Convergent and Divergent Sequence Evolution in the Surface Envelope Glycoprotein of Human Immunodeficiency Virus Type 1 within a Single Infected Patient

Edward C. Holmes; Lin Qi Zhang; Peter Simmonds; Christopher A. Ludlam; Andrew J. Leigh Brown

Proceedings of the National Academy of Sciences of the United States of America, Vol. 89, No. 11 (Jun. 1, 1992), 4835-4839.

Stable URL:
http://links.jstor.org/sici?sici=0027-8424%2819920601%2989%3A11%3C4835%3ACADSEI%3E2.0.CO%3B2-V

Proceedings of the National Academy of Sciences of the United States of America is currently published by National Academy of Sciences.

Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 4835–4839, June 1992
Evolution

Convergent and divergent sequence evolution in the surface envelope glycoprotein of human immunodeficiency virus type 1 within a single infected patient

(V3 loop evolution/phylogeny/escape mutants/neutralization)

EDWARD C. HOLMES*†, LIN QI ZHANG*, PETER SIMMONDS‡, CHRISTOPHER A. LUDLAM§, AND ANDREW J. LEIGH BROWN*

*Division of Biological Sciences, Institute of Cell, Animal and Population Biology, Crew Building, West Mains Road, Edinburgh EH9 3JN, Scotland; ‡Department of Medical Microbiology, University of Edinburgh, Teviot Place, Edinburgh EH8 9AG, Scotland; and §Department of Hematology, Royal Infirmary of Edinburgh, Lauriston Place, Edinburgh EH3 9YW, Scotland

Communicated by J. Maynard Smith, February 3, 1992

ABSTRACT **In an investigation of the evolution of the third hypervariable loop of gp120 (V3), the principal neutralization determinant of human immunodeficiency virus type 1, we have analyzed 89 V3 sequences of plasma viral RNA purified from peripheral blood samples donated over 7 years by an infected hemophiliac. Considerable sequence diversity in the V3 region was found at all time points after seroconversion. Phylogenetic analysis revealed that an important diversification had occurred by 3 years postinfection and that, subsequently, most sequences could be allocated to either one of two major lineages that persisted throughout the remainder of the infection. Rapid changes in frequency of the most common sequences and the observation that the same hexapeptide motif (GPGSAV) at the crown of the V3 loop has evolved convergently provide strong evidence that selective processes determine the evolutionary fate of sequence variants in this region.**

Analyses of nucleic acid sequences from human immunodeficiency virus type 1 (HIV-1) have revealed a large amount of variation both between and within patients (1–10). This variation is not evenly distributed across the HIV-1 genome; rates of nucleotide substitution are particularly high in hypervariable regions (V1–V5) of the *env* gene (11) and are lower in *pol* and the p24 coding region of *gag* (1, 11, 12). This suggests that the levels of sequence variation are not simply reflections of the intrinsic mutation rate.

There are good grounds for expecting that natural selection will be important in determining HIV-1 variability. This may be manifest as selective constraint against variation in amino acid sequence, as in the *pol* gene and the p24 region of *gag* (11, 12). On the other hand, studies of two other lentiviruses, equine infectious anemia virus and visna virus, suggest that much of the variability seen in *env* may be adaptive. In both of these viruses, antigenically distinct isolates were found to arise progressively during infection, always appearing before their associated neutralizing antibodies (refs. 13 and 14 and references therein). Differences between isolates could be assigned to mutations clustered in the *env* genes, the majority of which resulted in amino acid replacement. Thus, it was proposed that viral mutants arise that can "escape" from recognition by neutralizing antibodies. This change in the genetic composition of the viral population, referred to as "antigenic drift" although it is a selective process, is generally considered to be one of the principal mechanisms by which lentiviruses evade host immune systems (reviewed in ref. 15). However, high levels of variability are also seen in the *nef* and *tat* genes (2, 6).

