Journal of Virology, July 1992, p. 4622–4627
0022-538X/92/074622-06$02.00/0

Vol. 66, No. 7

Evolution of the V3 Envelope Domain in Proviral Sequences and Isolates of Human Immunodeficiency Virus Type 1 during Transition of the Viral Biological Phenotype

CARLA L. KUIKEN,[1] JEAN-JACQUES DE JONG,[1] ELLY BAAN,[1] WILCO KEULEN,[1] MATTHIJS TERSMETTE,[2] AND JAAP GOUDSMIT[1]*

Human Retrovirus Laboratory, Academic Medical Center, L1-157, Meibergdreef 15, 1105 AZ Amsterdam,[1] and Central Laboratory of the Netherlands Red Cross Blood Transfusion Service, 1066 CX Amsterdam,[2] The Netherlands

Received 23 December 1991/Accepted 3 April 1992

The third variable domain (V3) of the envelope gene of human immunodeficiency virus type 1 contains a major neutralization epitope and determinants of syncytium-inducing (SI) capacity and replication rate (reviewed by J. P. Moore and P. L. Nara, AIDS Suppl. 2:S21–S33, 1991). Sequences were generated from DNA of samples taken 3 months apart over a period of 24 and 30 months from peripheral blood mononuclear cells (PBMC) of two individuals, both before and after cocultivation with uninfected donor PBMC. The isolated virus shifted from the non-syncytium-inducing (NSI) phenotype to the SI phenotype during the study period. This shift was associated with distinct changes in the V3 domain in both patients. The association of the phenotype shift with the V3 sequence changes was confirmed by construction of viruses with chimeric V3 loops. The shift from NSI- to SI-associated V3 variants was also seen in the uncultured PBMC of both patients, but not until 3 and 9 months after the detection of SI virus in culture. In the samples of uncultured PBMC DNA, several subgroups of sequences were found, indicating that the process of evolution may not be gradual and that several distinct populations can coexist. The paucity of intermediate sequences indicated that strong selection pressure was exerted on this part of the envelope. The early emergence of disease-associated SI variants in cultured material indicates that virus culture may have relevance for the in vivo situation.

Two initially asymptomatic human immunodeficiency virus type 1-infected homosexual men, in whom the virus underwent a change from the non-syncytium-inducing (NSI) phenotype to the syncytium-inducing (SI) phenotype were selected for this study. The association of the emergence of SI virus with progression to AIDS has been established in the Amsterdam cohort of homosexual men by Tersmette et al. (11). Patient H479 developed AIDS 9 months after the phenotype of isolated virus shifted from NSI to SI; patient H168 did so 15 months after the phenotype switch. Characteristics of the two study patients, designated H168 and H479, and of the virus isolated from their peripheral blood mononuclear cells (PBMC) are shown in Table 1. The cocultivation procedure and the determination of SI capacity in PBMC have been described earlier (10). DNA was isolated from patient PBMC both before and after cocultivation with uninfected donor PBMC, amplified by polymerase chain reaction, and cloned, and individual clones were sequenced by using an automated sequencer. This procedure has been described earlier in detail (12). Direct sequences, omitting the cloning step, were obtained from the virus isolates of patient H168. These sequences have been published earlier (6) and are shown for comparison only.

The methods for constructing the chimeric viruses and for monitoring replication and syncytium formation have been described before (6). In brief, the V3 region was amplified by polymerase chain reaction using primers spanning the two cysteines bordering the V3 region. The amplified V3 region was inserted in a plasmid containing the *Nco*I-to-*Bam*HI fragment of HXB-2. Subsequently, this chimeric V3 envelope fragment was cloned into a full-length molecular HXB-2 clone lacking this envelope fragment. Thus, a complete molecular HXB-2 clone with a chimeric V3 region was created. The molecular clones were transfected by electroporation into SupT1 cells. To monitor viral replication, p24 core antigen was measured by using a capture enzyme-linked immunosorbent assay. Formation of syncytia was assessed by two independent workers.

