SCIENCE CHINA
Life Sciences

• COVER ARTICLE •

June 2010 Vol.53 No.6: 633–642
doi: 10.1007/s11427-010-4010-8

In silico characterization of the functional and structural modules of the hemagglutinin protein from the swine-origin influenza virus A (H1N1)-2009

SUN YePing[1,2,3,†], SHI Yi [1,2,3,†], ZHANG Wei [1,2,3], LI Qing [1,2,3,4], LIU Di [1], Christopher VAVRICKA[1], YAN JingHua[1,3] & GAO George F[1,2,3,5,*]

[1] *CAS Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China*;
[2] *Graduate University of Chinese Academy of Sciences, Beijing 100049, China*;
[3] *China-Japan Joint Laboratory of Molecular Immunology and Molecular Microbiology, Institute of Microbiology, Chinese Academy of Sciences, Beijing 100101, China*;
[4] *College of Life Science, Science and Technology University of China, Hefei 230027, China*;
[5] *Beijing Institutes of Life Science, Chinese Academy of Sciences, Beijing 100101, China*

Received January 15, 2010; accepted February 25, 2010

The 2009 swine-origin influenza virus (S-OIV, H1N1 subtype) has developed into a new pandemic influenza as announced by the World Health Organization. In order to uncover clues about the determinants for virulence and pathogenicity of the virus, we characterized the functional modules of the surface glycoprotein hemagglutinin (HA), the most important protein in molecular epidemiology and pathogenesis of influenza viruses. We analyzed receptor binding sites, basic patch, neutralization antibody epitopes and T cell epitopes in the HA protein of the current S-OIV according to the corresponding functional and structural modules previously characterized in other H1 HA molecules or HA molecules of other subtypes. We compared their differences and similarities systematically. Based on the amino acids defined as the functional and structural modules, the HA protein of 2009 S-OIV should specifically bind to the human 2,6-receptor. The D225G/E mutation in HA, which is found in some isolates, may confer dual binding specificity to the 2,3- and 2,6-receptor based on previously reported work. This HA variant contains two basic patches, one of which results in increased basicity, suggesting enhanced membrane fusion function. The 2009 S-OIV HA also has an extra glycosylation site at position 276. Four of the five antibody neutralization epitopes identified in A/RP/8/34(H1N1) were exposed, but the other was hidden by a glycosylation site. The previously identified cytotoxic T cell epitopes in various HA molecules were summarized and their corresponding sequences in 2009 S-OIV HA were defined. These results are critical for understanding the pathogenicity of the virus and host immune response against the virus.

2009 S-OIV, hemagglutinin, *in silico*, structural and functional modules

Citation: Sun Y P, Shi Y, Zhang W, *et al*. *In silico* characterization of the functional and structural modules of the hemagglutinin protein from the swine-origin influenza virus A (H1N1)-2009. Sci China Life Sci, 2010, 53: 633−642, doi: 10.1007/s11427-010-4118-x

Recently, a new strain of influenza A virus was detected that caused disease in and was transmitted among humans. On 11 June 2009, the World Health Organization (WHO) declared that the infections caused by the new strain had reached pandemic levels. Characterized as an influenza A virus of the H1N1 subtype, this new strain, called 2009 swine-origin influenza virus (S-OIV), was generated by reassortment of gene fragments from influenza viruses of

†Contributed equally to this work
*Corresponding author (email: gaof@im.ac.cn)

swine (HA, NP and NS), avian (PB2 and PA) and human H3N2 (PB1) circulating in North American swine. By contrast, the NA and M gene segments had their origin from the Eurasian avian-like swine H1N1 lineage [1–3].

The virulence of S-OIV seems to be higher than seasonal influenza virus. One of the first S-OIV isolates from the USA, A/California/04/09 (CA04), and other S-OIV isolates replicated more efficiently than a currently circulating seasonal human H1N1 virus in mice and ferrets. CA04 also replicates efficiently in non-human primates and causes more severe pathological lesions [4]. The efficiency of transmission of influenza A pandemic (H1N1) 2009 virus is uncertain. One study assumed that its transmissibility (R_0) is substantially higher than that of seasonal influenza and comparable with viruses that have caused previous influenza pandemics [5]. Recently, it has been revealed that the 2009 pandemic influenza A (H1N1) may pose an increased risk of severe illness in pregnant women, similar to previous epidemic and pandemic diseases [6].

Hemagglutinin (HA), the surface glycoprotein of influenza virus, is of primary importance in the epidemiology of the influenza virus. HA is responsible for viral binding to host receptors, enabling entry into the host cell through endocytosis and subsequent membrane fusion, and so it determines the virulence, host ranges, cell tropism and transmissibility of influenza viruses [7–9]. There are 16 subtypes of HA [10], and the high-resolution crystal structures of multiple subtypes of HA, including H1 [11–14], H2 [15], H3 [16–23], H5 [24–27], H7 [13], H9 [27] and H14 [19] have been determined, which facilitates detailed structural and functional studies. Some structurally and functionally important modules have been defined in these subtypes of HA. Based on these previous studies, we are able to identify the corresponding structural and functional modules in HA of the current 2009 pandemic virus. By comparing the differences and similarities of these deduced modules in HA of the 2009 virus and experimentally defined ones in other HA molecules, we speculate some potential determinants of the virulence of 2009 pandemic H1N1 virus.

1 Materials and methods

1.1 Selection of HA for analysis

For analysis of the mutations in 2009 S-OIV HA, all 1913 HA sequences from the 2009 H1N1 outbreak were downloaded from GenBank. Nucleotide sequence translation and editing were carried out with the EMBOSS package. The alignment of the translated peptides of all 1913 sequences was generated using Clustal W. The frequency of residues at each position was then calculated.

