Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 9061–9065, October 1993
Medical Sciences

Increasing antigenic and genetic diversity of the V3 variable domain of the human immunodeficiency virus envelope protein in the course of the AIDS epidemic

CARLA L. KUIKEN*, GABRIËL ZWART*, ELLY BAAN*, ROEL A. COUTINHO†, J. ANNEKE R. VAN DEN HOEK†, AND JAAP GOUDSMIT*‡

*Department of Virology, Academic Medical Center, Meibergdreef 15, 1105 AZ Amsterdam, The Netherlands; and †Municipal Health Service, Department of Public Health, Amsterdam, The Netherlands

Communicated by Manfred Eigen, June 21, 1993

ABSTRACT **Population-wide variation in genomic RNA of human immunodeficiency virus type 1 (HIV-1) encompassing the V3 loop of the envelope protein was studied in serum samples of 74 newly infected individuals from three Dutch cohorts: 30 homosexual men, 32 drug users, and 12 hemophiliacs. During acute infection, HIV-1 RNA sequences present in serum are relatively homogeneous, which makes direct sequencing feasible. This offered an opportunity to study the infecting virus variants before mutations had accumulated in the new host. The sampling dates ranged from 1980 to 1991, thus spanning the entire AIDS epidemic in the Netherlands. The diversity in the sequenced region increased over time in both the homosexual and the drug-user risk groups. Furthermore, this increase was associated with an increase in antigenic variation, as witnessed by serum reactivity to a V3 peptide panel. Despite this diversification, some 1990 sequences still closely resembled the earliest 1980 sequence, making ancestral inferences problematic. No evidence was found of a change in the master sequence of the virus quasispecies over time. At the amino acid level, no risk-group-associated variation was found, but at the nucleotide level, the drug-user and homosexual/hemophiliac sequences could be distinguished on the basis of a single silent nucleotide change in the sequence encoding the tip of the V3 loop. Hemophiliac sequences could not be distinguished from those of homosexuals. In spite of the large and increasing genetic variability, all sequences were more similar to the European/American HIV consensus sequence than to that of non-Western strains.**

The envelope protein of human immunodeficiency virus type I (HIV-1) contains five variable domains (V1–V5) separated by relatively conserved regions. The viral genome, and especially the envelope gene, is characterized by a high potential for variation without loss of functionality (e.g., ref. 1). The third variable domain (V3) has been shown to contain determinants of cell tropism (2), infectivity (3), and cytopathicity (4, 5). It is a target for neutralizing antibodies (6) and a recognition site for T cells (7). The V3 loop is immunogenic and subject to strong selective pressure (8). Based on sequence analysis, five genetic clusters of V3 sequences can be distinguished among viruses circulating in the current AIDS epidemic. European and American sequences form only one of these five types, the so-called type B (9). Analysis of the variation of the V3 region has also been used to establish the relatedness of the HTLV-IIIB and the LAI strains of HIV-1 (10), as well as to investigate epidemiological links between viruses occurring in apparent clusters, such as in the case of the Florida dentist (11). Upon transmission, some studies indicate that the major variant found in the donor is transmitted (12), while others report transmission of a minor variant (13, 14). It appears that, in the person newly infected through these routes, one variant multiplies quickly and thus establishes an initially rather homogeneous population (12, 15), irrespective of the heterogeneity of the population found in the donor. This homogeneity around antibody seroconversion allows for assessment of the dominant variant through direct sequencing. Until now, the assessment of V3 variation in the population was often based on comparison of sequence populations from samples taken late in infection, and frequently after cocultivation *in vitro*. The Amsterdam cohort studies (16–18) have made it possible to obtain samples from newly infected individuals at different points in time during the spread of the AIDS epidemic, and from three risk groups: homosexual men, drug users (DUs), and hemophilia patients. About 10 years of follow-up are available to study the development of the infecting virus variants. We obtained V3 sequences from viral genomic RNA from sera of these three risk groups, mostly taken at the time of seroconversion, spanning the AIDS epidemic in the Netherlands.

