doi:10.1016/j.jmb.2009.01.040 *J. Mol. Biol.* (2009) **387**, 465–491

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Distinct Glycan Topology for Avian and Human Sialopentasaccharide Receptor Analogues upon Binding Different Hemagglutinins: A Molecular Dynamics Perspective

Dong Xu[1]*, E. Irene Newhouse[2]*, Rommie E. Amaro[3,4], Hsing C. Pao[1], Lily S. Cheng[1], Phineus R. L. Markwick[5], J. Andrew McCammon[1,3,4,5,6], Wilfred W. Li[1] and Peter W. Arzberger[1]

[1]*National Biomedical Computation Resource, University of California San Diego, La Jolla, CA 92093-0505, USA*

[2]*Maui High Performance Computing Center, Kihei, Maui, HI 96753, USA*

[3]*Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, CA 92093-0365, USA*

[4]*National Science Foundation Center for Theoretical Biological Physics, University of California San Diego, La Jolla, CA 92093-0365, USA*

[5]*Howard Hughes Medical Institute, University of California San Diego, La Jolla, CA 92093-0365, USA*

[6]*Department of Pharmacology, University of California San Diego, La Jolla, CA 92093-0365, USA*

Hemagglutinin (HA) binds to sialylated glycans exposed on the host cell surface in the initial stage of avian influenza virus infection. It has been previously hypothesized that glycan topology plays a critical role in the human adaptation of avian flu viruses, such as the potentially pandemic H5N1. Comparative molecular dynamics studies are complementary to experimental techniques, including glycan microarray, to understand the mechanism of species-specificity switch better. The examined systems comprise explicitly solvated trimeric forms of avian H3, H5, and swine H9 in complex with avian and human glycan receptor analogues—LSTa (α-2,3-linked lactoseries tetrasaccharide a) and LSTc (α-2,6-linked lactoseries tetrasaccharide c), respectively. The glycans adopted distinct topological profiles with inducible torsional angles when bound to different HAs. The corresponding receptor binding domain amino acid contact profiles were also distinct. Avian H5 was able to accommodate LSTc in a tightly "folded umbrella"-like topology through interactions with all five sugar residues. After considering conformational entropy, the relative binding free-energy changes, calculated using the molecular mechanics-generalized Born surface area technique, were in agreement with previous experimental findings and provided insights on electrostatic, van der Waals, desolvation, and entropic contributions to HA–glycan interactions. The topology profile and the relative abundance of free glycan receptors may influence receptor binding kinetics. Glycan composition and topological changes upon binding different HAs may be important determinants in species-specificity switch.

*Corresponding authors. D. Xu is to be contacted at National Biomedical Computation Resource, University of California, San Diego, 9500 Gilman Drive, Mail Code 0505, La Jolla, CA 92093-0505, USA. E. I. Newhouse, 231 MRTC-A, 590 Lipoa Parkway, Kihei, HI 96743, USA. E-mail addresses: dxu@mccammon.ucsd.edu; einew@hotmail.com.

Abbreviations used: HA, hemagglutinin; MD, molecular dynamics; MM-GBSA, molecular mechanics-generalized Born surface area; NA, neuraminidase; RBD, receptor binding domain; SSE, secondary structure element; vdW, van der Waals.

Received 10 September 2008;
received in revised form
21 December 2008;
accepted 17 January 2009
Available online
5 February 2009

Edited by D. Case

Keywords: avian influenza; hemagglutinin; glycan topology; binding specificity; molecular dynamics

Introduction

Increased circulation of highly pathogenic type A avian influenza in poultry and wild birds in Asia has had severe economic consequences for the poultry and associated industries in the affected countries.[1–3] Subtypes of the influenza viruses are named based on the observed combinations of two viral surface membrane glycoproteins, hemagglutinin (HA) and neuraminidase (NA), with 16 and 9 types, respectively.[4] HA, a homotrimer, is involved in the attachment of viral particles to sialosaccharides on host cell membrane lipids or surface proteins. NA, a homotetramer, plays a role in the release of newly synthesized viral particles by cleaving terminal sialic acid (Sia1) residues to which HA would otherwise bind.[5–7] The current human epidemic strains, such as H3N2 and H1N1, have relatively low pathogenicity, causing illnesses in 30–50 million people around the world each year with an approximate 1% mortality rate. Three historical influenza pandemics have taken place, with H1N1 in 1918, H2N2 in 1957, and H3N2 in 1968. The 1918 H1N1 pandemic affected up to 1 billion people, killing an estimated 20–50 million people around the world, with a mortality rate of 2%–5%.[2] It is thought to have emerged from an avian progenitor without reassortment via an intermediate host.[8] Since the first fatal human case of H5N1 in 1997,[9,10] more than 371 cases have been reported, with a fatality rate over 65%.[11,12] It is thought that a new highly lethal H5N1 strain with the ability to easily transmit among humans could emerge through genome reassortment or adaptation.[13,14]

While other proteins, including the RNA polymerase PB2 subunit, nonstructural protein 1, HA activation proteases, and NA, are important factors affecting host selectivity, viral tropism, virulence, and transmissibility,[8,15,16] HA is directly involved in the initial stage of infection by binding to sialylated glycans exposed on the host cell surface.[17,18] Thus, extensive effort has been devoted to study HA's role in species specificity. The terminal sialic acid, also known as Neu5Ac (*N*-acetylneuraminic acid, or NANA), is often linked to a galactose subunit of the polysaccharide chain via an α-2,3 or α-2,6 glycosidic linkage (Fig. 1). A number of studies have established that avian viruses prefer glycans with the α-2,3 linkage, while human adapted viruses prefer those with α-2,6.[5–7] Human bronchial mucus contains a mucin, a heavily glycosylated protein with attached α-2,3-linked oligosaccharides, which may bind avian viruses and help prevent infection.[6] More recent studies have shown that the human lower respiratory tract also contains avian H5N1 receptors of the α-2,3 linkage type.[20–22] Thus, humans may be susceptible to direct infection by high concentrations of avian influenza viruses, even though few cases have been documented.[23] For rapid spread between humans, it is hypothesized that a change in the glycan preference is required, as the earliest isolates of viruses from the three 20th century pandemics all preferentially bound to glycans with the α-2,6 linkage.[15,24,25] The host cell receptor preference is reflected in the glycan distributions in birds and intermediate hosts. For example, duck trachea expresses predominantly the α-2,3 linkage type, whereas swine trachea expresses both.[26] The two most recent human influenza pandemics, the 1957 Asian flu and the 1968 Hong Kong flu, are believed to be caused by viruses that had reassorted in swine.[27] The glycan distribution in humans, however, has been found to be more complex than previously thought.[28–30] The natural diversity of glycans may hinder detection by existing assays,[31,32] whereas the human upper respiratory tract contains many complex types of glycans.[19]

The receptor binding domain (RBD) of HA comprises several key structural components (Fig. S1), including the 190 helix (HA1 188–190), the 130 loop (HA1 134–138), and the 220 loop (HA1 221–228), with a number of conserved residues for receptor binding and species specificity.[18] The interaction with glycans is often weak, in the millimolar range, nonetheless compensated by multivalent interactions (avidity) between multiple HAs and cell surface receptors.[33] Extensive crystal structural studies have revealed that the orientation of Sia is constrained by several conserved residues [e.g., Y98, S/T136, W153, H183, and L/I194 (H3 numbering)] that exhibit more variations in HAs from humans than from birds.[19,34] Furthermore, the larger RBD size of human H3, compared with that from avian H1 or H5, may be required to accommodate the larger α-2,6 receptors.[35] There is growing recognition that the optimization of molecular interaction may require significant conformational adjustments of the participating proteins, ligands, or substrates and carbohydrates.[19,34] While superimpositions of pentasaccharides using known crystal structures offer potential clues as to why α-2,3 or α-2,6 linkage may be preferred,[36,37] both HA protein and glycan flexibility are largely undetectable in crystallography studies, which can only capture one snapshot of the

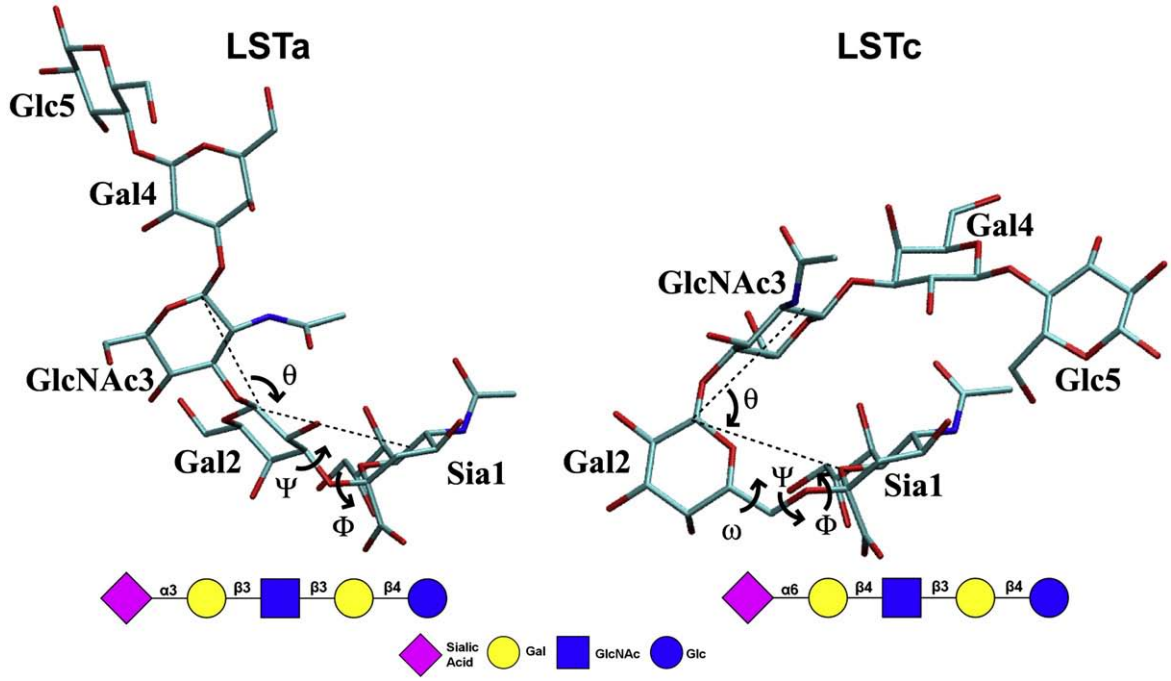

Fig. 1. Avian and human host cell receptor analogues, LSTa and LSTc, used in the simulations. The glycans are shown in both licorice representation and Consortium for Functional Glycomics representation with labels for the sugar units. Torsion angles (Φ, Ψ, and ω) in α-2,3 and α-2,6 linkages are also illustrated along with the θ angle used in defining glycan topology by Chandrasekaran *et al*.[19]

dynamic interactions between the proteins and ligands limited largely by crystallization conditions.

Besides biochemical and crystallographic methods, a number of other techniques have been applied to study the switch of species specificity. Glycan microarrays contain hundreds of synthetic glycans that mimic the natural glycan diversity and topology.[38] Artificially expressed HA trimers are used to sample receptor specificity and binding avidity[38] and help define receptor "fingerprints" for different strains of avian flu viruses.[27,38,39] However, glycan microarrays cannot yet fully represent the complex types of glycans. In chicken red blood cell-based hemagglutination assays, permutations of known mutations that enable avian to human receptor specificity switch in H5 strains[18,40] affected binding and cell entry to different degrees[41,42] as the resialylated chicken red blood cell may also lack complex branched glycans.[27,41] In a comparison of a ferret animal model for H1N1 transmissibility with glycan microarray data, it was found that mutants binding α-2,3 and α-2,6, in addition to long α-2,6, transmit efficiently, whereas those lacking long α-2,6 binding transmit inefficiently.[27,41]

The pentasaccharides LSTa (α-2,3-linked lactoseries tetrasaccharide a) and LSTc (α-2,6-linked lactoseries tetrasaccharide c), natural sialosides from human milk (Fig. 1), contain lactosamine (Gal2–GlcNAc3) and lactose (Gal4–Glc5) units and are often found in complex glycans on the cell surface. Based on a survey of available crystal structures, Chandrasekaran *et al*. recently proposed a θ angle parameter to define the topology adopted by the long α-2,3- and α-2,6-linked glycans[19] and that the glycan topology, rather than the glycosidic linkage difference, plays a crucial role in HA–glycan specificity of recognition. LSTa and LSTc have been extensively used as avian and human receptor analogues in influenza virus host selectivity studies, particularly in the crystallographically determined HA–glycan complexes.[34–36,43] These crystal structures provide ideal starting points for molecular dynamics (MD) simulations. Significant advances in carbohydrate force field development have been made over the years, primarily by GLYCAM06[44,45] for the AMBER force field,[46] CSFF,[47] and others[48–50] for the CHARMM force field.[51] Longer and more accurate MD simulations that sample greater glycan conformational space have become a feasible approach to probe protein–carbohydrate interactions. The inherent flexibility that challenged crystallography and limited earlier computational studies to short disaccharides or trisaccharides[52–54] can now be examined in detail. Previously reported HA affinity for sialyl oligosaccharides usually had dissociation constants in the millimolar range.[55–62] This behavior has been attributed to an enthalpy–entropy compensation phenomenon in which the entropy of binding is overall negative, offsetting the enthalpic gain.[63,64] The role of entropy in opposing binding is interpreted in terms of conformational distortion and freezing of flexible oligosaccharide ligands, as well as solvent reorganization accompanying binding. While the solvent-associated entropy contribution is still the least understood aspect, MD simulations of free and bound glycans make it possible to explore the conformational energetics of glycan–HA binding interactions.

Here, we report a comparative study based on explicitly solvated MD simulations for 80 ns of free LSTa and LSTc in solution and those for 40 ns in complex with avian influenza H3, H5, and H9 trimers (denoted as LSTx-H*n*, where "x" = "a" or "c" and $n = 3$, 5, or 9). The MD simulations complement the traditional view of static crystallography and provide essential dynamic information about glycan conformation and topology through the monitoring of time-dependent properties, such as RMSD, θ, and glycosidic torsional angles. Statistical clustering of the MD snapshots captured representative glycan conformations, offering insight to the key sugar and HA contact residues responsible for the interactions. The three-dimensional glycan topologies recognized by the HAs were characterized using volumetric analysis of the dominant glycan clusters. The conformational and topological dynamics were further quantified in terms of interaction energy, molecular mechanics-generalized Born surface area (MM-GBSA) binding free energy, and conformational entropy change. The results were in agreement with experimental data and offer new perspectives to viral host range selection by different HAs.

Results

System convergence of glycan in solution and HA–glycan complex simulations

To ensure that the conformational space was adequately sampled in the free glycan simulations, we calculated the average square projection of all glycosidic torsion angle eigenvectors onto those of the preceding time interval. When the average square projection is close to 1.0, the essential spaces spanned by the two sets of eigenvectors are the same.[65] Figure S9 shows that free LSTa conformational sampling converged after about 35 ns and that free LSTc converged after about 45 ns. Extended simulation up to 80 ns did not show significant signs of additional conformational space sampling using the GLYCAM06 force field. The convergence of the HA–glycan complex systems was examined using the MM-GBSA technique. Figure S10 shows the plateaus of major interaction energies, ΔE_{MM}, $\Delta G_{Solvation}$, and, in particular, $\Delta G_{GBTotal}$. The systems were well behaved and had reasonably converged within 40 ns of simulation.

RMSD analysis of glycans in solution and complex

The overall glycan conformational properties may be examined through RMSD analysis. The structural fluctuations of the glycans upon binding to different HAs were plotted and summarized (Fig. S2; Tables S1 and S2). Free LSTa had an average RMSD of 5.92 Å, and LSTc had an average RMSD of 11.61 Å. Upon binding to the different HAs, the RMSD of LSTc was reduced significantly. In particular, the RMSDs of LSTc-H9 were the lowest—1.09, 1.32, and 3.47 Å for the three monomers, respectively. The RMSDs of LSTc-H5 varied the least, with standard deviations as low as 0.67, 0.50, and 0.52. For LSTc-H3, the RMSD was higher than 4 Å but still less than half that of free LSTc. The bound LSTc exhibited more than 60% reduction in the average RMSD compared with the free state, whereas bound LSTa exhibited only minor changes in the average RMSD.

Glycan topology and θ angle variations

It has been proposed that the θ angle between the C2 atom of Sia1 and the C1 atoms of Gal2 and GlcNAc3 can be used as an indicator of glycan topology. The cone-like topology ($\theta > 110°$) is adopted by long α-2,3-linked glycans and preferred by avian HA; the umbrella-like topology ($\theta < 110°$) is adopted by long α-2,6-linked glycans and preferred by human adapted HA. This topological θ angle was monitored over the course of the MD simulations (Fig. 2). The free LSTa had an average θ angle of 152°, adopting a narrow cone-like topology. The average θ angle in the free LSTa was substantially reduced upon binding to H3 and H9. Surprisingly, one monomer of LSTa-H3 obtained an average θ angle of 105°, dropping below the 110° cutoff, and would be considered as having an umbrella-like topology. One monomer of LSTa-H9 obtained an average θ angle of 116°, close to the 110° cutoff. Binding to H5 had the least impact on LSTa's topology.

