JOURNAL OF VIROLOGY, Mar. 1994, p. 1586–1596
0022-538X/94/$04.00+0

Vol. 68, No. 3

Genomic Cloning and Complete Sequence Analysis of a Highly Divergent African Human Immunodeficiency Virus Isolate

MARLEEN VANDEN HAESEVELDE,[1] JEAN-LUC DECOURT,[1]† ROBERT J. DE LEYS,[1] BART VANDERBORGHT,[1]‡ GUIDO VAN DER GROEN,[2] HUGO VAN HEUVERSWIJN,[1] AND ERIC SAMAN[1]*

Innogenetics N.V., B-9052 Gent,[1] and Institute of Tropical Medicine, B-2000 Antwerp,[2] Belgium

Received 23 August 1993/Accepted 29 November 1993

Analysis of the complete sequence of a human immunodeficiency virus (HIV) isolate (Ant70) obtained from a Cameroonian patient indicates that this virus is the most divergent strain within the HIV-1 family hitherto described. Comparison of the Pol protein, usually highly conserved within the HIV-1 family, shows only about 73% similarity with the HIVmn isolate, whereas for the more variable proteins such as envelope, similarities of 50% or lower are found. The principal neutralizing determinant (V3 loop) and the immunodominant region within gp41 also contain some unusual substitutions, which may have implications for protein function as well as for serological assays based on these regions. Phylogenetic analyses show that this isolate occupies a unique position relative to the human HIV-1 isolates and the recently described SIVcpz virus, indicating that this Cameroonian isolate may provide us with new insights into the origins of the HIV-1 family.

Two distinct types of the human immunodeficiency viruses (HIV) have been described as the responsible etiologic agents of AIDS (5, 9, 10, 21, 22, 56, 61). Although both types, HIV-1 and HIV-2, cause a dysfunction of the immune system and induce similar clinical symptoms in infected persons, they are genetically distinct (11). In addition to the three major genes—*gag*, *pol*, and *env*—common to all retroviruses, genes coding for regulatory proteins are found in the HIV genome (3, 11, 27, 45). These supplementary genes are found in both virus types except for the double-spliced *vpu* and *vpx* open reading frames, which are unique to HIV-1 and HIV-2, respectively.

Epidemiological studies have shown that the prevalence of HIV-2 infection is confined mainly to West Africa (1, 4, 9, 10, 57), whereas HIV-1 infection is a worldwide health problem. Primarily because of the severity of the disease caused by HIV, there has been a large international effort to study the virus. Numerous HIV-1 isolates have been obtained from diverse geographic locations. At present, the HIV-1 family consists of at least five sequence subtypes, depending on the coding sequence considered (50). The variation observed in the nucleic acid sequence is not uniformly distributed throughout the viral genome. Some regions, such as the envelope sequence, are quite variable and can differ by as much as 30% between members of the different HIV-1 subtypes, whereas variability seems to be less pronounced in other parts of the genome.

The closest known relative to the members of the HIV-1 family is not the HIV-2 virus type but, as recently shown, the simian immunodeficiency virus (SIV) SIVcpz, isolated from a chimpanzee (32, 54, 55). The overall genomic organization of this virus is similar to that found in HIV-1, but its sequence is more divergent than any HIV-1 sequence reported thus far.

SIVs have previously also been isolated from African green monkeys (SIVagm), sooty mangabeys (SIVsmm), macaques (SIVmac), and mandrills (SIVmnd) (7, 21, 66, 67). These SIVs, in contrast to the SIVcpz isolate, are more closely related to the HIV-2 subtype (SIVmac, SIVsmm) or form separate groups within the HIV-SIV group (SIVagm, SIVmnd) (50). Recently, the isolation of still another SIV variant from African Sykes' monkeys was described; this virus is genetically equidistant from other primate lentiviruses (28). At present, five known types of primate immunodeficiency viruses can thus be distinguished (28): HIV-1 and SIVcpz making up one type; HIV-2, SIVsmm, and SIVmac forming a second type; and three other types designated SIVagm, SIVmnd, and SIVsyk. The relationship between these five viral groups, based mainly on viral gene sequence similarity, has led to the hypothesis that AIDS viruses arose evolutionarily from a simian virus progenitor (19, 36) and that humans acquired HIV as a result of transspecies infection from nonhuman primates, making SIVcpz a candidate as a possible precursor of HIV-1. In order to substantiate this hypothesis, it is important to continue the search for variant HIV-1 strains that display a more pronounced variability with HIV-1 reference strains and that may be more closely related to the SIV chimpanzee isolate than to the members of the different HIV-1 virus subtypes.

In a previous report we described the isolation of an unusual HIV from a person of West-Central African origin (17). This virus, designated Ant70, was partially characterized and shown to differ immunologically and biochemically from other HIV isolates. Sequence analysis of the viral U3 region revealed an unique enhancer organization not found in other HIV-1 viruses. Ant70 was shown to be more closely related to HIV-1 than to HIV-2 but clearly differed from the HIV-1 reference strains. We now report the molecular cloning of the complete Ant70 genome and present a detailed analysis of the sequence. We show that the Ant70 genomic sequence is more divergent than any HIV-1 sequence published thus far and that the virus is equidistantly related to the HIV-1 types and to SIVcpz.

MATERIALS AND METHODS

Virus isolation and characterization. Virus was isolated by cultivation of peripheral blood lymphocytes from two epidemiologically related HIV-infected individuals (spouses) together with phytohemagglutinin-stimulated lymphocytes from

* Corresponding author. Mailing address: Innogenetics N.V., Industriepark Zwijnaarde 7, Box 4, B-9052 Gent, Belgium. Phone: 32 9 241 08 04. Fax: 32 9 241 07 99.

† Present address: Upjohn N.V., 2670 Puurs, Belgium.

‡ Present address: Fundaçao Oswaldo Cruz, Rio de Janeiro, Brazil.

healthy donors as previously described (17, 44). At the time of virus isolation, one patient was classified in CDCII and the other in CDCIII (17). Once virus was detected in the culture supernatant by measuring reverse transcriptase activity and p24 antigen, the virus was transferred to MT-4 or Molt4 (clone 8) cells by cocultivation (38). The resulting stably infected cell lines were further used to prepare viral RNA and total chromosomal DNA. Ant70-NA was isolated as described for Ant70. Ant70-NA is the virus isolated from the sexual partner of the Ant70 donor.

Isolation of viral RNA and chromosomal DNA. Virus was harvested from 1 liter of culture supernatant by centrifugation and then disrupted in a buffer containing 6 M guanidium chloride. RNA was pelleted through a 5.5 M CsCl cushion, resuspended, and extracted with phenol. Oligo(dT)-primed cDNA synthesis was performed on one-fifth of the RNA by using a commercially available kit as specified by the manufacturer (Amersham Corp., Amersham, United Kingdom). Total cellular DNA was prepared as described by Maniatis et al. (48) from MT4 cells stably infected with Ant70 and Ant70-NA.

Molecular cloning. Standard molecular biological techniques were performed as described previously (48). Restriction endonucleases were used as recommended by the manufacturer (Boehringer GmbH, Mannheim, Germany). The construction of a cDNA library from Ant70 viral RNA in pUC13 was described previously (17). The recombinant plasmid pNEFU3R70 containing *nef* and *LTR* information was sequenced and subsequently used to screen a genomic library.

Chromosomal DNA prepared from MT4 cells persistently infected with the Ant70 virus isolate was used to construct a genomic library as described previously (71). Briefly, the DNA was partially digested with the endonuclease *Sau*3AI (Boehringer). Fragments between 9 and 23 kb, isolated on a 40 to 10% sucrose gradient, were partially filled in and ligated in the commercially available lambda GEM11 vector (Promega, Madison, Wis.) as specified by the manufacturer. The resulting products were packaged by using the PackageneR in vitro packaging system (Promega), and 0.5 M PFU was obtained when plated onto *Escherichia coli* MB406. Plaques were transferred to Hybond-N membranes (Amersham) and screened with the ^{32}P-labeled *Eco*RI fragment of the cDNA recombinant plasmid pNEFU3R70. The hybridization was performed overnight at 65°C in 5× SSPE (1× SSPE is 0.18 M NaCl, 0.01 M sodium phosphate, and 0.02 M EDTA [pH 8.3])–5× Denhardt solution–0.5% sodium dodecyl sulfate (SDS)–100 μg of denatured salmon sperm DNA per ml. Filters were subsequently washed four times (twice for 20 min in 2× SSC–0.1% SDS and twice for 20 min in 0.2× SSC–0.1% SDS) (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) at room temperature and autoradiographed at −70°C with an intensifying screen. Clones displaying positive signals were plaque purified and propagated on *E. coli* MB406. Ant70 information from two recombinant lambda phages (GEM70S4-16 and GEM70S7-17) was subcloned in the pUC12 and pUC13 vectors and further characterized by restriction mapping.

