JOURNAL OF VIROLOGY, Jan. 1995, p. 263–271 Vol. 69, No. 1
0022-538X/95/$04.00+0

Genetic Diversity of the Envelope Glycoprotein from Human Immunodeficiency Virus Type 1 Isolates of African Origin

JOOST LOUWAGIE,[1]* WOUTER JANSSENS,[2] JOHN MASCOLA,[3] LEO HEYNDRICKX,[2]
PATRICIA HEGERICH,[1] GUIDO VAN DER GROEN,[2] FRANCINE E. MCCUTCHAN,[1]
AND DONALD S. BURKE[3]

Henry M. Jackson Foundation Research Laboratory, Rockville, Maryland[1]; Walter Reed Army Institute of Research and the Military Medical Consortium for Applied Retroviral Research, Washington, D.C.[3]; and Prince Leopold Institute of Tropical Medicine, Antwerp, Belgium[2]

Received 5 July 1994/Accepted 17 October 1994

The genetic diversity of the envelope glycoprotein of human immunodeficiency virus type 1 (HIV-1) isolates was studied. HIV-1 isolates were obtained from eight countries in Africa: Djibouti, Gabon, Kenya, Senegal, Somalia, Uganda, Zaire, and Zambia. The DNA sequences encoding the complete HIV-1 envelope protein were PCR amplified and sequenced. Phylogenetic relationships among the 21 sequences from this study and the 32 previously published full-length *env* HIV-1 sequences were determined. Twenty of the newly sequenced African isolates could be assigned to envelope subtypes A, C, D, and G. One isolate, collected in Zambia, did not belong to any of the eight previously described subtypes and may represent a prototype sequence of its envelope subtype. The phylogenetic classification of these isolates was strongly supported by bootstrapping and the congruence of trees generated by either distance methods or maximum parsimony analysis. The data presented in this study confirm the existence of several genetic subtypes within the global HIV epidemic and broaden the genetic variability previously observed for envelope subtypes. The geographic spread of different subtypes was shown to be substantial, and the notion of cocirculation of subtypes was reinforced.

Soon after the discovery of human immunodeficiency virus type 1 (HIV-1) as the causative agent of AIDS, the significant genetic variability of this virus evoked intense interest. The broadest dimension of viral variability emerged from the initial comparisons of genetic sequences of HIV-1 viruses from North American and European patients with those from African patients (1–3, 18, 36). Significant but less dramatic variability was observed for virus isolates from individuals in a single geographic locale (37, 41). Within individual patients, a swarm of highly related but individually distinguishable viral variants was detected during the course of infection (11, 14, 27). These concepts have served as a durable framework for subsequent genetic analyses and underscored the need for an international focus in the collection of genetic data for the design of prophylactic vaccines.

Recently, an expanding genetic database for internationally collected HIV-1 isolates has added significant new dimensions to geographic aspects of virus variability (20). Concomitantly, concepts of local viral variability in population segments that may form early vaccine trial cohorts have undergone considerable refinement (5, 19, 24, 30). Among the major findings of this initiative, perhaps the most significant is the emergence of several genetically equidistant clades, or genetic subtypes, defined by both *env* and *gag* gene analyses. Some subtypes are not yet fully described, either because they are recognized solely on *gag* or *env* sequences or because few isolates are available for analysis (5, 12, 20, 25, 29). Evidence for local cocirculation of viruses from different subtypes continues to emerge. Some recent examples include the presence of subtype B and subtype E viruses in Thailand (24, 32, 33), subtype B and subtype F viruses in Brazil (19), and multiple subtypes cocirculating in several African countries (20, 28, 29). Geographic spreading and intermixing of genetic subtypes of HIV-1 may typify the epidemic in coming years (40).

Despite these advances, a full description of HIV-1 variants in many regions, including those with significant HIV-1 incidence and prevalence that may participate in early vaccine field trials, is often lacking. For sites harboring complex mixtures of variants, analysis of a large number of isolates from several different segments of the population may be required. Here we contribute 21 new full-length *env* gene sequences from isolates collected in eight countries in Africa, including some locales which are participating in early preparations for vaccine trials. These data contribute additional isolates to previously described *env* subtypes, provide evidence of more complex mixtures of genetic subtypes in some countries, and provide one sequence that may represent a prototype sequence of its respective *env* subtype.

MATERIALS AND METHODS

HIV-1 isolates. Virus isolates were from the peripheral blood mononuclear cells of HIV-1 enzyme-linked immunosorbent assay (ELISA)- and Western blot (immunoblot)-seropositive individuals and were collected during the period 1989 to 1991. The isolates, the geographic locales from which they were obtained, and the year of their isolation were DJ258, DJ259, DJ263, DJ264, and DJ373 (Djibouti, 1991); GA-VI525 (Gabon, 1990); KE89 and KE124 (Kenya, 1990); SN364 and SN365 (Senegal, 1990); SO145 (Somalia, 1989); UG266, UG268, UG269, UG273, UG274, and UG275 (Uganda, 1990); ZR-VI191 (Zaire, 1989); ZM18 and ZM20 (Zambia, 1989); and ZM184 (Zambia, 1990). HIV-1 isolates from Gabon, Kenya, and Zaire were contributed by Martine Peeters and Marleen Temmerman at the Prince Leopold Institute of Tropical Medicine (Antwerp, Belgium) and by Hugo Van Heuverswyn at Innogenetics (Ghent, Belgium). Isolates from Uganda were contributed by Hans Wigzell at the Karolinska Institute, Stockholm, Sweden. Isolates from Djibouti, Senegal, Somalia, and Zambia were from the collection of the Walter Reed Army Institute of Research (Washington, D.C.), whose contributors included Curtiss Hayes, Peter Perine, Beth Ungar, Douglas Watts, and other members of the Military Medical Consortium for Applied Retroviral Research.

Virus isolation and preparation of DNA for PCR. Ficoll-separated peripheral blood mononuclear cells from HIV-1-seropositive individuals were cocultivated

---

* Corresponding author. Present address: Innogenetics NV, Industriepark Zwijnaarde 7, Box 4, B-9052 Ghent, Belgium. Phone: (32) 9 241 0711. Fax: (32) 9 241 0907. Electronic mail address: joostlou@attmail.com.

with phytohemagglutinin-stimulated donor peripheral blood mononuclear cells in the presence of interleukin-2 as previously described (6). Cultures were supplemented with donor cells every 3 to 4 days, and p24 antigen production was monitored by ELISA. Positive cultures were expanded to 2×10^7 cells over a 2- to 3-week interval. Reverse transcriptase- or p24 antigen-positive cultures were harvested, and DNA was extracted, purified, and quantitated by optical density at 260 nm (22).

Molecular cloning and DNA sequencing. The full-length *env* (gp160) gene was amplified by PCR. Amplification primers were in *tat* (JL67, 5′-tag agc cct gga agc atc cag gaa gtc agc cta; JL68, 5′-ctt agg cat ctc cta tgg cag gaa gaa g; JL85, 5′-ccg tct aga tag atc cta gac tag agc cct gga a; and JL86, 5′-ccg tct aga tgc tgt tta ttc att tca gaa ttg g) and *nef* (JL89, 5′-tcc agt ccc ccc ttt tct ttt aaa aa); primer combination JL86-JL89 was the most successful (10 of 21). The conditions for PCR amplifications were 10 mM Tris–50 mM KCl (pH 8.3)–1.5 mM $MgCl_2$–0.01% gelatin–200 μM dATP, dCTP, dGTP, and dTTP–100 pmol of each primer–1 μg of DNA template–2.5 U of *Taq* polymerase (Perkin-Elmer) in a volume of 100 μl. Thermocycling was performed by rapid heating to 95°C, holding at 95°C for 30 s, cooling to 55°C over 1 min, holding at 55°C for 30 s, heating to 72°C over 30 s, and holding at 72°C for 7 min. Amplification was for 35 cycles. Two microliters of PCR product was ligated in TA vector (Invitrogen). After transformation of *Escherichia coli* and plating on X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) plates, white colonies were picked and screened with a p32-labeled probe (SK70) (31). Plasmid DNA of positive clones was purified (Qiagen plasmid kit) and used as a template for sequencing. Sequencing was done with fluorescent dye primers or terminators (Applied Biosystems) by using thermal cycle sequencing with *Taq* polymerase. Electrophoresis and data collection were done with an Applied Biosystems 373A DNA sequencer.

Analysis of data. DNA sequencing projects were assembled and edited by using Seqman II, Editseq, and Mapdraw (DNAStar). Multiple alignments of DNA or protein sequences were done with CLUSTAL V (16). Phylogenetic analysis was performed with PHYLIP 3.5 (10), the input sequences for isolates were randomized with the jumble option (where applicable), and all trees constructed were unrooted. For DNA distance methods, we used Dnadist. Neighbor-joining trees were built with Neighbor, and Fitch-Margoliash trees were built with Fitch. Maximum parsimony trees were generated by using Dnapars. The statistical robustness of both neighbor-joining and maximum parsimony trees was assessed by bootstrap resampling (100 data sets) of the multiple alignment; bootstrap consensus trees were built by using Consense. Protein neighbor-joining trees were constructed with Protdist (Dayhoff PAM matrix) and Neighbor.

Nucleotide sequence accession numbers. The GenBank accession numbers for the *env* sequences obtained in this study are L22939 through L22957, L23064, and L23065.

RESULTS

Sequencing of the gp160 gene. In this study, we determined the nucleotide sequence of a 2.5-kbp segment that encodes the *env* protein (gp160) of 21 HIV-1 isolates collected in diverse geographic locales in Africa. The isolates were not obtained as part of any systematic study of the diversity of HIV-1 within any country but were collected for various clinical and epidemiological studies. Isolates were from south-central Africa (Zambia), central Africa (Gabon and Zaire), East Africa (Kenya, Somalia, and Uganda), and West Africa (Senegal). A selection of genetically diverse isolates (21 of 63 isolates) was made on the basis of available *gag* sequence information or results obtained with subtype-specific amplification primers (21). The *env* sequences of the selected isolates were obtained by PCR amplification and molecular cloning of a 3.3-kbp DNA segment encompassing the regulatory genes *tat*, *rev*, *vpu*, and part of *nef* as well as the gp160 coding region. At least one full-length clone was obtained for each isolate studied. The entire gp160 region of one clone was sequenced for each isolate; sequencing encompassed both strands of DNA, and all sequence ambiguities were resolved. The majority (17 of 21) of the selected clones contained an intact open reading frame for gp160. The clone from isolate DJ258 had a premature stop codon at amino acid 22, isolate UG275 had a stop at position 285, and GA-VI525 had a stop at position 681. One isolate (UG274) contained a 1-bp deletion that resulted in a frameshift mutation at nucleotide position 1058. No unusual insertions or deletions were observed, and the lengths of gp160 sequences varied between 2,498 (SN364) and 2,659 bp (ZM20). The previously reported range was 2,512 to 2,604 bp.

Phylogenetic analysis of HIV-1 isolates. Previously, phylogenetic analysis of a 721-bp segment of the *env* gene established the existence of five different *env* sequence subtypes of HIV-1 (29). A sixth *env* subtype was identified among isolates from Brazil (19, 34) and Romania (9). Recently, *env* subtype G was attributed to isolates from Gabon and Russia (4) and *env* subtype H was documented for isolates from Zaire, Cameroon, and the Central African Republic (17). We performed detailed phylogenetic analysis of the 21 African *env* sequences obtained in this study to determine their relationships with previously published sequences.

Pairwise alignments of the 21 sequences showed various degrees of genetic distance between isolates. The gp160 nucleotide sequence differences between some Djibouti isolates were especially low, with only 5% difference between DJ263 and DJ258. The highest pairwise difference was observed for isolate ZM20, with 24% difference relative to KE124, ZR-VI191, and GA-VI525. Inspection of the distance matrix showed the clustering of isolates into at least three major groups (data not shown); however, some isolates showed little resemblance to any other isolate. To allow a more detailed analysis of genetic relationships, a multiple alignment of all known complete gp160 sequences was built. This alignment, which contained isolates from 13 countries, formed the basis for phylogenetic analyses. All sequences from this study and 32 previously published sequences of *env* subtypes A through G were included. The recently published (17) subtype H isolates were not included, as only partial *env* sequences are available. Phylogenetic analysis was done by a distance-based method (neighbor joining) and the maximum parsimony method. DNA distance methods use pairwise nucleotide distances to sequentially group pairs of isolates that are most closely related (35). Maximum parsimony or minimum evolution analysis constructs trees that require the smallest number of evolutionary changes to explain the differences observed between the sequences tested (38).