Frequent θ angle transitions were observed in LSTc, echoing its great structural variations demonstrated by the RMSD analysis. Free LSTc spent about 50% of the time in a tightly folded umbrella-like topology (θ of ~60°) and switched between open umbrella (θ of ~100°–110°) and cone-like (more than 140° at times) topologies during the other 50% of time. Thus, free LSTc often adopted a cone-like topology with the θ angles reaching much more than 110°. LSTc-H3 exhibited greater fluctuations than the other bound LSTc. In one monomer, it opened, folded, and then reopened. All three monomers of LSTc-H5 were extremely stable and stayed folded at a θ angle of ~60° over the entire 40 ns of simulation. LSTc-H9, on the other hand, stayed folded at 60° to 80° in two monomers, while the third monomer opened up to over 100° and gradually refolded back toward the end of the simulation.

Glycan RMSD-based clustering and volumetric analysis

To reduce the vast number of glycan conformations sampled by the MD simulations, we carried out RMSD-based clustering analyses using the average linkage method[66] to extract the glycan representative structures. The agglomerative clustering results are shown in dendrograms with three-cluster solutions (Fig. S3). The cluster representative structures of LSTa are drawn with the Sia1 *N*-acetyl group pointing toward the reader, and those of LSTc

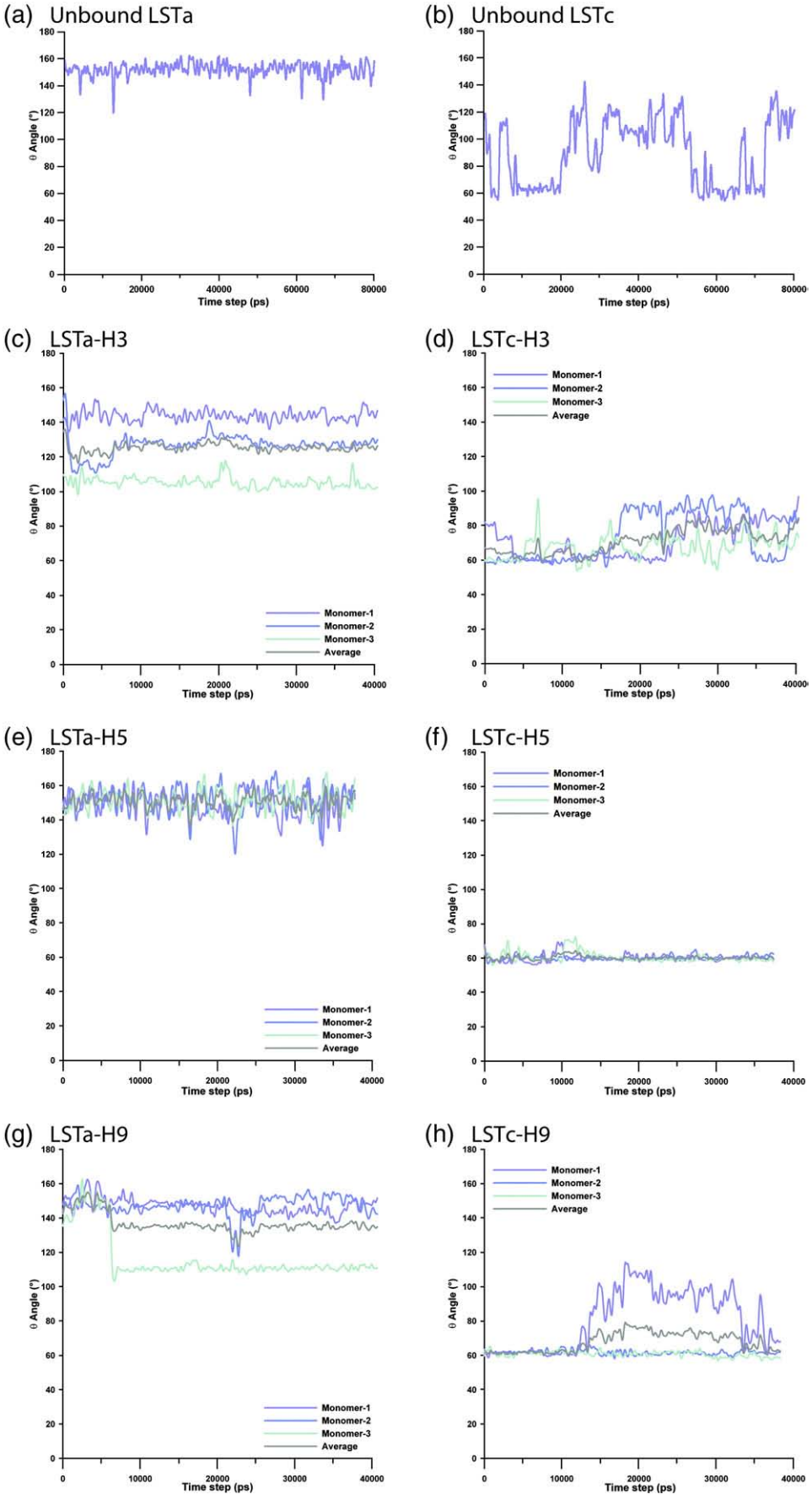

Fig. 2. Topological θ angle plots of the free and H3-, H5-, and H9-bound glycans.

are rotated 90° to provide a better viewing angle (Fig. 3). The relative cluster sizes are shown as bar graph insets.

The free LSTa topology was mostly cone-like with 81% of the population represented by a representative with a θ angle of 135° (Table 1; Table S1 has the population average data). When bound to H3, the populations spread out into three clusters of similar sizes. A representative of 29% of the LSTa population dropped its θ angle to 106° and took on the open umbrella-like topology. No reduction was observed for LSTa-H5 θ angles, with 86% of the

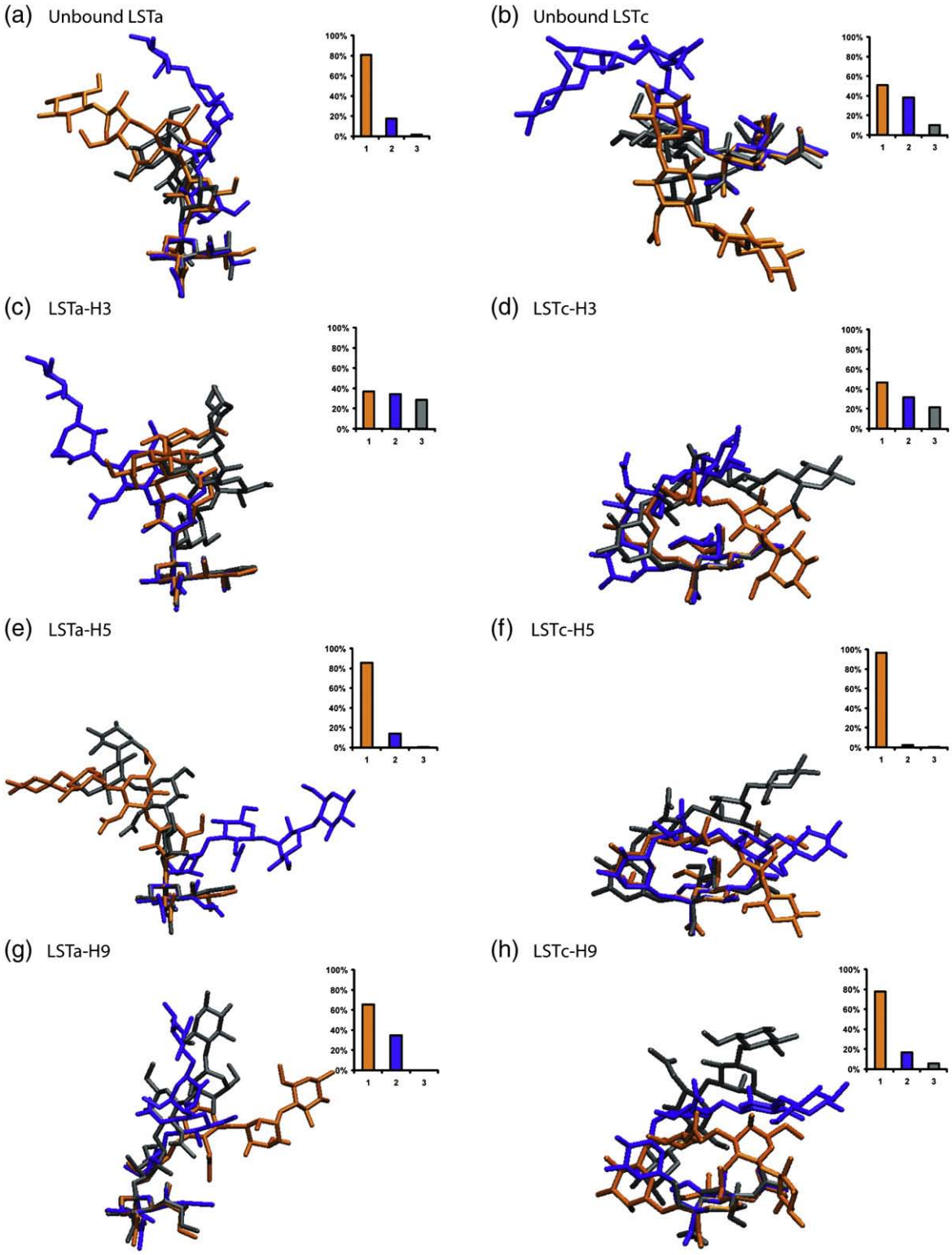

Fig. 3. Representative conformations adopted by the free and H3-, H5-, and H9-bound glycans. The conformations are aligned on the heavy atoms (C and O) of the Sia1 pyranose ring.

Table 1. Summary of relative abundance and Φ, Ψ, and θ angles in the LSTa cluster representatives

State	Cluster	Population (%)	θ	Sia1–Gal2		Gal2–GlcNAc3		GlcNAc3–Gal4		Gal4–Glc5	
				Φ	Ψ	Φ	Ψ	Φ	Ψ	Φ	Ψ
Unbound	1	80.7	135	179	−23	17	−18	15	49	53	8
	2	17.6	157	−96	−57	44	18	52	4	27	−24
	3	1.7	125	190	−57	−21	−30	38	59	47	−30
H3	1	36.9	128	189	−4	−26	−25	39	8	14	−23
	2	34.3	138	176	−6	−51	197	38	−37	73	4
	3	28.8	106	−87	−45	−27	−24	29	11	65	18
H5	1	85.7	143	189	−28	31	−20	54	30	45	13
	2	14.0	144	−64	−9	36	25	38	0	42	0
	3	0.3	158	151	−65	42	26	18	28	52	−5
H9	1	65.4	148	−74	23	52	0	31	48	20	−5
	2	34.6	111	−66	−2	−40	−21	43	21	44	−4
	3	0.08	134	−65	18	18	45	47	21	27	18

The Φ, Ψ, and θ values are expressed in degrees.

population represented by a glycan with a θ angle of 143°. LSTa-H9 had a representative of one-third of the population with a θ angle of 111°, closely resembling the umbrella-like topology.

The free LSTc adopted the tightly folded (a representative of 51% of the population with θ = 58°), the open umbrella-like (one of 10% at θ = 97°), and the cone-like (one of 39% at θ = 133°) topologies (Table 2). The representative with the cone-like topology was found to have a ω angle of −53°. More than half of LSTc-H3 covered by cluster 2 and cluster 3 representatives had large θ angles (70°–80°). Notably, these open umbrella LSTc topologies resemble the ones found in the swine H1 [Protein Data Bank (PDB) code 1RVT][35] and human H3 X-31[36] crystal structures (Table S4). LSTc-H5 covered by the cluster 1 representative adopted the tightly folded umbrella topology (97% at θ = 63°) compared with the dominant cluster representatives of LSTc-H3 (47% at θ = 60°) and LSTc-H9 (78% at θ = 60°). About 17% of LSTc-H9 (cluster 2 representative) opened up to the open umbrella topology (θ = 90°). A small population (6%) of LSTc-H9 (cluster 3 representative) had θ = 109°, almost cone-like.

When the LSTa and LSTc cluster representatives were represented in the same viewing angle [Fig. S4; see Fig. S1 for the secondary structure elements (SSEs)], the differences in the preferred topologies by different HAs were striking. LSTa-H3 exhibited favored orientation toward the 130 loop, and LSTa-H5 exhibited a similar profile to the free LSTa, except that the dominant representative also favored the 130 loop. LSTa-H9, however, predominantly favored orientation toward the 190 helix. All the LSTc bound to HA favored orientation toward the 130 loop, with the LSTc-H5 representatives having similar θ angles. Their distal lactose units, however, sampled distinct spaces.

The overall glycan topology may be examined using the volumes sampled by the glycans in motion, unlike the θ angle mostly determined by the first three sugar units. Figure 4 shows the volumetric maps of the dominant clusters, whereas Fig. S5 shows the volumetric maps of the entire population. Cone-like topologies were observed in the dominant cluster of the free and bound LSTa's and LSTc-H5, which has a stem-like feature. The orientations and shapes of the volumetric maps varied dramatically for all the free and bound glycans. LSTa-H3 bent toward the 190 helix, while LSTa-H9 leaned against the groove formed between the 220 loop and the 190 helix. The free LSTc, with the dominant cluster representative's θ angle at 58°, gave no hint of an umbrella-like topology. Bound LSTc adopted

Table 2. Summary of relative abundance and Φ, Ψ, ω, and θ angles in the LSTc cluster representatives

State	Cluster	Population (%)	θ	Sia1–Gal2			Gal2–GlcNAc3		GlcNAc3–Gal4		Gal4–Glc5	
				Φ	Ψ	ω	Φ	Ψ	Φ	Ψ	Φ	Ψ
Unbound	1	51.2	58	192	205	−53	69	187	45	1	39	11
	2	38.5	133	192	169	58	16	−22	52	−30	62	6
	3	10.3	97	164	180	179	48	−28	34	−49	−3	18
H3	1	46.7	61	−60	193	−53	75	194	26	21	38	15
	2	31.7	79	−66	166	−28	44	185	42	−11	24	26
	3	21.5	68	−55	200	40	64	186	37	−27	41	11
H5	1	96.8	58	−67	195	−65	60	193	30	16	35	25
	2	2.4	78	−77	192	−38	17	175	21	−40	59	7
	3	0.8	73	−73	188	−38	49	193	11	−38	64	9
H9	1	77.7	60	−53	205	−57	62	192	39	29	62	47
	2	16.7	90	214	200	36	43	195	23	−31	37	−5
	3	5.6	109	194	97	135	43	177	68	−39	36	−3

The Φ, Ψ, ω, and θ values are expressed in degrees.

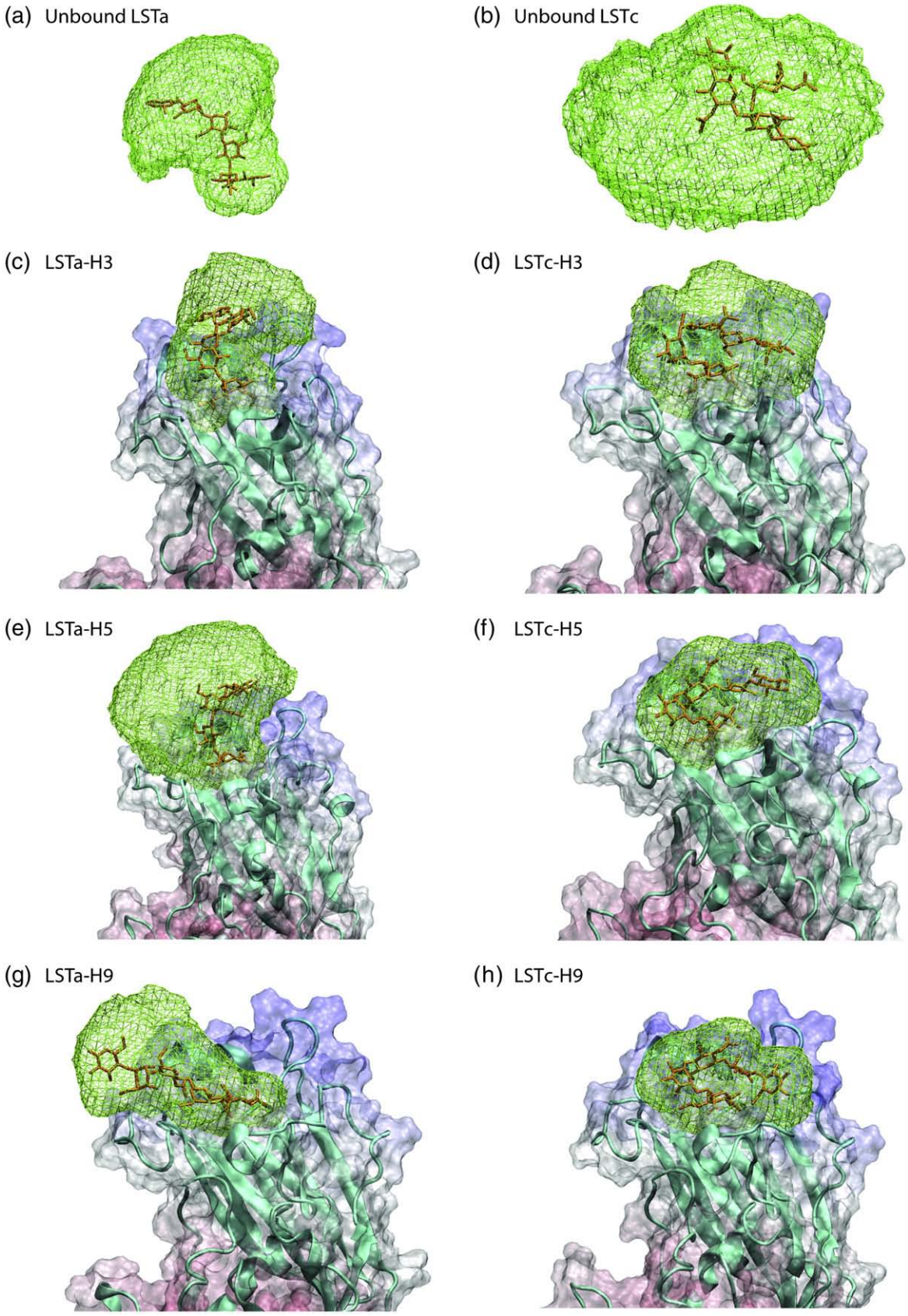

Fig. 4. Dominant cluster volumetric topologies of the free and H3-, H5-, and H9-bound glycans.

predominantly tightly folded umbrella conformations in the most populated clusters, showing little movement around the "umbrella stem" Sia1–Gal2. Their topologies were better described as "fishhook-like" because their lateral movement was very restricted. These glycans were also situated differently in the HA RBDs. LSTc-H5 topology made close contacts with the 130 loop and the 220 loop. LSTc-H9 shifted closer to the 130 loop and the 150 loop of H9 RBD, while LSTc-H3 was elevated, making contacts with the 190 helix.