Hamming distances between sequences were the basis for UPGMA cluster analysis and ratio-level multidimensional scaling (14).

Evolution of the V3 region in the viral isolates. The evolution of the V3 region in the viral isolates was remarkably similar in both patients (Fig. 1 and 2). Two distinct genotype changes occurred in the V3 region. The first change occurred between samples 1 and 3 in patient H479 and between samples 2 and 3 in patient H168. In H479, it consisted of two nucleotide changes at amino acids 306 (S→G) and 308 (N→H); in patient H168, one nucleotide changed, corresponding to amino acid 308 (H→P). The second, more drastic genotype change took place in both patients between samples 4 and 5 and coincided with the phenotype switch from NSI to SI (Table 1). In patient H479, this change consisted of the replacement of six nucleotides, encoding four amino acids at positions 314, 319, 320, and 324. In patient H168 too, six nucleotides changed, corresponding to five amino acid changes at positions 306, 308, 317, 320, and 324. The V3 regions from samples 5 and later most likely belonged to SI virus. Hereafter, the sequences up to sample 4 will be termed "NSI associated" and the sample 5 (and later) sequences will be termed "SI associated."

Composition of the groups of V3 clones. The composition of the groups of V3 clones from the viral isolates of patient

* Corresponding author.

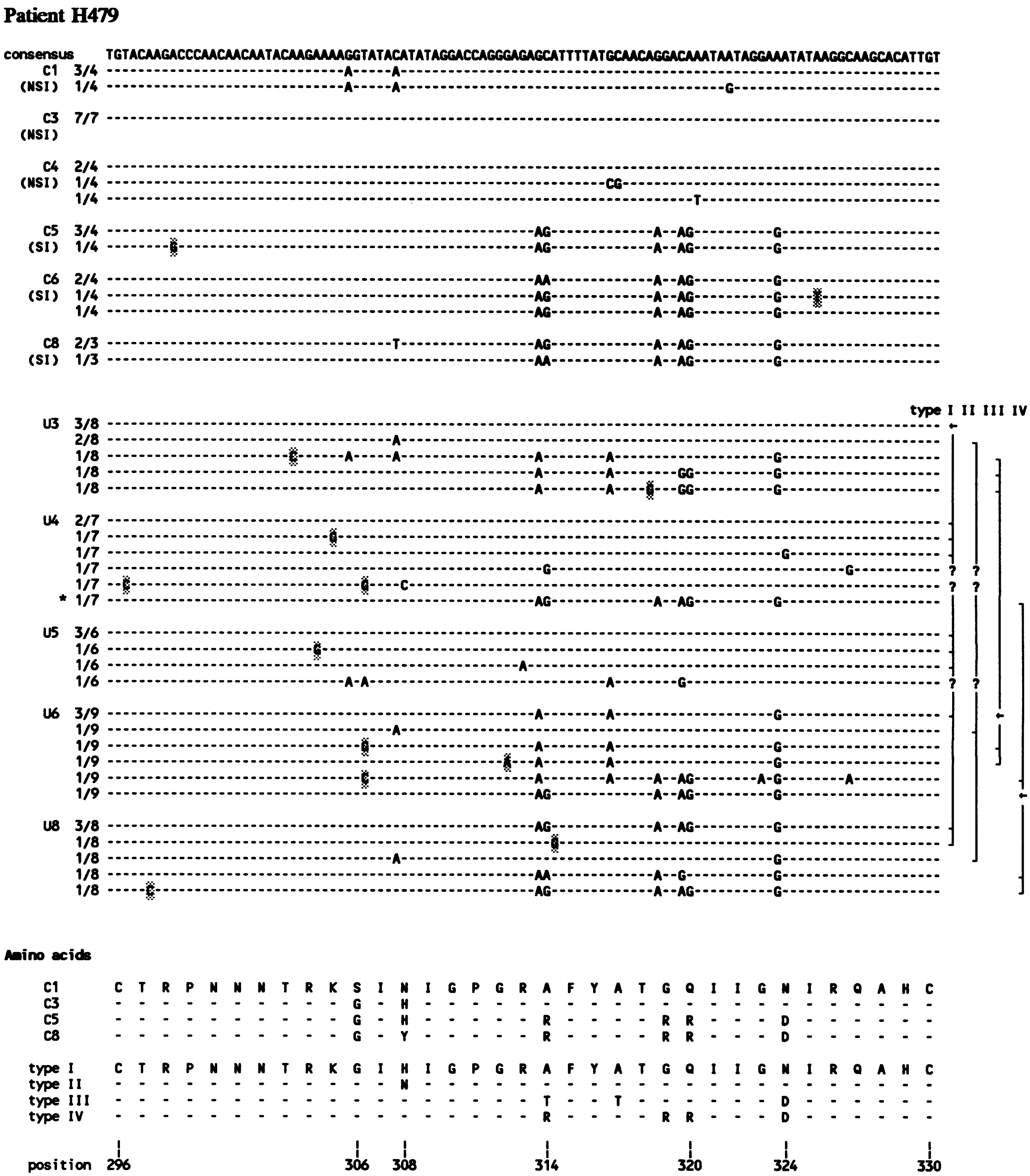

FIG. 1. V3 sequences recovered from material from patient H479. The consensus nucleotide sequence is taken from all sequences shown. All other sequences were compared with the consensus sequence, and identical nucleotides were replaced by hyphens. Stippled changes are silent. At the left-hand side of the figure, the letter U means the clones are from uncultured material, whereas C means the sequences were recovered from cocultivated material. For the cocultivation sequences, the biological phenotype of the bulk culture is indicated. The numbers after U and C are the sample numbers. The sample numbers represent 3-month intervals: samples with numbers differing by 1 were taken 3 months apart. The ratios are the number of identical clones per total number of clones in that sample. At the right-hand side of the figure, the lines (⊣) indicate the sequences that are classified as type I, II, III, or IV; the question marks indicate unclassifiable variants. The arrows indicate the sequences that are most characteristic of the indicated type. Amino acid sequences at the bottom of the figure represent four of five consensus sequences from the cultured material (consensus sequences of samples 3 and 4 were identical) and amino acid sequences corresponding to the four types distinguished in the uncultured material. The asterisk in sample U4 indicates the one clone with an SI-associated genotype before the NSI-to-SI switch.

