

Structural basis and sequence co-evolution analysis of the hemagglutinin protein of pandemic influenza A/H1N1 (2009) virus

Herman Tse^{1,2}, Richard Y T Kao^{1,2}, Wai Lan Wu¹, Wilian W L Lim³, Honglin Chen^{1,2}, Ming Yiu Yeung¹, Patrick C Y Woo^{1,2}, Kong-Hung Sze⁴ and Kwok-Yung Yuen^{1,2}

¹Department of Microbiology and State Key Laboratory for Emerging Infectious Diseases, The University of Hong Kong; ²Research Centre of Infection and Immunity, The University of Hong Kong; ³Public Health Laboratory Services, The Centre for Health Protection, Department of Health, The Government of Hong Kong Special Administrative Region; ⁴Department of Chemistry, The University of Hong Kong, Hong Kong

Corresponding authors: Professor Kwok-Yung Yuen, Department of Microbiology, Carol Yu Centre for Infection, University Pathology Building, The University of Hong Kong, Pokfulam, Hong Kong. Email: hkumicro@hkucc.hku.hk and Dr Kong-Hung Sze, Room 603, Chong Yuet Ming Chemistry Building, The University of Hong Kong, Pokfulam, Hong Kong. Email: khsze@hku.hk

Abstract

Severe pandemic influenza A H1N1 (2009) infection, especially in the lower respiratory tract, is often associated with the virus carrying a D222G substitution in the hemagglutinin (HA) protein of the virus. The mechanism for this association has not been fully explored. In the *in vitro* binding assay, it was found that clinical isolates carrying D222G substitution exhibit higher binding avidity to 2,3-linked sialic acids than the wild-type virus. The receptor binding pocket of the pandemic influenza (H1N1) HA was found to be smaller than those of other influenza A strains, allowing tighter binding of the virus with the receptor, yet also inducing steric stress for the binding. Our homology modeling and molecular docking calculations implicated that residue 222 may affect the positioning of the conserved Q223 residue, hence modulating flexibility of the binding pocket and steric hindrance during receptor binding. The molecular property of residue 222 can also directly influence the 'lysine fence' via the polarity of the amino acid residue where D222G substitution will enhance the electrostatic interactions between the receptor and the protein. The potential importance of residue 222 was illustrated by evolutionary analysis, which showed that this site is under intense selection pressure during adaptation of the virus to human host. Our findings provide a useful reference for follow-up studies in monitoring the ongoing evolution of the pandemic influenza A H1N1 (2009) virus.

Keywords: pandemic, influenza A, structural evolution, co-evolution, hemagglutinin, D222G

Experimental Biology and Medicine 2011; **236**: 915–925. DOI: 10.1258/ebm.2011.010264

Introduction

The influenza A virus has caused seasonal epidemics and some of the biggest pandemics in human history. Given the fact of the devastating impact of the pandemic 1918 'Spanish flu' and the most recent 2009 (H1N1) pandemic,^{1,2} preparation for the future influenza pandemic is warranted. Despite the significant effort in worldwide surveillance network for emerging influenza strains, the swine-origin influenza A/H1N1 virus emerged largely undetected until substantial numbers of people were infected and caused the pandemic in 2009. Indeed, transmission of the swine influenza virus strains to humans had already occurred in as early as 1976, and sporadic cases have been detected since 1998.^{3–5} Except for the event of 1976, the relatively limited scale of cross-species transmissions and the mild

clinical presentation has not drawn much attention from public health authorities and the scientific community. Arguably, the focus on avian influenza viruses, particularly the highly pathogenic A/H5N1 strains,⁶ was justified, in view of the significant mortality from reported human cases of influenza A/H5N1 infections.^{7,8} The preparedness for a potential pandemic potential of avian influenza viruses had been the major motivation behind international vigilance, which has paid off handsomely in the form of an effective and coordinated global public health response to the novel influenza A/H1N1 (2009) virus.

Earlier studies have identified that the pandemic (H1N1) 2009 virus was a result of multiple re-assortment events with viral genes originating from avian, swine and human influenza strains.⁹ Advanced phylogenetic analysis enabled the construction of a temporal profile of viral

evolution prior to and around the beginning of the pandemic, and possibly even before it had jumped the species barrier.^{10,11} However, these studies could not provide information regarding the particular genetic changes that allow efficient human infection and transmission caused by the pandemic (H1N1) 2009 virus. While there had been three recorded pandemics in the last century, this was the first time we were able to track the pandemic from its early beginning. It is essential to reconcile descriptions of the evolutionary changes with their functional significance for understanding virus adaptation during the early phase of a pandemic. Using influenza hemagglutinin (HA) as an example, functional mutations of this viral protein would be expected to result in variants in antigenicity, receptor binding properties and potentially leading to a change in clinical presentation. We and others found that a variant with D222G (D225G in H3 numbering) in the HA gene of the pandemic (H1N1) 2009 virus is strongly associated with increased disease severity in both humans and mice.^{12–15} The D222G variant is more frequently detected from specimens obtained in the lower respiratory tract of severe patients.¹² A receptor binding assay with a mouse-adapted virus containing the D222G substitution in the HA and its parental virus suggested that D222G may lead to an increase in binding to the α 2,3-linked sialic acids receptor by the virus.¹⁵ However, the structural basis for an altered receptor binding preference by the D222G substitution remains unclear. The D222G mutation in the pandemic influenza (H1N1) virus may be just one of the markers in the adaptation to humans. It is likely that other mutations may exist throughout the influenza genome and act functionally to facilitate the virus' adaptation to the new host environment. Several research groups have described sites in the viral genome under positive selection in an attempt to identify the key functional residues in proteins of the pandemic (H1N1) influenza 2009 virus.^{16,17}

In the present study, we extended the analysis to delineate the effect of the mutations on viral protein structure and function in sialic acid receptor binding and correlations among functionally important mutations. Our results reveal a potential mechanism for the effect of the widely reported D222G mutation in the HA molecule of the virus on receptor binding, which might help explain its association with clinical severity and tissue tropism.

Materials and methods

Virus-turkey erythrocyte binding assay

Clinical isolates of the pandemic (H1N1) 2009 virus were obtained from the Center for Health Protection of Hong Kong Special Administrative Region and isolates from severe cases were obtained from the Virus Unit of the Queen Mary Hospital, Hong Kong Special Administrative Region. The virus-turkey erythrocyte binding assay was performed as previously described with minor modifications.¹⁸ The optimal condition for neuraminidase treatment of turkey erythrocyte was tested using A/PR/8/34 isolate. Briefly, turkey erythrocytes were treated with different concentrations of *Vibrio cholerae* neuraminidase (Sigma,

St Louis, MO, USA) for 60 min at 37°C to remove the 2,3-linked sialic acid. The erythrocytes were washed twice with phosphate-buffered saline (PBS) and then diluted to 2% (v/v) erythrocyte solutions with PBS. The 2% erythrocyte solution (25 μ L) was mixed with eight HA units of influenza viruses (100 μ L) and incubated at room temperature for 60 min. Ten isolates with 222D and four isolates with 222G were tested in this study. Hemagglutination was measured and data expressed as the maximal concentration of neuraminidase that allowed for full hemagglutination.

Sequence and structural data

Complete gene and genome sequence data of the pandemic influenza A/H1N1 (2009) virus were downloaded from the National Center for Biotechnology Information Influenza Virus Resource (<http://www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html>).¹⁹ The crystal structural data of influenza viral proteins were obtained from Research Collaboratory for Structural Bioinformatics Protein Data Bank (<http://www.pdb.org>).

