Genes Genet. Syst. (2013) **88**, p. 225–232

Predictability of antigenic evolution for H3N2 human influenza A virus

Yoshiyuki Suzuki*

Graduate School of Natural Sciences, Nagoya City University, 1 Yamanohata, Mizuho-cho, Mizuho-ku, Nagoya-shi, Aichi-ken 467-8501, Japan

(Received 1 July 2013, accepted 14 November 2013)

Influenza A virus continues to pose a threat to public health. Since this virus can evolve escape mutants rapidly, it is desirable to predict the antigenic evolution for developing effective vaccines. Although empirical methods have been proposed and reported to predict the antigenic evolution more or less accurately, they did not provide much insight into the effects of unobserved mutations and the mechanisms of antigenic evolution. Here a theoretical method was introduced to predict the antigenic evolution of H3N2 human influenza A virus by evaluating *de novo* mutations through estimating the antigenic distance. The antigenic distance defined with the hemagglutination inhibition (HI) titer was estimated with antigenic models taking into account the volume, isoelectric point, relative solvent accessibility, and distances from receptor-binding sites (RBS) and *N*-linked glycosylation sites (NGS) for amino acids in hemagglutinin 1 (HA1). When the best model with the optimized parameter values was used to predict the antigenic evolution for the dominant strains, the prediction accuracy was relatively low. However, there appeared to be an overall tendency that the amino acid sites with larger potential net effect on antigenicity were more likely to evolve and the amino acid changes with larger potential effect were more likely to take place, suggesting that natural selection may operate to enhance the antigenic evolution of H3N2 human influenza A virus.

Key words: predictability, antigenic evolution, antigenic distance, hemagglutinin, influenza A virus

INTRODUCTION

Influenza A virus is an etiological agent of influenza (Shope, 1931; Smith et al., 1933), causing 3–5 million cases of severe illness and 250–500 thousand cases of deaths worldwide annually in humans (World Health Organization, 2009). In the infection of influenza A virus, humoral immunity is directed mainly against hemagglutinin (HA) and neuraminidase (NA), which are envelope proteins glycosylated at the *N*-linked glycosylation sites (NGS). HA and NA are classified into 17 (H1–H17) and 10 (N1–N10) subtypes according to the antigenic properties, respectively (World Health Organization, 1980; Tong et al., 2012). H3N2 has been a major subtype of influenza A virus circulating among humans since 1968.

HA exists ~10-times more abundantly than NA on the virion surface (Mitnaul et al., 1996). HA is a homotrimeric type I transmembrane glycoprotein consisting of 566 amino acid sites, which is cleaved into signal peptide (amino acid positions [–16]–[–1]), HA1 (positions 1–328), and HA2 (positions 330–550) (Skehel and Wiley, 2000). Signal peptide directs the co-translational transport of HA into the endoplasmic reticulum (ER). HA1 is the major target of neutralizing antibodies and contains the sialic-acid receptor binding sites (RBS) (positions 98, 136, 153, 183, 190, and 194), which are highly conserved across HA subtypes. HA2 is an anchor protein to the envelope and mediates fusion of the envelope and the endosomal membrane. HA1 and HA2 are covalently linked by a disulfide bond in a virion.

Vaccines consisting of inactivated or live-attenuated virions are available for prophylaxis against influenza A virus infection. However, this virus can evolve escape mutants rapidly because of a high mutation rate and a short generation time (Smith et al., 2004; Bedford et al., 2012). To counteract the evolution of escape mutants, it may be effective to develop vaccines targeting the amino acid sites of viral proteins under strong functional constraints, where mutations may not be tolerable (Suzuki, 2004, 2006; Han and Marasco, 2011). Alternatively, the

Edited by Yoko Satta

* Corresponding author. E-mail: yossuzuk@nsc.nagoya-cu.ac.jp

evolution of escape mutants may be predicted and vaccines may be prepared in advance.

Currently vaccines against influenza A virus are generated based on the latter strategy; vaccine seed strains are reformulated every year by the Global Influenza Surveillance and Response System (GISRS) of the World Health Organization (WHO) under the assumption that the wild dominant strain in the last epidemic may dominate in the next epidemic (Treanor, 2004). Since this assumption does not always hold, empirical methods have been proposed to predict the antigenic evolution of influenza A virus based on the ideas that one of the strains isolated in the last epidemic may dominate in the next epidemic (Bush et al., 1999; He and Deem, 2010; Ito et al., 2011) and the amino acid changes observed in the past may be repeated in the future (Xia et al., 2009).

It has been reported that these empirical methods can predict the antigenic evolution of influenza A virus more or less accurately (Bush et al., 1999; Xia et al., 2009; He and Deem, 2010; Ito et al., 2011). However, these methods may fail if the sampling of isolates was incomplete in the past or new antigenic variants were produced through *de novo* mutations after the last epidemic. In addition, the mechanisms of antigenic evolution were still unclear. The purpose of the present study was to introduce a theoretical method to predict the antigenic evolution of H3N2 human influenza A virus by evaluating *de novo* mutations through estimating the antigenic distance.

