

Glycan topology determines human adaptation of avian H5N1 virus hemagglutinin

Aarthi Chandrasekaran^{1,5}, Aravind Srinivasan^{1,5}, Rahul Raman^{1,5}, Karthik Viswanathan^{1,5}, S Raguram¹, Terrence M Tumpey², V Sasisekharan³ & Ram Sasisekharan^{1,3,4}

A switch in specificity of avian influenza A viruses' hemagglutinin (HA) from avian-like (α 2-3 sialylated glycans) to human-like (α 2-6 sialylated glycans) receptors is believed to be associated with their adaptation to infect humans¹⁻⁴. We show that a characteristic structural topology—and not the α 2-6 linkage itself—enables specific binding of HA to α 2-6 sialylated glycans and that recognition of this topology may be critical for adaptation of HA to bind glycans in the upper respiratory tract of humans. An integrated biochemical, analytical and data mining approach demonstrates that HAs from the human-adapted H1N1 and H3N2 viruses, but not H5N1 (bird flu) viruses, specifically bind to long α 2-6 sialylated glycans with this topology. This could explain why H5N1 viruses have not yet gained a foothold in the human population^{5,6}. Our findings will enable the development of additional strategies for effective surveillance and potential therapeutic interventions for H5N1 and possibly other influenza A viruses.

Until now, annual influenza outbreaks and epidemics have been caused by the human-adapted H1N1 and H3N2 viruses that circulate in human populations. However, the ability of H5N1, H7N7 and H9N2 avian virus subtypes to infect humans has raised concern over the possibility of a new flu pandemic equal in impact to the 1918 H1N1 Spanish influenza.

The key first step in the infection, transmission and virulence of these viruses is the binding of HA to sialylated glycans on the epithelial cell surface^{3,4,7-9}. Transmission from birds to humans is believed to be closely associated with the ability of the HA to switch its preference from α 2-3 sialylated glycans (α 2-3) to α 2-6 sialylated glycans (α 2-6), which are extensively expressed in the human upper respiratory epithelia^{1,8,10}. Crystal structures of HAs from H1 (human and swine), H3 (avian and human) and H5 (avian) and their complexes with α 2-3 and/or α 2-6 oligosaccharides have provided molecular insights into key residues involved in specific HA-glycan interactions^{2,3,11-15} and glycan arrays have been used to investigate the glycan binding specificity of wild-type and mutant H1, H3 and H5

HAs^{16,17}. The relationship between the HA glycan binding specificity and transmission efficiency has been demonstrated in ferrets, using the highly pathogenic and virulent 1918 H1N1 viruses¹⁸. The reliability of this animal model for studying human transmission is attributed to the predominance of human-like α 2-6 in the upper respiratory tract epithelium of ferrets^{6,9}. Switching the receptor binding specificity of the highly transmissible and pathogenic human H1N1 (A/South Carolina/1/18; SC18) virus from α 2-6 to α 2-3 has produced a virus (AV18) not transmissible in the ferrets¹⁸. Although A/New York/1/18 (NY18) H1N1 virus, which shows mixed α 2-3/ α 2-6 binding, does not transmit efficiently, the A/Texas/36/91 (Tx91) H1N1 strain—that also binds to both α 2-3 and α 2-6—transmits efficiently¹⁸.

These confounding results with respect to the correlation between glycan binding specificity of HA and transmissibility of the viruses led us to pose several questions. First, how diverse are the sialylated glycan receptors in the human upper respiratory tissues and could this diversity account for the specificity in the tissue tropism of the virus? Second, are there nuances of glycan conformation that might play a role in α 2-3 and/or α 2-6 binding specificity of HA? Taken together, what are the glycan binding requirements of HA, beyond binding to sialylated glycans with a specific linkage, necessary for human adaptation? Currently, avian H5N1 viruses have α 2-3 glycan specificity and also show some α 2-6 binding^{19,20}, but have not yet transmitted between humans. Answers to these questions would aid the understanding of human adaptation of these viruses and facilitate effective surveillance of the evolution of H5N1 into a potentially pandemic human virus. To address these issues, we developed a framework comprising four complementary analyses to investigate binding of human-adapted HA to α 2-3 and α 2-6. Integration of these complementary analyses led to the identification of the necessary glycan determinant for human adaptation of HA.

The first of the four complementary analyses aimed to delineate the diversity of sialylated glycans in the human upper respiratory tissues to answer, in part, a long-standing question on the different types of α 2-6 glycans expressed in these tissues. To elaborate the diversity of α 2-6, which predominate in the upper respiratory tissues⁸, human tracheal tissue sections were costained with Concanavalin/

¹Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ²Influenza Division, Mailstop G-16, National Center for Immunization and Respiratory Diseases, Coordinating Center for Infectious Diseases, Centers for Disease Control and Prevention, 1600 Clifton Road, NE, Atlanta, Georgia 30333, USA. ³Harvard-MIT Division of Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ⁴Center for Cancer Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ⁵These authors contributed equally to this work. Correspondence should be addressed to R.S. (rams@mit.edu).

Received 18 September 2007; accepted 3 December 2007; published online 6 January 2008; doi:10.1038/nbt1375

(ConA)/Jacalin and SNA-I/Jacalin (**Fig. 1a**). The glycan binding specificity of these lectins has been previously characterized using glycan arrays. As SNA-I binds to diverse α 2-6, its binding pattern indicated the distribution of α 2-6. Jacalin, which specifically binds to -Gal β 1-3GalNAc α - and -GlcNAc β 1-3GalNAc α - motifs, characteristic of O-linked glycans, predominantly stained goblet cells. On the other hand, ConA, which specifically binds to mannose commonly found in high mannose, hybrid and complex type N-linked glycans, predominantly stained ciliated cells. The lectin binding patterns in **Figure 1a** demonstrated a wide distribution of N-linked α 2-6, compared with the more localized distribution of O-linked α 2-6 on the apical side of the human tracheal epithelium. Matrix-assisted laser desorption/ionization– mass spectrometry (MALDI-MS) profiling of N-linked sialylated glycans of a representative upper respiratory epithelial cell line was performed using techniques that were specifically modified to

better suit the analysis of these complex acidic glycans. The results showed a substantial diversity (**Fig. 1b**), as well as predominant expression of α 2-6 (**Supplementary Fig. 1** online), in the human upper respiratory epithelium. Furthermore, desialylation and fragmentation of representative mass peaks using MALDI tandem time of flight (TOF-TOF) (**Fig. 1c,d**) demonstrated the presence of long oligosaccharide branches with multiple lactosamine repeats.

