Available online at www.sciencedirect.com
SCIENCE DIRECT®
Gene 328 (2004) 127–133
GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES
www.elsevier.com/locate/gene

Negative selection on neutralization epitopes of poliovirus surface proteins: implications for prediction of candidate epitopes for immunization

Yoshiyuki Suzuki*

Center for Information Biology and DNA Data Bank of Japan, National Institute of Genetics, 1111 Yata, Mishima-shi, Shizuoka-ken 411-8540, Japan

Received 10 October 2003; accepted 1 November 2003

Received by S. Yokoyama

Abstract

For development of effective vaccines against viruses, it is of importance to choose appropriate epitopes as the target for immunization. These epitopes should eventually be determined experimentally, but it would be helpful if we could predict candidate epitopes computationally because it accelerates the entire process. To predict candidate epitopes for immunization, it is of great interest to characterize the target epitopes of poliovirus vaccine, which has empirically proven to be the most effective among all vaccines available. Here I show that almost all amino acid sites of poliovirus surface proteins VP1, VP2, and VP3 including neutralization epitopes are negatively selected and no site is under positive selection. These results, together with those obtained in previous studies, indicate that vaccines directed against epitopes, which consist of negatively selected sites protect vaccinees more effectively than those directed against epitopes which contain positively selected sites. These observations suggest that candidate epitopes for immunization are predicted by the molecular evolutionary analysis of viral protein (and its coding nucleotide) sequences, as the epitopes which consist exclusively of negatively selected amino acid sites.

Keywords: Poliovirus; Epitope; Positive selection; Negative selection; Vaccine

1. Introduction

Modern viral vaccine was first established by Edward Jenner in 1796 against smallpox virus, and it was declared that naturally occurring smallpox viruses were eradicated in 1980. Vaccines have subsequently been licensed against adenovirus, poliovirus, hepatitis A and B, influenza A and B, Japanese encephalitis, measles, mumps, rabies, rubella, varicella-zoster, and yellow fever viruses (Murphy and Chanock, 2001). Among these vaccines, the most successful one appears to be that against poliovirus, although the vaccine itself has occasionally caused epidemics. Wild polioviruses have already been eradicated from the Americas, Western Pacific region, and Europe, and is planned to be eradicated globally by the end of 2005.

Vaccines protect vaccinees against viral infection by eliciting strong adaptive immune responses which are mediated by antibodies and cytotoxic T lymphocytes (CTLs). The antibodies mainly inhibit infection of viruses to cells (neutralization) through binding to antigenic sites (neutralization epitopes) of viral surface proteins, whereas CTLs eliminate infected cells through recognizing processed peptides (CTL epitopes) of viral proteins, which are presented on cell surface together with the heterodimer of major histocompatibility complex class I and β_2-microglobulin molecules. It is considered that antibodies play a major role in protection against viral infection (Murphy and Chanock, 2001).

Abbreviations: VP, viral protein; N-Ag, neutralization antigenic site; DDBJ, DNA Data Bank of Japan; $r_{S(N)}$, rate of synonymous (nonsynonymous) substitution; $c_{S(N)}$, number of synonymous (nonsynonymous) substitutions per codon site; $s_{S(N)}$, number of synonymous (nonsynonymous) sites per codon site; $d_{S(N)}$, number of synonymous (nonsynonymous) substitutions per synonymous (nonsynonymous) site; R, transition/transversion ratio.

* Tel.: +81-55-981-6847; fax: +81-55-981-6848.
E-mail address: yossuzuk@lab.nig.ac.jp (Y. Suzuki).

0378-1119/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2003.11.020

For development of effective vaccines against viruses, it is of importance to choose appropriate epitopes as the target for immunization. For example, it has been suggested that conserved epitopes are more appropriate as the target than variable epitopes from the mathematical modeling of viral infection dynamics (Nowak et al., 1991). These epitopes should eventually be determined experimentally, but it would be helpful if we could predict candidate epitopes computationally because it accelerates the entire process, which is particularly important when vaccines are developed against emerging viruses. Computational methods have been established for predicting epitope regions in the amino acid sequence of a given protein (Petrovsky and Brusic, 2002). It is however necessary to choose appropriate epitopes for immunization among them. To predict candidate epitopes for immunization, it may be useful to characterize the neutralization and CTL epitopes of various viruses. In particular, if we could extract common features of epitopes among viruses against which vaccines are empirically known to be highly effective, such information may be used to predict candidate epitopes for immunization against other viruses.