MATERIALS AND METHODS

Antigenic distance Difference in antigenicity between influenza A virus strains can be quantified using the hemagglutination inhibition (HI) titer. The HI titer t_{ij} is defined as the magnitude of the maximum dilution of anti-serum raised against strain i that inhibits the hemagglutination of strain j. It has been proposed that vaccination by strain i effectively prevents infection by strain j if $t_{ij}/t_{ii} \le 4$ (antigenic escape threshold). The antigenic distance between strains i and j (d_{ij}) is defined as the logarithm of the geometric mean for t_{ij}/t_{ii} and t_{ji}/t_{jj},

$$d_{ij} = \frac{1}{2}\log\left(\frac{t_{ij}}{t_{ii}}\frac{t_{ji}}{t_{jj}}\right)$$

(Lees et al., 2010).

Antigenic model The d_{ij} value may be inferred from the difference in the amino acid sequence of HA1 between strains i and j (Lee and Chen, 2004; Gupta et al., 2006; Lee et al., 2007; Liao et al., 2008; Huang et al., 2009; Lees et al., 2010). Mutations occurring at any amino acid site of HA1 containing a surface exposed area in the three-dimensional structure may alter the antigenicity (Lees et al., 2010). However, the extent of antigenic change may vary depending on the difference in the physical and chemical properties between the original and mutant amino acids (Hensley et al., 2009). The antigenic change may also be affected by the physical distance of the mutation site from RBS and NGS, because it is known that antibodies directed against the antigenic sites in close proximity to RBS neutralize virions efficiently (Whittle et al., 2011) and *N*-linked glycans attached to NGS shield antigenic sites from recognition by antibodies (Kobayashi and Suzuki, 2012b).

Based on these biological information, d_{ij} was estimated by model 1 as

$$\hat{d}_{1(ij)} = \sum_s \left[\left(\left| vol_{s(i)} - vol_{s(j)} \right| \cdot w_{vol} + \left| pI_{s(i)} - pI_{s(j)} \right| \cdot w_{pI} \right) \cdot f\left(d_{RBS_s}\right) \cdot g\left(\frac{d_{NGS_{s(i)}} + d_{NGS_{s(j)}}}{2} \right) \cdot \delta_s \right],$$

where s denotes the amino acid site, *vol* and *pI* the volume (Grantham, 1974) and isoelectric point (Nelson and Cox, 2008) of amino acids representing the physical and chemical properties, respectively, and w_{vol} and w_{pI} the weights for the volume and isoelectric point, respectively. The d_{RBS} and d_{NGS} indicate the Euclidean distances of the C_α atom of the amino acid from the closest C_α atom in RBS and NGS, respectively, in the three-dimensional structure of HA trimer for A/Hong Kong/19/1968 (Protein Data Bank [PDB] ID: 2HMG; corresponding to amino acid positions 1–328 and 330–504). RBS were highly conserved and therefore fixed at positions 98, 136, 153, 183, 190, and 194. In contrast, NGS may change during evolution and thus were determined for each sequence as asparagine [N] of the sequon, which consists of three consecutive amino acids of N, any amino acid except for proline [P], and serine [S] or threonine [T] ([N]–[X not P]–[S or T]). The functions $f(x)$, $g(y)$, and δ were defined as

$$f(x) = e^{-kx},\ k > 0,$$
$$g(y) = 1 - e^{-ly},\ l > 0, \text{ and}$$
$$\delta_s = \begin{cases} 1, & acc_s > 0 \\ 0, & acc_s = 0 \end{cases},$$

where *acc* denotes the relative solvent accessibility of the amino acid obtained from the computer program ASAVIEW (Ahmad et al., 2004) using 2HMG.

It should be noted that model 1 was an extension of the models proposed in the previous studies (Liao et al., 2008; Lees et al., 2010). In the present study, antigenic effects of amino acid changes were quantified taking into account the difference in the physical and chemical properties of the amino acids involved. These quantities were simply summed with weights as the first approximation, because their synergistic effects on antigenicity were unclear. In addition, the distances from RBS and NGS were used as weights on the antigenic effects for each amino acid site. The distance effects were assumed to change exponentially primarily because of the computational feasibility.

Table 1. Comparison of antigenic models

Model	Correlation coefficient	Sensitivity	Specificity	MCC	*vol*[a]	*pI*[b]	*rbs*[c]	*ngs*[d]	*acc*[e]
1	0.697	0.593	0.907	0.538	√	√	√	√	√
2	0.672	0.630	0.860	0.508		√	√	√	√
3	0.465	0.704	0.605	0.300	√		√	√	√
4	0.662	0.519	0.930	0.509	√	√		√	√
5	0.698	0.593	0.907	0.538	√	√	√		√
6	0.680	0.593	0.791	0.389	√	√	√	√	
7	0.673	0.630	0.860	0.508		√	√		√
8	0.479	0.704	0.605	0.300	√		√		√
9	0.684	0.519	0.884	0.440	√	√			√
10	0.683	0.593	0.791	0.389	√	√	√		

[a]Volume of amino acid (Grantham, 1974). [b]Isoelectric point of amino acid (Nelson and Cox, 2008). [c]Distance of amino acid from RBS. [d]Distance of amino acid from NGS. [e]Relative solvent accessibility of amino acid.

