The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.
Structural Determinants for Naturally Evolving H5N1 Hemagglutinin to Switch Its Receptor Specificity

Kannan Tharakaraman,1 Rahul Raman,1 Karthik Viswanathan,1 Nathan W. Stebbins,1 Akila Jayaraman,1 Arvind Krishnan,1 V. Sasishekaran,1 and Ram Sasishekaran1,*
1Department of Biological Engineering, Koch Institute of Integrative Cancer Research, Infectious Diseases Interdisciplinary Research Group, Singapore-MIT Alliance for Research and Technology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139, USA
*Correspondence: rams@mit.edu
http://dx.doi.org/10.1016/j.cell.2013.05.035

SUMMARY

Of the factors governing human-to-human transmission of the highly pathogenic avian-adapted H5N1 virus, the most critical is the acquisition of mutations on the viral hemagglutinin (HA) to “quantitatively switch” its binding from avian to human glycan receptors. Here, we describe a structural framework that outlines a necessary set of H5 HA receptor-binding site (RBS) features required for the H5 HA to quantitatively switch its preference to human receptors. We show here that the same RBS HA mutations that lead to aerosol transmission of A/Vietnam/1203/04 and A/Indonesia/5/05 viruses, when introduced in currently circulating H5N1, do not lead to a quantitative switch in receptor preference. We demonstrate that HAs from circulating clades require as few as a single base pair mutation to quantitatively switch their binding to human receptors. The mutations identified by this study can be used to monitor the emergence of strains having human-to-human transmission potential.

INTRODUCTION

The highly pathogenic H5N1 influenza A virus subtype poses a global health concern. This is evident from it having already led to several localized outbreaks in humans with a high case fatality ratio (~60%) since 2003 (Guan et al., 2009; Neumann et al., 2010). The H5N1 subtype, however, has not yet adapted to the human host and established sustained human-to-human transmission via respiratory droplets (or aerosol transmission). One of the key factors governing adaptation of virus to human host is the glycan receptor-binding specificity of its hemagglutinin (HA). The HA from avian subtypes typically binds to α2→3 sialylated glycans (or avian receptors) (Ge and Wang, 2011). A hallmark feature of human-adapted subtypes such as H1N1, H2N2, and H3N2 is the “quantitative switch” in their binding preference to α2→6 sialylated glycan receptors (or human receptors), which is defined by high relative binding affinity to human receptors over avian receptors. This quantitative switch in receptor specificity has been shown to correlate with the respiratory droplet transmissibility of the pandemic H1N1 and H2N2 viruses in ferrets (Pappas et al., 2010; Srinivasan et al., 2008; Tumpey et al., 2007; Viswanathan et al., 2010). A necessary determinant of human adaptation of avian-adapted H5N1 subtype therefore is for its HA to acquire mutations that quantitatively switch its binding preference to human receptors (Ge and Wang, 2011; Shriner et al., 2009). Structure and receptor complexes of the HA from different subtypes (H1, H2, H3, and H5) have shed light on the nature of mutations preferred by avian and human viruses (Gamblin et al., 2004; Liu et al., 2009; Ha et al., 2001; Lin et al., 2012). These studies showed that the residues at 226 and 228 are critical determinants of the receptor-binding specificity of H2 and H3 HA, with human viruses favoring L226 and S228 and avian viruses favoring Q226 and G228 (Lin et al., 2012; Lin et al., 2009). On the other hand, for H1 viruses, 190 and 225 are critical determinants of the receptor-binding specificity, with human viruses favoring D190 and D225 and avian viruses favoring E190 and G225 (Gamblin et al., 2004). Identifying mutations that switch glycan receptor-binding preference of H5 HA has been the focus of several previous studies (Chandrasekaran et al., 2008; Gambaryan et al., 2008; Stevens et al., 2006, 2008; Wang et al., 2009; Watanabe et al., 2012; Yamada et al., 2006). Some of these studies include analyses of glycan receptor binding of H5 HAs with natural variations in the receptor-binding site (RBS) (Yamada et al., 2006). Other studies have mutated H5 HA to include either the “hallmark” changes for human adaptation of H2/H3 HA (Q226L and G228S or LS) and/or H1 HA (E190D, G225D, or DD). More recently, two studies by Imai et al. (2012) and Herfst et al. (2012) demonstrated that specific sets of mutations in HA from A/Vietnam/1203/04 (Viet04) and A/Indonesia/5/05 (Ind05) viruses, respectively, confer respiratory droplet viral transmission in ferrets to the viruses possessing these mutant H5 HAs. From these studies, it is evident that differences in genetic background (using natural H5N1 isolate versus laboratory reassorted strain) and selection pressure strategies give rise to distinct sets of amino acid changes in Viet04 and Ind05 HAs that are associated with aerosol transmission in ferrets. None of the wild-type
(WT) natural variants or mutant H5 HAs from the aforementioned studies, however, have shown a quantitative switch in binding to human receptor in a fashion characteristic of “pandemic” strain HAs (such as 1918 H1N1, 1958 H2N2, and 2009 H1N1) (Figure S1 available online).

Sequences of H5 HA isolated after 2006 have diverged considerably from the prototypic strains Viet04 and Ind05. This sequence divergence has critical implications for identifying amino acid changes in the RBS required to quantitatively switch the binding preference of H5 HA to human receptors. In this scenario, an important unanswered question is how current H5 HA would quantitatively switch to human receptor binding in the context of other molecular changes in its RBS due to sequence divergence from prototypic strains such as Viet04 and Ind05 (Watanabe et al., 2011, 2012).

