JOURNAL OF VIROLOGY, June 1993, p. 3345–3356
0022-538X/93/063345-12$02.00/0

Vol. 67, No. 6

# Selection for Specific Sequences in the External Envelope Protein of Human Immunodeficiency Virus Type 1 upon Primary Infection

L. Q. ZHANG,[1]† P. MACKENZIE,[1] A. CLELAND,[1] E. C. HOLMES,[1] A. J. LEIGH BROWN,[1] AND P. SIMMONDS[2]*

*Centre for HIV Research, University of Edinburgh, Edinburgh EH9 3JN,[1] and Department of Medical Microbiology, Medical School, University of Edinburgh, Teviot Place, Edinburgh EH8 9AG,[2] United Kingdom*

Received 18 December 1992/Accepted 5 March 1993

**Viral RNA was extracted from plasma samples collected from five individuals during the period of viremia before seroconversion in primary infection with human immunodeficiency virus type 1 (HIV-1) and amplified by polymerase chain reaction. Nucleotide sequence analysis of amplified DNA from the V3 and V4 hypervariable regions indicated that the initial virus population of each acutely infected individual was completely homogeneous in sequence. No intrasample variability was found among the 44,090 nucleotides sequenced in this region of *env*, contrasting with the high degree of variability normally found in seropositive individuals. Paradoxically, substantial sequence variability was found in the normally highly conserved *gag* gene (encoding p17) in most of the preseroconversion samples. The diversity of p17 sequences in samples that were homogeneous in V3 and V4 can most readily be explained by the existence of strong selection for specific *env* sequences either upon transmission or in the interval between exposure and seroconversion in the exposed individual. Evidence that localizes the selected region upon transmission to V3 is provided by the similarity or identity of V3 loop sequences in five individuals with epidemiologically unrelated HIV-1 infections, while regions flanking the V3 loop and the V4 hypervariable region were highly divergent. The actual V3 sequences were similar to those associated with macrophage tropism in primary isolates of HIV, irrespective of whether infection was acquired by sexual contact or parenterally through transfusion of contaminated factor VIII. Proviral DNA sequences in peripheral blood mononuclear cells remained homogeneous in the V3 and V4 regions (and variable in $p17^{gag}$) for several months after seroconversion. The persistence of HIV sequences in peripheral blood mononuclear cells identical to those found at primary infection in the absence of continued virus expression provides an explanation for the previously observed differences in the composition of circulating DNA and RNA populations in sequential samples from seropositive individuals.**

Infection with human immunodeficiency virus type 1 (HIV-1) is characterized by a slow but inexorable course of disease leading ultimately to AIDS. A striking feature of HIV infection is the diversity and rapid turnover of sequence variants within an infected individual (28, 32, 49, 58). In particular, the region of the viral genome that encodes the external membrane glycoprotein (gp120) encoded by the HIV *env* gene shows an extremely rapid rate of sequence change, concentrated into five hypervariable domains (V1 to V5) interspersed among less variable regions (C1 to C4). The V3 and V2 hypervariable regions have been shown to be the targets of neutralizing antibody elicited by natural infection (17, 18, 21, 26, 31, 41), and it has been speculated that high rates of sequence change in these regions may have adaptive value in allowing immunological escape from the initially effective antiviral immune response upon primary infection and enable long-term persistence of HIV (29, 59, 63). Alternatively, the observed sequence turnover of distinct gp120 sequence variants in plasma (49, 58) may result from sampling of variants that are adapted for infection of different cell types in vivo, such as cells of the monocyte/macrophage lineage. Supporting this latter hypothesis is the mapping of several determinants of in vitro phenotype observed upon culture of HIV, such as cell tropism, cytopathology, and ability to grow in transformed T-cell lines, to parts of the *env* gene that include the V3 hypervariable region (5, 16, 20, 36, 46, 50, 55, 56) and the observation that the phenotype of virus isolated from seropositive individuals switches from non-syncytium inducing (NSI) to syncytium inducing (SI) during disease progression (16, 44, 52).

The relationship between the heterogeneous virus population in long-term-infected individuals with that actually transmitted to a sexual or parenteral contact remains unclear. In particular, it is not known whether infection is normally initiated by a single virus particle or whether there is cotransmission of several variants to the contact. Given the evidence for selection for specific variants in long-term-infected individuals (19), it is clearly important to determine whether similar processes operate in the newly infected individual before seroconversion. It is known that sequences of virus in plasma collected from individuals shortly after infection are less variable than those found in the contact at the time of transmission (30, 60). There is also some evidence for selective transmission of NSI and macrophage-tropic variants, although whether this results from specific transmission of HIV variants with the NSI phenotype, or whether there is elimination of cotransmitted SI variants at the time of primary infection, remains unclear (40, 54).

The aim of this study was to analyze the initial virus population in the acute viremic stage of primary infection before seroconversion for antibody to HIV. We have carried out nucleotide sequence analysis of amplified viral RNA

* Corresponding author.
† Present address: Aaron Diamond AIDS Research Center, New York, NY 10016.

from plasma and proviral DNA sequences from peripheral blood mononuclear cells (PBMCs) of the V3 and V4 hypervariable regions and part of the *gag* gene from three individuals infected by sexual contact and six hemophiliacs with acute infection from transfusion of HIV-contaminated factor VIII. This procedure has enabled an investigation of the extent of variation in different regions of the viral genome, the relationship between *env* and *gag* sequences of individuals infected from a common source (27), and the nature of the V3 and V4 sequences upon transmission. The results presented here confirm that selection for particular HIV *env* variants does occur upon primary infection and provide an insight into the nature of the selection process operating before seroconversion.

## MATERIALS AND METHODS

**Samples.** Heparinized plasma samples collected around the time of seroconversion in three hemophiliacs, p82, p74, and p84, and three heterosexual contacts of HIV-positive individuals, Sc1, Sc2, and Sc3, were separated and stored on the day of collection at −20°C. PBMC samples collected 3 to 25 weeks after seroconversion from p28, p77, p79, p84, and Sc3 were prepared by separation of heparinized blood over Ficoll (Nyegaard); cells at the interface were washed in phosphate-buffered saline (PBS), resuspended in 50% fetal calf serum–40% PBS–10% dimethyl sulfoxide, and stored in aliquots in liquid nitrogen until required. Samples were also available from p82 5 years after infection and from s44, a long-term-infected male homosexual with AIDS.

**HIV serology.** Pre- and postseroconversion samples were tested for anti-HIV and p24 antigen by Abbott second-generation enzyme immunoassay and Coulter or Dupont p24 antigen assay, respectively, carried out in accordance with the manufacturers' instructions. p24 antigen was quantified by reference to a standard dilution series of the positive control.

**Nested PCR.** Viral RNA was extracted from 50- to 200-μl volumes of plasma and reverse transcribed as previously described (62). For analysis of proviral sequences, DNA was extracted from stored PBMCs as previously described (48). Amplification of proviral DNA or cDNA (except Sc3) was carried out with primers spanning V3 to V5, a and h (first polymerase chain reaction [PCR]), followed by a second PCR using primers b and g (47). For Sc3, a new primer, 633 (sequence GGAGGGGCATACATTGC, 5′ base position 7083 in $HIV_{HTLV\text{-}IIIB}$, clone HXB2r [34]), was used in place of primer h, and primer 634 (ATTCTGCATGGGAGTGTC, position 7029 in HXB2r) was used in place of primer g. Nucleotide sequences could in all cases be read from positions 6614 to 6800 (positions 844 to 1029 in the *env* region; V3 and flanking regions) and in the V4 and flanking regions from 6938 to 7022 (1165 to 1272 in *env*). For analysis of sequence variability, some sequences were slightly longer than this, extending upstream to position 6581 and downstream to 6848 (V3) and between 6938 and 7115 (V4).

Sequences in the $p17^{gag}$ region were amplified with primers of sequences GCGAGAGCGTCAGTATTAAGCGG (sense outer primer, position 342) and TCTGATAATGCT GAAAACATGGG (position 865), followed by a second nested PCR with sense primer GGGAAAAAATTCGGT TAAGGCC (position 382) and antisense primer CTTCTAC TACTTTTACCCATGC (position 817). Nucleotide sequences between positions 451 and 793 (p17 coding sequence plus the 5′ 60 bp of p24) were used for sequence comparisons and evolutionary analysis of the *gag* region.

Quantitation of provirus in PBMCs was carried out by limiting dilution as described previously (48). Both negative and positive controls were included in all PCR amplification reactions to monitor the sensitivity of the PCR and to detect possible contamination from product carryover. For nucleotide sequence determinations, cDNA or proviral DNA was amplified at limiting dilution to separate single molecules of target nucleic acid as previously described (48). Samples were diluted sufficiently before amplification to ensure that each nucleotide sequence were derived from one copy of viral RNA or DNA; frequencies of positive reaction were in all cases less than 20%. Using standard Poisson formulae, it can be calculated than greater than 90% of positive reactions would originate from single molecules of viral nucleic acid (48). Any sequence showing ambiguities was discarded from the analysis.

**Sequence analysis.** Amplified DNA derived ultimately from single molecules of the HIV genome by limiting dilution was directly sequenced as previously described (47, 48). Sequences were aligned by using PILEUP and LINEUP programs in the University of Wisconsin Genetics Computer Group package (11). Sequence variability was expressed as (i) the number of different nucleotide sequences within a sample, (ii) the number of polymorphic sites divided by the number of nucleotides sequenced within a sample ($P_n$), and (iii) the mean distance for each pairwise comparison between nucleotide sequences within a sample ($d$). Sequence distances were estimated by using the program DNADIST from the PHYLIP package (14) under an evolutionary model assigning different rates of transition and transversion and different frequencies of the four nucleotides.