Homology modeling

Homology models of influenza HA protein were constructed using the SWISS-MODEL homology modeling server²⁰ using established protocols (<http://swissmodel.expasy.org>).²¹ The HA protein of the ligand-free crystal structure of the pandemic influenza A/H1N1 (2009), 3LZG,²² was used as structure template for homology modeling of the mutated sequences. To better access the effect of sialylated glycan ligand binding on the active site, homology modeling of the HA proteins was also carried out using the following glycan bound HA protein crystal structures as templates: 1JSI (swine influenza A/H9N2 [1998] with α 2,6-linked sialic acid ligand),²³ 1JSN (avian influenza A/H5N1 [1997] with α 2,3-linked sialic acid ligand)²³ and 3HTT (avian influenza A/H1N1 [2005] with α 2,3-linked sialic acid ligand).²⁴ Template models were edited and sequence aligned using the SwissPdb Viewer programme 'Deep-View'²⁵ (<http://spdbv.vital-it.ch/>) before submission to the SWISS-MODEL server for carrying out homology modeling.

Molecular docking

Molecular docking studies were performed on a Dell Dimension 5150 machine with a 3.0 GHz dual core Xenon central processing unit (CPU), 2048 MB RAM and running on Fedora 8 Linux operating system. Autodock Vina (<http://vina.scripps.edu/>)²⁶ was used to identify optimal binding modes for the binding of sialic acid ligands to the known binding site on the homology models of mutant influenza HA proteins. The structures were prepared for the docking study as follows: for the protein, waters were removed from the PDB file and hydrogen atoms were added; Gasteiger charges, atomic solvation parameters and fragmental volumes were merged to the receptor. The templates for modeling of α 2,6-linked sialic acid ligands were obtained from the following PDB entries: 1JSI, 1MQN, 1RVO, 1RVT, 1RVZ and 3HTQ. The templates for modeling of α 2,3-linked sialic acid ligands were obtained from the following PDB entries: 1JSN,

1MQM, 1RVX, 3HTP and 3HTT. For all ligand templates, structures were minimized to a local energy minimum using the CHARMM-like force field. Gasteiger charges were assigned and non-polar hydrogen atoms were merged. The ligands were allowed to have all torsions rotatable except the rings during docking simulations, whereas the protein molecular structure was either rigid or allowing flexible side chain torsions for the residues inside the known binding site. The docking search grid was estimated with the auxiliary program AutoDockTools 1.5.4. The grid dimensions were $40 \text{ \AA} \times 40 \text{ \AA} \times 40 \text{ \AA}$ and $45 \text{ \AA} \times 45 \text{ \AA} \times 45 \text{ \AA}$ along the x -, y - and z -axis for HA receptor proteins with rigid and flexible binding pockets, respectively. The grids were chosen to be sufficiently large to cover significant portions of the active pocket of the HA protein. The center of the docking ligand was initially positioned at the grid center. Grid searching was performed using the Lamarckian genetic algorithm with other parameters left at default values. Autodock Vina uses a sophisticated gradient optimization method in its local optimization procedure and it can further speed up the execution by taking advantage of multiple CPUs or CPU cores by using multithreading. The number of steps in a run was determined adaptively, according to the apparent complexity of the problem, and several runs starting from random conformations were performed concurrently using multithreading. The optimization algorithm maintained a set of diverse significant minima found that were then combined from the separate runs and used during the structure refinement and clustering stage. Comparison of docking results and generation of molecular graphics images were performed using the UCSF Chimera package.²⁷

Sequence analysis

All genes and translated protein sequences were aligned with ClustalW²⁸ using default parameters. Sequences with slight length variations due to indels were manually edited to maintain consistency of the data-set. Sequences containing ambiguous bases or residues were removed.

Identification of sites under positive selection was performed using the following methods as implemented in HyPhy:²⁹ single likelihood ancestor counting, fixed effects likelihood and directional evolution of protein sequences (DEPS).³⁰ Phylogenetic reconstruction was performed using the neighbor-joining method within HyPhy. The Consurf webserver (<http://consurf.tau.ac.il/>)³¹ was used for calculation and visualization of conserved and variable regions in the three-dimensional structure of the pandemic influenza HA protein.

Co-evolution among residues of the influenza HA protein was examined by mutual information (MI) analysis by using a normalized measure of the MI value to assess the correlation between any two sites. The MI matrix was computed as described in previous studies.^{32,33} Hierarchical cluster analysis was then performed with the dChip program to group sites showing highly correlated mutations. Identification of 'site fixation' in an attempt to predict future trends in mutation was performed in accordance to methods described in previous studies.^{32,34}

Statistical analysis

Statistical analysis was performed using R version 2.11.0. All P values reported are for a two-tailed test, and $P < 0.05$ is considered statistically significant.

Results

The D222G variant in the HA of the pandemic (H1N1) 2009 virus by virus-turkey erythrocyte binding assay

The clinical isolates that contained 222D served as the positive control, which still produced hemagglutination despite treatment of the turkey erythrocyte by $32 \mu\text{g/mL}$ of neuraminidase, whereas the four clinical isolates with D222G substitution had demonstrated hemagglutination at a neuraminidase concentration of not more than $16 \mu\text{g/mL}$ (Figure 1). Our result suggests that the four clinical isolates carrying D222G substitution were more sensitive to the

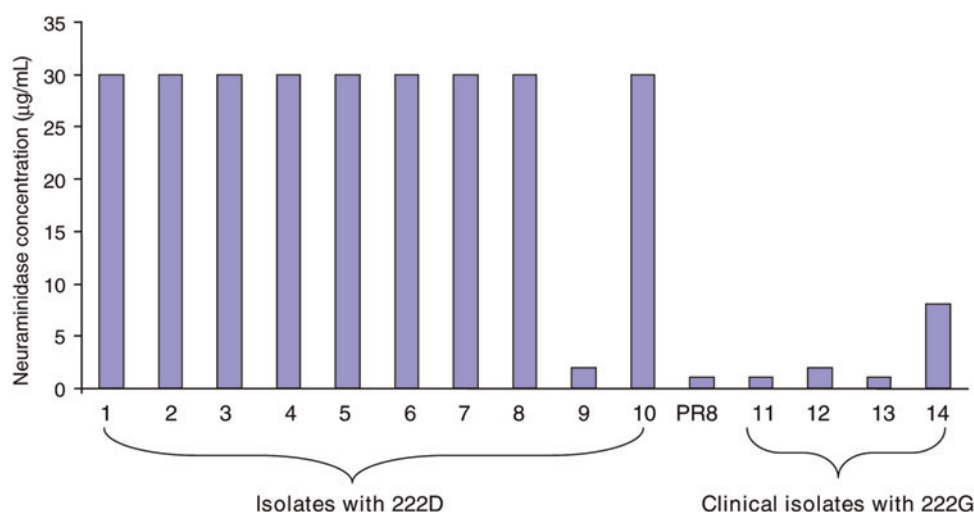


Figure 1 Cellular receptor binding avidity assay. Influenza virus cellular receptor binding avidity was estimated with turkey blood cells after treatment with *Vibrio cholerae* neuraminidase which mainly removes 2,3-linked sialic acids as described in the Materials and methods. The data represent the mean value of two independent experiments. PR8, A/Puerto Rico/8/34 is a reference virus used to validate the assay as described in a previous study¹⁶

treatment with neuraminidase in turkey red cells in the hemagglutination assay, suggesting that the D222G variant has a higher binding avidity to 2,3-linked sialic acids than the wild-type virus.