DNA amplification. PCR (40) was used to complete the Ant70 sequence information with the same cellular DNA utilized to construct the genomic library. The missing part of the genome was amplified with the HIV-1-specific primer SE41-1 (5′ GGGTTCTTGGGAGCAGCAGGAAGCACTAT GGGCG 3′) and the Ant70-specific primer SEN1 (5′ GGATC CAGTCCCCCCTTTTCTTTTAAAAA 3′) positioned in *gp41* and *nef*, respectively. To facilitate the subsequent cloning of the PCR products, a *Bam*HI restriction site was incorporated into the SEN1 oligonucleotide. For the amplification reactions the DNA Thermal Cycler (Perkin-Elmer Cetus, Emeryville, Calif.) was used. The reaction mixture consisted of 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 0.001% (wt/vol) gelatin, 200 μM each dNTP, 100 pmol of each oligonucleotide primer, and 2.5 U of *Taq* polymerase in a final volume of 100 μl. Each cycle consisted of a 1-min denaturation step at 94°C, a 2-min annealing step at 50°C, and a 135-s elongation step at 72°C. After 30 cycles, the samples were extended for another 10 min at 72°C. The resulting PCR products were detected by Southern blot hybridization with an HIV-1 ^{32}P-labeled probe containing HIV-1 envelope information. The DNA derived from the PCR was then chloroform extracted, *Bam*HI digested, purified on agarose gel, and cloned into *Eco*RV- and *Bam*HI-digested plasmid pSK$^+$, giving p70TM4.

The long terminal repeat (LTR) region of Ant70-NA was amplified with HIV-1-specific primers SEN1B (5′ TTTTTA AAAGAAAAGGGGGGACTGGA 3′) and SK30 (5′ GGTCT GAGGGATCTCTA 3′). The same reaction conditions as for the amplification of gp41 of Ant70 were used except for the temperatures, which were 1 min each at 95, 45, and 72°C. The amplified product was cloned into the pUC13 vector by using an internal *Hin*dIII restriction site of the amplified product, giving pNALTR0.

The region of the Ant70-NA envelope gene encompassing the V3 loop portion was amplified by using the Ant70-specific primers p70S7-05 (5′ ACTTGTACACATGGCATTAGGCC 3′) and p70S7-11 (5′ AATTGTTACATTGTTGTTGCTGC 3′), and the DNA was subsequently cloned as a blunt-ended fragment into the *Sma*I-digested pSK$^+$ vector, giving pNAV3-4.

Sequencing and genome analysis. DNA sequencing was performed on double-stranded DNA by using the dideoxy-chain termination procedure (60) with a Sequenase Sequencing Kit (United States Biochemicals, Cleveland, Ohio). Part of the genome was sequenced by subcloning the viral information; the remainder was sequenced by using specific oligonucleotides to prime DNA templates. Both strands were sequenced, and the data were processed by using a software package (IntelliGenetics, Mountain View, Calif.). Nucleotide sequence alignments were performed by using the PCgene software package; protein sequences were aligned by using the ALIGN program (version 1.01).

Phylogenetic tree analysis. Phylogenetic analyses were conducted with the aid of the PAUP 3.0 (Phylogenetic Analysis Using Parsimony; David Swofford) program for the *pol* and *env* genes (63). The trees are the result of a heuristic search performed by using random addition sequence with 10 replications. The MULPARS option and subtree-pruning-regrafting (SPR) branch swapping were used. The maximum number of equal length trees (MAXTREES) saved at each step was 300. SIVmndgb1 was used as the outgroup.

Nucleotide sequence accession number. The complete sequence of the Ant70 genome was deposited at the GenBank data base under accession number L20587. The V3 sequence of the Ant70-NA variant has the GenBank accession number L23119.

RESULTS

Molecular cloning of the complete Ant70 information. An oligo(dT)-primed library was constructed starting from Ant70 viral RNA. Screening under low-stringency conditions with a probe containing the 5′-end proviral HIV-1 sequences yielded one cDNA clone (pNEFU3R70) which contained an insert of 900 bp corresponding to the 3′ end of the Ant70 viral RNA including the *nef* gene and the 3′ U3 and R regions. This

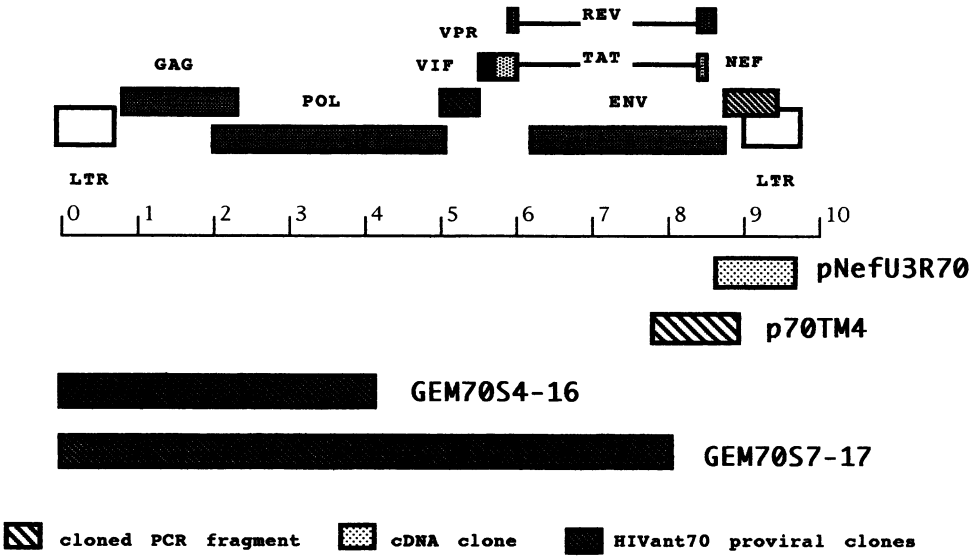

FIG. 1. Localization of different cloned fragments of the Ant70 genome relative to the standard HIV-1 genome. The positions of two lambda clones (GEM70S4-16 and GEM70S7-17) as well as the positions of the cDNA clone and the PCR-amplified fragment are shown. Sizes are indicated in kilobases.

cDNA fragment was further used to screen a lambda phage genomic library constructed from total DNA of Ant70-infected MT4 cells.

Five recombinant phages from this library contained viral information which extended from the 5′ LTR diversely into *gag*, *pol*, or *env*. Phages GEM70S4-16 and GEM70S7-17 with inserts of approximately 4,000 and 8,000 bp, respectively, were retained for further analysis. Their information was nonoverlapping with the cDNA clone, and the Ant70 sequence was completed by PCR. Primers based on a conserved region were designed in the HIV-1 *env* gene and on the Ant70-specific *nef* sequence. The 1144-bp amplification product was cloned in the pBluescript SK^+ plasmid, giving p70TM4.

The complete nucleotide sequence of Ant70 was finally obtained from lambda GEM70S4-16 and completed for the missing parts by sequences from lambda GEM70S7-17, p70TM4, and pNEFU3R70. The positions of these fragments are shown relative to a standard HIV-1 genome in Fig. 1. Sequence information obtained from overlapping regions of different cloned fragments was always completely identical.

Similarity displayed by Ant70 to members of the HIV-1 group. The complete Ant70 genome is 9,785 nucleotides in length and has the genetic organization 5′ LTR *gag-pol-vif-vpr-tat-rev-vpu-env-nef* LTR 3′, typical for members of the HIV-1 family. The nucleic acid and protein sequence homologies between Ant70 on the one hand, and HIV-1 (African and North American isolates), SIVcpz, and HIV-2 on the other, are shown in Table 1. An average overall nucleic acid sequence similarity of 70% was found with prototype HIV-1 strains (Lai, Mn, Eli, Mal). As is the case with all HIV-1 isolates, the *gag* and *pol* genes of Ant70 are more conserved than the *env* gene and the genes coding for regulatory proteins. The similarity

TABLE 1. Comparison of Ant70 and HIV-1/HIV2 genes and proteins

Isolate	% Identical residues between Ant70 and isolate for:																	
	Gag		Pol		Vif		Vpr		Vpu		Tat		Rev		Env		Nef	
	NA[a]	AA[a]	NA	AA	NA	AA	NA	AA	NA	AA	NA	AA	NA	AA	NA	AA	NA	AA
North American HIV-1 isolates																		
HIVlai	71	68	75	74	68	62	77	77	—[b]	27	—	38	—	46	62	50	63	59
HIVmn	71	68	74	73	69	60	77	61	—	27	—	42	—	43	62	50	64	53
African HIV-1 isolates																		
HIVu455	68	68	74	75	70	64	76	79[c]	—	24[d]	—	50	—	52	61	48[c]	65	61
HIVeli	70	67	75	74	67	60	77	82	—	33	—	45	—	48	62	52	61	60
HIVmal	72	70	75	74	68	61	74	79	—	28	—	43	—	46	61	53	61	59
Chimpanzee isolate																		
SIVcpz	72	70	76	77	67	61	78	78	—	39	—	46	—	51	63	53	63	59
HIV-2 isolate																		
HIV2Rod[e]	—	54	—	56	—	35	—	45	—	—	—	38	—	27	—	34	—	38

[a] NA, nucleic acids; AA, amino acids.
[b] —, the similarity was either too low or not relevant (e.g., Vpu and HIV-2).
[c] Premature stop ignored.
[d] To obtain the Vpu protein, a frameshift was overruled.
[e] For HIV-2Rod, only the protein similarity was calculated.

with HIV-1 proteins varies from 74% for Pol to 50% for Env and to about 30% for Vpu (Table 1). By comparison, one of the most divergent strains within the HIV-1 family (HIVmal) shows average protein similarities from 90% in Pol, over 86% in Gag, and 70% in Vpu relative to prototype strains such as HIVlai or HIVmn. These similarities were calculated as in Table 1. This observation leads to the conclusion that Ant70 is a highly divergent strain within the HIV-1 family.