In the phylogenetic tree based on neighbor-joining analysis, several major groups of related isolates were easily discernible (Fig. 1). We used the *env* subtype nomenclature proposed in the 1993 Los Alamos HIV database to refer to these groups (29). *env* subtype A previously contained two isolates from Rwanda (KIG93 and SF170), one from Zaire (Z321), and one from Uganda (U455). Seven isolates were added to this subtype: three had been collected in Djibouti (DJ258, DJ263, and DJ264), one had been collected in Kenya (KE89), two had been collected in Uganda (UG273 and UG275), and one had been collected in Zaire (ZR-VI191). *env* subtype B contained mainly isolates collected in the United States and Europe; none of the African isolates from this study belonged to this subtype. *env* subtype C included viruses from Djibouti (DJ259 and DJ373), Somalia (SO145), Senegal (SN364), Uganda (UG268), and Zambia (ZM18 and ZM20). The 1993 Los Alamos database lists five other viruses of this subtype; one was collected in South Africa, and the others were collected in India. Most of the early Zairian isolates belonged to *env* subtype D (Z2Z6, ELI, NDK, and MAL). Five isolates were added to this subtype, and their origins were Kenya (KE124), Senegal (SN365), and Uganda (UG266, UG269, and UG274). The five currently available full-length gp160 subtype E sequences derive from HIV-1 isolates collected in Thailand (23, 24). Partial *env* sequences from a study of Central African Republic isolates may provide additional members of this subtype (28). None of the African isolates from this study clustered with the E subtype. Similarly, none of the 21 isolates

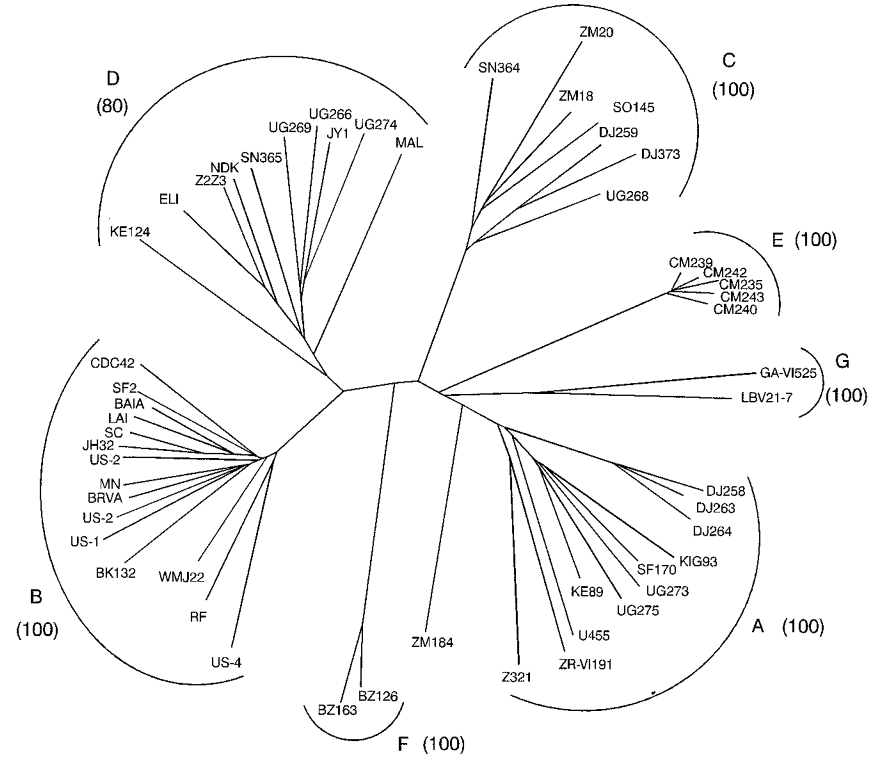

FIG. 1. Phylogenetic tree of HIV-1 on the basis of *env* DNA sequences. The gp160 multiple sequence alignment was resampled by the bootstrap method (100 data sets); a consensus tree generated by neighbor-joining analysis is shown. Branch lengths for the consensus tree were calculated by using the maximum likelihood algorithm. The horizontal and vertical orientations of branches are noninformative and for clarity only; branching patterns and branch lengths reflect phylogenetic distance relationships. The seven *env* (A to G) subtypes discussed in the text are delineated (arcs); numbers in parentheses are bootstrap values for the delineated subtypes.

clustered with subtype F isolates from Brazil. One of the isolates collected in Zambia (ZM184) did not belong to any of the above-mentioned groups.

In the phylogenetic tree based on maximum parsimony analysis (data not shown), the overall branching pattern was very similar to that observed by neighbor-joining analysis (Fig. 1). The same subtypes were easily identifiable, and the clustering of isolates in subtypes was identical. As in neighbor-joining analysis, isolate ZM184 did not cluster with any of the major *env* subtypes but appeared to be a genetic outlier. Overall, 52 of the 53 sequences were classified into seven *env* subtypes. A partial *env* sequence multiple alignment that contained subtype H isolates VI557 and CA13 failed to group isolate ZM184 within this subtype. These findings fortify the current classification and, concomitantly, lend significance to ZM184 as an outlier of the eight described *env* subtypes.

Recently, two groups independently described the isolation and sequencing of highly divergent HIV isolates (ANT70 and MVP5180) obtained from Cameroonian patients (13, 39). The genomic organization of these isolates, designated subtype O, is similar to that of HIV-1 isolates. However, *env* genetic difference from any previously described HIV-1 strain was about 50%. None of the isolates obtained in this study belonged to highly divergent subtype O. Figure 2 shows a phylogenetic analysis of the *env* protein of the HIV-1 isolates analyzed in this study, the two published subtype O isolates, and two isolates obtained from naturally infected chimpanzees. This analysis illustrates the clear distinction between any known HIV-1 subtype and subtype O or chimpanzee isolates.

Analysis of protein sequences. The gp160 DNA sequences of the 21 isolates obtained for this study were translated into protein sequences. The neighbor-joining protein-based trees showed a branching pattern similar to that from DNA-based analysis (data not shown). Seven protein *env* subtypes were identified, and the clustering of isolates in subtypes was very similar. In contrast to DNA-based phylogenetic analysis, isolate ZM184 weakly clustered within *env* subtype D (bootstrap value, 82 of 100); DNA-based phylogenetic tree analysis did not group it with any subtype.

Figure 3 shows a multiple protein alignment of all known gp160 sequences; the isolates were grouped according to subtype. A near-perfect conservation of the 18 cysteine resi-

