Variation in Simian Immunodeficiency Virus *env* Is Confined to V1 and V4 during Progression to Simian AIDS

JULIE OVERBAUGH,[1,2]* LYLE M. RUDENSEY,[1,2] MARK D. PAPENHAUSEN,[1,2]
RAOUL E. BENVENISTE,[3] AND W. R. MORTON[2]

Department of Microbiology, SC-42,[1] and Regional Primate Research Center,[2] University of Washington, Seattle, Washington 98195, and Laboratory of Viral Carcinogenesis, National Cancer Institute, Frederick, Maryland 21702[3]

We have monitored changes in the simian immunodeficiency virus (SIV) envelope (*env*) gene in two macaques which developed AIDS after inoculation with a molecular clone of SIV. As the animals progressed to AIDS, selection occurred for viruses with variation in two discrete regions (V1 and V4) but not for viruses with changes in the region of SIV *env* that corresponds to the immunodominant, V3 loop of human immunodeficiency virus. Within the highly variable domains, the vast majority of nucleotide changes encoded an amino acid change (98%), suggesting that these envelope variants had evolved as a result of phenotypic selection. Analysis of the biological properties of these variants, which have been selected for in the host, may be useful in defining the mechanisms underlying viral persistence and progression to simian AIDS.

Variability within the human immunodeficiency virus (HIV) genome begins with changes introduced by the highly error-prone HIV reverse transcriptase (26, 27). Synthesis of the viral RNA may also be highly error prone, so that typically the viral particles that leave a cell are genetically different from the one(s) that infected the cell. Presumably, certain variants are then selected because they are more successful at escaping the immune response or at infecting new populations of cells. Interestingly, the viruses present late in infection when there is severe immunodeficiency tend to be more virulent to cells in culture than the viruses present before symptoms have developed (2, 8, 10, 30, 35). The most extensive variation has been observed in the gene coding for the surface envelope glycoprotein, gp120 (1, 15, 34, 36), a protein that, because of its exposed location, is likely to be a key target for the immune system. This protein also determines the cell tropism of the virus (7, 18, 22, 37), and it is an important determinant for cytopathicity of the virus in CD4-positive T lymphocytes (7, 11, 17, 33, 37), suggesting that variation in the envelope gene (*env*) may affect pathogenicity.

Part of the difficulty in studying the effect of HIV variation on the progression of AIDS is that one is not certain of the sequence(s) of the infecting virus. Furthermore, lentivirus sequences typically are analyzed after virus has been propagated in cell culture, and it is not clear whether these cell culture-adapted viruses are representative of the viruses present in the host. A model system in which variation can be measured in vivo, starting from a molecularly defined virus stock, would eliminate some of the complexities presented in the analysis of HIV variation. Recent studies have presented analysis of variation in HIV type 2 (HIV-2) (12) and simian immunodeficiency virus (SIV) (6) sequences in macaques infected with cloned virus, but both of these studies focused only on analysis of viral genetic change in animals who were asymptomatic. To analyze virus variation as it relates to disease progression, we have used direct cloning and sequence analysis to study the development

of changes in virus sequence in two macaques (*Macaca nemestrina*) which developed AIDS after inoculation with a molecular clone of SIV (SIV Mne CL8 [3]). One of these animals (M87004) showed a marked drop in CD4+ lymphocytes 1 year after infection, but he is still alive, with a low level of CD4+ lymphocytes, 3 years after infection. The other animal (F87047) was euthanized 83 weeks postinfection (p.i.), just 2 months after her CD4+ lymphocytes level began to drop, because of bleeding due to severe thrombocytopenia (3). Peripheral blood lymphocytes (PBLs) were obtained from M87004 at five time points, between 9 weeks and 121 weeks p.i., and from F87047 at 81 weeks p.i. To look directly for SIV variants in these 6 samples, without any of the selection that might be imposed by cell culture, SIV envelope sequences were amplified by polymerase chain reaction (PCR).

To minimize bias against variant genomes, primers for PCR were chosen that recognized conserved sequences in the SIV/HIV-2 envelope gene. Two rounds of amplification were used to generate a sufficient amount of product for cloning. The oligonucleotides used as primers (along with their relative positions of homology to SIV Mne [GenBank accession number M32741], shown in brackets) are as follows: round 1, SIV-env1 (5'-*GGTA*CCCTCTTTGAGACCT CAATAAA [6357 to 6376]) and SIV-env8 (5'*ATAGAATTC* CCAATTGGAGTGATCTCTAC [7594 to 7575]); round 2, SIV-env7 (5'*GACGCTACC*TAAAGCCTTGTGTAAAATTA [6364 to 6383]) and SIV-env4 (5'-*CTTAAG*TCAAGACGGTG GAGACGTGA [7543 to 7524]). Non-SIV sequences are shown in italics. Extraction of DNA and preparation of PCR mixtures were carried out in a closed laboratory isolated from PCR products and plasmids. Contamination in the reagents used for genomic DNA preparation was controlled for by amplifying DNA from uninfected cells isolated in parallel to each sample. The controls were considered negative only if no product could be detected after two rounds of amplification at the level of sensitivity of Southern blot analysis using a SIV *env*-specific probe.

The amplified region included five variable domains in SIV/HIV-2 *env* (21) that correspond to the domains called V1 through V5 in the HIV *env* sequence (20). The product of

_____
* Corresponding author.

amplification was cloned into M13mp18, and multiple clones were prepared and sequenced for each of the six samples. We tried to ensure that these clones represented several distinct variants by choosing, in each case, clones that came from two or three separate amplification reactions and by cloning the products of reactions likely to have started with multiple SIV genomes. In preparation for analysis of SIV variation in vivo, we determined the error rate of *Taq* polymerase by amplifying a lambda clone of SIV Mne under our reaction conditions; of 7,144 nucleotides analyzed, a total of 15 errors in nucleotide incorporation, 6 of which would predict amino acid changes and 1 of which would result in a frameshift, occurred.

