Context-Specific Target Definition in Influenza A Virus Hemagglutinin-Glycan Receptor Interactions

Zachary Shriver,1 Rahul Raman,1 Karthik Viswanathan,1 and Ram Sasisekharan1,*

1Harvard-MIT Division of Health Sciences and Technology, Koch Institute for Integrative Cancer Research, Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
*Correspondence: rams@mit.edu
DOI 10.1016/j.chembiol.2009.08.002

Protein-glycan interactions are important regulators of a variety of biological processes, ranging from immune recognition to anticoagulation. An important area of active research is directed toward understanding the role of host cell surface glycans as recognition sites for pathogen protein receptors. Recognition of cell surface glycans is a widely employed strategy for a variety of pathogens, including bacteria, parasites, and viruses. We present here a representative example of such an interaction: the binding of influenza A hemagglutinin (HA) to specific sialylated glycans on the cell surface of human upper airway epithelial cells, which initiates the infection cycle. We detail a generalizable strategy to understand the nature of protein-glycan interactions both structurally and biochemically, using HA as a model system. This strategy combines a top-down approach using available structural information to define important contacts between glycans and HA, with a bottom-up approach using data-mining and informatics approaches to identify the common motifs that distinguish glycan binders from nonbinders. By probing protein-glycan interactions simultaneously through top-down and bottom-up approaches, we can scientifically validate a series of observations. This in turn provides additional confidence and surmounts known challenges in the study of protein-glycan interactions, such as accounting for multivalency, and thus truly defines concepts such as specificity, affinity, and avidity. With the advent of new technologies for glycomics—including glycan arrays, data-mining solutions, and robust algorithms to model protein-glycan interactions—we anticipate that such combination approaches will become tractable for a wide variety of protein-glycan interactions.

Introduction

Complex glycans at the surface of cells and on circulating signaling molecules play a fundamental role in determining how a cell “sees” and responds to external events. In this capacity, through interactions with proteins, complex glycans modulate a variety of biological processes including pathogen recognition, innate and acquired immunity (Alexopoulou et al., 2007; Crocker et al., 2007; Rudd et al., 2004; van Die and Cummings, 2006), glycoprotein targeting (Bhatia and Mukhopadhyay, 1999; Helenius and Aeberli, 2004; Varki et al., 2008), adhesion (Kawashima et al., 2005; Lowe, 2002; Taylor and Drickamer, 2007; Wang et al., 2005) and trafficking (Crocker, 2005; Smithson et al., 2001). Obtaining a complete picture of a biological process or understanding higher-level organization of a biological system accordingly requires decoding protein-glycan interactions. In recognition of this practical need, there has been a surge in the development of strategies for chemical and chemoenzymatic synthesis of diverse glycan structures (Blixt and Razi, 2006; Hanson et al., 2004; Seeberger and Werz, 2007) that represent the common terminal motifs displayed as part of N- and O-linked glycans and glycolipids. To properly present these glycans to their protein partners, synthetic glycan motifs have been anchored to a variety of platforms including polymeric backbones such as polyglutamic acid and polyacrylamide, on dendrimers, and more recently on glycan microarrays (Bovin et al., 2004; Collins and Paulson, 2004; Gambaryan et al., 2006; Mammen et al., 1995; Totani et al., 2003). Despite these advances in the development of platforms to study glycan-protein interactions, challenges remain in defining glycan targets for the purpose of bridging the biochemical and biophysical specificity of glycan-protein interactions with the biological functions modulated by these interactions. In the case of glycan-protein interactions, one size does not fit all. Strategies that have proven informative in the areas of genomics or proteomics may or may not facilitate our description of the glycome for several reasons. Glycan-protein interactions, leading to either the activation or inhibition of a biological response, are often not binary but rather involve more subtle mediation of a signaling pathway. In addition, glycan-protein interactions typically involve multivalency with respect to both the protein and the glycan, 1:1 monovalent complexes are often weak and display dissociation constants on the order of 1 to 1000 μM. Biochemical/biophysical descriptions of protein-glycan interactions therefore depend both on context and experimental design. In this framework, the careful description of experimental results, including an understanding of both the strengths and limitations of an approach, is appreciated. Finally, it is becoming clear that there are both finer and coarser determinants to the specificity of a given protein-glycan interaction than the simple monosaccharide sequence of a glycan. Describing a glycan-binding sequence by its monosaccharide composition and the linkages between the monosaccharides, although important, in most cases does not afford the same descriptive capacity as it does for other biopolymers.
Ensuring an accurate and complete picture of protein-glycan interactions despite these difficulties demands studies that integrate the basic biochemistry and molecular biology of the system with analytical approaches, while simultaneously enabling the appropriate translation to the biology. One of the best-studied systems of glycan-protein interactions modulating a biological function is that of the role of influenza A virus hemagglutinin (HA)-glycan interactions in viral pathogenesis. This review presents an overview of the various technologies that have been used for glycan receptor target definition for HA, their strengths and limitations, and how an integrated framework enabled the bridging of HA-glycan interactions with the host adaptation of the virus. This integrated approach could serve as a framework for target definition in other protein-glycan interactions.