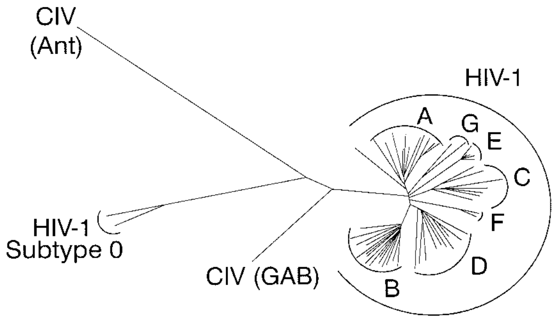

FIG. 2. Phylogenetic relationship of HIV-1 isolates to simian immunodeficiency viruses isolated from naturally infected chimpanzees [CIV (Ant) and CIV (GAB)] and to subtype O isolates.

```
            1                  20                  40                  60                  80                  100
                                   signal/gp120 +  +     +        *                   *                                            *
A  Consensus MRVKGIQRN.QHLWRWG.TMLLGMLMICSA.AENLWVTVYYGVPVWKEATTTLFCASDAKAYETEVHNVWATHACVPTDPNPQEIVLENVTENFNMWKNNMVEQMHEDIISLWDQSLKPC
        Z321 -K-----G-W-NW-K--.-LI--LVI----.----------------D-E------------D--K------------------LS-G----K-D-------------V-----------
       KIG93 ---M---M-C-N-----..-I----I---V.-G--------------D-E------------D--A-------------------L-K----E------G------T-------------
       SF170 ---M---M-C-N-L---.--I----I----.TSK-------------D-E-------------R----------------D----Y------G-------------T-------------
        KE89 ---M-T---C---L---.--I---II----.----------------N-E---------------K-------------------H------E-------------T-------------
    ZR-VI191 -------M-W----K--.-LI--LGI--N-..-------F-------D-E------------SE-K--------------D-K--N-V----D------K-----Q--------------
        U455 ---M-----YPC-----.--I--LII--N-..QQ-------------D-V------------DA---------------------D-V----E---------D-----------------
       UG273 ---------W---L---.-II--LIT---T.-K-------------RD-E---------------M-------------------R------E--I----------T-------------
       UG275 -KAM-T-M-C-NW----.--I----I---T.--D------------RD-E------------D--M-------------------N-G----K---------------------------
       DJ258 ---M---K-YPL-----.-IIF IMI--N-..KEM-----------RD-K------------D----------------------Y-K----K---------------------------
       DJ263 ---M-M---YPL--K--.-IIFWTMI--N-..-K------------RNTE------------DA---------------------H------K---------------------------
       DJ264 ---M-----YPLF-K--.MIIFWIMI--N-..-K------------RD-E------------DV---------------------H-K----K---------------------------

B       JH32 ------RK-Y-------.-----I------.--Q---------------A------------D---------------------V-------K---------------------------
         LAI ----EKYQHLWRWGWKWG-----I------.T-K----------------------------D---------------------V--V-----------D--------------------
          MN ------R--Y--W-G--.-----L------.T-K----------------------------D---------Q-----------VE-V--------------------------------
          RF ---MEMRK-C----K--.------------.--D-----------------------E----K--------K------------VL----------------------------------
          SC -----SG--Y-------.-----I------.--Q----------------------------D-----I---------------V--G--------------------------------
        BAIA ---TE-RKSY--W----.I---------N-.E-K----------------------------D---------------------VE----------------------------------
       BK132 ----E-RK-C-------.-L---I---S--.-------------------------------D-----------------S---VA-V-------I-T--I-------------------
        BRVA ------KK-Y-------GM----I------.TDK---------------N------------D--I------------------L-MG-----------D--------------------
       CDC42 --A---RK-C-------.------------.-A-----------------------------D--A------------N-----V-----------------------------------
         SF2 -K---TR--Y-------.-L----------.T-K-------------------------R--D---------------------V--G-----------------Q--------------
         US1 ----ETRK-Y-------.IL---T------..-K--------------------G-------D----------------------A------D---------------V-----------
         US2 ------KKTYR------TM.---I---S--.-GQ--------------------G-------DK----------------D-K-LP----------------------------------
         US3 ----E-RK-C-R--K--.------------.---S-------------------G-------D-----------------S---V-----------------------------------
         US4 --AMEMRKHC-----.-GIL---I----R-.TTV----------------------------KA-A------------------VN-T----------------------------R---
       WMJ22 ------M--C----I--.---F--W-----.V-Q----------------------------S--A------------------VI-G--------------------------------

C      DJ259 -K-M-T---WPRW-V--.ILGFW--L--NG..N.-------------D-NPP----------KV----------------S---LA-------------G--D---Q-------------
       DJ373 ---M---M-C-QW-IG-.ILGFW--LM-NG.MG-------------Q--NP-----------------------------S---L-M------------G--D---Q-------EG----
       SO145 ---R--L--W-QW-I--.ILGFW------G.GG-------------R--K--------S----R----------------S---ML-G-----------D--D-----------------
        ZM20 ---R--L--C-QW-I--.ILGFW-----NGVGD.---------------K------------------------------S---M--A-----------D--D-----------------
        ZM18 ---RE-L--W-QW-I--.ILGFW-V-NYNVVGD----------------K-------------R-----------------------G-----------D--D-------R---------
       UG268 ---M-----C-QW-I--.ILGFWI----NV.MG---------------------.-------------------------------------S------D--D---Q-V-----------
       SN364 ---T-ML--C-PW-I--.ILGFW-RL-YNV.GG----------------K-------------R-G------------------M-W-----Y------D--D-----------------

D        MAL ---RE----Y-NW----.M-------T--I.--D--------------------------S-------I----------------E------G---------------------------
         ELI --AR--E--C-NW-K--.I----I--T---.-D---------------------------S----A--I----------------A----------------------------------
         JY1 ---M--RM-Y----K--.I----I--T--V.--D--------------------------S--P-A--I-------------R--EM-----------------------------N---
         NDK --AREKE--C-N--K--.I-------T---.--D----------I-----------------KK-A--I----------------E----------------------------------
        Z2Z3 ---R--E--C-N--K--.I----I--T--N.-D---------------------------S-K--A--I----------------E-----------R----------------------
       KE124 ---M-----C-L-LGG-.I-I---II--N-.-N.------------RD-E------------K--M-------------------H--------------------T-------------
       SN365 --AREMK--Y-------.--------T--V.--K--------------------------S----K--I----------------E-A--------------------------------
       UG266 ------P--Y-N-L---.-II---II---T.-D--------------D-E----G-------D-------------------R--K-----------------R---------E------
       UG274 ---RE-E--Y----I--.--------TY-V.--Q--------------------------S-IA-A--I-----------D-R--R--------D-------------------------
       UG269 ---RE-E--Y-------.----------RV.TGE--------------------------S-KKGA--I-------------R--K----------------------------E-----

E      CM239 ----ET-M-WPN--K--.-LI--LVI----.SD-------------RD-D-------N---H-----------------------Y-------------------Q--V-----------
       CM240 ----AT-M-WPN--K--.-LI--LVI----.SD-------------RD-D----G------H-----------------------H-------------------Q--V-----------
       CM242 ----ET-M-WPN--K--.-LI--LVI----.SN-------------RD-D-------N---H-----------------------H-------------------Q--V-----------
       CM243 ----ET-M-WPN--K--.-LI--LVI----.SD-------------RD-D-----------H-----------------------Y-----------N-------Q--V-----------
       CM235 ----ET-M-WPN--K--.-LI--LVI----.SN-------------RD-D----F--N---H-----------------------H-------------------Q--V-----------

F      BZ126 ---R-M---W---GK--.LLF--I-I--N-..----------------------------S--K--------------------V---------D-----------T-------------
       BZ163 ---R-M---W---GK--.LLF--I-L--N-..---------------D------------S--R--------------------V---------D----S------T-------------

G    LBV21-7 --A---L--W----K--.ILI--LVI--N-.SG-----------A-ED-D------------S--S-------------------S-N--------------------------E-----
    GA-VI525 ---R---T-W-N--N--.-LI--LVI-S--.SN-------------ED-E-P--------S-S--S-------------------S-------------K--------------E-----

       ZM184 ---R--PK-W-QW-I--.ILGFW------V.VG-------------R--K-------------R-----------------------------------D--D-----------------

            121                140                 160                 180                 200                 220
                  *    *                         V1                 *                        V2                         *        *
A  Consensus VKLTPLCVTLNCTN.N.NNT...N...TT.SN..................KGEMKNCSFNITTELRDKKQKVYALFYKLDIVPIDN.................YRLINCNTSAITQACPK
        Z321 ----------S-H-ITIKDNNTN.........VDTEM.............-E-I----Y-M-------QR-I-S---R------GG..NSSNGDSS......K-----------------
       KIG93 -Q--------D-NYNITI-S...TKTPSPVT-ISEEM.............RE-I------M-----------HS--------HLND.SKPK...........Q-----------------
       SF170 ------------SHNITTHN.....IT-NNP-ITKYM.............E--I----Y-M------------S------V----K.NNGNNNIRT......Q-----------------
        KE89 ------------S-V-VTA-S....GT-HGGR...................-R-----Y--------RR----S---R--V---NE.NNSSNN.........E-----------------
    ZR-VI191 ----------E-H-STTPDNNTT..........QSTI.............ID-TT--TY-M-------V---HS---R--V-Q-NGINGTSNNNSA......Q-----------------
        U455 ----------D-H-ITI---NNNTN.......ITDGV.............RE--------M------------S---R----Q-NK.TDNNS...........---------T-------
       UG273 --------------A-LT-ANKNAKVNSATNGTIQSM.............ER-I------M-------TRR--S------V-Q-K-.SS.............E-----------------
       UG275 ------------SDV-N........VTV-NHSISSDM.............---I------M-------Q--IHP---R--V---NG.SNNR...........Q-----------------
       DJ258 ----------E-H-VSNS-.......TSSYR-ISSEM.............RE-I------M-------Q--MS----R--V-Q-SE.SNNS...........Q-----------------
       DJ263 ----------D-H-V-.....SS-SSTSINRTIDSKM.............QE-I------M-----------S----R--V-Q-NE.SSNS...........Q------------K----
       DJ264 -E--------D-H-NV..........TSNNRTISRDM.............Q--V------M---------Q-S----R--V-Q-NE.SSNSS..........Q-----------K-----

B       JH32 ------------ID.WG-D-.SP-ATN--S-GGEK............M.E-------------SI---V--EH-----H-V---N-STKDNIKNDNSTR....----S----V-------
         LAI --------S-K--D.LG-A-.NT-SSN-NS-SGEM............MME---I-------S-SI-G-V--E--F------I---..........DTTS....-T-TS----V-------
          MN -------------D.LR-T-NTN-STANNN--SEG............TIKG------------SI---M--E---L------S---.........DSTS....----S----V-------
          RF -------------D..A-LNGT....NV-S-SGGT............MMEN--I-----QV--SR---T--K--------V---EKGNISPKNNTSNNTSYGN-T--H--S-V-------
          SC --------------.LR-D-.ST-ATN--S--RGK............M.EG---T--------SI-S-V--E--------V-----...........TS....-T-------V-------
        BAIA -------------D.LR-A-.NG-DTN--S-SREM............M.GG--------K---NI-G-V--E-----E--------NSNN........R....----S----V-------
       BK132 --------------.LR-D-ST.......N-TAEE............MEA---------K---SI---L--E------------N-.........DNTS....----S----I-------
        BRVA ------------HD.F..-A.TN.....AT--SGK............MMEG------------SI---M--E--------------.......DKTNTR....----S----V-------
       CDC42 -------------D.L.-TNNTT-TTELSIIVVWE............QRG----R--------SI---V-RE--------VE---DNK........NTTNNTK---------V-------
         SF2 -------------D.LGKA-NT....NSSNWKEEI...............---I---------SI---I--EN---RN--V----........NASTTTNYTN----H--R-V-------
         US1 ---P---------D.LV-A-NTG-ATN--S-SWEK............L.E---I--------KNM---VERE--------V--M--.........DNTS....----S----VV------
         US2 -------------D.FSR-D.TN.....-NNSSGE............KME---I----------V---I--Q--------V---E-.......DNDNTS....----S----V-------
         US3 --------------.WK-......VTY-NS-SGEN............M.ER---------V--SIKN-V--E--------------.........DNTS....-T-------V-------
         US4 -------------DKLTGS-NGT-STSG-N-TSGTNSTSTNSTDSWEKMPE--I---------SV---V--E-S------V-----DNAS.............---------V-------
       WMJ22 ------------ID.......K....NI-DWKNTT............IIGG--V---------SR---VH-E--------V---KGDNNSSR...........---------V-------

C      DJ259 ------------S-.........VQSNA-YE-S.......T.........-E--------V-------R--A-----R-----LN.....ENNS........E-----------------
       DJ373 ------------S-....KVTATGNGSV-YNST.......E.........-D.-R----------------K-----------LG.....NGNE........E---------S-------
       SO145 -------------EA-I-I-SV...GNV-.--D.......N.......KE..-I------A----S-----------R-----LNEGEN...NSR.......K-----------------
        ZM20 --M-------E-G-VTVPKDNSTYNGDIN.--T.......T......MKE..--R-----A--VIN--TS--H----RV----LNDENSNKTNANSTKP..TK-I-------T-------
        ZM18 ----------E-G-V-VT..........H.E-S.......T.......KG..--------A----K----R------------LNENNNSSEDSS.......E-----------------
       UG268 --------------V-V-I-NNA-ATNSPYE-G.......K......LMEQ--I------V---I-----TAH-------V-SLEGNSNT.............-----------------
       SN364 ------------R-.......VTTSNNA-SNDN.......P.........N--I--------------RRNE-----R-----LSGSKNSSNSS........E-----------------

D        MAL --------------V-GTAVNGT-AGSNRTNAEL.............KMEI--V--------PVGS--R-E.--T--N--L-Q--DSDNS............S---------V-------
         ELI ------------SDELR--GTMG-NVT-EEKG...............M......------V--V-K----Q------R--------DSSTNS.........TN-----------------
         JY1 --------------AGG-K-TNG-NTTNQEEQ...............MME-------------VIS---KQ-H----R--V----DDNSANTSNTN....YTN-----------------
         NDK -------------DELR-S..KG-GKVEEEEK...............R......------V....---REQ---------------NNRTNS.........TN------D--T-------
        Z2Z3 ------------IDEVME-VTMK-NNV-EEIR...............M......---------VV---TKQ-H----R-------DNSTNS..........TN-----------------
       KE124 ------------S-V-GTGGNSTTGN......NI.............TREDIG-----------V--R-KQ-----------QMNENQD.NSSSGE....Y.R-----------------
       SN365 ------------RDISSDA-SNTISNV-G.IP...............MMG----Q---------I---R-N--S---R--L---G-SNK...NSTN....YTS---------V-K-----
       UG266 ------------SEGRVT...DT-G..-IGNG...............TIGD..-R---------I----KQ---------V----DNNSTN..........TS-------------T---
       UG274 -------------EWKNTTEQEK-T..AVPGS...............TTSDTD-R--------------EQ---------V-Q--DNNSTNS.........TN---V--T----------
       UG269 -------------EAKF-V-RAT-GKN-VRNA...............TKGGL---D-Y--V---I----KQ-R-------V-QMN-SESIN..........TS-------S---------

E      CM239 --------------A-LT-VN..-.....IT-VPNIIGNIT..........D-VR-----M-----------H---------Q-EG....NNSS........E---------V-K-----
       CM240 --------------A-LT-GS..SK.....T-VSNIIGNIT..........D-VR--T--M----T------H---------Q-ED....KKNSS.......E---------V-K-----
       CM242 ----------H---A-LT-GK..-.....ITKVSN...NIT..........D-VR-----M-----------H---------Q-ED....NQNSSN......E---------V-K-----
       CM243 ----------I---AKLT-ANLT-V..NNIT-VSNIIGNIT..........D-VR-----M-----------H---------Q-GD....KNSS........E---------V-K-----
       CM235 ----------H---AKLTTAN..-H..NNITSVSNTIGNIT..........D-VR-----M-----------H------------ED....NKTSS.......E---------V-K-----

F      BZ126 ----------H--SATP-A-ATQ-GTTLKE...................EP-AIQ-----MS--V-N-Q---H-----------.......SNSSNNSR...E----H-D--T-------
       BZ163 --------------AI.-A-AIQDGD.LKE...................GP-AIQ-----V-M-V---Q---H---.R------.......SNGRNGSGD..Y---------T-------

G    LBV21-7 --------------AIA-V-NSNK......TSNNSSLGNST......VSSIE------------V---SK-E-----R--V---N...GNENSSTT......D-------A-T-K-----
    GA-VI525 -E-----------KVTY-S-AAN-RSS-VTTPTSTVTNNST......VSG-E--T-----V-------QKAE--------V----...NNRNDS.........----H--V-T-K-----

       ZM184 ------------SDATST-SNTT............................E-I----YRT----K--TLR--S------V-Q--......PSNNT......M----S------A-----
```

