Towards a better understanding of the novel avian-origin H7N9 influenza A virus in China

Yongbo Wang[1,†], Han Cheng[1,†], Zexian Liu[2,†], Zhicheng Pan[1,†], Zhangyan Dai[3,†], Wankun Deng[1,†], Tianshun Gao[1,†], Xiaotong Li[1], Yuangen Yao[1], Jian Ren[4], Yu Xue[1,*]

[1]Department of Biomedical Engineering, College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China

[2]Hefei National Laboratory for Physical Sciences at Microscale and School of Life Sciences, University of Science and Technology of China, Hefei, Anhui 230027, China

[3]School of Chemistry and Molecular Biosciences, University of Queensland, Brisbane, Queensland 4027, Australia

[4]State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-sen University, Guangzhou, Guangdong 510275, China

***Running title*:** *Bioinformatic analysis of H7N9 virus*

[†]There authors shared first authorship.

*To whom correspondence should be addressed.

Yu Xue, Tel: +86-27-87793903, Fax: +86-27-87793172, E-mail: xueyu@hust.edu.cn.

Abstract

Recently, a highly dangerous bird flu has infected over 70 patients in China. Sequencing analysis revealed a novel avian-origin H7N9 virus to be responsible for the outbreak. In this work, we performed a systematically computational analysis of the virus. We clarified the controversial viewpoint on neuraminidase (NA) origin and confirmed it was reassorted from Korean wild birds with higher confidence, and raised an updated model of “three-step reassortant” for virus origins. Further analysis of NA sequences suggested that most variations are not drug resistant and current drugs are still effective for the therapy. We also identified a potentially optimal 9-mer epitope, which can be helpful for vaccine development. The interaction of hemagglutinin (HA) human receptor analog was confirmed by structural modeling, while NA might influence cellular processes through a PDZ-binding motif. A hypothetic virus infection model was also proposed. Taken together, our studies provide a better understanding of the newly reassorted H7N9 viruses.

Keywords: Influenza A virus, H7N9, hemagglutinin, neuraminidase, epitope, PDZ-binding

Supplementary materials: Supplementary Fig. S1 is attached.

Introduction

On 29 March 2013, Chinese Center for Disease Control and Prevention (China CDC) isolated and confirmed a new influenza A (H7N9) virus that had infected three Chinese patients, with two from Shanghai and one from Anhui Province. It's the first time that H7N9 viruses infect humans[1,2]. The Chinese media (Ecns.cn) reported the considerably rapid spread of the virus-caused flu. Before April 12, all infected cases were detected only in or near to the Yangtze river delta, while one seven-year-old girl in Beijing and two residents in Henan Province were newly confirmed with H7N9 bird flu on April 14. In particular, the husband of a Shanghai woman who died as a result of H7N9 infection on April 3 was also confirmed to be infected with the bird flu virus on April 11, and the human-to-human transmission can still not be fully excluded. In this regard, the anti-viral therapy and vaccines are urgently needed, whereas more analyses will be greatly helpful for better understanding the H7N9 subtype viruses[3-7]. To date (April 16), H7N9 viruses have infected up to 72 patients and killed 14 of them in China (Fig. 1).

Accurate determination of the origin of highly pathogenic avian influenza (HPAI) H7N9 viruses is the most crucial problem for preventing pandemic[3,6-8]. It's still not exactly known where H7N9 virus originated, but several lines of evidences have supported its avian-origin. On April 11, researchers from China and Japan published their back-to-back results for early findings of H7N9 viruses, separately[3,6]. In Gao's study, H7N9 viruses were isolated and sequenced from three Chinese patient samples, while phylogenetic analyses supported a triple reassortant model that H7N9 might be reassorted by hemagglutinin (HA) of H7N3 in Zhejiang duck, neuraminidase (NA) of H7N9 in Korean wild bird, and six internal genes (PB2, PB1, PA, NP, M and NS) of H9N2 in Beijing brambling[6]. Kageyama's results largely supported this model, whereas they drew a different conclusion that NA might be reassorted from mallard in Czech Republic[3]. Both analyses reported an R294K mutation of NA in A/Shanghai/1/2013 to be resistant to Oseltamivir (Tamiflu)[3,6,9]. In the NA stalk region of all H7N9 viruses, both studies detected a five amino acids deletion (69-73aa), which is associated with increased virulence[3,6,10]. In particular, both analyses revealed mutations in known receptor-binding sites (RBSs) of HA may increase the binding affinity of H7N9 viruses to human-type receptors[3,6].

In this work, we first performed phylogenetic analyses for each gene in newly sequenced

HPAI H7N9 viruses, respectively. The controversial viewpoint of Korea or Czech Republic-origin of NA gene was carefully evaluated and clarified. Based on previous studies[3,6] and our phylogenetic results, a more rigorous model of “three-step reassortant” was raised for the avian-origin H7N9 virus. By sequence alignment and analysis, genetic variations in HPAI H7N9 viruses were systematically characterized and can serve as a useful resource for further functional experiments. The structural modeling suggests that current anti-viral drugs such as Oseltamivir, Laninamivir, and Zanamivir are still effective for the therapy. Furthermore, combined with sequence-based prediction and structural docking approaches, we identified a potentially optimal 9-mer epitope, which might be helpful for vaccine development. Moreover, we detected a N9-conserved C-terminal class 2 PDZ-binding motif in H7N9 NA protein, which may regulate cellular processes through interacting with PDZ domain proteins. In addition, we confirmed the interaction between HA and avian/human receptor analog by structural modeling. Finally, a potential model for HPAI H7N9 virus infection was raised. Taken together, our analyses are helpful for better understanding the newly avian-origin H7N9 viruses, whereas the results can be useful for further experimental consideration.

Methods

Sequence data preparation

The nucleotide and protein sequences of 7 newly sequenced HPAI strains of H7N9 viruses were downloaded from the GISAID (Global Initiative on Sharing Avian Influenza Data) database on April 14. The 7 strains were isolated from 4 patients (A/Shanghai/1/2013 | EPI_ISL_138737, A/Shanghai/2/2013 | EPI_ISL_138738, A/Anhui/1/2013 | EPI_ISL_138739, and A/Hangzhou/1/2013 | EPI_ISL_138977), 2 birds (A/Chicken/Shanghai/S1053/2013 | EPI_ISL_138983, A/Pigeon/Shanghai/S1069/2013 | EPI_ISL_138985), and 1 environment (A/Environment/Shanghai/S1088/2013 | EPI_ISL_138984). Totally, we obtained 56 nucleotide sequences and 77 protein sequences for eight genes (HA, NA, PB2, PB1, PA, NP, M and NS) of H7N9 viruses. We also downloaded 253,527 nucleotide and 320,023 protein sequences of all influenza subtype viruses from the NCBI Influenza Virus Resource (ftp://ftp.ncbi.nih.gov/genomes/INFLUENZA/) on 9 April 2013[11]. To avoid any bias, only sequences annotated with “c” (complete sequence) were reserved. Moreover, CD-Hit[12], a program of clustering similar sequences, was used to clear redundancy with a threshold of 100% sequence identity. Finally, the non-redundant NCBI dataset contains 115,192 nucleotide sequences.

