Title

Mechanism of Glycan Receptor Recognition and Specificity Switch for Avian, Swine, and Human Adapted Influenza Virus Hemagglutinins: A Molecular Dynamics Perspective

Journal

Journal of the American Chemical Society, 131(47)

ISSN

0002-7863

Authors

Newhouse, E Irene
Xu, Dong
Markwick, Phineus RL
et al.

Publication Date

2009-12-02

DOI

10.1021/ja904052q

Peer reviewed

J|A|C|S ARTICLES

Published on Web 11/05/2009

Mechanism of Glycan Receptor Recognition and Specificity Switch for Avian, Swine, and Human Adapted Influenza Virus Hemagglutinins: A Molecular Dynamics Perspective

E. Irene Newhouse,*,† Dong Xu,*,‡ Phineus R. L. Markwick,§ Rommie E. Amaro,|| Hsing C. Pao,‡ Kevin J. Wu,‡ Maqsudul Alam,⊥ J. Andrew McCammon,‡,§,||,# and Wilfred W. Li‡

Maui High Performance Computing Center, Kihei, Maui, Hawaii 96753; National Biomedical Computation Resource, University of California—San Diego, La Jolla, California 92093-0505; Howard Hughes Medical Institute, University of California—San Diego, La Jolla, California 92093-0365; Department of Chemistry and Biochemistry and NSF Center for Theoretical Biological Physics (CTBP), University of California—San Diego, La Jolla, California 92093-0365; Department of Microbiology, University of Hawaii at Manoa, 2538 McCarthy Mall, Snyder 111, Honolulu, Hawaii 96822; and Department of Pharmacology, University of California—San Diego, La Jolla, California

Received May 19, 2009; E-mail: dxu@mccammon.ucsd.edu; einew@hotmail.com

Abstract: Hemagglutinins (HA's) from duck, swine, and human influenza viruses have previously been shown to prefer avian and human glycan receptor analogues with distinct topological profiles, pentasaccharides LSTa (α-2,3 linkage) and LSTc (α-2,6 linkage), in comparative molecular dynamics studies. On the basis of detailed analyses of the dynamic motions of the receptor binding domains (RBDs) and interaction energy profiles with individual glycan residues, we have identified ∼30 residue positions in the RBD that present distinct profiles with the receptor analogues. Glycan binding constrained the conformational space sampling by the HA. Electrostatic steering appeared to play a key role in glycan binding specificity. The complex dynamic behaviors of the major SSE and trimeric interfaces with or without bound glycans suggested that networks of interactions might account for species specificity in these low affinity and high avidity (multivalent) interactions between different HA and glycans. Contact frequency, energetic decomposition, and H-bond analyses revealed species-specific differences in HA−glycan interaction profiles, not readily discernible from crystal structures alone. Interaction energy profiles indicated that mutation events at the set of residues such as 145, 156, 158, and 222 would favor human or avian receptor analogues, often through interactions with distal asialo-residues. These results correlate well with existing experimental evidence, and suggest new opportunities for simulation-based vaccine and drug development.

Introduction

Influenza A virus (IAV) is classified based upon the serotypes of viral envelope glycoproteins hemagglutinin (HA) and neuraminidase (NA), of which 16 and 9 are known to date, respectively. The HA from the highly pathogenic avian influenza (HPAI) virus H5N1 was first isolated in 1961 from aquatic terns killed by the IAV subtype H5N3.[1] Since then, the HPAI virus H5N1 has killed or caused the extermination of millions of chickens and ducks. To date, more than 400 human cases have been confirmed, with an average fatality rate of 63%.[2] In contrast, the most devastating historic pandemic of influenza was the 1918 H1N1 outbreak with an estimated fatality rate of 2.5%. Subsequent pandemic outbreaks of the 1957 H2N2 and 1968 H3N2 killed about one million people each, but the estimated death rate is comparable to that of the seasonal flu, at 0.5−1%.[3] The most recent 2009 H1N1 pandemic influenza of swine origin is currently estimated to have killed thousands, with the death rate estimated to be higher than 1%.[2] The potential threat of a new highly lethal pandemic strain that may emerge through antigenic shift or genome reassortment has led to increased surveillance of IAV from various species, and continued interest in vaccine and drug development for prophylaxis and therapy.

HA is known to be responsible for the attachment of the virus to host cell membrane surface glycoproteins or glycolipids via multivalent interactions to the sialoglycans,[4−6] with measured binding affinity in the millimolar range.[7] Avian and human adapted IAV have been shown to prefer α-2,3 or α-2,6 linked glycans,

† Maui High Performance Computing Center.
‡ National Biomedical Computation Resource, University of California—San Diego.
§ Howard Hughes Medical Institute, University of California—San Diego.
|| Department of Chemistry and Biochemistry and NSF Center for Theoretical Biological Physics (CTBP), University of California—San Diego.
⊥ Department of Microbiology, University of Hawaii at Manoa.
Department of Pharmacology, University of California—San Diego.

(1) Becker, W. B. *J. Hyg. (London)* **1966**, *64*, 309–320.
(2) WHO 2009.
(3) Greger, M. *Crit. Rev. Microbiol.* **2007**, *33*, 243–299.
(4) Takemoto, D. K.; Skehel, J. J.; Wiley, D. *Virology* **1996**, *217*, 452–458.
(5) Skehel, J. J.; Wiley, D. C. *Annu. Rev. Biochem.* **2000**, *69*, 531–569.
(6) Wiley, D. C.; Skehel, J. J. *Annu. Rev. Biochem.* **1987**, *56*, 365–394.
(7) Sauter, N. K.; Bednarski, M. D.; Wurzburg, B. A.; Hanson, J. E.; Whitesides, G. M.; Skehel, J. J.; Wiley, D. C. *Biochemistry* **1989**, *28*, 8388–8396.

 10.1021/ja904052q CCC: $40.75

respectively.[8,9] HA interacts with the terminal SIA of glycan receptors at conserved residues 98, 136, 153, and to a lesser degree, 183 and 194 (H3 numbering, Figure S1 of the Supporting Information).[5,8] Mutations at residues 98, 183, and 194 are known to eliminate erythrocyte binding.[9] However, the human adaptation of influenza viruses may require different mutations in different HA subtypes, or even within the same subtype. For example, residues 190, 225 are known to control reversible binding specificity shifts for H1N1;[12] residues 226, 228 are important in avian to human adaptation for H2 and H3 subtypes;[13,14] residues 137, 192 increase α-2,6 receptor binding specificity in H5N1;[10] and residues 143, 186, 196 are important for clade-1 H5N1, but have variable effects in clade-2 H5N1 strains.[11] Similarly, residues 226 and 228 in H5 only enhance binding to biantennary α-2,6 linked glycans, and residues 190 and 225 in the same context abolished binding in glycan microarray studies.[12]

The specific SIA-GAL linkage types of glycan receptors do not always correlate well with host cell tropism, transmission, and virulence of IAV.[13,14] Recently, a mouse study using chimeric strains of avian HPAI H5N1 with mutually exchanged H5 and N1 indicated that increased affinity for synthetic α-2,6 linked sialoglycans actually decreased the systematic spread and lethality in mice, but not in chickens.[15] In a mouse model of the 1918 H1N1 infection, chimeric strains with the 1918 H1 and N1 within a contemporary human H1N1 background were highly pathogenic, regardless of the preference for sialoglycan linkages reported in glycan microarray studies.[16] In a ferret model of the 1918 H1N1, regardless of receptor binding specificity, the H1N1 strains have similar tissue pathogenicity to the infected primary hosts. Transmissibility from primary to secondary hosts, however, is found to depend not only upon binding of α-2,3 or α-2,6 receptors, but also the ability to bind long α-2,6 glycans.[17]

To accommodate such complexities, the topology of glycan receptors, not the specific linkage type, has been hypothesized to be an important determinant for the human adaptation of avian H5. Shorter glycans (di-, trisaccharides), regardless of linkage type, were proposed to adopt cone-like topology, whereas longer α-2,6 linked glycans were umbrella-like.[18] Cell surface glycans also contain complex modifications, including fucosylation, sulfation, and sialylation,[19] and the types and distribution of glycan receptors in different tissues are only beginning to be understood through the use of specific lectins,[14,20] and more recently mass spectroscopy studies.[18] Biochemical studies have suggested that the inner part (distal to the terminal SIA residue) of glycan receptors may play a role in the determination of species specificity as well.[21,22]

A better understanding of the mechanism of glycan receptor recognition and specificity requires new approaches that can take advantage of the increasing amount of complex biological evidence and petascale computational power. Molecular dynamics (MD) simulations offer a complementary, high-resolution, computational approach to investigate biological phenomena difficult to assess experimentally.[23] Both experiments[24] and MD simulations[25] of glycosylated HA indicated that viral- and host-glycan interactions may affect HA-receptor binding specificity and affinity. In one of our earlier works,[26] we completed an exhaustive set of comparative MD simulation of the pentasaccharides LSTa and LSTc, two known avian and human receptor analogues in complex with duck H3, H5, and swine H9.[8,27] While there was no direct experimental measurement of LSTa and LSTc binding affinity with the HA's studied, the predicted binding affinities reproduced known binding preferences to α-2,3 or α-2,6 linked glycans after taking into consideration of glycan conformational entropy.[26] In addition, a number of additional HA RBD residues interacted with the distal glycan residues of LSTa and LSTc in these simulations, with corresponding increases in interaction energy from these non-SIA interactions.[26] The secondary structural elements (SSE) of the receptor binding domains (RBD) of HA and Hn-LSTx (n = 3,3h,5,9 and x = a,c) are illustrated using H5-LSTx as an example in Figure 1. The RBD comprises residues 98 and 117 to 265 (H3 numbering, Figure S1 of the Supporting Information).

Here we report the computational analysis of the MD trajectories of duck and human H3, duck H5, and swine H9 using the pentasaccharides LSTa and LSTc as the avian (α-2,3 linked) and human (α-2,6 linked) receptor analogues. On the basis of detailed analyses of the dynamic motions of the RBD and interaction energy profiles with individual glycan residues, we have identified ~30 residue positions and SSE in the RBD that present distinct interaction profiles with the receptor analogues. These receptor-specific interaction profiles yield interesting clues to the mechanism of receptor recognition and species specificity switch. This study establishes a computational framework for the glycobiology of IAV, complementary to experimental studies, to elucidate the mechanistic details of the human adaptation of the influenza virus.

