BIOINFORMATICS

Vol. 21 Suppl. 3 2005, pages iii39–iii44
doi:10.1093/bioinformatics/bti1203

*Sequence analysis*

# Within-host evolution of CD8$^+$-TL epitopes encoded by overlapping and non-overlapping reading frames of simian immunodeficiency virus

Austin L. Hughes[1,*], Helen Piontkivska[1], Kendall C. Krebs[2], David H. O'Connor[2] and David I. Watkins[2]

[1]Department of Biological Sciences, University of South Carolina, Columbia, SC, USA; [2]Wisconsin Primate Research Center and Department of Pathology and Laboratory Medicine, University of Wisconsin, Madison, WI, USA

Received on May 10, 2005; accepted on August 16, 2005

**ABSTRACT**

**Summary:** In order to understand the impact of overlapping reading frames on natural selection by host CD8+ T lymphocytes (CD8$^+$-TL), we analyzed the pattern of nucleotide substitution in simian immunodeficiency virus (SIV) genomes sampled from populations at time of death in 35 rhesus monkeys. Both the mean number of nonsynonymous nucleotide substitutions per nonsynonymous site ($d_N$) and the mean number of synonymous nucleotide substitutions per synonymous site ($d_S$) were elevated in overlap regions in comparison to non-overlap regions. Mean $d_N$ exceeded mean $d_S$ in CD8$^+$-TL epitopes restricted by the host's class I major histocompatibility complex molecules. This pattern, which is indicative of positive Darwinian selection favoring amino acid changes in these epitopes, was seen in both overlap and non-overlap regions; but mean $d_N$ was particularly elevated in restricted CD8$^+$-TL epitopes encoded in overlap regions. Amino acid changes from the inoculum were defined as parallel if the same amino acid change occurred at the same site independently in two or more monkeys, and a surprisingly high proportion (71.9%) of observed amino acid changes throughout the SIV genome occurred in parallel in different monkeys. The proportion of parallel changes in restricted epitopes encoded by overlapping reading frames was still higher (80%), supporting the hypothesis that the interaction of positive selection and overlapping reading frames enhances the probability of convergent or parallel amino acid change.

**Contact:** austin@biol.sc.edu

## INTRODUCTION

Viral genomes often include overlapping reading frames, and the evolutionary implications of this phenomenon have frequently been the subject of speculation (Keese and Gibbs, 1992; Miyata and Yasunaga, 1978; Pavesi *et al*., 1997). One important question is how natural selection can act simultaneously on two different protein products encoded in different reading frames by the same DNA sequence (Rogozin *et al*., 2002). Simian immunodeficiency virus (SIV) provides an example of conflicting selection pressures on overlapping reading frames (Hughes *et al*., 2001). Rhesus monkey (*Macaca mulatta*) hosts expressing the A*01 class I MHC molecule mount a strong CD8$^+$-T lymphocyte (CD8$^+$-TL) response against a peptide (A*01-Tat SL8) from the Tat protein of SIV that is encoded by the portion of the *tat* gene that overlaps the *vpr* gene. This CD8$^+$-TL response gives rise to positive Darwinian selection favoring amino acid changes in the Tat peptide (Allen *et al*., 2000). The evidence for this selection is that the mean number of nonsynonymous nucleotide substitutions per nonsynonymous site ($d_N$) exceeds the mean number of synonymous nucleotide substitutions per synonymous site ($d_S$) in the region of the *tat* gene encoding the epitope (Allen *et al*., 2000). At the same time, the corresponding region of the Vpr protein is subject to strong purifying selection against amino acid changes (Hughes *et al*., 2001). As a consequence, nonsynonymous substitutions occur in the *tat* reading frame so as to cause synonymous changes in the *vpr* reading frame to a greater extent than expected under random mutation (Hughes *et al*., 2001).

Convergent or parallel evolution involves the independent evolution of similar traits in different lineages, and natural selection is thought likely to enhance the probability of convergent or parallel evolution (Doolittle, 1994; Hughes, 1999). In human immunodeficiency virus type 1 (HIV-1), sequence comparisons have shown that the same amino acid replacements occur in parallel in different hosts in epitopes recognized by the host immune system (Piontkivska and Hughes, 2004; Strunnikova *et al*., 1995). Parallel amino acid changes were observed in the previously mentioned CD8$^+$-TL epitope in Tat of SIV in different monkeys independently infected with the same inoculum (Hughes *et al*., 2001). Because the same amino acid changes occurred repeatedly in different monkeys, it was hypothesized that the constraints imposed by overlapping reading frames may enhance the probability of parallel evolution.