The observation of extensive diversity of the glycan structures in the human upper respiratory tissues prompted the second complementary analysis, involving examination of the conformational features of these diverse sialylated glycans and their role in HA-glycan interactions. Analyses of all the HA-glycan cocrystal structures indicated that the orientation of the Neu5Ac sugar is fixed relative to the HA glycan binding site. A highly conserved set of amino acids Y98, S/T136, W153, H183, L/I194 (numbered based on H3 HA) across different HA

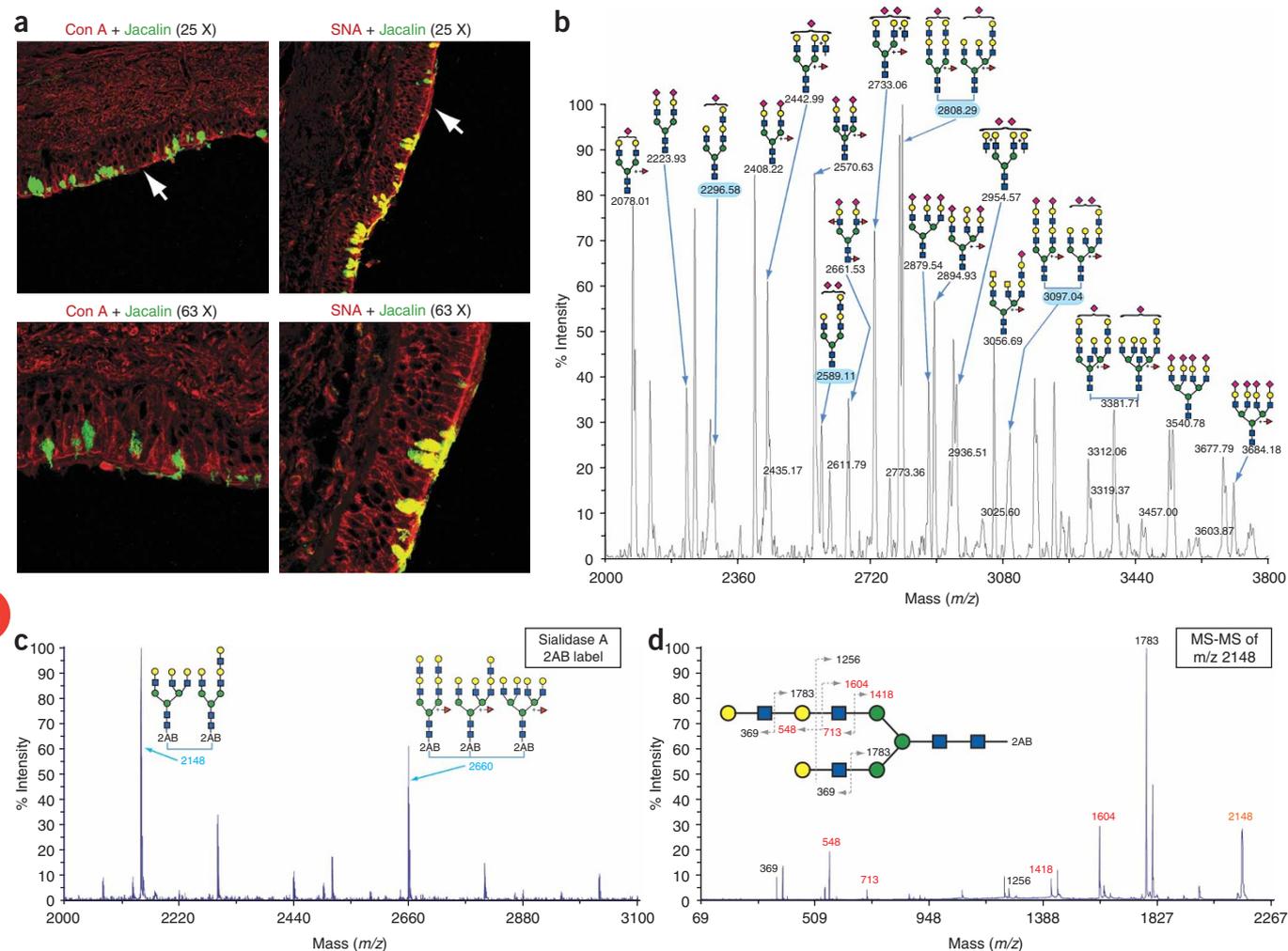
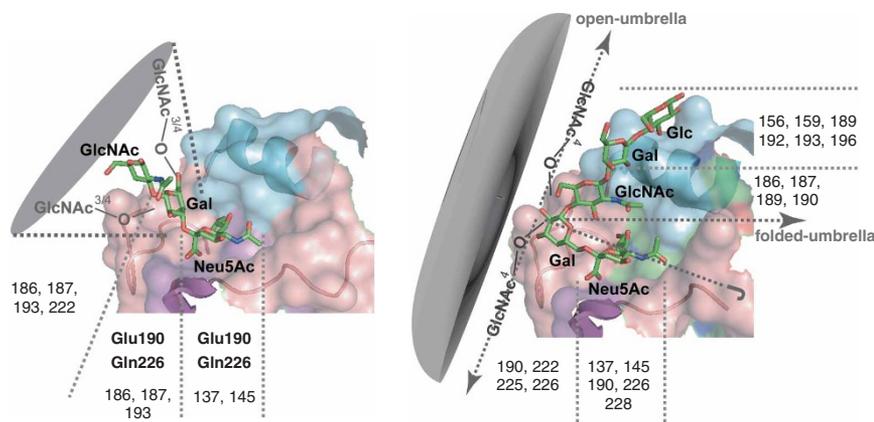


Figure 1 Glycan diversity in human upper respiratory tissues. **(a)** Costaining of tracheal tissue sections with ConA (red)/Jacalin (green) and SNA-I (red)/Jacalin (green). The localized regions of Jacalin binding correspond to goblet cells expressing O-linked glycans and the regions of ConA binding correspond to ciliated cells expressing N-linked glycans (white arrow) on the apical side of the tracheal epithelium. The extensive binding of SNA-I to both goblet cells (costain with Jacalin in yellow) and ciliated cells indicates predominant expression of O-linked and N-linked α 2-6 on the apical side. **(b)** MALDI-MS glycan profile of human bronchial epithelial (HBE) cells using graphical representation (without explicit linkage assignment) of possible sialylated glycan structures that satisfy the mass peaks (within ± 3.5 Daltons). HBEs predominantly express α 2-6 (in comparison with α 2-3) sialylated glycans (**Supplementary Fig. 1**). **(c)** Desialylation using Sialidase A and subsequent 2-AB labeling of the N-linked glycans observed in **b** to deconvolute the branching pattern from the number of sialic acids. The peaks highlighted in cyan in **b** and **c** were further analyzed using TOF-TOF MS. **(d)** The MS-MS profile of a representative peak at m/z 2148 shows critical fragment ions at m/z 548 and 713 and their corresponding counter ions (shown in red) that support the long oligosaccharide branch (with multiple lactosamine repeats) over multiple short lactosamine branches. MS-MS profile of m/z 2660 also supports a long oligosaccharide branch (data not shown). Glycans are represented using the graphical nomenclature adopted by the Consortium for Functional Glycomics (CFG).