Materials and Methods

MD Simulations. The details of the MD simulations using an avian-adapted H3,[28] avian H5,[8] and swine H9[8] have been previously described.[26] The human H3 X31[27] (H3h) is set up in a similar fashion. Briefly, each simulation was started from the corresponding crystal structure listed in Table S1 of the Supporting Information. In the case of duck H5 and swine H9 where the coordinates for the trimeric HA were not available, the trimer was generated from a monomer using the crystal

(8) Ha, Y.; Stevens, D. J.; Skehel, J. J.; Wiley, D. C. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 11181–11186.
(9) Martin, J.; Wharton, S. A.; Lin, Y. P.; Takemoto, D. K.; Skehel, J. J.; Wiley, D. C.; Steinhauer, D. A. *Virology* **1998**, *241*, 101–111.
(10) Yang, Z. Y.; Wei, C. J.; Kong, W. P.; Wu, L.; Xu, L.; Smith, D. F.; Nabel, G. J. *Science* **2007**, *317*, 825–828.
(11) Yamada, S.; Suzuki, Y.; Suzuki, T.; Le, M. Q.; Nidom, C. A.; et al. *Nature* **2006**, *444*, 378–382.
(12) Stevens, J.; Blixt, O.; Tumpey, T. M.; Taubenberger, J. K.; Paulson, J. C.; Wilson, I. A. *Science* **2006**, *312*, 404–410.
(13) Matrosovich, M. N.; Matrosovich, T. Y.; Gray, T.; Roberts, N. A.; Klenk, H. D. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 4620–4624.
(14) Nicholls, J. M.; Chan, M. C.; Chan, W. Y.; Wong, H. K.; Cheung, C. Y.; Kwong, D. L.; Wong, M. P.; Chui, W. H.; Poon, L. L.; Tsao, S. W.; Guan, Y.; Peiris, J. S. *Nat. Med.* **2007**, *13*, 147–149.
(15) Yen, H. L.; Aldridge, J. R.; Boon, A. C.; Ilyushina, N. A.; Salomon, R.; Hulse-Post, D. J.; Marjuki, H.; Franks, J.; Boltz, D. A.; Bush, D.; Lipatov, A. S.; Webby, R. J.; Rehg, J. E.; Webster, R. G. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 286–291.
(16) Qi, L.; Kash, J. C.; Dugan, V. G.; Wang, R.; Jin, G.; Cunningham, R.; Taubenberger, J. K. *J. Virol.* **2009**, *83*, 3754–3761.
(17) 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.; Garcia-Sastre, A. *Science* **2007**, *315*, 655–659.
(18) Chandrasekaran, A.; Srinivasan, A.; Raman, R.; Viswanathan, K.; Raguram, S.; Tumpey, T. M.; Sasisekharan, V.; Sasisekharan, R. *Nat. Biotechnol.* **2008**, *26*, 107–113.
(19) Varki, N. M.; Varki, A. *Lab. Invest.* **2007**, *87*, 851–857.
(20) Shinya, K.; Ebina, M.; Yamada, S.; Ono, M.; Kasai, N.; Kawaoka, Y. *Nature* **2006**, *440*, 435–436.
(21) Gambaryan, A. S.; Tuzikov, A. B.; Pazynina, G. V.; Desheva, J. A.; Bovin, N. V.; Matrosovich, M. N.; Klimov, A. I. *Virology J.* **2008**, *5*, 85–94.
(22) Gambaryan, A.; Yamnikova, S.; Lvov, D.; Tuzikov, A.; Chinarev, A.; Pazynina, G.; Webster, R.; Matrosovich, M.; Bovin, N. *Virology* **2005**, *334*, 276–283.
(23) Woods, R. J. *Glycoconjugate J.* **1998**, *15*, 209–216.
(24) Ohuchi, M.; Ohuchi, R.; Feldmann, A.; Klenk, H. D. *J. Virol.* **1997**, *71*, 8377–8384.
(25) Kasson, P. M.; Pande, V. S. *Biophys. J.* **2008**, *95*, L48–50.
(26) Xu, D.; Newhouse, E. I.; Amaro, R. E.; Pao, H.; Cheng, L. S.; McCammon, J. A.; Li, W. W.; W.; Arzberger, P. *J. Mol. Biol.* **2009**, *387*, 465–491.
(27) Eisen, M. B.; Sabesan, S.; Skehel, J. J.; Wiley, D. C. *Virology* **1997**, *232*, 19–31.
(28) Ha, Y.; Stevens, D. J.; Skehel, J. J.; Wiley, D. C. *Virology* **2003**, *309*, 209–218.

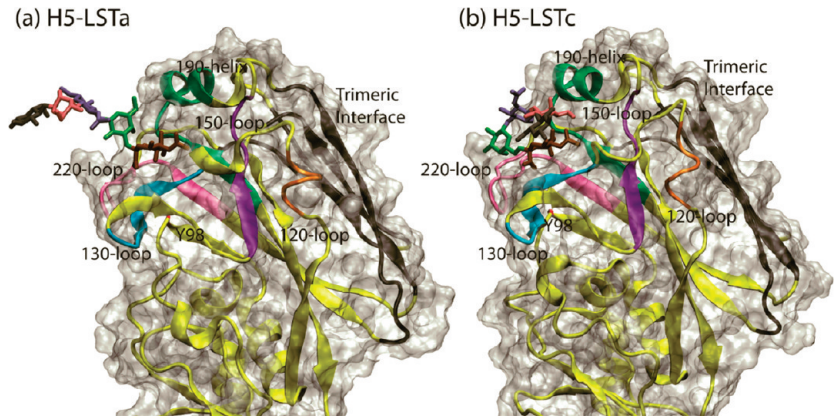

Figure 1. The RBD of H5 is shown in a yellow ribbon diagram with LSTa (a) or LSTc (b) bound. The major SSE are labeled and highlighted as follows: 120-loop (orange), 130-loop (cyan); 150-loop (blue/purple); 190-helix (green); and 220-loop (magenta). The trimeric interface β-strands are colored dark gray. The glycan residues are colored as follows: Sia1(ochre); Gal2 (lime); GlcNAc3 (blue); Gal4 (pink); and Glc5 (dark gray). Y 98 is shown in red (H3 Numbering).

symmetry in Chimera.[29] LSTa and LSTc are fully resolved in the crystal structures of H3h and H9, respectively, so the glycan structures were transferred to H3 and H5 by aligning the SIA of LSTx to that in the crystal structure using VMD.[30] MD was performed on the full trimeric HA using NAMD 2.6[31] with the FF99SB force field[32] for the protein and GLYCAM06 force field[33] for the glycans. Atom counts for the full systems are in Table S1 of the Supporting Information. Supercomputers at MHPCC, NBCR, and SDSC provided 0.5 (MHPCC) − 0.8 ns/day (NCBR and SDSC) on 64 or 128 processors, respectively. With 4 Hn-apo, and 8 Hn-LSTx systems, a total of two million CPU hours were used for the simulation and postanalysis.

MM-GBSA Binding Free Energy. Total binding free energy was calculated using the MM-GBSA scheme,[34,35] as reported previously.[26] A single trajectory approach was used for computational efficiency with 1000 snapshots at 40 ps interval. Surface tension (γ) of 0.0072 kcal/(mol Å), and the Hawkins−Cramer−Truhlar model (GB^{HCT})[36,37] with parameters described by Tsui and Case,[38] were used for the nonpolar solvation free energy calculation. To be consistent with experimental assay conditions, the salt concentration was set to 0.15 M, and the temperature to 310 K. Normal mode analysis (NMA) was used to compute solute translational, rotational, and vibrational entropy terms, using snapshots at the midpoint and end point of a trajectory. The ligand conformational entropy term was also included with 100 ns simulation of the free glycans to ensure adequate conformational sampling.

rmsd/RMSF/N−H Bond Reorientation Analyses. rmsd and RMSF analyses were carried using the Wordom utility[39] and VMD,[30] respectively, as previously described.[26] Errors were estimated from treating each monomer as a separate datum. The N−H bond reorientation order parameters were calculated using a previously described in-house program.[40] Errors were estimated as for RMSF.

Cluster Analysis. The average-linkage agglomerative method from the AMBER ptraj module was used to cluster the MD trajectories.[41] Snapshots were selected for analysis every 10 ps. To focus on the conformational changes including side-chain motions around the HA RBD, a clustering analysis was performed on all atoms (H3: 748 atoms, H5: 765 atoms, H9: 686 atoms) of the five SSEs (H3: 50, H5: 51, H9: 46 residues) in the RBD, as illustrated in Figure 1. A 3-cluster solution was obtained from each trajectory. Statistical criteria such as SSR/SST (sum of squares regression/total sum of squares) and critical distance show good clustering quality (SSR/SST > 98%) in the 3-cluster results.

Principal Component Analysis. For each Hn-apo system, the structures of the RDB domain collected across the 40-ns trajectory were superimposed in a mass-weighted backbone root-mean-square (rms) fitting procedure using the average structure as a reference. The covariance matrix was calculated and diagonalized to obtain the principal component eigenvectors using the coordinates of the C-α atoms. Each structure collected across the trajectory was then projected into the collective coordinate space defined by the two lowest principal component eigenvectors. In order to establish differences in the essential structural-dynamic properties of the RBD with or without the bound glycan, each Hn-LSTx system was superimposed onto the same Hn-apo reference structure and projected into the same collective coordinate space defined by the Hn-apo principal component eigenvectors. All calculations were performed using an in-house FORTRAN code. Errors in the RMSF and N−H

(29) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comput. Chem.* **2004**, *25*, 1605–1612.
(30) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graph* **1996**, *14* (33−8), 27–28.
(31) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781–802.
(32) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. *Proteins* **2006**, *65*, 712–25.
(33) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; Gonzalez-Outeirino, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. *J. Comput. Chem.* **2008**, *29*, 622–55.
(34) Karplus, M.; Kushick, J. N. *Macromolecules* **1981**, *14*, 325–332.
(35) Kollman, P.; Massova, A.; Reyes, I.; Kuhn, C.; Huo, B.; Chong, S.; Lee, L. T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E., III *Acc. Chem. Res.* **2000**, *33*, 889–897.
(36) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett.* **1995**, *246*, 122–129.
(37) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 19824–19839.
(38) Tsui, V.; Case, D. A. *Biopolymers: Nucleic Acid Sci.* **1995**, *56*, 275–291.
(39) Seeber, M.; Cecchini, M.; Rao, F.; Settanni, G.; Caflisch, A. *Bioinformatics* **2007**, *23*, 2625.
(40) Markwick, P. R. L.; Bouvignies, G.; Blackledge, M. *J. Am. Chem. Soc.* **2007**, *129*, 4724–4730.
(41) Shao, J.; Tanner, S. W.; Thompson, N.; Cheatham, T. E. *J. Chem. Theory Comput.* **2007**, *3*, 2312–2334.

reorientational analyses were estimated from treating each monomer as a separate datum.