Monkeys experimentally infected with SIV represent a particularly attractive system for studying viral evolution over the course of infection. Changes in the sequence of viral genes can be compared among monkeys with known MHC class I genotypes that have been infected with the same viral inoculum. Thus, unlike population studies of HIV-1, it is certain that sequence evolution has occurred independently in each monkey host. Here we use complete SIV genome sequence data obtained at time of death from a cohort of 35 rhesus monkeys experimentally infected with SIV (O'Connor *et al*., 2004) to examine the interaction of overlapping reading frames with natural selection driven by host CD8$^+$-TL recognition.

*To whom correspondence should be addressed at Department of Biological Sciences, University of South Carolina, Coker Life Sciences Building, 700 Sumter Street, Columbia SC 29208, USA.

Previous analysis of these data showed that $d_N$ significantly exceeds $d_S$ in known CD8$^+$-TL epitopes restricted by the host's MHC, implying that these epitopes are subject to positive selection favoring escape variants (O'Connor *et al*., 2004). On the other hand, other genomic regions generally do not show significantly higher $d_N$ than $d_S$, except for the variable loops of Env, which are targets for antibody recognition (Zwick *et al*., 2003). A number of the positively selected CD8$^+$-TL epitopes are encoded by overlapping reading frames in the *vpr*, *vpx*, *tat*, *env* and *nef* genes. Here we compare the pattern of evolution in CD8$^+$-TL epitopes encoded in overlapping and in non-overlapping reading frames in order to understand the effect of overlapping reading frames on the evolution of the viral CD8$^+$-TL evasion.

## METHODS

### Sequence data

Experimental methods are described by O'Connor *et al*. (2004). Complete viral genomes were sequenced from each of 35 SIV-infected rhesus monkeys at the time of death. Each sequence represented a mixture of viral genomes, and thus each sequence included a number of ambiguous nucleotides representing single-nucleotide polymorphisms (SNPs) within the viral population infecting the monkey. Each sequence was aligned individually with the inoculum sequence using the CLUSTALW program (Thompson *et al*., 1994).

### Synonymous and nonsynonymous substitution

The number of synonymous nucleotide substitutions per synonymous site ($d_S$) and the number of nonsynonymous nucleotide substitutions per nonsynonymous site ($d_N$) between each genomic sequence and the inoculum sequence were estimated by Nei and Gojobori's (1986) method. Any codon in which the alignment postulated a gap in either the individual genomic sequence or in the inoculum sequence, or for which sequence could not be determined for a given monkey was excluded from the computation for that comparison. Since information on the frequency of SNPs within the viral population inhabiting a given monkey was not available, we assumed equal frequency of all SNPs at a given site, and on this assumption we computed mean $d_S$ and $d_N$ between the inoculum and the population of sequences derived from each monkey. Note that this computation did not require that we know the phase of SNPs (i.e. which SNPs were linked in any given individual viral genome), information which was not available from the genomic sequence.

In preliminary analyses using a subset of the data, we applied a number of complex models for estimating $d_S$ and $d_N$ : Li's (1993) method, the modified Nei and Gojobori method (Zhang *et al*., 1998), and Yang and Nielsen's (2000) method. All models produced essentially the same result as the unmodified Nei and Gojobori (1986) method. Mean $d_S$ and $d_N$ were quite low (generally $<$1%), and with such a low level of divergence the differences among models of sequence evolution are generally undetectable (Nei and Kumar, 2000). Therefore, we report only results using the Nei and Gojobori (1986) method, which, because it makes fewer assumptions than the other models, is expected to have a lower variance (Nei and Kumar, 2000).

Known CD8$^+$-TL epitopes bound by the class I molecules Mamu-A*01, Mamu-A*02, and Mamu-B*17 (O'Connor *et al*., 2004) were mapped on the genomic sequences obtained from each monkey. For each monkey, $d_S$ and $d_N$ were computed for each of the nine genes in the SIV genome separately for the following regions: (1) restricted epitopes (A*01, A*02 or B*17 epitopes known to be presented by that monkey's class I MHC molecules) encoded by non-overlapping reading frames; (2) non-restricted epitopes (A*01, A*02 or B*17 epitopes not presented by that monkey's class I MHC molecules) encoded by non-overlapping reading frames; (3) non-epitope region (the remainder of the gene excluding restricted and non-restricted epitopes) encoded by non-overlapping reading frames; (4) restricted epitopes encoded by overlapping reading frames; (5) non-restricted epitopes encoded by overlapping reading frames; and (6) non-epitope regions encoded by overlapping reading frames. Because the virus from each monkey was compared independently with the inoculum, each such comparison was phylogenetically (Felsenstein, 1985) and thus statistically independent.