MATERIALS AND METHODS

Origin and Sampling of the Sera. In 1980 a prospective hepatitis B vaccine cohort study was started in the Netherlands (16). From this cohort, nine early HIV-1 antibody-positive serum samples were used. For eight of these, it is known that the window period, the time elapsed between the last antibody-seronegative and the first antibody-seropositive sample, was at most 3 months. Seven other early HIV-1-positive serum samples were identified, retrospectively, from stored serum samples from homosexual men. These had been submitted for viral diagnostics, mostly for suspected hepatitis, cytomegalovirus infection, or "gay immune compromised syndrome," which was later identified as acute HIV infection. We assume that these are early postseroconversion samples, both because it has been estimated that the virus was introduced into the Amsterdam homosexual community in the late 1970s or early 1980s (19) and because two of them were associated with an acute HIV-like syndrome. Fifteen samples taken from each of the years 1985 and 1990 were used from participants in the Amsterdam cohort of homosexual men (17). Since in 1985 seronegative participants in the cohort were screened every 3 months, the window period for the 1985 sera is 3 months; for the 1990 samples, it is 6 months.

Additional serum samples were obtained from 37 participants in the Amsterdam cohort of (mostly intravenous) DUs (18). Introduction of HIV into the DU risk group has been estimated to have occurred later than in the homosexual group, in the early or middle 1980s (20). Sampling years range from 1986 to 1991. For 25 of 32 samples the window period

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

Abbreviations: DU, drug user; SI, syncytium-inducing.

‡To whom reprint requests should be addressed.

was 6 months; the longest window period was 4 years. Three of the sequences are from samples of DUs who deny intravenous drug use. One of these was infected, in his own estimation, through a homosexual contact.

The hemophilia patients' samples were taken in 1984 and 1985, before blood and factor VIII/IX batches were screened for HIV. This group has been described (21). These sera were all taken within 6 months after the last seronegative sample.

Another study, using the MT2 cell assay (22), determined the syncytium-inducing (SI) capacity of virus present in samples of peripheral blood mononuclear cells taken at the same time as the sera from the homosexual men who seroconverted in 1985 and 1990 (23). SI virus was recovered from only one isolate.

RNA Isolation and Sequencing. Genomic RNA was isolated from the sera according to Boom *et al.* (24). The RNA was converted to cDNA with avian myeloblastosis virus reverse transcriptase, and the cDNA was used in an initial PCR as described (12). Direct sequences were generated by subjecting 1 μl of the first PCR product in a total volume of 50 μl to a second, asymmetric PCR with an SP6-extended 5'KSI primer (5'-GATTTAGGTGACACTATAGGCAGTCTAGCAGAAGAAG-3', 0.4 pmol; underlining indicates the SP6 extension) and the 3'KSI primer (20 pmol), and DNA for reverse sequencing was obtained with an SP6-extended 3'KSI primer (5'-GATTTAGGTGACACTATAGTGGGTCCCCTCCTGAGGA-3', 0.8 pmol) and the 5'KSI primer (20 pmol) under the same conditions as described above, except that the reaction mixture contained 2.2 mM $MgCl_2$ instead of 1.8 mM. The mixture was amplified in 20 cycles of 30 sec at 96°C, 1 min at 50°C, and 2 min at 72°C.

From some of the samples no RNA could be isolated. For the homosexual cohort, this was the case for 4 of 16 of the 1980/1981/1982 sera, 2 of 15 of the 1985 sera, and 2 of 15 of the 1990 sera. Two of 37 sera in the DU group and 1 of 13 sera in the hemophilia patient cohort yielded no RNA.

Peptide ELISAs. For these experiments, sera taken a year after the estimated seroconversion date were used when available, because early postseroconversion sera often have low antibody titers. The method and synthetic peptides used to determine the antibody reactivity of the sera have been described (25).

Data Analysis. For all analyses, positions containing alignment gaps or ambiguous nucleotides were excluded. Alignment of the sequences was straightforward and done manually. Phylogenetic analyses were done with the neighbor-joining method (26) as implemented in the PHYLIP package (27). Additional multivariate analysis of the sequences was done using the program PCOORD (28), which extracts the principal components from a pairwise distance matrix that has been transformed to make it Euclidean. Principal-component extraction is a method for projecting high-dimensional data sets onto a small number of axes, while conserving the maximum amount of information about the individual elements of the set. For bootstrapping, the SEQBOOT, DNADIST, and CONSENSE programs from this package were used. PHYLIP's DRAWTREE program was used to produce the plots. With all sequences present in the analysis, the number of valid positions was 250. For direct comparison of nucleotides, Hamming distances were used. Amino acid recoding to count only nonconserved changes was done according to the first recoding level in the scheme proposed by Smith and Smith (29). The distance matrix input for the neighbor-joining program was generated using either Kimura's two-parameter estimation for nucleotides (30), identity scoring for amino acid and recoded amino acid distances, or an estimation of the number of silent substitutions according to Nei and Gojobori's method (31). The analysis of the discriminatory power of individual positions was done using the CART (classification and regression trees) technique (32).