**Overview of Influenza A Viruses**

Influenza A virus is a negative strand RNA virus with eight gene segments. Three of the genes—hemagglutinin (HA), neuraminidase (NA), and the polymerase (PB)—have been shown to be critical for infection and human-to-human transmission (Palese, 2004; Pappas et al., 2008; Tumpey et al., 2007) (Figure 1). Within influenza A, five of the genome segments encoding the nucleocapsid protein (NP), the matrix proteins (M1 and M2), the nonstructural proteins (NS1 and NS2), and polymerase proteins (PB1, PB2 and PA) have maintained a relatively unbroken evolutionary history in humans. In contrast, the two genes encoding the major cell surface proteins (HA and NA) have been subjected to substantial evolutionary pressure, including mutation (antigenic drift) and wholesale reassortment (antigenic shift). Due to their variability, strains of influenza virus are identified based on their serotype of HA and NA. There are currently 16 known serotypes of HA and nine of NA.

At the moment, there is substantial public health concern surrounding the prospect of another influenza A pandemic and its associated potential global implications. The past century has seen four influenza pandemics. The first and most severe occurred in 1918, involved an H1N1 virus, and led to the death of at least 40 million people worldwide. Less serious pandemics occurred in 1957, 1968, and 1977. Notably, in real time has been the advent of a reassorted H1N1 virus, viz., 2009 H1N1 or “swine” flu, which is antigenically dissimilar from seasonal, circulating H1N1s and which has already been declared a pandemic by the World Health Organization. Viruses containing the 16 HA and nine NA serotypes are naturally present in wild aquatic bird populations where they exist commensally without causing disease, allowing birds to become a reservoir for influenza strains. This is of specific concern because of the influenza pandemics of the last century, those that arose from H2N2 (1957) and H3N2 (1968), were avian-human reassortments that resulted in the humanization of an avian-adapted virus and efficient human-to-human transmission. Those genetic reassortments that led to an avian-to-human switch yield a number of important scientific and medical questions, not the least of which is what changes lead away from infectivity and propagation in avian species and toward human transmissibility? In light of a particular influenza strain, H5N1 or the so-called bird flu, addressing these questions becomes more critical. Transmission of avian H5N1 influenza viruses to humans has been observed thus far only upon direct contact with infected poultry; the virus has not yet demonstrated efficient human-to-human transmission ability. Given that human infectivity has occurred and appreciating that this virus strain is highly lethal (estimates are as high as ~60% of infected individuals), the importance of comprehending the mechanism and specificity of viral entry and infection as well as identifying additional strategies for intervention is clear.

**HA-Glycan Interactions: Description and Biological Importance**

The influenza A infection cycle can be described as a three-stage process (Figure 1). The first stage is the attachment of HA of the virus to complex glycan receptors on the host cells. Following attachment, the virus is internalized by endocytosis where structural changes in HA produce the fusion of the viral membrane with the endosomal membrane. Internalization facilitates activation of the ion channel activity of M2 and transport of the viral RNA to the nucleus. In the nucleus, viral RNA undergoes replication and transcription. Newly synthesized proteins HA and NA are secreted through the Golgi to the cell surface. Other proteins

---

**Figure 1. The Influenza Infection Cycle**

HA on the surface of the virus binds glycans terminated by sialic acid with a specific linkage (green or red), initiating fusion of the virus with the host cell. This interaction is highly specific and is governed by the type of sialic acid linkage, the underlying sugars and branching pattern of the glycan receptors. The other viral surface proteins are the neuraminidase (NA) and the ion channel protein M2. Once the virus is internalized in the cell, the fusion between viral and nuclear membranes occurs, and complexes of RNA and proteins termed viral ribonucleoprotein complexes (vRNPs) are transported into the nucleus of the host cell. The transcription to mRNA takes place in the host cell nucleus followed by export and protein synthesis. Also within the nucleus there is transcription of the RNA genome. Assembly of progeny vRNPs then occurs, with export, assembly of the virus progeny, and finally budding of the newly formed virus particles.
are transported to the nucleus where they associate with the synthesized transcripts of viral RNA to form virions. These virions, with HA and NA on their surface, then bulge from the cell membrane. The action of NA, which cleaves the sialic-acid-capped glycan structure and eliminates the interaction of the host cell glycans with the newly formed virus particle, facilitates the release of progeny viruses from the host cells.

In the context of infectivity, the HA-glycan interaction is one of the most critical components governing virus selectivity. HA itself is a homotrimer (Figure 2); each monomer is synthesized as a single polypeptide that contains a proteolytic site, that is cleaved by host enzymes into two subunits (HA1 and HA2). Numerous crystal structures of different HAs have been solved, both alone and as co-crystals with various glycan structures (Eisen et al., 1997; Gamblin et al., 2004; Ha et al., 2001, 2003; Sauter et al., 1992; Skehel and Wiley, 2000; Stevens et al., 2006c; Weis et al., 1988; Yamada et al., 2006). This body of structural work offers valuable insights on the specificity of HA binding. In tandem, biochemical studies identified a key feature for binding of human-adapted HA to glycans terminated by N-acetyl neuraminic acid linked α2→6 to galactose (N-acetyl-D-neuraminic acid [Neu5Ac]α2→6 D-galactose [Gal], hereafter referred to as α2→6) present on human respiratory epithelia (Ibricevic et al., 2006; Russell et al., 2006; Shinya et al., 2006; Skehel and Wiley, 2000; van Riel et al., 2007). These biochemical findings have been correlated with the finding that human respiratory tissue contains epithelial cells with α2→6 sialic-acid-capped glycans (sites for attachment of human-adapted viruses), whereas cells that primarily express glycans terminated by N-acetyl neuraminic acid linked α2→3 to galactose (Neu5Acα2→3Gal, hereafter referred to as α2→3), such as the alveolar cells, are sites for attachment of avian-adapted viruses (Shinya et al., 2006). These and other findings suggest that for a virus to cross over from avian species to humans, its HA must switch binding preference from α2→3 sialylated glycan receptors, present in avian species, to α2→6 receptors (Connor et al., 1994; Glaser et al., 2006; Kumari et al., 2001; Matrosovich et al., 2004, 2007; Rogers et al., 1983; Russell et al., 2006; Tumpey et al., 2007), present in human upper airway epithelium. Recent developments question the conception that the preference of HA binding to α2→3 or α2→6 alone is sufficient to designate human- versus avian-adapted HA (Bewley, 2008).