1.2 Definition of the modules

For identification of the functional and structural modules, we used HA from CA04, one of the first S-OIV isolates from the USA. The structural and functional modules defined in this study included glycosylation sites, receptor binding sites, basic patches, antibody neutralization epitopes and T cell epitopes. The receptor binding site in CA04 HA was defined as the amino acids located in the corresponding positions of other HA molecules described in previous 3D structure and mutation analysis studies [11,26,28]. Basic patches in CA04 HA were defined as the amino acids located in the corresponding positions of the basic patches in HA molecules from the 1918 H1N1 and 2004 H5N1 viruses [11,26]. Neutralization antibody epitopes were defined according to the antibody neutralization epitopes identified in the HA molecule of the A/RP/8/34 (H1N1) strain. Cytotoxic T lymphocyte (CTL) epitopes previously identified in other HA molecules were summarized and their corresponding sequences in CA04 HA were defined. Glycosylation sites in CA04 are defined by the glycosylation motif "NXS/T", where X represents any amino acid excluding proline [29].

1.3 Structural modeling

The homology model for CA04 HA was generated using the SWISS-MODEL homology-modeling server [30]. All six individual chains of the 1RUY PDB structure were modeled separately against the representative S-OIV (A/California/04/2009) HA sequence (ACP41105.1). The 1RUY structure was selected for modeling by a BLAST search of the ACP 41105.1 sequence against the PDB, using the ''Advanced Search'' option on the PDB website. The chains were consolidated into one PDB file using UCSF Chimera [31]. The images were produced using the PyMOL software.

2 Results

2.1 Overall mutations in the HA genes

In order to analyze all the mutations in the 1913 sequences of HA, we used the amino acid sequence of CA04 HA as a reference. If the amino acid in a certain position of a given 2009 S-OIV HA was different from the amino acid in the corresponding position of CA04 HA, we defined it as a mutation. The mutation frequency was the total number of mutations at this position in all the sequences divided by 1913. The overall average mutation frequency was defined as the arithmetic mean value of the above mutation frequency of each position in the 1913 sequences. By this method, the overall average mutation frequency was 0.746%. Calculation of the frequency of mutations at each position in the 1913 sequences revealed that some positions had higher frequencies than the overall average mutation frequency. These positions and their mutation frequency were: 2 (4.18%), 8 (2.5%), 12 (2.3%), 48 (5.85%), 100 (1.09%), 138 (1.57%), 145 (1.25%), 220 (93.3%), 222 (0.836%), 239 (5.802), 310 (4.025%), 312 (1.254%), 321 (1.620%), 338

(2.352%), 387 (0.993%), 391 (1.150%), 428 (1.516%) and 477 (0.836%). A higher mutation frequency may suggest certain selective pressure at these positions.

The structure model for CA04 HA was established and is shown in Figure 1A. The receptor binding site, the two basic patches and Asn in the glycosylation sites are indicated in this model. All the functional and structural modules are shown in Figure 1B.

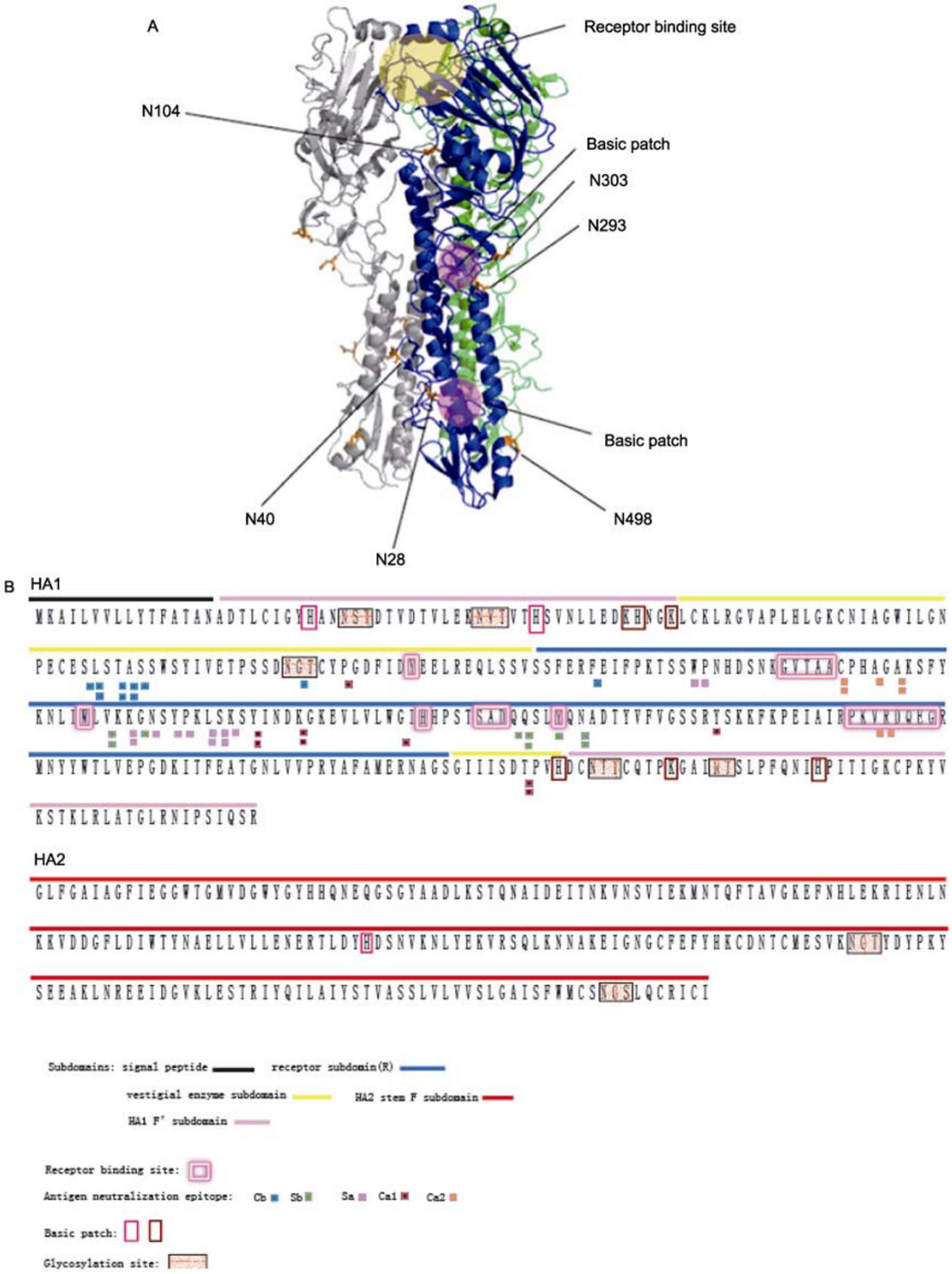

Figure 1 Functional and structural motifs defined in this study. A, Schematic view of the motifs defined in this study. The sequence ranges of the sub-domains of A/California/04/2009 HA are indicated with heavy lines of different colors above the sequence. The amino acids that make up the receptor binding site, basic patches, glycosylation sites, and antibody neutralization epitopes are indicated with different markers. B, The structural model of the A/California/04/2009 (CA04) HA. The receptor binding site, basic patches and Asn in the glycosylation sites are shown.

