

# A Likelihood-based Index of Protein–Protein Binding Affinities with Application to Influenza HA Escape from Antibodies

Teruaki Watabe,\* Hirohisa Kishino,† Leonardo de Oliveira Martins,† and Yasuhiro Kitazoe\*

\*Center of Medical Information Science, Kochi University, Kochi, Japan; and †Laboratory of Biometrics, Graduate School of Agriculture and Life Science, University of Tokyo, Tokyo, Japan

In many biological systems, proteins interact with other organic molecules to produce indispensable functions, in which molecular recognition phenomena are essential. Proteins have kept or gained their functions during molecular evolution. Their functions seem to be flexible, and a few amino acid substitutions sometimes cause drastic changes in function. In order to monitor and predict such drastic changes in the early stages in target populations, we need to identify patterns of structural changes during molecular evolution causing decreases or increases in the binding affinity of protein complexes. In previous work, we developed a likelihood-based index to quantify the degree to which a sequence fits a given structure. This index was named the sequence-structure fitness (SSF) and is calculated empirically based on amino acid preferences and pairwise interactions in the structural environment present in template structures. In the present work, we used the SSF to develop an index to measure the binding affinity of protein–protein complexes defined as the log likelihood ratio, contrasting the fitness of the sequences to the structure of the complex and that of the uncomplexed proteins. We applied the developed index to the complexes formed between influenza A hemagglutinin (HA) and four antibodies. The antibody–antigen binding region of HA is under strong selection pressure by the host immune system. Hence, examination of the long-term adaptation of HA to the four antibodies could reveal the strategy of the molecular evolution of HA. Two antibodies cover the HA receptor-binding region, while the other two bind away from the receptor-binding region. By focusing on branches with a significant decline in binding ability, we could detect key amino acid replacements and investigate the mechanism via conditional probabilities. The contrast between the adaptations to the two types of antibodies suggests that the virus adapts to the immune system at the cost of structural change.

## Introduction

Many biological functions are predominantly controlled by protein–protein interactions, and molecular recognition phenomena are essential to biological systems. These recognition phenomena involve the association of proteins to ligands or substrates. One of the most important molecular recognitions is that performed between lymphocytes and major histocompatibility complex (MHC) class I molecules. Natural killer (NK) cells—one of the lymphocyte classes—detect downregulation of MHC class I molecules by means of specific membrane receptors. A main category of these receptors is the killer cell immunoglobulin-like receptor (KIR) family. These KIR genes have evolved in primates to generate a diverse family of receptors with unique structures that enable them to recognize MHC class I molecules with locus and allele specificity (Vilches and Parham 2002). Their combinatorial expression creates a repertoire of NK cells that antagonize the spread of pathogens and tumors.

Proteins have evolved, keeping their functions or newly gaining other functions. For example, globins arose very early in evolution and are found in a wide range of organisms. The globins have maintained the ability to interact with each other to enable cooperative oxygen binding, and they show functional flexibility and realize oxygen binding in a number of ways. While vertebrate hemoglobins have evolved, conserving their ability to bind ligands and deliver oxygen molecules bound at the heme sites, many functionally important structural differences were found to exist among vertebrate hemoglobins (Naoi et al.

2001). Functional flexibility appears to be a distinctive feature of protein evolution.

A few amino acid substitutions sometimes cause drastic changes to the protein function or to its influence in a system. The severe acute respiratory syndrome (SARS) coronavirus is one of the most well-studied viruses. It was caught in the act of adapting to humans, and the viral spike glycoprotein was identified as a major determinant of the species' specificity of coronavirus infection. Only four amino acids in the receptor-binding domain differ between the human epidemic and the civet strains, but they cause more than a 1,000-fold difference in the binding affinity to the human angiotensin-converting enzyme 2, a specific receptor glycoprotein on the surface of host cells (Li et al. 2005). Adaptation of a virus to a homologous receptor of a new host species appears to require very few amino acid substitutions at the large receptor-binding interface (Holmes 2005).

Measuring the binding affinity of complexes formed between biological molecules is indispensable to monitoring and predicting adaptive evolution in a target population. Particularly in the arms race of the host–parasite system, binding affinity plays a crucial role. Watabe et al. (2006) developed a likelihood-based index, named sequence-structure fitness (SSF), which quantifies the degree to which a sequence fits in the given structure. The SSF is calculated empirically based on amino acid preferences and pairwise interactions in the structural environment present in templates' structure. In the present study, we used the SSF to develop an index to measure the binding affinity of protein–protein complexes defined as the log likelihood ratio, contrasting the fitness of the sequences to the structure of the complex and that of the uncomplexed proteins. This index enables us to quantify the binding affinity between proteins. We applied the developed index to systems of complexes between virus proteins containing epitopes (influenza A hemagglutinin [HA]) and antibodies of the host immune system.

Key words: protein structure, affinity of complex, influenza virus, antigenic drift, functional redundancy.

E-mail: twatabe-mi@umin.ac.jp.

*Mol. Biol. Evol.* 24(8):1627–1638. 2007

doi:10.1093/molbev/msm079

Advance Access publication May 2, 2007

We followed an evolutionary process of the evasion of HA from antibodies and analyzed the binding ability of the antibodies to HA at each node of the phylogenetic dendrogram of the HA sequence. On the one hand, we found that the binding ability to the HA surface of antibodies whose neutralization abilities are supplied by the indirect mechanism of blocking virus attachment decreased but was never completely lost. On the other hand, antibodies covering the receptor-binding region of HA by direct binding to the receptor binding sites completely lost their binding abilities along the HA evolutionary pathway. We also found that the binding ability of these antibodies to HA started to increase after a couple of years in an impotent stage, and this binding ability was restored to a magnitude comparable to the original, prompting HA to evade them again.

## Materials and Methods

### A Likelihood-based Index of the Binding Ability of Protein–Protein Complexes

To trace long-term changes in interactions between proteins, measuring the affinity of the complexes for many sequences along the evolutionary tree is indispensable. High-throughput experiments to directly measure affinity are currently impractical. Therefore, we estimated affinity by using structural information for the templates and information on amino acid preferences and pairwise interactions in the structural environment in the Protein Data Bank (PDB).

The binding ability of the complexes formed between proteins A and B is measured as the inverse of the dissociation constant of the complex:

$$\frac{1}{K_d} = \frac{[A \cdot B]}{[A][B]},$$

where  $[A]$ ,  $[B]$  and  $[A \cdot B]$  are the concentrations of proteins A and B and their complexes in equilibrium. To predict the ratio using information from the protein database, we interpret the ratio as reflecting the likelihood ratio

$$\frac{P(A \cdot B)}{P(A)P(B)}.$$

The binding ability can be characterized by the amino acid sequences in the binding region and the structure of the complexes. Hence, the likelihood is explained by the likelihood of the amino acid sequences given the structure:

$$\frac{P(A \cdot B)}{P(A)P(B)} = \frac{P(\text{seq}_A, \text{seq}_B | \text{str}_{A+B})P(\text{str}_{A+B})}{P(\text{seq}_A | \text{str}_A)P(\text{str}_A)P(\text{seq}_B | \text{str}_B)P(\text{str}_B)}.$$

If the structures were stable during evolution of the proteins, it could be thought that the likelihoods  $P(\text{str}_{A+B})$ ,  $P(\text{str}_A)$ , and  $P(\text{str}_B)$  were constant. Assuming stability of the structures during evolution of the proteins, we defined the index of the affinity of the complex as the log likelihood ratio of the amino acid sequences in the binding region of the complex to those in the binding region of the uncomplexed proteins (fig. 1),

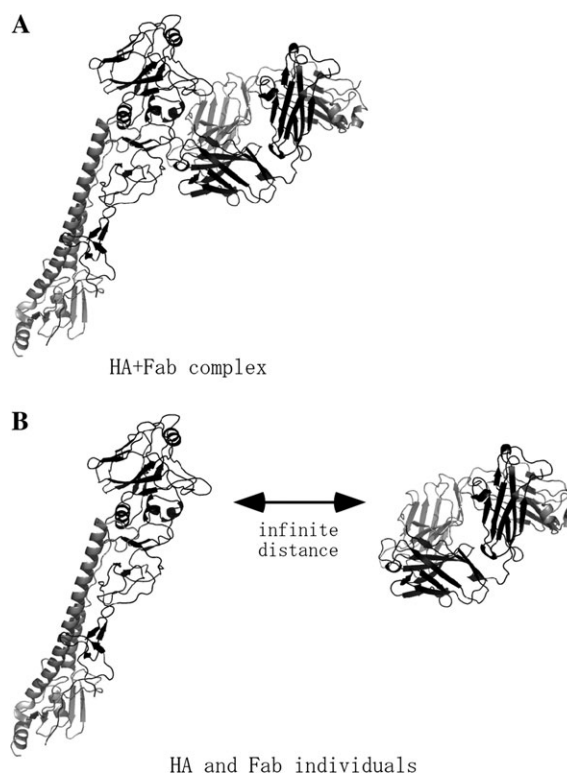


FIG. 1.—A, The complex formed between HA and fragment antigen binding (Fab) HC45, and (B) the uncomplexed individual proteins. In the analysis, it was considered that the individual uncomplexed proteins are far from each other.

$$[\text{affinity}] \approx \log \frac{P(\text{seq}_A, \text{seq}_B | \text{str}_{A+B})}{P(\text{seq}_A | \text{str}_A)P(\text{seq}_B | \text{str}_B)}, \quad (1)$$

which would be interpreted as  $-\log K_d$ .

