Chem. Res. Chin. Univ. 2013, 29(5), 934—940 doi: 10.1007/s40242-013-3200-x

Phylogenetic and Structural Analysis of Major Surface Proteins Hemagglutinin and Neuraminidase of Novel Avian Influenza Virus A H7N9 from Chinese Patient

YUAN Xiao-hui[1,2], WANG Ying-chen[1*], QU Zhang-yi[1*], REN Jia-yi[2], WU Xiao-min[3] and WANG Jing-fei[4]

1. Department of Microbiology, Public Health College, Harbin Medical University, Harbin 150081, P. R. China;
2. Institute of Pathogen Biology, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100176, P. R. China;
3. Key Laboratory of Resource and Plant Biology of Anhui Province, College of Life Science, Huaibei Normal University, Huaibei 235000, P. R. China;
4. Harbin Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Harbin 150081, P. R. China

Abstract This research reveals the phylogenetic history and structural information of the hemagglutinin(HA) and neuraminidase(NA) from novel avian influenza virus A/Hangzhou/1/2013(H7N9_2013) strain from human infected in China. Strains closely related to the H7N9_2013 strain were obtained from Nation Center for Biotechnology Information(USA)-basic local alignment search tool(NCBI-BLAST) searching, and the phylogenetic trees were constructed. The 3D structures of HA and NA from H7N9_2013 strain were built by homology modeling technology, and molecular dynamics(MD) simulations were performed on the high-performance computer cluster. Characteristic amino acid sites were then screened from multiple sequence alignment(MSA) *via* home-made Python script and mapped onto the 3D structures. The thermodynamic characteristic root-mean-square-fluctuation (RMSF) of these sites in the structure was also analyzed with MD trajectories. The HA of H7N9_2013 strain is closely related to the A/duck/Zhejiang/12/2011 strain isolated in China, while the NA of H7N9_2013 strain is mostly related to the A/mallard/Czech Republic/13438-29K/2010 strain isolated in Europe. The 3D structures of HA and NA from H7N9_2013 stain are mostly identical to the existing structure of H7 and N9. A total of 11 and 14 characteristic amino acid sites were identified in HA and NA, respectively, in H7N9_2013. Structural analysis indicates that certain sites in the top region of HA are important, at which the mutation of some amino acids can impact the receptor binding that may be related to its infection of human beings.

Keywords Avian influenza virus; Hemagglutinin; Neuraminidase; Homology modeling; Phylogenetic

1 Introduction

Avian influenza virus(AIV) mainly causes highly infectious diseases in birds[1]. Highly pathogenic strains cause sudden infant death syndrome in birds and can also infect human beings[2]. Since March, 2013, in Eastern China including Shanghai City and Anhui, Jiangsu, and Zhejiang provinces among other provinces, human infection cases with H7N9 subtype of AIV have been reported[3]. Until May 2, 2013, there were a total of 127 confirmed cases of human infection with H7N9 reported in mainland China, of which 27 people died.

On April 5, 2013, Hangzhou Center for Disease Control (China) submitted the hemagglutinin(HA), neuraminidase (NA) and matrix protein sequence data of a novel avian influenza virus strain of H7N9 subtype(A/Hangzhou/1/2013) isolated from a human infection case to the GenBank database[4].

Influenza viral genome is divided into 8 segments, encoding a total of 13 kinds of proteins[5]. Wherein the HA and NA are two major viral surface antigenic proteins, and are also the main judge basis for classifying influenza A virus into subtypes. HA and NA are grouped into 16 and 9 types respectively. The combination between HA and NA is not random. Only a specific part of the composition can be found in nature[6]. The AIV A H7N9 is usually weak pathogenic in birds, with no human infection report before 2013[3].

In order to explore the molecular evolution of the novel strains of HA and NA and the structural features of the characteristic sites, multiple sequence alignment(MSA) was performed, and a rooted phylogenetic tree was constructed to reveal the evolutionary history of the novel H7N9. The homology modeling technology was then used to build the new HA and NA structures, and 10 ns explicit solvent molecular dynamics(MD) simulations were performed subsequently. Finally, the characteristic sites of the HA and NA segments from the new H7N9_2013 were identified by running a program written

*Corresponding authors. E-mail: qurainbow@126.com; dino1987@163.com
Received May 3, 2013; accepted May 17, 2013.
Supported by the National Natural Science Foundation of China(Nos.81172725, 30271175).
© Jilin University, The Editorial Department of Chemical Research in Chinese Universities and Springer-Verlag GmbH

in Python script. Combined with the trajectories from MD simulations, the root-mean-square-fluctuations(RMSFs) of these sites were analyzed, too. This study provides new ideas for the new H7N9 inhibitors and vaccine development.

2 Materials and Methods

2.1 Sequence Data

All the sequence data used in this study were derived from the GenBank(GB) database[4]. The GB format sequence data were converted to FASTA format[7] *via* SeqVerter software[8] in order to facilitate the subsequent evolution analysis. The name of novel H7N9 subtype of AIV strain is A/Hangzhou/1/2013 (H7N9_2013), and the accession numbers for its HA and NA amino acid sequences are AGI60301(HA_2013) and AGI60300 (NA_2013), respectively. These two sequences were published in the GenBank database by Hangzhou Center for Disease Control(China) on April 5, 2013. Non-structure proteins of influenza virus such as NS1 also play a role in pathogen-host interaction[9]. And non-structure proteins from the novel H7N9 strains should not be overlooked.