Structural analysis of pandemic influenza A HA

To study the effects of amino acid polymorphism at residue 222 of the pandemic influenza A/H1N1 (2009) HA protein, three homology models were constructed to represent the D222G, D222E and D222N variants using the wild-type HA protein (D222-WT) as the template. As the four proteins are highly similar in sequence (>99% identity) and structure, the structural comparison is mainly focused around the receptor binding pocket which is located in close proximity to residue 222. Among the differences generated by the amino acid substitutions, the change in surface charge has the most apparent effect on the binding of the ligand to the pocket. The 'lysine fence' is a long stretch of basic surface charges created by the lysine residues at position 130, 142 and 219 (H1 numbering), which are positioned to anchor the Sia1 and galactose (Gal2) sugars of both α 2,3 and α 2,6 glycans.³⁵ This structural feature is considered more significant for the pandemic influenza A/H1N1 (2009) virus, as the lysine residue at position 142 is not found in the other H1N1 strains. Incidentally, the amino acid residue at position 222 is part of the 220-loop which interacts with this lysine fence. A neutral glycine or basic asparagine mutation at this position would extend and enhance the basic charge of the lysine fence, whereas the wild-type aspartic acid residue or the glutamic acid mutation would create a localized patch of acidic charge (Figure 2). The latter effect would result in a diminishing of the attractive effect of the lysine fence towards the sialylated glycans.

Comparison of the crystal structure of the pandemic influenza A/H1N1 (2009) HA protein (3LZG) with available

crystal structures of other influenza A HA proteins showed that the conformation of the receptor binding pocket is relatively well conserved. Despite the high overall structural similarity of the HA proteins, we noted a considerable variation in the widths of the opening of their receptor binding pockets. The shortest distance between the 190-helix and the 130-loop across the binding site for the N-acetylneuraminic acid (Sia1) ring is only 9.45 Å in the pandemic influenza A/H1N1 (2009) HA protein, whereas the corresponding distances are 10.54 Å for the avian influenza A/H1N1 HA (pdb 3HTT), 10.58 Å for the avian influenza A/H5N1 HA (pdb 1JSN) and 11.05 Å for the swine influenza A/H9N2 HA (pdb 1JSI) (Figure 3a). Compared with the ligand bound X-ray structures of other HA proteins shown in Figure 3b, *in silico* molecular docking of an α 2,3-linked sialylated oligosaccharide ligand to the conserved binding position onto the pandemic influenza A/H1N1 (2009) HA protein would show at least three regions of steric clashes between the ligand and protein in the absence of conformational flexibility of the binding site. The binding of the α 2,6-linked sialylated oligosaccharide ligand to the pandemic influenza A/H1N1 (2009) HA protein would be expected to experience similar steric stress. Therefore, a larger conformational change or flexibility of the receptor binding pocket of the pandemic influenza A/H1N1 (2009) HA protein will be required to allow the effective binding of sialylated oligosaccharide ligands.

The effects of the D222 polymorphisms were further explored by molecular docking of sialylated glycan ligands to the pandemic influenza A/H1N1 (2009) HA variants. Under the condition of no conformational flexibility of the HA molecule, the repertoire of *in silico* molecular docking results generated by Autodock Vina were mostly suboptimal and did not match the conserved binding position of the Sia1 ring known from crystal structures with bound sialylated glycan ligands (data not shown). Even allowing for flexible side-chains on the residues in the

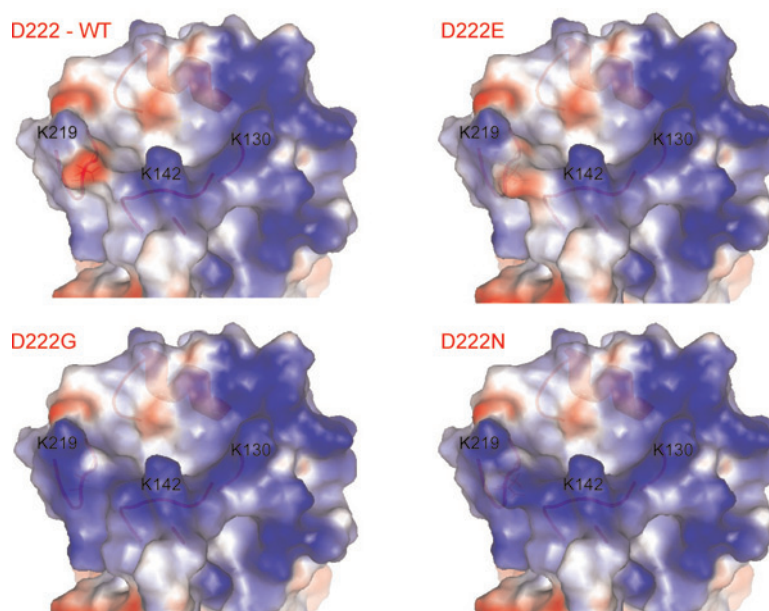


Figure 2 Distribution of surface charges around the sialylated glycan binding site on the pandemic influenza A/H1N1 (2009) hemagglutinin (HA) protein. Blue and red colors represent basic and acidic charges, respectively. Lysine residues at position 130, 142 and 219 (H1 numbering) can be shown to form a stretch of basic charges known as the 'lysine fence'. D222G and D222N mutations would significantly extend and enhance the basic charge of the lysine fence

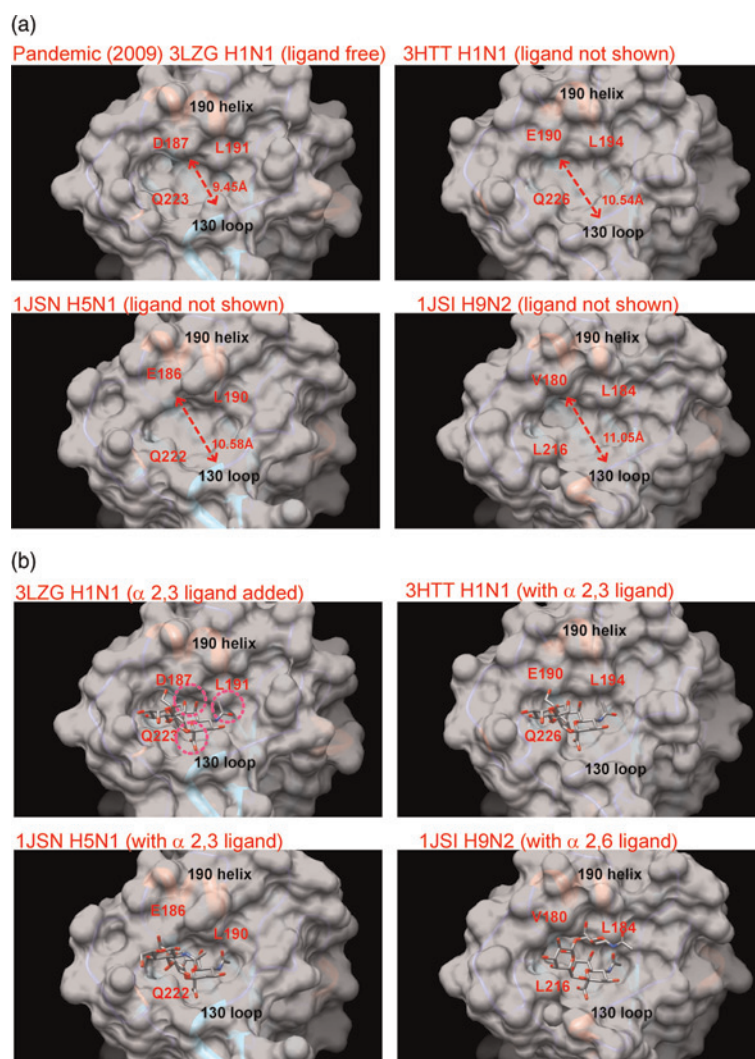


Figure 3 Structural comparison of the receptor binding pocket of the influenza A virus hemagglutinin (HA) protein. (a) The opening of the pocket of the pandemic influenza A/H1N1 (2009) HA protein is appreciably narrower than that of the other influenza viruses shown. Dotted line indicates the shortest distance between the 190-helix and the 130-loop with the distance measurement shown next to the line. (b) Molecular docking of an α 2,3-linked sialic acid ligand to the receptor binding pocket of pandemic influenza A/H1N1 (2009) HA protein showed regions of steric clashes between the two molecules. Regions of the steric clashes are highlighted with dotted circles. No such steric clashes are observed with the ligand bound crystal structures of the other influenza HA proteins (3HTT, 1JSN, 1JSI)

receptor binding pocket, only the α 2,3-linked sialic acid was able to produce a docking model with the D222G variant at the conserved Sia1 position, whereas all α 2,6-linked sialic acid still failed to dock at the conserved position.

