Nucleotide Sequence and Genome Organization of Biologically Active Proviruses of the Bovine Immunodeficiency-like Virus[1]

KEVIN J. GARVEY,* M. STEVEN OBERSTE,* JOHN E. ELSER,* MICHAEL J. BRAUN,† AND MATTHEW A. GONDA*,[2]

*Laboratory of Cell and Molecular Structure, Program Resources, Inc., NCI—Frederick Cancer Research Facility, Frederick, Maryland 21701; and †Laboratory of Molecular Systematics, Smithsonian Institution, Museum Support Center, Washington, D.C. 20560.

The complete nucleotide sequences and translations of major open reading frames (ORF) of two distinct, infectious, proviral molecular clones (106 and 127) of the bovine immunodeficiency-like virus (BIV), obtained from a single virus isolation, were determined and compared. The genomes of BIV 127 and 106 are 8482 and 8391 nucleotides (nt), respectively, in the form predicted for the viral RNA. The structural organization of the genomes of BIV 127 and 106 are identical to one another and most similar to that of the lentivirus subfamily of retroviruses. In addition to gag, pol, and env genes, the BIV genome contains five short ORFs between and overlapping pol and env in the "central region," a hallmark of the lentiviruses which is believed to play an important role in their pathogenesis. Three of the short ORFs in the central region of BIV have been identified by location and structural similarity to the nonstructural/regulatory genes (vif, tat, and rev) of other lentiviruses; we also discovered two unique ORFs, termed W and Y, which may serve as exons for novel genes. BIV does not have the nef gene found in primate lentivirus genomes. The proviral LTR of BIV 127 is 589 nt, contains regulatory signals for initiation, enhancement, and termination of viral transcription, and has sequences related to the Sp1 and NF-κB binding sites. A major deletion (87 nt) in the env gene and 2 minor deletions (2 nt each) in the R regions of the LTRs account for the smaller size of clone 106. Numerous point mutations were also present; some caused coding substitutions that were most prevalent in the env encoding ORF. These data suggest that, within a single virus isolate, BIV displays extensive genomic variation. These infectious clones of BIV represent well-defined tools with which to analyze the function of the various ORFs and to dissect the molecular mechanisms of replication and pathogenesis. © 1990 Academic Press, Inc.

INTRODUCTION

Lentiviruses are a widely disseminated group of exogenous, nononcogenic retroviruses, that cause chronic, multisystemic diseases with insidious outcomes; some of these viruses are of importance to veterinary and human medicine (Gonda et al., 1989; Haase, 1986). Lentiviruses have been isolated from a variety of mammalian species and include visna and progressive pneumonia viruses of sheep (Cutlip and Laird, 1976; Kennedy et al., 1968; Sigurdsson, 1954), equine infectious anemia virus (EIAV) (McGuire et al., 1987), caprine arthritis encephalitis virus (Crawford et al., 1980), simian immunodeficiency viruses (SIVs) from various primate species (Daniel et al., 1985; Kanki et al., 1985), feline immunodeficiency virus (FIV) (Pedersen et al., 1987), the human immunodeficiency viruses types 1 and 2 (HIV-1 and HIV-2) (Barré-Sinoussi et al., 1983; Clavel et al., 1986; Gallo et al., 1984), the etiologic agents of the acquired immunodeficiency syndrome (AIDS), and the bovine immunodeficiency-like virus (BIV), a recently characterized lentivirus from cattle (Braun et al., 1988; Gonda et al., 1987; Van Der Maaten et al., 1972).

The lentiviruses share certain biologic and pathologic characteristics (Gonda et al., 1989; Haase, 1986; McGuire et al., 1987). Their similarity is reinforced by the fact that they are genetically related (Chiu et al., 1985; Gonda et al., 1985, 1986, 1987; Sonigo et al., 1985; Yokoyama et al., 1988). Evidence of this relationship has been strengthened by demonstrations of cross-reactivity between antigens of different species of lentiviruses, including the HIVs (Pyper et al., 1984; Casey et al., 1985; Gonda et al., 1987; Montagnier et al., 1984; Olmsted et al., 1989a). Taken together, these findings suggest that the similarities between animal lentivirus-related diseases and AIDS have a genetic basis; thus, animal lentiviruses may prove useful as model systems for developing and testing novel approaches to prevent or abrogate lentivirus-induced lesions related to HIV infection.

The genomes of lentiviruses are some of the most complex among retroviruses (Gonda et al., 1989; Gonda, 1989, Haseltine, 1988). They typically and invariably have gag, pol, and env structural genes and, in addition, a number of nonstructural/regulatory genes.

Although the lentiviruses share certain characteristics of their biology and genome organization in common, the complement of nonstructural/regulatory genes found in the genomes of isolates from various species differ. Even between isolates of the same species, at the nucleotide level, the genomes are quite variable, particularly in the *env* sequences (Alizon *et al.*, 1986; Braun *et al.*, 1987; Benn *et al.*, 1985, Hahn *et al.*, 1986; Starcich *et al.*, 1986; Payne *et al.*, 1987; Fisher *et al.*, 1988; Saag *et al.*, 1988; Gonda, 1989). Investigations of the genetic and organizational diversity observed between viral isolates of the same and different species may provide for a better overall understanding of the molecular mechanisms of lentivirus replication, latency, and pathogenesis.

BIV is an infectious retrovirus that causes lymphadenopathy, lymphocytosis, central nervous system lesions, progressive weakness, and emaciation (Van Der Maaten *et al.*, 1972). Preliminary epidemiologic evidence suggests that BIV infection is widespread in cattle populations in the U.S. BIV has the morphology of a lentivirus, encodes a reverse transcriptase (RT) with a Mg^{2+} cation preference, and has immunologic cross-reactivity with HIV, SIV, and EIAV (Gonda *et al.*, 1987). Moreover, the detection of sequence homology in the highly conserved RT domain of *pol* conclusively demonstrates that BIV is a lentivirus, distinct from all previously characterized lentivirus isolates (Gonda *et al.*, 1987).

At present, there are no reports on the complete nucleotide sequence and genome organization of BIV. To further characterize BIV and to provide fundamental reagents for detailed molecular, pathogenetic, and epidemiologic studies, we recently derived two biologically functional proviral molecular clones, BIV 106 and 127 (Braun *et al.*, 1988). Both clones were isolated from a bacteriophage library prepared from DNA of cells that carried a BIV infection from a single isolation. Here we report on their nucleotide sequences, genomic organization and variability, and relationship to HIVs and other lentiviruses.

MATERIALS AND METHODS

Proviral molecular clones

The molecular cloning of biologically active proviruses from the genomic DNA of BIV-infected bovine cells has previously been described (Braun *et al.*, 1988). For sequence analysis, the lambda (λ) clones of BIV 106 and 127 were subcloned into plasmids. Briefly, λ BIV 106 was digested with the restriction enzymes *FspI* and *SpeI*. These enzymes do not cut in the proviral sequences, but do cut in the bovine host flanking sequences of the λ insert, adjacent to the viral long termi-

nal repeats (LTRs), to generate a fragment of 11 kilobases (kb). This insert was isolated from gels, the ends were made flush with Klenow fragment, and *ClaI* linkers were added using T4 DNA ligase. *ClaI* linkers were chosen for addition to the 5′ and 3′ termini because *ClaI* does not cut within the *FspI–SpeI* fragment (Braun *et al.*, 1988). The resulting modified DNA, containing the complete BIV 106 proviral DNA sequences, was cloned into the *ClaI* site of pBluescript (Stratagene) for further propagation and analyses. λ BIV 127 was digested with *ClaI* and a 17-kb fragment containing the BIV 127 proviral sequences was isolated from gels and cloned into the *ClaI* site of pBluescript (Stratagene) using a procedure similar to the one described above for BIV 106. The resulting plasmids contain functional BIV 106 and 127 proviruses and are called pBIV106[inf] and pBIV127[inf]. The biologic activity of the proviral inserts contained in these plasmids was determined by digesting the DNA with *ClaI* and microinjecting the digested material into permissive cells, as previously described for λ BIV clones (Braun *et al.*, 1988). Cells thus manipulated formed syncytia within 24–48 hr, and supernates from these cultures were positive for RT activity within 2 weeks. This biologic activity could be passed to other permissive bovine cell cultures by cell-free supernates, thus verifying the functionality of the pBIV106[inf] and pBIV127[inf] clones used for sequencing.

Nucleotide sequencing

The proviral genomes of pBIV106[inf] and pBIV127[inf] were sequenced by the dideoxynucleotide chain termination method (Sanger *et al.*, 1977) from random-shotgun libraries (Deininger, 1983) in the M13 vectors mp18 or mp19 (Norrander *et al.*, 1983) using an M13 universal primer and Sequenase version 1.0 or 2.0 (United States Biochemicals) according to the manufacturer's recommendations. The entire 11-kb *ClaI* fragment was used to generate a random-shotgun library for sequencing pBIV106[inf]. The 17-kb *ClaI* fragment of pBIV127[inf] was further digested by *SmaI* or *SmaI* and *XbaI* to generate smaller fragments free of excess flanking sequences (Braun *et al.*, 1988). The resulting fragments were a *SmaI* segment (1.2 kb) containing 5′ host flanking sequences and LTR, a *SmaI* to *XbaI* segment (5 kb) containing primarily *gag*, *pol*, "central region," and 5′ *env* sequences, and a *XbaI* to *XbaI* segment (3 kb) containing the remaining *env*, 3′ LTR, and host flanking sequences. Independent random-shotgun libraries for sequencing were made of the 5- and 3-kb segments and a previously described 9.5-kb *SmaI* clone (Braun *et al.*, 1988) encompassing nucleotides (nt) from both the 5- and 3-kb fragments of pBIV 127[inf]. The entire 1.2-, 3-, and 5-kb segments were also

cloned into the *Sma*I and/or *Xba*I site(s) of M13 mp18 or mp19 for site-directed sequencing and walking with oligonucleotide primers for sequence confirmation or to fill in gaps where necessary. Ambiguities and compressions were resolved by the use of *Taq* polymerase (Promega Corporation), with or without deaza-GTP, and Sequenase with dITP according to the manufacturers' recommendations. The entire proviral sequences for BIV 106 and 127 were determined on both strands and each base was sequenced an average of eight times. The possibility of clustered *Sma*I or *Xba*I sites at the junctions of the 1.2- and 5- or the 5- and 3-kb segments, respectively, of BIV 127 was ruled out by comparison to BIV 106 sequence generated from the random-shotgun library and by sequencing in M13 additional BIV 127 subclones, generated by digestion with restriction enzymes other than *Sma*I or *Xba*I.

Computer analyses

The nucleotide sequences of the proviral genomes were reconstructed using the computer programs of Staden (Staden, 1982). Nucleotide sequences and translations of BIV 106 and 127 were analyzed using the UWGCG suite of genetic analysis programs (Devereaux *et al.*, 1984) run on a DEC VAX 8600 (Advanced Scientific Computing Laboratory, NCI-FCRF). The published nucleotide sequences of the other lentiviruses [HIV-1 (Ratner *et al.*, 1987), HIV-2 (Guyader *et al.*, 1987), SIV$_{mac}$ (Chakrabarti *et al.*, 1987), SIV$_{agm}$ (Fukasawa *et al.*, 1988), visna virus (Braun *et al.*, 1987), and EIAV (Kawakami *et al.*, 1987)] used in analyses were obtained from GenBank. The peptide sequences were inferred from the nucleotide sequences using UWGCG Translate.

Virus and cell culture

Bovine leukocyte adherent cell (BLAC-20) cultures obtained from a 4-month-old calf were derived from the long-term passage of phytohemagglutinin-stimulated peripheral blood leukocytes (Gonda *et al.*, manuscript in preparation). BLAC-20 cells were used for the propagation of parental BIV stock [DNA from cells infected with the parental virus was used to isolate BIV 106 and 127 proviral molecular clones (Braun *et al.*, 1988)] and progeny obtained from cells productively infected with BIV 106 or 127 after introduction of the functional proviral DNA by microinjection. Cells were grown in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum, 1% penicillin/streptomycin, and 1% glutamine in a humidified atmosphere of 5% CO_2 at 37°. BIV-infected cultures were monitored for maximum syncytium induction before harvesting for total cellular RNA.

RNA isolation

Total cellular RNA was isolated from uninfected and from BIV 106-, BIV 127-, or parental stock-infected BLAC-20 cells by the guanidine isothiocyanate method (Chomczynski and Sacchi, 1987).

Primer extension

To determine the cap site of the viral RNA, a simulated strong stop cDNA was synthesized using an oligonucleotide (20-mer) primer with the sequence 5'-TGTTGGGTGTTCTTCACCGC-3', representing the complement of nt 186–205 of the BIV 127 viral RNA (186 nt from the predicted transcription initiation site), and total cellular RNA from BIV-infected cultures. Briefly, the oligonucleotide primer was labeled at the 5' end using T4 polynucleotide kinase and [γ-^{32}P]ATP (Maniatis *et al.*, 1982). Following phenol extraction and ethanol precipitation, 0.2 pmol of primer was combined with 5 μg total RNA from uninfected BLAC-20 cells or BLAC-20 cells infected with BIV 106, BIV 127, or parental stocks in annealing buffer consisting of 80% formamide, 0.4 *M* NaCl, 0.04 *M* Pipes, pH 6.4, and 0.001 *M* EDTA (Casey and Davidson, 1977). The primer and RNA were annealed by heating to 70° for 10 min, followed by slow cooling (60 min) to 37°. The annealing reactions were diluted 10-fold into RT buffer (0.05 *M* Tris–HCl, pH 8.5, 10 m*M* MgCl$_2$, 0.04 *M* KCl, 1 m*M* DTT, and 0.5 m*M* each dNTP), Mo-MuLV RT (200 U) (Bethesda Research Labs) was added, and the reactions were incubated at 37° for 60 min (Calzone *et al.*, 1987). The reactions were phenol extracted, ethanol precipitated, and analyzed on a 6% polyacrylamide sequencing gel containing 7 *M* urea. A sizing ladder of BIV 106 sequence was created by sequencing an M13 template that spanned the U$_3$–R junction, R and U$_5$ elements, and the U$_5$-untranslated region, using the same labeled 20-mer as in the primer extension reactions and electrophoresing it in lanes adjacent to the primer extension reactions (Sanger *et al.*, 1977).

Evolutionary analyses

The amino acid sequences from a highly conserved segment of the RT domain of the *pol* gene of a series of retroviruses were aligned by the method of Dayhoff (1978). The frequency of matching residues (*M*) was calculated from the optimal pairwise alignments, counting gaps as mismatches. Phylogenetic trees were constructed from the distance values (−log *M*) by the method of Fitch and Margoliash (1967) as previously described (Gonda *et al.*, 1986). The retroviral sequences used in these analyses are referenced in Gonda *et al.* (1986, 1987, 1989) with the exception of

SIV$_{sm}$ (Hirsch et al., 1989), SIV$_{mne}$ (Gonda et al., manuscript in preparation), FIV (Olmsted et al., 1989b), and ovine progressive pneumonia virus (OPPV) (Garvey et al., manuscript in preparation).

RESULTS

Nucleotide sequence of BIV 106 and 127

The complete nucleotide sequences and translations of BIV 106 and 127 are depicted in Fig. 1. The genome of BIV 127 is 8482 nt in the form of the viral RNA; BIV 106 is slightly smaller at 8391 nt. This difference is explained by the fact that BIV 106 has suffered a major (87 nt; nt positions 5819–5905) and two minor (4 nt; nt positions 99–100 and 8470–8471) deletions relative to the BIV 127 genome. Numerous substitutions are also prevalent as discussed below (Table 1). Because BIV 127 contains a significantly greater amount of genetic information, we tentatively have designated it to be the prototypic sequence from which BIV 106 was derived.