Figure 2 Cone-like (left) and umbrella-like (right) topologies of α 2-3 and α 2-6 sialylated glycans.

The topology of α 2-3 and α 2-6 is governed by the glycosidic torsion angles of the trisaccharide motifs—Neu5Ac α 2-3Gal β 1-3/4GlcNAc and Neu5Ac α 2-6Gal β 1-4GlcNAc, respectively (Supplementary Fig. 3 online). A parameter (θ)—the angle between the C2 atom of Neu5Ac and C1 atoms of the subsequent Gal and GlcNAc sugars in these trisaccharide motifs—characterizes the topology. Superimposition of the θ contour and the conformational maps of the α 2-3 and α 2-6 motifs showed that the α 2-3 motif adopts 100% cone-like topology and α 2-6 motif sampled both the cone-like and umbrella-like topologies (Supplementary Fig. 3). In the cone-like topology sampled by α 2-3 and α 2-6, the GlcNAc and subsequent sugars are

positioned along a region spanning a cone. The interactions of HA with the cone-like topology primarily involves contacts of amino acids at the numbered positions (based on H3 HA numbering) with Neu5Ac and Gal sugars. On the other hand, in the umbrella-like topology, which is unique to α 2-6, the GlcNAc and subsequent sugars bend toward the HA binding site (as observed in HA- α 2-6 cocrystal structures). Longer α 2-6 oligosaccharides (at least a tetrasaccharide) would favor this conformation as it is stabilized by intra-sugar van der Waals contact between acetyl groups of GlcNAc and Neu5Ac. HA interactions with sialylated glycans with the umbrella-like topology involve contacts of amino acids at the numbered positions (based on H3 HA numbering) with GlcNAc and subsequent sugars in addition to contacts with Neu5Ac and Gal sugars.



subtypes are involved in anchoring the Neu5Ac sugar. The specificity of HA to α 2-3 or α 2-6 is therefore governed by interactions of the HA glycan binding site with the glycosidic oxygen atom and sugars besides Neu5Ac.

The conformation of the Neu5Ac α 2-3Gal linkage is such that the Gal and sugars beyond Gal (at the reducing end, and in both linear and branched α 2-3) occupy a cone-like region of space. Hence, we define it as a cone-like glycan topology (Fig. 2). In the cone-like topology, HA interactions with the glycans primarily involve contacts with Neu5Ac and Gal sugars in a three-sugar (or trisaccharide) α 2-3 motif (Neu5Ac α 2-3Gal β 1-3/4GlcNAc-) and this observation is corroborated by the co-crystal structures^{2,3,12,13}. In addition to the two critical amino acids, E190 and Q226 (Fig. 2), the contacts with the α 2-3 motif and substitutions on this motif, such as sulfation and fucosylation, appear to involve key amino acid positions (Fig. 2). The variability of the amino acids in these positions potentially accounts for the differential binding specificity of HA to the diverse α 2-3 sialylated glycans (see below). Compared with the Neu5Ac α 2-3Gal linkage, the presence of the C6-C5 bond provides additional glycan conformational flexibility to the Neu5Ac α 2-6Gal linkage. This enables the α 2-6 motif to adopt a cone-like topology, as well as the potential to span a wider region of space analogous to the opening of an umbrella. We thus define the latter conformation—which spans a wider region on the HA surface than the cone-like topology—as an umbrella-like glycan topology (Fig. 2). In contrast to the cone-like topology, the length of the oligosaccharide and its degree of branching beyond a trisaccharide critically influence HA binding contacts in the umbrella-like topology. Therefore, the amino acids involved in interactions with the umbrella-like topology (key numbered positions shown in Fig. 2) are not conserved across the human-adapted H1 and H3 HAs. Depending upon the HA subtype, a combination of amino acids at these positions are involved in interacting with a long α 2-6. This observation is corroborated by the cocrystal structures of H1 and H3 with α 2-6 oligosaccharides^{3,11,14}. Therefore, whereas the cone-like topology is characteristic of α 2-3 as well as short α 2-6 glycans such as single lactosamine branches, the umbrella-like topology is unique to α 2-6 (Fig. 2) and is typically adopted by long glycans with multiple repeating lactosamine units.

Interrogation of the HA-glycan cocrystal structures highlights the fact that the human-adapted H1 and H3 HAs have mutated from their presumed avian counterparts to gain additional contacts with α 2-6 in the umbrella-like topology. Defining HA-glycan interactions based on *trans* and *cis* conformations (adopted by α 2-3 and α 2-6 linkages, respectively) is inadequate as this does not fully capture the structural features and conformational flexibility of the diverse sialylated glycans observed in human tissues. In contrast, the cone-like and umbrella-like classifications for HA binding represent the full extent of the structural diversity and conformational plurality of sialylated glycans and are able to distinguish the α 2-3 and α 2-6 binding of avian HAs from that of the α 2-6 binding of human-adapted H1 and H3 HAs.

The requirements for HA binding to the cone-like and umbrella-like topologies were further corroborated using the third complementary data mining approach, involving analysis of the extensive glycan array data available for H1, H3 and H5 HAs^{2,16}. The data mining analysis provided correlations between glycan features (Fig. 3a) abstracted from the glycans on the array and the HA binding to these glycans. These correlations are given as rules or classifiers (Fig. 3b) that verify the above structural constraints. Consistent with these observations, the distinct α 2-3 classifiers (Fig. 3b) indicated that variations around the trisaccharide α 2-3 motif primarily influence the differential α 2-3 binding of H1, H3 and H5 HAs. On the other hand, the length-dependent α 2-6 classifiers (Fig. 3b) support the critical role of oligosaccharide length in HA binding to α 2-6 in the umbrella-like topology. The α 2-6 classifier common to the human-adapted H1 and H3 HAs is consistent with its gain in ability to bind long α 2-6. Although the glycan binding of wild-type and mutant H5N1 HAs is not supported by the long α 2-6 classifier, it is consistent with both α 2-3 and short α 2-6 classifiers (Fig. 3b).

