play an important role in regulating the persistence of the enhanced response but not its amplitude. Since protein F1 appears to be identical to axonal growth-associated proteins [GAP-43 (*13*) and pp 46 (*14*)] and B-50, which is related to PI turnover (*15*), its direct relation to synaptic enhancement (*3, 5*) may require such mechanisms, particularly presynaptic growth (*16*).

The present results suggest a new mechanism for the long-term regulation of synaptic plasticity. The redistribution of PKC activity between membrane and soluble fractions after LTP may indicate that PKC is physically translocated from the cytosol to the membrane. Since ion chelation (with EDTA) was unable to dissociate PKC activity from LTP-stimulated membranes, it is likely that PKC became strongly attached to membranes after LTP. It has been proposed that PKC is activated after its association with the membrane (*17*), and strong attachment of PKC to synaptic membranes could result in prolonged activation following LTP. Increasing the proximity of PKC to its membrane-bound substrates could result in the persistent elevation of substrate phosphorylation, as has been observed with protein F1 phosphorylation 1 hour after LTP (*18*). Consistent with this scenario is the recent observation (*19*) that iontophoretic application of phorbol ester, known to associate PKC with membranes (*10, 11*), enhances the persistence of long-term synaptic plasticity in the dentate gyrus after LTP induction.

REFERENCES AND NOTES

1. D. L. Alkon, *Science* **226**, 1037 (1984); E. R. Kandel and J. H. Schwartz, *ibid.* **218**, 433 (1982); G. Lynch and M. Baudry, *ibid.* **224**, 1057 (1984); A. Routtenberg, *Prog. Neurobiol.* **12**, 85 (1979).
2. A. Routtenberg, in *Neurobiology of Learning and Memory*, G. Lynch, J. McGaugh, N. Weinberger, Eds. (Guilford, New York, 1984), pp. 479–493.
3. ———, D. Lovinger, S. Cain, R. Akers, O. Steward, *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **42**, 755 (1983); A. Routtenberg, D. Lovinger, O. Steward, *Behav. Neural Biol.* **43**, 3 (1985).
4. A. B. Oestreicher, C. J. Van Dongen, H. Zwiers, W. H. Gispen, *J. Neurochem.* **41**, 331 (1983).
5. D. Lovinger *et al.*, *Brain Res.* **343**, 137 (1985).
6. R. Akers and A. Routtenberg, *ibid.* **334**, 147 (1985).
7. R. B. Nelson and A. Routtenberg, *Exp. Neurol.* **89**, 213 (1985).
8. A. Kishimoto, Y. Takai, T. Mori, U. Kikkawa, Y. Nishizuka, *J. Biol. Chem.* **255**, 2273 (1980).
9. M. J. Berridge, *Biochem. J.* **212**, 849 (1983).
10. A. S. Kraft and W. B. Anderson, *Nature (London)* **301**, 621 (1983).
11. M. W. Wooton and R. W. Wrenn, *FEBS Lett.* **171**, 183 (1984).
12. A distinction between early and late events in the development of a plastic change in synaptic communication [B. L. McNaughton, *J. Physiol. (London)* **324**, 249 (1982)] suggests distinct molecular mechanisms, with the increase in membrane PKC activity being related to the later synaptic events underlying long-term synaptic change.
13. J. H. P. Skene and M. Willard, *J. Neurosci.* **1**, 419 (1981).
14. R. B. Nelson, A. Routtenberg, C. Hyman, K. H. Pfenninger, *Soc. Neurosci. Abstr.* **11**, 927 (1985).
15. J. Jolles *et al.*, *Nature (London)* **286**, 623 (1980).
16. A. Routtenberg, *Behav. Neural Biol.* **44**, 186 (1985).
17. Y. Takai, *J. Biol. Chem.* **254**, 3692 (1979).
18. A. Routtenberg *et al.*, *J. Neurochem.* **44**, S135 (1985).
19. D. Lovinger, P. Colley, D. J. Linden, K. Murakami, A. Routtenberg, *Soc. Neurosci. Abstr.* **11**, 927 (1985).

20. This work was supported by National Institute of Mental Health grant MH25281 and Air Force Office of Scientific Research AFOSR 83-0335. We thank Dr. K. Murakami for his assistance and advice.

22 July 1985; accepted 21 October 1985

Equine Infectious Anemia Virus *gag* and *pol* Genes: Relatedness to Visna and AIDS Virus

ROBERT M. STEPHENS, JAMES W. CASEY, NANCY R. RICE*

Comparison of HTLV-III, the putative AIDS virus, with other related viruses, may help to reveal more about the origin of AIDS in humans. In this study, the nucleotide sequence of the *gag* and *pol* genes of an equine infectious anemia virus (EIAV) proviral DNA clone was determined. The sequence was compared with that of HTLV-III and of visna, a pathogenic lentivirus of sheep. The results show that these viruses constitute a family clearly distinct from that of the type C viruses or the BLV–HTLV-I and -II group. Within the family, EIAV, HTLV-III, and visna appear to be equally divergent from a common evolutionary ancestor.

THE FINDING OF GONDA *et al.* (*1*) that cloned genomes of HTLV-III and visna were able to form stable heteroduplexes under conditions of low stringency was the first indication that HTLV-III, the putative cause of human acquired immune deficiency syndrome, should be classified as a lentivirus. Sequence analysis of visna proviral DNA has confirmed this conclusion, revealing that visna is the closest known relative of HTLV-III (*2*). More recent data show that caprine arthritis-encephalitis virus (CAEV), a close relative to visna, also contains regions able to form stable hybrids with HTLV-III DNA (*3*).