The sequence distribution of a structure  $P(\text{seq} | \text{str})$  in Equation 1 expresses the fitness of the amino acids given the structural environment. To predict the protein structure of an amino acid sequence, Simons et al. (1999a, 1999b) proposed to maximize:

$$P(a_1, \dots, a_n | \mathbf{X}) \cong \prod_i P(a_i | \mathbf{X}) \prod_{i < j} \frac{P(a_i, a_j | \mathbf{X})}{P(a_i | \mathbf{X})P(a_j | \mathbf{X})}. \quad (2)$$

Here, an amino acid sequence and the positions of the  $C^\alpha$  atoms of amino acid residues are denoted by  $\mathbf{A} = (a_1, \dots, a_n)$  and  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ , respectively. The first term represents the amino acid preference and the second term is the pairwise interaction given the structure. Expecting that the fitness of an amino acid depends mostly on the local structure surrounding each of the amino acid residues, Simons et al. (1999a, 1999b) categorized local environments and calculated amino acid frequencies in the proteins registered in PDB for each of the categories (see below for more detail).

Watabe et al. (2006) obtained evidence that implies that the structure of the crown region of the HIV V3 loop varies more in the patient where the viral population experienced larger sequence evolution. They traced the value of

$P(\text{seq}|\text{str})$  based on the information of the sequences and the structure of the templates in PDB. Assuming that the structures of the templates were not affected largely by immune systems, this measure expresses the extent of the structural change after infection indirectly. In the present study, we extended this to estimate the adaptive evolution via the change in affinity of the protein–protein complex by considering the log likelihood ratio (Equation 1).

A molecular dynamics approach with careful parameter specification may be useful. It is unclear at present how the errors in structural prediction may cause bias in the estimation of the structural variation if structures are predicted sequence by sequence. Our indirect approach has another advantage. By bypassing the step of structural prediction, it is possible to analyze thousands of sequences and trace structural changes or sequence preferences to the structure along evolutionary pathways on phylogenetic trees.

In Equation 2, the SSF was approximated by the first two terms appearing in the expansion formula (Watabe et al. 2006). The logarithm of the SSF could be correlated with the free energy of the protein:

$$\begin{aligned} \log[\text{SSF}] = & -\frac{E}{kT} \cong \sum_i \log P(a_i|\mathbf{X}) \\ & + \sum_{i < j} \log \frac{P(a_i, a_j|\mathbf{X})}{P(a_i|\mathbf{X})P(a_j|\mathbf{X})}. \end{aligned} \quad (2)$$

Here,  $k$  is the Boltzmann constant and  $T$  is the temperature. Ignoring the higher correlation between amino acid residues, we obtained the terms for the solvent accessibility and pairwise measures of sequence–structure compatibility. Rather than being the actual energy potential, the specific value of  $E/kT$  is derived from statistical analysis of a large number of known structures (Jones, Taylor, and Thornton 1992; Jones 1999; Robinson et al. 2003).

The numerator in Equation 1 is decomposed into three terms:

$$P(\text{seq}_A, \text{seq}_B|\text{str}_{A+B}) = P(\text{seq}_A|\text{str}_{A+B})P_{\text{int}}P(\text{seq}_B|\text{str}_{A+B}). \quad (3)$$

Here,  $P_{\text{int}}$  represents the interaction between proteins A and B. Combining Equations 1 and 3, the affinity of the complex is described by three terms. Using the approximation of Equation 2, the three terms are formulated as follows. The first term is the likelihood ratio of the protein A sequence in the complex to the protein A sequence in uncomplexed proteins:

$$\begin{aligned} \frac{P(\text{seq}_A|\text{str}_{A+B})}{P(\text{seq}_A|\text{str}_A)} \cong & \prod_{i \in \{A\}} \frac{P(a_i^{(A)}|\mathbf{X}^{(A+B)})}{P(a_i^{(A)}|\mathbf{X}^{(A)})} \\ \times \prod_{i, j \in \{A\}, i < j} & \frac{P(a_i^{(A)}, a_j^{(A)}|\mathbf{X}^{(A+B)})}{P(a_i^{(A)}|\mathbf{X}^{(A+B)})P(a_j^{(A)}|\mathbf{X}^{(A+B)})} \frac{P(a_i^{(A)}|\mathbf{X}^{(A)})P(a_j^{(A)}|\mathbf{X}^{(A)})}{P(a_i^{(A)}, a_j^{(A)}|\mathbf{X}^{(A)})}. \end{aligned} \quad (4)$$

This represents the structural preference of the protein A sequence to the complex compared with the uncomplexed

proteins. The second term expresses the interactions between proteins A and B:

$$P_{\text{int}} \cong \prod_{i \in \{A\}, j \in \{B\}} \frac{P(a_i^{(A)}, a_j^{(B)}|\mathbf{X}^{(A+B)})}{P(a_i^{(A)}|\mathbf{X}^{(A+B)})P(a_j^{(B)}|\mathbf{X}^{(A+B)})}. \quad (5)$$

The third term is the structural preference of the protein B sequence denoted by the likelihood ratio of the protein B sequence in the complex to the protein B sequence in the uncomplexed proteins:

$$\begin{aligned} \frac{P(\text{seq}_B|\text{str}_{A+B})}{P(\text{seq}_B|\text{str}_B)} \cong & \prod_{i \in \{B\}} \frac{P(a_i^{(B)}|\mathbf{X}^{(A+B)})}{P(a_i^{(B)}|\mathbf{X}^{(B)})} \\ \times \prod_{i, j \in \{B\}, i < j} & \frac{P(a_i^{(B)}, a_j^{(B)}|\mathbf{X}^{(A+B)})}{P(a_i^{(B)}|\mathbf{X}^{(A+B)})P(a_j^{(B)}|\mathbf{X}^{(A+B)})} \frac{P(a_i^{(B)}|\mathbf{X}^{(B)})P(a_j^{(B)}|\mathbf{X}^{(B)})}{P(a_i^{(B)}, a_j^{(B)}|\mathbf{X}^{(B)})}. \end{aligned} \quad (6)$$

We also defined the fitness of a single amino acid residue (FSR) in protein A to the complex by the contribution from the corresponding residue to Equations 4 and 5:

$$\begin{aligned} \text{FSR}(a_i^{(A)}) = & \frac{P(a_i^{(A)}|\mathbf{X}^{(A+B)})}{P(a_i^{(A)}|\mathbf{X}^{(A)})} \\ \times \prod_{j \in \{A\}, j \neq i} & \frac{P(a_i^{(A)}, a_j^{(A)}|\mathbf{X}^{(A+B)})}{P(a_i^{(A)}|\mathbf{X}^{(A+B)})P(a_j^{(A)}|\mathbf{X}^{(A+B)})} \frac{P(a_i^{(A)}|\mathbf{X}^{(A)})P(a_j^{(A)}|\mathbf{X}^{(A)})}{P(a_i^{(A)}, a_j^{(A)}|\mathbf{X}^{(A)})} \\ \times \prod_{j \in \{B\}} & \frac{P(a_i^{(A)}, a_j^{(B)}|\mathbf{X}^{(A+B)})}{P(a_i^{(A)}|\mathbf{X}^{(A+B)})P(a_j^{(B)}|\mathbf{X}^{(A+B)})}. \end{aligned} \quad (7)$$

## Contributing Residues and Background Residues

The residues in the binding regions are identified as those with structural environments in the complex different from the environments in the uncomplexed proteins. Here, the environment of a residue is defined as the number of  $C^\alpha$  atoms surrounding the  $C^\alpha$  atom of the corresponding residue within a distance of 1 nm which expresses the degree of residue burial. We consider the identified residues as the “contributing residues” to make a complex of protein A and B. Equation 5 includes interactions between the contributing residues of protein A and the contributing residues of B. The pairs with distances  $\leq 1$  nm are included.

The first term on the right-hand side of Equation 4 (Equation 6) reflects the changes in the residue environment on protein A (B) and is evaluated with contributing residues. To examine the effect of making a complex as accurately as possible, the second term includes not only the interactions between the contributing residues but also the interactions between the contributing residues and the surrounding residues inside the protein, termed the “background residues.” Here also, the pairs with distances  $\leq 1$  nm are included.