The final complementary analysis involved corroborating the above findings by establishing the tissue tropism and binding specificity of human-adapted HA to upper respiratory tissue glycans through investigating human tissue binding and direct glycan binding of recombinant HAs. The HAs from the hallmark pandemic SC18 and the human vaccine strain A/Moscow/10/99 H3N2 (Mos99) were chosen as representative human-adapted H1 and H3 candidates. SC18 and Mos99 HAs showed distinct binding patterns to human

upper respiratory (tracheal) and deep lung (alveolar) tissues (Fig. 4a). More importantly, both H1 and H3 HA showed substantial binding to the apical side of tracheal tissue sections. As noted earlier (Fig. 1), long-branch α 2-6 are predominantly expressed on the apical side of the upper respiratory epithelia. The α 2-6 binding specificity of the human-adapted HAs was verified using dose-dependent direct binding of HA to defined α 2-3 and α 2-6 oligosaccharides (Fig. 4a). The binding of SC18 and Mos99 HAs over a range of concentrations to 6'SLN-LN indicated their high-affinity binding to the long α 2-6 oligosaccharide. In addition to the long α 2-6, Mos99 also bound with high affinity to the short α 2-6 oligosaccharide (6'SLN) and at a relatively lower affinity to the long α 2-3 oligosaccharide (3'SLN-LN-LN).

Taken together, the above findings show that the human-adapted HAs bind specifically to the long α 2-6 (in the umbrella-like topology), which are predominantly expressed in the human upper respiratory tissues. The α 2-3 binding of Mos99 further suggests that a switch in the glycan binding preference (that is, loss of ability to bind α 2-3) is not a necessary determinant for human adaptation of HA. Tx91, also a mixed α 2-6/ α 2-3 binding virus that shows HA binding to long α 2-6 (Fig. 3b), is capable of efficient transmission¹⁸. On the other hand, NY18, another mixed α 2-3/ α 2-6 binding virus that does not have HA binding specificity to long α 2-6 (Fig. 3b), is not transmitted efficiently¹⁸. The efficient human adaptation of these viruses is, therefore, correlated with HA binding to sialylated glycans of a

characteristic umbrella-like topology, going beyond the specific α 2-3 or α 2-6 linkage.

Understanding the glycan binding specificity of influenza A virus HA is critical for surveillance of the evolution of highly pathogenic strains, such as H5N1, which threaten to gain a foothold in the human population. Various strains of the highly pathogenic α 2-3-specific H5N1 viruses show mixed α 2-3/ α 2-6 binding^{5,20} and yet these viruses have shown inefficient transmission in ferrets⁵. However, dose-dependent direct binding studies of representative H5N1 viruses—A/Vietnam/1203/04 (Viet0304) and A/Hong Kong/486/97 (HK486)—showed minimal binding affinity to long α 2-6 (Fig. 4b). The observed initiation of H5N1 infection in tissues predominantly

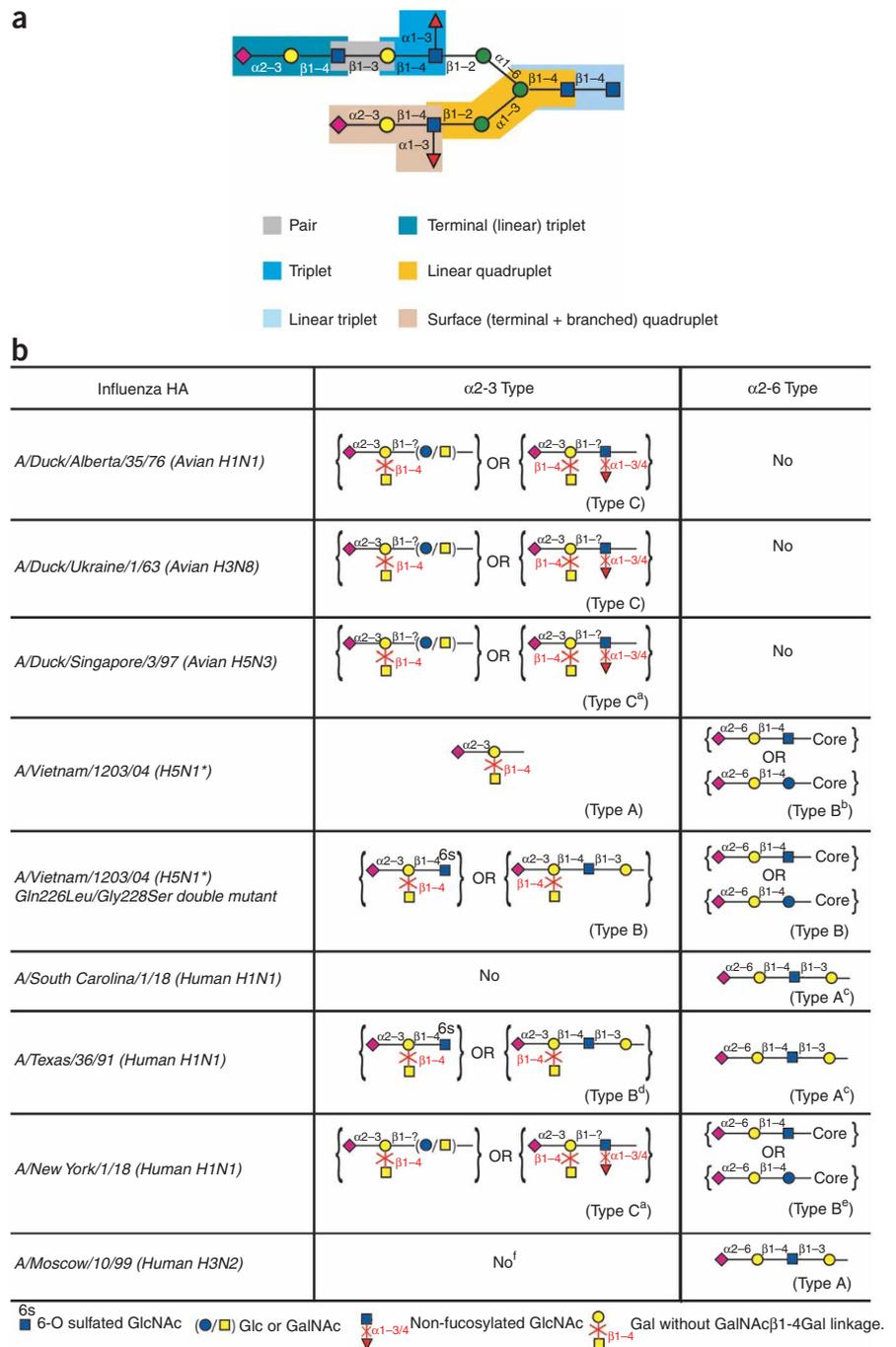


Figure 3 Data mining analysis of HA binding glycan array data. **(a)** Examples of the types of glycan features (e.g., pairs, triplets and quadruplets) abstracted from a representative complex glycan structure. A comprehensive set of these features was abstracted from the glycans in the glycan array. **(b)** Graphical representations of the complex classifier rules (Supplementary Table 1) for each HA analyzed using the glycan array. The α 2-3 Type A represents broadest specificity, whereas the Type B and Type C classifiers represent constraints imposed by structural variations around the trisaccharide α 2-3 motif. The α 2-6 Type A represents binding to long α 2-6, whereas Type B represents binding to short α 2-6 (linear or branched). The core corresponds to either the spacer attached to the reducing end or the trimannosyl core in case of the single α 2-6 biantennary glycan on the array. ^aBinding signals observed for fucosylated α 2-3 motif only if it has GlcNAc[6S]. ^bBinding signals observed only for 6'-sialyl lactose. ^cBinding signals also observed for short 6'-sialyl lactosamine (Type B). ^dBinding signals are significantly lower than α 2-3 Type B of H5N1 double mutant. ^eBinding signals observed only for short α 2-6 with GlcNAc[6S]. ^fBinding signals just above background observed for α 2-3 motif with GlcNAc[6S]. *The origin of A/Vietnam/1203/04 is avian but this viral strain was isolated from an infected human.