One of the prominent features of the neutralization and CTL epitopes of various viruses is that they contain positively selected sites, which is actually the case for almost all viruses so far examined. These include not only viruses against which vaccines are not available (Australian bat lyssavirus (Guyatt et al., 2003), human immunodeficiency virus type 1 (Suzuki and Gojobori, 1999), dengue (Twiddy et al., 2002), hepatitis C (Suzuki and Gojobori, 2001), respiratory syncytial (Woelk and Holmes, 2001), and Venezuelan equine encephalitis viruses (Brault et al., 2002)) but also those against which vaccines are available (influenza A (Suzuki and Gojobori, 1999) and measles viruses (Woelk et al., 2001)). However, it has been reported that measles virus continued to circulate even in fully vaccinated populations (Gustafson et al., 1987) and sera of vaccinees often failed to neutralize wild-type strains (Klingele et al., 2000). In addition, vaccines against influenza A virus have to be annually updated because of its rapid antigenic change but still sometimes fail to protect vaccinees (Kilbourne et al., 2002). By contrast, it was reported that the neutralization epitopes of rabies virus contained no positively selected site (Holmes et al., 2002). Interestingly, for rabies virus, there has been no report that vaccines failed to protect vaccinees due to antigenic variation (World Health Organization, 1992). These observations suggest the possibility that the candidate epitopes for immunization are predicted as the epitopes, which contain no positively selected site but consist exclusively of negatively selected sites. To test this hypothesis, it is of great interest to examine natural selection operating on the target epitopes of poliovirus vaccine, which has empirically proven to be the most effective among all vaccines available.

The purpose of this paper is to infer natural selection at each amino acid site of the neutralization epitopes of poliovirus surface proteins. Implications of the results for prediction of candidate epitopes for immunization are discussed.

2. Materials and methods

2.1. Neutralization epitopes of polioviruses

Poliovirus is the etiological agent of poliomyelitis. This virus belongs to the genus *Enterovirus* of the family *Picornaviridae* and possesses a nonsegmented, single-stranded, and positive-sense RNA genome with about 7400 nucleotide sites in a nonenveloped virion. The genome encodes a polyprotein of about 2200 amino acid sites, which is cleaved into structural proteins, viral protein (VP)1, VP2, VP3, and VP4, and nonstructural proteins 2A, 2B, 2C, 3A, 3B, 3C, and 3D. Inactivated and live-attenuated poliovirus vaccines were developed by Jonas Salk and Julius S. Youngner in 1955 and by Albert B. Sabin in 1961, respectively. In both cases, the target for immunization is neutralization epitopes N-AgI, N-AgIB, N-AgII, N-AgIII, and N-AgIV, which also determine serotypes 1, 2, and 3 of polioviruses. These epitopes are located in VP1, VP2, and VP3, which are exposed on the virion surface. The amino acid positions of these proteins comprising each epitope are as follows; positions 91–101 and 144 of VP1 for N-AgI, positions 96–104 and 141–152 of VP1 for N-AgIB, positions 221–226 of VP1 and 138, 142, 164–170, and 270 of VP2 for N-AgII, positions 286–290 of VP1 and 58–60, 70–73, 76, 77, and 79 of VP3 for N-AgIII, and positions 72 of VP2 and 76 of VP3 for N-AgIV (Fiore et al., 1997). All amino acid positions refer to those in serotype 1 poliovirus proteins throughout this paper.

2.2. Sequence data

Nucleotide sequences for VP1, VP2, and VP3 genes of serotypes 1, 2, and 3 polioviruses were extracted from the international nucleotide sequence database (DNA Data Bank of Japan [DDBJ] release 53). For VP1, I collected only the sequences of wild-type strains because vaccine-derived strains may have experienced biased natural selection during attenuation. I also eliminated recombinant sequences because it is inappropriate to construct a phylogenetic tree including both the recombinant and nonrecombinant sequences, which is required for inference of natural selection. After further excluding sequences with minor gaps and ambiguous nucleotides, I obtained 118, 2, and 23 VP1 sequences for serotypes 1, 2, and 3, respectively. It has been reported that a relatively large number of sequences are required for analysis of natural selection at single amino acid sites (Suzuki and Gojobori, 1999; Yang et al., 2000). I therefore analyzed only serotypes 1 and 3, separately. However, to obtain some insights into natural selection operating on serotype 2, I also analyzed the total of 143 sequences of all serotypes.