Morphologically, equine infectious anemia virus (EIAV) resembles the lentiviruses (*4*); it also shares with them the trait of rapid antigenic variation in the infected host (*5*). Consistent with this classification, a partial cross-reaction between the p24's of lymphadenopathy-associated virus (LAV) and EIAV has been observed (*6, 7*). To determine more precisely the extent to which EIAV is related to HTLV-III, we analyzed the nucleotide sequence of a cloned EIAV proviral genome. Since it is possible to detect even quite distant relatedness if one analyzes highly conserved regions of the genome, we concentrated on the *gag* and *pol* genes.

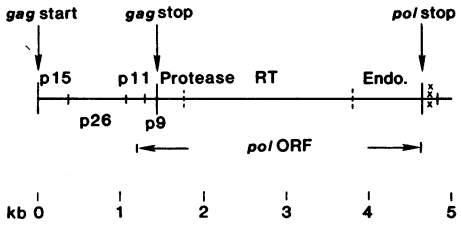


Fig. 1. Proteins encoded by the *gag* and *pol* genes of EIAV. Boundaries between the *gag* proteins are known from the work of Henderson *et al.* (*16*). Boundaries within *pol* are based on homology with other retroviruses and are approximate. In the 180 base pair region immedicately downstream of *pol* there are translation terminators in all three reading frames. RT, reverse transcriptase; Endo., endonuclease.

Full-length EIAV proviral DNA was cloned from infected equine fibroblasts (*8*), and after appropriate subcloning the 5′ half of the genome was sequenced by the method of Maxam and Gilbert (*9–11*). The organization of the coding regions is diagrammed in Fig. 1, and the full sequence from the primer binding site (PBS) to the 3′ end of the *pol* gene is shown in Fig. 2.

As in HTLV-III (*12*) and mouse mammary tumor virus (M-MTV) (*13*), the EIAV PBS is complementary to $tRNA_3^{Lys}$, while the visna PBS is complementary to the isoaccepting $tRNA_{1,2}^{Lys}$ (*2*). The two PBS sequences differ at five positions. In contrast, the mammalian type C viruses use $tRNA_{1,2}^{Pro}$ as a primer (*14*) and the avian type C viruses, $tRNA^{Trp}$ (*14*).

Beginning 116 bases after the PBS, there is an eight of nine match with the consensus splice donor sequence (*15*). This is the best

R. M. Stephens and N. R. Rice, LBI–Basic Research Program, Laboratory of Molecular Virology and Carcinogenesis, NCI–Frederick Cancer Research Facility, Frederick, MD 21701.
J. W. Casey, Section of Genetics, Laboratory of Viral Carcinogenesis, National Cancer Institute, Frederick, MD 21701.

*To whom correspondence should be addressed.