expressing α 2-6 (ref. 21) could possibly be explained by the α 2-6 binding at high viral loads. Although additional factors such as the neuraminidase and other viral gene products may play a role in H5N1 viral transmission, a necessary condition for the human adaptation of its HA is to acquire mutations that would provide binding specificity to long α 2-6. The glycan topology could also play an important role in the balance between the α 2-6 binding of HA and cleavage properties of neuraminidase^{22,23}. Thus, the identification of the α 2-6 sialylated glycans present on lung tissue can now be used to analyze any correlation between glycan topology and neuraminidase stalk length in virulence during human-to-human transmission.

Three related issues have confounded the interpretation of glycan receptor specificity leading to human adaptation of HA. The first arises from the association of 'human-like receptors' exclusively with the Neu5Ac α 2-6Gal linkage, as evidenced by using short α 2-6

oligosaccharides 6'SL and 6'SLN to define glycan binding specificity of the H5N1 viruses^{5,20}. These and other earlier studies have predominantly focused on α 2-3 and α 2-6 linkages, as well as a switch in the linkage preference. However, as both α 2-3 (short, long, linear and branched) and short α 2-6 adopt the cone-like topology as opposed to long α 2-6, (which adopts the characteristic umbrella-like topology), glycan topology—and not just linkage *per se*—is the key determinant for human adaptation of HAs. The second issue deals with the binding specificity of HA-glycan interactions. Many HA glycan array experiments¹⁷ have been performed at a relatively high HA concentration, without a dose-dependent relationship to establish specificity. Our dose-dependent binding studies enabled investigation of the HA-glycan binding affinities and thereby demonstrate the specificity of human-adapted HAs for binding long α 2-6 (Fig. 4a). These studies can be further extended to derive parameters such as binding affinity