## Calculation of Amino Acid Preferences and Pairwise Interactions

The amino acid preferences and interactions of the amino acid pairs were measured by empirical frequencies

**Table 1**  
**Categories of Pairwise Distances**

	Categories
Pairwise distance (small interval) <sup>a</sup>	0–7, 7–8, 8–9, 9–10
Pairwise distance <sup>a</sup>	0–7, 7–8, 8–9, 9–10

<sup>a</sup> For pairwise distance, each category represents a range of spatial distances 0.1 nm between the C<sup>α</sup> atoms of the corresponding residues. The category,  $x$ – $y$ , corresponds within the range of  $x < (\text{pairwise distance}) \leq y$ .

observed in the nonredundant set of protein structures in the PDB (Hobohm and Sander 1994; the latest library is available from [ftp://ftp.embl-heidelberg.de/pub/databases/protein\\_extras/pdb\\_select/](ftp://ftp.embl-heidelberg.de/pub/databases/protein_extras/pdb_select/)). The local environment of a residue is defined as the number of C<sup>α</sup> atoms surrounding the C<sup>α</sup> atom of the corresponding residue within a distance of 1 nm as explained above. The spatial distances between the C<sup>α</sup> atoms of amino acid pairs are categorized into each 1 nm from 0.7 nm to 1 nm (table 1). For the spatial distance  $\leq 0.7$  nm we compiled them on one category because amino acid pairs with such small distance were rare. The environment class  $E_i$  is defined by the degree of residue burial. The interaction between a residue and its neighborhood in a primary sequence plays a major role in making local conformations such as an  $\alpha$ -helix structure. Hence, we treated separately the interactions between residues with small site intervals  $j - i \leq 4$  from those with larger site intervals (Watabe et al. 2006).

We empirically calculated the conditional probabilities based on frequencies in the nonredundant set of protein structures in the PDB. To obtain a sufficient sample size for the estimation of amino acid preferences and pairwise interactions, Simons et al. (1999a, 1999b) categorized the local environments into a few classes. However, our log likelihood ratio may depend too greatly on the pattern of categorization to reliably estimate the binding ability and its variation. To avoid this problem, we adopted the sliding-window approach to achieve a balance between robust estimation of conditional probabilities and sensitive detection of changes in microstructure. In other words, we estimated the amino acid preference and the pairwise interaction in a local environment using the proportions of the corresponding amino acids or amino acid pairs out of all residues or pairs of residues in a similar environment. More specifically, we estimated a conditional probability given the local environment,  $E_i = n$ , from the proportions of amino acids or amino acid pairs within the local environment,  $n - 8 \leq E_i \leq n + 8$ . Some examples of amino acid preferences and interactions of the amino acid pairs are given in Supplementary Material online. We also show the details of calculation of the binding affinity of the complex between HA (A/Aichi/68) and fragment antigen binding (Fab) HC45 in the Supplementary Material online.

#### Validation of the Predicted Binding Ability

First, we examined if the predicted binding ability is consistent with the measured values for the case of the two antibodies to the influenza HA (Fleury et al. 2000). We further examined the correlation between our index and the measured value for the complex between tumor endothelial marker (TEM)-1  $\beta$ -lactamase and its inhibitory protein. The

hydrolysis of  $\beta$ -lactam antibiotics by class A  $\beta$ -lactamases is a common cause of bacterial resistance to these agents. The  $\beta$ -lactamase inhibitory protein (BLIP) binds and inhibits several class A  $\beta$ -lactamases, including TEM-1  $\beta$ -lactamase. To identify the residues on BLIP that contribute to its binding affinity, Zhang and Palzkill (2003) carried out alanine-scanning mutagenesis and contrasted the binding abilities of 23 mutants and one double mutant with that of the wild type. The X-ray crystallography data for the complex is also available in PDB (PDB code: 1JTG; Lim et al. 2001). Therefore we compared our predicted binding abilities with the measured values (Table 2 in Zhang and Palzkill 2003).

Furthermore we analyzed a set of 746 complex structural data and calculated the affinity of contacting chains in those complexes. The PDB codes of the complex data in this set were listed in the nonredundant set of protein structures (Hobohm and Sander 1994). In the nonredundant set, in most cases one chain was chosen for one PDB code. Some were from complex structures. We selected such chains and added the structural data of the remaining chains to make them complete complex data entries.

#### Application to Complexes Between Influenza HA and Antibodies

We applied the developed index to complexes formed between influenza HA and antibodies, because the co-circulation of multiple genetic lineages of influenza A viruses has been well documented (Cox and Bender 1995; Smith et al. 2004) and X-ray crystallography data for complexes are available. Influenza viruses continue to evolve and new antigenic variants constantly emerge (antigenic drift) (Webster et al. 1992). In the influenza A viruses of humans and other mammals, antibodies play a role in the selection of mutants (Webster 2000). This enables influenza A viruses to evade the host immune system, decreasing the ability to form antibody–antigen complexes. This causes re-infection of hosts with pre-existing antibodies and allows the influenza virus to persist in human populations. Gubareva, Novikov, and Hayden (2002) showed that amino acid changes accompanying transmission among humans accumulated in HA and that these changes might be related to antibody pressure (Lipatov et al. 2004). We measured the affinity of HA protein–antibody complexes and followed-up with affinity changes along a phylogenetic tree of an influenza A virus.

#### Influenza Hemagglutinin Sequences

We analyzed the HA protein sequences from 253 influenza A virus isolates from humans isolated from 1968 to 2003. We also included two sequences from swine and three sequences of avian influenza A in the analysis. The overall structure of the H3 HA protein of a duck virus (A/duck/Ukraine/63, PDB data: 1MQL, Ha et al. 2003) is very similar to that of a human (A/Aichi/68). In fact, the root-mean-square deviation, which measures the average spatial distance of corresponding amino acid residues, was only 0.06 nm between avian HA (1MQL) and human HA (1KEN) after structural alignment. The majority of the sequences (248 sequences) were analyzed by Smith et al.