Since the formulation was indeed arbitrary, models 2–10 were also proposed by eliminating some of the variables from model 1 (Table 1), and their performances were evaluated with reference to those of the previous models (Liao et al., 2008; Lees et al., 2010) in a similar manner as Lees et al. (2010).

Model comparison Antigenic distances between strains of H3N2 human influenza A virus, whose HA1-coding nucleotide sequences were available in the Influenza Virus Resource at the National Center for Biotechnology Information (Bao et al., 2008) and did not contain minor gaps, ambiguous nucleotides, or premature termination codons, were retrieved from the literature (Supplementary Table S1). Data were eliminated when the homologous HI titer of the strains isolated in 2001 or later (2001~) was <1280, following Lees et al. (2010). As a result, 540 antigenic distances between 204 pairs of 80 strains were available for the analysis. The antigenic distances between the same pair of strains were averaged.

The data obtained above were divided into three groups according to the isolation years for the pairs of strains compared; both strains isolated in 2000 or earlier (~2000) (114 pairs of 34 strains), both strains isolated in 2001~ (70 pairs of 26 strains), and one strain isolated in ~2000 and the other strain in 2001~ (20 pairs of 19 strains). The first and second groups of data were used as the training and validation data sets in the model comparison, respectively (Supplementary Table S2). It should be noted that this scheme was the same as that introduced in Lees et al. (2010), and was adopted in the present study to facilitate the comparison of performances for new antigenic models with previous ones (Liao et al., 2008; Lees et al., 2010), although the results may depend on the scheme.

Phylogenetic relationship of HA1 for 60 strains involved in the training and validation data sets was examined by making a multiple alignment of 984 nucleotide sites, which did not contain any gaps, using MAFFT (version 6.901b) (Katoh et al., 2002) and constructing neighbor-joining (NJ) trees (Saitou and Nei, 1987) with the maximum composite likelihood (MCL) (Tamura et al., 2004) and p (Nei and Kumar, 2000) distances, which were known to produce reliable trees when a large number of closely related sequences was analyzed, using MEGA (version 5.20) (Tamura et al., 2011).

In the model comparison, the training data set was used for optimizing the parameter values and the validation data set for examining the performance in estimating the antigenic distance. For each model, the parameter values were optimized by minimizing the sum of least squares residual S in the estimation of antigenic distance over all available pairs of strains i and j in the training data set

$$S = \sum_{training\ data\ set} \left(d_{ij} - \hat{d}_{ij} \right)^2,$$

using the genetic algorithm (Tomita et al., 2000). Three sets of random real numbers were used as the initial parameter values in the optimization. Each model with optimized parameter values was used to estimate the antigenic distance between all available pairs of strains i and j in the validation data set. The correlation coefficient was computed between the observed and estimated values of antigenic distances. In addition, the sensitivity, specificity, and Matthews correlation coefficient (MCC) (Matthews, 1975) for identifying antigenic variants (hereafter indicating the pairs of strains whose antigenic differences exceeded the antigenic escape threshold) were computed as follows; if the numbers of true positives, true negatives, false positives, and false negatives are denoted as tp, tn, fp, and fn, respectively,

$$sensitivity = \frac{tp}{tp + fn},$$

$$specificity = \frac{tn}{tn + fp}, \text{ and}$$

$$MCC = \frac{tp \cdot tn - fp \cdot fn}{\sqrt{(tp + fp) \cdot (tp + fn) \cdot (tn + fp) \cdot (tn + fn)}}$$

(Matthews, 1975). The model providing the best performance judged from these indicators was selected as the best one.

Prediction of antigenic evolution Information on the dominant epidemic strains of H3N2 human influenza A virus was available from the literature (Supplementary Table S3). To explore the predictability of antigenic evolution by evaluating *de novo* mutations through estimating the antigenic distance, the best model with the optimized parameter values obtained above was used to predict the changes in the amino acid sequence of HA1 for the dominant strains in 2001~. It should be noted that the dominant strains in successive epidemics may not necessarily be the direct ancestor and descendant of each other. However, since the epidemic strains of H3N2 human influenza A virus are known to branch off the single trunk lineage in the chronological order (Fitch et al., 1991), the dominant strains in successive epidemics were expected to be closely related to each other. Therefore, in the present study, the amino acid changes requiring only one nucleotide mutation from the dominant strain in the last epidemic were considered as candidates to take place in the next epidemic at each amino acid site.

