

# Structural Considerations in the Fitness Landscape of a Virus

Teruaki Watabe<sup>\*,1</sup> and Hirohisa Kishino<sup>2</sup>

<sup>1</sup>Center of Medical Information Science, Kochi University, Kochi, Japan

<sup>2</sup>Laboratory of Biometrics, Graduate School of Agriculture and Life Science, University of Tokyo, Tokyo, Japan

\*Corresponding author: E-mail: twatabe-mi@umin.ac.jp.

Associate editor: Naoko Takezaki

## Abstract

Viral fitness is determined by replication within hosts and transmission between them. We examine how pleiotropic mutations that have antagonistic effects (i.e., antibody evasion vs. receptor binding) on viral replication within hosts can impact viral immune escape in the host population. When the host population is vaccinated, the virus escapes from passive immunity by mutations in the antibody-binding region on the surface of the target protein. However, the reduced ability of the antibody to bind the virus is often accompanied by a reduced ability of the virus to bind the cell receptor because the antibody-binding region overlaps with the receptor-binding domain (RBD). The types of permitted mutations are limited. To investigate the causal relation between a mutation in a viral genome and adaptive evolution of a viral population, we developed a mathematical model that describes the population dynamics of viruses, antibodies, and normal/infected cells within a host. The coefficients describe the binding affinity between the virus and the induced antibody and that between the virus and its receptor. Our knowledge-based index enables us to estimate the effect of a mutation in a binding region on the binding affinity. Using population genetic theory, we evaluated the probability that a mutant is fixed in a host population. The mutations that can be fixed with high probabilities may determine how long a vaccine remains effective. We simulate the adaptive evolution of coronavirus, the etiological agent of severe acute respiratory syndrome, and show that some of mutations in the RBD may have high fixation probabilities in the vaccinated host population.

**Key words:** viral evolution, fitness landscape, protein structure, fixation probability.

## Introduction

Viruses provide an ideal opportunity to quantify the causal effect of a mutation in a genome on the adaptive evolution of the population. Phenotypic evolution is driven by selection pressure acting on viability and/or reproductive ability. Physiological and developmental processes of higher organisms are robust against genomic mutations (Wagner 2000; Li et al. 2004). Due to this buffering mechanism, genetic variations are preserved in phenotypically normal populations, and it is therefore difficult to establish a causal relationship between genomic mutations and phenotypic alterations (Rutherford 2000). On the other hand, viruses have a simple life cycle, so their adaptive evolution can be explained by a few factors (Beerenwinkel et al. 2002). In viral evolution, mutations in genetic coding sequences may correlate with phenotypic evolution through structural changes in proteins that function at various stages of the life cycle. This enables the simultaneous investigation of the phenotypic and genotypic evolution of viruses by monitoring the structural changes in proteins that are caused by genotypic changes and that are reflected in phenotypic differences.

We focus this study on the adaptive evolution of virus under the selection pressure of passive immunity induced by a vaccine. A mutation in a viral genome may change the structure of the antibody-binding region that is often overlapped with or close to the receptor-binding domain (RBD). The structural change of the antibody-binding re-

gion may affect the binding affinity between the virus and the host cell receptor as well as the binding affinity between the virus and the antibody. Therefore, the virus must decrease its affinity to antibodies while maintaining its ability to bind the receptor.

To quantify the causal effect of a mutation on the improved infectivity, we first develop a mathematical model to describe viral dynamics in a vaccinated host and to construct the fitness landscape of the virus. The model describes the dynamics of three interacting populations: host cells, viruses, and induced antibodies. Two types of binding affinities are incorporated into the model as coefficients in the differential equations. Our knowledge-based index of binding affinities, or association constants (Watabe et al. 2007), enables us to estimate changes in the values of the coefficients due to mutation. Assuming that epidemic factors are unchanged, fitness is measured as the size of the mutant population within a host relative to the size of the wild-type population. Adopting a population genetic approach, we calculate the probability that a mutant is fixed in a viral population within and between hosts.

As an empirical example, we analyze the adaptive evolution of a coronavirus, which was the etiological agent of severe acute respiratory syndrome (SARS; Ksiazek et al. 2003; Rota et al. 2003). SARS coronavirus (SARS-CoV) utilizes angiotensin-converting enzyme 2 (ACE2) to attach its spike (S) protein to the host cell (Li et al. 2003). The

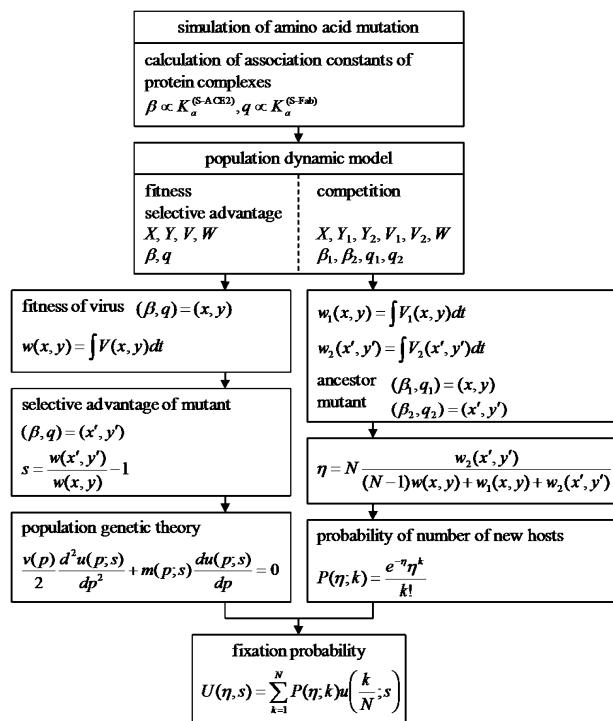


Fig. 1. Process flowchart of the method.

cross-species transmission of the virus was permitted by a few amino acid substitutions in the RBD of the S protein (Li W et al. 2005; Song et al. 2005). In order to prevent re-emergence of SARS, potent therapeutics and vaccines are needed. The S protein is a major antigenic determinant and is therefore an attractive target for neutralizing antibodies. It has been suggested that passive immunization with human monoclonal antibodies (mAbs) might be used to treat SARS (Holmes 2003), and human mAbs with a potent neutralizing effect against the RBD of the S protein have been identified (Sui et al. 2004; Prabakaran et al. 2006). Using our knowledge-based model of viral population dynamics and population genetics, we predict amino acid mutations that are most likely to fix in the viral population.

## Materials and Methods

### Summary of the Whole Process

Figure 1 summarizes the whole process described below. In this study, we determine a mutant virus by different values of the two parameters of the mathematical model and assume that the parameters are proportional to the association constants of the protein complexes: the S protein–ACE2 receptor complex and the S protein–antibody complex. We simulate the emergence of mutants by changing the amino acid sequence of the RBD of the S protein. The amino acid substitutions in the RBD affect the association constants of the protein complexes, and hence, the parameters of the mathematical model that characterize the virus change. We determine the amino acid mutation rate by structural constraints on the muta-

tions. In the simulation, the rate is incorporated by generating mutations according to the amino acid preferences in the local structural environment, and the association constants are evaluated by a knowledge-based sequence–structure fitness measure (Watabe et al. 2007).