To see the general pattern of variation, we began by analyzing a late sample taken more than a year after M87004 started to show symptoms of clinical AIDS. The products from three independent PCR amplifications of genomic DNA, isolated from a PBL sample taken 121 weeks p.i., were cloned, and the complete nucleotide sequence of the SIV *env* product (1,140 nucleotides) was determined for 18 clones. Two regions of extensive variation, corresponding to previously defined hypervariable domains in HIV *env* were found (20) (Fig. 1). The most dramatic variation was in variable region 1 (V1), where all 18 clones diverged significantly from the inoculated virus, SIV Mne CL8. Each clone had changed 7 to 12 of the 47 amino acids in the region of SIV that most closely corresponds to the V1 region of HIV. Almost half of the amino acids in the C-terminal portion of V1 had changed in some clones. In addition, several in-frame deletions and in-frame insertions (of 6, 9, 12, and 27 nucleotides), the majority of which resulted in insertion of additional serine and threonine residues, were detected in V1. The sequences which create the insertions have about 70% homology with the adjacent sequences. The divergence occurs at nucleotides in the first or second position of the codon, leading to the unique amino acid sequences of the insert. In general, the base substitutions in V1 were biased heavily towards codons for serine and threonine; 60% of the changes observed in this region were to these two amino acids. One other, small region of variation was detected and corresponds to variable region 4 (V4) in HIV *env*. Each of 18 clones showed a change of three or four out of five amino acids within V4, with the change often substituting a charged amino acid and being likely either to create or to remove a potential site for N-linked glycosylation (NXT or NXS). V4 is adjacent to a region of SIV *env* that is homologous to one of the regions important for binding of HIV-1 to the CD4 receptor (9, 16, 23), but that region itself, as well as the rest of the SIV *env* gene (including regions equivalent to HIV-1 V2, V3, and V5), showed very few changes. The number of frameshift mutations that occurred (three) is consistent with the number predicted for errors due to amplification with *Taq* polymerase, and these mutations were found in a subset of clones likely to have been generated from a single template. Thus, it is difficult to say which, if any, of the variants present in vivo actually encoded a frameshift mutation.

Having identified which regions were most variable late in the course of simian AIDS, we looked at earlier samples to determine whether there was a correlation between emergence of these variant genomes and the appearance of disease. Macaque M87004 developed clinical immunodeficiency about 55 weeks after infection (3), and we amplified sequences from PBLs taken at 9, 35, 78, and 81 weeks and determined the nucleotide sequences of three discrete regions of the genome, V1, V3, and V4.

Variation in V1 was evident as early as 35 weeks postinoculation and steadily increased from then onward (Fig. 2A). The functional significance of the V1 region of SIV and HIV has yet to be determined, but an epitope reactive to the serum of M87004 has been mapped to the SIV V1 region (5), so it is possible that changes in V1 allow the virus to escape the immune response. Indeed, one explanation for the selection of variants with added serine and threonine residues, which provide potential sites for O-linked glycosylation, is that the addition of carbohydrates could serve to mask a nearby epitope.

Within V4, where a cluster of amino acid changes had been found at 121 weeks, similar changes could be detected at 78 and 81 weeks but not at 9 and 35 weeks (Fig. 2A). The predominance of variants with an altered potential N-linked glycosylation site near a receptor-binding domain late in infection suggests that these variants can rapidly replicate, perhaps because of enhanced affinity for the CD4 receptor, in the absence of an active immune system. Similar variability in the N-linked glycosylation sites in V4 has been observed in HIV isolates from a group of patients infected by a single pool of contaminated factor VIII (31). During the evolution of SIVmac in vivo, a new potential N-linked glycosylation site in V4 was also observed (6), but in contrast to the results with SIV Mne, this change occurred in an asymptomatic macaque.

When we sequenced the clones present late in infection, at 121 weeks, we were surprised to find that the V3 loop region had not undergone variation, because it is the region that corresponds to the variable, immunodominant neutralization epitope of the HIV envelope (14, 19, 25, 29). Therefore we looked at the V3 region earlier in the course of infection (Fig. 2A). At 9, 35, and 78 weeks the sequence was well conserved. However, at 81 weeks, clones from one reaction showed significant variation in V3 (Fig. 2A). Several of the clones coming from this amplification also showed a high degree of variation, which extended outside the V3 loop and included several changes of cysteine residues and changes that generated stop codons (data not shown). In contrast, 12 clones from a separate amplification of the same DNA were conserved in V3 and in flanking regions, and V3 was similarly invariant at 78 weeks. These results could indicate that there was a minor population containing some unusual variants at 81 weeks, but we cannot exclude the possibility that the changes in one group of clones were the result of an unexplained burst of *Taq* polymerase errors.

