

Structure modeling and spatial epitope analysis for HA protein of the novel H1N1 influenza virus

Consortium for Influenza Study at Shanghai (WU Di^{1,2}, XU TianLei¹, SUN Jing¹, DAI JianXin^{4,6}, DING GuoHui⁹, HE YunGang⁹, ZHOU ZhengFeng¹⁰, XIONG Hui¹⁰, DONG Hui¹⁰, JIN WeiRong¹⁰, BIAN Chao⁹, JIN Li^{3,9,10}, WANG HongYan³, WANG XiaoNing^{3,5}, YANG Zhong^{2,3}, ZHONG Yang^{2,3}, WANG Hao^{4,6}, CHE XiaoYan⁷, HUANG Zhong⁸, LAN Ke⁸, SUN Bing⁸, WU Fan⁸, YUAN ZhenAn⁸, ZHANG Xi⁸, ZHOU XiaoNong¹¹, ZHOU JiaHai¹², MA ZhiYong¹³, TONG GuangZhi¹³, GUO YaJun^{4,6}, ZHAO GuoPing^{3,9,10,14†}, LI YiXue^{2,9†} & CAO ZhiWei^{1,2†})

¹ School of Life Sciences and Technology, Tongji University, Shanghai 200092, China;

² Shanghai Center for Bioinformation Technology, Shanghai 200235, China;

³ School of Life Sciences, Fudan University, Shanghai 200433, China;

⁴ Cancer Institute of the Second Military Medical University, Shanghai 200433, China;

⁵ School of Life Sciences and Technology, South China University of Technology, Guangzhou 510641, China;

⁶ National Engineering Research Center for Antibody Drugs, Shanghai 201203, China;

⁷ Zhujiang Hospital, Southern Medical University, Guangzhou 510282, China;

⁸ Shanghai Municipal Center for Disease Control & Prevention, Shanghai 200336, China;

⁹ Shanghai Institutes of Biological Sciences, Chinese Academy of Sciences (Lab of Computational Genomics, CAS-MPG Partner Institute of Computational Biology; Center of Bioinformatics, Key Lab of System Biology; Lab of Synthetic Biology, Institute of Plant Physiology & Ecology; Institute Pasteur), Shanghai 200031, China;

¹⁰ National Engineering Research Center for Biochip Technology, Shanghai Biochip Co. Ltd, Shanghai 201203, China;

¹¹ National Institute for Parasitic Diseases, Chinese Center for Disease Control and Prevention, Shanghai 200025, China;

¹² Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, Shanghai 200032, China;

¹³ Shanghai Veterinary Research Institute, Chinese Academy of Agricultural Sciences, Shanghai 200241, China;

¹⁴ National Engineering Research Center for Biochip Technology, Shanghai Biochip Co. Ltd, Shanghai 201203, China

In recent months, a novel influenza virus H1N1 broke out around the world. With bioinformatics technology, the 3D structure of HA protein was obtained, and the epitope residues were predicted with the method developed in our group for this novel flu virus. 58 amino acids were identified as potential epitope residues, the majority of which clustered at the surface of the globular head of HA protein. Although it is located at the similar position, the epitope of HA protein for the novel H1N1 flu virus has obvious differences in the electrostatic potential compared to that of HA proteins from previous flu viruses.

novel H1N1 influenza virus, HA protein, spatial epitope, bioinformatics

Since the outbreak of the novel H1N1 influenza virus in Mexico in March, 2009, the number of the infected cases has kept increasing daily in over 30 countries^[1-3]. It has been reported that human beings have not obtained resistance to this type of new virus^[4]. Although antibodies and related reagents are highly sensitive and specific to diagnosis and the treatment of flu viruses, progress in developing antibodies for this novel virus is

seriously restrained due to the scarcity and usage limitation of the imported virus strains. In this case, predicting the potential epitope only from the virus sequences re-

Received May 17, 2009; accepted May 22, 2009

doi: 10.1007/s11434-009-0429-3

†Corresponding authors (email: gpzhao@sibs.ac.cn, yxli@sibit.org, zwcao@sibit.org)
Supported by the National Key Basic Research and Development Program of China (Grant Nos. 2004CB720103, 2006AA02312) and Shanghai Education Foundation (Grant Nos. 000236018, 2000236016)

leased by the National Center of Biotechnology Information (NCBI) is highly desired to help the development of specific antibodies.