In this study, we have developed a distinct structural framework to systematically analyze RBS of H5 HA from the perspective of structural topology of its glycan receptor, residues that interact with this receptor, and their interresidue interactions in the RBS. Using this framework, we compared the RBS of H5 HA with that H2 HA—its phylogenetically closest neighbor—to define molecular features that are critical for quantitative switching H5 HA binding to human receptors. Analysis of sequences for naturally evolving H5 HAs show that the different H5 clades have evolved to acquire distinct features that make them closer to human adaptation. We demonstrate that a subset of rapidly evolving and currently circulating H5 clades require as few as a single base pair change to quantitatively switch to human receptor binding. However, the acquisition of these distinct features in the various clades appears to be somewhat nuanced. We show here that amino acid changes that led to aerosol transmission of Viet04 and Ind05, when introduced in the HA of some of the currently circulating H5 clades, did not quantitatively switch these H5 HA binding to human receptors. Our study highlights the critical need to investigate RBS amino acid sequence divergence of H5 HA in the context of RBS molecular features to delineate H5N1 human adaptation.

RESULTS

Defining Molecular Features for H5 HA RBS

We began by systematically analyzing the RBS of H5 HA from the perspective of structural topology of its natural avian receptor and that of the amino acids in the RBS. Previously, we had developed a framework to distinguish binding of HA to avian and human receptors on the basis of the three-dimensional structural topology of these receptors (Chandrasekaran et al., 2008). When bound to HA, the avian receptor sampled a conformational space that resembles a cone (thus the term cone-like topology was used to describe this receptor). The majority of contacts of H5 HA (using Viet04 crystal structure [Stevens et al., 2006, 2008]) with avian receptor adopting a cone-like topology involve the Neu5Acα2→6Galβ1→ motif. The key amino acids in the H5 HA RBS involved in this interaction predominantly lie in the base of the RBS, involving residues Ser-136 in 130 loop, Trp-153, Ile-155 in the 150 loop, Lys-222, and Gln-226 in the 220 loop with specific additional contacts from Glu-190, Lys-193, and Leu-194 in the 190 helix at the top of the RBS (Figure 1).

There are several human-adapted HAs, including various seasonal and pandemic strains from H1N1, H3N2, and the pandemic H2N2 subtypes. Based on the phylogenetic “closeness” of H5 HA to H2 HA (Figure S2), we selected human-adapted H2N2 HA (A/Albany/6/58 or Alb58) bound to human receptor to contrast with the structural model of H5 HA bound to avian receptor (Stevens et al., 2006, 2008). We chose Alb58 as a representative H2N2 strain because it is a prototypic pandemic strain, and we have already extensively characterized its quantitative glycan receptor binding and phenotypic properties (such as aerosol transmissibility) (Pappas et al., 2010; Viswanathan et al., 2010). As the X-ray crystal structure of Alb58 HA is not available, we constructed a homology-based model for this HA using the template crystal structure of another human-adapted H2 HA (A/Singapore/1/57) (cocrystallized with human receptor), which has high sequence identity to Alb58 HA (Liu et al., 2009).

The human receptor bound to HA samples a larger conformational space that resembles a fully closed to fully open umbrella (and thus we have used the term umbrella-like topology to define this receptor). There are two regions in the umbrella-like topology of the human receptor: the base region composed of Neu5Acα2→6Galβ1→ motif and an extension region comprising sugar residues beyond this motif (typically GlcNAcβ1→3Galβ1→1). These two regions span a wider range of interacting amino acids in the H2 HA RBS. A comparison of H5 HA bound to avian receptor in cone-like topology and H2 HA bound to human receptor in umbrella-like topology showed four key differences (Figure 1). First, the composition of the 130 loop of H2 HA is different from H5 HA in that there is a deletion in this loop (Figure S3). The deletion in the 130 loop in H5 HA relative to H2 HA critically influences the 130 loop. Second, amino acids in the “base” of the RBS (such as those in the 130 loop at positions 136–138 and the 220 loop at positions 219–228) are different. Third, the “top” of the RBS primarily comprising the “190 helix” (residues 188–196) that interacts with the extension region of human receptor in H2 HA is different (specifically at positions 188, 189, 192, and 193). Fourth, position 158 is glycosylated in H5 HA, but not in H2 HA. Glycosylation at this site could potentially interfere with the extension region of human receptor (Stevens et al., 2008).

To determine what mutations are needed to overcome the above differences and so that H5 HA can switch its specificity to human receptor, we developed a metric (RBS network or RBSN) to capture the network of interactions between the critical residues in the RBS and other residues in their close spatial environment that make contact with the glycan (see Experimental Procedures). The higher the network score of an amino acid within the RBS, the more structurally constrained it is to be mutated. For example, residues that make critical contacts with sialic acid such as Phe-98, Trp-153, and His-183 are highly networked (RBSN scores > 0.7) and, hence, are less constrained to mutate (Figure S4). On the other hand, residues that are at the interface of the RBS and antigenic sites such as in the 130 loop, 190 helix, and 220 loop are poorly to moderately networked (RBSN scores < 0.15) and can readily undergo mutations as a part of antigenic drift. Thus, we define the term—molecular feature (for each of the four differences) —that incorporates glycan topology, HA residues involved in the binding, and their interresidue interaction network in the RBS. We demonstrate
that such an approach provides a robust framework to investigate amino acid mutations (and hence matching features) that quantitatively switch binding of H5 HA to human receptors in a manner similar to what has been observed for pandemic HAs.