RESULTS

Variation in Nucleotide Sequences in Time. In Table 1, the mean pairwise intersequence Hamming distance, calculated over nucleotides, is shown per isolation period and risk group for the whole region. The first samples of homosexuals and DUs were less variable than the later ones [difference between earliest and latest samples, $P < 0.001$ for the homosexual (1980 vs. 1990) and $P < 0.02$ for the DU (1986 vs. 1991) sequences, Mann–Whitney statistic]. The variability of the latest samples was around 10%.

When the calculations were restricted to the V3 loop itself, contained between the cysteines, the sequences isolated from the homosexual group showed an increase of variation over time that appeared to be almost linear and in 1990 were the most variable. In the DU group, the diversity also increased after 1986, but there was no linear trend. Thus, in both these groups the variation increased with time. This increasing variability was not reflected in a change in consensus sequence. At the nucleotide level, the distance between the earliest and the latest consensus sequences was small, 1.8% in both groups. No evidence was found of fixation of amino acid changes.

The mean diversity did not differ much between the groups. The mean difference between the early (1985 and 1986/1987, respectively) homosexual and DU sequences was 8.9%, vs. 10.0% between the late (1990 and 1990/1991) sequences of these groups. These differences are of the same magnitude as the within-group differences in these years. Because hemophiliac infection incidence has decreased to almost zero since 1985 and the samples used in this study have a time span of <2 years, the data were too scarce to demonstrate temporal variation. For 1984 and 1985, the variation in this group is in the same order of magnitude as that in the other risk groups. Omitting the 3' flank of the sequences from the calculations gave results comparable to those obtained for the V3 loop. This was to be expected because of the moderate variation in the 5' region as compared to that in the 3' flank (data not shown). The ratio of silent to nonsilent changes [K_s/K_a (31)] was around 1.0 for all groups (range, 0.86–1.28). These values are commonly found in V3 sequences (13, 33) and are indicative of high selection pressure on the region; in a fragment of the *gag* gene, a value of 6.7 was reported (33).

Phylogenetic Analyses. In the dendrogram based on the nucleotides of all sequences (Fig. 1), a distinction can be seen between the sequences from DUs and those from the other risk groups. However, a bootstrap analysis showed that this dis-

Table 1. Variation of the sequences and distances to subtype consensus

	Homosexuals			DUs		Hemophiliacs
	1980–82	1985	1990	<1989	>1988	1984–85
1980–82	6.63					
1985	8.00	9.61				
1990	8.40	9.67	9.71			
<1989	7.38	9.06	9.33	7.27		
>1988	8.39	9.5	10.14	7.76	8.91	
1984–85	6.86	8.37	8.39	7.97	8.99	7.77
Type A	15.38	14.30	16.06	15.95	16.87	15.68
Type B	3.44	5.02	4.44	4.18	5.30	3.97
Type C	16.52	16.03	17.31	16.43	17.53	17.27
Type D	15.34	15.31	16.06	16.46	16.98	15.98
Type E	19.07	19.03	19.93	19.95	20.24	19.80

Values are mean Hamming distances (expressed as percentage of total sequence length) within and between groups of sequences. The lower five rows give distances of the sequences in each group to the consensus of the subtypes A–E. Intersequence distances per single year in the DU group were 6.75, 7.64, 9.08, 9.56, 8.31, and 10.15% for the years 1986 through 1991.

tinction was not reproducible. No other distinctions appear in the dendrogram in Fig. 1. Statistical geometry analysis (34) indicated that the amount of phylogenetic information in the data set was limited and that a high level of noise had accumulated (data not shown). This is also suggested by the shape of the tree: long branches lead to the "leaves" (the sequences themselves), whereas the internal branches are short. In dendrograms made separately for the DU and homosexual groups, there was once again no indication of any subclusters. The inferred ancestral relationships between the sequences in these two group-dendrograms did not agree with the known temporal ordering based on the year of sampling. The rank correlation (Spearman's ρ) of the pairwise differences between the sequences, on the one hand, and the difference between their respective years of sampling, on the other, was not statistically significant in either group. Because the diversity does increase over time, this can only mean that some 1990 sequences still closely resemble the 1980 sequences. This is in agreement with the short branches of some late sequences in the dendrogram in Fig. 1.