Genetic organization

The genetic organization of BIV (Fig. 2) is most similar to that of other lentiviruses. There are three large and several short open reading frames (ORFs). The overall topography for the BIV genome is

5'LTR-gag-pol-"central region"-env-3'LTR.

The three larger ORFs encode genes with structural or enzymatic function and are in the invariant order gag, pol, and env, present in all replication-competent retroviruses. The gag and pol genes are in different reading frames and overlap. The pol and env genes are in the same reading frame and are separated by the "central region," the hallmark of lentiviruses (Braun et al., 1987; Chakrabarti et al., 1987; Fukasawa et al., 1988; Guyader et al., 1987; Ratner et al., 1987; Sonigo et al., 1985). The 5'-most 2 nt deletion (R region of the 5' LTR) in BIV 106 has caused a frame shift in all ORFs relative to BIV 127. Despite the deletions and substitutions in BIV 106, it has exactly the same overall organization as BIV 127; none of the ORFs appears to be truncated and no additional ORFs were created (Fig. 2).

Structure of the LTR

The proviral LTRs of BIV 127 and 106 are 589 and 587 nt long, respectively. In addition to the 2 nt deletion in BIV 106, there are four substitutions that distinguish the LTRs of the two proviral clones. The boundaries of the LTR and its internal U$_3$, R, and U$_5$ elements were determined by sequence analysis and biochemical experiments to complement the structural analysis (see below). The BIV 127 LTR is diagrammatically depicted with annotations in Fig. 3. The boundary distinguishing the retroviral LTR from internal viral sequences is indicated by the presence of the short inverted repeat of the sequences 5'ACTG3' located after a polypurine tract in the 3' end of the genome and 5'CAGT3' located before a sequence complementary to the 3' end of the binding site for a tRNA that is used to initiate minus-strand synthesis by the viral RT. The short inverted repeats extend into the viral LTRs in the U$_3$ and U$_5$ regions for 7 nt as longer imperfect inverted repeats that correspond with sequences found at 5' and 3' ends of the LTRs at the cell DNA junction (Fig. 3). The tRNA used by BIV is tRNA $_{1,2}^{Lys}$ and is identical to that used by visna virus (Braun et al., 1987; Sonigo et al., 1985).

Retroviral LTRs classically contain a long direct repeat between the U$_3$ and U$_5$ elements. For BIV 127, this element is 111 nt (109 nt for BIV 106), and structurally defines the U$_3$–R and R–U$_5$ boundaries. Strong stop cDNA was synthesized using a radiolabeled oligonucleotide primer, RT, and viral RNA from cells independently infected by BIV 106, BIV 127, and parental virus. The primer extension reactions were compared to a dideoxy sequence ladder encompassing the BIV 106 LTR to the untranslated region 5' of the gag ORF (see Materials and Methods) to further deduce the size of U$_5$ + R elements and the putative position of the U$_3$–R junction corresponding to the cap site in the viral RNA (Fig. 4). The cap sites in BIV 106 and 127 were determined to be at the first T in the R element of the LTR. The length of R + U$_5$ is 203 ± 1 nt for BIV 106 and 205 ± 1 nt in BIV 127 and parental stocks. This value is in agreement with assignments based on structural analyses and confirms the deletion of 2 nt in the BIV 106 LTR. By computation, the deduced U$_3$ and U$_5$ elements are therefore 384 and 94 nt, respectively, for both BIV 127 and 106.

In the U$_3$ region (Fig. 3), putative transcription-regulating signals were identified by analogy to previously defined lentivirus LTR signal sequences. Potential promoter sequences for the TATA and CAT boxes were identified. We also identified by sequence homology a conserved sequence for a retroviral core enhancer and NF-κB binding site similar to the one within the enhancer of the κ immunoglobin chain gene. Sequence homology to Sp1 binding sites (nt 8294–8302) related to that of the human metallothionein gene were found between the CAT and TATA boxes and an additional sequence homologous to an Sp1 binding site in the enhancer elements of SV40 was found in U$_5$ (nt 181–189) (Fig. 1). Sp1 and NF-κB binding sites have been detected in the U$_3$ region of the LTRs of HIV-1, HIV-2, and/ or SIVs (Gaynor et al., 1988; Guyader et al., 1987) and an NF-κB site, in the U$_3$ region of FIV (Olmsted et al.,

1989a). Also identified were sequences (nt 8123–8131) related to the glucocorticoid receptor binding site observed in the mouse mammary tumor virus LTR (Wingender, 1988). The functional significance of these structurally identified transcription-modulating sequences remains to be determined.

In the R region (Fig. 3), the signal sequence for polyadenylation (AATAAA) was present 87 bp from the cap site and a poly(A) addition site (CA) was identified 18 nt from the signal sequence, marking the end of R and the R–U_5 boundary. Interestingly, surrounding the cap site is a series of AluI restriction enzyme recognition sequences. It is tempting to speculate that this repeated motif surrounding the cap site may be useful in providing structural conformation for regulation of transcription.

gag, pol, and env ORFs

The predicted molecular weights for the precursors of the major ORFs (Fig. 2) of BIV were calculated from the translation of the DNA sequence of BIV 127 (Fig. 1). The processed protein products of the gag, pol, and env ORFs were deduced by homology with known proteins of HIV and other lentiviruses, conserved structural features such as Cys amino acid (aa) residues, and hydropathy plots (Table 2). For conformity, the proposed nomenclature and acronyms for retroviral proteins recommended by Leis et al. (1988) have been used for BIV gag, pol, and env gene products.

The gag precursor of BIV has a calculated molecular weight of 53 kilodaltons (kDa) (Table 2), which is consistent with the p53 antigen detected on Western blots (Gonda et al., 1987) and by radioimmunoprecipitation of a similar sized protein from BIV-infected cells (Battles et al., manuscript in preparation) and insect cells infected by a recombinant baculovirus containing only the gag ORF of BIV (Rasmussen et al., manuscript in preparation) using rabbit polyvalent antisera raised to purified BIV and natural or experimental sera from BIV-infected cattle. The p53 N-terminal sequence begins with the sequence Met–Lys–Arg–Arg and is not compatible with myristoylation since the second residue of the N-terminal sequence required for addition of myristic acid is Gly, as found in HIV-1 (Mervis et al., 1988).

BIV p53 is predicted to be processed into three proteins, p15, p25, and p13 (Table 2), by analogy to cleavage products of HIV (Lillehoj et al., 1988; Mervis et al., 1988) and from analysis of alignments between BIV and HIV, EIAV, and other lentivirus gag proteins. Proteins of p17, p26, and p14 have been serologically recognized in purified virion preparations and BIV-infected cells using antisera that recognize gag proteins of BIV (Battles et al., manuscript in preparation; Rasmussen et al., manuscript in preparation). The molecular weights for the processed gag gene products determined in SDS–polyacrylamide gel electrophoresis experiments of immunoprecipitated BIV are close to those predicted (Table 2); therefore, to maintain consistency with protein studies, we have chosen to use the experimentally determined molecular weights when referring to these proteins by size.

The matrix protein (MA), p17, would reside on the amino terminus side of the capsid protein (CA), p26. At the carboxy terminus of the gag precursor and within the predicted p14 domain reside two copies of a Cys-rich motif ($Cys–X_2–Cys–X_4–H–X_4–Cys$) (gag aa 405–418 and 423–436) reminiscent of the "zinc finger" seen in nucleic acid-binding proteins of other retroviruses (Berg, 1986; Covey, 1986; Evans and Hollenberg, 1988; Henderson et al., 1981; Oroszlan and Copeland, 1985). Moreover, the nucleic acid-binding or nucleocapsid (NC) proteins of retroviruses usually reside at the carboxy terminus of the CA protein; therefore, we feel that p14 is the NC protein of BIV. Amino acid sequencing and functional studies of recombinant and native proteins of BIV will be helpful in providing additional confirmation of these assignments.

Segments of the pol gene, in particular RT and endonuclease/integrase (IN), are the most conserved in the evolution of retroviruses, and homologies between the amino acid sequences of the predicted translated proteins of various species are readily detected (Chiu et al., 1985; Gonda, 1989; Gonda et al., 1989, 1986; Sonigo et al., 1985). In the pol ORF are the protease (PR), RT, and IN coding regions of the virus (Table 2). The pol precursor of BIV is predicted to be 121 kDa and is synthesized as a polyprotein with the gag gene precursor presumably by ribosomal frameshifting. The predicted polyprotein comprising gag and pol precursors is 174 kDa; a protein of this size has been demonstrated in radioimmunoprecipitations of bovine cells infected with BIV (Battles et al., manuscript in preparation) and recombinant baculovirus-infected insect cells containing the entire gag/pol ORFs (Rasmussen et al., manuscript in preparation) using antisera with reactivity to BIV gag. The BIV PR (p11), RT (p72), and IN (p32) proteins have been identified in the translation of the pol ORF (Fig. 1) by homology to those of HIV and other lentiviruses. The pol ORF precursor (p121) and processed products (PR, RT, and IN) have not yet been immunologically demonstrated.