2.2 Basic Local Alignment Search Tool(BLAST)

The HA(AGI60301) and NA(AGI60300) sequences were submitted to Nation Center Biotechnology Information (USA)(NCBI)-BLAST server[10] to search for AIV strains closest to the new H7N9_2013 subtype in the GenBank. Seve- ral influenza virus HA and NA amino acid sequences were selected as references in phylogenetic analysis according to the NCBI-BLAST results. Screened sequences were divided into two parts; the first part was closest to the new strain sequence in the BLAST search results, and the second part was the known subtype of avian H7N9 influenza viral sequences. Eventually a total of 26 sequences were included in this analysis, with 13 for both HA and NA, respectively. All the sequences related are shown in Table 1.

Table 1 Sequence data and virus strains

Accession	Subtype	Year	Country	Protein	Strain
AGI60301.1	H7N9	2013	China	HA	A/Hangzhou/1/2013
AFI73274.1	H7N3	2011	China	HA	A/duck/Zhejiang/11/2011
AGE08098.1	H7N9	2011	USA	HA	A/northern shoverl/Mississippi/11OS145/2011
CAY39406.1	H7N9	2008	Spain	HA	A/*Anas crecca*/Spain/1460/2008
ACU15247.1	H7N1	2007	Netherlands	HA	A/mallard/Netherlands/22/2007
ACN39305.1	H7N7	2006	Korea	HA	A/wild bird feces/Korea/HDR22/2006
BAJ23224.1	H7N3	2006	Japan	HA	A/duck/Shimane/137/2006
BAH03372.1	H7N1	2002	Mongolia	HA	A/duck/Mongolia/867/2002
AEZ68716.1	H7N7	2001	Germany	HA	A/turkey/Germany/R11/2001
AAR02641.1	H7N3	2000	Netherlands	HA	A/mallard/Netherlands/12/2000
ABR37396.1	H7N1	1999	Italy	HA	A/chicken/Italy/1082/99
ABV01288.1	H7N1	1999	Italy	HA	A/turkey/Italy/1265/1999
ABI84694.1	H7N9	1988	USA	HA	A/turkey/Minnesota/1/1988
AGI60300.1	H7N9	2013	China	NA	A/Hangzhou/1/2013
AGE08101.1	H7N9	2011	USA	NA	A/northern shoverl/Mississippi/11OS145/2011
AEB89856.1	H11N9	2010	Czech	NA	A/mallard/Czech Republic/13438-29K/2010
ADN34729.1	H7N9	2009	Czech	NA	A/goose/Czech Republic/1848-T14/2009
ADN34740.1	H7N9	2008	Spain	NA	A/*Anas crecca*/Spain/1460/2008
BAK18440.1	H11N9	2008	Vietnam	NA	A/duck/Vietnam/G32/2008
ADG37202.1	H11N9	2007	Switzerland	NA	A/mallard/Switzerland/WV1071028/2007
ABW94402.1	H11N9	2006	Korea	NA	A/shorebird/Korea/S8/2006
ACR56279.1	H11N9	2005	China	NA	A/*Baikal teal*/Hongze/14/2005
BAF49418.1	H11N9	2004	Japan	NA	A/duck/Hokkaido/W245/2004
ABB20168.1	H2N9	2000	China	NA	A/Duck/Nanchang/2-0486/2000
AFM83137.1	H2N9	1999	Netherlands	NA	A/mallard/Netherlands/11/1999
ABI84697.1	H7N9	1988	USA	NA	A/turkey/Minnesota/1/1988

2.3 MSA and Phylogenetic Analysis

MSA was performed with above mentioned HA and NA sequences *via* MUSCLE software[11](Version 3.7) under a progressive algorithm[12] and adjusted manually with BioEdit (Version 7.09). The rooted phylogenetic trees were then constructed according to the alignment among HA and NA sequences *via* the molecular evolutionary genetics analysis (MEGA, Version 5.01)[13] package based on the Neighbor-Joining(NJ) method[14] and Jones-Taylor-Thornton(JTT) distance[15] which can estimate transition likelihoods from a large number of proteins, with H7N9 subtype of AIV isolated in the United States in 1988(A/turkey/Minnesota/1/1988) as the out-group. Bootstrap analysis[16] was done with 1000 replicates to test the robustness of the phylogenetic trees.

2.4 Homology Modeling

Homology modeling and molecular dynamics are the important methods of researches on protein structure[17,18]. The 3D structures of HA and NA were constructed by means of homology modeling method[19] and Discovery Studio 2.5 (Accelrys Software). Firstly web-FASTA program was performed in the protein data bank(PDB, http://www.rcsb.org) to determine the best templates for HA_2013 and NA_2013

modeling. Two structures were found to have a similarity of 96.2% to HA_2013 under the following PDB_ID: 4DJ6[20] and that of 81.9% to NA_2013 under the following PDB_ID: 1F8B[21]. Initial HA and NA molecules were built *via* the modeller module and energy minimizations(EMs) were then performed in a CHARMM27 force field[22] by a 500-step steepest descent(SD) minimization followed by conjugate gradient(CG) minimization until the final convergence was lower than 0.4184 $kJ\cdot mol^{-1}\cdot nm^{-1}$ to obtain the final HA and NA structures, which were finally tested on residue compatibility with the Profile-3D[23] program in Discovery Studio. All the calculations were performed on a Dell PowerEdge 2900 workstation.

