

# Differential neutralization efficiency of hemagglutinin epitopes, antibody interference, and the design of influenza vaccines

Wilfred Ndifon<sup>a,1</sup>, Ned S. Wingreen<sup>b,1</sup>, and Simon A. Levin<sup>a,1</sup>

Departments of <sup>a</sup>Ecology and Evolutionary Biology, and <sup>b</sup>Molecular Biology, Princeton University, Princeton, NJ 08544

Contributed by Simon A. Levin, April 1, 2009 (sent for review November 11, 2008)

**I**t is generally assumed that amino acid mutations in the surface protein, hemagglutinin (HA), of influenza viruses allow these viruses to circumvent neutralization by antibodies induced during infection. However, empirical data on circulating influenza viruses show that certain amino acid changes to HA actually increase the efficiency of neutralization of the mutated virus by antibodies raised against the parent virus. Here, we suggest that this surprising increase in neutralization efficiency after HA mutation could reflect steric interference between antibodies. Specifically, if there is a steric competition for binding to HA by antibodies with different neutralization efficiencies, then a mutation that reduces the binding of antibodies with low neutralization efficiencies could increase overall viral neutralization. We use a mathematical model of virus–antibody interaction to elucidate the conditions under which amino acid mutations to HA could lead to an increase in viral neutralization. Using insights gained from the model, together with genetic and structural data, we predict that amino acid mutations to epitopes C and E of the HA of influenza A/H3N2 viruses could lead on average to an increase in the neutralization of the mutated viruses. We present data supporting this prediction and discuss the implications for the design of more effective vaccines against influenza viruses and other pathogens.

antigenic distance | epidemic | epitope vaccine | evolution

Influenza viruses infect ≈5–15% of the world population each year (1). Infection leads to the production of antibodies that preferentially recognize the influenza viral hemagglutinin (HA) protein (2, 3). Most of these antibodies neutralize influenza viruses and, hence, limit infection by binding to specific regions of HA called (functional) epitopes, which are located within presumed topologically distinct sites called antigenic sites (denoted simply by epitopes) (Fig. 1A). Amino acid changes to HA have complex effects on viral neutralization by antibodies (4–6). For example, the antigenic similarity (a measure of the degree to which antibodies raised against one virus neutralize another virus) between certain pairs of influenza viruses actually increases after the introduction of additional amino acid differences between the HAs of the two viruses [see, e.g., ref. 6]. This observation could be explained by positing that the additional amino acid changes compensate for preexisting amino acid differences between the viruses. However, because some of the changes in question occur in entirely distinct HA epitopes from the preexisting differences, it is possible that there is another mechanism at play. Here, we propose such a mechanism based on steric interference between antibodies (Fig. 1B), and we discuss the implications for improving the effectiveness of influenza vaccines.

The ability of an antibody to neutralize a virus depends on the strength of the virus–antibody bond (i.e., the affinity of the antibody for the virus) and on the neutralization efficiency of the viral epitope bound by the antibody (4, 9–11).<sup>\*</sup> In the case of influenza virus, there is a simple physical explanation for such epitope dependence of viral neutralization. A large body of experimental work (11–15) suggests that occlusion of the receptor-binding site by antibodies bound to HA constitutes the dominant mechanism of influenza viral neutralization. Antibodies that bind to HA epitopes located at a distance from the receptor-binding site may therefore fail to occlude the site efficiently, thereby leading to a low degree of viral neutralization (4, 11, 12). Moreover, it has been shown that antibodies that bind to a given HA epitope can prevent further binding of antibodies to other epitopes of the same HA protein (Fig. 1B) and even to epitopes found on adjacent HA proteins (13, 14). [Note that some antibodies may neutralize influenza virus by inhibiting fusion of the viral envelope with the cellular membrane (11, 16).]

The above observations suggest that antibodies that bind to low-neutralization efficiency epitopes of HA might interfere with the binding of antibodies to high-neutralization efficiency epitopes, thereby impeding the neutralization of influenza viruses (Fig. 2). This raises the intriguing possibility that the influenza virus may have evolved to decrease the probability of its own neutralization by allowing itself to be bound preferentially by antibodies that recognize low-neutralization efficiency epitopes of HA. Indeed, steric interference by antibodies that bind low-neutralization efficiency epitopes could greatly reduce the extent of mutation required for a virus to evade neutralization by host antibodies. For example, if all epitopes had high neutralization efficiencies, then, for a virus to evade neutralization, viral mutations would have to reduce the affinities of antibodies for all epitopes to such a level that only a small fraction of epitopes would be bound by antibodies (Fig. 3A). By comparison, if some epitopes had low neutralization efficiencies, then, to evade neutralization, viral mutations would only need to decrease the affinities of antibodies for high-neutralization efficiency epitopes enough for the bound antibodies to consist predominantly of those for low-neutralization efficiency epitopes (Fig. 3B). A corollary to this observation is that a decrease in the affinities of antibodies for epitopes with low neutralization efficiencies could lead to an increase in viral neutralization (Fig. 2). This in turn suggests a possible approach to designing “low-interference” influenza vaccines that could greatly reduce the impact of antibody interference (Fig. 3C–E). Rather different approaches to vaccine design are suggested by the idea of “deceptive imprinting” (e.g., Ref. 17), which posits that vaccine effectiveness could be improved by limiting the antibody-induction potential of immunodominant epitopes. However, these approaches have not yet proven fruitful (18).

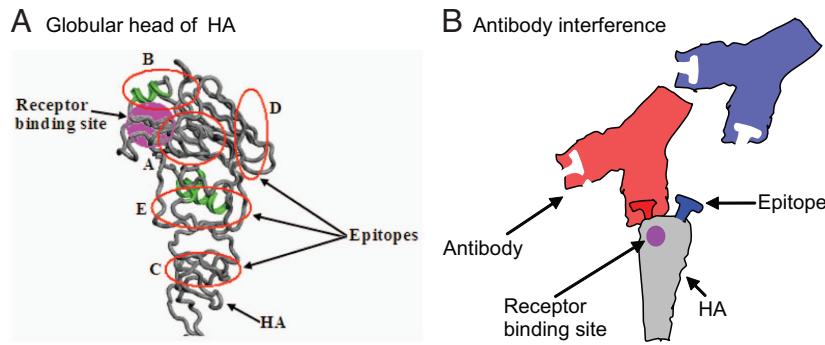
Author contributions: W.N., N.S.W., and S.A.L. designed research; W.N. performed research; W.N. and N.S.W. analyzed data; and W.N., N.S.W., and S.A.L. wrote the paper.

The authors declare no conflict of interest.

<sup>1</sup>To whom correspondence may be addressed. E-mail: [wndifon@princeton.edu](mailto:wndifon@princeton.edu), [wingreen@princeton.edu](mailto:wingreen@princeton.edu), or [slevin@princeton.edu](mailto:slevin@princeton.edu).

\*Neutralization efficiency refers to an intrinsic (e.g., independent of antibody affinity) capacity of an epitope to support viral neutralization by bound antibodies. Note that in cases when antibodies bind to influenza viruses multivalently, viral neutralization depends on both antibody avidity and the neutralization efficiency of the bound epitopes.

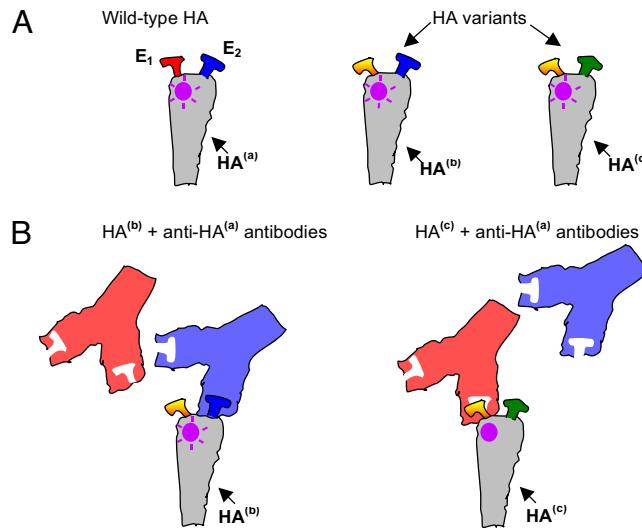
This article contains supporting information online at [www.pnas.org/cgi/content/full/0903427106/DCSupplemental](http://www.pnas.org/cgi/content/full/0903427106/DCSupplemental).



**Fig. 1.** HA and antibody interference. (A) Globular head of a monomer of HA (Protein Data Bank ID code 1hgf), showing five antibody-binding sites (or epitopes) and the receptor-binding site. The figure was drawn by using the PyMOL molecular graphics system. (B) Interference between antibodies that bind to two different HA epitopes. Illustrated are cross-sections of an IgG molecule and an HA trimer. The molecules were drawn approximately to scale. IgG is a Y-shaped molecule that can be separated into three fragments (two Fab fragments and one Fc fragment) of approximately the same size. A Fab fragment has approximate dimensions  $80 \times 50 \times 40 \text{ \AA}$  (7). In comparison, an HA trimer has a length of  $\approx 135 \text{ \AA}$  and a diameter of  $\approx 55 \text{ \AA}$  (8), approximately equal to the width of the  $40 \times 50\text{-\AA}$  distal surface of a Fab fragment, which contains the Fab-binding pocket.

## Results and Discussion

**Simple Model of Interference Between Antibodies, and Implications for Viral Neutralization.** Here, we consider an *in vitro* assay of the neutralization of an influenza virus by antibody-containing antiserum. We assume that the neutralization reaction reaches the equilibrium state and is therefore amenable to analysis by means of standard mass-action models of antibody–antigen binding (e.g., ref. 19). Let  $A_i, i = 1, \dots, m$ , denote antibody that binds to the  $i$ th HA epitope of the virus under consideration. Using the induced-fit model of antibody–antigen binding (20, 21) and assuming that the binding of antibody to an HA epitope interferes with the binding of antibodies to other epitopes of the



**Fig. 2.** Effects of antibody interference on viral neutralization. (A) Wild-type HA [denoted HA<sup>(a)</sup>] containing two epitopes, a high-neutralization efficiency epitope (denoted E<sub>1</sub>) located close to the receptor-binding site, and a low-neutralization efficiency epitope (denoted E<sub>2</sub>) located farther from the receptor-binding site. A variant of HA<sup>(a)</sup> [denoted HA<sup>(b)</sup>] contains mutations to E<sub>1</sub>, whereas another variant [denoted HA<sup>(c)</sup>] also contains mutations to E<sub>2</sub>. Antibodies raised against HA<sup>(a)</sup> bind to E<sub>1</sub> of HA<sup>(c)</sup> much more readily than they bind to E<sub>1</sub> of HA<sup>(b)</sup> because the additional mutations to E<sub>2</sub> remove antibody interference. Viruses carrying the more mutated HA<sup>(c)</sup> are therefore neutralized more efficiently than viruses carrying the less mutated HA<sup>(b)</sup> by wild-type antibodies. Lines emanating from the receptor-binding site indicate that the site is not completely occluded.

same HA protein (13, 14), it can be readily shown that the probability  $P_i$  that the  $i$ th epitope is bound at equilibrium is given by:

$$P_i = \frac{[A_i]/K_i}{1 + \sum_{j=1}^m [A_j]/K_j}, \quad i = 1, \dots, m, \quad [1]$$

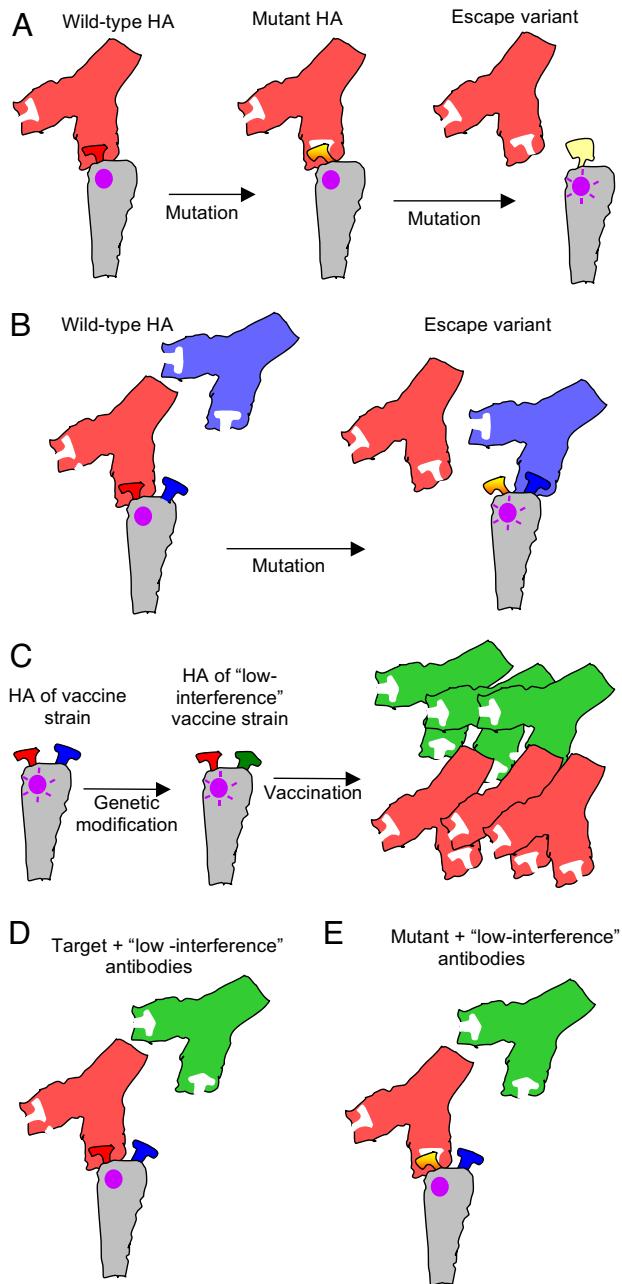
where  $[A_i]$  denotes the free concentration of antibodies that bind to the  $i$ th epitope and  $K_i$  is the equilibrium dissociation constant (or the reciprocal of the affinity) of those antibodies.

