The adaptability of H9N2 avian influenza A virus to humans: A comparative docking simulation study

Hengyue Xu[1], Jiaqiang Qian[2], Yifan Song[2] and Dengming Ming[1*]

[1]College of Biotechnology and Pharmaceutical Engineering, Nanajing Tech University, 30 South Puzhu Road, Jiangbei New District, Nanjing City, Jiangsu 211816, PR China, [2]Department of Physiology and Biophysics, School of Life Science, Fudan University, Shanghai 200438, China PR

Abstract

Influenza A virus, the H9N2 subtype, is an avian influenza virus that has long been circulating in the worldwide poultry industry and is occasionally found to be transmissible to humans. Evidence from genomic analysis suggests that H9N2 provides the genes for the H5N1 and H7N9 subtypes, which have been found to infect mammals and pose a threat to human health. However, due to the lack of a structural model of the interaction between H9N2 and host cells, the mechanism of the extensive adaptability and strong transformation capacity of H9N2 is not fully understood. In this paper, we collected 40 representative H9N2 virus samples reported recently, mainly in China and neighboring countries, and investigated the interactions between H9N2 hemagglutinin and the mammalian receptor, the polysaccharide α-2,6-linked lactoseries tetrasaccharide c, at the atomic level using docking simulation tools. We categorized the mutations of studied H9N2 hemagglutinin according to their effects on ligand-binding interactions and the phylogenetic analysis. The calculations indicated that all the studied H9N2 viruses can establish a tight binding with LSTc although the mutations caused a variety of perturbations to the local conformation of the binding pocket. Our calculations suggested that a marginal equilibrium is established between the conservative ligand-receptor interaction and the conformational dynamics of the binding pocket, and it might be this equilibrium that allows the virus to accommodate mutations to adapt to a variety of environments. Our results provided a comprehensive strategy for understanding the adaptive mechanisms of H9N2 viruses, which may help predict the propensity of H9N2 viruses to spread in mammals.

Key words: H9N2, hemagglutinin A, LSTc, sialylated penta-saccharide, Vina Docking, marginal contact equilibrium

INTRODUCTION

Avian influenza virus (AIV) H9N2 subtype is one of the most widely distributed subtypes of AIV in the world, causing huge economic losses to the poultry industry. It was first isolated at the turkey fork in the United States in 1966[1], and later reported in the Middle East, South Africa, Europe and the rest of the world[2, 3]. In China, the H9N2 influenza virus was first found in chicken farms in 1992, which was found in the report of the outbreak of avian influenza in mainland China[4]. Since then, sporadic outbreaks of avian influenza, mainly H9N2, have occurred in Guangdong Province[5-8]. In 1998, a nationwide outbreak of H9N2 avian influenza occurred in China in several months, first in Hebei Province, and then rapidly spread to most chicken farms across the country [9, 10]. Although H9N2 subtype has been proved to be a low virulent virus[11], H9N2 and other studies frequently found in mammals in recent years have shown that H9N2 is intrinsically related to H5N1, H7N9 and other highly pathogenic AIV and other epidemic strains[12]. For example, ferret-model experiments show that H9N2 has a moderate ability to promote cross-species transmission, thus infecting many species including humans, mammals and waterfowl[13, 14]. In recent years, due to the threat of H9N2 subtype avian influenza virus to the safety of poultry industry and public health, more and more attention has been paid to the study of its infection and transmission mechanism.

The structural and phylogenetic analysis of H9N2 have provided profound biochemical characteristics for the study of the possibility of H9N2 virus transmission in domestic poultry industry and its potential threat to public health. Serological data showed that the H9N2 avian influenza virus found in Eurasia is the offspring of three lineages: HK/G1(G1-like), BJ/94-like, and HK/Y439(Y439-like)[15, 16]. Both G1-like and BJ/94-like H9N2 viruses are now distributed in many parts of Asia. Due to the pressure of evolution and adaptation, some of them have developed into new subtypes, such as Iran- and Israel-subtypes[17]. These subtypes show significant stability in local wild- or poultry-breeding, especially on the molecular basis, and they are highly tolerant to mutations that maintain the 3D structure of haemagglutinin (HA) proteins[18]. The H9N2 avian influenza virus recently found in China and Hong Kong is either G1-like or BJ/94-like[19-21], for example, one of the early isolated strains in China, A/Brambling/Beijing/16/2012, belongs to BJ/94-like subtype. The data indicate that BJ/94 strain is mainly prevalent in northern China, [11], and it also provides an internal gene for a new H7N9 avian influenza virus found in mainland China. Data showed that strains of BJ/94-like lineage are primarily circulating epidemically in Northern China[11], and they also provide internal genes for the new H7N9 avian influenza virus, a highly lethal strain found in mainland China[22]. Phylogenetic analysis provides necessary resources for the interpretation of local and pandemic AIVs data and the establishment of influenza epidemic prediction model based on the accumulated data.

Three-dimensional atomic structure models provide valuable knowledge for the study of infection and transmission mechanism of avian influenza virus. Studies have shown that the infection of AIVs usually starts with the binding of the viral protein haemagglutinin A(HA) to the sialic acid receptor on the surfaces of the homoclinic cells, which causes the fusion of viral membrane and endoplasmic membrane, and finally causes the virus to enter the host cell[23]. Sialic acid receptors are located on the surface of respiratory epithelial cells of birds, mainly α-2,3-linked lactoseries tetrasaccharide a (LSTa), arranged in *trans* conformation. On the other hand, these receptors also

exist on the airway surface of mammals, but they have the form of α-2,6-linked lactoseries tetrasaccharide c (LSTc), which is *cis* conformation[24]. In H9N2 subtype AIVs, HA protein has a strong tendency to bind to α-2,3-linked receptors, but accumulated evidences show that they can also bind to α-2,6-linked receptors under certain conditions, which may directly infect mammals and humans[25, 26].

By comparing the structure and sequence of HA protein, receptor specificity and binding affinity, the adaptability of influenza virus to hosts of different species including mammals was studied. For example, the analysis of the escape mutant of H9 influenza virus highlighted a set of conserved residues at sites 98, 153, 183 and 195 (for consistency, all indices are numbered based on the amino acid sequence of H3 influenza virus) in the receptor binding site (RBS) and revealed the crucial role of these sites in the initiation of infection[23, 27-30]. The structural analysis and simulation study also showed that the mutation of 190 residue in 190-helix could cause significant structural change of glycan binding, and may drive the virus into the pathway of human adaptation[31, 32]. In addition to these conserved binding sites, residue 226 has been considered to be the key for viruses to adapt to different types of host cells. Mutation analysis suggested that L226 was helpful for H9 replication and transmission in human and ferret, while Q226 had a high affinity for poultry hosts[33]. In particular, it has been found that the growth rate of L226-type H9N2 virus in human airway epithelial cells is 100 times faster than that of Q226-type H9N2 virus[34]. Although 226 residues were found to be independent of air transmission, two other HA residues in H9N2 virus, L672 in PA and K363 in HA, were identified to play a role in airborne transmission [31]. Taking together, structural and related mutagenesis analysis provides important knowledge for understanding the infection, transmission and adaptive activities of the viruses.

However, due to the high mutation rate in the binding pocket and the flexibility of binding glycan, most of the structural mode HA proteins binding with ligands are still lacking, which makes it difficult to evaluate the effect of virus mutations on the infection and adaptation in the virus transmission[35]. In this paper, the structural models of 40 representative H9N2 strains circulated in China in last few years were analyzed, and the internal docking mode of binding pockets was studied. The microstructural changes caused by various mutations of HA protein of virus were characterized by phylogenetic tree analysis and structure comparison calculation, and the mutation principle of virus infection and transmission was deduced and discussed.