Previous investigations found that the antigenic determinants of flu virus were determined by the HA protein sitting in the virus envelopes^[5]. Based on the complete gene sequence of the HA protein from the first strain of novel H1N1 virus (FJ966082, A/California/04/2009(H1N1)) released by NCBI, the 3D structure of HA protein was obtained with homology modeling and the epitope residues were predicted with the group-grown tool SEPPA^[6].

Homology modeling was performed with program Modeller 9V6^[7]. Because there were two separate segments in the released HA structural files, separate templates were selected for each segment accordingly before joining into one complete protein. Six HA proteins of flu virus were selected as the templates, including HA

proteins of swine H9 influenza virus (PDB ID: 1JSD), avian H5 influenza virus (PDB ID: 1JSM), 1930 swine H1 influenza virus (PDB ID: 1RV0), 1934 H1 influenza virus (PDB ID: 1RVX), H3N2 influenza virus (PDB ID: 2VIU) and H5 influenza virus (PDB ID: 2IBX). These templates were chosen from the representative structures of HA proteins which had been grouped according to the sequence similarity higher than 95%. The homologous similarity between the template and the query sequence was 40%–84% for the HA1 segment with *E*-value of 0.0, and 63%–92% for the HA2 segment with *E*-value of 0.0. The result of sequential alignment is mapped in Figure 1(a). In order to avoid bias caused by single modeled structure, ten different initial structures for HA protein were built in the homology modeling process. CHARMM, InsightII program was used to eliminate bad contacts in the initial structures, and the modeled structures see the online supplementary material 1. PROC-