Glycosidic torsion angles

The earlier simulation analyses indicated that the topologies adopted by different glycans upon HA binding were distinct and not fully described by the θ angles, whose definition is limited to the first three sugar units commonly resolved in known crystal structures. The internal motions of the glycans, however, are well characterized by the glycosidic torsion angles, and it was straightforward to extract and examine the torsion angles of all glycosidic linkages in the representative glycan structures and the total glycan population averages.

Sia1–Gal2 linkage

The Sia1–Gal2 linkage Φ, Ψ plots of the free glycans were in reasonable agreement with the conformational maps obtained from GlycoMapsDB[67] (Fig. 5). The absence of *gauche* conformation (Φ of ~60°) was observed in free and bound LSTx simulations and in crystal structures, indicated by the encircled regions 1, 2, and 3[19] (Fig. 5a–d). The free LSTa was predominantly in the *anti*/*trans* conformation (Φ of ~180°, equivalent to the term *trans* conformation used in the original crystal structure report[34]), whereas the free LSTc was predominantly in the *gauche*/*cis* conformation (Φ of approximately −60°, equivalent to the term *cis* conformation used in the original crystal structure report).[34] The Ψ angles were distributed in LSTa from −80° to 40° and in LSTc from 80° and 240°.

During the MD simulations, the bound glycans displayed Φ and Ψ angles not seen in the starting crystal structures but present in the free glycans. On the other hand, torsion angles not found in the free glycan simulations were induced in some of the bound glycans. In at least one of the three HA monomers, a significant conformational transition from *anti* (starting conformation from crystal structure) to *gauche* was observed in LSTa-H3 (Fig. 5e) and LSTa-H5 (Fig. 5g). In contrast, a significant conformational transition from *gauche* (starting conformation from crystal structure) to *anti* was observed in LSTc-H3 and LSTc-H9. Interestingly, LSTa-H9 and LSTc-H5 were only found in the *gauche* conformation.

The Φ, Ψ angles in the representative clusters corroborated the observations in the population averages. The free LSTa adopted both the *anti* (cluster 1, 81%) and *gauche* (cluster 2, 18%) conformations. In the LSTa-H3 clusters, the *anti* conformation is the majority, and 22% of the LSTa-H3 shifted into the *gauche* conformation. LSTa-H5 adopted both the *anti* (cluster 1, 86%) and *gauche* (cluster 2, 14%) conformations. The *gauche* conformation was not found among the free LSTc cluster representatives, even though a small population existed in the *gauche* conformation (Fig. 5d). Only LSTc-H9 clusters 2 and 3 (23% total) retained the *anti* conformations (see also Table 2).

The ω angle

The ω angle is unique to the α-2,6 linkage. The ω angle histograms showed modal values at −45° (most populated), 52°, and 174° in free LSTc; −48° and 52° in LSTc-H3; −52° in LSTc-H5; and −50°, 54°, and 157° in LSTc-H9 (Fig. 6). The two minor peaks that were sampled in free LSTc were much reduced or absent in bound LSTc. The cluster representatives indicated that the majority of ω angles adopted a *gauche* conformation with a negative dihedral value, which was often associated with the tightly folded topology (θ = 55°–70°) (Table 2). Positive ω angles were found in free LSTc cluster 2 and cluster 3 at 58° and 179°, respectively; in LSTc-H3 cluster 3 at 40°; and in LSTc-H9 cluster 2 and cluster 3 at 36° and 135°, respectively.

Full-length glycan

The Φ, Ψ angles of all glycosidic linkages in populations and cluster representatives were also analyzed in conjunction with known crystal structures from human origins (Tables 1 and 2; Tables S1–S4). Radar plots of the cluster representatives were used to visualize and explore the potential correlation between the torsion and θ angles (Fig. 7). The θ angles of the two LSTa-H3 dominant clusters were similar, whereas the Gal2–GlcNAc3 Ψ angles, GlcNAc3–Gal4 Ψ angles, and Gal4–Glc5 Φ angles were fairly different (Fig. 7c). The θ angles of LSTa-H5 were similar too, albeit with different Sia1–Gal2 Φ, Ψ angles, Gal2–GlcNAc3 Ψ angles, and GlcNAc3–Gal4 Ψ angles (Fig. 7e).

The cluster 2 representative of LSTa-H3 (34% of population) had an induced Gal2–GlcNAc3 Ψ angle of 197°. The free LSTa representative clusters appeared to be similar to LSTa-H5 clusters, with the dominant cluster's θ angles differing by only 7°. The percentages of population covered by the free LSTa cluster representatives were also most similar to those of the LSTa-H5 (see Fig. 3a and e insets). LSTa-H9 was very restrictive in the Sia1–Gal2 Φ angle values.

The free LSTc cluster representatives differed with the bound LSTc in the Φ angles in the *anti* conformation, except for the minor clusters of LSTc-H9. Other differences in the ω angles were noted earlier. All the bound LSTc's had Gal2–GlcNAc3 Ψ angles (185°–195°) similar to that for the dominant cluster 1 in the free LSTc (208°). LSTc-H3 clusters allowed more variations in the ω angles. LSTc-H5 was most restrictive in the overall torsion angle requirements. LSTc-H9 clusters showed the most similarity to the

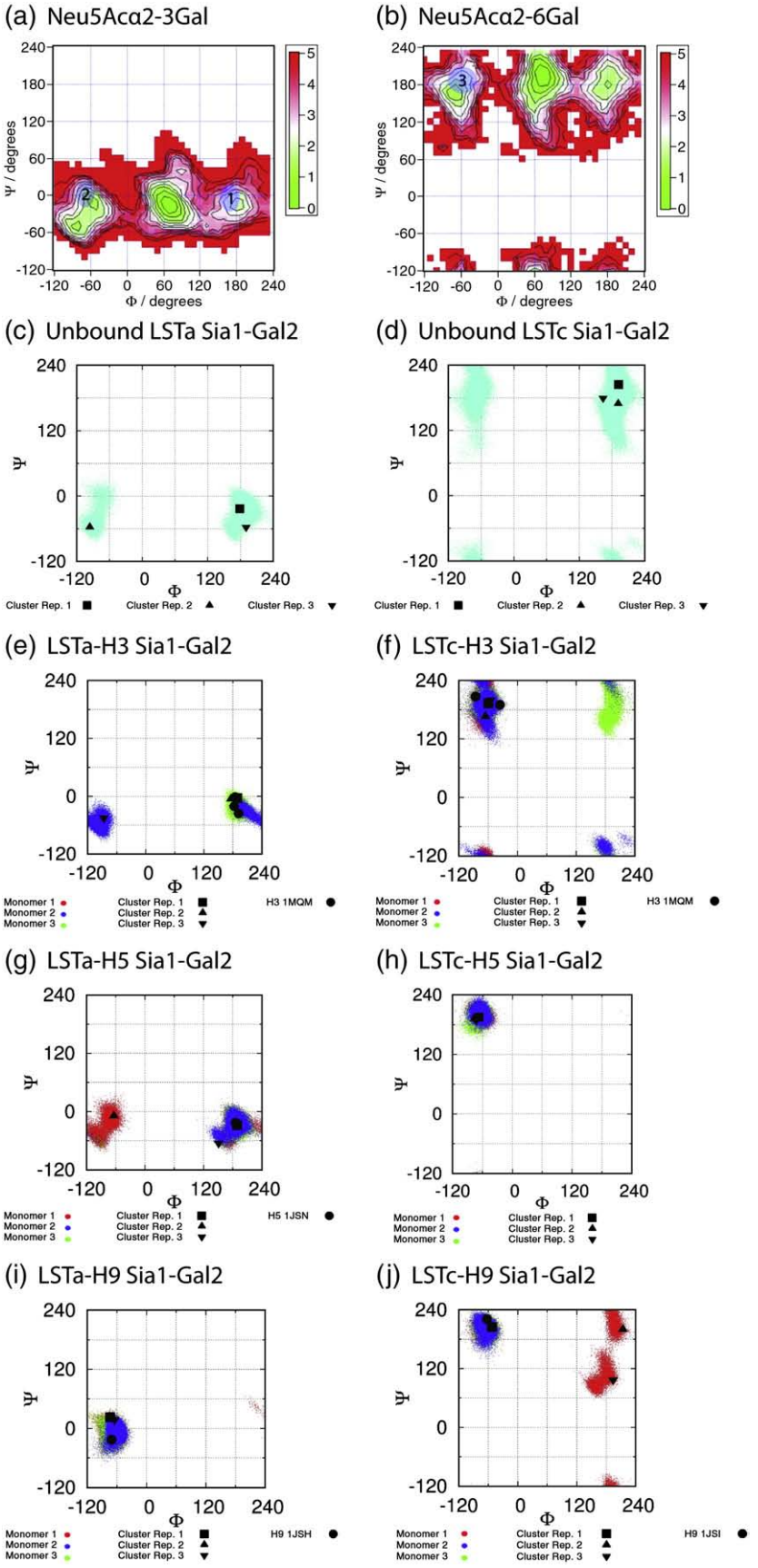

Fig. 5. The Φ, Ψ angle plots of the α-2,3 and α-2,6 linkages in LSTa and LSTc in solution and in complex with HAs. The GlycoMapsDB plots of the α-2,3- and α-2,6-linked disaccharides are shown in the top row for comparison; the encircled regions 1, 2, and 3 indicate the Φ, Ψ angles observed in the crystal structures.[19]

free LSTc clusters, with the exception of the Sia1–Gal2 Φ angles for the dominant clusters and the Gal2–GlcNAc3 Ψ angles for the minor clusters. LSTc-H3 and LSTc-H5 had almost identical torsion and θ angle distributions, except for variations in the ω and GlcNAc3–Gal4 Ψ angles.

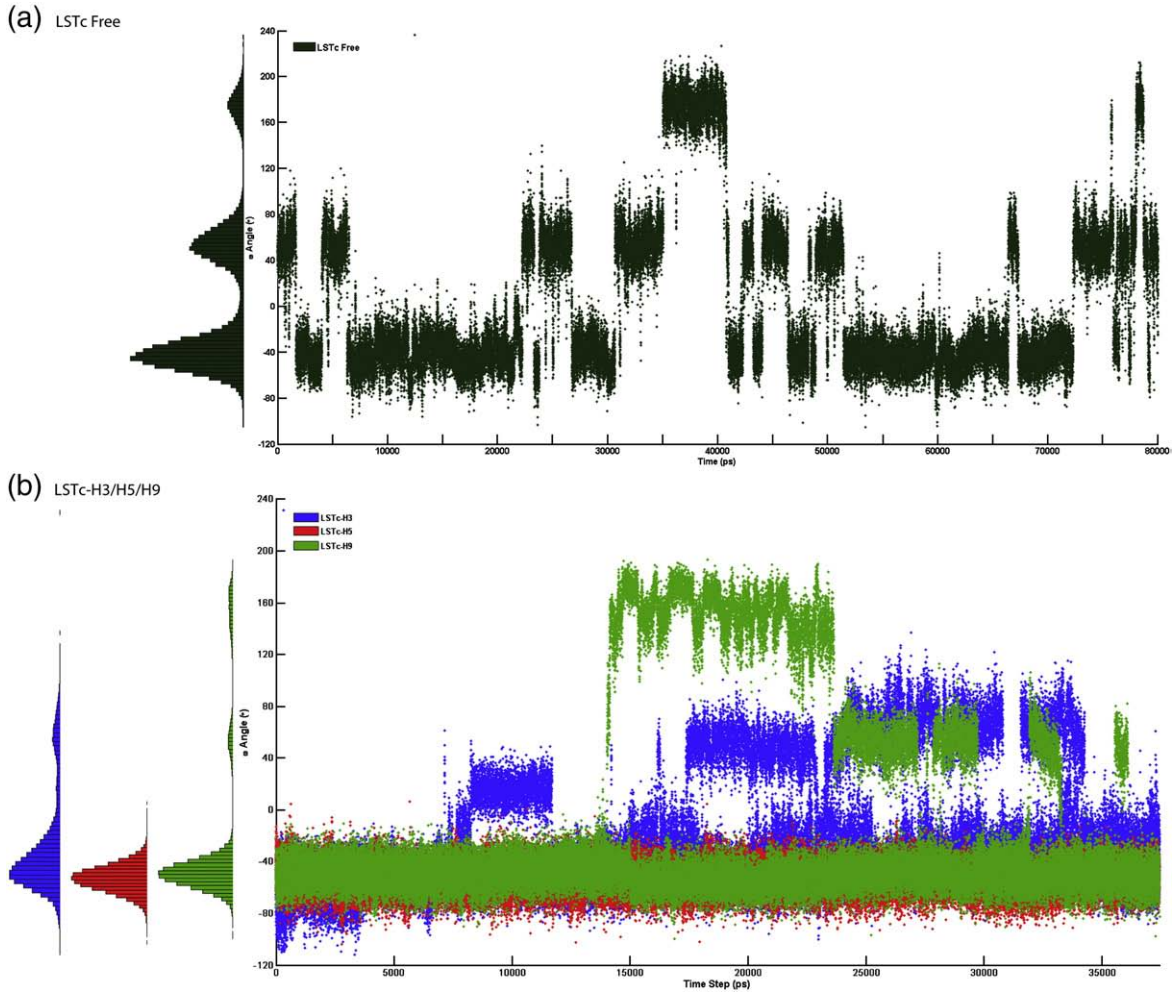

Fig. 6. The ω angle plot of the α-2,6 linkage in free and H3-, H5-, and H9-bound LSTc's along with their respective distribution histograms using a bin size of 5°.

Comparison with glycans bound to human HAs

The superposition of the free glycan torsional space and the representative bound structures from both the MD simulations and crystal structures revealed HA species-specific torsion angle preferences (Fig. 8). Of the Sia1–Gal2 glycosidic linkage, the Φ angles of the human LSTa-H1 (1RVX), the LSTa-H5 dominant cluster, and the first two clusters of LSTa-H3 (71%) showed the *anti* conformation found in ~81% of free LSTa. The Φ angles of the human LSTa-H3 (X-31) and swine LSTa-H9 were exclusively in the *gauche* conformation, found in ~18% of free LSTa (Fig. 8a). In contrast, even though the Φ angles were predominantly *anti* in the free LSTc, all the bound forms and crystal structures except for two minor clusters of LSTc-H9 were found in the *gauche* conformation (Fig. 8b).

Of the Gal2–GlcNAc3 glycosidic linkage, the Φ angles of the free LSTx adopted the *gauche* conformation and all the bound LSTx's are in the same conformation. One cluster of avian LSTa-H3 showed a Ψ angle in the *anti* conformation. The Ψ angles of the human LSTc-H3 and two of three LSTc-H1's were found in the *gauche* conformation, whereas those of LSTc bound to avian and swine HAs were in the less populated *anti* conformation.

There was no major difference between species at the distal GlcNAc3–Gal4 and Gal4–Glc5 linkages. Most torsion angles were located within the regions sampled in the free glycan simulations. More studies would be needed as only the human H3 (X-31) contains all the five sugar residues resolved.

Glycan–HA contact analysis

To investigate the distinctions among the glycan topologies, we carried out contact analyses on the glycan cluster representative–HA complexes to identify the key HA protein residues responsible for glycan recognition (Figs. 9 and 10; Figs. S7 and S8).