The *env* gene encodes the principal neutralization determinant of HIV-1 (5, 16), which has been mapped to an ≈35-amino acid residue disulfide-bonded loop structure in the third hypervariable region of surface envelope glycoprotein gp120 (V3) (16–18). There have been a number of attempts to assess *in vitro* the fitness of sequential V3 loop isolates by serological techniques (19) and to examine patterns of sequence change *in vivo* (9, 10). Simmonds *et al.* (20) analyzed sequence change in the adjacent V4 and V5 hypervariable regions of *env* over a 5-year period after seroconversion in an asymptomatic hemophiliac (9). It was concluded that analysis of HIV sequence change in *env* within a patient based only on the lymphocyte-associated proviral population could be misleading because of significant differences in the frequency and persistence of sequence variants compared with the plasma (considered to be recently replicated) virus population. Therefore, analysis of variation in the V3 region presented here is based on sequences detected in plasma virus particles.

To assess the role and nature of selection acting on V3 sequences, we have carried out a detailed analysis of the process of viral evolution in this region during the course of an infection.

MATERIALS AND METHODS

Nucleotide Sequences. Eighty-nine viral RNA sequences (plasma derived) of ≈240 base pairs were obtained directly from nested PCR amplified single molecules from an Edinburgh hemophiliac patient (p82) infected from a locally prepared batch of factor VIII in 1984 (1). Sequences were obtained in varying numbers (n) from seroconversion [year 0 (1984, $n = 12$); year 3 (1987, $n = 15$); year 4 (1988, $n = 11$); year 5 (1989, $n = 23$); year 6 (1990, $n = 15$); year 7 (1991, $n = 11$) postinfection]. These sequences have been assigned GenBank accession numbers M84240–M84317. Methods used for extraction and PCR amplification of single cDNA molecules are described by Zhang *et al.* (21) and those for nested PCR amplification and direct sequencing are described by Simmonds *et al.* (20). Clinical data for this patient are given in ref. 9. The patient has been asymptomatic throughout the period of study and has never received antiviral therapy.

Phylogenetic Analysis. V3 sequences from six isolates from the USA [HIV_{RF}, HIV_{WMJ2}, HIV_{MN}, HIV_{SF2}, HIV_{PV22}, and $HIV_{HTLVIIIB}$ (clone HXB2)] were used as outgroups (all were taken from the Los Alamos Human Retroviruses and AIDS

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: HIV-1, human immunodeficiency virus type 1; V3, third hypervariable region of surface envelope glycoprotein gp120.

†To whom reprint requests should be addressed.

data base). Sequences were aligned by hand, as the only length variation was the dipeptide insertion of QR between residues 309 and 310 in isolates HIV_{PV22} and $HIV_{HTLVIIIB}$ (HXB2).

All phylogenetic analysis programs used were taken from the PHYLIP package (version 3.4) provided by J. Felsenstein (Department of Genetics, University of Washington). The principal program used was DNAML, which implements a maximum likelihood method. Global branch swapping was used, as this increases the proportion of trees searched. The default settings were used for the other options. We also used the neighbor-joining distance matrix method of Saitou and Nei (22) as implemented in the PHYLIP program NEIGHBOR. Distances were estimated by using the same evolutionary model as underlies the DNAML program by the program DNADIST.

RESULTS

Phylogenetic Analysis of the V3 Region. Phylogenetic relationships can be inferred with greatest confidence when there is no homoplasy (convergence) in the data. As selection can cause convergence even at the sequence level (23), we divided the region under study into three sections—the 35-amino acid V3 loop itself and two flanking regions (a 19-residue 5′ region and a 23-residue 3′ region)—and carried out all phylogenetic analyses on nucleotide sequences. Trees were inferred from different combinations of these regions taken separately, as well as from the entire sequence, in order not to bias the analysis toward any spurious phylogenetic patterns found only in the loop.