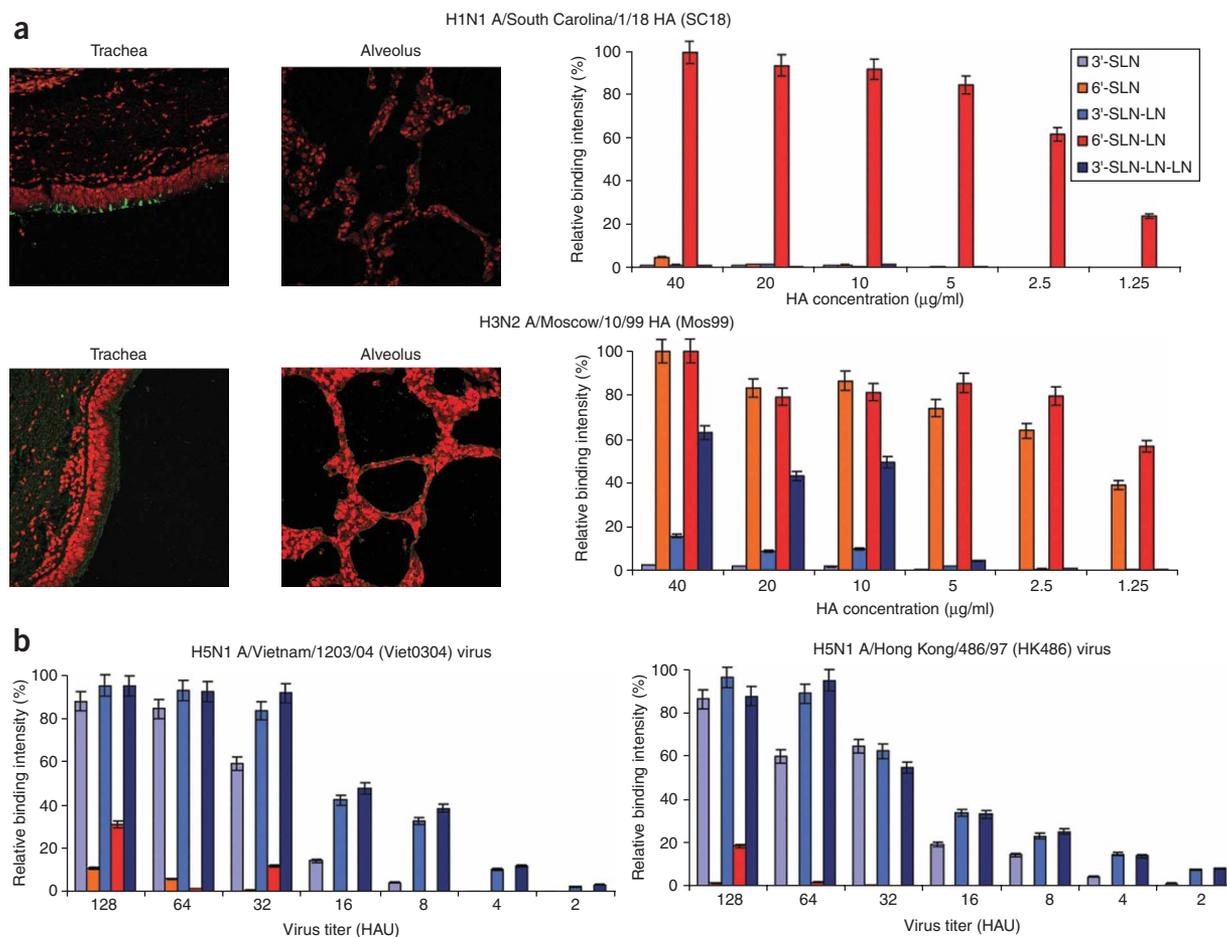


Figure 4 Glycan binding specificity of SC18 and Mos99 HAs, as well as H5N1 viruses. (a) Both SC18 and Mos99 HAs show substantial and preferential binding to the apical side of the tracheal tissue (green against as against propidium iodide staining in red) compared with the alveolar tissue sections, although binding of SC18 is more restricted than that of Mos99. The sialic acid-specific binding of HA is demonstrated by a substantial reduction in binding after pretreatment of tissues with Sialidase A (Supplementary Fig. 4 online). Moreover, competition experiments involving SC18 HA and SNA-I showed dramatic reductions in SC18 HA binding to tracheal sections (data not shown). Further AV18 HA does not bind to tracheal sections (data not shown). The binding specificity of SC18 and Mos99 HAs was demonstrated with dose-dependent direct binding to defined α 2-3 and α 2-6 oligosaccharides (shown on the right). The characteristic binding pattern of SC18 and Mos99 HAs is their binding at saturating levels to the long α 2-6 (6'SLN-LN) over a range of HA dilution from 40 to 5 μ g/ml. The narrow specificity of SC18 HA correlates with its restrictive tracheal tissue binding. On the other hand, the ability of Mos99 HA to bind to diverse sialylated glycans is consistent with its more extensive binding to tracheal tissue sections (as compared with SC18). The high affinity binding of SC18 and Mos99 HAs to long α 2-6 was confirmed using another human-adapted H3N2 (A/Wyoming/3/03) HA (Supplementary Fig. 5 and Supplementary Methods online). (b) In contrast to the dose-dependent binding profile of the human-adapted SC18 and Mos99 HAs, Viet0304 and HK486 H5N1 viruses bind with high affinity to the α 2-3 and minimal affinity to the long α 2-6 oligosaccharide. The dose-dependent direct binding of the whole virus corroborates with the binding of soluble HA protein and the deep lung tissue tropism of the H5N1 HA is consistent with its α 2-3 binding specificity (Supplementary Fig. 6 and Supplementary Methods online). HAU, Virus titer in Hemagglutinating Unit.

constants to quantify the relative glycan binding affinities of different HAs. The two representative H5N1 viruses showed binding signals for long α 2-6 at the highest viral concentrations. However, the binding affinity for long α 2-6 is minimal compared to the high affinity for α 2-3 over the entire viral concentration range (Fig. 4b). Interpretation of the glycan binding data at the highest concentration alone would have led to an erroneous conclusion that these viruses have acquired binding to human-like receptors. The third issue concerns the use of chicken red blood cells (cRBC) agglutination assays. Desialylated and α 2-6 resialylated cRBC have been extensively used in agglutination assays to study α 2-6 binding specificity of wild-type and mutant viruses^{18,24,25}. MALDI-MS glycan profile of cRBC (Supplementary Fig. 2 online) shows limited abundance of N-linked α 2-6 with long branches. Thus, α 2-6 resialylation of these cRBC is unlikely to provide the required human-like receptors defined in the approach presented in this study.

These findings suggest that using glycan arrays with long α 2-6 will be valuable for surveillance of the evolution of human adaptation of influenza A viral subtypes. Specifically, this can be achieved either by expanding the diversity of the glycans to include upper respiratory tissue-specific glycans on the current glycan microarrays, or through the development of a glycan array that includes representative long α 2-6 structures based on those observed in the upper respiratory tissues. Furthermore, these arrays need to be developed for rapid dose-dependent analyses of HA binding to establish specificity to long α 2-6 structures. Adaptation studies with the avian viruses using appropriate long α 2-6 structures might provide the needed selection pressure for human adaptation. The ability of H5N1 to bind cone-like topology glycans with high specificity may explain why selection pressure strategies based just on glycan linkages have been unsuccessful. Finally, a recent study reported that the mutations in the glycan binding site of HA (leading to human receptor binding) could facilitate the generation of antibodies with better neutralizing sensitivity²⁴. A sufficient understanding of the avian H5N1 HA mutations leading to long α 2-6 binding specificity offers an opportunity for intervention through vaccine development to negate the eventuality of a H5N1 pandemic.

METHODS

Lectin binding of human upper respiratory tissues. Normal human trachea tissue sections (US Biological) were deparaffinized and rehydrated, followed by blocking endogenous biotin using the streptavidin/biotin blocking kit (Vector Labs). Sections were then incubated with FITC-labeled Jacalin, biotinylated conA and biotinylated *Sambucus nigra* agglutinin (SNA-I) (Vector labs; 10 μ g/ml in PBS with 0.05% Tween-20) for 3 h. After washing with TBST (Tris-buffered saline with 0.1% Tween-20), the sections were incubated with Alexa fluor 546 streptavidin (Invitrogen) for 1 h. Slides were washed with TBST and viewed under a confocal microscope (Zeiss LSM510 laser scanning confocal microscopy). All incubations were performed at 22 °C. The glycan array data for the plant lectins can be accessed from the Consortium for Functional Glycomics (CFG) web site, <http://www.functionalglycomics.org/glycomics/publicdata/primaryscreen.jsp> by selecting Plant lectins in the "Analyte Category."

MALDI-MS and TOF-TOF MS analysis of N-linked glycans. Human bronchial epithelial (HBE) cells were chosen as representative upper respiratory ciliated epithelial cells based on the extensive attachment of human-adapted H1N1 and H3N2 viruses to these cells¹. These cells were harvested at >90% confluency with 100 mM citrate saline buffer and the cell membrane was isolated by homogenization. Washed and pooled RBCs (Rockland Immunochemicals) were lysed by resuspending in deionized water for 15 min and the cell membrane was isolated as described above. The cell membrane fractions were treated with PNGaseF (New England Biolabs) and the reaction mixture was incubated overnight at 37 °C. The reaction mixture was boiled for 10 min

to deactivate the enzyme and the deglycosylated peptides and proteins were removed using a Sep-Pak C18 SPE cartridge (Waters). The glycans were further desalted and purified into neutral (25% acetonitrile fraction) and acidic (50% acetonitrile containing 0.05% trifluoroacetic acid) fractions using graphitized carbon solid-phase extraction columns (Supelco) and lyophilized.

The neutral and acidic fractions were analyzed by MALDI-TOF MS in positive and negative linear modes, respectively, with soft ionization conditions (accelerating voltage 22 kV, grid voltage 93%, guide wire 0.3% and extraction delay time of 150 ns). Matrices used were as follows: 10 mg/ml 6-aza-thiothymine in ethanol spotted on perfluorinated Nafion resin for acidic glycans and 40 mg/ml 2,5-dihydroxybenzoic acid in ethanol for neutral glycans. The peaks were calibrated as nonsialylated species using external glycan standards. The predominant expression of α 2-6 sialylated glycans was confirmed by treatment of samples using Sialidase A and S. Briefly, the isolated glycans were incubated with 0.1 U of *Arthrobacter ureafaciens* sialidase (Sialidase A, cleaves both α 2-3 and α 2-6 sialic acid linkages) or *Streptococcus pneumoniae* sialidase (Sialidase S, specifically cleaves α 2-3 sialic acid linkage) in a final volume of 100 μ l of 50 mM sodium phosphate, pH 6.0, at 37 °C for 14 h. The glycans were purified as before and the sialidase-treated glycans were analyzed by MALDI-TOF MS.