Phylogenetic analysis

Because too many influenza virus sequences are present and multiple alignments of all these sequences are quite time-consuming, here we adopted a simple approach for the phylogenetic analysis. For each gene, we used its nucleotide sequences of 7 newly sequenced strains to perform homologous searches in the non-redundant NCBI dataset, separately. Only top 200 hits were reserved for each sequence-based search. The results of 7 rounds of searches for each gene were merged together, while the redundant sequences were cleared. Totally, we obtained 208, 209, 215, 247, 220, 239, 209, and 214 non-redundant nucleotide sequences for HA, NA, PB2, PB1, PA, NP, M and NS, respectively. The Clustal Omega 1.1.0[13] was chosen for multiple sequence alignments with default parameters for each gene, separately. The Neighbor-Joining (NJ) tree was constructed with MEGA 5.10[14]. In addition, the maximum parsimony (MP) trees were also constructed for HA and NA genes. The default parameters were used, while the number of Bootstrap Replications was selected

as 1000.

Non-synonymous variation analysis

We first analyzed non-synonymous variations between patient and non-pathogenic samples. Because the results of multiple sequence alignments exhibited that the sequences of A/Shanghai/1/2013 are quite different from other patient samples, we separated 4 patient samples into two groups, including the A/Shanghai/1/2013 group and the second group. The sequences of A/Anhui/1/2013 were selected as the representative sequences for the second group. Based on phylogenetic results, the nearest non-pathogenic samples were chosen as benchmark sequences for comparison. Again, because the NA sequences among Korean wild birds are highly divergent, we further aligned 5 NA sequences from Korean wild bird H7N9 viruses (A/wild bird/Korea/A14/11 | EPI_ISL_120868, A/wild bird/Korea/A3/11 | EPI_ISL_120869, A/spot-billed duck/Korea/447/11 | EPI_ISL_120871, A/wild bird/Korea/A9/11 | EPI_ISL_120881, and A/wild duck/Korea/SH20-27/2008 | EPI_ISL_133001). The identical positions in the alignment results were adopted as a benchmark for identifying NA variations in patient samples. We defined variation patterns among the A/Shanghai/1/2013 group, non-pathogenic group, and A/Anhui/1/2013 group as 4 types, including B-A-C (one variation in the A/Shanghai/1/2013 group and another variation in the A/Anhui/1/2013 group), B-A-A, B-A-B, and A-A-B. In addition, we aligned sequences of 4 patient samples together for each protein separately to identify non-synonymous variations among patient samples.

Structural modeling

Modeller 9.11[15,16] was employed for homology modeling the structures of HA, NA and the complexes. From the PDB database[17], previous reported H7 hemagglutinin structures including 4DJ6, 4DJ7, 4DJ8 from A/Netherlands/219/2003 (H7N7)[18] were downloaded and used as the templates for modeling HA monomer, HA in complex with avian receptor analog (3'-sialyl-*N*-acetyllactosamine [3'SLN]), and HA in complex with human receptor analog (6'-sialyl-*N*-acetyllactosamine [6'SLN]) in H7N9 virus. To dissect the key residues of NA for interacting drugs including Oseltamivir, Zanamivir and Laninamivir, known complex structures of 2QWK[19], 3TI3[20] and 3TI5[20] were selected as the templates for modeling NA-Oseltamivir, NA-Laninamivir and NA-Zanamivir complexes in H7N9 virus, respectively. The HA and NA protein

sequences of A/Anhui/1/2013 were chosen as the representative sequences for structural modeling. All the modeled structures were validated by Procheck[21].

Epitope prediction and docking

The CTL epitope predictor NetCTL 1.2 Server[22] was used for the prediction of MHC class I binding peptides for all HA protein sequences of 7 newly sequenced strains of H7N9 viruses. The B44 supertype was chosen, while the default parameters were adopted. Because the sequence-based prediction is not reliable and false positive hits can not be avoided, we additional performed structural modeling and docking analyses. Since the predicted epitopes are identical across the 7 strains, here the HA protein sequence of A/Anhui/1/2013 was chosen for further analyses. The 3D structures of HLA-B*4405 (3DX8) and DM1-TCR in complex with HLA-B*4405 (3DXA) were downloaded from the PDB database[17]. As previously described[23], we used ClusPro 2.0[24] to dock all predicted epitopes to HLA-B*4405, and then regarded the HLA-B*4405/epitope complex as an integrated ligand and further docked it to DM1-TCR. The binding energy scores were automatically calculated by ClusPro[24].

Results

A three-step reassortant model for avian-origin H7N9 viruses

In this work, NJ trees for eight genes of newly sequenced H7N9 viruses were constructed and visualized (Supplementary Fig. S1). To avoid any bias, MP trees were also performed for HA and NA genes (Fig. 2a,b, Supplementary Fig. S1a,b). The NJ trees of HA and NA generate the same results (Supplementary Fig. S1c,d). Although different tree-constructing methods were used, our analyses are highly consistent with Gao's results[6], which suggested that HA was reassorted from H7N3 in Zhejiang duck (Fig. 2a, Supplementary Fig. S1a) and NA was from H7N9 in Korean wild bird (Fig. 2b, Supplementary Fig. S1b). Also, our phylogenetic results of six internal genes do not fully support the Gao's hypothesis that all of them were originated from Beijing brambling[6], because only three genes of PB1 (Fig. 2C, Supplementary Fig. S1e), PB2 (Supplementary Fig. S1f), and PA (Supplementary Fig. S1g) in HPAI H7N9 viruses are nearest from the H9N2 sample of A/brambling/Beijing/16/2012. The other three genes of NP (Supplementary Fig. S1h), M (Supplementary Fig. S1i) and NS (Supplementary Fig. S1j) are not nearest from Beijing brambling, and might be originated from poultries in the Yangtze river delta.

Based on currently available data and phylogenetic results, here we refined the Gao's "triple reassortant" model[6] into a "three-step reassortant" model (Fig. 3): 1) The reassortment of six internal genes might occur in the Yangtze river delta, start before 11/2012 and still not finalize after the date, because A/brambling/Beijing/16/2012 was sequenced in November 2012 (From its annotation information); 2) HA of H7N3 subtype was reassorted with six genes of H9N2 subtype in Zhejiang Province after November 2012, because H7N9 subtype HA was not reassorted from A/brambling/Beijing/16/2012; 3) The reassortment of NA from H7N9 virus in Korean wild bird occurred in Shanghai before 19/02/2013, the date when the first Shanghai patient was infected. The final step would not occur at Zhejiang, because no HPAI strains of Zhejiang contain N9 NA genes.

A landscape of genetic variations in H7N9-infected patient samples

A number of genetic mutations or variations have been detected in recent studies[3,6]. Gao *et al*. observed a receptor specificity changing variation of Q226L in HA, an anti-viral resistance variation of R294K in NA, and a ferret-transmissible variation of I368V in PB1 among three patient samples[6].