**Table 2**  
**Epitopes on HA and Their Contacting Residues on HC45 Defined in the Two Schemes**

Fab residues <sup>a</sup> (Distance between C $\alpha$ s [0.1nm])	HA Residues (antigenic site)	Fab Residues <sup>b</sup> (number of atomic contacts)
<u>H53Y</u> <sub>2</sub> (8.9), <u>H54D</u> <sub>2</sub> (10.0)	48T (C)	<u>H53Y</u> <sub>2</sub> (2)
<u>H31T</u> <sub>1</sub> (9.4), <u>H53Y</u> <sub>2</sub> (9.3)	49G	<u>H53Y</u> <sub>2</sub> (2)
<u>H99T</u> <sub>3</sub> (9.1)	50K (C)	<u>H53Y</u> <sub>2</sub> (4), <u>H54D</u> <sub>2</sub> (1)
<u>H99T</u> <sub>3</sub> (7.0), <u>H100T</u> <sub>3</sub> (9.9)	52C	
<u>H99T</u> <sub>3</sub> (8.4), <u>H100T</u> <sub>3</sub> (9.1)	53N (C)	
<u>H99T</u> <sub>3</sub> (8.2), <u>H100T</u> <sub>3</sub> (7.2), <u>H100aI</u> <sub>3</sub> (7.8)	58I	
<u>H99T</u> <sub>3</sub> (7.3), <u>H100T</u> <sub>3</sub> (5.7), <u>H100aI</u> <sub>3</sub> (7.6)	59L (E)	<u>H100aI</u> <sub>3</sub> (1)
<u>H100T</u> <sub>3</sub> (8.3), <u>H100aI</u> <sub>3</sub> (9.2)	60D	<u>H32Y</u> <sub>1</sub> (5), <u>H100T</u> <sub>3</sub> (1)
<u>H27Y</u> (8.4), <u>H28T</u> (9.8), <u>H31T</u> <sub>1</sub> (8.9), <u>H32Y</u> <sub>1</sub> (9.1), <u>H96L</u> <sub>3</sub> (8.8), <u>H100T</u> <sub>3</sub> (7.8), <u>H100aI</u> <sub>3</sub> (9.3), <u>H100bI</u> <sub>3</sub> (9.9)	61G	
<u>H1Q</u> (9.0), <u>H2V</u> (8.4), <u>H26G</u> (8.4), <u>H27Y</u> (7.0), <u>H28T</u> (9.6)	62I (E)	<u>H32Y</u> <sub>1</sub> (6), <u>H94R</u> (1), <u>H96L</u> <sub>3</sub> (1)
<u>L56S</u> <sub>2</sub> (9.6)	63D (E)	<u>H2V</u> (2), <u>H27Y</u> (1), <u>H94R</u> (9)
<u>L55F</u> <sub>2</sub> (9.9), <u>L56S</u> <sub>2</sub> (6.3), <u>L57G</u> (6.3), <u>L58V</u> (9.7)	73D	
<u>L55F</u> <sub>2</sub> (8.5), <u>L56S</u> <sub>2</sub> (5.1), <u>L57G</u> (6.8)	74P	<u>L56S</u> <sub>2</sub> (1)
<u>L55F</u> <sub>2</sub> (9.6), <u>L56S</u> <sub>2</sub> (7.0), <u>L57G</u> (9.5)	75H (E)	<u>L56S</u> <sub>2</sub> (1)
<u>L54R</u> <sub>2</sub> (9.2), <u>L55F</u> <sub>2</sub> (8.2), <u>L56S</u> <sub>2</sub> (6.2), <u>L57G</u> (8.3)	76C	
<u>H100aI</u> <sub>3</sub> (8.0), <u>H100bI</u> <sub>3</sub> (8.4), <u>L53N</u> <sub>2</sub> (8.4), <u>L54R</u> <sub>2</sub> (7.7), <u>L55F</u> <sub>2</sub> (6.4), <u>L56S</u> <sub>2</sub> (6.0), <u>L57G</u> (9.3)	77D	
<u>H100aI</u> <sub>3</sub> (8.2), <u>L56S</u> <sub>2</sub> (9.6)	78V (E)	<u>H100aI</u> <sub>3</sub> (4), <u>H100bI</u> <sub>3</sub> (2), <u>L49Y</u> (4)
<u>H100aI</u> <sub>3</sub> (9.5)	79F	<u>H100aI</u> <sub>3</sub> (2)
<u>H27Y</u> (8.2), <u>H28T</u> (8.3), <u>H31T</u> <sub>1</sub> (8.9)	82E (E)	
<u>H27Y</u> (8.1), <u>H28T</u> (8.1)	90R	<u>H31T</u> <sub>1</sub> (3), <u>H32Y</u> <sub>1</sub> (3)
<u>H1Q</u> (9.5), <u>H2V</u> (9.3), <u>H25S</u> (9.4), <u>H26G</u> (6.6), <u>H27Y</u> (5.3), <u>H28T</u> (6.9)	91S (E)	
<u>H1Q</u> (9.1), <u>H2V</u> (9.5), <u>H26G</u> (8.1), <u>H27Y</u> (7.1), <u>H28T</u> (9.5)	92K (E)	<u>H26G</u> (2), <u>H27Y</u> (2), <u>H28T</u> (2)
<u>H1Q</u> (7.2), <u>H2V</u> (9.1), <u>H26G</u> (8.3), <u>H27Y</u> (9.0)	93A	
<u>H1Q</u> (8.6)	94F (E)	<u>H1Q</u> (10), <u>H2V</u> (1)
<u>H1Q</u> (7.6)	95S	
<u>L57G</u> (8.5)	96N (D)	
<u>L56S</u> <sub>2</sub> (8.3), <u>L57G</u> (5.4), <u>L58V</u> (8.1), <u>L59P</u> (8.6)	140K (A)	
<u>L56S</u> <sub>2</sub> (8.3), <u>L57G</u> (4.7), <u>L58V</u> (6.3), <u>L59P</u> (7.3)	141R	
<u>L57G</u> (8.4), <u>L58V</u> (9.2), <u>L59P</u> (9.7), <u>L81E</u> (8.8)	142G (A)	
<u>L57G</u> (9.8)	143P (A)	<u>L81E</u> (1)
<u>H31T</u> <sub>1</sub> (9.2)	144G (A)	
<u>H27Y</u> (9.0), <u>H28T</u> (6.8), <u>H29L</u> (9.3), <u>H30T</u> (7.8), <u>H31T</u> <sub>1</sub> (5.7), <u>H32Y</u> <sub>1</sub> (9.3)	270S	
<u>H28T</u> (9.2), <u>H30T</u> (7.7), <u>H31T</u> <sub>1</sub> (5.7), <u>H32Y</u> <sub>1</sub> (9.1), <u>H53Y</u> <sub>2</sub> (8.9)	271D	<u>H28T</u> (2), <u>H31T</u> <sub>1</sub> (7)
<u>H28T</u> (10.0), <u>H30T</u> (7.2), <u>H31T</u> <sub>1</sub> (4.7), <u>H32Y</u> <sub>1</sub> (6.6), <u>H33W</u> <sub>1</sub> (9.3), <u>H52D</u> <sub>2</sub> (9.7), <u>H52aP</u> <sub>2</sub> (9.9), <u>H53Y</u> <sub>2</sub> (7.8), <u>H97Q</u> <sub>3</sub> (9.2), <u>H99T</u> <sub>3</sub> (9.1)	272A	<u>H31T</u> <sub>1</sub> (1)
<u>H31T</u> <sub>1</sub> (7.9), <u>H32Y</u> <sub>1</sub> (8.7), <u>H97Q</u> <sub>3</sub> (8.3), <u>H98I</u> <sub>3</sub> (9.1), <u>H99T</u> <sub>3</sub> (6.3), <u>H100T</u> <sub>3</sub> (8.3)	273P (C)	<u>H30T</u> (1), <u>H53Y</u> <sub>2</sub> (2)
<u>H97Q</u> <sub>3</sub> (8.4), <u>H98I</u> <sub>3</sub> (8.5), <u>H99T</u> <sub>3</sub> (6.7), <u>H100T</u> <sub>3</sub> (9.9)	274I	<u>H99T</u> <sub>3</sub> (9)
<u>H97Q</u> <sub>3</sub> (9.5), <u>H98I</u> <sub>3</sub> (8.1), <u>H99T</u> <sub>3</sub> (6.9)	275D (C)	
	276T (C)	

<sup>a</sup> Residues on HC45 Fab of which the C $\alpha$  atom is within 1 nm from C $\alpha$  atoms of residues on HA. The amino acids are represented by the one-letter code appearing after the residue number, and the chain is indicated by "H" for heavy chain or "L" for light chain. The residues in the complementarity determining regions (CDRs) are underlined, with a subscript indicating the distinction of the CDR.

<sup>b</sup> Residues on fragment antigen binding (Fab) HC45 of which atoms make contact with atoms of residues on HA within 0.4 nm.

(2004), and 10 additional sequences were extracted from GenBank<sup>TM</sup> (Accession numbers: AF008665, AF008697, AF008711, AF008725, AF008755, AF008769, AF008828, AF008867, AF008886, AF008888, AF008903, AF008905, AF092062, AF131997, AF180570, AF180602, AF180643, AF201842, AF201874, AF201875, AF368444, AF368446, AJ311454, AY531037, AY660991-1018, AY661020-211, D21173, D21183, D49961, J02132, K03335, M16739, U08858, V01089, Z46405, Z46408, Z46413, Z46414, and the sequences from 1KEN and 1MQL in PDB). The amino acid replacements of the contributing and background residues of HA were reconstructed using the maximum parsimony method (PAUP<sup>\*</sup>; Swofford 2003). The tree was searched under topological constraints with the optimality criteria of "parsimony." The tree topology was the maximum-likelihood tree topology (PHYMLIP; Felsenstein 2005) of the whole amino acid sequences of hemagglutinin HA1.

### Template Structures of Complexes Between Hemagglutinin and Antibody Fragments

Virus entry steps begin with attachment to cell-surface receptors (Dimitrov 2004). The influenza A virus entry protein is the HA glycoprotein, and hence HA is a target of attack by the immune system. Several different neutralization mechanisms have been described (Knossow et al. 2002). In general, an antibody binds to the surface of the membrane-distal domain of HA. Knossow and collaborators have determined the structures of complexes formed between the influenza A (H3N2) virus (A/Aichi/68) HA and the monoclonal antibodies HC19 and HC63 (PDB code: 2VIR and 1KEN; Barbey-Martin et al. 2002; Bizebard et al. 1995). While bound HC63 Fab extends from HA within the space projected radially from the HA trimer, HC19 Fab binds on the side of the HA trimer (Knossow et al. 2002). Both antibodies cover the receptor binding region of HA. Fleury

et al. (2000) described the structures of complexes formed between HA of the influenza A (H3N2) virus (A/Aichi/68) and the antibody fragments of HC45 and BH151 (PDB code: 1QFU and 1EO8). Both HC45 and BH151 Fabs bind not to the receptor-binding region but to a region closer to the membrane. However, their large sizes make these antibodies highly efficient at neutralization (Fleury et al. 1999). The complex structural data for HC45 and BH151 suggested that definite flexibility exists in the selection of antibodies that bind to a given epitope (Fleury et al. 2000). Although variable domains of antibodies HC45 and BH151 display only 56% sequence identity, they bind to very closely related epitopes on HA with nearly identical dissociation constants.

The mechanisms for neutralizing virus infectivity by these four antibodies represent two particular types: one is by directly covering the receptor-binding region and the other is by indirect blocking, owing to the large size of the antibody. By using X-ray crystallography data for these four antibodies, we were able to compare the two different types of neutralization and analyze the mechanisms of complex formation. We used these four structural data sets of complexes as templates to estimate the binding ability between HA of viruses isolated from 1968 to 2003 and the four antibody fragments. Here, we assumed that the main chain structure of HA has been conserved during the evolution of the influenza virus.

The complementarity determining regions (CDRs) of HC19, HC45, HC63, and BH151 were identified through structural alignment with a Fab (PDB code: 1FDL) whose CDRs were well identified (Padlan, Abergel, and Tipper 1995).