FIG. 3. Amino acid sequence alignment of the HIV-1 *env* protein. −, identity with the consensus sequence; ·, gap in aligned sequences; X, corrected frameshift; a space, premature stop codon. Isolates that belong to the same subtype (A to G; Fig. 1) are grouped. *, cysteine residues; +, amino acids believed to be involved in gp120-gp41 interaction; ˆ, the leucine zipper motif; #, the CD4 binding site. Hypervariable domains (V1 to V5) in the gp120 molecule and an immunodominant domain (ID) in gp41 are delineated by solid lines.

```
            241                260                 280                 300                  320                 340
                      *         *          *       *                                                 *                  V3
A Consensus VSFEPIPIHYCAPAGFAILKCNDKKFNGTGPCKNVSTVQCTHGIKPVVSTQLLLNGSLAEE.EIIIRSENLTNNAKTIIVQLNESVEINCTRPNNNTRKSIR...IGPGQAFYATGDI..
       Z321 ---------------------R-EE-E-K---R-----------R---------------G.-VR-----F-D---I-----VKP-N-T-M----------S...----R--F-----..
      KIG93 ---------------------KRAN-T---------S-------R--I------------KDVQ------I----------FTKA-K-------KT---GV-...------W--R-NM..
      SF170 -----------------------AE-------------------R--I------------G.RVK-----I------------KT---------------V-...-------------..
       KE89 -T-------------------K----------T---------------------------G.-VM-----I-----N----FA-P-K-----------M---...-------------..
   ZR-VI191 ---------------------R--E-------R---S------------------------.K-M-----F-D-G-------V-P-K------S-----G-H...----R------Q-..
       U455 ---------------------K-PE---K---R---------------------------R.--R-----F-----------VNP-K---S--Y-.---N--RYS--S-----V--K-..
      UG273 ------------------------E-----L-----S-------R---------------T.-VK-K---ISD----L----TTP-K------G----T-V-...-------------..
      UG275 -----------------------QE-----I-N---------------------------G.KVK-----I-----N-L- -TTP-T-----------T-V-...-----S-------..
      DJ258 -T--------------------------------------------A-------------G.-VV-----I-----N-----VTP-R---S--G------V-...-----T-------..
      DJ263 -T----------------------------L------------------------------.-VV-----I----QN-----VAP-R-------------V-...-----T-------..
      DJ264 -T-----------------------------------------------------------.-VV-----I-----N-----VTP-G-------------V-...-----T-------..

B      JH32 I-----------------I-------------T----------------------------.-VV-----F-D---------K-P-V------SKT--RR-HI...---R---T-KQ-..
        LAI ----------------------N-T-------T-----------R----------------.-VV---A-F-D----------Q------------------I.QR---R--VTI-K-..
         MN I-------------------------S-K-S-------------R----------------.-VV-----F-D-------H-----Q-------Y-K--R-HI...---R---T-KN-..
         RF -----------T--------------------------------R----------------.-VV-----F-D-V--------A--Q--------------TK...---RVI----Q-..
         SC ------------RW.----N--N---------T-----------R-----H----------.-VVL----F-D---------K-A------------TR--HI...---R--------..
       BAIA I--------------------K------K---S-----------R----------------.-VV-----FAD----------------------------HI...---R-L-T--E-..
      BK132 ---------------------R----S-N--------------------------------.-VV-----F-D---------K-P----------Y---...I.TM---RVY-T--E-..
       BRVA -T--------------------N---------T-----------R----------------.-VV-----F---V------------------------R-TM...---RVY-T--Q-..
      CDC42 -----------T-T---L--------------T-----------R----------------.-VV-----F------------V-----------H---RVTL...---RVW-T--E-..
        SF2 -----------T----------N-T---K---T-----------R-I--------------.-VV---D-F---------------A--------------YI...---R--HT--R-..
        US1 ----------------G----K--------Q-------------R----------------.-VVV--A-F-----I---------Q---I----------HI...---R-I----G-..
        US2 -T-------------------K--T-------T-----------R----------------.-VV---V-F----R------R-A--------S-------HI...---R---T--N-..
        US3 -----------T-----------E--------T-------A-------------------K.-VV-----F-D----------------------------PI...---R--------..
        US4 ---------------------K----------------------R----------------.--VL----F-D-----------------I----------HI...---R--------..
      WMJ22 --------------------------------T-----------R----------------.--V-----F-D-------H------------Y--V-R-LSI...---R-.FR-RE-..

C     DJ259 --LD-----------Y-----R--N-------N---------------------------D.D-------------I---------A-----------E---...-----T-------..
      DJ373 ---------------Y-----R-T---------------------------------T---.D-------------I------Q--------------Q---...-----T-------..
      SO145 ---D-----------Y------NET-------RY--------------------------G.--M---------------H--Q----R---YA------V-...-----T--.-N--..
       ZM20 --LG------------G----TN-T-------H-------------------------S--.G------------Q-------K-I--V-A--G--------...-----T-F---A-..
       ZM18 -TLD-----------Y------N-T-------H----------------------------.------------------H-------V----S--------...-----------G-..
      UG268 -TLD-----------Y------N-T-------N--------------I---------I---.--------------I------K------A-------E---...-----T-------..
      SN364 --LD-----------Y------N-T-------N---------------------------G.------------------H----I--V-A---------M-...-----T-------..

D       MAL -T-D-------------------------EI------------------------------.--M-------D-T-N-------T-T------.G-NTRR.GI.HF-----L-T--I...
        ELI ---------------------R----------T-----------R----------------.-V------------N--AH-----K-T-A--.YQNTRQ.-T.P--L--SL-T-RSR..
        JY1 -T-------------------K-----------K----------R----------------.------------V-----H------------D-KITRQ.ST.P--L---L-T-RIK..
        NDK I--------F-----------R----------S-----------R----------------.------------V--------A-IV------.YKYTRQ.-T.S--LR-SL-TITGKKK
       Z2Z3 ---------------------R--R-------T-----------R----------------.--------------I---------A------.YRNIRQ.-T.S--L---L-T-KTR..
      KE124 ---------------------K--E-----------------------------------G.-VK---Q-------------VKP-R------.--NIRK.SI.R-------F-----..
      SN365 -N-------------------K---------------------------------------.-------------------F----M------.Y-NKRQ.-T.P--L--VLHT-RVK..
      UG266 -T------------------------------------------R----------------.--V-----------I---------I------Y--IKIQ.-T.P--R---LFT-..RRK
      UG274 -T-------------Y----------------Q-----------R----------------.--------------I---------T-D-A--YE-I-.Q.-T.P--T---L-T-QGRKK
      UG269 IT-------------------T------K--------------------------------.--------------I------A--T-T----YTK--.H.-A.Q...-R-WWT--.-..

E     CM239 I--D-------T---Y--------N-----------S------------------------.------------------H--K---------S----T--...T-----V--R----..
      CM240 I--D-------T---Y--------N-----------S------------------------.-------D----------H--K---------S----T--...T-----V--R----..
      CM242 M--D-------T---Y--------N-----------S------------------------.------------------H--K---------S----T--...T-----V--R----..
      CM243 I--D-------T---Y--F-----N-----------S------------------------.----------D-------H--K--G------S---PS.....IT----V--R----..
      CM235 I--D-------T---Y--------N-----------S------------------------.------------------H--K---------S----T--...P--------R----..

F     BZ126 --WD-----------Y---------------------------------------------.D-----Q-ISD-------HF----Q--------Y-----Y...F---R--HTA-K-..
      BZ163 --WG-----------Y-------------------.-------------------------.D-----Q-ISD-------HF----Q------------G-H...----R--------..

G   LBV21-7 -T-D---------------T-R---------------------------------------.--M-----F-----N----F-K-ID-V------------H...------L----A-..
   GA-VI525 ---------------------RE-N-T---L-----------------------------K.--------F-----V-------T----------------K...F-T-RVL----A-..

      ZM184 -----------T---------K-TE-----S-----S-------R---------------G.-VM-----I-D---N----FT-P-Q-S----S----R---...F-------.-N--..

            361                380                 400                 420                 440                 460
                     *                                               *      *                  V4                  *        ########
A Consensus .IGDIRQAHCNISRAKWNKTLQQVAKKLREHF..NNKTIIFNPSSGGDPEITTHSFNCGGEFFYCNTSGLFNST.WN..............N.TGSNDTITLPCRIKQIINMWQEVGKAM
       Z321 .----------V--TE--D--SK--AQ--K--VNTSTD---AN-----V-----------------------G-.-LNGTSN....N..TWKIDTV----I--------V----R--Q--
      KIG93 .-----K-Y--V--TE--Q---K--TQ-GK--...----T-TK-----L------------------T------.-K.NVNG....TMQESSS-KL------T-----------RA-Q--
      SF170 .-------Y--V---D------G--NQ-KSY-.S-.-----AS-----L----------------D--------.-GGNSTD....SIQES-S-E-----I-------------R--Q--
       KE89 .--N-------V---E--T---K-VT----Y-.G-----K-AN-----L-----------------S-------.-T.........NMQES-S-E------------.------RA-Q--
   ZR-VI191 .T------------TE-----H---IQ-----ITK-T-----S-----L----------------S--------.-ENGTQV....A....-D----------------VR---R--Q--
       U455 .----------V--RD--R-I----EQ-KKK-..-------AS-----I------------------------I.--GSMSN....DMGP......-G----Q-----------R--Q--
      UG273 .----------V--SE-KE---K-V-Q--T-W...------TN-----L------------------A------.-..NKND....SIK.S-D-K---I---Q-----------R--R--
      UG275 .----K---------E--E---K-VSQ--T--..G------GN-----I----------R-----D--------.--GNTSK......QESSS-E--GN---Q-----------RA---I
      DJ258 .----------V--S---E---K--TQ--K--...-T----AN-----I---------R---------S-----.--SS.......SIHTNYSSNDT-N---Q------V----K----I
      DJ263 .----------V--SE--R------TQ--K--...------AN-----I-------------------E-----.--STWDN....SSNHIESNHTEGN---Q------V----K--L--
      DJ264 .-----K----V--S---N------IQ--K-S...-T----ANP----I-------------------E-----.-....DN....SLNHTESNHTE-N---Q------V----K--R--

B      JH32 .A--L-------N--R--A--K-IVG---KQ-..V----V--R--------VM-------------STQ-----.-L.SNS.TWNDT.EGS-N--G------------------------
        LAI ..-NM-------------A--K-I-S----Q-.G-------KQ--------V--------------STQ-----.-F...NSTWST..EGS-N-EGS-----------F-----------
         MN .--T--------------D--R-IVS--K-Q-..K----V--Q--------VM---------------P-----.--GN..NTWNNT......----NN---Q-K---------------
         RF .-----K----L---Q--N--K--VT----Q-..D----V-TS--------VL--------------TQ-----.--.STEGS........-N--G-------------V----------
         SC .-----------------N--K-IVI---DQ-..E-------R--------VM-------------STQ--S--.--GT.........EGS-N--G-----------E------------
       BAIA .----------L------D--NKIVI----Q-..G----V-KH--------V--------------STQ-----.--VT.........EES-N-VE-N----------------K--R--
      BK132 .-----R------ST---N--G-IV---K-Q-..--N--V-KK--------VM---I-W----F---TK-----.--.STEG.........-DD-EERN----------V---------L
       BRVA .-----R----L--S--EN--K-IVT---VQ-..K----V--R--------VM----------F---TQ-----.-Y.RNTT........G-I-EG-SP---------------------
      CDC42 .L-N-----------Q--N----I-TT---Q-..G----A--Q--------VM-------------STQ----A.--VTSNGTWSVTR...KQKDTG-I------------R--V-----
        SF2 .-----K--------Q--N--E-IV-----Q-.G-----V--Q--------VM-----R--------TQ---N-.-R..LNHT........EG-KG----I-------------------
        US1 .-----R-Y-----T---N--K-IT---Q-Q-..GK---V-KQ-------VVM--------------TN-----.--.GTE.......EFDSTLSDAEN---Q-----------------
        US2 .-------------G---N--K-IVS----Q-..G----V--Q--------VM-------------STQ-----.-G.ANDT........WTESNN-E--------------R-------
        US3 .----------V-K----S--E-IV---G-Q-..G-N--V--Q--------VM-------------STQ-----.-V.NHNGTWNSTWSSN-N-DN-G--------R----R--------
        US4 .---------Y--K-N-TN--E-IVE----Q-.G--------S--------VF---------------Q-----.--IT.........EEV-K-KE----I-----S-----C-----P-
      WMJ22 .--I--------------N--K-IVE----Q-..K----V--H--------V--------------STQ-----.--.GTDIK........GDNKNSTL---------------G-----

C     DJ259 .-------------KE--E--RK-RE--KTL-..RTAN-T-Q------L-------------------D---N...................QSIT-EN---.--------------R--
      DJ373 .-------------Q--KE-----KG--K---..-KT.-K-A------I-------------------A---DS..ITN.........API-NSSTDAN---Q--------------R--
      SO145 .------------GD---R-----G---A---..P----K-E------L---------R---------------.Y-PN.........NTGS..E--NN---Q--------------Q--
       ZM20 .-------------SE--E--PN-SR--AG--..P-R----E------L---------R---------I---RS.HLGL.........FNGTENN-PQN--I----N-------------
       ZM18 .--N---------KEN------K-G---A---..P----K-DQH----L---------R---------N-----.YKPN.........DTNSTYNP------------------G--Q--
      UG268 .-------Y-----NE--I---W-RE--KR--..P----N-TQP----L---------R---------S----S..................DNNNSTI---------------G--R--
      SN364 .------------.GN--A--EK-KG--Q---..PG-N-S-E------L---------R------D--K---G-..THT.........A.......-SN--IQ-----------G--R-I

D       MAL .V----R-Y-T-NETE-D-------V--GSLL..-KTK----S---------------R---------K-----.-QNNG.ARLSNS..TESTG....S---------------KT----
        ELI ..SI-G---------Q-S-------R--GTLL..-KTI-K-K--------------------------------.--IS...AWNNI..TESNNST-TN---Q-------K-VAGR.--I
        JY1 ..------Y----A-A------------GDLL..-QT----K-PA-----------------------R-----.--..S.STWNND..TL-SE...G--K-------------G-----
        NDK KT-Y-G----K----E---A-----T--GNLL..-KT--T-K----------S-ML----D-------R-----.-.........NQ..TNSTGFN-G-V---------V-L--R-----
       Z2Z3 ..SI-G--Y----KNE---------I--GNLL..-KT----K--------------------------------.-DISK.SEWANS..TESDDKP...---Q-----------G-----
      KE124 .--N----Q--VNGTE--E----ISTQ-KK--..M------K----------------R---------K----V.-GNSS....NVT..KV-G-KVKE------K------------R--
      SN365 ..-----------KER--------VR--KDL-..-KT.---E---------A------R---------------.-KTKS.GD..............N-------------L--R-----
      UG266 ..-IKG--------EG-------------DLI..-KT----K------------------------------NSV-RN......DYP..SNSTGVNESV-----K------L--G--T--
      UG274 .--K-G-----V-K-G-E--------x-KDL-..KADN-T-K--------------------------R---NSG-.E......NYN..IS-S--G-E--R-Q-----------------
      UG269 .T------Y----K---QE---------GDLL..-KTQ---E--------V-K------------D--R---DS.Y-T......SAL..NSTQGNESEN---------------------

E     CM239 .--N--K-Y-E-NGT---EA-K--TE--K---..-------Q-P----L---M-H---R--------TQ-----............CIGNETME-C-G--I---K---------GA-Q--
      CM240 .--N--K-Y-E-NGT----V-K--TE--K---..-.-----Q-P----L---M-H---R--------TK---N-............CLGNETMA-C--------K---------G--Q--
      CM242 .-----K-Y-E-N-T---EA-K--T---K---..-----L-Q-P----L---M-H---R--------TQ---N-............CIGNETME-C-S------K---------GA-QP-
      CM243 .-----.-Y-E-NGT---RV-K--TE--K---..-------Q-P----L---M-H---R--------TR---N-............CIGNETMN-C-G------K---------GA-Q--
      CM235 .-----K-Y-E-NGT---EV-T--TE--K---..-------Q-P----L---M-H---R--------TR---N-............CIENGTMG-C-G--I---K---------GA-Q--

F     BZ126 .-----K----V-GTQ-----ER-RA--KS--..P-T--K--S-----L---M-----R-------------D-G....................--G----L---R--V-------R--
      BZ163 .-----K----V-GTQ-----E--RA--KS--..P-A--K--S-----L---M-----R-------------D-G....................--G----L-----MV-------R--

G   LBV21-7 .----------V-ETD-RDM--K-KAQ-QGIY...--S-T--S-----L---A----R--------------S.ILSNNNA....PSNSS.....KN.----------VR---R--Q--
   GA-VI525 .--N-------V-TE--G-M--N--A--QKE-...--N-T--S-----L---------R------K--E-----.-LFNSTG....LVNSTWNDTRK-.---Q-K----VR---R--Q--

      ZM184 .-------Y-S-NET---A---K---Q-QNY-..P----S-AN-----L-------------------N-----.Y-PNSTE....S.......N-DLN---------------R--R--
```