### Amino acid changes

Individual amino acid changes (residues different from the inoculum, including residues present as polymorphisms in the viral population within a given monkey) were classified with respect to a number of classificatory variables, and the resulting model-dimensional contingency tables were analyzed by log-linear analysis (Everitt, 1977). This statistical method makes it possible to test associations between classificatory variables while controlling for other variables. The following classifications were considered: (1) An amino acid change was classified as parallel if the same amino acid change was seen at the same site in two or more monkeys. (Note that we could not rule out the possibility that some parallel changes involved parallel increase in frequency in separate monkeys of rare variants already present in the inoculum.) (2) An amino acid change was classified as polymorphic if more than one amino acid residue occurred at a given site in a monkey; otherwise, the amino acid change was classified as fixed. (3) An amino acid change was classified as involving a polarity change if the amino acid change involved a change of polarity category, with F, L, I, M, V, P, A and W categorized as non-polar and the other 12 amino acid residues as polar. (4) Amino acid changes were classified with respect to whether they were encoded by non-overlapping or overlapping reading frames; (5) Amino acid changes were classified with respect to occurrence in restricted epitopes, in non-epitopes, or in non-epitope regions.

## RESULTS

### Synonymous and nonsynonymous substitution

Mean $d_S$ and mean $d_N$ were analyzed by factorial analysis of variance, which tested for differences between overlap and non-overlap regions, for differences among epitope categories (non-epitope regions, non-restricted epitopes and restricted epitopes), and for the interaction between these factors (Fig. 1). Mean $d_S$ was significantly greater in overlap regions (0.00746 $\pm$ 0.00066 SE) than in non-overlap regions (0.00192 $\pm$ 0.00028) ($F_{1,1506} = 63.50$; $P < 0.001$). But mean $d_S$ did not differ significantly among non-epitope regions (0.00499 $\pm$ 0.00086), non-restricted epitopes (0.00530 $\pm$ 0.00106), and restricted epitopes (0.00471$\pm$0.00086) ($F_{2,1506} = 1.17$; n.s.). There was a significant interaction between overlap and epitope categories (Fig. 1; $F_{2,1506} = 4.70$; $P = 0.009$). This interaction was explained by the different patterns of mean $d_S$ across epitope categories in overlap and non-overlap regions (Fig. 1). In non-overlap regions, mean $d_S$ was highest in non-epitope regions, whereas in over-lap regions, mean $d_S$ was highest in restricted epitopes (Fig. 1).

In the case of mean $d_N$, there was significantly higher mean value in overlap regions (0.00790 $\pm$ 0.00063) than in non-overlap regions (0.00413 $\pm$ 0.00028) ($F_{1,1506} = 64.99$; $P < 0.001$). Likewise, there was a highly significant difference among epitope categories, with mean $d_N$ in restricted epitopes (0.01152 $\pm$ 0.00114) nearly twice as high as that in non-restricted epitopes (0.00604 $\pm$ 0.00072) and over three times as high as that in non-epitope regions (0.00346 $\pm$ 0.00018) ($F_{2,1506} = 55.24$; $P < 0.001$). There was also a highly significant interaction between overlap and epitope category (Fig. 1; $F_{2,1506} = 28.75$; $P < 0.001$). This interaction was explained by the different patterns of mean $d_N$ across epitope categories in overlap and non-overlap regions, particularly the fact that mean $d_N$ in restricted epitopes in overlap regions (0.01875 $\pm$ 0.000231) was