### Mathematical Model of Viral Dynamics in a Vaccinated Host

To gain insight into the population dynamics of infectious viruses in a vaccinated host, we here develop a mathematical model that describes the dynamics of three interacting populations: host cells, virus particles, and neutralizing antibodies (Nowak and Bangham 1996). For vaccine-mediated protection against viral diseases, the crucial agents are antibodies, which mostly neutralize free viral particles (Burton 2002). Hence, the model contains four variables: uninfected cells ( $X$ ), infected cells ( $Y$ ), free viral particles ( $V$ ), and neutralizing antibodies ( $W$ ). These variables obey the following differential equations:

$$\begin{aligned}\dot{X} &= \lambda - dX - \beta XV \\ \dot{Y} &= \beta XV - aY \\ \dot{V} &= kY - uV - \beta XV - qVW \\ \dot{W} &= rV - hW - qVW.\end{aligned}\quad (1)$$

Uninfected cells are constantly produced at rate  $\lambda$  and infected by free viruses at rate  $\beta XV$ . Infected cells produce free viral particles at rate  $kY$ . The half-lives of uninfected cells, infected cells, and free viral particles are  $1/d$ ,  $1/a$ , and  $1/u$ , respectively. The basic reproductive ratio of the virus is given by  $R_0 = \beta \lambda k / (adu)$ . In the work of Nowak and Bangham (1996), the variable  $W$  represents B cells, and its production rate is set to  $rVW$ . In the present study, the variable  $W$  represents neutralizing antibodies; therefore, its production rate is set to be proportional to the amount of free viral particles,  $rV$ . The half-life of neutralizing antibodies is  $1/h$ . Free viral particles are neutralized by antibodies at rate  $qVW$ . In order to avoid ambiguity caused by the large number of parameters, we normalized the variables by the parameters in equation (1) and solved the equations numerically (see Supplementary Material online).

The half-life of uninfected cells,  $1/d$ , is set to 100 days (Nowak and Bangham 1996) and that of neutralizing antibodies (immunoglobulin G),  $1/h$ , is set to 20 days (Insel and Looney 2004). The virion half-life has been estimated for several viruses. In an analysis of viral dynamics during antiviral therapy, the hepatitis C virus (HCV) has been estimated to have an average half-life of 2.7 h (Neumann et al. 1998). For the hepatitis B virus (HBV), the estimated virion half-life is about 1 day (Nowak et al. 1996). The half-life of the human immunodeficiency virus has been estimated to be 5.8 h (Perelson et al. 1996). The estimated half-life of infected cells not only depends on the virus but also shows considerable interpatient variation. Half-lives of 1.7–70 days for HCV (Neumann et al. 1998) and 10–100 days for HBV (Nowak et al. 1996) have been obtained. In the present study, we employ two values

(2 h and 1 day) for virion half-life and two values (2 and 10 days) for infected cell half-life.

### Extended Mathematical Model to Deal with Viral Mutations

We introduce viral mutations that result in different binding affinities to the receptor and/or to the antibody. Assuming that a mutant virus emerges immediately after infection, we extend the mathematical model (eq. 1) as follows. Until a mutant emerges ( $0 \leq t < \hat{t}$ ), the form of the extended equation is the same as the original, except that subscripts are used:

$$\begin{aligned}\dot{X} &= \lambda - dX - \beta_1 X V_1 \\ \dot{Y}_1 &= \beta_1 X V_1 - a Y_1 \\ \dot{V}_1 &= k Y_1 - u V_1 - \beta_1 X V_1 - q_1 V_1 W \\ \dot{W} &= r V_1 - h W - q_1 V_1 W.\end{aligned}\quad (2)$$

Following emergence of the mutant ( $t \geq \hat{t}$ ), mutant viruses  $V_2$  infect cells, and cells infected by mutant viruses, denoted as  $Y_2$ , produce free viral particles of the mutant:

$$\begin{aligned}\dot{X} &= \lambda - dX - \beta_1 X V_1 - \beta_2 X V_2 \\ \dot{Y}_i &= \beta_i X V_i - a Y_i \\ \dot{V}_i &= k Y_i - u V_i - \beta_i X V_i - q_i V_i W \\ \dot{W} &= r V_1 + r V_2 - h W - q_1 V_1 W - q_2 V_2 W.\end{aligned}\quad (3)$$

Here the subscript  $i$  is 1 or 2. We consider a mutant emerges on a time scale of a half-life of the virus. We set the time  $\hat{t}$  equal to the half-life of free viral particles  $1/u$ . As with equation (1), we normalize the variables by the parameters in equations (2) and (3) (see Supplementary Material online).

### Adaptation to Vaccinated Hosts and Simulating the Viral Fitness Landscape in Vaccinated Hosts

We define the fitness of a viral population (characterized by a parameter set  $(\beta, q) = (x, y)$  in eq. 1) within vaccinated hosts by integrating the variable  $V$  over time:

$$w(x, y) = \int_0^\tau V(x, y; t) dt. \quad (4)$$

We assume that transmission between hosts is directly related to viral load within infected hosts (i.e., eq. 4). The incubation period of SARS is usually 3–5 days (World Health Organization 2003), and the virus can infect new hosts until medical treatment occurs. The medical treatment may start within a few days of onset. Hence we should consider 5–8 days as a time period that the virus can infect to other hosts. Therefore, we set the range of integration as 0 to  $\tau$ . On the  $(\beta, q)$ -plane, a viral fitness landscape is given based on changes in  $w$ . Viruses with higher values of  $w$  can infect new hosts more efficiently. On the other hand, the final results of our analysis do not depend much on the value of  $\tau$  (data not shown). The

aspect of the fitness landscape is characterized by the truncation of the integration itself rather than the time point of the truncation. Hence, we set the value of  $\tau$  equal to 10 days.

### Fixation of a Mutant Virus within a Vaccinated Host Population

In order to evaluate the fixation probability of a mutant virus within a vaccinated host population, we must determine the selective advantage and initial frequency of the mutant. The selective advantage can be defined by a ratio between the fitness of the mutant virus and that of the ancestral virus:

$$s = \frac{w(x', y')}{w(x, y)} - 1. \quad (5)$$

Here,  $w$  in the numerator is the fitness of the mutant virus characterized by a parameter set  $(\beta, q) = (x', y')$ , whereas  $w$  in the denominator is the fitness of the ancestor characterized by a parameter set  $(\beta, q) = (x, y)$ . When the fitness of the mutant virus is the same as that of the ancestral virus ( $w(x', y') = w(x, y)$ ), the value of its selective advantage is zero. On the other hand, a positive (or negative) value of the selective advantage corresponds to the mutant virus with higher (or lower) fitness than that of the ancestor. The fixation probability of an allele with frequency  $p$  is obtained as the solution of the ordinary differential equation

$$\frac{v(p)}{2} \frac{d^2 u(p)}{dp^2} + m(p) \frac{du(p)}{dp} = 0,$$

with boundary conditions  $u(0) = 0$  and  $u(1) = 1$  (Kimura 1962). Here, the second term represents the change in the frequency that is driven by the selective advantage and the first term corresponds to the drift of the frequency.  $m(p)$  and  $v(p)$  are expressed as

$$\begin{aligned}m(p) &= \frac{sp(1-p)}{1+sp} \\ v(p) &= \frac{p(1-p)}{N}.\end{aligned}$$