All three samples have a 5aa deletion in the NA stalk region[6], and this result was also confirmed by Kageyama's analysis[3]. Because the time was limited and only three strains were analyzed, mutations or variations might not be fully demonstrated. Here, we performed a more comprehensive analysis of non-synonymous variations between patient and non-pathogenic samples (Fig. 4). Totally, we characterized that point non-synonymous variations occur at up to 38 positions of 9 viral proteins, while M1 and NS2 are not mutated (Fig. 4). The 5aa deletion in NA was also detected. The variation patterns were analyzed, while the B-A-A type variations are predominant. Thus, the variation rate of A/Shanghai/1/2013 is greater than other patient samples[6] (Fig. 4). From multiple alignment results of 5 NA sequences in Korean wild bird samples, non-synonymous variations occur at 66 positions and implicate a considerably high evolutionary speed. However, analysis of multi-alignments of 4 patient samples only identified 3 non-synonymous variations. Thus, NA gene does not exhibit an accelerated evolution after infecting humans. Furthermore, we identified that non-synonymous variations occur at 5 additional positions among 4 patient samples (Fig. 4). All well-documented drug resistant, virulence increasing and viral protein function-changing variations were fully covered by our results[3,6]. Other uncharacterized variations can be useful for further experimental consideration.

Structural modeling suggests most variations are not drug-resistant

Anti-viral therapy with drugs might be the most effective treatment for H7N9 bird flu before vaccine development[3-7]. Because NA evolves rapidly at least in wild birds, it will be very dangerous if anti-viral resistance is conferred by genetic variations. Indeed, an R294K variation of NA in A/Shanghai/1/2013 is known to influence the NA-Oseltamivir (Tamiflu) interaction through disrupting the internal salt link between E276 and R224, and generate anti-viral resistance[3,6]. To test whether there are more drug-resistant variations, a previously characterized 3D structure of N9 NA-Oseltamivir complex (2QWK, Fig. 5a)[19] was used as a template for modeling H7N9 NA-Oseltamivir interaction (2QWK, Fig. 5b). Also, two structures of 2009 pandemic H1N1 NA-Laninamivir (3TI3), NA-Zanamivir were adopted (3TI5) for modeling the interactions of H7N9 NA and the two drugs, respectively (Fig. 5c,d)[20]. As previously described[20], key residues for drug interaction were directly determined and visualized from the structures, with a distance of ≤ 4 Å to drugs (Fig. 5). Totally, we identified 16 drug-interacting residues that are present in at least one complex (Table 1). The comparison of NA-Oseltamivir and H7N9 NA-Oseltamivir complexes

suggested that most interacting residues are not changed, and the distances to drug are slightly increased (Table 1). Thus, it can be expected that the binding affinity of H7N9 NA-Oseltamivir will be only moderately weaken. From the results, we did not observe any additional drug resistant variations, except the R294K in A/Shanghai/1/2013 (Table 1). Our results can be used as a reference for further detection of potential drug resistance if new variations appear.

Prediction of potential peptide epitopes for further vaccine design

As an important surface protein, HA is the major molecule for vaccine developed[23,25]. In a recent study, Su *et al.* performed a structural docking analysis of H1N1 and H5N1 epitopes that bind with HLA-B*4405 and DM1-TCR/HLA-B*4405 complex[23], because T cell-mediated responses can be greatly enhanced by DM1-TCR/HLA-B*4405/epitope complex[26]. Using a sequence-based tool[22], we totally predicted H7N9 HA with 8 T cell epitopes that potentially interact with HLA class I B44 supertype molecules (Table 2). As previously described[23], each predicted peptide was first docked to HLA-B*4405, and HLA-B*4405/peptide complex was then docked to DM1-TCR. The results were ranked based on calculated binding energy scores for docking to HLA-B*4405 (Table 2). Thus, although epitope 441 achieved the highest score from the sequence-based prediction, it's not an optimal epitope for its high binding energy (Table 2). From the docking results, we observed that both epitope 88 and 353 are embedded in the binding groove of HLA-B*4405 (Fig. 6a,b). However, epitope 88 does not directly interact with DM1-TCR (Fig. 6c), while epitope 353 interacts with both DM1-TCR and HLA-B*4405 via E354, G355, L356 and D358 (Fig. 6d). Thus, it can be expected the DM1-TCR binding will increase the binding energy for epitope 88. Indeed, the binding energy of HLA-B*4405/epitope 88 was significantly enhanced from -904.6 to -643.9 in DM1-TCR/HLA-B*4405/epitope 88, while the results for epitope 353 are slightly changed (Table 1). Taken together, the epitope 353 is an optimal candidate for further optimization or vaccine design.

H7N9 virus may increase virulence through a PDZ-binding motif at NA C-terminus

In Kageyama's results, a 218-230 deletion was detected at the H7N9 NS1 C-terminus[3]. This deletion removes a well characterized PDZ-binding motif, which determines the virulence by interacting with cellular PDZ domain proteins at least in H1N1 and H5N1, but does not influence the viral replication[27-29]. Thus, lack of the PDZ-binding motif will be expected to decrease the

pathogenicity of H7N9 viruses[3]. However, this result can not interpret why the fatal rate is high in China. From ELM (Eukaryotic Linear Motif) resource, we obtained three known C-terminal PDZ-binding motifs, including S/T-X-A/C/V/I/L/F$ (class 1, $: the end of the sequence), V/L/I/F/Y-X-A/C/V/I/L/F$ (class 2), and D/E-X-A/C/V/I/L/F$ (class 3)[30]. We scanned all proteins in H5N1 and H7N9 viruses with the three motifs. We observed that all HA proteins contain C-terminal class 2 PDZ-binding motifs, while class 1 motifs are not present in H7N9 NS1 sequences (Fig. 7a). Unexpectedly, H7N9 NA has a class 2 motif (YFL) (Fig. 7a), which is conserved in almost all N9 subtype NA proteins and may still regulate the virulence even if the NS1 does not interact with PDZ domain proteins any longer. The 3D structure of a typical NS1-binding PDZ domain in PDlim2 (3PDV) was visualized, while the binding residues were shown (Fig. 7b). Also, the PDZ-binding motif in H7N9 NA structure modeled from NA-Oseltamivir complex was highlighted (Fig. 7c). The secondary structure of H7N9 NA C-terminal was determined as a 3aa helix by a secondary structure assignment software of STRIDE (Fig. 7d)[31].

Discussion

In this work, we first clarified a controversial viewpoint on the origin of HPAI H7N9 NA[3,6]. Using 7 newly sequenced H7N9 NA sequences to perform BLAST searches will always identify the NA sequences in A/mallard/Czech Republic/13438-29K/2010(H11N9) as the best hit, with an average identity of 96%. The A/wild bird/Korea/A3/2011(H7N9) NA only exhibits an average identify of 93% with HPAI H7N9 NAs. However, estimation of the origin based on the BLAST search is not reliable, because a point variation can be changed back to the original state by another variation, and mutation rates are usually not equal for the transition (exchange of A <-> G or C <-> T) and transversion (C/T <-> A/G)[32]. Thus, the origins of HPAI H7N9 viruses should be determined by phylogenetic analysis, and our studies confirmed that Gao's results are more reliable for NA origin[6].

Gao *et al*. and Kageyama *et al*. observed that genetic variations in HPAI H7N9 viruses have conferred the interaction between HA and avian-like receptors in humans[3,6]. Since the complex structures of H7N7 NA with avian and human receptor analogs were resolved separately[18], we used the two structures as templates to model H7N9 NA in complex with two types of receptors (Fig. 8a,b). The results suggested that two modeled structures are highly similar with original complexes. In this regard, our results proposed a hypothetic model that how HPAI H7N9 viruses infect human cells (Fig. 9). First, HA binds to oligosaccharide 6'SLN of human receptors, mediates H7N9 virus uptake, and orchestrates membrane fusion. After viruses entered cells, NA will further interact with PDZ domain proteins through a C-terminal class 2 PDZ-binding motif to regulate the cellular processes (Fig. 9).