HA-Glycan Interactions: Answering Questions and Generating Others

The specificity of HA binding to cell surface glycan structures has been addressed not only by crystal structure analysis, but also by the advent of powerful biochemical and biological tools that provide significant insight and raise additional questions regarding HA-glycan interactions.

Tools to Study HA-Glycan Interactions

A variety of biochemical methods have traditionally been used to characterize the specificity of HA-glycan interactions. One of the earliest methods, still in use today for probing the glycan-binding specificity of a particular virus (through its HA) involves measuring its ability to agglutinate red blood cells (RBCs). RBCs from species such as chicken, turkey, horse, guinea pigs and humans have been used (Connor et al., 1994; Paulson and Rogers, 1987; Rogers and Paulson, 1983; Tumpey et al., 2007; Yang et al., 2007). Equine RBCs primarily contain α2→3 glycans at their cell surface, whereas turkey and guinea pig RBCs primarily contain α2→6 glycans. To augment the usefulness and specificity of these models, the agglutination method has been modified to include a step of complete desialylation of species-specific RBCs followed by specific resialylation by either α2→3 or α2→6 sialyltransferase, creating a more specific assay (Carroll et al., 1981). Agglutination and other traditional hemagglutination assays permit the definition of viruses’ specificity according to the sialic acid linkage, as described above. These assays prove useful in the classification of virus specificity and virulence. However, in its present form, this assay does not allow the examination of specificity beyond the sialic acid linkage, so probing the fine specificity of HA binding is not possible. There are other drawbacks to the use of RBC agglutination to define glycan-binding specificity of a virus strain. First, there is possibly substantial variability in the N- and O-linked cell surface glycans between different batches of RBCs and related potential for variable response. Additionally, the glycan structures on the cell surface of RBCs are not similar to those present on the surface of cells of the

Figure 2. Structure of Trimeric HA

The cartoon rendering of a representative H1N1 HA (Protein Data Bank ID 1RU7) is shown where each chain (HA1 or HA2) is colored distinctly. The region in HA1 involved in binding to sialylated glycan receptor and the region between HA1 and HA2 involved in membrane fusion are also highlighted using dotted circles.
upper airway (Chandrasekaran et al., 2008). Viruses might bind to, and elicit agglutination through glycan receptors that are not present in the upper airway and therefore not physiologically relevant.

Subsequent development of solid-phase fetuin capture assays has provided a wealth of information on the glycan-binding properties of influenza A viruses. In these assays, viruses are immobilized on fetuin-coated surfaces and their binding to various sialylated glycans (including polyvalent compounds) is evaluated (Gambaryan et al., 1995, 2005, 2008; Gambaryan and Matrosovich, 1992; Matrosovich et al., 2000). The presentation of the viruses is heterogeneous because the amount of virus captured on the plate depends on the binding of the viral HA to the sialylated glycans on fetuin. Furthermore, measuring the binding of fixed viruses to glycans in solution is opposite to the physiological event where glycans are less mobile on the cell surface as compared with the virus.

Although not commonly used to investigate HA-glycan interactions, isothermal titration calorimetry and surface plasmon resonance have been widely employed to determine equilibrium binding affinity constants and thermodynamic parameters for glycan-protein interactions (Dam et al., 2009; Duverger et al., 2003; Karamanska et al., 2008). These methodologies thus offer promising tools for quantifying the binding affinity of HA-glycan interactions.

Recent advances in the chemical and chemoenzymatic synthesis of glycans have allowed for the development of glycan array platforms. These platforms consist of hundreds of synthetic glycan motifs (typically present on N- and O-linked glycoproteins and glycolipids) displayed on the surface of the array. Multiple types of arrays have been developed that utilize different strategies including the formation of neoglycolipids (Fukui et al., 2002), neoglycoproteins (Gildersleeve et al., 2008; Huang et al., 2008), or the direct application of glycans to various surfaces (Blixt et al., 2004; Grun et al., 2006; Karamanska et al., 2008; Liang et al., 2008; Mercey et al., 2008; Xia et al., 2005). Several fine reviews outline advances that have occurred recently in the development of array technologies (Blixt et al., 2004; Feizi et al., 2002; Houseman and Mrksich, 2002; Liang et al., 2008; Oyelaran and Gildersleeve, 2007; Seeberger and Werz, 2007; Stevens et al., 2006b). Among the various glycan array platforms, those developed by the Consortium for Functional Glycomics (CFG) are arguably the most accessible. More than 2000 samples have been screened on CFG glycan arrays by the scientific community, and data generated on the CFG glycan array platform are disseminated freely to the public via web-based interfaces (www.functionalglycomics.org/static/consortium/resources/resourcescore.shtml). Recent studies have begun to adapt these technologies toward the ultimate presentation of natural glycans by harvesting glycans from the surface of cells and imprinting these on a glycan array format (Song et al., 2009), thus allowing one to probe the glycan repertoire of a biological system. In this manner, it is possible to create arrays specific to a disease or biological process that permit the interpretation of glycan-binding information in relation to that process.