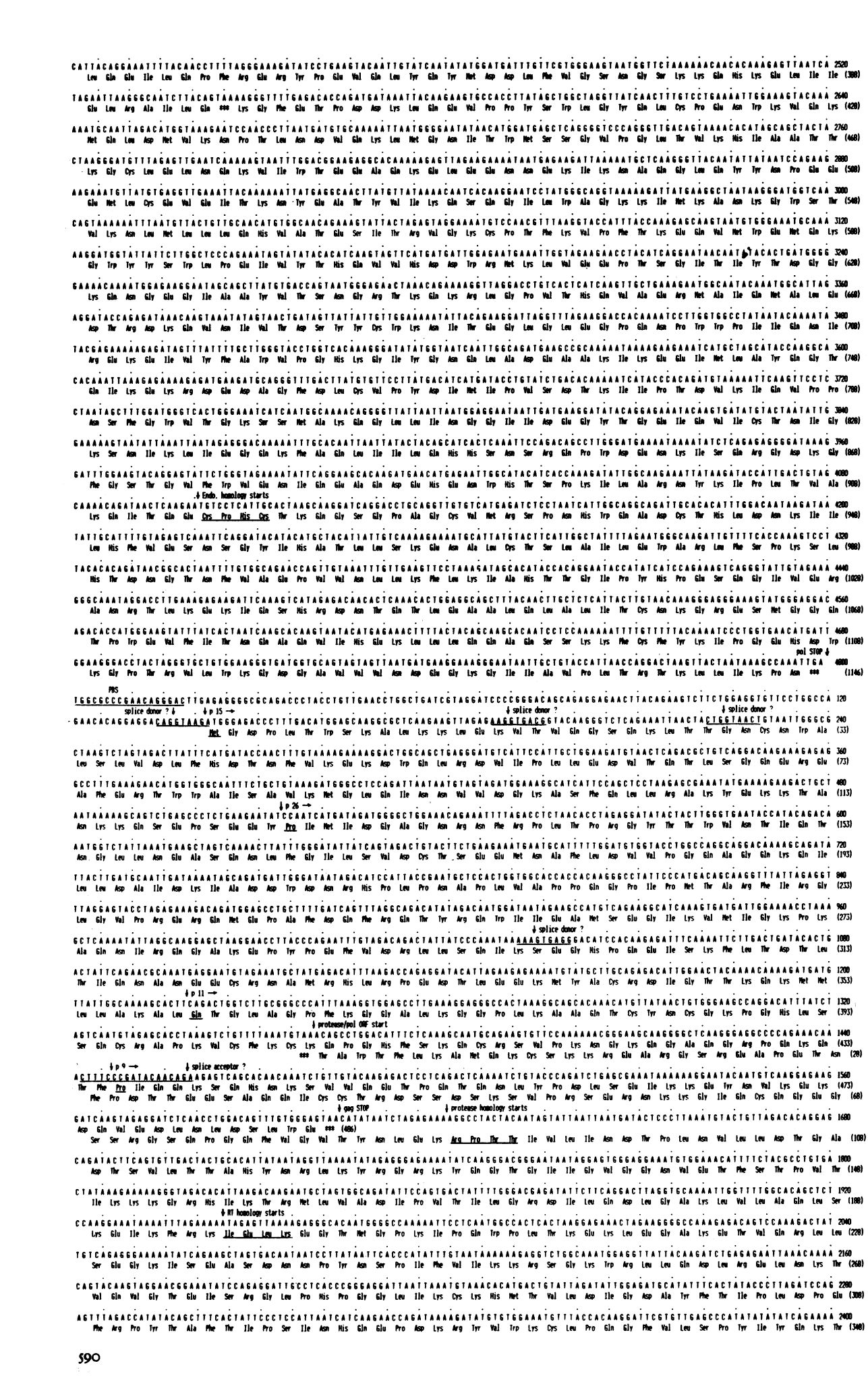


Fig. 2. Nucleotide sequence of the *gag* and *pol* genes of an EIAV proviral DNA clone. Nucleotide number beginning at the primer binding site (PBS) is given on the right at the end of each line; amino acid number within *gag* and within *pol* is shown in parentheses. Within *gag*, possible splice donors with at least a seven of nine match to the consensus sequence (*15*) are indicated. In the region between the 5′ end of the protease-*pol* open reading frame and the protease coding sequence, the best match with the consensus splice acceptor (*15*) is also shown. ***, termination codon.

Table 1. Relatedness of EIAV *gag* and *pol* proteins to those of other retroviruses, as indicated by ALIGN scores. The indicated EIAV proteins (or segments thereof) were compared with the corresponding proteins (or segments) of HTLV-III (*12*), visna (*2*), BLV (*21*), M-MuLV (*22*), and RSV (*24*) by using the computer program ALIGN. Since the program limits the number of residues in each test sequence to a maximum of 180, segmentation of a long sequence such as *pol* is necessary. The ALIGN scores given above represent the number of standard deviations by which the score of the test comparison exceeds the average score of a random comparison. Numbering for p26 segments is derived from Fig. 3; for *pol* segments, from Fig. 2. ND, not done. NA, not applicable; HTLV-III has a 132-residue deletion time in this region. BLV–HTLV-I (*25*) scores are included for comparison; for protease the HTLV-II sequence (*26*) was used.

	EIAV–HTLV-III	EIAV-visna	Visna–HTLV-III	EIAV-BLV	EIAV–M-MuLV	EIAV-RSV	BLV–HTLV-I
p15	3.7	5.7	4.1	0.8	ND	ND	8.0
p26							
1. 1–83	11.2	8.0	6.1	4.0	5.0	4.0	13.5
2. 84–128	0.7	3.2	−1.0	0.2	0.3	0.0	6.6
3. 129–171	9.6	8.7	7.8	4.9	7.7	5.4	11.0
4. 172–235	12.4	9.1	9.7	4.6	−0.3	0.2	17.4
p11	8.3	9.4	7.4	3.7	2.3	4.1	ND
Protease	8.1	7.5	9.3	5.9	7.4	6.6	10.1
pol							
1. 211–313	18.9	21.7	24.8	12.2	10.2	11.8	23.9
2. 313–420	18.9	15.2	16.0	12.5	9.4	14.4	18.8
3. 420–569	11.9	9.9	10.1	6.0	4.0	5.9	ND
4. 570–688	9.8	8.1	6.9	0.9	3.0	4.0	ND
5. 736–867	NA	17.7	NA	ND	1.5	ND	ND
6. 915–1037	14.4	20.2	12.4	10.2	10.2	12.5	24.4
7. 1038–1146	10.7	9.6	7.1	2.8	1.3	−1.7	16.7

match not only within this region but also within the long terminal repeat and within the *gag* region. Since both HTLV-III and visna also have eight of nine matches in this region with the consensus donor, we consider it likely that it is a functional site in EIAV. Three other sites that have seven of nine matches with the consensus are located within *gag* and are indicated in Fig. 2.

The presumed initiation codon for the *gag* gene begins 124 bases downstream from the PBS. From this initiator there is an open reading frame that extends for 1458 bases. Four EIAV *gag* proteins (with approximate molecular weights of 26,000, 15,000, 11,000, and 9,000) have been described (*16, 17*). From the results of Henderson *et al.* (*16*), who have performed NH_2-terminal and COOH-terminal amino acid sequence analysis of these proteins, we were able to locate their coding regions within the *gag* open reading frame. The sizes of the proteins are predicted to be: p15 (NH_2-terminal *gag* protein), 124 residues; p26, 235 residues; p11 (nucleic acid binding protein), 76 residues; and p9, 51 residues.

We used the computer program ALIGN (*18, 19*), which employs a scoring matrix and a gap penalty, to assess the relatedness of the EIAV proteins to those of other viruses. A computer-assisted alignment of p15, p26, and p11 with the corresponding proteins of HTLV-III and visna (*20*) is shown in Fig. 3, and the ALIGN scores (AS) for comparisons through *gag* and *pol* are given in Table 1.