The immunologic cross-reactivity of the gag gene products of BIV and HIV-1 has been localized to the CA proteins, p24 and p26 of HIV and BIV, respectively (Gonda et al., 1987). In molecular hybridization experiments, positive hybridization of cloned lentiviral sequences has been found mainly with the pol gene

```
4351 GAGTTGTAGAAAGAGCCCAGAGATCTAAAAGACAGATTGGCAGCTTATCAGGGAGATGTGAAACCGTAGAAGCAGCCCTTAGCCTCGCATTAGTTTCTTTAAATAAAAAAAAGAGGGGGAATAGGGGGCCATACACCATATGAAATAT 4500
          V V E R A H R D L K D R L A A Y Q G D C E T V E A A L S L A L V S L N K K R G G I G G H T P Y E I Y

4501 ACCTAGAATCAGAACATACCAAATACCAAGACCAACTAGAACAACAATTTTCAAAACAAAAAATTGAAAAGTGGTGTTACGTAAGGAACAGAAGAAAGGAATGGAAAGGACCCTACAAAGTGTTGTGGGACGGAGACGGGGCAGCAGTAA 4650
                                                                                   start vif→ M E R T L Q S V V G R R R G S S H
          L E S E H T K Y Q D Q L E Q Q F S K Q K I E K W C Y V R N R R K E W K G P Y K V L W D G D G A A V I

4651 TAGAGGAAGAGGGAAAAACAGCCTTATATCCACACCGTCATATGCGCTTCATCCCCCCCCCCAGATTCAGATATCCAAGATGGGAGTTCGTGAGGCAGACAGAATACAGCATGACCGCGTGCGTAAGAAAAGGGAAATTAGTCCTTACTTA 4800
                                                                                ↓sa
                                                       start W→ M G V R E A D R I Q H D R V R K K R E I S P Y L
          R G R G K N S L I S T F S Y A L H P P P F R Y P R W E F V R Q T E Y S M T A C V R K G K L V L T Y
          E E E G K T A L Y P H R H M R F I P P P D S Q I Q D G S S * end pol

4801 CCAGTACGCGATCTGGAAAAGAGTCTGGACGATAGAAACAGGATTTACAGATCCAAGTCTGTTTATGACCCCAGCTGGAACACACACCACTGAAGAAATAGGTCACTTAGATCTCTTTTGGCTTAGGTACTGTTCATGTCCGCATGAGAT 4950
                                                                              A
          P V R D L E K S L D D R N R I Y R S K S V Q D P S W N T H H * end W
          Q Y A I W K R V W T I E T G F T D P S L F M T P A G T H T T E E I G H L D L F W L R Y C S C P H E M

4951 GCCCCCCGTGGCTAGACTTCCTTAGAGGCACCCTCAATCTACGCATTTCCTGTCGACGCGCTCTTCAAGCGTCAGTGTTGACTAGCACCCCTAGACACTCCCTCCAACGCTTAGCTGCACTTCAGCTGTGCACTAACGCATGTCTCTGTTG 5100
                                                                              A                                    ↓sa .
          P P W L D F L R G T L N L R I S C R R A L Q A S V L T S T P R H S L Q R L A A L Q L C T N A C L C W
                                                                                        start Y→ M S L L

5101 GTACCCGTTAGGACGCATCAACGACACCACCCCGTTGTGGTTGAACTTTTCGTCTGGGAAGGAACCAACGATCCAACAACTGAGTGGCCACCCCTAACTCGTCGTAACATTCATAGATTGTGGCAATATGCCCGGACCTTGGGTGGCGAT 5250
          V P V R T H Q R H H P V V V E L F V W E G T N D P T T E W P P L T R R N I H R L W Q Y A R T L G G D
          Y P L G R I N D T T P L W L N F S S G K E P T I Q Q L S G H P * end vif   start tat exon I→ M P G P W V A M

5251 GATAATGTTGCCACAGCCCAAAGAAAGCTTTGGAGGAAAGCCAATTGGCTGGCTTTTCTGGAACACGTGCAAAGGACCTAGGCGGGACTGTCCACATTGTTGTTGTCCCATATGTAGTTGGCATTGTCAGCTTTGCTTTTTGCAGAAAAA 5400
                                                                                                         ↓sa
          D N V A T A Q R K L W R K A N W L A F L E H V Q R T * end Y
          I H L P Q P K E S F G G K P I G W L F W N T C K G P R R D C P H C C C P I C S W H C Q L C F L Q K N

5401 TCTAGGAATCAACTATGGATCAGGACCTAGACCGCGCGGAACGCGGGGAAAGGGGAGGAGGATCCGAAGAACTGCTTCAGGAGGAGATCAACGAAGGGAGGCTGACAGCCAGAGAAGCTTTACAAACATGGATCAATAACGGTGAGATCC 5550
            G                                           ↓sd                                            A
          L G I N Y G S G P R P R G T R G K G R R I R R T A S G G D Q R R E A D S Q R S F T N H D Q * end tat exon I
          start env→ M D Q D L D R A E R G E R G G G S E E L L Q E E I N E G R L T A R E A L Q T W I N N G E I H
          start rev exon 1→

5551 ACCCTTGGGTCCTGGCAGGAATGCTGTCCATGGGAGTAGGAATGCTACTAGGAGTATATTGTCAGTTACCAGACACACTGATTTGGATACTAATGTTTCAATTATGCCTTTATTGGGGTTTGGGTGAAACATCTAGAGAATTAGACAAGG 5700
          P W V L A G M L S H G V G M L L G V Y C Q L P D T L I W I L M F Q L C L Y W G L G E T S R E L D K D

5701 ATAGTTGGCAGTGGGTCAGAAGTGTATTTATAATAGCAATATTGGGAACTCTCACTATGGCAGGAACTGCTTTGGCCGACGACGATCAAAGTACTTTAATCCCCAATATCACAAAAATTCCTACAAAGGACACGGAACCCGGTTGCACCT 5850
          S W Q W V R S V F I I A I L G T L T M A G T A L A D D D Q S T L I P N I T K I P T K D T E P G C T Y

5851 ATCCGTGGGATATTAATCCTCTTGATTTTGGCTTTCACTCTGGGAATTCTGGGTATAATACTTGTCTTGGAGACGCAGCAACTCGGAGGATATATTGGCAGCCAGAGATACCATAGATTGGTGGCTCTCAGCTAATCAGGAAATACCTCCAA 6000
          P W I L I L L L I L A F I L G I L G I I L V L R R S N S E D I L A A R D T I D W W L S A N Q E I P P K

6001 AGTTTGCTTTCCCAATAATATTAATATCTTCCCCTCTAGCAGGCATAATAGGATATATTATGTCATGGAAAGGCACTTAGATGACTTCAAAAAGGGATGTCAAATTTGTGGGAGCCTGAGCCATGTGGGGAATGCTTTTGGAAGAAATTG 6150
                                                                                                  CT
          F A F P I I L I S S P L A G I I G Y Y V M E R H L E I F K K G C Q I C G S L S S M W G M L L E E I G

6151 GCAGGTGGCTCGCACGTAGGGAATGGAATTGTTAGTAGAGTAATGGTTTATCCTCTTAATCAGCTTCAGTTGGGGAATGTATGTCAATAGGGTAAATGCCTCAGGGTCACATGTAGCCATGGTCACCAGCCTCCAGGGTACCGCATAGTGA 6300
            C A              G                        G                        C                    G
          R W L A R R E W N V S R V M V I L L I S F S W G M Y V N R V N A S G S H V A M V T S P P G Y R I V N

6301 ATGATACCAGCCAGGCACCTTGGTATTGCTTCTCCTCGGCACCAATCCCAACGTGTAGTTCCTCTCAGTGGGGAGACAAATATTTGAGGAGAAAATAAACGAGACACTGGTCAAACAGGTGTATGAACAGGCCGCGAAACATTCGAGAG 6450
                                                                                                    C
          D T S Q A P W Y C F S S A P I P T C S S S Q W G D K Y F E E K I N E T L V K Q V Y E Q A A K H S R A

6451 CCACATGGATTGAACCTGATCTATTGGAGGAAGCAGTCTATGAGCTAGCTCTGTTATCAGCTAATGACAGTCGTCAGGTGGTGGTAGAAAATGGTACAGACGTATGTAGCTCACAGAACTCGAGCACAAACAAAGGCCACCCAATGACGC 6600
            G              G                      T                                T
          T W I E P D L L E E A V Y E L A L L S A N D S R Q V V V E N G T D V C S S Q N S S T N K G H P M T L

6601 TTCTAAAGTTGAGAGGGCAGGTGTCAGAAACTTGGATAGGGAATTCCTCCCTCCAGTTTTGTGTCCAGTGGCCATATGTCTTGGTAGGTCTTAATAATAGTGATAGTAATATTAGCTTCAATTCGGGAGATTGGATAGCAACCAATTGTA 6750
          L K L R G Q V S E T W I G N S S L Q F C V Q W P Y V L V G L N N S D S N I S F N S G D W I A T N C M

6751 TGCACCCAATTACACTAAATAAAAGTGCACAAGATCTAGGAAAAAATTTTCCGAGACTAACATTTCTTGACGGACAACGTGTCCCAGTTGAAGAACACACTGTGCGGACATAACACAAACTGTTTGAAATTTGGAAACAAGTCCTTCAGTA 6900
                                                     G                                                  C
          H P I T L N K S A Q D L G K N F P R L T F L D G Q L S Q L K N T L C G H N T N C L K F G N K S F S T

6901 CAAATTCCCTAATACTATGCCAAGACAACCCCATCGGCAACGACACCTTTATAGCCTAAGTCATTCCTTCTCAAAACAGGCCTCTGCCCGGTGGATTCTTGTAAAGGTCCCCAGCTATGGGTTTGTGGTAGTAAATGACACAGATACAC 7050
                                                                G
          N S L I L C Q D N P I G N D T F Y S L S H S F S K Q A S A R W I L V K V P S Y G F V V V N D T D T P

7051 CACCCATCCCTCCGCATCCGAAAGCCTCGAGCAGTCGGACTAGCAATATTCCTGCTTGTGCTGGCTATCATGGCCATCACATCTCCTTGGTGGCAGCTACAACGCTCGTGAACCAGCACACGACGGCTAAGGTTGTGGAGAGGGGTTGTGC 7200
          P S L R I R K P R A V G L A I F L L V L A I M A I T S S L V A A T T L V N Q H T T A K V V E R V V Q
                       SU —┴— TM

7201 AAAATGTGTCATATATTGCTCAAACCCCAGGACCAATTCACCCACCTGTTCAGGAATATAAACAACAGATTAAATGTCCTACACCATAGAGTTTCATACTTGGAGTATGTAGAGGGAAATCAGACAAAAACAAGTATTCTTTGGTTGCAAAC 7350
                                                                                                  G
          N V S Y I A Q T Q Q Q F T H L F R N I N N R L N V L H H R V S Y L E Y V E E I R Q K Q V F F G C K P

7351 CTCATGGAAGGTATTGCCACTTTGACTTTGGACCAGAGGAAGTTGGATGGAACAATAGTTGGAATAGCAAAACTTGGAATGATCTACAAGATGAGTATGATAAGATAGAAGAAAAAAATATTAAAAATTCGAGTGGACTGGCTCAATAGCT 7500
                                                                      G
          H G R Y C H F D F G P E E V G W N N S W N S K T W N D L Q D E Y D K I E E K I L K I R V D W L N S S

7501 CCCTGAGTGACACACAGGACACCTTTGGCCTGGAGACCTCTATTTTTGACCATTTAGTGCAATTGTTTGATTGGACTTCTTGGAAAGACTGGATAAAAATCATTATAGTAATCATTGTACTTTGGCTTCTGATAAAGATTCTCCTAGGTA 7650
                                      start rev exon 2→ L D F L E R L D K N H Y S N H C T L A S D K D S P R Y
          L S D T Q D T F G L E T S I F D H L V Q L F D W T S W K D W I K I I I V I I V L W L L I K I L L G H

7651 TGTTAAGAAGCTGCGCCAAGGTCAGCCAGAATTACCAACATCTCCCGGCGGAGGAGGAGGACGGGGACACAGAGCCAGAAAGCTCCCCGGCGGAGGAGGACCCGGCTTCTGGAAGTCTCTACGAGAATTGGTTGAACAAAATAGGAGAAA 7800
          tat exon 2→ E A A P R S A R I T N I S R R R R R T G T Q S Q K A P R R E E T R L L E V S T R I G * end tat exon 2
          V K K L R Q G Q P E L P T S P G G G G G R G H R A R K L P G E R R P G F W K S L R E L V E Q N R R K
          L R S C A K V S Q N Y Q H L P A E E E D G D T E P E S S P A R G D P A S G S L Y E N W L N K I G E S

7801 GCAAGAACGACGCCTATCGGGTCTGGACAGAAGAATACAACAGCTTGAGGATCTTGTTCGCCACATGTCGCTGGGATCTCCTGACCCCTCAACTCCTTCAGCTTCCGTTCTTTCTGTTAACCCTCCTGCTCAAACTCCTTTGGGACATCT 7950
                                                                                    C
          Q E R R L S G L D R R I Q Q L E D L V R H M S L G S P D P S T P S A S V L S V N P P A Q T P L G H L
          K N D A Y R V W T E E Y N S L R I L F A T C R W D L L T P Q L L Q L P F F L L T L L L K L L W D I F

7951 TCCGCCACGCTCCTATTTTAAACTTAAAAGGGTGGACTGTGTGGGCAGGGTGGGACCTCAGGACAACAGCAGCCCCGGACTTCCCATATGTGAATTGGACTGGATCCAGGGAACAAAATAACCCAGAAGGGGGATTAGACTCTGGGGCTT 8100
                   polypurine tract    →● U₃                              NF-κB sequence
          P P R S Y F K L K R V D C G A G W D L R T T A A P G L P E L D W I Q G T K * end rev exon 2
          R H A P I L N L K G W T V G Q G Q Q P P D F P Y V N W T G S R E Q N N P E G G L D S G A W

8101 GGTATGAAGGCCTGAGAGGTTCTCAGTAGATTGTAAGTCTTCGGCGGACTGCATGTCTGCACGTGAGACAGGAAATGTTTATCTTCTCAGCTGATGTGGTTAGGCCGATTACTGGAAACTAGACAACCTGATTCATTAGTGGTTAAGAT 8250
                                                                              core enhancer
          Y E G L R G S Q * end env                                           ┌──────┐
                                                                              │GTGGTTAG│

                                                                              TATA sequence   U₃ ↔● R
          CAAT sequence                                                       ┌──────────┐
          ┌────────┐
8251 TATGCATAAGTGCTCGCAATGATGTAGCTGCTTACGCTTGCTTACTCCGCCCTGAAACGCCTACCTTAACACGCAACACGCCCACCTGTAAGAATATATAACCATATCTTCACTCTGTACTTCAGCTCGTGTAGCTCATTAGCTCCGAG 8400
                 A                                                                                    R ●←
8401 CTCCCCCAACCTACAGCCTGAGAGGCACTGGCTCGGTTGGGTAGCCAGCCTTTCGGGTAATAAGGCTTGTTGGCATTCGGCA 8482
          poly(A) signal ▬    R ←●
          ┌──────┐
          │AATAAA│
```

FIG. 1—Continued

TABLE 1

Nucleotide Sequence Differences between BIV Strains 127 and 106

Functional unit	Nucleotide position	Base in strain 127	Base in strain 106	Amino acid in strain 127	Amino acid in strain 106	Functional unit	Nucleotide position	Base in strain 127	Base in strain 106	Amino acid in strain 127	Amino acid in strain 106
$R-U_b$	48	G	A	NC[a]	NC	env (SU) (continued)					
	99	T	—	NC	NC		6120, 6121	A, G	C, T	Ser	Leu
	100	T	—	NC	NC		6156, 6158	T, G	C, A	Trp	Arg
	167	T	G	NC	NC		6178	A	G	Asn	Ser
gag	327	G	A	Arg	Arg		6192	A	G	Met	Val
	365	C	T	Pro	Leu		6236	T	C	Asn	Asn
	666	T	A	Asp	Glu		6263	A	G	Val	Val
	1176	C	A	Pro	Pro		6443	T	C	His	His
	1338	T	C	Pro	Pro		6455	A	G	Thr	Thr
	1350	C	T	Asn	Asn		6473	A	G	Leu	Leu
	1395	T	C	Ser	Ser		6518	C	T	Asp	Asp
pol	2036	C	A	Thr	Thr		6560	C	T	Ser	Ser
	2390	T	C	Thr	Thr		6782	A	G	Gln	Gln
	3140	G	A	Glu	Glu		6894	T	C	Phe	Leu
	3305	A	T	Pro	Pro		6976	A	G	Lys	Arg
	3989	T	G	Pro	Pro	env (TM)	7285	A	G	His	Arg
	3991	A	G	Lys	Arg		7429	A	G	Asn	Ser
	4031	T	C	Cys	Cys		7698	G	A	Ala	Thr
	4044	A	G	Thr	Ala		7724	G	A	Glu	Glu
vif	4850	G	A	Asp	Asn		7877	T	C	Asp	Asp
	5029	G	A	Leu	Leu	rev (exon 2)	7698	G	A	Gly	Asp
W	4850	G·	A	Arg	Lys		7724	G	A	Ala	Thr
tat (exon 1)	5408	A	G	Ile	Val		7877	T	C	Ser	Pro
	5484	G	A	Gly	Glu	tat (exon 2)	7698	G	A	Arg	Arg
env (SU)	5484	G	A	Glu	Lys		7724	G	A	Ser	Asn
	5688	G	A	Glu	Lys	U_3-R	8344	A	G	NC	NC
	5819–5905[b]	87 nt	—	29 aa	—		8419	G	A	NC	NC
	6060	G	A	Val	Ile		8470	T	—	NC	NC
							8471	T	—	NC	NC

[a] No change.

[b] This represents the 87-bp deletion (29 aa) in BIV106 relative to BIV127 (see Fig. 1).

(Gonda et al., 1987; Olmsted et al., 1989a). Thus, it was anticipated that the predicted gag and pol proteins of BIV would have detectable amino acid homology to those of other lentiviruses. However, when BIV 127 translation products for gag and pol ORFs were globally compared to those of other lentiviruses, we found only a small amount of similarity when identical amino acid residues were calculated (range for gag = 23–29%; pol = 36–38%). The percentage of matching residues increased significantly when conservative amino acid substitutions were included (range for gag = 48–55%; pol = 59–61%). Nevertheless, at least for the pol ORF, there were regions of locally strong sequence conservation, especially in the RT and the IN domains of pol, as reported previously (Chiu et al., 1985; Gonda,

1989; Gonda et al., 1986, 1987, 1989; Sonigo et al., 1985).

To account for the immunologic cross-reactivity between BIV and HIV-1 gag proteins, a local comparison of the predicted amino acids for the CA proteins of this ORF was made. Only one stretch of 10 amino acids within the p26 of BIV was observed to have a high degree of similarity to HIV-1 p24; this area was also compared to several other lentiviruses (Fig. 5). Within this stretch, BIV shared 8 of 10 residues with EIAV, 7 of 10 with HIV-1, HIV-2, and SIV_agm, 6 of 10 with SIV_mac, and only 3 of 10 with visna virus. It is worth noting that only SIV_mac, HIV-1, and EIAV compete in heterologous competitive radioimmunoassays using radiolabeled HIV p24 and rabbit polyclonal antisera to BIV (HIV-2 and
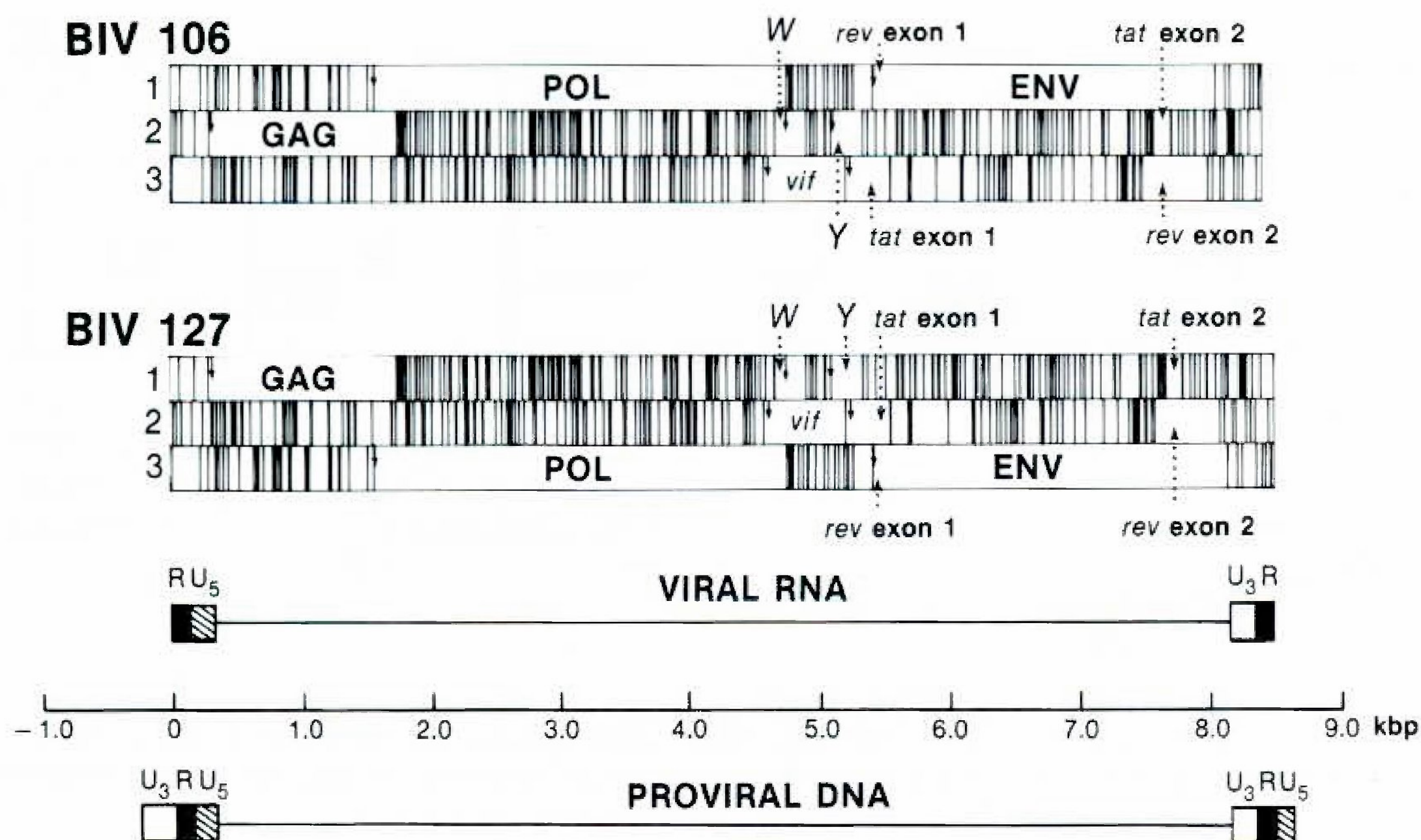
FIG. 2. Comparison of the genomic organization of BIV 106 and 127 biologically active clones deduced from ORF analysis of the nucleotide sequence presented in Fig. 1. Translation of the sequences in each reading frame commenced at nucleotide 1. Vertical bars represent stop codons. Arrows within the ORFs indicate the initiator AUG codon in viral genes or putative genes. ORFs annotated are *gag, pol, env, vif, tat* exons 1 and 2, *rev* exons 1 and 2, *W,* and *Y.* Location of LTR boundaries in the viral RNA and proviral DNA are indicated above and below the kb marker, respectively.