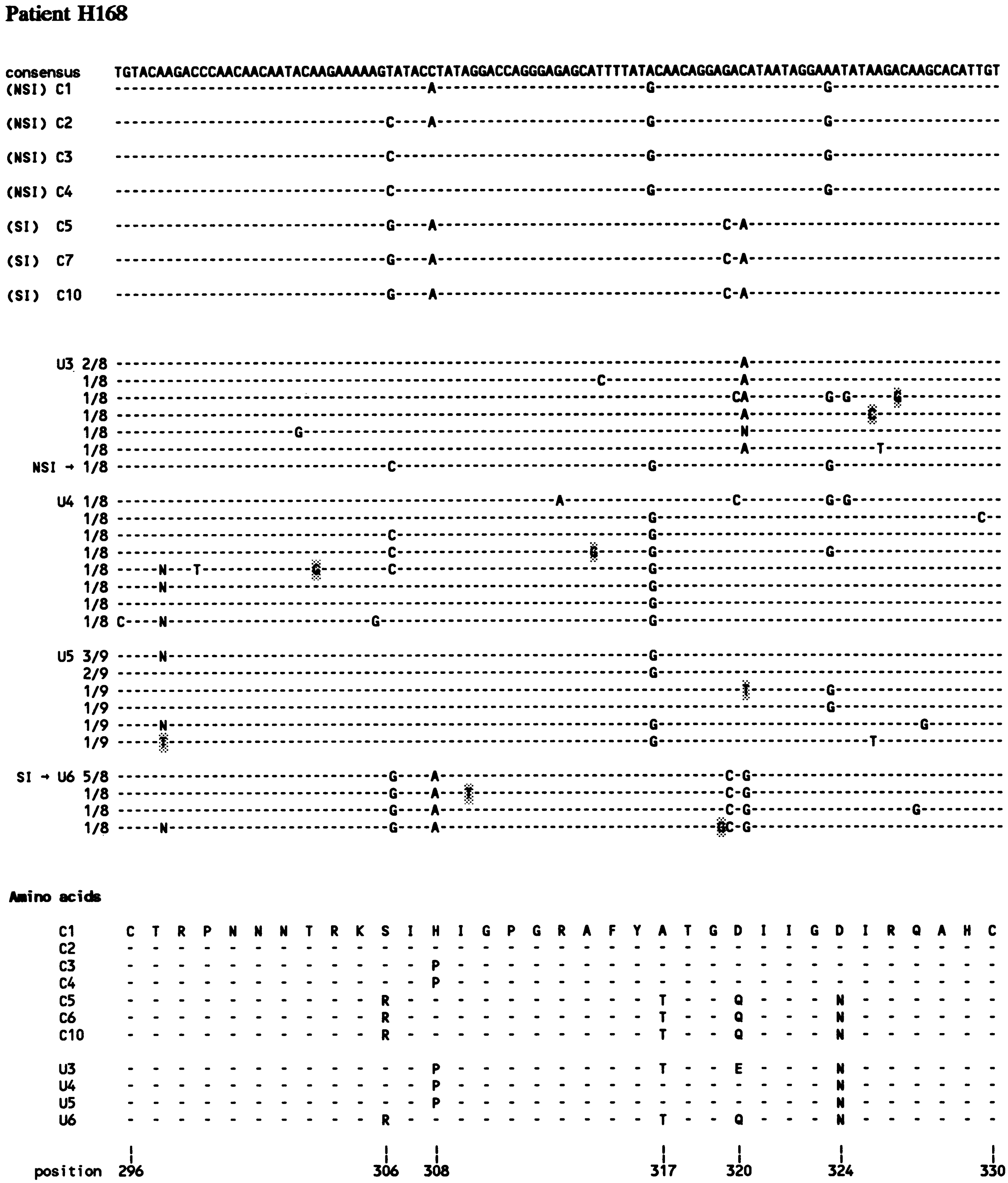

FIG. 2. V3 sequences recovered from material from patient H168. In samples U3 and U6, sequences identical to the NSI- and SI-associated V3 sequences in the cocultivated material are indicated by arrows. Derived amino acid sequences from uncultured material in this figure are sample consensus sequences. All other symbols are described in the legend to Fig. 1.

TABLE 1. Patient and virus characteristics during the follow-up period[a]

Sample no.	Patient H168				Patient H479			
	SI capacity	H9 tropism	p24 antigen (pg/ml)	No. of $CD4^+$ cells (10^9/liter)	SI capacity	H9 tropism	p24 antigen (pg/ml)	No. of $CD4^+$ cells (10^9/liter)
1	–	–	NT	0.2	–	–	–	0.4
2	–	–	49	NT	NT	NT	–	0.4
3	–	–	82	0.3	–	–	–	0.4
4	–	–	117	0.6	–	–	–	0.3
5	+	+	99	0.4	+	+	–	0.3
6	NT[b]	NT	NT	NT	+	+	–	0.3
7	+	+	185	0.3	+	+	–	0.2
8	NT	NT	NT	NT	+	+	–	0.1
10	+	+	NT	0.0	NT	NT	NT	NT

[a] Adapted with permission (11).
[b] NT, not tested.

H479 was very homogeneous: the Hamming distances from the clone sequences to the consensus sequence ranged from 0 to 0.71%. Since widely divergent African strains could be efficiently detected using the same primer set, this homogeneity was probably no primer selection artifact (15). To make sure that we sequenced different variants and not polymerase chain reaction replicas of the same variant, we compared a longer stretch with 79 additional nucleotides upstream and 37 nucleotides downstream of the V3 region in samples 4 to 6. We found that in these 221 nucleotides, no single sequence was identical to another. Since isolates of patient H168 were sequenced without a cloning step, no homogeneity data were available for this patient. The sequences directly derived from patient PBMC showed much more variability, the mean clone-consensus distance ranging from 0.95 to 2.7% in patient H479 and from 0.35 to 1.4% in patient H168.