Y98 and 130 loop

Y98, a residue on the floor of the HA RBD cavity, interacted with both glycans in all HAs. The G/V/TSS sequence of the 130 loop interacted with both glycans in all the HAs studied here. The 130 loop provided more contact with Sia1 than any other SSE of the HA RBD. S133a, an insertion specific to H5,

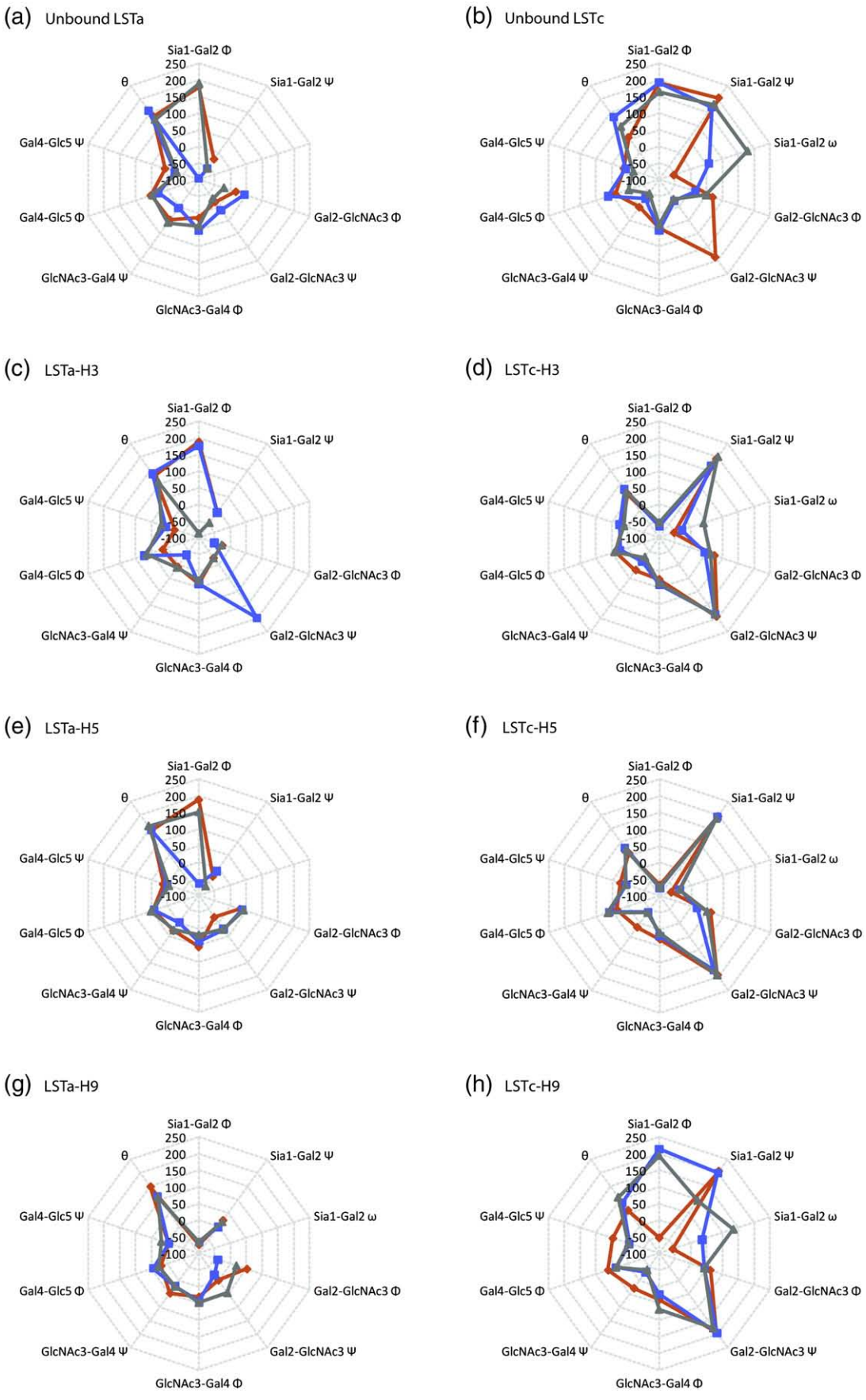

Fig. 7. Radar diagrams of the glycosidic torsion and θ angles in all glycan cluster representatives. Coloring scheme is consistent with Fig. 3. Cluster 1, orange; cluster 2, blue; and cluster 3, gray.

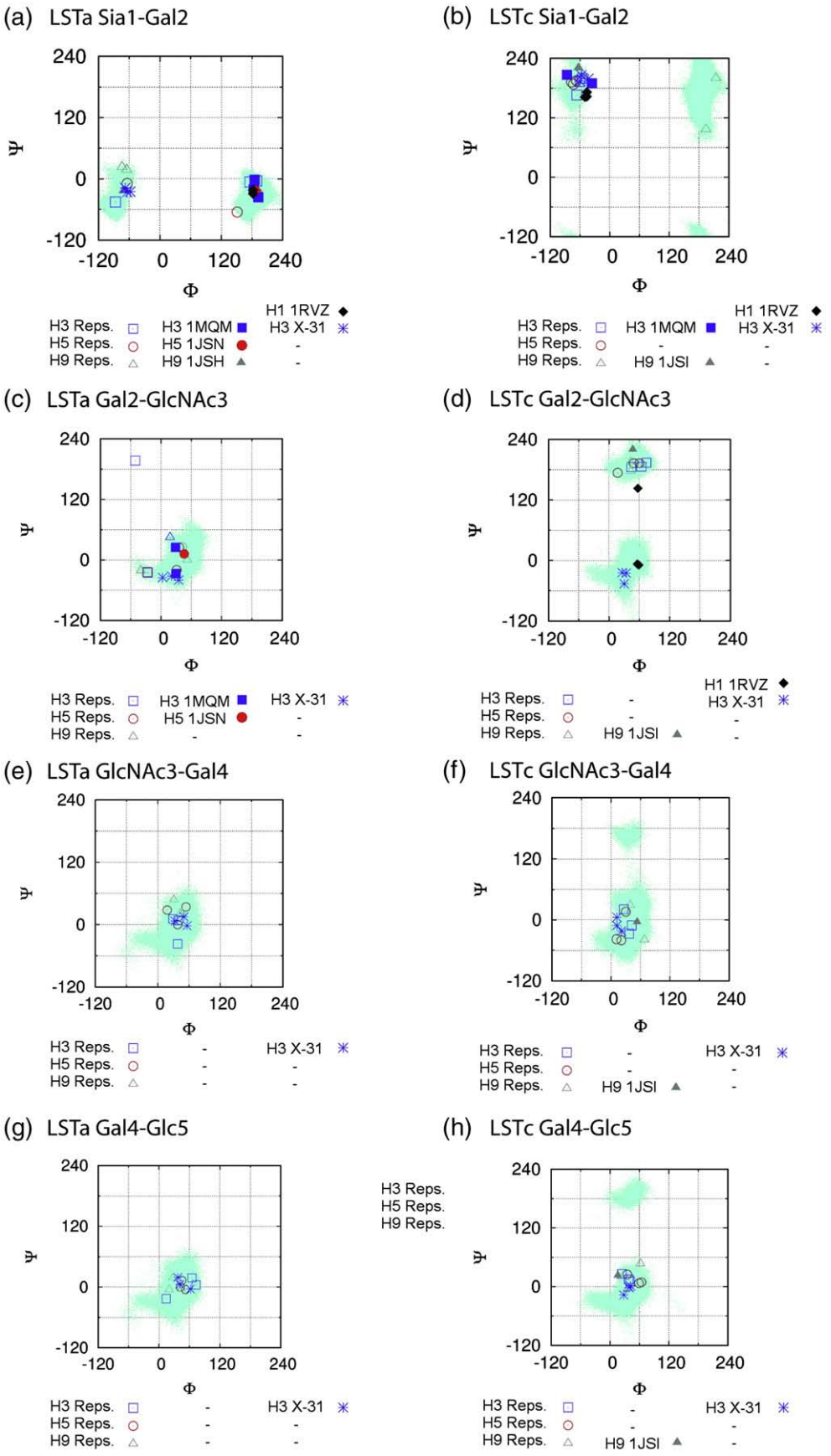

Fig. 8. The Φ, Ψ angle plots of all glycosidic linkages in LSTa and LSTc in solution (shown as background scatter plot) and in complex with HAs (shown as symbols for cluster representatives and crystal structures).

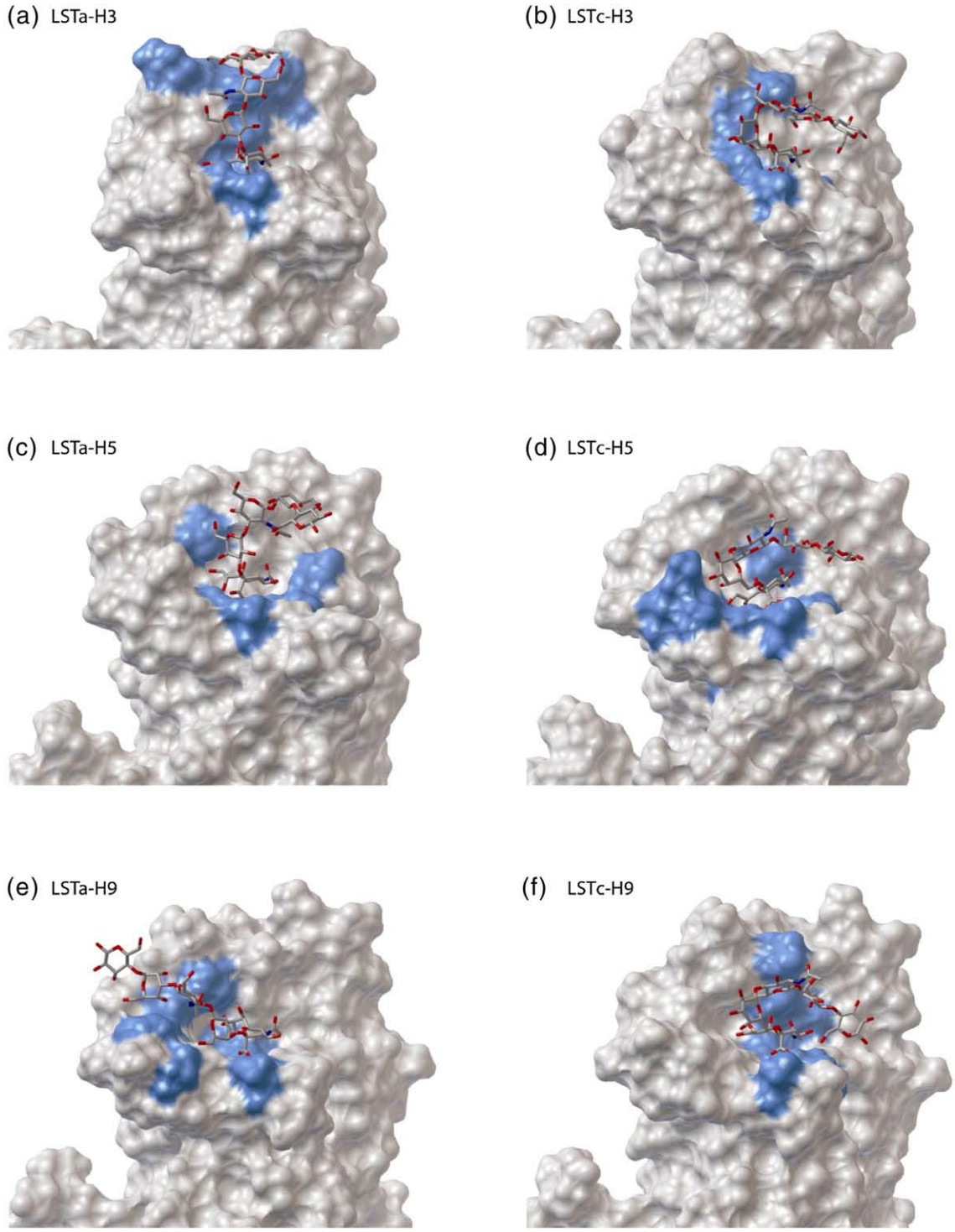

Fig. 9. The glycan dominant cluster representatives in complex with H3, H5, and H9. Contacted HA RBD residues are also highlighted.

had exclusive contacts with LSTa at Sia1. S136–GlcNAc3 and T135–Glc5 interactions were found in LSTc-H3 and LSTc-H9, respectively.

S145 and 150 loop

Usually far away from the bound glycans, S145 was contacted only by LSTa-H5 cluster 2. W153 was the main contact point between glycans and the 150 loop in all the HAs except for LSTa-H5. Residues 155–158 were favored by glycans Gal4–Glc5, particularly in LSTc. They also interacted with LSTa-H3 and LSTa-H5.

The 190 helix

The contact spectrum between the 190 loop and the glycans was broader and more extensive. H/

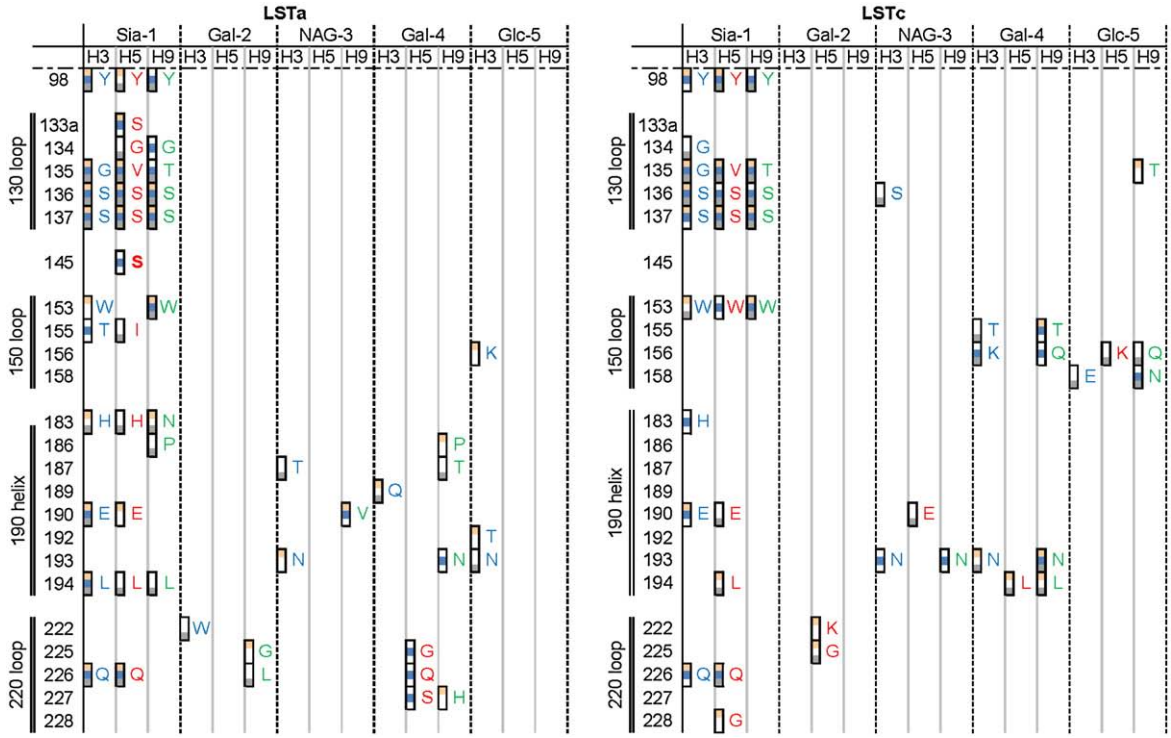

Fig. 10. Contacts between the sugar units of the glycan cluster representatives and the HA RBD residues. Clusters 1, 2, and 3 are shown in yellow, blue, and gray, respectively. HA residues are colored by strains: H3, blue; H5, red; and H9, green. Residue numbers are in human H3 numbering.

N183, E190, and L194 were found to make significant contacts with Sia1, particularly for LSTa. P186, T187, and Q189 were preferred by GlcNAc3 and Gal5 of LSTa. LSTa and LSTc contacted T192, N193, and L194 with GlcNAc3–Glc5.

The 220 loop

Q226–Sia1 interaction was favored in all HAs but H9. L226–Gal4 interaction was found in LSTa-H9. W/K222 and G225 were contacted by Gal2 and Gal4 in H5 and H9. S/H227 made contacts with Gal4 of LSTa-H5 and LSTa-H9. G228 was only found to interact with Sia1 of LSTc-H5.

Glycan–HA interaction energy

The interaction energies [electrostatic + van der Waals (vdW)] of the MD trajectories were computed and decomposed into individual sugar contributions in order to identify the key sugar units responsible for glycan–HA binding (Table 3).

Sia1 was clearly the predominant interacting glycan residue, accounting for more than 75% and 60% of the total interaction energies in the bound LSTa and LSTc, respectively. Substantially more contacts between the HAs and Gal2–Glc5 were found in bound LSTc than LSTa due to the preferential tightly folded umbrella or fishhook-like topology. This resulted in a significant increase in the interaction energy contributions from Gal2–Glc5 in bound LSTc than LSTa. In the context of the overall interaction energy, H3 preferred LSTa over LSTc and vice versa for H5 and H9. The interaction energy difference is more pronounced with H5, with a difference of almost 90 kcal/mol, compared with that of about 20–30 kcal/mol for H3 or H9.