SIV$_{agm}$ have not been tested); and BIV and HIV-1 CA proteins do not exhibit immunological cross-reactivity with visna virus in Western blots or radioimmunopre-cipitations (Gonda *et al.,* 1987). The homologous re-gion recognized in the p26 of BIV *gag* (Fig. 5) may be the epitope(s) responsible for the immunologic cross-reactivity between BIV, EIAV, and the primate lentivi-ruses. Using computer algorithms to lend support to this hypothesis (Jameson and Wolf, 1988), we found this conserved segment of BIV 127 to have a high sur-face probability, a feature of peptides predicted to be potential antigenic determinants (data not shown).

The BIV *env* precursor is predicted to be 102 kDa in its unmodified form (Table 2). The first hydrophobic stretch of amino acids which might represent the leader sequence is located 50 amino acids into the translation of the *env* ORF in the vicinity of the second Met as deduced from hydropathy plots (data not shown). It has not been determined whether the sec-ond or the first Met in this ORF represents the begin-ning of the translation of the *env* protein (see discus-sion below on *rev*). The addition of sugar residues by N-linked glycosylation substantially increases the mo-lecular weight of the envelope protein backbone in len-tiviruses and may account for up to 50% of the appar-ent molecular weight, making them the largest retrovi-rus *env* products characterized (Pyper *et al.,* 1984; Robey *et al.,* 1985). The BIV 127 *env* precursor has 21

potential glycosylation sites. It is predicted to be cleaved into surface protein (SU) and transmembrane protein (TM) after the Arg–Lys–Pro–Arg site (aa 552–555; Figs. 1 and 6A). This putative cleavage site and the area around it are partially conserved (5 of 7 aa) with that of HIV-1 (Fig. 6A). The cleavage products of BIV 127 *env* are predicted to be 62 and 40 kDa for the SU and TM, respectively. Proteins in the range of 100 and 45 kDa have been detected in Western blots and 100- and 45-kDa proteins have been immunoprecipi-tated using serum from naturally or experimentally in-fected cattle (Gonda, unpublished data; Battles *et al.,* manuscript in preparation). The larger is believed to be the glycosylated SU and the smaller the TM of the virus by analogy to other lentiviruses. The precursor from which they are derived has not been identified.

The SUs of BIV 127 and HIV-1 are identical at about 13% of their amino acids. This limited degree of relat-edness is not unexpected considering the great amount of genetic variability which exists between different HIV-1 isolates, where the preponderance of variability falls within the SU (Alizon *et al.,* 1986; Star-cich *et al.,* 1986) and the limited identity (39%) ob-served between HIV-1 and HIV-2 SUs (Guyader *et al.,* 1987). The use of the SU–TM cleavage site as the benchmark with which to align the *env* precursors of lentiviruses has proven to be a useful means for iden-tifying the TMs of retroviruses (Braun *et al.,* 1987). We
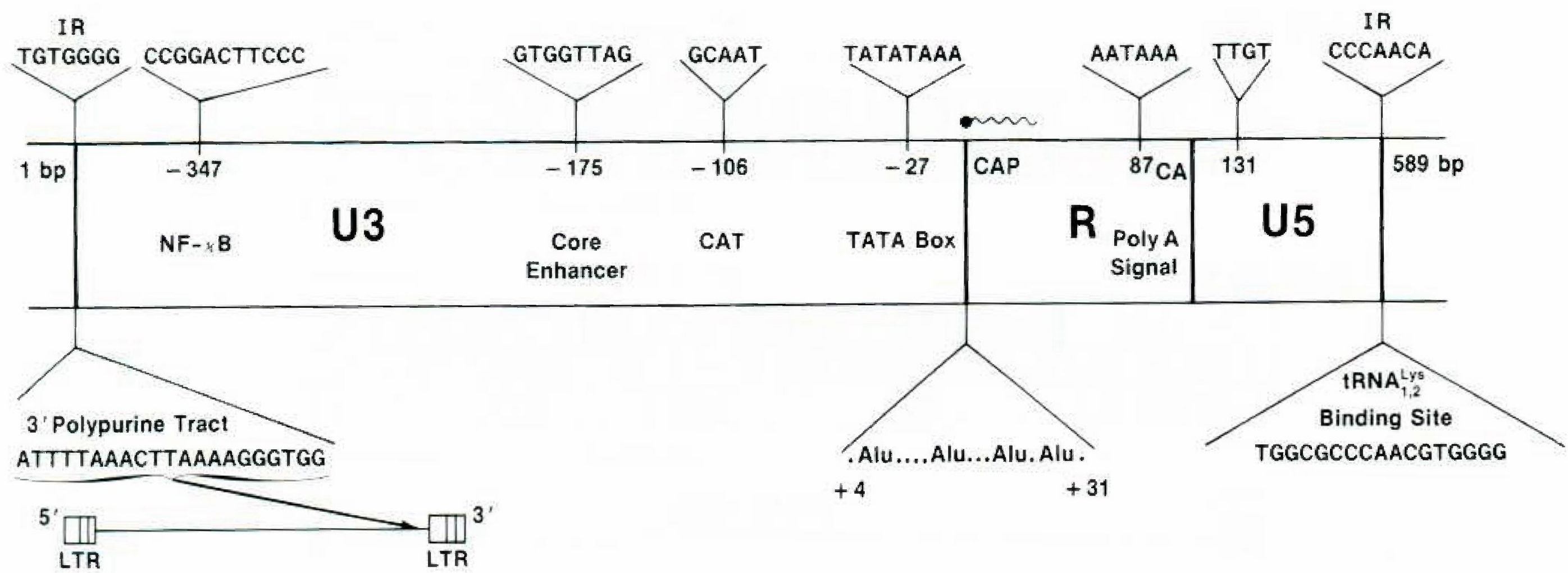
FIG. 3. Organization of BIV 127 LTR. Boundaries of the U_3, R, and U_5 regions are indicated as well as the nucleotide positions of the TATA and CAT boxes, core enhancer, polyadenylation signal, and poly(A) addition site (CA) at the R–U_5 boundary, cap site at the U_3–R boundary with repeated Alu recognition sites, tRNA primer binding site, polypurine tract, potential NF-κB site, and imperfect inverted repeat (IR) beginning at the dinucleotide (5'TG and 3'CA) of LTR boundaries.

have used a similar strategy to look for the TM of BIV by making structural comparisons between BIV and HIV-1. The TMs of lentiviruses have extracellular, transmembrane, and intracellular domains. Figure 6 shows the alignment of the exterior and transmembrane domains of the BIV and HIV-1 TMs, as determined by amino acid similarity and hydropathy plots. Overall, the amino acid identity is 25% for this TM segment, but is more limited in the intracellular domain (data not shown). Nevertheless, the hydropathy plots of the two TMs (Fig. 6B) show a striking resemblance. Moreover, there are several Cys residues conserved between BIV and HIV (Fig. 6A), as has been seen in other HIV-lentivirus comparisons (Braun et al., 1987; Kawakami et al., 1987).

Central region ORFs

The central regions of lentiviruses, especially those of the primate lentiviruses, are quite complex as they contain numerous nonstructural/regulatory genes (Haseltine, 1988). In HIV-1 and HIV-2, at least six different ORFs (vif, tat, rev, X, U, and R) that are actively transcribed and translated have been identified (Haseltine, 1988; Henderson et al., 1988; Strebel et al., 1988; Wong-Staal et al., 1987; Niederman et al., 1989). The X, U, and R products are termed vpx, vpu, and vpr; vpx is found only in HIV-2 and simian lentiviruses (with the exception of SIV_agm which lacks an X gene) and U is found only in HIV-1. A uniform nomenclature for the ORFs of the central region has been proposed (Gallo et al., 1988; Laurence, 1988); we follow that nomenclature here with BIV ORFs, where structural analogy or sequence similarity exist.

BIV has five small ORFs in this region that contain a potential translation initiation signal. The first is derived from sequences in the central region and overlaps the 3' end of pol at its 3' end (nt 4601–5194). Its position and size are similar to those of the gene encoding the viral infectivity factor (vif) described for HIV-1, HIV-2, SIVs, and visna virus (Braun et al., 1987; Chakrabarti et al., 1987; Fukasawa et al., 1988; Guyader et al., 1987; Ratner et al., 1987). There is little amino acid similarity to the vif ORFs of other lentiviruses; however, there is reasonably good conservation in hydropathy plots (data not shown).

The second short ORF initiates in the central region and overlaps the 5' end of the large env ORF (nt 5228–5536) in frame with sequences we have identified as vif. This ORF has an AUG start site that is preceded by a potential splice acceptor site. The translation of this ORF contains a Cys-rich region found in nucleic acid-binding proteins and the transactivator (tat) proteins of most lentiviruses (Fig. 7) (Braun et al., 1987; Chakrabarti et al., 1987; Guyader et al., 1987; Kawakami et al., 1987; Ratner et al., 1987). This Cys-rich region in BIV 127 shares 56% amino acid identity (no gaps) with HIV-1; identity is significantly lower when the rest of the translation of this ORF is evaluated. In addition, residues downstream of the Cys motif are very basic, consistent with the structure of a nucleic acid-binding protein. We take these two features, and the location of this ORF, as evidence that it represents a BIV tat ORF.

The tat genes of primate lentiviruses consist of two coding exons. The first is contained in the central region and the second in the 3' end of the genome over-

lapping env, but in a different reading frame from that of env and rev exon 2. The second tat exon is not considered necessary for tat function. The first tat exon is larger than the second. The mature tat message is derived from the primary transcript by a complex splicing mechanism (Haseltine, 1988). We searched for a second tat exon in BIV 127 in the translation of all three reading frames of the 3′ end of the BIV 127 env and found some amino acid similarity for a small ORF (nt 7657–7782) in reading frame 1 of the translation (Fig. 2) with the translation of the HIV-2 second tat exon (Guyader et al., 1987):

```
BIV 127    ... TNISRRRRRTGTQSQKAPRREETRLLEVSTR I G
               ||  |    ||   ||        |    | | |  |
HIV-2          KSISTR . . . TG . DSQPTKKQKKTVEATVETDTGPGR
```

There is an additional transactivating protein/regulator of viral gene expression found in lentivirus genomes called rev (Haseltine, 1988; Sodroski et al., 1986; Mazarin et al., 1988). It, like tat, is made up of two coding exons that are spliced into the mature mRNA from a primary transcript. The first rev coding exon usually is located near the 3′ end of the central region, often overlapping the 5′ end of env; the second exon is found in the 3′ end of env, overlapping, but in a different reading frame from that of the TM. The first exon of rev is usually smaller than the second exon. The rev protein is also a nucleic acid binding protein that contains stretches of basic amino acids highly enriched for Lys + Arg residues as has been seen in the rev of HIV-1 (Ratner et al., 1987; Sodroski et al., 1986) and in the hinge domain of nucleic acid binding proteins (Adler et al., 1988). The rev gene of visna virus, like that of the primate lentiviruses, is also transcribed as a large RNA precursor that is processed into mature form by multiple splicing (Davis et al., 1987; Davis and Clements, 1989; Mazarin et al., 1988). Thus, the rev gene of visna virus has two exons that contain Arg + Lys-rich regions, but without the characteristic Cys-rich region seen in tat. The visna virus rev exon 1 has been localized to the 5′ end of the env ORF and in frame with it; exon 2 overlaps the TM coding sequences, but is in a different reading frame (Mazarin et al., 1988).

There were no obvious short ORFs in, or overlapping, the 3′ end of the central region of BIV that appeared to be good candidates for the first rev exon of BIV. However, a potential ORF reminiscent of the second rev exon of visna virus and primate lentiviruses was identified in the 3′ end of the genome (nt 7571–8068) and in a different reading frame from that of env. It also is rich in Arg + Lys residues. These sequences proba-

bly represent BIV rev exon 2. The presence of the visna virus rev exon 1 in the env ORF prompted us to scan the analogous region in BIV for sequences that could represent rev exon 1. In the translation of the first 39 nt (nt 5415–5452) of the env ORF, there is some amino acid identity to rev exon 1 of HIV-2 (Guyader et al., 1987). The sequence for this 39 amino acid stretch begins Met–Asp–Gln–Asp–Leu and precedes the predicted hydrophobic leader sequence for env as in visna virus (Mazarin et al. 1988; Gonda, 1989). This rev exon is preceded by a potential splice acceptor site (nt 5396) and followed by a splice donor site (nt 5452), which would put these sequences in frame with those proposed for the BIV rev exon 2, when using a splice acceptor site at nt 7648. The predicted size for the product of this gene is 17 kDa (153 aa) (Table 2), which is about the same size as the 18-kDa (167 aa) rev gene product of visna virus, predicted from the translation of a cDNA clone (Mazarin et al., 1988).

There are two additional small ORFs in the BIV central region that contain a potential AUG start site. We have designated these two unique ORFs W (nt 4729–4890) and Y (nt 5089–5328). The predicted products for the W and Y ORFs are 54 and 80 amino acids (Table 2), respectively. ORF W contains a strong consensus sequence signal for initiation of translation at the first AUG in this ORF, while that of Y is weak (Kozak, 1986, 1989). In addition, there is a potential splice acceptor site immediately upstream of both W and Y. The presence of a strong consensus signal sequence in W suggests that this ORF may be actively translated, possibly from a spliced message. The fate of Y is more speculative at the moment. Interestingly, W and Y are located in a position in the central region analogous to vpr-, vpx-, and vpu-encoding ORFs of HIV-1 and/or HIV-2. No amino acid or nucleotide similarities for BIV ORFs W or Y were found with the ORFs of the central region of primate or ruminant lentiviruses. However, we noted that Y has a significant number of basic amino acids and thus, based on this property, appears to bear some resemblance to a nucleic acid-binding protein.