Let  $A_0$  denote the total concentration of antibodies found in the antiserum and  $f_i$  the fraction of those antibodies specific to the  $i$ th epitope. Because the concentration of antibodies produced during influenza virus infection tends to be substantially greater than the estimated concentration of viral HA (*SI Methods*), we can accurately approximate the free concentration of antibodies as  $[A_i] = f_i A_0$ , such that Eq. 1 becomes:

$$P_i = \frac{f_i A_0 / K_i}{1 + A_0 \sum_{j=1}^m f_j / K_j}, \quad i = 1, \dots, m. \quad [2]$$

Previous experimental results suggest that the fraction of neutralized influenza viruses is described well by a sigmoidal Hill function of the average number of bound antibodies (proportional to  $P_i$  for a monoclonal antibody) and that the efficiency of viral neutralization depends on the epitopes bound by the antibodies (*SI Methods* and Fig. S1). In the following, we show that antibody interference combined with such epitope-dependent neutralization efficiency could explain certain puzzling aspects of the antigenic effects of amino acid changes to the HA of influenza viruses.

The standard experimental method of quantifying the antigenic effects of amino acid changes to the HA of influenza viruses is the hemagglutination-inhibition (HI) assay (22). The HI assay measures the maximum amount  $h$  (also called the HI titer) by which antiserum raised against one virus (the “homologous” virus) can be diluted without losing its capacity to neutralize a second virus (the “heterologous” virus). Neutralization is indicated by the inability of the heterologous virus to agglutinate a standardized amount of red blood cells. The normalized HI titer is given by the ratio of the HI titer obtained with the heterologous virus to the HI titer obtained with the



**Fig. 3.** Deleterious effects of antibody interference on the host and proposed strategy for influenza vaccine design. (A) If viral HA contained only epitopes with high-neutralization efficiencies, then only viruses with large epitopic changes could escape from antibodies. (B) Antibody interference from low-neutralization efficiency epitopes enables viruses with small epitopic changes also to escape from antibodies. (C) Proposed low-interference vaccine strain is genetically modified from viral target at low-neutralization efficiency epitopes of HA. Vaccine-induced antibodies only recognize high-neutralization efficiency epitopes of target. (D) Antibodies induced by low-interference vaccine strain have low affinity for low-neutralization efficiency epitopes of the target and therefore do not interfere with antibodies to high-neutralization efficiency epitopes, implying better neutralization. (E) Without antibody interference the target virus cannot escape from vaccine-induced antibodies via small epitopic changes. Lines emanating from the receptor-binding site indicate that the site is not completely occluded.

homologous virus. We define the antigenic similarity between two influenza viruses as the geometric mean of the normalized HI titer of each virus relative to antiserum raised against the other virus; we call this the Archetti–Horsfall measure (AHM)

of antigenic similarity (6, 23, 24). Empirical data show that the heterologous HI titer is occasionally greater than the homologous HI titer by more than a factor of 2 (the dilution factor used in the HI assay) and that the AHM of the antigenic similarity between two influenza viruses sometimes increases after additional amino acid changes to the HA of one of those viruses (ref. 6 and Appendix A of ref. 24).

We propose a mechanistic explanation for the above counterintuitive observations: Consider a virus containing an HA [denoted  $\text{HA}^{(a)}$ ] against which antibodies have been raised. Let  $\text{HA}^{(a)}$  contain two epitopes, one (denoted  $E_1$ ) having high neutralization efficiency and the other (denoted  $E_2$ ) having much lower neutralization efficiency. Also, let the antibodies that bind to  $E_1$  and  $E_2$  be denoted by  $A_1$  and  $A_2$ , respectively, and let the affinities of these antibodies for their respective epitopes be given by  $1/K_1$  and  $1/K_2$ . If  $[A_2]/K_2$  is greater than  $[A_1]/K_1$ , then at equilibrium  $P_2 > P_1$  (see Eq. 2), and, depending on the neutralization efficiency of  $E_2$ , a substantial fraction of the virus could escape neutralization. Now, if a variant of  $\text{HA}^{(a)}$  [denoted  $\text{HA}^{(b)}$ ] differs from  $\text{HA}^{(a)}$  by an antigenically important mutation to  $E_1$ , then, all else being equal, the affinity of  $A_1$  antibodies for  $E_1$  of  $\text{HA}^{(b)}$  would be lower than the affinity of  $A_1$  antibodies for  $E_1$  of  $\text{HA}^{(a)}$ , and the fraction of antibodies bound to  $E_1$  of  $\text{HA}^{(b)}$  at equilibrium would decrease, resulting in lower neutralization of virus carrying  $\text{HA}^{(b)}$  compared with the parent virus (Fig. 2).

If in addition to the difference at  $E_1$ , another variant of  $\text{HA}^{(a)}$  [denoted  $\text{HA}^{(c)}$ ] also differs from  $\text{HA}^{(a)}$  at  $E_2$ , then, all else being equal, the affinity of  $A_2$  antibodies for  $E_2$  of  $\text{HA}^{(c)}$  would decrease, meaning that the fraction of  $A_1$  antibodies bound at equilibrium would increase, resulting in greater neutralization of virus carrying  $\text{HA}^{(c)}$  compared with virus carrying  $\text{HA}^{(b)}$  (Fig. 2). Consequently, the AHM measure of the antigenic similarity between the parent virus [carrying  $\text{HA}^{(a)}$ ] and a virus carrying the more mutated  $\text{HA}^{(c)}$  would be greater than between the parent virus and a virus carrying the less mutated  $\text{HA}^{(b)}$ . Following a similar line of reasoning, the antigenically important difference at  $E_2$  would naturally cause the HI titer of an  $\text{HA}^{(c)}$  virus relative to antiserum raised against an  $\text{HA}^{(b)}$  virus to be greater than the corresponding homologous HI titer, as seen in empirical data. Therefore, antibody interference could provide a mechanistic explanation for some of the puzzling empirical observations regarding antigenic similarity.

**Practical Applications of the Theory of Antibody Interference.** *Antigenic effects of amino acid changes to A/H3N2 viruses.* Here, we combine our theoretical model for antibody interference with genetic and structural data to make experimentally testable predictions for the antigenic effects of amino acid changes to the HA epitopes of influenza A/H3N2 viruses. We chose to consider A/H3N2 viruses because of the wealth of available genetic and antigenic data on the evolution of these viruses and because of their epidemiological importance. The HA of A/H3N2 viruses contains five epitopes, denoted A–E (the amino acid positions found in each of these epitopes are given in Table S1). Several lines of evidence suggest that epitopes A and B contain functional epitopes with high neutralization efficiencies. First, the physical locations of these epitopes on the HA protein are much closer to the receptor-binding site compared with the locations of the other three epitopes (Fig. 1A and Table 1). This fact, in light of the aforementioned occlusion mechanism of antibody-mediated neutralization (11, 15), suggests that functional epitopes found within epitopes A and B could indeed have high neutralization efficiencies.

Moreover, the vast majority of HA codons predicted to be under positive selection occurs in epitopes A, B, and, to a lesser extent, D (25). One expects the selective pressure to acquire antigenically consequential amino acid changes to be greater for

**Table 1.** Spatial distance between epitopes of HA and the receptor-binding site

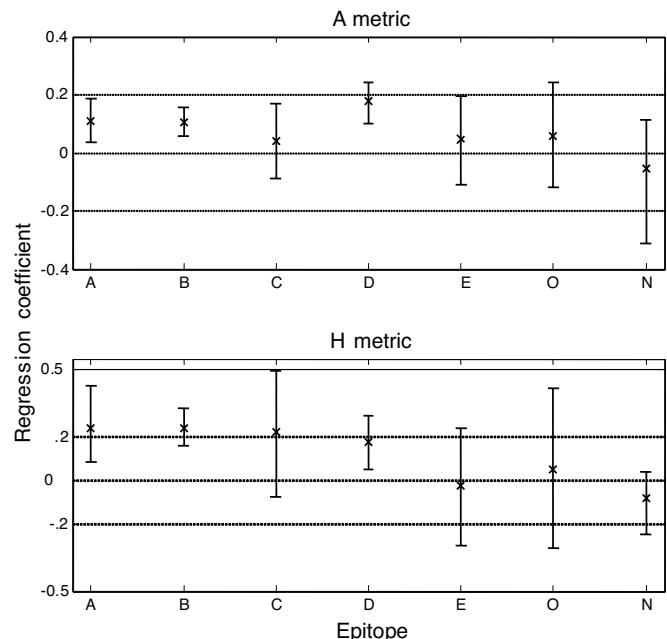
Epitope	Spatial distance, Å
A	10.7
B	13.0
C	44.5
D	17.7
E	25.5
O	35.8

Shown are the distances between the centroid of atomic positions of amino acids found in each epitope and the centroid of atomic positions of amino acids found in the receptor-binding site. The distances are based on chain A of the HA 3D structure (PDB ID: 1hgf) of A/Aichi/2/68 (A/H3N2). A, B, C, D, E, and O are defined in the legend of Fig. 4.

epitopes with high neutralization efficiencies compared with epitopes with lower neutralization efficiencies because changes to the high-neutralization efficiency epitopes have a greater probability of effecting viral escape from neutralization. In contrast to epitopes A, B, and D, epitopes C and E, which are the furthest of the five epitopes from the receptor-binding site (Table 1), accrue amino acid changes at a comparatively lower rate (25), suggesting that these epitopes could have lower neutralization efficiencies. (Note that the above-mentioned pattern of positive selection could still occur if all epitopes had the same neutralization efficiencies, provided epitopes A, B, and D were much more immunogenic than epitopes C and E.)

It follows from the above observations and from a consideration of the effects of antibody interference that amino acid changes to epitopes A, B, and D could lead on average to a decrease in neutralization and, hence, an increase in the antigenic distances between influenza A/H3N2 viruses, whereas amino acid changes to epitopes A and C could lead on average to an increase in neutralization (*cf.* Fig. 2, with epitope A corresponding to E<sub>1</sub> and epitope C corresponding to E<sub>2</sub>). To test these predictions, we estimated the contribution of amino acid changes to individual HA epitopes to antigenic change (*Materials and Methods*). We quantified antigenic change by the reciprocal of the AHM measure of antigenic similarity, whereas amino acid change was quantified by the metric of Atchley et al. (26), denoted the A metric, and by the Hamming metric, denoted the H metric (*Materials and Methods*). The results (Fig. 4 and **Table S2**) suggest that the hypothesis that amino acid changes to (dominant) epitopes A, B, and D could lead on average to an increase in antigenic distance, whereas amino acid changes to (subdominant) epitopes C and E could lead on average to a decrease in antigenic distance cannot be rejected at the 95% confidence level. Thus, changes to epitopes A, B, and D could be highly favored by natural selection, whereas changes to epitopes C and E could be disadvantageous to influenza viruses.

In addition, the results suggest that nonepitopic sites and sites that do not belong to any of the five epitopes of HA but are nevertheless experimentally confirmed targets of monoclonal antibodies (**Table S1**) do not make a significantly positive contribution to antigenic change. The latter sites are also located further from the receptor-binding site than epitope E (Table 1), suggesting that they could have lower neutralization efficiencies than the dominant epitopes. Similar results were obtained when the computed genetic distances were normalized to account for differences in the number of amino acid positions found in each epitope and when genetic distances were quantified by using other popular metrics (**Table S2**). Note that when we summed the number of amino acid changes to all low-neutralization efficiency epitopes (A, B, and D) and the number of amino acid changes to all high-neutralization efficiency epitopes (C and E), we found a statistically significant ( $P < 0.05$ ) antagonistic

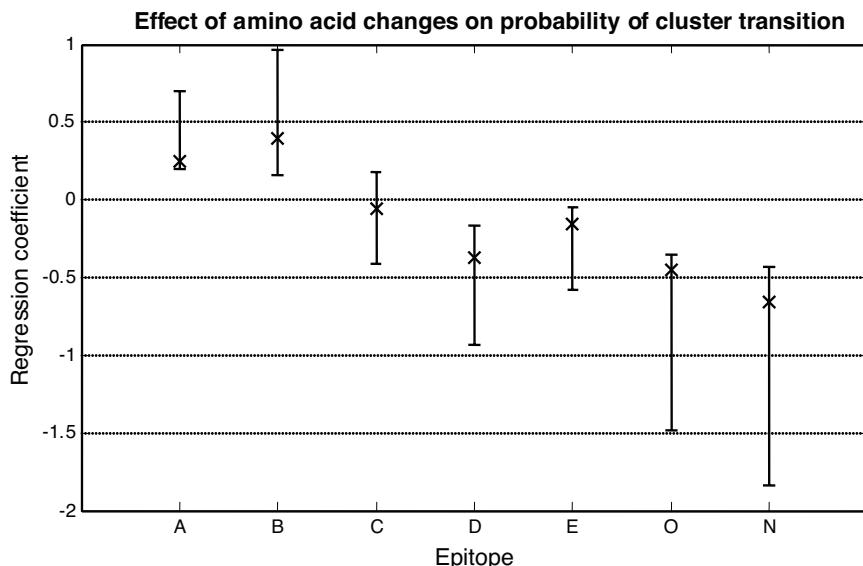


**Fig. 4.** Antigenic effects of amino acid changes to individual epitopes of influenza virus HA. Antigenic effects (regression coefficients) were estimated as described in *Materials and Methods*. The 95% confidence limits for each estimated regression coefficient, obtained by bootstrap resampling, are shown. A, B, C, D, and E denote the five epitopes of HA. O are epitopic sites that do not belong to any of the five epitopes, and N are sites not known to be bound by antibodies (see **Table S1** for additional details).

interaction between changes to the low- and high-neutralization efficiency epitopes. Although this result is consistent with the above predictions, it relies on the assumption that the interaction between changes to low- and high-neutralization efficiency epitopes is multiplicative.