## Results

### The Case of Two Antibodies to the Influenza HA

The estimated binding abilities of the antibody fragments HC45 and BH151 to the HA glycoprotein of influenza A/Aichi/68 were 3.20 for HC45 and 1.38 for BH151. We cannot compare these values directly to the dissociation constants because we ignored the term that is constant during the evolution of HA but that depends on the conformation of the complex. Nevertheless, we found that the estimated binding abilities were consistently proportionate to the dissociation constants of the complexes ( $K_d = 3.4 \pm 1.5$  nM for HC45 and  $8.9 \pm 2$  nM for BH151; Fleury et al. 2000).

### TEM-1 $\beta$ -Lactamase and Its Inhibitory Protein

Figure 2 shows the correlation between the predicted binding affinity and the value measured by experiment for the complex between enzyme TEM-1  $\beta$ -lactamase and its inhibitory protein. It should be noted that the value of experiment depends on the scale (in this case, the scale is nM). Hence, only the relative relations are informative. The data points represent 23 single amino acid mutations and one double amino acid mutation. The structural data (PDB code: 1JTG) included two measurements. Pairs connected by solid lines represent the variation of our binding affinity caused by the variation of structural measurement. The root-mean-square deviations between the two structural data were 0.02 nm for TEM-1  $\beta$ -lactamase and 0.01 nm

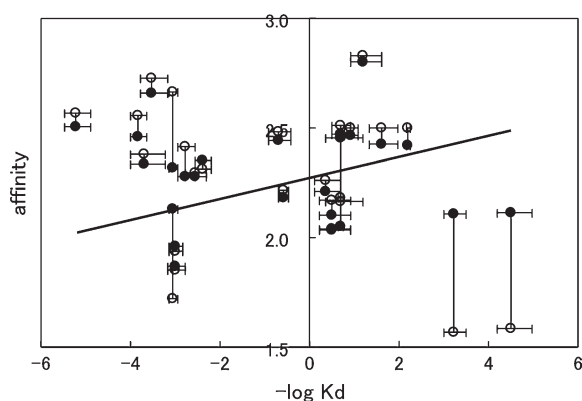


FIG. 2.—The correlation between the calculated affinity and the measured dissociation constant. The results of the analysis with the chains A and B in 1JTG data (solid circles) and those with the chains C and D (open circles) are plotted, and each set of two data points of corresponding mutated inhibitory protein is connected by solid line. The error bars were obtained by the experiments. The bold line represents the result of regression analysis with weighted least squares. The relation  $[\text{affinity}] = \alpha + \beta(-\log K_d)$  was fitted, where  $K_d$  is the dissociation constant measured by the experiment. The correlation coefficient was obtained as 0.33 and the regression coefficients were  $\alpha = 2.27(0.04)$  and  $\beta = 0.048(0.020)$  (the numbers in parentheses are standard errors). The  $P$  values were  $1.29 \times 10^{-43}$  for  $\alpha$  and 0.018 for  $\beta$ .

for the inhibitory protein. Although the differences between two structural data were small, amino acid mutations on some residues caused non-negligible differences between the calculations with the two structural data. In order to take account of those variables, we performed the regression analysis with weighted least squares to estimate the relation,

$$[\text{affinity}] = \alpha + \beta(-\log K_d). \quad (7)$$

Here  $K_d$  is the dissociation constant measured by the experiment. The best weight should be inversely proportional to the variance of the error term. Therefore we defined the weight as the inverse of squared difference between the two predicted values of the binding affinity. We observed positive correlation ( $r = 0.33$ ). The regression coefficients were  $\alpha = 2.27(0.04)$  and  $\beta = 0.048(0.020)$  (the numbers in parentheses are standard errors). The  $P$  values were obtained as  $1.29 \times 10^{-43}$  for  $\alpha$  and 0.018 for  $\beta$ . We obtained the positive correlation between the results of our method and the measured values by the experiments.

### Binding Ability and the Size of Binding Region

We also evaluated the method by analyzing a set of 746 complex structural data. In figure 3, the dependence of the calculated affinity on the number of contacting sites (corresponding to sites at which contributing residues as defined in the *Materials and Methods* were located) is shown. The affinity was scaled by the number of contacting sites. Here, the number of contacting sites is a sum of contacting sites for each chain. The complexes between HA and the Fab had contacting sites numbering about 70. The affinity per contacting site increases with the size of the binding region. Therefore the total binding affinity becomes increasingly stronger. By nonlinear regression analysis of

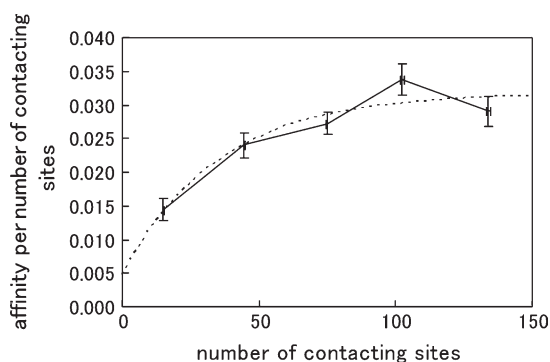


FIG. 3.—The dependence of the calculated affinity on the number of contacting sites is shown. The affinity was scaled by the number of contacting sites. The number of contacting sites was categorized (2 to 30, 31 to 60, 61 to 90, 91 to 120, and 121 to 150). The average of the number of contacting sites and the average affinity in each category are shown, with the error bars representing the standard errors. The dashed line represents the result of nonlinear regression analysis. The relation  $[\text{affinity}] = n_{\text{cs}}(a - b \exp(-cn_{\text{cs}}))$  was fitted, where  $n_{\text{cs}}$  is the number of contacting sites and the parameters  $a=0.0317(0.0035)$ ,  $b=0.0268(0.0074)$ , and  $c=0.0287(0.0190)$  were obtained (the numbers in parentheses are standard errors).

the size dependency, we fitted the relation,

$$[\text{affinity}] = n_{\text{cs}}(a - b \exp(-cn_{\text{cs}})), \quad (7)$$

where  $n_{\text{cs}}$  is the number of contacting sites, and the parameters obtained were  $a=0.0309(0.0031)$ ,  $b=0.0209(0.0033)$ , and  $c=0.0222(0.0095)$  (again, the numbers in parentheses are standard errors). Our method can also evaluate the relationship between the structural environment and the amino acid sequence.

#### Amino Acid Replacements at Residues Contributing to the Interaction Between HA and the Fab

As shown in the Supplementary Material online, we showed the phylogenetic dendrograms of amino acid replacement at the contributing and background residues of HA for each of the four complexes (figure S1 in the online Supplementary Material). The numbers of sites used for the analysis were 103 for the complex with HC45, 87 for the complex with BH151, 81 for the complex with HC19, and 105 for the complex with HC63. On all the dendrograms, the A/Aichi/68 isolate was on the trunk-branches, although on the phylogenetic tree of HA the A/Aichi/68 isolate was at a tip (figure S1E). Here, the term “trunk-branches” was defined as the set of interior branches leading from the root to the most distal tip (Bush et al. 1999). In the present study, we also borrowed another term, “twig-branches” (Bush et al. 1999), and used that term to identify branches other than trunk-branches.

#### Affinity of the Antibodies with Binding Regions Away from the Receptor-Binding Sites

We identified 30 contributing residues (see *Materials and Methods*) for the antibody fragments HC45 (table 2) and 27 residues for BH151 (table S1A in the Supplementary

Material online). We also identified the contributing residues of HA: 37 residues for the epitope region of HC45 (table 2) and 28 residues for the epitope region of BH151 (table S1A). The contributing and background residues are shown in figure 4A (for the complex between HA and HC45) and figure S2A in the Supplementary Material online (for the complex between HA and BH151). The HA contributing residues (red), the HA background residues (light purple), and the HC45 Fab contributing residues (blue) are shown as balls representing  $C^\alpha$  atoms. The Fab background residues are not shown. In figure 4A and in figure S2A online, the receptor binding sites are also shown as yellow balls. In table 2 and table S1A, we listed Fab residues in which atoms contact with atoms of residues in HA within a 0.4 nm distance, and it was found that the contributing residues obviously included such residues.

*Binding ability of the antibody fragments.* For the complex between HC45 and HA, the contributions from environmental changes in HA (Equation 4), that of Fab (Equation 6), and the contributions from interactions (Equation 5) were 0.07, 0.94, and 2.19, respectively. For the complex between BH151 and HA, contributions were 0.01, 0.27, and 1.11, respectively. The logarithm of Equations 4 and 6 consisted of the sum of the contributions from the contributing and background residues. The signs of these contributions varied from term to term, and these cancellations reflected balance of hydrophathy. The results suggest that the contributing residues on HA prefer to be buried and contribute positively to the affinity of the complex between HA and the Fab.

*Evolutionary change in the binding ability and key amino acid replacement.* Figure 5 shows the changes in the binding abilities of the antibody fragments HC45 (fig. 5A) and BH151 (fig. 5B) along the trunk-branches of the phylogenetic dendrogram of HA.