2.2 Receptor binding and D225G/E mutations

Viral infection and transmission begins with a critical interaction between HA and sialic acid (SA) containing glycans on the cell surface. All HA structures have similarly configured receptor binding domains (RBDs). The binding site comprises three structural elements, namely a α-helix (190-helix, HA1 188–190) and two loops (130-loop, HA1 134–138; 220-loop, HA1 221–228) [26]. These structural elements form the sides of the receptor binding site, and the residues Tyr98, Trp153, His183, and Tyr195 are conserved in all subtypes of HA (Figure 2) [32].

Previous studies using whole virus, identified a number of key RBD mutations that were implicated in avian-human receptor specificity switching of H1, H2, H3, and H5 serotypes. However, adaptation of avian H1, H2/H3 and H5 serotypes for human receptor binding occurs by different mechanisms. For H2 and H3, mutation of Gln226 and Gly228 in avian strains to Leu226 and Ser228 in human viruses correlates with a shift to human receptor specificity [33,34]. But for H5, residues 139, 182, 192, 193, 223, 226 and 228 have been associated with receptor specificity [35].

HA1 138, 186, 190, 194, 225, 226 and 228 have been previously shown to affect receptor binding to H1 HA [36,37]. Among them, two residues, HA1 190 and 225, play predominant roles in determining the receptor-binding specificity of H1 HA. In humans the residues D190/D225 determine specificity for the α-2,6 receptors; D190/G225 are the respective residues for α-2,6 and α-2,3 receptors in swine; and E190/G225 are the residues for α-2,3 receptors in avian species [38–41]. According to these previous studies, the receptor binding specificity of 2009 S-OIV HA should be specific α-2,6 human-type receptors, because it has D190/D225 in its RBD. The critical amino acids that affect receptor binding specificity in different HA serotypes are listed in Table 1.

Recently, a mutation (D225G/E) in the HA protein from 2009 H1N1 S-OIV has been identified. This mutation may have important functional significance because it may switch the receptor specificity and allow it to be a dual binding receptor for both the 2,3- and 2,6-receptors.

The receptor specificity and the impact of the D225G/E mutation can also be elucidated by comparing crystal structures of 1934-human HA in complex with the human receptor analog LSTC (PDB ID: 1RVZ), 1934-human HA in complex with the avian receptor analog LSTA (PDB ID: 1RVX) and 1930-swine HA in complex with the human receptor analog LSTC (PDB ID: 1RV0). In all these structures, the 130-loop interacts with sialic acid using three hydrogen bonds. But the interaction modes of the sialic acid with the 190-helix or 220-loop are different in the three structures. The residue Glu190 in the 190-helix in 1934-human HA forms a hydrogen bond with the 9-hydroxyl of sialic acid in LSTC or LSTA, while in 1930-swine HA, the residues Asp190 and Ser193 form two hydrogen bonds with the GlcNac-3 and Gal-4 of LSTC. The side chains of the residues Lys222 and Asp225 of the 220-loop in 1934-human HA form four hydrogen bonds with the 2- and 3-hydroxyls of Gal-2 in LSTC or LSTA, and a fifth hydrogen bond between the 4-hydroxyl of Gal-2 and the main-chain amide of residue 227 is mediated by a water molecule. Similarly in 1930-swine HA, the side chain of the residue Lys222 and the main-chain carbonyl of the residue Gly225 form four hydrogen bonds with the 2- and 3-hydroxyls of Gal-2, and a fifth hydrogen bond forms as in 1934-human HA complexes. For avian receptor binding, the residue Glu190 of the 190-helix in the 1934-human HA plays a key role in adjusting the position of Gln226. This position adjusting effect is stronger with the presence of two water molecules in its complex, especially for the avian receptor when compared

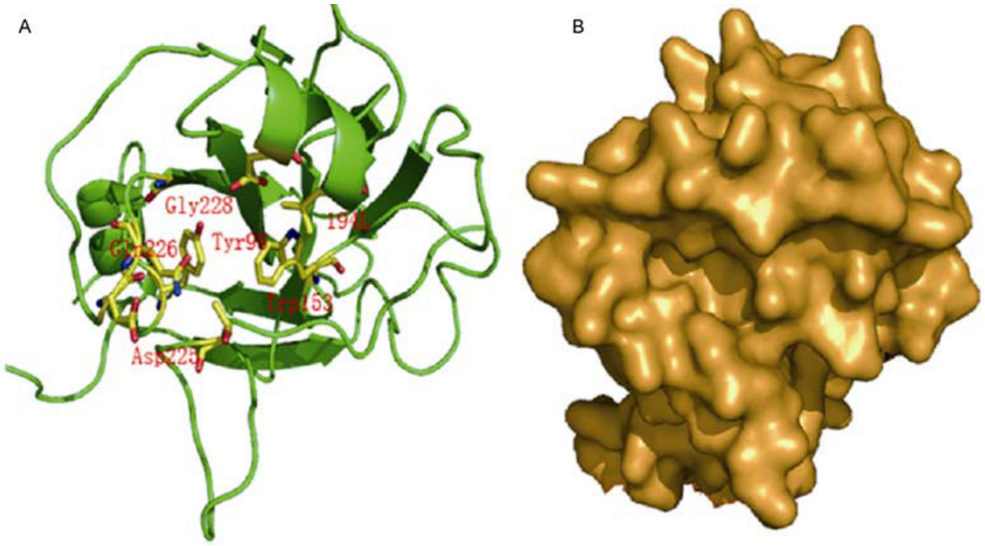

Figure 2 Receptor binding amino acids shown in the structural model of the A/California/04/2009 HA. A, Representation of the receptor binding site, with the key conserved amino acids indicated. B, The surface cavity of the sialic acid receptor binding site.