To better access the structural features of receptor-bound HA protein of the pandemic influenza A/H1N1 (2009) in the context of a conformationally widened binding pocket, homology modeling of the wild-type HA protein as well as its D222G, D222E and D222N variants were also carried out using the following glycan bound HA protein crystal structures as templates: 1JSI (swine influenza A/H9N2 [1998] with α 2,6-linked sialic acid ligand),²⁰ 1JSN (avian influenza A/H5N1 [1997] with α 2,3-linked sialic acid ligand)²² and 3HTT (avian influenza A/H1N1 [2005] with α 2,3-linked sialic acid ligand).²³ The resulting homology models of the pandemic influenza A/H1N1 (2009) HA protein have acquired wider receptor binding pockets resulting at higher successful docking rates for both α 2,3- and α 2,6-linked sialic acid ligands. In particular, the

α 2,3-linked sialic acid ligand is able to dock to the conserved position for D222 WT (with 3HTT and 1JSN pockets), D222G (with 3HTT and 1JSN pockets), D222N (with 1JSN pocket) and D222E (with a 1JSN pocket allowing flexible side-chains); whereas the α 2,6-linked ligand is able to dock to D222 WT (with 1JSI pocket), D222N (with a 1JSI pocket allowing flexible side-chains) and D222E (with a 1JSI pocket allowing flexible side-chains). These results support that a flexible or wider receptor binding pocket of the pandemic influenza A/H1N1 (2009) viral HA is critical for its binding to sialylated glycans.

Sequence and co-evolutionary analysis of pandemic influenza A HA

A total of 795 non-redundant full-length pandemic influenza HA protein sequences from the period of April 2009 to March 2010 were included for analysis. MI analysis was performed as described above. Using the arbitrary cut-off

of 0.5 from the previous study on H3N2 HA protein,³¹ 45,199 (28.3%) out of 159,895 possible pairs were found to have possible correlations in mutational patterns. This number is reduced to 816 (0.5%) if the cut-off is raised to 2.0 (Figure 4). The amino acid residues found to have correlation with mutations at residue 222 are shown in Table 1. The most significant correlation among these was observed for residue 298 of HA1, although a surprising number of the other residues were located in HA2. Examination of the mutations in residues 222 and 298 identified the D222E and P298S mutations as significantly correlated ($\phi = 0.62$; $P < 0.00001$ by Fisher's exact test; Table 2). While the two residues are not in close proximity with each other, the location of residue 298 near the C-terminal of HA1 led us to speculate that it may affect the flexibility of the protein. Proline is known to have a rigid conformation and may confer local structural stability and low flexibility.^{36,37} Hence, its substitution by serine may permit increased flexibility at the interface between HA1 and HA2, with downstream effects on the receptor binding by the protein.

Examination of amino acid residue pairs with the highest MI scores (Table 3) showed that correlations of mutations between residues often occur over relatively large distances. Only four out of the 10 pairs were residues from the same mature peptide fragment. The most intriguing finding is the presence of two amino acid residue pairs with one member located in the signal peptide. Given the temporal and spatial compartmentalization of the peptides, it is implausible for any direct interaction between the signal peptide and the HA1/HA2 molecules. Whether the apparent correlation is an artifact introduced during nucleic acid amplification and sequencing or a result of an uncharacterized mutational bias remains to be seen. Nonetheless, it is a clear reminder that 'positive' results from data mining have to be placed in a relevant biological context and be confirmed experimentally.

Hierarchical cluster analysis of the MI matrix revealed several small clusters of correlated sites which were also

Table 1 Correlation of mutations at amino acid residue 222 (H1 numbering) of the pandemic influenza A/H1N1 (2009) hemagglutinin (HA) protein with other residues of the same protein

Position on full-length consensus HA sequence (H1 numbering)	HA1/HA2	Normalized mutual information score
18 (2)	HA1	3.212
33 (17)		3.893
36 (20)		3.408
85 (69)		3.318
149 (133)		3.805
154 (138)		10.859
179 (163)		2.304
226 (210)		8.743
314 (298)		31.839
366	HA2	4.048
373		4.135
377		4.003
412		2.885
429		3.761
465		11.948
537		2.105
541		2.196

Table 2 Correlation of D222E and P298S mutations in the pandemic influenza A/H1N1 (2009) hemagglutinin (HA) protein

Polymorphisms	222D/G/N	222E
298P	753	22
298S	2	18

Numbers shown are the sequence counts for the specified polymorphisms

closely related in spatial distribution, such as the first eight amino acids of the signal peptide region. The D222 residue did not form a distinct cluster with other residues. A large number of sites were found to be grouped in a significant cluster but all of the sites showed only a weakly positive MI score of less than 2.0. Further examination revealed that most of these sites correspond to the second

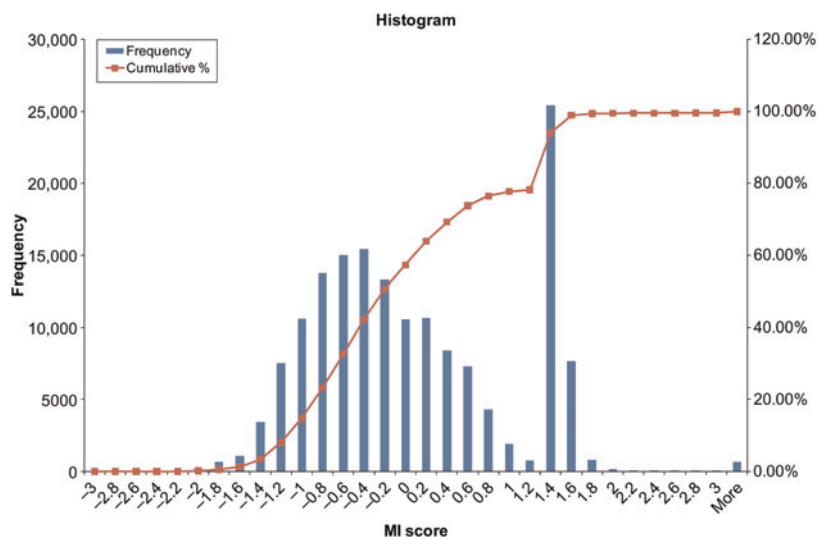


Figure 4 Histogram of the normalized mutual information (MI) scores from MI analysis of the pandemic influenza A/H1N1 (2009) hemagglutinin (HA) protein. It shows a bimodal distribution of MI scores with a broad peak centered around -0.4 and a sharp peak located at around 1.4

Table 3 Top 10 pairs of amino acid residues of pandemic influenza A/H1N1 (2009) hemagglutinin (HA) protein with highest mutual information (MI) scores

Rank	1st residue	2nd residue	MI score
1	52 (HA1)	146 (HA1)	33.233
2	311 (HA1)	495 (HA2)	33.022
3	2 (signal peptide)	310 (HA1)	32.534
4	239 (HA1)	314 (HA1)	31.836
5	15 (signal peptide)	466 (HA2)	31.346
6	472 (HA2)	527 (HA2)	29.780
7	302 (HA1)	544 (HA2)	29.270
8	163 (HA1)	494 (HA2)	28.665
9	293 (HA1)	543 (HA2)	28.546
10	399 (HA2)	490 (HA2)	28.534

Residue positions refer to the full-length consensus HA sequence

peak of the histogram shown in Figure 4. The association among these sites could not be confirmed with statistical tests of association such as the Fisher's exact test, and our assessment is that the degree of correlation, if any, does not meet the criteria for statistical significance.