For the complex between HA and antibody HC45, a drastic decrease in the binding ability was observed from the 1970s to the 1980s. This corresponds to branches between A/Amsterdam/1609/77 and A/Rotterdam/577/80 on the phylogenetic dendrogram. After A/Rotterdam/577/80, the binding ability of HC45 to HA decreased gradually. Other drastic decreases of similar magnitude occurred on the twig-branches: between A/Aichi/68 and A/Hong Kong/1/68 and between A/Amsterdam/1609/77 and A/Bilt-hoven/3895/77. These drastic decreases in affinity were mostly caused by a single amino acid replacement: I62K on HA1. The interaction between the sixty-second site on HA1 and the CDR3H region of the Fab largely decreased its strength. Furthermore lysine (K) is a basic amino acid and does not prefer to be buried. In figure 4B, the  $C^\alpha$  atom at the sixty-second site is shown as a red ball. The contributing residues in the CDR3H region of HC45 Fab, contacting the residue at the sixty-second site in HA1, are shown as orange balls, and the other contributing residues of the Fab are shown as blue balls. In fact, on the trunk-branches, the FSR of the sixty-second site in HA1 decreased from  $\log\text{FSR}(62\text{I})=0.80$  to  $\log\text{FSR}(62\text{K})=-0.56$ . Smith et al. (2004) also found that a single amino acid replacement on the influenza HA could cause significant antigenic changes. On the trunk-branches, it took 3 years for HA to gain such a reduced binding ability for

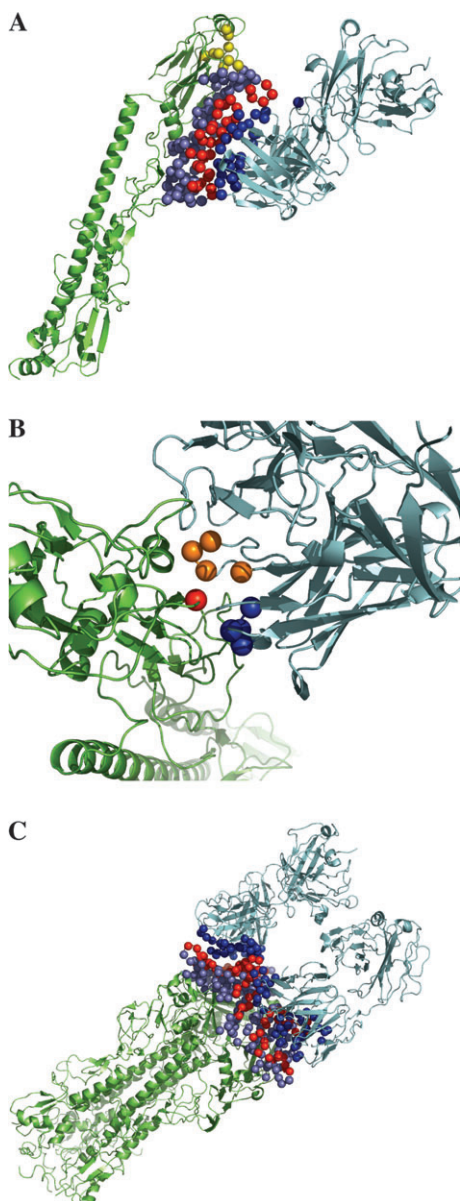


FIG. 4.—(A) The contributing and background residues of the complex between HA (green) and HC45 Fab (cyan) are shown as balls representing the  $C^{\alpha}$  atoms. The contributing residues and the background residues of HA and the contributing residues of Fab are shown as red, light purple, and blue balls, respectively. The receptor-binding sites are also shown as yellow balls. (B) The contributing residue at the sixty-second site in HA1 is shown as a red ball, and its partners of pairwise interactions in Fab are shown as orange and blue balls. The orange balls represent the contributing residues in the CDR3H region, and the blue balls correspond to the other residues. (C) The contributing and background residues of the complex between HA (green) and HC63 Fab (cyan) are shown as balls representing the  $C^{\alpha}$  atoms. The contributing and background residues of HA and the contributing residues of Fab are shown as red, light purple, and blue balls, respectively.

the Fab. However, this occurred on the twig-branches within a year.

In the case of antibody BH151, the binding ability to the A/Aichi/68 isolate was already low in comparison to that of HC45. Along the evolutionary pathway of HA, the binding ability of antibody BH151 decreased gradually.

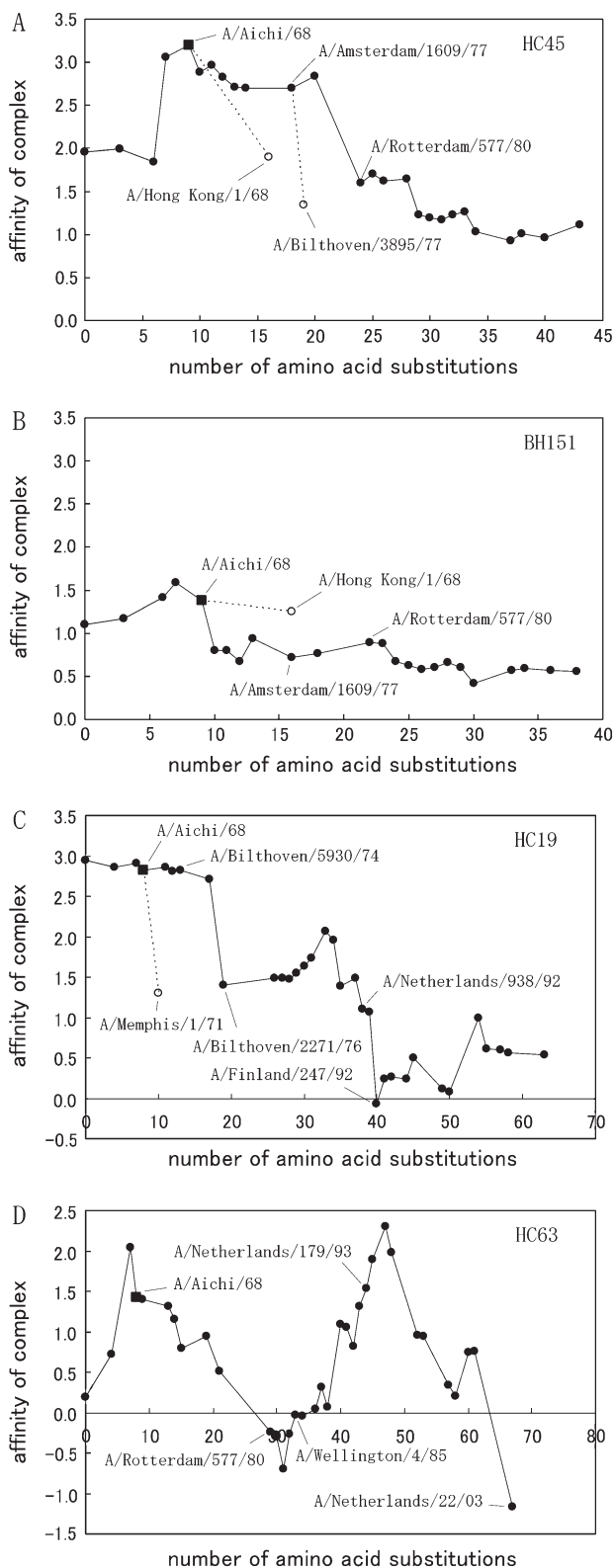


FIG. 5.—The change of the affinity of the complex between Fab (A: HC45, B: BH151, C: HC19, D: HC63) and HA along the trunk-branches on the phylogenetic dendrogram of HA is shown by a solid line. The horizontal axis represents the number of amino acid substitutions of the contributing residues and background residues of HA reconstructed by the maximum-parsimony method. Several changes of the affinity on twig-branches are shown by dotted lines.



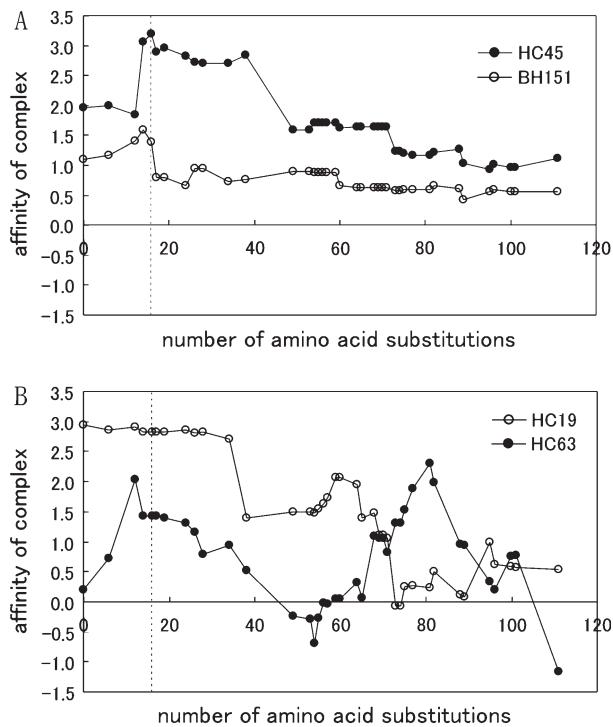


FIG. 6.—The change in the affinity of the complexes between Fab and HA along the trunk-branches on the phylogenetic tree of the full sequence of HA1 is shown. The horizontal axis represents the number of amino acid substitutions of the full sequence of HA1 reconstructed by the maximum-parsimony method. On each line, the point where the dashed line crosses corresponds to the node connecting with A/Aichi/68. (A) The line with solid circles is for HC45 and the line with open circles is for BH151. (B) The line with solid circles is for HC63 and the line with open circles is for HC19.