Table 1 Amino acid residues critical for determination of 2,6- or 2,3-receptor specificity

	139	182	190	192	193	223	225	226	228
H1 serotype									
α-2,6 specificity			D				D		
α-2,3 specificity			E				G/E		
H2 serotype									
α-2,6 specificity								L	S
α-2,3 specificity								Q	G
H3 serotype									
α-2,6 specificity								L	S
α-2,3 specificity								Q	G
H5 serotype									
α-2,6 specificity	R	K		R	K	N		L	S
α-2,3 specificity	G	N		Q	N	S		Q	G

to the human receptor complex or the uncomplexed HA. Therefore Glu190 facilitates binding to the avian receptor. However in 1930-swine HA, residue 190 is an aspartic acid, which does not interact with the 9-hydroxyl of sialic acid or Gln226, and is thus unable to facilitate binding to the avian receptor (Figure 3).

According to the above discussion and Table 1, we can predict the likely binding specificity of 2009 S-OIV HA. If interactions with Lys222 and Asp225 are formed with human receptors, as in the 1934-human HA complex, then the overall orientation of the oligosaccharide in the 2009 HA binding sites may also be similar. In this case, the Asp190 and Ser193 of 2009-swine HA would not contact the receptor. Due to a shorter side chain containing Asp190 instead of the Glu190 of 1934-human HA, the 2009 HA has weaker binding specificity to the avian receptor. Regarding the D225G mutation, the interaction mode would resemble the 1930-swine HA, and the structure may have a moderately different overall orientation of the oligosaccharide compared with 1934-human HA. In terms of the D225E mutation, it is more difficult to deduce receptor specificity because there is no previous experimental data from a HA with an E at position 225. Due to the longer side chain, Glu as opposed to Asp, it may generate one or two hydrogen bonds with the Gal-2 of the avian receptor and therefore would facilitate binding to the avian receptor. The longer side chain of E225 could possibly result in steric hindrance to the 2,3-linked receptor, thereby reducing the binding ability of this HA mutant to the avian receptor (Figure 3).

2.3 Basic patch

In the 1918 HA0 structure, a pH-sensitive histidine patch

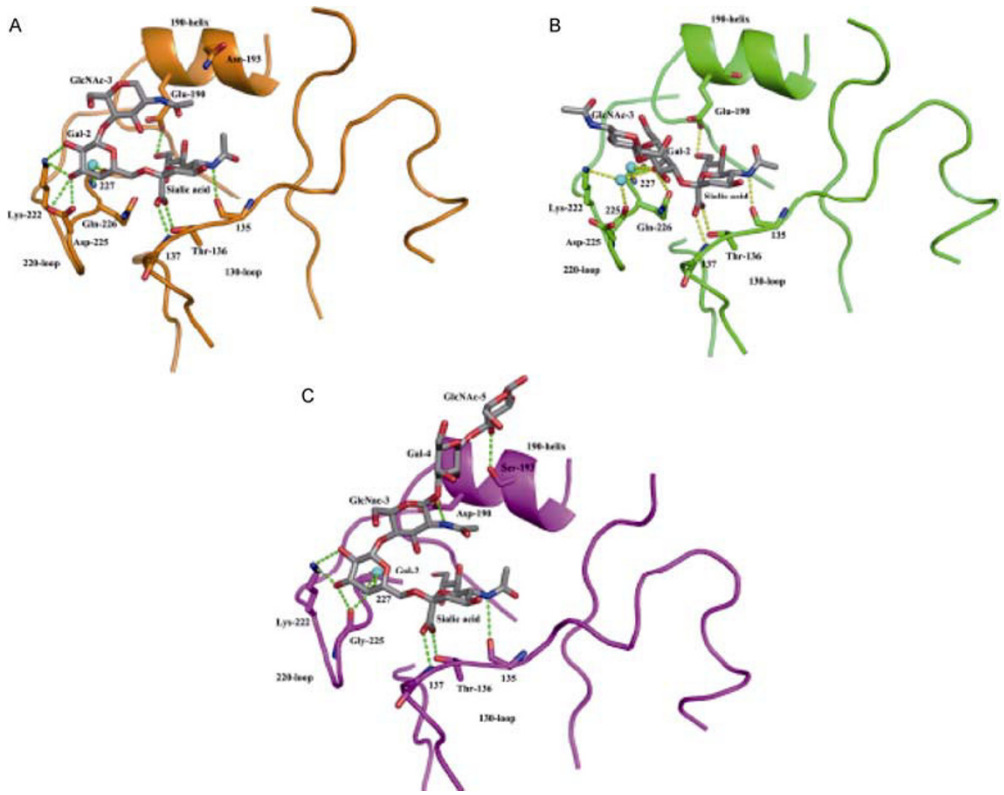

Figure 3 Detailed modeling of receptor binding specificity for CA04 HA and D225G/E mutants. A, 1934-human-H1 in complex with human receptor analog LSTC; B, 1934-human-H1 in complex with avian receptor analog LSTA; C, 1930-swine-H1 in complex with human receptor analog LSTC.

(HA1-His18, HA1-HisA38, and HA2-His111), together with the adjacent HA2 TrpB21, is proposed to play a role in fusion peptide destabilization and release [11]. This structural feature is conserved in other avian and human H1, H2, and H5 serotypes, as well as in Viet04 H5 HA. In the 1918 HA0, a second patch of four exposed histidines within the vestigial esterase domain, together with a nearby lysine, are also implicated in pathogenicity via enhanced membrane fusion. Of the five HA1 residues in this basic patch (HA1-His47, HA1-Lys50, HA1-His275, HA1-His285 and HA1-His298) in the 1918 HA, only three are conserved in avian H5 structures (HA1-His47, HA1-Lys50 and His298), but Viet04 and Sing97 HAs have an additional lysine (HA1-Lys45) and histidine (HA1-His295). Furthermore, Viet04 has yet another lysine (Lys46), which renders this patch even more basic and is found in two strains (1203/ 1204) that were isolated from the same patient (Table 2). The contribution of this region to virulence, if any, is as yet unknown, but is worthy of further investigation [26].