A total of 1476 non-redundant full-length pandemic influenza HA coding sequences from the period of April 2009 to March 2010 were downloaded for sequence analysis. Positive selection detection was performed using the different algorithms implemented in HyPhy. A total of 12 sites in the pandemic influenza A/H1N1 (2009) HA sequence were identified to be under selection by one or more of the algorithms (Tables 4 and 5). Although directional selection and positive selection act on different levels (amino acids and codons, respectively), a common set of amino acid residues/codon sites was identified to be under both types of selection. The most significant degree of selection was noted for residues 203 and 222 (H1 numbering) of HA1, with estimated dN/dS (ratio of the rate of substitutions at synonymous sites to the rate of substitutions at non-synonymous sites) of greater than 15 and Bayes factors of greater than 10.¹⁶ The dominant mutations at each position by

Table 5 Sites under directional selection as detected using the DEPS method as implemented in HyPhy

Codon position on consensus HA sequence (H1 numbering for HA1 if applicable)	Preferred residues	Maximum Bayes factor
49 (32)	I	1.603×10^6
56 (39)	E	232.5
103 (86)	G	5.691×10^6
203 (186)	T	9.128×10^4
204 (187)	G	5.330×10^3
214 (197)	S, T	2.498×10^5
220 (203)	T, (S*)	2.929×10^{20}
232 (215)	T	108.8
239 (222)	D, E, G, N	1.999×10^{20}
252 (235)	K	1.008×10^3
258 (241)	I	373
278 (261)	T	110.3
291 (274)	N	685.3
300 (283)	N	3.435×10^3
338 (312)	I	3.085×10^3
391	G, K	4.172×10^3
468	N	115.1
472	T	100.3
558	G	2.104×10^3
561	S	184.4
564	K	1.232×10^9

DEPS, directional evolution of protein sequences; IUPAC, International Union of Pure and Applied Chemistry

A Bayes factor of more than 20 is considered evidence for directional selection. Standard IUPAC abbreviations are used for amino acids

*Directionality of evolution for this site may have been reversed due to an artifact resulting from data sampling. The correct preferred residue should be threonine (T)

substitution count were T203S and D222G, although the D222E mutation was a close second to D222G for that position. The sites identified generally correspond to regions that are surface accessible or located in close proximity to the receptor binding pocket, and are responsible for an appreciable proportion of variation observed in the HA sequences (Figure 5). In addition, five 'site fixations' were identified in HA: V36I, V47A, S220T, E391K and I564K.

Table 4 Sites under positive selection as detected using HyPhy

HA segment	Codon position on consensus HA sequence (H1 numbering if applicable)	Estimated dN/dS with SLAC	Estimated dN/dS with 2FEL
Signal peptide	8	3.517	4.701
HA1	203 (186)	7.000	(dS = 0)
	214 (197)	7.009	(dS = 0)
	220 (203)	19.67	(dS = 0)
	239 (222)	15.41	4.083
	278 (261)	5.999	(dS = 0)
HA2	391	ND	2.343
	428	3.497	(dS = 0)
	468	4.844	(dS = 0)
	477	3.328	(dS = 0)
	544	3.994	ND
	558	6.164	7.572

ND, positive selection not detected using this method

Methods used are: single likelihood ancestor counting (SLAC) and two-rate fixed effects likelihood (2FEL). Significance level is set at 0.10 for both methods

Discussion

While the pandemic (H1N1) 2009 virus was found antigenically stable during the first year of circulation in humans, genetic variants were observed. Notably, a variant carrying D222G (H1 numbering) was found mainly in severe cases of patients as reported in several studies from different countries.^{12–14} A previous study found D222G substitution associated with change of receptor specificity from Gal 2,6-linked to 2,3-linked sialic acids in the pandemic 1918 virus.³⁸ We also found a mouse-adapted virus containing a D222G mutation in the HA gene, and which exhibits enhanced virulence in mice.¹⁵ It is suggested that the pandemic (H1N1) 2009 virus with D222G HA may be more likely to infect lower respiratory tract epithelial cells which express predominantly 2,3-linked sialic acids. To further verify if D222G is a phenotypic variant, we examined the receptor binding avidity of wild-type and mutant viruses using turkey blood cells treated with neuraminidase which mainly cleavage 2,3-linked sialic acids. We confirm that D222G variants had a higher preference for binding to

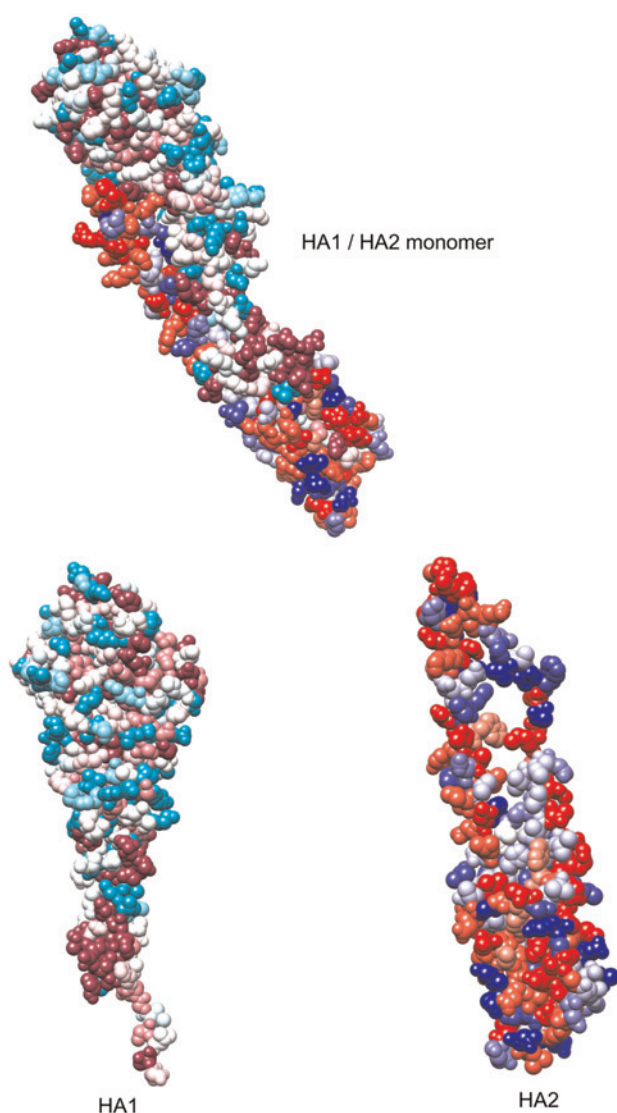


Figure 5 Identification of conserved and variable regions on the pandemic influenza A/H1N1 (2009) hemagglutinin (HA) protein. The most conserved regions are colored magenta (HA1)/red (HA2) and the most variable regions are colored cyan (HA1)/blue (HA2). For HA1, the variable regions are distributed throughout the surface of the molecule, while the constant regions can be found as isolated pockets in the internal regions and C-terminal region in contact with HA2. For HA2, regions of high variability can be found to a lesser extent throughout the surface of the molecule

2,3-linked sialic acids, as illustrated by its higher sensitivity to the treatment of neuraminidase, while the wild-type virus, with its preference for binding to 2,6-linked sialic acids, will exhibit resistance to neuraminidase treatment.