Within the NH_2-terminal protein, relatedness among EIAV, HTLV-III, and visna is weak but significant. The best match is between EIAV and visna (35 identities, 28 percent of EIAV residues; AS, 5.7), but the EIAV–HTLV-III and visna–HTLV-III scores are not much lower. There is no apparent homology between the EIAV protein and the NH_2-terminal *gag* proteins of HTLV-I (AS, −0.1) or BLV (AS, 0.8). In addition, the computer program RELATE (*19*) failed to detect homology between the EIAV protein and either of the two *gag* proteins upstream of p30 in M-MuLV or Rous sarcoma virus (RSV). From these results it appears that EIAV, HTLV-III, and visna are more closely related to each other than to the type C viruses or to the HTLV-I–BLV family. Nevertheless, the relatedness

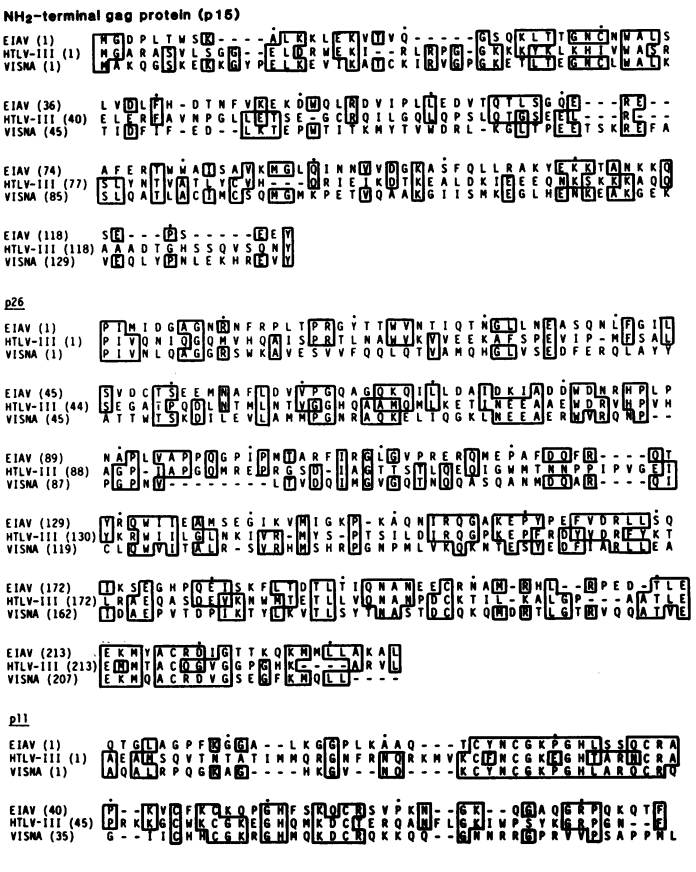
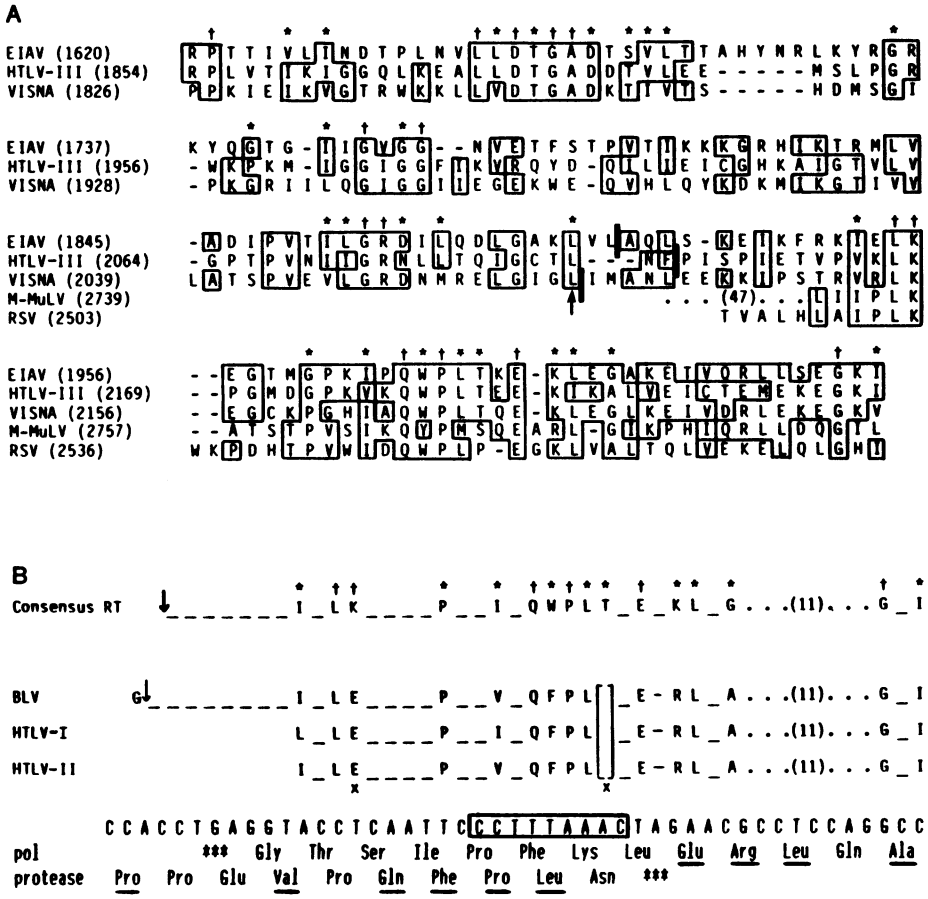