The glycan array platforms developed by the CFG have already been used to screen wild-type and mutant forms (mutations in HA) of intact viruses and recombinant HAs belonging to the H1, H2, H3, H5, H7, and H9 subtypes (Belser et al., 2008; Kumari et al., 2007; Stevens et al., 2006a, 2008; Wan et al., 2008). These studies have increased our understanding of HA binding to sialylated glycan receptors by mapping the effect of glycan modifications, such as sulfation and fucosylation, on the HA-glycan interactions. The scope of most of these studies is intended to serve as a primary screen where high viral titers or HA concentrations are used to define the binding preference of HA in terms of $\alpha^2 \rightarrow 3$ and/or $\alpha^2 \rightarrow 6$ binding.

**Tools to Bridge HA-Glycan Interactions with Viral Pathogenesis**

In 1989 Palese and colleagues first demonstrated the ability to manipulate the influenza virus genome by developing a system that allowed the use of standard recombinant DNA technology to modify the genome of influenza virus to express foreign genes (Luytjes et al., 1989). Their advance formed the basis for the development of methods—termed reverse genetics—which permit synthesis of the whole virus from the cDNAs of individual virus genes (Hoffmann et al., 2000; Pleschka et al., 1996) (Figure 3).

In addition to the advent and widespread adoption of reverse genetics, the emergence of ferrets as a model system to study pathogenesis and contact and respiratory droplet modes of transmission (Lowen and Palese, 2007; Tumpey et al., 2007; van der Laan et al., 2008) has proven of equal importance to the investigation of various aspects of influenza biology. Studies have demonstrated that ferrets possess similar glycan structures to humans, including a predominance of human-like $\alpha^2 \rightarrow 6$ glycans in their upper respiratory tract epithelium (Maines et al., 2006). The ability to completely reconstruct the pandemic 1918 H1N1 viruses through reverse genetics and to test its virulence in ferrets permitted a systematic exploration of the roles for various viral genes in the virulence and transmissibility of influenza A strains (Palese, 2004; Tumpey et al., 2005). Single gene reassortants of the highly virulent pandemic human H1N1 (A/South

---

**Figure 3. The Eight-Plasmid Polymerase(pol) I–II System for the Generation of Influenza A Virus**

Expression plasmids containing the eight viral cDNAs, inserted between the human pol I promoter and the pol II promoter are transfected into cells. Because each plasmid contains two different promoters, both cellular pol I and pol II will transcribe the plasmid template in different orientations, which will result in the production of vRNAs and mRNAs, respectively. Synthesis of the viral polymerase complex proteins (PB1, PB2, PA, and nucleoproteins) initiates the viral replication cycle. Ultimately, the assembly of viral molecules results in the interaction of all synthesized molecules to generate infectious influenza A virus.
Carolina/1/18 or SC18) virus with a contemporary epidemic human H1N1 (A/Texas/36/91 or Tx91) virus showed that the HA of SC18 had the most profound effect on the virulence of human H1N1 (A/Texas/36/91 or Tx91). In other studies using H7 and H9 viruses (Tumpey et al., 2007), it was observed that substantial $\alpha_2\rightarrow 6$ binding did transmit efficiently. These studies suggest that defining glycan receptor targets of HA in terms of $\alpha_2\rightarrow 3$ / $\alpha_2\rightarrow 6$ alone (using RBC agglutination, glycan arrays, etc.) is inadequate to bridge with the biological role of HA in human adaptation and viral virulence. In light of the critical role of HA with respect to virulence and human adaptation of the virus, it is clear that additional factors govern the specificity of HA binding and determine whether a virus is able to efficiently infect or transmit in humans.

**Integrating Bottoms-Up and Top-Down Approaches to Bridge Structure and Biology**

Addressing the above issues required an integration of information from complementary approaches that permit a move from the biological to the structural space in an iterative and transitive fashion (Figure 4). Each endeavor addressed a series of specific questions, and collectively the data provide complementary, overlapping sets of information that define the underlying specificity of influenza A HA. The bottom-up approach starts by examining the structure of glycans present on the cell surface.
Many glycan-protein interactions are multivalent and thus are the product of multiple interactions between a glycan structure and a protein. Qualitative description of the strength of this interaction is termed avidity and an apparent equilibrium constant is used to quantitatively describe the interaction. Whitesides and others have explicitly dealt with the relationship between affinity ($K_{\text{mono}}^\text{poly}$), a description of a 1:1 complex and avidity ($K_{\text{ave}}^\text{poly}$).

\[ K_{\text{ave}}^\text{poly} = (K_{\text{mono}}^\text{poly})^\alpha \]