FIG. 3—*Continued.*

```
              481                500                 520                 540                 560                 580
              ##         *                  V5                                    + + + +++ +       gp120/gp41
A   Consensus YAPPIE.GQIRCSSNITGLLLTRDGG...........NNNNTNETFRPGGGDMRDNWRSELYKYKVVKIEPLGVAPTKAKRRVVER..EKRAVG..GAVFLGFLGAAGSTMGAASITLTV
         Z321 -----K.-V-K-V-----I------V...........G--TS------------------------------------------A-..----I-.M--F---------------------
        KIG93 ------.-I---D-------------VG........D-ST-GT--------N-------------------------T---K----..----I-.I------------------------
        SF170 -----Q.-V-K-I-------------...........D--SA------T--N-------------------------P------Q-..------.I----I-------------------
         KE89 -----Q.-I-K-V------I------...........D--SES-------------------------------------------..------.I------------------------
     ZR-VI191 -----Q.-M---E------I-----V...........G--S-----------------------------------SR------A-..------.L-------------------L----
         U455 -----Q.-V---E-------------...........T--TK---------------K-------------------R--------..------.L--I---------------------
        UG273 -----K.-V---E-------------...........-GS------------------N--------Q---------P------Q-..------.L----I-------------------
        UG275 -----K.-V---V------I------...........G--SE--------R--------------------------R-R------..------.V------------------------
        DJ258 -----S.-E---E------I------...........S--S---I--------------------------------R--------..------.L--F---------------------
        DJ263 -----S.-E---K-------------...........S--S-----S------------------------------R--------..------.L------------------------
        DJ264 -----P.-E---E-------------...........S--S--------------------------------I---H--G-----..------.L----F-------------------

B        JH32 ------.-------------------...........D-Q-ET--------N------------------L-------------Q-..------.I-----------------.-M----
          LAI -----S.-------------------...........----GS-I---------------------------------------Q-..------.I--L--------------R-M----
           MN ------.-------------------.........KDTDT-DT-I----------------------T----------------Q-..----..AI--L----------------V----
           RF -----S.---K-I-------------..........EDTT--T-I--L---N---------------R---------R------Q-..------TI--M--------------G------
           SC -----K.--VK---------------......NSKNGSK-ENT-I---------------------------------------Q-..------TI--M--------------T-M----
         BAIA -----R.-------------------...........PEA-KT-V---------------------------------------Q-..------.I-------------------M----
        BK132 -D---G.-----T-------------.........NQ-GT-ET-I------N--------------------------------Q-..------.I--ASP-------------PT----
         BRVA -----R.---K---------------.........NN-ETTDT-I------N--------------------------------Q-..------.L--L----------------L----
        CDC42 --L--K.-L-----------------...........GE-Q-T-I---------------------------------------Q-..------ML--M--------------T-MA---
          SF2 -----G.---S---------------..........T-VT-DT-V---------------------I------I----------Q-..------IV--M--------------V-L----
          US1 -----R.-H---V-------------.......MNNGTKD-ET-V---------------------------------------Q-..---T--.I--M----------------L----
          US2 -----A.-------------------.........NK-DS-GT-I----------------------------I----------Q-..------.L--M---------------------
          US3 -----R.-------------------.........NN--T-RT-I----------------------R---------R------Q-..------TI--M----------------M----
          US4 -----R.---I---------------........TNN-RT-DT--------N-K-------------R---------QP-----Q-..------.L--L-I--------------V----
        WMJ22 -----Q.-------------------...........-SSSRE-I------N---------------R----------------Q-..------TI--M--------------G-L----

C       DJ259 -----K.-I-T-K--V----------..........T-TTTEG-I----------------------E-K-V-I----------G-..----.A.L--M---------------------
        DJ373 -----Q.-I-A-K-------------............T-S-D------------------------E-K---L------------..----.A.LA-----------------------
        SO145 -----K.-N-T-T--------L----.....NDPKRTEEQ-NT-R----------------------E-K---I---G--------..------.I------------------------
         ZM20 -P---A.-N-T-I-------------IDTNSTKITHET-GTR-------------------------E-K---I---GP------KQI------.I-------------------L----
         ZM18 -----A.-N-T-K-------------..........S-DTTNT------------------------E-K---I---A-------TR.------.I-----------------------A
        UG268 -----K.-K-T-R-------------.........ETSET-ST------------------------EVK----------------..------.I------------------------
        SN364 -----A.-N-T-K-------------...........TT--NT-I----------------------E-K---I------------..------.I------------------------

D         MAL -----A.-V-N-L------I------.........NSSD-SD---L-----------I---------R------------------..----I-.L--M----------------L----
          ELI ------.RN-L---------------...........I--S--------------------------Q---------R--------..----I-.L--M--------------R-V----
          JY1 ------.-L-K-T-------------...........V--S------------K----N--------R-----I---R--------..----I-.L-----------------V-VA--G
          NDK ------.-L-K---------------...........A--SSH--I-------------------------I-------R------..----I-.L-------------------V----
         Z2Z3 ------.---N---------------...........T--SS-----------------------------------R--------..----I-.L--M--------------R-L----
        KE124 ------.-L-N---------------..........E--SSQ-------------------------------L---S--------..----I-.L--M----------------M----
        SN365 ------.-V-S----------L----.........IH-TS-ET------------------------RL--------T------K-..----I-.L-VM----------------MA---
        UG266 ------.-L-K---------------...........V--SQ------------------------IRL----L---R--------..----I-.L-------------------L----
        UG274 -----I.---K---------------.....VNNSSNETSSK--I----------------------R-----R---R--------..----I-.L--L------T---------M----
        UG269 ------.-I-----------------...........IT--S---------N---------------R-----L---T--------..----I-.L--L----------------L----

E       CM239 -----R.-R-N-V-----I-------...........A--T----------NIK-------------Q-----I---RP-------..------.I--MIF-------------------
        CM240 -----S.-R-N-V-----I-------...........V--TD---------NIK-------------Q-----I---R--------..------.I--MIF-------------------
        CM242 -----S.-K-N-V-----I-------...........A--TNS--------NIK-------------Q-----I---RP-------..------.I--MIF-------------------
        CM243 -----S.-K-N-V-----I-------...........A-TT----------NIK-------------Q-----I---R--------..------.I--MIF-------------------
        CM235 -----S.-R-N-V-----I-------...........AI-T----------NIK-------------Q-----I---R--------..------.I--MIF-------------------

F       BZ126 --A---.-N-T-N-------------...........Q-S-QT--------N-K-------------E---------R---Q--K-..-----VGI--L-----S---------------
        BZ163 --A--S.EN-T-N-------------...........Q---QT----------K-Y---V-------E---------R---Q--R-..------.M--L-----S---------------

G     LBV21-7 -----A.-N-T-K------I------...........-ST-G------A-----------------I--K---------R----G-..----IT..-------------------LA---
     GA-VI525 -----AA-N-E-R-S----I------...........-----------------G--------------KS------R-R------..----I-.L------------------------

        ZM184 -----A.-V-K-T-----II-S----........NSG-ST-GT-------------------------L---------PR------..------.L------------------------


              601                620                 640                 660                 680                 700
                                 ^      ^      ^       ^     ^   ID     *     *
A   Consensus QARQLLSGIVQQQNNLLRAIEAQQHLLQLTVWGIKQLQARVLAVERYLKDQQLLGIWGCSGKLICTTAVPWNSSWSNKSLEEIWDNMTWMEWEREISNYTGLIYSLIEESQNQQEKNEQ.
         Z321 ---R-----------------------K------------I---------------------I--P-N-----------QSD---K---L--DK-V----QV--N------T---I--RD
        KIG93 -------------S-------------K-------------------------------------P-T----------T-S--------LQ-DK------D---Q---D----------D
        SF170 -------------S-------------K---------------------------------------N---------RTQS---N----LQ-DK------DI--N------I-------E
         KE89 -------------S-------------K--------------------R-------------I----N-----------QS---E----LQ-DK------DI--N-L------------D
     ZR-VI191 -------------S---------H-------------------L----R----------------P-N----Y------YDD--N----LQ-DK--D---Q---G--------------D
         U455 -------------S-------------K--------------------Q------------------T-----------Q-D--N----LQ--K---S---I--Q-------------LD
        UG273 -------------S-------------K------------L-------R------------------T-----------QS--------LQ-DK------NI--D----A---------D
        UG275 -------------S---K---------K--------------------R------------------N-----------QSD-------LQ-DK--N---E---R-L------------D
        DJ258 -------------S-------------K---------------L-S--R------------------T----------TFND-------LQ-DK------DI--T--------------D
        DJ263 -------------S-------------K---------------L-S--R------------------N----------TYND-------LQ-DK-V----DI--N--------------D
        DJ264 -------------S-------------K---------------L----R------------------T----A-----TYND--E----LQ-DK------DI--N-------------KD

B        JH32 ---L-----------------G------------------I-------------------------------A-----------------------D---S---T--------------E
          LAI ----------------------------------------I-------------------------------A--------Q--N-------D---N---S--H---------------E
           MN ---L---------------------M-----------------------------F-----------T----A-------DD--N-----Q-----D---S-----L-K--T-------E
           RF ---H--------------------------------------------R------------------T----A-------NM--N-----Q-----D----I--N-L------------E
           SC ---L--------------------------------------------R------------------T----T-------DK--G-----------D---S---T--------------E
         BAIA ---L--------------------------------------------R-----------------------A-------NK----------D---N---SI-----------------E
        BK132 -P-----------------------------------------------------------------T----A-------D---N-----------N-------T------S------LD
         BRVA ---L-------------M-------M-E--------------------------------------------A-------SD--------------D---N-------D--I------KE
        CDC42 --------------------K-------------------I--------------F----------------A-----T-DQ--N-------D---D---H---T------------Q-E
          SF2 ------------------------------------------------R-----------------------A--------D--------Q-----D---NT--T-L------------E
          US1 ------------------------------------------------------------------------D-----TMDQ--N-----D--K--D---S---N---------D----E
          US2 ------------------------------------------------------------------------T-------DQ--------------D---NT--T-L----L-------E
          US3 ---L---------------------M-----------------------------------------T----T-------DK----------D---N---T---T-L------------E
          US4 ---------------------P----------------P-I--------------------------T------------T---------------G-------N---IA---------E
        WMJ22 --------------------D---------------------------R------------------T----A------MNQ----L---------D---SI------------G----E

C       DJ259 -------------S-----------M------------T----I-----------------------N----A----R-QK--------IQ-----N-------G-L-D----------D
        DJ373 -------------S---K---VH--M------------T----I-----------------------------------Q----------Q-D---N---ET--R-L-V--T---Q--KD
        SO145 -------------S-----------M-----------R-----I----E----------------------------RTQ----E-L---Q-DK------DT--R-L-V---------KD
         ZM20 ----V--------S-----------I-----------------I-------------------------------T--TY-Y--N-----Q--K-----SDT--K-L-D--T---Q--KE
         ZM18 ----V--------S------------------------T----M-------------------------S---------QTD--------Q-D-------DT--R-L----I---R--KD
        UG268 -------------------------M------------T----I----Q-------------------------------GD--------Q-D--------T--R-L-D---------KD
        SN364 -------------S-----------M------------T----I-------------------------------------D--------Q-D---N---SI----L-----------ND

D         MAL ------------------------------------------------Q--R---M------H----F---------R--DD--N-----Q--K-------I--N------I------KE
          ELI -----M----------------------------------I---------------------H----N---------R--N---Q-----------D--------------T------KE
          JY1 -------------------------M-------------------S----------------H----T----------------N----I------D----V------N--I-------D
          NDK -----M----H-------------------------------------R------------RH----N---------R--D---Q-----------D--------------I------KE
         Z2Z3 ----------------------------------------I--------------------------T---------R--ND--Q-----------D-------R------T-------E
        KE124 -----M--------------D------------------------S----------------H----N------------Q-T------K------D-----------K--T-------E
        SN365 ----------------------------------------I-------R------------RH----N------------T---Q-----------D--------------T-------E
        UG266 ---L-------------M---------------------------S----------------H----S----------------N-----------D--------------V-------E
        UG274 ----V---------------K-------------------I---------------------H----T---------R------N-----------D-------T--------------E
        UG269 ----V---------------------------------------S--G------------RH----N---------R------H-----------D--------------T-------E

E       CM239 -------------S-------------------------------------KF--L------I--------------R-F----N----I----------NQ--EILTK--D--DR--KD
        CM240 -------------S-------------------------------------KF--L------I----------T---R-F----N---RI----------NQ--EILT------DR--KD
        CM242 -------------S-------------------------------------KF--L------I--------------R-F----N----I----------NQ--EILTK--D--DR--KD
        CM243 ------------.S-----.-------------..------.---------K.--L-.----I----------T---R-F----N--.-I--.-------NQ--EILT------DR--KD
        CM235 -------------S-------------------------------------KF--L------I----------T---R-Y----N----I----------NQ--EILT------DR--KD

F       BZ126 -------E----------------------------------------Q------L-----------N-----------Q----E--------K-----SYE--R---Q----------E
        BZ163 -------------S----------------------------------Q------L-----------N-----------Q----E-------QK-----SNEV-R---K----------G

G     LBV21-7 -V----------------------------------------------Q------------------N----T------FN--------I------N---HQ----L----I-------D
     GA-VI525 -V-----------S--------------------------------F-----P--------------N----T------YND------------------KQ---------I-------D

        ZM184 -------------S---K---------K------------A--M----Q------------------T----------TQD----K---LQ-DK------EI--Q-L-D----------D
```