Fig. 3. Relatedness of *gag* proteins of EIAV, HTLV-III (*12*), and visna (*2*). All identities are boxed. Numbering begins with the first amino acid in each protein. The NH_2-terminus of HTLV-III p26 has been reported (*7, 12*); other HTLV-III and visna cleavage sites will be reported by Oroszlan *et al.* (*20*).

between EIAV and HTLV-III is distant. For comparison, the NH_2-terminal *gag* proteins of BLV and HTLV-I share 40 identical residues (37 percent) and yield an ALIGN score of 8.0 (*21*)

Highly significant relatedness is seen among the p26 proteins of EIAV, HTLV-III, and visna. Overall, the best match is between EIAV and HTLV-III (71 identities, 30 percent of EIAV's residues), but the EIAV-visna and HTLV-III–visna alignments are nearly as good. As in p15, this relatedness is more distant than that between the p24's of BLV and HTLV-I (44 percent identities) (*21*). Unlike p15, EIAV p26 is also related to the corresponding protein of other retroviruses. The first 83 residues are highly related to those of HTLV-III and visna (AS, 6 to ll), and also to those of BLV, M-MuLV, and RSV (AS, 4 to 5). Between residues 84 and 128 of EIAV p26, there is little apparent relatedness even to HTLV-III and visna, but the region between residues 129 and 171 is highly conserved. Homology with HTLV-III and visna is highest (AS, 8.7 to 9.6), but it is significant with BLV, RSV, and M-MuLV (AS, 4.9 to 7.7). In segments 1 and 3, therefore, the EIAV protein appears about equally divergent from those of the BLV–HTLV-I family and the avian and mammalian type C viruses. This is not so in the COOH-terminal segment (EIAV residues 172 to 235), where relatedness is apparent to HTLV-III and visna (AS, 9.1 to 12.4), and to BLV and HTLV-I (AS, 4.1 to 4.6), but not to M-MuLV or RSV. This the only region throughout *gag* and *pol* where the lentiviruses appear more closely related to the BLV–HTLV-I and -II group than to the type C viruses. It suggests the possibility of a recombination event between early members of the lentivirus and BLV–HTLV-I and -II families.



Following p11 is the p9 protein mapped by Henderson *et al.* (*16*) to the COOH-terminus of the *gag* precursor. Although there is little or no homology with the corresponding region in HTLV-III (AS, −0.6), both proteins consist of about 30 percent charged residues and contain segments of 23 to 27 amino acids where the level of charged residues is near 50 percent. The function of these proteins is unknown. In contrast to the EIAV–HTLV-III arrangement, the termination codon for the visna *gag* gene falls at the COOH-terminus of p11. Consequently, visna is not expected to encode a protein comparable to p9.

The *pol* gene, which contains regions clearly homologous to other retroviral proteases, reverse transcriptases, and endonucleases, is in a different reading frame from *gag*. The *pol* open frame overlaps *gag* by 251 bases, so that a splice or a frameshift anywhere within this rather sizable area could

From these results with p15, p26, and p11, we conclude (i) that the closest relatives of EIAV are HTLV-III and visna; (ii) that EIAV, HTLV-III, and visna are roughly equidistant from each other; (iii) that the three lentiviruses are less closely related than are BLV and HTLV-I; and (iv) that in the region spanning the COOH-terminus of p26, the lentivirus sequence is significantly more closely related to that of BLV and HTLV-I than it is to that of the type C viruses.

Fig. 4. (A) Relatedness of presumptive proteases of EIAV, HTLV-III, and visna. All identities are boxed. The NH_2-termini of the proteases are unknown; in BLV, RSV, and M-MuLV there are 8, 9, and 14 amino acids, respectively, upstream of the first residue shown above (*23, 24, 29*). HTLV-III (*12*) and visna (*2*) are numbered starting from the cap site, and EIAV, from the PBS. A dagger (†) indicates a constant residue in EIAV, HTLV-III, visna, RSV (*24*), M-MuLV (*22*), BLV (*21*), and HTLV-II (*26*); an asterisk (*) indicates that only conservative substitutions (*35*) have occurred at this site in all these viruses. The boundary between protease and RT is not known for the lentiviruses. The last residue conserved in all the retroviral proteases is the leucine indicated by an arrow. The sequence of the NH_2-terminus of RSV RT (*27*) is shown, as is the sequence of a region 47 residues downstream from the NH_2-terminus of M-MuLV RT (*28*). In this region a dagger (†) indicates a constant residue among all the sequences shown above, and an asterisk (*), only conservative substitutions. The bars indicate possible cleavage sites in the lentivirus polyproteins. (B) Translation of the *pol* gene in BLV, HTLV-I, and HTLV-II is predicted to require a frameshift. The consensus RT sequence is taken from (A), and daggers and asterisks are used as in (A). The arrow indicates the position of the NH_2-terminus of RSV RT. Sequences from BLV (*21*), HTLV-I (*25*), and HTLV-II (*26*) are shown below, with the arrow indicating the known COOH-terminus of the BLV protease (*29*). The two "x" marks indicate deviations of the BLV group from the consensus sequence. At the second "x," the lentivirus proteins will contain K or N, depending on precisely where the frameshift occurs. The BLV vase sequence in this region is also shown, with translations of both the protease and *pol* frames. Note that there is a termination codon in the protease frame immediately downstream of the predicted frameshift, and one in the *pol* frame immediately upstream of the predicted frameshift. The nine-base sequence conserved in BLV, HTLV-I, and HTLV-II is boxed.

result in *pol* gene translation. There are several sequences in this area which give a mediocre match to the consensus splice acceptor, and since there are multiple possible splice donors in *gag*, a splicing mechanism cannot be ruled out. This arrangement of the *gag*, protease, and *pol* coding sequences in EIAV, HTLV-III, and visna is different from that of M-MuLV (where all are in the same reading frame) (*22*, *23*), of RSV (where *gag* and protease are in a different frame from *pol*) (*24*), and of BLV (*21*) and HTLV-I (*25*) and HTLV-II (*26*) (where all three are in different frames).