All analyses based on measures that did not involve amino acid changes, such as third codon positions and the estimated number of silent changes between the sequences, yielded dendrograms where the DU sequences were clearly distinct from the homosexuals' and hemophiliacs' sequences. This distinction was lost in all analyses that used measures associated with amino acid changes, such as the first or second codon positions in the sequences, differences between the amino acids, or distances based only on nonconserved amino acid differences (data not shown). The hemophiliacs' and homosexuals' sequences were intermixed in all analyses. We found no evidence of any temporal ordering of the sequences, which again shows that there are no systematic time-related changes in these sequences.

A final analysis was done to relate the variation in this data set to that found worldwide. We included the five HIV-1 subtypes that were recently presented by the Los Alamos group (9). The results indicated that all sequences in this data set still were much closer to the B subtype, which is most commonly found in Europe and the United States of America, than to subtypes representing sequences found elsewhere in the world. This can be seen immediately in Table 1.

Silent Mutations That Discriminate Between the Risk Groups. Other analysis methods were used to further investigate the existence of two distinct groups in the sequences. The separation of the DU sequences from the rest was prominent as shown by principal-coordinate analysis, a method similar to principal-component analysis (28). The method attempts to reduce the (transformed) matrix of pairwise distances to a small number of dimensions (called components), while retaining as much information as possible. The distinction was present in the first axis extracted. The most important axis in the data set explained only 7.6% of the total variation in the matrix, which was expected because of the large extent of randomization in the data. However, it separated the two groups efficiently, misclassifying only four DU sequences. Though the amount of explained variation is small, the fact that the first component extracted is the one separating the groups indicates that this group information is the most "visible" above the noise in the data set.

The next question was which of the third-codon-position residues contributed most to the differentiation of the DU group. For this purpose we used the CART (classification and regression trees) technique, which is a method to achieve optimal discrimination between groups based on categorical (non-numerical) data. This analysis showed that two positions could be used to discriminate between the groups. The first and most important of these was the third position of the GGN codon that encoded the second glycine in the GPGR sequence that is located at the tip of the V3 loop (position 312). This codon is GGC in 28 of 32 of the DU sequences, whereas it was formed by GGA (2 of 42) or GGG in the other risk groups. The second distinguishing nucleotide is in the third position of the codon that encodes serine 274, encoded by a TCT in 20 of 30 of the DU sequences (2 had an amino acid-changing mutation in that codon), and by TCC in all of