Rates of silent substitution per silent site ($K_s$) and replacement substitution per replacement site ($K_a$) were estimated by the method of Nei and Gojobori (35).

**Nucleotide sequence accession numbers.** All sequences reported in this publication have been submitted to GenBank (accession numbers L13488 to L13589).

## RESULTS

**Analysis of HIV sequences upon primary infection.** Plasma samples were collected from three hemophiliacs acutely infected after transfusion of separate batches of HIV-contaminated factor VIII and from three heterosexual contacts with epidemiologically unrelated primary infections. In five cases (p74, p84, Sc1, Sc2, and Sc3), samples were available in the interval of transient viremia immediately prior to seroconversion for anti-HIV antibody. All samples were p24 antigen positive. Anti-HIV-positive, p24 antigen-negative samples were collected 2 to 12 weeks later from each, providing evidence for proximity to seroconversion (Table 1). The sixth sample (from p82) was an anti-HIV-positive, p24 antigen-negative sample collected shortly after seroconversion. Evidence for proximity to seroconversion in this case was provided by the finding of an anti-HIV- and p24 antigen-negative sample collected 76 days previously and restricted reactivity to HIV proteins in an immunoblotting assay (anti-p24 only; data not shown).

To investigate the extent of sequence diversity in the *env* region of initial virus population before seroconversion, cDNA was synthesized from primers downstream of the V3 and V4 hypervariable regions and titrated to a suitable limiting dilution to isolate single viral sequences. After amplification with nested primers, DNA from positive reactions was directly sequenced to prevent in vitro copying errors by *Taq* polymerase from influencing our genetic

TABLE 1. p24 and anti-HIV screening of samples collected during primary infection from five individuals

| No. | Route of transmission[a] | Anti-HIV | p24 antigen (pg/ml) | Time (wk) of first antibody-positive sample |
|---|---|---|---|---|
| p74 | P | Negative | 50 | 4 |
| p84 | P | Negative | 20 | 6 |
| Sc1 | S | Negative | 15 | 12 |
| Sc2 | S | Negative | 500 | 2 |
| Sc3 | S | Negative | 50 | 1 |
| p82 | P | Positive | 0 | NA[b] |

[a] P, parenteral; S, sexual contact.
[b] NA, not applicable.

analysis (1, 48). Nucleotide sequence diversity within each sample was estimated in three ways. First, variation was expressed as the number of different nucleotide sequences within a sample. The frequency of polymorphic sites per nucleotide sequenced ($P_n$) and the mean pairwise distance between nucleotide sequences within a sample ($d$; see Materials and Methods) were also estimated.

In the *env* gene, each preseroconversion sample was homogeneous in sequence in the V4 region (nucleotide sequences between positions 1165 and 1272 in the *env* gene are shown in Fig. 1A) and in V3 (positions 844 and 1029; Fig. 2A), yielding maximum estimates of $P_n$ of as little as <1/3,276 in the V4 region and <1/9,880 in V3 for the RNA population in p74 (Table 2). Taking all five of the preseroconversion samples together, we found no intrasample sequence variability in over 44,000 nucleotides sequenced. This low frequency of substitution contrasts with that found previously in samples from p82 taken 5 years later, in which there was extensive variability in both V3 and V4 regions (1, 19, 48). One typical postinfection sample is included in Table 2 for comparison. Several distinct nucleotide sequences were observed (8 different sequences of 23 obtained in V3 and 12 of 43 obtained in V4), and there was a high frequency of polymorphic sites ($P_n$ values of 1/186 to 1/127 for V3 and V4, respectively). Sequence variability was also observed in the plasma virus population in the V4 (but not V3) region immediately shortly after seroconversion from p82 (Table 2).

Samples from the six patients with primary infection yielded distinct V4 region sequences, varying considerably in both the composition and length of the amino acid sequence between the conserved flanking regions (Fig. 1). In common with sequences from long-term-infected individuals, all of the six sequences contained frequent (three to six) potential sites for N-linked glycosylation, none of which were conserved among all patients. However, in the V3 region, variability was far more restricted, being confined to nine amino acid polymorphisms in the 35 residues of the V3 loop, compared with 19 in the 27 residues of the flanking regions shown in Fig. 2. Three potential N-linked glycosylation sites (amino acid positions 295, 301, and 331) are conserved among all patients, while those at positions 289 and 338 outside the V3 loop are variable between the sequences analyzed here. Remarkably, two patients with epidemiologically unrelated infections (p82 [infected in early 1984 from contaminated factor VIII] and Sc3 [heterosexual contact, 1992]) who had distinct V4 sequences shared exactly the same amino acid (but not nucleotide) sequence in the V3 loop. Furthermore, these V3 loop sequences were identical to the global consensus sequence of North American and European isolates (23, 24). Other seroconversion

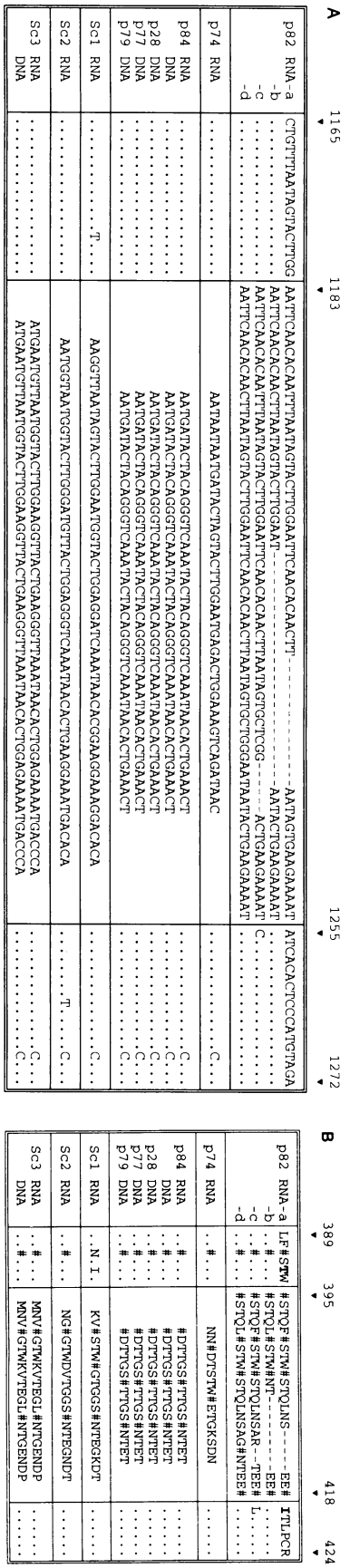



FIG. 1. Comparison of nucleotide (A) and inferred amino acid (B) sequences of the V4 region of viral RNA amplified from plasma preseroconversion or proviral DNA postseroconversion from individuals with primary HIV infection. The previously published sequences for p82 (49) are included for comparative purposes. Nucleotide and amino acid sequences are numbered from the start of the *env* gene of pHXB2r. Alignment of sequences in the V4 region was not attempted. Symbols: ., sequence identity with the p82 RNA-a sequence (flanking regions only); -, gap introduced to preserve alignment of sequences from p82; #, asparagine residue (N) forming a potential site for N-linked glycosylation (sites in flanking regions are shown irrespective of whether they are conserved or variable). Variable sites in flanking regions are indicated in bold.

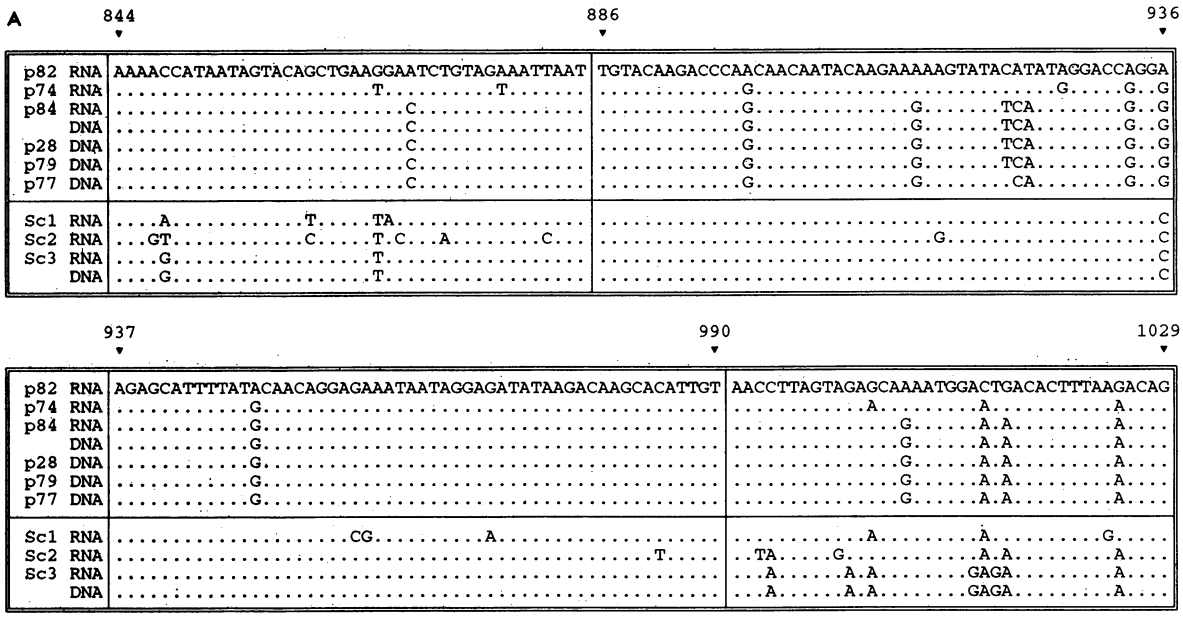



FIG. 2. Comparison of nucleotide (A) and inferred amino acid (B) sequences of viral RNA and proviral DNA in the V3 loop and flanking regions from individuals with primary HIV infection. The previously published sequence for p82 is included for comparative purposes only. Sequence numbering and symbols are as described in the legend to Fig. 1. Variable sites and nonconserved glycosylation sites are indicated in bold.