Genomic variability in the BIV 106 and 127 molecular clones

The sequencing of the two functional proviral clones of BIV presented us with the opportunity to examine the amount and location of diversity between them. There are 137 nucleotide changes (Table 1); 91 of these are deletions in BIV 106 and 87 of these occur at a single site in the envelope. The BIV 106 U_3, R, and U_5 elements (Fig. 1) have 8 nucleotide changes and 4 of these are deletions. There are no deletions in BIV 127 relative to BIV 106. Of the 46 nucleotide substitutions, 42 are in putative structural or nonstructural/reg-
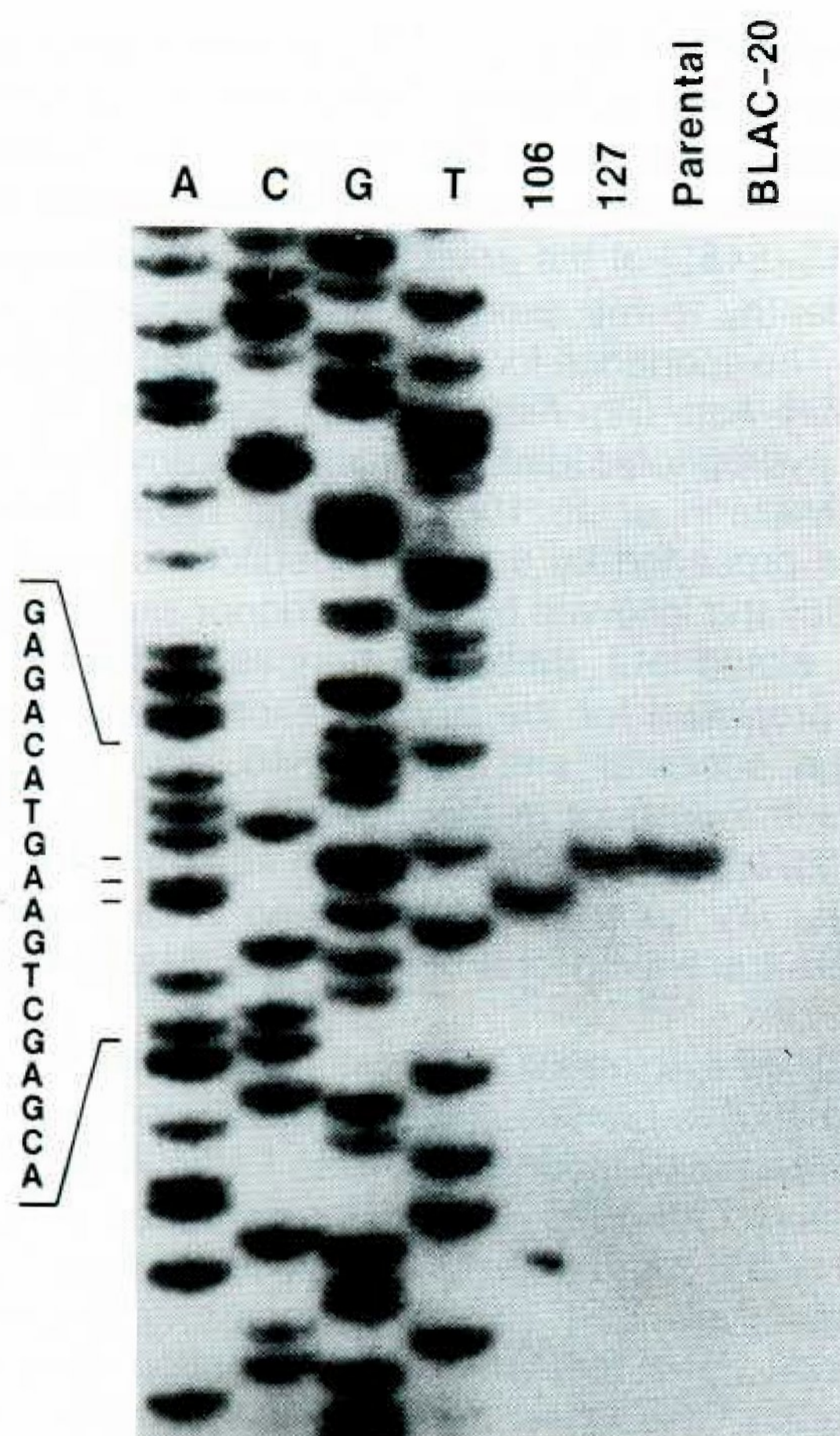
FIG. 4. Determination of the transcription initiation site by primer extension from viral RNAs. Lanes 1–4: Sequencing ladder (A, C, G, and T, respectively) using the nt determined for the noncoding strand of BIV 106, included for accurate estimation of the size of the extension products; lane 5: RNA from BIV 106-infected cells; lane 6: RNA from BIV 127-infected cells; lane 7: RNA from parental BIV-infected cells; lane 8: control RNA from uninfected cells. The RNA isolation, primer extension, and sequencing reactions were performed as described under Materials and Methods.

larger number of substitutions in *env* and that there are local regions of greater variability within *env*. Our analysis indicates that the *env* gene contains the greatest number of coding substitutions (*gag* = 2 of 7; *pol* = 2 of 8; and *env* = 12 of 24). In addition, within *env*, SU has the highest number of substitutions, suggesting that the SU is the most variable region of the genome. We subjected this observation to further scrutiny using a χ^2 test to determine whether the number of observed nucleotide changes in the *env* gene, and, in particular, the SU region, exceeded what would be expected if the changes (excluding the 87 nt deletion) were random over the lengths of each functional unit (*gag*, *pol*, SU, and TM). The results ($P = 0.0003$) of this χ^2 analysis

TABLE 2

PUTATIVE PROTEIN PRODUCTS OF BIV 127 ORFS

ORF	HIV-1 equivalent	BIV protein position	MW (Da)
gag[a]	*gag* precursor (p55)	nt 316–1743	53,440
	MA (p16)	aa 1–133	15,369
	CA (p24)	134–360	25,420
	NC (p15)	361–476	12,687
pol[b]	*pol* precursor (p110)	nt 1581–4739	120,658
	PR (p12)	aa 51–143	10,564
	RT (p66)	144–774	72,157
	IN (p32)	775–1053	32,054
env[c]	*env* precursor	nt 5415–8126	102,269
	SU	aa 1–555	62,098
	TM	556–904	40,189
vif	*vif*	4601–5194	22,828
W	?	4729–4890	6,620
Y	?	5089–5328	9,549
tat exon 1	*tat* exon 1	5228–5536	11,689[d]
tat exon 2	*tat* exon 2	7657–7782	ND[e]
rev exon 1	*rev* exon 1	5415–5452	ND
rev exon 2	*rev* exon 2	7571–8068	17,098[f]

[a] Alignment of BIV 127 *gag* and HIV-1$_{HXB2}$ *gag* by PIR Align program using a gap penalty of 6 and 300 random runs (Needleman and Wunsch, 1970). Processing products were derived from comparison with HIV-1$_{HXB2}$ (Mervis *et al.*, 1988).

[b] Alignment of BIV 127 *pol* and HIV-1$_{HXB2}$ *pol* by PIR Align and UW-GCG Gap programs. Protease-RT junction inferred from BIV-HIV-1$_{HXB2}$ comparison (Lillehoj *et al.*, 1988); endonuclease homology from BIV-visna virus and visna virus-HIV comparisons (Braun *et al.*, 1987).

[c] Alignment of BIV 127 *env* and HIV-1$_{HXB2}$ *env* by PIR Align program. Processing products from comparison of BIV and HIV-1$_{HXB2}$ (Muesing *et al.*, 1985).

[d] Molecular weights for *tat* exons 1 and 2 translated from a predicted spliced message.

[e] Not determined.

[f] Molecular weights for *rev* exons 1 and 2 translated from a predicted spliced message.

ulatory genes and are responsible for 24 coding substitutions. Several substitutions cause coding changes in overlapping ORFs. Most of the predicted replacements are relatively conservative; 6 however, result in charge changes (Table 1).

Figure 8 shows diagrammatically the distribution of nucleotide substitutions between BIV clones 127 and 106; the substitutions were graphed cumulatively for each 100 nt segment. In the structural genes, there are 7 substitutions in *gag*, 8 in *pol*, and 24 (111 if the 87 nt deletion is counted) in *env* (19 in SU; 5 in TM). In the nonstructural/regulatory genes, *vif* has 2, *tat* exon 1 has 2, *tat* exon 2 has 2, *rev* exon 2 has 3, and *W* has 1 substitution(s). By visual analysis of Fig. 8, there appears to be a nonrandom distribution of substitutions even when the major deletion in *env* is ignored. For the structural genes, the graph suggests that there are a

BIV   N I H Q G P K E P Y
HIV-1   D I R Q G P K E P F
HIV-2   D I K Q G P K E P F
SIVmac D V K Q G P K E P F
SIVagm D I R Q G P K E P F
EIAV   N I R Q G A K E P Y
Visna   L V K Q K N T E S Y

Consensus   B I b Q G P K E P a

FIG. 5. Alignment of the highly conserved region of lentivirus *gag* proteins. A global alignment was performed using the PIR Align program with a gap penalty of 6 (Needleman and Wunsch, 1970). The region shown represents residues 293–302 of the predicted BIV 127 *gag* protein sequence (BIV 106 and 127 are identical in this region). Boxed residues are those which are the same as BIV. The consensus line shows residues which are the same in at least five of the seven viruses. Lowercase symbols indicate conservation of amino acid character, but not a specific amino acid at that position: B, aspartate or asparagine; a, aromatic; b, basic.

statistically support the contention that the SU of the *env* gene is the most variable segment of the BIV genome used in the comparison.

Evolutionary relationship of BIV and other retroviruses

Despite the great diversity which exists between isolates of different retrovirus subfamilies (Gonda *et al.*, 1989), there are certain regions of the genome that are highly conserved and upon which evolutionary relationships can be readily derived. One such region that we have focused on is the amino terminus of the RT domain of *pol* (Gonda *et al.*, 1986, 1987, 1989; Gonda, 1989). Since the first report on the molecular cloning of BIV (Gonda *et al.*, 1987), wherein BIV was determined to be a member of the lentivirus subfamily of retroviruses, several new lentiviruses (FIV, OPPV, HIV-2, SIVmac, SIVagm, SIVmne, and SIVsm) have been cloned and sequenced. The relationship of BIV to these new viruses has not been determined. To investigate further the evolutionary relationship of BIV and other lentiviruses, we constructed phylogenetic trees for the retrovirus family from comparisons of amino acid sequences obtained from the RT domain of *pol*.

Figure 9A shows an alignment of 90 consecutive amino acid residues from a spectrum of retroviruses including all of the lentiviruses for which sequences are available to date, as well as representative members of the oncoviruses and spumaviruses. Pairwise compari-

sons between these residues show strong sequence conservation among the lentiviruses (50–53% or greater matching residues; no gaps); the sequence identity weakens when other retrovirus subfamily members are compared to the lentiviruses (38–42% or fewer matching residues; with gaps). It is obvious from an inspection of this alignment that BIV clusters with the lentiviruses.
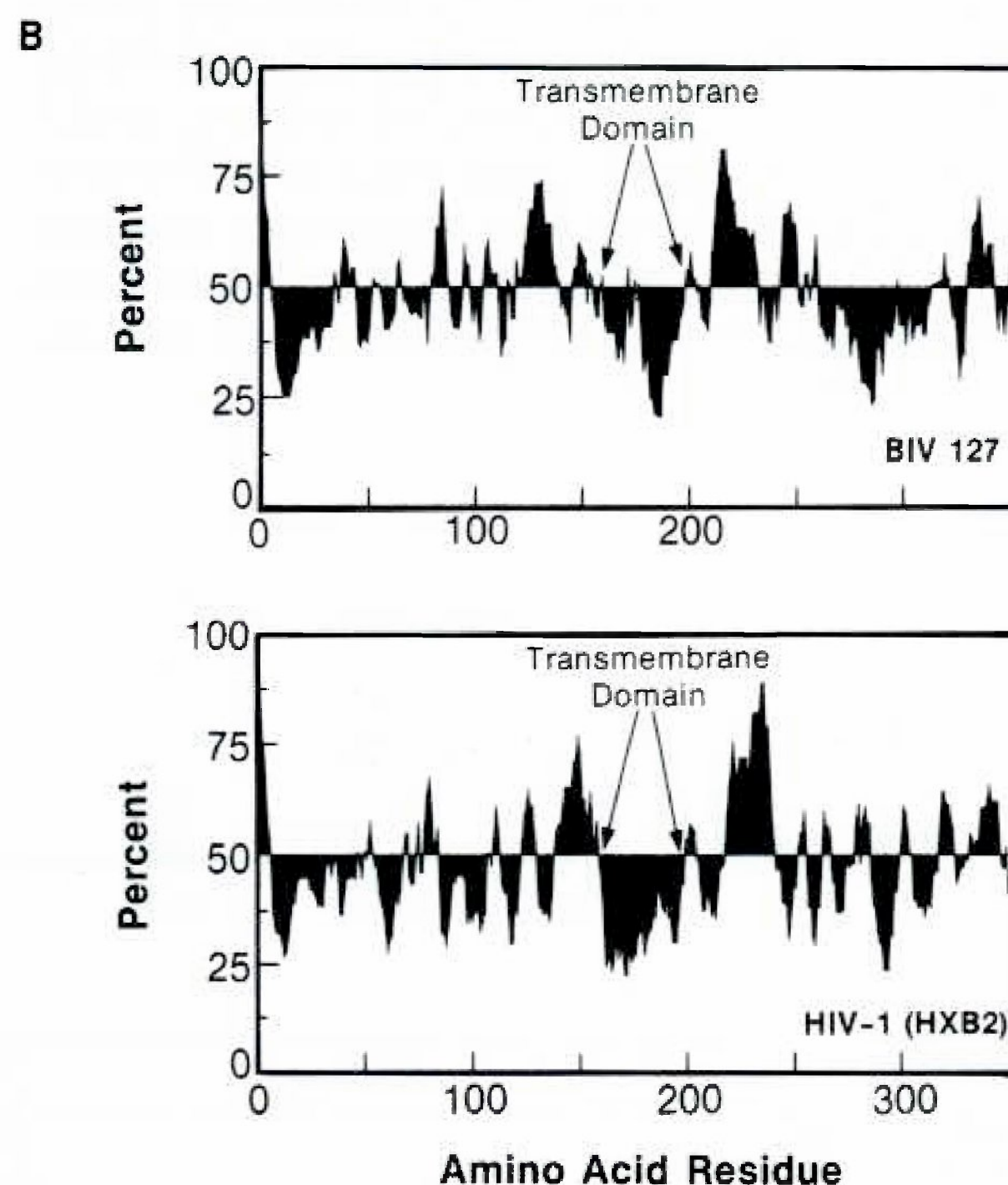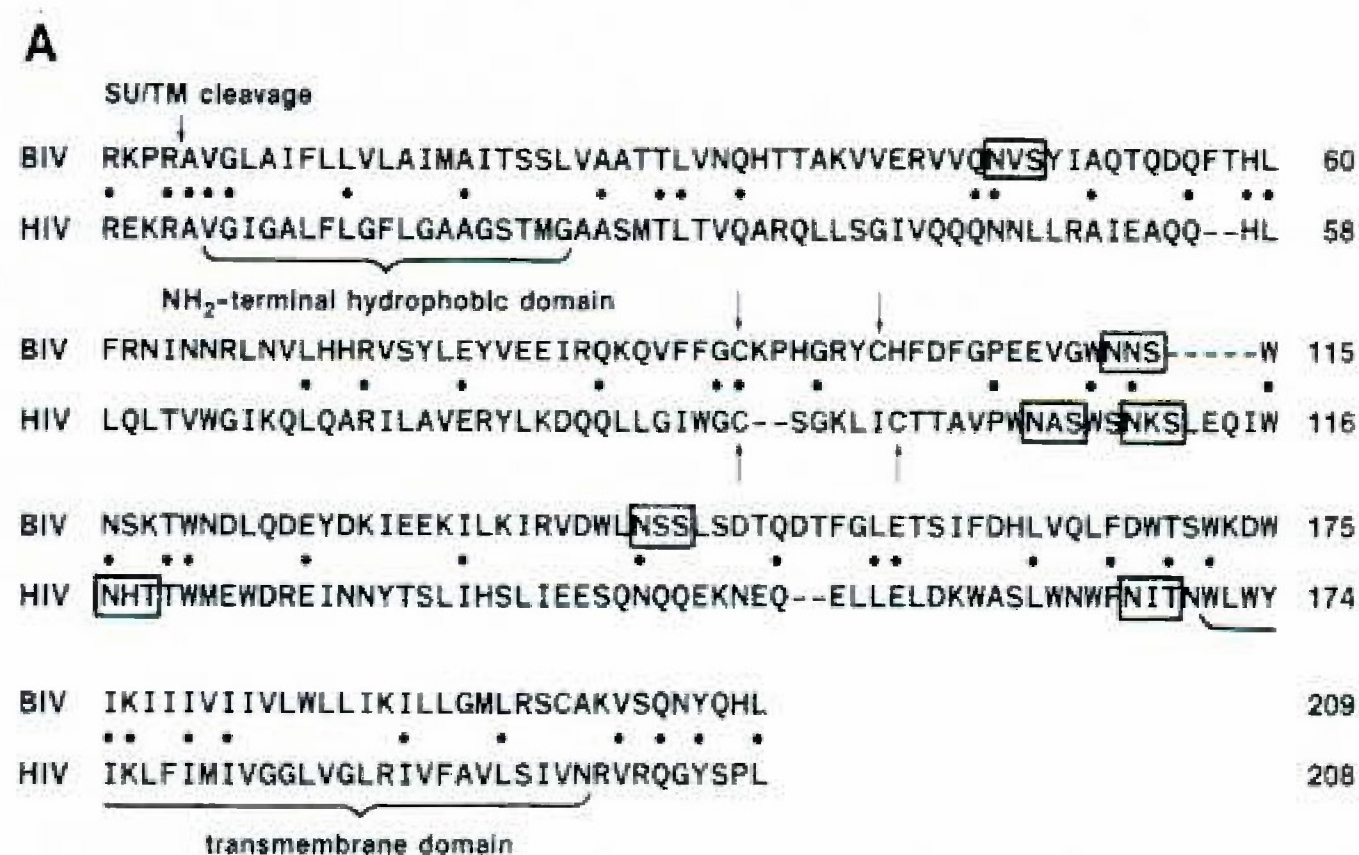
A