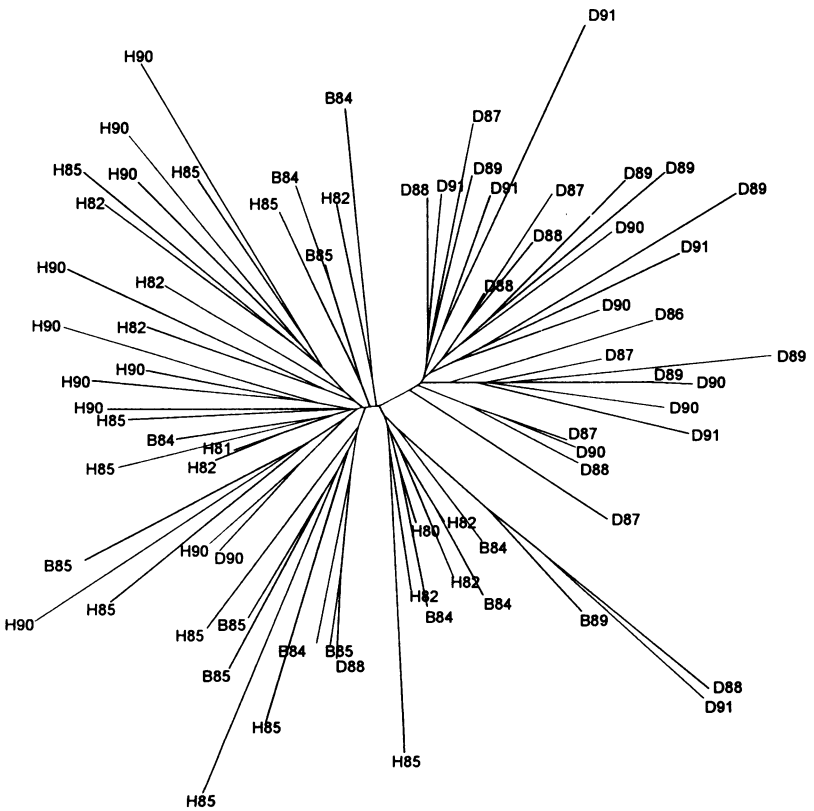

FIG. 1. Dendrogram produced on the basis of all 74 nucleotide sequences. The dendrogram was calculated by the neighbor-joining method and drawn with the DRAWTREE program. Sequence names have been replaced by symbols for legibility. Letters H, D, and B indicate sequences from homosexuals, DUs, and hemophiliacs, respectively; the number indicates the year of sampling.

12 hemophiliac and 28 homosexual-derived sequences (2 containing an amino acid-changing mutation). This second nucleotide did not add any discriminating power, because it was completely nested within the first. None of the DU sequences that were misclassified on the basis of the GGN codon contained the other DU-associated codon, TCT.

Four DU sequences were consistently classified with the homosexual/hemophiliac sequences in the phylogenetic analyses and in the principal-coordinate analysis. These were also the sequences lacking the C in the GGN codon.

Antigenic and Phenotype-Associated Variation. Previous work from Fouchier *et al.* (5) and from our group (35) has shown that two amino acid residues at positions 306 and 320 in the V3 loop are associated with the SI capacity of the whole virus. A positively charged residue (arginine or lysine) at one of these positions, combined with a non-negative one at the other, is associated with a virus that has the SI phenotype. In this data set, judging by these two positions, 4 sequences of the 74 (5.4%) are from SI variants. Two of these are from the DU group, one from the homosexual group, and one from the hemophiliac group. All except the last are from 1990 seroconversions. Results of coculturing peripheral blood mononuclear cells of 20 homosexual men from this group with MT2 cells showed MT2 tropism in one isolate, which is in exact agreement with the serum sequence data.

Antigenic reactivity to synthetic peptides mimicking part of the V3 loop is strongly associated with a single amino acid at position 308 (8, 25). We determined the highest antibody reactivity of sera from all persons in this study to three peptides that represent the most common amino acids found at position 308 in Europe and the United States, embedded in the European/United States consensus sequence (Table 2). Fifty-five of sixty-three samples with specific peak reactivity reacted most strongly with the peptide that contains the homologous amino acid at position 308; 11 samples showed either no reactivity above the cutoff or no specific reactivity to any of the three peptides (highest reactivity less than 10% higher than the next highest).

In the homosexual cohort, there appeared to be a diversification at this position: in the early 1980s, 9 of 10 were histidines; in 1985, 8 of 10; and in 1990, only 4 of 10 were histidines, with the rest of the sequences having either proline (4 sequences), arginine, or glutamine at that position. In the DU group, the trend was much less clear, possibly because of the reduced time span, but more of the early (before 1989) sequences contained histidine (9 of 12) than the later ones (12 of 19). The hemophiliac population again consisted of 85% histidine, but in spite of its apparent phylogenetic relation to the homosexuals' sequences, the two different amino acids were not the prolines that make up the rest of the 1985 homosexual sample, but a threonine and a glutamine.

Table 2. Relation between serum reactivity and amino acid 308

Amino acid at 308	No. of sera			
	Homologous	Heterologous	Nonreactive	Aspecific
Histidine	42	2	3	4
Proline	11	0	0	0
Asparagine	4	1	1	0
Other	—	2	2	1

Antibody response to three hexadecapeptides containing different amino acids at envelope position 308 in a consensus background [RKSI(H/N/P)IGPGRAFYTTG] was assayed by ELISA. Numbers indicate number of sera showing peak reactivity with the peptide containing the homologous amino acid in a consensus background, showing peak reactivity with a peptide containing a different amino acid, and showing no or multiple reactivity.