METHOD

Collection of HA protein sequence of H9N2 strains in China In order to study the human infection and animal-to-human/human-to-human transmission of H9N2 subtype avian influenza virus in China, 40 strains of H9N2 influenza virus were collected, which often appeared in recent studies[11, 33]. Three quarters of the selected strains were reported to have occurred in China from 1997 to 2012. The rest included three strains from South Korea, three strains from Japan, one strain from Israel, one strain from South Africa and two strains from the United States. Of all the known H9N2 strains, one of the two strains from the United States was the first one discovered in 1966. The selected strains are listed in Table S1.

Structural homology modeling of selected HA1 proteins The key to understand the mechanism of virus infection and animal-human/human-human transmission is to build a three-dimensional structure model of HA protein complexed with its binding partner LSTc. HA protein consists of two consecutive subunits, HA1 and HA2. It is the HA1 domain, especially the 190-loop and 220-loop of HA1 subunits of H5-, H7-, H9-sublineage viruses, that physically bind to SIA on the surface of target cells, thus causing virus particles to attack host cells[36]. Therefore, in this work, only the HA1 domain of studied proteins is considered for structural homology modeling and docking analysis. The reference structure model of H9N2-HA1 protein binding the penta-saccharide LSTc (α-2,6) was taken from a swine H9N2 influenza virus hemagglutinin (from Protein Data Bank[37], PDB entry code 1JSI[24]). The sequence similarity of HA1 proteins in different strains were determined using CLUSTAL2.0[38]. The calculation shows that the selected HA1 proteins have a very similar sequence to the reference swine H9 influenza virus, with an alignment score of 83 to 94 and mutations of 2 to 10. This level of similarity enables us to reasonably use homology model to build 3D model for HA1 protein based on reference structure. Structure homology modeling is implemented by software package Modeller version 9.4[39]. Then, all the structures are translated and rotated to maximize the overlap with the swine crystal model structure. In each model, a water molecule was carefully preserved because it was proved to mediate the formation of hydrogen bonds between the sialic acid and Gly-228[24].

Docking the LSTc to the binding pocket in H9N2 HA1 protein Since the focus of this work is to study the infection and transmission of virus in human body, we only used LSTc molecule as the ligand, and dock it with the selected target protein HA by using Autodock Vina (version 1.1.2)[40]. The Gasteiger partial charges were assigned to each atom of the protein and the LSTc ligand by the AutoDockTool version 4.2[41], a 30 × 30 × 30Å grid centered on atom C6 of the sialic acid and a parameter file for docking simulation were also generated by AutoDockTool. A total of 18 rotatable bonds were assigned to the molecules of LSTc penta-saccharide heterocyclic compounds, and the hydrogen atoms were fixed on their attached heavy atoms. Based on these parameters, Autodock can use Lamarckian genetic algorithm to search LSTc conformation on the binding site of HA1 protein through up to 200000 times of energy evaluation.

Considering the high structural similarity between the studied H9N2 proteins and the swine model, we expected them to have similar LSTc binding patterns. However, the preliminary docking simulation shows that AutoDock Vina often produced a very different conformation from the expected LSTc ligand, in which there is no identical conformation with the reference swine HA1-LSTc model. Therefore, for a given HA1 protein, we selected an LSTc conformation in the docking decoy set, which has the smallest atomic root mean square deviation (RMSD) compared with the reference swine LSTc model. Considering the importance of the binding of different units of the penta-saccharide to HA1 protein is also different，it is reasonable to assign the weight functions of 1.0, 1.0, 0.4, 0.4 and 0.2 to the five subunits of SIA-1, GAL-2, NAG-3, GAL-4, GLC-5 respectively when determining the RMSD value of the ligand. For each HA1 protein selected, two independent docking simulation calculations were carried out. Each docking simulation gave an optimal conformation based on the minimum RMSD calculation, and a total of two LSTc conformations were collected and compared. The resulting conformations were then used to analysis the effect of mutations on HA1-LSTc binding features.

Characterizing the binding microenvironment in the framework of local atom geometry In docking, contact distances are very important because they usually measure the overall strength of ligand binding affinity. In addition, other geometric features of conformation, such as relative contact orientation and convergence of orientation distribution, can also play an important role in the evaluation of ligand protein binding ability. In order to characterize these features, we constructed local affine coordinates by using CG, CD1 and CD2 of Leu226. The first two axes were vectors CG-CD1 and CG-CD2, and the third axis was formed by the cross product of these two vectors. A similar affine coordinate system was also built by using CD, OE1 and NE2 of Q226. The distribution of SIAO8, SIAO1A and SIAO1B were analyzed with these coordinates, and the binding modes of SIA with L226 and those with Q226 were compared.

Phylogenetic analysis of using key residual arrays Sequence alignment analysis revealed that only half of the sequences of the H9N2 strains studied were completely conserved, and 34% of the remaining half are partially conserved, that is, they have the same polarity but not the same amino acids. Among the many mutations, we noted in particular that only a few residues are actually involved in interacting with LSTc either directly or indirectly through some intermediate water molecules. To assess the importance and representativeness of these key mutations, we extracted them in each of the studied strains and produced a short amino acid sequence consisting of these mutations. We then constructed a new phylogenetic tree for these short sequences using the neighborhood-join method in MEGA5[42]. The phylogenetic tree thus constructed basically reflects the evolution process of the infection and transmission capacity of H9N2 strain as defined by LSTc binding features. On the other hand, we can measure the representativeness and completeness of the selected sites in determining the characteristics and functions of the studied strains by comparing the resulting phylogenetic tree with that obtained by full-length sequence alignment[11].

Results and Discussions

Homologous structures of the H9N2 HA1 proteins are very conservative

The sequence similarity between the HA1 proteins of H9N2 strained studied and the swine influenza HA1 protein was evaluated by the program Clustal W 2.0[43]. The alignment scores were determined to be between 83 and 94, and the alignment score for the identical sequences was set to 100. The three-dimensional structure superposition calculations showed that the homologous structures generated by the MODELLER program[39] is very similar to the X-ray crystal structure (PDB code 1JSI[24]) (see Figure 1). The overall structure of the SIA-binding domain of the HA1 protein, i.e. the residues between Ala109 and Gly252 as in PDB 1JSI, was found to be very conservative (see Figure S1), and the averaged backbone atomic root-mean-square-derivation (RMSD) between the models and the X-ray crystal structure is 0.22 Å with a standard error of 0.03Å. In particular, the side chain atoms of residues (such as Trp143) that are closely bound to SIA overlap well with the X-ray structure with the RMSDs smaller than 0.3Å. These results suggest that the overall 3D structures of H9N2 HA1 proteins, especially those portions bound to SIA, is very conservative, and that most mutations occur on residues with less exposure to SIA. Thus, it is reasonable to assume that the homologous modeling results in structures that bind to SIA in a similar manner, as observed in the X-ray structure of the H9 swine HA1-LSTc complex.