SU/TM cleavage

BIV RKPRAVGLAIFLLVLAIMAITSSLVAATTLVNQHTTAKVVERVVQNVSYIAQTQDQFTHL 60
HIV REKRAVGIGALFLGFLGAAGSTMGAASMTLTVQARQLLSGIVQQQNNLLRAIEAQQ--HL 58

NH₂-terminal hydrophobic domain

BIV FRNINNRLNVLHHRVSYLEYVEEIRQKQVFFGCKPHGRYCHFDFGPEEVGWNNS-----W 115
HIV LQLTVWGIKQLQARILAVERYLKDQQLLGIWGC--SGKLICTTAVPWNASWSNKSLEQIW 116

BIV NSKTWNDLQDEYDKIEEKILKIRVDWLNSSLSDTQDTFGLETSIFDHLVQLFDWTSWKDW 175
HIV NHTTWMEWDREINNYTSLIHSLIEESQNQQEKNEQ--ELLELDKWASLWNWFNITNWLWY 174

BIV IKIIIVIIVLWLLIKILLGMLRSCAKVSQNYQHL 209
HIV IKLFIMIVGGLVGLRIVFAVLSIVNRVRQGYSPL 208

transmembrane domain

B

FIG. 6. Comparison of the transmembrane proteins (TM) of BIV 127 and HIV-1HXB2. (A) Alignment of a portion of the BIV 127 and HIV-1HXB2 TM using UWGCG Gap (gap weight 2.0, gap length weight 0.2) (Devereaux *et al.*, 1984). The transmembrane and NH₂-terminal hydrophobic domains of the HIV-1 TM are indicated below the sequence. Arrows indicate conserved Cys residues. Potential sites of N-linked glycosylation are boxed. SU-TM cleavage site is indicated. (B) Hydropathy plots of the BIV 127 (top panel) and HIV-1HXB2 (lower panel) TMs using UWGCG Peptidestructure and Plotstructure (Jameson and Wolf, 1988). Hydrophobic regions appear below the mean; hydrophilic regions appear above the mean.
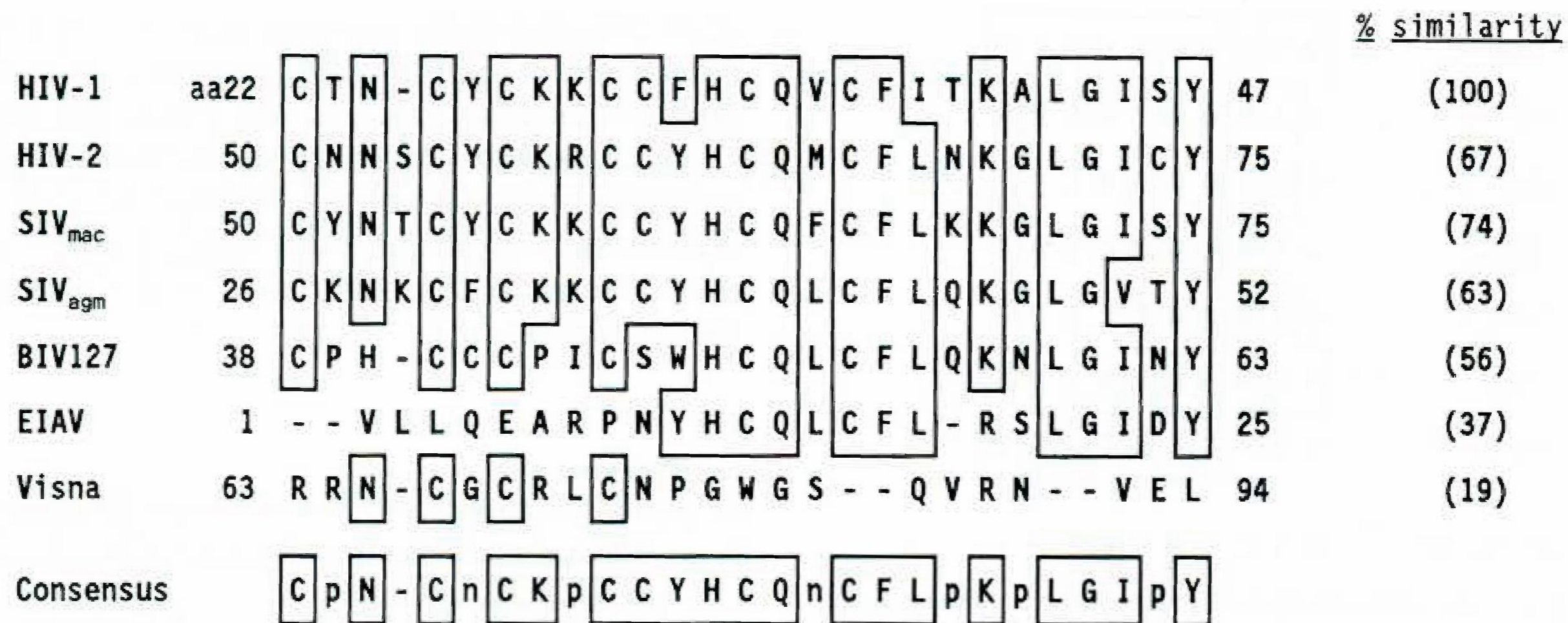
% similarity

HIV-1   aa22   C T N - C Y C K K C C F H C Q V C F I T K A L G I S Y   47   (100)

HIV-2   50   C N N S C Y C K R C C Y H C Q M C F L N K G L G I C Y   75   (67)

SIV_mac   50   C Y N T C Y C K K C C Y H C Q F C F L K K G L G I S Y   75   (74)

SIV_agm   26   C K N K C F C K K C C Y H C Q L C F L Q K G L G V T Y   52   (63)

BIV127   38   C P H - C C C P I C S W H C Q L C F L Q K N L G I N Y   63   (56)

EIAV   1   - - V L L Q E A R P N Y H C Q L C F L - R S L G I D Y   25   (37)

Visna   63   R R N - C G C R L C N P G W G S - - Q V R N - - V E L   94   (19)

Consensus   C p N - C n C K p C C Y H C Q n C F L p K p L G I p Y

FIG. 7. Alignment of a highly conserved region of the *tat* (exon 1) ORF of several lentiviruses. The nucleotide sequences of the indicated viruses were obtained from GenBank and the *tat* region was translated using the program Translate (UWGCG) (Devereaux et al., 1984). Each peptide sequence was optimally aligned with the HIV-1$_{HXB2}$ sequence using UWGCG Gap (gap weight 5.0, gap length weight 0.3). Multiple sequence alignment was generated using Genalign (Sobel and Martinez, 1985). p, polar; n, nonpolar.

Figure 9B is a phylogenetic tree of retroviral relationships for the RT segment of the *pol* gene shown in Fig. 9A. Among the retroviruses whose RTs have a preference for Mg^{2+} cations, there are two main branches, one which leads to the oncoviruses and one which leads to the lentiviruses. The branching order of lentiviruses has BIV diverging first from an ancestor that gave rise to all of the other lentiviruses. BIV appears to be more closely related to EIAV; however, from calcula-

tions of the evolutionary distances between lentiviruses on the tree, our overall impression of the BIV relationship is that it is relatively equidistant from all of the other lentiviruses for the RT segment tested. Furthermore, sequence comparisons from other ORFs of the genome suggest alternate protein-dependent tree topologies with lentiviruses other than BIV diverging first (data not shown). Moreover, the complement of non-structural/regulatory genes in the various isolates

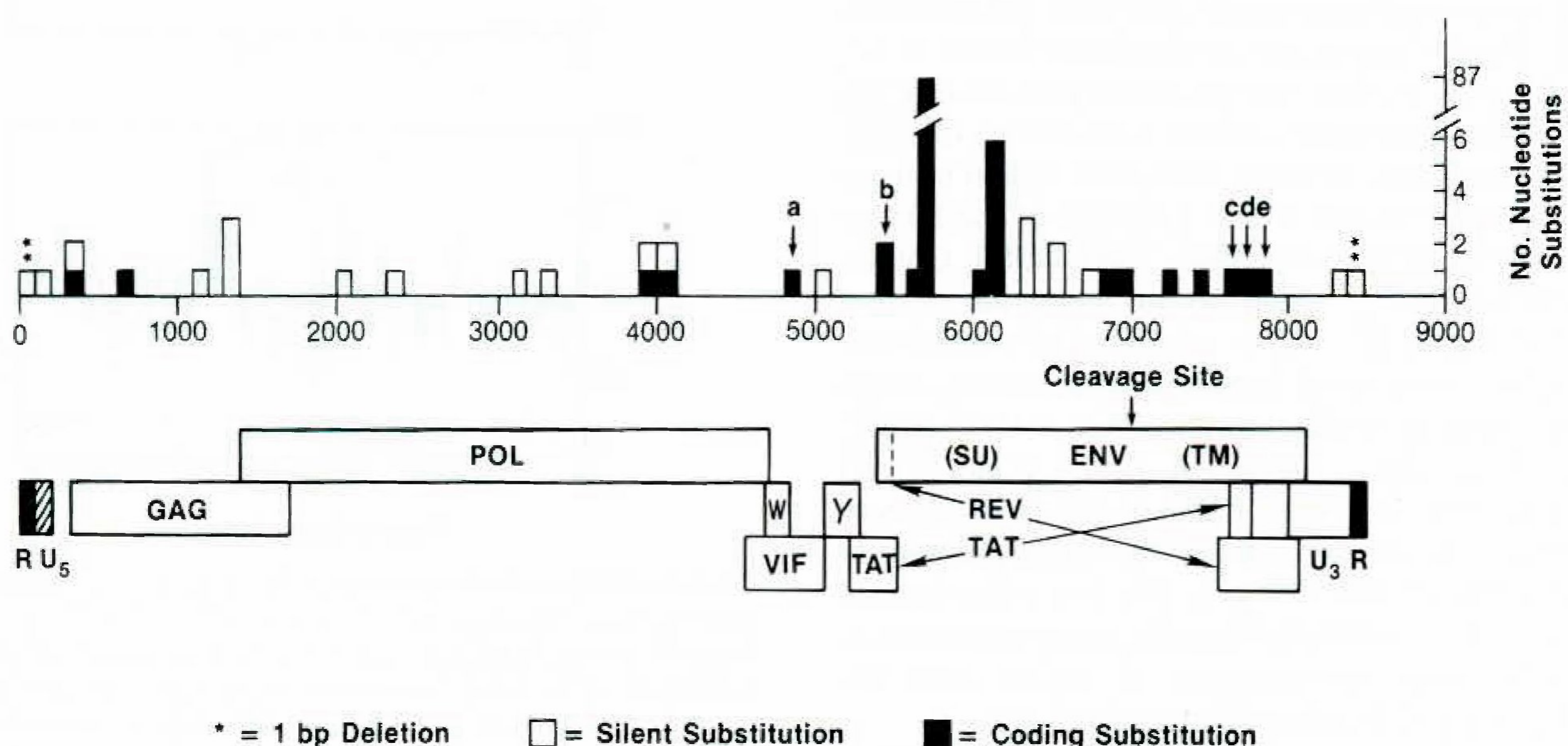* = 1 bp Deletion   □ = Silent Substitution   ■ = Coding Substitution