In this case, $\alpha$, is a measure of the cooperativity of the system. Moving from a description of equilibrium constants to a measure of $\Delta G$, $\Delta G_{\text{ave}}^\text{poly} = \alpha \Delta G_{\text{ave}}^\text{mono}$ is defined as the average change in free energy of the interaction:

\[ \Delta G_{\text{ave}}^\text{poly} = \Delta G_{\text{ave}}^\text{mono} \]

Where $\Delta G_{\text{ave}}^\text{poly}$ is the degree of cooperativity, $\alpha$, in such systems is greater than one. Accordingly, in the case where $\Delta G_{\text{ave}}^\text{poly} = \Delta G_{\text{ave}}^\text{mono}$, $\alpha$ is 1, and in the case where $\Delta G_{\text{ave}}^\text{poly} < \Delta G_{\text{ave}}^\text{mono}$, the system experiences negative cooperativity.

of human airway epithelial cells, and then by using this information to quantitatively probe the affinity of HA to individual glycan species.

Because human upper respiratory epithelia are the primary targets for human-to-human transmission of influenza A virus, it is essential to identify the types of sialylated glycans present on these tissues. The characterization of the glycan repertoire for a given tissue type is challenging due to the complexity associated with accessing and elucidating these structures. While it is recognized that sialylated glycans of epithelial cells are the receptor for HA (and accordingly virus) binding and internalization, epithelial cells themselves possess a broad array of structurally diverse glycan structures. Addressing this diversity requires orthogonal biochemical and biological tools that are specific, sensitive, and robust. One approach that has been employed is to bridge lectin staining of different tissue sections with glycan profiling of representative human cell lines using matrix-assisted laser desorption ionization (MALDI) mass spectrometry (MS) and MS/MS (to fragment glycans and provide structural information) analyses (Comelli et al., 2006; Uematsu et al., 2005). Other information-rich techniques have also been employed, including capillary high-performance liquid chromatography, liquid chromatography MS, and nuclear magnetic resonance (Manzi et al., 2000; Norgard-Sumnicht et al., 2000).

Regardless of the particular analytical techniques employed, the use of these complementary approaches provides a superior quality of information because (1) data sets can be cross-checked with one another, and (2) no single analytical approach, no matter how powerful, has equally descriptive power for all structures, especially glycans. Notably, in one approach, Sambucus nigra agglutinin (SNA-I), a prototypic lectin that specifically binds to α2→6 glycans, and jacalin and concanavalin A (Con A), which bind to specific motifs in O-linked (Galβ1→3GalNAcα2→6 (on ciliated cells) glycans and localized distribution of O-linked α2→6 glycans on goblet cells in human tracheal epithelia (Chandrasekaran et al., 2008). This analysis was then confirmed and extended using detailed MALDI analysis of glycans derived from representative human upper epithelia cells to identify glycan composition, and MS/MS fragmentation to determine their structural features, including the presence of a lactosamine extension. Thus, MS and MS/MS information is interpreted in the context of the information obtained from lectin analysis (Chandrasekaran et al., 2008). This is but one example of the application of different, overlapping analytical techniques to obtain complementary data sets; other, equally valid strategies can be devised.

Importantly, this coupled approach can provide a powerful complement to the construction of natural arrays and increase their utility to identify glycan-binding partners to proteins. Alternatively, such analysis can provide an important framework for the development of quantitative array and array-like binding assays using selected synthetic structures that are representative of cell surface glycans. For an analysis of this nature to be valid, three key variables must be addressed and controlled to ensure accurate interpretation. First, as is the case in biological systems, glycan-protein interactions are typically multivalent and the strength of such a contact should be described based on its avidity. Numerous experimental setups have been proposed to measure avidity, and Kiessling and colleagues provide an excellent overview of many of these (Kiessling et al., 2000). Regardless of platform, the most useful quantitative information is obtained when $N$, the number of binding sites is known, though even without $N$, useful thermodynamic information can be determined (Mammen et al., 1998) (Figure 5). The spatial arrangement of glycans and the glycan-binding sites on proteins influence the structural valency (i.e., number of
binding sites on a protein that are occupied by the glycan motifs of glycan-protein interactions (Dam and Brewer, 2008; Dam et al., 2009). The spatial arrangement of glycan motifs depends on the extent of branching (bi, tri, tetra, penta-antennary N-linked glycans) and spatial distribution of glycosylation sites (clustered, linear, and globular) in different glycoprotein structures. The spatial arrangement of glycan-binding sites in a protein instead depends on the quaternary association of individual domains. In the case of HA, there is a homotrimeric association of identical domains resulting in three glycan-binding sites per trimeric HA unit. Second, similar to biological systems, the glycan structures on the surface should be in excess. If this is not the case, ambiguous results might be obtained, particularly if this situation is not made explicit when plotting and calculating an association constant. Finally, the response should be measured at various concentrations to determine a dissociation constant (K_d), which is inverse of the affinity constant, K_d = 1 / K_a.