FIG. 3—*Continued.*

```
             721                740                 760                 780                 800                 820
                                                                                                                          *
A  Consensus LLALDKWASLWNWFDITNWLWYIKIFIMIVGGLIGLRIVFAVLSIVNRVRQGYSPLSFQTLLPAPRGP.DRPEGIEEEGGEQDRDRSIRLVSGFLALAWDDLRSLCLFSYHRLRDLILIA
        Z321 --------N-------S----------------------------I---------------THHQ-E-.----R---G-----------------P------------C------CA---
       KIG93 ----------------S------R--------------I------I-------------IHT-N-G--.---GR--------G----------------E---N-----------F----
       SF170 --------N-------SK-----R---------------------I-------------IHT-N----.----R--------G--------------------------------F----
        KE89 --------N-------S-----------------------T----I--------------HT-----L.----R-----------------T-----------------------F----
    ZR-VI191 -----------S----S------R---------------------I---------------PHHQ-E-.--------G------G------------------------------F---V
        U455 --------N-.---N-S------RL-VI------------T----I---------------A-I-E-L.G--GR--------GK------------I------N-----------FA--V
       UG273 -----------S----S------RL--------------L--F-VI----K----V----HT-N--DL.----R-A-----------.---------------------------F---V
       UG275 --------N-----G-S------RL-------------------VI--------------HT-N---L.---GR------K-------H--------------------------F----
       DJ258 --------G--S--G--------R---I---------------T-I---------------THHW-E-P----R---G-----K---------------------------Q---FV---
       DJ263 -----Q--N-----G--K-------------------------T-I---------------THHQ---.----R---G---------G---------------------------FV---
       DJ264 -------E------N--K---------------------L---TVIK--------------THHQ-E-.----R---G-----K---V-----------------------H---FV---

B       JH32 --G-----------T--------R---------V------T-------------------R-------.-----------DR-----GQ--D-L--II-V----------------L--V
         LAI --E-----------N------------------V--------------------------H--T----.------------R---------N-S---I------------------L--V
          MN --E------------------------------V-----------------------L--RP-V----.------------R---T-G---H----II-V-----F-----H.---L---
          RF --E-----N--------Q-----R---------V--K-----------------------H-------.-------G----R-----GGA-N---T-I----WT--S---------L--V
          SC --E-----------N------------------V------T-------------------R--SQ---.------------R-----G---D----II-V-.--------------L--V
        BAIA --E--------------E-----------I----------S----M--------------H---S---.---G--------R-----GP--N-----I-V---G------------L--V
       BK132 --E------------------------------V----I-T-------------------H-------.---G--------R---T-G---D----IF-V---N------------L-VV
        BRVA --E-----------N---------------------------------------------R--GR---.------------R-----SP--D-----F-V-----F----------L--V
       CDC42 --Q--------T-S---K---------------------------------------------N----.-----T--G---RG--G-T---H-----V------------------L--V
         SF2 --E-----------S------------------V--------------------------R--V----.---D--------R-----V---D-----I-E----------R-----L---
         US1 --E-----------N--------------I---V--------V-----------------R-------.------------------G---D-L---IRV----------------L--L
         US2 --E----T---S--S--H------------E---------------K----------------V----.-----T------R---------D-----I------------------L--V
         US3 --E-----N--T-----Q---------------A------T------------P------R-------.---A--------------SP--H-----I------------------L--V
         US4 --E--------------------R-------------------------------I-L--R---Q---.------------R-----N---H-L-V-I------------------L--C
       WMJ22 --E-------------------------------------T-------------------H--T----.------------R-----V---H-----I------------------L--V

C      DJ259 -------QN-----S--K--------------------I----------------------T-N----.---G-----------G-------------------------R---------
       DJ373 -------QN--S-----K--------------------I----------------------I-N----.---R----------KG--T-----------------F-------------V
       SO145 -----S-G---------K---------I----------I-V--------------------T-N----.--LGR---------K-------N-----------------------F--V-
        ZM20 -------NI-------SK-----R--------------I----------------------T-NI--EP--LGR-----------------N-----V--------R----Q---FT---
        ZM18 -----S-KN-----------------------------S----------------------T-N----.--LGR-----------K-----D-----F-----------------F----
       UG268 -------QN--S-----K--------------------I-----------K----------T-T----.---GE-------------V--I-------------------R-----L---
       SN364 -----N-NN-----N--K--------------------I----------------------I-N----.--LGR---------------------SI------------------F---V

D        MAL --E-----------S-SK-----R---IV---------I-----L------------L-----T----P-------------G-G------N--S--I-----N------------L---
         ELI --E-----------S--Q-----------I--------------L-----------------------.-----T------RG----V--LN--S--I----------------------
         JY1 --Q-----------S--K----------------------T---L-----------------------.-------------G--------N--S--F-----N----------------
         NDK --E-----------S--K------L-------------------V------------------V----.----E-------RG--------N-LF--F-----N-----------S----
        Z2Z3 --E-----------N--Q--------------------------L-----------------------.------------RG--------N--S--I-----N----------------
       KE124 --Q-----------S--K--------V--------------M--L-----------------------.-------------G--------N--S--I-----D----------------
       SN365 ----------------------------V-E-------I-----L-----------------------.-------------G--------N--S--I--G--N----------------
       UG266 --Q-----------N--Q--------------------------L------------L----------.-----T-------G-G-------------------------------L---
       UG274 --Q------V----S--K-----R--------------------L---------------------E-.----E--------G-G------N--S--I-----N------R-----L---
       UG269 --Q-----------S--------R-S----------------F-L-----------------------.-------------G-G---------S--I-----N----

E      CM239 --E---------------------------------------------------------PSHHQ-E-.----R---G----G----V---------------------------F----
       CM240 --E-----------------------------------I---------------------PSHHQKE-.--------G----G----V----------------------------T---
       CM242 --E-----------------------------------I---------------------PTHHQ-E-.----R---G----G----V---------------------------F----
       CM243 --E--------S--------------------------I--.---------------L--PTHHQ-EL.----R---G----G-E--V---------------------------F-S--
       CM235 --E--------------K-----------I--------I---------------------PFHHQ-E-.--S-R---G----G----V---------------------------F----

F      BZ126 ---------------MS---------------------------------K------L--HI-S--E-.--------G----GNE--V---N-----V-----N------RH--------
       BZ163 --------------S-S---------------------------------KA--------HI-S--E-.----E---G----G----V---D-----V-----N-------H----L--T

G    LBV21-7 -------E---S--S--Q------------------K------------------------AHHQ-E-.----R---G---------V-------SI--------------Q---F---V
    GA-VI525 ---------- -----SS----------V--------------------------------THHQGE-.--L-R---G---H--T--T-SAN---P-V----W-------RL----L--V

       ZM184 --------N--------R---------I-----------I-II-V--------------IPT-H-E-L.---GR---G----G--------------------------------C----


             841                860                 880                 900

A  Consensus AR.IVELLGR.......RGWEALKYLWNLLQYWGQELKNSAISLLDTTAIAVAEGTDRVIEVVQRAGRAILHIPRRI
        Z321 --.---T-I-R......----T----G--VI-----------N----V-----DW--------------F-N-----
       KIG93 --.T-----HSSLKGLRL---G--------L---R---I---N-F--I--V--GW-------G--I---F-N-----
       SF170 --.T----K-SSLRGLRL---G--------L---R-------N----I-----GW------IG-AIC----N-----
        KE89 TE.T-----HSSLRGLRL---G--------L-------I--SN-V--I--V--GW------IA-GI-----------
    ZR-VI191 --.---T---R......-------------L---R---I---T---A------RW---I--IG--IV----N-----
        U455 --.A------SSLKGLRL---G--------L---R---I---T---AV-V---GWI-----IG-TI-----N-----
       UG273 --.T-K---HR......----GI-------LF--R---I---N----I-----GW-------IG--I---F-N-----
       UG275 --.T-----Q...........G--------A---R---I-----F--I-VVI-GW------LG--LC----------
       DJ258 V-.A-----HSSLKGLRL-----Q--G---L----------VN----I--V--NW------IG-EV-----N-----
       DJ263 V-.A-----HSSLKGLRL-----Q--G---L-----------N----I-L---NW------IG--V-----N-----
       DJ264 V-.A-----HSSLKGLRL--G-----G---T-----------N----I-----NW------IG--V----RN--V--

B       JH32 T-.-------.......--------W-------S-------V--FNAI-----------LKIL---F-------T--
         LAI T-.-------.......--------W-------S-------V---NA-----------------G-C---R------
          MN --.-------.......----V---W-------S----S--V---NA---------------L-----------T--
          RF V-.-------.......--------W-------S-------V---N------------I---A--IL--F-------
          SC T-.-------.......--------W-------S---R---V-FVNA--------------LL---F-------T--
        BAIA T-.-------.......----F---W-------S-------V---N-I--------------L---V----------
       BK132 T-.-----E-.......--------W-------S-------V---NA---------------L--VY----N--T--
        BRVA T-.-------.......----V---W-------S-------V---NA-----G-R---A-------F----------
       CDC42 --.-------.......----V---W-------S-------V--VNV------------------IY--F-------
         SF2 --.T--I--H.......--------W-S-----I-------V-W-NA-----T---------A---Y------H---
         US1 T-.-------.......--------W-------S----T--VN--NA-------------------Y--L----T--
         US2 T-.T------.......--------W-------S-------VN--N----------------L--IY--FCN-----
         US3 --.-------.......--------W-------S-------V---NA------------------I--------T--
         US4 K-G-------.......----P---W-------S----S--V--FNA-----------I--I---IF--VI------
       WMJ22 K-.-------.......--------W-------SK------VG--NAI-----------------IC---I------

C      DJ259 --.A-----.......Q----T----GS-V----L---K------N----T-------I-KFI-GLW---CN-----
       DJ373 --.V-----.......Q----T----GS-V----L---K--V----------------I--II--IW--FCN-----
       SO145 --.A------SSLRELQ---------G--G----L---K--------I--T-------I--LA--IC-GVRN--T--
        ZM20 --.A------SSLRGLQ---------GS-V----L---K--------I----------I--L---IC---RN-----
        ZM18 --.AA-----SSLRGLQ---------GS-V----L---K-----V-AI--T-------I--FI--IC---RN-----
       UG268 --.A------SSLRGLQ---------GS-V--------K--------I.---S-----I---G-GI-----------
       SN364 --.A----I.......Q----I----GS-A----L-I-K--------I--T-------I--L---IC---SN-----

D        MAL T-.-------.......----------------------------N--------C------IG--F-----------
         ELI V-.-------.......---DI-----------S---R---S--F-AI-------------II---C--V-N-----
         JY1 T-.-------.......-----I----S-----T------F----NA-----------I--LIR--F--V------V
         NDK --.-------.......----------------S---R---S-----I------R-----------C----NV----
        Z2Z3 --.-------.......----------------SR------S-----I-------------I-R--C--V----T--
       KE124 T-.-------.......---------G------TR---------V--I-VV--GW--SI-KIG--L---F-------
       SN365 --.-------.......----V-----S-----I-K-----------I----------I-D-----C-------T--
       UG266 T-.-------.......--------W-------I--------N-FN------------A--I----V----N--T--
       UG274 --.---I---.......---K-I----------I----------F-------------I--I----V--V-N--T--
       UG269 T-.-------.......-----I-------K--I----------F--I--V-----------I---V--V-------

E      CM239 --.T-----HSSLKGLR----G----G---L-------I------N-------GW-------A-G-W----------
       CM240 --.T-----HSSLKGLR----G----G---L-------I-------A----A-GW-------A-G-W----------
       CM242 --.T------SSLKGLR----G----G---L-------I-----F-A------GW-------A-G-W----------
       CM243 --.A-----HSSLKGLR----G----G---L-------I------NA------GW--K----A-G-W----------
       CM235 --.T-K----SSLKGLR----G----G---L-------I-------A---I--GW-------A-G-W----------

F      BZ126 --.--NRGL-.......-------L-G--TL--------------N----V-T-----I-AAL--I-----N--T--
       BZ163 --.--NRGL-.......-------L-G--TL-----------G-FN----V-------I-AAL--I-----N--T--

G    LBV21-7 T-.T------NSLKGLRL---G--------L-----------N----V---T-NW-------A---C----N-----
    GA-VI525 --.T------.......-------------------------H--N---------------I--I-C--V-------

       ZM184 --.T-----HSSLKGLRL---G--H-----L----------------I------W---I--IG---C---RN--T--
```