The observed narrower receptor binding pocket of the pandemic influenza A/H1N1 (2009) HA protein may have several implications for its function. Firstly, the sialylated glycans binding to the pocket will experience steric hindrance from the neighboring protein residues. This is demonstrated by the molecular docking of an α 2,3-linked sialylated oligosaccharide ligand to the pandemic influenza A/H1N1 (2009) HA protein (Figure 3b), which showed at least three regions of steric clashes between the ligand and protein in the absence of conformational flexibility of the binding site. Secondly, the narrower opening of the pocket

will likely impact the kinetics of ligand binding in a negative manner, as a larger conformational change of the receptor binding pocket will be required to allow the ligand to adopt the normal bound position and conformation. Thirdly, from a mechanistic point of view, the probability of receptor binding will be decreased but this might be partially offset by a 'tighter' grip of the bound sialylated glycan assuming no further conformation changes occur. Therefore our structural analysis of the pandemic influenza A/H1N1 (2009) viral HA suggested that a flexible or wider receptor binding pocket is critical for its binding to sialylated glycans. While it could be difficult to accurately model protein flexibility without performing a computationally expensive molecular dynamics study, it is possible to infer the potential effects of specific amino acid residues on protein-ligand interaction. It has been reported that the HA protein of the pandemic (H1N1) 2009 influenza virus possesses residues that can be positioned to make optimal contacts with both α 2,3- and α 2,6-linked sialic acid ligands.³⁵

Visual inspection of the *in silico* docking results confirmed that the Sia1 ring of the ligand was unable to fit into the narrowed gap between the 190-helix and the 130-loop if a rigid binding pocket was assumed (Figure 6a). The effect was observed to be even greater for α 2,6-linked sialic acid because of the additional steric clash between Gal2 of the sialylated oligosaccharide ligand and the 220-loop of HA (Figure 6b), in particular with the Q223 residue, which had been shown to be one of the highly conserved residues involving in anchoring Sia1.³⁵ From our analysis, the amino acid residue at position 222 of HA1 is postulated to affect the positioning of the Q223 residue and modulate the steric hindrance of sialylated glycan binding. Additionally, the amino acid at position 222 can either reinforce or antagonize the basic charges of 'lysine fence' depending on the amino acid substitution, hence providing another means of affecting the receptor binding properties of the pandemic influenza HA protein. It is no surprise that D222G/N mutations were quickly found to be genotypic markers associated with disease severity during the pandemic, and the site was also shown to be under great selection pressure by our sequence analysis. In particular, the D222G mutation in the HA protein of pandemic (H1N1) influenza 2009 virus has reported to link to patients with severe disease.¹² Although it is generally considered that the D222 residue can provide optimal contact of the Gal2 on the α 2,6-linked sialic acid ligand,³⁵ this loss can be compensated with the substitution of a less bulky glycine residue because it can better ease the steric stress on Q223 and also provide more favorable charge attraction towards the sialylated glycan.

The findings of the MI analysis are perhaps more surprising with significant correlations found between polymorphisms of distant residues of the HA protein. The number and degree of such correlations are too great to attribute all of them to be experimental or statistical artifacts. Some residue pairs may appear distantly in the same monomer but could be placed more closely in the HA trimer structure. A more rigorous examination of the MI results could be achieved with a hierarchical cluster analysis as correlated

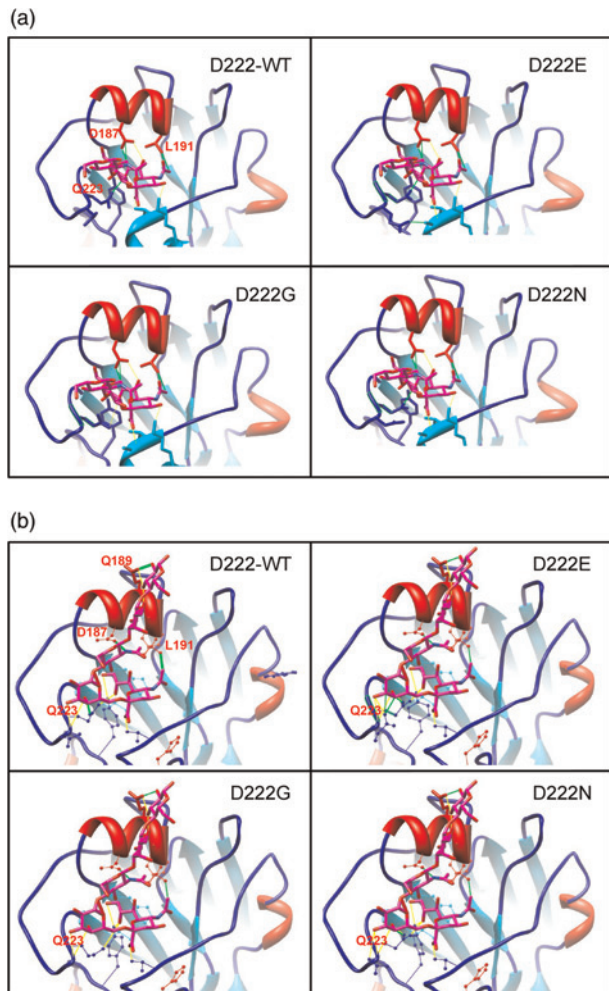


Figure 6 Results of the molecular docking of (a) α 2,3-linked sialic acid ligand and (b) α 2,6-linked sialic acid ligand to the receptor binding pocket of the pandemic influenza A/H1N1 (2009) hemagglutinin (HA) protein, showing steric hindrance by the narrowed opening. Solid green and yellow lines represent specific instances of steric clashes and hydrogen bonds, respectively

residues often form an interaction network, but the results suggest a lack of sufficiently significant clusters to work on. Our interpretation of the situation is that the amount of evolution that had occurred in one year of sampled data was too limited to provide a good description of the interactions within HA. Repeating the analysis with a larger data-set in the future might reveal further details of the nature of the correlations.

We also noted that the direction of selection at residue 203 appears to be inconsistent under different analysis. DEPS suggested that the preferred residue is serine and not threonine while fixation analysis suggested the opposite. The source of this apparent contradiction probably stems from the timing of the occurrence of the selection, as we found that the proportion of sequences with residue S203 was in sharp decline during the first few months of the sampling period. In the absence of a reliable outgroup, this might affect the accuracy of the root placement in the phylogenetic tree supplied for DEPS analysis, which in turn affects the correction determination of directionality.³⁰

Alternatively, the presence of recombination might confound the analysis although we were unable to detect any recombination signal in the data. We contend that the amino acid selected at residue 203 should be threonine but the DEPS analysis had been skewed by the phylogenetic uncertainty introduced as an artifact of the sampling period cut-off. This illustrates the need for applying multiple independent methods during analysis, examination of the actual data and some form of experimental verification whenever possible.³⁹

The occurrence of residue 203 in our results is also interesting in another way, as it ranks as the residue under greatest directional and positive selection in our analysis. This residue is located in an accessible area on the surface of the HA1 molecule and could presumably play a role in the antigenic escape from antibody-mediated immunity.⁴⁰ The S203T mutation has been reported to be associated with cases of severe and fatal pandemic influenza infections in Greece,⁴¹ although the authors did not clearly establish the proportion of mild cases carrying the same mutation. Most intriguingly, the mutation had been reported to co-occur with two other mutations (V106I and N248D) in the viral neuraminidase.⁴² This correlation in polymorphism across different segments of the influenza virus genome deserves to be further investigated, though it is not totally surprising given the extent of long-range correlations within HA shown in the present study.

