JOURNAL OF VIROLOGY, Aug. 1990, p. 3716–3725
0022-538X/90/083716-10$02.00/0

Vol. 64, No. 8

Cloning and Characterization of cDNAs Encoding Equine Infectious Anemia Virus Tat and Putative Rev Proteins

ROBERT M. STEPHENS,[1] DAVID DERSE,[2] AND NANCY R. RICE[1]*

Laboratory of Molecular Virology and Carcinogenesis, ABL-Basic Research Program,[1] and Laboratory of Viral Carcinogenesis,[2] National Cancer Institute-Frederick Cancer Research and Development Center, Frederick, Maryland 21701

Received 21 February 1990/Accepted 10 May 1990

We isolated and characterized six cDNA clones from an equine infectious anemia virus-infected cell line that displays a Rev-defective phenotype. With the exception of one splice site in one of the clones, all six cDNAs exhibited the same splicing pattern and consisted of four exons. Exon 1 contained the 5′ end of the genome; exon 2 contained the *tat* gene from mid-genome; exon 3 consisted of a small section of *env*, near the 5′ end of the *env* gene; and exon 4 contained the putative *rev* open reading frame from the 3′ end of the genome. The structures of the cDNAs predict a bicistronic message in which Tat is encoded by exons 1 and 2 and the presumptive Rev protein is encoded by exons 3 and 4. *tat* translation appears to be initiated at a non-AUG codon within the first 15 codons of exon 1. Equine infectious anemia virus-specific *tat* activity was expressed in transient transfections with cDNA expression plasmids. The predicted wild-type Rev protein contains 30 *env*-derived amino acids and 135 *rev* open reading frame residues. All of the cDNAs had a frameshift in exon 4, leading to a truncated protein and thus providing a plausible explanation for the Rev-defective phenotype of the original cells. We used peptide antisera to detect the faulty protein, thus confirming the cDNA sequence, and to detect the normal protein in productively infected cells.

Equine infectious anemia virus (EIAV) belongs to the lentivirus group of the retrovirus family. It is related in morphology (19), genome organization (27, 49, 56), and sequence (9, 27, 49, 56) to the other members of that group, including the human, simian, feline, and bovine immunodeficiency viruses, visna virus, and caprine arthritis-encephalitis virus. Like the other lentiviruses, EIAV infects cells of the monocyte-macrophage lineage (28, 36), causing a persistent, lifelong infection of its host. EIAV infection can, in some cases, be controlled by the host, resulting in asymptomatic periods that may last for many years. It would obviously be of great interest to understand the mechanism of this host control.

Replication of human immunodeficiency virus type 1 (HIV-1) is influenced by at least six genes in addition to *gag*, *pol*, and *env*. These are located either between *pol* and *env* or at the 3′ end of the genome, and in most cases their existence is advertised by open reading frames (ORFs) of substantial length. In contrast, EIAV appears to be simpler in genetic structure. Its genomic RNA is shorter than that of HIV-1 by about 1.2 kilobases (kb), and it completely lacks the large *vif* and *nef* genes of the latter. Three auxiliary ORFs have been pointed out in EIAV (49). The first (S1) is between *pol* and *env*, and its translated sequence is strikingly similar to that of HIV-1 *tat*. The HIV-1 Tat protein acts on sequence elements in the viral long terminal repeat (LTR) to increase viral gene expression (1, 25, 26, 39, 41, 47, 54, 58). In keeping with this interpretation of S1, transactivation of the EIAV LTR has been observed in infected cells (13, 53), and the gene responsible for this effect was localized to a 0.8-kb fragment containing ORF S1 (14, 15, 53). The second EIAV ORF, S2, is immediately downstream from S1 and overlaps the start of *env*, a position similar to that of *vpu* in HIV-1. There is no sequence homology with *vpu*, however, and no other clues to the nature of this EIAV ORF. S3 overlaps the 3′ end of the *env* gene and may therefore be the equivalent of HIV-1 *rev*. The HIV-1 Rev protein affects the partitioning of specific viral RNAs between the nucleus and cytoplasm. In the presence of Rev, cytoplasmic RNA consists predominantly of unspliced genomic and partially spliced *env* mRNAs; in the absence of Rev, cytoplasmic RNA consists of small multiply spliced mRNAs (16, 17, 22, 32, 33).