Note that the selective advantage is not necessarily small in our case (Wright 1931).  $N$  is the host population size, in which all hosts are vaccinated and infected by the ancestral virus before the mutant emerges. We assume that the size of the host population is constant until the mutant virus becomes fixed in the population although the individuals in the population change. In the present study, we consider a million scale of vaccinee and assume that infected vaccinee population stands at 1% ( $N = 10,000$ ). As a result, we obtain the following formula for the fixation probability of the mutant virus:

$$u(p) \equiv u(p; s) = \frac{1 - e^{-(2N-1)\log(1+sp)}}{1 - e^{-(2N-1)\log(1+s)}}. \quad (6)$$

### Population Dynamics of an Emerging Mutant Virus within Its First Host and the Initial Frequency in the Host Population

An emerging mutant virus within a host, which is characterized by the parameter set  $(\beta_2, q_2) = (x', y')$  in equation (3), inevitably competes with the existing viral strain characterized by the parameter set  $(\beta_1, q_1) = (x, y)$ . The result of this competition determines whether the emerging mutant can infect new hosts. To infect new hosts, the mutant must achieve a sufficient viral load within its first host. We define the integrals of the variables in equation (3) ( $V_1$  for the existing virus within the host and  $V_2$  for the mutant virus) as follows:

$$w_1(x, y) = \int_0^\tau V_1(x, y; t) dt$$

$$\text{and } w_2(x', y') = \int_0^\tau V_2(x', y'; t) dt.$$

We consider the ratio between  $w_2$  and the sum of  $w$ 's of all hosts within the population,  $(N - 1)w(x, y) + w_1(x, y) + w_2(x', y')$ , as a frequency that determines the number of new hosts infected by the mutant virus. We employ the Poisson distribution:

$$P(\eta; k) = \frac{e^{-\eta} \eta^k}{k!} \quad (7)$$

$$\text{with } \eta = N \frac{w_2(x', y')}{(N - 1)w(x, y) + w_1(x, y) + w_2(x', y')}.$$

Here,  $k$  represents the number of new hosts infected by the mutant in the next generation, and  $P(\eta; k)$  is the probability that the number of new hosts is  $k$ .  $\eta$  represents the expected value of the number of the new hosts. Combining the fixation probability with a given initial frequency (eq. 6) and the probability of the number of new hosts infected by the mutant (eq. 7), the fixation probability of the mutant virus in the host population is given by:

$$U(\eta, s) = \sum_{k=1}^N P(\eta; k) u\left(\frac{k}{N}; s\right).$$

### Amino Acid Mutation Rate in the RBD

The rate at which one amino acid residue changes to another is the product of two factors: the rate at which mutations occur at the corresponding residue ( $\alpha^i$  at residue  $i$ ) and the frequency with which the new amino acid is selected at the residue ( $f_{aa'}^i$  for a change from amino acid  $a$  to  $a'$  at residue  $i$ ).

$$R_{aa'}^i = \alpha^i f_{aa'}^i.$$

The rate  $\alpha^i$  is determined by several factors, including the error rate of the viral polymerase. We assume that

the rate  $\alpha^i$  is constant among residues in the RBD. Here, we do not explicitly consider the time span of occurrence of a resistant virus to the vaccine. Hence, we do not fix the explicit value of the rate. The frequency  $f_{aa'}^i$  should be determined mostly by structural constraints on mutations. In a simulation of amino acid substitutions, we incorporated the frequency by generating mutations according to the amino acid preferences in the local structural environment ( $P(a_i | \mathbf{X})$ ; Watabe et al. 2006, 2007).

### Footprint of Adaptive Evolution on the Fitness Landscape

In the present study, we examine the adaptive evolution of the virus by using the viral fitness landscape within a vaccinated host population that was obtained from the mathematical model of viral population dynamics in vaccinated hosts (eq. 1). The virus starts from a point on the fitness landscape and moves through the landscape based on changes of binding affinities resulting from particular mutations in the amino acid sequence of the RBD. The parameters of the mathematical model,  $(\beta, q)$ , represent the strengths of couplings between virus particles and uninfected cells and between virus particles and antibodies. Those couplings are mediated by proteins: the S proteins, the ACE2 receptors, and the fragments of antigen-binding (Fabs) of antibody. Hence, we assume that the parameters of the mathematical model,  $(\beta, q)$ , have the following relationships to the association constants of the protein complexes ( $K_a^{(S-ACE2)}$  for the S protein–ACE2 receptor complex and  $K_a^{(S-Fab)}$  for the S protein–Fab complex):  $\beta \propto K_a^{(S-ACE2)}$  and  $q \propto K_a^{(S-Fab)}$ . The trajectory of mutants on the fitness landscape was characterized by the changes in association constants calculated for the mutated amino acid sequence of the RBD.

### A Likelihood-Based Index of the Binding Affinity of Protein–Protein Complexes

*In silico* modeling based on molecular dynamics is a tool to observe structural changes caused by mutations occurring in genome. In order to carry out *in silico* calculations, information about the structure of the corresponding molecular complex is needed as a template. In the case of SARS-CoV, the structure of the complex formed between the RBD and the peptidase domain of human ACE2 has been determined (Li F et al. 2005), and X-ray crystallography data are available in the Protein Data Bank (PDB). The structure of the S protein when bound by a neutralizing antibody has also been described (Hwang et al. 2006; Prabhakaran et al. 2006). In order to estimate the binding free energy of the protein complex, the protein–protein interactions must be investigated. Several computational methods have been proposed (Elcock et al. 2001). Molecular dynamics simulations have increasingly been used to study protein–protein interactions. However, because of long convergence time, molecular dynamics simulations are prohibitive for large systems such as protein complexes (Cui et al. 2008). In this study, we employ a



knowledge-based sequence–structure fitness measure (Watabe et al. 2007) to evaluate the binding affinities between the S protein of SARS-CoV and the Fabs of mAbs and also those between the S protein and its receptor. The sequence–structure fitness measure is a likelihood-based index, which quantifies a degree to which a sequence fits in a given structure. The measure is calculated empirically based on amino acid preferences and pairwise interactions in structural environments present in template structures. By bypassing structural prediction, it is possible to analyze thousands of sequences and correlate high-fitness sequences with structural changes along evolutionary pathways.