Taken together, beyond clarifying the controversial results of NA origin, our analyses further demonstrated that most variations in HPAI H7N9 viruses are not anti-viral resistant, whereas the current drugs are still effective for therapy. Moreover, the prediction and structural docking of epitopes will be useful for vaccine development. In addition, the high virulence of HPAI H7N9 viruses might be attributed to a newly evolved PDZ-binding motif at NA C-terminus. Our study is helpful for better understanding the virus and further experimental design.

Acknowledgements

The authors thank Prof. Jiankui He (SUSTC) for kindly providing their pre-printed manuscript on H7N9 virus analysis. We also thank Prof. Anyuan Guo (HUST) for his computing server. This work was supported by grants from the National Basic Research Program (973 project) (2010CB945401, 2012CB911201, and 2012CB910101), and Natural Science Foundation of China (31171263, 81272578, 31071154, 30900835, 30830036, and 91019020).

References

1 Parry, J. H7N9 avian flu kills seven and infects 23 in China. *BMJ* **346**, f2222 (2013).

2 Parry, J. H7N9 avian flu infects humans for the first time. *BMJ* **346**, f2151 (2013).

3 Kageyama, T. *et al.* Genetic analysis of novel avian A(H7N9) influenza viruses isolated from patients in China, February to April 2013. *Euro Surveill.* **18** (2013).

4 Nicoll, A. & Danielsson, N. A novel reassortant avian influenza A(H7N9) virus in China – what are the implications for Europe. *Euro Surveill.* **18** (2013).

5 Uyeki, T. M. & Cox, N. J. Global Concerns Regarding Novel Influenza A (H7N9) Virus Infections. *N. Engl. J. Med.* doi:10.1056/NEJMp1304661 (in the press).

6 Gao, R. *et al.* Human Infection with a Novel Avian-Origin Influenza A (H7N9) Virus. *N. Engl. J. Med.* doi:10.1056/NEJMoa1304459 (in the press).

7 Cohen, J. New Flu Virus in China Worries and Confuses. *Science* **340**, 129-130 (2013).

8 Butler, D. Urgent search for flu source. *Nature* **496**, 145-146 (2013).

9 McKimm-Breschkin, J. L. *et al.* Mutations in a conserved residue in the influenza virus neuraminidase active site decreases sensitivity to Neu5Ac2en-derived inhibitors. *Journal of Virology* **72**, 2456-2462 (1998).

10 Matsuoka, Y. *et al.* Neuraminidase stalk length and additional glycosylation of the hemagglutinin influence the virulence of influenza H5N1 viruses for mice. *Journal of Virology* **83**, 4704-4708 (2009).

11 Bao, Y. *et al.* The influenza virus resource at the National Center for Biotechnology Information. *Journal of Virology* **82**, 596-601 (2008).

12 Fu, L., Niu, B., Zhu, Z., Wu, S. & Li, W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* **28**, 3150-3152 (2012).

13 Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* **7**, 539 (2011).

14 Tamura, K. *et al.* MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution* **28**, 2731-2739 (2011).

15 Eswar, N. *et al.* Comparative protein structure modeling using MODELLER. *Current protocols in protein science / editorial board, John E. Coligan ... [et al.]* **Chapter 2**, Unit 2 9 (2007).

16 Eswar, N. *et al.* Comparative protein structure modeling using Modeller. *Current protocols in bioinformatics / editoral board, Andreas D. Baxevanis ... [et al.]* **Chapter 5**, Unit 5 6, (2006).

17 Velankar, S. *et al.* PDBe: Protein Data Bank in Europe. *Nucleic Acids Res.* **40**, D445-452 (2012).

18 Yang, H., Carney, P. J., Donis, R. O. & Stevens, J. Structure and receptor complexes of the hemagglutinin from a highly pathogenic H7N7 influenza virus. *Journal of Virology* **86**, 8645-8652 (2012).

19 Varghese, J. N. *et al.* Drug design against a shifting target: a structural basis for resistance to inhibitors in a variant of influenza virus neuraminidase. *Structure* **6**, 735-746 (1998).

20 Vavricka, C. J. *et al.* Structural and functional analysis of laninamivir and its octanoate prodrug reveals group specific mechanisms for influenza NA inhibition. *PLoS Pathogens* **7**, e1002249 (2011).

21 Laskowski, R. A., Rullmannn, J. A., MacArthur, M. W., Kaptein, R. & Thornton, J. M. AQUA

and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *Journal of Biomolecular NMR* **8**, 477-486 (1996).

22 Larsen, M. V. *et al.* Large-scale validation of methods for cytotoxic T-lymphocyte epitope prediction. *BMC Bioinformatics* **8**, 424 (2007).

23 Su, C. T., Schonbach, C. & Kwoh, C. K. Molecular docking analysis of 2009-H1N1 and 2004-H5N1 influenza virus HLA-B*4405-restricted HA epitope candidates: implications for TCR cross-recognition and vaccine development. *BMC Bioinformatics* **14 Suppl 2**, S21 (2013).

24 Kozakov, D. *et al.* Achieving reliability and high accuracy in automated protein docking: ClusPro, PIPER, SDU, and stability analysis in CAPRI rounds 13-19. *Proteins* **78**, 3124-3130 (2010).

25 Subbramanian, R. A., Basha, S., Shata, M. T., Brady, R. C. & Bernstein, D. I. Pandemic and seasonal H1N1 influenza hemagglutinin-specific T cell responses elicited by seasonal influenza vaccination. *Vaccine* **28**, 8258-8267 (2010).

26 Archbold, J. K. *et al.* Natural micropolymorphism in human leukocyte antigens provides a basis for genetic control of antigen recognition. *J. Exp. Med.* **206**, 209-219 (2009).

27 Golebiewski, L., Liu, H., Javier, R. T. & Rice, A. P. The avian influenza virus NS1 ESEV PDZ binding motif associates with Dlg1 and Scribble to disrupt cellular tight junctions. *Journal of Virology* **85**, 10639-10648 (2011).

28 Jackson, D., Hossain, M. J., Hickman, D., Perez, D. R. & Lamb, R. A. A new influenza virus virulence determinant: the NS1 protein four C-terminal residues modulate pathogenicity. *Proc. Natl Acad. Sci. USA* **105**, 4381-4386 (2008).

29 Soubies, S. M. *et al.* Deletion of the C-terminal ESEV domain of NS1 does not affect the replication of a low-pathogenic avian influenza virus H7N1 in ducks and chickens. *The Journal of General Virology* **94**, 50-58 (2013).

30 Puntervoll, P. *et al.* ELM server: A new resource for investigating short functional sites in modular eukaryotic proteins. *Nucleic Acids Res.* **31**, 3625-3630 (2003).

31 Heinig, M. & Frishman, D. STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res.* **32**, W500-502 (2004).

32 Nei, M. Phylogenetic analysis in molecular evolutionary genetics. *Annual Review of Genetics* **30**, 371-403 (1996).

Figures

Fig. 1 – The numbers of confirmed infection and death of H7N9 patients in China (till April 16, 2013).