For each of VP2 and VP3, however, the total number of sequences of wild-type strains including all serotypes was only four, which was obviously too small for analysis of natural selection. Here, to obtain sufficient numbers of sequences for analysis, I added sequences of vaccine-derived strains for the following rationales. First, although the vaccine-derived strains may have experienced biased natural selection during attenuation, the time required for attenuation is considerably small compared with the evolutionary history of serotypes 1, 2, and 3 polioviruses. Accordingly, the number of nucleotide substitutions, which occurred during attenuation, should also be small, and thus the bias may not affect the inference of natural selection to a large extent. Second, the bias, if any, should be directed largely toward positive selection because mutants which were more adapted to the cell culture were repeatedly selected for during attenuation starting with the wild-type strain which was originally not adapted to it (Roivainen et al., 1993). It is unlikely that negative selection is enhanced during attenuation because the probability that a given amino acid site of wild-type strain is occupied by the amino acid which is not optimum to the wild but to the cell culture appears to be low. Therefore, we may have to be cautious when positive selection is inferred but still obtain some insights into negative selection operating in the wild. Third, nucleotide substitutions, which actually occurred during attenuation, are sometimes well characterized through comparison between the sequences of vaccine and its progenitor (wild-type) strains (Nomoto et al., 1982). It has been reported that nonsynonymous substitutions occurred at codon positions 142 and 165 of VP2 and 60 of VP3, which are parts of neutralization epitopes. Therefore, even if positive selection is inferred at these sites, it may be considered as artifacts. The total numbers of VP2 and VP3 sequences for wild-type and vaccine-derived strains of all serotypes were 25 and 31, respectively. These sequences were analyzed collectively because each serotype did not contain enough number of sequences for analysis.

2.3. Data analysis

For each set of nucleotide sequences, a multiple alignment was made using the computer program CLUSTAL W (version 1.81) (Thompson et al., 1994). No gap was introduced into each of the alignments of serotypes 1 and 3 VP1 sequences, which included 906 and 900 nucleotide sites, respectively. However, the alignment of VP1 sequences of all serotypes (serotype 2 VP1 sequences had 903 nucleotide sites) contained gaps, some of which were apparently accompanied by frame-shifts. To eliminate these gaps, I also made an alignment using amino acid sequences and extracted only the codon sites, which were consistently aligned using the nucleotide and amino acid sequences. As a result, 816 nucleotide sites were useful for analysis of natural selection. For VP2, serotype 1 sequences (816 nucleotide sites) were longer than serotypes 2 and 3 sequences (813 nucleotide sites) by three nucleotide sites without frameshift, so that these sites were removed from the alignment. The lengths of VP3 sequences were the same (714 nucleotide sites) for all serotypes and the alignment did not contain any gap.

Natural selection can be detected by comparing the rate of synonymous nucleotide substitution (r_S) with that of nonsynonymous nucleotide substitution (r_N). The relationship $r_S < r_N$ indicates positive selection whereas $r_S > r_N$ indicates negative selection. The parsimony-based (Suzuki, 1999) and likelihood-based (Yang et al., 2000) methods are available for inferring natural selection at single amino acid sites. Both methods utilize a phylogenetic tree of nucleotide sequences. In the parsimony-based method, we compute the total numbers of synonymous (c_S) and nonsynonymous (c_N) substitutions as well as the average numbers of synonymous (s_S) and nonsynonymous (s_N) sites for the entire phylogenetic tree at each codon site according to the maximum parsimony principle. The numbers of synonymous (d_S) and nonsynonymous (d_N) substitutions per synonymous and nonsynonymous sites at a codon site are given by c_S/s_S and c_N/s_N, respectively. The probability (p) of obtaining the observed or more biased numbers for c_S and c_N was computed under the assumption that c_S and c_N are binomially distributed with the probabilities of occurrence of synonymous and nonsynonymous substitutions given by $s_S/(s_S + s_N)$ and $s_N/(s_S + s_N)$, respectively (selective neutrality). The null hypothesis of selective neutrality is rejected if p is smaller than 5%. If neutrality is rejected, $d_S < d_N$ indicates positive selection, whereas $d_S > d_N$ indicates negative selection.

