Yao, H., Wo, J., Fang, Q., Cui, D., Li, Y., Yao, X., Zhang, Y., Wu, H., Zheng, S., Diao, H., Xia, S., Chan, K. H., Tsoi, H. W., Teng, J. L., Song, W., Wang, P., Lau, S. Y., Zheng, M., Chan, J. F., To, K. K., Chen, H., Li, L., and Yuen, K. Y. (2013) Human infections with the emerging avian influenza A H7N9 virus from wet market poultry: Clinical analysis and characterisation of viral genome. *Lancet* 381, 1916–1925.
- (7) Fouchier, R. A., Schneeberger, P. M., Rozendaal, F. W., Broekman, J. M., Kemink, S. A., Munster, V., Kuiken, T., Rimmelzwaan, G. F., Schutten, M., Van Doornum, G. J., Koch, G., Bosman, A., Koopmans, M., and Osterhaus, A. D. (2004) Avian influenza A virus (H7N7) associated with human conjunctivitis and a fatal case of acute respiratory distress syndrome. *Proc. Natl. Acad. Sci. U.S.A.* 101, 1356–1361.
- (8) Gao, R., Cao, B., Hu, Y., Feng, Z., Wang, D., Hu, W., Chen, J., Jie, Z., Qiu, H., Xu, K., Xu, X., Lu, H., Zhu, W., Gao, Z., Xiang, N., Shen, Y., He, Z., Gu, Y., Zhang, Z., Yang, Y., Zhao, X., Zhou, L., Li, X., Zou, S., Zhang, Y., Yang, L., Guo, J., Dong, J., Li, Q., Dong, L., Zhu, Y., Bai, T., Wang, S., Hao, P., Yang, W., Han, J., Yu, H., Li, D., Gao, G. F., Wu, G., Wang, Y., Yuan, Z., and Shu, Y. (2013) Human infection with a novel avian-origin influenza A (H7N9) virus. *N. Engl. J. Med.* 368, 1888–1897.
- (9) Subbarao, K., and Katz, J. (2000) Avian influenza viruses infecting humans. *Cell. Mol. Life Sci.* 57, 1770–1784.
- (10) Wan, H., Sorrell, E. M., Song, H., Hossain, M. J., Ramirez-Nieto, G., Monne, I., Stevens, J., Cattoli, G., Capua, I., Chen, L. M., Donis, R. O., Busch, J., Paulson, J. C., Brockwell, C., Webby, R., Blanco, J., Al-Natour, M. Q., and Perez, D. R. (2008) Replication and transmission of H9N2 influenza viruses in ferrets: Evaluation of pandemic potential. *PLoS One* 3, e2923.
- (11) Watanabe, T., Kiso, M., Fukuyama, S., Nakajima, N., Imai, M., Yamada, S., Murakami, S., Yamayoshi, S., Iwatsuki-Horimoto, K., Sakoda, Y., Takashita, E., McBride, R., Noda, T., Hatta, M., Imai, H., Zhao, D., Kishida, N., Shirakura, M., de Vries, R. P., Shichinohe, S., Okamatsu, M., Tamura, T., Tomita, Y., Fujimoto, N., Goto, K., Katsura, H., Kawakami, E., Ishikawa, I., Watanabe, S., Ito, M., Sakai-Tagawa, Y., Sugita, Y., Uraki, R., Yamaji, R., Eisfeld, A. J., Zhong, G., Fan, S., Ping, J., Maher, E. A., Hanson, A., Uchida, Y., Saito, T., Ozawa, M., Neumann, G., Kida, H., Odagiri, T., Paulson, J. C., Hasegawa, H., Tashiro, M., and Kawaoka, Y. (2013) Characterization of H7N9 influenza A viruses isolated from humans. *Nature* 501, 551–555.