Intramolecular hydrogen bonding in free glycans

LSTa and LSTc contain many polar groups, with ample opportunities for hydrogen-bond formation. However, hydrogen-bond analysis showed that hydrogen bonding played a minor role in intramolecular interactions. Neither glycan had hydrogen bonds with occupancies more than 50%. The RMSD plot in Fig. S2 shows that both glycans are extremely flexible in part due to the transient nature of the intraglycan hydrogen bonds.

Glycan conformational entropy

The abovementioned θ and torsion angles and volumetric analyses all suggested that significant conformational changes occurred in the glycans upon binding by different HAs. Unlike small ligands, conformational changes in the long and flexible carbohydrates may play a significant role in glycan–HA binding. The conformational entropy change ($-T\Delta S_{\mathrm{Conf}}$) associated with the loss of torsional degrees of freedom in the glycosidic linkages may have a substantial effect on binding affinity.[68,69] The conformational entropy values calculated using the

Table 3. Ensemble-averaged interaction energy decomposition between the individual glycan sugar units and the HAs (kilocalories per mole)

	H3	Percentage	H5	Percentage	H9	Percentage
LSTa						
Sia1	−131.33 (0.43)	72.90	−113.60 (0.27)	77.17	−75.43 (0.33)	75.16
Gal2	−12.16 (0.09)	6.75	−22.88 (0.38)	15.54	−3.20 (0.11)	3.19
GlcNAc3	−13.70 (0.15)	7.60	−5.36 (0.12)	3.64	−3.52 (0.01)	3.51
Gal4	−8.60 (0.17)	4.77	−3.68 (0.12)	2.50	−12.16 (0.08)	12.12
Glc5	−14.36 (0.41)	7.97	−1.68 (0.08)	1.14	−6.05 (0.08)	6.03
Total	−180.15 (0.13)		−147.2 (0.10)		−100.36 (0.07)	
LSTc						
Sia1	−100.30 (1.22)	61.53	−146.30 (1.19)	61.87	−75.70 (0.20)	57.31
Gal2	−10.38 (0.08)	6.37	−26.70 (0.21)	11.29	−4.64 (0.05)	3.51
GlcNAc3	−10.26 (0.12)	6.29	−26.70 (0.20)	11.29	−10.33 (0.03)	7.82
Gal4	−18.50 (0.17)	11.35	−21.30 (0.24)	9.01	−19.43 (0.15)	14.71
Glc5	−23.57 (0.40)	14.46	−15.46 (0.18)	6.54	−22.00 (0.13)	16.65
Total	−163.01 (0.26)		−236.46 (0.25)		−132.1 (0.06)	

Standard errors are given in parentheses.

torsional covariance method are listed in Table 4. $-T\Delta S_{Conf}$ increased in the order of LSTa-H5<LSTa-H3<LSTa-H9 and that of LSTc-H3<LSTc-H9<LSTc-H5. LSTc-H5 binding was opposed by the highest conformational entropy penalty of 5.52 kcal/mol, while LSTa-H5 experienced virtually no penalty. LSTc-H9 experienced a slightly lower penalty than did LSTc-H5, but LSTa-H9 entropy is the highest among the three LSTa bound systems. LSTc-H3 entropy was about half of LSTc-H5 and LSTc-H9, but LSTa-H3 entropy was much higher than LSTa-H5. The $\Delta(-T\Delta S_{Conf})$ difference between LSTc and LSTa bound to the same HA shows that LSTc-H3 binding was only marginally penalized relative to LSTa-H3 binding, whereas H5 was severely penalized for binding LSTc over LSTa by 5.24 kcal/mol. A moderate difference was found in H9.

MM-GBSA binding free-energy calculation

The MM-GBSA formalism has been successfully applied to carbohydrate–protein binding studies.[68,69] The energy components calculated from the MM-GBSA scheme using Eqs. (1) and (2) (Materials and Methods) are reported in Table 5. Electrostatic interactions ($\Delta E_{Electrostatic}$) were strong in H3 and H5 but also led to greater desolvation energy penalty (see also Fig. S10). An anticorrelation between electrostatic interactions and $-T\Delta S_{RTV}$ (entropic penalty due to loss of rotational, translational, and vibrational degrees of freedom) was also observed. LSTa-H3 and LSTx-H9 had relatively favorable vdW interactions ($\Delta E_{vdW}<-40$ kcal/mol). LSTc-H5 had better vdW interactions than LSTa-H5. $\Delta G_{GBTotal}$, the sum of enthalpic and desolvation terms, remained favorable across all glycan–HA complexes. LSTa-H5 had $\Delta G_{GBTotal}$ at −27 kcal/mol, whereas the others are all about −36 kcal/mol. However, LSTc-H3 and LSTc-H5 had high entropic penalties, 39 and 37 kcal/mol, respectively. In addition, LSTc-H5 and LSTc-H9 also had high conformational entropy penalties, 5.5 and 4.9 kcal/mol, respectively. After taking these entropic terms into consideration, we found that H3 had the biggest binding free-energy difference between LSTa and LSTc (almost 8 kcal/mol), followed by H5 (6 kcal/mol) and H9 (0.5 kcal/mol). Thus, based upon the energetics calculations, LSTc affinity was LSTc-H3<LSTc-H5<LSTc-H9, and LSTa affinity was LSTa-H5<LSTa-H3<LSTa-H9.

Table 4. Glycan conformational entropic changes between bound and free states

	H3	H5	H9
LSTa			
$-T\Delta S_{Conf}$	1.59 (0.34)	0.28 (0.01)	2.46 (0.02)
$-T\Delta S_{a\text{-}2,3}$	0.79 (0.18)	0.13 (0.03)	1.89 (0.03)
Percentage of contribution	50	46	77
LSTc			
$-T\Delta S_{Conf}$	2.47 (0.80)	5.52 (0.11)	4.85 (1.64)
$-T\Delta S_{\alpha\text{-}2,6}$	1.39 (0.45)	3.55 (0.06)	2.49 (0.82)
Percentage of contribution	56	64	38
$\Delta(-T\Delta S_{Conf})$	0.88 (0.43)	5.24 (0.06)	2.39 (0.82)

The percentage of contribution is $-T\Delta S_{\alpha\text{-}2,3}$ or $-T\Delta S_{\alpha\text{-}2,6}$ relative to $-T\Delta S_{Conf}$. All units are given in kilocalories per mole. Temperature is 310 K. $\Delta(-T\Delta S_{Conf})$ is the difference between $-T\Delta S_{Conf}$ of LSTc and that of LSTa bound to the same HA. Standard errors are given in parentheses.

Discussion

Glycan topological diversity and profiling

Free glycan topology and Sia1–Gal2 linkage types

The torsional angles in Sia1–Gal2 and Gal2–GlcNAc3 linkages are major determinants of sialoglycan topology. The θ angle, defined using atoms from these three residues, was proposed as a parameter to distinguish between two glycan topologies based upon a survey of existing crystal structures. HAs from human viruses appear to prefer long α-2,6 glycans with an umbrella-like topology (θ<110°), whereas those from avian viruses prefer long α-2,3 glycans with a cone-like topology (θ>110°).[19] The volumetric analysis, which presents the overall spatial volume

Table 5. MM-GBSA analysis of the glycan–HA binding energetics

Energy component[a]	LSTa-H3	SE	LSTc-H3	SE	LSTa-H5	SE	LSTc-H5	SE	LSTa-H9	SE	LSTc-H9	SE
$\Delta E_{\text{Electrostatic}}$	−147.98	0.5	−157.37	0.6	−105.28	0.5	−138.65	0.6	−59.85	0.4	−81.52	0.4
ΔE_{vdW}	−43.32	0.2	−37.36	0.2	−24.84	0.1	−36.46	0.2	−40.83	0.2	−43.50	0.2
ΔE_{MM}	−191.30	0.5	−194.73	0.5	−130.12	0.5	−175.11	0.6	−100.68	0.4	−125.02	0.4
$\Delta G_{\text{nonpolar}}$	−7.16	0.0	−6.63	0.0	−4.65	0.0	−7.09	0.0	−7.15	0.0	−6.90	0.0
ΔG_{polar}	163.33	0.4	164.24	0.5	107.76	0.4	144.54	0.4	72.30	0.3	94.64	0.3
$\Delta G_{\text{Solvation}}$	156.17	0.4	157.62	0.4	103.11	0.4	137.45	0.4	65.16	0.3	87.74	0.3
$\Delta G_{\text{GBTotal}}$	−35.13	0.2	−37.12	0.2	−27.01	0.1	−37.66	0.2	−35.52	0.1	−37.28	0.2
$-T\Delta S_{\text{RTV}}$	30.18	17.7	39.16	2.1	25.28	0.3	36.99	9.2	28.37	8.5	27.24	7.1
$-T\Delta S_{\text{Conf}}$	1.59	0.3	2.47	0.8	0.28	0.0	5.52	0.1	2.46	0.0	4.86	1.6
$\Delta G_{\text{Binding}}$	−3.37	5.9	4.51	0.8	−1.45	0.1	4.85	3.1	−4.70	2.8	−5.19	2.4

[a] Values are expressed in kilocalories per mole, averaged over 40 ns. ΔE_{MM} is the total gas phase energy as the sum of electrostatic and vdW contributions; $\Delta G_{\text{Solvation}}$, the GB solvation free energy as the sum of polar and nonpolar contributions; $\Delta G_{\text{GBTotal}}$, the total binding free energy before entropic connection; ΔS_{RTV}, the rotational, translational, and vibrational entropy calculated from normal-mode analysis; and ΔS_{Conf}, the conformational entropic penalty calculated from quasi-harmonic analysis of the glycosidic internal rotations.

sampled by the glycans, showed that the free long α-2,3 glycans, such as LSTa, did have the cone-like topology. However, the umbrella-like topology for the free long α-2,6 glycans, such as LSTc, did not hold. There was no stem-like structure at all as the glycan residues distal to the sialic acid sampled virtually all the spaces around. This was also corroborated by the absence of persistent intramolecular hydrogen bonds in the LSTc, as well as the greater RMSD observed for LSTc. Moreover, while two clusters of the free LSTc had the tightly folded (51%) or open umbrella-like (10%) topology, the cluster 2 representative, covering 39% of the population, had a θ angle of 133°, adopting a cone-like topology.

Inducible glycan conformations upon HA binding

The θ angle plots over the simulation time scale suggested that the θ angles of the different glycans could change significantly upon interactions with HAs. The differences existed even among different glycans interacting with different monomers within the same HA trimeric complex. By clustering the glycan conformations using the average linkage method, we obtained and examined the representative conformations using volumetric torsion angle analyses and radar plots. All the HAs examined induced conformational changes upon the glycans to various extents along the entire glycan. Even though free LSTa cluster representatives all had θ angles well above 110°, the bound LSTa-H3 and LSTa-H9 had cluster representatives, representing one-third of the populations, with θ angles below or close to 110°, adopting the umbrella-like topology. Similarly, LSTc-H9 contained a small population that had θ angles close to 110°, resembling a cone-like topology. Therefore, long glycans, such as LSTx, may adopt either an umbrella-like or a cone-like topology when bound to different HAs.

Which angles best describe glycan topological variations?

Taking the free and bound glycan simulation data together, the θ angle is not very useful in distinguishing between glycan topologies, even though it is a simpler parameter designed to represent conformational properties of the first three glycan residues. The torsional angles of the first three residues, and even the entire glycan, may need to be examined as a set to better understand the relationship between glycan topology and species-specificity shift. Based on our observations, the Sia1–Gal2 Φ angles, the Gal2–GlcNAc3 Ψ angles, and the ω angles of Sia1–Gal2 showed greater variations and have well-defined ranges. For example, the ω angle values varied between three modal values, the Gal2–GlcNAc3 Ψ angles varied between two modal values, and the Sia1–Gal2 Φ angles varied between three modal values. Further experiments with different types of glycans are required to develop better ways to describe glycan topological profiles.

Biological implications of glycan topological diversity

HA–glycan interactions fall into the scope of carbohydrate–protein interactions, which are often characterized to be of high avidity and low affinity.[33] It has been hypothesized that the most frequently seen glycan shape, not the average conformation, may be essential for recognition by carbohydrate binding proteins.[70] This also implies that free glycan conformations most similar to the bound conformations may greatly enhance the binding kinetics. However, the flexibility of the carbohydrate binding proteins may also influence the binding equilibrium by inducing glycan conformations to be much more favorable in the final protein–glycan complexes. In this case, both glycan and RBD sequence compositions become important. Therefore, topology may be important in terms of selectivity, and specificity is most likely established through favorable interactions between complementary interactions at a sequence-specific level.

The presence of long α-2,6 glycans in the cone-like topology may allow selective recognition by avian HAs. Clearly, avian HAs may also accommodate long α-2,6 glycans in the tightly folded umbrella topology, albeit with much lower affinity based upon the MM-GBSA results. These observations

could help explain cases in which viral entry is observed even though no α-2,3 type of glycan is detected.[23] The relative abundance of the preferred glycan topology may also play a role in species specificity, possibly by influencing the binding kinetics and/or equilibrium shift. The free LSTa and LSTc differed in the relative abundance of the conformations preferred by avian H3, H5, swine H9, and human H3 and H1 viruses (shown in Fig. 8). The human H1 and H3 appeared to favor the Gal2–GlcNAc3 Ψ angles that were more abundant in the free LSTc, whereas the avian and swine ones favored the less abundant ones. With LSTa, the distinction appeared to be mostly at the Sia1–Gal2 Φ angles, with the human H3 preferring the less abundant ones. However, human H1 preferred the same Sia1–Gal2 Φ angles as the avian and swine HAs. More studies are necessary to characterize the correlation of topological profiles and species-specificity shift.

In addition, glycan length may certainly influence the glycan topology and, possibly, species specificity. The Φ, Ψ plots of the free pentasaccharides were comparable with those of the disaccharide (Sia–Gal) simulations from GlycoMapsDB, except for the *gauche* conformation (Φ of ~60°) (Fig. 5). The absence was also observed in the sialyl-α-2,6-lactose, a trisaccharide, in aqueous solution and in other glycans co-crystallized with HA[19,71] (Fig. 5). Even though earlier studies focused on disaccharides or trisaccharides, our glycan contact analysis and interaction energy analysis showed that Gal4 and Glc5 may still contribute to HA–glycan interactions in a significant way. It is quite possible that these distal interactions may be important in helping define the specificity in low-affinity interactions between HA and glycans. For example, the Gal4–Glc5 interactions account for more than 31% of the interaction energies in LSTc-H9 and for 18% in LSTa-H9.

Glycan flexibility and dual specificity of HA

A number of human isolates have been found to have dual specificity, with human adapted viruses showing reduced α-2,3 glycan binding and enhanced α-2,6 preference.[72–74] The duck H3 used in this simulation is an avian precursor to a human H3.[43] LSTa-H3 could also adopt an umbrella-like topology (θ = 106°). In LSTa-H3 and LSTc-H3, the three clusters were evenly distributed and have topological profiles similar to those for free glycans (Fig. 7). Avian H3 may be able to interact with both types of glycans, consistent with experimental studies.[4,8] The duck H5 was quite stable with the tightly folded umbrella-like topology of LSTc or the narrow cone-like topology of LSTa. Swine H9 is known to recognize both avian and human receptors.[34] The LSTc bound to one of the monomers in H9 opened to θ angles close to 110° but gradually closed again, suggesting that H9 contained residues that favored interactions with LSTc in the tightly folded umbrella-like topology. Similarly, the ability for H9 to reduce the θ angle of LSTa from a narrow cone-like topology to one that resembled an umbrella-like topology (θ = 111°) suggests that H9 is capable of inducing or accommodating drastic changes in glycan topology. Therefore, swine H9 may interact with a cone-like topology or induce an umbrella-like topology in long glycans with α-2,3 or α-2,6 linkages. Our energetics analysis did indicate that while swine H9 could bind LSTa and LSTc equally well, avian H3 and H5 did prefer LSTa over LSTc. Therefore, while these HAs may be dual specific, the differential binding affinity would limit their spread in humans, notwithstanding the relative abundance of glycan receptors in the human respiratory tract.

These data suggest that the flexibility inherent in the glycans, naturally adopted or induced upon HA binding, is consistent with the dual-specific interactions observed in previous studies. The selective pressure over the course of evolution may also favor viruses that may interact with more varieties of glycans through multivalent interactions. However, the distinct dominant cluster topologies exhibited by different glycans upon HA binding indicate that receptor abundance, final topology, binding kinetics, and accessibility may be important factors for efficient viral attachment and entry.

Contact and energetics analysis of glycan–HA interactions

Understanding the glycan–HA binding mechanism requires the elucidation of both structural and energetics aspects of the process.