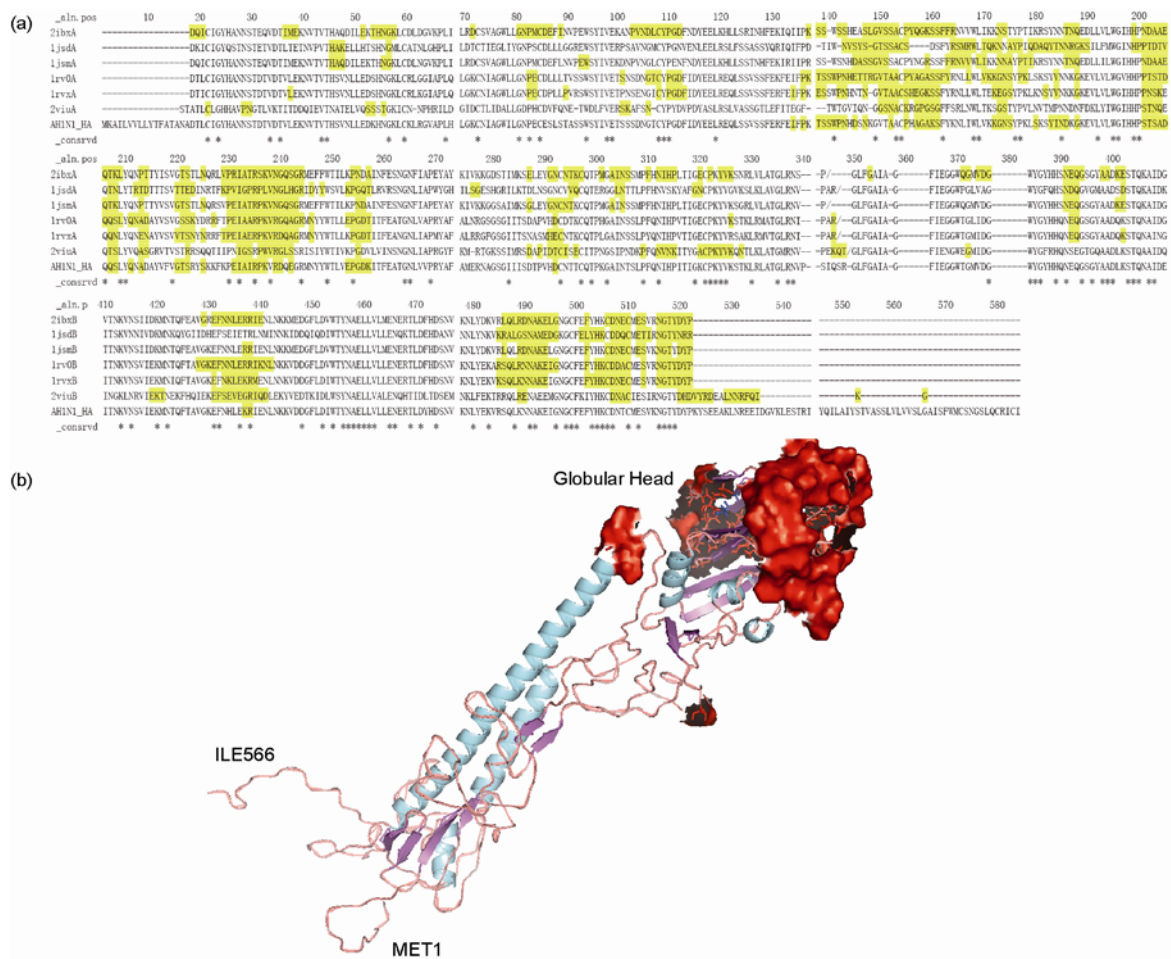


Figure 1 Modeled structure of HA protein for novel H1N1 influenza virus. (a) The result of sequential alignment for HA proteins. The residues of the potential spatial epitope are highlighted in yellow in the sequences. The symbol of asterisks is used to indicate the conserved residues for template proteins. (b) Modeled structure of HA protein and potential epitope. HA protein is shown with the solid ribbon mode; spatial epitope is shown with the surface mode.

HECK program was then applied to check the modeled structures^[8,9]. The detailed results of analysis see the online supplementary material 2.

The modeled structures of HA protein were submitted to SEPPA for predicting the possible spatial epitope, with the default threshold of 1.8. Among all 10 modeled structures, a residue was considered as an epitope residue only if it was a consensus. The results of epitope prediction showed that the potential spatial epitope residues mainly assembled at the surface of globular head and tail segments of HA protein. Because the tail part was usually buried inside the virus envelope^[5], only the epitope residues at the globular head segment were retained for further analysis. In the end, 58 residues were identified as the potential epitope residues which are highlighted in yellow in Figure 1(a), the majority of which clustered at the surface of the globular head, as shown in Figure 1(b).

The predicted epitope was also compared to the corresponding ones of the HA protein templates from pre-

vious flu viruses identified by SEPPA under the same parameters. The detailed results of epitope prediction for the six templates of HA proteins can be found in online supplementary material 3. As shown in Figure 2, the epitope was located at the similar position of HA proteins, but the chemical features of epitope patches were distinct among them. According to the calculated Gasteiger charge parameter, the electrostatic potential has been mapped to the surface of epitope residues. For the HA proteins of previous flu viruses, the epitope contained residues with strong positive potential (blue), or contained residues with both strong positive and strong negative potential (red). As to the HA protein of novel H1N1, the predicted epitope was found to have light positive or neutral potential, whereas residues with negative potential were not found there. The difference of electrostatic potential in the epitope area may affect the recognition and binding between the antibody and antigen. Further experiments are going on to validate the results in this work.