TABLE 2. Sequence diversity of viral RNA sequences during primary infection and on follow-up

| No. | Stage of infection[a] | V3 region | | | V4 region | | |
|---|---|---|---|---|---|---|---|
| | | Total no. of sequences obtained | No. of different sequences observed | $P_n$[b] (maximum) | Total no. of sequences obtained | No. of different sequences observed | $P_n$ (maximum) |
| p74 | Pr | 40 | 1 | <1/9,880 | 21 | 1 | <1/3,276 |
| p84 | Pr | 11 | 1 | <1/1,628 | 11 | 1 | <1/1,650 |
| Sc1 | Pr | 26 | 1 | <1/5,980 | 26 | 1 | <1/3,880 |
| Sc2 | Pr | 29 | 1 | <1/6,960 | 11 | 1 | <1/1,716 |
| Sc3 | Pr | 24 | 1 | <1/6,240 | 24 | 1 | <1/2,880 |
| p82 | Po | 12 | 1 | <1/2,880 | 8 | 3 | 1/126 |
| p82 | Fo | 23 | 8 | 1/186 | 43 | 12 | 1/127 |

[a] Pr, preseroconversion; Po, postseroconversion; Fo, follow-up (3 to 5 years).
[b] $P_n$, number of polymorphic sites/number of nucleotides sequenced. When sequences within a sample are uniform, only a maximum value can be calculated.

TABLE 3. Sequence diversity of provirus 3 to 6 months after primary infection with HIV-1

| No. | No. of copies/ $10^6$ cells[a] | V3 region | | | V4 region | | |
|---|---|---|---|---|---|---|---|
| | | Total no. of sequences obtained | No. of different sequences observed | Maximum $P_n$[b] | Total no. of sequences obtained | No. of different sequences observed | Maximum $P_n$ |
| p28 | 166 | 11 | 1 | <1/2,310 | 11 | 1 | <1/1,969 |
| p77 | 50 | 11 | 1 | <1/2,200 | 11 | 1 | <1/1,969 |
| p79 | 45 | 9 | 1 | <1/1,890 | 9 | 1 | <1/1,611 |
| p84 | 60 | 11 | 1 | <1/1,650 | 11 | 1 | <1/1,969 |
| Sc3 | 550 | 29 | 1 | <1/7,540 | 29 | 1 | <1/3,480 |

[a] Provirus in PBMC DNA quantified by limiting dilution (48).
[b] See Table 2, footnote *b*.

sequences are also similar in sequence to the global consensus sequences; those from Sc1 and Sc2 showed two differences, while those from p74 and p84 showed three and four, respectively. In contrast, sequences outside the V3 loop differed considerably between the seroconversion samples. Indeed, sequences of p82 and Sc3 that were identical in the V3 loop differed from each other at 8 of the 27 flanking region sites. Possible reasons for the similarity of amino acid sequences in the V3 loop but not in flanking regions are discussed below.

**Persistence of preseroconversion sequences in PMBCs.** We have previously found that seroconversion-type sequences can persist as a (minority) component of proviral sequences in PBMCs several years after primary infection (49). To investigate this phenomenon of persistence in more detail, we obtained samples collected after seroconversion from p84 (estimated interval from seroconversion, 15 weeks) and Sc3 (6 weeks). Neither sample contained any circulating RNA sequences detectable by PCR, but both contained high frequencies of provirus-bearing PBMCs (60 to 550 copies of provirus per $10^6$ cells; Table 3), comparable to those of long-term-infected asymptomatic individuals (48). Proviral sequences from both samples were completely homogeneous in V3 and V4 regions of *env* (combined maximum $P_n$ values of <1/11,380 and <1/3,619 for Sc3 and p84, respectively). Furthermore, in both cases, proviral nucleotide sequences were identical to those of the corresponding preseroconversion plasma RNA samples (Fig. 1 and 2), indicating that the two populations are for practical purposes equivalent. Similarly, the proviral sequences of postseroconversion samples from three more hemophiliacs (p28, p77, and p79) were homogeneous in the V3 and V4 regions (Table 3). Taking these five samples together, no intrasample sequence diversity was observed in the 26,588 proviral nucleotides analyzed, providing substantial evidence that whatever diversification may be taking place in the plasma or elsewhere in the infected individual, such changes are not represented in the PBMC population for many months (see Discussion).

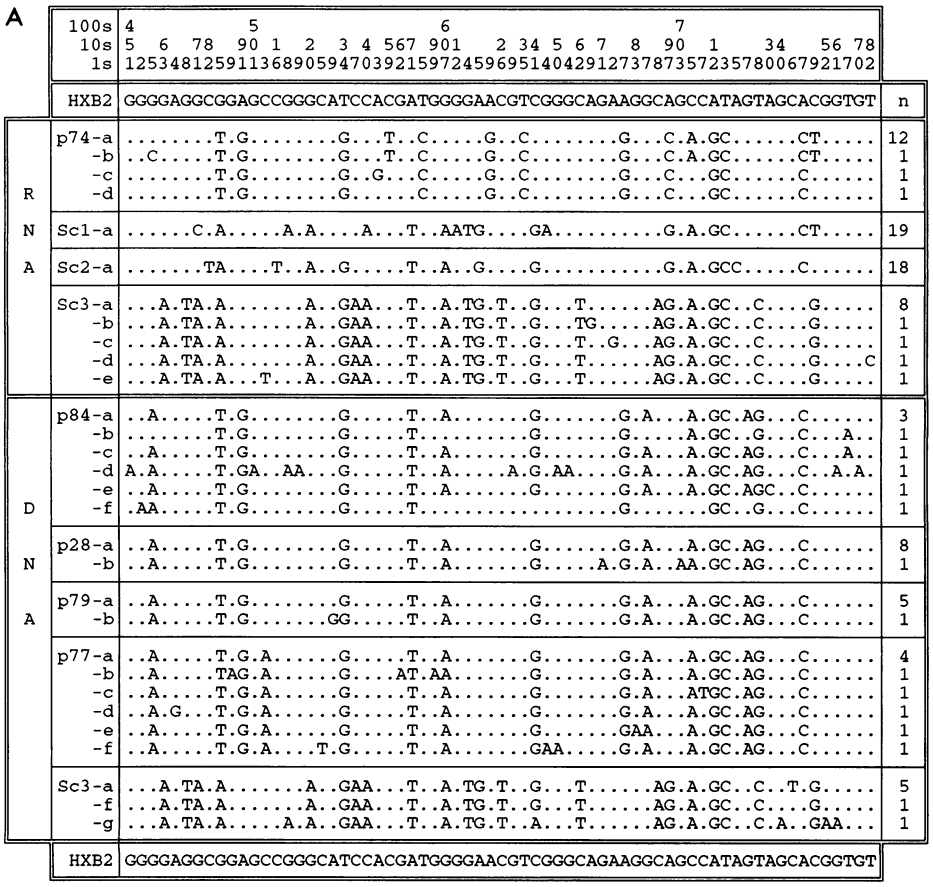
**Evidence for selection of specific *env* sequences before seroconversion.** The complete homogeneity in sequences in the *env* region could be explained in two different ways. First, it is possible that the initial virus population consisted of the progeny from a single infecting virus particle and that too few rounds of replication occurred between exposure and seroconversion for sequence diversity to develop. Alternatively, the lack of diversity in the *env* region could have resulted from selection for specific sequence variants before seroconversion which outcompeted other cotransmitted variants (see Discussion). We have carried out two investigations to distinguish between the two possibilities.

If the homogeneity of *env* sequences was the result of selection, then we might be able to observe variability in other regions of the genome not subject to the same selective pressures. To investigate this possibility, sequence analysis of part of $p17^{gag}$ gene was carried out on the preseroconversion RNA samples and postseroconversion DNA samples (Fig. 3), and the diversity was compared with that of the *env* sequences (Table 4). In contrast to V3 and V4, we found frequent nucleotide substitutions in seven of nine samples analyzed. Variability was found in both preseroconversion RNA and postseroconversion DNA samples and in cases of both sexual and parenteral transmission of HIV. Comparison of the combined results allows us to discount sampling as a possible explanation for the differences between the two regions. In the *env* region, a total of 190 sequences in the V3 and V4 regions combined yielded no intrasample variation ($P_n$ < 1/67,900), while analysis of 103 p17 sequences yielded 21 variant sequences and a frequency of intrasample variable sites of approximately 1/430. Indeed the range of variability in the *gag* region for the patients with primary infection ($P_n$ = 1/305 to 1/1,542) overlaps that of long-term-infected individuals (1/205 to 1/405; Table 4).

Taking all of the $p17^{gag}$ sequences together, only 34 of the 71 observed nucleotide differences led to an amino acid change, lower than that expected for randomly mutating DNA. Even lower proportions of nonsynonymous nucleotide changes are observed within individual populations; of 43 intrasample variable sites, only 15 led to amino acid substitutions. The ratio of synonymous ($K_s$) to nonsynonymous ($K_a$) substitutions in the preseroconversion samples was 3.2, similar to that found between isolates (25) and in two long-term-infected patients (ratios of 6.1 and 2.7 for s44 and p82, respectively). The frequency of G ↔ A transitions in the region analyzed was 52% (37 of 71).