In the 2009 S-OIV HA, the first basic patch is conserved compared to 1918 and other H1 HAs, but the basicity of the second basic patch may be increased because of the introduction of Lys46 and substitution of H285 for a lysine. The membrane fusion function may be enhanced compared with other H1 viruses, including the 1918 virus.

2.4 Antibody neutralization epitopes

In previous studies, the antigenic structure of H1 HA (A/PuertoRico/8/1934) has been found to include five distinct antigenic sites on the globular domain: Sa, Sb, Ca1, Ca2 and Cb. These sites are mapped onto the 3D structure model of the 2009 HA (Figure 4). Among these antibody neutralization epitopes, only position 239 in Ca2 has a mutation frequency (5.802%) higher than the overall mutation frequency, suggesting possible selective pressure of the antibodies. In HA of 2009 S-OIV, with the exception of Cb, all antigenic sites are exposed for antibody recognition. The Cb site is proximal to the oligosaccharide at HA1 Asn104. The glycosylation here may cover the Cb site and disrupt antibody recognition.

Recently a monoclonal antibody, CR6261, was isolated by phage display selection on recombinant H5 HA. CR6261 neutralizes multiple influenza subtypes, including H1, H2, H5, H6, H8 and H9, as well as protecting mice from lethal challenge with the 1918 H1N1 and H5N1 viruses. In the crystal structure of the CR6261 in complex with 1918 H1 HA, the interactions between H1 HA and C6261 are mediated by eight hydrogen bonds formed between C6262 and HA2 in addition to non-polar contacts that can be divided into two groups. The amino acids in 1918 H1 HA that form hydrogen bonds with C6261 include HA2-Gln42, HA2-Thr49, HA2-sn53, HA2-Asp46, HA2-Ile18, HA2-Asp19 and HA2-Gln42. The first region that forms non-polar contacts with C2621 consists of HA1-Val40, HA1-Leu42 and HA1-Leu292, along with HA2-Thr49, HA2-Val52 and HA2-Ile56. The second cluster includes His18 and His38 from HA1 and Trp21, Thr41 and Ile45 from HA2. These amino acids are totally conserved in HA 2009 S-OIV. Therefore, we speculated that the CR6261 monoclonal antibody can also effectively target 2009 S-OIV.

2.5 CTL epitopes

In addition to an antibody-induced humoral immune response, CTL mediated cellular immunity also plays an im-

Table 2 Amino acid residues contained within the basic patches of HA in different subtypes

	A18	A38	B111	A47	A45	A46	A50	A275	A285	A295	A298
H1 subtype											
A/California/04/2009 (H1N1)	H	H	H	H	D	K	K	H	K	Q	H
A/Brevig_Mission/1/1918 (H1N1)	H	H	H	H	D	S	K	H	H	Q	H
A/Puerto_Rico/8/34	H	H	H	H	D	S	K	H	L	Q	H
A/Brisbane/59/2007	H	H	H	H	N	S	K	D	Q	Q	H
H5 subtype											
A/Duke/Singapore/97	H	H	H	H	K	T	K	T	M	H	H
A/Vietnam/1203/2004	H	H	H	H	K	K	K	K	M	H	H
A/Vietnam/1194/2004(H5N1)	H	H	H	H	K	T	K	K	M	H	H
H7 subtype											
A/turkey/214845/02(H7N5)	H	N	A	N	R	T	P	A	G	Q	N
A Canada rv504 2004(H7N3)	H	N	A	N	T	V	K	S	G	Q	N
H9 subtype											
A/Quail/Hong Kong/AF157/ 92(H9N2)	Q	H	H	H	T	E	G	G	K	H	S
A HK 2108 2003(H9N2）	Q	Q	H	H	T	E	G	G	R	H	S

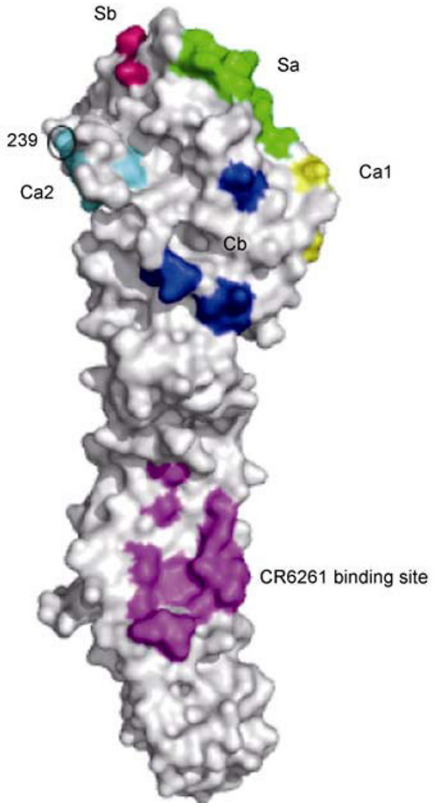

Figure 4 Antibody neutralization epitopes mapped in 3D structure of CA04. The five antibody neutralization epitopes are shown in different colors: Sa (green), Sb (red), Ca1 (yellow), Ca2 (cyan) and Cb (blue). The amino acids recognized by the monoclonal antibody CR6261 are labeled in magenta, and the amino acid residues with high mutation frequency are indicated by a circle.

portant role in defending against influenza virus infection [42,43]. Numerous CTL epitopes have been identified in influenza virus, but most of them are in the internal proteins of influenza virus, with CTL epitopes identified in HA occurring relatively rarely. However, there are indeed some important CTL epitopes that have been discovered in HA. For example, in A/Japan/305/57(H2N2) HA, an immunodominant H2 K^d-restricted CTL epitope (H2 204–212, LYQNVGTYV) was identified. It induced a robust CTL response in A/Japan/305/57(H2N2) virus challenged H2 K^d mice, but a CTL epitope overlapping with it (H2 210–219, TYVSVGTSTL) was sub-dominant [44]. In A/PR/8/34 (H1N1), two H-2-K^k restricted CTL epitopes were identified (HA1 259–266, FEANGNLI and HA2 10–18, IEGGW TGMI) [45] in 1991. Later, an H-2-K^k restricted CTL epitope (H1 533–541, IYATVASSL) was shown to cross-react with an H-2 K^k restricted CTL epitope (H2 533–541, IYATVAGSL) in A/Japan/305/57(H2N2) HA, and it was subsequently shown to be a sub-immunodominant CTL epitope in A/PR/8/34(H1N1) [46]. In 2003, a H2 K^b restricted CTL epitope (457–482, KEIGNGCFEF) was identified in A/PR/8/34 by screening the whole viral genome [47]. Other CTL epitopes identified in mice include two H2 K^b sequences in A/Udorn/72(H3N2) HA2 (HA2 93–102, SYNAELLVAL and HA2 181–189, GYKDWILWI).