System-level interaction energy analysis

The interaction energy of the individual sugar units and the entire HA populations shown in Table 3 clearly demonstrated the crucial role of Sia1 in anchoring the glycans in the HA RBD. The contact analysis showed that Sia1 made more contacts with the HA RBD residues than any other sugar units (Fig. 10), and it accounted for the majority of the total interaction energy among all HAs. Although Sia1 is the dominant factor, the asialo-sugar units Gal2 to Glc5 exhibit increased contribution when the appropriate topology of a glycan leads to advantageous contacts, which may be important for viral recognition of different hosts. For example, the representative clusters, especially in LSTa, demonstrated spatial orientation preferences with respect to the SSE of HA (Fig. 4).

The flexible LSTa-H5 mainly benefited from Sia1 and Gal2 interactions, which constituted 93% of the total energy. The rest of LSTa-H5 stretched out of the RBD binding site, away from the proximity of any protein residues except for some rare occurrences of Gal4–HA contacts. Such cone-like conformations would presumably facilitate HAs access to the Sia1–Gal2 units of the glycans attached to membrane proteins or lipids. On the other hand, all the sugar units of LSTc-H5 made contacts with one or more H5 RBD residues, and this tightly folded binding mode was rather rigid. For a productive encounter

with HA, this conformation is likely to require a longer asialo-polysaccharide stem to provide the necessary steric clearance and proper orientation. Accessibility to the HA RBD could also be blocked by glycosylation events nearby.

Although the LSTc was originally obtained from the same swine H9 crystal structure, the amino acid interaction profiles with different HAs were very different over the course of the simulation. The absence of interaction with E190, L194, Q226, and G228 in H9 by LSTc may explain why the avian H5 showed stronger interaction with LSTc. Besides the steric interference and accessibility issue discussed earlier, it is also possible that overly strong interactions might not be desirable for efficient viral entry or transmission and present a negative selection pressure. This may help explain why swine is suitable as an intermediate reassortment host.

Glycan conformational entropy analysis

The average square inner product plots and the occurrence of multiple low-frequency ω angle transitions (Fig. 6) demonstrated that both free LSTa and LSTc had reached adequate convergence in conformational space within 80 ns of simulation. The sampling convergence provided the basis for accurate determination of glycan conformational entropy using the quasi-harmonic torsional covariance method. Table 4 shows that $-T\Delta S_{\mathrm{Conf}}$ is indeed significant, consistent with other carbohydrate–protein binding studies.[68,69] It also shows that the contribution from the α-2,3 or α-2,6 linkages to the overall conformational entropy penalty is relatively large (40%–80%), confirming the importance of the Sia1–Gal2 linkage in glycan binding. However, contributions from other linkages are not negligible. Omitting these contributions would lead to incorrect conformational entropy prediction.

The conformational entropy results are consistent with the conformational changes shown in the radar plots (Fig. 7), the Sia1–Gal2 Φ/Ψ plots (Fig. 5), and the glycan cluster representative data (Tables 1 and 2). The radar plots summarize the overall behavior of all torsions in the glycan cluster representatives. LSTa-H9 had the largest deviation from free LSTa, particularly in Sia1–Gal2 Φ/Ψ angles. This justified the highest entropic penalty that LSTa-H9 received, 77% of which was from Sia1–Gal2 linkage (Table 4). H9's inversion of the generally preferred *anti* conformation in free LSTa to the *gauche*-only conformation in bound LSTa also resulted in the high percentage of contribution to the entropic penalty (Fig. 5). In contrast, the LSTa-H5 cluster representatives closely resembled those of free LSTa, and LSTa-H5 entropic loss was rather minimal. Comparatively, LSTa-H3 lost a moderate amount of entropy on Sia1–Gal2 Φ and Gal2–GlcNAc3 Ψ angles, with changes in not only angles but also the relative percentage of population.

Compared with LSTa, there was considerable entropy loss across all bound LSTc's because their predominant cluster representatives deviated markedly from those of the free LSTc. The entropic penalty was, to some extent, anticorrelated with the dominant cluster abundance. The higher the percentage of population of the bound glycan dominant cluster, the lower the degree of conformational freedom and thus the higher the entropy penalty. The more rigid LSTc-H5 (96.8%) lost a substantial amount of entropy (5.52 kcal/mol), followed by LSTc-H9 (77.7%; 4.85 kcal/mol) and LSTc-H3 (46.7%; 2.47 kcal/mol). The elevated Sia1 in LSTc-H3 brought additional conformational freedom[43] and alleviated its entropic loss.

All LSTc-H3 and LSTc-H5 cluster representatives deviated from free LSTc in terms of Sia1–Gal2 Φ and ω angles. LSTc-H9 was more flexible in this aspect; its second and third cluster representatives were more similar to free LSTc over these angles. This contrast was precisely reflected in the percentage of contributions of LSTc Sia1–Gal2 linkage to the conformation entropy penalty. LSTc-H9 had the lowest contribution (38%), down from LSTc-H5 (64%) and LSTc-H3 (56%). The slightly more flexible ω angle in LSTc-H3 might contribute to its lower percentage compared with LSTc-H5.

In summary, LSTa-H5 seemed as flexible as the free LSTa. LSTc became much more rigid upon binding, resulting in larger overall entropic penalties than LSTa. This is consistent with the observation that LSTa-HA interactions occur mainly on the first three sugar units Sia1–Gal2–GlcNAc3, whereas LSTc has substantial interaction between all sugar units and HA RBD. All HAs caused the free glycans to adopt unique conformations upon binding in order to create favorable interactions. The large 5.34-kcal/mol entropic difference significantly disadvantaged LSTc binding over LSTa on H5, whereas $-T\Delta S_{\mathrm{Conf}}$ did not make much difference in H3.

Glycan binding free-energy analysis

The inclusion of conformational entropy has been shown to help achieve better protein–carbohydrate binding energy prediction using the MM-GBSA formalism.[68,69] The binding free-energy results indeed captured the experimentally established trend characterized by these HA strains (Table 5; Table S5). H9 effectively accommodated and bound to LSTa and LSTc because of its large RBD and key interacting residues reflecting swine as an intermediate host between avian and human. Avian H3 and H5 preferred LSTa over LSTc. The overall predicted binding free energies suggested that these are low-affinity interactions and would be compensated due to multivalent interactions between multiple viral HAs and cell surface glycan receptors.[33]

Nonbonded vdW interactions are often found to have a strong correlation with binding since ΔE_{vdW} represents intermolecular packing and is a good indicator of the level of intermolecular shape complementarity.[75] The vdW interaction energies correlated well with the glycan–HA shape complementary observed in Figs. 9 and 10. H9 had favorable

vdW energies for LSTa and LSTc, indicating that H9 RBD provided a good fit for both glycans. Indeed, Figs. 9 and 10 showed that LSTa and LSTc Sia1 were shifted toward the H9 130 loop. As a result, Gal4–Glc5 of LSTa-H9 was able to make favorable contacts with the groove formed by the 220 loop and the 190 helix while simultaneously accommodating the extended length of LSTc-H9 introduced by the α-2,6 linkage and the placement of Gal2. In LSTa-H3, the elevated Sia1 helped achieve advantageous shape complementarity, with Gal4–Glc5 reaching to the top of the 190 helix with substantial interactions. However, the elevated Sia1 had less effect on LSTc-H3 due to the relatively smaller RBD size compared with H9. LSTa-H5 had the lowest vdW energy resulting from its great conformational flexibility. LSTc-H5 had moderate vdW energy, indicating that shape complementarity was moderate despite the rigidity displayed by bound LSTc, which in fact worked against binding due to the substantial loss of configurational and conformational entropy.

The desolvation penalty also correlates with vdW interactions and shape complementarity. In general, the better the fit, the higher the desolvation energy, as shown in H5 and H3 binding. However, if an HA has a larger RBD, the desolvation penalty is reduced because not all solvent molecules need to be expelled from the RBD upon binding. This may explain why H9 had not only favorable vdW interaction but also significantly less desolvation penalty than H3 and H5.

The MM-GBSA binding energies suggested that a large RBD may facilitate the binding of such glycans as LSTa and LSTc, particularly for LSTc and similarly other long α-2,6-linked glycans. A large RBD not only accommodates long glycans well but also effectively reduces desolvation and entropic penalties. It benefits glycan binding in two ways—that is, having more freedom to make favorable binding conformational adjustment and lowering conformational entropy penalty. Of course, this may also explain why H9 is less selective, since its electrostatic terms are much smaller compared with those of H3 and H5.

Experimental binding affinity data were not directly available for LSTx and the HAs used in this study. Selected earlier studies using trisaccharides and glycan microarray data were collected and are listed in Table S5 for indirect comparison. The trisaccharides may have different regiochemistries (Galβ_{1-4}Glc *versus* Galβ_{1-3}Glc) and sugar compositions (Glc3 *versus* GlcNAc3) in 3′SL (sialyl lactose) and 3′SLN (sialyl lactosamine), respectively. Glycan microarray studies on two closely related H5 mutants indicated that there could be significant differences in their binding to glycans that vary in composition, such as Gal5 *versus* Glc5, length, stereochemistry, and regiochemistry, such as Galβ_{1-3}GlcNAc *versus* Galβ_{1-4}GlcNAc[42] (Table S5). Therefore, the results may not be directly comparable. The MM-GBSA binding energies reflected the general trend exhibited in the binding constants of trisaccharides. A/Duck/Singapore/3/1997 (DK97) H5 shows no binding to α-2,6-linked 6′ SLN in glycan microarray experiments, whereas A/Vietnam1204/2004 (Viet04), a human H5, showed strong binding. Shtyrya *et al*. found no binding between A/Duck/Ukraine/1963 H3 and 6′SLN.[76] Swine H9 is known to bind both α-2,3- and α-2,6-linked glycans, and avian HAs are known to have stronger preference to α-2,3-linked glycans.

A limitation of MM-GBSA calculations is the use of a simple proportional surface area term to account for the nonpolar solvation contribution, which has been pointed out in the literature. The use of a more sophisticated treatment of nonpolar solvation could further improve the agreement with experiments.[77] Since solvent is treated implicitly in MM-GBSA calculations, water-mediated binding interactions are not captured in detail. Configurational entropy normal-mode calculations are computationally quite expensive for large protein–ligand complexes. This approach also neglects the contributions to the entropy change resulting from a change in the number of thermodynamically accessible energy minima.[78] Application of free-energy pathway methods, such as thermodynamic integration and free-energy perturbation, which involve explicit treatment of solvents, could improve predictive accuracy.[79] However, due to the high computational cost and convergence difficulties associated with these free-energy pathway methods, MM-GBSA has remained an attractive approach.

Monitoring binding specificity switch and drug design opportunities

The contact analysis illustrated the relative probability of Sia1 contacting with different HA SSEs (Fig. 10). In particular, swine H9 had the lowest interaction energies with LSTa and LSTc. The absence of Sia1 contacts with L226 and V190 in swine H9 was reminiscent of the Q226L and E190D mutations in H2 and H3, which are responsible for the avian-to-human specificity switch.[18,40] In H1, the E190D and G225D substitutions reduce α-2,3 binding and enhance α-2,6 binding.[80,81] Using glycan microarray, Stevens *et al*. showed that a D190E reverse mutation is sufficient to switch human H1 (NY18) back to avian specificity.[27] Gamblin *et al*. showed that the 1-Å difference in Q226 placement may prevent human H1 from interacting with the avian receptor, in contrast to avian H5 and H7.[35,82] A number of steric clashes and conformational differences in the RBD are postulated to account for the inability of human receptor to interact with avian H5/H7.[37] While we have not examined H1 simulations with these mutations, the loss of interactions with these HA residues may be an important indicator for the specificity switch. More recent studies using human isolates of H5 (KAN-1) add support for the role of S137A, T192I, N186K, E190D, Q196R K193S, G225D, Q226L, and G228S mutations in affecting recognition of glycans and cell entry.[42,73]

The individual glycans examined in our study contacted these amino acid residues, albeit with distinct interaction profiles (Fig. 10). More contact residues were identified from this simulation than from crystal structures because the full glycans were not resolved. The contact patterns displayed by the dominant cluster representatives may be important for monitoring of changes in binding specificity. For example, L194 and G228 were new interactions observed only in LSTc-H5, not in LSTa-H5. However, the interaction with S133 was only present in LSTa-H5. One caveat is that these contact residues were observed using a representative structure of the MD snapshots and may not be exhaustive in identifying all possible contacts. Some deviations from crystal structure-based examinations were noted.[34] More detailed interaction energy analyses based upon the individual clusters are needed to further characterize the individual contributions from the different RBD residues.

Most complex glycans are biantennary, triantennary, or mixed-antennary glycans, with the antenna length varying from 3 to 7 sugar units.[19] Glycan microarray studies on two closely related H5 mutants indicated that there can be significant differences between their binding to glycans that vary in composition and length[42] (Table S5). The avian DK97 H5 strain differs from the human isolate Viet04 strain in only two residues (Viet04/Sing97: R/E216, S/P221) in the 220 loop of the RBD. DK97 exhibits reduced binding of α-2,3 glycans compared with Viet04, and only H1 G225D, H3 G228S, and H3 G228S+Q226L receptor switch mutations confer Viet04 the ability to interact with biantennary α-2,6 glycans.[16] The same study showed that the type of GAL2–NAG3 linkage (β3 or β4) may also affect the ability of HA to bind the pentasaccharides examined, which has a GAL5 (in a second lactosamine unit) instead of a GLC5 found in LSTa/c. In short, the specific composition of a glycan may be critical in the determination of binding specificity.

The fact that Viet04 binds to both avian and human cell receptors but DK97 only binds to avian cell receptor might imply that receptor specificity of Viet04 is halfway to switching from avian to human type. However, glycan microarrays are still difficult to synthesize and do not contain complex glycans. The multivalent interactions were mimicked through the use of secondary antibodies. However, such approaches have been criticized for using nonphysiological concentrations of the glycans. The MM-GBSA plus entropic analysis showed good correlation with experimental data and could be used to perform *in silico* studies to help monitor binding specificity switch. While the MM-GBSA analysis showed that the interaction between LSTc and avian H5 is highly unfavorable with heavy entropic penalties, it does point out the potential for synthetic glycomimetic inhibitors that could block viral infections by blocking the HA RBD. In fact, LSTc is found in natural breast milk, and it is tempting to speculate that its presence may help prevent flu infections in infants. The energetics analyses reported in this study would facilitate the design of such inhibitors by taking into account the effect of desolvation and entropy.

Conclusions

Characterization of the glycan–HA binding interaction is challenging due to the flexible nature of polysaccharides, which could populate many energetically favorable conformational states. The key structural and dynamic properties of the free and bound glycans were studied through monitoring the RMSD, glycosidic torsion, topological θ angles, atomic density volume, and RMSD-based clustering of glycan trajectories over the course of 40 ns of simulations of various HA trimers and 80 ns of simulations of free glycans. α-2,6-Linked long glycans, such as LSTc, could adopt a cone-like topology previously thought to be unique to α-2,3-linked long glycans, such as LSTa. On the other hand, α-2,3-linked long glycans might adopt umbrella-like conformations upon binding to HAs. Our results revealed that the long glycans, such as LSTx, exhibited distinct topologies in solution and in complex with different HAs and defied the cone- or umbrella-like topology classification based on the θ angle definition. A comparison of three-dimensional atomic density volumes suggested the bound LSTc topology was restrained in spatial orientations that may be better characterized as fishhook-like. The interaction energy decomposition of individual glycan sugar units reflected the contacts associated with the representative binding modes, underscoring the indispensability of Sia1 and the significance of sugar composition for the glycan receptor binding. With the inclusion of glycan conformational entropy, the MM-GBSA analysis predicted the correct binding characteristics of avian H3, H5, and swine H9 with avian and human cell receptor analogues LSTa and LSTc. The energy components, such as electrostatics, vdW, desolvation, and configurational and conformational entropy, provided quantitative descriptions of the glycan–HA binding process and can be utilized as parameters for future studies. By introducing variations in both glycan chemical properties and HA RBD residues, the MD-based computational studies in this work can be readily extended to complement glycan microarray and other experimental efforts and will play an important role in the surveillance and prevention of future pandemics.

Materials and Methods

System preparation

The coordinates of LSTa and LSTc bound to H3, H5, and H9 were obtained from the PDB and are listed in Table S9. Complete LSTc was extracted from the LSTc-H9 crystal structure complex (PDB code 1JSI). LSTa was only available in its trisaccharide (Sia1–Gal2–GlcNAc3) form

in the PDB crystal structures. The missing Gal4 and Glc5 units were added using Maestro GUI (Schrödinger, Inc.). The glycans were then superimposed on the sialic acid moiety in their respective crystal structures. Segments of the complete pentasaccharides were substituted by available counterparts in the crystal structure to ensure the accuracy of the initial glycan conformation. The Φ and Ψ torsional angles of the added Gal4 and Glc5 units were assigned automatically by Maestro and manually adjusted to remove any steric clashes. The starting glycan conformations are shown in Fig. 1; the sequence of monosaccharides and their linkages were illustrated using Consortium for Functional Glycomics† nomenclature.