FIG. 8. Distribution of nucleotide substitutions between BIV 106 and 127. The positions of the major ORFs are shown. The arrow over the *env* segment represents the SU-TM cleavage site. Each box represents one substitution and substitutions are cumulatively displayed every 100 nt. a, coding substitutions in both *W* and *vif*. b, two nucleotide substitutions which cause coding changes in *tat* exon 1, but only one of the two causes a coding change in *env*. c, coding changes in *env* and *rev* exon 2, but not *tat* exon 2. d, nucleotide substitution which causes coding changes in *rev* exon 2 and *tat* exon 2, but not in *env*. e, nucleotide substitution which causes a coding change in *rev* exon 2, but not in *env*.
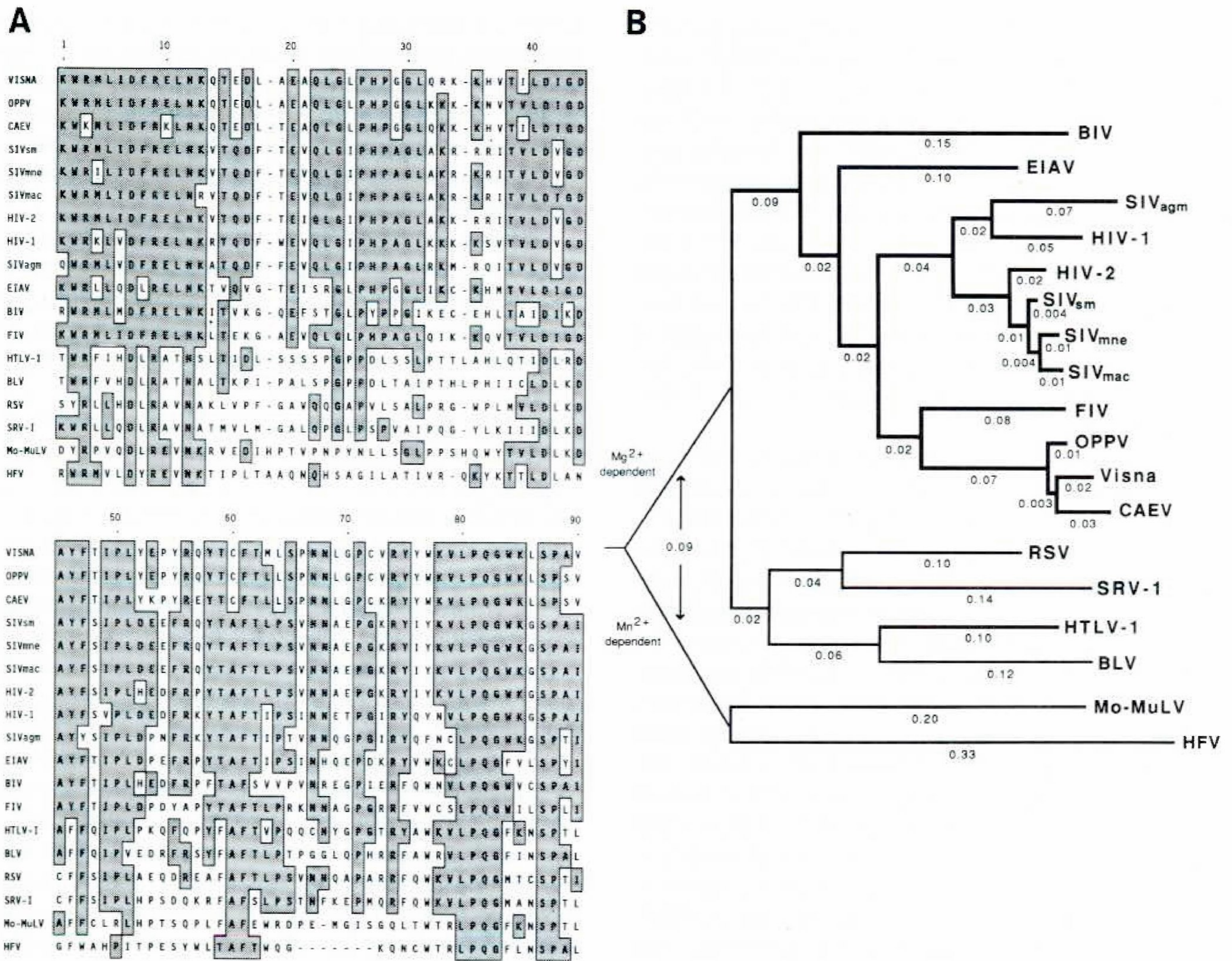
FIG. 9. (A) Alignment of 90 consecutive amino acids in the conserved RT domain of *pol* genes from lentiviruses with those of other retroviruses. Retroviruses used are visna virus, OPPV, CAEV, SIVsm, SIVmne, SIVmac, HIV-2, HIV-1, SIVagm, EIAV, BIV, FIV, HTLV-1 (human T-cell leukemia virus type 1), BLV (bovine leukemia virus), RSV (Rous sarcoma virus), SRV-1 (simian aids retrovirus type 1), Mo-MuLV (Moloney murine leukemia virus), and HFV (human foamy virus). The alignment shown is generally that optimal with visna virus; slight improvements in the other pairwise alignment scores can be made by minor shifts in the placement of gaps. Shaded boxes are drawn around identical residues when seven or more lentiviruses share that residue. (B) Fitch–Margoliash phylogenetic tree of retroviral relationships based on the *pol* gene sequences shown in (A). Branch lengths are in units of $-\log M$, where M is the frequency of matching residues. The tree was rooted with HFV and Mo-MuLV as the outgroup taxons because they consistently had the lowest alignment scores and because their RTs preferentially use Mn^{2+} cations as cofactors. The average percentage standard deviation of the tree was 5.17.

differs, thus further complicating the derivation and analysis of evolutionary trees for certain gene segments.

The RT tree has the majority of the lentiviruses clustering into two main groups, the primate lentiviruses and the FIV, OPPV, CAEV, and visna virus tetrad, wherein members of each group are more closely related to each other and appear to have arisen from a lentivirus predecessor after BIV and EIAV diverged. Immunologic cross-reactivity of *gag* gene products within each cluster support the identity of these two groups (Gonda *et al.*, 1987; Olmsted *et al.*, 1989a). Our results further indicate that BIV is not closely related to and did

not recently arise from any other characterized lentivirus; thus, it is a unique and distinct member of the lentiviruses. The topology of our tree is in general agreement with those recently derived by others using subsets of the viruses used here (Doolittle, 1989; Gonda *et al.*, 1989; Olmsted *et al.*, 1989b).

DISCUSSION

In the present study, the complete nucleotide sequences of two functional proviruses of BIV were determined. Their genomic organization is identical. The BIV

genome is 9.0 kb in the form of the integrated proviral DNA. Within the lentivirus family, only the EIAV provirus is smaller at 8.2 kb (Kawakami et al., 1987). Despite BIV's small size, its genome is quite complex since we have localized eight potential genes. The three largest ORFs encode the gag, pol, and env genes common to all replication-competent retroviruses. BIV also has the lentivirus central region in which we found five additional ORFs. Three of these are similar in structure and location to the exons for vif, tat, and rev; two unique ORFs W and Y were also found. ORFs similar to W and Y have not been described for visna virus or EIAV genomes (Kawakami et al., 1987; Sonigo et al., 1985; Braun et al., 1987).

The predicted first exon of rev was found in the env ORF, in frame with the env gene translation, and that of tat was found in the central region. The second exons for tat and rev appear to reside in the 3' end of the genome overlapping env, suggesting that the first and second encoding exons for tat and rev are brought together by the multiple splicing of a larger transcript to produce a mature mRNA, reminiscent of HIV and visna virus (Rabson et al., 1985; Davis et al., 1987; Mazarin et al., 1988). Potential splice acceptors and donors were recognized in the vicinity of these exons, which enhances the likelihood of this hypothesis. In fact, recent analysis of BIV transcripts using viral RNA from cells infected by the infectious clones or parental stock has demonstrated that BIV has a complex transcriptional pattern and encodes at least five size classes of mRNA (Oberste and Gonda, manuscript in preparation). The definitive identification and functional demonstration of the nonstructural/regulatory ORFs of BIV will require the isolation and characterization of cDNA clones and their products.

One striking feature of the BIV genome was the lack of a nef ORF which is located post-env in the genomes of primate lentiviruses (Chakrabarti et al., 1987; Fukasawa et al., 1988; Guyader et al., 1987; Niederman et al., 1989; Ratner et al., 1987). nef is postulated to be a transcriptional silencer and important in down regulating the expression of virus (Niederman et al., 1989). Interestingly, visna virus and EIAV also lack a nef ORF (Braun et al., 1987; Sonigo et al., 1985; Kawakami et al., 1987). The env ORFs of the nonprimate lentiviruses are very large. It is possible that the nef for these viruses is in frame with the 3'-most end of the TM domain of the env ORF, similar to the incorporation of rev exon 1 of visna virus and BIV in the 5' end of the env ORF. We searched for sequence similarity in BIV with various nef sequences from primate lentiviruses and found no reasonable identity to support this idea. Of the lentiviruses studied thus far, EIAV appears to have the simplest central region complexity since it also lacks an

ORF for vif (Kawakami et al., 1987). The nonstructural/regulatory genes of HIV are purported to play an important role in the pathogenesis, infectivity, and latency of HIV (Haseltine et al., 1988). It will be important to correlate and compare the biological significance of the presence and/or absence of these genes in BIV pathogenesis.

Genomic variability is a striking feature of lentivirus infections. Our nucleotide sequence comparisons of the two functional proviral molecular clones here have demonstrated that BIV also exhibits extensive genomic variation, even though the two clones were derived from a single virus isolation. The possible reasons proposed for the genomic diversity of lentiviruses include infidelity of the lentivirus RT during reverse transcription of the viral RNA (Preston et al., 1988; Roberts et al., 1988) and constraints in the secondary structure of the viral RNA that make the RT error prone (Braun et al., 1987). The genomes of BIV 106 and 127 differ by 1.7% when all deletions and substitutions are taken into consideration; however, 75% of the nucleotide substitutions and deletions occur in the SU. In the case of BIV 106 and 127, all of the changes were due to substitutions and deletions, no obvious duplications were observed. The largest number of coding changes occurs in the SU; there may be a selective advantage for the SU of lentiviruses to be hypervariable (Braun et al., 1987; Gonda, 1989). It is not known whether the nucleotide substitutions resulting in coding changes are reflected in the antigenic properties of the SU or are responsible for the differences in cytopathogenicity observed in vitro (Braun et al., 1988; unpublished data). The 29 amino acids encoded by the 87 nt deletion in the SU were predicted by computer algorithms to have a high surface probability and, thus, to be antigenic (data not shown). We assume that all of the changes between clones 127 and 106 reported here occurred in vivo prior to the isolation of the parental stock of BIV. It is unlikely that the changes occurred in vitro, since such changes have not been reported to occur for any lentivirus studied thus far (Saag et al., 1988), and our clones, at least in the hypervariable region of the SU of the env ORF, have remained constant during numerous propagations in culture.

In previous experiments using restriction enzyme mapping (assaying for the presence or absence of a unique BIV 127 EcoR1 site in the env gene), the predominant genome type in parental stocks of BIV propagated in bovine cells was determined to be one with sequences related to clone 106 (Braun et al., 1988). We were unable to detect BIV 127 sequences using this physical mapping technique in Southern transfers, even though a BIV 127 proviral clone was obtained. In the present report, the primer extension data demon-

strated that parental stocks of BIV used a BIV 127-like LTR. This suggests that there may be other BIV 106-like proviruses in cells infected with parental stocks which have BIV 127-like LTRs. The functional significance of this finding is at present not known, but suggests that each BIV isolate is comprised of a heterogeneous population of genetically related but molecularly distinct genomes. Individual functional proviral genomes within this population may differ in biological properties. Our present study supports the idea that lentiviruses are a quasi-species composed of many variant genomes; thus, the sequence and biologic properties of a single provirus may not be definitive of the *in vivo* and *in vitro* macrobiology of the isolate (Wain-Hobson *et al.,* 1989; Saag *et al.,* 1988; Fisher *et al.,* 1988; Tersmette *et al.,* 1989). The extent and biological significance of the variability seen in BIV genomes is presently under investigation using these and additional proviral molecular clones derived from the same bacteriophage library used to obtain BIV 106 and 127.

ACKNOWLEDGMENTS

REFERENCES

ADLER, S., WATERMAN, M. L., HE, X., and ROSENFELD, M. G. (1988). Steroid receptor-mediated inhibition of rat prolactin gene expression does not require the DNA-binding domain. *Cell* **52,** 685–695.

ALIZON, M., WAIN-HOBSON, S., MONTAGNIER, L., and SONIGO, P. (1986). Genetic variability of the AIDS virus: Nucleotide sequence analysis of two isolates from African patients. *Cell* **46,** 63–74.

BARRÉ-SINOUSSI, F., CHERMANN, J. C., REY, F., NEGEYRE, M. T., CHAMARET, S., GRUEST, J., DAUGUET, C., and MONTAGNIER, L. (1983). Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* **220,** 868–871.

BENN, S., RUTLEDGE, R., FOLKS, T., GOLD, J., BAKER, L., McCORMICK, J., FEORINO, P., PIOT, P., QUINN, T., and MARTIN, M. (1985). Genomic heterogeneity of AIDS retroviral isolates from North America. *Science* **230,** 949–951.

BERG, J. M. (1986). Potential metal-binding domains in nucleic acid binding proteins. *Science* **232,** 485–486.

BRAUN, M. J., CLEMENTS, J. E., and GONDA, M. A. (1987). The visna virus genome: Evidence for a hypervariable site in the *env* gene

and sequence homology among lentivirus envelope proteins. *J. Virol.* **61,** 4046–4054.

BRAUN, M. J., LAHN, S., BOYD, A. L., KOST, T. A., NAGASHIMA, K., and GONDA, M. A. (1988). Molecular cloning of biologically active proviruses of bovine immunodeficiency-like virus. *Virology* **167,** 515–523.

CALZONE, F. J., BRITTEN, R. J., and DAVIDSON, E. H. (1987). Mapping of gene transcripts by nuclease protection assays and cDNA primer extension. *In* "Methods in Enzmology" (S. L. Berger and A. R. Kimmel, Eds.), Vol. 152, pp. 611–632. Academic Press, San Diego.

CASEY, J., and DAVIDSON, N. (1977). Rates of formation and thermal stabilities of RNA:DNA and DNA:DNA duplexes at high concentrations of formamide. *Nucleic Acids Res.* **4,** 1539–1552.

CASEY, J. M., KIM, Y., ANDERSEN, P. R., WATSON, K. F., FOX, J. L., and DEVARE, S. G. (1985). Human T-cell lymphotropic virus type III: Immunologic characterization and primary structure analysis of the major internal protein, p24. *J. Virol.* **55,** 417–423.

CHAKRABARTI, L., GUYADER, M., ALIZON, M., DANIEL, M. D., DESROSIERS, R. C., TIOLLAIS, P., and SONIGO, P. (1987). Sequence of simian immunodeficiency virus from macaque and its relationship to other human and simian retroviruses. *Nature (London)* **328,** 543–547.

CHIU, I.-M., YANIV, A., DALBERG, J. E., GAZIT, A., SKUNTZ, S. F., TRONICK, S. R., and AARONSON, S. A. (1985). Nucleotide sequence evidence for relationship of AIDS retrovirus to lentiviruses. *Nature (London)* **317,** 366–368.

CHOMCZYNSKI, P., and SACCHI, N. (1987). Single-step method of RNA isolation by guanidinium thiocyanate–phenol–chloroform extraction. *Anal. Biochem.* **162,** 156–159.

CLAVEL, F., GUETARD, D., BRUN-VÉZINET, F., CHAMARET, S., REY, M.-A., SANTOS-FERREIRA, M. O., LAURENT, A. G., and DAUGUET, C. (1986). Isolation of a new human retrovirus from West African patients with AIDS. *Science* **233,** 343–346.

COVEY, S. N. (1986). Amino acid sequence homology in *gag* region of reverse transcribing elements and the coat protein gene of cauliflower mosaic virus. *Nucleic Acids Res.* **14,** 623–633.

CRAWFORD, T. B., ADAMS, D. S., CHEEVERS, W. P., and CORK, L. C. (1980). Chronic arthritis in goats caused by a retrovirus. *Science* **207,** 997–999.

CUTLIP, R. C., and LAIRD, G. A. (1976). Isolation and characterization of a virus associated with progressive pneumonia (Maedi) of sheep. *Amer. J. Vet. Res.* **37,** 1377–1382.

DANIEL, M. D., LETVIN, N. L., KING, N. W., KANNAGI, M., SEHGAL, P. K., HUNT, R. D., KANKI, P. J., and DESROSIERS, R. C. (1985). Isolation of T-cell lymphotropic HTLV-III-like retrovirus from Macaques. *Science* **228,** 1201–1204.

DAVIS, J. L., and CLEMENTS, J. E. (1989). Characterization of a cDNA clone encoding the visna virus transactivating protein. *Proc. Natl. Acad. Sci. USA* **86,** 414–418.

DAVIS, J. L., MOLINEAUX, S., and CLEMENTS, J. E. (1987). Visna virus exhibits a complex transcriptional pattern: One aspect of gene expression shared with the acquired immunodeficiency syndrome retrovirus. *J. Virol.* **61,** 1325–1331.

DAYHOFF, M. O. (1978). Survey of new data and computer methods of analysis. *In* "Atlas of Protein Sequence and Structure" (M. O. Dayhoff, Ed.), Vol. 5, Suppl. 3, pp. 1–8. Natl. Biomed. Res. Found., Washington, DC.

DEININGER, P. L. (1983). Random subcloning of sonicated DNA: Application to shotgun DNA sequence analysis. *Anal. Biochem.* **129,** 216–223.

DEVEREAUX, J., HAEBERLI, P., and SMITHIES, O. (1984). A comprehensive set of sequencing analysis programs for the VAX. *Nucleic Acids Res.* **12,** 387–395.

DOOLITTLE, R. F. (1989). The simian–human connection. *Nature (London)* **339**, 338–339.

EVANS, R. M., and HOLLENBERG, S. M. (1988). Zinc fingers: Gilt by association. *Cell* **52**, 1–3.