There are four CTL epitopes identified in human leukocyte antigen (HLA) background, and all of them are HLA-A11-restricted and found in A/PR/8/34. They are H1 63–71 (GIAPLQLGK), H1 149–158 (VTAACSHAGK), H1 450–460 (RTLDFHDSNVK) [48] and H1 458–467 (NVKNLY EKVK) [49].

Recently, we identified an HLA-A*0201-restricted CTL epitope, RI-10 or KI-10 (H5 HA 205–214, R/KLYQN PTTYI, RI-10) in H5 HA derived from H5N1 virus. It induced high levels of a CTL epitope-specific response both in HLA-A*0201/K^b transgenic mice and patients recovered

Table 3 CTL epitopes in HA and their corresponding sequences

CTL epitope	Source strain	MHC restriction	Corresponding sequence in 2009 S-OIV HA	Position in 2009 S-OIV HA
LYQNVGTYV	A/Japan/305/57(H2N2)	H2 K^d	LYQNADTYV	208-217
TYVSVGTSTL	A/Japan/305/57(H2N2)	H2 K^d	TYVFVGSSRY	224-233
IYATVASSL	A/Japan/305/57(H2N2)	H-2 K^k	IYSTVASSL	533-531
IYATVASSL	A/PR/8/34(H1N1)	H-2 K^k	IYSTVASSL	533-531
FEANGNLI	A/PR/8/34(H1N1)	H-2 K^k	FEATGNLV	259-256
KEIGNGCFEF	A/PR/8/34(H1N1)	H2 K^b	KEIGNGCFEF	475-484
SYNAELLVAL	A/Udorn/72(H3N2)	H2 K^b	TYNAELLVLL	437-446
GYKDWILWI	A/Udorn/72(H3N2)	H2 K^b	IYQILAIYS	607-616
GIAPLQLGK	A/PR/8/34(H1N1)	HLA-A11	GVAPLHLGK	63-71
VTAACSHAGK	A/PR/8/34(H1N1)	HLA-A11	VTAACPHAGA	149-158
RTLDFHDSNVK	A/PR/8/34(H1N1)	HLA-A11	RTLDYHDSNVK	450-460
NVKNLYEKVK	A/PR/8/34(H1N1)	HLA-A11	NVKNLYEKVR	458-467
RLYQNPTTYI	A/Bar-headed goose/Qinghai/1/2005	HLA-A*0201	SLYQNADTYV	207-216

from H5N1 infection [50].

Most of the peptide sequences, or at least the anchor residues usually at positions 2 and 9 or 10 of the peptides which are most importance for MHC binding, in CA04 HA are conserved compared to their corresponding CTL epitopes previously identified (Table 3). No mutations with high frequency were found among these sequences in the 1913 sequences of 2009 S-OIV HA that we examined.

3 Discussion

The current S-OIV pandemic starting in North America has aroused much panic around the world. Although some studies suggested that the new H1N1 is more virulent and transmits more efficiently than seasonal flu, it actually causes a disease no more severe than seasonal flu, although it is unusually dangerous for the young and pregnant women. However, scientists have repeatedly warned that this relatively mild virus could become deadlier if mutations (antigenic drift) or gene reassortment (antigen shift) with other influenza viruses occurs. But for now, it would appear as if this H1N1 virus, which mysteriously jumped from swine to humans will go down in history as causing confusion as opposed to a catastrophe [51].

Molecular characterization of the new virus is worthwhile however. HA is the most important surface protein of influenza viruses and it is also the most studied among all the viral proteins. Until the crystal structure of 2009 S-OIV HA is solved, structural modeling and *in silico* prediction is the most effective way to study the properties of this protein. In the present study, we deduced the structural and functional motifs based on primary sequence alignment and structure modeling. Through our analysis, we identified some special characteristics of the HA from 2009 S-OIV.

Comparison of the primary sequence and analysis of previously reported crystal structures for H1 HA in complex with its receptor, sialic acid, revealed to us that 2009 S-OIV HA should have an α-2,6-receptor binding specificity. Again by structural analysis, we deduced that the recently identified D225G/E mutation should confer α-2,3-receptor specificity in addition to α-2,6-receptor binding.

Recently, it was demonstrated that there was a different anatomical distribution between α-2,6- and α-2,3-linked sialoside receptors. The former is mainly distributed in epithelial cells in the bronchi and to a lesser degree in alveolar cells. In contrast, the latter translocates extensively to alveolar cells but to a lesser extent in bronchial epithelial cells [52]. Accordingly, the D225G/E mutants may exhibit different tissue tropism and cause different pathogenic outcomes with the CA04 isolate.

A critical clue with respect to the virulence and pathogenicity of 2009 S-OIV can be seen in the increased basicity of the basic patch within the HA molecule. The second basic patch of 2009 S-OIV HA (K46, H47, K50, H275, K285, H298) is obviously more basic than the corresponding basic patch in 1918 HA (H47, K50, H275, H285, H298), not to mention the corresponding basic patch in seasonal influenza A/Brisbane/59/2007 HA (H47, K50, H298). The increased basicity in the second basic patch has been implicated in enhanced membrane fusion activity and pathogenicity, which needs to be investigated further.

We analyzed both antibody neutralization and CTL epitopes in 2009 S-OIV HA. Although almost a year has passed since the first 2009 H1N1 virus was isolated, no obvious immune escape mutations have been detected except maybe at position 239. The lack of immune escape mutations in 2009 S-OIV HA suggests that the population has not developed specific antibodies against this virus, and this may explain why the pandemic is still currently ongoing [53]. This also reinforces the statement that S-OIV is new to the human population.