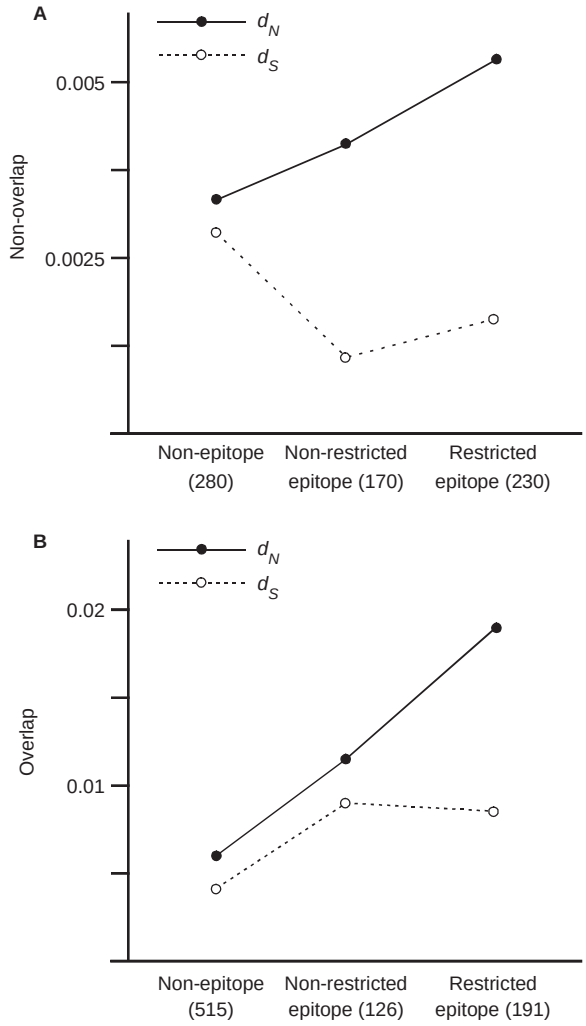


**Fig. 1.** Mean numbers of synonymous substitutions per synonymous site ($d_S$) and of nonsynonymous substitutions per nonsynonymous site ($d_N$) in non-overlap (**A**) and overlap (**B**) regions. Numbers in parentheses are numbers of comparisons (of regions in SIV genomes from individual monkeys versus the inoculum).

over three times the mean $d_N$ in restricted epitopes in non-overlap regions ($0.00552 \pm 0.000059$) (Fig. 1).

We compared mean $d_S$ and mean $d_N$ in the different overlap and epitope regions by means of paired $t$-tests. In non-overlap, non-epitope regions, mean $d_S$ and mean $d_N$ were not significantly different ($t = 0.78$; 279 df ). In overlap, non-epitope regions, mean $d_S$ was significantly greater than mean $d_N$ ($t = 4.16$; 514 df; $P < 0.001$). In non-restricted epitopes in non-overlap regions, mean $d_N$ was significantly greater than mean $d_S$ ($t = 4.08$; 169 df; $P < 0.001$); but in non-restricted epitopes in overlap regions, mean $d_S$ and mean $d_N$ were not significantly different ($t = 0.78$; 125 df ). In restricted epitopes, mean $d_N$ was significantly greater than mean $d_S$ both in non-overlap regions ($t = 5.70$; 229 df; $P < 0.001$) and in overlap regions ($t = 3.56$; 190 df; $P < 0.001$).

These results supported the hypothesis that positive selection has acted to favor amino acid replacements in restricted epitopes (O'Connor *et al*., 2004). In addition, they showed that overall levels of both synonymous and nonsynonymous substitutions were higher in overlap regions than in non-overlap regions.

## Differences among genes

The difference between overlap and non-overlap regions was particularly striking in the case of $d_N$ in restricted epitopes. In order to understand the causes of this pattern, we examined mean $d_S$ and mean $d_N$ in restricted epitopes separately in non-overlap and overlap regions of each gene (Table 1). In restricted epitopes in non-overlap regions, mean $d_N$ was significantly greater than mean $d_S$ in *gag*, *env* and *nef* (Table 1). By contrast, in overlap regions, mean $d_N$ was significantly greater than mean $d_S$ only in *tat* (Table 1). In fact the extraordinarily high mean $d_N$ in restricted epitopes in the overlap region of *tat* was far higher than mean $d_N$ in restricted epitopes in overlap or non-overlap regions of any other gene (Table 1). This high mean $d_N$ was due to the effect of a single epitope, A*01-Tat SL8 (1).

Although mean $d_S$ and mean $d_N$ in restricted epitopes in overlap regions did not differ significantly in genes other than *tat* (Table 1), this pattern may not always imply an absence of positive selection on these epitopes. Rather, unusually high $d_S$ in the overlap region may have obscured the effects of positive selection. This appears to have occurred in the case of *env*, where mean $d_N$ in restricted epitopes in overlap regions was actually significantly higher than that in restricted epitopes in non-overlap regions (Table 1). However, in *env* mean $d_S$ in restricted epitopes in overlap regions was also significantly higher than that in restricted epitopes in non-overlap regions (Table 1). A similar phenomenon may have occurred in *nef*, although in the latter case neither mean $d_S$ nor mean $d_N$ differed significantly between restricted epitopes in overlap regions and those in non-overlap regions (Table 1). In *nef* mean $d_S$ in restricted epitopes in overlap regions was higher than that in restricted epitopes in non-overlap regions, although the difference was not significant due to the high variance of the former (Table 1).

## Amino acid changes

We examined 1169 amino acid changes (differences from the inoculum) that occurred in the virus infecting the 35 monkeys. A high proportion (841/1169 or 71.9%) of amino acid changes that were parallel in the sense that the same amino acid change at the same site occurred independently in two or more monkeys. Overall, parallel amino acid changes were significantly more likely to occur in non-overlap regions than in overlap regions ($P = 0.005$; Fig. 2A). Of 831 changes in non-overlap regions, 612 (73.6%) were parallel, whereas of 338 changes in overlap regions, 229 (67.8%) were parallel (Fig. 2A). The proportion of fixed amino acid changes was significantly higher among parallel changes (205/328 or 62.5%) than among non-parallel changes (333/841 or 39.5%) ($P < 0.001$; Fig. 2B).