Most of the DNA sequence diversity in Ant70 results in amino acid alterations of the viral proteins. The similarity percentages obtained for the coding sequences and their corresponding proteins are quite similar for the conserved domains of the genome (*gag*, *pol*, and *vpr*) but differ for the more variable parts (*env*, *vif*, *vpu*, *tat*, *rev*, and *nef*). The unusual divergence of Ant70 is reflected not only in regions that are known to be highly variable but also in the considerable number of amino acid residues which, although normally considered invariant within the HIV-1 strains, are found to be different in this isolate. For the Gag and Pol proteins, up to 21 and 17%, respectively, of "invariant" residues (50) are changed in the corresponding Ant70 proteins. This is even more pronounced in the p17 part of the Gag protein, where 28% of residues are changed, a percentage which is even higher than the one obtained for the variable gp120 protein (26%). These results illustrate that Ant70 is the most divergent HIV-1 strain described to date.

Analysis of individual Ant70 genes. To better appreciate the degree of divergence displayed by Ant70, we investigated regulatory elements in the genome and functionally relevant domains in Ant70 proteins.

The complete LTR region (U5, R, U3) is 664 nucleotides long, and the sequence of the U3R region of the Ant70 viral RNA has already been described (17). In addition to the presence of the principal regulatory elements such as the TATA box, the Sp1-binding sites, the NF-κB-binding sites, and the TAR region, the U3 region was shown to contain a unique 18-bp insert sequence which separates the two NF-κB sites. This insert sequence is highly homologous to two elements located just upstream from these two NF-κB-binding sites and is strongly conserved. In addition to its presence in the Ant70 viral RNA (cDNA library), as shown previously (17), it is also found in the genomic DNA (genomic library) and in the LTR region of the Ant70-NA isolate.

Investigation of the untranslated region upstream from the *gag* information showed that the primer-binding site is perfectly conserved. This untranslated domain is somewhat longer for the Ant70 genome as a result of a 25-bp insertion found upstream from the RNA-packaging signal sequence. A 25-bp insert sequence is present at the same location in the SIVcpz genome (32). The usually well-conserved RNA packaging signal sequence AAAAAATTTTTGACTAGCGGA, essential for packaging unspliced genomic RNA into infectious particles (2, 12), is only partially conserved in the Ant70 genome as TAATTTTGCTGG. This indicates that conservation of the complete signal sequence is not absolutely required to obtain packaging of viral RNA.

gag sequences coding for the p17, p24, p9, and p7 proteins are similar in length to those previously characterized in HIV-1 proviruses. Like all other HIV-1 isolates, Ant70 contains a myristoylation site at the amino-terminal glycine of its p17 matrix protein (6). Whereas the RNA packaging signal is only partially conserved, the zinc finger motifs (Cys-X_2-Cys-X_4-His-X_4-Cys) of the p7 nucleocapsid protein involved in the packaging process (25) are both very well conserved, indicating that these structural features of the p7 protein are more important than the exact sequence conservation of the packaging sequence.

The overlap between the *gag* and *pol* open reading frames starts 72 nucleotides earlier than commonly found in HIV-1. Optimal Gag-Pol translational frameshifting in HIV depends on the presence of an A+T-rich heptanucleotide sequence, followed by G+C-rich inverted-repeat sequences which can form a stem-loop structure (53). Investigation of these regions in Ant70 shows that the frameshift site (TTTTTTA) is completely conserved. The inverted-repeat sequences, however, are shorter than usual, consisting of only 8 nucleotides instead of the usual 11, but retain the potential to form a stem-loop structure.

The *env* gene has a coding capacity for 863 amino acid residues and shows the typical pattern of alternating constant and hypervariable domains. The total envelope protein gp160 contains as many as 29 potential N-linked glycosylation sites (sequons), 15 of which are considered as being highly conserved among HIV-1 strains (underlined in Fig. 2) (16, 42, 50). Of these 15 sites, only 10 are conserved within the Ant70 glycoprotein. On the other hand, four new glycosylation sites are generated at locations not found in other HIV-1 strains (boxed in Fig. 2). The distribution of these sites is different for the external and the transmembrane parts of the protein. In gp120, only 7 of the 11 HIV-1 major conserved sites (64%) are preserved, while in gp41, 3 of the 4 conserved sites (75%) are present (Fig. 2). The new sequons are all localized in gp120.

The number of cysteine residues present in the envelope protein also warrants some comment. The 19 cysteine residues considered as highly conserved in the HIV-1 gp120 protein are also found in the Ant70 surface protein (boxed in Fig. 2). In addition, two cysteines are found in the V4 domain, where most other HIV-1 isolates do not contain cysteines. Since cysteines greatly contribute to the overall structure of proteins, the presence of these additional residues could influence the folding and epitope structure of the viral envelope in this region (65). On the other hand, a rather well-conserved cysteine in the C-terminal half of the transmembrane protein is not found in the Ant70 protein. Both cysteines enclosing the major immunodominant site in gp41 are conserved, and two extra cysteine residues are found further downstream in close proximity to each other.

The amino acid residues known to be involved in gp120-CD4 binding (marked with black dots in Fig. 2) (52) are well conserved except at position 427, where a glycine replaces a valine, and at position 391 where serine is replaced by tyrosine.

The principal neutralizing determinant within the third hypervariable region is known to be highly divergent in HIV-1 isolates with a tendency to conservation of the central part of the loop (34). Compared with the consensus V3 sequences A, B, and D as proposed by Myers et al. (50), the V3 sequence of Ant70 differs in 60 to 70% of the amino acids (Fig. 3A), the highest similarity (40%) being found with the Uganda-like sequence (consensus D). In most HIV-1 isolates worldwide, the central region of this domain contains a conserved sequence (GPGQ) (50) which is not found in Ant70. Instead a GPMAW motif is found here, which appears to be unique within the HIV-1 family. Because the V3 domain is an important target for neutralizing antibodies (33, 49, 69), and because of the high divergence within known V3 sequences, the principal neutralizing determinant of the related Ant70-NA virus was determined and compared with the V3 sequence of Ant70 (Fig. 3B). A variation of 28% was found between the two V3 domains, situated mainly in sequences flanking the central portion of the loop structure, whereas the