DISCUSSION

It has been suggested that V3 variability may be limited in HIV variants that are easily transmitted (15). The results reported here do not support that notion. The interperson heterogeneity of sequences taken shortly after infection (around 10% in the latest samples) is not much lower than that commonly found in later stages of infection. In a group of unrelated asymptomatic HIV-1 carriers from Florida, a mean interpatient variation of 13% was found for the same region, with a range of 8–15% (11). We conclude that there is no clear indication in this data set that only a small range of different V3 variants is allowed in viruses that are able to initiate infection. Even though it is by no means certain that infection is established by macrophage-tropic virus variants, it has been established with some certainty that early in infection, the majority of isolates are macrophage-tropic (36). The region surrounding V3 has been shown to contain a determinant of macrophage tropism (37), but individual amino acids that convey this tropism have not been pinpointed. From our data, there seems to be no definite amino acid pattern associated with V3 variants obtained shortly after transmission. Lack of a "common signature" in transmitted virus V3 sequences was also reported in a recent study (14). Two amino acids that were found to be indicative of macrophage-tropic and—by implication—transmissible virus (arginine-305 and alanine-317) (15) were present in this data set but did not constitute a majority.

In none of the groups was any evidence found of an accumulation of mutations over time leading to a change in the genetic constitution of the quasispecies. The diversity of the V3 region of the viral envelope protein does increase as the epidemic progresses, but the variation seems to be random. This is a discouraging trend, especially since it is mirrored in the antigenic variation: the antigenically important position 308 also contains more diverse amino acids in the later samples, as opposed to a majority of histidines in the earlier ones.

A subdivision of widely divergent HIV-1 variants into five subtypes (A–E) has been constructed recently (9). The amino acid consensus sequence of this data set is also identical to the type B consensus. This may mean that it takes longer than the 15 years that the epidemic has existed in Amsterdam before this region is changed so much that it turns into a different subtype. Alternatively, it may be an indication that the virus circulating in Europe and the United States has a limited scope of V3 variation—for example, because its secondary structure or other factors restrict the type of variation that it can support. At the epidemiological level, it shows that variants of non-Western origin have not entered these risk-group samples to date, and we take this as an indication that in Amsterdam, mixing between European/American and other variants (or subtype B vs. other subtypes) has occurred on only a small scale, if at all.

A number of silent mutations differentiated between the DU sequences and the homosexual/hemophiliac ones. The distinction is difficult to detect because of the high noise level in the data, but a number of subsequent analyses give us some confidence in its reality. The results from principal-coordinate analysis clearly show the same distinction and indicate that it is the strongest trend in the data set. By using a non-numerical discriminant analysis, the distinction was found to be neatly preserved in two silent mutations. This makes sense given the high immune pressure directed against the V3 loop at the amino acid level, which would be expected to quickly wipe out any nonfunctional group differences at the amino acid level. That so few positions support the distinction may also partly explain its virtual disappearance in a bootstrap analysis, in which the same phylogenetic analysis is repeated many times while deleting some of the positions in the sequences and replacing them by duplicates of others. However, sequencing a region that is less exposed to selective pressures will be the only way to establish the

distinction between the risk groups beyond any doubt. The way in which the group information is preserved is intriguing. It seems most likely that the infection of the DU risk group started with a single virus variant that had this characteristic. The fact that a second codon also reflects the distinction provides indirect support for this heredity hypothesis.

All hemophiliac sequences were consistently grouped with the homosexual sequences. This may be attributed to a lack of reliable phylogenetic information, but it is also the most likely situation from an epidemiological point of view. Blood donations by DUs are rare in the Netherlands (C. van der Poel; personal communication). It may also indicate that the homosexual HIV variants are more representative of variants circulating in the population as a whole, and the DUs' variants are mavericks. The DU-associated GGC codon is found infrequently in the Los Alamos sequence collection: only 1 of 30 sequences had a C at that position, the macrophage-tropic virus HIV-BAL (9).

The inferred ancestral ordering of the sequences did not agree with the known time of sampling. This indicates that evolutionary relationships cannot be reliably determined on the basis of these V3-region sequences. This may be because the sequences are too short and contain too much noise (e.g., caused by immune pressure) to show these relationships. Alternatively it may be that "old" variants stay around for a long time or reemerge from a reservoir where they are able to maintain the same genotype for many years. A third possibility that has to be considered is that of convergent evolution, where V3 regions from viruses with different ancestry come to closely resemble each other. Again, this can only be resolved by sequencing another part of the genome.