Contact Frequency Analysis. VMD and shell scripts were used to count the occurrence of a HA RBD residue within a 5 Å cutoff of any glycan residue at each time step. The B-factor column of the corresponding HA PDB file was replaced by the calculated percent contact frequency for each residue. The color coding and visualization were done in VMD,[30] PMV,[42] or UCSF Chimera[29] where appropriate. The graphics was generated using Chimera.

Ensemble-Averaged Electrostatic Calculations. Ensemble-averaged electrostatics calculations were performed in VMD with the PME Electrostatics Plugin. Snapshots were taken every 5 ps over the course of the 40 ns simulation (8000 snapshots) and used in the calculations. The resulting images were created using VMD. Electrostatic surface maps were produced using Chimera[29] with the dominant cluster representatives from the cluster analysis. The partial charges and the electrostatics potential were calculated using PDB2PQR[43] and APBS[44] servers at NBCR. The molecular surface was calculated using MSMS.[45,46]

Energy Decomposition. To determine the individual enthalpic contribution to the interaction energy between HA and glycan residue pairs, complete trajectories were analyzed with NAMD, using a configuration file modeled after the sample configuration file provided in the NAMD User Guide. The interaction energy profiles for the glycans with different HA RBD residues were visualized using the Cytoscape software,[47] and the heat map feature of the ClusterMaker plugin.

H-Bond Analysis. Hydrogen bond analyses were carried out using the criteria that the H to acceptor distance is less than or equal to 2.5 Å and that the three atoms form an angle between 120° and 180°,[48] using a custom VMD script.[49] Only hydrogen bonds with occupancies >50% in at least one system were considered in the hydrogen bond analysis. The average occupancy from the three monomeric units with standard errors was reported.

Propensity Index. The propensity index was computed as the interaction energy difference with the interaction energy of LSTc subtracted from that of LSTa. If the index was less than the sum of the standard errors of the two measurements, then it was set to 0. This means that there is no significant preference for LSTa or LSTc, regardless of the actual strength of the interaction. A positive propensity index indicates that LSTc is favored (computed interaction is stronger) over LSTa, whereas a negative index indicates that LSTa is favored over LSTc.

Results

HA Receptor Binding Domains. HA is a heterogeneous protein with sequence identities of 38% for avian H3/H5, 37% for avian H3/H9, 45% in H5/H9, and 95% in avian H3/human H3h. The RBD of HA bound with LSTa (α-2,3) and LSTc (α-2,6) are illustrated with H5 in Figure 1. The RBD comprises several key structural components of HA1: the 130-loop (residue number 134−142, H3 numbering), 150-loop (150−156), 190-helix (181−193), and the 220-loop (220−230), with a number of conserved residues implicated in receptor binding and species specificity.[5] Figure S2 of the Supporting Information contains the alignments for all 148 RBD residues (residues 98 and 117 to 265, H3 numbering), only 40 of which are identical. Besides conserved residues in the SSE mentioned above, other conserved residues are mostly in the trimeric interface.[50] Residues 247−254 are most conserved with 7 out of 8 identical residues. In this study, we focus mainly on the functional roles and characteristics of the residues and SSE on the receptor binding interface.

System Stability and Correlation with Experimental Data. The convergence of the simulation systems within the limitation of computing power was examined using the MM-GBSA technique, and the total binding free energy was stable over the course of the simulation of 40 ns for each system.[26] The backbone rmsd of the HA's were generally under 2.5 Å (Figure S2 of the Supporting Information). Even though the simulations would have taken more than 200 years on a single core, the time scale achieved was still short compared to experimental time scale of milliseconds to seconds for HA-glycan interactions.

The H3h and LSTa binding constant has been determined experimentally.[7] As shown in Table S2 of the Supporting Information, the experimental K_d was equivalent to a free energy of binding of −3.30 kcal/mol. The predicted free energy of binding for H3-LSTa is −2.05 kcal/mol. No experimental data are available for H3h and LSTc, but α-2,6 sialyllactose is known to have a binding free energy of −3.65 kcal/mol with H3h. The predicted free energy of binding for H3h-LSTc was −2.37 kcal/mol. The relative difference between H3h-LSTx is small. In addition, we have found previously that the free energy binding difference from the current computational scheme is able to predict the relative binding preference for the avian and human receptor analogues (Table S1, S2 of the Supporting Information).[26] While the earlier study focused on the glycans in terms of topological differences when free or bound to different HA, we focus on changes in the proteins in the current work.

Principal Component Analysis and Cluster Analysis. We examined the different principal components plots of Cα atoms in the RBD (see Figure S3 of the Supporting Information for an example of human H3), and chose PC1 and PC2 for comparative purposes. Principal components 1 (PC1) and 2 (PC2) of the RBD for Hn-LSTx projected onto the components for HA-apo provide information on the effect of glycan binding on the conformational space sampling by the RBDs (Figure 2). H3-apo, H3h-apo, and H5-apo monomers (black) sampled separate conformational spaces, while H9-apo monomers sampled similar conformational space. H3 and H3h, with 95% identical sequence identity, had similar HA-apo plots (Figure 2, Column 1). All of the HAs with LSTx (LSTa, green; LSTc, red) bound were restricted in the sampled conformational spaces, compared to those of the respective HA-apo. There were subtle shifts in the conformational spaces sampled by Hn-LSTa (Figure 2, Column 2) and Hn-LSTc (Figure 2, Column 3). When the individual plots are overlapped (Figure 2, Column 3), the

(42) Sanner, M. F. *Structure (Cambridge, MA, U. S.)* **2005**, *13*, 447–462.
(43) Dolinsky, T. J.; Czodrowski, P.; Li, H.; Nielsen, J. E.; Jensen, J. H.; Klebe, G.; Baker, N. A. *Nucleic Acids Res.* **2007**, *35*, W522–525.
(44) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 10037–10041.
(45) Bhandarkar, M.; Brunner, R.; Chipot, C.; Dalke, A.; Dixit, S. It's a document distributed with NAMD, and is also available online from http://www.ks.uiuc.edu/Research/namd/2.7b1/ug/ there is no paper version published commercially.
(46) Sanner, M. F.; Olson, A. J.; Spehner, J. C. *Biopolymers* **1996**, *38*, 305–320.
(47) Cline, M. S.; Smoot, M.; Cerami, E.; Kuchinsky, A.; Landys, N.; et al. *Nat. Protoc.* **2007**, *2*, 2366–2382.
(48) Chachra, R.; Rizzo, R. C. *J. Chem. Theory Comput.* **2008**, *4*, 1526–1540.
(49) Gumbart, J. C.; Lo, D.; Trabucco, L.; Villa, E. *personal communication from J. C. Gumbart* 2008.
(50) Ha, Y.; Stevens, D. J.; Skehel, J. J.; Wiley, D. C. *Embo J.* **2002**, *21*, 865–875.

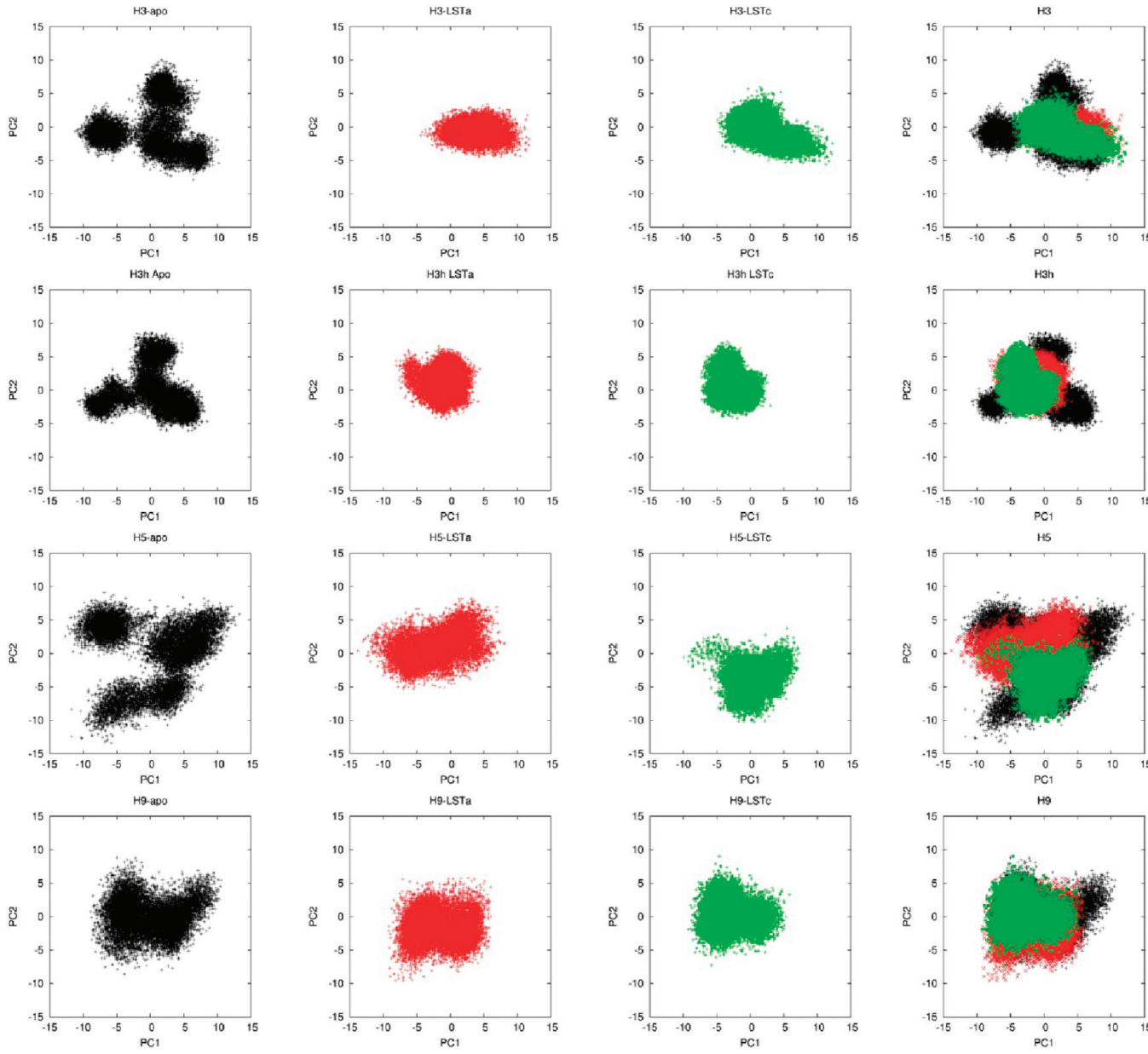

Figure 2. The projections of principal component 1 and 2 for the RBDs of Hn-LSTx (n = 3,5,9 and x = a,c) are shown separately or superimposed over those of H3, H3h, H5, and H9. The conformational space sampled by Hn is in black, Hn-LSTa in red, and Hn-LSTc in green.

conformational spaces sampled by H9-LSTx differed the least from their apo-forms. The conformational spaces sampled in H5-LSTx exhibited intermediate level of overlap, compared to H3-LSTx and H9-LSTx.