In order to test whether the pattern of variation observed in M87004 was typical of the evolution of SIV in vivo, PBLs from a second macaque (F87047) inoculated with SIV Mne CL8 were analyzed 81 weeks p.i., just a few weeks after depletion of CD4$^+$ lymphocytes began. The nucleotide sequences of 10 clones were determined for V1, V3, and V4 (Fig. 2B). Like the sequences obtained from M87004, the V3 region was conserved. Extensive variation in V4 was not observed, consistent with the suggestion that V4 variants begin to emerge after the immune system is depressed. Once again, the greatest variation was found in V1, with most of the changes coding for serine or threonine residues. Thus, similar variants were found in the two macaques. We cannot exclude the possibility that some of the variants were present in the inoculum, because the molecular clone was briefly expressed in cell culture to generate infectious virus for inoculation. However, we think this is unlikely because none of the variants present in F87047 were identical to those in M87004. Furthermore, if the cluster of variation observed in the macaques simply reflected sequences in

```
                                        V1                              V2
env a. a.    105
          I.........I.........I........     .I.........I.........I.........I.........I.........I.........I.........I.........I.........I
mcSIVMne  TPLCITMKCNKSETDKWGLTKSSTTTAP---------TAIPTKAEAIKVVNENSPCINHDNCTGLEQEPMISCKFNMTGLKRDKRREYNETWYSADLVCEQGNSTENESRCYMNHCNTSVIQESCDKHY

1-2     ┌ ....V..........R...........S..........A..GT........S.....L.....................K..........................................
1-11    │ ....V..........R...........S..........A..GT........S.....L.....................K..........................................
1-1     └ ....V....D.....R...........S..........A..GT........S.....L.....................K..........................................
1-12    ┌ ...............R...........TT.........T.P--........S....L......................K..........................................
1-3     └ ...............R...........TT.........T.P--........S....L......................K..........................................
1-6     ┌ .................I...ATKP.....T.S.RT.T......S....DY.....................K.........T........................................
1-9     └ .................I...ATKP.....T.S.RT.T......S....DY.....................K.........T........................................
1-7     ┌ ..........N............A.TTTT.......TTA.PT...E...K..S...Y...............K..........................................
1-10    └ .......................A.TTTT.......TTA.PT...E...K..S...Y...............K..........................................
1-5     ⊏ ...............R...........I..TTTSTTTTS..A.IGK.......S.....L...................K...........................I..............
1-8     ⊏ ...............I..TTTSTTTTS..A..GT..TE....S.S...Y...............K..........................................
1-4     ⊏ .................R......S..........A.KET..E-...K..S...Y...............K..........................................

2-5     ┌ ...............R...........TT.........T.P--........S....T......................K.........T.......................
2-7     │ ...............R...........TT.........T.P--........S....T......................K.........T.......................
2-8     └ ...............R...........TT.........T.P--........S....T......................K.........T.......................
2-2     ┌ ...............I..TTTSTTTTS..A..GT...E...K..S...Y.....................................
2-3     │ ...............I..TTTSTTTTS..A..GT...E...K..S...Y.....................................
2-1     └ ...............I..TTTSTTTTS..A..GT...E...K..S...Y.....................................
```

```
                                                                    V3
          225
          I.........I.........I.........I.........I.........I.........I.........I.........I.........I.........I.........I.........I
mcSIVMne  WDAIRFRYCAPPGYALLRCNDTNYSGFMPNCSKVVVSSCTRMMETQTSTWFGFNGTRAENRTYIYWHSKDNRTIISLNKYYNLTMKCRRPGNKTVLPVTIMSGLVFHSQPINDRPKQAWCRFEGNWKEAIK

1-2       ...........................................................................................................................
1-11      ...........................................................................................................................
1-1       ...........................................................................................................................
1-12      .................................................................................................................W.........
1-3       ....................................................................................................P.....................
1-6       ...........................................................................................................................
1-9       .........................................................K...............................................W.................
1-7       ...........................................................................................................................
1-10      ...........................................................................................................................
1-5       ...........................................................................................................................
1-8       .........................................................................................................W.................
1-4       ...........................................................................................................................

2-5       ...........................................................................................................................
2-7       ....K......................................................................................................................
2-8       ..........................................................................................A................................
2-2       ...........................................................................................................................
2-3       ...........................................................................................................................
2-1       .............................................................frameshift....................................................
```

```
                                            V4                CD4              V5
          355
          I.........I.........I.........I.........I.........I.........I.........I.........I.........I.........I.........I.........I
mcSIVMne  EVKQTIVKHPRYTGTNNTDKINLTAPGGGDPEVTFMWTNCRGEFLYCKMNWFLNWVEDKNLTGTTQKPQERHKRNYVPCHIRQIINTWHKVGRNVYLPPREGDLTCNSTVTSLIANIDWIDGNQTNITM

1-2       ....................................................R..NE....T....................K.....................................
1-11      ....................................S....................R..NE....T....................K.....................................
1-1       ....................................................R..NE....T....................K.....................................
1-12      ....................................................RD.NR.........................K.....................................
1-3       ....................................F...............RD.NR..............................................................
1-6       ....................................................D.NR............................................................
1-9       ....................................................RD.NR..............................................................
1-7       ....................................................R..NE..............................K.....................................
1-10      ....................................................R..NE....T.........................K.....................................
1-5       ....................................................D.NR............................................................
1-8       ....................................F...............RD.NR..............................................................
1-4       ................................D...................R..NE.............................K.....................................

2-5       ....................................H...............R..NK............................................................
2-7       ......frameshift
2-8       ......frameshift
2-2       ....................................................D.NR.................S...........................................
2-3       ...............................................A...D.NR..............................................................
```
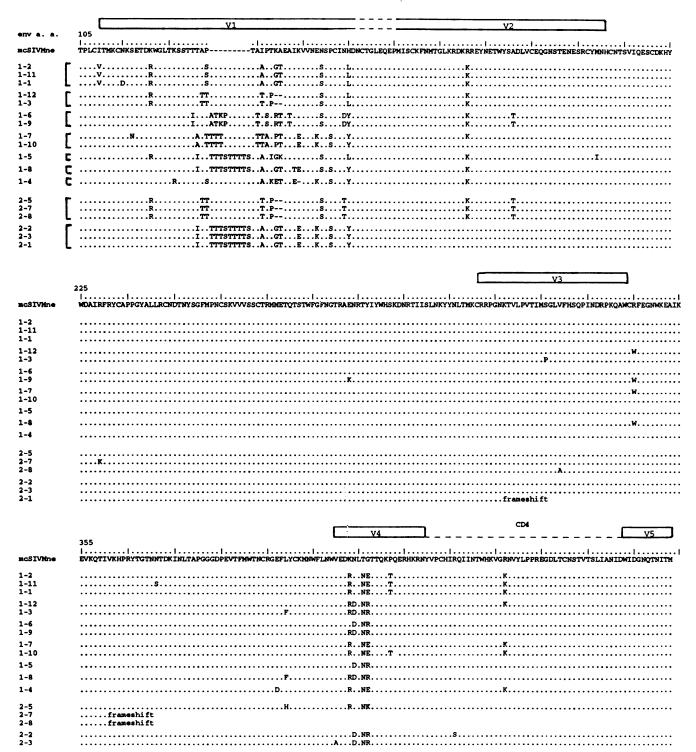
FIG. 1. Comparison of the predicted amino acid sequence of SIV Mne CL8 with that of SIV variants present in the lymphocytes of M87004 at 121 weeks p.i. The amino acid (a. a.) sequence of the SIV Mne CL8 envelope (GenBank number M32741), from amino acid 105 to 485, is shown in the single-letter amino acid code. Above the sequence, the location of the regions in SIV *env* that align with V1 through V5 and with a region important for binding to the cellular CD4 receptor (CD4) in HIV *env* (9, 16, 24) are shown. Predicted amino acid differences in each of 18 clones, representing the products of PCR amplification of SIV *env* sequences in DNA from PBLs taken at 121 weeks p.i., are shown below. Dots indicate no change in amino acid. Clones 1-1 through 1-12 represent products from a reaction mixture containing 150 ng of total lymphocyte DNA, and clones 2-1 through 2-8 represent product pooled from mixtures from two separate amplification reactions that each included 25 ng of lymphocyte DNA. From our analysis of *Taq* error rate, approximately one amino acid change per clone is predicted to be due to *Taq* error. For this analysis, clones with three or fewer amino acid differences were considered as possibly the same variant, so seven is a minimum estimate of the number of variants that was detected in the first set of clones. The grouping of variants that might be from the same template genome is shown in brackets at the beginning of the first line of each sequence. The second set of clones appears to represent just two distinct variants, suggesting that each 25-ng sample contained a single template molecule. Consistent with this interpretation is that reactions with less than 25 ng of lymphocyte DNA did not yield SIV *env* product. We believe that the frameshift mutations are due to errors in amplification because they occurred in a subset of an otherwise identical group of clones and because one frameshift was observed in 7,144 nucleotides amplified from the clone DNA. The variants in the second set of clones appear to be distinct from all of those in the first set.