Our structural analysis is limited by the lack of information on the binding kinetics of the HA protein to the sialylated glycans. Unfortunately, *in silico* prediction of binding kinetics, especially in cases involving protein flexibility and conformational change, requires computationally expensive molecular dynamics study, much like assessment of protein flexibility itself. On the other hand, experimental determination of binding kinetics requires dedicated equipment to provide quantitative characterization of the interaction, and so far relatively few direct studies had been published.⁴³ As a narrowed opening of the receptor binding pocket on the HA protein may hinder the approach of the sialylated glycan receptor, we anticipate that the dynamic aspect of the interaction can be an important determining factor in the effective transmission and entry of the pandemic influenza A/H1N1 (2009) virus. Another important structural aspect of the HA protein is its trimer formation, which is essential for the biological function of the triggering of membrane fusion between the virus and endosome to facilitate cell entry. The HA of influenza viruses have evolved extensively to contain antigenically diverse subtypes. For group 2 HAs, which include the H3 subtype but not the H1, H2 and H5 subtypes, a *tert*-Butylhydroquinone (TBHQ) binding site exists at the interface between HA monomers and the binding of TBHQ at this site can inhibit the conformational change of HA leading to reduced cell entry of the virus.^{44,45} The fusion peptide at the N-terminus of HA2 was found to be highly conserved with strong negative selection as expected, but other regions of HA2 contained numerous variable sites under positive selection. It may be possible that polymorphisms at these positions can promote trimer formation, enhance conformational change of the trimer and facilitate membrane fusion. Similar to TBHQ, understanding the

functional importance of the other HA2 regions may enable the development of novel antiviral compounds.

In summary, we investigated the flexibility of the receptor binding pocket and the correlation of mutations in influenza HA during viral evolution. Our *in silico* analysis results are also complemented by the findings of the *in vitro* assay on cellular receptor binding avidity by HA mutants. Taken together, these results show that a simple mutation at a key site can simultaneously impact different aspects of protein function and evolution, thus potentially contributing to the human adaptation of the pandemic influenza (H1N1) virus. It should be emphasized that our present work does not represent an exhaustive analysis of the functional significance of the D222 residue and associated mutations. Other aspects of host-virus interaction, including adaptive host immunity responses to viral proteins, have not been explored in the present study, and their potential interactions with structural flexibility may yet be discovered. For example, it may be possible for the flexibility of receptor binding pocket in the HA of this virus to facilitate interactions between viral antigen and host immune cells, leading to the rapid establishment of an enhanced immune response. In our opinion, the integration of structural and evolutionary analysis is a powerful and efficient means for the identification of viral determinants important for virulence and host adaptation. Further application of these methods to pandemic influenza and other pathogens can help unravel the mystery behind the appearance of pandemics caused by the novel viruses and hopefully enable us to develop effective preventive strategies.

Author contributions: K-YY and K-HS conceived and supervised the study; HT, WLW and MYY performed data collection and sequence analysis; WWLL provided the viral strains, K-HS performed homology modeling and molecular docking; PCYW, HT and RYTK participated in the interpretation of the results; and HT, RYTK and K-YY wrote the manuscript.

ACKNOWLEDGEMENTS

This work is partly supported by the Ted Sun Foundation, Providence Foundation Limited in memory of the late Dr Lui Hac Minh, the Hong Kong Special Administrative Region Research Grant Council, Research Fund for the Control of Infectious Diseases of the Food and Health Bureau of the Hong Kong Special Administrative Region, and The Shaw Foundation. We thank Dr Kelvin To for reading the manuscript.

REFERENCES

- 1 Franco-Paredes C, Hernandez-Ramos I, Del Rio C, Alexander KT, Tapia-Conyer R, Santos-Preciado JI. H1N1 influenza pandemics: comparing the events of 2009 in Mexico with those of 1976 and 1918–1919. *Arch Med Res* 2009;**40**:669–72
- 2 Taubenberger JK, Reid AH, Janczewski TA, Fanning TG. Integrating historical, clinical and molecular genetic data in order to explain the origin and virulence of the 1918 Spanish influenza virus. *Philos Trans R Soc Lond B Biol Sci* 2001;**356**:1829–39
- 3 Vincent AL, Swenson SL, Lager KM, Gauger PC, Loiacono C, Zhang Y. Characterization of an influenza A virus isolated from pigs during an outbreak of respiratory disease in swine and people during a county fair in the United States. *Vet Microbiol* 2009;**137**:51–9
- 4 Lessler J, Cummings DA, Fishman S, Vora A, Burke DS. Transmissibility of swine flu at Fort Dix, 1976. *J R Soc Interface* 2007;**4**:755–62
- 5 Gray GC, McCarthy T, Capuano AW, Setterquist SF, Olsen CW, Alavanja MC. Swine workers and swine influenza virus infections. *Emerg Infect Dis* 2007;**13**:1871–8
- 6 Guan Y, Poon LL, Cheung CY, Ellis TM, Lim W, Lipatov AS, Chan KH, Sturm-Ramirez KM, Cheung CL, Leung YH, Yuen KY, Webster RG, Peiris JS. H5N1 influenza: a protean pandemic threat. *Proc Natl Acad Sci USA* 2004;**101**:8156–61
- 7 Peiris JS, Yu WC, Leung CW, Cheung CY, Ng WF, Nicholls JM, Ng TK, Chan KH, Lai ST, Lim WL, Yuen KY, Guan Y. Re-emergence of fatal human influenza A subtype H5N1 disease. *Lancet* 2004;**363**:617–9
- 8 Yuen KY, Chan PK, Peiris M, Tsang DN, Que TL, Shorridge KF, Cheung PT, To WK, Ho ET, Sung R, Cheng AF. Clinical features and rapid viral diagnosis of human disease associated with avian influenza A H5N1 virus. *Lancet* 1998;**351**:467–71
- 9 Shinde V, Bridges CB, Uyeki TM, Shu B, Balish A, Xu X, Lindstrom S, Gubareva LV, Deyde V, Garten RJ, Harris M, Gerber S, Vagasky S, Smith F, Pascoe N, Martin K, Dufficy D, Ritger K, Conover C, Quinlisk P, Klimov A, Bresee JS, Finelli L. Triple-reassortant swine influenza A (H1) in humans in the United States, 2005–2009. *N Engl J Med* 2009;**360**:2616–25
- 10 Smith GJ, Bahl J, Vijaykrishna D, Zhang J, Poon LL, Chen H, Webster RG, Peiris JS, Guan Y. Dating the emergence of pandemic influenza viruses. *Proc Natl Acad Sci USA* 2009;**106**:11709–12
- 11 Smith GJ, Vijaykrishna D, Bahl J, Lycett SJ, Worobey M, Pybus OG, Ma SK, Cheung CL, Raghvani J, Bhatt S, Peiris JS, Guan Y, Rambaut A. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature* 2009;**459**:1122–5
- 12 Chen H, Wen X, To KK, Wang P, Tse H, Chan JF, Tsoi HW, Fung KS, Tse CW, Lee RA, Chan KH, Yuen KY. Quasispecies of the D225G substitution in the hemagglutinin of pandemic influenza A(H1N1) 2009 virus from patients with severe disease in Hong Kong, China. *J Infect Dis* 2010;**201**:1517–21
- 13 Mak GC, Au KW, Tai LS, Chuang KC, Cheng KC, Shiu TC, Lim W. Association of D222G substitution in haemagglutinin of 2009 pandemic influenza A (H1N1) with severe disease. *Euro Surveill* 2010;**15**:pii:19534
- 14 Miller RR, MacLean AR, Gunson RN, Carman WF. Occurrence of haemagglutinin mutation D222G in pandemic influenza A(H1N1) infected patients in the West of Scotland, United Kingdom, 2009–10. *Euro Surveill* 2010;**15**:pii:19546
- 15 Zheng B, Chan K-H, Zhang AJX, Zhou J, Chan CCS, Poon VKM, Zhang K, Leung VHC, Jin D-Y, Woo PCY, Chan JFW, To KKW, Chen H, Yuen K-Y. D225G mutation in hemagglutinin of pandemic influenza H1N1 (2009) virus enhances virulence in mice. *Exp Biol Med* 2010;**235**:981–8
- 16 Shen J, Ma J, Wang Q. Evolutionary trends of A(H1N1) influenza virus hemagglutinin since 1918. *PLoS One* 2009;**4**:e7789
- 17 Valli MB, Meschi S, Sella M, Zaccaro P, Ippolito G, Capobianchi MR, Menzo S. Evolutionary pattern of pandemic influenza (H1N1) 2009 virus in the late phases of the 2009 pandemic. *PLoS Curr* 2010;**2**:RRN1149
- 18 Hensley SE, Das SR, Bailey AL, Schmidt LM, Hickman HD, Jayaraman A, Viswanathan K, Raman R, Sasisekharan R, Bennis JR, Yewdell JW. Hemagglutinin receptor binding avidity drives influenza A virus antigenic drift. *Science* 2009;**326**:734–6
- 19 Bao Y, Bolotov P, Dernovoy D, Kiryutin B, Zaslavsky L, Tatusova T, Ostell J, Lipman D. The influenza virus resource at the National Center for Biotechnology Information. *J Virol* 2008;**82**:596–601
- 20 Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 2006;**22**:195–201
- 21 Bordoli L, Kiefer F, Arnold K, Benkert P, Battey J, Schwede T. Protein structure homology modeling using SWISS-MODEL workspace. *Nat Protoc* 2009;**4**:1–13
- 22 Xu R, Ekiert DC, Krause JC, Hai R, Crowe JE Jr, Wilson IA. Structural basis of preexisting immunity to the 2009 H1N1 pandemic influenza virus. *Science* 2010;**328**:357–60
- 23 Ha Y, Stevens DJ, Skehel JJ, Wiley DC. 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**:11181–6