Comparison of LSTc conformations in two dockings
We noticed that due to the large size of the ligand binding molecule LSTc and the fact that the binding pocket is not deep inside the protein but on the surface, the three units of the second half of the penta-saccharide ligand, namely NAG-3, GAL-4 and GLC-5, fall on the outside of the binding pocket and are less constrained. A visual observation showed that the conformation of these three units predicted by docking varies greatly, while the calculated binding energy remains almost constant. Therefore, in order to determine the ligand-binding conformation in docking calculations, in addition to the low binding energy, we also require that the conformation of LSTc is the most similar to its conformation in the crystal structure. This requirement ensures the repeatability of the Vina docking outputs. Table 1 lists the results of the twice docking of LSTc with H9-HA1 proteins from studied strains. The results show that the LSTc conformations given by the two docking calculations are very close to each other. The following analyses on the ligand-binding features were then based on the complex structures of the second docking calculations.

Mutations in the HA1 protein cause extensive perturbation of LSTc binding Although the overall backbone structures of the H9 HA1 proteins studied are very conservative, we found that many of the mutations that occur in the binding pocket have a significant impact on the interaction between the protein and the sialic acid ligand. We categorized the strains studied into four groups based on the type and location of the mutations and their effects on protein-ligand interactions (Table 2).

The first group had only one member, A/brambling/Beijing/16/2012. In this model[24], the binding pocket has four highly conservative residues Tyr98, Ser136, Trp153, Leu190while there key mutations locate around LSTc: S137K, H227M and V190A (Figure 2a). In particular, the mutation S137K introduced a salt bridge between the negatively charged sialic acid unit of LSTc and the positively charged side chain of residue Lys137, which significantly stabilized the conformation of sialic acid in the pocket. This mutation also eliminated the hydrogen bond initially established between H227 and R220, which in turn weakened the contact between the two hemicycles of the 220-loop ring, and formed two new hydrogen bonds between the ring and a water molecule previously bound to the ligand. Compared to the hydrogen bond conformation in the X-ray structure (PDB code 1JSI), the water molecule rotated 90º and the bonding atom was shifted from SIA-1-O9 of LSTc to SIA-1-O8. And as a result, 220-loop moves closer to the ligand and establishes stronger interactions with LSTc. Mutation V190A reduces the Van der Waals radius of the side-chain that drive 190-helix close to the GlcNAc-3, however, the introduction of V190A did not bring observable change in the overall structure.

The second group had the largest number of strains studied, which were mainly characterized by the introduction of mutant strains H227Q and S137K, while the key residues L226 remained unchanged. These mutations introduce strong interactions between the amine groups on the side chain and the hydroxyl group of Gal-2, bringing the most twisted region of the ligand (between Gal-2 and SIA-1) closer to 220-loop (Figure 2b). Interestingly, this mutation had little effect on the SIA-1 conformation in the binding pocket and had very limited perturbation to the interaction between the key residue L226 and the ligand. This group also included strains with a V190A mutation that did not cause significant perturbation of the ligand conformation, as observed in group 1.

In the third group, both of the two adjacent residues L226 and H227 were replaced by glutamine, and all strains had the same amino acid distribution at the nine key sites except for some different mutations at site 190 (Figure 2c). Similar to group 2, the second saccharine unit, Gal2, was pulled towards Q227 in most members of the group 3. Compared with the X-ray structure, mutation L226Q made the interaction between side-chain of residue at 226 and SIA-1 more intense. In particular, in the case of A/duck/Fujian/T14/2007, SIA-1 was even pushed to the edge of the binding pocket， where SIA-1 was observed deep inside the bonding pocket in most other model structures, including the X-ray crystal structure. Taken together, the double mutation of L226Q and H227Q introduce relatively large conformation changes to the ligand in the binding pocket. Since all members of this group had nearly identical sequences at all nine key sites, the conformational instability of the LSTc can be attributed to mutations outside the binding pocket, especially the resides at 226 and 227.

All the remaining H9N2 strains studied formed the fourth group, which had two key mutations compared to the three above: A/V190E and N183H. Calculations showed that mutation A/V190E enables the residual 190 to form various contacts with the ligand, building hydrogen bonds between E190-OE and SIA1-O7/O9, GAL2-O2/O3/O4, NAG3-O3. These contacts made the ligand-binding conformation more complex than in the third group. The mutation N183H introduced the big imidazole side chain, which in many cases pushed SIA towards residue 226, making it easier for Q226 to come into contact with SIA (Figure 2d).

The geometric characteristics of the binding of residue 226 to LSTc

Residue 226 of HA1 protein has long been considered key to the recognition of particular types of host cells by influenza virus. Studies have shown that L226 is involved in the spread of H9 in humans and mink, while Q226 shows a higher affinity for avian hosts. For example, Wan and colleagues reported that H9N2 could be cultured in human respiratory epithelial cells [34]. The 226 sites of the HA protein are LEU amino acids, which are more infectious than GLN. The former can grow at a faster rate, up to 100 times the concentration of the latter. In order to understand the key role of these two amino acids on residue 226 in controlling the adaptability of influenza virus to different hosts, we used the local coordinated to characterize the detailed binding conformation between penta-saccharide ligands (especially the first two subunits SIA-1 and Gal-2) and these two residues.

Calculations revealed that the side chain atoms of Q226 and L226 formed a similar hydrogen bond and van der Waals interactions with the three atoms of LSTc, SIA-O1A, SIA-O1B and SIA-O18. However, there was a statistical difference in the interaction depending on the type of residues: the average contact distance between L226 and LSTc was smaller than that between Q226 and LSTc, indicating that the contact between Leu226 and SIA was stronger. Furthermore, the three heavy atoms of the Leu226 side chain, CD2, CD1 and CG, were all found within 5Å of the four SIA-1 heavy atoms, while only NE atoms of Gln226 could be found within this same range.

There were also statistical differences in the spatial orientation distribution of the atoms on the side chains of residue 226 of different types when interacting with the ligand (Figure 3). The relative spatial orientation of the SIA atom interacting with L226 was determined by the local coordinate system constructed by CG-CD1 and CG-CD2 of L226. The calculations showed that

SIA atoms in contact with L226 are concentrated in a small region and their orientations relative to L226 are fairly conservative, indicating that the interaction between L226 and the ligand is conservative and stable. One the contrary, as indicated by the coordinate system built by the side chains CD-OE1 and CD-NE2 of Q226, the distribution of SIA heavy atoms in contact with Q226 are very divergent and random, indicating that the interaction between Q226 and peta-saccharide is not conservative and unstable.

Phylogeny tree built based on the mutations of the key residues

We noted that the HA protein of the studied H9N2 strains contained a large number of mutations, far beyond the 9 mutations carefully examined in above grouping study (Figure S1). Given that these nine mutations are located around LSTc-binding pocket, we wondered whether the sequence of the nine residues largely determined the ability of the virus to infect. To address this problem, we compared the phylogenetic tree built by the sequence of the nine mutations and that constructed from the sequence of the entire H9N2-HA1 protein (Figure 4). First, the fourth group, characterized by two mutations of A/V190E and N183H, was well separated from the other groups in both phylogenetic trees. Second, the strain A/brambling/Beijing/16/2012 of group 1 located in the middle of the same sub-set of group 2 strains at a deeper level in both phylogenetic trees. Third, all of the members of group 3 had intricate relationships with those of group 2 in both trees, however, two members from group 2 (A/quail/Hong Kong/G1/1997 and A/chicken/Israel/178/2006) were well separated from the others on the first branch. Finally, in both trees, the deepest branches were composed of group 2 members, most of which were samples from 2007 and 2008. Overall, the two trees have very similar topologies on a number of key pieces of information, suggesting that the nine key mutation sequences we selected play a crucial role in determining the relationships between the strains.