PCA eigenmode analyses of HA-apo forms showed that the maximum collective motions occurred in the 120-loop, 130-loop, and the 260-loop regions (Figure S4 of the Supporting Information). H3h also exhibited additional collective motion in the 220-loop, and in the trimeric interface T1 (residues 161−173), T2 (200−206) regions (Figure S1 of the Supporting Information). Similarly, swine H9 also exhibited greater collective motion in the trimeric interface region. The exact role of the changes in the collective motion at the trimeric interface is under further examination.

Representative Conformation Cluster Analysis. To examine the representative conformations adopted by the different HAs, we performed an all-atom clustering analysis of the RBD's. We selected a three cluster solution using the average linkage method (Table S3 of the Supporting Information) since the PCA analysis indicated that only a small number of conformational spaces were sampled. H3-apo and H9-apo were dominated by one cluster, whereas H5-apo spread into three clusters. In H3-LSTx, there is much higher occupation of the second cluster. The situation is similar for H3h. In H5-LSTa the first cluster has higher occupation than in H5-apo, and H5-LSTc spread out evenly among three clusters. The average rmsd between the three clusters for all the HA systems varied between 1.3 to 1.9 Å. The dominant cluster representatives are visualized with the four major SSEs colored separately (Figure S5 of the Supporting Information). These surface maps of the major SSEs of the RBD's indicated topological variations that are possible due to natural protein motions, as well as possible changes in the presence of the glycan receptors.

In the subsequent sections, we examined the differences between Hn-LSTx and Hn-apo using RMSF, backbone amide bond reorientation order parameter analysis, and compared with

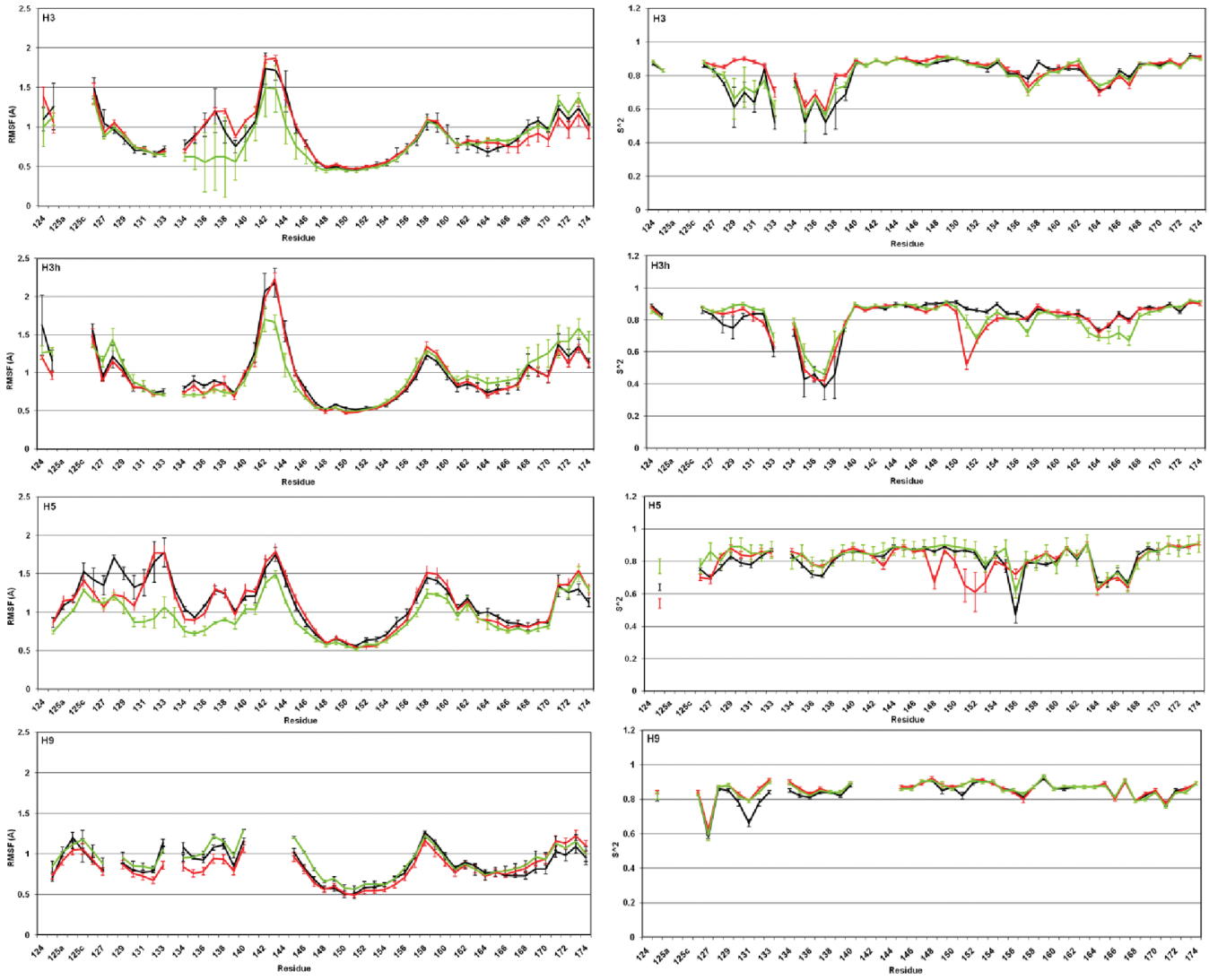

Figure 3. Comparison of backbone Cα RMSF variation and N−H bond reorientation analyses with different receptors bound or unbound. The region shown includes the 130 loop, 150 loop, and the T1 (Figure S1 of the Supporting Information) trimeric interface strand.

the experimental B-factors reported in the crystal structures (Figures S6, S7, and S8 of the Supporting Information).

Backbone Cα RMSF and N−H Bond Reorientational Motion Analysis. We used Cα atom RMSF analysis to observe the range of fluctuations for each residue along the RBD (Figure S6 of the Supporting Information). The order parameter, S^2, is given by $<P_2\cos\phi(t)>_{\lim t\to\infty}$. It is a value that varies from 1 (rigid) to 0 (rotation by 2π rads). The backbone N−H bond order parameter provides additional information compared to the RMSF analysis, describing the reorientation of the amide bond of the protein backbone (Figure S7 of the Supporting Information). These may be indicative of possible formation or breakage of hydrogen bonds that involve the backbone amide bond hydrogen atoms. Cα RMSF reduction may incur greater entropic penalty compared to backbone amide bond reorientation (BAR) without the corresponding RMSF reduction.

The RBD residues exhibiting large RMSF values are in good agreement with those with large crystal thermal factors—maxima and mimina occur at nearly the same residues (Figure S8 of the Supporting Information). However, as already suggested by the PCA analysis, there existed differences between the Hn-apo and Hn-LSTx in several subregions over the RBD. Most notably, the residue 133 in H5-LSTc exhibited 0.7 Å reduction in RMSF, whereas there were no differences (<0.1 Å) in the corresponding position in the other 3 HA-LSTc (Figure 3, panel A). Also, the presence of glycan ligands induced very little change in RMSF variation in the 150 loop region. However, H3h exhibited significant BAR at residues 151, 152 in the presence of LSTx, whereas H5 only did so in the presence of LSTa. Another significant BAR event occurred at residue 243 in H3h, not seen in any of the other HA (Figure S7 of the Supporting Information).

Contact Frequency Analysis. We set out to identify key residues that participate in glycan binding by plotting the contact frequency of individual RBD residues with any of the glycan residues (using a distance cutoff of 5 Å). Using the monomeric structures, the dominant cluster representatives of the Hn-LSTx/Apo were used to illustrate the per residue contact frequencies (Figure 4, Table S4 of the Supporting Information). The contact frequency offered a high level view of interactions possible through H-bonds, van der Waals interactions, and electrostatic interactions, though the last may be capable of interactions of a longer range than 5 Å. Each glycan formed a distinct contact pattern with individual HA, with key Sia1 contacts conserved. Residues 134−137, 153, 155, 190, and 194 had a contact frequency of 1, suggesting constant contact with the glycans

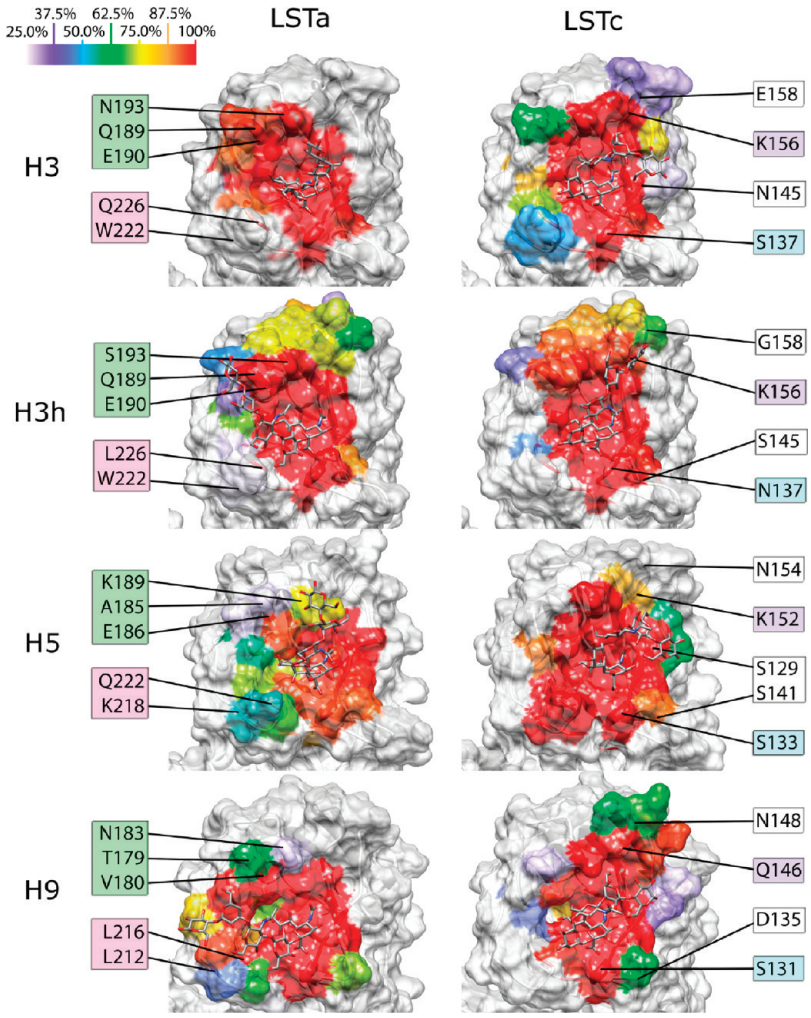

Figure 4. Contact frequencies with LSTa and LSTc by the RBD residues in the dominant cluster representative of H3, H5, or H9. Contact frequency from one monomer is illustrated, with values from 25 to 100%. Averaged values and standard errors are in Table S4 of the Supporting Information. These images may not reflect the average contact frequencies for some of the residues.