The proportion of amino acid changes causing a polarity change was significantly higher in restricted epitopes (58/106 or 54.7%) than in non-restricted epitopes (56/125 or 44.8%) or in non-epitope regions (263/938 or 28.0%) ($P < 0.001$; Fig. 3). There was a significant three-way association among the factors of parallelism, occurrence in an epitope, and occurrence in an overlap region ($P = 0.007$; Fig. 4). Parallel amino acid changes accounted for a higher proportion of changes in restricted epitopes in overlap regions (40/48 or 83.3%) than in restricted epitopes in non-overlap regions (38/58 or

**Table 1.** Mean $d_S$ and $d_N$ in restricted CD8$^+$-TL epitopes of different SIV genes

| Gene | Non-overlap regions $d_S \pm$ SE | $d_N \pm$ SE | $P^a$ | Overlap regions $d_S \pm$ SE | $d_N \pm$ SE | $P^a$ |
|---|---|---|---|---|---|---|
| *gag* | $0.0027 \pm 0.0011$ | $0.0063 \pm 0.0010$ | 0.011 | — | — | — |
| *pol* | $0.0030 \pm 0.0018$ | $0.0010 \pm 0.0005$ | N.S. | — | — | — |
| *vif* | $0.0000 \pm 0.0000$ | $0.0034 \pm 0.0018$ | N.S. | — | — | — |
| *vpx* | — | — | — | $0.0000 \pm 0.0000$ | $0.0000 \pm 0.0000$ | N.S. |
| *vpr* | — | — | — | $0.0107 \pm 0.0018$ | $0.0094 \pm 0.0011$ | N.S. |
| *tat* | | | | $0.0040 \pm 0.0040$ | $0.0921 \pm 0.0072$ | <0.001 |
| *env* | $0.0019 \pm 0.0007$ | $0.0045 \pm 0.0010$ | 0.011 | $0.0075 \pm 0.0026^b$ | $0.0102 \pm 0.0020^b$ | N.S. |
| *nef* | $0.0001 \pm 0.0001$ | $0.0092 \pm 0.0016$ | <0.001 | $0.0079 \pm 0.0047$ | $0.0091 \pm 0.0029$ | N.S. |

[a]Two-tailed $t$-tests of the hypothesis that mean $d_S$ = mean $d_N$ .
[b]Significantly different from the corresponding value for the non-overlap regions at 0.05 level ($t$-test, two-tailed).

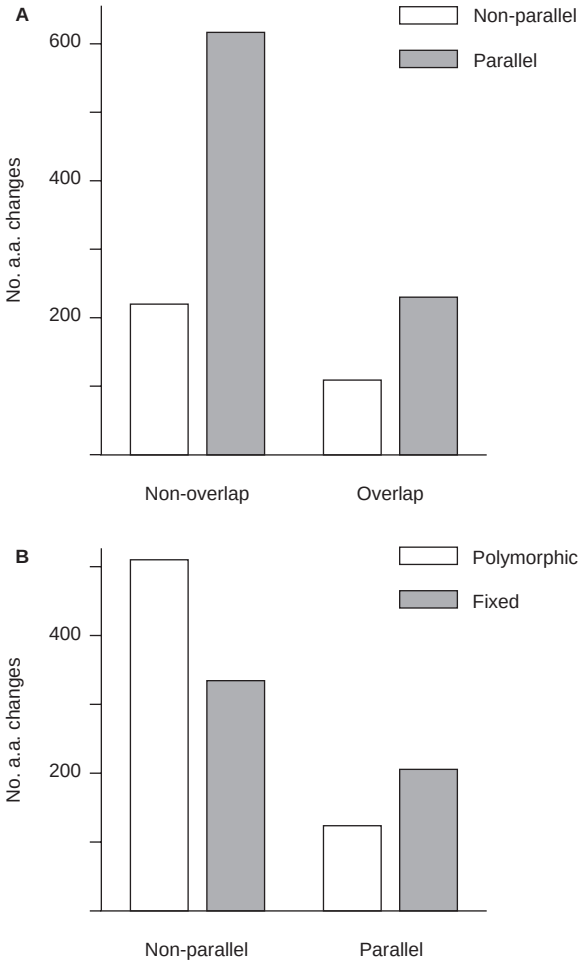


**Fig. 2.** (**A**) Numbers of parallel and non-parallel amino acid changes in non-overlap and overlap regions. The proportions differed significantly between non-overlap and overlap ($P = 0.005$). (**B**) Numbers of polymorphic and fixed (non-polymorphic) amino acid changes in non-overlap and overlap regions. The proportions differed significantly between non-overlap and overlap ($P < 0.001$).