**Amino acid changes responsible for transitions between viral antigenic clusters.** A previous empirical study (5) of the evolutionary dynamics of influenza A/H3N2 led to the intriguing observation that whereas the genetic evolution of the virus is continuous, its antigenic evolution is characterized by discontinuous transitions between clusters of antigenically similar viruses. The transition from one antigenic cluster to another (containing viruses capable of escaping from induced antibodies) was generally associated with amino acid changes to two or more HA epitopes but, occasionally, with a single amino acid change. The results of the preceding section suggest that cluster transition should be favored by amino acid changes to epitopes with high neutralization efficiencies, but disfavored by changes to epitopes with low neutralization efficiencies. To test this hypothesis, we quantified the correlation between amino acid changes to individual epitopes of HA and transitions between each pair of temporally adjacent antigenic clusters K and K', where K' is the cluster that replaced K (*Materials and Methods*).

The results (Fig. 5) show that amino acid differences at epitopes A and B correlate positively with transitions from K to K'. In contrast, transitions from K to K' are negatively correlated with amino acid differences at epitopes D and E, at epitopic sites that do not belong to any of the five epitopes of HA, and at sites not known to be bound by antibodies. The effect of amino acid differences at epitope C on cluster transitions could not be quantified accurately. With the exception of epitope D, the observed effects of amino acid differences are consistent with results presented in the preceding section and with the predictions of our antibody-interference model. Similar results were



**Fig. 5.** Correlation between amino acid differences at individual HA epitopes and transitions between 10 pairs of temporally adjacent influenza A/H3N2 viral antigenic clusters. The correlation between amino acid differences and the transition from antigenic cluster K to K' was quantified as described in *Materials and Methods*. In the figure, a positive regression coefficient indicates that an increase in the number of amino acid differences at the associated epitope increases the probability that a given virus belongs to K'. Error bars indicate standard errors of estimated regression coefficients, and they were computed by assuming that the number of viruses belonging to either cluster K or K' is binomially distributed. A, B, C, D, E, O, and N are defined in the legend of Fig. 4.

obtained when the number of amino acid differences at each epitope was adjusted to account for differences in the lengths of individual HA epitopes (Fig. S2).

### Conclusion

We have argued that antibodies that bind to influenza virus HA epitopes located at a distance from the receptor-binding site could be less efficient at occluding the site (and hence at effecting viral neutralization) than antibodies that bind closer to the receptor-binding site. Importantly, the antibodies that bind at a distance from the receptor-binding site could sterically interfere with the antibodies that bind closer to the receptor-binding site (13, 14), thereby impeding viral neutralization. We suggested that the deleterious effects of such antibody interference on viral neutralization could explain some puzzling aspects of the observed antigenic effects of amino acid changes to HA epitopes. As an example, we predicted that amino acid changes to epitopes C and E of the HA of influenza A/H3N2 viruses would lead on average to a decrease in the antigenic difference between A/H3N2 viruses. We showed that existing genetic and antigenic data on A/H3N2 virus evolution support this prediction.

Considering the enormous evolutionary potential of influenza viruses it is possible that the antibody-interference mechanism described above has been exploited to destructive effect by these viruses, particularly as a means to escape from antibodies induced by prior infection and/or vaccination. Experimental studies of antibody interference aimed at the design of more effective, “low-interference” influenza vaccines may therefore be of value. For example, the HA of vaccine strains could be mutated in all regions except those that contain epitopes with high neutralization efficiencies (e.g., epitopes A and B of A/H3N2 HA). Such a vaccine would still raise antibodies against low-neutralization efficiency epitopes (e.g., epitopes C and E of A/H3N2 HA), but these low-neutralization efficiency antibodies would not interact with the target wild-type virus strains (Fig. 3 C–E). Alternatively, influenza vaccines could be designed to include only those regions of HA that correspond to epitopes with high neutralization efficiencies. Peptide-based influenza

vaccines were proposed (27), but there has been little focus on peptides that correspond to epitopes with high neutralization efficiencies. Furthermore, antiinfluenza viral drugs could be designed to include HA proteins carrying modifications to high-neutralization efficiency epitopes; these modified HA proteins would compete with virus for binding to low-neutralization efficiency antibodies in a manner akin to the role played by neuraminidase inhibitors (28).

Because the antibody-interference mechanism described here is general in nature it could have implications for the evolution, neutralization, and management of other pathogens, including HIV type 1, for which there is evidence of competition between antibodies with different neutralization efficiencies (29). It is worth noting that our analyses of influenza viral neutralization were based on *in vitro* experimental data. Research needs to be done to determine precisely how much information these data provide about the neutralization of influenza viruses *in vivo*. Also, note that we did not account specifically for amino acid changes to HA resulting from interactions between HA and other viral proteins. The most important of these interactions, involving the viral neuraminidase protein (3), would most likely result in amino acid changes occurring primarily at the receptor-binding site of HA, and not at the epitopic sites that were of interest to us. (Note that HA must not bind to its receptor so strongly that the HA–receptor bond cannot be cleaved efficiently by the influenza viral neuraminidase protein.)

### Materials and Methods

**Quantifying the Antigenic Effects of Amino Acid Changes.** We quantified the antigenic effects of amino acid changes to individual HA epitopes by using the linear model  $\log_2(r\text{AHM}^{ab}) = \sum_i \beta_i d_i^{ab} + \varepsilon$ , where  $r\text{AHM}^{ab}$  denotes the reciprocal of the AHM measure of the antigenic similarity between viruses  $a$  and  $b$  (a measure of antigenic distance) (23),  $d_i^{ab}$  is the genetic distance between the  $i$ th epitope of the HAs of  $a$  and  $b$ ,  $\beta_i$  is the antigenic effect of genetic changes to the  $i$ th epitope, and  $\varepsilon$  is the residual. We used 72 pairs of influenza A/H3N2 viruses (Tables S3 and S4) to compute  $r\text{AHM}^{ab}$  and  $d_i^{ab}$ , with  $d_i^{ab}$  given by the metric of Atchley et al. (26), which accounts for antigenically relevant physicochemical properties of amino acids, and by the Hamming metric (see SI Methods for additional details). We computed 95% confidence limits for  $\beta_i$  by ( $i$ ) resampling (with replacement) the 72 combinations of  $r\text{AHM}^{ab}$  and  $d_i^{ab}$   $10^4$

times, and (ii) assuming that  $\log_2(rAHM^{ab}) = N(\sum_i \beta_i d_i^{ab}, \sigma)$ , where  $\sigma$  denotes the standard deviation of antigenic distances and  $N$  denotes the Gaussian distribution.

**Quantifying the Effects of Amino Acid Changes on Transitions Between Antigenic Clusters.** We compiled the HA amino acid sequences of influenza A/H3N2 viruses belonging to 11 antigenic clusters (ref. 5 and Table S5). For each pair of temporally adjacent antigenic clusters K and K', with K' the cluster that replaced K, we computed the number of amino acid differences, epitope by epitope, between the HA sequence of each virus and the consensus sequence

of K. For each virus, we normalized the computed number of amino acid differences by the total number of amino acid differences between the viral HA sequence and the consensus sequence of K. We then used logistic regression to quantify the effect of amino acid differences at each epitope on the probability that the virus belongs to either K or K'.

**ACKNOWLEDGMENTS.** We thank Nigel Dimmock, Lynn Enquist, Eddie Holmes, and Adel Mahmoud for very constructive comments on this manuscript and Sandi Milburn for excellent administrative assistance. This work was supported by a Burroughs–Wellcome graduate fellowship (to W.N.) and by Defense Advanced Research Projects Agency Grant HR0011-05-1-0057.

1. World Health Organization (2003) Influenza Factsheet, ([www.who.int/mediacentre/factsheets/fs211/en/](http://www.who.int/mediacentre/factsheets/fs211/en/)), accessed September 14, 2008.
2. Flint SJ, Enquist LW, Racaniello VR, Skalka AM (2004) *Principles of Virology: Molecular Biology, Pathogenesis, and Control* (Am Soc Microbiol, Washington DC).
3. Palese P, Shaw ML (2006) In *Field's Virology*, ed Knipe DM, et al. (Lippincott, Philadelphia).
4. Brown LE, Murray JM, White DO, Jackson DC (1990) An analysis of the properties of monoclonal antibodies directed to epitopes of influenza virus hemagglutinin. *Arch Virol* 114:1–26.
5. Smith DJ, et al. (2004) Mapping the antigenic and genetic evolution of influenza virus. *Science* 305:371–376.
6. Lee MS, Chen JS (2004) Predicting antigenic variants of influenza A/H3N2 viruses. *Emerg Infect Dis* 10:1385–1390.
7. Poljak RJ, et al. (1973) Three-dimensional structure of the Fab' fragment of a human immunoglobulin at 2.8 Å resolution. *Proc Natl Acad Sci USA* 70:3305–3310.
8. Wilson IA, Skehel JJ, Wiley DC (1981) Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution. *Nature* 289:366–373.
9. Jackson DC, Murray JM, White DO, Gerhard WU (1982) Enumeration of antigenic sites of influenza virus hemagglutinin. *Infect Immun* 37:912–918.
10. Schofield DJ, Stephenson JR, Dimmock NJ (1997) High and low efficiency neutralization epitopes on the haemagglutinin of type A influenza virus. *J Gen Virol* 78:2441–2446.
11. Knossow M, et al. (2002) Mechanism of neutralization of influenza virus infectivity by antibodies. *Virology* 302:294–298.
12. Fleury D, et al. (1999) A complex of influenza hemagglutinin with a neutralizing antibody that binds outside the virus receptor binding site. *Nat Struct Biol* 6:530–534.
13. Taylor HP, Dimmock NJ (1994) Competitive binding of neutralizing monoclonal and polyclonal IgG to the HA of influenza A virions in solution: Only one IgG molecule is bound per HA trimer regardless of the specificity of the competitor. *Virology* 205:360–363.
14. Taylor HP, Armstrong SJ, Dimmock NJ (1987) Quantitative relationships between an influenza virus and neutralizing antibody. *Virology* 159:288–298.
15. Bizebard T, et al. (2002) Structure of influenza virus hemagglutinin complexed with a neutralizing antibody. *Nature* 376:92–94.
16. Sui J, et al. (2009) Structural and functional bases for broad-spectrum neutralization of avian and human influenza A viruses. *Nat Struct Biol* 16:265–273.
17. Garrity RR, et al. (1997) Refocusing neutralizing antibody response by targeted dampening of an immunodominant epitope. *J Immunol* 159:279–289.
18. Pantophlet R, Burton DR (2006) GP120: Target for neutralizing HIV-1 antibodies. *Annu Rev Immunol* 24:739–769.
19. Fazekas de St Groth S, Webster RG (1963) The neutralization of animal viruses. IV. Parameters of the influenza virus-antibody system. *J Immunol* 90:151–164.
20. Gibson RE, Levin SA (1976) Distinctions between the two-state and sequential models for cooperative ligand binding. *Proc Natl Acad Sci USA* 74:139–143.
21. Rini JM, Schulze-Gahmen U, Wilson IA (1992) Structural evidence for induced fit as a mechanism for antibody–antigen recognition. *Science* 255:959–965.
22. Webster R, Cox N, Stohr K (2002) *WHO Manual on Animal Influenza Diagnosis and Surveillance*. WHO/CDS/CSR/NCS/2002.5 Rev. 1 (WHO, Geneva, Switzerland).
23. Archetti I, Horsfall FL (1950) Persistent antigenic variation of influenza A viruses after incomplete neutralization *in ovo* with heterologous immune serum. *J Exp Med* 92:441–462.
24. Ndifon W, Dushoff J, Levin SA (2009) On the use of hemagglutination-inhibition for influenza surveillance: Surveillance data are predictive of influenza vaccine effectiveness. *Vaccine* 27:2447–2452.
25. Bush RM, Fitch WM, Bender CA, Cox NJ (1999) Positive selection on the H3 hemagglutinin gene of human influenza virus A. *Mol Biol Evol* 16:1457–1465.
26. Atchley WR, Zhao J, Fernandes AD, Druke T (2005) Solving the protein sequence metric problem. *Proc Natl Acad Sci USA* 102:6395–6400.
27. Naruse H, et al. (1994) A potential peptide vaccine against two different strains of influenza virus isolated at intervals of about 10 years. *Proc Natl Acad Sci USA* 91:9588–9592.
28. Gubareva LV, Hayden FG (2006) In *Influenza Virology: Current Topics*, ed Kawaoka Y (Caister Academic, Norwich, UK).
29. Verrier F, Nadas A, Gorny MK, Zolla-Pazner S (2001) Additive effects characterize the interaction of antibodies involved in neutralization of the primary dualtropic human immunodeficiency virus type 1 isolate 89.6. *J Virol* 75:9177–9186.

# Supporting Information

Ndifon et al. 10.1073/pnas.0903427106

## SI Methods

**Amino Acid Difference Metrics.** Amino acid differences were quantified by the Atchley (1), Hamming, BLOSUM62 (2), and Miyata (3) metrics. The Atchley metric accounts for structural and physicochemical differences between amino acids that are relevant to antibody binding (see below); the BLOSUM62 metric accounts for the likelihood that one amino acid will be replaced by another amino acid on evolutionary time scales; the Miyata metric accounts for differences in the volume and hydrophobicity of amino acids. The Atchley metric was constructed from 494 different metrics that quantified various properties of amino acids, such as size, volume, and structure (1). The statistical method of factor analysis was used to extract a set of five vectors (or “factors”) that collectively captured most of the variation embedded in the original 494 metrics. The resulting factors were subjected to an oblique rotation to improve their explanatory power with respect to a variety of amino acid attributes.