during the simulation. Those with contact frequencies less than 80% had as much as 30% standard error. This implied that the monomers within each trimer had quite different contact frequencies with those residues. Alternatively, more long-range electrostatic interactions might have compensated for the variations in distance. The distal asialoglycans contributed to the diverse interactions between HA and glycans.

Hydrogen Bonds. We analyzed the hydrogen bonds that formed dynamically during the MD simulation, and compared them with H-bonds previously reported from crystal structural studies[5,8] (Table 1). Six systems (all the HA-LSTa complexes and both H3 and H3h in complex with LSTc) possessed a hydrogen bonding interaction network limited to the primary Sia1/135−137 interaction sites, 98, and where applicable, 190. H5-LSTc had three unique persistent hydrogen bonds with Gal2 and GlcNAc3, and swine H9 had high occupancy H-bonds with the distal glycan residues Gal4 and Glc5. The standard errors for most of the hydrogen bond occupancies >80% were ∼10%, but for occupancies between 20 and 80% the errors were significantly larger. A number of H-bond interactions were identified by crystal structure studies, but not through this simulation study, and these were noted below Table 1.

Electrostatic Surface Maps. Electrostatics is predicted to play a major role in the glycan-HA binding energetic analysis.[26] We examined the possible contribution from the charged residues and mutations observed in the different HA (Figure 6). In H3, the change from E158 to G158 in H3h removed negative potential from H3h near the trimeric interface. In H5, K189 introduced positive potential, whereas D183 introduced negative potential in the 190 helix. Also, K218 introduced positive potential in the structurally equivalent position of W222 in H3. In H9 D135 introduced negative potential in the N145 position of H3.

Interaction Energy Profile. In order to assess the effect of the contacts made between the glycans and HA RBD quantitatively, we performed an interaction energy profile analysis for those residues that have been reported as important for glycan binding and/or species specificity. These include residues 98, 135−138, 153, 155, 186, 190−194, and 226−228.[8,12,28,51] Additional residues were studied if the contact frequency exceeded 50% in at least one of the eight systems studied. The interaction energy includes contributions from both electrostatic and van der Waals interactions. The values calculated, with standard errors, are listed in Tables S5−S8 of the Supporting Information. For ease of comparison, the data were visualized

(51) Russell, R. J.; Stevens, D. J.; Haire, L. F.; Gamblin, S. J.; Skehel, J. J. *Glycoconjugate J.* **2006**, *23*, 85–92.

Table 1. H-Bond Occupancies

			H-bond occupancies (%)							
			H3		H3 X-31		H5		H9	
glycan	role[a]	protein residue	LSTa	LSTc	LSTa	LSTc	LSTa	LSTc	LSTa	LSTc
Sia1 N5[b]	D	135main O	**85 (5)**[c]	**62 (27)**	**97 (0)**	**94 (3)**	**62 (19)**	**80 (8)**	**95 (2)**	**98 (1)**
Sia1 O7[d]	D	190 OE	36 (25)	27 (15)	43 (35)	2 (2)	**73 (23)**	0	n/a	n/a
Sia1 O8[d]	D	190 OE	41(24)	37 (15)	**57 (24)**[e]	31 (28)[e]	**70 (23)**[e]	0[e]	n/a	n/a
Sia1 O9[b]	D	190 OE	**52 (24)**	5 (5)	**61 (14)**	**55 (26)**	5 (3)	26 (21)	n/a	n/a
Sia1 O9[b]	D	226 OE	0	39 (21)	n/a	n/a	41 (16)[e]	**56 (20)**[e]	n/a	n/a
Sia1 O8[b]	A	98 OH	**56 (19)**	37 (20)	45 (25)	**70 (23)**	39 (14)	**92 (4)**	**69 (15)**	22 (11)
Sia1 O9	A	98 OH	**51 (21)**	26 (18)	26 (21)	1 (1)	33 (19)	4 (4)	43 (21)	29 (15)
Sia1 O1A	A	136 OG	25 (14)	12 (8)	36 (7)	49 (14)	**54 (7)**	0	37 (6)	12 (4)
Sia1 O1B[b]	A	136 OG	**75 (19)**	**52 (19)**	40 (7)	63 (23)	**39 (20)**	45 (27)	**63 (18)**	**64 (20)**
Sia1 O1A	A	137 OG	**79 (13)**	**57 (21)**	n/a	n/a	34 (11)	**69 (13)**	**70 (9)**	**81 (7)**
Sia1 O1B[b]	A	137 OG	36 (25)	21 (12)	n/a	n/a	39 (18)	**55 (25)**	30 (14)	28 (9)
Sia1 O1B[d]	A	137main N	**72 (8)**	**54 (26)**	**62 (2)**	**50 (21)**	47 (15)	**69 (12)**	34 (15)	**59 (7)**
Sia1 O1A[b]	A	137main N	**64 (1)**	**54 (24)**	**72 (5)**	**83 (0)**	**61 (12)**	**92 (3)**	**92 (5)**	**81 (7)**
Sia1 O9[b]	A	228 sideOG	0	0	12 (5)	31 (23)	n/a	n/a	n/a	n/a
Sia1 O9[d]	A	228main N	2 (0)	0	0[e]	0[e]	1 (0)	**58 (25)**	0	0
Gal2 O4	D	225main O	0	0	0[e]	0	0	**63 (12)**	0[e]	2 (1)
Gal2 O3	A	222 NZ	n/a	n/a	n/a	n/a	0	**58 (5)**	n/a	n/a
GlcNAc O6	D	190 OE	13 (13)	16 (13)	0	0	0	**71 (6)**	n/a	n/a
Gal4 O6[d]	D	156main O	0	0	0	0	0	0	0	**55 (7)**[e]
Gal4 O6[d]	A	155OG1	0	0	0	0	n/a	n/a	0	**70 (6)**
Glc5 O6	A	189main N	0	0	22 (2)	0	0	0	0	0
Glc5 O6[d]	A	135main N	0	0	0	0	0	0	0	**63 (19)**[e]

[a] Role: D = glycan is donor, A = glycan is acceptor. n/a = HA did not have this atom. [b] All glycan atom names are GLYCAM06 names, all protein atom names are standard PDB atoms names. Hydrogen-bonding with these residues described in ref 5. [c] Standard error in (), occupancies >50% are bold. [d] Hydrogen-bonding described in ref 8. [e] These H-bonds identified by ref 8 in some HA. Missing H-bonds when compared to these sources were with 131, 158, Gal4−156, 133-Glc5, in swine H9 with LSTc; 156-Glc5, 158-Glc5, 193-Glc5 in human H3 with LSTc; and 183-Sia1 in human H3 and avian H5 with LSTx.

with a heat map (Figures 5 and 6). Here we examined the enthalpic contributions to HA-glycan interactions; the calculated total free energy of binding were reported previously for H3, H5, and H9,[26] and for H3h in this study (Tables S1and S2 of the Supporting Information).

Sialic Acid Binding Sites. Y98 is critical to the binding of Sia1 and in our computations the interaction was conserved in all four HA. The primary interaction in all systems was between Sia1 and residues 136−137, which concurred with the structural analysis of Ha et al.[8] Other residues also calculated to be important for Sia1 binding included 145, 153, and 226, with smaller contributions from 156, 190, 193, and 222. Our energy decomposition indicates that Sia1-D145 in swine H9 was an unfavorable interaction (positive interaction energy, colored red in Figure 5); an effect unique to H9 among the HA studied, the result of electrostatic repulsion due to the negative charge on D145.

Asialoglycan Binding Sites. Gal2 had only a few interactions computed to be ∼−5 kcal/mol or greater: with E190 in H3-LSTa, H3h-LSTa, and H5-LSTa. GlcNAc3 had interactions of similar magnitude with 193 in H3-LSTc, 190 in H3h-LSTa, 193 in H5-LSTc, and 190 in H9-LSTc. In Gal4, the residues with the greatest interactions were 193 in H3-LSTc, 193 in H5-LSTc, and 155, 156, 190 in H9-LSTc. Glc5 had the strongest interactions with 156 with both LST in both H3 and H3h.

Propensity Index. The propensity index is shown in Figure 7. The values used to compute the propensity index, as well as their standard errors, are shown in Tables S5−S8 of the Supporting Information. Only those residues that had an index value with magnitude greater than the sum of the standard errors are shown. The characteristic patterns for each HA are immediately obvious in Figure 7. HA-Sia1 interactions were generally about equal in the LST, but those which were not, were, with one exception (H5−226), stronger with LSTa. With Gal4 and Glc5, most of the interactions differing between LST were calculated to be stronger in LSTc.

Ensemble-Averaged Electrostatic Potential Isosurface Maps. These maps showed the aggregate potential acting on approaching ligands. Extensive MD simulations have been shown to reproduce accurately experimentally observed changes in protein electrostatic field.[52] Furthermore, protein−ligand complexes whose association is enhanced many fold over diffusion control by electrostatic steering have been identified.[53] The RBD for all the HA studied here had positive electrostatic potential, though that of H9 was weaker than those of the other HA. The HA were significantly dipolar, with well-separated positively and negatively charged regions, though again, the separation was less in H9.