In the likelihood-based method, the ratio r_N/r_S is assumed to follow a certain probability distribution among codon sites in the sequence. Fourteen different distributions were initially proposed to be used, but model M8 is empirically known to produce more reliable results than others. In this model, codon sites are classified into categories 0 and 1, which exist with proportions p_0 and p_1, respectively. In category 0, r_N/r_S is assumed to follow a beta distribution with parameters a and b within the range $0 \le r_N/r_S \le 1$, whereas it takes a particular value in category 1. Free parameters are estimated by the maximum likelihood method. If the estimate of r_N/r_S in category 1 was larger than unity, positively selected sites may exist in the sequence. We compute the posterior probability that a given codon site belongs to categories 0 and 1 and if it was higher than 95%, negative and positive selection are inferred, respectively. Note that if the estimate of r_N/r_S in category 1 was smaller than unity, all codon sites in the sequence are inferred as negatively selected with 100% posterior probability.

For each set of sequences, a phylogenetic tree was constructed by the neighbor-joining method (Saitou and Nei, 1987) with the number of nucleotide substitutions estimated using Kimura's (1980) model. ADAPTSITE (version 1.3) (Suzuki et al., 2001) was used for inferring natural selection at single amino acid sites by the parsimony-based

method. I assumed Kimura's model as the pattern of nucleotide mutation. The transition/transversion ratio (R) was estimated as the ratio of the transitional to transversional nucleotide diversity. The R-values estimated were 3.23, 3.00, and 2.01 for serotypes 1, 3, and all serotypes of VP1 sequences, respectively, and 1.41 and 1.33 for VP2 and VP3 sequences, respectively. PAML (version 3.13) (Yang, 1997) was used for inferring natural selection at single amino acid sites by the likelihood-based method. R was also estimated in this method.

3. Results

3.1. Inference of natural selection by the parsimony-based method

The results for inference of natural selection at each amino acid site of VP1, VP2, and VP3 of polioviruses by the parsimony-based method are summarized in Table 1. Interestingly, there was no signal for positive selection at all sites of all proteins including neutralization epitopes. Instead, most sites were inferred as negatively selected in all proteins. For serotype 1 VP1 sequences, d_S was larger than d_N at 293 out of 302 codon sites (97.0%) and negative selection was inferred at 284 sites (94.0%). Among the remaining nine sites, $d_S < d_N$ at only two sites (0.7%) and no nucleotide substitution was observed at seven sites (2.3%). The former sites were positions 5 and 175, which were outside the neutralization epitopes. Since the p-value for rejecting selective neutrality was unity at both sites, these results could be explained by the stochastic error under neutral evolution. Among 37 sites which were involved in the neutralization epitopes, $d_S > d_N$ at all sites (100%) and negative selection was inferred at 34 sites (91.9%).

Similarly, $d_S > d_N$ at 94.3% of all codon sites for serotype 3 VP1 sequences. However, the proportion of negatively selected sites was 65.0%, which was smaller than that observed for serotype 1. This result was obtained probably because the power of the test of selective neutrality for serotype 3 was lower than that for serotype 1, due to the smaller number of sequences analyzed for the former than the latter. $d_S < d_N$ at no site (0%). Among the codon sites that were involved in the neutralization epitopes, 97.3% showed $d_S > d_N$ and 54.1% were inferred as negatively selected.

When VP1 sequences of all serotypes were analyzed, $d_S > d_N$ at 97.8% of all sites and negative selection was inferred at almost all sites (97.1%) where $d_S > d_N$. By contrast, $d_S < d_N$ at only one site (0.4%), which was again position 175 with $p = 1$. Importantly, all sites (100%) in the neutralization epitopes were inferred as negatively selected. In summary, the results for serotypes 1, 3, and all serotypes of VP1 sequences indicate that almost all sites, including those in the neutralization epitopes, are negatively selected. A few sites may evolve neutrally, but no site is under positive selection.