In most of the earlier studies that focused on screening different HAs on glycan arrays, the binding event was designated as a single point “on” or “off.” This designation, though potentially useful, necessarily misses the context of the interaction and the relative biological importance of the interaction. More recently, biochemical assays have been designed to screen HA-glycan interactions over an entire range of HA concentrations (Srinivasan et al., 2008) that take into consideration the aforementioned factors, including spacing of the glycan motifs in an array platform and the spatial arrangement of glycan-binding sites. These assays have permitted quantification of the relative binding affinity of HA to different glycan motifs. Using such quantitative assays, it has been demonstrated that the human-adapted HAs share a high binding affinity to α2→6 glycans (particularly those that have multiple lactosamine repeats) that in turn correlate with their efficient airborne transmissibility (Maines et al., 2009; Srinivasan et al., 2008). Viruses of avian or swine origin although known to infect humans could be distinguished from the human-adapted viruses based on their quantitative α2→6 binding affinity of their respective HAs (Chandrasekaran et al., 2008; Maines et al., 2009).

Analysis employing the top-down approach originates from consideration of the molecular and structural aspects of HA-glycan interactions and correlates these aspects with the biochemical binding affinities. Using glycan conformational analysis, examination of HA-glycan cocrystal structures demonstrated that glycan topology (or three-dimensional shape of the glycan) plays a critical role in distinguishing binding of α2→3 and α2→6 glycans to avian and human-adapted HAs (Bewley, 2008; Chandrasekaran et al., 2008). Analysis of the various crystal structures indicated that a highly conserved set of amino acids Tyr98, Ser/Thr136, Trp153, His183, Leu/Ile194 (numbered based on H3 HA) across different HA subtypes are involved in anchoring the sialic acid. The specificity of HA to either α2→3 or α2→6 is governed by an extended range of interactions within the glycan-binding site—not only with the sialic acid, but also with the glycosidic oxygen atom and monosaccharides beyond sialic acid.

The ensemble of conformations sampled by the α2→3 and α2→6 glycans in the binding site of HA was described using a shape-based topological description. In the case of α2→3 glycans, the conformations sampled by the Neu5Acα2→3Gal linkage (keeping the Neu5Ac anchored) and the sugars beyond this linkage (at the reducing end) span a region on the binding surface of HA that resembles a “cone.” The assembly of these conformations is therefore described by the term cone-like topology (Figure 6).

When contrasted with the Neu5Acα2→3Gal linkage, the presence of the C6-C5 bond within the Neu5Acα2→6Gal linkage
provides additional conformational flexibility. The different conformations sampled by Neu5Acα2→6Gal linkage (keeping the Neu5Ac anchored) and the sugars beyond this linkage (at the reducing end) thus span a wider region on the HA binding surface. One part of this wider region is similar to the cone-like surface and the other part resembles a space that is readily described by the opening of an umbrella from a fully folded to a fully open form. In contrast to the cone-like topology, the set of conformations that sample this other region is better described using the term umbrella-like topology (Figure 6). In this case, the stem of the umbrella is occupied by the Neu5Acα2→6Gal-motif and the spokes of the umbrella (that are the flexible part causing the opening and closing) are occupied by the sugars at the reducing end of Gal.

The defining characteristic of glycan conformations that span a cone-like topology is that the majority of the interactions with the HA are made by a three-sugar (or trisaccharide) α2→3 (Neu5Acα2→3Galβ1→3GlcNAcβ) or α2→6 (Neu5Acα2→6Galβ1→4GlcNAcβ) motif. However, the glycan conformations that sample the umbrella-like topology are such that longer oligosaccharides (beyond a trisaccharide) make substantial contacts with the binding site on HA.

Using these shape-based definitions of the flexible glycan conformation, it has been shown that the umbrella-like topology is predominantly adopted by α2→6 glycans that possess at least four sugars including the Neu5Ac, for example, poly-lactosamine branches terminated by α2→6-linked Neu5Acα (long α2→6). The cone-like topology can instead be adopted by both α2→3 and α2→6 glycans. In the case of α2→6 glycans, those with a three-sugar α2→6 motif, such as N-linked glycans having single lactosamine branches terminated by α2→6 linked Neu5Ac (short α2→6), are more likely to adopt cone-like topology as compared with the long α2→6 branch.

Applying the topology-based description of glycan conformation to the analysis of HA-glycan cocrystal structures revealed that the umbrella-like topology was characteristic of long α2→6 motifs interacting with human adapted H1 and H3 HAs and the cone-like topology was characteristic of α2→3 glycans interacting with avian HA. Because the avian HA binding pocket is well-suited for maximum contacts with cone-like topology, it is probable that α2→6 glycans, particularly those with short α2→6 motifs, will adopt a cone-like topology in the glycan-binding site of avian HA. More recently, the role of glycan topology in molecular HA-glycan interactions was investigated using molecular dynamics (MD) simulations (Xu et al., 2009). Findings from this study support the notion that the umbrella-like topology is energetically preferred by glycans upon binding to human-adapted HA. The adoption of an umbrella-like topology by a long α2→6 motif in an avian H5 HA-binding site was associated with a high conformational entropy penalty in comparison with the adoption of this topology in the human-adapted H3 HA-binding site. These studies have further extended the definition of glycan topology using other parameters such as glycosidic torsion angles of sugars beyond the terminal sialic acid linkage and volume occupied by the different glycan topologies.