Discussion

MD Simulation of Glycan−Protein Interactions. Oligosaccharide binding to the lectin Concanavalin A or bacterial polysaccharide binding to antibody have been simulated and validated against experimental data using MM-PBSA[54] and MM-GBSA techniques.[55] Glycan binding to HA has a dissociation constant $K_d \approx 2-3$ mM.[56] Some MD on glycan interactions in HA have been published,[25,57−59] but these studies were for

(52) Suydam, I. T.; Snow, C. D.; Pande, V. S.; Boxer, S. G. *Science* **2006**, *313*, 200–204.
(53) McCammon, J. A. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 7683–7684.
(54) Bryce, R. A.; Hillier, I. H.; Naismith, J. H. *Biophys. J.* **2001**, *81*, 1373–1388.
(55) Kadirvelraj, R.; Gonzalez-Outeirino, J.; Foley, B. L.; Beckham, M. L.; Jennings, H. J.; Foote, S.; Ford, M. G.; Woods, R. J. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 8149–8154.
(56) Hanson, J. E.; Sauter, N. K.; Skehel, J. J.; Wiley, D. C. *Virology* **1992**, *189*, 525–533.
(57) Frank, M.; Lieth, C. W. *J. Mol. Model.* **1997**, *3*, 408–414.

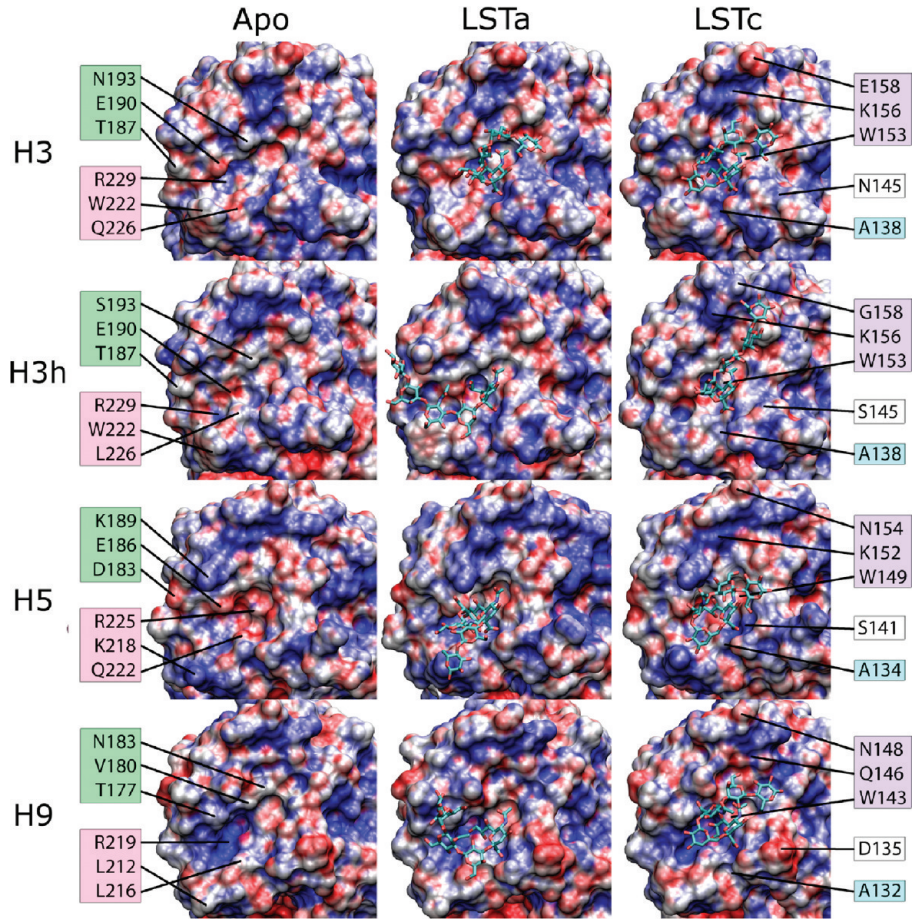

Figure 5. Electrostatic surface of major cluster representatives. Blue indicates positive on the electrostatic potential, and red indicates negative electrostatic potential. The residues are labeled according to the SSE where they are located using the same coloring scheme as Figure S4 of the Supporting Information.

shorter times scales, shorter glycans, or a single HA. MD has also been used to elucidate interactions with comparable strength to glycan-HA, for instance, the effects of anesthetics[60] and melatonin interaction with calmodulin.[61] Although we are unable to simulate these systems on physiologically relevant time scales (milliseconds), our approach attempts to combine the known sampling advantages from multicopy MD simulations[62,63] (i.e., simulating the full trimeric HA structures), while simulating long enough to sample dynamical events that may occur on 40 ns time scales, such as hydrogen bond reorientation, side chain motion, and substantial ligand, solvent, and flexible loop rearrangements, which are typically inaccessible via static crystal structures. Although we are unable to simulate the twelve systems presented here on physiologically relevant time scales, they are part of a productive cycle of interactions between "dry" and "wet" laboratories to gain insight into infectious diseases and other health related issues.

The HA RBD comprises four major SSEs and four β-strands forming the trimeric interface. This region includes three of the four known epitopes, with site C near the trimeric interface region, and part of the newly discovered epitope that offered cross-subtype protective antibodies.[64,65] Significant sequence variations in the RBD SSE, some likely coordinated due to selective pressure from the host immune response, help the virus evade antibody recognition successfully.[66] Some of these changes include the introduction of *N*-linked glycosylation sites that help mimic host glycome. The trimeric interface regions are also variable between subtypes. Both the HA RBD and the trimeric interface regions remain potential sites for small molecule inhibitor design, including the use of glycomimetic compounds, to complement vaccine development for the prophylaxis and therapy of influenza infection. In this study, we focused on HA in an attempt to characterize the behaviors of HA from different species interacting with glycan receptor analogues.

Human Adaptation and Swine Flu. Swine H9 RBD exhibited the least change in conformational space sampling in the PCA analysis (Figure 2) and predominantly showed a single cluster in the all-atom average linkage cluster analysis (Table S3 of

(58) Li, M.; Wang, B. *Biochem. Biophys. Res. Commun.* **2006**, *347*, 662–668.
(59) Auewarakul, P.; Suptawiwat, O.; Kongchanagul, A.; Sangma, C.; Suzuki, Y.; et al. *J. Virol.* **2007**, *81*, 9950–9955.
(60) Liu, Z.; Tang, P. *Biophys. J.* **2005**, *88*, 3784–3791.
(61) Turjanski, A. G.; Estrin, D. A.; Rosenstein, R. E.; McCormick, J. E.; Martn, S. R.; Pastore, A.; Biekofsky, R. R.; Martorana, V. *Protein Sci.* **2004**, *13*, 2925–2938.
(62) Caves, L. S.; Evanseck, J. D.; Karplus, M. *Protein Sci.* **1998**, *7*, 649–666.
(63) Monticelli, L.; Sorin, E. J.; Tieleman, D. P.; Pande, V. S.; Colombo, G. *J. Comput. Chem.* **2008**, 29.
(64) Ekiert, D. C.; Bhabha, G.; Elsliger, M. A.; Friesen, R. H.; Jongeneelen, M.; Throsby, M.; Goudsmit, J.; Wilson, I. A. *Science* **2009**, *324*, 246–251.
(65) Sui, J.; Hwang, W. C.; Perez, S.; Wei, G.; Aird, D.; et al. *Nat. Struct. Mol. Biol.* **2009**, *16*, 265–273.
(66) Kasson, P. M.; Pande, V. S. *Pac. Symp. Biocomput.* **2009**, 492–503.

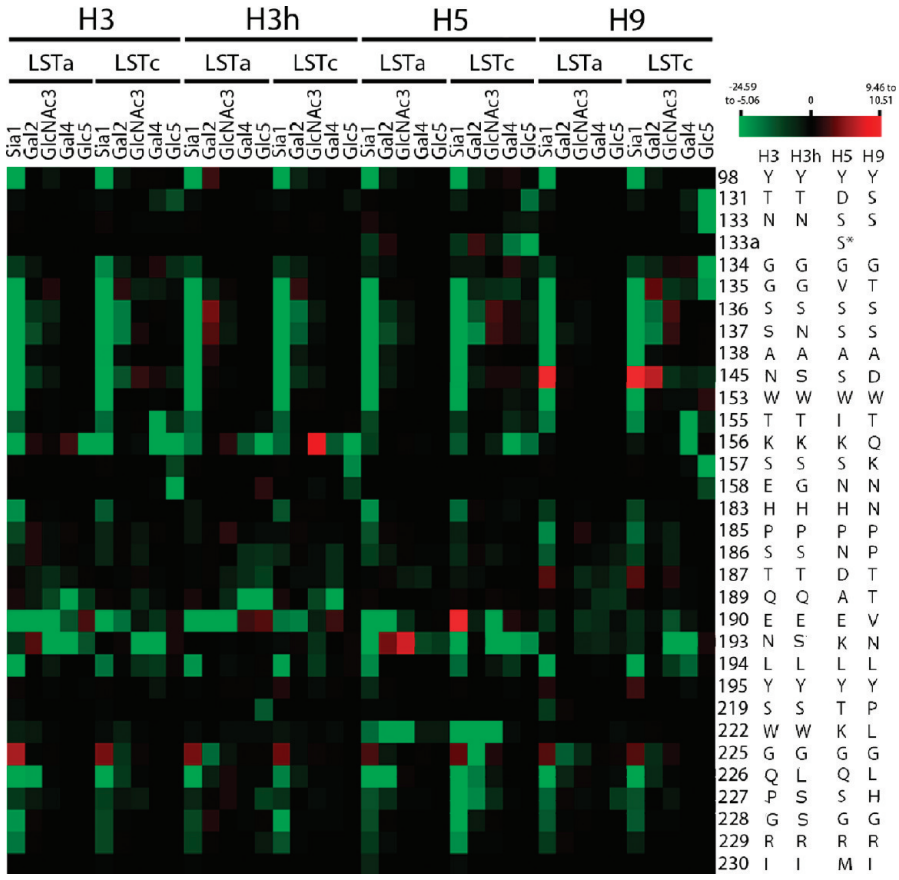

Figure 6. Heat map visualization of interaction energy profile for individual sugar residues and individual RBD residues. Green indicates negative interaction energy values (favorable), and red indicates positive interaction energy values (unfavorable). RBD residues are indicated as residue number (H3 numbering. NAG3 designates GlcNAc3 for space reasons. Residue 133a is an insertion unique to H5.

the Supporting Information), with or without bound glycan. Relative to the other HAs, swine H9 also had the fewest residues with changes in BAR order parameter, and they were also smaller than in the other HA (Figure 3, Figure S7 of the Supporting Information). The small standard errors for the RMSF and BAR analyses also indicated that the motions of the H9 RBD residues were minimal. The swine H9 RBD contained fewer charged residues in the four SSE of the RBD. In addition to the conserved 130-loop interactions, only residues 155, 156, 157, and 193 also had significant interaction energy with at least one of the glycan residues (Figure 6, Table S8 of the Supporting Information). D145 stood out with predicted interaction ∼+10 kcal/mol (Figure 6, H9 D135 in Figure 5), due to its negative charge. In addition, the ensemble-averaged electrostatic potential computation also showed much weaker positive electrostatic potential at the RBD (Figure 8).