Only 4 of the 74 sequences contained amino acids that were previously shown to convey the SI phenotype in *in vitro* mutagenesis experiments (35). This number is small when compared with that in a cross-sectional sample from the homosexual cohort, where 22 of 188 seropositives (11.7%) were found to harbor SI variants (22), but it agrees with the data reported by Keet *et al.* (23), who found that 2 of 96 seroconversion samples from this cohort contained SI variants, as determined by cocultivation. A subset of this study overlapped with the present study and confirms the dominance of non-SI virus in seroconverters. There are other reports of low prevalence of SI virus early in infection and preferential transmission of non-SI virus (36). Since the outgrowth of SI virus in the host later in infection is often accompanied by an overall accumulation of mutations (8, 38), this again suggests that sequences in this seroconversion set are more representative of virus variants circulating in the population than are randomly selected sequences from long-term-infected individuals.

We wish to express our gratitude to Margreeth Valk, Nel Albrecht, Gerard Mientjes, Harry van Haastrecht, and Nico Lelie for help in collecting the data; to Bette Korber for her NEI program, and to Kay Nieselt and Jack Leunissen for technical assistance and helpful discussions. This study is part of the Amsterdam Cohort Studies and was funded in part by a grant from the National Foundation for Preventive Medicine.

1. Pedrosa Martins, L., Chenciner, N., Asjö, B., Meyerhans, A. & Wain-Hobson, S. (1991) *J. Virol.* **65**, 4502–4507.
2. Shioda, T., Levy, J. A. & Cheng-Mayer, C. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 9434–9438.
3. Freed, E. O., Delwart, E. L., Buchschacher, G. L., Jr., & Panganiban, A. T. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 70–74.
4. de Jong, J. J., de Ronde, A., Keulen, W., Tersmette, M. & Goudsmit, J. (1992) *J. Virol.* **66**, 6777–6780.
5. Fouchier, R. A. M., Groenink, M., Kootstra, N. A., Tersmette, M., Huisman, H. G., Miedema, F. & Schuitemaker, H. (1992) *J. Virol.* **66**, 3183–3187.
6. Goudsmit, J., Debouck, C., Meloen, R. H., Smith, L., Bakker, M., Asher, D. M., Wolff, A. V., Gibbs, C. J., Jr., & Gajdusek, D. C. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 4478–4482.
7. Takahashi, H., Nakagawa, Y., Pendleton, C. D., Houghten, R. A., Yokomuro, K., Germain, R. N. & Berzofsky, J. A. (1992) *Science* **255**, 333–336.
8. Wolfs, T. F. W., Zwart, G., Bakker, M., Valk, M., Kuiken, C. L. & Goudsmit, J. (1991) *Virology* **185**, 195–205.
9. Myers, G., Korber, B., Berzofsky, J. A., Smith, R. F. & Pavlakis, G. N. (1992) *Human Retroviruses and AIDS 1992. A Compilation and Analysis of Nucleic Acid and Amino Acid Sequences* (Theor. Biol. Biophys., Los Alamos, NM).
10. Wain-Hobson, S., Vartanian, J.-P., Henry, M., Chenciner, N., Cheynier, R., Delassus, S., Pedrosa Martins, L., Sala, M., Nugeyre, M.-T., Guetard, D., Klatzmann, D., Gluckman, J.-C., Rozenbaum, W., Barre-Sinoussi, F. & Montagnier, L. (1991) *Science* **252**, 961–965.
11. Ou, C. Y., Clesieisky, C. A., Myers, G., Bandes, C. I., Luo, C.-C., Korber, B. T. M., Mullins, J. I., Schochetman, G., Berkeiman, R. L., Economou, A. N., Witte, J. J., Furman, L. J., Satten, G. A., MacInnes, K. A., Curran, J. W., Jaffe, H. W. & Epidemiologic Investigation Group Laboratory Investigation Group (1992) *Science* **256**, 1165–1171.