**Quantifying Genetic and Antigenic Distances.** The genetic distance between two hemagglutinin (HA) epitopes was quantified by summing the differences between corresponding amino acids found in each epitope. In the case of the BLOSUM62 metric (2), similarity measures were converted into measures of differences; for each entry  $s_{ij}$ , denoting the similarity between amino acids  $i$  and  $j$ , we defined the difference:  $d_{ij} = s_{jj} - s_{ij}$ . The difference between each amino acid and itself was therefore equal to 0. For the Atchley metric (1), the difference between amino acids  $i$  and  $j$  was defined as  $d_{ij} = \sum_{k=1}^2 |f_{ik} - f_{jk}|$ , where  $f_{ik}$  denotes the  $k$ th factor loading of amino acid  $i$ . The factors were accorded equal weights by setting the  $k$ th factor loading of amino acid  $i$  equal to  $100 \times f_{ik}/\max\{f_{mk}\}_m$ . Observe that for each amino acid only the first two factor loadings, which account for such antigenically relevant amino acid attributes as secondary structure, hydrophobicity, surface exposure, and polarity, were used (results qualitatively similar to those reported here were obtained by using all five factor loadings, with or without the above-mentioned factor weighting). Antigenic distances between pairs of viruses were quantified by the logarithm (base 2) of the reciprocal of the Archetti–Horsfall measure (AHM) of antigenic similarity (4), using antigenic data that we collected in our previous work. The antigenic distances were based on results of the hemagglutination-inhibition assay (5), which tend to be highly variable. To minimize this variability, we only used pairs of viruses for which at least two independent measurements of antigenic distance were available; the geometric mean of such independently measured antigenic distances was used.

**Some Remarks on Estimated Antigenic Effects.** The sign (as well as the magnitude) of the antigenic effect or regression coefficient (RC) associated with a particular epitope depends on the dataset used to estimate it. For example, assume that we have data for only three viruses whose HAs are denoted by  $\text{HA}^{(a)}$ ,  $\text{HA}^{(b)}$ , and  $\text{HA}^{(c)}$ . Let  $\text{HA}^{(a)}$  and  $\text{HA}^{(b)}$  differ genetically at the high-neutralization efficiency epitope A, with corresponding antigenic distance  $d(\text{HA}^{(a)}, \text{HA}^{(b)}) > 0$ . Also, let  $\text{HA}^{(a)}$  differ from  $\text{HA}^{(c)}$  at both epitope A and the low-neutralization efficiency epitope C, with corresponding antigenic distance  $d(\text{HA}^{(a)}, \text{HA}^{(c)}) < d(\text{HA}^{(a)}, \text{HA}^{(b)})$  (see Fig. 2 in the main text). In this simple example, the RC for epitope C (obtained by regressing genetic differences against corresponding antigenic differences) is likely to be negative, all else being equal. However, if  $\text{HA}^{(a)}$  and  $\text{HA}^{(b)}$  differ at epitope C (and not A), then the estimated RC

for epitope C need not be negative. As the amount of data increases, the estimated sign of a given RC would become more representative of the average contribution of the associated epitope to antigenic change.

## Concentration of Antibody and of Influenza Virus HA During Infection.

The concentration of antibodies found in antisera extracted from a variety of influenza virus-infected animals can be high, on the order of  $10^{15}$  molecules per mL (e.g., ref. 6). In contrast, nasal washes taken from influenza virus-infected individuals contain on the order of  $10^6$  50% tissue culture-infective doses (TCID<sub>50</sub>) per mL (e.g., ref. 7). The TCID<sub>50</sub> can be thought of as a measure of the number of infectious viral particles (IVPs) per cell required to infect 50% of the cells found in a virus-cell mixture. There is no exact way to extrapolate from TCID<sub>50</sub> to the number of IVPs. However, a good approximation (8) is obtained by assuming that the fraction of cells infected by  $r$  virus particles in a virus-cell mixture with known average multiplicity of infection  $c$  follows a Poisson distribution:

$$P(r; c) = \frac{e^{-c} c^r}{r!}. \quad [\text{s1}]$$

The value of  $c$  at which 50% of the cells become infected (i.e., by receiving one or more virus particles) is  $-\ln(0.5) = 0.69$ , suggesting that the number of IVPs per TCID<sub>50</sub> is 0.69. This simple calculation, which assumes single-hit kinetics of influenza virus infection (8), suggests that nasal washes contain on the order of  $10^6$  IVPs per mL. Assuming a ratio of IVPs to total virus particles of 1:50 (8), this corresponds to a total virus concentration of  $5 \times 10^7$  virus particles per mL. Even considering the number of HA trimers per virus [i.e.,  $\approx 345$  (9)] and the dilution factor of nasal lining fluid found in nasal washes [i.e., 5–64 (10)], the estimated concentration of HA,  $\approx 10^{11}$  to  $10^{12}$  molecules per mL, is still substantially lower than the measured antibody concentration of  $10^{15}$  molecules per mL.

**Model of Influenza Viral Neutralization.** The results of previous experimental studies (11–14) suggest that the fraction of neutralized influenza viruses increases steadily with the average number of bound antibodies per virus ( $N_{\text{Abs}}$ ) and also depends on the bound epitope. More specifically, the relationship between  $N_{\text{Abs}}$  and the fraction of neutralized viruses,  $F$ , is described well by the following sigmoidal Hill function:

$$F = \frac{(N_{\text{Abs}})^h}{\lambda^h + (N_{\text{Abs}})^h}, \quad [\text{s2}]$$

where  $h$  is the Hill coefficient and  $\lambda$  is the  $N_{\text{Abs}}$  required for half-maximal neutralization.

**Fig. S1** shows a sigmoidal relationship between  $N_{\text{Abs}}$  and  $F$ , for two monoclonal antibody-virus mixtures, namely HC2-A/fowl plaque/Rostock/34 (H7N1) and HC45-A/Hong Kong/X-31/68 (H3N2). We describe in detail how this relationship was measured experimentally in the case of HC45 antibodies. The average number of bound HC45 antibodies per virus was estimated by incubating a fixed amount of virus with different amounts of iodine-labeled antibodies for 1 h at room temperature (12). The antibody–virus mixture was then vortexed, and the radioactivity of the resulting pellet (made up of antibody–virus complexes) was counted. The radioactivity of antibody–virus complexes was quantified as the difference in radioactivity

between experimental and control samples (i.e., samples that do not contain virus) and was used to estimate  $N_{\text{Abs}}$  (12). In addition, mixtures containing a fixed amount of virus and different amounts of HC45 antibodies were used to infect Madin–Darby canine kidney cells, and the number of plaques formed in each virus–cell mixture was determined.  $F$  was defined as the ratio of the number of plaques formed in the presence of antibody to the number of plaques formed in the absence of antibody.

For the HC2–virus mixture, the estimated values of the parameters of (S2) were  $h = 3.3 \pm 2.5$  and  $\lambda = 51.4 \pm 10.9$ , whereas for the HC45–virus mixture, the parameter estimates were  $h = 12.5 \pm 1.9$  and  $\lambda = 96.8 \pm 2.1$ . The data fit sigmoidal Hill functions much better than the data fit linear functions (the mean-squared errors for the sigmoidal and linear fits were:  $6.4 \times 10^{-3}$  versus  $3.5 \times 10^{-2}$  for the HC2–virus mixture, and  $4.0 \times 10^{-4}$  versus  $1.2 \times 10^{-2}$ , for the HC45–virus mixture). Using the above information, we can construct an explicit model of viral neutralization that takes into account differences in the neutralization efficiencies of the epitopes bound by antibodies (15). Let us assume that there are  $m$  epitopes with neutralization efficiencies  $\sigma_i$ ,  $i = 1, \dots, m$ , and  $0 \leq \sigma_i \leq 1$ . To account for the

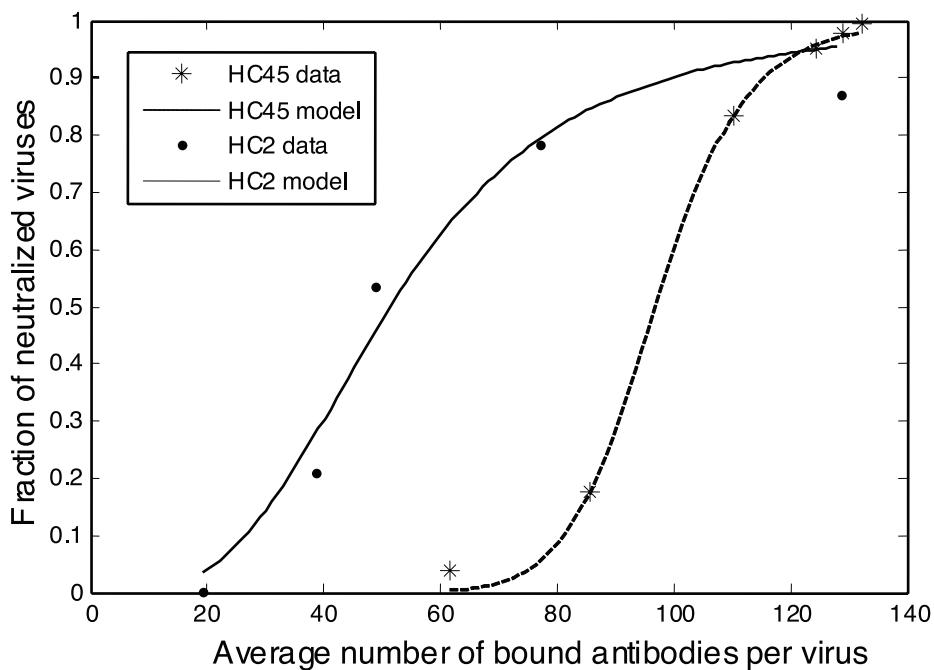
different neutralization efficiencies of the epitopes, we weight the contribution to neutralization of antibody bound to the  $i$ th type by  $\sigma_i$ , yielding the following expression for the fraction of neutralized viruses:

$$F = \frac{\left( n \sum_{i=1}^m \sigma_i P_i \right)^h}{\lambda^h + \left( n \sum_{i=1}^m \sigma_i P_i \right)^h}, \quad i = 1, \dots, m, \quad [\text{s3}]$$

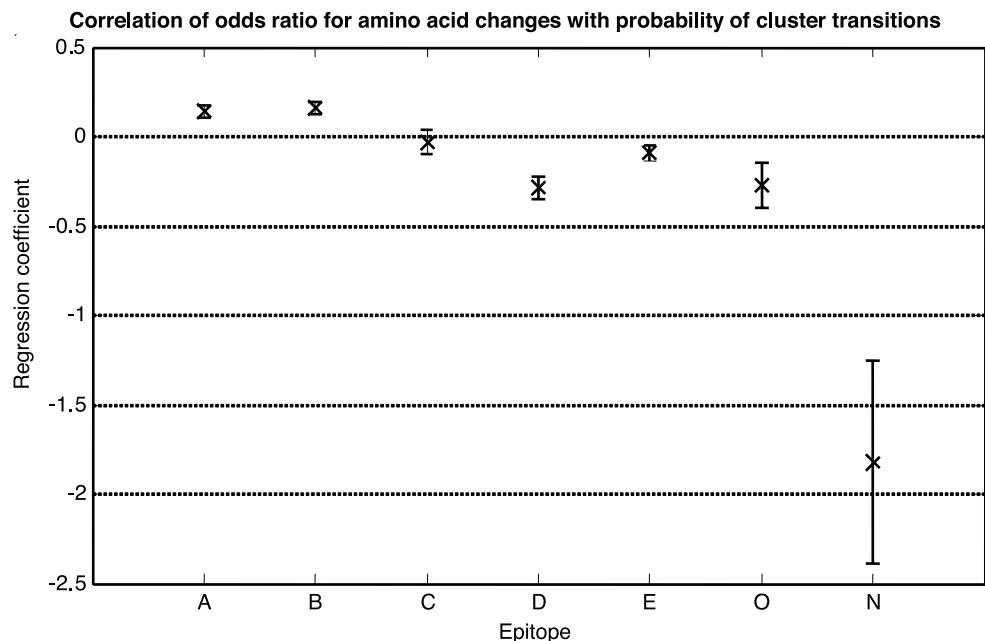
where  $P_i$  is the probability that the  $i$ th epitope is bound at equilibrium (see Eq. 2 in the main text), and  $n$  is the total number of binding sites per virus.

Note that the cooperativity of antibody-mediated viral neutralization, which is suggested by the above results, may derive from the fact that viruses interact with target cells by forming bonds between HA monomers and cell surface receptors. Because of their low strength (the affinity of each of these bonds is on the order of a millimolar), many HA–receptor bonds are required to stabilize a given virus–cell interaction (16). Therefore, multiple antibody molecules must bind to virus to prevent a stable virus–cell interaction.