Sialidase A-treated glycans were 2-aminobenzamide (2-AB) derivatized, as described previously²⁶, purified²⁷ and lyophilized. MS/MS mass spectra were acquired on an Applied Biosystems 4700 TOF/TOF Proteomics analyzer, equipped with delayed extraction and a UV laser (355 nm) controlled by AB 4700 Data Explorer software. 1 μ l of a 5 mg/ml solution of α -cyano-4-hydroxycinnamic acid prepared in acetonitrile/water mixture (60:40, vol/vol) was seeded on a standard 192-well stainless steel MALDI sample plate and allowed to air dry. Subsequently, 1 μ l of 2-AB-labeled glycan sample was spotted on the matrix and allowed to dry. Typically, 1,000–5,000 shots/spectrum were collected in the positive reflector mode for MS/MS spectra. CID mode (with air as the collision gas) with a chamber pressure of $1-6 \times 10^{-6}$ Torr and acceleration voltage of 2 kV was used for glycan fragmentation and internal peptide standards were used for calibration.

Cloning, baculovirus synthesis, expression and purification of HA. The soluble form of HA was expressed using the Baculovirus Expression Vector System (BEVS). H5N1 (A/Vietnam/1203/2004; Viet0304) baculovirus was created from a pAcGP67-Viet04-HA construct² using Baculogold system (BD Biosciences) according to manufacturer's instructions. The crystal structure of the BEVS expressed uncleaved HA0 from SC18 (ref. 15) is identical to the structure of the SC18 HA isolated from the whole virus after bromelain treatment¹⁴. Studies have also shown that glycan binding specificity of BEVS-expressed HAs is in agreement with that of whole viruses¹⁷. H1, H3 and H5 baculoviruses were used to infect 500 ml suspension cultures of Sf9 cells (Invitrogen) cultured in Sf900 II SFM medium (Invitrogen). The infection was monitored and the cells were harvested 3–4 d after infection. The soluble form of HA was purified from the supernatant of the infected cells using the protocol described previously¹⁵. Briefly, the supernatant was concentrated using Centricon Plus-70 centrifugal filters (Millipore) and the soluble HA was recovered from the concentrated cell supernatant by performing affinity chromatography using Ni-NTA beads (Qiagen). Eluting fractions containing HA were pooled and dialyzed overnight with a 10 mM Tris-HCl, 50 mM NaCl buffer (pH 8.0). Subsequently, ion exchange chromatography was performed on the dialyzed samples using a Mono-Q HR10/10 column (GE Healthcare). The fractions containing HA were pooled together and subjected to ultrafiltration using Amicon Ultra 100 K NMWL membrane filters (Millipore). The protein was concentrated and reconstituted in PBS. The purified protein was quantified using Bio-Rad's protein assay (Bio-Rad).

Binding of H1, H3 and H5 HA to human lung tissues. Normal human trachea (US Biological) and lung (US Biomax) tissue sections were deparaffinized, rehydrated and incubated with 1% BSA in PBS for 30 min to prevent nonspecific binding. SC18 and Mos99 were precomplexed with primary antibody (mouse anti-6 \times His tag, Abcam) and secondary antibody (Alexa fluor 488 goat anti-mouse, Invitrogen) in a ratio of 4:2:1, respectively, for 20 min on ice. The tissue binding was performed over different HA concentrations by diluting the precomplexed HA in 1% BSA-PBS. Tissue sections were then incubated with

the HA-antibody complexes for 3 h at 22 °C. Sections were counterstained with propidium iodide (Invitrogen; 1:100 in TBST), washed extensively and then viewed under a confocal microscope (Zeiss LSM510 laser scanning confocal microscopy). In the case of sialidase pretreatment, tissue sections were incubated with 0.2 units of Sialidase A (recombinant from *Arthrobacter ureafaciens*, Prozyme) for 3 h at 37 °C before incubation with the proteins.

Dose-dependent direct binding of H1, H3 HA and H5 viruses. Streptavidin-coated high binding capacity 384-well plate (Pierce) was rinsed with PBS and each well was incubated with 50 µl of 2.4 µM solution of biotinylated glycans (3'SLN, 6'SLN, 3'SLN-LN, 6'SLN-LN and 3'SLN-LN-LN) in PBS overnight at 4 °C. LN corresponds to lactosamine (Galβ1-4GlcNAc) and 3'SLN and 6'SLN, respectively, correspond to Neu5Acα2-3 and Neu5Acα2-6 linked to LN. These glycans were obtained from the Consortium for Functional Glycomics (CFG). The plate was subsequently washed with PBS to remove excess glycan and used without further processing. Appropriate amounts of His-tagged HA protein, primary (mouse anti 6× His tag IgG, Abcam) and secondary (HRP-conjugated goat anti Mouse IgG, Santacruz Biotechnology) antibodies were mixed in the ratio 4:2:1 and incubated on ice for 20 min. The highest HA concentration of 40 µg/ml was chosen based on typical concentrations used in CFG glycan array analysis¹⁷. The mixture (precomplexed HA) was made up to a final volume of 250 µl with 1% BSA in PBS buffer and 50 µl of precomplexed HA was added to each glycan-coated well and incubated at 22 °C for 2 h. The wells were extensively washed with PBS containing 0.05% Tween-20 followed by washes with PBS.

Virus stocks (from CDC) were propagated in the allantoic cavity of 10-d-old embryonated hens' eggs at 37 °C. The allantoic fluids were harvested 24 h after inoculation and inactivated by treatment with B-propiolactone (BPL; 1/1,000) for 3 d at 4 °C. Virus binding to the glycan-coated wells was performed as described²⁰ by adding appropriate amount of virus to each well after diluting in PBS containing 1% BSA and incubating overnight at 4 °C. After rinsing excess virus with PBS containing 0.05% Tween-20 and PBS, the wells were incubated with antibody against the virus (ferret-anti-influenza A raised against A/Hong Kong/213/03 H5N1) for 5 h at 4 °C. After extensive washing of the antibody, the plate was incubated with HRP-linked goat-anti-ferret antibody (Rockland Immunochemicals) for 2 h at 4 °C. After extensive washes with PBS containing 0.05% Tween-20 and PBS, in all cases, HRP activity was estimated using Amplex Red Peroxidase Assay Kit (Invitrogen) according to manufacturer's instructions. Appropriate negative controls were included and all assays were done in triplicate.