2.5 MD Simulation

MD simulations were performed with the fast MD simulation software package Gromacs(Version 4.5.3)[24] in the CHARMM27 force field[22]. The initial models were dissolved in triclinic boxes containing SPC/E(SPC=simple-point-charge) water molecules, which are sufficient to immerse the proteins(about 3.3×10^4 water molecules for HA and 1.38×10^4 water molecules for NA). To neutralize the system, the NA^+ counter-anions were added. After removing bad contacts by EM and relaxing water solvent by position-restrained MD simulations, the final 10 ns production MD simulations for HA and NA were performed under periodic boundary conditions with a time step of 2 fs at 310 K(*ca*. 37 °C). The normal pressure & temperature(NPT) ensemble[25] was applied and long-range electrostatics was described by the particle-mesh Ewald(PME) method[24]. The covalent bonds involving hydrogen were constrained with the LINCS algorithm[24]. Finally, the MD trajectories were analyzed in terms of potential energy(PE) and root mean square deviation(RMSD)[26] with the Gromacs suite of programs. In order to estimate the flexibility of particular structures, especially those of HA and NA containing the H7N9 characteristic sites during the last equilibrium of MD simulation, the RMSF[25] of Cα was calculated simultaneously. All the MD simulations were performed in the DAWNING Supercomputer Center(128 Cores).

2.6 Screening for Characteristic Sites

The characteristic sites mean the amino acid sites in H7N9_2013 which are significantly different from those of other AIV types in MSA result previously. Because of the relatively high site-homology of HA or NA within the same type of influenza virus, we designed a novel characteristic site screen algorithm to find H7N9_2013 strain's characteristic sites in the HA and NA sequences as follows. In the aligned HA or NA sequences, all, but the out-group ones, sequences were selected as the reference group. For each amino acid site in the HA or NA's reference group, site homology was calculated. Those sites whose homologies were more than 80% and the majority amino acid residue in the reference group was different from that of the novel strain(A/Hangzhou/1/2013) would be regarded as the characteristic sites. The RMSF of these sites was also analyzed with the trajectories from MD simulations. The characteristic sites were screened *via* programs written in Python script language.

3 Results and Discussion

3.1 Phylogenetic Analysis

The NCBI-BLAST searching results showed that the HA_2013 sequence is closest with the HA of H7N3 subtype influenza virus A/duck/Zhejiang/12/2011 isolated in Zhejiang, China, in 2011, with amino acid sequence homology of 98%(548/560), while the NA_2013 sequence is closest with the NA of H11N9 subtype influenza virus A/mallard/Czech Republic/13438-29K/2010 isolated in Czech Republic in 2010 with amino acid sequence homology of 97%(457/470). The phylogenetic analysis was performed using MEGA software, and the phylogenetic trees were constructed. The phylogenetic tree of HA_2013 protein as shown in Fig.1(A) reveals that all the H7

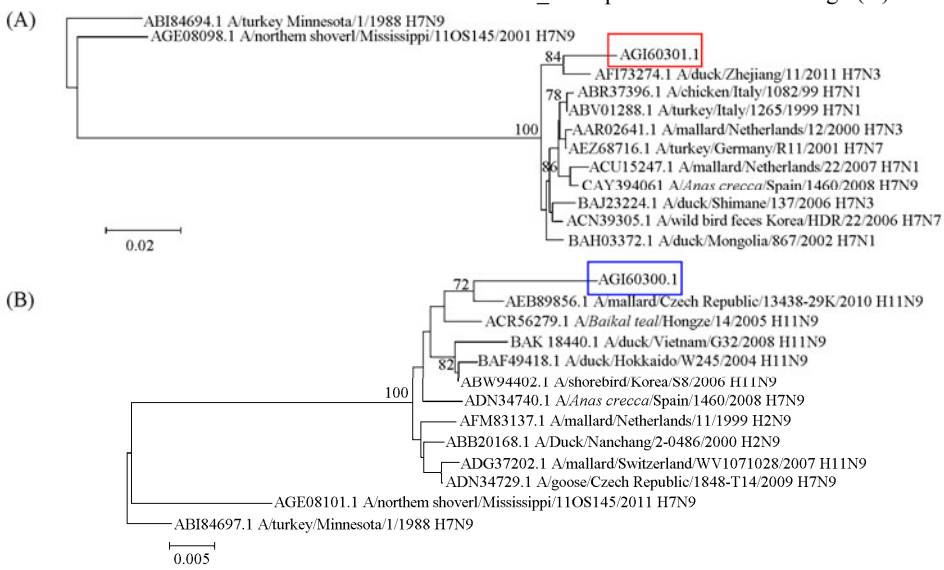

Fig.1 Phylogenetic analysis of HA_2013(A) and NA_2013(B)