The electrostatic contribution to the H9-LSTx interaction is the weakest and the solvation free energy penalty for H9 is the lowest in the previous published MM-GBSA analysis on this set of systems (Table 5 in[26]). Swine H9-LSTc is the only HA examined that had significant interaction with Glc5 in the 130 loop and K157 (Figure 6). Of all the H-bonds identified in this simulation, only swine H9-LSTc formed H-bonds with Gal4 at residue 156, and Glc5 at residues 135, 155. The propensity index indicated that most of the interactions calculated to be stronger in H9-LSTc were with the distal glycan residues Gal4 and Glc5 (Figure 7). Therefore, the swine H9 is unique in its adaptation for recognition of glycan receptors through nonbonded and hydrogen-bond interactions with distal glycans. The difference between the binding free energies of H9-LSTx was smaller than for H3-LSTx or H5-LSTx. (Table S1 of the Supporting Information).

These results may help explain in part the dual specific recognition of receptors with α-2,6 (human) or α-2,3 (avian) linkages by swine flu viruses. It may also give a clue to the current pandemic H1N1 flu of swine origin in that the swine H9 may be quite well adapted for binding both glycan receptors. Further simulations are ongoing to test this hypothesis, and our current results provide preliminary theoretical support for the notion that swine flu subtypes should be closely monitored.

Human Adaptation and Hong Kong Pandemic Flu H3N2. The PCA and all-atom RBD cluster analysis indicated that the conformational spaces sampled by H3h and H3 have significant overlaps, with distinct differences (Figure 1, Table S3 of the Supporting Information). While H3-LSTc is still dominated by one cluster, H3h-LSTc is distributed into three clusters. H3-LSTc had reduced RMSF near position 138, but with a huge standard error. H3-apo and H3h-apo have RMSF difference of ∼0.5 Å at residue 137, possibly due to the S137N mutation (Figure 6, Figure S1 of the Supporting Information). In contrast, the 150 loop had no differences in both H3 (Figure 3). Looking at the entire RBD region, H3-LSTa had more RMSF fluctuation (∼0.25 Å) in the 190 helix, the trimeric interface T3 region around residue 213 (Figures S1 and S6 of the Supporting Information), and the 220 loop. H3h-LSTc was similar, with minor RMSF fluctuations in the trimeric interface T1 region, 190 helix, T2, T3, and the region between the 220 loop and T4 (Figures S1, S6 of the Supporting Information). Previously

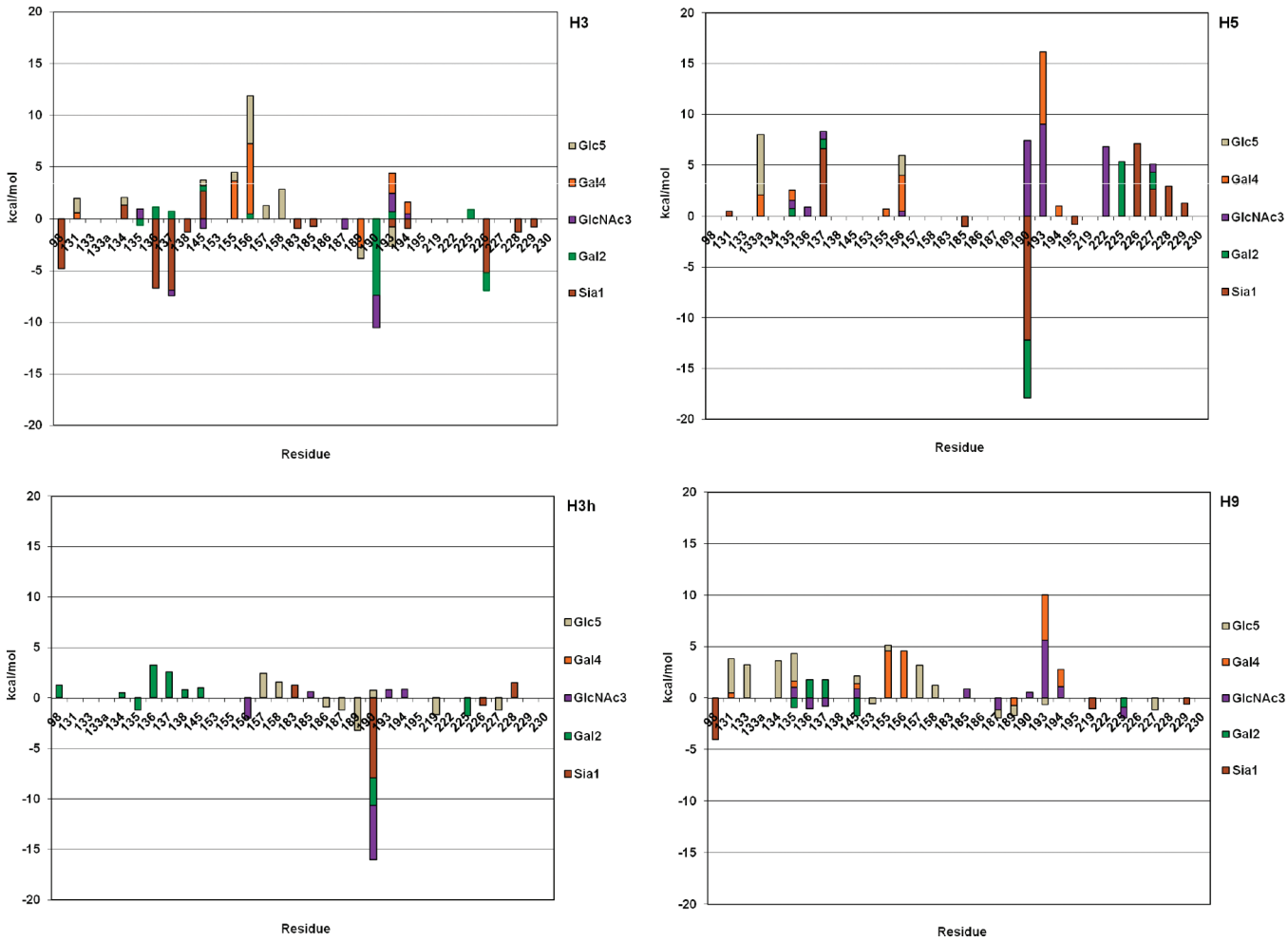

Figure 7. Propensity index for RBD residues to favor LSTa or LSTc in bound state. Negative values favor bound LSTa, and vice versa for bound LSTc.

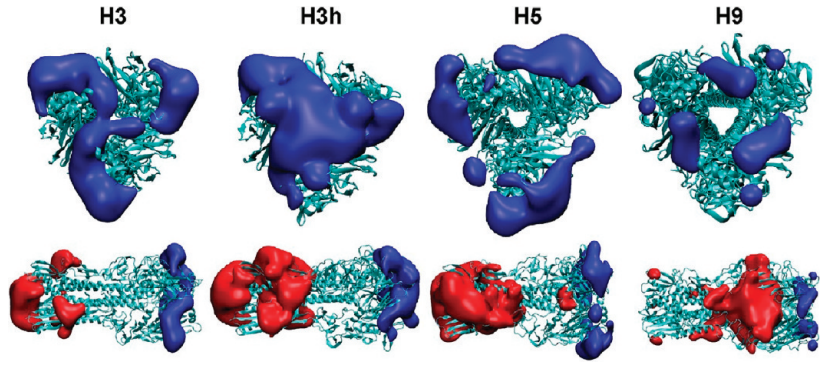

Figure 8. The 40 ns ensemble-averaged electrostatic potential fields. For all systems, the positive (blue lobes) and negative (red lobes) potential field strengths were ± 200 kT/e, with the exception of H9, which had a positive potential field drawn at +130 kT/e and a negative potential at −100 kT/e. The top panel shows the electrostatic potential shown looking down on the RBD; bottom panel shows the side view of the HA proteins.

identified trimer interface mutations apparently can affect species specificity,[6] suggesting that these calculations associating bound glycans with motions in the trimer interface may be worth further investigation.

In the BAR analysis, the 130 loop, 150 loop, and T3 showed increased flexibility (150 loop for LSTx, and T3 for LSTa) or decreased (130 loop for LSTa).In the 127 to 133 region, the BAR showed increased rigidity in H3-LSTa, and slightly more rigidity for H3h- LSTc. Residues exhibiting differences in motion with bound glycan could be markers for human adaptation of H3.

The 220 loop contains residue 226 and 228, known to be important for the receptor specificity switch.[5] All three H3h systems had the same RMSF in the 220 loop (Figure S6 of the Supporting Information). H3h-LSTc was the only system in which the contact frequency of both 157 and 158, which are near the trimeric interface T1 (Figure 4 and Table S4 of the Supporting Information), was greater than 50%.

What could be driving these changes in behavior? In avian H3, negatively charged E158 is positioned near the trimeric interface T1 region (Figure 4, Figure 5). This is mutated to G158 in human H3. 137 is part of the SIA binding site,[8] and N137 is associated with a slightly repulsive Gal2 interaction in H3h-LSTa, as well as a stronger interaction with Sia1. (Figure 6, 7, Tables S5−S6 of the Supporting Information). Q226 had an intermittent H-bond (Table 1) between Sia1 O9 and 226 OE, but the interaction energies between Sia1, Gal2 of LSTa with Q226 was higher than those from LSTc, as indicated by the propensity index (Figure 7). Several Sia1 interactions with H3 were stronger in H3-LSTa, quite different from H3h, where only the Sia1−190 interaction is stronger in H3h-LSTa (Figure 7). In addition to more equal Sia1 interactions with the 130-loop, H3h also had increased interaction with LSTa toward the end of the 150 loop, near the trimeric interface T1 and the 190 helix. L226 in H3h removed an interaction energy difference that favored LSTa; the interaction energies for both H3h-LSTx were small. Overall, our findings are consistent with the notion that loss of binding to α-2,3 linked receptors, and an increase to α-2,6 linked receptors is a mechanism of species specificity switch in H3 subtype viruses.