```
                        ┌─────────────V1─────────────┐         ┌──────────V3──────────┐        ┌──────V4──────┐
env a.a.      112                                           311                                408
              ITMKCNKSETDKWGLTKSSTTTAPTAIPTKAEAIKVVNENSPCINHD---//---CRRPGNKTVLPVTIMSGLVFHSQPINDRPKQAWC---//---WVEDKNLTGTTQKPQERHKRN

A. M87004
9wk:1-1       ....................................................    ..............................    ...................
9wk:1-2       ....................................................    ..............................    ...................
9wk:1-4       ....................................................    ..............................    ...................
9wk:1-5       ....................................................    ..............................    ...................
9wk:1-6       ....................................................    ..............................    ...................
9wk:1-3       ...........................L....................S...    .................T............    ...................

9wk:2-1       ....................................................    ..............................    ...................
9wk:2-2       ....................................................    ..............................    ...................
9wk:2-3       ....................................................    ..............................    ...................
9wk:2-4       ....................................................    ..............................    ...................
9wk:2-5       ....................................................    ......................K.......    ...................
9wk:2-6       ....................................................    .................T....K.......    ...................

35wk:1-1      ...........................TS..T.........S.....         ..............................    ...................
35wk:1-3      ...........................TS..T.........S.....         ..............................    ...................
35wk:1-4      ...........................TS..T.........S.....         ..............................    ...................
35wk:1-5      ...........................TS..T.........S.....         ..............................    ...................
35wk:1-6      ...........................TS..T.........S.....         ..............................    ...................
35wk:1-2      T..........................TS..T.........S.....         ..............................    ...................

35wk:2-1      ....................................................    ..............................    ...................
35wk:2-2      ....................................................    ..............................    ...................
35wk:2-10     ...........................TT..RT.........L....         ..............................    ...................
35wk:2-3      ...........................TT..RT.........L....         ..............................    ...................
35wk:2-4      ...........................TT..RT.........L....         ..............................    ...................
35wk:2-5      ...........................TT..RT.........L....         ..............................    ...................

78wk:1-7      ....................................................    ..............................    ...................
78wk:1-1      ...........................I....TT..ST.........S...     ..............................    .....D.N...........
78wk:1-3      ...........................I....TT..ST.........S...     ..............................    .....D.N...........
78wk:1-8      ...........................I....TT..ST.........S...     ..............................    .....D.N...........
78wk:1-10     ...........................I....TT..ST.........S...     ..............................    .....D.N...........
78wk:1-11     ...........................I....TT..ST.........S...     ..............................    .....D.N...........
78wk:1-12     ...........................I....TT..ST.........S...     ..............................    .....D.N...........
78wk:1-2      ...........................P.I..TT..ST.........S...     ..............................    .....D.N...........
78wk:1-5      ...........................I....TT..ST.........S...     .S............................    .....D.N...........
78wk:1-4      ...........................I....TT..ST.........S...     ..............................    .....D.N...R.......
78wk:1-6      ...........................I....TT..ST.........S...     ..............................    .....D.NR..........
78wk:1-9      ...........................I....TT..ST.........S...     ..............................    .....G.N...........

78wk:2-5      ....................................................    ..............................    ...................
78wk:2-6      ....................................................    ..............................    ...................
78wk:2-7      ....................................................    ..............................    ...................
78wk:2-11     ....................................................    ..............................    ...................
78wk:2-12     ....................................................    ..............................    ...................
78wk:2-1      ....................................................    ..............................    G..................
78wk:2-2      ...........................V....TT..ST.........S...     ..............................    .....D N...........
78wk:2-3      ...........................I....TT..ST.........S...     ..............................    .....D N.....L.....
78wk:2-10     ...........................I....TT..ST.........S...     ..............................    .....D N...........
78wk:2-4      ...........................I....TT..ST.........S.S.     ..............................    .....D N...........
78wk:2-8      ...........................I....TT..ST.........S.S.     ..............................    .....D N...........
78wk:2-9      ...........................I....K-..ST.........S...     ..............................    .....D.N...........

81wk:1-6      ...........................I....TT..ST.........S...     ..............................    .....D.N...........
81wk:1-9      ...........................STA.PT.............S...      ....................L.........    ..K....N...........
81wk:1-2      .....................T..TT..RT.T.............S...       ....................L.........    .V.....N...........
81wk:1-5      .....................T.ATT..RT.T.............S...       ....................L.........    .....D.N...........
81wk:1-7      ...........................S..A..GT...........S...      ..............................    .... D.NR..........
81wk:1-8      ...........................S..A..GT...........S...      ..............................    .....D.N...........
81wk:1-10     ...........................S..A..GT...........S...      .......A......................    *....D.N...........
81wk:1-12     ...........................S..A..GT...........S...      .......A......................    .....D.N...........
81wk:1-3      ...........................S..A..GT...........S...      .....T.........K.............Y    IKN...NK.....R.....
81wk:1-11     ..I.Y.........K...........S..A..GT.........KS....N      ...............L.K...........Y    .IKN...NK.....framesh
81wk:1-4      ..I.Y...A....K...........S..A..GT.........KS....N       ..............................    .....D.N...........
81wk:1-1      ..I.Y.........K....L.......S..A..GT.........KS....N      ...............K.............Y    .....D.N...........

81wk:2-9      ..........................TG.--...........S.......      ..............................    .....D.N...........
81wk:2-4      ...........................I....TT..ST.........S...     ..............................    .....D.N...........
81wk:2-6      ...........................TTS.RT.............S...      ..............................    .......N...........
81wk:2-7      ...........................TTS.RT.............S...      ..............................    .......N.........R..
81wk:2-5      ...........................TTS.RT.............S...      ..............................    .....D.N...........
81wk:2-8      ...........................TTS.RT.............S...      ..............................    .....D.N...........
81wk:2-10     ...........................TTS.RT.............S...      .................T............    .....D.N...frameshift
81wk:2-1      ...........................TTS.RT...........S..S.       ..............................    .....D.N...frameshift
81wk:2-2      ...........................TTSITT.............S...      ..............................    .......N...........
81wk:2-12     ...........................TTSITT.............S...      ..............................    .......N...........
81wk:2-3      ...........................TTSITT.............S...      ..............................    .....D.N...........
81wk:2-11     ...........................TTSITT.............S...      ..............................    .....D.N...........

B. F87047
81wk:1-11     ..........................TT...T...N....S......         ..............................    ...................
81wk:1-1      ..........................TT...T...N....S......         ..............................    ...............W....
81wk:1-4      ..........................TT...T...N....S......         ..............................    ...............W....
81wk:1-5      ..........................TT...T...N....S......         ..............................    ...............W....
81wk:1-7      ..........................TT...T...N....S......         ..............................    ...............W....
81wk:1-8      ..........................TT...T...N....S......         ..............................    ...............W....
81wk:1-9      ..........................TT...T...N....S......         ..............................    ...............W....
81wk:1-10     ..........................TT...T...N....S......         ..............................    ...............W....
81wk:1-12     ..........................TT...T...N....S......         ..............................    ...............W....
81wk:1-2      ..........................TT...T...N....S......         .................T............    ...............W....
```