At the branch between A/Amsterdam/1609/77 and A/Rotterdam/577/80, the binding ability did not decrease; it even increased slightly. The substitution of I62K did not significantly affect the binding ability of BH151. Antibody BH151 could neutralize infection by isolate A/Hong Kong/1/68 and A/Aichi/68. Figure 6A shows the changes in the binding abilities of the two antibody fragments along the trunk-branches of the phylogenetic tree of the full sequence of HA1. Nodes along the trunk-branches of the phylogenetic tree are represented by circles. The binding ability decreasing gradually and remaining a positive value of affinity might suggest that the two antibodies HC45 and BH151 are not a great enough threat to the influenza virus.

#### Affinity of Antibodies Covering the Receptor-Binding Region

We identified the contributing residues of the antibody fragments HC19 (table S1B in the Supplementary Material online) and HC63 (tables S1C and S1D in the Supplementary Material online). The identified HC19 residues comprised 25 residues. The complex between HA and HC63 explained in the PDB structure data (1KEN) consisted of two HC63 Fabs and a HA trimer. Both Fabs binding to the HA trimer contained 35 contributing residues, although a few discrepancies exist. The contributing residues of HA

were also identified (tables S1B, S1C, and S1D online). The contributing and background residues are shown in figure S2B in the online Supplementary Material (for the complex between HA and HC19) and figure 4C (for the complex between HA and HC63).

*Binding ability of the antibody fragments.* The binding abilities of antibody fragments HC19 and HC63 to the HA glycoprotein of influenza A/Aichi/68 were calculated using Equation 1. The structure data for the complex with HC19 Fab (PDB code: 2VIR) contains a few mutations in the HA1 sequence. Hence, we superimposed the HA sequence of the A/Aichi/68 isolate to the corresponding structure. The results were 2.82 for HC19 and 1.42 for HC63. For the complex between HC19 and HA, the contributions from the environmental change in HA (Equation 4), that in Fab (Equation 6), and the contribution from the interaction (Equation 5) were  $-1.47$ ,  $1.14$ , and  $3.15$ , respectively. Conversely, for the complex between HC63 and HA, the contributions were  $-2.78$ ,  $0.01$  and  $4.52$ , respectively. The structural data for the complex between HA and HC63 (PDB code: 1KEN) contained two sets of Fab. Hence, we obtained an additional contribution ( $-0.32$ ) from clustering the Fabs. The contributing residues in HA prefer not to be buried and contribute negatively to the affinity of the complex between HA and the Fab. This is a different situation from the complex between HA and the Fab binding at a region distant from the receptor binding sites.

*Evolutionary change in the binding ability and key amino acid replacement.* Figure 5 shows the changes in the binding abilities of the antibody fragments HC19 (fig. 5C) and HC63 (fig. 5D) along the trunk-branches of the phylogenetic dendrogram of HA. For the complex between HA and antibody HC19, drastic decreases in binding ability were observed at the branch between A/Bilthoven/5930/74 and A/Bilthoven/2271/76 and at the branch between A/Netherlands/938/92 and A/Finland/247/92. The binding ability to the A/Finland/247/92 isolate was completely lost (the result of Equation 1 was negative). Along these two branches, the binding ability decreased by a different mechanism. At the first branch, the decrease in binding ability was caused by the amino acid replacement G158E in HA1. The interaction between the one hundred fifty-eighth site in HA1 and the CDR2H and CDR3L regions of the Fab largely decreased their strength. Furthermore, glutamic acid (E, acidic amino acid) does not prefer to be buried. As a result, the FSR of the one hundred fifty-eighth site in HA1 decreased from  $\log\text{FSR}(158\text{G})=0.83$  to  $\log\text{FSR}(158\text{E})=-0.77$ . By the same mechanism, another drastic decrease occurred on the twig-branch between A/Aichi/68 and A/Memphis/1/71. At the second mentioned branch on the trunk, the decrease in the binding ability was caused by the amino acid replacement S157L in HA1. The interaction between the one hundred fifty-eighth site in HA1 and the CDR1H and CDR2H regions of the Fab largely decreased their strength. In this case, leucine (L, a nonpolar amino acid) prefers to be buried. However, the decrease in the interaction strength was larger than the contribution of the one hundred fifty-eighth site's burial preference. The FSR of the one hundred fifty-eighth site in HA1 decreased from  $\log\text{FSR}(157\text{S})=0.65$  to  $\log\text{FSR}(157\text{L})=-0.49$ .

In the case of antibody HC63, the binding ability was completely lost over a period of 5 years (from A/Rotterdam/577/80 to A/Wellington/4/85). Along the evolutionary pathway of HA, after complete loss, the binding ability started to increase and against A/Netherlands/179/93 was restored to almost the same strength as against A/Aichi/68. The binding ability then decreased again and was completely lost against A/Netherlands/22/03. This repetitive decrease in the binding ability was not a trivial regression. At the first complete loss of the binding ability in figure 5D, the complex between the A/Rotterdam/577/80 isolate and the HC63 Fab significantly decreased the affinity between the one hundred thirty-seventh residue and the CDR3H region by the amino acid replacement N137Y when comparing the complex between A/Aichi/68 and HC63. During the regression, the tyrosine (Y) at the one hundred thirty-seventh residue was retained, and the balance of hydrophathy on HA increased its contribution to the affinity of the complex (result of Equation 4). At the second complete loss of binding ability, the dominant contribution to the decrease of the affinity of the complex between A/Netherlands/22/03 and HC63 was from the amino acid replacement G225D in HA1. The affinity between the two hundred twenty-fifth residue and the CDR3H region decreased significantly when comparing the complex between A/Aichi/68 and HC63. At that time, the amino acid at the one hundred thirty-seventh residue was replaced by serine, and the affinity between this residue and the CDR3H region was almost restored.

Figure 6B shows the changes in the binding abilities of the two antibody fragments along the trunk-branches of the phylogenetic tree of the full sequence of HA1. The circles on the lines represent the nodes on the trunk-branches of the tree. When the binding ability of HC63 was completely lost, the binding ability of HC19 was still preserved. Conversely, when the binding ability of HC19 was completely lost, the binding ability of HC63 was restored.

## Discussion

We developed an index to measure the binding affinity of protein-protein complexes and applied it to the investigation of the long-term escape of the influenza virus from the host immune system by tracing the changes of the binding ability between the HA glycoprotein and four monoclonal antibodies. The overview of the global trend enabled us to contrast the affinities with different types of antibodies. By focusing on branches with a significant decline in binding ability, we could detect key amino acid replacements and further investigate the mechanism via conditional probabilities representing the fitness of the amino acids to the structural environment.

Our index of affinity is based on the positions of the  $C^\alpha$  atoms. Comparing the atomic positions of the backbone atoms of HA in complexes with HC45 and BH151 Fabs (Fleury et al. 2000) uncovered no change in the effect of overall HA structure upon complex formation when compared with either the uncomplexed molecule or with HA in the complex with the other Fab. Conversely, few side chains changed their orientation and, hence, local changes

occurred in the HA structure. This implies that the organization of hydrogen bonds and the strength of the hydrophobic interaction are the result of the complicated process of protein docking. Because of these changes in microstructure via rotation of side chains, accurately predicting the complex structure from the structure of uncomplexed HA and Fab is difficult (Jenwitheesuk and Samudrala 2003). Therefore, we decided to extract grossly the change in microstructures reflected by the extent of the amino acid preference in the local environment. The change in the ability of rotation of the side chains is expected to be at least partly detected as a change in the log likelihood ratios.

The decline in antibody binding activity is a well-known behavior of H3N2 viruses. However, compared with the two antibodies HC45 and BH151, the monoclonal antibodies HC19 and HC63, which bind to the receptor-binding region and cover the receptor binding sites, have totally different features. Both of these antibodies completely lost their binding ability for a period of time on the evolutionary pathway of HA. After this complete loss of binding ability, antibody HC63 started to increase its ability, and over 10 years, the ability to bind to HA was restored. When the virus was free from antibody HC63, amino acid substitutions accumulated, and they enabled antibody HC63 to again bind to the same epitope region. However, the amino acid substitutions were not simply back mutations of the substitutions causing reduced binding ability. Irreversible sequence evolution appears to have occurred through coordinated or covariational substitutions at different sites that compensated for the effects on the structural characteristics (Nakajima et al. 2005).