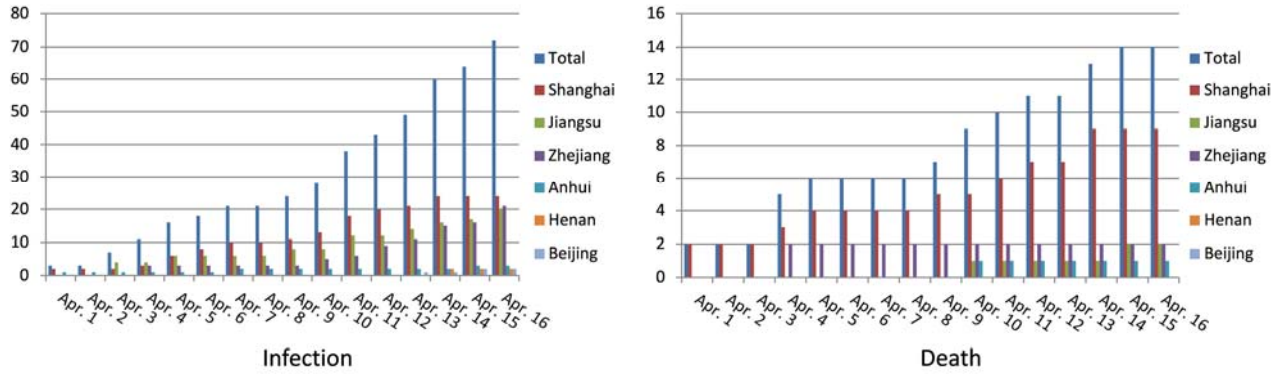


Fig. 2 – Phylogenetic analyses of H7N9 virus genes. The MP trees were present for (a) HA and (b) NA. (c) The NJ tree was visualized for PB1. More detailed phylogenetic trees were shown in Supplementary Fig. S1.

Fig. 3 – A “three-step reassortant” model of avian-origin H7N9 viruses. The six internal genes were reassorted from brambling and poultries in the Yangtze river delta before November 2012. HA of H7N3 was reassorted with the internal genes in poultries after November 2012. The final step for the reassortment of H7N9 NA from Korean wild bird occurred in Shanghai before 19/02/2013.

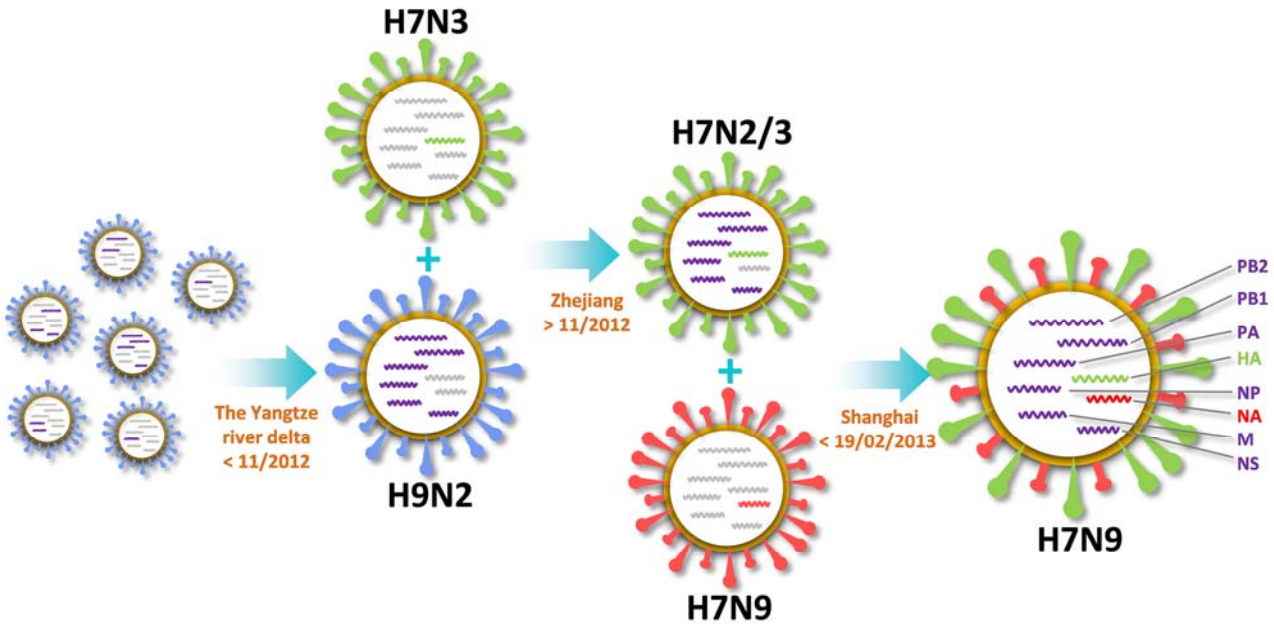


Fig. 4 – A variation map of HPAI H7N9 genes in patient samples. We analyzed variations between patients and non-pathogenic samples and among 4 patients, respectively. The variation patterns were also counted. *a*. The benchmark HA sequence was taken from A/duck/Zhejiang/11/2011(H7N3); *b*. Using identical positions of alignment results for 5 Korean wild bird H7N9 NA sequences as the background; *c*. From A/chicken/Hebei/1102/2010(H5N2); *d*. From A/chicken/Zhejiang/607/2011 (H9N2); *e*. The statistics of variation patterns confirmed that H7N9 virus in A/Shanghai/1/2013 evolves more rapidly.

Variations between patient and non-pathogenic samples

HA	146 (138)	183	195 (186)	230	285	292	298	335	410	455	541
Anhui/1	A (GCA)	S (AGC)	V (GTA)	P (CCA)	N (AAT)	H (CAT)	I (ATA)	I (ATT)	N (AAT)	D (GAC)	V (GTA)
Shanghai/1	S (TCA)	N (AAC)	G (GGA)	T (ACA)	D (GAT)	Y (TAT)	I (ATA)	I (ATT)	T (ACT)	D (GAC)	A (GCA)
H7N3 [a]	A (GCA)	D (GAC)	G (GGA)	P (CCA)	D (GAT)	H (CAT)	V (GTA)	T (ACT)	T (ACT)	N (AAT)	A (GCA)

NA	68-69	79	289	396	239	375	406	408	371	NP
Anhui/1	deletion	N (AAT)	R (AGG)	A (GCT)	M (ATG)	D (GAC)	V (GTC)	V (GTT)	M (ATG)	Anhui/1
Shanghai/1	deletion	N (AAT)	K (AAG)	A (GCT)	V (GTG)	E (GAA)	I (ATC)	I (ATT)	I (ATA)	Shanghai/1
H7N9 [b]	QISNT	G (GGT)	R (AGG)	T (ACT)	M (ATG)	D (GAC)	V (GTC)	V (GTT)	I (ATA)	H5N2 [c]

PB1	368	525	598	637	395	461	559	627	PB2
Anhui/1	V (GTA)	V (GTT)	L (CTG)	I (ATT)	A (GCC)	I (ATC)	N (AAT)	K (AAG)	Anhui/1
Shanghai/1	I (ATA)	V (GTT)	M (ATG)	V (GTT)	S (TCC)	V (GTC)	T (ACT)	K (AAG)	Shanghai/1
H9N2 [d]	V (GTA)	I (ATT)	L (CTG)	I (ATT)	A (GCC)	I (ATC)	T (ACT)	E (GAG)	H9N2