The binding affinity of a complex formed between proteins A and B is measured as the association constant of the complex,  $K_a = \frac{[A \cdot B]}{[A][B]}$ , where [A], [B], and [A·B] are the concentrations of proteins A, B, and their complex in equilibrium. We can relate the ratio to the likelihood ratio,  $K_a \propto \frac{P(A \cdot B)}{P(A)P(B)}$  (see Supplementary Material online), and predict the ratio using information from the protein database. Using the fact that the amino acid sequences in the binding region and the structure of the complex characterize the binding affinity of the complex, we represent the likelihood as the likelihood of the amino acid sequences given the structure:

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

Because we do not consider structural changes in the main chains of the proteins, the likelihoods  $P(\text{str}_{A+B})$ ,  $P(\text{str}_A)$ , and  $P(\text{str}_B)$  are constant. Thus, the index of the affinity of the complex is defined as the likelihood ratio of the amino acid sequences in the binding region of the complex to those in the binding region of the uncomplexed proteins:

$$K_a \propto \frac{P(\text{seq}_A, \text{seq}_B | \text{str}_{A+B})}{P(\text{seq}_A | \text{str}_A)P(\text{seq}_B | \text{str}_B)}. \quad (8)$$

The sequence distribution of a structure  $P(\text{seq} | \text{str})$  in equation (8), which we call the sequence–structure fitness measure, expresses the fitness of the amino acids given the structural environment and can be represented by the amino acid preference and the pairwise interaction (Watabe et al. 2006, 2007):

$$P(a_1, \dots, a_n | \mathbf{X}) \cong \prod_i P(a_i | \mathbf{X}) \prod_{i < j} \frac{P(a_i, a_j | \mathbf{X})}{P(a_i | \mathbf{X})P(a_j | \mathbf{X})}. \quad (9)$$

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. Expecting that the fitness of an amino acid depends primarily on the local structure surrounding each amino acid residue, Simons et al. (1999a, 1999b) have categorized local environments and calculated amino acid frequencies in the proteins registered in the PDB for each category.

Applying equation (9) to equation (8), the index of binding affinity is expressed by the product of the three parts:

$$K_a \propto \frac{P(\text{seq}_A | \text{str}_{A+B})}{P(\text{seq}_A | \text{str}_A)} \times P_{\text{int}} \times \frac{P(\text{seq}_B | \text{str}_{A+B})}{P(\text{seq}_B | \text{str}_B)}.$$

The first and third parts on the right-hand side of the above equation can be interpreted as contributions from hydrophobic interactions. The second part corresponds to contributions from the pairwise interactions between residues of the proteins A and B. Its explicit formula is

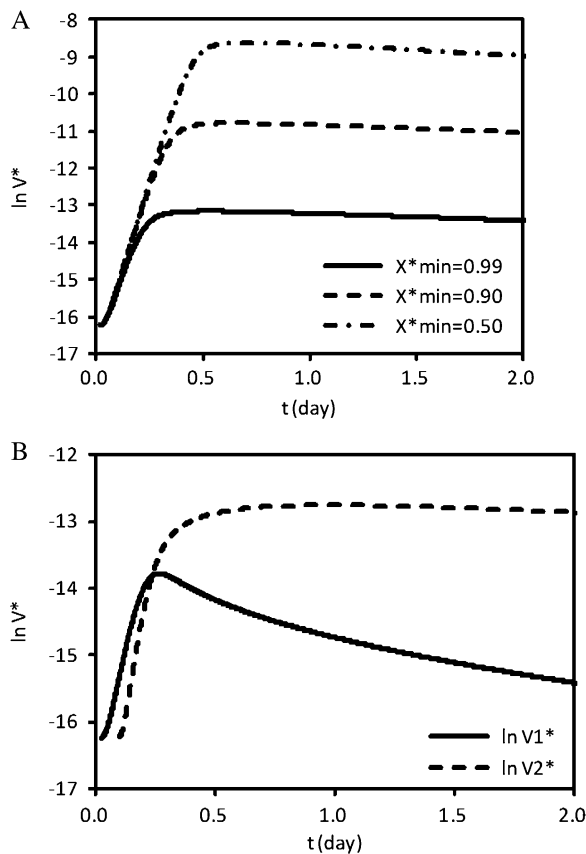
$$P_{\text{int}} = \prod_{\substack{i \in A \\ j \in B}} \frac{P(a_i, a_j | \mathbf{X}_{A+B})}{P(a_i | \mathbf{X}_{A+B})P(a_j | \mathbf{X}_{A+B})}. \quad (10)$$

### Calculation of Amino Acid Preferences and Pairwise Interactions

The amino acid preferences and interactions of the amino acid pairs are measured by empirical frequencies 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 degree of residue burial characterizes the local environment of a residue, which is defined as the number of  $C^\alpha$  atoms surrounding the  $C^\alpha$  atom of the corresponding residue within a distance of 1 nm (10 Å) (Watabe et al. 2007). The spatial distances between the  $C^\alpha$  atoms of each amino acid pair are categorized into intervals of 0.1 nm from 0.7 to 1 nm. Pairs with spatial distances of  $\leq 0.7$  nm are compiled into one category because such small distances are rare (Watabe et al. 2007). The interaction between a residue and its neighborhood in the primary sequence plays a major role in making local structural conformations, such as an  $\alpha$ -helix. Hence, we treat interactions between residues with small site intervals  $j - i \leq 4$  separately from those with larger site intervals (Watabe et al. 2006). To achieve a balance between robust estimation of amino acid preferences and pairwise interactions and sensitive detection of changes in microstructure, we adopt a sliding window approach (Watabe et al. 2007).

### Binding Affinities of Spike Protein–Receptor and Spike Protein–Antibody Complexes

We estimate the binding affinities of the S protein–receptor and S protein–antibody complexes using the index developed above (eq. 8) along with X-ray crystallography data from the PDB. Li et al. (2005) have determined the structure of the complex formed between the RBD of the S protein of SARS-CoV and the peptidase domain of human ACE2 (PDB code: 2AJF). Hwang et al. (2006) have described the structure of the complex formed between the S protein and a neutralizing antibody, 80R (PDB code: 2GHW). The 80R-binding epitope on the RBD overlaps very closely with the ACE2-binding site (Hwang et al. 2006). Prabhakaran et al. (2006) have reported the crystal structure of



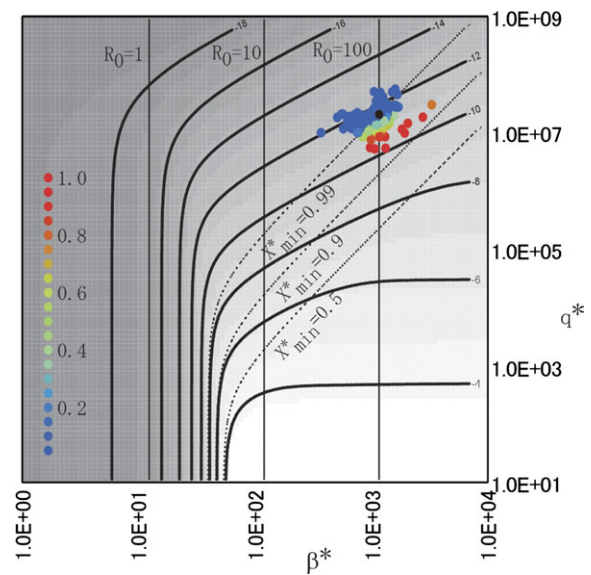
**FIG. 2.** Viral time courses of an ancestor and a mutant. In (A), the viral time courses were obtained with the basic reproductive ratio  $R_0 = 100$  and the three values of the parameter  $q$  that bring the three minimum values of the variable for the uninfected cell  $X$  ( $Xd/\lambda|_{\min} = 0.99, 0.90$ , and  $0.50$ ). In (B), the ancestor is characterized by the basic reproductive ratio  $R_0 = 100$  and the minimum value of the variable for the uninfected cell  $X$  ( $Xd/\lambda|_{\min} = 0.99$ ). The basic reproductive ratio of the mutant is  $R_0 = 200$ .

the RBD of the S protein in complex with a neutralizing antibody, m396 (PDB code: 2DD8). This antibody recognizes a region that overlaps half of the receptor-binding region. We used these three structural data sets on protein complexes as templates to estimate the binding affinity of mutated viruses in our simulation. Here, we assume that the main chain structure of the S protein is conserved for the mutated viruses.