Our findings seem to suggest that influenza HA endeavors to keep some of its structural features, such as its physicochemical characteristics, that do not require the specific single amino acid sequence of HA. This is functional redundancy, and antibodies employ this feature to bind to the HA surface. One of the remarkable properties of binding between proteins is redundancy (Dang, Nikolajczyk, and Sen 1998), a term for the feature where mutations at several protein-binding sites do not affect the affinity of the complex (Fields et al. 1995). In general, the physicochemical characteristics of protein are known to be evolutionarily conserved (Afonnikov, Oshchepkov, and Kolchanov 2001). By functional redundancy of antibodies, different antibodies whose variable domains display low sequence identities can bind to the same epitope region on HA. Hence, it may be true that antibodies do not require specific amino acid sequences of the epitope region on HA. The HA sequence, particularly in the receptor-binding region, changes under the selective pressure of the host immune system. After disarmament of the immune system, the HA protein regresses to the functionally same structural state as the original protein. Then the disarmed antibody can again bind to its epitope on HA, and HA sequence evolution is again under selective pressure. During the antibody's period of impotency, the other antibody that bound to the closely related epitope region preserved its ability to bind to the epitopes. This "arms race" is a competition between the virus's changing the sequences of the epitope regions at the cost of structural change and the immune system's producing sets of functionally redundant antibodies.

The final remark regards future studies. By modifying our index, it may be possible to analyze systems consisting of proteins and other organic molecules. One potential application of this index is the monitoring and prediction of the arms race between natural ligands and inhibitors that are bound by enzymes, where enzyme activity depends on the relative strengths of its binding affinities to the ligand and the inhibitor. Therefore, comparing the two indices will be essential. We hope that the efficiency of our likelihood-based index of binding affinities increases with the growth of structural information.

### Supplementary Material

Supplementary materials are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

### Acknowledgments

We thank Keiko Udaka for useful comments on the long-term variation of antibody–antigen affinity. This work was supported by the Japan Society for the Promotion of Science (JSPS).

### Literature Cited

- Afonnikov DA, Oshchepkov DY, Kolchanov NA. 2001. Detection of conserved physico-chemical characteristics of proteins by analyzing clusters of positions with co-ordinated substitutions. *Bioinformatics*. 17:1035–1046.
- Barbey-Martin C, Gigant B, Bizebard T, Calder LJ, Wharton SA, Skehel JJ, Knossow M. 2002. An antibody that prevents the hemagglutinin low pH fusogenic transition. *Virology*. 294:70–74.
- Bizebard T, Gigant B, Rigolet P, Rasmussen B, Diat O, Bösecke P, Wharton SA, Skehel JJ, Knossow M. 1995. Structure of influenza virus haemagglutinin complexed with a neutralizing antibody. *Nature*. 376:92–94.
- 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.
- Cox NJ, Bender CA. 1995. The molecular epidemiology of influenza viruses. *Semin Virol*. 6:359–370.
- Dang W, Nikolajczyk BS, Sen R. 1998. Exploring functional redundancy in the immunoglobulin  $\mu$  heavy-chain gene enhancer. *Mol Cell Biol*. 18:6870–6878.
- Dimitrov DS. 2004. Virus entry: molecular mechanisms and biomedical applications. *Nat Rev Microbiol*. 2:109–122.
- Felsenstein J. 2005. PHYLIP (Phylogeny inference package) version 3.6. Department of Genome Sciences, University of Washington, Seattle.
- Fields BA, Goldbaum FA, Ysem X, Poljak RJ, Mariuzza RA. 1995. Molecular basis of antigen mimicry by an anti-idiotypic. *Nature*. 374:739–742.
- Flery D, Barrère B, Bizebard T, Daniels RS, Skehel JJ, Knossow M. 1999. A complex of influenza hemagglutinin with a neutralizing antibody that binds outside the virus receptor binding site. *Nat Struct Biol*. 6:530–534.
- Flery D, Daniels RS, Skehel JJ, Knossow M, Bizebard T. 2000. Structural evidence for recognition of a single epitope by two distinct antibodies. *Proteins*. 40:572–578.
- Gubareva LV, Novikov DV, Hayden FG. 2002. Assessment of hemagglutinin sequence heterogeneity during influenza virus transmission in families. *J Infect Dis*. 186:1575–1581.
- Ha Y, Stevens DJ, Skehel JJ, Wiley DC. 2003. X-ray structure of the hemagglutinin of a potential H3 avian progenitor of the 1968 Hong Kong pandemic influenza virus. *Virology*. 309:209–218.
- Hobohm U, Sander C. 1994. Enlarged representative set of protein structures. *Protein Sci*. 3:522–524.
- Holmes KV. 2005. Adaptation of SARS coronavirus to humans. *Science*. 309:1822–1823.
- Jenwitheesuk E, Samudrala R. 2003. Improved prediction of HIV-1 protease-inhibitor binding energies by molecular dynamics simulations. *BMC Struct Biol*. 3:2.
- Jones DT, Taylor WR, Thornton JM. 1992. A new approach to protein fold recognition. *Nature*. 358:86–89.
- Jones DT. 1999. GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. *J Mol Biol*. 287:797–815.
- Knossow M, Gaudier M, Douglas A, Barrère B, Bizebard T, Barbey C, Gigant B, Skehel JJ. 2002. Mechanism of neutralization of influenza virus infectivity by antibodies. *Virology*. 302:294–298.
- Li W, Zhang C, Sui J. (16 co-authors). 2005. Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *EMBO J*. 24:1634–1643.
- Lim D, Park HU, LDe Castro, Kang SG, Lee HS, Jensen S, Lee KJ, Strynadka NCJ. 2001. Crystal structure and kinetic analysis of  $\beta$ -lactamase inhibitor protein-II in complex with TEM-1  $\beta$ -lactamase. *Nat Struct Biol*. 8:848–852.
- Lipatov AS, Govorkova EA, Webby RJ, Ozaki H, Peiris M, Guan Y, Poon L, Webster RG. 2004. Influenza: emergence and control. *J Virol*. 78:8951–8959.
- Nakajima K, Nobusawa E, Nagy A, Nakajima S. 2005. Accumulation of amino acid substitutions promotes irreversible structural changes in the hemagglutinin of human influenza AH3 virus during evolution. *J Virol*. 79:6472–6477.
- Naoi Y, Chong KT, Yoshimatsu K, Miyazaki G, Tame JRH, Park SY, Adachi S, Morimoto H. 2001. The functional similarity and structural diversity of human and cartilaginous fish hemoglobins. *J Mol Biol*. 307:259–270.
- Padlan EA, Abergle C, Tipper JP. 1995. Identification of specificity-determining residues in antibodies. *FASEB J*. 9:133–139.
- Robinson DM, Jones DT, Kishino H, Goldman N, Thorne JL. 2003. Protein Evolution with dependence among codons due to tertiary structure. *Mol Biol Evol*. 20:1692–1704.
- Simons KT, Bonneau R, Ruczinski I, Baker D. 1999a. Ab initio protein structure prediction of CASP III targets using ROSETTA. *Proteins*. 37(Suppl. 3):171–176.
- Simons KT, Ruczinski I, Kooperberg C, Fox BA, Bystroff C, Baker D. 1999b. Improved recognition of native-like protein structures using a combination of sequence-dependent and sequence-independent features of proteins. *Proteins*. 34:82–95.
- Smith DJ, Lapedes AS, de Jong JC, Bestebroer TM, Rimmelzwaan GF, Osterhaus ADME, Fouchier RAM. 2004. Mapping the antigenic and genetic evolution of influenza virus. *Science*. 305:371–376.
- Swofford DL. 2003. PAUP\* version 4.0: Phylogenetic analysis using parsimony (\*and other methods). Sunderland (MA): Sinauer Associates.
- Vilches C, Parham P. 2002. KIR: diverse, rapidly evolving receptors of innate and adaptive immunity. *Annu Rev Immunol*. 20:217–251.
- Watabe T, Kishino H, Okuhara Y, Kitazoe Y. 2006. Fold recognition of the human immunodeficiency virus type 1 V3 loop and flexibility of its crown structure during the course of adaptation to a host. *Genetics*. 172:1385–1396.

Webster RG, Bean WJ, Gorman OT, Chambers TM, Kawaoka Y. 1992. Evolution and ecology of influenza A viruses. *Microbiol Rev.* 56:152–179.

Webster RG. 2000. Immunity to influenza in the elderly. *Vaccine.* 18:1686–1689.

Zhang Z, Palzkill T. 2003. Determinants of binding affinity and specificity for the interaction of TEM-1 and SME-1  $\beta$ -

lactamase with  $\beta$ -lactamase inhibitory protein. *J Biol Chem.* 278:45706–45712.

Naoko Takezaki, Associate Editor

Accepted April 18, 2007