Neighbor-Joining trees *via* JTT distance was rooted to strain A/turkey/Minnesota/1/1988, and bootstrap values greater than 70% are shown above the corresponding nodes. The numbers in the tree show the distance between different sequences. HA_2013 and NA_2013 are labelled by two boxes on (A) and (B), respectively.

subtype strains formed a closely related lineage with a boot strap value of 100% except for the reference strain A/northern shoverl/Mississippi/11OS145/2011 when the tree was rooted to strain A/turkey/Minnesota/1/1988. In this lineage, novel strain H7N9_2013 and its closest relative strain A/duck/Zhejiang/12/2011 form a small clade with a bootstrap value of 84%, indicating that the HA protein of novel strain H7N9_2013 should be regarded as H7 subtype and closely related to that of strain A/duck/Zhejiang/12/2011.

The phylogenetic tree of NA_2013 as shown in Fig.1(B) reveals that all the N9 subtype strains formed a closely related lineage with a bootstrap value of 100% except for the reference strain A/northern shoverl/Mississippi/11OS145/2011 when the tree was rooted to strain A/turkey/Minnesota/1/1988. In this lineage, novel strain H7N9_2013 and A/mallard/Czech Republic/13438-29K/2010 form a clade with a bootstrap value of 72%, implying that the NA protein of novel strain H7N9_2013 should be grouped into N9 subtype and is closely related to that of strain A/mallard/Czech Republic/13438-29K/2010.

According to the above analysis, the closer strains to the novel A/Hangzhou/1/2013 are grouped into the H7 subtypes, and the closest one was from Zhejiang, China, which constituted a more reliable clade with the novel strain. The closer strains to the novel A/Hangzhou/1/2013 also belong to the N9 subtypes according to the NA sequence, and the closest one is from Czech Republic of the Europe and constituted a reliable clade with the novel strain. These results prompt that the genotype of the novel strain is indeed H7N9 subtype, and the closest HA and NA proteins were derived from poultry samples isolated in China and Czech Republic respectively, suggesting that the novel strain is likely derived from a gene reassortment event.

3.2 Structures of HA_2013 and NA_2013

The initial structures of HA_2013 and NA_2013 were built *via* Modeler program, and EM was performed in a CHARMM27 force field. Fig.2(A) shows that HA_2013 forms stable structures of heterologous hexamer molecule consisting of three HA1 subunits and three HA2 subunits. The three HA1 subunits form an outer structure of HA, with β sheets as the main secondary structure. In the top region of HA, there is a pocket-like structure that forms the receptor-binding domain with the host cells. HA2 subunit is embedded inside HA; its structure and sequence are more conservative, with 6 helix structures important for assembling with each other *via* hydrophobic interactions. Fig.2(B) shows the structure of NA_2013, which adopts a typical box-shaped NA tetramer structure, with each monomer containing a propeller-like arrangement of six-bladed β sheets that are important for assembling with each other *via* hydrophobic interaction.

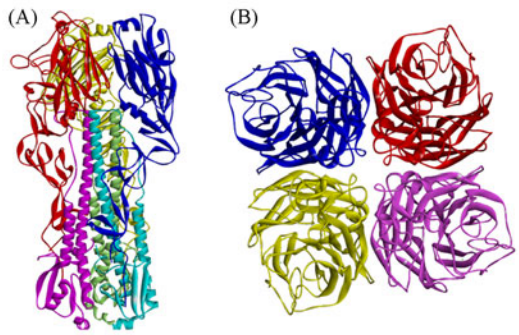

Fig.2 Structures of HA_2013(A) and NA_2013(B)

The 3D structures of heterologous hexamer HA_2013(A) and box-shaped tetramer NA_2013(B) with monomers presented in different colors.

Profiles-3D was used to evaluate the compatibility of all amino acids on the 3D structure of proteins, especially on a hypothetical protein structure. As shown in Fig.3(A) and (B), all the residues received positive verifying scores, indicating that the primary sequence is compatible with that of the 3D structure.

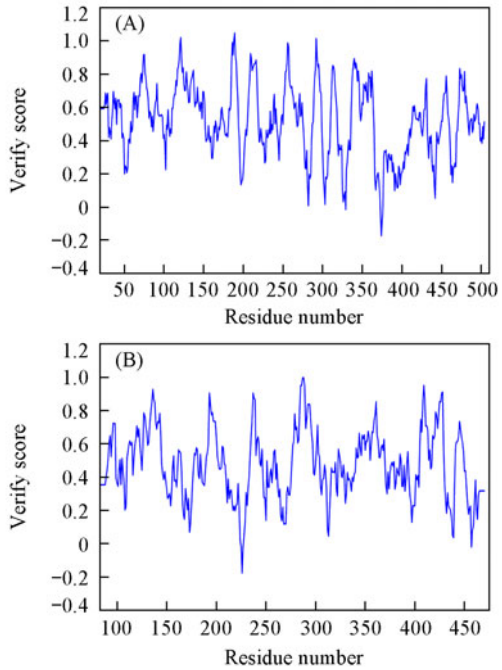

Fig.3 Profile-3D scores of HA_2013 monomer(A) and NA_2013 monomer(B)

3.3 Characteristic Sites of HA and NA from H7N9_2013

In order to indentify the characteristic sites of HA and NA from H7N9_2013 based on the MSA result, we designed a novel characteristic site screen algorithm implemented by Python script language. The results show that 11 characteristic sites were indentified in HA_2013 sequence which were distributed in the middle and C-terminal of HA, while 14 characteristic sites were indentified in NA_2013 sequence of the novel strain H7N9_2013. There was a characteristic deletion mutation of 5 consecutive residues, from residue 68 to 72, indentified in NA of H7N9_2013, which is different from that in other N9 subtype of influenza virus, as shown in Tables 2 and 3.