An unusual feature of the data was that, whereas in most phylogenetic analyses sequences are contemporaneous, sequences in this data set were sampled from different time points during the evolution of the viral population within an infected patient. Therefore, trees were also inferred for each year separately, in addition to analyses of the whole data set.

Relationships of V3 Sequences. The most striking and consistent result from the phylogenetic analyses of the nucleotide sequence data was that a major division into two distinct lineages had occurred by 1987. This is depicted in the global maximum likelihood tree for all distinct V3 sequences of plasma virions from p82 (Fig. 1). For reasons of clarity, only branches connecting the 46 p82 sequences that were separated by branch lengths found to be significantly different from 0 under the DNAML model are shown. Individual sequences are represented by a dot. A single USA outgroup sequence, $HIV_{HTLVIIIB}$ clone HXB2, is also included. The subdivision into two major descendent lineages, D and E (see below), is indicated. Although not drawn so, this tree may be rooted by HXB2 (as labeled), in which case evolution may be thought of as running from left to right. Twelve identical V3 sequences were obtained at seroconversion. All trees showed this sequence to be the ancestor of all others in that it is closest to the root (as labeled in Fig. 1). Finally, all branch lengths are drawn to scale, which permits an assessment of the relative amounts of evolutionary change along different lineages.

Other, smaller, bifurcations were also consistently observed and these, like the major subdivision, were found to accord with the pattern of amino acid replacements in the V3 loop (see below). Conversely, a few sequences were more difficult to place and frequently changed position in different analyses. These were mainly sequences from year 3, a sample in which most sequences had diverged to a similar extent (see below). By year 4, however, the major lineages were clearly distinguished.

Evolution of the V3 Loop. Having established the phylogenetic relationships between sequences from the V3 region from this patient, we wished to determine, because of the functional importance of the V3 loop itself, whether the lineages we have identified correspond to changes in the antigenic structure of the loop. In the entire data set, there were a total of 24 different V3 loop amino acid sequences. These sequences are listed in Table 1, together with the frequencies at which they are found in each year. Each can be seen as having one or more amino acid differences from that found at seroconversion (designated sequence A). If we superimpose the V3 loop amino acid sequences onto the trees deduced from the nucleotide sequences (Fig. 1, for example), we obtain the evolutionary "framework" depicted in Fig. 2. Distinct evolutionary lineages of V3 loop sequences have been identified and assigned a letter (A–F) and each different sequence within these lineages has been assigned a number (Fig. 2 and Table 1). The proposed relationships and the directions of changes between the sequences of a lineage are indicated by arrows (Fig. 2). Sequences which have persisted through time points are indicated by dashed arrows. Relationships between some lineages were difficult to assign, as different analyses produced different phylogenies, so affiliations between lineages have only been indicated when they were unequivocal. For example, the relationships between lineages B, E, and F are uncertain and the symbol ? that connects them is used to signify this uncertainty. Finally, unlike most molecular phylogenies, real ancestors may be