To understand the control of viral gene expression in EIAV-infected cells, it is necessary to delineate the structure, expression, and activity of the viral regulatory proteins, including those encoded by ORFs S1, S2, and S3. Therefore, we analyzed cDNAs from a library derived from infected cells. We found that, as in HIV-1, both the Tat and putative Rev proteins are products of two exons. Unlike HIV-1, one of the EIAV Tat-coding exons is derived from the 5′ end of the genome, and translation initiation evidently occurs without an AUG codon. Also unlike HIV-1, coding exon 1 of the putative EIAV Rev protein is derived from the *env* gene.

MATERIALS AND METHODS

Cells. EIAV-producing cultures of equine fetal kidney cells were grown in Eagle minimal essential medium containing 15% fetal bovine serum, 1% sodium pyruvate, and nonessential amino acids. EIAV-infected canine thymus cells (Cf2Th), and EIAV-infected feline (Fea) cells, and uninfected canine osteosarcoma cells (D17) were grown in Dulbecco modified Eagle medium containing 10% fetal bovine serum. Single-cell clones of Cf2Th(EIAV) were established by plating cells into 96-well dishes at 0.6 cell per well. Cells from 25 wells that contained apparent single-cell colonies were propagated. Primary horse macrophages were cultured and infected with the Wyoming strain of EIA as previously described (45a).

RNA isolation and analysis. RNAs from tissue culture cell lines were isolated by using the guanidine isothiocyanate procedure of Chirgwin et al. (8). RNAs from infected pri-

* Corresponding author.

mary horse macrophages was isolated on successive days postinfection by using RNAzol as recommended by the manufacturer (Biotecx Laboratories, Friendswood, Tex.). In some cases, the RNA was poly(A) selected by a single pass through an oligo(dT)-cellulose column (2). For Northern (RNA) analysis, RNA samples were fractionated electrophoretically in 1% agarose gels in 2.2 M formaldehyde (34). Samples were blotted onto nitrocellulose and hybridized with nick-translated, ^{32}P-labeled EIAV DNA at 2×10^6 cpm/ml (specific activity, about 5×10^8 cpm/μg of DNA). Hybridization was performed at 42°C for 24 h in buffer containing 40% formamide and 5× SSC (1× SSC is 0.15 M sodium chloride plus 0.015 M sodium citrate). The most stringent wash was 0.4× SSC at 68°C. The probe was a full-length molecular clone of EIAV (56).

cDNA library construction and screening. A 4-μg sample of poly(A)$^+$ RNA from cell clone 6 was used as starting material with the cDNA synthesis kit from Invitrogen, Inc. (San Diego, Calif.). The resulting cDNA was not size selected. The total library size was 1.6×10^6 recombinants. A sample of the product was ligated into the plasmid vector supplied (pTZ18R), and after transformation of DH5αF′ cells a total of 200,000 colonies were screened with a nick-translated full-length EIAV probe, as described above. DNA minipreparations of the approximately 140 positive colonies revealed that 14 had DNA inserts of at least 1.2 kb, as expected for a *tat* or *rev* message. DNA sequencing with universal and reverse primers showed that six of these contained sequences from both the 3′ and 5′ ends of the EIAV genome. DNAs of these six were prepared by the alkaline extraction method (6), followed by centrifugation in CsCl density gradients with ethidium bromide.

DNA sequencing. DNAs were sequenced by the dideoxy chain termination method (51), using the Sequenase system and protocol (United States Biochemicals).

Transactivation by *tat*. The assay and vectors of Dorn et al. (14) were used to study transactivation by *tat*. Briefly, the assay measures the activity of the gene for chloramphenicol acetyltransferase (CAT), which is dependent for its transcription upon the EIAV LTR; these are linked in plasmid pEIcat (15). cDNAs to be tested were excised from library vector pTZ18R by using *Pst*I and *Hin*dIII and inserted into expression plasmids pRSPA-K (sense) and pRSPA-S (antisense). These plasmids consist of Bluescript-KS (Stratagene) with the Rous sarcoma virus promoter and the simian virus 40 polyadenylation signal inserted into the multiple cloning site. pRSPA-K and pRSPA-S differ only in the orientation of these signals relative to the remaining cloning sites. To measure transactivating activity, canine (D17) cells were cotransfected with 5 μg of pEIcat and 1 μg of a pRSPA-K or pRSPA-S cDNA expression plasmid. CAT activity was measured after 48 h as previously described (14).