These characteristic sites are mapped onto the 3D structures of the HA and NA built by homology modeling, while part of them was not included because of the molecular structure being not a full-length structure prediction. As shown in Fig.4(A), 9 characteristic sites(Ser183, Val188, Val195, Ala198,

Table 2 Characteristic sites of hemagglutinin*

Accesion	Strain	Site 18	Site 190	Site 195	Site 202	Site 205	Site 218	Site 242	Site 314	Site 417	Site 462	Site 548
AGI60301.1		I	S	V	V	A	V	I	D	N	D	V
AFI73274.1	A/duck/Zhejiang/11/2011 H7N3	I	D	I	G	A	I	Q	D	T	N	A
BAJ23224.1	A/duck/Shimane/137/2006 H7N3	V	D	I	G	T	V	Q	N	T	N	A
BAH03372.1	A/duck/Mongolia/867/2002 H7N1	V	D	I	G	T	I	Q	N	T	N	A
ACU15247.1	A/mallard/Netherlands/22/2007 H7N1	V	D	I	G	T	I	Q	N	T	N	A
ACN39305.1	A/wild/bird/feces/Korea/ 2006 H7N7	V	D	I	G	T	I	Q	N	T	N	A
CAY39406.1	A/*Anas crecca*/Spain/1460/2008 H7N9	V	D	I	G	T	I	Q	N	T	N	A
ABR37396.1	A/chicken/Italy/1082/99 H7N1	V	D	I	V	T	I	Q	N	T	N	A
ABV01288.1	A/turkey/Italy/1265/1999 H7N1	V	D	I	G	T	I	Q	N	T	N	A
AAR02641.1	A/mallard/Netherlands/12/2000 H7N3	M	D	I	G	T	I	Q	N	T	N	A
AEZ68716.1	A/turkey/Germany/R11/2001 H7N7	V	D	I	G	T	I	Q	N	T	N	A
ABI84694.1	A/turkey/Minnesota/1/1988 H7N9	C	K	V	G	T	I	Q	N	S	N	A
AGE08098.1	A/northern/ Mississippi/2011 H7N9	C	K	I	G	T	I	Q	N	S	N	A

* A: Alanine; C: cysteine; D: aspartic acid; G: glycine; I: isoleucine; K: lysine; M: methionine; N: asparagine; Q: glutamine; S: serine; T: threonine; V: valine.

Table 3 Characteristic sites of neuraminidase*

Accession	Strain	Site 16	Site 19	Site 40	Site 50	Site 68	Site 69	Site 70	Site 71	Site 72	Site 81	Site 84	Site 112	Site 359	Site 401
AGI60300.1		I	A	G	T	—	—	—	—	—	T	N	S	A	A
AEB89856.1	A/mallard/Czech Republic/2010 H11N9	V	A	S	A	T	Q	I	S	N	A	G	S	V	T
ACR56279.1	A/*Baikal teal*/Hongze/14/2005 H11N9	V	T	S	A	T	Q	I	S	N	E	G	N	V	T
BAK18440.1	A/duck/Vietnam/G32/2008 H11N9	V	T	G	A	T	Q	I	S	N	A	G	N	V	T
BAF49418.1	A/duck/Hokkaido/W245/2004 H11N9	V	T	S	A	T	Q	I	S	N	A	G	N	V	T
ABW94402.1	A/shorebird/Korea/S8/2006 H11N9	V	T	S	A	T	Q	I	S	N	A	G	N	V	T
ADN34740.1	A/*Anas crecca*/Spain/1460/2008 H7N9	V	T	S	A	T	Q	I	S	N	A	S	N	V	T
AFM83137.1	A/mallard/Netherlands/11/1999 H2N9	V	T	S	A	T	Q	I	S	N	A	G	N	I	T
ABB20168.1	A/Duck/Nanchang/2-0486/2000 H2N9	V	T	S	A	T	Q	I	S	N	A	G	N	V	T
ADG37202.1	A/Switzerland/WV1071028/2007 H11N9	I	T	S	E	T	Q	I	S	N	A	G	N	V	T
ADN34729.1	A/Czech Republic/1848-T14/2009 H7N9	V	T	S	E	T	Q	I	S	N	A	G	N	V	T
ABI84697.1	A/turkey/Minnesota/1/1988 H7N9	V	T	N	A	T	Q	I	S	N	A	E	D	G	T
AGE08101.1	A/northern/Mississippi/2011 H7N9	V	T	N	T	T	Q	I	S	N	A	E	N	G	T

* A: Alanine; D: aspartic acid; E: glutamic acid; G: glycine; I: isoleucine; N: asparagines; Q: glutamine; S: serine; T: threonine; V: valine; —: delation.