1. Atchley WR, Zhao J, Fernandes AD, Dröge T (2005) Solving the protein sequence metric problem. *Proc Natl Acad Sci USA* 102:6395–6400.
2. Henikoff S, Henikoff JG (1992) Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci USA* 89:10915–10919.
3. Miyata T, Miyazawa S, Yasunaga T (1979) Two types of amino acid substitutions in protein evolution. *J Mol Evol* 12:219–236.
4. Archetti I, Horsfall FL (1950) Persistent antigenic variation of influenza A viruses after incomplete neutralization in ovo with heterologous immune serum. *J Exp Med* 92:441–462.
5. Fazekas de St Groth S, Webster RG, Davenport FM (1969) The antigenic subunit of influenza viruses. I. The homologous antibody response. *J Immunol* 103:1099–1106.
6. Webster R, Cox N, Stohr K (2002) WHO Manual on Animal Influenza Diagnosis and Surveillance. WHO/CDS/CSR/NCS/2002.5 Rev. 1 (WHO, Geneva, Switzerland).
7. Treanor JJ, et al. (2000) Effectiveness and safety of the oral neuraminidase inhibitor oseltamivir in treating acute influenza: A randomized controlled trial. *J Am Med Assoc* 283:1016–1024.
8. Flint SJ, Enquist LW, Racaniello VR, Skalka AM (2004) *Principles of Virology: Molecular Biology, Pathogenesis, and Control* (Am Soc Microbiol, Washington, DC).
9. Taylor HP, Dimmock NJ (1994) Competitive binding of neutralizing monoclonal and polyclonal IgG to the HA of influenza A virions in solution: Only one IgG molecule is bound per HA trimer regardless of the specificity of the competitor. *Virology* 205:360–363.
10. Hendley J, Gwaltney J (2004) Viral titers in nasal lining fluid compared to viral titers in nasal washes during experimental rhinovirus infection. *J Clin Virol* 30:326–328.
11. Taylor HP, Armstrong SJ, Dimmock NJ (1987) Quantitative relationships between an influenza virus and neutralizing antibody. *Virology* 159:288–298.
12. Knossow M, et al. (2002) Mechanism of neutralization of influenza virus infectivity by antibodies. *Virology* 302:294–298.
13. Edwards MJ, Dimmock NJ (2001) A haemagglutinin (HA1)-specific Fab neutralizes influenza A virus by inhibiting fusion activity. *J Gen Virol* 82:1387–1395.
14. Fleury D, et al. (1999) A complex of influenza hemagglutinin with a neutralizing antibody that binds outside the virus receptor-binding site. *Nat Struct Biol* 6:530–534.
15. Schofield DJ, Stephenson JR, Dimmock NJ (1997) High and low efficiency neutralization epitopes on the haemagglutinin of type A influenza virus. *J Gen Virol* 78:2441–2446.
16. Collins BE, Paulson JC (2004) Cell surface biology mediated by low affinity multivalent protein-glycan interactions. *Curr Opin Chem Biol* 8:617–625.
17. Bui HH, Peters B, Assarsson E, Mbawuike I, Sette A (2007) Ab and T cell epitopes of influenza A virus, knowledge and opportunities. *Proc Natl Acad Sci USA* 104:246–251.
18. Centers for Disease Control and Prevention (2003) Information for the vaccines and related biological products advisory committee. Available at [www.fda.gov/ohrms/dockets/ac/03/briefing/3922B1\\_2.pdf](http://www.fda.gov/ohrms/dockets/ac/03/briefing/3922B1_2.pdf).
19. Smith DJ, et al. (2004) Mapping the antigenic and genetic evolution of influenza virus. *Science* 305:371–376.



**Fig. S1.** Relationship between the average number of bound antibodies ( $N_{Abs}$ ) per virus and the fraction of neutralized viruses. Shown are data from two monoclonal antibody–virus mixtures, namely H7N1 and H3N2. Data for the HC2–virus mixture were extracted from Fig. 5 of ref. 11, whereas data for the HC45–virus mixture were extracted from Fig. 2 of ref. 12. Note that in the latter figure, HC45 is the only antibody for which there is a usable overlap between the plots of antibody concentration versus fraction of neutralized viruses and antibody concentration versus  $N_{Abs}$ . Also, in that figure the  $N_{Abs}$  per 10 hemagglutinating units of virus was given. We estimated the  $N_{Abs}$  per virus by using the fact that 10 hemagglutinating units contain  $\approx 2.2 \pm 8 \times 10^{10}$  HA monomers (12), of which there are  $\approx 3 \times 345 = 1,035$  per virus (9). The model denotes the curve drawn by using Eq. S2, with parameter values estimated from the data.



**Fig. S2.** Correlation between the odds ratio for amino acid differences at individual HA epitopes and transitions between 10 pairs of temporally adjacent influenza A/H3N2 viral antigenic clusters. For each pair of temporally adjacent antigenic clusters K and K', with K' the cluster that replaced K, we computed the number of amino acid differences, epitope by epitope, between the HA sequence of each virus and the consensus sequence of K, normalized by the total number of amino acid differences between the viral HA sequence and the consensus sequence of K. Odds ratios were obtained by multiplying the normalized number of amino acid differences at the  $i$ th epitope by the quantity: (number of amino acids found in HA)/(number of amino acids found in the  $i$ th epitope). We then used logistic regression to quantify the effect of the odds ratio for amino acid differences at each epitope on the probability that a considered virus belongs to either K or K'. In the figure, a positive regression coefficient indicates that an increase in the odds ratio for amino acid differences at the associated epitope increases the probability that the virus belongs to the new cluster, K'. Error bars were computed based on the assumption that the number of viruses belonging to either K or K' is distributed binomially. The amino acid sequences belonging to each antigenic cluster are given in Table S5. A, B, C, D, E, and O are defined in the legend of Table S1, whereas N denotes HA amino acid positions that are currently not known to be bound by antibodies.

**Table S1. Amino acid positions found in each HA epitope and in the receptor-binding site**

Epitope	Amino acid positions
A	122, 124, 126, 130, 131, 132, 133, 135, 137, 138, 140, 142, 143, 144, 145, 146, 150, 152, 168
B	128, 129, 155, 156, 157, 158, 159, 160, 163, 164, 165, 186, 187, 188, 189, 190, 192, 193, 194, 196, 197, 198
C	44, 45, 46, 47, 48, 50, 51, 53, 54, 273, 275, 276, 278, 279, 280, 294, 297, 299, 300, 304, 305, 307, 308, 309, 310, 311, 312
D	96, 102, 103, 117, 121, 167, 170, 171, 172, 173, 174, 175, 176, 177, 179, 182, 201, 203, 207, 208, 209, 212, 213, 214, 215, 216, 217, 218, 219, 226, 227, 228, 229, 230, 238, 240, 242, 244, 246, 247, 248
E	57, 59, 62, 63, 67, 75, 78, 80, 81, 82, 83, 86, 87, 88, 91, 92, 94, 109, 260, 261, 262, 265
O	49, 60, 74, 79, 90, 274, 151, 52, 277, 220, 134, 136, 153, 17, 199, 2, 3, 4, 31, 112, 205, 220, 271
R	98, 135, 136, 137, 138, 153, 155, 183, 190, 194, 195, 224, 225, 226

Given are the amino acid positions found in individual antibody-binding sites (also called epitopes) and in the receptor-binding site of influenza A/H3N2 virus HA. A, B, C, D, and E refer to the five epitopes of HA; O is the collection of sites that are experimentally verified targets of monoclonal antibodies (17) but that do not belong to any of the five HA epitopes; and R is the receptor-binding site.

**Table S2. Antigenic effects of amino acid changes to HA**

Epitope	Regression coefficients and 95% confidence limits			
	Un-normalized genetic distance		Normalized genetic distance	
	Bootstrap re-sampling	Gaussian approximation	Bootstrap re-sampling	Gaussian approximation
Metric of Atchley et al. (2005)				
A	1.1141e-1 (3.4849e-2 1.8491e-1)	(2.6548e-2 1.9627e-1)	2.1168 (7.0100e-1 3.5476)	(5.0440e-1 3.7291)
B	1.0696e-1 (6.0047e-2 1.5702e-1)	(4.5911e-2 1.6800e-1)	2.3531 (1.3166 3.4575)	(1.0100 3.6961)
C	3.9832e-2 (-8.9038e-2 1.7234e-1)	(-1.1790e-1 1.9757e-1)	1.0755 (-2.2531 4.7490)	(-3.1834 5.3343)
D	1.7751e-1 (1.0021e-1 2.4405e-1)	(8.3383e-2 2.7163e-1)	7.2779 (3.9769 9.9725)	(3.4187 1.1137e+1)
E	4.8446e-2 (-1.0755e-1 1.9449e-1)	(-1.2960e-1 2.2649e-1)	1.0658 (-2.2056 4.2498)	(-2.8511 4.9828e)
O	5.7090e-2 (-1.1871e-1 2.4101e-1)	(-1.7048e-1 2.8466e-1)	1.3131 (-2.5835 5.5975)	(-3.9209 6.5471)
N	-5.2285e-2 (-3.1291e-1 1.1609e-1)	(-1.6832e-1 6.3751e-2)	-9.0975 (-2.6615e+1 5.0545)	(-2.9288e+1, 1.1093e+1)
Hamming metric				
A	2.3718e-1 (8.3187e-2 4.3092e-1)	(5.2863e-2 4.2150e-1)	4.5065 (1.8535 7.8598)	(1.0044 8.0085)
B	2.3621e-1 (1.5627e-1 3.2580e-1)	(1.2799e-1 3.4444e-1)	5.1967 (3.4024 7.1893)	(2.8158 7.5777)
C	2.1934e-1 (-7.3614e-2 4.9546e-1)	(-1.5611e-1 5.9480e-1)	5.9222 (-1.5116 13.5921)	(-4.2151 1.6060e+1)
D	1.7741e-1 (4.8042e-2 2.9422e-1)	(7.0012e-3 3.4781e-1)	7.2736 (1.5746 12.6321)	(2.8705e-1 1.4260e+1)
E	-2.3436e-2 (-2.9644e-1 2.3839e-1)	(-4.1098e-1 3.6411e-1)	-0.5156 (-6.8583 5.4725)	(-9.0415 8.0103)
O	4.9409e-2 (-3.0804e-1 4.1786e-1)	(-3.7556e-1 4.7438e-1)	1.1364 (-6.9225 9.4313)	(-8.6379 1.0911e+1)
N	-7.7575e-2 (-2.4251e-1)	(-2.7504e-1 1.1989e-1)	-13.4980 (-40.9438)	(-4.7857e+1 2.0861e+1)

Table S2. Continued

	4.1641e-2)		6.4581)	
BLOSUM62 metric (Henikoff & Henikoff, 1992)				
A	4.6858e-2 (2.2340e-2 7.4417e-2)	(1.5801e-2 7.7915e-2)	8.9030e-1 (4.2885e-1 1.4115)	(3.0021e-1 1.4804)
B	3.8415e-2 (2.4016e-2 5.2745e-2)	(2.0300e-2 5.6531e-2)	8.4514e-1 (5.2535e-1 1.1783)	(4.4660e-1 1.2437)
C	7.0744e-3 (-3.6647e-2 4.4634e-2)	(-5.1212e-2 6.5361e-2)	1.9101e-1 (-1.0251 1.2060)	(-1.3827 1.7647)
D	6.1962e-2 (3.4215e-2 8.8459e-2)	(2.8426e-2 9.5497e-2)	2.5404 (1.4366 3.6323)	(1.1655 3.9154)
E	1.5177e-2 (-2.9954e-2 6.3283e-2)	(-4.7402e-2 7.7757e-2)	3.3390e-1 (-6.4317e-1 1.4058)	(-1.0428 1.7106)
O	1.1041e-3 (-5.6136e-2 5.3411e-2)	(-6.6809e-2 6.9017e-2)	2.5395e-2 (-1.2590 1.2308)	(-1.5366 1.5874)
N	-1.4927e-2 (-4.1359e-2 7.1387e-3)	(-4.7879e-2 1.8026e-2)	-2.5972 (-7.1568 1.3622)	(-8.3310 3.1365)
Miyata metric (Miyata et al., 1979)				
A	6.8550e-3 (3.7789e-3 9.4570e-03)	(3.0739e-3 1.0636e-2)	1.3024e-01 (7.0719e-02 1.7968e-01)	(5.8405e-2 2.0208e-1)
B	5.1250e-3 (2.7244e-3 8.3327e-03)	(1.1955e-3 9.0546e-3)	1.1275e-01 (5.8538e-02 1.8438e-01)	(2.6300e-2 1.9920e-1)
C	1.0887e-2 (-6.3834e-5 2.2536e-02)	(-1.1661e-3 2.2941e-2)	2.9396e-01 (2.0602e-02 6.1338e-01)	(-3.1485e-2 6.1941e-1)
D	5.3908e-3 (2.7376e-3 8.1081e-03)	(1.6663e-3 9.1153e-3)	2.2102e-01 (1.1306e-01 3.3528e-01)	(6.8320e-2 3.7373e-1)
E	7.7503e-4 (-7.9728e-3 1.1749e-02)	(-1.0988e-2 1.2538e-2)	1.7051e-02 (-1.7897e-01 2.5507e-01)	(-2.4174e-1 2.7584e-1)
O	5.8913e-3 (-4.5886e-3 1.5021e-02)	(-5.9567e-3 1.7739e-2)	1.3550e-01 (-1.0581e-01 3.5320e-01)	(-1.3700e-1 4.0800e-1)
N	1.8142e-3 (-2.9759e-3 5.4448e-03)	(-5.9702e-3 9.5987e-3)	3.1568e-01 (-5.1658e-01 9.5189e-01)	(-1.0388 1.6702)

The regression coefficients (RCs) were obtained by regressing the genetic distances between corresponding hemagglutinin epitopes of 72 pairs of influenza A/H3N2 viruses (Table S3) and the antigenic distances between those viruses. Genetic distances were quantified by the different metrics indicated, and they were either unnormalized or normalized (by dividing by the total number of amino acid positions found in the considered epitope). Antigenic distances were quantified by the logarithm (base 2) of the reciprocal of the Archetti–Horsfall measure (AHM) of antigenic similarity (4). The best-fit RCs and associated 95% confidence limits are given. The confidence limits were estimated by means of: (i) bootstrap resampling of the 72 pairs of genetic and antigenic distances, and (ii) a Gaussian approximation of the distribution of deviations between antigenic distances predicted by the best-fit RCs and the actual antigenic distances. A, B, C, D, E, and O are defined in the legend of Table S1; N denotes HA amino acid positions that are currently not known to be bound by antibodies.