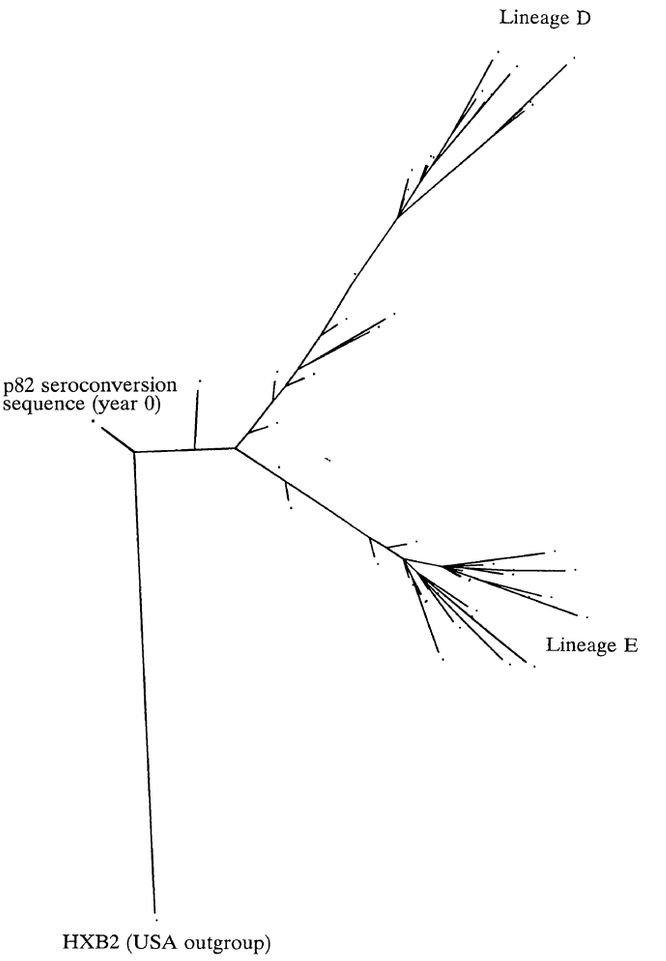

FIG. 1. Maximum likelihood tree depicting evolutionary relationships among V3 sequences obtained from plasma virus from p82. This tree may be rooted by the HXB2 outgroup sequence. Given this rooting, the first p82 sequence to appear is that found at seroconversion, as indicated. The major evolutionary subdivision is clearly visible and the two major descendent lineages, D and E, are also labeled. Dots at the tips of the lineages represent individual sequences; all branch lengths are drawn to scale.

Table 1. The 24 V3 loop amino acid sequences and their sample frequencies in the plasma

Lineage	Sequence	Frequency in year 0	3	4	5	6	7
	296 CTRPNNNTRKSIHIGPGRAFYTTGEIIGDIRQAHC 330						
A		1.000					
B	.....D.....		0.067				
C1	.....P.....D.....		0.067				
C2	.....P.....Q.....		0.267				
C3	.....G.....Q.....		0.267				
C4	.....P.....D.T.....		0.067				
C5	.....D.T.....		0.067				
D1	.....V....Q.....			0.091			
D2	.....R.....V...EQ...N.....			0.455			
D3	.....G.....V...EQ...N.....			0.091			
D4	.....R.Y.....V...EQ...N.....				0.087		
D5	.....R.Y....S.V...EQ...N.....				0.043		
D6	.....R.Y.....V...DQ.....					0.067	
D7	.....R.Y.....V...DQ...N.....					0.200	0.077
D8	.....Y...R.G.....SV..AEQ...N.....					0.200	0.077
E1	.....G.....S...A..D.....		0.067		0.043	0.333	0.769
E2	.....G.....S...A..G.....			0.182	0.696		
E3	.....G.....S...A..R.....			0.091			
E4	.....G.....S.V.A..G.....				0.043		
E5	.....G.....S...A..G...N.....				0.087		
E6	.....G.....S.V.A..D.....					0.067	
E7	.....G.....A..D.....					0.067	
E8	.....G.....V....D.....					0.067	
F	.....G.....		0.133	0.091			0.077
	Total	12	15	11	23	15	11

Amino acid sequences are listed according to their evolutionary lineage (A–F). Only residues that differ from those in sequence A (detected at seroconversion) are shown, with a dot denoting identical residues. Amino acid positions are numbered according to Wolfs *et al.* (10). The frequencies of each sequence in each year from which a sample was available are presented, with the total number of sequences obtained given at the bottom. Blank space indicates that the sequence was not detected in that particular year.

present in the data and the framework expresses the postulated ancestor–descendent relationships.

Many of the evolutionary patterns observed at the nucleotide level, and depicted in Fig. 1, correspond to amino acid changes in the loop itself. From Figs. 1 and 2 and Table 1, it is clear that evolutionary lineages D and E dominate the data set. Lineage D can be distinguished by 2 amino acid replacements from its apparent ancestor—C2. These are valine for phenylalanine at position 315 and glutamine for glutamic acid at position 320. A number of other replacements are also found in most of the lineage D variants, such as aspartic acid to asparagine (N) at position 324. The relationship of the lineage E sequences to each other is characterized by a serine to glycine substitution at position 306. Most members of this lineage also have a serine at position 313 and an alanine at position 317.