## Results

### Within-Host Dynamics

In figure 2A, we show how the viral dynamics depends on the basic reproductive ratio  $R_0 = 100$  and the three values of the parameter  $q$  that yield the three minimum values of the variable for the uninfected cell  $X$  ( $Xd/\lambda|_{\min} = 0.99, 0.90$ , and  $0.50$ ). Because the variables were normalized, the value of  $Xd/\lambda$  corresponds to the proportion of uninfected cells to its maximum amount when the virus does not invade. During the early stage of infection, viral load increases linearly on a logarithmic scale and does not depend on the



**FIG. 3.** Fitness landscape and the positions of the 1,000 SAA mutations. The fixation probability of each mutation within the host population vaccinated with the antibody 80R is indicated by color. Warm colors represent higher probabilities and cold colors indicate lower probabilities. We performed simulations of mutations with parameters of 2 h for virion half-life and 2 days for infected cell half-life. We defined parameters with asterisks as  $\beta^* = \beta\lambda k/(ad)$  and  $q^* = q\lambda kr/(adu)$ . The values on the contours correspond to logarithms of fitness:  $\ln w - \ln(\lambda/d)(k/a)$ .

ability of the neutralizing antibody ( $q$ ). In figure 2B, the viral time courses of an ancestor and a mutant are shown. The ancestor is characterized by  $R_0 = 100$  and the minimum value of the variable for the uninfected cell  $X$  ( $Xd/\lambda|_{\min} = 0.99$ ). The basic reproductive ratio of the mutant is  $R_0 = 200$ . It is shown that the mutant drives out the ancestor.

### Within-Host Fitness Landscape

We obtained the fitness landscapes by solving the mathematical model (fig. 3 and supplementary fig. S1, Supplementary Material online). The vertical axis and horizontal axis represent the model parameters  $q$  and  $\beta$ , respectively. The parameter  $\beta$  is proportional to the basic reproductive ratio:  $\beta = R_0 adu/(\lambda k)$ . We show vertical lines characterized by the basic reproductive ratios of 1, 10, and 100. Moreover, we show curves expressing minimum values of the variable for the uninfected cell  $X$  ( $Xd/\lambda|_{\min} = 0.99, 0.90$ , and  $0.50$ ). The fitness landscapes have common features. Fitness is defined by the integral of  $V$  (in eq. 1) over time, which is truncated at time  $\tau = 10$  days to reflect the infectious period. Furthermore, as predicted from equation (1), the dynamics of a virus with slow replication ( $R_0 \sim 1$ ) does not depend on the binding affinity of neutralizing antibody during the early stages of infection (see supplementary material, Supplementary Material online). Hence, when  $R_0$  is low, the fitness landscape does not depend much on binding affinity of the neutralizing antibody, whereas when  $R_0$  is high, fitness varies with  $q$  and  $\beta$ .

### Changes in Binding Affinity due to Single-Amino Acid Mutations

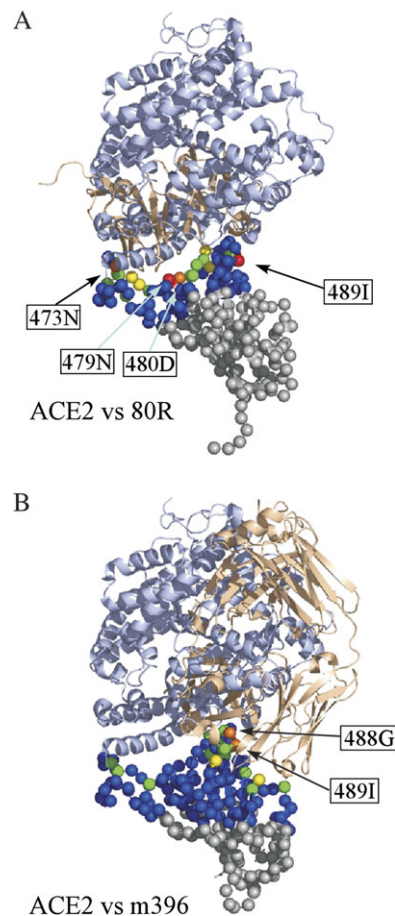
In the structural data available for the three protein complexes in the PDB (2AJF, 2GHW, and 2DD8), there are 180 common amino acid residues in the RBD. Among these, 78 amino acid residues (45%) are related to receptor binding (supplementary fig. S2A, Supplementary Material online). In the case of the complex with the antibody 80R, 79 amino acid residues (46%) of the RBD are utilized by the mAb to bind, completely covering the amino acid residues for receptor binding (supplementary fig. S2B, Supplementary Material online). In the case of the complex with the antibody m396, 84 amino acid residues (48%) of the RBD contribute to mAb binding. Half of these (46 amino acid residues or 26%) overlap with the region for receptor binding (supplementary fig. S2C, Supplementary Material online).

To see the distribution of the selection intensities of mutants, we performed 1,000 single amino acid (SAA) mutations at random at each of amino acid positions related to receptor binding and evaluated the association constants for each mutant. We plot the positions of the mutated sequences on the landscape for the antibodies 80R (fig. 3) and m396 (supplementary fig. S3, Supplementary Material online). Here we employed a point,  $(\beta^*, q^*) = (1.20 \times 10^3, 2.25 \times 10^7)$ , such that the minimum value of the variable for the uninfected cell is  $Xd/\lambda|_{\min} = 0.99$  and the basic reproductive ratio is  $R_0 = 100$ , as the point for the original sequence. Within the range of variation of the SAA mutations on the fitness landscape, some mutants gain large increases of fitness that achieve large selective advantages.

### Major Amino Acid Residues Contributing to Adaptive Evolution

We simulated SAA mutations at all residues of the RBD and evaluated the fixation probability of each SAA mutation. In figure 4, we indicate the amino acid residues at which mutations can be fixed with high probabilities within the host populations. In table 1, we list the fixation probabilities of mutations at the amino acid residues indicated in figure 4. At the starting point on the fitness landscape, mutants realizing greater affinity to the receptor and/or lower affinity to the antibody gain a higher fixation probability. In the case of vaccination using the antibody 80R (fig. 4A), the average fixation probability at residue 489I of the S protein is 0.71. However, at the residues on either side of 489I, the average fixation probabilities are less than 0.10. On the other hand, in the case of vaccination using the antibody m396 (fig. 4B), mutations to charged amino acid residues at residues 488G and 489I have high fixation probabilities, and the average fixation probabilities are greater than 0.40. These differences are caused by structural differences in the S protein–antibody binding.