Val211, Ile235, Asp307, Asn410 and Asp455) are mapped onto HA structure. Three mutations Val195, Ala198 and Ile235 were identified in the top region that is the human host cell receptor binding region of HA, which may be associated with viral infection to humans.

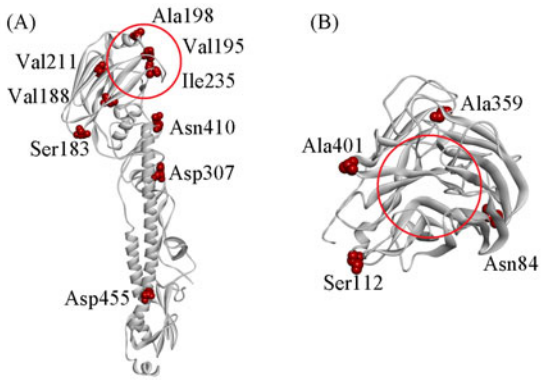

Fig.4 Characteristic sites mapped on HA_2013(A) and NA_2013(B)

The characteristic sites of HA_2013(A) and NA_2013(B) are shown in red CPK presentation. Red circle frames present the receptor binding sites of HA and active pocket region of NA, respectively.

As shown in Fig.4(B), 4 characteristic sites(Asn84, Ser112, Ala359 and Ala401) are mapped onto NA structure. Because the NA active pocket is relatively clear, further analysis shows that these 4 sites are not in the active pocket area, suggesting that existing anti-flu drugs such as Zanamivir and Oseltamivir may still be effective to the novel H7N9_2013. There are more mutations found in its N-terminal, especially the deletion mutation of 5 consecutive residues from 68 to 72, which may change the binding property of NA with the viral envelope in the novel strain[27]. The influence of this deletion mutation on host specificity of the novel H7N9 deserves further studies. It was demonstrated that several components from traditional Chinese herbs such as Shikonin[28], Cinnamaldehyde[29] and Astragaloside IV[30] could block the replication of adenovirus, so further researches may reveal their ability to combat influenza virus.

3.4 MD Simulations

EM cannot solve the problem of energy barriers. When the EM of the initial theoretically calculated structures was done, it is necessary to perform a long-range MD simulation in an explicit water solvent environment to obtain a stable and reasonable final conformation. On the other hand, MD simulation is also a powerful simulation technology to probe the molecular thermodynamic properties in a thermodynamic steady-state. PE and RMSD are two important indicators to examine whether MD system is steady, which can reflect the changes of the system energy and backbone chain of the molecules. As shown in Fig.5, when the simulation lasted for a certain period of time in a 10 ns dynamic simulation, PE and RMSD reached

equilibration in HA and NA in 3 and 4 ns respectively. After the system reached equilibration, the MD trajectory can be used to study the molecular structure characteristics in such a thermodynamic state.

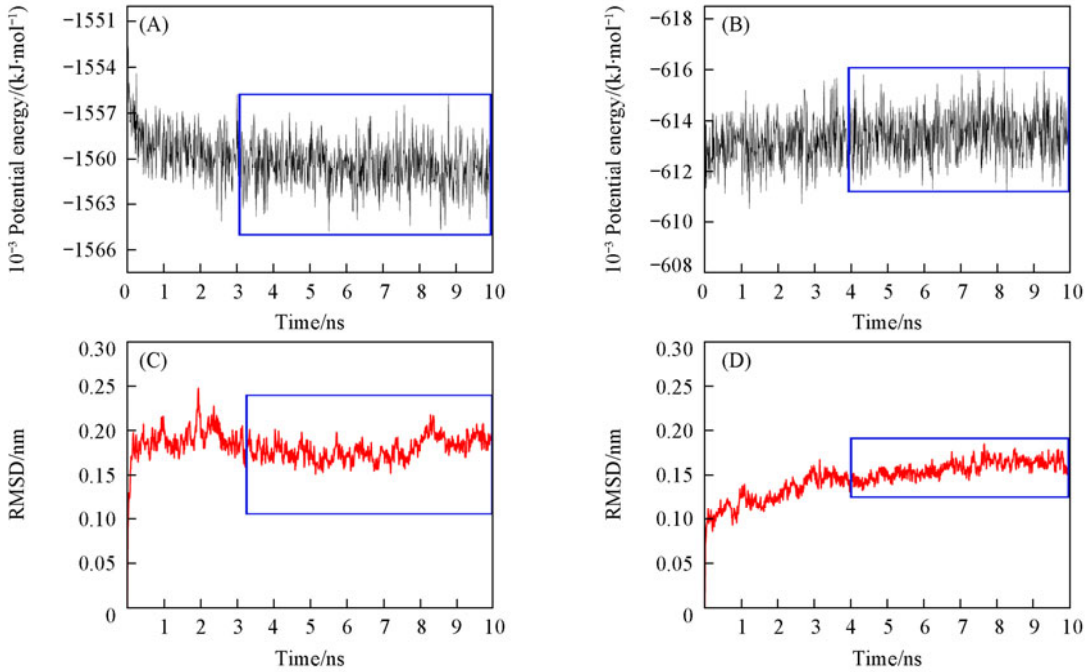

Fig.5 MD simulations of HA_2013(A, C) and NA_2013(B, D)

When the simulation lasted for a certain period of time in a 10 ns dynamic simulation, PE(A, B) and RMSD(C, D) of HA and NA reached equilibration in 3 and 4 ns, respectively. The solid frames contain equilibrium stages of two MD simulations.