For VP2, $d_S > d_N$ at 253 out of 271 codon sites (93.4%) and negative selection was inferred at 164 sites (60.5%) (Table 1). This pattern was similar to that for serotype 3 VP1 sequences, and probably caused by the relatively small number of sequences analyzed. $d_S < d_N$ at six sites (2.2%). They were positions 70, 142, 165, 171, 173, and 187, of which 142 and 165 were located in N-AgII. Note that

Table 1
Results of parsimony-based method for inferring natural selection at single amino acid sites of VP1, VP2, and VP3 of polioviruses

Protein (serotype)	Region	$d_S = d_N = 0$	$d_S < d_N$			$d_S > d_N$			All
			Positive[a]	N.S.[b]	All	Negative[a]	N.S.	All	
VP1 (1)	epitope	0 (0%)[c]	0 (0%)	0 (0%)	0 (0%)	34 (91.9%)	3 (8.1%)	37 (100%)	37 (100%)
	other	7 (2.6%)	0 (0%)	2 (0.8%)	2 (0.8%)	250 (94.3%)	6 (2.3%)	256 (96.6%)	265 (100%)
	all	7 (2.3%)	0 (0%)	2 (0.7%)	2 (0.7%)	284 (94.0%)	9 (3.0%)	293 (97.0%)	302 (100%)
VP1 (3)	epitope	1 (2.7%)	0 (0%)	0 (0%)	0 (0%)	20 (54.1%)	16 (43.2%)	36 (97.3%)	37 (100%)
	other	16 (6.1%)	0 (0%)	0 (0%)	0 (0%)	175 (66.5%)	72 (27.4%)	247 (93.9%)	263 (100%)
	all	17 (5.7%)	0 (0%)	0 (0%)	0 (0%)	195 (65.0%)	88 (29.3%)	283 (94.3%)	300 (100%)
VP1 (all)	epitope	0 (0%)	0 (0%)	0 (0%)	0 (0%)	33 (100%)	0 (0%)	33 (100%)	33 (100%)
	other	5 (2.1%)	0 (0%)	1 (0.4%)	1 (0.4%)	231 (96.7%)	2 (0.8%)	233 (97.5%)	239 (100%)
	all	5 (1.8%)	0 (0%)	1 (0.4%)	1 (0.4%)	264 (97.1%)	2 (0.7%)	266 (97.8%)	272 (100%)
VP2 (all)	epitope	0 (0%)	0 (0%)	2 (18.2%)	2 (18.2%)	1 (9.1%)	8 (72.7%)	9 (81.8%)	11 (100%)
	other	12 (4.6%)	0 (0%)	4 (1.5%)	4 (1.5%)	163 (62.7%)	81 (31.2%)	244 (93.8%)	260 (100%)
	all	12 (4.4%)	0 (0%)	6 (2.2%)	6 (2.2%)	164 (60.5%)	89 (32.8%)	253 (93.4%)	271 (100%)
VP3 (all)	epitope	0 (0%)	0 (0%)	0 (0%)	0 (0%)	2 (20%)	8 (80%)	10 (100%)	10 (100%)
	other	14 (6.1%)	0 (0%)	1 (0.4%)	1 (0.4%)	154 (67.5%)	59 (25.9%)	213 (93.4%)	228 (100%)
	all	14 (5.9%)	0 (0%)	1 (0.4%)	1 (0.4%)	156 (65.5%)	67 (28.2%)	223 (93.7%)	238 (100%)
All[d] (all)	epitope	0 (0%)	0 (0%)	2 (3.7%)	2 (3.7%)	36 (66.7%)	16 (29.6%)	52 (96.3%)	54 (100%)
	other	31 (4.3%)	0 (0%)	6 (0.8%)	6 (0.8%)	548 (75.4%)	142 (19.5%)	690 (94.9%)	727 (100%)
	all	31 (4.0%)	0 (0%)	8 (1.0%)	8 (1.0%)	584 (74.8%)	158 (20.2%)	742 (95.0%)	781 (100%)

[a] Positive and negative selection were inferred when $d_S < d_N$ and $d_S > d_N$ with $p < 0.05$, respectively.
[b] Not significant.
[c] Percentage relative to the total number of amino acid sites examined as indicated in the right-most column.
[d] Results for VP1, VP2, and VP3 sequences of all serotypes were summed.

positions 142 and 165 are the sites where occurrence of nonsynonymous substitutions during attenuation has been reported, as mentioned above. It is therefore possible that positive selection operated at these sites during attenuation. However, $p = 1$ at all six sites except for position 187, where p was still 0.754, indicating that the results can be explained by the stochastic error under neutral evolution. These observations suggest that even if positive selection operated at positions 142 and 165 during attenuation, it was cancelled out by negative selection which had probably operated over the evolutionary history of serotypes 1, 2, and 3 polioviruses. Indeed, $d_S > d_N$ at the remaining nine sites in the neutralization epitopes. Therefore, most sites, including those in the neutralization epitopes, appear to be negatively selected in VP2. Positive selection is unlikely to operate at any site.