PB1-F2	37	42	51	71	180	NS1
Anhui/1	Q (CAA)	Y (TAC)	M (ATG)	E (GAA)	I (ATC)	Anhui/1
Shanghai/1	Q (CAA)	C (TGC)	T (ACG)	K (AAA)	I (ATC)	Shanghai/1
H9N2	R (CGA)	C (TGC)	M (ATG)	D (GAC)	V (GTC)	H9N2

PA	100	356	394
Anhui/1	A (GCC)	R (AGG)	N (AAC)
Shanghai/1	A (GCC)	R (AGG)	N (AAC)
H9N2	V (GTC)	K (AAG)	D (GAC)

M2	10	24	97
Anhui/1	P (CCT)	E (GAG)	K (AAG)
Shanghai/1	P (CCT)	D (GAT)	K (AAG)
H9N2	L (CTT)	D (GAT)	E (GAG)

Variations among patient samples

Site	NA (26)	HA (226/235)	PB2 (224)	PB2 (292)	NS1 (80)
Hangzhou/1	I (ATA)	I (ATA)	S (TCA)	V (GTT)	G (GGT)
Anhui/1	I (ATA)	L (CTA)	T (ACA)	V (GTT)	S (AGT)
Shanghai/2	M (ATG)	L (CTA)	T (ACA)	I (ATT)	S (AGT)
Shanghai/1	I (ATA)	Q (CAA)	T (ACA)	V (GTT)	S (AGT)

Variation patterns [e]

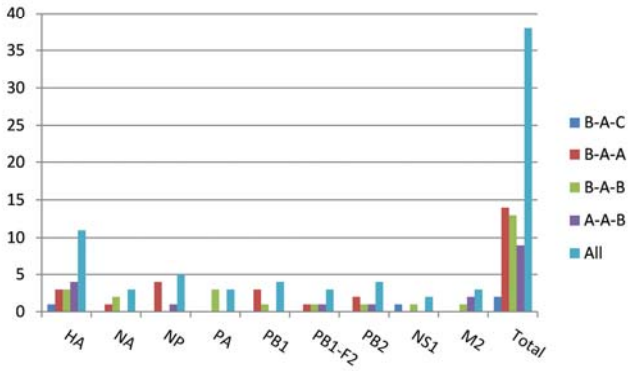


Fig. 5 - The structure modeling of NA-drug interactions. (a) The original NA-Oseltamivir complex from 2QWK[19]; Other complex structures are modeled for H7N9 subtype, including (b) NA-Oseltamivir, (c) NA-Laninamivir and (d) NA-Zanamivir.

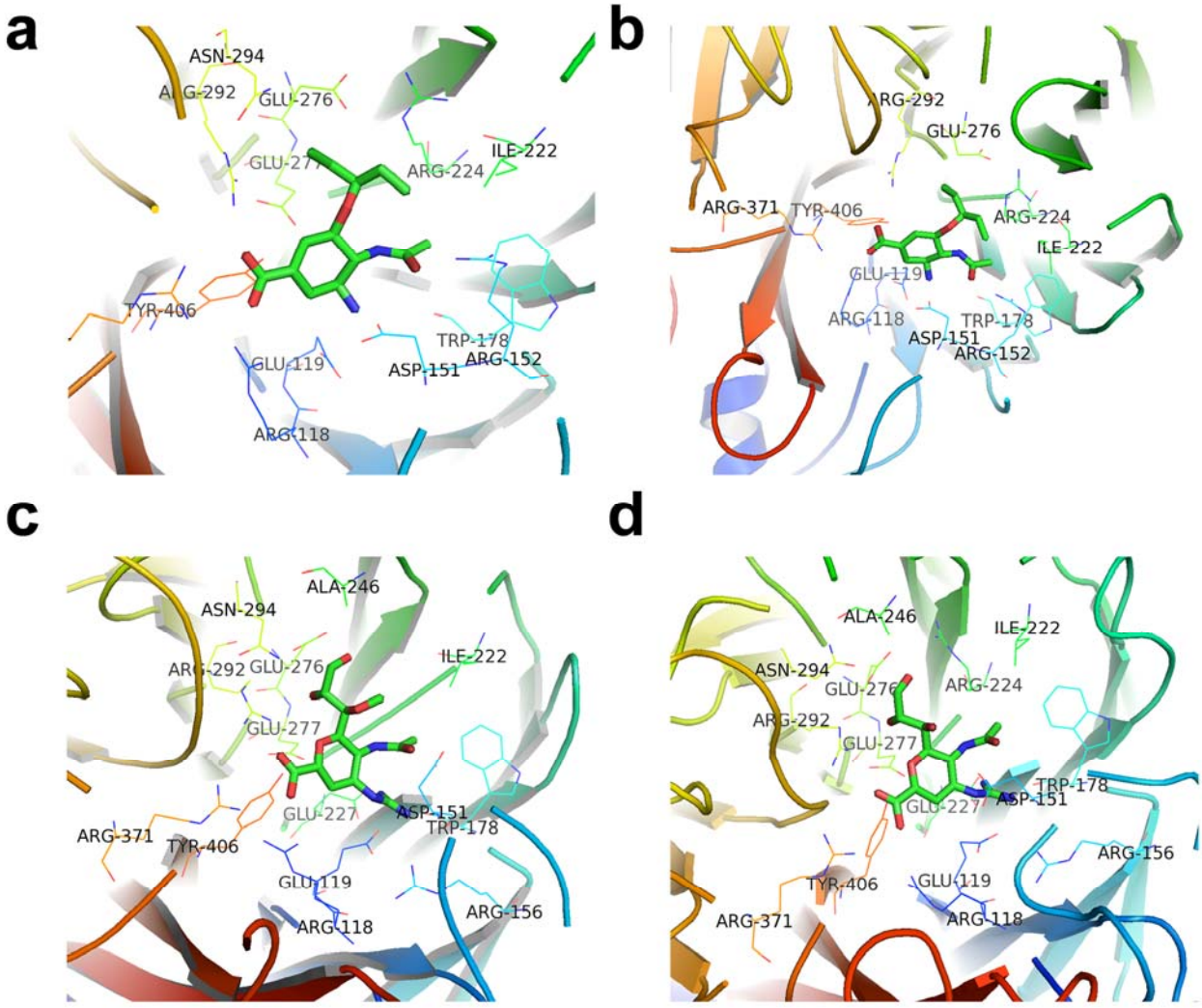


Fig. 6 – The molecule docking of HLA-B*4405, DM1-TCR and predicted epitopes. The key interacting residues in epitopes were labeled, while their binding sites in HLA-B*4405 or DM1-TCR were marked in golden. (a) HLA-B*4405/epitope 88; (b) HLA-B*4405/epitope 353; (c) DM1-TCR/HLA-B*4405/epitope 88; (d) DM1-TCR/HLA-B*4405/epitope 353.