For each amino acid site of the dominant strains in 2001~, the antigenic distances estimated between the original amino acid and those accessible with single nucleotide mutations were summed with the weights proportional to the relative rates of nucleotide mutations. The pattern of nucleotide mutations was assumed to follow the two-parameter model (Kimura, 1980) with the transition/transversion rate ratio (κ) of 4 (Suzuki, 2011, 2013). The amino acid sites in HA1 were ranked according to the sum of potential antigenic changes; those ranked higher were predicted as more likely to evolve in the next epidemic than those ranked lower. In addition, the amino acid changes requiring single nucleotide mutations were ranked at each amino acid site according to the antigenic distance estimated between the original and mutant amino acids; those ranked higher were predicted as more likely to take place in the next epidemic than those ranked lower.

Amino acid changes were classified as stabilizing or destabilizing according to the thermodynamic stabilities of original and mutant proteins estimated by FOLDX (version 3.0 beta 5.1) (Guerois et al., 2002; Schymkowitz et al., 2005) using 2HMG.

RESULTS AND DISCUSSION

Comparison of antigenic models Antigenic models 1–10 were compared by optimizing the parameter values with the training data set and evaluating the performance in estimating the antigenic distance with the validation data set. In both NJ trees constructed using the MCL and p distances, the strains involved in the training and validation data sets were separated into distinct clusters though the bootstrap probability was relative low (39% and 36% in the MCL and p distance trees, respectively) (Supplementary Figs. S1 and S2), suggesting that these data sets were independent. Although three sets of initial parameter values were used in the optimization for each model, similar values were always obtained after optimization (Supplementary Table S4), indicating that the parameter values obtained were reliable. Therefore, the optimized parameter values associated with the smallest S were employed in the subsequent analysis.

In model 1, the antigenic distance was estimated using the volume, isoelectric point, accessibility, and distances from RBS and NGS for the amino acids in HA1. In the analysis of the validation data set, the correlation coefficient between the observed and estimated values of antigenic distances was 0.697 and the sensitivity, specificity, and MCC for identifying antigenic variants were 0.593, 0.907, and 0.538, respectively (Table 1). The effectiveness of each factor constituting model 1 in estimating the antigenic distance was examined by eliminating it from model 1 (models 2–6) (Table 1). The performance was decreased when the volume (model 2), isoelectric point (model 3), distance from RBS (model 4), and accessibility (model 6) were omitted. However, the performance did not alter to any large extent when the distance from NGS (model 5) was omitted from model 1, suggesting that this factor did not significantly contribute to the estimation of antigenic distance. Indeed, it has been reported that the effect of *N*-linked glycans to shield antigenic sites from recognition by antibodies was relatively weak (Kobayashi and Suzuki, 2012b). Reduction in the performance was again observed by eliminating each factor constituting model 5 (models 7–10) (Table 1). From these results, model 5 was judged as the best model in the present study.

The performance of model 5 appeared to be comparable to those of the best models in the previous studies; MCC for the best model was 0.54 in Liao et al. (2008) and 0.55 in Lees et al. (2010). Apparently, a smaller number of parameters (4) was necessary to be estimated in model 5 compared with the previous models (≥ 10). In the previous models, the sensitivity was relatively high (0.83 in Liao et al. [2008] and 0.97 in Lees et al. [2010]), whereas the specificity was relatively low (0.73 in Liao et al. [2008] and 0.57 in Lees et al. [2010]). In contrast, in model 5, the sensitivity was relatively low (0.593), whereas the

specificity was relatively high (0.907) (Table 1). These observations suggest that model 5 and previous models may complement in identifying antigenic variants.

Predictability of antigenic evolution To explore the predictability of antigenic evolution by evaluating *de novo* mutations through estimating the antigenic distance, model 5 with optimized parameter values obtained above was used to predict the amino acid changes in HA1 for the dominant strains of H3N2 human influenza A virus in 2001~. The amino acid sites in HA1 were ranked according to the sum of potential antigenic changes and the amino acid changes requiring single nucleotide mutations were ranked at each amino acid site according to the antigenic distance estimated between the original and mutant amino acids (Table 2).

During the evolution of H3N2 human influenza A virus in 2001~, the dominant strains appeared to have been replaced six times accompanied with 67 amino acid changes (Table 2). In the phylogenetic tree of HA1, the dominant strains in successive epidemics did not appear to be the direct ancestor and descendant of each other (Supplementary Figs. S1 and S2); indeed 24 (35.8%) of 67 amino acid changes were those in the opposite direction of the preceding changes (Table 2). However, the dominant strains in successive epidemics were closely related to each other; among 67 amino acid changes, 65 (97.0%) were those between the amino acids accessible with single nucleotide mutations (Table 2). These observations supported the adequacy of the prediction strategy in the present study, where the amino acid changes requiring only one nucleotide mutation were considered as candidates at each amino acid site.