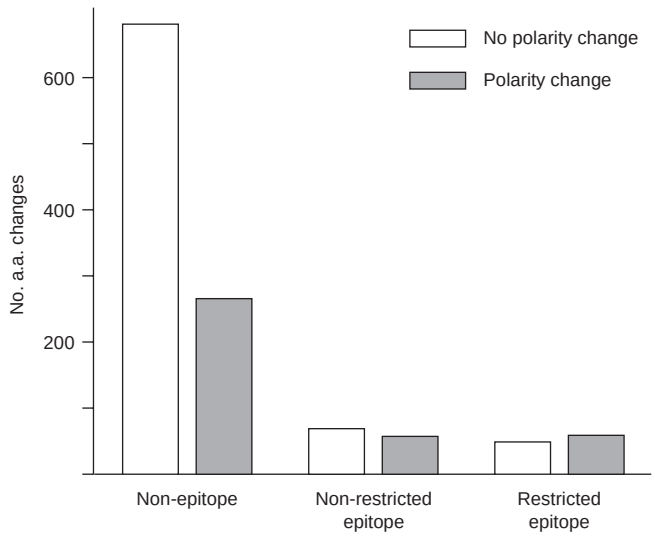


**Fig. 3.** Amino acid changes in different genomic regions categorized as to whether or not a change of amino acid residue polarity was involved. The proportions differed significantly among genomic regions ($P < 0.001$).

65.5%) (Fig. 4). By contrast, parallel amino acid changes accounted for a higher proportion of changes in non-overlap non-epitope regions (522/696 or 75.0%) than in overlap non-epitope regions (151/242 or 62.4%) (Fig. 4). Likewise, in the case of nonrestricted epitopes, parallel amino acid changes accounted for a higher proportion of changes in non-overlap regions (52/77 or 67.5%) than in restricted epitopes in non-overlap regions (38/58 or 65.5%) (Fig. 4).

## DISCUSSION

We analyzed the pattern of nucleotide substitution in SIV populations infecting 35 rhesus monkeys in order to assess the impact of overlapping reading frames on natural selection driven by host CD8$^+$ TL recognition. Both the mean number of nonsynonymous nucleotide substitutions per nonsynonymous site ($d_N$) and the mean number of synonymous nucleotide substitutions per synonymous site ($d_S$) were elevated in overlap regions in comparison to non-overlap regions. It has been previously shown that natural selection favors amino

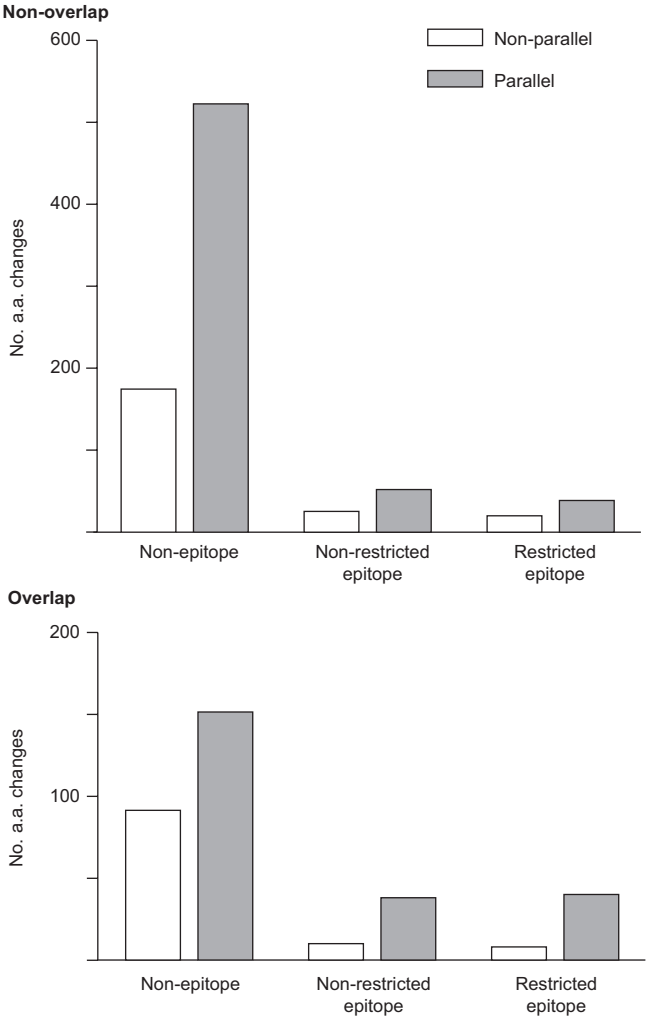


**Fig. 4.** Parallel and non-parallel amino acid changes in different genomic regions. There was a significant three-way association among parallelism, overlap and genomic region ($P = 0.007$), explained by the relatively high proportion of parallel changes in restricted epitopes in overlap regions.