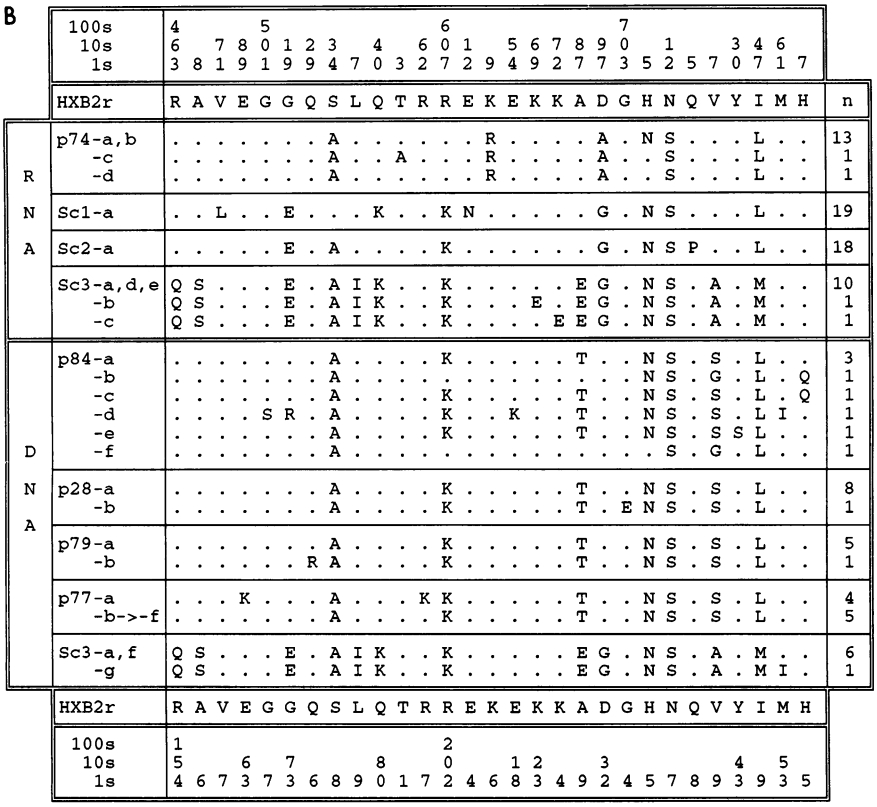
The second piece of evidence for specific selection of particular *env* sequences is provided by a comparison of V3 and V4 sequences at seroconversion of four hemophiliacs exposed to the same infectious source, a batch of HIV-contaminated factor VIII (27). In the absence of selection, seroconversion sequences in recipients should be a random selection of *env* variants present in factor VIII. However, comparison of the V3 and V4 sequences from p28, p77, p79, and p84 reveals far greater sequence similarity than would be expected from nonselective transmission (Table 5). All four hemophiliacs show identical sequences in V4 and flanking regions (Fig. 1), while three of the four hemophiliacs are identical in the V3 region (Fig. 2). The fourth, p79, shows a single nucleotide substitution in the V3 loop, leading to the replacement of proline for serine at position 308. This degree of variability in the *env* region is far lower than has ever been observed in viral RNA sequences of seropositive individuals

A



| | | 100s | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|
| | | 10s | 5 6 78 | 90 1 2 3 4 567 | 901 2 34 5 6 7 8 | 90 1 34 56 78 | |
| | | 1s | 12534812591136890594703921597245969514042912737873572357800679217 02 | | | | |
| | HXB2 | | GGGGAGGCGGAGCCGGGCATCCACGATGGGGAACGTCGGGCAGAAGGCAGCCATAGTAGCACGGTGT | | | | n |
| R | p74-a | | ........T.G........G...T..C.....G..C........G...C.A.GC......CT..... | | | | 12 |
| | -b | | ..C.....T.G.......G...T..C.....G..C.....TG.....G...C.A.GC......CT..... | | | | 1 |
| | -c | | ........T.G........G..G...C.....G..C........G...C...GC......C..... | | | | 1 |
| | -d | | ........T.G........G......C.....G..C........G...C...GC......C..... | | | | 1 |
| N | Sc1-a | | ......C.A.....A.A....A...T..AATG....GA..........G.A.GC......CT..... | | | | 19 |
| A | Sc2-a | | .......TA....T..A..G.....T..A..G....G...........G.A.GCC.....C..... | | | | 18 |
| | Sc3-a | | ...A.TA.A.......A..GAA...T..A.TG.T..G...T......AG.A.GC..C....G..... | | | | 8 |
| | -b | | ...A.TA.A.......A..GAA...T..A.TG.T..G...TG.....AG.A.GC..C....G..... | | | | 1 |
| | -c | | ...A.TA.A.......A..GAA...T..A.TG.T..G...T..G...AG.A.GC..C....G..... | | | | 1 |
| | -d | | ...A.TA.A.......A..GAA...T..A.TG.T..G...T......AG.A.GC..C....G....C | | | | 1 |
| | -e | | ...A.TA.A...T...A..GAA...T..A.TG.T..G...T......AG.A.GC..C....G..... | | | | 1 |
| D | p84-a | | ..A.....T.G........G.....T..A.......G.......G.A...A.GC.AG...C..... | | | | 3 |
| | -b | | ........T.G........G.....T..........G.......G.....A.GC..G...C...A.. | | | | 1 |
| | -c | | ..A.....T.G........G.....T..A.......G.......G.A...A.GC.AG...C...A.. | | | | 1 |
| | -d | | A.A.....T.GA..AA...G.....T..A.....A.G.AA....G.A...A.GC.AG...C..A.A. | | | | 1 |
| | -e | | ..A.....T.G........G.....T..A.......G.......G.A...A.GC.AGC..C..... | | | | 1 |
| | -f | | .AA.....T.G........G.....T..................G.......GC..G...C..... | | | | 1 |
| N | p28-a | | ..A.....T.G........G.....T..A.......G.......G.A...A.GC.AG...C..... | | | | 8 |
| | -b | | ..A.....T.G........G.....T..A.......G.....A.G.A..AA.GC.AG...C..... | | | | 1 |
| A | p79-a | | ..A.....T.G........G.....T..A.......G.......G.A...A.GC.AG...C..... | | | | 5 |
| | -b | | ..A.....T.G.......GG.....T..A.......G.......G.A...A.GC.AG...C..... | | | | 1 |
| | p77-a | | ..A.....T.G.A......G.....T..A.......G.......G.A...A.GC.AG...C..... | | | | 4 |
| | -b | | ..A.....TAG.A......G....AT.AA.......G.......G.A...A.GC.AG...C..... | | | | 1 |
| | -c | | ..A.....T.G.A......G.....T..A.......G.......G.A...ATGC.AG...C..... | | | | 1 |
| | -d | | ..A.G...T.G.A......G.....T..A.......G.......G.A...A.GC.AG...C..... | | | | 1 |
| | -e | | ..A.....T.G.A......G.....T..A.......G.......GAA...A.GC.AG...C..... | | | | 1 |
| | -f | | ..A.....T.G.A....T.G.....T..A.......GAA.....G.A...A.GC.AG...C..... | | | | 1 |
| | Sc3-a | | ...A.TA.A.......A..GAA...T..A.TG.T..G...T......AG.A.GC..C..T.G..... | | | | 5 |
| | -f | | ...A.TA.A.......A..GAA...T..A.TG.T..G...T......AG.A.GC..C....G..... | | | | 1 |
| | -g | | ...A.TA.A.....A.A..GAA...T..A.TG.T..A...T......AG.A.GC..C.A..GAA... | | | | 1 |
| | HXB2 | | GGGGAGGCGGAGCCGGGCATCCACGATGGGGAACGTCGGGCAGAAGGCAGCCATAGTAGCACGGTGT | | | | |

B



| | | 100s | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|
| | | 10s | 6 7 8 0 1 2 3 4 | 6 0 1 5 6 7 8 9 0 | 1 3 4 6 | | |
| | | 1s | 3 8 1 9 1 9 9 4 7 0 3 2 7 2 9 4 9 2 7 7 3 5 2 5 7 0 7 1 7 | | | | |
| | HXB2r | | R A V E G G Q S L Q T R R E K E K K A D G H N Q V Y I M H | | | | n |
| R | p74-a,b | | . . . . . A . . . . . R . . . . A . N S . . . L . . | | | | 13 |
| | -c | | . . . . . A . . A . . . R . . . . A . . S . . . L . . | | | | 1 |
| | -d | | . . . . . A . . . . . R . . . . A . . S . . . L . . | | | | 1 |
| N | Sc1-a | | . . L . . E . . . K . . K N . . . . . G . N S . . . L . . | | | | 19 |
| A | Sc2-a | | . . . . . E . A . . . . K . . . . . G . N S P . . L . . | | | | 18 |
| | Sc3-a,d,e | | Q S . . . E . A I K . . K . . . . . E G . N S . A . M . . | | | | 10 |
| | -b | | Q S . . . E . A I K . . K . . . E . E G . N S . A . M . . | | | | 1 |
| | -c | | Q S . . . E . A I K . . K . . . . E E G . N S . A . M . . | | | | 1 |
| D | p84-a | | . . . . . A . . . . K . . . . . T . . N S . S . L . . | | | | 3 |
| | -b | | . . . . . A . . . . . N S . G . L . Q | | | | 1 |
| | -c | | . . . . . A . . . . K . . . . . T . . N S . S . L . Q | | | | 1 |
| | -d | | . . . . S R . A . . . . K . . K . . T . . N S . S . L I . | | | | 1 |
| | -e | | . . . . . A . . . . K . . . . . T . . N S . S S L . . | | | | 1 |
| | -f | | . . . . . A . . . . . S . G . L . . | | | | 1 |
| N | p28-a | | . . . . . A . . . . K . . . . . T . . N S . S . L . . | | | | 8 |
| | -b | | . . . . . A . . . . K . . . . . T . E N S . S . L . . | | | | 1 |
| A | p79-a | | . . . . . A . . . . K . . . . . T . . N S . S . L . . | | | | 5 |
| | -b | | . . . . . R A . . . . K . . . . . T . . N S . S . L . . | | | | 1 |
| | p77-a | | . . . K . . . A . . . K K . . . . . T . . N S . S . L . . | | | | 4 |
| | -b->-f | | . . . . . A . . . . K . . . . . T . . N S . S . L . . | | | | 5 |
| | Sc3-a,f | | Q S . . . E . A I K . . K . . . . . E G . N S . A . M . . | | | | 6 |
| | -g | | Q S . . . E . A I K . . K . . . . . E G . N S . A . M I . | | | | 1 |
| | HXB2r | | R A V E G G Q S L Q T R R E K E K K A D G H N Q V Y I M H | | | | |
| | 100s | | 1 | 2 | | | |
| | 10s | | 5 6 7 8 | 0 1 2 3 | 4 5 | | |
| | 1s | | 4 6 7 3 7 3 6 8 9 0 1 7 2 4 6 8 3 4 9 2 4 5 7 8 9 3 9 3 5 | | | | |