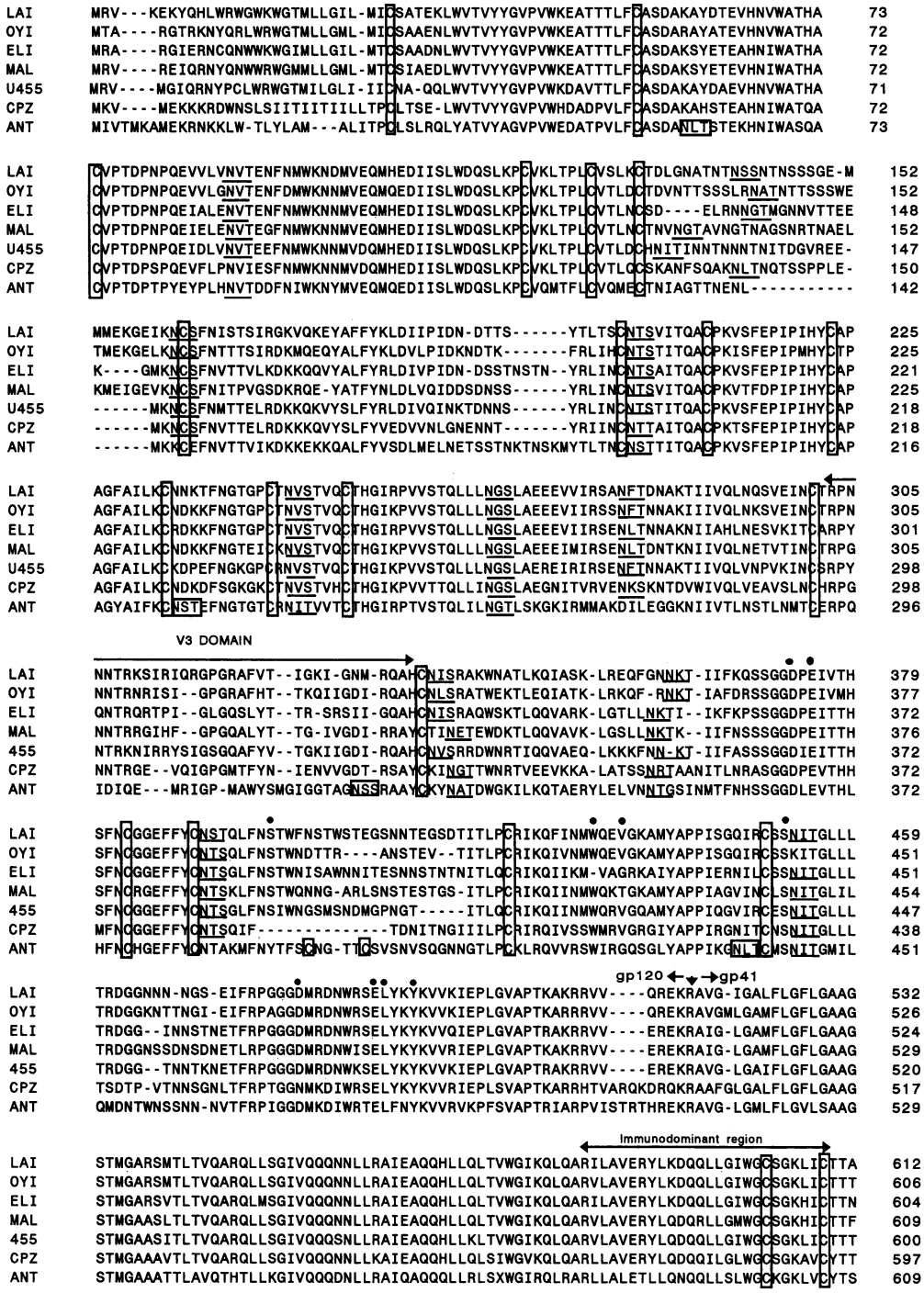

FIG. 2. Alignment of several HIV-1 Env protein sequences with the Env protein of Ant70. The putative conserved N-glycosylation sites are underlined, and the cysteine residues are boxed.The V3 loop region is overlined, and the residues involved in CD4 binding are indicated (black dots). In gp41, the major antigenic site is shown. N-glycosylation sites unique to Ant70 are boxed. Numbers to the right of the sequence indicate the number of amino acids for each isolate. In the U455 isolate, the premature stop codon was ignored and represented as X. ANT designates the Ant70 protein, and CPZ designates the SIVcpzgab protein.

GPMAW sequence was found to be unchanged in Ant70-NA. These results are in keeping with data obtained from studies on the variability in this domain, where the central part of the loop has been shown to be much less susceptible to mutations than are the flanking regions in strains isolated from sexual partners or even from the same individual at different stages of the disease (29, 69, 70). The general structural features of the V3 loop have been investigated by using the Chou and Fasman rules (8) (PCGENE; IntelliGenetics): a striking difference emerged between the V3 loop of the well-known HIV-1 isolates and those of Ant70 and Ant70-NA. Whereas the V3 of most HIV-1 isolates displays a high probability for a β-turn at

```
LAI   VPWNASWS-NKSLEQIWNNMTWMEWDREINNYTSLIHSLIEESQNQQEKNEQELLELDKWASLWNWFNITNWLWYIKIFI   691
OYI   VPWNASWS-NKSLNEIWDNMTWMQWEREIDNYTHLIYTLIEESQNQQEKNEQELLELDKWAGLWSWFSITNWLWYIRIFI   685
ELI   VPWNSSWS-NRSLNEIWQNMTWMEWEREIDNYTGLIYSLIEESQTQQEKNEKELLELDKWASLWNWFSITQWLWYIKIFI   683
MAL   VPWNSSWS-NRSLDDIWNNMTWMQWEKEISNYTGIIYNLIEESQIQQEKNEKELLELDKWASLWNWFSISKWLWYIRIFI   688
455   VPWNSSWS-NKSQEDIWNNMTWLQWEKEISSYTGIIYQLIEESQNQQEKNELDLLALDKWANLXNWFNISNWLWYIRLFV   679
CPZ   VPWNNSWPGSNSTDDIWGNLTWQQWDKLVSNYTGKIFGLLEEAQSQQEKNERDLLELDQWASLWNWFDITKWLWYIKIFL   677
ANT   VKWNRTWIGNES---IWDTLTWQEWDRQISNISSTIYEEIQKAQVQQEQNEKKLLELDEWASIWNWLDITKWLWYIKIAI   686

LAI   MIVGGLVGLRIVFAVLSIVNRVRQGYSPLSFQTHLPTPRGP-DRPEGIEEEGGERDRDRSIRLVNGSLALIWDDLRSLCL   770
OYI   IIVGGLVGLRIVFAVLSIVNRVRQGYSPLSFQTRLPTQRGP-DRPEGIEEEGGERDRDRSGRLVDGFLALIWDDLRSLCL   764
ELI   MIIGGLIGLRIVFAVLSLVNRVRQGYSPLSFQTLLPAPRGP-DRPEGTEEEGGERGRDRSVRLLNGFSALIWDDLRSLCL   762
MAL   IVVGGLIGLRIIFAVLSLVNRVRQGYSPLSLQTLLPTPRGPPDRPEGIEEEGGEQGRGRSIRLVNGFSALIWDDLRNLCL   768
455   IIVGGLIGLRIVFTVLSIINRVRQGYSPLSFQTLAPIPEGL-GRPGRIEEEGGEQGKDRSIRLVSGFLAIAWDDLRNLCL   758
CPZ   MAVGGIIGLRIIMTVFSVVRRVRQGYSPLSLQTLIPVQREQ-GRLGEIDEGGGEQDRSRSVRLVEGCLPLIWDDLRNLGI   756
ANT   IIVGALVGVRVIMIVLNIVKNIRQGYQPLSLQ-IPNHHQEEAGTPGRTGGGGGEEGRPRWIPSPQGFLPLLYTDLRTIIL   765

LAI   FSYHRLRDLLLIVTRIVELLGRR-------GWEALKYWW---NLLQYWSQELKNSAVSLLNATAIAVAEGTDRVIEVVQG   840
OYI   FSYHRLRDLILIVARIVELLGRR-------GWEVLKYWW---NLLQYWSQELKNSVISLLNATAIAVAEGTDRVIEIVQR   834
ELI   FSYHRLRDLILIAVRIVELLGRR-------GWDILKYLW---NLLQYWSQELRNSASSLFDAIAIAVAEGTDRVIEIIQR   832
MAL   FSYHRLRDLLLIATRIVELLGRR-------GWEALKYLW---NLLQYWGQELKNSAISLLNTTAIAVAECTDRVIEIGQR   838
455   FSYHRLRDFALIVARAVELLGRSSLKGLRLGWEGLKYLW---NLLLYWGRELKISAITLLDAVAVAVAGWIDRVIEIGQT   835
CPZ   WSYQSLTSLACNVWRQLKTLGHLILHSLRLLRERLCLLG---GIIQYWGKELKISAISLLDATAIAVAEGTDRIIEAFQV   833
ANT   WTYHLLSNLASGIQKVISYL-RLGL--WILGQKIINVCRICAAVTQYWLQELQNSATSLLDTLAVAVANWTDGIIAGIQR   842

LAI   ACRAIRHIPRRIRQGLERILL   861
OYI   AYRAFLNIPRRIRQGLERALL   855
ELI   ACRAVLNIPRRIRQGLERSLL   853
MAL   FGRAILHIPRRIRQGFERALL   859
455   IGRAILNIPRRIRQGLERALL   856
CPZ   TLRIIRNIPRRIRQGLERALL   854
ANT   IGTGIRNIPRRIRQGLERSLL   863
```

FIG. 2—*Continued.*

the central proline, no such configuration is predicted at this position for the Ant70 isolates, a feature also found for some African strains (e.g., HIVeli and some Ugandan strains) (Fig. 3C). Although these predictions have to be considered with caution, one can nevertheless assume that an important structural difference between the two types of V3 loops is likely to occur.

The V3 loop peptide is known to be involved in syncytium formation by HIV in cell culture, and the computed charge of the V3 peptide has been described as a measure of the syncytium-inducing capacity of the virus (20). Calculations (CHARGEPRO, PCGENE) of the V3 loop charge of both Ant70 viruses show that for Ant70, a small negative value (−0.09) is found, whereas Ant70-NA has a rather high positive charge at pH 7 (4.91). According to the authors, this indicates that the Ant70 virus is likely to be non-syncytium inducing and nonmonocytotropic, whereas the Ant70-NA isolate would have the opposite properties. Although the monocytotropic properties of both viruses have not been studied, the syncytium-inducing character of both viruses is exactly as predicted from these calculations (unpublished observations).

Two unique substitutions are seen within the immunodominant epitope (WGCSGKLIC) of the transmembrane protein (24): a lysine replaces the serine, and a valine is found instead of the isoleucine preceding the second cysteine (Fig. 2).

The Rev-responsive element is highly divergent compared with Rev-responsive element sequences of other HIV-1 strains (13) since only 68% of the nucleotides are conserved. Despite the multiple differences found in the sequence, the Rev-responsive element of Ant70 retains the capacity to form the complex RNA stem-loop structure (calculated ΔG, −124.8 kcal [−522.2 kJ]) necessary for efficient transport of unspliced mRNAs from the nuclear compartment to the cytoplasm of the infected cell (15, 47).

Like all members of the HIV-1 group, Ant70 contains genes encoding the regulatory proteins Vif, Vpr, Tat, Rev, Vpu, and Nef. The deduced Vif, Vpr, and Nef proteins are comparable in length to those of previously described HIV-1 isolates and show no remarkable difference apart from their low similarity.