acid changes in $CD8^+$-TL epitopes restricted by the monkey's class I MHC (O'Connor *et al*., 2004), as evidenced by a pattern of $d_N$ exceeding $d_S$ in genomic regions encoding restricted epitopes. Our results show that this positive Darwinian selection can act on $CD8^+$-TL epitopes in both overlap and non-overlap regions of the SIV genome, but that this selection causes a more marked elevation of nonsynonymous substitution in overlap regions than in non-overlap regions.

The A*01-Tat SL8 epitope (Allen *et al*., 2000) contributed strongly to the enhanced mean $d_N$ in epitopes encoded in overlap regions. By contrast, a single $CD8^+$TL epitope in an overlapping region of the *vpx* gene was possibly subject to strong functional constraint, showing no synonymous or nonsynonymous changes from the inoculum in any monkey restricting that epitope. Alternatively, it is possible that, while this epitope formally exists, it was not recognized by cytotoxic lymphocytes (CTL) during infection.

In comparing individual amino acid changes from the inoculum sequence in different regions, we found that a higher proportion of amino acid replacement involving a change of residue polarity occurred in restricted epitopes than in other genomic regions. Thus, changes in restricted epitopes were disproportionately likely to involve a radical amino acid change. This result provides additional support for the hypothesis that positive selection on restricted $CD8^+$-TL epitopes is driven by escape from host immune recognition, since more radical changes of amino acid residue are more likely to interfere with recognition based on protein–protein interactions.

Amino acid changes were defined as parallel if the same amino acid change occurred at the same site independently in two or more monkeys, and surprisingly 71.9% of observed amino acid changes were parallel. Some of these, of course, may have been present as rare variants in the inoculum. Significantly more parallel changes occurred in non-overlap than in overlap regions. Thus, an overlapping reading frame alone did not in itself enhance the probability of parallel change; in fact, it decreased the probability of parallel change. Surprisingly, many parallel amino acid changes occurred in non-epitope regions, where there was no evidence of positive selection. Evolutionary biologists have predicted that convergent or parallel evolution is much more likely to occur as a result of positive selection than in the absence of positive selection (Doolittle, 1994; Hughes, 1999). Parallel changes in non-epitope regions may represent changes that occur in response to some other sources of selection besides that exerted by $CD8^+$-TL recognition of peptides bound by the class I MHC molecules A*01, A*02 and B*17.

Such sources of selection might include the following: selection on $CD8^+$-TL epitopes bound by class I molecules other than A*01, A*02 or B*17; selection on $CD4^+$-TL epitopes; selection on neutralizing antibody epitopes; and compensatory changes accompanying $CD8^+$-TL escape variants (Friedrich *et al*., 2004). Some such unidentified source of selection presumably accounts for the fact that in our analyses mean $d_N$ was significantly greater than mean $d_S$ in non-restricted epitopes in non-overlap regions. In addition, the fact that parallel changes were more likely to be fixed than polymorphic is consistent with the hypothesis that many of them were selectively favored, since polymorphic variants may be expected to include some slightly deleterious mutations in the process of being eliminated by purifying selection. On the other hand, many parallel changes may not be selectively favored, but may simply represent mutations that are neutral for the virus during infection and are fixed by chance.

In spite of the overall higher rate of parallel amino acid changes in non-overlap regions than in overlap regions, 80% of amino acid changes in restricted epitopes in overlap regions were parallel. This was the highest rate of parallel change observed, and it occurred in a region known to be subject to positive selection. Thus, in these epitopes, the theoretical expectation of an association between positive selection and a high rate of parallel change was supported. In addition, these results support the hypothesis that the rate of parallel change will be enhanced when there is positive selection on a region encoded by an overlapping reading frame, as a result of constraints on possible amino acid changes imposed by the other reading frame (Hughes *et al*., 2001). Furthermore, amino acid changes in restricted epitopes tended to be radical, in the sense of causing amino acid residue polarity change. These results suggest that parallel radical amino acid changes are a signature of $CD8^+$-TL selection and that identifying such changes may help guide the search for as yet unknown epitopes.