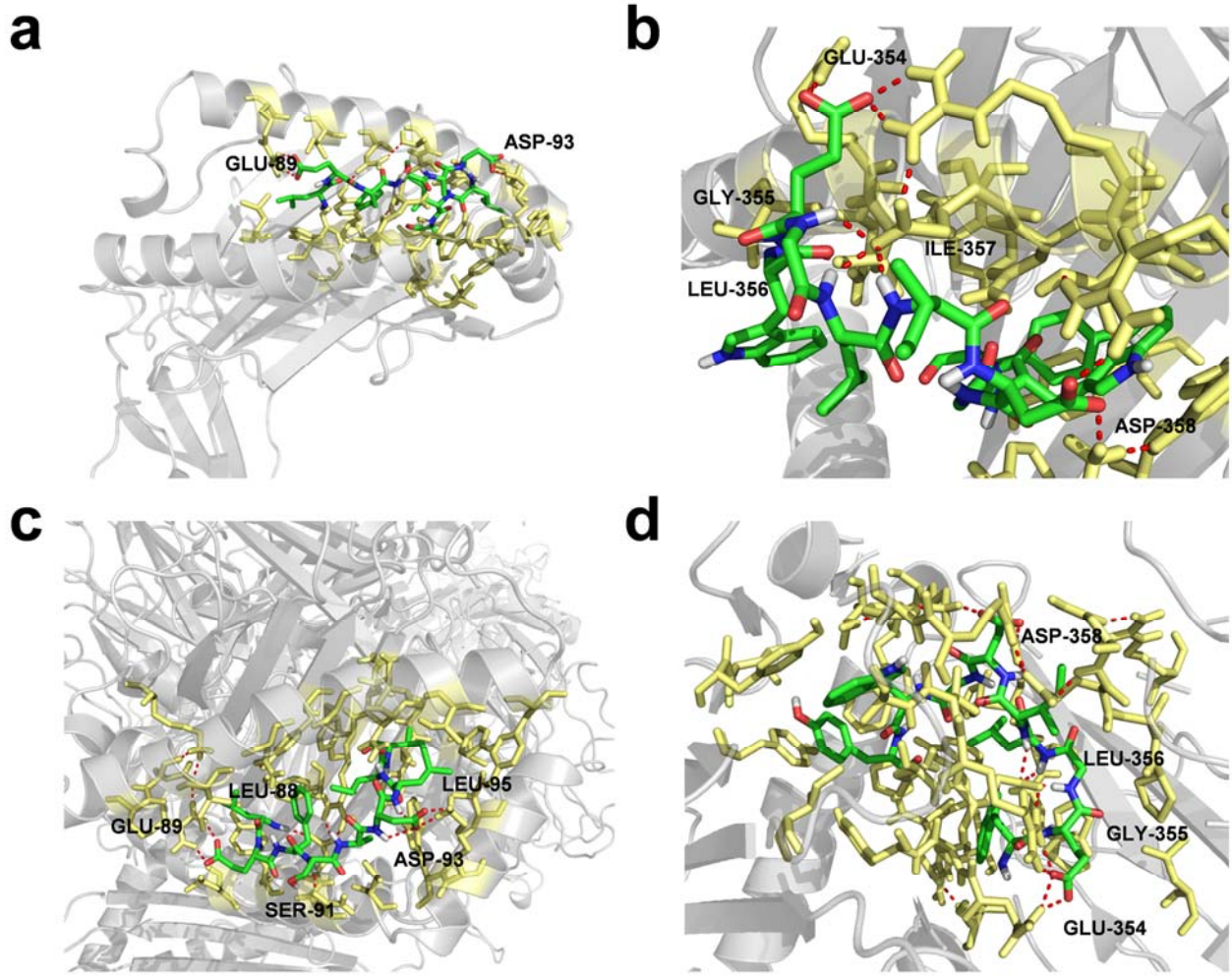
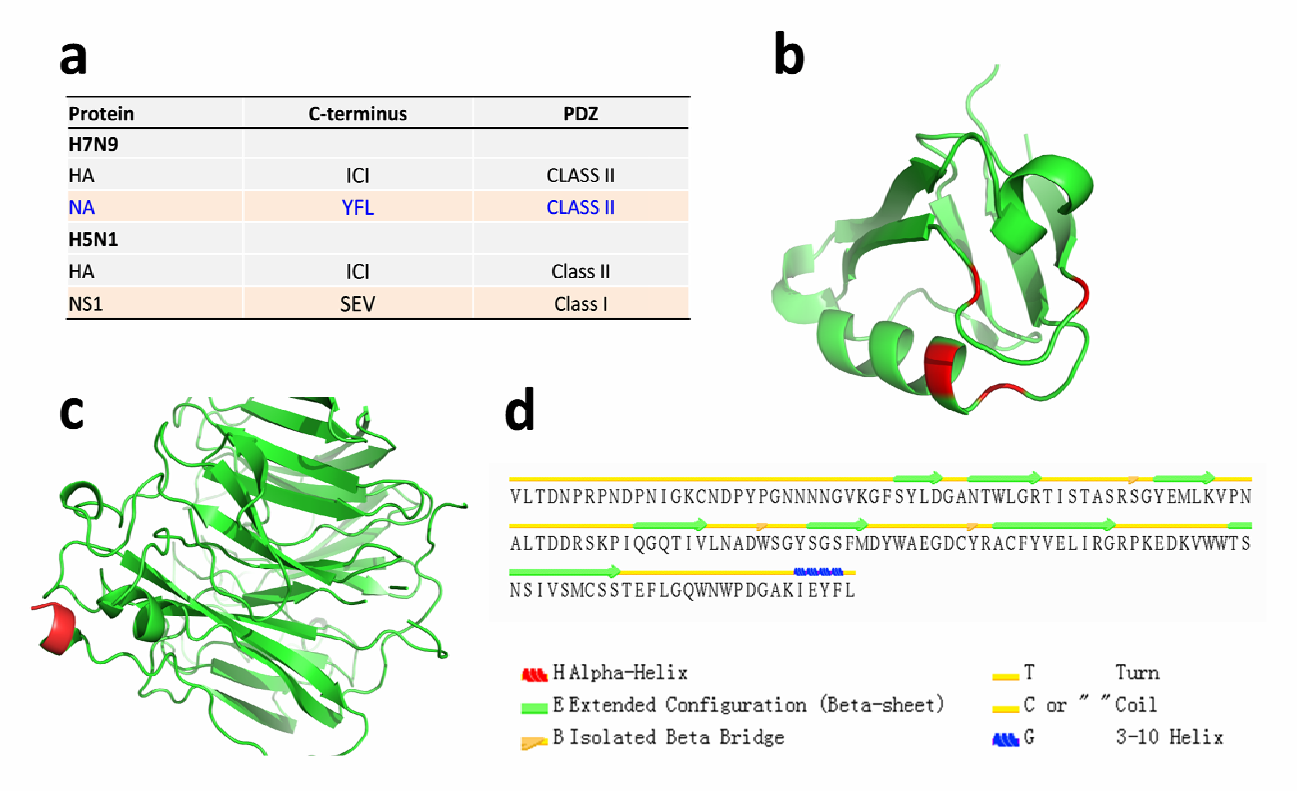


Fig. 7 – The analyses of PDZ-binding in influenza A virus. (a) The predicted PDZ-binding motifs in H7N9 and H5N1; (b) Interacting residues in NS1-binding PDZ domain (red); (c) The PDZ-binding motif in H7N9 NA (red); (d) Secondary structures of H7N9 NA C-terminus.