The protein encoded by the double-spliced *tat* gene is 10 amino acids longer than most Tat proteins owing to an elongated second exon. The acidic and cysteine-rich domains are both perfectly conserved. In addition to the seven cysteine residues in the functional domain (58, 59), three extra cysteines are found at the carboxy end of the protein. This probably has no impact on the activity of the protein since it is known from studies with other isolates that only the first 57 amino acids are required for full Tat function (35). However, this domain has recently been implicated in the regulation of a major histocompatibility complex class I promoter activity (31). The basic domain of the *trans*-activation protein is much less arginine rich than usual and lacks three of the six arginine residues (RKKRGRP instead of RKKRRQRRR); the nuclear localization signal (RKKR) is perfectly conserved (14).

In contrast to Tat, the polypeptide sequence of Rev is shorter than usual (101 amino acids instead of 116). The domain important for nuclear localization of the protein (14) lacks 2 of the 10 arginine residues; however, the acidic region is well conserved.

The most divergent among these small regulatory proteins is undoubtedly Vpu. Despite its normal length of 85 amino acids, the sequence shows nearly no similarity to other Vpu sequences (Fig. 4). This is not unusual, since this protein is generally rather divergent within the HIV-1 group. The poor conservation of the ERAEDSGNESEG peptide subunit in Ant70 (EIRDDSDYESNG), a region which is highly conserved among different HIV-1 isolates and found even in the HIV-1-related chimpanzee virus SIVcpz, is, however, noteworthy (62). It should nevertheless be mentioned that the two serine residues in this domain (underlined), which are important sites for phosphorylation (62), are conserved. In contrast to its high degree of divergence, the hydropathy profile (30) of

A

```
Ant70  CERPQIDIQE-MRIGP-MAWYSMGIGGTAGNSSRAAYC  36
ConsA  CTRPNNNTRKSVHIGPGQAFYA---TGDIIGDIRQAHC  35
ConsB  CTRPNNNTRKSIHIGPGRAFYT---TGEIIGDIRQAHC  35
ConsD  CTRPYNNTRQRTHIGPGQALYT---T-RIIGDIRQAHC  34
       *.** ....  .***  * *.  .     ..*.*.*
```

B

```
V3ANT70    - CERPQIDIQEMRIGPMAWYSMGIGGTAGNSSRAAYC -36
             : ::    : :::::::::::: :: :  ::::::::
V3ANT70NA  - CKRPAMQVQKMRIGPMAWYSMRIGSTGENSSRAAYC -36
```

C

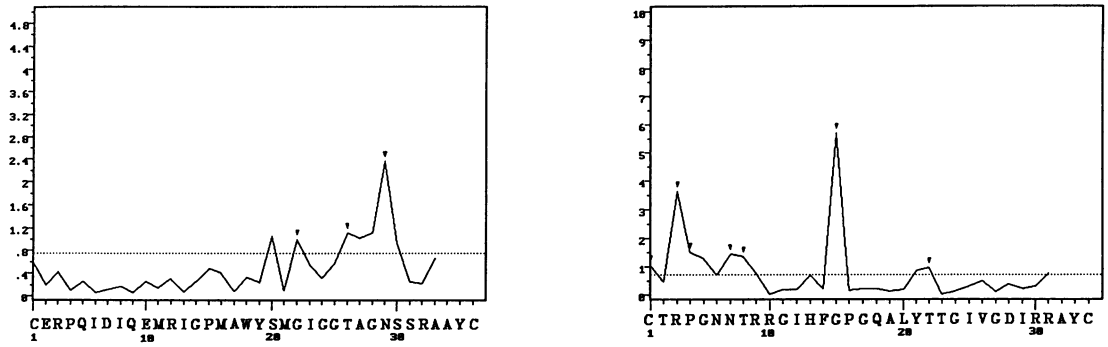

FIG. 3. Comparison of the V3 loop region of the Ant70 isolate with the V3 consensus sequences as defined by Myers et al. (50) (A) and between Ant70 and the closely related isolate Ant70-NA (B). Gaps were introduced for maximum similarity. The Ant70-specific N-glycosylation site is underlined. The β-turn probability was calculated by the method of Chou and Fasman (8) and plotted along the peptide sequence for Ant70 (left) and for HIVmal (right) (C). Small arrowheads indicate probabilities above the cutoff (dotted line).

the Ant70 Vpu protein is quite similar to the Vpu profiles of other HIV-1 isolates.

Phylogenetic relationship of Ant70. Phylogenetic tree analysis for the three major genes, *gag*, *pol*, and *env*, was performed to illustrate the relationship of Ant70 to previously characterized HIV-1 strains. Prototype HIV-2 and SIVsmm strains are also included in the analysis. Figure 5 shows the phylogenetic tree obtained for the *pol* region, but the topology of the tree was quite similar, regardless of the gene considered. Until now, the SIVcpz isolate branched before any of the HIV-1 strains. The finding that Ant70 diverges more from other HIV-1 strains that the SIVcpz isolate does and thus branches before any of these viruses is quite unexpected and relates to a topology found for some HIV-2 strains (23).

Prevalence of Ant70. To investigate the prevalence of Ant70-like isolates in different geographical areas (Central and West Africa, Brazil, Europe), we performed a small serological study (to be published elsewhere). A panel of 326 HIV-1-positive serum samples were tested for their reactivity with a peptide (QIDIQEMRIGPMAWYSMGIGG) representing the V3 domain of Ant70. Only the sera obtained from the patients from whom the Ant70 and the Ant70-NA were isolated reacted with the Ant70 peptide. The presence of antibodies to the V3 loop in these sera was confirmed by using a peptide based on the consensus V3 sequence (Fig. 3A, consB). However, the spread of this virus subtype may presently be limited to the Cameroon region since 5 to 10% of HIV-positive serum samples from that geographical area were recently found to be reactive with the Ant70 V3 peptide (51).

```
LAI    MQPIQI-AIAALVVAIIIA--IVVWSIVII-EYRKILRQRKIDRLIDRLI----  46
MAL    IQPLVILAIVALVVTLIIA--IVVWTIVFI-EIRKIRRQRKIDRLIDRIR----  46
ELI    MQPLGIIAIAALVVAIILA--IVVWTIVFI-EYRRIKKQRRIDCLLDRIT----  47
CPZ    M-TLLVGLVLILV-GLI-AWNICIWGYIIKWGYRRYKRHRLETEI-ERLNLILR  50
ANT    MHHRDLLA-IIIISA-LLFINVILWGFILRKYLEQKEQDRKEREILERLRRI-R  51
                              *          *       *

LAI    ERAEDSGNESEGEISA-LVEMGVEMGHHAPWDID--DL   81
MAL    ERAEDSGNESEGDTEE-LSKL-VEMGHDAPWDVD--DL   80
ELI    ERAEDSGNESEGDREK-LSKL-VEMGHHAPWDID--DL   81
CPZ    ERAEDSGNESNGEEEERLEQLIHNYNHNNHFANPMFDL   88
ANT    EIRDDSDYESNGEEEQEVMDLVL----SHGFDNPMFEP   85
```

FIG. 4. Alignment between Vpu proteins of different HIV-1 strains. The conserved motif as well as the important serine residues are indicated (arrows). The Ant70 strain has a full-length Vpu protein, unlike some other HIV-1 strains (e.g., HIVmn). ANT indicates the Ant70 sequence, and CPZ indicates the SIVcpzgab protein. Conserved residues are marked with an asterisk. Numbers to the right indicate the number of amino acids in each protein.

DISCUSSION

A detailed genetic analysis of the Ant70 isolate shows that it is the most divergent HIV-1 strain described until now, a conclusion based on the unusually high degree of variation found throughout its entire genome. Moreover, this observation is not confined only to regions known to be variable in HIV-1, since a fraction of "invariant" residues are changed as well.

The LTR region of the HIV genome is known to regulate viral transcription, and several transcription factors have been shown to interact with distinct sequences within this region. A previous analysis of the U3 and R region of the Ant70 LTR region showed that the major regulatory elements are conserved (17). In addition to these elements, an insert fragment resembling two sequences situated upstream from the two NF-κB sites was shown to be present in the U3 region of the viral RNA genome. Investigation of this region in Ant70 proviral DNA has confirmed the presence of this insertion.