Mutagenesis of cDNA 9. cDNA 9 was digested with *Bam*HI, which cuts within the *tat* ORF of exon 1. The staggered ends were filled in with nucleoside triphosphates by using the Klenow fragment of *Escherichia coli* DNA polymerase, and the resulting blunt ends were ligated. The creation of a four-base insertion was verified by DNA sequencing of the product.

Stable transfection of cDNA 11. D17 cells were cotransfected by the calcium phosphate procedure (15) with 10 μg of pRS-cDNA 11 and 0.1 μg of pNeoHiC, which contains the neomycin resistance gene linked to the simian virus 40 early promoter in a pUC18 backbone. Cells resistant to 700 μg of G-418 (GIBCO Laboratories, Grand Island, N.Y.) per ml were selected and propagated.

Synthetic peptide antisera. Production of rabbit antisera to three *env*-derived peptides is described elsewhere (45a). Peptide sgp-1 represents residues 41 to 55 of the mature surface glycoprotein; peptide tm-2 represents residues 247 to 259 of the transmembrane protein; and peptide tm-3 represents residues 402 to 415 of the transmembrane protein. In addition, antiserum to peptide sgp-3, whose sequence is Gly-Ile-Ser-Thr-Pro-Ile-Thr-Gln-Gln-Ser-Glu-Lys-Ser-Lys-Cys (residues 7 to 21 of the mature surface glycoprotein), was raised. All of these sera precipitate the *env* precursor polyprotein, as well as the mature viral protein.

Three *rev* ORF-derived peptides were also used. The sequence of peptide rev-1 is Lys-Ile-Asp-Pro-Gln-Gly-Pro-Leu-Glu-Ser-Asp-Gln-Cys (residues 29 to 40 of the predicted Rev protein); the sequence of peptide rev-2 (residues 45 to 58) is Cys-Leu-Arg-Gln-Ser-Leu-Pro-Glu-Glu-Lys-Ile-Pro-Ser-Gln-Thr; and the sequence of peptide rev-3 (residues 66 to 76) is Cys-Gly-Pro-Gly-Pro-Thr-Gln-His-Thr-Pro-Ser-Arg. sgp-3 and rev-1 were purchased from Multiple Peptide Systems (Pasadena, Calif.); rev-2 was purchased from Peninsula Laboratories (San Carlos, Calif.); and rev-3 was synthesized by Terry Copeland (ABL-Basic Research Program, Frederick, Md.).

sgp-3, rev-1, rev-2, and rev-3 were coupled to keyhole limpet hemocyanin through their cysteine residues (the cysteines of rev-1, rev-2, and rev-3 do not occur in the EIAV sequence and were added to facilitate coupling). Coupling and production of antisera have been described previously (45).

Immune precipitations. Cells were grown for 3 to 5 h in methionine- and cysteine-free medium containing 10% dialyzed fetal bovine serum, [^{35}S]methionine, and [^{35}S]cysteine, each at 50 μCi/ml (600 to 1,000 Ci/mmol; Amersham Corp. Arlington Heights, Ill.). Cells were lysed in radioimmune precipitation assay buffer (20 mM Tris [pH 7.5], 2 mM EDTA, 150 mM sodium chloride; 1% sodium deoxycholate, 1% Triton X-100, 0.25% sodium dodecyl sulfate) containing 1 mM phenylmethylsulfonyl fluoride and 200 Kallikrein units of aprotinin per ml. Lysates were centrifuged for 10 min at 10,000 × *g*, and supernatant samples were incubated with antiserum at 4°C for 90 min in protein A-Sepharose (Pharmacia, Inc., Piscataway, N.J.). Precipitates were collected by centrifugation, washed three times in radioimmune precipitation assay buffer, boiled in gel loading dye, and loaded onto sodium dodecyl sulfate–10 to 20% polyacrylamide gels. After electrophoresis, the gels were enhanced with diphenyloxazole-dimethyl sulfoxide as previously described (45), dried, and exposed to X-ray film.