Similarly to the cases for serotype 3 VP1 and for VP2, $d_S > d_N$ at 223 out of 238 codon sites (93.7%) of VP3 and negative selection was inferred at 156 sites (65.5%) (Table 1). $d_S < d_N$ at only one site (0.4%). This site was position 236, which was outside the neutralization epitopes. This result could again be explained by the stochastic error under neutral evolution because $p = 1$. $d_S > d_N$ at all 10 sites involved in the neutralization epitopes. These results indicate that most sites, including those in the neutralization epitopes, are negatively selected in VP3. Positive selection is again unlikely to operate at any site. In total, $d_S > d_N$ at 95.0% and negative selection was inferred at 74.8% of all sites of VP1, VP2, and VP3. By contrast, $d_S < d_N$ at only 1.0% of all sites and positive selection was not inferred at any site (0%). Similarly, $d_S > d_N$ at 96.3% and negative selection was inferred at 66.7% of sites in the neutralization epitopes of VP1, VP2, and VP3. $d_S < d_N$ at only 3.7% of the sites and positive selection was not inferred at any site (0%).

3.2. Inference of natural selection by the likelihood-based method

Interestingly, a more extreme picture of natural selection operating at each amino acid site of VP1, VP2, and VP3 was obtained by the likelihood-based method (Table 2). That is, not only was there no signal for positive selection at all sites of all proteins including neutralization epitopes, but also was every site of all proteins inferred to be negatively selected with 100% posterior probability. For serotype 1 VP1 sequences, the R-value was estimate to be 3.46, which was similar to that estimated for the parsimony-based method (3.23), indicating that these values were more or less reliable. The estimates of the proportions of categories 0 and 1 were 0.975 and 0.025, respectively. In category 0, a and b were estimated to be 0.409 and 16.8, respectively, suggesting that the beta distribution is L-shaped and highly skewed toward $r_N/r_S = 0$. In addition, r_N/r_S in category 1 was estimated to be 0.342, which was much smaller than unity. Therefore, all sites including those in the neutralization epitopes were inferred as negatively selected with 100% posterior probability.

Similar results were obtained when I analyzed serotype 3 VP1, all serotypes of VP1, VP2, and VP3 sequences. Therefore, every site of VP1, VP2, and VP3 including neutralization epitopes was inferred as negatively selected with 100% posterior probability. Similar results were again obtained when the branch lengths of the phylogenetic tree were also estimated in this method (data not shown).

4. Discussion

4.1. Reliability of the results

The reliability of the parsimony-based and likelihood-based methods for inferring natural selection at single amino acid sites has been assessed by computer simulation and actual data analysis (Suzuki and Nei, 2001, 2002). It has been reported that the parsimony-based method is generally conservative for inferring both positively and negatively selected sites. By contrast, the likelihood-based method sometimes becomes liberal and tends to give false-positive results of positively selected sites, although the reliability of negatively selected sites inferred has not been studied well. These observations indicate that the results obtained in this study, namely the absence of positively selected sites and the abundance of negatively selected sites are conservative and reliable. This assertion is even strengthened by the fact that the vaccine-derived strains, which may have experienced extra positive selection during attenuation, were included in the analysis of VP2 and VP3. I conclude that almost all amino acid sites of VP1, VP2, and VP3 including neutral-

Table 2
Results of likelihood-based method for inferring natural selection at single amino acid sites of VP1, VP2, and VP3 of polioviruses

Protein	Serotype	R	Category 0			Category 1		Positive[a]	Negative[a]
			p_0	a	b	p_1	r_N/r_S		
VP1	1	3.46	0.975	0.409	16.8	0.025	0.342	0 (0%)[b]	302 (100%)
	3	2.77	0.948	0.008	0.232	0.052	0.250	0 (0%)	300 (100%)
	all	2.72	0.986	0.388	17.9	0.014	0.295	0 (0%)	272 (100%)
VP2	all	1.38	0.931	0.278	13.6	0.069	0.354	0 (0%)	271 (100%)
VP3	all	1.61	0.983	0.111	3.79	0.017	0.590	0 (0%)	238 (100%)