Dynamics of progression toward the SI-associated genotype. The progression toward the SI-associated genotype appeared to be relatively straightforward in patient H168: the order of appearance of the variants in the uncultured material was the same as that in the viral isolates. The uncultured material of this patient had a delay of one sample (3 months) compared with that of the virus isolate. In contrast, the dynamics of the uncultured PBMC sequences in patient H479 were much more complex and seemed to depend on more factors than just chronology. Cluster analysis revealed the existence of subgroups of sequences spanning the different samples (types I to IV in Fig. 1). Though type I seemed to consist mainly of early sequences and type IV consisted mostly of late ones, sequences of the same type showed up in different samples. In patient H168, the clustering agreed much better with the chronological ordering of the sequences. The NSI- and SI-associated sequences in the isolates are both present in uncultured DNA: type I and IV, respectively, in patient H479; one clone in sample 3 and the sample 6 consensus in patient H168 (Fig. 2, indicated by arrows).

To further investigate the evolution of the SI phenotype, multidimensional scaling was done on the matrix of Hamming distances between the consensus sequences of all samples from patient H479. This statistical technique can be used to assess the number of mathematical dimensions that are minimally required to give an accurate description of the data. If one dimension is sufficient, the datum points can be represented as lying on a straight line. If many acquired mutations are lost again and there is no unidirectional evolutionary line, more than one dimension is needed to accurately represent the Hamming distance matrix.

Two dimensions were needed to accurately represent our data set, consisting of 11 consensus sequences. In Fig. 3, the change in the virus of H479 over time is represented graphically. The consensus sequences were separated into what could be termed an early group and a late group. Sequences from virus isolates from sample 5 were in the late group, whereas those from uncultured cells from sample 5 were in the early group. This corresponds to the observation that the cultured material shows an evolutionary change before the uncultured material does.

The uncultured sample 6, corresponding to type III, was the main cause of deviation from the one-dimensional representation (Fig. 3). When this sequence was omitted from the data set, a one-dimensional model produced adequate fit.

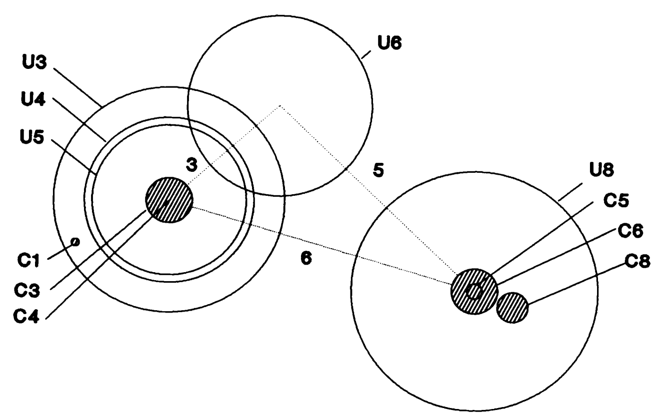

FIG. 3. Two-dimensional representation of the changes in the sequence populations of patient H479, based on the multidimensional scaling of the matrix of Hamming distances between nucleotide consensus sequences. Distances between the circle centers represent these Hamming distances. Thus, the difference between cultured samples 3 (C3) and 6 (C6) is 6 nucleotides, while between uncultured samples 5 and 6, it is 3 nucleotides. The radius of each circle is proportional to the mean Hamming distance between the clones and the consensus sequence of each sample. The circles were added to the figure manually. Hatched circles represent cultured material (C), and open circles represent uncultured patient material (U).