Conclusion

In this study, we compared 40 HA protein structures of AIV H9N2 virus, mostly found in China, to understand its ability to infect humans. The three-dimensional structures of the protein areHA proteins of 40 H9N2 virus and analyzed interactions between these proteins and LSTc, in particular by characterizing key mutations around the ligand-binding pocket, by using homologous modeling and molecular docking programs. We categorized the mutations according to the perturbations they produced, which are mainly divided into those involved in direct ligand-binding interaction and those involved in indirect interaction. Both 200-loop and 190-Helix are intermediate elements that regulate the effect of mutations on LSTc-binding. The importance of the selected mutations was demonstrated by the fact that the phylogenetic tree constructed from the selected mutant sequences had a similar topological structure to that made from the entire protein sequence. Our calculations suggested that examining the molecular microenvironment perturbed by key mutations might provide a way to understand how the virus acquires the ability to infect humans through mutation at the molecular level.

Acknowledgements

This work was supported, in part, by the National Key Research and Development Program of China (Grant No. 2017YFC1600900, 2019YFA0905701) and by the Key University Science Research Project of Jiangsu Province (Grant No. 17KJA180005). X.H. was supported by National Students' platform for innovation and entrepreneurship training program (Grant No. 201910291055Z).

Reference:

1. Homme, P.J. and B.C. Easterday, *Avian influenza virus infections. I. Characteristics of influenza A-turkey-Wisconsin-1966 virus.* Avian Dis, 1970. **14**(1): p. 66-74.
2. Alexander, D.J., *A review of avian influenza in different bird species.* Vet Microbiol, 2000. **74**(1-2): p. 3-13.
3. Suss, J., et al., *Influenza virus subtypes in aquatic birds of eastern Germany.* Arch Virol, 1994. **135**(1-2): p. 101-14.
4. Chen, B., Z. Zhang, and W. Chen, *The study of avian influenza: I. The isolation and preliminary serological identification of avian influenza virus in chicken.* Chin J Vet Med, 1994. **20**: p. 3-5.
5. Zhang, Z., B. Chen, and W. Chen, *The study of avian influenza: II. The inci- dence and serological survey of avian influenza.* Chin J Vet Med, 1994. **20**( ): p. 6-7.
6. Chen, F. and C. Xia, *Cloning and analysis of avian influenza nucleo- protein gene from A/Chicken/Beijing/1/96(H9N2).* Chin J Prev Vet Med, 1999. **167**( ): p. 1-28.
7. Liu, J.H., et al., *Phylogenetic analysis of neuraminidase gene of H9N2 influenza viruses prevalent in chickens in China during 1995-2002.* Virus Genes, 2003. **27**(2): p. 197-202.
8. Tang, X., G. Tian, and C. Zhao, *Isolation and characterization of prevalent strains of avian influenza viruses in China.* Chin J Vet Med, 1998. **37**( ): p. 100-102.
9. Liu, H., et al., *Phylogenetic analysis of the hemagglutinin genes of twenty-six avian influenza viruses of subtype H9N2 isolated from chickens in China during 1996-2001.* Avian Dis, 2003. **47**(1): p. 116-27.
10. Zhang, P., et al., *A novel genotype H9N2 influenza virus possessing human H5N1 internal genomes has been circulating in poultry in eastern China since 1998.* J Virol, 2009. **83**(17): p. 8428-38.
11. Yuan, J., et al., *Characterization of an H9N2 avian influenza virus from a Fringilla montifringilla brambling in northern China.* Virology, 2015. **476**: p. 289-297.
12. Lam, T.T., et al., *Dissemination, divergence and establishment of H7N9 influenza viruses in China.* Nature, 2015. **522**(7554): p. 102-5.
13. Yu, H., et al., *Genetic diversity of H9N2 influenza viruses from pigs in China: a potential threat to human health?* Vet Microbiol, 2011. **149**(1-2): p. 254-61.

14. Zhang, K., et al., *Domestic cats and dogs are susceptible to H9N2 avian influenza virus.* Virus Res, 2013. **175**(1): p. 52-7.
15. Guo, Y.J., et al., *Characterization of the pathogenicity of members of the newly established H9N2 influenza virus lineages in Asia.* Virology, 2000. **267**(2): p. 279-88.
16. Sun, Y.P. and J.H. Liu, *H9N2 influenza virus in China: a cause of concern.* Protein & Cell, 2015. **6**(1): p. 18-25.
17. Fusaro, A., et al., *Phylogeography and evolutionary history of reassortant H9N2 viruses with potential human health implications.* J Virol, 2011. **85**(16): p. 8413-21.
18. Brown, C.J., et al., *Evolutionary rate heterogeneity in proteins with long disordered regions.* J Mol Evol, 2002. **55**(1): p. 104-10.
19. Bi, J., et al., *Phylogenetic and molecular characterization of H9N2 influenza isolates from chickens in Northern China from 2007-2009.* PLoS One, 2010. **5**(9).
20. Sun, Y., et al., *Genotypic evolution and antigenic drift of H9N2 influenza viruses in China from 1994 to 2008.* Vet Microbiol, 2010. **146**(3-4): p. 215-25.
21. Zhang, Y., et al., *Molecular and antigenic characterization of H9N2 avian influenza virus isolates from chicken flocks between 1998 and 2007 in China.* Vet Microbiol, 2012. **156**(3-4): p. 285-93.
22. Liu, D., et al., *Origin and diversity of novel avian influenza A H7N9 viruses causing human infection: phylogenetic, structural, and coalescent analyses.* Lancet, 2013. **381**(9881): p. 1926-32.
23. Mair, C.M., et al., *Receptor binding and pH stability - how influenza A virus hemagglutinin affects host-specific virus infection.* Biochim Biophys Acta, 2014. **1838**(4): p. 1153-68.
24. Ha, Y., et al., *X-ray structures of H5 avian and H9 swine influenza virus hemagglutinins bound to avian and human receptor analogs.* Proc Natl Acad Sci U S A, 2001. **98**(20): p. 11181-6.
25. Matrosovich, M.N., S. Krauss, and R.G. Webster, *H9N2 influenza A viruses from poultry in Asia have human virus-like receptor specificity.* Virology, 2001. **281**(2): p. 156-62.
26. Saito, T., et al., *Characterization of a human H9N2 influenza virus isolated in Hong Kong.* Vaccine, 2001. **20**(1-2): p. 125-33.
27. Eisen, M.B., et al., *Binding of the influenza A virus to cell-surface receptors: structures of five hemagglutinin-sialyloligosaccharide complexes determined by X-ray crystallography.* Virology, 1997. **232**(1): p. 19-31.
28. Weis, W., et al., *Structure of the influenza virus haemagglutinin complexed with its receptor, sialic acid.* Nature, 1988. **333**(6172): p. 426-31.
29. Kaverin, N.V., et al., *Structural differences among hemagglutinins of influenza A virus subtypes are reflected in their antigenic architecture: analysis of H9 escape mutants.* J Virol, 2004. **78**(1): p. 240-9.
30. Newhouse, E.I., et al., *Mechanism of glycan receptor recognition and specificity switch for avian, swine, and human adapted influenza virus hemagglutinins: a molecular dynamics perspective.* J Am Chem Soc, 2009. **131**(47): p. 17430-42.
31. Zhong, L., et al., *Molecular mechanism of the airborne transmissibility of H9N2 avian influenza A viruses in chickens.* J Virol, 2014. **88**(17): p. 9568-78.
32. Steinhauer, D.A., *Influenza: Pathways to human adaptation.* Nature, 2013. **499**(7459): p. 412-3.