Mutations at residues 488–490 do not substantially change the binding affinity of the complex between the S protein and the antibody 80R. In contrast, mutations



**FIG. 4.** Amino acid residues at which mutations can be fixed with high probabilities within the host populations. The colored spheres are at the positions of the  $C^\alpha$  atoms of the corresponding amino acid residues. The red sphere indicates the residue at which the average fixation probability over the 20 possible mutations is within the category of  $0.6 \leq \bar{U} < 0.8$ , whereas the orange sphere indicates the residue at which the average fixation probability is within  $0.4 \leq \bar{U} < 0.6$  (similarly, the yellow sphere corresponds to  $0.2 \leq \bar{U} < 0.4$ , the green sphere to  $0.1 \leq \bar{U} < 0.2$ , and the blue sphere to  $\bar{U} < 0.1$ ). We performed simulations of adaptive evolution on the fitness landscape with parameters of 2 h for virion half-life and 2 days for infected cell half-life. We employed a starting point located along the line expressing the minimum value of the variable for the uninfected cell  $X(Xd/\lambda|_{\min} = 0.99)$ . The point was at  $(\beta^*, q^*) = (1.20 \times 10^3, 2.25 \times 10^7)$ , where the basic reproductive ratio is  $R_0 = 100$ . The receptor ACE2 (light blue) and the antibodies (flesh color) are superimposed.

at residue 489I increase the binding affinity of the S protein–receptor complex; hence, their fixation probabilities within hosts vaccinated with the antibody 80R are high (supplementary fig. S4A, Supplementary Material online). Mutations at the residues on either side of 489I decrease the binding affinity of the S protein–receptor complex; hence, their fixation probabilities are very low (supplementary fig. S5A, Supplementary Material online). On the other hand, the binding affinity of the complex between the S protein and the antibody m396 is affected by mutations at residues 488–490. Mutations that increase (or decrease) the binding affinity of the S protein–receptor complex also

**Table 1.** The Fixation Probabilities of Mutations at the Amino Acid Residues Indicated in Figure 4.

| Amino acid residue | Mutation |      |        |      |        |        |      |      |        |        |      |        |      |      |      |      |      |      |      |        |
|--------------------|----------|------|--------|------|--------|--------|------|------|--------|--------|------|--------|------|------|------|------|------|------|------|--------|
|                    | G        | A    | V      | L    | I      | F      | P    | M    | W      | S      | T    | N      | Q    | Y    | C    | D    | E    | K    | R    | H      |
| ACE2 versus 80R    |          |      |        |      |        |        |      |      |        |        |      |        |      |      |      |      |      |      |      |        |
| 473 N              | 0.24     | 0.92 | 0.95   | 0.96 | 0.77   | 0.47   | 0.84 | 0.82 | 0.67   | 0.57   | 0.08 | —      | 0.75 | 0.27 | 0.11 | 0.03 | 0.18 | 0.53 | 0.65 | 0.21   |
| 479 N              | 0.48     | 0.90 | 0.99   | 0.97 | 0.95   | 0.66   | 0.63 | 0.95 | <0.005 | <0.005 | 0.50 | —      | 0.33 | 0.84 | 0.90 | 0.41 | 0.98 | 0.91 | 0.79 | 0.65   |
| 480 D              | <0.005   | 0.73 | 0.97   | 0.94 | 0.96   | 0.69   | 0.30 | 0.74 | 0.58   | <0.005 | 0.23 | <0.005 | 0.33 | 0.54 | 0.51 | —    | 0.35 | 0.91 | 0.78 | 0.18   |
| 489 I              | 0.99     | 0.91 | 0.19   | 0.15 | —      | 0.73   | 0.72 | 0.81 | 0.77   | 0.84   | 0.96 | 0.86   | 0.92 | 0.76 | 0.93 | 0.91 | 0.64 | 0.92 | 0.89 | 0.93   |
| ACE2 versus m396   |          |      |        |      |        |        |      |      |        |        |      |        |      |      |      |      |      |      |      |        |
| 488 G              | —        | 0.04 | <0.005 | 0.21 | <0.005 | 0.62   | 0.77 | 0.07 | 0.80   | 0.03   | 0.43 | 0.91   | 0.66 | 0.36 | 0.17 | 0.79 | 0.81 | 0.81 | 0.09 | 0.54   |
| 489 I              | 0.40     | 0.53 | 0.30   | 0.59 | —      | <0.005 | 0.07 | 0.28 | 0.50   | 0.34   | 0.42 | 0.85   | 0.24 | 0.10 | 0.85 | 0.71 | 0.83 | 0.95 | 0.56 | <0.005 |

increase (or decrease) the binding affinity of the S protein–antibody complex. Among these mutations, mutations to charged residues decrease the binding affinity of the S protein–antibody complex due to hydrophobic interaction. In the S protein–antibody m396 complex, the degrees of residue burial for residues 488–490 are greater than those in the S protein–receptor complex. This makes the influence of hydrophobic interaction stronger for residues 488–490 in the S protein–antibody m396 complex than in the S protein–receptor complex. Hence, their fixation probabilities within hosts vaccinated with the antibody m396 are high (supplementary figs. S4B and S5B, Supplementary Material online).

Closer Examination of Pairwise Interactions

The mutations at residue 489I that increase the binding affinity of the S protein–receptor complex primarily affect the pairwise interactions with the two glycines whose C<sup>α</sup> atoms are 0.7–0.8 nm away from the C<sup>α</sup> atom of residue 489I ((*r*, aa, *m*′, Δlog<sub>10</sub>*K*<sub>a</sub>) = (0.77, G, 17, 0.25) and (0.73, G, 21, 0.26) in table 2 and (0.77, G, 17, 0.19) and (0.73, G, 21, 0.19) in supplementary table S1, Supplementary Material online). The pairwise interaction of isoleucine at residue 489 (whose degree of residue burial is 20) with the glycine (whose degree of residue burial is about 20) is the weakest interaction among the potential amino acid residues at position 489. Hence, most mutations at residue 489 increase the binding affinity of the S protein–receptor complex. On the other hand, the pairwise interaction of glycine at residue 488 (whose degree of residue burial is 20) with the glycine on the receptor (whose degree of residue burial is about 20) is the strongest interaction among the potential amino acid residues at position 488. Hence, most mutations at residue 488 decrease the binding affinity of the S protein–receptor complex (supplementary table S2, Supplementary Material online). Here, the pairwise interaction is denoted by the ratio of the pairwise frequency to the product of individual frequencies in the same local structural environment (eq. 10). The strength of pairwise interaction is measured by amino acid preferences in the local structural environment, and the amino acid preferences are evaluated by using the nonredundant set of protein structures in the PDB.