The above molecular and structural analyses of HA-glycan interactions offer a framework for the interpretation of glycan array data (Figure 4). Data-mining tools have been developed to identify patterns among binders (candidate ligands identified on the array) and nonbinders. These patterns are defined using glycan features that are abstracted from the glycan structures tested, for example as part of a glycan array. The combination of patterns from binders and nonbinders provides rules, or classifiers, that define glycan-binding motifs for a given HA. These classifiers can be used to corroborate the structural aspects of glycan-HA interactions obtained from analysis of the cocrystal structures. For instance, the α2→3 classifiers define motifs in terms of substitutions around a trisaccharide Neu5Acα2→3Galβ1→3GlcNAcβ1→6 whereas α2→6 classifiers define length-dependent motif that terminates in Neu5Acα2→6Galβ1→4GlcNAcβ1→6 structure.

Additionally, from an analysis of crystal structures and docking of glycan structures, it is possible to estimate ΔGmono, or the strength of a monovalent glycan-protein interaction (Figure 5). Several advances have improved the force fields used to model glycans and protein-glycan interactions (Case et al., 2005; Kirschner et al., 2008). These theoretical methods have become remarkably effective at predicting carbohydrate 3D structures, since the advent of accurate carbohydrate-specific force fields (Woods, 1998) and the advances in timescales that are accessible to MD simulations. All commonly employed biomolecular simulation packages (AMBER, CHARMM, GROMOS, etc.) now include high-quality carbohydrate force fields. There are currently several methods to calculate binding free energies (ΔG). The most accurate of these computes relative binding affinities for two related ligands by mutating one molecule onto another. Such “computational alchemy” can be performed by thermodynamic integration, but is limited to predicting relative binding energies for structurally similar ligands. Although the predicted interaction energies from these docking methods might only be qualitative, the resultant structures of the ligand-protein complexes provide additional information on the interactions.

In conclusion, the recent development of biochemical and biophysical tools truly enables the scientific community to rapidly and thoroughly address some fundamental aspects of describing glycan-protein interactions. By combining quantitative binding affinity information obtained from the array with molecular simulations, it is possible to correlate changes in the enthalpic and entropic contributions of the monovalent glycan-protein interaction with the quantitative differences in relative glycan-binding affinity. By identifying biochemical systems where multivalency can be taken into account (and hence define N in a rigorous manner), it is possible to define avidity and provide additional context to the glycan-protein interactions, which can be carried through to biological studies.

**Framework for Role Of HA-Glycan Interactions in Influenza A Virus Biology**

The intersection of the bottom-up and top-down approaches highlight the fact that for a virus to efficiently infect humans, it must bind to glycans that can adopt the umbrella-like topology (for example sialylated structures containing multiple lactosamine repeats, i.e., long α2→6). Thus, the structural topology of the glycan and not just the linkage appears to govern HA-binding specificity. This framework not only takes into account glycan structure in the context of protein binding, but also resolves the apparent inconsistencies between the binding and transmission data for SC18, NY18, AV18, and Tx91. SC18 and
**Glycan-Binding Properties of the 2009 A/H1N1 “Swine Flu” Hemagglutinin**

The integration of the top-down and bottom-up approaches was recently employed to provide key insights into the glycan-binding properties of HA from the 2009 swine flu H1N1 influenza A virus (henceforth referred to as 2009 A/H1N1) and their correlation with the transmissibility of these viruses in ferrets (Maines et al., 2009). One of the characteristics of a human-adapted virus is in its ability to achieve efficient transmission between humans via respiratory droplets (airborne) (Tumpey et al., 2007). Comparison of the respiratory droplet transmissibility of representative 2009 A/H1N1 viruses and a recent seasonal H1N1 influenza outbreak strain (A/Brisbane/59/07 or Brl/07) in ferrets showed that the 2009 A/H1N1 viruses transmitted less efficiently (Maines et al., 2009).

It has been demonstrated previously (Srinivasan et al., 2008; Tumpey et al., 2007) that the efficiency of respiratory droplet transmission in ferrets correlates with the α2-6 binding affinity of the viral HA. In fact, a single amino acid mutation in the HA of the efficiently transmitting SC18 virus led to a virus (NY18) that transmitted inefficiently. The α2-6 binding affinity of NY18 HA was substantially lower than that of SC18 HA (Srinivasan et al., 2008). In a bottom-up approach, using a quantitative dose-dependent direct glycan receptor-binding and human lung tissue-binding assays (Maines et al., 2009), the binding pattern of a representative 2009 A/H1N1 HA (A/California/04/09; or CA/04) was found to be similar to that of the pandemic SC18 H1N1 HA. Like SC18 HA, CA/04 showed substantial binding (in a dose-dependent fashion) to the single 8′SLN-LN α2-6 oligosaccharide motif. However, the binding affinity of CA/04 HA to 8′SLN-LN was substantially lower than that of SC18 HA. CA/04 HA also showed binding to the apical surface of human tracheal tissue section (in a sialic-acid-specific manner), and the binding pattern correlates with the predominant distribution of α2-6 sialylated glycans on the apical surface of the tracheal tissue (Chandrasekaran et al., 2008) and the α2-6 binding of CA/04 HA in the direct glycan receptor-binding assay.