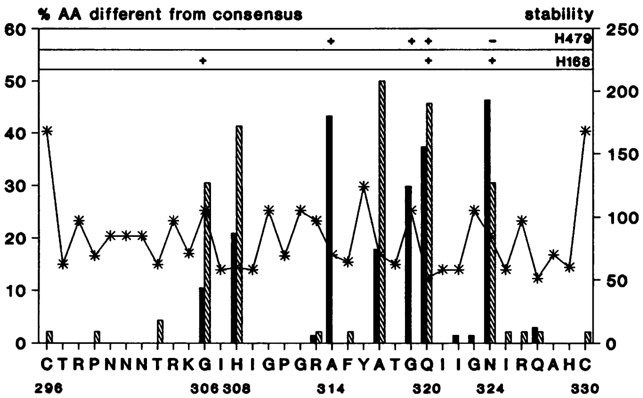

FIG. 4. Amino acid variability of all sequences made from uncultured material. Amino acid differences were scored with respect to the consensus sequence over all sequences, which was identical to the consensus sequence for patient H479 (shown at the bottom of the figure). Bars indicate the percentages of sequences with deviant amino acids (AA) at that position for patients H479 (■) and H168 (▧). The line indicates the amino acid stability index (mean stability = 100) for the consensus amino acid at each position. Numbers below the amino acids indicate their locations on the viral envelope. At the top of the graph, + and − indicate the change in charge at the NSI-to-SI switch for each patient.

It appears in the figure as a deviation from the path from the early samples to the late samples.

The delayed outgrowth of the SI-associated V3 variants in the uncultured DNA may be explained as follows. If SI variants are present in the patient's PBMC in low copy numbers, they may become detectable after cocultivation because they outgrow the NSI variants. In vivo, where dominance of a variant does not depend solely on replication rate, detection of the SI-associated variants may be delayed because the cells containing NSI DNA still form a majority, surviving for several months. One clone from an NSI sample (from the uncultured PBMC from sample 4 of patient H479) also had the SI-associated V3 sequence. That the virus isolated from this sample did not produce syncytia could be a sampling effect. The fraction of infected CD4$^+$ cells in asymptomatic people is estimated to be about 1/80,000 (9). A total of 3.5 million cells were used to infect the cultures, so by chance the cell sample used to infect the culture may not have contained this variant. A second possibility is that this V3 region belonged to a defective genome.

Location of the variability of the V3 region. The location of the variability of the V3 region in the uncultured DNA of the two patients is shown in Fig. 4. There are several highly variable amino acid sites (306, 308, 317, 319, 320, and 324). It is obvious that some variability occurred at positions with a high intrinsic stability (5), indicating that replacement of amino acids did not necessarily occur at the inherently most likely positions (as indicated by the lowest amino acid stability). The GPGR sequence, forming the tip of the loop and containing amino acid 311, which Freed et al. (4) found to influence the SI capacity in HeLa T4 cells, also remained intact. There was some variability at position 308, which McKeating et al. (7) and Wolfs et al. (13) found to be an important amino acid for specific antibody recognition of the V3 loop. However, the bulk of the variation was located C terminal of the GPGR sequence in both patients. Comparison of the switches in the two patients showed that only two positions where amino acids changed were common to both events. Of these, one amino acid change occurred in the reverse direction (N→D versus D→N-324). The other position that both events had in common was 320 D→Q-320 in H168, Q→R-320 in H479. Not one identical amino acid change was found for both events. However, we did note that of four amino acid changes that occurred at the NSI-to-SI switch in the viral isolates of H479, three resulted in a higher positive charge of the V3 loop by the introduction of a positively charged arginine (R) and one (N→D) resulted in a lower positive charge. In H168, the net charge difference was +3 (S→R, D→Q, and D→N). Recently Fouchier et al. (3) sequenced V3 regions of 17 SI and 20 NSI isolates and found that the SI phenotype appeared to depend on the charge of two residues at positions 306 and 320 in the V3 loop. If either of these had a positive charge, the isolate almost always had the SI phenotype. Callahan (1) hypothesized that a feature critical for the functionality of the V3 region in mediating envelope binding and subsequent fusion of viral and cell membranes is its high positive charge relative to the rest of gp120. Our data also suggest that the charge of the V3 region may be important to its functionality.