According to the thermodynamic stabilities estimated for the original and mutant proteins, 39 of 67 amino acid changes were classified as stabilizing whereas 28 as destabilizing, which were not significantly different (p = 0.222 by the two-tailed binomial test). It should be noted that, in general, the rate of destabilizing mutations is much higher than that of stabilizing mutations and proteins are only marginally stable (DePristo et al., 2005; Tokuriki et al., 2007, 2009; Tokuriki and Tawfik, 2009). In addition, the average antigenic distances between the original and mutant amino acids derived from the stabilizing and destabilizing changes were estimated to be 0.185 and 0.172, respectively. These observations suggest that both stabilizing and destabilizing changes contribute to the antigenic evolution while maintaining the overall stability of HA1 during evolution of H3N2 human influenza A virus.

For each of six replacements of the dominant strains, 284 amino acid sites, which contained a surface exposed area, of HA1 were ranked according to the sum of potential antigenic changes. It was observed that 60 amino acid sites (7 were eliminated because of the lack of surface exposed area) where amino acid changes appeared to have occurred in the dominant strains in 2001~ were not usually ranked top (Table 2). These amino acid sites were ranked in the top 10 only for 3 cases (5%), suggesting a difficulty in predicting the exact amino acid sites to evolve in HA1 for the dominant strains. However, the amino acid sites where the changes were observed tended to be ranked higher in two (A/California/7/2004 → A/Wisconsin/67/2005 and A/Wisconsin/67/2005 → A/Brisbane/10/2007) of six replacements of the dominant strains ($0.01 < p < 0.05$ and $p < 0.01$, respectively, by the Mann-Whitney U test). In addition, when the rankings of amino acid sites were divided into two categories (ranked 1–142 as highly ranked and 143–284 as lowly ranked), 40 of 60 sites were categorized as highly ranked and 20 as lowly ranked (p = 0.0135 by the two-tailed binomial test) (Table 2). Therefore, there appeared to be an overall tendency that the amino acid sites with a higher potential of causing antigenic changes were more likely to evolve than those with a lower potential in HA1 for dominant strains of H3N2 human influenza A virus.

For each of the dominant strains of H3N2 human influenza A virus in 2001~, amino acid changes requiring only one nucleotide mutation were ranked at each amino acid site of HA1 according to the antigenic distance estimated between the original and mutant amino acids. Four-to-seven amino acids were ranked at 58 sites (7 sites were eliminated due to the lack of accessibility and 2 because amino acid changes required more than one nucleotide mutation), but the observed amino acid changes were ranked top only at 11 sites (19.0%), suggesting a difficulty in predicting the exact amino acid changes to take place in HA1 for the dominant strains (Table 2). However, when the rankings of amino acid changes were divided into two categories (upper half and lower half) at each amino acid site, 37 of 58 changes were categorized in the upper half and 14 in the lower half (7 were in the middle) (p = 0.00177 by the two-tailed binomial test) (Table 2). Therefore, there appeared to be an overall tendency that the amino acid changes causing a greater antigenic evolution were more likely to take place than those causing a smaller antigenic evolution in HA1 for dominant strains of H3N2 human influenza A virus.

Since influenza A virus can evolve escape mutants rapidly, it was desirable to predict the antigenic evolution for developing effective vaccines. Although empirical methods have been proposed and reported to predict the antigenic evolution more or less accurately (Bush et al., 1999; Xia et al., 2009; He and Deem, 2010; Ito et al., 2011), they did not provide much insight into the effects of unobserved mutations and the mechanisms of antigenic evolution. In the present study, a theoretical method was introduced to predict the antigenic evolution by evaluating the effects of *de novo* mutations through estimating the antigenic distance. It was observed that the amino acid

Table 2. Predictability of amino acid sites and changes causing antigenic evolution