TABLE 4. Comparison of *env* and *gag* sequence diversity during primary infection

| No. | Sample[a] | Stage[b] | V3/V4 region | | | | $p17^{gag}$ region | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Total no. of sequences obtained | No. of different sequences observed | $P_n$[c] (maximum) | $d$ (%)[d] | Total no. of sequences obtained | No. of different sequences observed | $P_n$ (maximum) | $d$ (%) |
| p74 | RNA | Pr | 40 | 1 | <1/13,156 | 0.0 | 15 | 4 | 1/1,029 | 0.3 |
| Sc1 | RNA | Pr | 26 | 1 | <1/9,860 | 0.0 | 19 | 1 | <1/6,517 | 0.0 |
| Sc2 | RNA | Pr | 29 | 1 | <1/8,676 | 0.0 | 18 | 1 | <1/6,174 | 0.0 |
| Sc3 | RNA | Pr | 24 | 1 | <1/9,120 | 0.0 | 12 | 5 | 1/1,029 | 0.2 |
| p28 | DNA | Po | 11 | 1 | <1/4,279 | 0.0 | 9 | 2 | 1/1,542 | 0.1 |
| p77 | DNA | Po | 11 | 1 | <1/4,169 | 0.0 | 9 | 6 | 1/343 | 0.6 |
| p79 | DNA | Po | 9 | 1 | <1/3,501 | 0.0 | 6 | 2 | 1/1,029 | 0.1 |
| p84 | DNA | Po | 11 | 1 | <1/3,619 | 0.0 | 8 | 6 | 1/305 | 0.9 |
| Sc3 | DNA | Po | 29 | 1 | <1/11,020 | 0.0 | 7 | 3 | 1/400 | 0.5 |
| p82 | RNA | Fo | 43 | 18 | 1/160 | 2.5 | 14 | 14 | 1/205 | 1.7 |
| s44 | RNA | Fo | | | | | 13 | 10 | 1/405 | 1.0 |

[a] RNA, viral RNA sequences in plasma preseroconversion; DNA, proviral sequences from PBMCs postseroconversion.
[b] See Table 2, footnote *a*.
[c] See Table 2, footnote *b*.
[d] *d* (%), mean percentage nucleotide distance between sequences within a sample.

and many orders of magnitude lower than that of the follow-up sample from p82 (Table 2). It is therefore likely that only the subset of virus variants within factor VIII bearing the observed V3 and V4 sequences at seroconversion were capable of virus transmission. In contrast to the envelope region, high degrees of sequence variability were observed in *gag*; inter- and intrapatient variability was comparable to that of long-term-infected individuals, consistent with our observations of other patients in this study.

## DISCUSSION

**Homogeneous *env* sequences at seroconversion.** In this study, we have carried out extensive sequence analysis of two normally hypervariable regions of the *env* gene (V3 and V4) to characterize the initial virus population associated with primary infection by HIV. A striking feature of the data obtained was the absence of any sequence heterogeneity in these regions in any of the nine acutely infected individuals. Among viral RNA sequences obtained before the first appearance of anti-HIV antibody, we found no variability in a total of 44,090 nucleotides sequenced, while among the postseroconversion proviral DNA sequences, none was observed in 26,588 nucleotides.

The detection of such low rates of sequence variability within the initial virus populations was made possible by the use of a method for sequencing PCR-amplified DNA (47, 48) that avoids the problems of copying errors by *Taq* polymerase, in vitro recombination, and selective amplification that other methods inevitably encounter (32, 33). In contrast to our results, other studies have found far greater sequence heterogeneity of viral populations around the time of primary infection (37, 60). In the former study, substitutions, frameshifts, and stop codons were found scattered throughout V4 but also in the normally well conserved C4 region in proviral sequences taken within weeks of seroconversion. In the latter, some of the postseroconversion sequences obtained from amplification of plasma virus showed substitutions in the normally invariant cysteine residues spanning the V3 loop. In both studies, nucleotide sequences were obtained by cloning PCR products and are therefore subject to potentially high rates of in vitro copying artifacts. The frequency of copying errors associated with this method ranges from as high as 1/350 in one study (12) to 1/1,800 (13). These rates are considerably higher than the heterogeneity in the in vivo virus populations that we found in the *env* gene. As neither of the two previous studies of primary infection by HIV carried out control experiments to determine the error rate associated with cloning, we cannot rule out the possibility that some or all of the sequence variability observed in proviral sequences shortly after seroconversion (37) and a proportion of the amino acid substitutions in the V3 region (60) originated from in vitro copying artifacts rather than representing genuine viral sequence heterogeneity.

**Evidence for selection in primary infection.** A monotypic population of *env* sequences on primary infection could arise in three ways. First, each primary infection with HIV may be the result of infection from a single viable virus particle. The uniformity of the *env* sequences would then simply reflect the fact that there were insufficient numbers of replicative cycles to allow measurable diversification of the initial virus population. Alternatively, in the case of transmission and coinfection with several different viruses, one virus might outgrow the others by chance and lead to the observed uniform population. In neither of these models is it necessary to invoke specific selection to explain the observed reduction or absence of sequence variability of *env* sequences found in this or previous studies of sexual transmission (60). Similarly, the observed reduction of sequence variability of the *env* gene found upon mother-to-child trans-

FIG. 3. Comparison of variable sites of nucleotide (A) and inferred amino acid (B) sequences within part of the $p17^{gag}$ gene (343 bp) from individuals with primary HIV infection with the sequence of pHXB2r. Nucleotide positions (numbered from the start of the pHXB2r sequence) of polymorphic sites are shown at top of each panel; variable amino residues are indicated at the bottom of panel B. The number of observed occurrences of each sequence variant (n) is recorded at right of each panel.

TABLE 5. Combined diversity of *env* and *gag* sequences in four individuals exposed to the same source of infection

| Region | Total no. of sequences obtained[a] | No. of different sequences observed | $P_n$[b] (maximum) | $d$ (%)[c] |
|---|---|---|---|---|
| V3 | 42 | 2 | 1/4,025 | 0.02 |
| V4 | 42 | 1 | <1/7,518 | 0.00 |
| p17$^{gag}$ | 32 | 14 | 1/499 | 0.60 |

[a] Combined analysis of sequences from four hemophiliacs (p28, p77, p79, and p84) exposed to the same source of HIV infection analyzed together.
[b] See Table 2, footnote *b*.
[c] See Table 4, footnote *d*.

mission (61) clearly does not in itself provide any evidence for selective transmission of particular variants of HIV.

However, neither of these simple explanations can explain why there should be at least 100-fold-greater sequence diversity in the *gag* region in the samples from seven of the nine individuals with primary infection. There is no a priori reason to suppose that the mutation rate of the *gag* region should be so much higher than that of the *env* region, particularly when sequence variability in the *gag* region is generally much less than that in the *env* region in long-term-infected individuals. Furthermore, the observation that two patients showed no variability in p17$^{gag}$ (combined $P_n$ of <1/12,691) also shows that any such diversification process in *gag* does not necessarily always occur.

Sequence change to prevent binding to major histocompatibility complex class I molecules has been reported to occur in the p17 region of *gag* (38), and it possible in principle that the heterogeneity of *gag* sequences observed at seroconversion was the result of positive selection to evade immunological recognition. However, the majority of sequence changes in the *gag* region were silent, in that they did not affect the encoded amino acid sequence (Fig. 3), and therefore could have no conceivable effect on class I or T-cell receptor binding or on viral phenotype. As we cannot account for the diversity of *gag* sequences in terms of the simple models described above, we infer that the contrasting homogeneity of *env* sequences is the result of strong selection before seroconversion, and we propose that only a small proportion of *env* sequence variants are capable of causing primary infection.

From the data that we have presented, it is not possible to conclude where the process of population expansion and generation of sequence diversity in the *gag* region occurs. One possibility is that transmission of a single virus particle was responsible for primary infection and that extremely rapid diversification leads to the observed variability in the *gag* region upon primary infection, while selection for a specific determinant close to V3 or V4 eliminates detectable sequence heterogeneity in this region in the initial virus population.