12. Wolfs, T. F. W., Zwart, G., Bakker, M. & Goudsmit, J. (1992) *Virology* **189**, 103–110.
13. Wolinsky, S. M., Wike, C. M., Korber, B., Hutto, C., Parks, W. P., Rosenblum, L. L., Kunstman, K. J., Furtado, M. R. & Munoz, J. (1992) *Science* **255**, 1134–1137.
14. Zhu, T., Mo, H., Wang, N., Nam, D. S., Cao, Y., Koup, R. A. & Ho, D. D. (1993) *Science*, in press.
15. Zhang, L. Q., MacKenzie, P., Cleland, A., Holmes, E. C., Leigh Brown, A. J. & Simmonds, P. (1993) *J. Virol.* **67**, 3345–3356.
16. Coutinho, R. A., Lelie, N., Albrecht-van Lent, P., Reerink-Bongers, E. E., Stoutjesdijk, L., Dees, P., Nivard, J., Huisman, J. G. & Reesink, H. W. (1983) *Br. Med. J.* **286**, 1305–1308.
17. de Wolf, F., Goudsmit, J., Paul, D. A., Lange, J. M. A., Hooykaas, C., Schellekens, P., Coutinho, R. A. & van der Noordaa, J. (1987) *Br. Med. J.* **295**, 569–572.
18. van den Hoek, J. A. R., Coutinho, R. A., van Haastrecht, H. J. A., van Zadelhoff, A. W. & Goudsmit, J. (1988) *AIDS* **2**, 55–60.
19. Coutinho, R. A., Krone, W. J. A., Smit, L., Albrecht-van Lent, P., van der Noordaa, J., Schaesberg, W. & Goudsmit, J. (1986) *Genitourin. Med.* **62**, 38–43.
20. van Haastrecht, H. J. A., van den Hoek, J. A. R., Mientjes, G. H. & Coutinho, R. A. (1992) *AIDS* **6**, 131–132.
21. Wolfs, T. F. W., Breederveld, C., Krone, W. J. A., van der Hoek, L., Bakker, M., Smit, L., Goudsmit, J. & the Dutch Haemophilia Group (1988) *Thromb. Haemostasis* **59**, 396–399.
22. Koot, M., Keet, I. P. M., Vos, A. H. V., de Goede, R. E. Y., Roos, M. T. L., Coutinho, R. A., Miedema, F., Schellekens, P. T. A. & Tersmette, M. (1993) *Ann. Int. Med.*, in press.
23. Keet, I. P. M., Krijnen, P., Koot, M., Lange, J. M. A., Miedema, F., Goudsmit, J. & Coutinho, R. A. (1993) *J. AIDS* **7**, 51–57.
24. Boom, R., Sol, C. J. A., Salimans, M. M. M., Jansen, C. I., Wertheim-van Dillen, P. M. E. & van der Noordaa, J. (1990) *J. Clin. Microbiol.* **28**, 495–503.
25. Zwart, G., Wolfs, T. F. W., Valk, M., van der Hoek, L., Kuiken, C. L. & Goudsmit, J. (1992) *AIDS Res. Hum. Retroviruses* **8**, 1897–1908.
26. Saitou, N. & Nei, M. (1987) *Mol. Biol. Evol.* **4**, 406–425.
27. Felsenstein, J. (1990) PHYLIP Manual Version 3.2 (Univ. Herbarium, Univ. of California, Berkeley).
28. Higgins, D. G. (1992) *Comput. Appl. Biosci.* **8**, 15–22.
29. Smith, R. F. & Smith, T. F. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 118–122.
30. Kimura, M. (1980) *J. Mol. Evol.* **16**, 111–120.
31. Nei, M. & Gojobori, T. (1986) *Mol. Biol. Evol.* **3**, 418–426.
32. Breiman, L., Friedman, J. H., Olshen, R. A. & Stone, C. J. (1984) *Classification and Regression Trees* (Wadsworth & Brooks, Pacific Grove, CA).
33. Balfe, P., Simmonds, P., Ludlam, C. A., Bishop, J. O. & Leigh Brown, A. J. (1990) *J. Virol.* **64**, 6221–6233.
34. Eigen, M., Winkler-Oswatitsch, R. & Dress, A. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 5913–5917.
35. de Jong, J. J., Goudsmit, J., Keulen, W., Klaver, B., Krone, W. J. A., Tersmette, M. & de Ronde, A. (1992) *J. Virol.* **66**, 757–765.
36. Roos, M. T. L., Lange, J. M. A., de Goede, R. E. Y., Coutinho, R. A., Schellekens, P. T. A., Miedema, F. & Tersmette, M. (1992) *J. Infect. Dis.* **165**, 427–432.
37. Westervelt, P., Gendelman, H. E. & Ratner, L. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 3097–3101.
38. Kuiken, C. L., de Jong, J. J., Baan, E., Keulen, W., Tersmette, M. & Goudsmit, J. (1992) *J. Virol.* **66**, 4622–4627.