Effect of V3 region mutations on SI capacity and replication rate. To assess the effect of V3 region mutations on SI capacity and the replication rate, HXB-2 viruses with chimeric V3 regions were constructed as described before (6). It was shown that replacement of the original HXB-2 V3 region by the H168 SI-associated V3 region resulted in a virus that had SI capacity in SupT1 cells, whereas replacement with a sequence differing by one amino acid from the H168 NSI-associated V3 region resulted in marked reduction of SI capacity. To test whether the H479-4 V3 region derived from an NSI virus and the H479-6 V3 region from an SI virus (differing in four amino acids) could influence the phenotype of HXB-2, chimeric viruses containing these V3 regions were constructed. The chimeric clone containing the "NSI" V3 region showed markedly less syncytia (although some syncytia were observed) and a lower replication rate than the one containing the "SI" H479 or the original HXB-2 V3 region (Table 2 and Fig. 5). The replication rate of the virus and the appearance of syncytia were somewhat interrelated: even NSI virus produced some syncytia after prolonged

TABLE 2. V3 amino acid sequences and observed SI capacities

Construct	V3 amino acid sequence[a]	Syncytia[b] on day: 3	4	7
H479-4	CTRPNNNTRKGIHI..GPGRAFYATGQIIGNIRQAHC	−	−	+
H479-6	--------------..-----R----RRK--D------	+	+++	++++
HXB-2	-----------R-R-QR-------VTI-K-.--M-----	+	++	++++

[a] SI capacity observed for 7 days in SupT1 cells infected with HXB-2 viruses carrying these V3 regions.
[b] The presence (+) or absence (−) of syncytia observed (the number of plus signs indicates the relative abundance of syncytia).

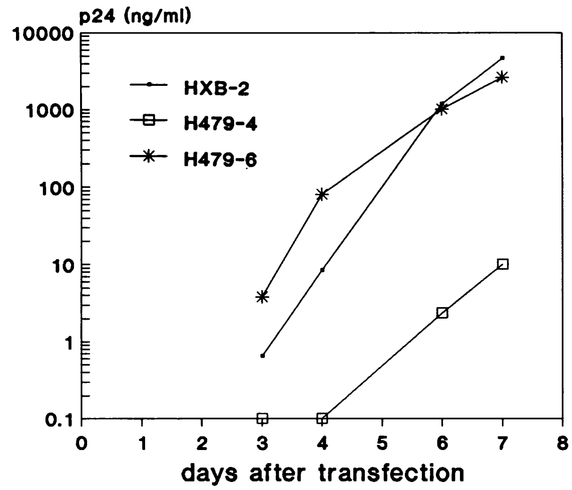

FIG. 5. p24 core antigen production measured during a 7-day period using 5×10^6 SupT1 cells and 5 μg of DNA of the molecular clones bearing the V3 regions shown in Table 2.

culture. This may be because the virus takes longer to reach sufficient concentration, or because mutants with different biological properties have been generated in the culture. However, even when p24 concentrations are kept at comparable or higher levels, the NSI virus produces smaller and less syncytia than the SI virus (5a). Thus, we conclude that changes in the V3 region particular to H479 can influence the SI capacity and replication rate of a virus (6) and, thus, that the changes we observed both in vitro and in vivo represent a phenomenon with biological impact.