FIG. 2. (A) Comparison of the predicted amino acid sequence of SIV Mne CL8 with those of variants present in lymphocytes of M87004 at 9, 35, 78, and 81 weeks p.i. The amino acid (a.a.) sequences of three regions of SIV Mne CL8 *env* are shown in line 1 with double slashes to indicate where the sequence is discontinuous. The envelope amino acid number for each region is shown above the line, and variable regions V1, V3, and V4 are noted. Clones from the earliest time point are shown first, with progressively later time points following. The time after infection when the sample was taken is indicated in the name of the clone, and clones are clustered by sample date and by independent PCRs (1 or 2). An asterisk indicates a stop codon. The amounts of total genomic DNA used in the PCRs were as follows (with the minimal amount of DNA from the sample which yielded SIV *env* product given in parentheses): at 9 weeks p.i., clones 1-1 to 1-6 had 1 µg and clones 2-1 to 2-6 had 2 µg (500 ng); at 35 weeks p.i., clones 1-1 to 1-6 had 1 µg and clones 2-1 to 2-12 had 2 µg (500 ng); at 78 weeks p.i., clones 1-1 to 1-12 had 300 ng and clones 2-1 to 2-12 had 1 µg (150 ng); and at 81 weeks p.i., clones 1-1 to 1-12 had 150 ng and clones 2-1 to 2-12 had 50 ng (25 ng). (B) Similar analysis for clones of PCR product from F87047 at 81 weeks. The clones were obtained from a single PCR.