Amino acid changes related to feature 1 involve altering the length of the 130 loop specifically by introducing a deletion, which, in turn, affects the RBSN diagram involving 131, 133, and 155 positions. Residues at positions 131 and 133 had low RBSN scores (<0.04) in H5 HA and therefore could be readily mutated in context of the 130 loop deletion such that it contributes to human receptor specificity. Changing the base of the RBS (feature 2), which plays a critical role in governing glycan receptor specificity, involves alteration of a combination of residue positions in the 130 loop and 220 loop. In H5 HA, Gln-226 plays a critical role in contacts with Neu5Ac$_2$$_3$Gal$_2$$_3$ motif of avian receptor, and Ser-137 and Gln-226 are involved in the interresidue interaction network. Conversely, in H2 HA, the corresponding Leu-226 and Arg-137 are not related. Arg-137 and Ser-228 in H2 HA provide additional stabilizing contacts with sialic acid. Therefore, one way to match feature 2 (for human receptor binding) involves changing residues at 137 and 226 positions in H5 HA. Residue position at 137 is readily mutable given its low RBSN score in H2 and H5 HA (<0.01). However, residue at 226 has a much higher RBSN score in H2 and H5 HA (>0.25). Making changes to this residue therefore would also involve making other changes—specifically, changing Gly-228/Asp in addition to Ser-137/Arg mutation. Although Gln-226 → Leu mutation governs switch in contacts from Neu5Ac$_2$ → 3Gal → to Neu5Ac$_2$ → 6Gal → motif, Ser-137 → Arg and Gly-228 → Ser mutation provides additional stabilization to the 130 and 220 loop at the base of the H5 RBS from the standpoint of interamino acid networking and improved contacts with glycan receptor (Stevens et al., 2006, 2008). This stabilization can also be accomplished by mutation Asn-224 to Lys or Arg (naturally observed in some pandemics), as this would enhance its interamino acid interaction network with Asp-96, Leu-97, and Pro-99.

Figure 1. Comparing Molecular Features in the RBS of H5 and H2 HA

The RBS of H5 and its phylogenetically closest H2 HA is rendered as a cartoon. The key residues are labeled, and their side chains are shown in stick representation. The four key features are indicated based on coloring the carbon atom in different colors. The 130 loop (feature 1) is shown in cyan. The base of the RBS (feature 2), including residue positions 136–138, 153, and 221–228, is shown in gray. The side chain of Arg in the 224 position is shown at 50% transparency. The top of the RBS (feature 3), including the 190 helix and residue positions 155, 156, and 219, is shown in green. The glycosylation sequon at the 158 position (feature 4) in H5 HA is shown in orange. The RBSN diagram is shown adjacent to the residue positions in that network. The circular nodes are colored according to their RBSN score (pink representing a low score to bright red representing a high score) and their connectivity to other nodes. The asterisk next to residue positions indicates that these positions belong to the adjacent HA1 domain in the HA trimer. The HA structure images were generated using Pymol (http://www.pymol.org/).

See also Figures S2, S3, and S4.
(Figure 1). Therefore, feature 2 can also be matched by fewer mutations at 224 and 226 positions. The RBSN diagram of the residue at the 221 position in H5 HA is identical to that in H2 HA, although this position has a Ser in H5 HA and a Pro in H2 HA (Figure 1). It is likely for the Pro to govern the conformation and side-chain orientation of the adjacent Lys-222 residue, which plays a key role in making contacts with the human receptor. Therefore, changing Ser-221 → Pro in H5 HA would permit more optimal conformation of the 220 loop for contact with the human receptor. On the other hand, mutations at positions 188, 189, 192, and 193 and the RBSN diagrams depicting their interaction networks (Figure 1) will impact feature 3. The residues Ala-188, Ala-189, and Thr-192 in H5 HA do not have any interresidue contacts with other residues in the RBS. However, residues Glu-188, Thr-189, and Arg-192 are involved in multiple interaction networks. Therefore, amino acid changes in positions 188, 189, 192, and 193 (hence altering feature 3) of H5 HA RBS would also impinge on its human receptor-binding property. Given that the RBSN scores of all these residue positions are low (<0.1) in H5 HA, they are readily mutable. Finally, removal of glycosylation sequon at position 158 relates to feature 4. This could be accomplished by mutating Asn-158 to a residue such as Asp or by mutating Thr-160 to Ala.

Among these four distinct features, feature 2 critically influences the base of the RBS, which plays an important role in distinguishing contacts between H5 and Neu5Ac2→3Gal motif (avian receptor) and H2 HA and Neu5Ac2→6Gal motif (human receptor). The difference in the 130 loop length (feature 1) would potentially affect the base of the RBS, which, in turn, is critical for glycan-receptor specificity. Changes related to feature 2 to achieve the necessary glycan receptor specificity therefore would critically depend on the 130 loop length (feature 1). Conversely, feature 3, which involves the 190 helix, plays an important role in contacts with avian receptor in H5 (specifically Lys or Arg at 193 position) and in contacts with extension region of human receptor. Therefore, amino acid changes in H5 HA related to feature 3 would reduce binding to avian receptor and enhance contact with extension region of human receptor. Feature 4 relates to the state of glycosylation at position 158, which impinges on the interaction with the extension region of human receptor (Stevens et al., 2008).