All glycan–HA bound systems were prepared as complete trimeric complexes for the MD simulations. UCSF Chimera[83] was used to perform symmetry transformation from monomer to trimer using crystal records for H5 and H9. Protonation states of the protein residues were determined at pH 7.4 by the National Biomedical Computation Resource PDB2PQR‡ Web service.[84] The HA protein and glycan systems were parameterized using the FF99SB AMBER force field[46] and GLYCAM06 force field,[44] respectively. Each system was solvated in a TIP3P water box,[85] leaving 10 Å between the solute surface and the box boundary. All crystallographically resolved water molecules were retained in the glycan–HA bound systems. The systems were neutralized, and an ionic concentration of 0.15 M NaCl was introduced to mimic experimental assay conditions. The abovementioned preparative procedures were performed using the LEaP module of AMBER9 suite.[86] VMD[87] was used to measure the solvation box center and dimensions for MD simulation setup, and Autoionize plug-in was used to verify the ionic concentration.

MD simulations

All MD simulations were carried out using NAMD 2.6[88] with a 1-fs time step, and periodic boundary conditions were applied in conjunction with particle-mesh Ewald summation[89] to treat long-range electrostatics. Bonds involving hydrogen atoms were constrained by the RATTLE algorithm.[90] The temperature was maintained with Langevin dynamics, and a hybrid Nosé–Hoover Langevin piston method[91] was used to maintain pressure at 1 atm. A multiple-time-stepping algorithm was employed, in which bonded interactions were evaluated at every time step and short-range nonbonded interactions and long-range electrostatic interactions were evaluated every two time steps at a cutoff of 14 Å.[89,92] The one to four electrostatic and vdW interactions were scaled by the standard AMBER values (SCEE = 1.2, SCNB = 2.0) in the simulations of glycan–HA complexes, while these scaling factors were removed in the unbound glycan simulations as recommended by GLYCAM06.[44,45] Energy minimization was performed on the free glycan and bound glycan–HA systems for 50,000 steps. The following glycan and protein atoms were gradually relaxed in four phases: (1) hydrogen atoms for 5000 steps; (2) hydrogen atoms, water molecules, and ions for 5000 steps; (3) all atoms except for protein backbone for 5000 steps; and (4) all atoms for the final 25,000 steps. The systems were then equilibrated at 310 K in the isothermal–isobaric (NPT) ensemble[93] using a harmonic constraint force constant of 4.0 kcal/mol/Å^2 applied to protein backbone atoms with scaling factors of 1.0, 0.75, 0.50, and 0.25 for the sequential segments, respectively. Production runs were subsequently performed at a total of 40 ns for each bound system (LSTa-H3, LSTc-H3, LSTa-H5, LSTc-H5, LSTa-H9, and LSTc-H9) and that of 80 ns for the free LSTa and LSTc. No restraint was applied to any of the simulations. The simulations were carried out on the Maui High Performance Computer Center's Dell Power Edge 1955 system using 64 processors (benchmark = 0.5 ns/day), the National Center for Supercomputing Applications' Xeon cluster (Tungsten) using 128 processors (0.5 ns/day), and the National Biomedical Computation Resource's AMD Opteron cluster (Oolite) using 64 processors (0.8 ns/day).

MD trajectory postprocessing

The three bound glycan trajectories were extracted from each of the solvated glycan–HA trimeric systems by removing the solvent and ions. The free glycan trajectories were obtained in a similar fashion.

Convergence of free glycan conformational sampling

To examine the free glycan simulation convergence in terms of conformational properties, we carried out an iterative principal component analysis procedure on all glycosidic torsion angles along trajectory time steps of increasing length (e.g., snapshots 1–1000, snapshots 1–2000, etc.) to evaluate the significance of sampling. Essential space was compared through the inner product of the corresponding principal component analysis eigenvectors, similar to the technique for calculating protein essential dynamics proposed by Amadei *et al.*[94] and Van Aalten *et al.*[65] This procedure was implemented in a custom MatLab script.

Glycan RMSD calculation

In the context of glycan–HA binding interactions, most of the glycan conformational changes occur relative to their Sia1 units. It has been observed that Sia1 placement is relatively stable in the HA RBD. In analyzing the glycan conformations, a global RMSD alignment of the glycans is obviously inappropriate. We decided to align the glycan trajectory frames on the heavy atoms (C and O) of the Sia1 pyranose ring in order to remove the overall rotation and translation. The RMSD was measured on the heavy atoms of Gal2–Glc5 six-member pyranose rings between the individual glycan trajectory and the glycan MD starting structure. The superimposition and RMSD calculations were performed using the Wordom program.[95]

RMSD clustering

The Sia1-aligned glycan trajectories of the same HA were concatenated into a single trajectory. The hierarchical average linkage clustering analysis was then carried out on the RMSD distance matrix of the single trajectory using custom Bio3D/R[96] scripts. The hierarchical average linkage clustering method is widely used because of its superior performance in producing clusters with the smallest within-cluster variance and large between-cluster separation compared with many other clustering algorithms.[66] It also provides statistical support for rational cluster structure determination and has been successfully applied in MD trajectory analysis, [66] microarray gene expression,[97] and other biomedical data analyses.[98] A

† http://www.functionalglycomics.org
‡ http://nbcr.sdsc.edu/pdb2pqr

three-cluster solution was selected for the sake of simplicity. The structures of the glycan cluster representatives and the cluster percentage of population were extracted (Fig. 4). The agglomerative clustering process and the RMS distances at which clusters were merged are illustrated in the dendrograms.

Glycosidic torsion angles

Glycan internal motions can be described using the glycosidic torsion angles Φ, Ψ, and ω (Fig. 1). The torsion angles of all glycan linkages were extracted using the Wordom program[95] according to the atoms defined in the system setup: Φ = C1–C2–O3–C3 and Ψ = C2–O3–C3–H3 for LSTa and Φ = C1–C2–O6–C6, Ψ = C2–O6–C6–C5, and ω = O6–C6–C5–H5 for LSTc. The ω angle was defined in accordance with the International Union of Pure and Applied Chemistry[69] and NMR definition, which was used to create the GlycoMaps torsion database§. The range of −120° to 240° was selected for better illustration of the torsion angles due to their cyclic nature and direct comparison with previous studies.[19] The glycan cluster representative torsion angles were extracted, and the average and standard deviation values of Φ, Ψ torsion angles were also computed by regions, which were separated by nonpopulated zones on the Φ, Ψ plots.

Topological θ angle

The θ angle, defined as the angle between the C2 atom of Sia1 and the C1 atoms of Gal2 and GlcNAc3[19] (Fig. 1), was extracted from each glycan trajectory using the Wordom program.[95]

Glycan volumetric topology

Three bound glycan–HA complex monomeric trajectories were extracted from each of the solvated glycan–HA trimeric systems by removing solvent and ions. The monomeric trajectories of the same HA were then concatenated into a single trajectory. Using the Wordom program, we performed RMSD alignment on HA RBD residues 117–265 (H3 numbering) as defined by Ha *et al.*[34] The frames of the resulting RBD-aligned glycan–HA complex trajectories that belong to the dominant glycan cluster members were extracted. The volumetric maps were constructed by averaging atomic density of the extracted MD frames using the VMD Volmap Tool plug-in with 1-Å resolution grids.

Hydrogen-bond analysis

Intramolecular hydrogen bonding was measured using a custom VMD script. A distance of 2.5 Å and angle cutoffs of 120°–180° between the hydrogen atom and the acceptor/donor heavy atoms were used as the hydrogen-bonding criteria.

Glycan–HA contact analysis

The glycan–HA complex structures that contain the glycan cluster representatives were extracted from the trajectories, and contact analysis was carried out using AutoDockTools 1.51.[99] The close contact of pairs of protein–ligand atoms was detected when their distances were closer than the sum of the vdW radii. Hydrogen-bond interactions were detected using the distance and angle criteria from the Atlas of Side-Chain and Main-Chain Hydrogen Bonding‖.

Glycan conformational entropy

Adopting a quasi-harmonic approximation approach[100] that has been successfully employed in previous glycan–protein binding studies,[68,69] we determined the relative glycan conformational entropy between the bound and free states by

$$\Delta S_{\mathrm{Conf}} = \frac{R}{2}\ln\left[\frac{\det(\sigma_{\mathrm{bound}})}{\det(\sigma_{\mathrm{unbound}})}\right],$$

where R is the gas constant, σ is the covariance matrix evaluated over a given trajectory, and det(σ) is the determinant of the covariance matrix. The quasi-harmonic approximation algorithm was implemented in MatLab and carried out on the covariance matrices of the torsion angles of all glycosidic linkages extracted from the free and bound glycan trajectories. Individual linkage contributions to the conformational entropy change, such as α-2,3 and α-2,6 linkages, were obtained by partitioning the covariance matrix into selected torsion blocks and applying the above equation to the determinant of each block.

MM-GBSA binding free energy

Using equidistant snapshots extracted from the MD trajectories, we computed the total binding free energy using the MM-GBSA scheme.[101,102] For computational efficiency, a single trajectory approach was used and the calculations were performed on ~1000 snapshots with a sampling interval of 40 ps.

$$G = E_{\mathrm{MM}} + \Delta G_{\mathrm{Solvation}} - T\Delta S \text{ and} \quad (1)$$

$$\Delta G_{\mathrm{binding}} = G_{\mathrm{complex}} - \left(G_{\mathrm{receptor}} + G_{\mathrm{ligand}}\right), \quad (2)$$

where E_{MM} represents the sum of electrostatic, vdW, and internal energies. $\Delta G_{\mathrm{Solvation}}$ is the desolvation free-energy penalty, estimated from GB and solvent-accessible surface area calculations, which yield G_{polar} and G_{nonpolar}. A surface tension coefficient (γ) of 0.0072 kcal/(mol Å) is used to calculate the nonpolar solvation free-energy contribution. The Hawkins–Cramer–Truhlar pairwise GB model[103,104] ($\mathrm{GB}^{\mathrm{HCT}}$) was used, with the parameters described by Tsui and Case.[105] A total of 0.15 M salt concentration and 310 K temperature were adopted, consistent with the explicit solvent MD simulations. $T\Delta S$ is the product of temperature and solute entropy, which includes the solute translational, rotational, and vibrational entropy derived from normal-mode analyses of the solute coordinates after energy minimization of solute structures to within an RMSD of 1.0^{-5} kcal/mol. Due to its prohibitive computational cost on large protein–ligand systems, normal-mode analysis was carried out on snapshots at mid- and end-points of the sampling. To account for glycan distortion and loss of internal rotational degrees

§ http://www.glycosciences.de/tools/glytorsion/

‖ http://www.biochem.ucl.ac.uk/bsm/sidechains/index.html

of freedom upon binding, we augmented $T\Delta S$ with the conformational entropy $-T\Delta S_{\mathrm{Conf}}$ to compute the overall binding free energy in Eq. (2).

Acknowledgements

D.X., E.I.N., L.S.C., H.C.P., W.W.L., and P.W.A. were supported by the National Institutes of Health (NIH) [P41 RR08605 (to P.W.A.) and TATRC W81XWH-07-2-0014 (to W.W.L. and P.W.A.)]. L.S.C. was also supported by the National Science Foundation [INT 0407508 (to the PRIME program)]. R.E.A. was funded in part by the NIH (GM077729 and MRAC CHE060073N). This work was also funded by the NIH through grant GM31749 and the National Science Foundation through grants 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. Barry Grant from the University of California San Diego Department of Chemistry and Biochemistry and Dr. Axel Kohlmeyer from the University of Pennsylvania Center for Molecular Modeling for helpful discussions; Dr. S. J. Gamblin from MRC, UK, for providing the HA X-31 crystal structure; and Dr. J. C. Gumbart, Dr. L. Trabuco, and Dr. E. Villa from the University of Illinois at Urbana–Champaign for their help with the hydrogen-bond analysis with the VMD script. We also thank the reviewers of this manuscript for their careful review and very valuable technical suggestions.

Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2009.01.040

References

1. Tiensin, T., Chaitaweesub, P., Songserm, T., Chaisingh, A., Hoonsuwan, W., Buranathai, C. *et al.* (2005). Highly pathogenic avian influenza H5N1, Thailand, 2004. *Emerging Infect. Dis.* **11**, 1664–1672.
2. Greger, M. (2007). The human/animal interface: emergence and resurgence of zoonotic infectious diseases. *Crit. Rev. Microbiol.* **33**, 243–299.
3. Lee, Y. J., Choi, Y. K., Kim, Y. J., Song, M. S., Jeong, O. M., Lee, E. K. *et al.* (2008). Highly pathogenic avian influenza virus (H5N1) in domestic poultry and relationship with migratory birds, South Korea. *Emerging Infect. Dis.* **14**, 487–490.
4. Webster, R. G., Bean, W. J., Gorman, O. T., Chambers, T. M. & Kawaoka, Y. (1992). Evolution and ecology of influenza A viruses. *Microbiol. Rev.* **56**, 152–179.
5. Baum, L. G. & Paulson, J. C. (1990). Sialyloligosaccharides of the respiratory epithelium in the selection of human influenza virus receptor specificity. *Acta Histochem. Suppl.* **40**, 35–38.
6. Couceiro, J. N., Paulson, J. C. & Baum, L. G. (1993). Influenza virus strains selectively recognize sialyloligosaccharides on human respiratory epithelium; the role of the host cell in selection of hemagglutinin receptor specificity. *Virus Res.* **29**, 155–165.
7. Rogers, G. N. & Paulson, J. C. (1983). Receptor determinants of human and animal influenza virus isolates: differences in receptor specificity of the H3 hemagglutinin based on species of origin. *Virology*, **127**, 361–373.
8. Russell, C. J. & Webster, R. G. (2005). The genesis of a pandemic influenza virus. *Cell*, **123**, 368–371.
9. Claas, E. C., Osterhaus, A. D., van Beek, R., De Jong, J. C., Rimmelzwaan, G. F., Senne, D. A. *et al.* (1998). Human influenza A H5N1 virus related to a highly pathogenic avian influenza virus. *Lancet*, **351**, 472–477.
10. Subbarao, K., Klimov, A., Katz, J., Regnery, H., Lim, W., Hall, H. *et al.* (1998). Characterization of an avian influenza A (H5N1) virus isolated from a child with a fatal respiratory illness. *Science*, **279**, 393–396.
11. Normile, D. (2008). Avian influenza. Flu virus research yields results but no magic bullet for pandemic. *Science*, **319**, 1178–1179.
12. WHO (2008). Communicable disease surveillance and response.
13. Centers for Disease Control and Prevention. (2007). The influenza (flu) viruses. In *Influenza (Flu)* (Diseases, C. C. f. I., ed.), *Influenza (Flu)*, vol. 2007, pp. Centers for Disease Control and Prevention, Atlanta, GA. http://www.cdc.gov/flu/about/viruses/index.htm.
14. WHO (2003). Influenza A (H5N1) in Hong Kong Special Administrative Region of China.
15. Neumann, G. & Kawaoka, Y. (2006). Host range restriction and pathogenicity in the context of influenza pandemic. *Emerging Infect. Dis.* **12**, 881–896.
16. Stevens, J., Blixt, O., Tumpey, T. M., Taubenberger, J. K., Paulson, J. C. & Wilson, I. A. (2006). Structure and receptor specificity of the hemagglutinin from an H5N1 influenza virus. *Science*, **312**, 404–410.
17. Wiley, D. C. & Skehel, J. J. (1987). The structure and function of the hemagglutinin membrane glycoprotein of influenza virus. *Annu. Rev. Biochem.* **56**, 365–394.
18. Skehel, J. J. & Wiley, D. C. (2000). Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. *Annu. Rev. Biochem.* **69**, 531–569.
19. Chandrasekaran, A., Srinivasan, A., Raman, R., Viswanathan, K., Raguram, S., Tumpey, T. M. *et al.* (2008). Glycan topology determines human adaptation of avian H5N1 virus hemagglutinin. *Nat. Biotechnol.* **26**, 107–113.
20. Shinya, K., Ebina, M., Yamada, S., Ono, M., Kasai, N. & Kawaoka, Y. (2006). Avian flu: influenza virus receptors in the human airway. *Nature*, **440**, 435–436.
21. van Riel, D., Munster, V. J., de Wit, E., Rimmelzwaan, G. F., Fouchier, R. A., Osterhaus, A. D. & Kuiken, T. (2006). H5N1 virus attachment to lower respiratory tract. *Science*, **312**, 399.
22. van Riel, D., Munster, V. J., de Wit, E., Rimmelzwaan, G. F., Fouchier, R. A., Osterhaus, A. D. & Kuiken, T. (2007). Human and avian influenza viruses target different cells in the lower respiratory tract of humans and other mammals. *Am. J. Pathol.* **171**, 1215–1223.