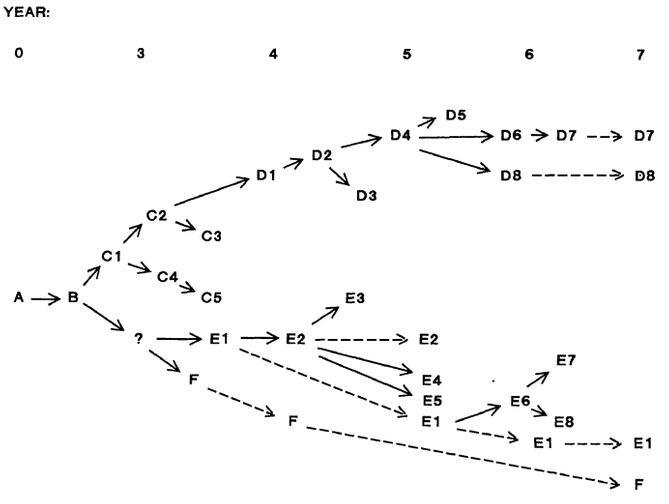

FIG. 2. Evolutionary framework relating the 24 different amino acid sequences found in the V3 loop. Distinct evolutionary lineages are designated by letters A–F and sequences within lineages are designated by numbers. Proposed relationships are indicated by arrows. Lineages that persist through years are indicated by dashed arrows. Time scale is given along the top.

Comparison of lineages D and E reveals some interesting features in the evolution of the 6 amino acids that correspond to the crown of the V3 loop (10). In the seroconversion sample, these were GPGRAF (Table 1). The same motif has also been found in 146/256 (60%) isolates of HIV-1 from the USA examined by La Rosa *et al.* (5). Lineages D and E have undergone a number of changes in this sequence; GPGRAV is most frequently found in lineage D [although only 12 times (5%) in the set of isolates from the USA (5)], while GPGSAF [not reported by La Rosa *et al.* (5)] characterizes the early evolution of lineage E. Remarkably, both lineages acquired the motif GPGSAV independently.

As already noted, the relationships of lineages found in year 3 (B, C, E, and F) and of the sequences within the C lineage are harder to interpret. All analyses suggested that sequence B, characterized by a single amino acid change (glutamic acid to aspartic acid at position 320) from the seroconversion sequence A and found only once in the data set, is the ancestor of all later variants. The similarity between sequences A and B, in the face of the diversity found in year 3, suggests that B probably arose early in infection.

All members of the C lineage can be derived from sequence B. C4 and C5 are grouped through their possession of a threonine at position 322, while the relationship between C2 and C3 is suggested by their possession of a glutamine at position 320. The remaining intersequence relationships are even harder to define but, as depicted (Fig. 2), it seems most likely that C1 is the ancestor of all others.

Changes in Sequence Frequency. The frequencies at which the sequences are found in the plasma are listed in Table 1. This table highlights the dramatic changes in frequency of the most abundant sequences. The most notable examples are provided by sequence A, the only sequence found at seroconversion but not seen subsequently; E2, which constituted almost 70% of the sample in year 5 but was not found in any later samples; and D2, which made up 45% of the sample in year 4 but was not found afterward.

It is unlikely that these changes in frequency are simply sampling artifacts for two reasons. First, we have shown in a study of the V4 and V5 sequences from the same patient (9) that the sample frequencies estimated from nucleotide sequence data (where n was $\approx$10) accorded well with the relative frequencies of length variants observed by scanning densitometry in much larger samples of the same material (where n, the input copy number, was $\approx$100). Second, random sampling would affect most strongly those variants found at low frequency. The striking feature of the evolution described here (Table 1) is that the most dramatic changes involve those sequences that reach high frequency.