Analyzing Natural Sequence Evolution of H5 HA from the Perspective of RBS Features

Having described and contrasted the H5 HA RBS features (henceforth referred to as RBS features) in the context of H2 HA RBS, we then investigated the acquisition of these features in the context of sequence evolution of H5 HA over time instead of searching for specific hallmark human-adaptive mutations described in previous studies (Maines et al., 2011; Neumann et al., 2012; Russell et al., 2012; Stevens et al., 2008). A phylogeny map of A/H5N1 HA sequences was constructed, and the branches were color-coded based on the features present (Figure 2). Feature 4 has been observed in almost all of the clades; however, amino acid changes characteristic of features 1, 2, and 3 were observed more recently. Many of the clades, including currently circulating clade 1, clade 2.2, clade 2.2.1, and clade 7, have acquired amino acid changes characteristic of one or two of features 1, 3, and 4. However, only a subset of the rapidly evolving and currently circulating clades 2.2.1 and 7 have acquired amino acid changes— which are critical features of the RBS base—to match feature 1 and/or part of feature 2. In the context of the key structural features of HA RBS, the deletion in the 130 loop with a concurrent loss of glycosylation (features 1 and 4) in the same HA was the most critical change observed in the evolution of 2.2.1 H5 HA since 2007. A recent study (Russell et al., 2012) compared the sequence evolution of H5N1 simply in the context of hallmark changes reported by the two H5N1 transmission studies involving Viet04 (Imai et al., 2012) and Ind05 (Herfst et al., 2012) and indicated that clade 2.2.1 viruses were most similar to human-transmissible virus. In this earlier analysis, clade 7 does not appear to be significant. On the other hand, apart from a subset of clade 2.2.1 strains that have acquired one or more RBS features, many clade 2.2.1 strains have not yet acquired any of the RBS features, hence they are almost indistinguishable from noncirculating clades (3, 5, 6, 8, and 9). Taken together, the above findings suggest that acquisition of these distinct features in the various clades appears to be somewhat nuanced and that subsets of the currently circulating H5 HA strains have not only diverged considerably from older human isolates (such as Viet04) but have also acquired the key molecular RBS features necessary for human receptor preference.

Design and Validation of Amino Acid Changes Required to Quantitatively Switch Receptor Specificity of Currently Circulating H5 HA

Based on our analyses of the rapidly evolving and currently circulating H5 HAs, we chose representative examples from a subset of clades 2.2.1 and 7, which have already acquired amino acid changes to match one or two of the RBS features, to validate the framework and predictions. We chose A/chicken/Vietnam/NCVD-093/08 (ckViet08; clade 7 HA that acquired changes in feature 3), A/Egypt/N03450/2009 (Egy09; clade 2.2.1 HA that acquired changes in features 1 and 4), and A/duck/Egypt/10185SS/2010 (dkEgy10; another clade 2.2.1 HA that acquired changes in features 1, 4, and part of feature 2 based on Asn224→Lys natural mutation). On these HAs from these distinct recent H5 isolates, we introduced amino acid changes to match one or more of the remaining features. We recombinantly expressed WT and mutant forms of these HAs and

Figure 2. Phylogenetic Tree of Representative A/H5N1 HA1 Protein Sequences Showing Clustering of Sequences Based on Antigenic Clades

The sequences are color-coded by the features observed in the H1A domain. Scale bar represents 0.001 amino acid substitutions per site. Feature 1 has evolved in clade-2.2, -2.2.1, and -7 strains after 2006, whereas feature 2 has evolved in clade-1 and -2.2.1 after 2007. Feature 3 has evolved in clade-2.2 and -7 strains. Some of the currently circulating clades (1, 2.1.3, 2.2, 2.2.1, 2.3.2, 2.3.4, and 7) have acquired mutations to match multiple RBS features and hence appear to be closer to human adaptation. Among these H5 HAs, the relative order of dominant circulating clades found to have acquired multiple features includes clade-2.2.1, -2.2, -1, and -7. A subset of clade 2.2.1 and clade 7 has acquired amino acid changes in the RBS base toward matching feature 1 and/or feature 2. See also Table S1 and Figure S3.
evaluated their glycan-binding properties in dose-dependent direct glycan array assay. A summary of WT and mutant HAs with their corresponding RBS features and glycan-binding properties is shown in Table 1. The details of the amino acid changes to match RBS features are summarized in Table 1. The dose-dependent glycan array data for the WT and mutant H5 HAs are shown in Figure 3.

From Table 1 and Figure 3, it is evident that avian receptor binding of WT H5 HAs that have already acquired one or more of the RBS features is not as high as (specifically binding to 3′ sialyl-lactosamine (SLN) what we have observed for other avian-adapted HAs (Chandrasekaran et al., 2008; Viswanathan et al., 2010). In the case of ckViet08 (Figure 3A), the natural acquisition of changes in the 190 helix (feature 3), especially Met-193, is consistent with lowering of avian receptor binding because Lys or Arg at 193 in other H5 HAs (which have not acquired feature 3) make optimal contact with the avian receptors. In the cases of Egy09 and dkEgy10 (Figures 3C and 3E), the 130 loop deletion (feature 1) does appear to also have some detrimental effects on avian receptor binding.