Position	Rank[a]	Predicted[b]	Observed[c]	Rank[d]
A/Sydney/5/1997→A/Moscow/10/1999				
3	N.A.[e]	N.A.	I→L	N.A.
57	N.A.	N.A.	R→Q	N.A.
137	N.A.	N.A.	Y→S	N.A.
142	N.A.	N.A.	S→R	N.A.
145	N.A.	N.A.	K→N	N.A.
160	N.A.	N.A.	K→R	N.A.
194	N.A.	N.A.	I→L	N.A.
196	N.A.	N.A.	A→T	N.A.
233	N.A.	N.A.	H→Y	N.A.
A/Moscow/10/1999→A/Fujian/411/2002				
25	150	L→P	L→I	5/5
50	53	R→G[f]	R→G	1/5
75	172	H→R	H→Q	4/7
83	122	E→K	E→K	1/6
131	179	A→D	A→T	3/6
144	210	I→T	I→N	3/7
145	96	N→D	N→K[g]	2/7
155	12	H→R	H→T	N.A.
156	30	Q→R	Q→H	3/6
160	85	R→G	R→K	5/5
172	155	D→G	D→E	7/7
183	N.A.	N.A.	H→L	N.A.
186	43	S→R	S→G	2/6
192	141	T→I	T→I	1/5
196	151	T→I	T→A	2/5
202	N.A.	N.A.	V→I	N.A.
222	99	W→R	W→R	1/5
226	206	I→T	I→V	2/7
309	N.A.	N.A.	V→I	N.A.
A/Fujian/411/2002→A/California/7/2004				
138	136	A→D	A→S	5/6
145	31	K→E	K→N	2/6
159	118	Y→H	Y→F	6/6
183	N.A.	N.A.	L→H	N.A.
188	10	D→G	D→N	2/7
189	24	S→R	S→N	3/6
196	129	A→D	A→T	3/6
225	111	G→D	G→D	1/6
226	190	V→D	V→I	3/6
227	161	S→F	S→P	2/6
309	N.A.	N.A.	I→V	N.A.
A/California/7/2004→A/Wisconsin/67/2005				
8	59	N→D	N→D	1/7
122	44	N→D	N→D	1/7
186	39	G→D	G→V	5/6
188	25	N→D	N→D	1/7
193	23	S→R	S→F	N.A.
196	154	T→I	T→A	2/5
223	231	V→E	V→I	3/5
225	77	D→G	D→N	2/7
A/Wisconsin/67/2005→A/Brisbane/10/2007				
8	27	D→G	D→N	2/7
50	59	G→R	G→E	2/4
122	18	D→G	D→N	2/7
138	115	S→F	S→A	6/6
140	33	K→E	K→I	6/6
190	1	D→G	D→N	2/7
223	181	I→R	I→V	5/6
A/Brisbane/10/2007→A/Perth/16/2009				
62	110	E→K	E→K	1/6
144	92	N→D	N→K	2/7
158	26	K→E	K→N	2/6
173	156	K→E	K→Q	4/6
183	N.A.	N.A.	H→L	N.A.
186	126	V→D	V→G	5/6
189	17	N→D	N→K	2/7
190	3	N→D	N→D	1/7
214	184	I→T	I→S	4/7
A/Perth/16/2009→A/Victoria/361/2011				
45	165	S→R	S→N	3/6
48	188	T→R	T→I	3/5
62	110	K→E	K→E	1/6
144	29	K→E	K→N	2/6
156	23	H→R	H→Q	4/7
183	N.A.	N.A.	L→H	N.A.
186	37	G→D	G→V	5/6
198	167	A→E	A→S	5/6
212	227	T→I	T→A	2/5
214	73	S→R	S→I	4/6
219	161	S→F	S→Y	3/6
223	231	V→E	V→I	3/5
312	111	N→D	N→S	4/7

[a]Rank of amino acid sites in HA1. [b]Predicted amino acid change. [c]Observed amino acid change. [d]Rank of observed amino acid change/number of amino acid changes requiring single nucleotide mutations from the original amino acid. [e]Not available. [f]Gray shaded when the predicted amino acid change was the same as the observed one. [g]Gray shaded when the observed amino acid change was that in the opposite direction of the preceding change at the same site.

sites with larger potential net effect on antigenicity were more likely to evolve and the amino acid changes with larger potential effect were more likely to take place, suggesting that natural selection may operate to enhance the antigenic evolution of H3N2 human influenza A virus. Although the prediction accuracy for the amino acid sites and changes causing antigenic evolution was still relatively low, the method may be improved by modeling and integrating additional biological factors in the estimation of antigenic distance, e.g., interaction among amino acids within HA1 (Wiley et al., 1981; Kryazhimskiy et al., 2011) and between HA1 and antibodies (Wilson et al., 1981; Xia et al., 2012), and in the prediction of antigenic evolution, e.g., cross-immunity and -neutralization between different strains (Ekiert et al., 2012; Omori and Sasaki, 2013) and multifunctional and compensatory effects of amino acid changes on antigenicity and receptor binding affinity in HA1 (Soundararajan et al., 2011; Kobayashi and Suzuki, 2012a).

The author thanks Kimihito Ito for valuable suggestions in collecting information on the dominant epidemic strains of H3N2 human influenza A virus from the literature, and anonymous reviewers for valuable comments. This work was supported by the Grant-in-Aid for Research in Nagoya City University to Y. S.

REFERENCES

Ahmad, S., Gromiha, M. M., Fawareh, H., and Sarai, A. (2004) ASAView: database and tool for solvent accessibility representation in proteins. BMC Bioinformatics **5**, 51.

Bao, Y., Bolotov, P., Dernovoy, D., Kiryutin, B., Zaslavsky, L., Tatusova, T., Ostell, J., and Lipman, D. (2008) The Influenza Virus Resource at the National Center for Biotechnology Information. J. Virol. **82**, 596–601.

Bedford, T., Rambaut, A., and Pascual, M. (2012) Canalization of the evolutionary trajectory of the human influenza virus. BMC Biol. **10**, 38.

Bush, R. M., Bender, C. A., Subbarao, K., Cox, N. J., and Fitch, W. M. (1999) Predicting the evolution of human influenza A. Science **286**, 1921–1925.

DePristo, M. A., Weinreich, D. M., and Hartl, D. L. (2005) Missense meanderings in sequence space: a biophysical view of protein evolution. Nat. Rev. Genet. **6**, 678–687.