[a] Numbers of positively and negatively selected sites, respectively. Note that all negatively selected sites were inferred with 100% posterior probability.
[b] Percentage relative to the total number of amino acid sites examined.

ization epitopes of polioviruses are negatively selected and no site is under positive selection. The amino acid sites in the neutralization epitopes are negatively selected probably because they are also important for viral infectivity (Spriggs and Fields, 1982) and thereby functionally constrained. In fact, it has been reported that the neutralization epitopes of polioviruses are also involved in receptor recognition and formation of altered-particles, which are the intermediates for entry into cells (Domingo et al., 1993). Since almost all codon sites of VP1, VP2, and VP3 outside the neutralization epitopes were also inferred as negatively selected, the conclusion would hold even if there were additional and yet undetermined neutralization epitopes in these proteins.

4.2. Negatively selected epitopes as the candidate for immunization

The results obtained in this study, together with those obtained in the previous studies as mentioned above, indicate that vaccines directed against epitopes (of poliovirus and rabies virus) which consist of negatively selected sites protect vaccinees more effectively than those directed against epitopes (of influenza A and measles viruses) which contain positively selected sites. These observations appear to be consistent with the results from the mathematical modeling of viral infection dynamics, where it was suggested that conserved epitopes were more appropriate as the target for immunization than variable epitopes, as mentioned above. It should be noted, however, that viral proteins usually contain multiple epitopes, and even when some of them contained positively selected sites, others may consist of negatively selected sites. In fact, some epitopes of influenza A and measles viruses did not appear to contain positively selected sites (Suzuki and Gojobori, 1999; Woelk et al., 2001). However, vaccines against these viruses appeared to be less effective probably because the epitopes, which contained positively selected sites, were immunodominant and viruses escaped from immune responses by producing antigenic variation at positively selected sites. In addition, it has been suggested by the mathematical modeling that antigenic variation in the immunodominant epitopes can shift immune responses toward epitopes with low antigenicity (Nowak et al., 1995). These observations suggest that only negatively selected epitopes should be used for immunization.

In conclusion, candidate epitopes for immunization against viruses may be predicted by the molecular evolutionary analysis, as the epitopes which consist exclusively of negatively selected amino acid sites. Since computational methods have also been established for predicting epitope regions in the amino acid sequence of a given protein as mentioned above, it may be possible to predict candidate epitopes only from the sequence data. This information may be useful for development of vaccines particularly when experimental data are scarce, as in the cases against emerging viruses.

Acknowledgements

The author thanks two anonymous reviewers for their valuable comments.

References

Brault, A.C., Powers, A.M., Holmes, E.C., Woelk, C.H., Weaver, S.C., 2002. Positively charged amino acid substitutions in the E2 envelope glycoprotein are associated with the emergence of Venezuelan equine encephalitis virus. J. Virol. 76, 1718–1730.

Domingo, E., Diez, J., Martinez, M.A., Hernandez, J., Holguin, A., Borrego, B., Mateu, M.G., 1993. New observations on antigenic diversification of RNA viruses: antigenic variation is not dependent on immune selection. J. Gen. Virol. 74, 2039–2045.

Fiore, L., Ridolfi, B., Genovese, D., Buttinelli, G., Lucioli, S., Lahm, A., Ruggeri, F.M., 1997. Poliovirus Sabin type 1 neutralization epitopes recognized by immunoglobulin A monoclonal antibodies. J. Virol. 71, 6905–6912.

Gustafson, T.L., Lievens, A.W., Brunell, P.A., Moellenberg, R.G., Buttery, C.M., Sehulster, L.M., 1987. Measles outbreak in a fully immunized secondary-school population. N. Engl. J. Med. 316, 771–774.

Guyatt, K.J., Davis, T.J., Holmes, E.C., Smith, G.A., Smith, I.L., Mackenzie, J.S., Young, P.L., 2003. A molecular epidemiological study of Australian bat lyssavirus. J. Gen. Virol. 84, 485–496.

Holmes, E.C., Woelk, C.H., Kassis, R., Bourhy, H., 2002. Genetic constraints and the adaptive evolution of rabies virus in nature. Virology 292, 247–257.