3.5 RMSF of Characteristic Sites

RMSF of Cα atoms is a residue-based property defined over a certain time to reflect the differences in residue mobility within and between simulations. Fig.6 shows the RMSF of per residue of HA and NA in the final balancing stage of the MD simulation, in which the identified characteristic sites are labeled in number lines.

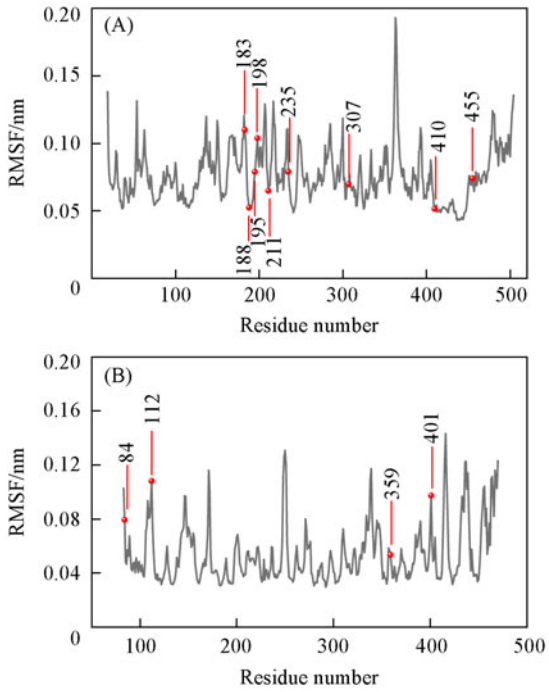

Fig.6 RMSF of HA_2013(A) and NA_2013(B)

RMSF of per residue of HA_2013 monomer and NA_2013 monomer over the final balancing stage of the MD simulation, in which the identified characteristic sites are labeled in number lines. The flexibility of the amino acid residues fluctuated less than 0.1 nm, indicating that in equilibrium these residues were very stable.

Overall, we found that for most of the amino acids in the MD process, RMSF is within about 0.1 nm, indicating that the overall structure is relatively stable. However, some amino acids wave greatly, implying that these are flexible sites. Further studies find that the RMSF of characteristic site Ser183 is greater than 0.1 nm, for Ser183 is a surface amino acids with second structure β-turn, and the RMSF of characteristic sites Val195, Ala198 and Ile235 in the top region of HA_2013 is smaller than 0.1 nm, indicating that these amino acid mutations form stable conformations. Therefore, these 3 variations might be the direct cause of the novel H7N9 infecting human. The 4 characteristic sites on NA_2013 are almost with lower fluctuations and far from the active pocket. Preliminary analysis suggests that these sites may not affect the structure of active pocket.

4 Conclusions

Novel avian influenza virus A/Hangzhou/1/2013 strain (H7N9_2013) from human infected in China was derived from a gene reassortment event, the HA_2013 and NA_2013 proteins were derived from poultry samples isolated in China and Czech Republic respectively. The 3D structures of HA and NA from H7N9_2013 stain are mostly identical to the existing structure of H7 and N9. A total of 11 and 14 characteristic amino acid sites are identified in HA and NA respectively in H7N9_2013. There was no mutation in the active pocket area of NA_2013, suggesting that existing anti-flu drugs such as Zanamivir and Oseltamivir may still be effective on the novel

H7N9_2013. Certain sites in the top region of HA_2013 can impact the receptor binding that may be related to its infection of human beings.

References

[1] Kuiken T., Holmes E. C., McCauley J., Rimmelzwaan G. F., Williams C. S., Grenfell B. T., *Science*, **2006**, *312*, 394

[2] Bertran K., Perez-Ramirez E., Busquets N., Dolz R., Ramis A., Darji A., Abad F. X., Valle R., Chaves A., Vergara-Alert J., Barral M., Hofle U., Majo N., *Vet. Res.*, **2011**, *42*, 24

[3] Gao R., Cao B., Hu Y., Feng Z., Wang D., Hu W., Chen J., Jie Z., Qiu H., Xu K., Xu X., Lu H., Zhu W., Gao Z., Xiang N., Shen Y., He Z., Gu Y., Zhang Z., Yang Y., Zhao X., Zhou L., Li X., Zou S., Zhang Y., Yang L., Guo J., Dong J., Li Q., Dong L., Zhu Y., Bai T., Wang S., Hao P., Yang W., Han J., Yu H., Li D., Gao G. F., Wu G., Wang Y., Yuan Z., Shu Y., *N. Engl. J. Med.*, **2013**, *368*(20), 1888