An alternative possibility is that the exposed individual is nonselectively infected with several viruses bearing a range of *gag* and *env* sequences. Specific selection for viruses bearing the putative *env* determinant then takes place in the contact in the interval between exposure and seroconversion. The observed variability of *gag* sequences upon acute primary infection arises by recombination between *env* and *gag* sequences. Such recombination could take place within the acutely infected individual between exposure and seroconversion; alternatively, recombination may have already occurred in the individual who was the source of infection, and there was cotransmission of a series of hybrid viruses to the contact. In this model, it is likely that the predominant sequence in the *gag* region is the one originally associated with the selected *env* sequences, while the single occurrences of the alternative *gag* variants represent those of viruses generated by recombination. The frequency with which this process is observable varies considerably between individuals; in p84 and p77, 5 of 8 and 5 of 9 viral sequences may be such recombinants, while none were observed in 18 to 19 *gag* sequences from Sc1 or Sc2, perhaps because of less frequent recombination. However, it is also possible that the homogeneous *gag* sequences in Sc1 and Sc2 at seroconversion arose from greater sequence similarity in the transmitted variants or from the transmission of relatively few or even a single virus particle.

Although we cannot distinguish between these or other possible series of preseroconversion events, they all imply strong selection for specific determinants close to V3 or V4 upon primary infection. This conclusion is supported by our observation that recipients of the same batch of factor VIII seroconverted with identical V4 and almost identical V3 sequences following exposure to what was probably a heterogeneous mixture of *env* sequence variants such as is found in plasma of long-term-infected individuals (Table 4).

**Localization of the determinant selected upon transmission.** In this study, we observed no sequence variability in either the V3 or V4 region of *env*. However, this observation does not in itself localize the actual determinant of selection, as strong selection tends to eliminate variation at linked neutral sites. Thus, we cannot conclude from these data alone that there is selection for specific V3 or V4 sequences; all we can say is that the determinant that is selected is much closer to V3 and V4 than it is to p17$^{gag}$, as the frequency of recombination between two sites is likely to be proportional to their physical separation in the viral genome.

Comparison of primary amino acid sequences reveals no similarity in the V4 or flanking regions between the patients with epidemiologically unrelated infections that would localize such a determinant to this region of the *env* gene. Indeed, the V4 sequences are so different from each other that simply attempting an alignment of the sequences is difficult to justify (Fig. 1). Apart from the homogeneity within each sample, they differ little in composition from those found in long-term-infected individuals; potential N-linked glycosylation sites are frequently found in all six patients, as is sequence duplication and deletion (47).

On the other hand, variability in the V3 region appears to be restricted in the initial virus population of different patients. Despite being epidemiologically unrelated, the amino acid, but not the nucleotide, sequences of the V3 loops from variants infecting p82 and Sc3 were identical. Variability was also highly restricted among viruses characterized from the other four epidemiologically unrelated transmissions, differing from each other by a maximum of 6 from 35 amino acid residues. It is also significant that the two identical sequences from p82 and Sc3 were identical to the global consensus sequence of North American and European sequences (23, 24), providing some evidence that virus variants in most infected individuals derive from this specific sequence. The restriction in sequence variability observed in the V3 loop does not extend to the flanking regions of V3. Most revealingly, the sequences in the flanking regions of variants from p82 and Sc3 differ substantially from each other, hinting at a very sharp boundary between the region that is subject to the putative selection process and regions

outside the loop where amino acid substitutions appear not to be constrained.

The outcome of the selection process is that the initial virus population is essentially clonal in origin in the selected region (*env*) but polyclonal and hence variable elsewhere. The apparent requirement for a specific *env* determinant upon primary infection may account for the stochastic nature of virus transmission. For example, only 18 of the 32 recipients of an HIV-contaminated batch of factor VIII actually became infected (27), and the risk of infection correlated with the number of bottles transfused, implying that fully infectious virions were relatively rare in the factor VIII. This is despite the fact that each of the 32 hemophiliacs receiving the batch was probably exposed to viable virus if the model of multiple transmission outlined above is correct. The absence of infection in half of the cohort can thus be explained on the basis that none of the transmitted variants in these patients had *env* sequences to permit seroconversion. The fate of transmitted viruses in such cases of nonproductive infection is unclear. However, it should be noted that transient T-cell proliferative responses have been detected several months before seroconversion in individuals exposed to HIV (7), suggesting substantial virus replication before the onset of primary infection. Furthermore, evidence for cell-mediated immunity has been found in individuals who are at risk for infection by HIV from unprotected sexual intercourse but who remain seronegative (8). In such individuals, it is possible that there is indeed transmission and limited viral replication of HIV, but that the infection does not necessarily progress to systemic spread and seroconversion.

**Phenotype of the selected *env* sequences.** It is now reasonably clear that there are differences in cytopathology and cell tropism between isolates obtained during the early asymptomatic stage of infection and those from patients with AIDS or AIDS-related complex (3, 15, 51). Isolates from early stages of infection generally grow more slowly, do not induce syncytia between infected cells, and retain a capability to replicate in primary macrophages as well as in activated $CD4^+$ lymphocytes. Later in infection, variants that are highly cytopathic upon virus culture (SI) and have lost their ability to grow in macrophages may appear (45). Consistent with this time course, it has been found that HIV isolates collected at the time of primary infection are generally NSI and macrophage tropic (6, 9, 40). However, in one of the studies (40), 3 of the 19 virus isolates collected before seroconversion were SI and non-macrophage tropic. The detection of this more virulent phenotype at seroconversion correlated with a subsequent more rapid progression to AIDS than observed in those infected with NSI variants (40).

The status of the V3 loop as a determinant of biological phenotype is uncertain. For certain virus isolates, in vitro recombination experiments have implicated regions including V3 as necessary and sufficient determinants of macrophage tropism and cytopathology (4, 5, 15, 16, 20, 36, 46, 50, 55, 56). Comparison of V3 sequences of SI and NSI virus isolates indicated a significant association between the presence of acidic amino acid residues at position 306 and/or 320 and the SI phenotype (16); these predictions have since been experimentally verified by site-directed mutagenesis of an HXB2-like isolate (10). Although the properties of the determinant in *env* selected upon virus transmission cannot be established with any certainty, it could be predicted that virus isolates from p82, p84, p74, Sc2, and Sc3 should be NSI, while Sc1 should be SI.

In another study, V3 sequences of noncytopathic and macrophage-tropic variants showed substantially restricted sequence variability compared with the sequences of T-cell-tropic variants and were also similar or identical to the global consensus of 245 North American virus isolates (5). There is a striking similarity between the overall sequences of preseroconversion sequences reported in this study with macrophage-tropic variants, in terms of both the positions where substitutions occurred and the actual amino acid residues substituted. For example, substitution of lysine by arginine and of threonine by alanine occurred at positions 305 and 317, respectively, in two patients in our data set and frequently among macrophage-tropic variants (5, 16).

We and others have described the slow diversification of PBMC sequences over several years of infection (30, 49). In both studies, variants found early in infection had V3 sequences similar to the macrophage-tropic/global consensus, but these were replaced by a range of sequence variants with predicted SI phenotypes (16). These observation support the hypothesis that the determinant associated with primary infection is indeed macrophage tropism or an associated property, although it is intriguing that this appears to be the case irrespective of whether infection was acquired by sexual contact or parenterally.

**Virological events during primary infection and seroconversion.** We have previously found that V3, V4, and V5 sequences found at seroconversion may persist unchanged as a increasingly small component of the proviral population over the years following primary infection (49). In contrast, such sequences disappear completely from the plasma virus population immediately after seroconversion. From other studies, it has been found that plasma RNA sequences in early stages of infection (60) are orders of magnitude more heterogeneous than RNA sequences analyzed preseroconversion and the population of proviral DNA sequences found in PBMCs several months after seroconversion in this study.

The persistence of sequences in PBMCs identical to those of plasma virus preseroconversion is therefore not necessarily associated with their continued active expression. Indeed, we found that the frequencies of provirus-bearing cells shortly after seroconversion were typical of those of asymptomatic, HIV-seropositive individuals in the clinically silent stage of HIV infection (39, 49), even though plasma virus had fallen to low or undetectable levels. Furthermore, the V3 and V4 sequences of the DNA samples taken shortly after infection from p84 and Sc3 were homogeneous and identical to those found in each corresponding preseroconversion RNA sample (Fig. 1 and 2), while *gag* sequences showed equivalent degrees of variability (Table 4).

It is possible that the provirus-bearing cells detected after seroconversion represent cells nonproductively infected during primary infection, either through integration of provirus at a site in a chromosome unfavorable for RNA transcription or from infection with a mutated virus incapable of cytopathic virus replication. Although we did not identify the cell type bearing provirus in the immediate postseroconversion period, it is significant that the most frequently infected cell type in both HIV and simian immunodeficiency virus infection is the $CD45RO^+$ memory cell (43, 57). Indeed, it is possible that those naive $CD4^+$ T cells that become activated are the main targets HIV during the period of acute viremia that precedes seroconversion. The overwhelming majority of such cells would probably have been destroyed at this stage, as shown by the transient drop in total $CD4^+$ counts and the extremely high numbers of HIV-infected cells during primary infection (9, 22). However, those that do survive proceed to differentiate into

memory cells, each bearing a single copy of an inactive integrated provirus and an unaltered phenotype. These cells or their descendants would continue to circulate within the seropositive individual for several years. Therefore, in the very early stages of HIV infection, the only DNA sequences obtainable by PCR are simply copies of the initial virus population. The fact that such viruses are those that failed to productively infect the host cell may explain the well-known difficulty in isolating HIV from some individuals early in infection, even when substantial numbers of PBMCs contain proviral DNA (2, 42, 53).