In addition to antibody response, the CTL response plays an important role in defending the body against the influenza virus. It has been suggested that the CTL response against the influenza virus is widely cross-reactive because there are numerous CTL epitopes in viral internal proteins that are conserved among different subtypes of influenza virus. Cross-reactive memory CTLs established in seasonal influenza may have a dampening effect on pandemic influenza [54,55]. A pre-existing CTL response against 2009 S-OIV was observed in a healthy population by virtue of conserved CTL epitopes that were internal to the virus [56]. The majority of CTL epitopes specific to influenza viruses were identified as internal viral proteins, with CTL epitopes identified in HA rarely occurring. However some previous studies and our recent results suggest that there may also be CTL epitopes with strong immunogenicity in the HA molecules [44,50]. Therefore it is worthwhile to identify the CTL response of the current S-OIV HA.

Glycosylation has been shown to have important impact on many functions of HA molecules. For example, glycans near antigenic peptide epitopes interfere with antibody recognition [57], and glycans near the proteolytic activation site of HA modulate cleavage and influence the infectivity of influenza viruses. Among different subtypes of influenza viruses, there is extensive variation in the glycosylation sites of the head region, and accordingly these have a differential effect on the functional properties of the HA molecules. We found seven potential glycosylation sites in the 2009 S-OIV HA, and any one of them may interfere with antibody recognition of the Ca2 site.

This work was supported by the National Basic Research Program of China (*Grant No. 2010CB534004*) *and an intramural grant from the Chinese Academy of Sciences* (*Grant No. KSCX2-YW-R-158*).

1 Smith G J, Vijaykrishna D, Bahl J, *et al*. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. Na-

ture, 2009, 459: 1122–1125
2 Gao G F, Sun Y P. It is not just AIV: From avian to swine-origin influenza virus. Sci China Life Sci, 2010, 53: 151–153
3 Trifonov V, Khiabanian H, Rabadan R. Geographic dependence, surveillance, and origins of the 2009 influenza A (H1N1) virus. N Engl J Med, 2009, 361: 115–119
4 Itoh Y, Shinya K, Kiso M, *et al*. *In vitro* and *in vivo* characterization of new swine-origin H1N1 influenza viruses. Nature, 2009, 460: 1021–1025
5 Yang Y, Sugimoto J D, Halloran M E, *et al*. The transmissibility and control of pandemic influenza A (H1N1) virus. Science, 2009, 326: 729–733
6 Louie J K, Acosta M, Jamieson D J, *et al*. Severe 2009 H1N1 Influenza in pregnant and postpartum women in California. N Engl J Med, 2010, 362: 27–35
7 Wilson I A, Cox N J. Structural basis of immune recognition of influenza virus hemagglutinin. Annu Rev Immunol, 1990, 8: 737–771
8 Gao G F, Shaw P C. The challenges of avian influenza virus: Mechanism, epidemiology and control. Sci China Ser C-Life Sci, 2009, 52: 405–406
9 Liu D, Liu X, Yan J, *et al.* Interspecies transmission and host restriction of avian H5N1 influenza virus. Sci China Ser C-Life Sci, 2009, 52: 428–438
10 Fouchier R A, Munster V, Wallensten A, *et al*. Characterization of a novel influenza A virus hemagglutinin subtype (H16) obtained from black-headed gulls. J Virol, 2005, 79: 2814–2822
11 Stevens J, Corper A L, Basler C F, *et al*. Structure of the uncleaved human H1 hemagglutinin from the extinct 1918 influenza virus. Science, 2004, 303: 1866–1870
12 Lin T, Wang G, Li A, *et al*. The hemagglutinin structure of an avian H1N1 influenza A virus. Virology, 2009, 392: 73–81
13 Russell R J, Gamblin S J, Haire L F, *et al*. H1 and H7 influenza haemagglutinin structures extend a structural classification of haemagglutinin subtypes. Virology, 2004, 325: 287–296
14 Gamblin S J, Haire L F, Russell R J, *et al*. The structure and receptor binding properties of the 1918 influenza hemagglutinin. Science, 2004, 303: 1838–1842
15 Liu J, Stevens D J, Haire L F, *et al*. Structures of receptor complexes formed by hemagglutinins from the Asian Influenza pandemic of 1957. Proc Natl Acad Sci USA, 2009, 106: 17175–17180
16 Fleury D, Wharton S A, Skehel J J, *et al*. Antigen distortion allows influenza virus to escape neutralization. Nat Struct Biol, 1998, 5: 119–123
17 Chen J, Lee K H, Steinhauer D A, *et al*. Structure of the hemagglutinin precursor cleavage site, a determinant of influenza pathogenicity and the origin of the labile conformation. Cell, 1998, 95: 409–417
18 Ha Y, Stevens D J, Skehel J J, *et al*. X-ray structure of the hemagglutinin of a potential H3 avian progenitor of the 1968 Hong Kong pandemic influenza virus. Virology, 2003, 309: 209–218
19 Russell R J, Kerry P S, Stevens D J, *et al*. Structure of influenza hemagglutinin in complex with an inhibitor of membrane fusion. Proc Natl Acad Sci USA, 2008, 105: 17736–17741
20 Fleury D, Daniels R S, Skehel J J, *et al*. Structural evidence for recognition of a single epitope by two distinct antibodies. Proteins, 2000, 40: 572–578
21 Fleury D, Barrere B, Bizebard T, *et al.* A complex of influenza hemagglutinin with a neutralizing antibody that binds outside the virus receptor binding site. Nat Struct Biol, 1999, 6: 530–534
22 Barbey-Martin C, Gigant B, Bizebard T, *et al*. An antibody that prevents the hemagglutinin low pH fusogenic transition. Virology, 2002, 294: 70–74
23 Bullough P A, Hughson F M, Skehel J J, *et al*. Structure of influenza haemagglutinin at the pH of membrane fusion. Nature, 1994, 371: 37–43
24 Ekiert D C, Bhabha G, Elsliger M A, *et al.* Antibody recognition of a highly conserved influenza virus epitope. Science, 2009, 324: 246–251
25 Sui J, Hwang W C, Perez S, *et al*. Structural and functional bases for broad-spectrum neutralization of avian and human influenza A viruses. Nat Struct Mol Biol, 2009, 16: 265–273
26 Stevens J, Blixt O, Tumpey TM, *et al*. Structure and receptor specificity of the hemagglutinin from an H5N1 influenza virus. Science, 2006, 312: 404–410
27 Ha Y, Stevens D J, Skehel J J, *et al*. H5 avian and H9 swine influenza virus haemagglutinin structures: possible origin of influenza subtypes. EMBO J, 2002, 21: 865–875
28 Shen J, Ma J, Wang Q. Evolutionary trends of A (H1N1) influenza virus hemagglutinin since 1918. PLoS One, 2009, 4: e7789
29 Kornfeld R, Kornfeld S. Assembly of asparagine-linked oligosaccharides. Annu Rev Biochem, 1985, 54: 631–664
30 Schwede T, Kopp J, Guex N, *et al*. SWISS-MODEL: An automated protein homology-modeling server. Nucleic Acids Res, 2003, 31: 3381–3385
31 Pettersen E F, Goddard T D, Huang C C, *et al*. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem, 2004, 25: 1605–1612
32 Wilson I A, Skehel J J, Wiley D C. Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. Nature, 1981, 289: 366–373
33 Connor R J, Kawaoka Y, Webster R G, *et al*. Receptor specificity in human, avian, and equine H2 and H3 influenza virus isolates. Virology, 1994, 205: 17–23
34 Rogers G N, Paulson J C, Daniels R S, *et al*. Single amino acid substitutions in influenza haemagglutinin change receptor binding specificity. Nature, 1983, 304: 76–78
35 Yamada S, Suzuki Y, Suzuki T, *et al*. Haemagglutinin mutations responsible for the binding of H5N1 influenza A viruses to human-type receptors. Nature, 2006, 444: 378–382
36 Rogers G N, D'Souza B L. Receptor binding properties of human and animal H1 influenza virus isolates. Virology, 1989, 173: 317–322
37 Matrosovich M N, Gambaryan A S, Teneberg S, *et al*. Avian influenza A viruses differ from human viruses by recognition of sialyloligosaccharides and gangliosides and by a higher conservation of the HA receptor-binding site. Virology, 1997, 233: 224–234
38 Reid A H, Janczewski T A, Lourens R M, *et al*. 1918 influenza pandemic caused by highly conserved viruses with two receptor-binding variants. Emerg Infect Dis, 2003, 9: 1249–1253
39 Skehel J J, Wiley D C. Receptor binding and membrane fusion in virus entry: The influenza hemagglutinin. Annu Rev Biochem, 2000, 69: 531–569
40 Tumpey T M, Maines T R, Van Hoeven N, *et al*. A two-amino acid change in the hemagglutinin of the 1918 influenza virus abolishes transmission. Science, 2007, 315: 655–659
41 Srinivasan A, Viswanathan K, Raman R, *et al*. Quantitative biochemical rationale for differences in transmissibility of 1918 pandemic influenza A viruses. Proc Natl Acad Sci USA, 2008, 105: 2800–2805
42 Thomas P G, Keating R, Hulse-Post D J, *et al*. Cell-mediated protection in influenza infection. Emerg Infect Dis, 2006, 12: 48–54
43 Doherty P C, Brown L E, Kelso A, *et al*. Immunity to avian influenza A viruses. Rev Sci Tech, 2009, 28: 175–185
44 Cao W, Myers-Powell B A, Braciale T J. The weak CD8+ CTL response to an influenza hemagglutinin epitope reflects limited T cell availability. J Immunol, 1996, 157: 505–511
45 Gould KG, Scotney H, Brownlee GG. Characterization of two distinct major histocompatibility complex class I Kk-restricted T-cell epitopes within the influenza A/PR/8/34 virus hemagglutinin. J Virol, 1991, 65: 5401–5409
46 Chen W, Anton L C, Bennink J R, *et al*. Dissecting the multifactorial causes of immunodominance in class I-restricted T cell responses to viruses. Immunity, 2000, 12: 83–93
47 Zhong W, Reche P A, Lai C C, *et al*. Genome-wide characterization of a viral cytotoxic T lymphocyte epitope repertoire. J Biol Chem, 2003, 278: 45135–45144
48 Gianfrani C, Oseroff C, Sidney J, *et al*. Human memory CTL response specific for influenza A virus is broad and multispecific. Hum Immunol, 2000, 61: 438–452
49 Alexander J, Oseroff C, Sidney J, *et al*. Derivation of HLA-A11/Kb