- (12) Belser, J. A., Blixt, O., Chen, L. M., Pappas, C., Maines, T. R., Van Hoeven, N., Donis, R., Busch, J., McBride, R., Paulson, J. C., Katz, J. M., and Tumpey, T. M. (2008) Contemporary North American influenza H7 viruses possess human receptor specificity: Implications for virus transmissibility. *Proc. Natl. Acad. Sci. U.S.A.* 105, 7558–7563.
- (13) Maines, T. R., Chen, L. M., Matsuoka, Y., Chen, H., Rowe, T., Ortin, J., Falcon, A., Nguyen, T. H., Mai le, Q., Sedyaningsih, E. R., Harun, S., Tumpey, T. M., Donis, R. O., Cox, N. J., Subbarao, K., and Katz, J. M. (2006) Lack of transmission of H5N1 avian-human reassortant influenza viruses in a ferret model. *Proc. Natl. Acad. Sci. U.S.A.* 103, 12121–12126.
- (14) Gambaryan, A. S., Tuzikov, A. B., Piskarev, V. E., Yamnikova, S., Lvov, D. K., Robertson, J. S., Bovin, N. V., and Matrosovich, M. N. (1997) Specification of receptor-binding phenotypes of influenza virus isolates from different hosts using synthetic sialylglycopolymers: Non-egg-adapted human H1 and H3 influenza A and influenza B viruses share a common high binding affinity for 6'-sialyl(N-acetyllactosamine). *Virology* 232, 345–350.
- (15) Russell, R. J., Stevens, D. J., Haire, L. F., Gamblin, S. J., and Skehel, J. J. (2006) Avian and human receptor binding by hemagglutinins of influenza A viruses. *Glycoconjugate J.* 23, 85–92.
- (16) Shinya, K., Ebina, M., Yamada, S., Ono, M., Kasai, N., and Kawaoka, Y. (2006) Avian flu: Influenza virus receptors in the human airway. *Nature* 440, 435–436.
- (17) Stevens, J., Blixt, O., Glaser, L., Taubenberger, J. K., Palese, P., Paulson, J. C., and Wilson, I. A. (2006) Glycan microarray analysis of the hemagglutinins from modern and pandemic influenza viruses reveals different receptor specificities. *J. Mol. Biol.* 355, 1143–1155.
- (18) Chandrasekaran, A., Srinivasan, A., Raman, R., Viswanathan, K., Raguram, S., Tumpey, T. M., Sasisekharan, V., and Sasisekharan, R. (2008) Glycan topology determines human adaptation of avian H5N1 virus hemagglutinin. *Nat. Biotechnol.* 26, 107–113.
- (19) Shriver, Z., Raman, R., Viswanathan, K., and Sasisekharan, R. (2009) Context-specific target definition in influenza A virus hemagglutinin-glycan receptor interactions. *Chem. Biol.* 16, 803–814.
- (20) Srinivasan, A., Viswanathan, K., Raman, R., Chandrasekaran, A., Raguram, S., Tumpey, T. M., Sasisekharan, V., and Sasisekharan, R. (2008) Quantitative biochemical rationale for differences in transmissibility of 1918 pandemic influenza A viruses. *Proc. Natl. Acad. Sci. U.S.A.* 105, 2800–2805.
- (21) Jayaraman, A., Chandrasekaran, A., Viswanathan, K., Raman, R., Fox, J. G., and Sasisekharan, R. (2012) Decoding the distribution of glycan receptors for human-adapted influenza A viruses in ferret respiratory tract. *PLoS One* 7, e27517.
- (22) Bewley, C. A. (2008) Illuminating the switch in influenza viruses. *Nat. Biotechnol.* 26, 60–62.
- (23) Tumpey, T. M., Maines, T. R., Van Hoeven, N., Glaser, L., Solorzano, A., Pappas, C., Cox, N. J., Swayne, D. E., Palese, P., Katz, J. M., and Garcia-Sastre, A. (2007) A two-amino acid change in the hemagglutinin of the 1918 influenza virus abolishes transmission. *Science* 315, 655–659.