## ACKNOWLEDGMENTS

We are grateful to the following for supplying samples for sequence analysis: C. A. Ludlam and staff at the Haemophilia Centre, Department of Haematology, Royal Infirmary of Edinburgh; R. Robertson and staff at the Muirhouse Surgery, Edinburgh; A. J. France and G. E. D. Urquhart and staff at the Department of Medical Microbiology, Ninewells Hospital, Dundee; and staff at the Regional Infectious Disease Unit, City Hospital, Edinburgh. We also acknowledge the help provided by Pamela Barr, Selma Rebus, and other laboratory staff in the Department of Medical Microbiology, University of Edinburgh, and staff in the MRC Human Genetics Unit, Western General Hospital, Edinburgh, for the separation, storage, and serological testing (anti-HIV, p24 antigen) of the samples. Finally, we are grateful to Hanneka Schuitemaker and Frank Mediema for valuable discussion of the results.

The work was supported by the Medical Research Council AIDS Directed Program, grant G9025030.

## REFERENCES

1. **Balfe, P., P. Simmonds, C. A. Ludlam, J. O. Bishop, and A. J. Leigh Brown.** 1990. Concurrent evolution of human immunodeficiency virus type 1 in patients infected from the same source: rate of sequence change and low frequency of inactivating mutations. J. Virol. **64**:6221–6233.
2. **Bieniasz, P. D., K. Ariyoshi, and J. N. Weber.** 1992. Comparison of different quantitative techniques for the assessment of HIV viraemia. Abstr. VIII Int. Conf. AIDS **1**:Tu38.
3. **Cheng-Mayer, C., D. Seto, M. Tateno, and J. A. Levy.** 1988. Biological features of HIV-1 that correlate with virulence in the host. Science **240**:80–82.
4. **Chesebro, B., J. Nishio, S. Perryman, A. Cann, W. OBrien, I. S. Chen, and K. Wehrly.** 1991. Identification of human immunodeficiency virus envelope gene sequences influencing viral entry into CD4-positive HeLa cells, T-leukemia cells, and macrophages. J. Virol. **65**:5782–5789.
5. **Chesebro, B., K. Wehrly, J. Nishio, and S. Perryman.** 1992. Macrophage-tropic human immunodeficiency virus isolates from different patients exhibit unusual V3 envelope sequence homogeneity in comparison with T-cell-tropic isolates: definition of critical amino acids involved in cell tropism. J. Virol. **66**:6547–6554.
6. **Cichutek, K., S. Norley, R. Linde, W. Kreuz, M. Gahr, J. Lower, G. von Wangenheim, and R. Kurth.** 1991. Lack of HIV-1 V3 region sequence diversity in two haemophiliac patients infected with a putative biologic clone of HIV-1. AIDS **5**:1185–1187.
7. **Clerici, M., J. Berzofsky, G. M. Shearer, and C. O. Tackett.** 1991. Exposure to human immunodeficiency virus type 1-specific T helper cell responses before detection of infection by polymerase chain reaction and serum antibodies. J. Infect. Dis. **164**:178–182.
8. **Clerici, M., J. V. Giorgi, C. C. Chou, V. K. Gudeman, J. A. Zack, P. Gupta, H. N. Ho, J. A. Berzofsky, and G. M. Shearer.** 1992. Cell-mediated immune response to human immunodeficiency virus (HIV) type 1 in seronegative homosexual men with recent sexual exposure to HIV-1. J. Infect. Dis. **165**:1012–1019.
9. **Daar, E. S., T. Moudgil, R. D. Meyer, and D. D. Ho.** 1991. Transient high levels of viremia in patients with primary human immunodeficiency virus type 1 infection. N. Engl. J. Med. **324**:961–964.
10. **de Jong, J. J., A. de Ronde, W. Keulen, M. Tersmette, and J. Goudsmit.** 1992. Minimal requirements for the human immunodeficiency virus type 1 V3 domain to support the syncytium-inducing phenotype: analysis by single amino acid substitution. J. Virol. **66**:6777–6780.
11. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–395.
12. **Dunning, A. M., P. Talmud, and S. E. Humphries.** 1988. Errors in the polymerase chain reaction. Nucleic Acids Res. **16**:10393.
13. **Ennis, P. D., J. Zemmour, R. D. Salter, and P. Parham.** 1990. Rapid cloning of HLA-A,B cDNA by using the polymerase chain reaction: frequency and nature of errors produced in amplification. Proc. Natl. Acad. Sci. USA **87**:2833–2837.
14. **Felsenstein, J.** 1989. PHYLIP-phylogeny inference package (version 3.2). Cladistics **5**:164–166.
15. **Fisher, A. G., B. Ensoli, D. Looney, A. Rose, R. C. Gallo, M. S. Saag, G. M. Shaw, B. H. Hahn, and F. Wong Staal.** 1988. Biologically diverse molecular variants within a single HIV-1 isolate. Nature (London) **334**:444–447.
16. **Fouchier, R. A. M., M. Groenink, N. A. Kootstra, M. Tersmette, H. G. Huisman, F. Miedema, and H. Schuitemaker.** 1992. Phenotype-associated sequence variation in the third variable domain of the human immunodeficiency virus type 1 gp120 molecule. J. Virol. **66**:3138–3187.
17. **Fung, M. S., C. R. Sun, W. L. Gordon, R. S. Liou, T. W. Chang, W. N. Sun, E. S. Daar, and D. D. Ho.** 1992. Identification and characterization of a neutralization site within the second variable region of human immunodeficiency virus type 1 gp120. J. Virol. **66**:848–856.
18. **Goudsmit, J., C. Debouck, R. H. Meloen, L. Smit, M. Bakker, D. M. Asher, A. V. Wolff, C. J. Gibbs, Jr., and D. C. Gajdusek.** 1988. Human immunodeficiency virus type 1 neutralization epitope with conserved architecture elicits early type-specific antibodies in experimentally infected chimpanzees. Proc. Natl. Acad. Sci. USA **85**:4478–4482.
19. **Holmes, E. C., L. Q. Zhang, P. Simmonds, C. A. Ludlam, and A. J. Leigh Brown.** 1992. Convergent and divergent sequence evolution in the surface envelope glycoprotein of human immunodeficiency virus type 1 within a single infected patient. Proc. Natl. Acad. Sci. USA **89**:4835–4839.
20. **Hwang, S. S., T. J. Boyle, H. K. Lyerly, and B. R. Cullen.** 1991. Identification of the envelope V3 loop as the primary determinant of cell tropism in HIV-1. Science **253**:71–74.
21. **Javaherian, K., A. J. Langlois, C. McDanal, K. L. Ross, L. I. Eckler, C. L. Jellis, A. T. Profy, J. R. Rusche, D. P. Bolognesi, and S. D. Putney.** 1989. Principal neutralizing domain of the human immunodeficiency virus type 1 envelope protein. Proc. Natl. Acad. Sci. USA **86**:6768–6772.
22. **Jurriaans, S., J. T. Dekker, and A. de Ronde.** 1992. HIV-1 viral DNA load in peripheral blood mononuclear cells from seroconverters and long-term infected individuals. AIDS **6**:635–641.
23. **LaRosa, G. J., J. P. Davide, K. Weinhold, J. A. Waterbury, A. T. Profy, J. A. Lewis, A. J. Langlois, G. R. Dreesman, R. N. Boswell, and P. Shadduck.** 1990. Conserved sequence and structural elements in the HIV-1 principal neutralizing determinant. Science **249**:932–935.
24. **LaRosa, G. J., K. Weinhold, A. T. Profy, A. J. Langlois, G. R. Dreesman, R. N. Boswell, P. Shadduck, D. P. Bolognesi, T. J. Matthews, and E. A. Emini.** 1991. Conserved sequence and structural elements in the HIV-1 principal neutralizing determinant: further clarifications. Science **253**:1146.
25. **Leigh Brown, A., and P. Monaghan.** 1988. Evolution of the structural proteins of human immunodeficiency virus: selective constraints on nucleotide substitution. AIDS Res. Hum. Retroviruses **4**:399–407.
26. **Looney, D. J., A. G. Fisher, S. D. Putney, J. R. Rusche, R. R. Redfield, S. D. Burke, R. C. Gallo, and F. Wong Staal.** 1988. Type restricted neutralization of molecular clones of HIV. Science **241**:357–359.