At residue 480D, the number of interacting residues on the surface of the receptor is three and the distances to those residues are larger than 0.9 nm (supplementary tables S3 and S4, Supplementary Material online). Hence, at residue 480D, the interaction with residues on the surface of the receptor seems to be weak. Except for a mutation to cysteine, no mutation substantially changes the ability to bind to the receptor (supplementary figs. S6A and B, Supplementary Material online). Residue 480D does not seem to contribute substantially to binding to the receptor. However, mutations to hydrophobic or positively charged amino acids at this residue decrease the ability to bind to the antibody 80R (supplementary fig. S6A, Supplementary Material online), so residue 480D may play a key role in

**Table 2.** The Effect of the Mutation I489G on Hydrophobic Interactions and Pairwise Interactions (fixation probability: 0.99 [80R], 0.40 [m396]).

| ACE2                          |          |                      |                      | 80R      |          |                      |                      | m396     |          |                      |                      |
|-------------------------------|----------|----------------------|----------------------|----------|----------|----------------------|----------------------|----------|----------|----------------------|----------------------|
| Total of $\Delta\log_{10}K_a$ |          |                      |                      |          |          |                      |                      |          |          |                      |                      |
| 0.61                          |          |                      |                      | −0.10    |          |                      |                      | 0.48     |          |                      |                      |
| Hydrophobic interaction       |          |                      |                      |          |          |                      |                      |          |          |                      |                      |
| <i>n</i>                      | <i>m</i> | $\Delta\log_{10}K_a$ |                      | <i>n</i> | <i>m</i> | $\Delta\log_{10}K_a$ |                      | <i>n</i> | <i>m</i> | $\Delta\log_{10}K_a$ |                      |
| 14                            | 20       | −0.07                |                      | 15       | 18       | −0.21                |                      | 15       | 27       | −0.21                |                      |
| Pairwise interactions         |          |                      |                      |          |          |                      |                      |          |          |                      |                      |
| <i>r</i>                      | aa       | <i>m</i> ′           | $\Delta\log_{10}K_a$ | <i>r</i> | aa       | <i>m</i> ′           | $\Delta\log_{10}K_a$ | <i>r</i> | aa       | <i>m</i> ′           | $\Delta\log_{10}K_a$ |
| 0.82                          | T        | 16                   | 0.02                 | 0.76     | S        | 17                   | 0.14                 | 0.91     | T        | 25                   | 0.04                 |
| 0.80                          | Q        | 14                   | 0.05                 | 0.90     | Y        | 22                   | −0.01                | 0.81     | D        | 25                   | 0.12                 |
| 0.77                          | G        | 17                   | 0.25                 | 0.94     | Y        | 22                   | −0.01                | 0.79     | T        | 21                   | 0.03                 |
| 0.88                          | K        | 21                   | 0.002                |          |          |                      |                      | 0.60     | V        | 20                   | −0.12                |
| 0.73                          | G        | 21                   | 0.26                 |          |          |                      |                      | 0.62     | M        | 23                   | 0.01                 |
| 0.85                          | D        | 23                   | 0.09                 |          |          |                      |                      | 0.52     | G        | 24                   | 0.14                 |
|                               |          |                      |                      |          |          |                      |                      | 0.84     | G        | 27                   | 0.15                 |
|                               |          |                      |                      |          |          |                      |                      | 0.80     | W        | 21                   | 0.09                 |
|                               |          |                      |                      |          |          |                      |                      | 0.97     | D        | 22                   | 0.07                 |
|                               |          |                      |                      |          |          |                      |                      | 0.95     | S        | 19                   | 0.03                 |
|                               |          |                      |                      |          |          |                      |                      | 0.95     | D        | 24                   | 0.09                 |
|                               |          |                      |                      |          |          |                      |                      | 0.99     | Y        | 23                   | 0.01                 |

NOTE.—*n*, the degree of residue burial of residue 489 in the individual S protein; *m*, the degree of residue burial of residue 489 in the S protein within a protein–protein complex; *r*, the distance (in nm) between the C<sup>α</sup> atom of residue 489 and that of the target residue; aa, the target amino acid residue of the pairwise interaction; and *m*′, the degree of residue burial of the target residue in the target protein within a protein–protein complex.



binding to the antibody 80R (fig. 4A). In the S protein–antibody m396 complex, there is no residue on the surface of the antibody m396 whose C $\alpha$  atom is within a 1-nm sphere around the C $\alpha$  atom of residue 480 of the S protein (figs. 4B and supplementary fig. S6B, Supplementary Material online).

At residue 480D, a mutation to valine, whose residue is hydrophobic, affects the hydrophobic interaction with residues of the antibody 80R, increasing the binding affinity. However, the pairwise interactions with the four residues (R, S, N, D) whose C $\alpha$  atoms are 0.7–0.8 nm away from the C $\alpha$  atom of residue 480 become weaker. As a consequence, the mutation D480V decreases the binding affinity of the complex between the S protein and the antibody 80R (supplementary table S3, Supplementary Material online). At the same time, the mutation D480V does not substantially change the binding affinity of the S protein–receptor complex. Hence, its fixation probability within hosts vaccinated with the antibody 80R is high. The mutation to lysine, whose residue is positively charged, affects the hydrophobic interaction with residues of the antibody 80R, decreasing the binding affinity (supplementary table S4, Supplementary Material online). Therefore, the fixation probability of this mutation within hosts vaccinated with the antibody 80R is also high.

As we have seen previously, residue 480D located at the center of the binding region of the S protein does not contribute substantially to binding to the receptor. On the other hand, the residues located at the edge of the binding region (e.g., 489I, 488G, and 473N) contribute to binding to the receptor. Residue 473N has five contacting sites on the surface of the receptor and also on the surface of the antibody 80R. Most mutations at residue 473N increase the ability to bind to the receptor and decrease the ability to bind to the antibody 80R (supplementary fig. S7, Supplementary Material online). The mutation N473L increases the strength of the hydrophobic interaction both with the receptor and with the antibody 80R. However, the binding affinity of the S protein–antibody 80R complex is decreased by weakened interactions with the two serines whose C $\alpha$  atoms are 0.70–0.75 nm away from the C $\alpha$  atom of residue 473 (supplementary table S5, Supplementary Material online). The fixation probability of this mutation within hosts vaccinated with the antibody 80R is high.

## Discussion

Within a host population vaccinated with the antibody 80R, the mutation D480V shows a high fixation probability. This is a result of decreased binding affinity of the antibody 80R to the S protein. Residue 480D has been previously suggested as a key residue determining binding to the antibody 80R (Sui et al. 2008). It has also been suggested that mutations at this residue do not affect binding to the receptor (Wong et al. 2004). By a mutation to a hydrophobic amino acid from another type of amino acid, the hydrophobic interaction becomes stronger, increasing the binding affinity. However, the binding affinity of the antibody 80R to the S

protein is dominated by pairwise interactions. The mutation D480V decreases the binding affinity. In protein–protein associations, hydrophobic interactions seem to be less dominant (Gohlke et al. 2003). The mutation I489K reduces the intensity of hydrophobic interactions. However, the binding affinity of the S protein–receptor complex increases, whereas that of the S protein–antibody 80R complex decreases. This difference is due to the different effects of the mutation on the pairwise interactions.

We used a mathematical model that describes the dynamics of interacting populations: host cells, virus particles of an ancestor and a mutant, and mAbs. We examined how immune escape occurs when mutations have antagonistic effects on binding antibody and receptors. We simplified this complex, multiscale problem by modeling response to mAbs in order to gain a fundamental understanding of viral evolutionary response to adaptive immunity. Our simplified model takes advantage of the rich supply of protein structural data and sequence data that are readily available for several viruses, providing a framework that can be adapted for other viruses. In future studies, we plan to extend our model to deal with polyclonal antibodies, which is a more realistic immune response.