We next sought to understand the relationship between molecular features in recent H5 HAs and the hallmark amino acid changes in Viet04 and Ind05 HA that conferred respiratory droplet transmission to viruses having these mutant HAs (Herfst et al., 2012; Imai et al., 2012). The LS mutation and loss of glycosylation sequn at the 158 position were the RBS mutations previously reported for Ind05 (Herfst et al., 2012). The LS amino acid mutations only partially matched RBS base feature 2. Introducing the LS mutation with loss of 158 glycosylation sequn on ckViet08 (V4.2) showed some increased human receptor binding while retaining most of the avian receptor binding and, therefore, did not quantitatively switch this mutant HA binding to human receptors. Introducing the 130 loop deletion and amino acid changes in addition to LS and loss of glycosylation at 158 (see V4.5 in Table S1) completely switched binding to the human receptor. Therefore, the LS mutation alone is not sufficient to completely match (feature 2). The Asn-224→Lys/Gln226→Leu and loss of the 158-glycosylation sequn were the RBS mutations previously reported for Viet04 (Imai et al., 2012). As an example, introducing these mutations on ckViet08 (V4.3) wholly abolished binding to both avian and human receptors. However, introducing the deletion in the 130 loop of V4.3 resulted in a mutant (V4.4) that quantitatively switched binding to human receptors (Figure 3B) (apparent binding affinity constant [Kd] for 6′ SLN-LN ~100 pM), highlighting the critical importance of the RBS base (feature 2) and the 130 loop deletion (feature 1) in modulating human receptor-binding specificity. Jointly, these results clearly demonstrate that the same amino acid mutations that led to aerosol transmission of Viet04 and Ind05 are not sufficient to quantitatively switch current circulating H5 HA binding to human receptors.

Egy09 already acquired amino acid changes to match features 1 and 4. Similar to what was observed with ckViet08 V4.2 mutant, introducing the LS mutations in the RBS base (partially matches feature 2) even with the 130 loop deletion in Egy09 (E4.1) did not quantitatively switch its binding to human receptor. Instead, matching feature 2 by introducing the Asn-224→Lys/Gln-226→Leu mutations on Egy09 (E4.3) quantitatively switched its binding to the human receptor (Kd’ ~50 pM) (Figures 3C and 3D). The dkEgy10 is one of a kind because it has evolved to be the closest matching RBS features 1, 2, and 4. The glycan receptor binding of dkEgy10 HA (Figure 3E) shows that the WT HA still predominantly binds to avian receptors (3′ SLN-LN and 3′ SLN-LN-LN with minimal binding to 3′ SLN and human receptors). The predominant avian receptor-binding property of dkEgy10 can be attributed to the 226 position, which still has a Gin (and not Leu) that is needed to match feature 2. In fact, introducing this single Gln226→Leu amino acid mutation (which involves a single base pair mutation) on dkEgy10 (E5.1) quantitatively switched its binding to human receptors (Kd’ ~100 pM) (Figure 3F). The above mutant H5 HAs that show a switch in receptor preference demonstrate the necessary apparent binding affinity to human receptor (similar to the 2009 H1N1 pandemic HA) and hence pass the threshold for potential aerosol viral transmission (Figure S1).

To extend the quantitative characterization of glycan receptor specificity to binding to physiological glycan receptors in human respiratory tissues, we analyzed the binding of dkEgy10 WT and

---

**Table 1. Summary of RBS Features and Glycan-Binding Properties of WT and Mutant H5 HAs**

<table>
<thead>
<tr>
<th>Human Receptor →</th>
<th>Base (Neu5Ac(\beta2\rightarrow6)Gal(\alpha1\rightarrow))</th>
<th>Extension (→4GlcNAc(\beta1\rightarrow3)Gal(\beta1\rightarrow4\rightarrow))</th>
<th>Glycan-Binding Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>ckViet08 (V4.0)</td>
<td>F1 (Δ130 loop) F2 (RBS base)</td>
<td>K192/M193</td>
<td>++ +</td>
</tr>
<tr>
<td>V4.2</td>
<td>Q226L/G228S</td>
<td>K192/M193</td>
<td>T160A ++ +</td>
</tr>
<tr>
<td>V4.3</td>
<td>N224K/Q226L</td>
<td>K192/M193</td>
<td>T160A</td>
</tr>
<tr>
<td>V4.4</td>
<td>L133Δ</td>
<td>K192/M193</td>
<td>N158D + +++</td>
</tr>
<tr>
<td>Egy09 (E4.0)</td>
<td>Δ133</td>
<td>A160</td>
<td>+++</td>
</tr>
<tr>
<td>E4.3</td>
<td>Δ133</td>
<td>N224K/Q226L</td>
<td>A160 + +++</td>
</tr>
<tr>
<td>dkEgy10 (E5.0)</td>
<td>Δ133</td>
<td>K224</td>
<td>D158/A160 ++</td>
</tr>
<tr>
<td>E5.1</td>
<td>Δ133</td>
<td>K224/Q226L</td>
<td>D158/A160 +++</td>
</tr>
</tbody>
</table>

The key amino acids in WT H5 that contribute to acquisition of the corresponding feature are indicated by WT amino acid (one letter code) followed by the HA position, whereas mutations introduced to match the features are indicated by WT amino acid followed by the HA position and the mutant amino acid. The blank feature columns indicate absence of H2-like RBS feature in the H5 HAs. The apparent avian and human receptor-binding affinities are indicated in the +2→3 and +2→6 columns in which highest = ++++, high = ++, moderate = +, and low = +, and no observable binding = blank column. See also Table S1.
E5.1 mutant on human tracheal and alveolar tissues. Studies previously have demonstrated that the apical surface of human tracheal tissue predominantly expresses human receptors, and the human alveolar tissue predominantly expresses avian receptors (Chandrasekaran et al., 2008; Mansfield, 2007; Matrosovich et al., 2004; Shinya et al., 2006). Consistent with the quantitative switch to human receptor binding, E5.1 mutant showed the expected staining of the apical surface of the human tracheal tissue section, whereas the WT dkEgy10 showed extensive staining of the human alveolar section (Figure 4A).