FIG. 3—*Continued.*

dues in gp120 was observed (cysteine residues are denoted by asterisks in Fig. 3). Two cysteine residues in gp41 protein were conserved; the carboxy-terminal gp41 cysteine residue was replaced with phenylalanine in three isolates (BRVA, MN, and DJ373). Hypervariability was observed for the signal peptide, variable domains V1 through V5, and the carboxy-terminal region that flanks the V3 loop. Almost all length variation occurred in those domains. In general, gp41 was less variable than gp120 and showed more restricted length variation. A near-perfect conservation was found for gp120 amino acids believed to be involved in gp120-gp41 interaction (denoted by plus signs in Fig. 3) (15) and for the leucine zipper (denoted by superscript carets) (8). Conservation was more limited for the gp120-gp41 cleavage position, CD4 binding site (denoted by number signs) (26), and immunodominant domain in gp41. Most potential N glycosylation sites were very well conserved for isolates of all subtypes (not shown in Fig. 3). The V3 loop crown tetrapeptide was predominantly GPGR for isolates that belonged to subtypes B and F. GPGQ was most common for all other subtypes. GLGQ was observed for isolates ELI, JY1, Z2Z6, and SN365; isolate U455 had an unusual sequence of GSGQ. Sequences GTGR, GTGQ, GRGQ, and GLRQ were found for GA-VI525, UG274, UG266, and NDK respectively. The regions that flanked this tetrapeptide had more-variable sequences and showed considerable length variation.

DISCUSSION

The population of the African continent is burdened with a significant portion of the global AIDS epidemic and, in the absence of effective measures to prevent new infections, at risk for increased incidence and prevalence of HIV-1 in coming years. The genetic diversity of HIV-1 in Africa is known to be substantial, but most current descriptions of the regional prevalence of variants lack a broad and representative sample base and are insufficiently detailed to provide a firm foundation for planning antiviral interventions. Here we have reported a genetic analysis of 21 additional HIV-1 isolates collected in Africa.

The genetic variability of internationally collected HIV-1 isolates in general, and of African isolates in particular, has been confirmed and extended by the data presented here. As a framework for analysis, we used sequences from the seven genetic subtypes of HIV-1 (designated A through G) for which full-length *env* DNA sequences were available. Most isolates in this study belonged to subtypes A, C, or D, while subtypes B, E, and F were not found. One isolate was classified as subtype G. Other studies have reported on *env* subtype G isolates collected in southern Russia (4) and Gabon (17). However, one isolate, ZM184, did not belong to any previously recognized subtype. Evidence for an additional *env* subtype (subtype H) has previously been obtained from isolates collected in Zaire and Cameroon (17); none of the isolates in this study belonged to this subtype. Thus, a relatively small collection of isolates contained members of four known subtypes and an isolate not related to currently recognized subtypes. However, no claims about the proportions of subtypes in any locale can be made since the initial sampling was not random and an additional selection for divergent viruses was made on the basis of *gag* sequencing information and PCR typing.

The impression that subtypes of HIV-1 are continuing their geographic spread has been reinforced. The geographic separation of the locations from which isolates of subtypes A, C, and D have been recovered has been broadened. Previously, subtype A isolates had been found in Rwanda, Uganda, and Zaire. We found subtype A isolates not only in Uganda and Zaire but also in neighboring Kenya and Djibouti, 1,000 miles to the northeast. Subtype C isolates, previously recovered from Zambia, Malawi, and South Africa, were also found in Uganda and, interestingly, 2,000 miles away in Djibouti and more than 4,000 miles away in Senegal. Similarly, subtype D isolates, first recognized among isolates from Zaire, were also found in Uganda, Kenya, and Senegal. The number of genetic subtypes recovered from some of the locales studied has also been extended. For example, isolates of three subtypes were recovered from Uganda, and isolates from Zambia included subtype C and, potentially, an additional subtype represented by ZM184. The expanding and overlapping geographic ranges of multiple genetic subtypes add new complexity to planning antiviral interventions that are influenced by genetic variation.

Data concerning genetic variation of HIV-1 are currently focused on two structural genes, *gag* and *env*, that differ widely in function and overall sequence conservation among HIV-1 and related primate lentiviruses. The full-length *env* sequences reported here greatly expand the compendium of such sequences and provide an opportunity to contrast and compare *gag* and *env* full-length sequence databases. Two main points emerge from this comparison. First, *gag* and *env* phylogenetic trees have essentially identical topologies, each with seven or more distinct, well-separated subtypes. Second, *gag* and *env* subtype assignments are largely congruent. Among the 38 isolates for which complete *gag* and *env* sequences could be found, 28 were assigned the same subtype. Five isolates from Thailand clustered with subtype A isolates when the *gag* gene was analyzed (20) but formed a different subtype, called E, in *env* analysis (24, 29). Similarly, three viruses from Africa exhibit a subtype A *gag* sequence but a subtype D *env* sequence; these are the early Zaire isolate MAL and two viruses from this study, KE124 and UG266, from Kenya and Uganda, respectively. Isolate ZR-VI191 was classified as subtype G for *gag* and subtype A for *env*; isolate GA-VI525 was classified as subtype H for *gag* and subtype G for *env* (29). Current data do not permit a complete interpretation of these findings, but dual infection of a single individual with HIV-1 viruses of more than one subtype or recent or historical intersubtype recombination events are among the possibilities.

It appears that HIV-1 subtypes may be unevenly distributed geographically and/or in different population subsets and may be present at widely different frequencies, making adequate sampling difficult. Epidemiologic surveillance of HIV-1 subtypes is an important and ongoing element of preparation for global antiviral interventions. The recent introduction of rapid subtyping methods, such as the heteroduplex mobility assay (7), may facilitate collection of this information in coming years.