RESULTS

Northern analysis of EIAV transcripts. If EIAV *tat* and *rev* transcripts are generated by splicing patterns similar to those observed for HIV-1, these transcripts are expected to be about 1.5 kb long (1, 16, 38, 50). We therefore examined RNAs from EIAV-producing cultures for transcripts of this size. We found that Northern blots of total RNAs from the equine, feline, and canine cell lines showed no detectable hybridizing 1.5-kb RNA, although they did have both genomic and envelope mRNAs. Poly(A) selection of the RNA before electrophoresis occasionally enabled detection of 1.5-kb RNA in the canine cells. However, in contrast to the results of others (14, 43), we did not detect 1.5-kb EIAV RNA in poly(A)$^+$ RNA from infected equine or feline cells

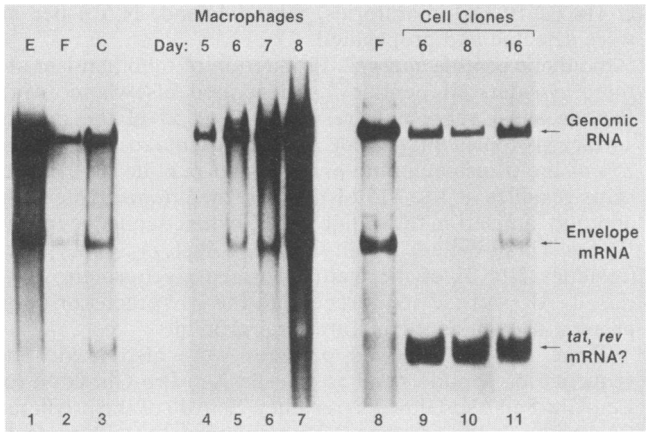

FIG. 1. Northern analysis of RNAs from EIAV-infected cells. A 1-μg sample of poly(A)$^+$ RNA (lanes 1 to 3) or a 15-μg sample of total RNA (lanes 4 to 11) was fractionated by electrophoresis in a denaturing agarose gel. RNA was blotted onto nitrocellulose and hybridized with a nick-translated probe representing the entire EIAV genome. Lanes 1 to 3 and 8 contained RNAs from cells that produced attenuated tissue culture-adapted EIAV. Lanes 4 to 7 contained RNAs from primary horse macrophages 5 to 8 days postinfection with the Wyoming strain of EIAV. Cell clones 6, 8, and 16 were derived from the EIAV-producing canine thymus culture. E, Equine kidney cells; F, feline cells; C, canine thymus cells. The exposure time for lanes 1 to 7 was 18 h with no intensifying screen, and that for lanes 8 to 11 was 18 h with an intensifying screen.

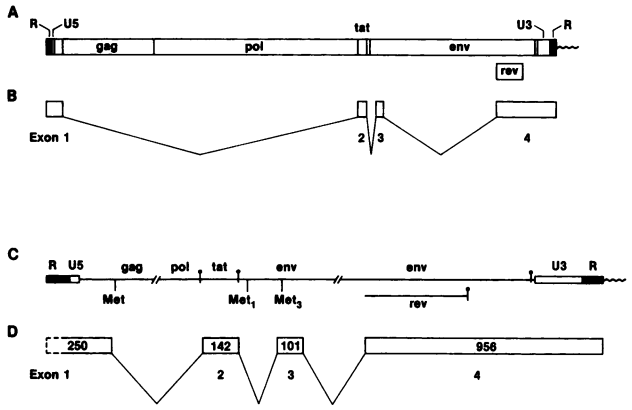

FIG. 2. Schematic representation of EIAV cDNA clones. (A) Diagram of genomic RNA (27, 49, 56). Boundaries of *gag*, *pol*, and *env* coding regions are indicated by vertical lines. The *tat* (S1) ORF lies immediately downstream of *pol*. The presumptive *rev* ORF (S3) overlaps that of *env*. (B) Locations of the exons relative to the genome in diagram A. (C) Diagram of genomic RNA emphasizing regions of the exons. ♀, Termination codon. Met_1 is the initiator of the envelope polyprotein (3), and Met_3 is methionine 3 in the *env* ORF. (D) Locations and sizes of the exons relative to the genome in diagram C. We presume, but did not determine, that exon 1 begins at the 5′ end of the R region and contains 250 bases. The cDNAs varied in their 5′ termini, with the longest containing 232 bases in exon 1.

(Fig. 1, lanes 1 to 3). Similarly, primary horse macrophages infected with the virulent Wyoming strain virus contained no detectable low-molecular-weight transcripts (lanes 4 to 7). This meant that isolation of *rev* and *tat* cDNAs would be very difficult, for nearly all cDNAs would be synthesized from genomic and *env* messages.