**Table S3.** Reciprocal of the AHM measure of antigenic similarity (rAHM) between virus pairs

Virus 1	Virus 2	rAHM	Virus 1	Virus 2	rAHM
1	2	2.3784	19	17	39.1918
4	1	1.7818	19	16	22.6274
8	9	2.2251	18	17	4.4645
8	7	2.4481	18	16	5.6569
8	36	4.9246	17	16	5.0397
8	10	2.2974	17	14	22.6274
8	35	1.4142	16	15	4.7568
8	37	2	13	12	14.2544
9	7	2.3469	22	23	4.4898
9	36	10.5561	22	13	5.6569
9	10	1.8661	23	13	1
9	35	2	24	25	16
9	37	2.3784	24	26	22.6274
7	36	2.4623	25	26	2.3784
7	10	1.8661	26	22	2
7	35	1.4142	24	27	4.7568
7	37	1.4142	28	24	5.0397
36	10	3.4822	16	29	4.7568
36	35	2	18	30	2.3784
36	37	5.6569	17	30	2
10	35	2	30	16	2.3784
10	37	1.1892	20	31	2.1104
35	37	1.4142	31	19	4.6458
9	6	13.4543	32	20	7.3890
7	6	8	32	31	4.7679
11	9	11.3137	32	19	8.6366
11	8	6.7272	19	33	70.8276
12	11	1.1892	17	33	7.6255
12	8	13.4543	17	34	1.5651
21	20	5.2872	33	34	22.6274
21	19	38.0546	18	33	8
21	16	38.0546	38	39	1.4142
20	19	8.5450	39	1	4
20	17	35.4138	38	1	2.3784
20	16	38.0546	39	2	1.6818
19	18	21.3997	38	40	2

The logarithm (base 2) of the rAHM is used as a measure of antigenic distance. The HA sequences of the viruses are given in [Table S4](#).

**Table S4. HA sequences of influenza A/H3N2 viruses**

ID	Name	HA amino acid sequence	Reference
1	California/7/04	QKLPGNDNSTATLCLGHHAVPNGTIVKTITNDQIEVTNATELVQSSSTGGICDS PHQILDGENCTLIDALLGDPQCDGFQNKKWDLFVERSКАSNCYPDVPDYA SLRSLVASSGTLEFNESFNWTGVTQNGTSSACKRRSNNSFFSRLNWLTHLF KYPALNVTMPNEKFQDFKLYIWGVHHPGTNDQISLYAQASGRITVSTKRSQQ TVIPNIGSRPRVRDIPSRIYWTIVKPGDILLINSTGNLIAPRGYFKIRSGKSSIM RSDAPIGKCNCSECITPNGSIPNDKPFQNVNRITYGACPRYVKQNTLKLATGMR NVPEKQTR	DQ86597 3
2	Wisconsin/67/05	QKLPGNDNSTATLCLGHHAVPNGTIVKTITNDQIEVTNATELVQSSSTGGICDS PHQILDGENCTLIDALLGDPQCDGFQNKKWDLFVERSКАSNCYPDVPDYA SLRSLVASSGTLEFNESFNWTGVTQNGTSSACKRRSNNSFFSRLNWLTHLF KYPALNVTMPNEKFQDFKLYIWGVHHPGTNDQISLYAQASGRITVSTKRSQQ TVIPNIGSRPRVRNIPSRIYWTIVKPGDILLINSTGNLIAPRGYFKIRSGKSSIM RSDAPIGKCNCSECITPNGSIPNDKPFQNVNRITYGACPRYVKQNTLKLATGMR VPEKQTR	DQ86594 7
3	Kentucky/3/06	QKLPGNDNSTATLCLGHHAVPNGTIVKTITNDQIEVTNATELVQSSSTGEICDS PHQILDGENCTLIDALLGDPQCDGFQNKKWDLFVERSКАSNCYPDVPDYA SLRSLVASSGTLEFNESFNWTGVTQNGTSSACKRRSNNSFFSRLNWLTHSKF KYPALNVTMPNEKFQDFKLYIWGVHHPGTNDQISLYAQASGRITVSTKRSQQ TVIPNIGSRPRVRDIPSRIYWTIVKPGDILLINSTGNLIAPRGYFKIRSGKSSIM RSDAPIGKCNCSECITPNGSIPNDKPFQNVNRITYGACPRYVKQNTLKLATGMR NVPEKQTR	EF473496
4	Wellington/1/04	QKLPGNDNSTATLCLGHHAVPNGTIVKTITNDQIEVTNATELVQSSSTGGICDS PHQILDGENCTLIDALLGDPQCDGFQNKKWDLFVERSКАSNCYPDVPDYA SLRSLVASSGTLEFNESFNWTGVTQNGTSSACKRRSNNSFFSRLNWLTHLF KYPALNVTMPNEKFQDFKLYIWGVHHPGTNDQISLYAQASGRITVSTKRSQQ TVIPNIGSRPRVRDIPSRIYWTIVKPGDILLINSTGNLIAPRGYFKIRSGKSSIM RSDAPIGKCNCSECITPNGSIPNDKPFQNVNRITYGACPRYVKQNTLKLATGMR NVPEKQTR	EF566073
5	Singapore/37/04	QKLPGNDNSTATLCLGHHAVPNGTIVKTITNDQIEVTNATELVQSSSTGGICDS PHQILDGENCTLIDALLGDPQCDGFQNKKWDLFVERSКАSNCYPDVPDYA SLRSLVASSGTLEFNESFNWTGVTQNGTSSACKRRSNNSFFSRLNWLTHLF KYPALNVTMPNEKFQDFKLYIWGVHHPGTNDQISLYAQASGRITVSTKRSQQ TVIPNIGSRPRVRDIPSRIYWTIVKPGDILLINSTGNLIAPRGYFKIRSGKSSIM RSDAPIGKCNCSECITPNGSIPNDKPFQNVNRITYGACPRYVKQNTLKLATGMR NVPEKQTR	EF566176
6	Fujian/411/02	QKLPGNDNSTATLCLGHHAVPNGTIVKTITNDQIEVTNATELVQSSSTGGICDS PHQILDGENCTLIDALLGDPQCDGFQNKKWDLFVERSКАSNCYPDVPDYA SLRSLVASSGTLEFNESFNWTGVTQNGTSSACKRRSNKSFFSRLNWLTHLK YKYPALNVTMPNEKFQDFKLYIWGVHHPGTNDQISLYAQASGRITVSTKRSQ QTIPNIGSRPRVRDVSSRIYWTIVKPGDILLINSTGNLIAPRGYFKIRSGKSSI MRSDAPIGKCNCSECITPNGSIPNDKPFQNVNRITYGACPRYVKQNTLKLATGM RNPEKQTR	EF541397
7	Panama/2007/99	QKLPGNDNSTATLCLGHHAvgSNGTLVKTITNDQIEVTNATELVQSSSTGRICD SPHQILDGENCTLIDALLGDPHCDGFQNKEWDLFVERSКАSNCYPDVPDY ASLRSLVASSGTLEFNESFNWTGVAQNGTSSACKRRSNNSFFSRLNWLHQL KYKYPALNVTMPNEKFQDFKLYIWGVHHPSTSDQTSIYAQASGRVTVSTKRS QQTVIPNIGSIPWVRGVSSRIYWTIVKPGDILLINSTGNLIAPRGYFKIRSGKS SIMRSDAPIGKCNCSECITPNGSIPNDKPFQNVNRITYGACPRYVKQNTLKLATG MRNVPEKQTR	DQ48734 0
8	Sydney/5/97	QKIPGNDNSTATLCLGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGRICDS PHRILDGENCTLIDALLGDPHCDGFQNKEWDLFVERSКАSNCYPDVPDYA SLRSLVASSGTLEFNESFNWTGVAQNGTSYACKRSSIKSFFSRLNWLHQLKY KYPALNVTMPNDKFQDFKLYIWGVHHPSTSDQTSIYAQASGRVTVSTKRSQQ TVIPNIGSRPWRGISSRIYWTIVKPGDILLINSTGNLIAPRGYFKIRSGKSSIM RSDAPIGKCNCSECITPNGSIPNDKPFQNVNRITYGACPRYVKQNTLKLATGMR NVPEKQTR	EF566075
9	Moscow/10/99	QKLPGNDNSTATLCLGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGRICD SPHQILDGENCTLIDALLGDPHCDGFQNKEWDLFVERSКАSNCYPDVPDY ASLRSLVASSGTLEFNESFNWTGVAQNGTSSACKRRSIKSFFSRLNWLHQL NRYPALNVTMPNDKFQDFKLYIWGVHHPSTSDQTSIYAQASGRVTVSTKRS QQTVIPNIGSIPWVRGVSSRIYWTIVKPGDILLINSTGNLIAPRGYFKIRSGKS SIMRSDAPIGKCNCSECITPNGSIPNDKPFQNVNRITYGACPRYVKQNTLKLATG MRNVPEKQTR	DQ48734 1
10	Chile/6416/01	QKLPGNDNSTATLCLGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGRICD SPHQILDGENCTLIDALLGDPHCDGFQNKEWDLFVERSКАSNCYPDVPDY VSLRSLVASSGTLEFNESFNWTGVAQNGTSSACKRRSDKSFFSRLNWLHQL	DQ86597 2

Table S4. Continued

		KYKYPALNVTMPNNEKFKLYIWGVHHPGTDSDQISIYAQASGRVTVSTKRS QQTVIPNIGSRHWVRGXSSRISIYWTIVKPGDILLINSTGNLIAPRGYFKIRSGKS SIMRSADAPIGKCNCSECITPNGSIPNDKPFQNVNRITYGACPRYVKQNTLKLATG MRNVPEKQTR	
11	Nanchang/93 3/95	QKLPGNDNSTATLCLGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGRICD SPHRILDGKNCTLIDALLGDPHCDGFQNKEWDLFVERSKEYNCYPYDVPDY ASLRSLVASSGTLEFTNEGFWTGVQAQDGTSYACKRGSVKSFFSRLNWHLKL EYKYPALNVTMPNNDKFKLYIWGVHHPTSDVQTSLYVQASGRVTVSTKR SQQTIPNIGSRPWRGISSRISIYWTIVKPGDILLINSTGNLIAPRGYFKIRSGKS SIMRSADAPIGNCSSECITPNGSIPNDKPFQNVNRITYGACPRYVKQNTLKLATG MRNVPEKQTR	AF008725
12	Wuhan/359/9 5	QKLPGNDNSTATLCLGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGRICD SPHRILDGKNCTLIDALLGDPHCDGFQNKEWDLFVERSKEYNCYPYDVPDY ASLRSLVASSGTLEFTNEGFWTGVQAQDGTSYACKRGSVKSFFSRLNWHLKL EYKYPALNVTMPNNDKFKLYIWGVHHPTSDQTSIYVQASGRVTVSTKR SQQTIPNIGSRPWRGISSRISIYWTIVKPGDILLINSTGNLIAPRGYFKIRSGKS IMRSADAPIGNCSSECITPNGSIPNDKPFQNVNRITYGACPRYVKQNTLKLATG MRNVPEKQTR	AF008722
13	Johannesburg /33/94	QKLPGNDNSTATLCLGHHAVPNGTLVKTITNDQIEVTNATELVQSSPTGRICD SPHRILDGKNCTLIDALLGDPHCDGFQNKEWDLFVERSKEYNCYPYDVPDY ASLRSLVASSGTLEFTENFNWTGVQAQDGTSYACKRGSVNSFFSRLNWHLKL EYKYPALNVTMPNNNGFDKLYIWGVHHPTSDQTSIYVQASGRVTVSTKR SQQTIPNIGSRPWRGLSSRISIYWTIVKPGDILLINSTGNLIAPRGYFKIRNG KSSIMRSADAPIGNCSSECITPNGSIPNDKPFQNVNRITYGACPRYVKQNTLKL TGMRNVPEKQTR	AY66118 0
14	Shanghai/11/ 87	QKLPGNDNSTATLCLGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGRICD SPHRILDGKNCTLIDALLGDPHCDGFQNKEWDLFVERSKEYNCYPYDVPDY ASLRSLVASSGTLEFTENFNWTGVQTSGGSYACKRGSVNSFFSRLNWHLHESE YKYPALNVTMPNNNGFDKLYIWGVHHPTDREQTNLVYVQASGRVTVSTKR SQQTIPNIGSRPWRGLSSRISIYWTIVKPGDILLINSTGNLIAPRGYFKIRTGKS SIMRSADAPIGTCSECITPNGSIPNDKPFQNVNKITYGACPRYVKQNTLKLATG MRNVPEKQTR	AF008886
15	Mississippi/1 /85	QKLPGNDNSTATLCLGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGRICD SPHRILDGKNCTLIDALLGDPHCDGFQNKEWDLFVERSKEYNCYPYDVPDY ASLRSLVASSGTLEFTINEGFWTGVQTSGGSSTCKRGSNNSFFSRLNWLYESE SKYPALNVTMPNNNGFDKLYIWGVHHPTDKEQTNLVYVQASGRVTVSTKR SQQTIPNIGSRPWRGLSSRISIYWTIVKPGDILLINSTGNLIAPRGYFKIRTGKS SIMRSADAPIGTCSECITPNGSIPNDKPFQNVNKITYGACPRYVKQNTLKLATG MRNVPEKQTR	L19003
16	Philipines/2/8 2	QNLPGNDNSTATLCLGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGRICD SPHRILDGKNCTLIDALLGDPHCDGFQNKEWDLFVERSKEYNCYPYDVPDY ASLRSLVASSGTLEFTINEGFWTGVQTSGGSSTCKRGSNNSFFSRLNWLYESE SKYPVLNVTMPNNNGFDKLYIWGVHHPTDKEQTNLVYVQASGRVTVSTKR SQQTIPNIGSRPWRGLSSRISIYWTIVKPGDILLINSTGNLIAPRGYFKIRTGSSI MRSADAPIGTCSECITPNGSIPNDKPFQNVNKITYGACPRYVKQNTLKLATGM RNVPEKQTR	L19002
17	Bangkok/1/7 9	QNLPGNDNSTATLCLGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGRICD SPHRILDGKNCTLIDALLGDPHCDGFQNKEWDLFVERSKEYNCYPYDVPDY ASLRSLVASSGTLEFTINEGFWTGVQTSGGSYACKRGPDNGFFSRLNWLYESE SKYPVLNVTMPNNNGFDKLYIWGVHHPTDKEQTNLVYVQASGRVTVSTKR SQQTIPNIGSRPWRGLSSRISIYWTIVKPGDILLINSTGNLIAPRGYFKIRTGKS SIMRSADAPIGTCSECITPNGSIPNDKPFQNVNKITYGACPRYVKQNTLKLATG MRNVPEKQTR	DQ50882 5
18	Texas/1/77	QNLPGNDNSTATLCLGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGRICD SPHRILDGKNCTLIDALLGDPHCDGFQNKEWDLFVERSKEYNCYPYDVPDY ASLRSLVASSGTLEFTINEGFWTGVQTQNGGSYACKRGPDNGFFSRLNWLYKS ESTYPVLNVTMPNNNGFDKLYIWGVHHPTDKEQTNLVYVQASGRVTVSTKR SQQTIPNIGSRPWRGLSSRISIYWTIVKPGDILLINSTGNLIAPRGYFKIRTGK SSIMRSADAPIGTCSECITPNGSIPNDKPFQNVNKITYGACPRYVKQNTLKLAT GMRNVPEKQTR	EF626623
19	Victoria/3/75	QDLPGNDNSTATLCLGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGRICD NPHRILDGINCTLIDALLGDPHCDGFQNKEWDLFVERSKEYNCYPYDVPDY ASLRSLVASSGTLEFTINEGFWTGVQTQNGGSACKRGPDNGFFSRLNWLYKL GSTYPVQNVTMPNNNDNSDKLYIWGVHHPTDKEQTDLVYVQASGRKVTVSTKR NQQTIPNIGSRPWRGLSSRISIYWTIVKPGDILVINSNGNLIAPRGYFKMR TGKSSIMRSADAPIGTCSECITPNGSIPNDKPFQNVNKITYGACPRYVKQNTLKL LATGMRNVPBKQTR	EF626609
20	England/42/7 2	QDLPGNDNSTATLCLGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGRICD NPHRILDGIDCTLIDALLGDPHCDGFQNNETWDLFVERSKEYNCYPYDVPDYA	EF626613