The above results together underscore the importance of delineating key structural RBS features in H5 in conferring the quantitative switch in binding to human receptors (Figure 4B). The base of the RBS captured by feature 2, which plays a critical role by making contacts with the terminal sialic acid linkage, appears to be a key determinant. Our data show (and are consistent with RBSN) that a combination of two amino acid changes, Asn-224/Lys and Gln-226/Leu in the RBS base, is able to match feature 2 effectively in specific H5 clades when compared to the LS combination in which additional mutations match the same feature. Our results also support a critical role for the “130 loop length” in augmenting the RBS base for optimal contacts with the human receptor because changes to the RBS base alone with Asn-224→Lys/Gln-226→Leu completely abolished binding of ckViet08 V4.3 mutant. Many strains in the H5 clade 2.2.1 do not possess the “130 loop” deletion (feature 1). In these strains, matching the critical feature 2, with feature 4 already matched, does not confer a switch in H5 receptor specificity. A representative clade 2.2.1 example (A/Egypt/2786-NAMRu3/06 [Egy06] and its mutant [E3.1]) is shown in Table S1. These strains are further away from human adaptation, and it cannot be generalized, therefore, that all clade 2.2.1 strains are closest to human adaptation. The 130 loop deletion has been naturally acquired by a subset of clade 2.2.1 HA and shows a lowering in binding to avian receptors as observed in both Egy09 and dkEgy10 strains. Our structural analysis and data also indicate that involvement of features 3 and 4—which impact contacts with the extension region of the human receptor—in human receptor switch depends on clade-specific H5 HA RBS. Consequently, in the context of the current HA sequence evolution of rapidly evolving H5N1, features 1 and 2 appear to be necessary and additionally appear to be sufficient to match either feature 3 or 4 for human adaptation of distinct H5 HA clades.

**DISCUSSION**

In summary, we have developed a distinct approach to define the molecular features that characterize RBS of HA and have used these features to compare RBS of H5 HA with that of its closest phylogenetic neighbor—pandemic H2 HA. This network approach permits us to understand how amino acid changes (resulting from the extensive sequence divergence of H5 HA as a part of its natural evolution) in the RBS modulate glycan-binding specificity. Using this approach, we demonstrate that some of the recent H5 HAAs require as few as one or two amino acid mutations to quantitatively switch their receptor preference. Importantly, our approach permitted us to scan the RBS of HAs from various H5 isolates from 2007 for the natural
acquisition of necessary RBS features for switch in receptor specificity. Different phylogenetic clades of H5 have important nuances to their RBS structural features, which, in specific instances, dynamically change with the natural sequence evolution. Some features that are present in the parent clade might be lost as this clade diversifies, or this diversification could lead to the addition of features that are critical for human adaptation of H5 HA. Even within a clade, not all sequences have an identical set of structural features. For example, in the case of 2.2.1 HAs, only ~87% have loss of 158 glycosylation. Also, not all 2.2.1 HAs have a deletion in the 130 loop. Data show that the same amino acid mutations that lead to aerosol transmission of Viet04 and Ind05, when introduced in more recent H5 HA, give distinct results (depending on the H5 clade) and, importantly, do not quantitatively switch any of the mutant HAs binding to human receptors. These residues alone cannot be used as reference points to analyze the switch in receptor specificity of currently circulating and evolving H5N1 strains (Russell et al., 2012).

A question arises as to the relationship between RBS of H5 and H1 HA—its second-nearest human-adapted phylogenetic neighbor in group 1 HAs (Figure S2). We have performed a similar structural analysis and have demonstrated that amino acid changes in H5 HA to match features with RBS of pandemic H1N1 (A/South Carolina/1/18) HA led to a quantitative switch in binding to human receptors (Figure S5). However, the current natural evolution of H5 HA has led to acquisition of RBS features along an H2-like path. Human adaptation of H5 HA is one of the key factors involved in the adaptation of the H5N1 virus to the human host for a sustained circulation. Other hallmark factors and genetic signatures such as Glu627→Lys in PB2, PB1-F2 length, activity, and stalk length of neuraminidase have been associated with increased pathogenicity and efficient transmission in humans and ferret animal models. However, it is unclear as to how many of these additional hallmark factors are required for the human adaptation of this subtype. Nevertheless, from a surveillance standpoint, it is critical to investigate amino acid sequence divergence of H5 HA in the context of RBS molecular features in addition to monitoring changes in other viral genes to delineate H5N1 human adaptation.

**EXPERIMENTAL PROCEDURES**

**Cloning, Baculovirus Synthesis, and Mammalian Expression and Purification of HA**

H5 WT and mutant HA sequences were codon optimized for insect cell expression and were synthesized at DNA2.0 (Menlo Park, CA). The synthesized genes were then subcloned into pAcGP67A plasmid, and baculoviruses were created using Baculogold system (BD Biosciences, San Jose, CA) according to the manufacturer’s instructions. The recombinant baculoviruses were then used to infect suspension cultures of Sf9 cells cultured in BD Baculogold Max-XP SFM (BD Biosciences, San Jose, CA). The infection was monitored, and the conditioned media were harvested 3–4 days postinfection. The soluble HA from the harvested conditioned media was purified using Nickel affinity chromatography (HiTrap HP columns, GE Healthcare, Piscataway, NJ). Eluting fractions containing HA were pooled, concentrated, and buffer exchanged.