- 24 Lin T, Wang G, Li A, Zhang Q, Wu C, Zhang R, Cai Q, Song W, Yuen KY. The hemagglutinin structure of an avian H1N1 influenza A virus. *Virology* 2009;**392**:73–81
- 25 Guex N, Peitsch MC. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 1997;**18**:2714–23
- 26 Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* 2010;**31**:455–61
- 27 Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* 2004;**25**:1605–12
- 28 Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. Clustal X version 2.0. *Bioinformatics* 2007;**23**:2947–8
- 29 Pond SL, Frost SD, Muse SV. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* 2005;**21**:676–9
- 30 Kosakovsky Pond SL, Poon AF, Leigh Brown AJ, Frost SD. A maximum likelihood method for detecting directional evolution in protein sequences and its application to influenza A virus. *Mol Biol Evol* 2008;**25**:1809–24
- 31 Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res* 2010;**38** Suppl:W529–33
- 32 Xia Z, Jin G, Zhu J, Zhou R. Using a mutual information-based site transition network to map the genetic evolution of influenza A/H3N2 virus. *Bioinformatics* 2009;**25**:2309–17
- 33 Faith JJ, Hayete B, Thaden JT, Mogno I, Wierzbowski J, Cottarel G, Kasif S, Collins JJ, Gardner TS. Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS Biol* 2007;**5**:e8
- 34 Shih AC, Hsiao TC, Ho MS, Li WH. Simultaneous amino acid substitutions at antigenic sites drive influenza A hemagglutinin evolution. *Proc Natl Acad Sci USA* 2007;**104**:6283–8
- 35 Soundararajan V, Tharakaraman K, Raman R, Raguram S, Shriver Z, Sasisekharan V, Sasisekharan R. Extrapolating from sequence—the 2009 H1N1 ‘swine’ influenza virus. *Nat Biotechnol* 2009;**27**:510–3
- 36 Lin X, Ma ZM, Yao X, He LF, Yuan ZH, Ding JP, Wen YM. Substitution of proline 306 in the reverse transcriptase domain of hepatitis B virus regulates replication. *J Gen Virol* 2005;**86**: 85–90
- 37 Lascu I, Chaffotte A, Limbourg-Bouchon B, Veron M. A Pro/Ser substitution in nucleoside diphosphate kinase of *Drosophila melanogaster* (mutation killer of prune) affects stability but not catalytic efficiency of the enzyme. *J Biol Chem* 1992;**267**:12775–81
- 38 Tumpey TM, Maines TR, Van Hoeven N, Glaser L, Solórzano A, Pappas C, Cox NJ, Swayne DE, Palese P, Katz JM, García-Sastre A. A two-amino acid change in the hemagglutinin of the 1918 influenza virus abolishes transmission. *Science* 2007;**315**:655–9
- 39 Nozawa M, Suzuki Y, Nei M. Reliabilities of identifying positive selection by the branch-site and the site-prediction methods. *Proc Natl Acad Sci USA* 2009;**106**:6700–5
- 40 Potdar VA, Chadha MS, Jadhav SM, Mullick J, Cherian SS, Mishra AC. Genetic characterization of the influenza A pandemic (H1N1) 2009 virus isolates from India. *PLoS One* 2010;**5**:e9693
- 41 Melidou A, Gioula G, Exindari M, Chatzidimitriou D, Diza E, Malisiovas N. Molecular and phylogenetic analysis of the haemagglutinin gene of pandemic influenza H1N1 2009 viruses associated with severe and fatal infections. *Virus Res* 2010;**151**:192–9
- 42 World Health Organization (WHO). Preliminary review of D222G amino acid substitution in the haemagglutinin of pandemic influenza A (H1N1) 2009 viruses. Geneva: WHO, 2009
- 43 Hidari KI, Shimada S, Suzuki Y, Suzuki T. Binding kinetics of influenza viruses to sialic acid-containing carbohydrates. *Glycoconj J* 2007;**24**:583–90
- 44 Yu KL, Torri AF, Luo G, Cianci C, Grant-Young K, Danetz S, Tiley L, Krystal M, Meanwell NA. Structure-activity relationships for a series of thiobenzamide influenza fusion inhibitors derived from 1,3,3-trimethyl-5-hydroxy-cyclohexylmethylamine. *Bioorg Med Chem Lett* 2002;**12**:3379–82
- 45 Russell RJ, Kerry PS, Stevens DJ, Steinhauer DA, Martin SR, Gamblin SJ, Skehel JJ. Structure of influenza hemagglutinin in complex with an inhibitor of membrane fusion. *Proc Natl Acad Sci USA* 2008;**105**:17736–41