Through a top-down approach, involving a combination of sequence analysis and homology-based structural modeling of CA/04 HA, it was predicted that this HA would have a substantially higher α2-6 as compared with α2-3 binding (Soudararajan et al., 2009). This is consistent with the observed binding pattern in the bottom-up approach. Building on this study, a more recent investigation offered a possible structural rationale for the observed reduced α2-6 binding affinity of CA/04 HA in comparison with that of SC18 HA (Maines et al., 2009). The receptor-binding site (RBS) of CA/04 HA shares many similar or analogous residues with that of SC18 HA, providing an explanation for the similarity in their binding pattern (Maines et al., 2009; Soudararajan et al., 2009). These analogous residues include Asp190 and Asp225, which are “signature” amino acids of human adapted H1N1 HAs that make optimal contacts with the α2-6 glycans (Gamblin et al., 2004). The main differences in the RBS between SC18 and CA/04 HA are at positions 145, 186, 189, 219, and 227. The structural analysis of the interresidue and residue-glycan contacts involving these residues points to some key differences between CA/04 and SC18 HA (Figure 7). Specifically, in SC18, a network of interactions involving hydrophobic contacts among Ala219, Ala227, and Pro186, and ionic contacts between Thr187, Thr189, and Asp190 are involved in positioning Asp190 for optimal contacts with α2-6 glycans. However, the unique combination of residues Ile219 and Glu227 in CA/04 HA had an disrupting effect on the network of interactions given that the contacts between these residues are neither hydrophobic nor ionic. This disrupting effect on the network could have a negative effect on the positioning of Asp190 and hence offering a possible structural explanation for the observed lower α2-6 binding affinity of CA/04 in comparison with that of SC18 HA. Taken together, the substantially lower α2-6 binding affinity of CA/04 HA as compared with that of SC18 HA correlates with its less efficient respiratory transmission in ferrets.

Figure 7. Structural Rationale for α2-6 Binding Affinity of 2009 A/H1N1 HA

Shown on the left are residues in positions 186, 187, 189, 209, and 227 of SC18 HA (cartoon representation in gray) in complex with an α2-6 oligosaccharide (Neu5Acα2-6Galβ1-4GlcNAcβ1-3Galβ1-4Glc) (Srinivasan, et al., 2008). The oligosaccharide is shown in stick representation colored orange (carbon atoms). The network of interactions is shown in dotted gray lines. Shown on the right are the same residue positions in structural complex of CA/04 HA (cartoon representation in violet) with the same α2-6 oligosaccharide (Maines, et al., 2009). The presence of the unique combination of Ile219 and Glu227 is not favorable for the optimal positioning of Asp190 (as seen on the left) for contacts with the α2-6 oligosaccharide. Lys222 is also shown as it is positioned to interact with Glu227.

Tx91, both efficient transmitters, bind with high affinity to glycans that adopt an umbrella-like topology, whereas NY18 and AV18, which are inefficient transmitters, bind with higher affinity to cone-like glycans, including glycans that either have α2-3 or α2-6 sialic acid. This framework demonstrates that description of HA-glycan interactions based on trans and cis conformations (adopted by α2-3 and α2-6 linkages, respectively) alone does not fully capture the structural features and conformational flexibility of the diverse sialylated glycans observed in human tissues. However, cone-like and umbrella-like classifications are able to fully capture the conformational plurality of these glycans and their binding to HA. Significantly, these classifications are able to distinguish the α2-3 and α2-6 binding of avian- and swine- adapted HAs from that of the α2-6 binding of human adapted HAs (Figure 7). Moreover, H1, H5, and H9 are of different structural clades and hence have distinct spatial arrangement of residues in the glycans-binding sites. The conformational flexibility of the α2-6 permits different types of umbrella-like topologies in the glycans-binding sites to accommodate the diverse constraints imposed by the different structural clades of HA (Figure 6). These observations suggest that it is challenging to design mutations in either H5 or H7 HA based...
simply on the characteristic changes in a few residues in H1 and H3 HAs that lead to their human adaptation. A set of mutations must instead occur that accommodate umbrella-like glycans in the context of the binding pocket, which is subtly distinct for each HA structural clade.

**Conclusion**

Lessons learned from the study of glycans, and particularly the HA example presented here, highlight: (1) the need to employ multiple biophysical, bioanalytical, and biological approaches to ultimately define the binding specificity of HA, and (2) the fact that because different cell types have a distinct cell surface glycan repertoire, studies employed using cell culture or any animal model must be carefully interpreted, because glycan structures present on cells in culture or in a given animal might not be reflective of the structures present on the surface of primary cells within the tissue. For example, the use of mice and Madin-Darby canine kidney cells to characterize virus infectivity (Hatakeyama et al., 2005; Stray et al., 2000) might not be ideal owing to differences in the glycan receptors present in them as compared with physiologically relevant cells.

The development of a systematic understanding of the receptor binding specificity of the HA from various influenza strains is anticipated to help address a number of critical questions, including what defines an avian influenza strain versus one that has become humanized, and what mutations in HA enable the conversion of a strain from an avian virus to one capable of efficient human-to-human transmission. Understanding the receptor specificity of HA as well as the set of mutations that allow a virus to gain the ability to recognize human-like glycans of the upper respiratory tract provides insight into the epidemic and pandemic potential of various strains. Addressing these questions is not only of great scientific significance, but also is immediately pertinent in light of the rise of H5N1 and other strains is anticipated to help address a number of critical questions.

**ACKNOWLEDGMENTS**

The authors would like to acknowledge support from National Institute of General Medical Sciences of the National Institutes of Health (GM 57073 and U54 GM62116 to RS) and the Singapore–Massachusetts Institute of Technology Alliance for Research and Technology (SMART). The authors thank V. Sasisekharan for critical reading of the manuscript and constructive discussion.

**REFERENCES**