Table S4. Continued

		SLRSLVASSGTLEFINEGFTWTGVTQNGGSNACKRGPDGFFSRLNWLYKSGS TYPVLNVTPNNDNFKLYIWGVHHPSTNQEQTSLYVQASGRVTVSTKGSQ QTIPNIGSRPVVRGLSSRISIYWTIVKPGDILVINSNGNLIAPIRGYFKMRTGKS SIMRSADAPIGTCISECITPNGSIPNDKPFQNVNKITYGACPKYVKQNTLKLATGM RNVPEKQTR	
21	Hong kong/1/68	QDLPGNDNSTATLCGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGRICN SPHRILDGKNDCTLIDALLGDPHCDGVFKNETWDLFVERSKAFCNSCYPYDVPDY ASLRSLVASSGTLEFINEGFTWTGVTQDGGSNAFKRGPGKRRFSRLNWLTKGSG STYPVLNVTPNNDNFKLYIWGVHHPSTNQEQTSLYVQAPGRVTVSTRSQ QTIPNIGSRPVVRGLSSRISIYWTIVKPGDILVINSNGNLIAPIRGYFKMRTGKS SIMRSADAPIGTCISECITPNGSIPNDKPFQNVNKITYGACPKYVKQNTLKLATGM RNVPEKQTR	AF201874
22	Shangdong/9/93	QKLPGNDNSTATLCGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGRICG SPHRILDGKNCNTLIDALLGDPHCDGFQNKEWDLFVERSKAYSNCYPYDVPDY ASLRSLVASSGTLEFINEDFNWTGVAQDGGSYACKRGSVNSFFSRLNWHLHK EYKPALNVTPMPNNNGFDKLYIWGVHHPSTSDQTSLYVRASGRVTVSTKRS SQQTVPNIGSRPVVRGQSSRISIYWTIVKPGDILLINSTGNLIAPIRGYFKIRNGKS SIMRSADAPIGNCSSECITPNGSIPNDKPFQNVNKITYGACPRYVKQNTLKLATGM RMRNVPEKQTR	L76037
23	Guangdong/2/5/93	QKLPGNDNSTATLCGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGRICD SPHRILDGKNCNTLIDALLGDPHCDGFQNKEWDLFVERSKAYSNCYPYDVPDY ASLRSLVASSGTLEFINEDFNWTGVAQDGGSYACKRGSVNSFFSRLNWHLHK EYKPALNVTPMPNNNGFDKLYIWGVHHPSTSDQTSLYVRASGRVTVSTKRS SQQTVPNIGSRPVVRGQSSRISIYWTIVKPGDILLINSTGNLIAPIRGYFKIRNGKS SIMRSADAPIGNCSSECITPNGSIPNDKPFQNVNKITYGACPRYVKQNTLKLATGM RNVPEKQTR	AF008828
24	Beijing/353/89	QKLPGNDNSTATLCGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGRICD SPHRILDGKNCNTLIDALLGDPHCDGFQNKEWDLFVERSKAYSNCYPYDVPDY ASLRSLVASSGTLEFINEDFNWTGVAQSGESYACKRGSVNSFFSRLNWHLHESE YKPALNVTPMPNNNGFDKLYIWGVHHPITDREQTNLVVRASGRVTVSTKRS SQQTVPNIGSRPVVRGQSSRISIYWTIVKPGDILLINSTGNLIAPIRGYFKIRNGKS SIMRSADAPIGTCSEECITPNGSIPNDKPFQNVNKITYGACPRYVKQNTLKLATGM RNVPEKQTR	Z46391
25	Hong kong/34/90	QKLPGNDNSTATLCGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGRICD SPHRILDGKNCNTLIDALLGDPHCDGFQNKEWDLFVERSKAYSNCYPYDVPDY ASLRSLVASSGTLEFINEDFNWTGVAQDGGSYACKRGSVNSFFSRLNWHLHK EYKPALNVTPMPNNKGFDKLYIWGVHHPSTDRDQTSLYVRASGRVTVSTKRS SQQTVPNIGSRPVVRGQSSRISIYWTIVKPGDILLINSTGNLIAPIRGYFKIRNGKS SIMRSADAPIGTCSEECITPNGSIPNDKPFQNVNKITYGACPRYVKQNTLKLATGM RMRNVPEKQTR	Z46409
26	Beijing/32/92	QKLPGNDNSTATLCGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGRICD SPHRILDGKNCNTLIDALLGDPHCDGFQNKEWDLFVERSKAYSNCYPYDVPDY ASLRSLVASSGTLEFINEDFNWTGVAQDGGSYACKRGSVNSFFSRLNWHLHK EYKPALNVTPMPNNKGFDKLYIWGVHHPSTDRDQTSLYVRASGRVTVSTKRS SQQTVPNIGSRPVVRGQSSRISIYWTIVKPGDILLINSTGNLIAPIRGYFKIRNGKS SIMRSADAPIGTCSEECITPNGSIPNDKPFQNVNKITYGACPRYVKQNTLKLATGM RMRNVPEKQTR	AF008812
27	Washington/15/91	QKLPGNDNSTATLCGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGRICD SPHRILDGKNCNTLIDALLGDPHCDGFQNKEWDLFVERSKAYSNCYPYDVPDY ASLRSLVASSGTLEFINEDFNWTGVAQSGESYACKRGSVNSFFSRLNWHLHESD YKPALNVTPMPNNKGFDKLYIWGVHHPITDREQTNLVVRASGRVTVSTKRSQ QTIPNIGSRPVVRGLSSRISIYWTIVKPGDILLINSTGNLIAPIRGYFKIRTGSSI MRSADAPIGTCSEECITPNGSIPNDKPFQNVNKITYGACPRYVKQNTLKLATGM RNVPEKQTR	AF008676
28	England/427/88	QKLPGNDNSTATLCGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGRICD SPHRILDGKNCNTLIDALLGDPHCDGFQNKEWDLFVERSKAYSNCYPYDVPDY ASLRSLVASSGTLEFINEDFNWTGVAQSGESYACKRGSVNSFFSRLNWHLHK EYKPALNVTPMPNNKGFDKLYIWGVHHPSTDRDQTSLYVRASGRVTVSTKRS SQQTVPNIGSRPVVRGQSSRISIYWTIVKPGDILLINSTGNLIAPIRGYFKIRTGK SSIMRSADAPIGTCSEECITPNGSIPNDKPFQNVNKITYGACPRYVKQNTLKLATGM RNVPEKQTR	L18997
29	Caen/1/84	QKLPGNDNSTATLCGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGRICD SPHRILDGKNCNTLIDALLGDPHCDGFQNKEWDLFVERSKAYSNCYPYDVPDY ASLRSLVASSGTLEFINEGFNWTGVTQSGGSYACKRGSVNSFFSRLNWLYKSE SKYPVLNVTPMPNNKGFDKLYIWGVHHPSTDKEQTNLVVRASGRVTVSTKRS SQQTVPNIGSRPVVRGLSSRISIYWTIVKPGDILLINSTGNLIAPIRGYFKIRTGKS SIMRSADAPIGTCSEECITPNGSIPNDKPFQNVNKITYGACPRYVKQNTLKLATGM RNVPEKQTR	AF008867
30	Belgium/2/81	QKLPGNDNSTATLCGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGRICD SPHRILDGKNCNTLIDALLGDPHCDGFQNKEWDLFVERSKAYSNCYPYDVPDY ASLRSLVASSGTLEFINEGFNWTGVTQSGGSYACKRGSVNSFFSRLNWLYKSE SKYPVLNVTPMPNNKGFDKLYIWGVHHPSTDKEQTNLVVRASGRVTVSTKRS SQQTVPNIGSRPVVRGQSSRISIYWTIVKPGDILLINSTGNLIAPIRGYFKIRNGKS SIMRSADAPIGTCSEECITPNGSIPNDKPFQNVNKITYGACPRYVKQNTLKLATGM RNVPEKQTR	AF201844

Table S4. Continued

		SPHRILDGKNCNTLVDALLGDPHCDGFQNEKWDLFVERSKAFNSNCYPYDVPDYA ASLRSLVASSGTLEFINESFNWTGVTQSGGSYACKRGSDKSFFSRLNWLYESE SRYPVLNVTMPNNGNFDKLYIWGVHHPSTDEQTNLYVRASGRVTVSTKRS QTIPNIGSRPVWRGLSSRISIYWTIVKPGDILLINSNGNLIAPRGYFKMRTGK SIMRSDAPIGTCSECITPNGSIPNDKPFQNVNKITYGACPKYVKQNTLKLATGM MRNVPEKQTR	
31	Port chalmers/1/73	QDFPGNDNSTATLCLGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGKICNNPHRILDGINCTLIDALLGDPHCDGFQNETWDLFVERSKAFNSNCYPYDVPDYA ASLRSLVASSGTLEFINEGFTWTGVTQNGGSNACKRGPDGFFSRLNWLYKSGS AYPVLNVTMPNNDNFDKLYIWGVHHPSTDEQTNLYVQTSGRVTVSTKRSQ QTIPNIGSRPVWRGLSSRISIYWTIVKPGDILVINSNGNLIAPRGYFKMRTGKSS IMRSDAPIGTCSECITPNGSIPNDKPFQNVNKITYGACPKYVKQNTLKLATGM MRNVPEKQTR	EF626618
32	Hong kong/107/71	QDLPGNDNSTATLCLGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGKICNNPHRILDGINCTLIDALLGDPHCDVFQNETWDLFVERSKAFNSNCYPYDVPDYA ASLRSLVASSGTLEFINEGFTWTGVTQNGGSNACKRGPDGFFSRLNWLYKSGS RTYPVLNVTMPNNDNFDKLYIWGVHHPSTDEQTSLYVQTSGRVTVSTRRSQ QTIPNIGSRPVWRGLSSRISIYWTIVKPGDILVINSNGNLIAPRGYFKMRTGKSS IMRSDAPIGTCSECITPNGSIPNDKPFQNVNKITYGACPKYVKQNTLKLATGM MRNVPEKQTR	EF626615
33	Bangkok/2/79	QNLPGNDNSTATLCLGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGRICDSPHRILDGKNCNTLIDALLGDPHCDGFQNEKWDLFVERSKAFNSNCYPYDVPDYA ASLRSLVASSGTLEFINEGFTWTGVTQNGGSYACKRGSDNSFFSRLNWLYESE SKYPVLNVTMPNNGNFDKLYIWGVHHPSTYKEQTLYVRASGRVTVSTKRS QTIPNIGSRPVWRGLSSRISIYWTIVKPGDILLINSNGNLIAPRGYFKIRTGKSS IMRSDAPIGTCSECITPNGSIPNDKPFQNVNKITYGACPKYVKQNTLKLATGM MRNVPEKQTR	EF626611
34	Shanghai/31/80	QDLPGNDNNTATLCLGHHAVPNGTLVKTITNDQIEVTNATEXLVQSSSTGRICDSPHRILDGKNCNTLIDALLGDPHCDGFQNEKWDLFVERSKAFNSNCYPYDVPDYA ASLXSLVASSGTLEFINEGFTWTGVTQNGGSYACKRGSDNSFFSRLXWLTKSG STYPVLNVTMPNNDNFDKLYIWGVHHPSTDEQTNLYVRXSGRVTVSTRXQ QTIPNIGPRPVWRGXSSRISIYWTIVKPGDVLVINSNGNLIAPRGYFKMRTGKSS KSSIMRSDAPIGTCSECITPNGSIPNDKPFQNVNKITYGACPKYVKQNTLKLATGM MRNVPEKQTR	EF626620
35	Hong kong/1550/02	QKLPGNDNSTATLCLGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGGICDSPHQILDGENCTLIDALLGDPHCDGFQNKWWDLFVERSKAYSNCYPYDVPDYA ASLRSLVASSGTLEFNNEKFQDKLYIWGVHHPTSDQISIYQAQASGRITVSTKRS KYKPALNVTPNNEKFQDKLYIWGVHHPTSDQISIYQAQAPGRVTVSTKRS QTIPNIGSRPVWRDVSSRISIYWTIVKPGDILLINSTGNLIAPRGYFKIRSGKSS IMRSDAPIGKCNCSECITPNGSIPNDKPFQNVNRITYGACPRYVKQNTLKLATGM MRNVPEKQTR	(18)
36	Fujian/140/00	QKLPGNDNSTATLCLGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGGICDSPHQILDGENCTLIDALLGDPHCDGFQNKWWDLFVERSKAYSNCYPYDVPDYA ASLRSLVASSGTLEFNNEFSNWTGVAQNGTSSACKRRSNKSFFSRLNWHLHQL KYKPALNVTPNNEKFQDKLYIWGVHHPTSDQISIYQAQAPGRVTVSTKRS QTIPNIGSRPVWRGISSRISIYWTIVKPGDILLINSTGNLIAPRGYFKIRSGKSS IMRSDAPIGKCNCSECITPNGSIPNDKPFQNVNRITYGACPRYVKQNTLKLATGM MRNVPEKQTR	(18)
37	New york/55/01	QKLPGNDNSTATLCLGHHAVPNGTLVKTITNDQIEVTNATELVQSSSTGGICDSPHQILDGENCTLIDALLGDPHCDGFQNKWWDLFVERSKAYSNCYPYDVPDYA VSLRSLVASSGTLEFNNEFSNWTGVAQNGTSSACKRRSDKSFFSRLNWHLHQL KYKPALNVTPNNEKFQDKLYIWGVHHPTSDQISIYQAQASGRITVSTKRS QTIPNIGSRPVWRGVSSGSIYWTIVKPGDILLINSTGNLIAPRGYFKIRSGKSS IMRSDAPIGKCNCSECITPNGSIPNDKPFQNVNRITYGACPRYVKQNTLKLATGM MRNVPEKQTR	(18)
38	Brisbane/10/07	QKLPGNDNSTATLCLGHHAVPNGTIVKKTITNDQIEVTNATELVQSSSTGEICDSPHQILDGENCTLIDALLGDPQCDGFQNKWWDLFVERSKAYSNCYPYDVPDYA SLRSLVASSGTLEFNNEFSNWTGVTQNGTSSACKRRSNNSFFSRLNWLTBLKF KYPALNVTPNNEKFQDKLYIWGVHHPTDNDQIFLYQAQASGRITVSTKRSQQ TVIPNIGSRPVRNPSRISIYWTIVKPGDILLINSTGNLIAPRGYFKIRSGKSSIM RSDAPIGKCNCSECITPNGSIPNDKPFQNVNRITYGACPRYVKQNTLKLATGM NVPEKQTR	ACD62305
39	Nepal/921/06	QKLPGIDNSTATLCLGHHAVPNGTIVKKTITNDQIEVTNATELVQSSSTGEICDSPHQILDGENCTLIDALLGDPQCDGFQNKWWDLFVERSKAYSNCYPYDVPDYA SLRSLVASSGTLEFNNEFSNWAGVTQNGTSSACKRGSNNSFFSRLNWLTQSKF KYPALNVTPNNEKFQDKLYIWGVHHPTDNDQIFLYQAQASGRITVSTKRSQQ TVIPNIGSRPVRNPSRISIYWTIVKPGDILLINSTGNLIAPRGYFKIRSGKSSIM RSDAPIGKCNCSECITPNGSIPNDKPFQNVNRITYGACPRYVKQNTLKLATGM NVPEKQTR	ABU50564