33. Wan, H., et al., *Replication and transmission of H9N2 influenza viruses in ferrets: evaluation of pandemic potential.* PLoS One, 2008. **3**(8): p. e2923.
34. Wan, H. and D.R. Perez, *Amino acid 226 in the hemagglutinin of H9N2 influenza viruses determines cell tropism and replication in human airway epithelial cells.* J Virol, 2007. **81**(10): p. 5181-91.
35. de Graaf, M. and R.A. Fouchier, *Role of receptor binding specificity in influenza A virus transmission and pathogenesis.* EMBO J, 2014. **33**(8): p. 823-41.
36. Russell, R.J., et al., *Structure of influenza hemagglutinin in complex with an inhibitor of membrane fusion.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(46): p. 17736-17741.
37. Berman, H.M., et al., *The Protein Data Bank.* Nucleic Acids Research, 2000. **28**(1): p. 235-242.
38. Larkin, M.A., et al., *Clustal W and clustal X version 2.0.* Bioinformatics, 2007. **23**(21): p. 2947-2948.
39. Fiser, A. and A. Sali, *Modeller: generation and refinement of homology-based protein structure models.* Methods Enzymol, 2003. **374**: p. 461-91.
40. Trott, O. and A.J. Olson, *Software News and Update AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading.* Journal of Computational Chemistry, 2010. **31**(2): p. 455-461.
41. Morris, G.M., et al., *AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility.* Journal of Computational Chemistry, 2009. **30**(16): p. 2785-2791.
42. Tamura, K., et al., *MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods.* Molecular Biology and Evolution, 2011. **28**(10): p. 2731-2739.
43. Larkin, M.A., et al., *Clustal W and Clustal X version 2.0.* Bioinformatics, 2007. **23**(21): p. 2947-8.

Figures & Tables

Figure 1. a) the homologous model (red) built by Modeller is very similar to the X-ray crystal structure (green, PDB code 1JSI), especially at the LSTc binding domain. b) the conservative binding pocket where LSTc is shown with ball-and-stick model. This plot was prepared by VMD.

Figure 2. typical mutations and their effects on the HA1-LSTc binding interactions. a: type I, b: type II, c: type III and d: type IV.

Figure 3. the statistical distribution of the orientation between residue 226 and LSTc is dependent on the mutation. The orientation of the hydroxyl groups hanging outside the SIA-1 sugar ring is: a) variable with respect to the side chain of Q226 and b) relatively stable with respect to the side chain of L226.

Figure 4. Comparison of phylogeny tree of the studied H9N2 strains. The phylogenetic tree is: a) built based on the mutations of the key residues involving LSTc binding, b) using the whole sequence HA proteins.

Table 1. Comparison of the twice Vina docking

Strains	RMSD1 [a)]	RMSD2 [b)]	RMSF [c)]
1.A/brambling/Beijing/16/2012	1.97	1.88	0.20
2.A/chicken/Fujian/FH14/2007	3.04	3.08	0.18
3.A/chicken/Shandong/22/2008	2.86	2.93	0.22
4.A/chicken/hunan/HB25/2007	2.78	3.01	0.38
5.A/chicken/Anhui/Al16/2008	2.94	2.81	0.32
6.A/chicken/Shandong/lx316/2007	2.79	2.78	0.04
7.A/duck/Shantou/2144/2000	3.01	3.05	0.21
10.A/swine/Shandong/fNY/2003(2)	2.20	2.55	0.23
14.A/chicken/Hunan/HG29/2007	3.81	2.17	0.21
15.A/chicken/Yunnan/YC12/2007	4.44	4.52	0.33
16.A/chicken/Zhejiang/HJ/2007	3.69	2.85	0.24
17.A/chicken/Anhui/AH16/2008	3.10	2.54	0.09
18.A/chicken/Nanchang/4-301/2001	4.12	4.37	0.17
19.A/chicken/Shandong/LY/2003	4.06	4.39	0.34
23.A/swine/Shandong/fLS/2003	4.45	4.03	0.33
24.A/environment/Hunan/1-70/2007	3.38	3.11	0.22

25.A/environment/Hunan/1-70/2007	2.36	2.14	0.19
26.A/pigeon/Nanchang/2-0461/2000	3.05	4.11	0.31
27.A/Gf/HK/SSP607/2003	2.79	3.29	0.12
28.A/HK/2108/2003	3.68	3.08	0.15
29.A/quail/Hong Kong/G1/1997	4.30	3.34	0.38
30.A/chicken/Israel/178/2006	2.99	2.50	0.15
8.A/bird/Guangxi/A1/2006	2.53	3.09	0.10
9.A/quali/Guangxi/B1/2006	3.10	2.58	0.15
11.A/chicken/Shanghai/F/1998	4.98	4.94	0.22
12.A/chicken/Hubei/C1/2007	1.94	1.91	0.17
20.A/duck/Hubei/W1/2004	2.03	2.29	0.20
21.A/duck/Fujian/T14/2007	2.06	3.41	0.36
22.A/duck/Xuzhou/07/2003	3.34	4.18	0.23
31.A/n. s. /*Int. Ala.*/8BM3470/2008	2.75	3.21	0.20
32.A/chicken/Korea/S1/2003	3.23	3.46	0.17
33.A/chicken/Korea/S21/2004	3.66	3.36	0.23
34.A/chicken/Korea/S4/2003	3.95	3.69	0.25
35.A/duck/Hokkaido/49/1998	3.01	3.21	0.11
36.A/duck/Hong Kong/Y439/1997	3.43	3.13	0.31
37.A/ostrich/S. Africa/AI1586/2008	4.34	3.53	0.38
38.A/duck/Hokkaido/9/1999	4.45	4.05	0.39
39.A/duck/Hokkaido/238/2008	3.25	2.07	0.11
40.A/turkey/Wisconsin/1/1966	3.08	3.59	0.27

a)The RMSD between LSTc in crystal structure and that given by the first Vina docking.

b) The RMSD between LSTc in crystal structure and that given by the second Vina docking.

c) The RMSD between LSTcs derived from the two Docking calculations.

Table 2. The H9N2 strains are grouped according to the mutations around the binding pocket