23. Beigel, J. H., Farrar, J., Han, A. M., Hayden, F. G., Hyer, R., de Jong, M. D. *et al.* (2005). Avian influenza A (H5N1) infection in humans. *N. Engl. J. Med.* **353**, 1374–1385.
24. Connor, R. J., Kawaoka, Y., Webster, R. G. & Paulson, J. C. (1994). Receptor specificity in human, avian, and equine H2 and H3 influenza virus isolates. *Virology*, **205**, 17–23.
25. Matrosovich, M., Tuzikov, A., Bovin, N., Gambaryan, A., Klimov, A., Castrucci, M. R. *et al.* (2000). Early alterations of the receptor-binding properties of H1, H2, and H3 avian influenza virus hemagglutinins after their introduction into mammals. *J. Virol.* **74**, 8502–8512.
26. Suzuki, Y. (2005). Sialobiology of influenza: molecular mechanism of host range variation of influenza viruses. *Biol. Pharm. Bull.* **28**, 399–408.
27. Stevens, J., Blixt, O., Glaser, L., Taubenberger, J. K., Palese, P., Paulson, J. C. & 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.
28. Matrosovich, M. N., Matrosovich, T. Y., Gray, T., Roberts, N. A. & Klenk, H. D. (2004). Human and avian influenza viruses target different cell types in cultures of human airway epithelium. *Proc. Natl Acad. Sci. USA*, **101**, 4620–4624.
29. Nicholls, J. M., Chan, M. C., Chan, W. Y., Wong, H. K., Cheung, C. Y., Kwong, D. L. *et al.* (2007). Tropism of avian influenza A (H5N1) in the upper and lower respiratory tract. *Nat. Med.* **13**, 147–149.
30. Glaser, L., Conenello, G., Paulson, J. & Palese, P. (2007). Effective replication of human influenza viruses in mice lacking a major alpha2,6 sialyltransferase. *Virus Res.* **126**, 9–18.
31. Varki, A. (2007). Glycan-based interactions involving vertebrate sialic-acid-recognizing proteins. *Nature*, **446**, 1023–1029.
32. Varki, N. M. & Varki, A. (2007). Diversity in cell surface sialic acid presentations: implications for biology and disease. *Lab. Invest.* **87**, 851–857.
33. Collins, B. E. & Paulson, J. C. (2004). Cell surface biology mediated by low affinity multivalent protein–glycan interactions. *Curr. Opin. Chem. Biol.* **8**, 617–625.
34. Ha, Y., Stevens, D. J., Skehel, J. J. & Wiley, D. C. (2001). X-ray structures of H5 avian and H9 swine influenza virus hemagglutinins bound to avian and human receptor analogs. *Proc. Natl Acad. Sci. USA*, **98**, 11181–11186.
35. Gamblin, S. J., Haire, L. F., Russell, R. J., Stevens, D. J., Xiao, B., Ha, Y. *et al.* (2004). The structure and receptor binding properties of the 1918 influenza hemagglutinin. *Science*, **303**, 1838–1842.
36. Eisen, M. B., Sabesan, S., Skehel, J. J. & Wiley, D. C. (1997). Binding of the influenza A virus to cell-surface receptors: structures of five hemagglutinin–sialyloligosaccharide complexes determined by X-ray crystallography. *Virology*, **232**, 19–31.
37. Russell, R. J., Stevens, D. J., Haire, L. F., Gamblin, S. J. & Skehel, J. J. (2006). Avian and human receptor binding by hemagglutinins of influenza A viruses. *Glycoconj. J.* **23**, 85–92.
38. Stevens, J., Blixt, O., Paulson, J. C. & Wilson, I. A. (2006). Glycan microarray technologies: tools to survey host specificity of influenza viruses. *Nat. Rev. Microbiol.* **4**, 857–864.
39. Belser, J. A., Blixt, O., Chen, L. M., Pappas, C., Maines, T. R., Van Hoeven, N. *et al.* (2008). Contemporary North American influenza H7 viruses possess human receptor specificity: implications for virus transmissibility. *Proc. Natl Acad. Sci. USA*, **105**, 7558–7563.
40. Rogers, G. N., Paulson, J. C., Daniels, R. S., Skehel, J. J., Wilson, I. A. & Wiley, D. C. (1983). Single amino acid substitutions in influenza haemagglutinin change receptor binding specificity. *Nature*, **304**, 76–78.
41. Tumpey, T. M., Maines, T. R., Van Hoeven, N., Glaser, L., Solorzano, A., Pappas, C. *et al.* (2007). A two-amino acid change in the hemagglutinin of the 1918 influenza virus abolishes transmission. *Science*, **315**, 655–659.
42. Yang, Z. Y., Wei, C. J., Kong, W. P., Wu, L., Xu, L., Smith, D. F. & Nabel, G. J. (2007). Immunization by avian H5 influenza hemagglutinin mutants with altered receptor binding specificity. *Science*, **317**, 825–828.
43. Ha, Y., Stevens, D. J., Skehel, J. J. & Wiley, D. C. (2003). X-ray structure of the hemagglutinin of a potential H3 avian progenitor of the 1968 Hong Kong pandemic influenza virus. *Virology*, **309**, 209–218.
44. Kirschner, K. N., Yongye, A. B., Tschampel, S. M., Gonzalez-Outeirino, J., Daniels, C. R., Foley, B. L. & Woods, R. J. (2008). GLYCAM06: a generalizable biomolecular force field. Carbohydrates. *J. Comput. Chem.* **29**, 622–655.
45. Tessier, M. B., DeMarco, M. L., Yongye, A. B. & Woods, R. J. (2008). Extension of the GLYCAM06 biomolecular force field to lipids, lipid bilayers and glycolipids. *Mol. Simul.* **34**, 349–364.
46. Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A. & Simmerling, C. (2006). Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins*, **65**, 712–725.
47. Kuttel, M., Brady, J. W. & Naidoo, K. J. (2002). Carbohydrate solution simulations: producing a force field with experimentally consistent primary alcohol rotational frequencies and populations. *J. Comput. Chem.* **23**, 1236–1243.
48. Eklund, R. & Widmalm, G. (2003). Molecular dynamics simulations of an oligosaccharide using a force field modified for carbohydrates. *Carbohydr. Res.* **338**, 393–398.
49. Guvench, O., Greene, S. N., Kamath, G., Brady, J. W., Venable, R. M., Pastor, R. W. & Mackerell, A. D., Jr. (2008). Additive empirical force field for hexopyranose monosaccharides. *J. Comput. Chem.* **29**, 2543–2564.
50. Kamath, G., Guvench, O. & MacKerell, A. D. (2008). CHARMM additive all-atom force field for acyclic carbohydrates and inositol. *J. Chem. Theory Comput.* **4**, 1990.
51. MacKerell, A. D., Jr., Bashford, D., Bellott, M., Dunbrack, R. L., Jr., Evanseck, J. D., Field, M. J. *et al.* (1998). All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, **102**, 3586–3616.
52. Frank, M. & Lieth, C. W. (1997). Comparison of the conformational behavior of sialyllactose complexed with the two viral attachment proteins influenza A hemagglutinin and the murine polyomavirus. *J. Mol. Model.* **3**, 408–414.
53. Li, M. & Wang, B. (2006). Computational studies of H5N1 hemagglutinin binding with SA-alpha-2, 3-Gal and SA-alpha-2, 6-Gal. *Biochem. Biophys. Res. Commun.* **347**, 662–668.
54. Auewarakul, P., Suptawiwat, O., Kongchanagul, A., Sangma, C., Suzuki, Y., Ungchusak, K. *et al.* (2007).

An avian influenza H5N1 virus that binds to a human-type receptor. *J. Virol.* **81**, 9950–9955.
55. Sauter, N. K., Bednarski, M. D., Wurzburg, B. A., Hanson, J. E., Whitesides, G. M., Skehel, J. J. & Wiley, D. C. (1989). Hemagglutinins from two influenza virus variants bind to sialic acid derivatives with millimolar dissociation constants: a 500-MHz proton nuclear magnetic resonance study. *Biochemistry*, **28**, 8388–8396.
56. Mochalova, L., Gambaryan, A., Romanova, J., Tuzikov, A., Chinarev, A., Katinger, D. *et al.* (2003). Receptor-binding properties of modern human influenza viruses primarily isolated in Vero and MDCK cells and chicken embryonated eggs. *Virology*, **313**, 473–480.
57. Sauter, N. K., Hanson, J. E., Glick, G. D., Brown, J. H., Crowther, R. L., Park, S. J. *et al.* (1992). Binding of influenza virus hemagglutinin to analogs of its cell-surface receptor, sialic acid: analysis by proton nuclear magnetic resonance spectroscopy and X-ray crystallography. *Biochemistry*, **31**, 9609–9621.
58. Hanson, J. E., Sauter, N. K., Skehel, J. J. & Wiley, D. C. (1992). Proton nuclear magnetic resonance studies of the binding of sialosides to intact influenza virus. *Virology*, **189**, 525–533.
59. Glick, G. D., Toogood, P. L., Wiley, D. C., Skehel, J. J. & Knowles, J. R. (1991). Ligand recognition by influenza virus. The binding of bivalent sialosides. *Biochemistry*, **266**, 23660–23669.
60. Takemoto, D. K., Skehel, J. J. & Wiley, D. (1996). A surface plasmon resonance assay for the binding of influenza virus hemagglutinin to its sialic acid receptor. *Virology*, **217**, 452–458.
61. Critchley, P. & Dimmock, N. J. (2004). Binding of an influenza A virus to a neomembrane measured by surface plasmon resonance. *Bioorg. Med. Chem.* **12**, 2773–2780.
62. Hidari, K., Shimada, S., Suzuki, Y. & Suzuki, T. (2007). Binding kinetics of influenza viruses to sialic acid-containing carbohydrates. *Glycoconj. J.* **24**, 583–590.
63. Toone, E. J. (1994). Structure and energetics of protein–carbohydrate complexes. *Curr. Opin. Struct. Biol.* **4**, 719–728.
64. Ambrosi, M., Cameron, N. R. & Davis, B. G. (2005). Lectins: tools for the molecular understanding of the glycocode. *Org. Biomol. Chem.* **3**, 1593–1608.
65. Van Aalten, D. M. F., De Groot, B. L., Findlay, J. B. C., Berendsen, H. J. C. & Amadei, A. (1997). A comparison of techniques for calculating protein essential dynamics. *J. Comput. Chem.* **18**, 169–181.
66. Shao, J., Tanner, S. W., Thompson, N. & Cheatham, T. E. (2007). Clustering molecular dynamics trajectories: 1. Characterizing the performance of different clustering algorithms. *J. Chem. Theory Comput.* **3**, 2312–2334.
67. Frank, M., Lütteke, T. & von der Lieth, C. -W. (2007). GlycoMapsDB: a database of the accessible conformational space of glycosidic linkages. *Nucleic Acids Res.* **35**, 287–290.
68. Kadirvelraj, R., Gonzalez-Outeirino, J., Foley, B. L., Beckham, M. L., Jennings, H. J., Foote, S. *et al.* (2006). Understanding the bacterial polysaccharide antigenicity of *Streptococcus agalactiae versus Streptococcus pneumoniae*. *Proc. Natl Acad. Sci. USA*, **103**, 8149–8154.
69. Bryce, R. A., Hillier, I. H. & Naismith, J. H. (2001). Carbohydrate–protein recognition: molecular dynamics simulations and free energy analysis of oligosaccharide binding to concanavalin A. *Biophys. J.* **81**, 1373–1388.
70. DeMarco, M. L. & Woods, R. J. (2008). Structural glycobiology: a game of snakes and ladders. *Glycobiology*, **18**, 426–440.
71. Poppe, L., Stuike-Prill, R., Meyer, B. & van Halbeek, H. (1992). The solution conformation of sialyl-α(2–6)-lactose studied by modern NMR techniques and Monte Carlo simulations. *J. Biomol. NMR*, **2**, 109–136.
72. Neumann, G. & Kawaoka, Y. (2001). Reverse genetics of influenza virus. *Virology*, **287**, 243–250.
73. Yamada, S., Suzuki, Y., Suzuki, T., Le, M. Q., Nidom, C. A., Sakai-Tagawa, Y. *et al.* (2006). Haemagglutinin mutations responsible for the binding of H5N1 influenza A viruses to human-type receptors. *Nature*, **444**, 378–382.
74. World Health Organization Global Influenza Program Surveillance Network. (2005). Evolution of H5N1 avian influenza viruses in Asia. *Emerging Infect. Dis.* **11**, 1515–1521.
75. Chachra, R. & Rizzo, R. C. (2008). Origins of resistance conferred by the R292K neuraminidase mutation via molecular dynamics and free energy calculations. *J. Chem. Theory Comput.* **4**, 1526–1540.
76. Shtyrya, Y., Mochalova, L., Voznova, G., Rudneva, I., Shilov, A., Kaverin, N. & Bovin, N. (2009). Adjustment of receptor-binding and neuraminidase substrate specificities in avian–human reassortant influenza viruses. *Glycoconj. J.* **26**, 99–109.
77. Levy, R. M., Zhang, L. Y., Gallicchio, E. & Felts, A. K. (2003). On the nonpolar hydration free energy of proteins: surface area and continuum solvent models for the solute–solvent interaction energy. *J. Am. Chem. Soc.* **125**, 9523–9530.
78. Gilson, M. & Zhou, H. (2007). Calculation of protein–ligand binding affinities. *Annu. Rev. Biophys. Biomol. Struct.* **36**, 21.
79. Pearlman, D. (2005). Evaluating the molecular mechanics Poisson–Boltzmann surface area free energy method using a congeneric series of ligands to p38 MAP kinase. *J. Med. Chem.* **48**, 7796–7807.
80. Taubenberger, J. K., Reid, A. H. & Fanning, T. G. (2000). The 1918 influenza virus: a killer comes into view. *Virology*, **274**, 241–245.
81. Glaser, L., Stevens, J., Zamarin, D., Wilson, I. A., Garcia-Sastre, A., Tumpey, T. M. *et al.* (2005). A single amino acid substitution in 1918 influenza virus hemagglutinin changes receptor binding specificity. *J. Virol.* **79**, 11533–11536.
82. Russell, R. J., Gamblin, S. J., Haire, L. F., Stevens, D. J., Xiao, B., Ha, Y. & Skehel, J. J. (2004). H1 and H7 influenza haemagglutinin structures extend a structural classification of haemagglutinin subtypes. *Virology*, **325**, 287–296.
83. Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612.
84. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. A. (2004). PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. *Nucleic Acids Res.* **32**, W665–W667.
85. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. (1983). Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935.
86. Case, D. A., Cheatham, E., III, Darden, T., Gohlke, H., Luo, R., Merz, K. M., Jr. *et al.* (2005). The Amber

biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688.
87. Humphrey, W., Dalke, A. & Schulten, K. (1996). VMD: visual molecular dynamics. *J. Mol. Graphics*, **14**, 33–38.
88. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E. *et al.* (2005). Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802.
89. Grubmuller, H., Heller, H., Windemuth, A. & Schulten, K. (1991). Generalized Verlet algorithm for efficient molecular dynamics simulations with long-range interactions. *Mol. Sim.* **6**, 121–142.
90. Andersen, H. (1983). RATTLE: a "velocity" version of the SHAKE algorithm for molecular dynamics calculations. *J. Comp. Phys.* **52**, 24–34.
91. Scott, E. F., Yuhong, Z., Richard, W. P. & Bernard, R. B. (1995). Constant pressure molecular dynamics simulation: the Langevin piston method. *J. Chem. Phys.* **103**, 4613–4621.
92. Schlick, T., Skeel, R. D., Brunger, A. T., Kalé, L. V., Board, J. A., Jr., Hermans, J. & Schulten, K. (1999). Algorithmic challenges in computational molecular biophysics. *J. Comput. Phys.* **151**, 9–48.
93. Hans, C. A. (1980). Molecular dynamics simulations at constant pressure and/or temperature. *J. Chem. Phys.* **72**, 2384–2393.
94. Amadei, A., Linssen, A. B. M. & Berendsen, H. J. C. (1993). Essential dynamics of proteins. *Proteins*, **17**, 412–425.
95. Seeber, M., Cecchini, M., Rao, F., Settanni, G. & Caflisch, A. (2007). Wordom: a program for efficient analysis of molecular dynamics simulations. *Bioinformatics*, **23**, 2625.
96. Grant, B. J., Rodrigues, A. P., Elsawy, K. M., McCammon, J. A. & Caves, L. S. (2006). Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics*, **22**, 2695–2696.
97. Tracy, L. F., David, B., Patrick, O. B. & Rosenzweig, R. F. (1999). Systematic changes in gene expression patterns following adaptive evolution in yeast. *Proc. Natl Acad. Sci. USA*, **96**, 9721–9726.
98. Xu, D. & Redman-Furey, N. (2007). Statistical cluster analysis of pharmaceutical solvents. *Int. J. Pharm.* **339**, 175–188.
99. Sanner, M. F., Duncan, B. S., Carrillo, C. J. & Olson, A. J. (1999). Integrating computation and visualization for biomolecular analysis: an example using Python and AVS. Proceedings of the Pacific Symposium on Biocomputing, pp. 401–412.
100. Karplus, M. & Kushick, J. N. (1981). Method for estimating the configurational entropy of macromolecules. *Macromolecules*, **14**, 325–332.
101. Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L. *et al.* (2000). Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* **33**, 889–897.
102. Massova, I. & Kollman, P. (2000). Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspect. Drug Discov. Des.* **18**, 113–135.
103. Hawkins, G. D., Cramer, C. J. & Truhlar, D. G. (1995). Pairwise solute descreening of solute charges from a dielectric medium. *Chem. Phys. Lett.* **246**, 122–129.
104. Hawkins, G. D., Cramer, C. J. & Truhlar, D. G. (1996). Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J. Phys. Chem.* **100**, 19824–19839.
105. Vickie Tsui, D. A. C. (2000). Theory and applications of the generalized Born solvation model in macromolecular simulations. *Biopolymers*, **56**, 275–291.