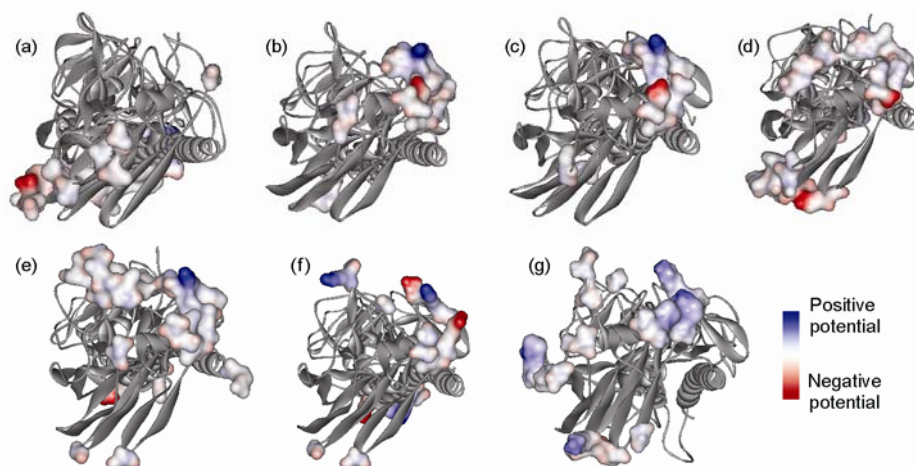


Figure 2 Potential spatial epitope of HA protein for novel H1N1 influenza virus and previous flu viruses. (a) H3N2 Influenza virus (PDB ID: 2VIU); (b) H5 influenza virus (PDB ID: 2IBX); (c) avian H5 influenza virus (PDB ID: 1JSM); (d) swine H9 influenza virus (PDB ID: 1JSD); (e) 1930 Swine H1 influenza virus (PDB ID: 1RV0); (f) 1934 H1 influenza virus (PDB ID: 1RVX); (g) novel H1N1 influenza virus (top-view of spatial epitope). Structure of the whole protein is shown with the solid ribbon mode, and potential epitope is shown with the surface mode. The surface is colored according to the electrostatic potential, red for negative potential and blue for positive potential.

- 1 WHO Influenza A (H1N1)-update 29. http://www.who.int/csr/don/2009_05_15/en/index.html, 15 May 2009 accessed 15 May 2009
- 2 Jonathan R K. Swine influenza. *J Clin Pathol*, 2009, doi:10.1136/jcp.2009.067710
- 3 Fraser C, Donnelly C A, Cauchemez S, et al. Pandemic potential of a strain of influenza A (H1N1): Early findings. *Science*, 2009, doi: 10.1126/science.1176062
- 4 Irvine R M, Brown I H. Novel H1N1 influenza in people: Global spread from an animal source. *Vet Rec*, 2009, 164: 577–578
- 5 Ekiert D C, Bhabha G, Elsliger M A, et al. Antibody recognition of a highly conserved influenza virus epitope. *Science*, 2009, 324: 246–251, doi: 10.1126/science.1171491
- 6 Sun J, Wu D, Xu T L, et al. SEPPA: A computational server for spatial epitope prediction of protein antigens. *Nucleic Acids Res*, 2009, doi: 10.1093/nar/gkp417
- 7 Sali A, Potterton L, Yuan F, et al. Evaluation of comparative protein modeling by MODELLER. *Proteins: Struct Funct Bioinformatics*, 1995, 23: 318–326, doi: 10.1002/prot.340230306
- 8 Laskowski R A, MacArthur M W, Moss D S, et al. PROCHECK: A program to check the stereochemical quality of protein structures. *J Appl Cryst*, 1993, 26: 283–291, doi: 10.1107/S0021889892009944
- 9 Morris A L, MacArthur M W, Hutchinson E G, et al. Stereochemical quality of protein structure coordinates. *Proteins*, 1992, 12: 345–364, doi: 10.1002/prot.340120407