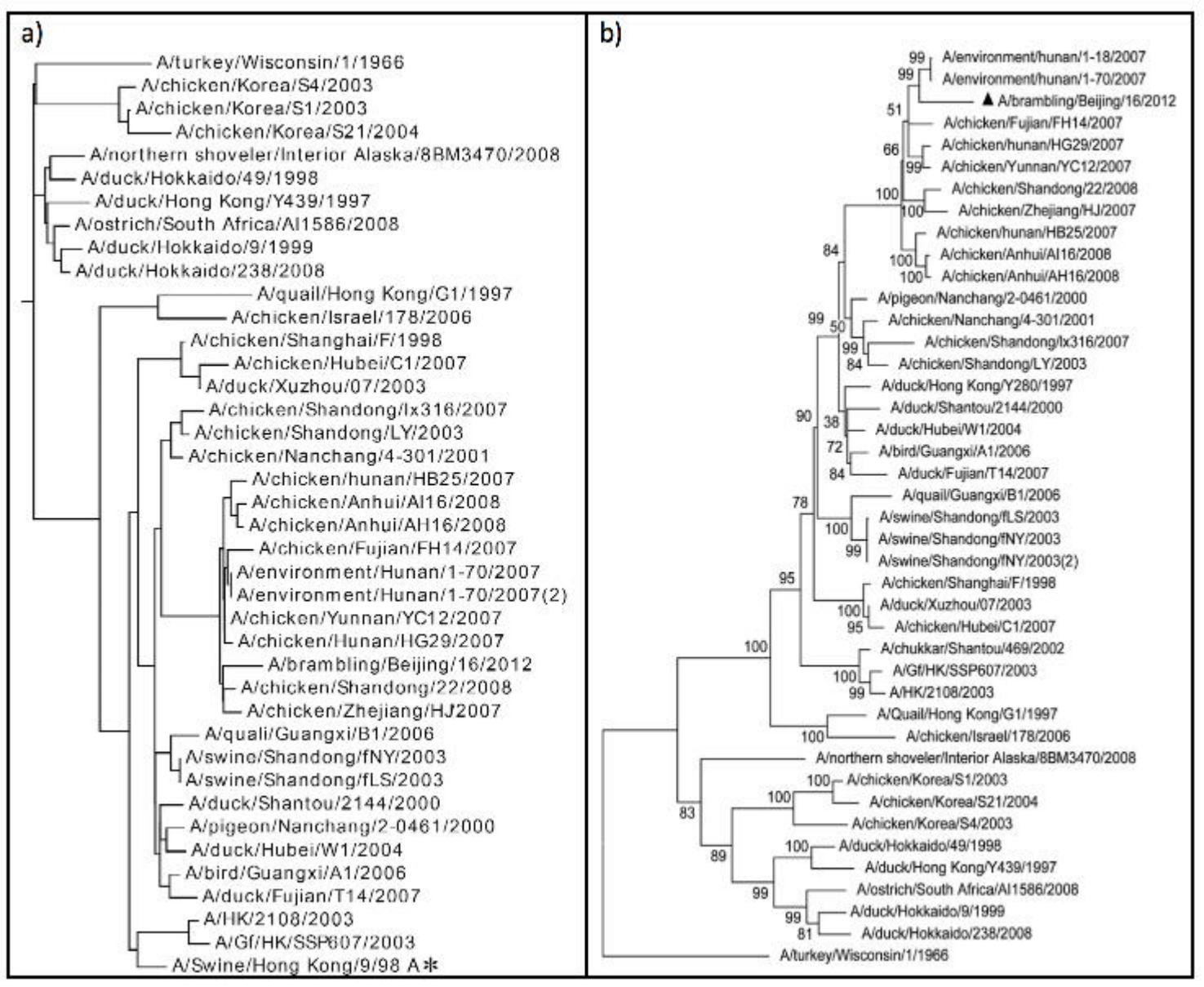
				130Loop	140loop	180-Helix		190-Helix	220-loop			
Group	H9N2 Strain	Lineage	Host	137	145	183	189	190	225	226	227	228
	A/Swine/Hong Kong/9/98 (1JSI)	BJ/94	Human	S	D	N	T	V	G	L	H	G
I	1.A/brambling/Beijing/16/2012	BJ/94	Brambling	K	D	N	T	A	G	L	M	G
II	2.A/chicken/Fujian/FH14/2007	BJ/94	Chicken	K	D	N	T	A	G	L	Q	G
	3.A/chicken/Shandong/22/2008	BJ/94	Chicken	K	D	N	T	A	G	L	Q	G
	4.A/chicken/hunan/HB25/2007	BJ/94	Chicken	K	D	N	T	A	G	L	Q	G
	5.A/chicken/Anhui/Al16/2008	BJ/94	Chicken	K	D	N	T	A	G	L	Q	G

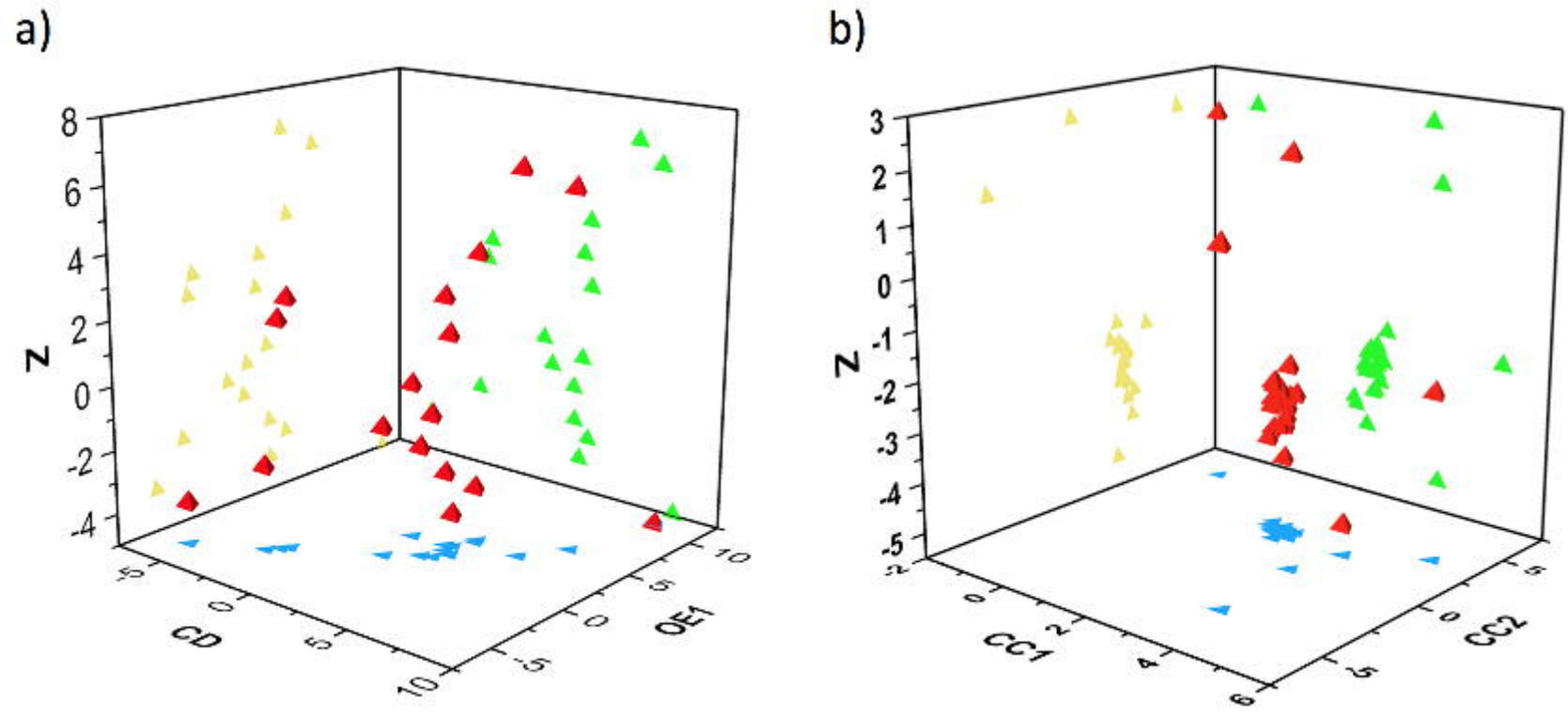
	6.A/chicken/Shandong/lx316/2007	BJ/94	Chicken	K	D	N	T	V	G	L	Q	G
	7.A/duck/Shantou/2144/2000	BJ/94	Duck	K	D	N	T	A	G	L	Q	G
	10.A/swine/Shandong/fNY/2003(2)	BJ/94	Swine	K	D	N	T	A	G	L	Q	G
	14.A/chicken/Hunan/HG29/2007	BJ/94	Chicken	K	D	N	T	A	G	L	Q	G
	15.A/chicken/Yunnan/YC12/2007	BJ/94	Chicken	K	D	N	T	A	G	L	Q	G
	16.A/chicken/Zhejiang/HJ/2007	BJ/94	Human	K	D	N	T	T	G	L	Q	G
	17.A/chicken/Anhui/AH16/2008	BJ/94	Chicken	K	D	N	T	A	G	L	Q	G
	18.A/chicken/Nanchang/4-301/2001	BJ/94	Human	K	D	N	T	A	G	L	Q	G
	19.A/chicken/Shandong/LY/2003	BJ/94	Human	K	D	N	T	A	G	L	Q	G
	23.A/swine/Shandong/fLS/2003	BJ/94	Swine	K	D	N	T	A	G	L	Q	G
	24.A/environment/Hunan/1-70/2007	BJ/94	Human	K	D	N	T	A	G	L	Q	G
	25.A/environment/Hunan/1-70/2007	BJ/94	Human	K	D	N	T	A	G	L	Q	G
	26.A/pigeon/Nanchang/2-0461/2000	Y439	Pigeon	K	D	N	T	V	G	L	Q	G
	27.A/Gf/HK/SSP607/2003	Y439	Human	N	N	N	T	V	G	L	Q	G
	28.A/HK/2108/2003	Y439	Human	S	N	N	T	V	G	L	Q	G
	29.A/quail/Hong Kong/G1/1997	BJ/94	Quail	R	G	H	T	E	D	L	Q	G
	30.A/chicken/Israel/178/2006	G1	Chicken	K	D	H	T	A	G	L	Q	G
III	8.A/bird/Guangxi/A1/2006	BJ/94	Bird	K	D	N	T	A	G	Q	Q	G
	9.A/quali/Guangxi/B1/2006	BJ/94	Quali	K	D	N	T	T	G	Q	Q	G
	11.A/chicken/Shanghai/F/1998	BJ/94	Human	K	D	N	T	A	G	Q	Q	G
	12.A/chicken/Hubei/C1/2007	BJ/94	Chicken	K	D	N	T	V	G	Q	Q	G
	20.A/duck/Hubei/W1/2004	BJ/94	Duck	K	D	N	T	V	G	Q	Q	G
	21.A/duck/Fujian/T14/2007	BJ/94	Duck	K	D	N	T	A	G	Q	Q	G
	22.A/duck/Xuzhou/07/2003	BJ/94	Duck	K	D	N	T	V	G	Q	Q	G
IV	31.A/n. s. /*Int. Ala.*/8BM3470/2008	BJ/94	Duck	K	S	H	T	E	G	Q	Q	G
	32.A/chicken/Korea/S1/2003	BJ/94	Chicken	K	D	H	T	E	G	Q	Q	G
	33.A/chicken/Korea/S21/2004	G1	Chicken	K	G	H	T	E	G	Q	Q	G
	34.A/chicken/Korea/S4/2003	G1	Chicken	K	N	H	T	E	G	Q	Q	G
	35.A/duck/Hokkaido/49/1998	Y439	Duck	K	N	H	T	E	G	Q	Q	G
	36.A/duck/Hong Kong/Y439/1997	Y439	Duck	R	N	H	T	E	D	Q	Q	G
	37.A/ostrich/S. Africa/AI1586/2008	Y439	Ostrich	K	D	H	T	E	G	Q	Q	G
	38.A/duck/Hokkaido/9/1999	Y439	Duck	K	N	H	T	E	G	Q	Q	G
	39.A/duck/Hokkaido/238/2008	Y439	Duck	K	N	H	T	E	G	Q	Q	G
	40.A/turkey/Wisconsin/1/1966	Y439	Turkey	R	N	H	T	E	G	Q	Q	G