which errors in reverse transcription frequently occurred, similar variants would be observed when SIV Mne CL8 replicated in cell culture, and they were not (28). Whatever the origin of the variants, it is clear from the samples obtained at different time points that selection for the variants was occurring in the infected animal.

We have suggested that the standard methods for isolating HIV and SIV from a patient, which involve propagation of the virus in cell culture prior to cloning, may select for just a subset of variants that grow best in cell culture (24). However, direct amplification of viral sequences by PCR might be misleading because it could pick up sequences that are not expressed and therefore are probably unimportant. The data presented here suggest that most of the proviruses detected by PCR were in fact expressed because there had obviously been selection for particular envelope proteins. This is evident in several ways. First, several of the predominant variants, such as the variants bearing nine amino acid insertions in V1 at 121 weeks p.i. (1-5, 1-8, and 1-1), clearly represent distinct genomes (Fig. 1). In the absence of their expression, this could occur only if the identical insertion, as well as several other single-amino-acid changes common to each variant, had occurred many times during reverse transcription. Second, several of the variants detected early after infection appear to be progenitors of variants detected later in infection. For example, the variation in clones from the first PCR of the 35-week sample is seen in the clones from the second PCR of the 81-week sample, but the latter clones have additional changes (Fig. 2A).

The most compelling evidence that the cloned *env* genes were expressed is the strong selection that is evident for nucleotide changes that encoded an amino acid change in the two variable regions. If the mutations in the nucleotide sequence of V1 and V4 were random, then approximately 79% of the nucleotide changes would be predicted to yield an amino acid change. However, about 98% of the 818 nucleotide changes found in V1 and V4 encoded an amino acid change (Table 1). Approximately 8% (64 of 799) of the amino acid changes in these two variable regions were the result of a double mutation in a single codon; in all these cases, both nucleotide changes were needed to effect the particular amino acid change observed. In comparison, in the rest of the *env* sequence (excluding V1 and V4) in the clones from 121 weeks p.i., only 52% of the nucleotide changes caused amino acid changes. Similar results (57%) were observed when we analyzed the percentage of nucleotide changes that encoded an amino acid change in a relatively conserved region of SIV *env* (mid-*env*) in all the clones obtained from samples 9 through 121 weeks p.i. While the changes observed in V1 and V4 appear to impart a strong selective advantage to the variants, it seems that in other regions of *env* there are some constraints against amino acid changes, perhaps because they disrupt function of the *env* protein. Taken together, these observations strongly support the

hypothesis that the variants detected in the present study represent viruses that had evolved in vivo as a result of phenotypic selection.

To better understand the epidemiology of AIDS, attempts to document the evolution of primate lentiviruses have been made (32). This analysis has been limited to some extent by lack of information on the rate of change of lentivirus sequences (32). From the data at 121 weeks, we calculated that there is approximately 0.56% nucleotide change and 1.5% amino acid change per year in the SIV Mne gp120 in an infected macaque. It was not possible to determine whether this rate was linear throughout asymptomatic infection and disease progression because at early time points, we primarily analyzed regions of *env* that were variable (V1 and V4). The rate of change of SIV Mne was similar to that reported for asymptomatic rhesus macaques infected with SIVmac (0.73% nucleotide and 1.8% amino acid change per year [6]). However, in the study of SIVmac, variation was more evenly distributed throughout the variable regions, although variation was slightly more pronounced in V1 and V4 (6). In this study, we observed a high percentage of G→A (29%) and A→G (29%) transitions, which is consistent with the predominant types of nucleotide change observed in the evolution of SIVmac (6) and HIV-1 (13).

By direct analysis of the viral genomes present in two macaques that developed AIDS after inoculation with a cloned virus, two regions of hypervariation were clearly delineated. Surprisingly, the V3 loop region, which is variable among HIV isolates and has been described as a principal neutralizing epitope for HIV (14, 19, 25, 29), was conserved. It is important to note that the V3 loops of SIV and HIV-2 diverge significantly from the V3 loop of HIV-1

TABLE 1. Analysis of the percentage of nucleotide changes that code for amino acid changes

Region	No. of nt changes coding for aa change[a]	No. of nt changes not coding for aa change	% of total nt changes coding for aa change
V1[b]	626	14	98
V4[b]	173	5	97
121-wk *env* [−(V1 + V4)][c]	43	40	52
Mid *env*[b,d]	77	59	57
Random (V1 or V4)[e]	79	21	79

[a] nt, nucleotide; aa, amino acid.
[b] Analysis of 94 clones from the six PBL samples of M87004 and F87047.
[c] Includes the sequences outside of V1 and V4 in the clones analyzed at 121 weeks (Fig. 1).
[d] Stretch of approximately 220 bases in the middle of *env* which includes V3.
[e] For the calculation of random mutations, the numbers are arbitrarily presented for 100 total mutations. The numbers were derived from the exact sequence of V1 and V4.

and that the role of V3 in neutralization of SIV has not been carefully analyzed. However, recent studies demonstrated that the V3 loop of HIV-2, which is more closely related to SIV than to HIV-1, performs a similar role in neutralization (4). These SIV *env* variant sequences, isolated from a macaque that developed AIDS after infection with a molecularly cloned virus, can now be incorporated into the parental SIV genome. Such chimeric viruses, with defined differences in the envelope, can be used to determine the functional, immunogenic, and pathogenic functions of these variants that were selected for in the host.