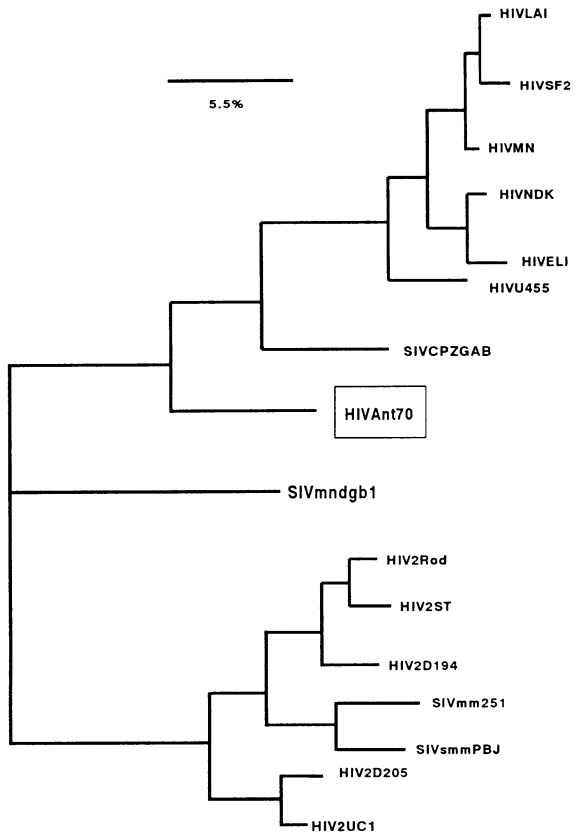

FIG. 5. Phylogenetic tree analysis of the HIV-1 and HIV-2/SIV *pol* coding sequence. Analysis was done as described in Materials and Methods on 1,242 sites, of which 795 were varied. Horizontal branch lengths, indicative of single-base changes between sequences, can be read by using the scale bar. The vertical distances are for clarity only. This tree was congruent with similarly constructed trees for other coding regions, with or without the Ant70 sequence. SIVmndgb1 was the designated hypothetical ancestor (63).

Analysis of the same region in Ant70-NA, the virus isolated from the sexual partner, shows that this triplicated sequence was largely conserved when the virus was transmitted horizontally. These results clearly indicate that the presence of this sequence insertion is not due to a cloning artifact. The occurrence of CTG duplications upstream from the NF-κB sites as well as the presence of a fourth SP1 site has been described for other HIV-1 isolates (37). In fact, duplication of this short DNA sequence has also been observed in the LTR of the SIVcpz isolate, albeit in a somewhat different position from that in the other HIV-1 strains (50). The insertion in Ant70 clearly differs from these in two aspects: first, Ant70 and Ant70-NA contain the repeated CTG motif in triplicate instead of the usual duplication; second, the third copy is positioned between the two NF-κB sites, which is never the case in the other isolates, where both tandem copies are found adjacent to each other upstream from the NF-κB sites. Although no regulatory role has yet been demonstrated for these repeated motifs, the strong conservation leads us to speculate that their presence must somehow be beneficial to the virus.

Some of the regulatory proteins of Ant70 show no striking differences from proteins of previously described HIV-1 viruses, whereas others, such as the Tat protein, display important differences in their primary sequence (e.g., additional cysteine residues). The regulatory proteins Tat and Rev are less arginine rich in their nuclear localization domains, indicating that some variation can be tolerated without jeopardizing their function. This is also illustrated in the equine infectious anemia virus, in which the basic region was also shown to have a much lower arginine content than in HIV-1 or HIV-2 (18).

The Vpu protein of HIV-1 isolates is known to be divergent but normally contains a highly conserved peptide. In Ant70, however, this peptide is remarkably different except for both serine residues, which are well conserved. Since these residues are known to be phosphorylated, their conservation is indicative of the importance of this posttranslational modification for protein function. Although the role of the conserved motif in Vpu function remains unknown, the changes observed in this region in Ant70 may be related to changes in other proteins. Since Vpu has been described as being involved in the maturation of the viral glycoproteins (68), the variation observed in this otherwise conserved region may be an adaptation to alterations in the membrane proteins. In contrast to the very variable primary sequence, the general hydropathic profile of Vpu is maintained, preserving the putative membrane-anchoring domain which has been postulated recently (46).

Among the structural proteins of Ant70, the envelope proteins are undoubtedly the most divergent. In particular, the outer membrane protein shows major differences. Of the 15 highly conserved potential N-linked glycosylation sites, 4 are missing in gp120, while 4 of a total of 29 are rare or even unique compared with sites in prototype HIV isolates. These alterations in the glycosylation pattern can influence the structure and function of this polypeptide. Indeed, mutagenesis experiments affecting asparagine residues in potential glycosylation sites were shown to have significant effects on envelope functions in other HIV-1 strains (42). More precisely, the presence of an extra sequon at the C-terminal end of the V3 loop may have profound implications for the structure and function of this important region (see also below).

Although the envelope protein is known to be extremely variable among HIV-1 isolates, its cysteine residues are usually well conserved. This is also the case for the Ant70 isolate, indicating that the overall structure of the glycoproteins probably is similar to that found for other isolates. However, the presence of two extra cysteine residues in the V4 domain may confer a local difference in structure, possibly by forming a small loop upon linking both residues via a disulfide bridge, since it is known that all cysteines within gp120 are involved in intramolecular bridges (43). The occurrence of extra cysteine residues in gp120 of SIV isolates has been described before, but in these cases they are all localized near the first variable domain (28). The conservation of both cysteine residues flanking the major epitope of gp41 was expected since it has been shown that these residues are essential for the processing of gp160 (64). It is remarkable that the third conserved cysteine in gp41 is missing in the Ant70 isolate, whereas two extra cysteines are present further downstream. Both features are also present in the SIVcpzgab isolate.

The most striking observation is the extremely divergent principal neutralizing determinant, which differs considerably in the regions flanking the central region of the domain but surprisingly also in the central tip of the loop, which usually remains rather constant among HIV-1 isolates (GPGQAF). This sequence is altered to GPMAW, which is, to our knowledge, not found in any other HIV-1 V3 domain, except in the related Ant70-NA isolate. Structure predictions for the V3 region by using the Chou and Fasman rules (8) indicate that, at

variance with most HIV-1 isolates, the tip of the Ant70 loop has no tendency to support the formation of a β-turn.

The degree of variation between Ant70 and Ant70-NA V3 regions that was observed is in keeping with the variability found in this domain between viruses isolated from sexual partners. The sequences flanking the central domain were more susceptible to variation, while the tip of the V3 loop was found to remain constant (29, 69). The presence of an extra N-glycosylation site at the C-terminal end of the V3 loop in both isolates, as well as the aberrant sequence at the tip of the loop, may have implications for the overall structure of this immunologically and functionally important region.

Since the V3 loop is known to induce isolate-specific antibody responses in infected patients, peptide serology was used to investigate the seroprevalence of antibodies specific for the Ant70 V3 loop sequences in Africa and elsewhere. Several hundred serum samples from different geographical locations (Europe, Africa, and Brazil) were examined for their reactivity with a peptide representing the Ant70 V3 region. Simultaneously, the same serum samples were analyzed for their reactivity with the V3 consensus peptide described by LaRosa et al. (41). The limited reactivity found with the Ant70 V3 peptide indicates that this sequence does not occur frequently in African HIV-1 strains isolated to date. However, the spread of this virus subtype may at present be limited to the Cameroonian area, where 5 to 10% of HIV-positive sera were recently found to be reactive with the Ant70 V3 peptide. Moreover, during the preparation of this paper, the results of another group were brought to our attention. These investigators isolated and characterized an HIV isolate very similar to Ant70 and originating from the same geographical region in Africa (25a, 26). Taken together, these data indicate that isolates similar to Ant70 can be found in a limited geographical area in Africa. A regular survey to monitor the prevalence and spread of this virus is of epidemiological interest.

The immunodominant epitope localized in the gp41 protein has been described as highly conserved in the HIV-1 group. In particular, the sequence between both cysteine residues (CS GKLIC) is known to be invariant, and the corresponding peptide has been used in different serological assays for diagnostic purposes. The observation that two residues from this region are changed in the Ant70 isolate indicates that even this part of the gp41 protein can accommodate sequence changes. The implications of the observed substitutions for serology have been investigated and will be published elsewhere.

Comparative analysis of the three structural genes of the virus shows a similar distance between Ant70 and the HIV-1 isolates on the one hand and between Ant70 and the chimpanzee isolate SIVcpz on the other. It is also remarkable that Ant70 does not show a closer relationship to African isolates than to isolates from other parts of the world. Phylogenetic tree analysis indicates a unique position for the Ant70 isolate. From the position of this Cameroonian isolate, one can speculate that it could represent the closest relative to the HIV ancestor that is presently known. In particular, the branching of this isolate before any other HIV-1 isolate, even the SIVcpz variant, illustrates the very special position of this Ant70 isolate within the HIV-1 family. From these data, it may be inferred that human and chimpanzee lentiviruses probably have a common origin. The phylogenetic tree configuration is reminiscent of the topology found in the HIV-2 tree, in which some human isolates (HIV-2D205; HIV-2UC1) branch before the SIVmm and the SIVsmm strains (23, 39). The discovery of viruses such as Ant70 may allow us to establish a similar relationship between human and simian HIV-1-like isolates and lead us to the origin of the HIV-1 family. Analysis of more SIVcpz and other divergent human strains may well provide us with the necessary information to achieve that goal.

ACKNOWLEDGMENTS

We thank Betty Willems, Leo Heyndrickx, Conny van Loon, and Geertrui van Eynde for expert technical assistance. The help of Gerald Myers in the phylogenetic tree analysis is greatly appreciated. We also thank F. Shapiro for editorial comments.

This work was partially funded by the Belgian Institute for Scientific Research in Agriculture and Industry (IWONL/IRSIA grant 89.014I-I630-28).