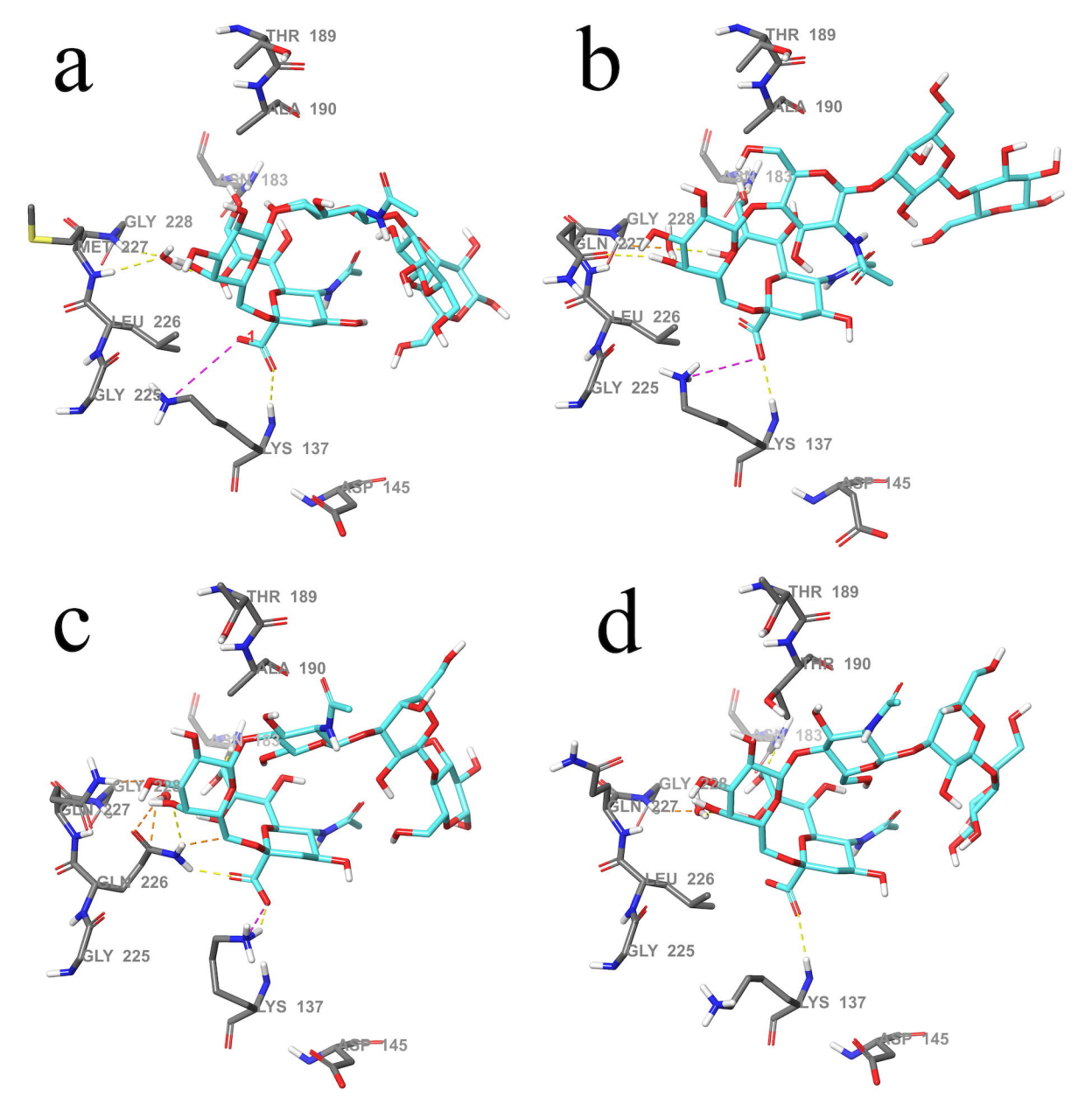
Supporting information

Figure S1. The alignment of the selected 40 H9N2 influenza A virus strains.