Ekiert, D. C., Kashyap, A. K., Steel, J., Rubrum, A., Bhabha, G., Khayat, R., Lee, J. H., Dillon, M. A., O'Neil, R. E., Faynboym, A. M., et al. (2012) Cross-neutralization of influenza A viruses mediated by a single antibody loop. Nature **489**, 526–532.

Fitch, W. M., Leiter, J. M., Li, X. Q., and Palese, P. (1991) Positive Darwinian evolution in human influenza A viruses. Proc. Natl. Acad. Sci. USA **88**, 4270–4274.

Grantham, R. (1974) Amino acid difference formula to help explain protein evolution. Science **185**, 862–864.

Guerois, R., Nielsen, J. E., and Serrano, L. (2002) Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. J. Mol. Biol. **320**, 369–387.

Gupta, V., Earl, D. J., and Deem, M. W. (2006) Quantifying influenza vaccine efficacy and antigenic distance. Vaccine **24**, 3881–3888.

Han, T., and Marasco, W. A. (2011) Structural basis of influenza virus neutralization. Ann. N. Y. Acad. Sci. **1217**, 178–190.

He, J., and Deem, M. W. (2010) Low-dimensional clustering detects incipient dominant influenza strain clusters. Protein Eng. Des. Sel. **23**, 935–946.

Hensley, S. E., Das, S. R., Bailey, A. L., Schmidt, L. M., Hickman, H. D., Jayaraman, A., Viswanathan, K., Raman, R., Sasisekharan, R., Bennink, J. R., and Yewdell, J. W. (2009) Hemagglutinin receptor binding avidity drives influenza A virus antigenic drift. Science **326**, 734–736.

Huang, J.-W., King, C.-C., and Yang, J.-M. (2009) Co-evolution positions and rules for antigenic variants of human influenza A/H3N2 viruses. BMC Bioinformatics **10**, S41.

Ito, K., Igarashi, M., Miyazaki, Y., Murakami, T., Iida, S., Kida, H., and Takada, A. (2011) Gnarled-trunk evolutionary model of influenza A virus hemagglutinin. PLoS ONE **6**, e25953.

Katoh, K., Misawa, K., Kuma, K.-i., and Miyata, T. (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. **30**, 3059–3066.

Kimura, M. (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16**, 111–120.

Kobayashi, Y., and Suzuki, Y. (2012a) Compensatory evolution of net-charge in influenza A virus hemagglutinin. PLoS ONE, **7**, e40422.

Kobayashi, Y., and Suzuki, Y. (2012b) Evidence for N-glycan shielding of antigenic sites during evolution of human influenza A virus hemagglutinin. J. Virol. **86**, 3445–3451.

Kryazhimskiy, S., Dushoff, J., Bazykin, G. A., and Plotkin, J. B. (2011) Prevalence of epistasis in the evolution of influenza A surface proteins. PLoS Genet. **7**, e1001301.

Lee, M.-S., and Chen, J. S.-E. (2004) Predicting antigenic variants of influenza A/H3N2 viruses. Emerg. Infect. Dis. **10**, 1385–1390.

Lee, M.-S., Chen, M.-C., Liao, Y.-C., and Hsiung, C. A. (2007) Identifying potential immunodominant positions and predicting antigenic variants of influenza A/H3N2 viruses. Vaccine **25**, 8133–8139.

Lees, W. D., Moss, D. S., and Shepherd, A. J. (2010) A computational analysis of the antigenic properties of haemagglutinin in influenza A H3N2. Bioinformatics **26**, 1403–1408.

Liao, Y.-C., Lee, M.-S., Ko, C.-Y., and Hsiung, C. A. (2008) Bioinformatics models for predicting antigenic variants of influenza A/H3N2 virus. Bioinformatics **24**, 505–512.

Matthews, B. W. (1975) Comparison of the predicted and observed secondary structure of T4 phage lysozyme. Biochim. Biophys. Acta **405**, 442–451.

Mitnaul, L. J., Castrucci, M. R., Murti, K. G., and Kawaoka, Y. (1996) The cytoplasmic tail of influenza A virus neuraminidase (NA) affects NA incorporation into virions, virion morphology, and virulence in mice but is not essential for virus replication. J. Virol. **70**, 873–879.

Nei, M., and Kumar, S. (2000) Molecular Evolution and Phylogenetics. Oxford University Press, Oxford, New York.

Nelson, D. L., and Cox, M. M. (2008) Lehninger Principles of Biochemistry. 5th edition. W. H. Freeman and Company, New York.

Omori, R., and Sasaki, A. (2013) Timing of the emergence of new successful viral strains in seasonal influenza. J. Theor. Biol. **329**, 32–38.

Saitou, N., and Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol.