transgenic mice: Functional CTL repertoire and recognition of human A11-restricted CTL epitopes. J Immunol, 1997, 159: 4753–4761
50 Sun Y, Liu J, Yang M, *et al*. Identification and structural definition of H5-specific CTL epitopes restricted by HLA-A*0201 derived from H5N1 subtype of influenza A viruses. J Gen Virol, 2010, 91: 919–930
51 Enserink M, Cohen J. Virus of the year. The novel H1N1 influenza. Science, 2009, 326: 1607
52 Shinya K, Ebina M, Yamada S, *et al*. Avian flu: Influenza virus receptors in the human airway. Nature, 2006, 440: 435–436
53 Kilbourne E D. Influenza pandemics of the 20th century. Emerg Infect Dis, 2006, 12: 9–14
54 Lee L Y, Ha do L A, Simmons C, *et al*. Memory T cells established by seasonal human influenza A infection cross-react with avian influenza A (H5N1) in healthy individuals. J Clin Invest, 2008, 118: 3478–3490
55 Kreijtz J H, de Mutsert G, van Baalen C A, *et al*. Cross-recognition of avian H5N1 influenza virus by human cytotoxic T-lymphocyte populations directed to human influenza A virus. J Virol, 2008, 82: 5161–5166
56 Greenbaum J A, Kotturi M F, Kim Y, *et al*. Pre-existing immunity against swine-origin H1N1 influenza viruses in the general human population. Proc Natl Acad Sci USA, 2009, 106: 20365–20370
57 Skehel J J, Stevens D J, Daniels R S, *et al*. A carbohydrate side chain on hemagglutinins of Hong Kong influenza viruses inhibits recognition by a monoclonal antibody. Proc Natl Acad Sci USA, 1984, 81: 1779–1783