Concluding remarks. The evolution of the V3 region is not always gradual, as witnessed by the multidimensional scaling of the sequences from PBMC of patient H479, where the composition and thus the master sequence of the population changes very quickly over time. In patient H168, the evolution is relatively straightforward: only one mutation is acquired and then lost again (T→A-317). In contrast, the pattern we see in patient H479 is that of a "fitness landscape" (2). Several fitness peaks (concentrations of similar variants) are formed by viruses that are almost as fit as the "master" virus but differ appreciably from it. The variants in the fitness lowlands between the peaks generate less offspring and therefore exist only at low frequencies. Evolving from one fitness peak to another, a virus has to cross these lowlands by acquiring mutations that reduce fitness by themselves but are beneficial when combined with others, leading to a new fitness peak. In these crossings, it may be advantageous for a variant to take a longer route (consisting of more mutations) if that enables it to avoid a particularly crippling mutation. The sequences from patient H479's uncultured PBMC may be a case in point. It appeared that an intermediate variant developed (type III) that was not on the shortest path from the NSI- to the SI-associated genotype. The shortest route has a length of six mutations, a route through type III requires two more mutations (step 1 from type I to type III, three nucleotide changes; step 2 from type III to type IV, five nucleotide changes). The virus could be taking a longer route toward the "goal" (the SI-associated variant) because of associated fitness advantages. If it does indeed behave like this, then the calculation of a minimum-length phylogenetic tree may be problematic, because the assumption that the observed changes are the result of a minimum number of mutations may not hold. The computation of evolutionary rate, being based on the same assumption, also becomes hazardous.

We are grateful to Gabriël Zwart and Tom Wolfs for their data and helpful comments along the way and to Ron Fouchier and Hanneke Schuitemaker for sharing unpublished results.

This research was supported by grant 281026 from The Netherlands Foundation for Preventive Medicine and by the "HIV-1 variants" grant from the Dutch Ministry of WVC.

REFERENCES

1. **Callahan, L. N., M. Phelan, M. Mallinson, and M. A. Norcross.** 1991. Dextran sulfate blocks antibody binding to the principal neutralizing domain of human immunodeficiency virus type 1 without interfering with gp120-CD4 interactions. J. Virol. **65:**1543–1550.
2. **Eigen, M.** 1987. Stufen zum Leben. Die fruehe Evolution im Visier der Molekularbiologie. Piper Verlag, Munich, Germany.
3. **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:**3183–3187.
4. **Freed, E. O., D. J. Myers, and R. Risser.** 1991. Identification of the principal neutralizing determinant of human immunodeficiency virus type 1 as a fusion domain. J. Virol. **65:**190–194.
5. **Graur, D.** 1985. Amino acid composition and the evolutionary rates of protein coding genes. J. Mol. Evol. **22:**53–62.

5a. **de Jong, J.-J.** Personal communication.
6. **de Jong, J.-J., J. Goudsmit, W. Keulen, B. Klaver, W. Krone, M. Tersmette, and A. de Ronde.** 1992. Human immunodeficiency virus type 1 clones chimeric for the envelope V3 domain are distinct in syncytium formation and replication capacity. J. Virol. **66:**757–765.
7. **McKeating, J. A., J. Gow, J. Goudsmit, L. H. Pearl, C. Mulder, and R. A. Weiss.** 1989. Characterization of HIV-1 neutralization escape mutants. AIDS **3:**777–784.
8. **Moore, J. P., and P. L. Nara.** 1991. The role of the V3 domain of gp120 in HIV infection. AIDS Suppl. **2:**S21–S33.
9. **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.
10. **Tersmette, M., R. E. Y. de Goede, B. J. M. Al, I. N. Winkel, R. A. Gruters, H. T. 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.
11. **Tersmette, M., R. A. Gruters, F. de Wolf, R. E. Y. de Goede, J. M. A. Lange, P. T. A. Schellekens, J. Goudsmit, H. G. Huisman, and F. Miedema.** 1989. Evidence for a role of virulent human immunodeficiency virus (HIV) variants in the pathogenesis of acquired immunodeficiency syndrome: studies on sequential HIV isolates. J. Virol. **63:**2118–2125.
12. **Wolfs, T. F. W., J. J. de Jong, H. van den Berg, J. M. G. H. Tijnagel, W. J. A. 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.
13. **Wolfs, T. F. W., 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.
14. **Young, F. W., Y. Takane, and R. Lewyckyj.** 1978. ALSCAL: a nonmetric multidimensional scaling program with several differences options. Behav. Res. Methods Instrum. **10:**451–453.
15. **Zwart, G.** Personal communication.