Although the NH_2-terminus of the presumptive protease is not yet known, relatedness to other retrovirus proteases is apparent in the translated sequence beginning 18 bases after the *gag* termination codon and extending for at least 95 amino acids (Fig. 4). The EIAV sequence is most highly related to that of HTLV-III (AS, 8.1), but the EIAV-visna and HTLV-III–visna scores are comparable (AS, 7.5 to 9.3). As in p26, the EIAV protease is also related to that of BLV, M-MuLV, and RSV (AS, 5.9 to 7.4).

The boundary between protease and reverse transcriptase (RT) is not known but can be estimated. Only 14 residues after the last residue known to be conserved in all the viral proteases (the leucine is indicated by an arrow in Fig. 4), clear homology with M-MuLV and RSV RT has begun. The protease COOH-terminus must therefore fall within this short span. Possible cleavage sites can be identified in this region in EIAV, HTLV-III, and visna, based on the known cleavage sites within *gag* (*16*, *20*).

Having established that a sequence at the NH_2-terminus of RSV RT (*27*), near the NH_2-terminus of M-MuLV RT (*28*), and at the presumptive NH_2-terminus of lentivirus RT is quite conserved, one can use this information to predict the NH_2-terminus in other viruses where it is not immediately obvious, namely, the BLV–HTLV-I and -II group. In that group, as in RSV, protease and *pol* are in different reading frames, and the mechanism of translation of the *pol* gene is unknown. If we examine the translated BLV sequence for homology with the conserved RT sequence, we find that the first half of the sequence exists in the protease reading frame, immediately downstream from the known COOH-terminus of the mature protease (*29*) (Fig. 4B). This arrangement is seen in both HTLV-I and -II, though the COOH-terminus of the protease is not known for those viruses. The second half of the conserved RT sequence is contiguous with the first half but is in the *pol* reading frame in all three viruses. This suggests that in order to translate an RT homologous to those of RSV, M-MuLV, and the lentiviruses, a frameshift must occur between the first and second halves of this conserved sequence. It is of interest that there is a sequence of nine bases perfectly conserved in BLV, HTLV-I, and HTLV-II at the site of the presumptive frameshift (Fig. 4B). The prediction is that most of the time translation proceeds to the first stop codon in the protease frame (just downstream of the frameshift region) and terminates. But occasionally a frameshift occurs, somehow promoted by the conserved base sequence, resulting in *pol* gene translation.

We also analyzed the relatedness of EIAV *pol* to that of other viruses by dividing it into segments and comparing each with the corresponding region of HTLV-III, visna, BLV, M-MuLV, and RSV. As expected (*2*, *21*, *30*), we found a highly conserved region (segments 1 and 2) in the NH_2-terminal half of the presumptive RT. Pairwise comparisons of EIAV, HTLV-III, and visna in segment 1 yielded ALIGN scores of 18.9 to 24.8 (best match, visna–HTLV-III), and in segment 2, 15.2 to 18.9 (best match, EIAV–HTLV-III). In these two segments, 53 percent of EIAV's 211 residues are identical in HTLV-III. EIAV is also highly related in this region to BLV, M-MuLV, and RSV (AS, 10.2 to 14.4). EIAV, unlike HTLV-III and visna, has a termination codon in segment 2 (residue 397 in Fig. 2), indicating that the clone we are sequencing is probably defective.

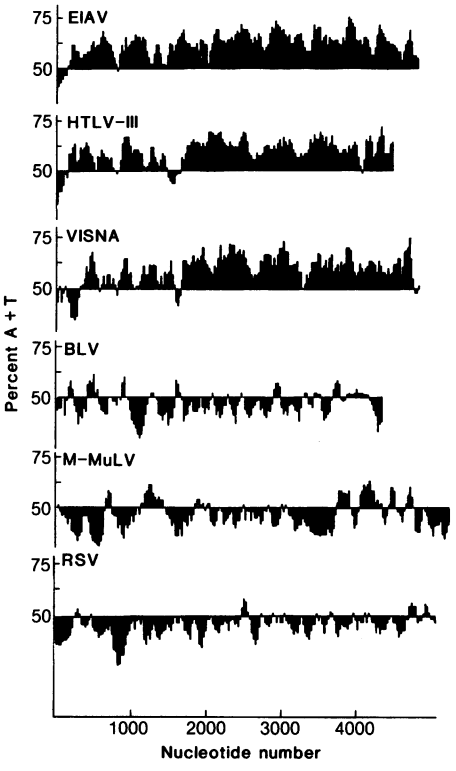


Fig. 5. Base composition of retroviral coding strands in the *gag-pol* region. A + T content was computed by using a window size of 100 bases and a slide of 25 bases. The M-MuLV sequence begins at the 5′ end of the *gag* gene; all other begin at the primer binding site. All extend to the 3′ end of the *pol*.

The same pattern of relatedness is observed for EIAV segment 6 which, on the basis of homology with M-MuLV and RSV, is located near the NH_2-terminus of the presumptive endonuclease (*31*, *32*). Pairwise comparisons of EIAV, HTLV-III and visna give scores of 12.4 to 20.2 (best match, EIAV-visna). Scores for comparisons between EIAV and BLV, M-MuLV, and RSV are also very high (10.2 to 12.5).