27. **Ludlam, C. A., J. Tucker, C. M. Steel, R. S. Tedder, R. Cheingsong Popov, R. A. Weiss, D. B. M. McClelland, I. Philp, and R. J. Prescott.** 1985. Human T-lymphotropic virus type III (HTLV-III) infection in seronegative haemophiliacs after transfusion of factor VIII. Lancet **ii**:233–236.
28. **Martins, L. P., N. Chenciner, B. Asjo, A. Meyerhans, and S. Wain Hobson.** 1991. Independent fluctuation of human immunodeficiency virus type 1 *rev* and gp41 quasispecies in vivo. J. Virol. **65**:4502–4507.
29. **Masuda, T., S. Matsushita, M. J. Kuroda, M. Kannagi, K. Takatsuki, and S. Harada.** 1990. Generation of neutralization-resistant HIV-1 in vitro due to amino acid interchanges of third hypervariable env region. J. Immunol. **145**:3240–3246.
30. **McNearney, T., Z. Hornickova, R. Markham, A. Birdwell, M. Arens, A. Saah, and L. Ratner.** 1992. Relationship of human immunodeficiency virus type 1 sequence heterogeneity to stage of disease. Proc. Natl. Acad. Sci. USA **89**:10247–10251.
31. **Meloen, R. H., R. M. Liskamp, and J. Goudsmit.** 1989. Specificity and function of the individual amino acids of an important determinant of human immunodeficiency virus type 1 that induces neutralizing activity. J. Gen. Virol. **70**:1505–1512.
32. **Meyerhans, A., R. Cheynier, J. Albert, M. Seth, S. Kwok, J. Sninsky, L. Morfeldt Manson, B. Asjo, and S. Wain Hobson.** 1989. Temporal fluctuations in HIV quasispecies in vivo are not reflected by sequential HIV isolations. Cell **58**:901–910.
33. **Meyerhans, A., J. P. Vartanian, and S. Wain Hobson.** 1990. DNA recombination during PCR. Nucleic Acids Res. **18**:1687–1691.
34. **Myers, G., B. Korber, J. Berzofsky, T. F. Smith, and G. N. Pavlakis.** 1991. Human retroviruses and AIDS 1991. Los Alamos National Laboratory, Los Alamos, N.Mex.
35. **Nei, M., and T. Gojobori.** 1986. Simple methods for estimating the numbers of synonymous and non-synonymous nucleotide substitutions. Mol. Biol. Evol. **3**:418–426.
36. **O'Brien, W. A., Y. Koyanagi, A. Namazie, J. Q. Zhao, A. Diagne, K. Idler, J. A. Zack, and I. S. 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.
37. **Pang, S., Y. Schlesinger, E. S. Daar, T. Moudgil, D. D. Ho, and I. S. Chen.** 1992. Rapid generation of sequence variation during primary HIV-1 infection. AIDS **6**:453–460.
38. **Phillips, R. E., S. Rowland Jones, D. F. Nixon, F. M. Gotch, J. P. Edwards, A. O. Ogunlesi, J. G. Elvin, J. A. Rothbard, C. R. Bangham, and C. R. Rizza.** 1991. Human immunodeficiency virus genetic variation that can escape cytotoxic T cell recognition. Nature (London) **354**:453–459.
39. **Psallidopoulos, M. C., S. M. Schnittman, L. M. Thompson, M. Baseler, A. S. Fauci, H. C. Lane, and N. P. Salzman.** 1989. Integrated proviral human immunodeficiency virus type 1 is present in $CD4^+$ peripheral lymphocytes in healthy seropositive individuals. J. Virol. **63**:4626–4631.
40. **Roos, M. T. L., J. M. A. Lange, R. E. Y. Goede, R. A. Coutinho, P. T. A. Schellekens, F. Miedema, and M. Tersmette.** 1992. Viral phenotype and immune response in primary human immunodeficiency virus type 1 infection. J. Infect. Dis. **165**:427–432.
41. **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.
42. **Schneweis, K. E., J. P. Kleim, E. Bailly, D. Niese, N. Wagner, and H. H. Brackmann.** 1990. Graded cytopathogenicity of the human immunodeficiency virus (HIV) in the course of infection. Med. Microbiol. Immunol. **179**:193–203.
43. **Schnittman, S. M., H. C. Lane, J. Greenhouse, J. S. Justement, M. Baseler, and A. S. Fauci.** 1990. Preferential infection of CD4+ memory T cells by human immunodeficiency virus type 1: evidence for a role in the selective T-cell functional defects observed in infected individuals. Proc. Natl. Acad. Sci. USA **87**:6058–6062.
44. **Schuitemaker, H., M. Koot, N. Kootstra, M. W. Dercksen, R. T. D. Goede, R. Steenwijk, J. M. A. Lange, J. K. Schattenkerk, F. Miedema, and M. Tersmette.** 1992. Biological phenotype of human immunodeficiency virus type 1 clones at different stages of infection: progression of disease is associated with a shift from monocytotropic to T-cell-tropic virus populations. J. Virol. **66**:1354–1360.
45. **Schuitemaker, H., N. A. Kootstra, R. E. de Goede, F. de Wolf, F. Miedema, and M. Tersmette.** 1991. Monocytotropic human immunodeficiency virus type 1 (HIV-1) variants detectable in all stages of HIV-1 infection lack T-cell line tropism and syncytium-inducing ability in primary T-cell culture. J. Virol. **65**:356–363.
46. **Shioda, T., J. A. Levy, and C. Cheng Mayer.** 1991. Macrophage and T cell-line tropisms of HIV-1 are determined by specific regions of the envelope gp120 gene. Nature (London) **349**:167–169.
47. **Simmonds, P., P. Balfe, C. A. Ludlam, J. O. Bishop, and A. J. Leigh Brown.** 1990. Analysis of sequence diversity in hypervariable regions of the external glycoprotein of human immunodeficiency virus type 1. J. Virol. **64**:5840–5850.
48. **Simmonds, P., P. Balfe, J. F. Peutherer, C. A. Ludlam, J. O. Bishop, and A. J. Leigh Brown.** 1990. Human immunodeficiency virus-infected individuals contain provirus in small numbers of peripheral mononuclear cells and at low copy numbers. J. Virol. **64**:864–872.
49. **Simmonds, P., L. Q. Zhang, F. McOmish, P. Balfe, C. A. Ludlam, and A. J. Leigh Brown.** 1991. Discontinuous sequence change of human immunodeficiency virus (HIV) type 1 *env* sequences in plasma viral and lymphocyte-associated proviral populations in vivo: implications for models of HIV pathogenesis. J. Virol. **65**:6266–6276.
50. **Takeuchi, Y., M. Akutsu, K. Murayama, N. Shimizu, and H. Hoshino.** 1991. Host range mutant of human immunodeficiency virus type 1: modification of cell tropism by a single point mutation at the neutralization epitope in the *env* gene. J. Virol. **65**:1710–1718.
51. **Tersmette, M., R. E. Y. de Goede, B. J. M. Al, I. N. Winkel, R. A. Gruters, H. T. M. Cuypers, H. G. Huisman, and F. Miedema.** 1988. Differential syncytium-inducing capacity of human immunodeficiency virus isolates: frequent detection of syncytium-inducing isolates in patients with acquired immunodeficiency syndrome (AIDS) and AIDS-related complex. J. Virol. **62**:2026–2032.
52. **Tersmette, M., R. A. Gruters, F. de Wolf, R. E. de Goede, J. M. Lange, P. T. Schellekens, J. Goudsmit, H. G. Huisman, and F. Miedema.** 1989. Evidence for a role of virulent human immunodeficiency virus (HIV) variants in the pathogenesis of acquired immunodeficiency syndrome: studies on sequential HIV isolates. J. Virol. **63**:2118–2125.
53. **Ulrich, P. P., M. P. Busch, T. El-Bienk, J. Shiota, J. Vennari, K. Shriver, and G. N. Vyas.** 1988. Assessment of human immunodeficiency virus expression in co-culture peripheral blood mononuclear cells from healthy seropositive subjects. J. Med. Virol. **25**:1–10.
54. **Weiss, S. H., J. J. Goedert, S. Gartner, M. Popovic, D. Waters, P. Markham, F. Di Marzo, M. H. Gail, W. E. Barkley, J. Gibbons, F. A. Gill, M. Leuther, G. M. Shaw, R. C. Gallo, and W. A. Blattner.** 1988. Risk of human immunodeficiency virus (HIV-1) infection among laboratory workers. Science **239**:68–71.
55. **Westervelt, P., H. E. Gendelman, and L. Ratner.** 1991. Identification of a determinant within the human immunodeficiency virus 1 surface envelope glycoprotein critical for productive infection of primary monocytes. Proc. Natl. Acad. Sci. USA **88**:3097–3101.
56. **Westervelt, P., D. B. Trowbridge, L. G. Epstein, B. M. Blumberg, Y. Li, B. H. Hahn, G. M. Shaw, R. W. Price, and L. Ratner.** 1992. Macrophage tropism determinants of human immunodeficiency virus type 1 in vivo. J. Virol. **66**:2577–2582.
57. **Willerford, D. M., M. J. Gale, R. E. Benveniste, E. A. Clarke, and W. M. Gallatin.** 1990. Simian immunodeficiency virus is restricted to a subset of blood CD4+ lymphocytes that includes memory cells. J. Immunol. **144**:3779–3783.

58. **Wolfs, T. F., J. J. de Jong, H. Van den Berg, J. M. Tijnagel, W. J. Krone, and J. Goudsmit.** 1990. Evolution of sequences encoding the principal neutralization epitope of human immunodeficiency virus 1 is host dependent, rapid, and continuous. Proc. Natl. Acad. Sci. USA **87:**9938–9942.
59. **Wolfs, T. F., G. Zwart, M. Bakker, M. Valk, C. L. Kuiken, and J. Goudsmit.** 1991. Naturally occurring mutations within HIV-1 V3 genomic RNA lead to antigenic variation dependent on a single amino acid substitution. Virology **185:**195–205.
60. **Wolfs, T. F. W., G. Zwart, M. Bakker, and J. Goudsmit.** 1992. HIV-1 genomic RNA diversification following sexual and parenteral virus transmission. Virology **189:**103–110.
61. **Wolinsky, S. M., C. M. Wike, B. T. M. Korber, C. Hutto, W. P. Parks, L. A. Rosenblum, K. J. Kunstman, M. R. Furtado, and J. L. Munoz.** 1992. Selective transmission of human immunodeficiency virus type-1 variants from mothers to infants. Science **255:**1134–1137.
62. **Zhang, L. Q., P. Simmonds, C. A. Ludlam, and A. J. Leigh Brown.** 1991. Detection, quantification and sequencing of HIV-1 from the plasma of seropositive individuals and from factor VIII concentrates. AIDS **5:**675–681.
63. **Zwart, G., H. Langedijk, L. van der Hoek, J. J. de Jong, T. F. Wolfs, C. Ramautarsing, M. Bakker, A. de Ronde, and J. Goudsmit.** 1991. Immunodominance and antigenic variation of the principal neutralization domain of HIV-1. Virology **181:**481–489.