- (24) Lakdawala, S. S., Shih, A. R., Jayaraman, A., Lamirande, E. W., Moore, I., Paskel, M., Sasisekharan, R., and Subbarao, K. (2013) Receptor specificity does not affect replication or virulence of the 2009 pandemic H1N1 influenza virus in mice and ferrets. *Virology* 446, 349–356.
- (25) Roldos, V., Canada, F. J., and Jimenez-Barbero, J. (2011) Carbohydrate-protein interactions: A 3D view by NMR. *ChemBioChem* 12, 990–1005.
- (26) Stevens, J., Blixt, O., Tumpey, T. M., Taubenberger, J. K., Paulson, J. C., and Wilson, I. A. (2006) Structure and receptor specificity of the hemagglutinin from an H5N1 influenza virus. *Science* 312, 404–410.
- (27) Stevens, J., Corper, A. L., Basler, C. F., Taubenberger, J. K., Palese, P., and Wilson, I. A. (2004) Structure of the uncleaved human H1 hemagglutinin from the extinct 1918 influenza virus. *Science* 303, 1866–1870.
- (28) Sasaki, G. L., Elli, S., Rudd, T. R., Macchi, E., Yates, E. A., Naggi, A., Shriner, Z., Raman, R., Sasisekharan, R., Torri, G., and Guerrini, M. (2013) Human ( $\alpha$ 2→6) and avian ( $\alpha$ 2→3) sialylated receptors of influenza A virus show distinct conformations and dynamics in solution. *Biochemistry* 52, 7217–7230.
- (29) Zhang, W., Shi, Y., Qi, J., Gao, F., Li, Q., Fan, Z., Yan, J., and Gao, G. F. (2013) Molecular basis of the receptor binding specificity switch of the hemagglutinins from both the 1918 and 2009 pandemic influenza A viruses by a D225G substitution. *J. Virol.* 87, 5949–5958.
- (30) Gamblin, S. J., Haire, L. F., Russell, R. J., Stevens, D. J., Xiao, B., Ha, Y., Vasishth, N., Steinhauer, D. A., Daniels, R. S., Elliot, A., Wiley, D. C., and Skehel, J. J. (2004) The structure and receptor binding properties of the 1918 influenza hemagglutinin. *Science* 303, 1838–1842.

- (31) Liu, J., Stevens, D. J., Haire, L. F., Walker, P. A., Coombs, P. J., Russell, R. J., Gamblin, S. J., and Skehel, J. J. (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.
- (32) Amadei, A., Linssen, A. B., and Berendsen, H. J. (1993) Essential dynamics of proteins. *Proteins* **17**, 412–425.
- (33) Kitao, A., and Go, N. (1999) Investigating protein dynamics in collective coordinate space. *Curr. Opin. Struct. Biol.* **9**, 164–169.
- (34) Gotsev, M. G., and Ivanov, P. M. (2009) Molecular dynamics of large-ring cyclodextrins: principal component analysis of the conformational interconversions. *J. Phys. Chem. B* **113**, 5752–5759.
- (35) Xu, D., Newhouse, E. I., Amaro, R. E., Pao, H. C., Cheng, L. S., Markwick, P. R., McCammon, J. A., Li, W. W., and Arzberger, P. W. (2009) Distinct glycan topology for avian and human sialopenta-saccharide receptor analogues upon binding different hemagglutinins: A molecular dynamics perspective. *J. Mol. Biol.* **387**, 465–491.
- (36) Azzalini, A., Menardi, G., and Rosolin, T. (2012) *R package pdfCluster: Cluster analysis via nonparametric density estimation*, version 1.0-0, Università di Padova, Padua, Italy.
- (37) Matrosovich, M. N., Matrosovich, T. Y., Gray, T., Roberts, N. A., and Klenk, H. D. (2004) Human and avian influenza viruses target different cell types in cultures of human airway epithelium. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 4620–4624.