With EIAV segments 3, 4, and 7, which represent COOH-terminal regions of RT and endonuclease, a different pattern is seen. Pairwise comparisons between EIAV, HTLV-III, and visna give highly significant scores (6.9 to 11.9; best match is EIAV–HTLV-III in each case), but relatedness to the corresponding regions of BLV, M-MuLV, and RSV varies considerably. In segment 3 this homology, while lower than in segments 1, 2, and 6, is still significant (AS, 4.0 to 6.0). Segment 4 appears to be weakly related to M-MuLV and RSV (AS, 3.0 and 4.0) and not to BLV, and in segment 7 there is no significant homology. The remaining EIAV segment (No. 5) is related only to that of visna, for the HTLV-III *pol* sequence has a 132–amino acid deletion in this region. On the basis of the alignment with other *pol* genes, this falls within the highly variable region at the COOH-terminus of RT and NH_2-terminus of endonuclease (*28*, *31*, *32*).

Similarity between EIAV and HTLV-III and visna is also apparent on the relatively gross level of overall base composition. In the *gag-pol* region the coding strand of each has very high adenine (A) content (38 to 39 mole percent) and very low cytosine (C) content (15 to 18 mole percent). This results in an unusually high A + T (T, thymidine) content (60 to 62 mole percent), which is very different from that of the type C viruses of the BLV–HTLV-I family (45 to 47 mole percent). In consequence there should be little stable secondary structure in lentivirus genomic RNA, and this may account for the relative ease with which the EIAV RT synthesizes full-length complementary DNA in the endogenous reaction (*33*).

In conclusion, EIAV appears to resemble HTLV-III rather than visna in the use of $tRNA_3^{Lys}$ and in encoding a p9 at the COOH-terminus of *gag*. However, both EIAV and visna have a sizable insertion (132 amino acids in EIAV) relative to HTLV-III near the COOH-terminus of RT. Depending on the region of the

genome, EIAV appears slightly more related to HTLV-III (p26, protease, segments 2, 3, 4, and 7 of *pol*), or visna (p15, p11, segments 1 and 6 of *pol*), but with the possible exception of *pol* segment 6, these differences are not large. Since visna and HTLV-III appear no closer to each other than either is to EIAV, we conclude that the three viruses are about equidistant from each other.

Chiu *et al.* (*34*) recently compared the translated nucleotide sequence at the NH_2-terminus of RT (corresponding to our *pol* segments 1 and 2) in EIAV, CAEV, and HTLV-III. They concluded that the three viruses are about equidistant from each other. Since CAEV is a close relative of visna (81 percent amino acid identities in the region sequenced by Chiu *et al.*), our results are in good agreement with theirs. They also concluded that divergence among the lentiviral *pol* genes is about the same as that observed between the *pol* genes of BLV and HTLV-I. This conclusion, however, while true for *pol* segments 1 and 2 does not apply to *pol* segments 6 and 7 (see Table 1). There, as throughout *gag* and protease, BLV and HTLV-I are more highly related to each other than are EIAV, visna, and HTLV-III to each other. Similarly, within RT the BLV–HTLV-I and -II group appears more closely related to the avian type C viruses than to the mammalian type C viruses (*2*, *21*, *34*), but in other genomic regions closer relatedness to the mammalian viruses is seen (*21*). Thus a phylogenetic tree that accurately reflects relationships in one segment of the genome may not apply to other segments. This may result from real differences in evolutionary rate, from limitations in our methods of analyzing relatedness, from viral recombination, or from some combination of these factors. Until this issue is clarified further, we view the lentiviruses, the BLV–HTLV-I and -II group, the mammalian type C viruses, and the avian type C viruses as four major retroviral groups that are approximately equidistant from each other.