REFERENCES

1. **Alizon, M., S. Wain Hobson, L. Montagnier, and P. Sonigo.** 1986. Genetic variability of the AIDS virus: nucleotide sequence analysis of two isolates from African patients. Cell **46:**63–74.
2. **Barre-Sinoussi, F., J. C. Chermann, F. Rey, M. T. Nugeyre, S. Chamaret, J. Gruest, C. Dauguet, C. Axler-Blin, F. Vezinet-Brun, C. Rouzioux, W. Rozenbaum, and L. Montagnier.** 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immunodeficiency syndrome (AIDS). Science **220:**868–871.
3. **Benn, S., R. Rutledge, T. Folks, J. Gold, L. Baker, J. McCormick, P. Feorino, P. Piot, T. Quinn, and M. Martin.** 1985. Genomic heterogeneity of AIDS retroviral isolates from North America and Zaire. Science **230:**949–951.
4. **Bobkov, A., R. Cheingsong-Popov, M. Garaev, A. Rzhaninova, P. Kaleebu, S. Beddows, M. H. Bachmann, J. I. Mullins, J. Louwagie, W. Janssens, G. van der Groen, F. McCutchan, and J. Weber.** Identification of a new env G subtype and heterogeneity of HIV-1 strains in the former Soviet Union. AIDS, in press.
5. **Bruce, C., C. Clegg, A. Featherstone, J. Smith, and J. Oram.** 1993. Sequence analysis of the gp120 region of the env gene of Ugandan human immuno-

deficiency proviruses from a single individual. AIDS Res. Hum. Retroviruses **9:**357–363.

6. **Burke, D. S., A. K. Fowler, R. R. Redfield, S. Dilworth, and C. N. Oster.** 1990. Isolation of HIV-1 from the blood of seropositive adults: patient stage of illness and sample inoculum size are major determinants of a positive culture. J. Acquired Immune Defic. Syndr. **3:**1159–1167.
7. **Delwart, E., E. G. Shpaer, J. Louwagie, F. McCutchan, M. Grez, H. Rübsamen-Waigmann, and J. I. Mullins.** 1993. Genetic relationships determined by a DNA heteroduplex mobility assay analysis of HIV-1 env genes. Science **262:**1257–1261.
8. **Dubay, J. W., S. J. Roberts, B. Brody, and E. Hunter.** 1992. Mutations in the leucine zipper of the human immunodeficiency virus type 1 transmembrane glycoprotein affect fusion and infectivity. J. Virol. **66:**4748–4756.
9. **Dumitrescu, O., M. L. Kalish, S. C. Kliks, C. I. Bandea, and J. A. Levy.** Unpublished data.
10. **Felsenstein, J.** 1989. PHYLIP—phylogenetic interference package (version 3.2). Cladistics **5:**164–166.
11. **Goodenow, M., T. Huet, W. Saurin, S. Kwok, J. Sninsky, and S. Wain Hobson.** 1989. HIV-1 isolates are rapidly evolving quasispecies: evidence for viral mixtures and preferred nucleotide substitutions. J. Acquired Immune Defic. Syndr. **2:**344–352.
12. **Guo, H. G., M. S. Reitz, R. C. Gallo, Y. C. Ko, and K. S. S. Chang.** 1993. A new subtype of HIV-1 gag sequence detected in Taiwan. AIDS Res. Hum. Retroviruses **9:**925–927.
13. **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 (MPV-5180) from Cameroon. J. Virol. **68:**1581–1585.
14. **Hahn, B., G. M. Shaw, M. E. Taylor, R. R. Redfield, P. H. Markham, S. Z. Salahuddin, F. Wong-Staal, R. C. Gallo, E. S. Parks, and W. P. Parks.** 1986. Genetic variation in HTLV-III/LAV over time in patients with AIDS or at risk for AIDS. Science **232:**1548–1553.
15. **Helseth, E., U. Olshevsky, C. Furman, and J. Sodroski.** 1991. Human immunodeficiency virus type 1 gp120 envelope glycoprotein regions important for association with the gp41 transmembrane glycoprotein. J. Virol. **65:**2119–2123.
16. **Higgins, D. G., A. J. Bleasby, and R. Fuchs.** 1992. CLUSTAL V: improved software for multiple sequence alignment. Comput. Appl. Biosci. **8:**189–191.
17. **Janssens, W., L. Heyndrickx, K. Fransen, J. Motte, M. Peeters, J. N. Nkengasong, P. M. Ndumbe, E. Delaporte, J. L. Perret, C. Atende, P. Piot, and G. van der Groen.** 1994. Genetic and phylogenetic analysis of env subtypes G and H in central Africa. AIDS Res. Hum. Retroviruses **10:**877–879.
18. **Laure, F., R. Leonard, K. Mbayo, Z. Lurhuma, N. Kayembe, C. Brechot, P. S. Sarin, M. Sarngadharan, F. Wong-Staal, and R. C. Gallo.** 1987. Genomic diversity of Zairian HIV isolates: biological characteristics and clinical manifestations of HIV infection. AIDS Res. Hum. Retroviruses **3:** 343–353.
19. **Louwagie, J., E. Delwart, F. McCutchan, E. Sanders-Buell, T. Brennan, R. Redfield, B. Ungar, G. Eddy, and D. Burke.** 1994. Genetic analysis of HIV-1 isolates from Brazil reveals the presence of two distinct genotypes. AIDS Res. Hum. Retroviruses **10:**561–567.
20. **Louwagie, J., F. McCutchan, M. Peeters, T. Brennan, E. Sanders-Buell, G. Eddy, G. van der Groen, K. Fransen, G. Gershy-Damet, R. Deleys, and D. Burke.** 1993. Phylogenetic analysis of gag genes from seventy international HIV-1 isolates provides evidence for multiple genotypes. AIDS **7:**769–780.
21. **Louwagie, J., F. E. McCutchan, G. van der Groen, M. Peeters, K. Fransen, P. Piot, G. M. Gershy-Damet, G. Roelants, H. Van Heuverswyn, G. Eddy, B. Ungar, and D. Burke.** 1992. Genetic comparison of HIV-1 isolates from Africa, Europe, and North America. AIDS Res. Hum. Retroviruses **8:**1467–1469.
22. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
23. **Mascola, J., J. Louwagie, F. McCutchan, C. Fischer, P. Hegerich, K. Wagner, A. Fowler, J. McNeil, and D. Burke.** 1994. Two antigenically distinct subtypes of HIV-1: viral genotype predicts neutralization immunotype. J. Infect. Dis. **169:**48–54.
24. **McCutchan, F. E., P. A. Hegerich, T. P. Brennan, P. Phanuphak, P. Singharaj, A. Jugsudee, P. W. Berman, A. M. Gray, A. K. Fowler, and D. S. Burke.** 1992. Genetic variants of HIV-1 in Thailand. AIDS Res. Hum. Retroviruses **8:**1887–1895.
25. **McCutchan, F. E., B. L. Ungar, P. Hegerich, C. R. Roberts, A. K. Fowler, S. K. Hira, P. L. Perine, and D. S. Burke.** 1992. Genetic analysis of HIV-1 isolates from Zambia and an expanded phylogenetic tree for HIV-1. J. Acquired Immune Defic. Syndr. **5:**441–449.
26. **McKeating, J. A., J. P. Moore, M. Ferguson, H. S. Marsden, S. Graham, J. W. Almond, D. J. Evans, and R. A. Weiss.** 1992. Monoclonal antibodies to the C4 region of human immunodeficiency virus type 1 gp120: use in topological analysis of a CD4 binding site. AIDS Res. Hum. Retroviruses **8:** 451–459.
27. **Meyerhans, A., R. Cheynier, J. Albert, M. Seth, S. Kwok, J. Sninsky, L. Morfeldt-Manson, B. Asjö, and S. Wain-Hobson.** 1989. Temporal fluctuations in HIV quasispecies in vivo are not reflected by sequential HIV isolations. Cell **58:**901–910.
28. **Murphy, E., B. Korber, M. C. Georges-Courbot, B. You, A. Pinter, D. Cook, M. P. Kieny, A. Georges, C. Mathiot, F. Barré-Sinoussi, and M. Girard.** 1993. Diversity of V3 region sequences of human immunodeficiency viruses type-1 from the Central African Republic. AIDS Res. Hum. Retroviruses **9:**997–1006.
29. **Myers, G., B. Korber, S. Wain-Hobson, R. Smith, and G. Pavlakiss.** 1993. Human retroviruses and AIDS 1993, a compilation and analysis of nucleic acid and amino acid sequences. Los Alamos National Laboratory, Los Alamos, N.Mex.
30. **Oram, J. D., R. G. Downing, M. Roff, N. Serwankambo, J. C. S. Clegg, A. S. R. Featherstone, and J. C. Booth.** 1991. Sequence analysis of the V3 loop regions of the env genes of Ugandan human immunodeficiency proviruses. AIDS Res. Hum. Retroviruses **7:**605–614.
31. **Ou, C. Y., S. Kwok, S. W. Mitchell, D. H. Mack, J. J. Sninsky, J. W. Krebs, P. Feorino, D. Warfield, and G. Schochetman.** 1988. DNA amplification for direct detection of HIV-1 in DNA of peripheral blood mononuclear cells. Science **239:**295–297. (Erratum, **240:**4849.)
32. **Ou, C. Y., Y. Takebe, C. C. Luo, M. Kalish, W. Auwanit, C. Bandea, N. de la Torre, H. D. Gayle, N. L. Young, and B. G. Weniger.** 1992. Wide distribution of two subtypes of HIV-1 in Thailand. AIDS Res. Hum. Retroviruses **8:** 1471–1472.
33. **Ou, C. Y., Y. Takebe, B. G. Weniger, C. C. Luo, M. L. Kalish, W. Auwanit, S. Yamazaki, H. D. Gayle, N. L. Young, and G. Schochetman.** 1993. Independent introduction of two major HIV-1 genotypes into distinct high-risk populations in Thailand. Lancet **341:**1171–1174.
34. **Potts, K. E., M. Kalish, T. Lott, G. Orloff, C. C. Luo, M. A. Bernard, C. B. Alves, R. Badaro, J. Suleiman, O. Ferreira, G. Schochetman, D. Johnson, C. Y. Ou, J. L. Ho, and Brazilian Collaborative AIDS Research Group.** 1993. Genetic heterogeneity of the V3 region of the HIV-1 envelope glycoprotein in Brazil. AIDS **7:**1191–1197.
35. **Saitou, N., and M. Nei.** 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:**406–425.
36. **Shaw, G. M., B. H. Hahn, S. K. Arya, F. E. Groopman, R. C. Gallo, and F. Wong-Staal.** 1984. Molecular characterization of human T-cell leukemia (lymphotropic) virus type III in the acquired immune deficiency syndrome. Science **226:**1165–1171.
37. **Starcich, B. R., B. H. Hahn, G. M. Shaw, P. D. McNeely, S. Modrow, H. Wolf, E. S. Parks, W. P. Parks, S. F. Josephs, R. C. Gallo, and F. Wong-Staal.** 1986. Identification and characterization of conserved and variable regions in the envelope gene of HTLV-III/LAV, and retroviruses of AIDS. Cell **45:**637–648.
38. **Stewart, C. B.** 1993. The powers and pitfalls of parsimony. Nature (London) **361:**603–607.
39. **Vanden Haesevelde, M., J.-L. Decourt, R. J. De Leys, B. Vanderborght, G. van der Groen, H. van Heuverswijn, and E. Saman.** 1994. Genomic cloning and complete sequence analysis of a highly divergent African human immunodeficiency virus isolate. J. Virol. **68:**1586–1596.
40. **Weniger, B. G., Y. Takebe, C. Y. Ou, and S. Yamazaki.** 1994. The molecular epidemiology of HIV in Asia. AIDS **8**(Suppl.)**:**13–28.
41. **Wong-Staal, F., G. M. Shaw, B. H. Hahn, S. Z. Salahuddin, M. Popovic, P. Markham, R. Redfield, and R. C. Gallo.** 1985. Genomic diversity of human T-lymphotropic virus type III (HTLV-III). Science **229:**759–762.