We found that some amino acid mutations in the RBD may have a high fixation probability (greater than 0.9). Within the range of variation of the SAA mutations on the fitness landscape, some mutants produce sufficient changes in binding strengths to achieve high fixation probabilities within the host population. The variation in binding strengths corresponds to the variation in association constants calculated by using the structural preference of the amino acid sequence, which was estimated using structural data stored in the PDB. Hence, the appearance of the mutants that may be fixed with high probability within the host population is not a grandiose figment but in reality it can happen. The time span of the occurrence of the mutants fixed within the host population can be measured by setting an explicit value of the rate  $\alpha^i$  at which mutations occur at the corresponding residue. By measuring the time span, we can determine the effective period of the vaccine. In development of vaccine, it may be crucial to address the emergencies of such mutations in the RBD.

## Supplementary Material

Supplementary figures S1–S7, tables S1–S5, and other materials are available at *Molecular Biology and Evolution* online (<http://www.mbe.oxfordjournals.org/>).

## Acknowledgments

We thank the three anonymous reviewers for their constructive comments. We thank Keiko Uda for useful comments on the half-life periods of viruses and antibodies. This study was supported by JSPS KAKENHI Grant-in-Aid for Scientific Research (B) 19300094 and Scientific Research (C) 20570223.

## References

- Beerenwinkel N, Schmidt B, Walter H, Kaiser R, Lengauer T, Hoffmann D, Korn K, Selbig J. 2002. Diversity and complexity of HIV-1 drug resistance: a bioinformatics approach to predicting phenotype from genotype. *Proc Natl Acad Sci USA*. 99: 8271–8276.
- Burton DR. 2002. Antibodies, viruses and vaccines. *Nat Rev Immunol*. 2:706–713.
- Cui Q, Sulea T, Schrag JD, Munger C, Hung M-N, Naïm M, Cygler M, Purisima EO. 2008. Molecular dynamics—solvated interaction energy studies of protein-protein interactions: the MP1-p14 scaffolding complex. *J Mol Biol*. 379:787–802.
- Elcock AH, Sept D, McCammon JA. 2001. Computer simulation of protein-protein interactions. *J Phys Chem B*. 105:1504–1518.
- Gohlke H, Kiel C, Case DA. 2003. Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes. *J Mol Biol*. 330:891–913.
- Hobohm U, Sander C. 1994. Enlarged representative set of protein structures. *Protein Sci*. 3:522–524.
- Holmes KV. 2003. SARS coronavirus: a new challenge for prevention and therapy. *J Clin Invest*. 111:1605–1609.
- Hwang WC, Lin Y, Santelli E, Sui J, Jaroszewski L, Stec B, Farzan M, Marasco WA, Liddington RC. 2006. Structural basis of neutralization by a human anti-severe acute respiratory syndrome spike protein antibody, 80R. *J Biol Chem*. 281: 34610–34616.
- Insel RA, Looney RJ. 2004. The B-lymphocyte system: fundamental immunology. In: Stiehm ER, Ochs HD, Winkelstein JA, editors. *Immunologic disorders in infants & children*, 5th ed. Philadelphia (PA): Elsevier. p. 53–84.
- Kimura M. 1962. On the probability of fixation of mutant genes in a population. *Genetics* 47:713–719.
- Ksiazek TG, Erdman D, Goldsmith CS, et al. (26 co-authors). 2003. A novel coronavirus associated with severe acute respiratory syndrome. *N Engl J Med*. 348:1953–1965.
- Li F, Li W, Farzan M, Harrison SC. 2005. Structure of SARS coronavirus spike receptor-binding domain complexed with receptor. *Science* 309:1864–1868.
- Li F, Long T, Lu Y, Ouyang Q, Tang C. 2004. The yeast cell-cycle network is robustly designed. *Proc Natl Acad Sci USA*. 101: 4781–4786.
- Li W, Moore MJ, Vasilieva N, et al. (12 co-authors). 2003. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. *Nature* 426:450–454.
- Li W, Zhang C, Sui J, et al. (16 co-authors). 2005. Receptor and viral determinants of SARS-coronavirus adaptation to human ACE2. *EMBO J*. 24:1634–1643.
- Neumann AU, Lam NP, Dahari H, Gretch DR, Wiley TE, Layden TJ, Perelson AS. 1998. Hepatitis C viral dynamics in vivo and the antiviral efficacy of interferon- $\alpha$  therapy. *Science* 282:103–107.
- Nowak MA, Bangham CRM. 1996. Population dynamics of immune responses to persistent viruses. *Science* 272:74–79.
- Nowak MA, Bonhoeffer S, Hill AM, Boehme R, Thomas HC, McDade H. 1996. Viral dynamics in hepatitis B virus infection. *Proc Natl Acad Sci USA*. 93:4398–4402.
- Perelson AS, Neumann AU, Matkowitz M, Leonard JM, Ho DD. 1996. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science* 271:1582–1586.
- Prabakaran P, Gan J, Feng Y, Zhu Z, Choudhry V, Xiao X, Ji X, Dimitrov DS. 2006. Structure of severe acute respiratory syndrome coronavirus receptor-binding domain complexed with neutralizing antibody. *J Biol Chem*. 281:15829–15836.
- Rota PA, Oberste MS, Monroe SS, et al. (35 co-authors). 2003. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* 300:1394–1399.
- Rutherford SL. 2000. From genotype to phenotype: buffering mechanisms and the storage of genetic information. *Bioessays* 22:1095–1105.
- Simons KT, Bonneau R, Ruczinski I, Baker D. 1999a. Ab initio protein structure prediction of CASP III targets using ROSETTA. *Proteins*. 37(3 Suppl):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.
- Song H-D, Tu C-C, Zhang G-W, et al. (59 co-authors). 2005. Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. *Proc Natl Acad Sci USA*. 102: 2430–2435.
- Sui J, Aird DR, Tamin A, et al. (16 co-authors). 2008. Broadening of neutralization activity to directly block a dominant antibody-driven SARS-coronavirus evolution pathway. *PLoS Pathog*. 4:e1000197.
- Sui J, Li W, Murakami A, et al. (14 co-authors). 2004. Potent neutralization of severe acute respiratory syndrome (SARS) coronavirus by a human mAb to S1 protein that blocks receptor association. *Proc Natl Acad Sci USA*. 101:2536–2541.
- Wagner A. 2000. Robustness against mutations in genetic networks of yeast. *Nat Genet*. 24:355–361.
- Watabe T, Kishino H, de Oliveira Martins L, Kitazoe Y. 2007. A likelihood-based index of protein-protein binding affinities with application to influenza HA escape from antibodies. *Mol Biol Evol*. 24:1627–1638.
- 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.
- Wong SK, Li W, Moore MJ, Choe H, Farzan M. 2004. A 193-amino acid fragment of the SARS coronavirus S protein efficiently binds angiotensin-converting enzyme 2. *J Biol Chem*. 279: 3197–3201.
- World Health Organization. 2003. Outbreak news—severe acute respiratory syndrome (SARS). *Wkly Epidemiol Rec*. 78:81–83.
- Wright S. 1931. Evolution in Mendelian populations. *Genetics* 16:97–159.