REFERENCES

1. **Alizon, M., S. Wain-Hobson, L. Montagnier, and P. Sonigo.** 1986. Genetic variability of the AIDS virus: nucleotide sequence analysis of two isolates from African patients. Cell **46**:63–74.
2. **Asjo, B., J. Albert, A. Karlsson, L. Morfeldt-Manson, G. Biberfeld, K. Lidman, and E. M. Fenyo.** 1986. Replicative capacity of human immunodeficiency virus from patients with varying severity of HIV infection. Lancet **ii**:660–662.
3. **Benveniste, R. E., M. Gonda, J. Greenwood, P. Dorn, E. Clark, W. Morton, J. Overbaugh, and G. Heidecker.** Unpublished observations.
4. **Böttiger, B., A. Karlsson, P.-A. Andreasson, A. Nauclér, C. M. Costa, E. Norrby, and G. Biberfeld.** 1990. Envelope cross-reactivity between human immunodeficiency virus types 1 and 2 detected by different serological methods: correlation between cross-neutralization and reactivity against the main neutralizing site. J. Virol. **64**:3492–3499.
5. **Buchanan, T., G. Goodwin, and J. Overbaugh.** Unpublished observations.
6. **Burns, D. P. W., and R. C. Desrosiers.** 1991. Selection of genetic variants of simian immunodeficiency virus in persistently infected rhesus monkeys. J. Virol. **65**:1843–1854.
7. **Cheng-Mayer, C., M. Quiroga, J. W. Tung, D. Dina, and J. A. Levy.** 1990. Viral determinants of human immunodeficiency virus type 1 T-cell or macrophage tropism, cytopathogenicity, and CD4 antigen modulation. J. Virol. **64**:4390–4398.
8. **Cheng-Mayer, C., D. Seto, M. Tateno, and J. A. Levy.** 1988. Biologic features of HIV-1 that correlate with virulence in the host. Science **240**:80–82.
9. **Cordonnier, A., L. Montagnier, and M. Emerman.** 1989. Single amino-acid changes in HIV envelope affect viral tropism and receptor binding. Nature (London) **340**:571–574.
10. **Fenyö, E. M., L. Morfeldt-Månson, F. Chiodi, B. Lind, A. von Gegerfelt, J. Albert, E. Olausson, and B. Åsjö.** 1988. Distinct replicative and cytopathic characteristics of human immunodeficiency virus isolates. J. Virol. **62**:4414–4419.
11. **Fisher, A., L. Ratner, H. Mitsuya, L. M. Marsells, M. E. Harper, S. Broder, R. C. Gallo, and F. Wong-Staal.** 1986. Infectious mutants of HTLV-III with changes in the 3' region and markedly reduced cytopathic effects. Science **233**:655–659.
12. **Franchini, G., P. Markham, E. Gard, K. Fargnoli, S. Keubaruwa, L. Jagodzinski, M. Robert-Guroff, P. Lusso, G. Ford, F. Wong-Staal, and R. C. Gallo.** 1990. Persistent infection of rhesus macaques with a molecular clone of human immunodeficiency virus type 2: evidence of minimal genetic drift and low pathogenetic effects. J. Virol. **64**:4462–4467.
13. **Goodenow, M., T. Huet, W. Saurin, S. Kwok, J. Sninsky, and S. Wain-Hobson.** 1989. HIV-1 isolates are rapidly evolving quasispecies: evidence for viral mixtures and preferred nucleotide substitutions. J. Acquired Immune Defic. Syndr. **2**:344–352.
14. **Goudsmit, J., C. Debouck, R. H. Meloen, L. Smit, M. Bakker, D. M. Asher, A. V. Wolff, C. J. Gibbs, Jr., and C. Gajdusek.** 1988. Human immunodeficiency virus type 1 neutralization epitope with conserved architecture elicits early type-specific antibodies in experimentally infected chimpanzees. Proc. Natl.