Convergent Evolution. One of the most unexpected observations stemming from the phylogenetic analysis of V3 sequences was that the same amino acid has been fixed independently at the same site in more than one sequence. The most spectacular examples involve the hexapeptide at the crown of the V3 loop. The evolution of variants of this motif, as inferred from the evolutionary framework (Fig. 2), is shown in Fig. 3. We infer the motif found at seroconversion (GPGRAF) to be the ancestor of all later motifs, including GPGSAV (indicated by an asterisk in Fig. 3), which appears to have evolved independently three times within p82—in sequences D5, E4, and E6—and twice in 1989 alone. Other instances of convergent evolution in this motif are GPGRAV, found in most of the D lineage sequences and in E8, and the original GPGRAF motif, which reappears in E7. Convergent evolution in this hexapeptide is evidently frequent. In a second hemophiliac patient from Edinburgh also infected in 1984 with a virus of the same sequence (9), the independent evolution of the GPGSAV motif has also been observed (data not shown), making four times in all. Further convergence occurs outside this motif; the glycine at position 308 appears to have arisen twice (in D3 and the E lineage) as does the glycine at position 320 (in three members of the E lineage—E2, E4, and E5—as well as in genotype F) and the asparagine at position 324 (in the D lineage and E5). Overall, this level of convergence seems to be most compatible with the interpretation that sequence change in this region represents adaptive evolution (23).

DISCUSSION

We have used a number of different methods of phylogenetic inference. This is particularly necessary when natural selection may be important. From simulation studies, it has become clear that maximum likelihood inference (as implemented in DNAML) performs extremely well, in terms of the proportion of times it retrieves the correct tree, under a neutral (Poisson based) model of molecular evolution (24). It is unclear, however, how any method fares under evolutionary processes that deviate from neutrality, and thus it is probably best not to rely solely on one.

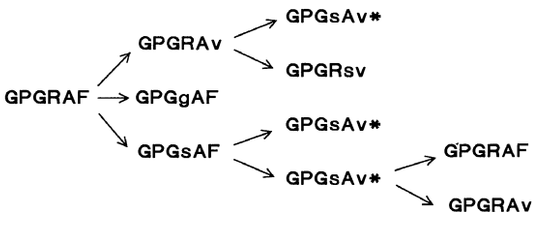

FIG. 3. Evolutionary relationships between the different hexapeptide motifs found at the crown of the V3 loop. The motif found at seroconversion (year 0) is GPGRAF and, as this is the inferred ancestor of all later motifs, is written in capital letters. From this, five different motifs evolve during the course of the infection, with the amino acids that differ from that found at seroconversion written in lowercase letters. Convergent evolution of the motif GPGSAV is denoted by an asterisk.

Evolution of HIV Sequences in p82. During the infection of p82 there has been an evolutionary diversification of V3 sequences. This diversification can be assigned to evolution along a number of major lineages, themselves characterized by successive substitutions at residues of (presumed) antigenic importance. Selection for viral mutants that can escape from the neutralizing capability of host antibodies is likely to be the principal mechanism that drives this evolutionary change (25). The rapid change in the composition of the V3 sequence population—the rise and fall of variants—is a consequence of this continual process of neutralization–escape. This is also reflected as alterations in the relative prominence of different lineages, specifically D and E, in samples from successive years. Thus, antigenic drift, instead of being the sequential replacement of one antigenically distinct variant by another, may involve a complex interaction between the different, and competing, evolutionary lineages present in the plasma.

The rapid changes in the plasma population also suggest that antigenic evolution has an important frequency-dependent element. Standard host–parasite models (26) suggest that the higher the frequency a viral variant reaches in the plasma, the higher the probability of its recognition and neutralization. Table 1 shows that in each year it is generally the most frequent sequence that appears to be neutralized (i.e., shows the greatest reduction in frequency). Consequently, variants found at low frequency will have a greater selective advantage and will increase in frequency until they too are countered with an effective immune response.