REFERENCES AND NOTES

1. M. A. Gonda *et al.*, *Science* **227**, 173 (1985).
2. P. Sonigo *et al.*, *Cell* **42**, 369 (1985).
3. M. Gonda *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, in press
4. M. Gonda *et al.*, *Am. J. Vet. Res.* **39**, 731 (1978).
5. O. Narayan *et al.*, *J. Gen. Virol.* **41**, 343 (1978); Y. Kono *et al.*, *Arch. Virol.* **41**, 343 (1973); R. Montelara *et al.*, *J. Biol. Chem.* **259**, 10539 (1984).
6. L. Montagnier *et al.*, *Ann. Virol.* **135E**, 119 (1984).
7. J. M. Casey *et al.*, *J. Virol.* **55**, 417 (1985).
8. L. S. Levy *et al.*, unpublished results.
9. A. M. Maxam and W. Gilbert, *Methods Enzymol.* **65**, 499 (1980).
10. For sequencing, fragments were subcloned into pUC8 or modified pUC8 vectors. Reactions for G, G + A, C, C + T, and A + C were performed. For about 64 percent of the DNA, the sequence of both strands was determined. There were no compressions within the remaining 36 percent. Storage and analysis of the sequence data were performed on an IBM PC as described (*11*)
11. R. Stephens, *Gene Anal. Tech.* **2**, 67 (1985).
12. L. Ratner *et al.*, *ibid.*, p. 450; S. Wain-Hobson *et al.*, *Cell* **40**, 9 (1985); R. Sanchez-Pescador *et al.*, *Science* **227**, 484 (1985).
13. G. Peters and C. Glover, *J. Virol.* **35**, 31 (1980).
14. F. Harada *et al.*, *J. Biol. Chem.* **250**, 3487 (1975); G. Peters *et al.*, *J. Virol.* **21**, 1031 (1977); G. Peters and C. Glover, *ibid.* **33**, 708 (1980).
15. S. M. Mount, *Nucleic Acids Res.* **10**, 459 (1982).
16. L. Henderson *et al.*, in preparation.
17. B. Parekh *et al.*, *Virology* **107**, 520 (1980); R. C. Montelaro *et al.*, *J. Virol.* **42**, 1029 (1982).
18. The computer program ALIGN (*19*) brings two sequences into optimal alignment based on the mutation data matrix, a matrix bias of 6, and a gap penalty of 6, and then assigns a score to the comparison. A score for a random comparison of two sequences of the same length and composition is obtained by jumbling the two and comparing them. This latter operation was performed 300 times, yielding an average score for a random comparison. The align score (AS) expresses in standard deviation (SD) units how far the test score is above the average score. A score of 5 SD (probability of chance occurrence 3×10^{-7}) has been taken to imply evolutionary relatedness of two sequences, and a score between 3 and 5 SD is suggestive of such relatedness.
19. M. O. Dayhoff, in *Atlas of Protein Sequence and Structure*, M. O. Dayhoff, Ed. (National Biomedical Research Foundation, Washington, DC, 1979), vol. 5, suppl. 3, pp. 1–8.
20. S. Oroszlan *et al.*, unpublished results.
21. N. R. Rice *et al.*, *Virology* **142**, 357 (1985).
22. T. Shinnick *et al.*, *Nature (London)* **293**, 543 (1981)
23. Y. Yoshinaka *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 1618 (1985).
24. D. E. Schwartz *et al.*, *Cell* **32**, 852 (1983).
25. M. Seiki *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3618 (1983).
26. K. Shimotohno *et al.*, *ibid.*, **82**, 3101 (1985).
27. T. D. Copeland *et al.*, *J. Virol.* **36**, 115 (1985).
28. T. D. Copeland *et al.*, *Virology* **143**, 676 (1985).
29. Y. Yoshinaka *et al.*, *J. Virol.*, in press.
30. H. Toh *et al.*, *Nature (London)* **305**, 827 (1983).
31. P. Hippenmeyer and D. Grandgenett, *Virology* **137**, 358 (1984).
32. Y. Yoshinaka *et al.*, *J. Virol.* in press.
33. N. R. Rice *et al.*, *ibid.* **29**, 907 (1979).
34. I.-M. Chiu *et al.*, *Nature (London)* **317**, 366 (1985).
35. Conservative groupings are A, G, P, S, and T; I, L, M, and V; D, E, N, and Q; H, K, and R; F, Y, and W; R. M. Schwartz and M. O. Dayhoff, in *Atlas of Protein Sequence and Structure*, M. O. Dayhoff, Ed. (National Biomedical Research Foundation, Washington, DC, 1978), vol. 5, suppl. 3, pp.353–358.

36. We thank J. Elser for help in sequencing and S. Gregory for manuscript preparation. We thank L. Henderson and S. Oroszlan for sharing results with us. Supported by NCI DHHS (contract NO1-C)-23909 with Litton Bionetics).

2 October 1985; accepted 14 November 1985

Putative Reverse Transcriptase Intermediates of Human Hepatitis B Virus in Primary Liver Carcinomas

H. WILL,* J. SALFELD, E. PFAFF, C. MANSO, L. THEILMANN, H. SCHALER

Nucleocapsid-pol fusion proteins have been detected by serological screening hepatocellular carcinoma tissues that contain hepatitis B virus (HBV) DNA. The existence of these fusion proteins suggests that HBV may synthesize its reverse transcriptase in a fashion analogous to the way that retroviruses synthesize and process a precursor. The accumulation of HBV reverse transcriptase intermediates in tumorous tissues and not in other tissues may be related to the absence of viral core particles and possibly contributes to tumor development.

HUMAN HEPATITIS B VIRUS (HBV) contains a DNA genome but resembles retroviruses (i) in its mode of replication involving reverse transcription of an RNA pregenome, (ii) in its gene organization, and (iii) in its association with tumor development (*1*). At the level of transcription, both differences and similarities between HBV and retroviruses exist (*2*). In contrast to retroviruses (*3*), all transcripts identified so far are unspliced and originate from different promoters; however, as in retroviruses, they are coterminal. For HBV and the related duck hepatitis B virus (DHBV) the major messenger RNA's (mRNA's) identified have been related to the hepatitis B surface antigen (HBsAg) and nucleocapsid protein, hepatitis B core antigen (HBcAg), but not to the reverse transcriptase believed to be encoded by the pol frame (*4*). Only the HBcAg mRNA covers the complete pol frame; for expression of the pol protein from this mRNA, internal translation initiation would have to occur. Alternatively, in analogy to retroviruses (*3*), the HBV reverse transcriptase could be synthesized via a HBcAg-Pol (c-pol) fusion protein. This is conceivable since the HBcAg coding region (C gene) of HBV overlaps with that of the pol frame reminiscent of the gag-pol frame arrangement in most retroviruses. As confirmation of this hypothesis, we have identified and charac-

Mikrobiologie and ZMBH, University of Heidelberg, Im Neuenheimer Feld 230 and 282, 6900 Heidelberg, Federal Republic of Germany.

*Present address: Max-Planck-Institut für Biochemie, Am Klopferspitz, 8033 Martinsreid/München, Federal Republic of Germany.