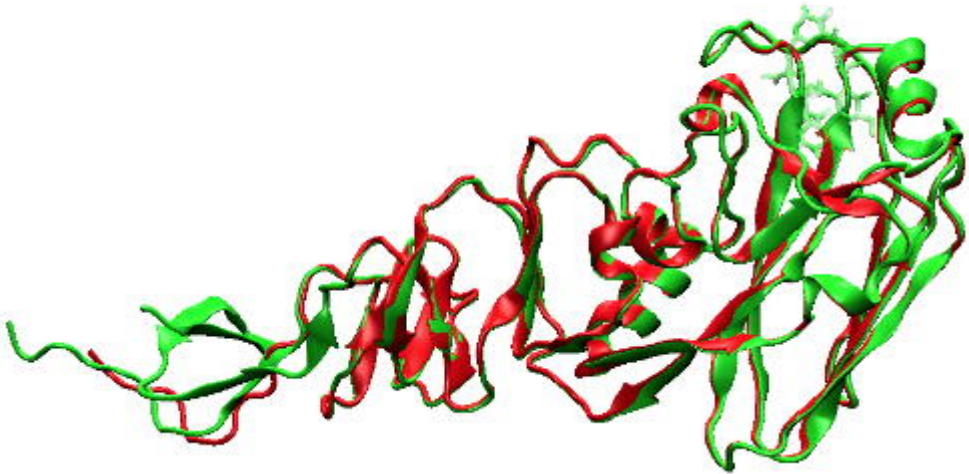
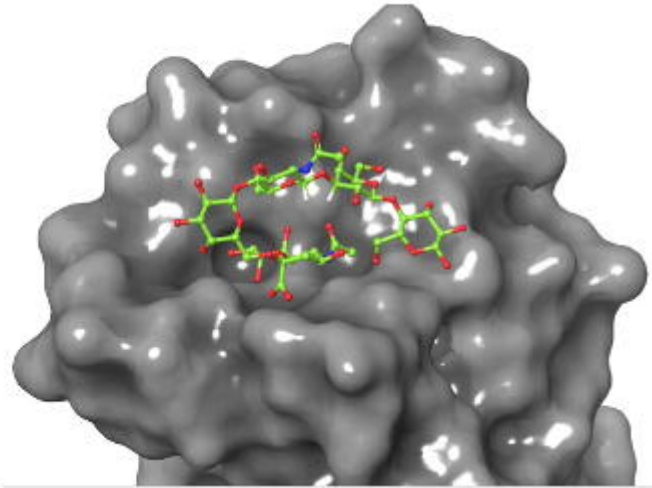
Table S1. list of the studied H9N2 strains


a)
A/turkey/Wisconsin/1/1966
A/chicken/Korea/S4/2003
A/chicken/Korea/S1/2003
A/chicken/Korea/S21/2004
A/northern shoveler/Interior Alaska/8BM3470/2008
A/duck/Hokkaido/49/1998
A/duck/Hong Kong/Y439/1997
A/ostrich/South Africa/AI1586/2008
A/duck/Hokkaido/9/1999
A/duck/Hokkaido/238/2008
A/quail/Hong Kong/G1/1997
A/chicken/Israel/178/2006
A/chicken/Shanghai/F/1998
A/chicken/Hubei/C1/2007
A/duck/Xuzhou/07/2003
A/chicken/Shandong/lx316/2007
A/chicken/Shandong/LY/2003
A/chicken/Nanchang/4-301/2001
A/chicken/hunan/HB25/2007
A/chicken/Anhui/AI16/2008
A/chicken/Anhui/AH16/2008
A/chicken/Fujian/FH14/2007
A/environment/Hunan/1-70/2007
A/environment/Hunan/1-70/2007(2)
A/chicken/Yunnan/YC12/2007
A/chicken/Hunan/HG29/2007
A/brambling/Beijing/16/2012
A/chicken/Shandong/22/2008
A/chicken/Zhejiang/HJ2007
A/quali/Guangxi/B1/2006
A/swine/Shandong/fNY/2003
A/swine/Shandong/fLS/2003
A/duck/Shantou/2144/2000
A/pigeon/Nanchang/2-0461/2000
A/duck/Hubei/W1/2004
A/bird/Guangxi/A1/2006
A/duck/Fujian/T14/2007
A/HK/2108/2003
A/Gf/HK/SSP607/2003
A/Swine/Hong Kong/9/98 A*
b)
99
99
51
66
99
100
100
84
100
100
99
50
99
84
90
38
72
84
78
100
99
95
100
95
100
99
100
83
100
100
89
100
99
99
81
A/environment/hunan/1-18/2007
A/environment/hunan/1-70/2007
A/brambling/Beijing/16/2012
A/chicken/Fujian/FH14/2007
A/chicken/hunan/HG29/2007
A/chicken/Yunnan/YC12/2007
A/chicken/Shandong/22/2008
A/chicken/Zhejiang/HJ/2007
A/chicken/hunan/HB25/2007
A/chicken/Anhui/AI16/2008
A/chicken/Anhui/AH16/2008
A/pigeon/Nanchang/2-0461/2000
A/chicken/Nanchang/4-301/2001
A/chicken/Shandong/lx316/2007
A/chicken/Shandong/LY/2003
A/duck/Hong Kong/Y280/1997
A/duck/Shantou/2144/2000
A/duck/Hubei/W1/2004
A/bird/Guangxi/A1/2006
A/duck/Fujian/T14/2007
A/quail/Guangxi/B1/2006
A/swine/Shandong/fLS/2003
A/swine/Shandong/fNY/2003
A/swine/Shandong/fNY/2003(2)
A/chicken/Shanghai/F/1998
A/duck/Xuzhou/07/2003
A/chicken/Hubei/C1/2007
A/chukkar/Shantou/469/2002
A/Gf/HK/SSP607/2003
A/HK/2108/2003
A/Quail/Hong Kong/G1/1997
A/chicken/Israel/178/2006
A/northern shoveler/Interior Alaska/8BM3470/2008
A/chicken/Korea/S1/2003
A/chicken/Korea/S21/2004
A/chicken/Korea/S4/2003
A/duck/Hokkaido/49/1998
A/duck/Hong Kong/Y439/1997
A/ostrich/South Africa/AI1586/2008
A/duck/Hokkaido/9/1999
A/duck/Hokkaido/238/2008
A/turkey/Wisconsin/1/1966


a)
Z
8
6
4
2
0
-2
-4
CD
-5
0
5
10
OE1
-5
0
5
10
b)
Z
3
2
1
0
-1
-2
-3
-4
-5
CC1
-2
0
2
4
6
CC2
-5
0
5


a
THR 189
ALA 190
ASN 183
GLY 228
MET 227
LEU 226
GLY 225
LYS 137
ASP 145
b
THR 189
ALA 190
ASN 183
GLY 228
GLN 227
LEU 226
GLY 225
LYS 137
ASP 145
c
THR 189
ALA 190
ASN 183
GLY 228
GLN 227
GLN 226
GLY 225
LYS 137
ASP 145
d
THR 189
THR 190
ASN 183
GLY 228
GLN 227
LEU 226
GLY 225
LYS 137
ASP 145

a)



b)