REFERENCES

1. **Albert, J., U. Bredberg, F. Chiodi, B. Böttiger, E. M. Fenyö, E. Norrby, and G. Biberfeld.** 1987. A new human retrovirus isolate of West African origin (SBL-6669) and its relationship to HTLV-IV, LAV-II, and HTLV-IIIB. AIDS Res. Hum. Retroviruses **3:**3–10.
2. **Aldovini, A., and R. A. Young.** 1990. Mutations of RNA and protein sequences involved in human immunodeficiency virus type 1 packaging result in production of noninfectious virus. J. Virol. **64:**1920–1926.
3. **Alizon, M., P. Sonigo, F. Barré-Sinoussi, J. C. Chermann, P. Tiollais, L. Montagnier, and S. Wain-Hobson.** 1984. Molecular cloning of lymphadenopathy-associated virus. Nature (London) **312:**757–760.
4. **Barin, F., S. M'Boup, F. Denis, P. Kanki, J. S. Allan, T. H. Lee, and M. Essex.** 1985. Serological evidence for virus related to simian T-lymphotropic retrovirus III in residents of West Africa. Lancet **ii:**1387–1389.
5. **Barré-Sinoussi, F., J. C. Chermann, F. Rey, M. T. Nugeyre, S. Chamaret, J. Gruest, C. Dauguet, C. Axler-Blin, F. Brun-Vézinet, C. Rouzioux, W. Rosenbaum, and L. Montagnier.** 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science **220:**868–871.
6. **Bryant, M., and L. Ratner.** 1990. Myristoylation-dependent replication and assembly of human immunodeficiency virus 1. Proc. Natl. Acad. Sci. USA **87:**523–527.
7. **Chakrabarti, L., M. Guader, M. Alizon, M. D. Daniel, R. C. Desrosiers, P. Tiollais, and P. Sonigo.** 1987. Sequence of simian immunodeficiency virus from macaque and its relationship to other human and simian retroviruses. Nature (London) **328:**543–547.
8. **Chou, P. Y., and G. D. Fasman.** 1979. Prediction of β-turns. Biophys. J. **26:**367–384.
9. **Clavel, F.** 1987. HIV-2, the West African AIDS virus. AIDS **1:**135–140.
10. **Clavel, F., D. Guetard, F. Brun-Vézinet, S. Chamaret, M. A. Rey, M. D. Santos-Ferriera, A. G. Laurent, C. Dauguet, C. Katlama, C. Rouzioux, D. Klatzmann, J. C. Champalimaud, and L. Montagnier.** 1986. Isolation of a new human retrovirus from West-African patients with AIDS. Science **233:**343–346.
11. **Clavel, F., M. Guyader, D. Guétard, M. Sallé, L. Montagnier, and M. Alizon.** 1986. Molecular cloning and polymorphism of the human immune deficiency virus type 2. Nature (London) **324:**691–694.
12. **Clavel, F., and J. M. Orenstein.** 1990. A mutant of human immunodeficiency virus with reduced RNA packaging and abnormal particle morphology. J. Virol. **64:**5230–5234.
13. **Cochrane, A. W., C.-H. Chen, and C. A. Rosen.** 1990. Specific interaction of the human immunodeficiency virus REV protein with a structured region in the env mRNA. Proc. Natl. Acad. Sci. USA **87:**1198–1202.
14. **Cochrane, A. W., A. Perkins, and C. A. Rosen.** 1990. Identification of sequences important in the nucleolar localization of human immunodeficiency virus *rev*: relevance of nucleolar localization to function. J. Virol. **64:**881–885.
15. **Dayton, E. T., D. M. Powell, and A. I. Dayton.** 1989. Functional analysis of CAR, the target sequence for the rev protein of HIV-1. Science **246:**1625–1629.
16. **Dedera, D. A., R. Gu, and L. Ratner.** 1992. Role of asparagine-linked glycosylation in human immunodeficiency virus type 1

transmembrane envelope function. Virology **187:**377–382.
17. **De Leys, R., B. Vanderborght, M. Vanden Haesevelde, L. Heyndrickx, A. van Geel, C. Wauters, R. Bernaerts, E. Saman, P. Nijs, B. Willems, H. Taelman, G. vander Groen, P. Piot, T. Tersmette, J. G. Huisman, and H. Van Heuverswyn.** 1990. Isolation and partial characterization of an unusual human immunodeficiency retrovirus from two persons of West-Central African origin. J. Virol. **64:**1207–1216.
18. **Derse, D., and S. H. Newbold.** 1993. Mutagenesis of EIAV TAT reveals structural features essential for transcriptional activation and TAR element recognition. Virology **194:**530–536.
19. **Desrosiers, R. C.** 1990. A finger on the missing link. Nature (London) **345:**288–289.
20. **Fouchier, R. A. M., M. Groenink, N. A. Kootstra, M. Tersmette, H. G. Huisman, F. Miedema, and H. J. 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.
21. **Fukasawa, M., T. Miura, A. Hasegawa, S. Morikawa, H. Tsujimoto, K. Miki, T. Kitamura, and M. Hayami.** 1988. Sequence of simian immunodeficiency virus from African green monkey, a new member of the HIV/SIV group. Nature (London) **333:**457–461.
22. **Gallo, R. C., S. Z. Salahuddin, M. Popovic, G. M. Shearer, M. Kaplan, B. F. Haynes, T. J. Palker, R. Redfield, J. Oleske, B. Safai, G. White, P. Foster, and P. D. Markham.** 1984. Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. Science **224:**500–503.
23. **Gao, F., L. Yue, A. T. White, P. G. Pappas, J. Barchue, A. P. Hanson, B. M. Greene, P. M. Sharp, G. M. Shaw, and B. H. Hahn.** 1992. Human infection by genetically diverse SIVsm-related HIV-2 in West Africa. Nature (London) **358:**495–499.
24. **Gnann, J. W., J. B. Cormick, S. Michell, J. A. Nelson, and M. B. A. Oldstone.** 1987. Synthetic peptide immunoassay distinguishes HIV type 1 and HIV type 2 infections. Science **237:**1346–1349.
25. **Gorelick, R. J., S. M. Nigida, J. W. Bess, L. O. Arthur, L. E. Henderson, and A. Rein.** 1990. Noninfectious human immunodeficiency virus type 1 mutants deficient in genomic RNA. J. Virol. **64:**3207–3211.
25a. **Gürtler, L. G., P. H. Hauser, J. Eberle, A. von Brunn, S. Knapp, L. Zekeng, J. M. Tsague, and L. Kaptue.** 1994. A new subtype of human immunodeficiency virus type 1 (MVP-5180) from Cameroon. J. Virol. **68:**1581–1585.
26. **Gürtler, L., L. Zekeng, J. Eberle, S. Knapp, H. Hauser, and L. Kaptue.** 1993. Further characterization of a new HIV-1 subtype from Cameroon, Poster PO-A10-0147. IXth Int. Conf. AIDS.
27. **Hahn, B. H., G. M. Shaw, K. S. Ayra, M. Popovick, and R. C. Gallo.** 1984. Molecular cloning and characterization of the HTLV-III virus associated with AIDS. Nature (London) **312:**166–168.
28. **Hirsch, V. M., G. A. Dapoliti, S. Goldstein, H. McClure, P. Emau, P. N. Fultz, M. Isahakia, R. Lenroot, G. Myers, and P. R. Johnson.** 1993. A distinct African lentivirus from Sykes' monkeys. J. Virol. **67:**1517–1528.
29. **Holmes, E. C., L. Q. Zhang, P. Simmonds, C. A. Ludlam, and A. J. Brown.** 1992. Convergent and divergent 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.
30. **Hopp, T. P., and K. R. Woods.** 1981. Prediction of protein antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci. USA **78:**3824–3828.
31. **Howcroft, K. T., K. Strebel, M. A. Martin, and D. S. Singer.** 1993. Repression of MHC class I gene promoter activity by two-exon Tat of HIV. Science **260:**1320–1322.
32. **Huet, T., R. Cheynier, A. Meyerhans, G. Roelants, and S. Wain-Hobson.** 1990. Genetic organization of a chimpanzee lentivirus related to HIV-1. Nature (London) **345:**356–359.
33. **Javaherian, K., A. J. Langlois, G. J. LaRosa, A. T. Profy, D. P. Bolognesi, W. C. Herlihy, S. D. Putney, and T. J. Matthews.** 1990. Broadly neutralizing antibodies elicited by the hypervariable neutralizing determinant of HIV-1. Science **250:**1590–1593.
34. **Javaherian, K., A. J. Langlois, C. McDanall, K. L. Ross, L. I. Eckler, C. L. Jellis, A. T. Profy, J. R. Rusche, D. P. Bolognesi, S. D. Putney, and T. J. Matthews.** 1989. Principal neutralizing domain of the human immunodeficiency virus type 1 envelope protein. Proc. Natl. Acad. Sci. USA **86:**6768–6772.
35. **Jeyapaul, J., M. R. Reddy, and S. A. Kahn.** 1990. Activity of synthetic tat peptides in the human immunodeficiency virus type 1 long terminal repeat-promoted transcription in a cell-free system. Proc. Natl. Acad. Sci. USA **87:**7030–7034.
36. **Karpas, A.** 1990. Origin and spread of AIDS. Nature (London) **348:**578.
37. **Koken, S. E. C., J. L. B. van Wamel, J. Goudsmit, B. Berkhout, and J. L. M. C. Geelen.** 1992. Natural variants of the HIV-1 long terminal repeat: analysis of promoters with duplicated DNA regulatory motifs. Virology **191:**968–972.
38. **Koyanagi, Y., S. Harada, M. Takahashi, F. Ushino, and N. Yamamoto.** 1985. Selective cytotoxicity of AIDS virus infection towards HTLV-1 transformed cell lines. Int. J. Cancer **36:**445–451.
39. **Kreutz, R., U. Dietrich, H. Kuehnel, K. Nieselt-Struwe, M. Eigen, and H. Ruebsamen-Waigmann.** 1992. Analysis of the envelope region of the highly divergent HIV-2alt isolate extends the known range of variability within the primate immunodeficiency viruses. AIDS Res. Hum. Retroviruses **8:**1619–1629.
40. **Kwok, S., D. H. Mack, K. B. Mullis, B. Poiesz, G. Ehrlich, D. Blair, A. Friedman-Kien, and J. J. Sninsky.** 1987. Identification of human immunodeficiency virus sequences by using in vitro enzymatic amplification and oligomer cleavage detection. J. Virol. **61:**1690–1694.
41. **LaRosa, G. J., J. P. David, K. Weinhold, J. A. Waterbury, A. T. Profy, A. Lewis, A. J. Langlois, G. R. Dreesman, R. N. Boswell, P. Shadduck, L. H. Holley, M. Karplus, D. P. Bolognesi, T. J. Matthews, E. A. Emini, and S. D. Putney.** 1990. Conserved sequence and structural elements in the HIV-1 principal neutralizing determinant. Science **249:**932–935.
42. **Lee, W.-R., X.-F. Yu, W.-J. Syu, M. Essex, and T.-H. Lee.** 1992. Mutational analysis of conserved N-linked glycosylation sites of human immunodeficiency virus type 1 gp41. J. Virol. **66:**1799–1803.
43. **Leonard, C. K., M. W. Spellman, L. Riddle, R. J. Harris, J. N. Thomas, and T. J. Gregory.** 1990. Assignment of intrachain disulfide bonds and characterization of potential glycosylation sites of the type 1 recombinant human immunodeficiency virus envelope glycoprotein (gp120) expressed in Chinese hamster ovary cells. J. Biol. Chem. **265:**10373–10382.
44. **Levy, J. A., and J. M. Shimabukuro.** 1985. Recovery of AIDS-associated retroviruses from San Francisco patients with AIDS. Science **225:**840–842.
45. **Luciw, P. A., S. J. Potter, K. Steimer, and D. Dina.** 1984. Molecular cloning of AIDS-associated retrovirus. Nature (London) **312:**760–763.
46. **Maldarelli, F., M. Y. Chen, R. L. Willey, and K. Strebel.** 1993. Human immunodeficiency virus type 1 Vpu protein is an oligomeric type I integral membrane protein. J. Virol. **67:**5056–5061.
47. **Malim, M. H., J. Hauber, S. Y. Le, J. V. Maizel, and B. R. Cullen.** 1989. The HIV-1 rev transactivator acts through a structured target sequence to activate nuclear transport of unspliced viral mRNA. Nature (London) **338:**254–257.
48. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
49. **Matsushita, S., M. Robdert-Guroff, J. Rusche, A. Koito, T. Hattori, H. Hoshino, K. Javaherian, K. Takasuki, and S. Putney.** 1988. Characterization of a human immunodeficiency virus neutralizing monoclonal antibody and mapping of the neutralizing epitope. J. Virol. **62:**2107–2114.
50. **Myers, G., B. Korber, S. Wain-Hobson, R. F. Smith, and G. N. Pavlakis.** 1993. Human retroviruses and AIDS 1993. Los Alamos National Laboratory, Los Alamos, N.M.
51. **Nkengasong, J. N., M. Peeters, M. Vanden Haesvelde, S. S. Musi, B. Willems, P. M. Ndumbe, E. Delaporte, J. L. Perret, P. Piot, and G. van der Groen.** 1993. Antigenic evidence of the presence of the aberrant HIV-1ant70 virus in Cameroon and Gabon. AIDS **7:**1536–1538.
52. **Olshevsky, U., E. Helseth, C. Furman, J. Li, W. Haseltine, and J. Sodroski.** 1990. Identification of individual human immunodefi-