However, the propensity index predicted H3h to have more favorable interactions with LSTa. Without entropic considerations, the MM-GBSA analysis of the total free energy of binding also indicated that LSTa was favored over LSTc (Table S2 of the Supporting Information) for H3h. After the entropic contribution from glycan receptors was taken into consideration, LSTc is slightly favored over LSTa by H3h. Entropic contributions to the overall binding energy have already been shown by several studies to play an important role glycan-protein interactions,[55] including the HA[26] and related NA[67] systems. The conformational dynamics indicated that protein conformational entropy may contribute to the final free energy binding difference, which is still an area of active research.

The ensemble-averaged electrostatic potential isosurface map pointed out another potential mechanism that could influence receptor recognition, and be related to mutations that occur in the RBD. All of the HA have positive electrostatic potential over the RBD. H3h is unique in that its trimeric interface is also electrostatically positive, a visible contrast to H3 and H5. This may be due to the neutral G158 in H3h, in contrast to the negatively charged E158 in H3. The enhanced electropositive isosurface might facilitate the interaction with negatively charged sialoglycans, or other subsitutuents on glycan molecules, such as sulfation, additional sialylation or naturally occurring biantennary glycans. Although the positive potential would not distinguish between glycan receptor linkage types, it certainly could increase the efficiency of infection, or transmissibility when suitable receptors are present.

The Pandemic H5N1 Threat. H5 showed the most differences in conformational space sampling (Figure 2). It is unique among the HA in this study in having extended 120 to 130 loop regions (Figure 3). The BAR effect in the trimeric interface T1, particularly, residues 163 to 168, is pronounced compared to swine H9. H5 also shared similar BAR effect in the 150 loop with H3h (Figure 3, right panel). The region around residue 137, implicated as important for α-2,6 linked glycan specificity shift in H3/H3h, had similar RMSF differences to H3. H3h does not have S137; it has N137, which, like S137 had backbone amide H-bonds, but no side chain H-bonds as in the other three HAs (Table 1). Although H5 has N158 rather than E158, it did not have a similar electrostatic potential isosurface to H3h. This is likely due to the number of negatively charged residues in the H5 RBD: D131, D187, and E190, compared to only E190 in H3h, or only D145 in H9 (Figures 5 and 6).

The propensity index calculations suggested that the RBD environment is favorable for LSTc binding (Figure 7), with the unique 133a insertion and 156 interacting more with the distal Gal4 and Glc5 of LSTc. Residue 156 did exhibit reduced BAR motion in the presence of bound glycans, with the most reduction in H5-LSTa (Figure 3, right panel). The contact frequency analysis showed that the bound LSTc is almost in full contact with the RBD residues, with consistent reduction of RMSF in the major SSE of the RBD (Figure S6 of the Supporting Information). This suggests a likely strong entropic penalty for H5-LSTc. E190 had H-bonds with Sia1 O7 and O8 in H5-LSTa, but not LSTc, but one significant H-bond with GlcNac O6 in H5-LSTc (Table 1). The overall interaction energy for Sia1 of LSTc at residue E190 was 4.5 kcal/mol, highly unfavorable. As in the case of D135 in swine H9, this unfavorable interaction was compensated by the overall interaction energy at other positions, indicated by the propensity index (Figure 7).

Electrostatic Steering. Extensive MD simulations have been shown to accurately reproduce the experimentally observed changes in protein electrostatic field[52] and provide additional insights not ascertained with electrostatics surface maps alone. Our analyses revealed a large, positive electrostatic lobe around the RBDs for all the HA systems, which clearly indicates an electrostatic driving force for attracting the negatively charged terminal SIA moieties on host receptor cells (Figure 8). H5 and H3 exhibit the most similar potentials both in terms of shape and magnitude, whereas H3h is overall more positive and H9 has the weakest positive electrostatic potential. Whether the different field strengths relate to rate of association of the SIA receptors is an intriguing question that is not accessible with current MD simulations. Many examples of protein−ligand complexes whose association rates are enhanced many-fold due to electrostatic steering have been identified.[53] Furthermore, the ends of each HA near the membrane all exhibit a negative electrostatic potential. It is possible that this configuration may help guide glycan receptors toward the RBD. We note that the actual membrane-bound components of the HA structures are cleaved to facilitate crystallization and therefore not present in our analyses. Again, H9 exhibits an overall weaker negative electrostatic potential that is different in character, including a large lobe much closer to the center of the molecule. It is tempting to speculate that mutations could be made to introduce a more electropositive isosurface for H5 or other HAs, and tested in silico or in vitro to see whether this affects the binding kinetics.

HA Interaction Energy Profiles and Propensity Index. The heat map allowed the identification of distinct interaction patterns that would be otherwise unnoticed. For example, removal of noted unfavorable contacts between the glycan and HA could potentially enhance the binding affinity of glycan receptors. D145 stood out in swine H9 as an unfavorable interaction due to the negative charge. This could be a species barrier, and a mutation at this position might increase the chance of swine H9 crossing the species barrier. Residue 156 had an unfavorable interaction energy calculated at $+2.5 \pm 0.6$ kcal/mol in H3h-LSTc, and has not been studied systematically

(67) Amaro, R. E.; Cheng, X.; Ivanov, I.; Xu, D.; McCammon, J. A. *J. Am. Chem. Soc.* **2009**, *131*, 4702–4709.

before. A possible role for 156 in α-2,6 specificity was proposed from the crystal structure.[8] Residues 156 and 157 also form antigenic site B.[12,68] Other notable residues with unfavorable interactions include residues 190, 193, and could be markers to monitor for receptor binding specificity or species specificity switch.

The propensity index offers an easy way to examine whether a particular RBD contains residues favorable for one receptor or another, based upon enthalpic considerations. While Sia1 interactions contributed from 60 to 75% of the free energy of binding, the distal or inner glycan residues may contribute through interactions with other RBD residues. The flexible nature of the glycan receptors suggest that many of these interactions may be transient, and through weak forms of interactions such as dipole−dipole interactions. The abundance of Glc5 interactions in the propensity index analysis suggests that it may contribute significantly to the multivalent, avidity-based glycan receptor and HA protein interactions.

Species Specificity Switch and Cooperative Interaction Network. The propensity index strongly suggests that receptor binding specificity and species specificity shift studies should be conducted through systematic profiling. The same residues may participate in different interactions and perform different roles. This may apply particularly to low affinity and high avidity interactions between the HAs and their glycan receptors. A S137A mutation in a human isolate of H5 was found to greatly increase α-2,6 glycan binding, whereas S137A and T192I mutations decreased α-2,3 glycan binding relative to α-2,6 glycan.[10] While we did not detect strong interactions at T192, we could not rule out that compensatory changes at other locations might have occurred. Direct simulation experiments are required to assess the effects of these mutations within the H5 subtype, with reference to the experimental studies. Between avian and swine subtypes, three mutations at charged residues occurred involving S145D, K156Q, and E190 V. Nine mutations occurred in the RBD between H3 and H3h (Figure S2 of the Supporting Information), and six are shown in this study and earlier experiments to be involved in glycan interactions: N137S, G158E, S193N, L226Q, S227P, and S228G. More coordinated mutagenesis studies, performed both in silico and in vitro, would be required to tease apart the intricate network of interactions that determine species specificity switch.

Conclusions

We have presented several complementary modes of analyses for about 30 residues of the HA RBD which form an intricate network of receptor binding surface. Some of these interactions were calculated for all four HA, many are unique to a given system, and a few may be diagnostic for binding preference and species specificity. Our MD simulations revealed that there are modest differences in the dynamics between H3, H5, and H9. Most notably, the glycan binding domain for H3, H3h, and H5 is substantially more flexible than that in H9. However, the dynamic region of the binding domain is not coincident with the specific glycan interaction sites, which exhibit a high level of rigidity, as had been determined from structural comparison of apo-HA to HA-LSTx crystals.[8,27,28] Comparative dynamics analysis of the apo and glycan bound systems revealed that while the dynamic motions of the flexible region of the binding domain in H3, H3h, and H5 are dampened in the presence of both glycan types, no substantial backbone structural or dynamic changes are observed in the RBD in all four HA systems. Our extensive energetic decomposition analysis revealed some differences in the constituent specific energetic interactions between avian-type and human-type glycans with the various HA-mediated contacts. The dominant attractive force between glycan and HA is probably electrostatic, not hydrogen-bonding. The results of this theoretical study have allowed a unique dissection of the delicate balance of complex and diverse interactions that determine the system specific glycan binding affinities, and provide more detailed insight into the different energetic contributions that control the process of initial attachment to host cells in different viral strains of HA.

Acknowledgment. E.I.N., D.X., and W.W.L. are supported by TATRC W81XWH-07-2-0014, and in part, by NIH P41 RR08605. H.C.P., K.J.W. are also supported by NSF INT 0407508 to the PRIME program. R.E.A. is funded in part by NIH GM077729 and MRAC CHE060073N. P.R.L.M. is funded by HHMI (to J.A.M.). The work is also funded by NIH GM31749, NSF MCB-0506593 and MCA93S013 (to J.A.M.). Additional support from the Howard Hughes Medical Institute, San Diego Supercomputing Center, Maui High Performance Computing Center, Accelrys Inc., the W.M. Keck Foundation, the National Biomedical Computation Resource, and the Center for Theoretical Biological Physics is gratefully acknowledged. We thank Dr. S. J. Gamblin from MRC, UK for providing the H3h LSTa and LSTc crystal structure coordinates, and the referees for their helpful comments.

Supporting Information Available: Structure based sequence alignments; rmsd for all systems studied; PCA plots for PC1−6 for H3h; principal component eigenmodes; topological variatins in dominant representative cluster RBDs; RMSF plots for all HA; H3h Backbone N−H order parameter plots for all HA; Cα thermal factor for all HA; experimental crystal information and previously computed MM-GBSA free energies; MM-GBSA binding free energy of H3h-LSTx; representative conformational clusters of HA; contact frequency of all systems; H3, Hh3, H5, H9 interaction energy decompositions; full citations for refs 11, 45, 47, 59, and 65 This material is available free of charge via the Internet at http://pubs.acs.org.

JA904052Q

(68) Yassine, H. M.; Lee, C. W.; Suarez, D. L.; Saif, Y. M. *Vaccine* **2008**, *26*, 966–977.