The evolution of the viral population is made more complicated by the fact that selection pressures may not be constant during the course of the infection. Nowak *et al.* (27, 28) suggest that the move from the asymptomatic to the symptomatic stage of infection is triggered by the loss of a specific immune response that contains a highly diverse viral population during the early stages and that, in AIDS, new antigenic variants may no longer be favored. It is interesting to note that no new V3 amino acid sequences were found in year 7—at which time the $CD4^+$ cell count of p82 had fallen below 200 (9), although the patient remained asymptomatic.

Constraints on V3 Loop Sequences. We interpret the observation of extensive convergent evolution to be due to the interplay between selection for variability and for conservation. Specifically, although there is positive selection for replacement of amino acids that remove B-cell (or T-cell) epitopes once these are recognized, there is also a selective constraint as to which amino acids are functionally viable within this region. Interestingly, many of the differences between the V3 loop sequences from p82 only involve a limited number of amino acid replacements. This is most clearly seen in the case of sequence E1, which replaces E2 as the most frequent in the population, although distinguished by only a single amino acid change (glycine to aspartic acid at position 320; Table 1). It is also possible that there are conformational changes elsewhere in the protein, which may compensate for changes in the loop itself.

Finally, the high proportion of nucleotide substitutions that lead to amino acid replacements (K_a) compared to silent substitutions (K_s) (9) is also consistent with the action of selection (data not shown).

Rate of Sequence Evolution. Several studies, including our own, have attempted to estimate a mean rate of sequence

evolution for HIV-1 (1, 4, 10, 11, 29). If, as we suggest, natural selection is the major mechanism of sequence change in this region of the HIV-1 genome, then any attempt to estimate a mean rate of evolution within a patient is likely to be misleading. In p82, the mean nucleotide distance between samples from years 5 and 6 was actually greater than that between years 4 and 6 (data not shown), and similar phenomena have been observed before (4, 10). This can now be seen to arise from the alternation between the two predominant evolutionary lineages in p82. Even within a lineage, we observe substantial variation in the rate of change. The concept of a mean rate of sequence evolution is almost meaningless in the context of a region such as V3, which is so strongly influenced by selection.

We thank the staff of the Hemophilia Centre (Royal Infirmary of Edinburgh) for taking samples, Sarah Ashelford for assisting with the sequencing, and two referees for helpful comments. This work was supported by the AIDS Directed Programme and the Committee on Epidemiological Studies of AIDS of the Medical Research Council.