FISHER, A. G., ENSOLI, B., LOONEY, D., ROSE, A., GALLO, R. C., SAAG, M. S., SHAW, G. M., HAHN, B. H., and WONG-STAAL, F. (1988). Biologically diverse molecular variants within a single HIV-1 isolate. *Nature (London)* **334**, 444–447.

FITCH, W. M., and MARGOLIASH, E. (1967). Construction of phylogenetic trees: A method based on mutation distances as estimated from cytochrome c sequences is of general applicability. *Science* **155**, 279–284.

FUKASAWA, M., MIURA, T., HASEGAWA, A., MORIKAWA, S., TSUJIMOTO, H., MIKI, K., KITAMURA, T., and HAYAMI, M. (1988). Sequence of simian immunodeficiency virus from African green monkey, a new member of the HIV/SIV group. *Nature (London)* **333**, 457–461.

GALLO, R. C., SALAHUDDIN, S. Z., POPOVIC, M., SHEARER, G. M., KAPLAN, M., HAYNES, B. F., PALKER, T. J., and MARKHAM, P. D. (1984). Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science* **224**, 500–503.

GALLO, R., WONG-STAAL, F., MONTAGNIER, L., HASELTINE, W. A., and YOSHIDA, M. (1988). HIV/HTLV gene nomenclature. *Nature (London)* **333**, 504.

GAYNOR, R. B., KUWABARA, M. D., GARCIA, J. A., HARRICH, D., BRISKIN, M., WALL, R., and SIGMAN, D. S. (1988). Repeated B motifs in the human immunodeficiency virus type I long terminal repeat enhancer region do not exhibit cooperative factor binding. *Proc. Natl. Acad. Sci. USA* **85**, 9406–9410.

GONDA, M. A. (1989). Visna virus genome: Variability and relationship to other lentiviruses. *In* "Applied Virology Research: Vol. 2—Virus Variation and Epidemiology." Plenum, New York, in press.

GONDA, M. A., BOYD, A. L., NAGASHIMA, K., and GILDEN, R. V. (1989). Pathobiology, molecular organization, and ultrastructure of HIV. *Arch. AIDS Res.* **3**, 1–42.

GONDA, M. A., BRAUN, M. J., CARTER, S. G., KOST, T. A., BESS, JR., J. W., ARTHUR, L. O., and VAN DER MAATEN, M. J. (1987). Characterization and molecular cloning of a bovine lentivirus related to human immunodeficiency virus. *Nature (London)* **330**, 388–391.

GONDA, M. A., BRAUN, M. J., CLEMENTS, J. E., PYPER, J. M., GALLO, R. C., WONG-STAAL, F., and GILDEN, R. V. (1986). Human T-cell lymphotropic virus type III shares sequence homology with a family of pathogenic lentiviruses. *Proc. Natl. Acad. Sci. USA* **83**, 4007–4011.

GONDA, M. A., WONG-STAAL, F., GALLO, R. C., CLEMENTS, J. E., NARAYAN, O., and GILDEN, R. V. (1985). Sequence homology and morphologic similarity of HTLV-III and visna virus, a pathogenic lentivirus. *Science* **227**, 173–177.

GUYADER, M., EMERMAN, M., SONIGO, P., CLAVEL, F., MONTAGNIER, L., and ALIZON, M. (1987). Genome organization and transactivation of the human immunodeficiency virus type 2. *Nature (London)* **326**, 662–669.

HAASE, A. T. (1986). Pathogenesis of lentivirus infections. *Nature (London)* **322**, 130–136.

HAHN, B. H., SHAW, G. M., TAYLOR, M. E., REDFIELD, R. R., MARKHAM, P. D., SALAHUDDIN, S. Z., WONG-STAAL, F., and GALLO, R. C. (1986). Genetic variation in HTLV-III/LAV over time in patients with AIDS or at risk for AIDS. *Science* **232**, 1548–1553.

HASELTINE, W. A. (1988). Replication and pathogenesis of the AIDS virus. *J. Acquired Immune Def. Syn.* **1**, 217–240.

HENDERSON, L. E., COPELAND, T. D., SOWDER, R. C., SMYTHERS, G. W., and OROSZLAN, S. (1981). Primary structure of the low molecular weight nucleic acid-binding proteins of murine leukemia viruses. *J. Biol. Chem.* **256**, 8400–8406.

HENDERSON, L. E., SOWDER, R. C., COPELAND, T. D., BENVENISTE, R. E., and OROSZLAN, S. (1988). Isolation and characterization of a novel protein X-orf product) from SIV and HIV-2. *Science* **241**, 199–201.

HIRSCH, V. M., OLMSTED, R. A., MURPHEY-CORB, M., PURCELL, R. H., and JOHNSON, P. R. (1989). An African primate lentivirus (SIV$_{sm}$) closely realted to HIV-2. *Nature (London)* **339**, 389–392.

JAMESON, B. A., and WOLF, H. (1988). The antigenic index: A novel algorithm for predicting antigenic determinants. *Comput. Appl. Biosci.* **4**, 181–186.

KANKI, P. J., ALROY, J., and ESSEX, M. (1985). Isolation of T-lymphotropic retrovirus related to HTLV-III/LAV from wild-caught African green monkeys. *Science* **230**, 951–954.

KAWAKAMI, T., SHERMAN, L., DAHLBERG, J., GAZIT, A., YANIV, A., TRONICK, S. R., and AARONSON, S. A. (1987). Nucleotide sequence analysis of equine infectious anemia virus proviral DNA. *Virology* **158**, 300–312.

KENNEDY, R. C., EKLUND, C. M., LOPEZ, C., and HADLOW, W. J. (1968). Isolation of a virus from the lungs of Montana sheep affected with progressive pneumonia. *Virology* **35**, 483–484.

KOZAK, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**, 283–292.

KOZAK, M. (1989). The scanning model for translation: An update. *J. Cell Biol.* **108**, 229–241.

LAURENCE, J. (1988). HIV-1 gene nomenclature. *AIDS Res. Hum. Retrovirus.* **4**, vii–viii.

LEIS, J., BALTIMORE, D., BISHOP, J. M., COFFIN, J., FLEISSNER, E., GOFF, S. P., OROSZLAN, S., ROBINSON, H., SKALKA, A. M., TEMIN, H. M., and VOGT, V. (1988). Standardized and simplified nomenclature for proteins common to all retroviruses. *J. Virol.* **62**, 1808–1809.

LILLEHOJ, E. P., SALAZAR, F. H. R., MERVIS, R. J., RAUM, M. G., CHAN, H. W., AHMAD, N., and VENKATESAN, S. (1988). Purification and structural characterization of the putative gag–pol protease of human immunodeficiency virus. *J. Virol.* **62**, 3053–3058.

MANIATIS, T., FRITSCH, E. F., and SAMBROOK, J. (1982) "Molecular Cloning: A Laboratory Manual," pp. 125–126. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

MAZARIN, V., GOURDOU, I., QUÉRAT, G., SAUZE, N., and VIGNE, R. (1988). Genetic structure and function of an early transcript of visna virus. *J. Virol.* **62**, 4813–4818.

McGUIRE, T. C., O'ROURDE, K., and CHEEVERS, W. P. (1987). A review of antigenic variation by the equine infectious anemia virus. *In* "Contr. Microbiol. Immunol., Antigenic Variation: Molecular and genetic mechanisms of relapsing disease" (J. M. Cruse and R. E. Lewis, Jr., Eds.), Vol. 8, pp. 77–89. Karger, Basel.

MERVIS, R. J., AHMAD, N., LILLEHOJ, E. P., RAUM, M. G., SALAZAR, F. H. R., CHAN, H. W., and VENKATESAN, S. (1988). The gag gene products of human immunodeficiency virus type 1: Alignment within the gag open reading frame, identification of posttranslational modifications, and evidence for alternative gag precursors. *J. Virol.* **62**, 3993–4002.

MONTAGNIER, L., DAUGUET, C., AXLER, C., CHAMARET, S., GRUEST, J., NUGEYRE, M. T., REY, F., BARRÉ-SINOUSSI, F., and CHERMANN, J. C. (1984). A new type of retrovirus isolated from patients presenting with lymphadenopathy and acquired immune deficiency syndrome: Structural and antigenic relatedness with equine infectious anaemia virus. *Ann. Virol.* **135**, 119–134.

MUESING, M. A., SMITH, D. H., CABRADILLA, C. D., BENTON, C. V., LASKY, L. A., and CAPON, D. J. (1985). Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovirus. *Nature (London)* **313**, 450–458.

NARAYAN, O., GRIFFIN, D. E., and CHASE, J. (1977). Antigenic shift of visna virus in persistently infected sheep. *Science* **197**, 376–378.

NEEDLEMAN, S. B., and WUNSCH, C. D. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **48**, 443–453.

NIEDERMAN, T. M. J., THIELAN, B. J., and RATNER, L. (1989). Human immunodeficiency virus type 1 negative factor is a transcriptional silencer. *Proc. Natl. Acad. Sci. USA* **86**, 1128–1132.

NORRANDER, J., KEMPE, T., and MESSING, J. (1983). Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. *Gene* **26**, 101–106.

OLMSTED, R. A., BARNES, A. K., YAMAMOTO, J. K., HIRSCH, V. M., PURCELL, R. H., and JOHNSON, P. R. (1989a). Molecular cloning of feline immunodeficiency virus. *Proc. Natl. Acad. Sci. USA* **86**, 2448–2452.

OLMSTED, R. A., HIRSCH, V. M., PURCELL, R. H., and JOHNSON, P. R. (1989b). Nucleotide sequence analysis of feline immunodeficiency virus: Genome organization and relationship to other lentiviruses. *Proc. Natl. Acad. Sci. USA* **86**, 8088–8092.

OROSZLAN, S., and COPELAND, T. D. (1985). Primary structure and processing of *gag* and *env* gene products of human T-cell leukemia viruses HTLV-I$_{cr}$ and HTLV-I$_{atk}$. *Curr. Top. Microbiol. Immunol.* **115**, 221–233.

PAYNE, S., FANG, F. L. C., DHRUVA, B. R., RWAMBO, P., ISSEL, C. J., and MONTELARO, R. C. (1987). Antigenic variation and lentivirus persistence: Variations in the envelope gene sequences during EIAV infection resemble changes reported for sequential isolates of HIV. *Virology* **161**, 321–331.

PAYNE, S., PAREKH, B., MONTELARO, R. C., and ISSEL, C. J. (1984). Genomic alterations associated with persistent infections by equine infectious anaemia virus, a retrovirus. *J. Gen. Virol.* **65**, 1395–1399.

PEDERSEN, N. C., HO, E. N., BROWN, M. L., and YAMAMOTO, J. K. (1987). Isolation of a T-lymphotropic virus from domestic cats with an immunodeficiency-like syndrome. *Science* **235**, 790–793.

PRESTON, B. D., POIESZ, B. J., and LOEB, L. A. (1988). Fidelity of HIV-1 reverse transcriptase. *Science* **242**, 1168–1171.

PYPER, J. M., CLEMENTS, J. E., MOLINEAUX, S. M., and NARAYAN, O. (1984). Genetic variation among lentiviruses: Homology between visna virus and caprine arthritis-encephalitis virus is confined to the 5' *gag–pol* region and a small portion of the *env* gene. *J. Virol.* **51**, 713–721.

RABSON, A. R., DAUGHERTY, D. F., VENKATESEN, S., BOULUKOS, K. E., BENN, S. I., FOLKS, T. M., FEORINO, P., and MARTIN, M. A. (1985). Transcription of novel open reading frames of AIDS retrovirus during infection of lymphocytes. *Science* **229**, 1388–1390.

RATNER, L., FISHER, A., JAGODZINSKI, L. L., MITSUYA, H., LIOU, R.-S., GALLO, R. C., and WONG-STAAL, F. (1987). Complete nucleotide sequences of functional clones of the AIDS virus. *AIDS Res. Hum. Retrovirus.* **3**, 57–69.

ROBERTS, J. D., BEBENEK, K., and KUNKEL, T. A. (1988). The accuracy of reverse transcriptase from HIV-1. *Science* **242**, 1171–1173.

ROBEY, W. G., SAFAI, B., OROSZLAN, S., ARTHUR, L. O., GONDA, M. A., GALLO, R. C., and FISCHINGER, P. J. (1985). Characterization of envelope and core structural gene products of HTLV-III with sera from AIDS patients. *Science* **228**, 593–595.

SAAG, M. S., HAHN, B. H., GIBBONS, J., LI, Y., PARKS, E. S., PARKS, W. P., and SHAW, G. M. (1988). Extensive variation of human immunodeficiency virus type-1 *in vivo*. *Nature (London)* **334**, 440–444.

SANGER, F., NICKLEN, S., and COURSON, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463–5468.

SIGURDSSON, B. (1954). Maedi, a slow progressive pneumonia of sheep: An epizoological and a pathological study. *Brit. Vet. J.* **110**, 225–270.

SOBEL, E., and MARTINEZ, H. (1985). A multiple sequence alignment program. *Nucleic Acids Res.* **14**, 363–372.

SODROSKI, J., GOH, W. C., ROSEN, C., DAYTON, A., TERWILLIGER, E., and HASELTINE, W. (1986). A second post-transcriptional *trans*-activator gene required for HTLV-III replication. *Nature (London)* **321**, 412–417.

SONIGO, P., ALIZON, M., STASKUS, K., KLATZMANN, D., COLE, S., DANOS, O., RETZEL, E., TIOLLAIS, P., HAASE, A., and WAIN-HOBSON, S. (1985). Nucleotide sequence of the visna lentivirus: Relationship to the AIDS virus. *Cell* **42**, 369–382.

STADEN, R. (1982). Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. *Nucleic Acids Res.* **10**, 4731–4751.

STARCICH, B. R., HAHN, B. H., SHAW, G. M., MCNEELY, P. D., MODROW, S., WOLF, H., PARKS, E. S., and PARKS, W. P. (1986). Identification and characterization of conserved and variable regions in the envelope gene of HTLV-III/LAV, the retrovirus of AIDS. *Cell* **45**, 637–648.

STREBEL, K., KLIMKAIT, T., and MARTIN, M. A. (1988). A novel gene of HIV-1 *vpu*, and its 16-kilodalton product. *Science* **241**, 1221–1223.

TERSMETTE, M., GRUTERS, R. A., DEWOLF, F., DEGOEDE, R. E. Y., LANGE, J. M. A., SCHELLEKENS, P. T. A., GOUDSMIT, J., HUISMAN, H. G., and MIEDEMA, F. (1989). Evidence for a role of virulent human immunodeficiency virus (HIV) variants in the pathogenesis of acquired immunodeficiency syndrome: Studies on sequential HIV isolates. *J. Virol.* **63**, 2118–2125.

VAN DER MAATEN, M. J., BOOTHE, A. D., and SEGER, C. L. (1972). Isolation of a virus from cattle with persistent lymphocytosis. *J. Natl. Cancer Inst.* **49**, 1649–1657.

WAIN-HOBSON, S., GOODENOW, M., HUET, T., KWOK, S., and SNINSKY, J. (1989). HIV viruses are rapidly evolving quasi-species. In "Retroviruses of Human A.I.D.S. and Related Animal Diseases" (M. Girard and L. Valette, Eds.), pp. 18–19. Pasteur Vaccins, Paris.

WINGENDER, E. (1988). Compilation of transcription regulating proteins. *Nucleic Acids Res.* **16**, 1879–1901.

WONG-STAAL, F., CHANDA, P. K., and GHRAYEB, J. (1987). Human immunodeficiency virus: The eighth gene. *AIDS Res. Hum. Retrovirus.* **3**, 33–39.

YOKOYAMA, S., CHUNG, L., and GOJOBORI, T. (1988). Molecular evolution of the human immunodeficiency and related viruses. *Mol. Biol. Evol.* **5**, 237–251.