---

![Figure 4. Physiological Glycan Receptor Binding and Summary of RBS Features in Current H5 HAs](image-url)

(A and B) In (A), the left panel shows staining of human alveolar section by dkEgy10. The middle panel shows staining of human tracheal tissue section by the E5.1 mutant. The right panel shows staining of human tracheal tissue section by CA04. For all the tissue sections, the HA staining is shown in green against propidium iodide shown in red. Apical surface of trachea is indicated by a white arrow (B). Surface rendering of H5 RBS with the region corresponding to features 1–4 is colored cyan, gray, green, and orange, in that order. The human receptor is shown as a stick. Features 1 and 2 are colored dark red to indicate their necessary requirement for human adaptation of H5 HA in the context of its current sequence evolution. See also Table S1 and Figure S3.
into 1 × PBS pH 8.0 (GIBCO) using 100 kDa MWCO spin columns (Millipore, Billerica, MA). The purified protein was quantified using BCA method (Pierce).

The gene was codon optimized for mammalian expression, synthesized (DNA2.0, Menlo Park, CA), and subcloned into modified pcDNA3.3 vector for expression under CMV promoter. Recombinant expression of HA was carried out in HEK293-F FreeStyle suspension cells (Invitrogen, Carlsbad, CA) cultured in 293-F FreeStyle Expression Medium (Invitrogen, Carlsbad, CA) maintained at 37 °C, 80% humidity, and 8% CO2. Cells were transfected with Poly-ethylene-imine max (PEI-MAX, Polysciences, Warrington, PA) with the HA plasmid and were harvested 7 days postinfection. The supernatant was collected by centrifugation, filtered through a 0.45 μm filter system (Nalgene, Rochester, NY), and supplemented with 1:1,000 diluted protease inhibitor cocktail (Calbiochem filtration) and supplemented with 1:1,000 diluted protease inhibitor cocktails (EMD Millipore, Billerica, MA). HA was purified from the supernatant using His-trap columns (GE Healthcare) on an AKTA Purifier FPLC system. Eluting fractions containing HAs were pooled, concentrated, and buffer exchanged into 1× PBS pH 7.4 using 100 kDa MWCO spin columns (Millipore, Billerica, MA). The purified protein was quantified using BCA method (Pierce, Rockford, IL).

Both expression systems were used in this study. Importantly, no differences were observed in the glycan-binding properties of the HA derived from baculovirus when compared to that of the material derived from mammalian expression.

Homology Modeling of HA

A structural model of Ab58 HA trimer was built using the MODELER homology modeling software. To build the model, the solved crystal structure of A/Singapore/1/57 hemagglutinin with human receptor (PDB: 2WRT), which has 99% sequence identity in HAH to Ab58, was used as a template. During modeling, the ligand (human receptor) was copied from the template structure into the model structure. The final model was minimized to release internal constraints.

Dose-Dependent Direct Binding of WT and Mutant HA

To investigate the multivalent HA-glycan interactions, a streptavidin plate array comprising representative biotinylated s2→3 and s2→6 sialylated glycans was used. 3′ SLN, 3′ SLN-LN, and 3′ SLN-LN-LN are representative avian receptors. 6′ SLN and 6′ SLN-LN are representative human receptors. LN corresponds to lactosamine (Galα1→4GalNAc), and 3′ SLN and 6′ SLN, respectively, correspond to Neu5Acα2→3 and Neu5Acα2→6 linked to LN. The biotinylated glycans were obtained from the Consortium of Functional Glycomics through the resource request program. We have chosen a defined set of representative avian and human receptors given that the focus of our experimental studies is to quantitatively characterize relative binding affinity to human versus avian receptors and not to define specificity on the basis of number of human versus avian receptors using a larger array of glycans. The quantitative affinity defined using our defined glycan array platform has been used in several previous studies to correlate glycan-binding properties of HA with physiological properties of the virus such as respiratory droplet transmission in ferrets (Jayaraman et al., 2011; Maines et al., 2009; Pearce et al., 2012) and antibody response in mice (Hensley et al., 2009). Streptavidin-coated high-binding-capacity 384-well plates (Pierce) were loaded to the full capacity of each well by incubating the well with 50 μl of 2.4 μM of biotinylated glycans overnight at 4 °C. Excess glycans were removed through extensive washing with PBS. The trimeric HA unit is composed of three HA monomers. The spatial arrangement of the glycans in the plate array favors binding to only one of the three HA monomers in the trimeric HA unit. Therefore, in order to specifically enhance the multivalency in the HA-glycan interactions, the recombinant HA proteins were precomplexed with the primary and secondary antibodies in the molar ratio of 4:2:1 (HA:primary:secondary). The identical arrangement of four trimeric HA units in the precomplex for all the HAs permits comparison between their glycan binding affinities. A stock solution containing appropriate amounts of histidine-tagged HA protein, primary antibody (Mouse anti-6 × His tag IgG from Abcam), and secondary antibody (HRP conjugated goat anti-mouse IgG from Santa Cruz Biotechnology) in the ratio 4:2:1 was incubated on ice for 20 min. Appropriate amounts of precomplexed stock HA were diluted to 250 μl with 1% BSA in PBS. 50 μl of this precomplexed HA was added to each of the glycan-coated wells and incubated at room temperature (RT) for 3 hr followed by the wash steps with PBS and PBST (1× PBS + 0.05% Tween-20). The binding signal was determined based on HRP activity using Amplex Red Peroxidase Assay Kit (Invitrogen, Carlsbad, CA) according to the manufacturer’s instructions. The experiments were done in triplicate. Minimal binding signals were observed in the negative controls, including binding of precomplexed unit to wells without glycans and binding of the antibodies alone to the wells with glycans. The binding parameters, cooperativity (n), and KD for HA-glycan binding were calculated by fitting the average binding signal value (from the triplicate analysis) and the HA concentration to the linearized form of the Hill equation:

\[
\log\left(\frac{y}{1-y}\right) = n \log K_d + \log [HA] \quad (n < 1)
\]

where y is the fractional saturation (average binding signal maximum observed binding signal). In order to compare KD values, the values reported in this study correspond to the appropriate representative avian (3′ SLN-LN or 3′ SLN-LN-LN) and human (6′ SLN-LN) receptors that gave the best fit to the above equation and the same slope value (n value is ~1.3).