It has been proposed that a $CD8^+$-TL vaccine against HIV-1 should target conserved epitopes in order to prevent escape from vaccine-primed immunity (Nabel, 2002). Alternatively, epitopes showing the strongest $CD8^+$-TL selection and thus of escape mutants might be chosen for a vaccine, since immune responses targeting these epitopes might be most effective in eliminating the virus (O'Connor *et al*., 2001). In either case an understanding of the pattern of evolutionary change of $CD8^+$-TL epitopes during infection will be an important prerequisite for effective vaccine design. The present results show, in an SIV model in rhesus macaques, that patterns of $CD8^+$-TL selection are influenced by a variety of factors, including overlapping reading frames, as well as differences among genes. Understanding the interaction of these factors will make it increasingly possible to predict the pattern of evolutionary change in amino acid sequences over the course of infection with SIV, HIV-1 and related viruses.

## ACKNOWLEDGEMENTS

This research was supported by NIH grants R01 GM43940 (to A.L.H.) and P51 RR000167 (to the Wisconsin National Primate Research Center, University of Wisconsin-Madison).

*Conflict of Interest:* none declared.

## REFERENCES

Allen,T.M. *et al*. (2000) Tat-specific CTL select for SIV escape variants during resolution of primary viremia. *Nature*, **407**, 386–390.

Doolittle,R.F. (1994) Convergent evolution: the need to be explicit. *Trends Biochem. Sci.*, **19**, 15–18.

Everitt,B.S. (1977) *The Analysis of Contingency Tables*. Chapman and Hall, London.

Felsenstein,J. (1985) Phylogenies and the comparative method. *Am. Nat*., **125**, 1–15.

Friedrich,T.C. *et al*. (2004) Extraepitopic compensatory substitutions partially restore fitness to simian immunodeficiency virus variants that escape from an immunodominant cytotoxic-T-lymphocyte response. *J. Virol.*, **78**, 2581–2585.

Hughes,A.L. (1999) *Adaptive Evolution of Genes and Genomes*. Oxford University Press, New York.

Hughes,A.L. *et al*. (2001) Simultaneous positive and purifying selection on overlapping reading frames of the *tat* and *vpr* genes of simian immunodeficiency virus. *J. Virol.*, **75**, 7966–7972.

Keese,P.R. and Gibbs,A. (1992) Origins of genes: "big bang" or continuous creation? *Proc. Natl Acad. Sci. USA*, **89**, 9480–9493.

Li,W.-H. (1993) Unbiased estimates of the rates of synonymous and nonsynonymous substitution. *J. Mol. Evol*., **36**, 96–99.

Miyata,T. and Yasunaga,T. (1978) Evolution of overlapping genes. *Nature*, **272**, 532–535.

Nabel,G.J. (2002) HIV vaccine strategies. *Vaccine*, **20**, 1945–1947.

Nei,M. and Gojobori,T. (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol*., **3**, 418–426.

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

O'Connor,D. *et al*. (2001) Vaccination with CTL epitopes that escape: an alternative approach to HIV vaccine development? *Immunol. Lett.*, **79**, 77–84.

O'Connor,D.H. *et al*. (2004) A dominant role for $CD8^+$-T-lymphocyte selection in simian immunodeficiency virus sequence variation. *J. Virol*., **78**, 14012–14022.

Pavesi,A. *et al*. (1997) On the informational content of overlapping genes in prokaryotic and eukaryotic viruses. *J. Mol. Evol.*, **44**, 625–631.

Piontkivska,H. and Hughes,A.L. (2004) Between-host evolution of CTL epitopes in human immunodeficiency virus type 1 (HIV-1): an approach based on phylogenetically independent comparisons. *J. Virol*., **78**, 11758–11765.

Rogozin,I.B. *et al*. (2002) Purifying and directional selection in overlapping prokaryotic genes. *Trends Genet.*, **18**, 228–232.

Strunnikova,N. *et al*. (1995) Convergent evolution within the V3 loop domain of human immunodeficiency virus type 1 in association with disease progression. *J. Virol.*, **69**, 7548–7558.

Thompson,J.D. *et al*. (1994) CLUSTALW: 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.

Yang,Z. and Nielsen,R. (2000) Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol. Biol. Evol.*,**17**, 32–43.

Zhang,J. *et al*. (1998) Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc. Natl Acad. Sci USA*, **98**, 3708–3713.

Zwick,M.B. *et al*. (2003) A novel human antibody against human immunodeficiency virus type 1 gp120 is V1, V2, and V3 loop dependent and helps delimit the epitope of the broadly neutralizing antibody immunoglobulin G1 b12. *J. Virol*., **77**, 6965–6978.