15. **Hahn, B. H., M. A. Gonda, G. M. Shaw, M. Popovic, J. A. Hoxie, R. C. Gallo, and F. Wong-Staal.** 1985. Genomic diversity of the acquired immune deficiency syndrome virus HTLV-III: different viruses exhibit greatest divergence in their envelope genes. Proc. Natl. Acad. Sci. USA **82**:4813–4817.
16. **Lasky, L. A., G. Nakamura, D. H. Smith, C. Fennie, C. Shimasaki, E. Patzer, P. Berman, T. Gregory, and D. J. Capon.** 1987. Delineation of a region of the human immunodeficiency virus type 1 gp120 glycoprotein critical for interaction with the CD4 receptor. Cell **50**:975–985.
17. **Lifson, J. D., G. R. Reyes, M. S. McGrath, B. S. Stein, and E. G. Engleman.** 1986. AIDS retrovirus induced cytopathology: giant cell formation and involvement of CD4 antigen. Science **232**:1123–1127.
18. **Liu, Z., C. Wood, J. A. Levy, and C. Cheng-Mayer.** 1990. The viral envelope gene is involved in macrophage tropism of a human immunodeficiency virus type 1 strain isolated from brain tissue. J. Virol. **64**:6148–6153.
19. **Matsushita, S., M. Robert-Guroff, J. Rusche, A. Koito, T. Hattori, H. Hoshino, K. Javaherian, K. Takatsuki, and S. Putney.** 1988. Characterization of a human immunodeficiency virus neutralizing monoclonal antibody and mapping of the neutralizing epitope. J. Virol. **62**:2107–2114.
20. **Modrow, S., B. H. Hahn, G. M. Shaw, R. C. Gallo, F. Wong-Staal, and H. Wolf.** 1987. Computer-assisted analysis of envelope protein sequences of seven human immunodeficiency virus isolates: prediction of antigenic epitopes in conserved and variable regions. J. Virol. **61**:570–578.
21. **Myers, G., J. A. Berzofsky, A. B. Rabson, T. F. Smith, and F. Wong-Staal.** 1990. Human retroviruses and AIDS. Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, Mex.
22. **O'Brien, W. A., Y. Koyanagi, A. Namazie, J. Zhao, A. Diagne, K. Idler, J. A. Zack, and I. S. Y. Chen.** 1990. HIV-1 tropism for mononuclear phagocytes can be determined by regions of gp120 outside the CD4-binding domain. Nature (London) **348**:69–73.
23. **Olshevsky, U., E. Helseth, C. Furman, J. Li, W. Haseltine, and J. Sodroski.** 1990. Identification of individual human immunodeficiency virus type 1 gp120 amino acids important for CD4 receptor binding. J. Virol. **64**:5701–5707.
24. **Overbaugh, J., P. R. Donahue, S. L. Quackenbush, E. A. Hoover, and J. I. Mullins.** 1988. Molecular cloning of a feline leukemia virus that induces fatal immunodeficiency in cats. Science **239**:906–910.
25. **Palker, T., M. E. Clark, A. J. Langlois, T. J. Matthews, K. J. Weinhold, R. R. Randall, D. P. Bolognesi, and B. F. Hayes.** 1988. Type-specific neutralization of the human immunodeficiency virus with antibodies to env-encoded synthetic peptides. Proc. Natl. Acad. Sci. USA **85**:1932–1936.
26. **Preston, B. D., B. J. Poiesz, and L. A. Loeb.** 1988. Fidelity of HIV-1 reverse transcriptase. Science **242**:1168–1171.
27. **Roberts, J. D., K. Bebenek, and T. A. Kunkel.** 1988. The accuracy of reverse transcriptase from HIV-1. Science **242**:1171–1173.
28. **Rudensey, L., and J. Overbaugh.** Unpublished observations.
29. **Rusche, J. R., K. Javaherian, C. McDanal, J. Petro, D. L. Lynn, R. Grimaila, A. Langlois, R. C. Gallo, L. O. Arthur, P. J. Fischinger, D. P. Bolognesi, S. D. Putney, and T. J. Matthews.** 1988. Antibodies that inhibit fusion of human immunodeficiency virus-infected cells bind a 24-amino acid sequence of the viral envelope, gp120. Proc. Natl. Acad. Sci. USA **85**:3198–3202.
30. **Schwartz, S., B. K. Felber, E. Fenyo, and G. N. Pavlakis.** 1989. Rapidly and slowly replicating human immunodeficiency virus type 1 isolates can be distinguished according to target-cell tropism in T-cell and monocyte cell lines. Proc. Natl. Acad. Sci. **86**:7200–7203.
31. **Simmonds, P., P. Balfe, C. A. Ludlam, J. O. Bishop, and A. J. L. Brown.** 1990. Analysis of sequence diversity in hypervariable regions of the external glycoprotein of human immunodeficiency virus type 1. J. Virol. **64**:5840–5850.
32. **Smith, T. F., A. Srinivasan, G. Schochetman, M. Marcus, and G. Myers.** 1988. The phylogenetic history of immunodeficiency

viruses. Nature (London) **333**:573–575.

33. **Sodroski, J., W. C. Goh, C. Rosen, and W. H. K. Campbell.** 1986. Role of HTLV-III/LAV envelope in syncytium formation and cytopathicity. Nature (London) **322**:470–474.

34. **Starcich, B. R., B. H. Hahn, G. M. Shaw, P. D. McNeely, S. Modrow, H. J. Wolf, E. S. Parks, W. P. Parks, S. F. Josephs, R. C. Gallo, and F. Wong-Staal.** 1986. Identification and characterization of conserved and variable regions in the envelope gene of HTLV-III/LAV, the retrovirus of AIDS. Cell **45**:637–648.

35. **Tersmette, M., R. A. Gruters, F. de Wolf, R. E. Y. de Goede, J. M. A. Lange, P. T. A. Schellekens, J. Goudsmit, H. G. Huisman, and F. Miedema.** 1989. Evidence for a role of virulent human immunodeficiency virus (HIV) in the pathogenesis of acquired immunodeficiency syndrome: Studies on sequential HIV isolates. J. Virol. **63**:2118–2125.

36. **Willey, R. L., R. A. Rutledge, S. Dias, T. Folks, T. Theodore, C. E. Buckler, and M. A. Martin.** 1986. Identification of conserved and divergent domains within the envelope gene of the acquired immunodeficiency syndrome retrovirus. Proc. Natl. Acad. Sci. USA **83**:5038–5042.

37. **York-Higgins, D., C. Cheng-Mayer, D. Bauer, J. A. Levy, and D. Dina.** 1990. Human immunodeficiency virus type 1 cellular host range, replication, and cytopathicity are linked to the envelope region of the viral genome. J. Virol. **64**:4016–4020.