As noted above, there were no differences in the glycan-binding properties for HA derived from baculovirus when compared to that of HA produced via mammalian expression.

Binding of WT and Mutant HAs to Human Tissue Sections

Paraffinized human tracheal and alveolar (US BioChain) tissue sections were deparaffinized, rehydrated, and incubated with 1% BSA in PBS for 30 min to prevent nonspecific binding. HA was precomplexed with primary antibody (mouse anti-6 × His tag, Abcam) and secondary antibody (Alexa Fluor 488 goat anti-mouse, Invitrogen) in a molar ratio of 4:2:1, respectively, for 20 min on ice. The tissue binding was performed over two different HA concentrations (40 μg/ml and 20 μg/ml) by diluting the precomplexed stock HA in 1% BSA-PBS. Tissue sections were then incubated with HA-antibody complexes for 3 hr at RT. The tissue sections were counterstained by propidium iodide (Invitrogen; 1,100 in TBST). The tissue sections were mounted and then viewed under a confocal microscope (Zeiss LSM 700 laser scanning confocal microscope). Sialic-acid-specific binding of HAs to tissue sections was confirmed by loss of staining after pretreatment with Sialidase A (recombinant from Arthrobacter ureaficiens, Prozyme). This enzyme has been demonstrated to cleave the terminal Neu5Ac from both Neu5Acα2→3Gal and Neu5Acα2→6Gal motifs. In the case of sialidase pretreatment, tissue sections were incubated with 0.2 units of Sialidase A for 3 hr at 37 °C prior to incubation with the proteins. Pretreatment of human tracheal and alveolar tissue sections with Sialidase A resulted in complete loss of HA staining.

Capturing Network of RBS Residues

The coordinates of the HI HA-avian receptor and Ab58 HA-human receptor complex were uploaded into the PDBePISA server (http://www.ebi.ac.uk/msd-srv/prot_int/pistart.html) to determine key residues in the HA RBS that make contact with the corresponding glycan receptor (interface cutoff of 30% was used). For these residues, their environment was defined using a distance threshold of 7 Å, and the contacts, including putative hydrogen bonds (which include water-bridged ones), disulfide bonds, π bonds, polar interactions, salt bridges, and Van der Waals interactions (nonhydrogen), occurring between pairs of residues within this threshold distance were computed as described previously (Soundararajan et al., 2011). These data were assembled into an array of eight atomic interaction matrices. A weighted sum of the eight atomic interaction matrices was then computed to produce a single matrix that accounts for the strength of atomic interaction between residue pairs within the RBS, using weights derived from relative atomic interaction energies (Soundararajan et al., 2011). The residue-residue interaction network calculated in this fashion generates a matrix that describes all the contacts made by critical RBS residues with spatial proximal neighboring residues in their environment. For each element i, j is the sum of the path scores of all paths between residues i and j. The degree of networking score for each residue was computed by summing across the rows of the matrix, which was meant to correspond to the extent of “networking” for each residue. The interactional relationship between critical
RBS residues and their environment is represented using a two-dimensional open connectivity network diagram (RBSN diagram). The degree of networking score was normalized (RBSN score) with the maximum score for each protein so that the scores varied from 0 (absence of any network) to 1 (most networked). Although previous studies (Herfst et al., 2012; Imai et al., 2012) have reported mutations in stalk and other regions of HA outside the RBS, the RBSN does not include these residue positions, and hence, any amino acid changes in these positions would not affect the glycan receptor-binding property of HA.

**Sequence Analysis of H5 HA and Estimation of Key Features**

A total of 6,014 H5 HA sequences were downloaded from the EpiFlu database. From this, only those sequences that had complete coding regions, including start and stop codons, were considered. In order to avoid estimation errors due to multiply represented sequences, all groups of identical sequences in the data set were represented by the oldest sequence in the group. The remaining 2,959 sequences were ordered by isolation time and aligned, and the occurrence rate of each feature (defined as the percent fraction of sequences from a given year that contains that feature) was calculated.

**Phylogeny Tree Construction and Calculation of Sequence-Based Distance Measure for RBS Analysis**

Phylogeny tree was constructed for 2,959 H5 HA amino acid sequences by neighbor-joining method using MEGA 5.10 software (http://www.megasoftware.net/). The branches were colored based on the different features present (Table 1) using the Rainbow Tree software (http://www.hiv.lanl.gov/content/sequence/RAINBOWTREE/rainbowtree.html).

**SUPPLEMENTAL INFORMATION**

Supplemental Information includes five figures and one table and can be found with this article online at http://dx.doi.org/10.1016/j.cell.2013.05.035.

**ACKNOWLEDGMENTS**

This work was funded in part by the National Institutes of Health (R37 GM057073-13) and the National-Research-Foundation-supported Interdisciplinary Research Group in Infectious Diseases of SMART (Singapore-MIT Alliance for Research and Technology).

Received: April 5, 2013
Revised: May 9, 2013
Accepted: May 20, 2013
Published: June 6, 2013

**REFERENCES**