Biol. Evol. **4**, 406–425.
Schymkowitz, J. W. H., Rousseau, F., Martins, I. C., Ferkinghoff-Borg, J., Stricher, F., and Serrano, L. (2005) Prediction of water and metal binding sites and their affinities by using the Fold-X force field. Proc. Natl. Acad. Sci. USA **102**, 10147–10152.
Shope, R. E. (1931) Swine influenza. III. Filtration experiments and etiology. J. Exp. Med. **54**, 373–385.
Skehel, J. J., and Wiley, D. C. (2000) Receptor binding and membrane fusion in virus entry: the influenza hemagglutinin. Annu. Rev. Biochem. **69**, 531–569.
Smith, D. J., Lapedes, A. S., de Jong, J. C., Bestebroer, T. M., Rimmelzwaan, G. F., Osterhaus, A. D. M. E., and Fouchier, R. A. M. (2004) Mapping the antigenic and genetic evolution of influenza virus. Science **305**, 371–376.
Smith, W., Andrewes, C. H., and Laidlaw, P. P. (1933) A virus obtained from influenza patients. Lancet **222**, 66–68.
Soundararajan, V., Zheng, S., Patel, N., Warnock, K., Raman, R., Wilson, I. A., Raguram, S., Sasisekharan, V., and Sasisekharan, R. (2011) Networks link antigenic and receptor-binding sites of influenza hemagglutinin: mechanistic insight into fitter strain propagation. Sci. Rep. **1**, 200.
Suzuki, Y. (2004) Negative selection on neutralization epitopes of poliovirus surface proteins: implications for prediction of candidate epitopes for immunization. Gene **328**, 127–133.
Suzuki, Y. (2006) Natural selection on the influenza virus genome. Mol. Biol. Evol. **23**, 1902–1911.
Suzuki, Y. (2011) Positive selection for gains of N-linked glycosylation sites in hemagglutinin during evolution of H3N2 human influenza A virus. Genes Genet. Syst. **86**, 287–294.
Suzuki, Y. (2013) Detection of positive selection eliminating effects of structural constraints in hemagglutinin of H3N2 human influenza A virus. Infect. Genet. Evol. **16**, 93–98.
Tamura, K., Nei, M., and Kumar, S. (2004) Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc. Natl. Acad. Sci. USA **101**, 11030–11035.
Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. **28**, 2731–2739.
Tokuriki, N., and Tawfik, D. S. (2009) Stability effects of mutations and protein evolvability. Curr. Opin. Struct. Biol. **19**, 596–604.
Tokuriki, N., Stricher, F., Schymkowitz, J., Serrano, L., and Tawfik, D. S. (2007) The stability effects of protein mutations appear to be universally distributed. J. Mol. Biol. **369**, 1318–1332.
Tokuriki, N., Oldfield, C. J., Uversky, V. N., Berezovsky, I. N., and Tawfik, D. S. (2009) Do viral proteins possess unique biophysical features? Trends Biochem. Sci. **34**, 53–59.
Tomita, M., Hashimoto, K., Takahashi, K., Matsuzaki, Y., Matsushima, R., Saito, K., Yugi, K., Miyoshi, F., Nakano, H., Tanida, S., et al. (2000) The E-CELL project: towards integrative simulation of cellular processes. New Gener. Comput. **18**, 1–12.
Tong, S., Li, Y., Rivailler, P., Conrardy, C., Castillo D. A. A., Chen, L.-M., Recuenco S., Ellison, J. A., Davis, C. T., York, I. A., et al. (2012) A distinct lineage of influenza A virus from bats. Proc. Natl. Acad. Sci. USA **109**, 4269–4274.
Treanor, J. (2004) Weathering the influenza vaccine crisis. N. Engl. J. Med. **351**, 2037–2040.
Whittle, J. R. R., Zhang, R., Khurana, S., King, L. R., Manischewitz, J., Golding, H., Dormitzer, P. R., Haynes, B. F., Walter, E. B., Moody, M. A., et al. (2011) Broadly neutralizing human antibody that recognizes the receptor-binding pocket of influenza virus hemagglutinin. Proc. Natl. Acad. Sci. USA **108**, 14216–14221.
Wiley, D. C., Wilson, I. A., and Skehel, J. J. (1981) Structural identification of the antibody-binding sites of Hong Kong influenza haemagglutinin and their involvement in antigenic variation. Nature **289**, 373–378.
Wilson, I. A., Skehel, J. J., and Wiley, D. C. (1981) Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. Nature **289**, 366–373.
World Health Organization (1980) A revision of the system of nomenclature for influenza viruses: a WHO memorandum. Bull. W. H. O. **58**, 585–591.
World Health Organization (2009) Influenza (Seasonal). http://www.who.int/mediacentre/factsheets/fs211/en/index.html.
Xia, Z., Jin, G., Zhu, J., and Zhou, R. (2009) Using a mutual information-based site transition network to map the genetic evolution of influenza A/H3N2 virus. Bioinformatics **25**, 2309–2317.
Xia, Z., Huynh, T., Kang, S.-g., and Zhou, R. (2012) Free-energy simulations reveal that both hydrophobic and polar interactions are important for influenza hemagglutinin antibody binding. Biophys. J. **102**, 1453–1461.