Data mining analysis of glycan microarray data. The data from glycan microarray screening of H1, H3 and H5 HA were obtained from the Consortium for Functional Glycomics (CFG) web site, <http://www.functionalglycomics.org/glycomics/publicdata/primaryscreen.jsp>. From this web page, the "Other" option in the "Analyte category" was selected to obtain a list of microbial lectins screened on the glycan array. The data for all influenza virus HAs were obtained by sorting this list by investigator and searching the sorted list using "Jim Paulson" as the investigator name. The principle and methodologies for data mining analysis are well established in many fields and involve two main aspects: feature abstraction and classification. A variety of features, including all possible di-, tri- and tetrasaccharide combinations, present in each glycan in the glycan array were abstracted (Fig. 3a). The rationale behind choosing these features is based on the binding of di-, tri- or tetrasaccharides to the glycan binding site of HA. The final data set has features from the glycans, as well as the binding signals for each of the HAs screened on the array. The rule-induction classification method was used. One of the main advantages of this method is that it generates *IF-THEN* rules, which can be interpreted more easily when compared to the other statistical or mathematical methods (Supplementary Table 1 online).

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

This work was supported by the National Institute of General Medical Sciences glue grant U54 GM62116 and National Institutes of Health (NIH) grants GM57073. The authors thank James Stevens for his generous help in providing

the H1, H3 baculoviruses and H5 HA clones. Human bronchial epithelial (HBE) cells (16HBE14o-; were a gift from D.C. Gruenert, University of California, San Francisco. The authors thank the CFG for the glycan array data and for the supply of glycans for direct binding experiments. The authors acknowledge Maha Venkataraman and Subu Ramakrishnan at the Bioinformatics Core of the CFG for their help with data preparation in the data mining analysis and Dr. Shu-hua Nong for his help with purification of the HA proteins. Confocal microscopy was performed using the WM Keck Foundation Biological Imaging Facility at the Whitehead Institute. The MALDI TOF-TOF analysis was performed at the Barnett Institute of Chemical and Biological Analysis (Northeastern University, Boston).

Published online at <http://www.nature.com/naturebiotechnology/>

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions>

- van Riel, D. *et al.* Human and avian influenza viruses target different cells in the lower respiratory tract of humans and other mammals. *Am. J. Pathol.* **171**, 1215–1223 (2007).
- Stevens, J. *et al.* Structure and receptor specificity of the hemagglutinin from an H5N1 influenza virus. *Science* **312**, 404–410 (2006).
- Russell, R.J., Stevens, D.J., Haire, L.F., Gamblin, S.J. & Skehel, J.J. Avian and human receptor binding by hemagglutinins of influenza A viruses. *Glycoconj. J.* **23**, 85–92 (2006).
- Skehel, J.J. & Wiley, D.C. Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. *Annu. Rev. Biochem.* **69**, 531–569 (2000).
- Yen, H.L. *et al.* Inefficient transmission of H5N1 influenza viruses in a ferret contact model. *J. Virol.* **81**, 6890–6898 (2007).
- Maines, T.R. *et al.* Lack of transmission of H5N1 avian-human reassortant influenza viruses in a ferret model. *Proc. Natl. Acad. Sci. USA* **103**, 12121–12126 (2006).
- Kuiken, T. *et al.* Host species barriers to influenza virus infections. *Science* **312**, 394–397 (2006).
- Shinya, K. *et al.* Avian flu: influenza virus receptors in the human airway. *Nature* **440**, 435–436 (2006).
- van Riel, D. *et al.* H5N1 Virus Attachment to Lower Respiratory Tract. *Science* **312**, 399 (2006).
- Nicholls, J.M., Bourne, A.J., Chen, H., Guan, Y. & Peiris, J.S. Sialic acid receptor detection in the human respiratory tract: evidence for widespread distribution of potential binding sites for human and avian influenza viruses. *Respir. Res.* **8**, 73 (2007).
- Eisen, M.B., Sabesan, S., Skehel, J.J. & Wiley, D.C. 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 (1997).
- Ha, Y., Stevens, D.J., Skehel, J.J. & Wiley, D.C. 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 (2001).
- Ha, Y., Stevens, D.J., Skehel, J.J. & Wiley, D.C. X-ray structure of the hemagglutinin of a potential H3 avian progenitor of the 1968 Hong Kong pandemic influenza virus. *Virology* **309**, 209–218 (2003).
- Gamblin, S.J. *et al.* The structure and receptor binding properties of the 1918 influenza hemagglutinin. *Science* **303**, 1838–1842 (2004).
- Stevens, J. *et al.* Structure of the uncleaved human H1 hemagglutinin from the extinct 1918 influenza virus. *Science* **303**, 1866–1870 (2004).
- Stevens, J. *et al.* Glycan microarray analysis of the hemagglutinins from modern and pandemic influenza viruses reveals different receptor specificities. *J. Mol. Biol.* **355**, 1143–1155 (2006).
- Stevens, J., Blixt, O., Paulson, J.C. & Wilson, I.A. Glycan microarray technologies: tools to survey host specificity of influenza viruses. *Nat. Rev. Microbiol.* **4**, 857–864 (2006).
- Tumpey, T.M. *et al.* A two-amino acid change in the hemagglutinin of the 1918 influenza virus abolishes transmission. *Science* **315**, 655–659 (2007).
- Gambaryan, A. *et al.* Evolution of the receptor binding phenotype of influenza A (H5) viruses. *Virology* **344**, 432–438 (2006).
- Yamada, S. *et al.* Haemagglutinin mutations responsible for the binding of H5N1 influenza A viruses to human-type receptors. *Nature* **444**, 378–382 (2006).
- Nicholls, J.M. *et al.* Tropism of avian influenza A (H5N1) in the upper and lower respiratory tract. *Nat. Med.* **13**, 147–149 (2007).
- Neumann, G. & Kawaoka, Y. Host range restriction and pathogenicity in the context of influenza pandemic. *Emerg. Infect. Dis.* **12**, 881–886 (2006).
- Matrosovich, M., Zhou, N., Kawaoka, Y. & Webster, R. The surface glycoproteins of H5 influenza viruses isolated from humans, chickens, and wild aquatic birds have distinguishable properties. *J. Virol.* **73**, 1146–1155 (1999).
- Yang, Z.Y. *et al.* Immunization by avian H5 influenza hemagglutinin mutants with altered receptor binding specificity. *Science* **317**, 825–828 (2007).
- Kumari, K. *et al.* Receptor binding specificity of recent human H3N2 influenza viruses. *Viol. J.* **4**, 42–53 (2007).
- Bigge, J.C. *et al.* Nonselective and efficient fluorescent labeling of glycans using 2-amino benzamide and anthranilic acid. *Anal. Biochem.* **230**, 229–238 (1995).
- Grun, C.H. *et al.* One-step biotinylation procedure for carbohydrates to study carbohydrate-protein interactions. *Anal. Biochem.* **354**, 54–63 (2006).