Kilbourne, E.D., Smith, C., Brett, I., Pokorny, B.A., Johansson, B., Cox, N., 2002. The total influenza vaccine failure of 1947 revisited: major intrasubtypic antigenic change can explain failure of vaccine in a post-World War II epidemic. Proc. Natl. Acad. Sci. U. S. A. 99, 10748–10752.

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

Klingele, M., Hartter, H.K., Adu, F., Ammerlaan, W., Ikusika, W., Muller, C.P., 2000. Resistance of recent measles virus wild-type isolates to antibody-mediated neutralization by seroconverted vaccines with antibody. J. Med. Virol. 62, 91–98.

Murphy, B.R., Chanock, R.M., 2001. Immunization against viral diseases. In: Knipe, D.M., Howley, P.M., Griffin, D.E., Lamb, R.A., Martin, M.A., Roizman, B., Straus, S.E. (Eds.), Fields Virology, 4th ed. Lippincott Williams and Wilkins, Philadelphia.

Nomoto, A., Omata, T., Toyoda, H., Kuge, S., Horie, H., Kataoka, Y., Genba, Y., Nakano, Y., Imura, N., 1982. Complete nucleotide sequence of the attenuated poliovirus Sabin 1 strain genome. Proc. Natl. Acad. Sci. U. S. A. 79, 5793–5797.

Nowak, M.A., Anderson, R.M., McLean, A.R., Wolfs, T.F., Goudsmit, J., May, R.M., 1991. Antigenic diversity thresholds and the development of AIDS. Science 254, 963–969.

Nowak, M.A., May, R.M., Phillips, R.E., Rowland-Jones, S., Lalloo, D.G., McAdam, S., Klenerman, P., Koppe, B., Sigmund, K., Bangham, C.R.M., McMichael, A.J., 1995. Antigenic oscillations and shifting immunodominance in HIV-1 infections. Nature 375, 606–611.

Petrovsky, N., Brusic, V., 2002. Computational immunology: the coming of age. Immunol. Cell Biol. 80, 248–254.

Roivainen, M., Piirainen, L., Hovi, T., 1993. Persistence and class-specificity of neutralizing antibody response induced by trypsin-cleaved type 3 poliovirus in mice. Vaccine 11, 713–717.

Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.

Spriggs, D.R., Fields, B.N., 1982. Attenuated reovirus type 3 strains generated by selection of haemagglutinin antigenic variants. Nature 297, 68–70.

Suzuki, Y., 1999. Molecular Evolution of Pathogenic Viruses. PhD dissertation, Department of Genetics, School of Life Science, The Graduate University of Advanced Studies, Hayama, Japan.

Suzuki, Y., Gojobori, T., 1999. A method for detecting positive selection at single amino acid sites. Mol. Biol. Evol. 16, 1315–1328.

Suzuki, Y., Gojobori, T., 2001. Positively selected amino acid sites in the entire coding region of hepatitis C virus subtype 1b. Gene 276, 83–87.

Suzuki, Y., Nei, M., 2001. Reliabilities of parsimony-based and likelihood-based methods for detecting positive selection at single amino acid sites. Mol. Biol. Evol. 18, 2179–2185.

Suzuki, Y., Nei, M., 2002. Simulation study of the reliability and robustness of the statistical methods for detecting positive selection at single amino acid sites. Mol. Biol. Evol. 19, 1865–1869.

Suzuki, Y., Gojobori, T., Nei, M., 2001. ADAPTSITE: detecting natural selection at single amino acid sites. Bioinformatics 17, 660–661.

Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.

Twiddy, S.S., Woelk, C.H., Holmes, E.C., 2002. Phylogenetic evidence for adaptive evolution of dengue viruses in nature. J. Gen. Virol. 83, 1679–1689.

Woelk, C.H., Holmes, E.C., 2001. Variable immune-driven natural selection in the attachment (G) glycoprotein of respiratory syncytial virus (RSV). J. Mol. Evol. 52, 182–192.

Woelk, C.H., Holmes, E.C., Brown, D.W.G., 2001. Immune and artificial selection in the haemagglutinin (H) glycoprotein of measles virus. J. Gen. Virol. 82, 2463–2474.

World Health Organization, 1992. WHO Expert Committee on Rabies, 8th Report. World Health Organization, Geneva, Switzerland.

Yang, Z., 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. 13, 555–556.

Yang, Z., Nielsen, R., Goldman, N., Pedersen, A.M., 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155, 431–449.