Protein	C-terminus	PDZ
H7N9		
HA	ICI	CLASS II
NA	YFL	CLASS II
H5N1		
HA	ICI	Class II
NS1	SEV	Class I

Fig. 8 – The modeled HA structures in complex with (a) avian receptor analog 3'SLN and (b) human receptor analog 6'SLN. The sticks represent the analog (3'SLN or 6'SLN) while the lines represent the HA. The complexes of NA-analog for A/Netherlands/219/2003 and A/Anhui/2013 were aligned together for comparison. The hydrogen bonds between analog and HA were showed in green (A/Netherlands/219/2003) and black (A/Anhui/2013).

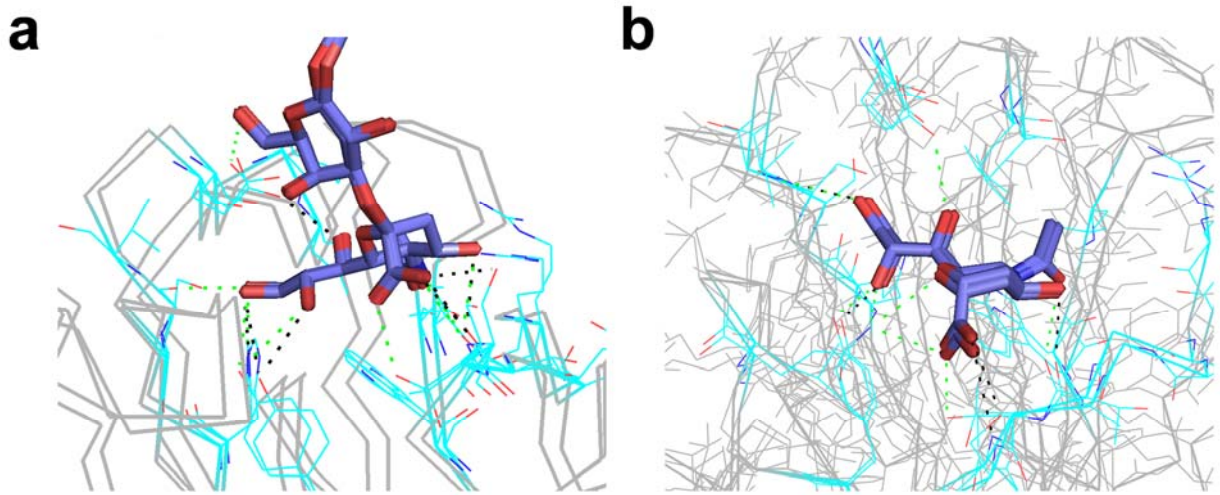
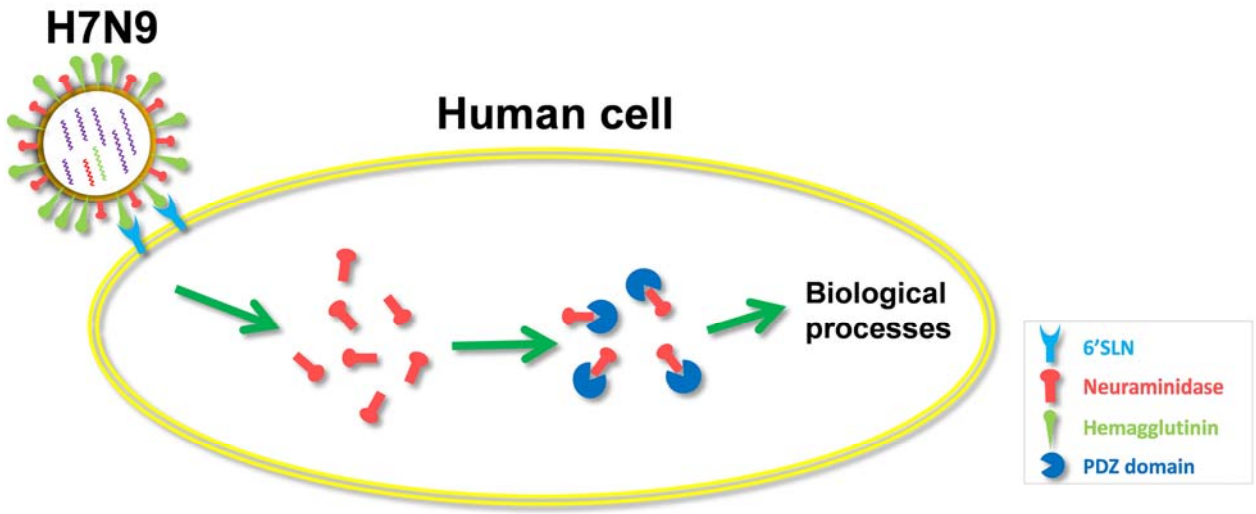


Fig. 9 – A model of HPAI H7N9 virus infection. The viruses enter human cells through the HA-6'SLN interaction, while both HA and NA further interact with PDZ domain proteins to influence biological processes.



Tables

Table 1 – The key residues of NA-drug interaction. The numbers indicate the distances which are given in Å. *a*. From the original NA-Oseltamivir complex (2QWK); *b*. From the H7N9 NA-Oseltamivir model; *c*. The variation at interacting residues; d. The R292K variation in A/Shanghai/1/2013 patient sample.

Residue	Oseltamivir[a]	Oseltamivir[b]	Zanamivir	Laninamivir	Variation[c]
R118	2.8	3.4	3.4	3.0	No
E119	2.8	3.3	3.2	2.8	No
D151	2.6	2.9	3.0	3.0	No
R152	2.6	2.9		3.0	No
R156			3.6		No
W178	3.7	3.9	3.2	2.7	No
I222	3.9	3.8	3.7	3.7	No
R224	3.9	3.4	4.0		No
E227			3.2	3.1	No
A246			3.4	3.2	No
E276	3.5	3.4	2.9	2.9	No
E277	3.7		4.0	3.8	No
R292	3.2	3.3	3.3	2.9	R->K[d]
N294	3.8		3.2	3.3	No
R371	2.7	3.4	3.1	3.2	No
Y406	3.4	3.2	3.1	3.2	No

Table 2 – The predicted epitopes of HA and their docking results to HLA-B*440 and DM1-TCR. The predicted score of NetCTL 1.2[22] was shown for each epitope. The binding energy scores were given by ClusPro[24]. A lower energy value means a more stable docking complex[23].

Start	Epitope	NetCTL	To HLA-B*4405	To DM1-TCR
			Energy	Energy
88	LEFSADLII	1.7827	-904.6	-643.9
353	WEGLIDGWY	0.8101	-821.1	-804.3
333	PEIPKGRGL	1.0328	-683	-703.4
502	REEAMQNRI	1.4055	-637.4	-757.9
412	VEKQIGNVI	1.0442	-609	-756.3
112	NEEALRQIL	1.525	-508.5	-756.4
441	MENQHTIDL	1.9779	-534.5	-752.3
469	AEEDGTGCF	1.3287	-454.1	-652.3

Supplementary materials

Supplementary Fig. S1 – Phylogenetic analyses of H7N9 virus genes. The detailed MP trees for (a) HA and (b) were presented, while the NJ trees were also constructed for (c) HA and (d) NA. The NJ trees were also shown for (e) PB1, (f) PB2, (g) PA, (h) NP, (i) M and (j) NS.