1. Balfe, P., Simmonds, P., Ludlam, C. A., Bishop, J. O. & Leigh Brown, A. J. (1990) *J. Virol.* **64,** 6221–6233.
2. Delassus, S., Cheynier, R. & Wain-Hobson, S. (1991) *J. Virol.* **65,** 225–231.
3. Epstein, L. G., Kuiken, C., Blumberg, B. M., Hartman, S., Sharer, L. R., Clement, M. & Goudsmit, J. (1991) *Virology* **180,** 583–590.
4. Hahn, B., Shaw, G. M., Taylor, M. E., Redfield, R. R., Markham, P. D., Salahuddin, S. Z., Wong-Staal, F., Gallo, R. C., Parks, E. S. & Parks, W. P. (1986) *Science* **232,** 1548–1553.
5. La Rosa, G. J., Davide, J. P., Weinhold, K., Waterbury, J. A., Profy, A. T., Lewis, J. A., Langlois, A. J., Dreesman, G. R., Boswell, N., Shadduck, P., Holley, L. H., Karplus, M., Bolognesi, D. P., Matthews, T. J., Emini, E. A. & Putney, S. D. (1990) *Science* **249,** 932–935.
6. Meyerhans, A., Cheynier, R., Albert, J., Seth, M., Kwok, S., Sninsky, J., Morfeldt-Manson, L., Asjo, B. & Wain-Hobson, S. (1989) *Cell* **58,** 901–910.
7. Saag, M. S., Hahn, B. H., Gibbons, Y., Li, Y., Parks, E. S., Parks, W. P. & Shaw, G. M. (1988) *Nature (London)* **344,** 440–444.
8. Simmonds, P., Balfe, P., Ludlam, C. A., Bishop, J. O. & Leigh Brown, A. J. (1990) *J. Virol.* **64,** 5840–5850.
9. Simmonds, P., Zhang, L. Q., McOmish, F., Balfe, P., Ludlam, C. A. & Leigh Brown, A. J. (1991) *J. Virol.* **65,** 6266–6276.
10. Wolfs, T. F. W., De Jong, J.-J., Van Den Berg, H., Tunagel, J. M. G. H., Krone, W. J. A. & Goudsmit, J. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 9938–9942.
11. Li, W.-H., Tanimura, M. & Sharp, P. (1988) *Mol. Biol. Evol.* **5,** 313–330.
12. Leigh Brown, A. J. & Moynaghan, P. (1988) *AIDS Res. Hum. Retroviruses* **4,** 399–407.
13. Scott, J. V., Stowring, L., Hasse, A. T., Narayan, O. & Vigne, R. (1979) *Cell* **18,** 321–327.
14. Salinovich, O., Payne, S., Montelaro, R., Hussain, K., Issel, C. & Schnorr, K. (1986) *J. Virol.* **57,** 71–80.
15. Clements, J. E., Gdovin, S. L., Montelaro, R. C. & Narayan, O. (1988) *Annu. Rev. Immunol.* **6,** 139–159.
16. Javaherian, K., Langlois, A. J., McDanal, C., Ross, K. L., Eckler, L. I., Jellis, C. L., Profy, A. T., Rusche, J. R., Bolognesi, D. P., Putney, S. D. & Matthews, T. J. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 6768–6772.
17. Modrow, S., Hahn, B. H., Shaw, G., Gallo, R. C., Wong-Staal, F. & Wolf, H. (1987) *J. Virol.* **61,** 570–578.
18. Leonard, C. K., Spellman, M. W., Riddle, L., Harris, R. J., Thomas, J. N. & Gregory, T. J. (1990) *J. Biol. Chem.* **265,** 10373–10382.
19. McKeating, J. A., Gow, J., Goudsmit, J., Pearl, L. H., Mulder, C. & Weiss, R. (1989) *AIDS* **3,** 777–784.
20. Simmonds, P., Balfe, P., Peutherer, J. F., Ludlam, C. A., Bishop, J. O. & Leigh Brown, A. J. (1990) *J. Virol.* **64,** 864–872.
21. Zhang, L. Q., Simmonds, P., Ludlam, C. A. & Leigh Brown, A. J. (1991) *AIDS* **5,** 675–681.
22. Saitou, N. & Nei, M. (1987) *Mol. Biol. Evol.* **4,** 406–425.
23. Stewart, C. B., Schilling, J. W. & Wilson, A. C. (1987) *Nature (London)* **330,** 401–404.
24. Saitou, N. & Imanishi, T. (1989) *Mol. Biol. Evol.* **6,** 514–525.
25. Wolfs, T. F. W., Zwart, G., Bakker, M., Valk, M., Kuiken, C. L. & Goudsmit, J. (1991) *Virology* **185,** 195–205.
26. Anderson, R. M. & May, R. M. (1982) *Parasitology* **85,** 411–426.
27. Nowak, M. A., May, R. M. & Anderson, R. M. (1990) *AIDS* **4,** 1095–1103.
28. Nowak, M. A., Anderson, R. M., McLean, A. R., Wolfs, T. F. W., Goudsmit, J. & May, R. M. (1991) *Science* **254,** 963–969.
29. Yokoyama, S. & Gojobori, T. (1987) *J. Mol. Evol.* **24,** 330–336.