[4] Benson D. A., Cavanaugh M., Clark K., Karsch-Mizrachi I., Lipman D. J., Ostell J., Sayers E. W., *Nucleic Acids Res.*, **2013**, *41*, D36

[5] Jagger B. W., Wise H. M., Kash J. C., Walters K. A., Wills N. M., Xiao Y. L., Dunfee R. L., Schwartzman L. M., Ozinsky A., Bell G. L., Dalton R. M., Lo A., Efstathiou S., Atkins J. F., Firth A. E., Taubenberger J. K., Digard P., *Science*, **2012**, *337*, 199

[6] Zhang Y., Lin X., Wang G., Zhou J., Lu J., Zhao H., Zhang F., Wu J., Xu C., Du N., Li Z., Wang X., Bi S., Shu Y., Zhou H., Tan W., Wu X., Chen Z., Wang Y., *Plos One*, **2010**, *5*, e9167

[7] Barton M. D., Barton H. A., *Source Code Biol. Med.*, **2012**, *7*, 4

[8] Rodríguez F., Escalera L., Reguera B., Rial P., Riobó P., De Jesús Da Silva T., *Harmful Algae*, **2012**, *13*, 26

[9] Zhu C. Y., Sun T. T., Zhao J., Wang N., Zheng F. L., Ai H. X., Zhu J. F., Wang X. Y., Zhu Y., Wu J. G., Liu H. S., *Chem. Res. Chinese Universities*, **2012**, *28*(1), 103

[10] Camacho C., Coulouris G., Avagyan V., Ma N., Papadopoulos J., Bealer K., Madden T. L., *BMC Bioinformatics*, **2009**, *10*, 421

[11] Edgar R. C., *Nucleic Acids Res.*, **2004**, *32*, 1792

[12] Yuan X. H., Wang Y. C., Qu Z. Y., Ren J. Y., Wang J. F., Guo Y. Y., Wang Y. X., Hua D., Wu X. M., Yang Z. W., Gao H., *Chem. J. Chinese Universities*, **2011**, *32*(8), 1838

[13] Tamura K., Peterson D., Peterson N., Stecher G., Nei M., Kumar S., *Mol. Biol. Evol.*, **2011**, *28*, 2731

[14] Saitou N., Nei M., *Mol. Biol. Evol.*, **1987**, *4*, 406

[15] Jones D. T., Taylor W. R., Thornton J. M., *Comput. Appl. Biosci.*, **1992**, *8*, 275

[16] Felsenstein J., *Evolution*, **1985**, *39*, 783

[17] Lan H. N., Wang Y. X., Zheng M. Z., Han W. W., Zheng X., *Chem. Res. Chinese Universities*, **2013**, *29*(1), 139

[18] Yuan X., Qu Z., Wu X., Wang Y., Liu L., Wei F., Gao H., Shang L., Zhang H., Cui H., Zhao Y., Wu N., Tang Y., Qin L., *Vaccine*, **2009**, *27*, 5103

[19] Yuan X. H., Qu Z. Y., Wu X. M., Wang Y. C., Wei F. X., Zhang H. Y., Liu L., Yang Z. W., *Chem. J. Chinese Universities*, **2009**, *30*(8), 1636

[20] Kageyama T., Fujisaki S., Takashita E., Xu H., Yamada S., Uchida Y., Neumann G., Saito T., Kawaoka Y., Tashiro M., *Eurosurveill.*, **2013**, *18*, 20453

[21] Smith B. J., Colman P. M., von Itzstein M., Danylec B., Varghese J. N., *Protein Sci.*, **2001**, *10*, 689

[22] Sapay N., Tieleman D. P., *J. Comput. Chem.*, **2011**, *32*, 1400

[23] Eisenberg D., Luthy R., Bowie J. U., *Methods Enzymol.*, **1997**, *277*, 396

[24] van Der Spoel D., Lindahl E., Hess B., Groenhof G., Mark A. E., Berendsen H. J., *J. Comput. Chem.*, **2005**, *26*, 1701

[25] Yuan X. H., Wang Y. C., Jin W. J., Zhao B. B., Chen C. F., Yang J., Wang J. F., Guo Y. Y., Liu J. J., Zhang D., Gong L. L., He Y. W., *Plos One*, **2012**, *7*, e32938

[26] Wang X., Snoeyink J., *IEEE/ACM Trans. Comput. Biol. Bioinform.*, **2008**, *5*, 525

[27] Li J., zu Dohna H., Cardona C. J., Miller J., Carpenter T. E., *Plos One*, **2011**, *6*, e14722

[28] Gao H., Liu L., Qu Z. Y., Wei F. X., Wang S. Q., Chen G., Qin L., Jiang F. Y., Wang Y. C., Shang L., Gao C. Y., *Biol. Pharm. Bull.*, **2011**, *34*, 197

[29] Liu L., Wei F. X., Qu Z. Y., Wang S. Q., Chen G., Gao H., Zhang H. Y., Shang L., Yuan X. H., Wang Y. C., *Labmedicine*, **2009**, *40*, 669

[30] Shang L., Qu Z.Y ., Sun L. H., Wang Y. C., Liu F. H., Wang S., Gao H., Jiang F. Y., *Journal of Pharmacy and Pharmacology*, **2011**, *63*, 688