ciency virus type 1 gp120 amino acids important for CD4 receptor binding. J. Virol. **64:**5701–5707.

53. **Parkin, N. T. L., M. Chamorro, and H. E. Varmus.** 1992. Human immunodeficiency virus type 1 *gag-pol* frameshifting is dependent on downstream mRNA secondary structure: demonstration by expression in vivo. J. Virol. **66:**5147–5151.
54. **Peeters, M., K. Fransen, E. Delaporte, M. Vanden Haesvelde, G. M. Gershy-Damet, L. Kestens, G. van der Groen, and P. Piot.** 1992. Isolation and characterization of a new chimpanzee lentivirus (simian immunodeficiency virus isolate cpz-ant) from a wild-captured chimpanzee. AIDS **6:**447–451.
55. **Peeters, M., C. Honoré, T. Huet, L. Bedjabaga, S. Ossari, P. Bussi, R. W. Cooper, and E. Delaporte.** 1989. Isolation and partial characterization of an HIV-related virus occurring naturally in chimpanzees in Gabon. AIDS **3:**625–630.
56. **Popovic, M., M. G. Sarngadharan, E. Reed, and R. C. Gallo.** 1984. Detection, isolation, and continuous production of cytopathic human T-lymphotropic retrovirus (HTLV-III) from patients with AIDS and pre-AIDS. Science **224:**497–500.
57. **Quinn, T. C., J. M. Mann, J. W. Curran, and P. Piot.** 1986. AIDS in Africa: an epidemiologic paradigm. Science **234:**955–963.
58. **Rice, A. P., and F. Carlotti.** 1990. Mutational analysis of the conserved cysteine-rich region of the human immunodeficiency virus type 1 *tat* protein. J. Virol. **64:**1864–1868.
59. **Ruben, S., A. Perkins, R. Purcell, K. Joung, R. Sia, R. Burghoff, W. A. Haseltine, and C. A. Rosen.** 1989. Structural and functional characterization of human immunodeficiency virus *tat* protein. J. Virol. **63:**1–8.
60. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
61. **Sarngadharan, M. G., M. Popovic, L. Ruch, J. Schuepbach, and R. C. Gallo.** 1984. Antibodies reactive with human T-lymphotropic retrovirus (HTLV-III) in serum of patients with AIDS. Science **224:**506–508.
62. **Strebel, K., T. Klimkait, F. Maldarelli, and M. A. Martin.** 1989. Molecular and biochemical analyses of human immunodeficiency virus type 1 *vpu* protein. J. Virol. **63:**3784–3791.
63. **Swofford, D. L.** 1991. PAUP: Phylogenetic Analysis Using Parsimony, version 3.0s. Computer program distributed by the Illinois Natural History Survey, Champaign, Ill.
64. **Syu, W., W. R. Lee, B. Du, Q. C. Yu, M. Essex, and T. H. Lee.** 1991. Role of conserved gp41 cysteine residues in the processing of human immunodeficiency virus envelope precursor and viral infectivity. J. Virol. **65:**6349–6352.
65. **Tschachler, E., H. Buchow, R. Gallo, and M. S. Reitz.** 1990. Functional contribution of cysteine residues to the human immunodeficiency virus type 1 envelope. J. Virol. **64:**2250–2259.
66. **Tsujimoto, H., R. W. Cooper, T. Kodama, M. Fukasawa, T. Miura, Y. Ohta, K. I. Ishikawa, M. Nakai, E. Frost, G. E. Roelants, J. Roffi, and M. Hayami.** 1988. Isolation and characterization of simian immunodeficiency virus from mandrills in Africa and its relationship to other human and simian immunodeficiency viruses. J. Virol. **62:**4044–4050.
67. **Tsujimaoto, H., A. Hasegawa, N. Maki, M. Fukasawa, T. Miura, S. Speidel, R. W. Cooper, E. N. Moriyama, T. Gojobori, and M. Hayami.** 1989. Sequence of a novel simian immunodeficiency virus from a wild-caught African mandrill. Nature (London) **341:**539–541.
68. **Willey, R. L., F. Maldarelli, M. A. Martin, and K. Srebel.** 1992. Human immunodeficiency virus type 1 Vpu protein regulates the formation of intracellular gp160-CD4 complexes. J. Virol. **66:**226–234.
69. **Wolfs, T. F. W., J. J. DeJong, 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.
70. **Wolfs, T. F. W., G. Zwart, M. Bakker, and J. Goudsmit.** 1992. HIV-1 genomic RNA diversification following sexual and parenteral transmission. Virology **189:**103–110.
71. **Zabarousky, E. R., and R. L. Allikmets.** 1986. An improved technique for the efficient construction of gene libraries by partial filling-in of cohesive ends. Gene **42:**119–123.