**Table S4. Continued**

40	Brisbane/9/0 6	QKLPGNDNSTATLCLGHHAVPNTIVKTITNDQVEVTNATELVQSSSTGEICD SPHQILDGENCTLIDALLGDPQCDGFQNKKWDLFVERSKAYSNCYPYDVPDY ASLRSLVASSGTLEFNESFNWTGVTQNGTSSACKRRSNNSFFSRLNWLTHL KFKYPALNVTMPNEKFDKLYIWGVHHPGTNDQRFLYAQASGRITVSTKRS QQTVIPNIGSRPRVRNIPSRIISIYWTIVKPGDILLINSTGNLIAPRGYFKIRSGKSS IMRSDAPIGKCNSECITPNGSIPNDKPFQNVNRITYGACPRYYVKQNTLKLATG MRNVPEKQTR	ACC6707 9
----	-------------------	---	--------------

The GenBank references of the displayed HA sequences are given, where applicable.

**Table S5. Antigenic clusters of influenza A/H3N2 viruses**

Cluster name	Viruses and GenBank references of corresponding hemagglutinin amino acid sequences
HK68	BI/6022/72 (AAT64674); BI/15793/68 (AY661038); BI/16190/68 (AY661039); BI/16398/68 (AY660991); HK/1/68 (AF201874); BI/808/69 (AAT64715); BI/908/69 (AAT64667); BI/17938/69 (AAT64668); BI/93/70 (AAT64669); BI/2668/70 (AAT64670); BI/6449/71 (AAT64671); BI/21438/71 (AAT64672); BI/21801/71 (AAT64673); HK/107/1971 (ABQ58932)
EN72	BI/7398/74 (AY661027); BI/9459/74 (AY661005); EN/42/72 (AF201875); BI/3517/73 (AAT64679); BI/2600/75 (AAT64718); BI/23337/72 (AAT64676); BI/748/73 (AAT64678); BI/23290/72 (AAT64675); BI/552/73 (AAT64677); BI/5146/74 (AAT64693); BI/2813/75 (AAT64703); PO/1/73 (AAC78096); BI/5931/74 (AAT64692); BI/21793/72 (AAT64716); BI/5930/74 (AAT64717)
VI75	VI/3/75 (V01098); BI/1761/76 (AAT64681); BI/5029/76 (AAT64683); BI/3895/77 (AAT64686); BI/5657/76 (AAT64684); RO/5828/77 (AAT64687); RO/8179/77 (AAT64688); BI/6545/76 (AAT64685); AM/1609/77 (AAT64720)
TX77	BI/628/76 (AY661044); BI/2271/76 (AAT64682); TX/1/77 (AAL62329)
BK79	BK/1/79 (AAF18090); CA/1/84 (AAB69809); NL/233/82 (AAT64700); CH/2/88 (AAB69847); OS/13676/83 (AAT64691); BI/4791/81 (AAT64689); PH/2/82 (AAF60285); LE/360/86 (AAB69845); BI/10684/82 (AAT64690); NL/209/80 (AAT64722); NL/241/82 (AAT64723); WE/4/85 (AAT64725); ST/10/85 (AAT64727); RO/577/80 (AAT64721); GU/V728/85 (AAT64726); NL/330/85 (AAT64724); CO/2/86 (AAT64728)
SI87	SI/02/87 (D10161); SI/35/89 (AY661061); HK/1/89 (AAT64734); NL/738/89 (AAT64704); EN/427/88 (AAT64730); VI/1/89 (AAT64740); NL/650/89 (AAT64743); NL/450/88 (AAT64729); ST/12/88 (AAT64731); GE/5007/89 (AAT64732); SI/40/89 (AAT64738); SI/53/89 (AAT64739); ME/2/90 (AAT64744); AT/21/89 (AAT64746); WE/5/89 (AAT64733); SI/34/89 (AAT64735); NL/620/89 (AAT64742); ME/5/90 (AAT64745); CA/1/91 (AAT64751); SH/11/1987 (AF008886); GU/54/89 (BAA01026); SI/36/89 (AAT64737); SH/24/90 (AAT64749); SU/1/90 (Smith (et al.); VI/7/87 (AF008888))
BE89	SO/23/92 (AY661127); SO/27/92 (AY661128); SO/8/92 (AY661126); ST/7/92 (AY661120); ST/8/92 (AY661121); MA/G12/91 (AY661079); NL/816/91 (AY661078); ST/13/92 (AY661137); ST/20/91 (AY661119); TI/5957/92 (AY661116); UM/1982/92 (AY661134); UM/2000/92 (AY661135); VI/2/90 (AY661073); VI/68/92 (AF008697); BE/352/89 (D49961); BE/353/89 (AF008684); PA/320/92 (AAT64757); PA/407/92 (AAT64759); PA/548/92 (AAT64766); LY/1149/91 (AAT64773); LY/ONS1594/91 (AAT64776); LY/ONS1276/91 (AAT64778); LY/CHU23672/91 (AAT64780); LY/CHU24222/91 (AAT64782); RO/100540/92 (AAT64787); NI/3129/92 (AAT64789); HK/56829/92 (AAT64797); HK/56798/92 (AAT64799); PA/287/93 (AAT64807); MA/G116/93 (AAT64817); NL/819/92 (AAT64755); EN/261/91 (AAT64756); LY/ONS1189/91 (AAT64775); NL/823/92 (AAT64783); FI/218/92 (AAT64784); AM/4112/92 (AAT64790); MA/G58/92 (AAT64792); MA/OV31/92 (AAT64793); HK/56941/92 (AAT64798); EN/5458/93 (AAT64825); GE/6447/91 (AAT64752); PA/325/92 (AAT64758); PA/417/92 (AAT64760); PA/512/92 (AAT64765); PA/564/92 (AAT64767); NL/935/92 (AAT64772); LY/1182/91 (AAT64774); LY/ONS1373/91 (AAT64777); LY/ONS1337/91 (AAT64779); LY/CHU24103/91 (AAT64781); NI/3126/92 (AAT64788); EN/260/91 (AAT64750); PA/424/92 (AAT64761); PA/457/92 (AAT64762); PA/467/92 (AAT64763); PA/490/92 (AAT64764); PA/583/92 (AAT64768); PA/597/92 (AAT64769); PA/614/92 (AAT64770); EN/1285/92 (AAT64771); FI/220/92 (AAT64785); GE/5113/92 (AAT64786); MA/G101/93 (AAT64818); NL/165/93 (AAT64821)
BE92	SC/160/93 (Z46414); HK/2/94 (Z46408); HK/55/94 (AF008773); HK/56/94 (AY661178); JO/33/94 (AF008774); JO/47/94 (AY661179); NL/18/94 (AY661020); SC/142/93 (Z46413); SO/15/94 (AY661176); SO/25/94 (AY661177); YA/62/93 (AY661154); HK/1/94 (AAB63709); GU/25/93 (AAB63765); HK/3/95 (AAB63692); HK/42/96 (AAB63706); NL/101/93 (AAT64814); MA/G122/93 (AAT64822); NL/179/93 (AAT64824); YA/61/93 (AAT64828); NL/440/93 (AAT64840); LY/23602/93 (AAT64847); OS/2352/93 (AAT64849); NL/1/95 (AAT64856); HK/32/95 (AAT64860); HK/38/95 (AAT64861); NL/17/93 (AAT64808); ST/12/92 (AAT64811); FI/247/92 (AAT64813); LY/1803/93 (AAT64844); VI/75/95 (AAT64863); NL/938/92 (AAT64800); NL/3/93 (AAT64806); NL/115/93 (AAT64815); NL/126/93 (AAT64816); MA/G130/93 (AAT64823); YA/56/93 (AAT64827); SH/9/93 (AAT64834); NL/398/93 (AAT64841); LY/672/93 (AAT64843); LY/1815/93 (AAT64845); OS/2219/93 (AAT64848); SE/C273/92 (AAT64804); BE/32/92 (AAT64805); MA/G102/93 (AAT64819); MA/G109/93 (AAT64820); AK/4/93 (AAT64830); SH/6/93 (AAT64831); VI/104/93 (AAT64832); WE/59/93 (AAT64833); NL/241/93 (AAT64835); ST/20/93 (AAT64836); NL/357/93 (AAT64837); NL/371/93 (AAT64839); LY/22686/93 (AAT64846); NL/91/96 (AAT64869); HK/434/96 (AAF16422); SI/3/93 (AAT64886)
WU95	LY/1781/96 (AF131998); OS/21/97 (AF368444); OS/244/97 (AF368446); SI/1/96 (AY661199); EN/7/94 (Z46405); FI/338/95 (AF368436); FI/339/95 (AY661184); FI/381/95 (AY661196); GE/AI9509/95 (AY661182); HK/358/96 (AF180602); HK/1/97 (AY661202); HK/49/95 (AY661187); HK/55/95 (AY661189); LY/2279/95 (AY661192); NA/933/95 (AF008725); NL/271/95 (AY661191); NL/5/98 (AY661209); WU/359/95 (AF008722); HK/357/96 (AAB66778); BR/8/96 (AAB66800); JO/10/97 (AAF16495); MA/G252/93 (AAT64826); HK/20/96 (AAT64868); NL/372/93 (AAT64838); HK/280/97 (AAT64878); NI/491/97 (AAT64885); NL/399/93 (AAT64842); GE/3958/96 (AAT64870)
SY97	AU/10/97 (AY661201); NL/300/97 (AY661205); SY/5/97 (AF180584); NL/124/01 (AY661023); NL/126/01 (AY661024); NL/414/98 (AY661206); NL/427/98 (AY661207); NL/118/01 (AAT64697); NL/3/00 (AAT64696); NL/301/99 (AAT64701); NL/120/02 (AAT64705); NL/1/02 (AAT64706); NL/20/03

**Table S5. Continued**

	(AAT64708); NL/462/98 (AAT64883); PA/2007/1999 (ABE73114); MO/10/99 (ABU50516)
FU02	NL/222/03 (AAT64712); FI/170/03 (AAT64707); NL/22/03 (AAT64709); NL/213/03 (AAT64710); NL/217/03 (AAT64711); FU/411/02 (ABH01014)

The identities of viruses belonging to each antigenic cluster were taken from Fig. S1 of ref. 19.