For reasons unrelated to cDNA cloning, we have established single-cell clones from the EIAV-producing canine thymus cell line. Many of these clones do not produce infectious virus, and they release few or no particles into the medium. When total RNAs from several such clones were examined, the EIAV mRNA was found to be predominantly about 1.5 kb long, with some genomic and very little *env* mRNA (Fig. 1, lanes 9 to 11). Cytoplasmic RNA contained no detectable full-length or *env* mRNA (data not shown). Assuming similar strategies for the control of EIAV and HIV-1 gene expression, these cells appear to display a Rev-defective phenotype characterized by accumulation of fully spliced transcripts in the cytoplasm. The cell clones thus provide an abundant source of the multiply spliced RNAs that encode regulatory proteins. Analysis of cDNA clones that represent these RNAs would allow characterization of the splicing pattern and prediction of the amino acid sequences of the regulatory proteins. We reasoned that the basis for the Rev-defective phenotype would be obvious from the sequence of the cDNAs and that the authentic Rev product could be predicted. These cells provide an excellent reagent for testing of the putative product, since some HIV-1 *rev* mutants have been rescuable when a competent Rev protein was supplied in *trans* (22, 23, 31, 46).

cDNA clones. A plasmid cDNA library was constructed by using poly(A)-containing RNA from cell clone 6 (Fig. 1, lane 9). Screening of about 2×10^5 colonies with a full-length EIAV probe resulted in about 140 positive clones. Fourteen had an insert of at least 1.2 kb, and DNA sequencing revealed that six of these contained sequences from both the 5′ and 3′ ends of the EIAV genome. All six of these were analyzed in detail. Three (no. 6, 9, and 11) were completely sequenced, and the splice junctions were established in the other three (no. 1, 3, and 5). The splicing pattern was deduced by alignment of the cDNA sequences with the genomic sequence (27, 49, 56).

With one exception, noted below, all of the analyzed clones showed the same splicing pattern and contained four exons (Fig. 2; Table 1). Exon 1 contains 250 bases and consists of the R and U5 regions, the primer-binding site, and 118 bases between the primer-binding site and *gag*. The splice donor lies 5 bases upstream of the initiator methionine codon of *gag*. Exon 2 (142 bases) contains most of the S1 ORF, which immediately follows the *pol* gene termination codon in the same frame. In clones 1, 3, 5, 6, and 11, the first splice acceptor is located 13 bases downstream of the *pol* terminator; in clone 9, the acceptor is four additional bases downstream (this is the only difference in splicing pattern among all six clones). At the 3′ end of exon 2, the splice donor immediately follows a termination codon that ends the *tat* ORF. Exon 3 (101 bases) is derived from the *env* gene, near (but not at) its 5′ terminus. It does not contain the *env* gene initiator or signal peptide but does contain methionine codons 3 and 4 in the *env* ORF. Exon 4 (956 bases) contains the 3′ half of the transmembrane protein-coding sequence, as well as the entire S3 ORF, which overlaps that of *env*. These cDNAs contain two nonoverlapping ORFs. ORF 1 begins in exon 1 and continues into S1 in exon 2; this ORF is expected to encode Tat. ORF 2 is derived from exons 3 and 4 and fuses a short section of *env* with S3; this ORF may encode Rev. The total size of the message [excluding poly(A)] is 1,449 bases.

tat. The splicing pattern of five of the six cDNAs showed that the *tat* ORF of exon 2 is fused in frame with an ORF of 114 bases in exon 1. The translated sequence of the resulting

TABLE 1. EIAV splice sites[a]

Donor		Acceptor	
CAG	GTAAGA ..... intron 1a	TGAGTATTGTTGCAG	G
	..... intron 1b	TATTGTTGCAGGAAG	C
AAG	GTTTGA ..... intron 2	TTATAATAATGACAG	T
TAG	GTATGT ..... intron 3	ACCTCTTCGCCTAAG	A
Consensus			
CAG	GTAAGT	T XCAG	G
A	G	C_{11} T	A

[a] The consensus eucaryotic splice donor and acceptor sites shown were described by Mount (37) and Shapiro and Senapathy (52). Bases that occur in at least 95% of the known cases are marked with asterisks. EIAV bases which do not match the consensus are underlined. The acceptor following intron 1a reflects the cDNA sequences of clones 1, 3, 5, 6, and 11; the acceptor following intron 1b reflects the cDNA sequence of clone 9.

combined ORF is shown in Fig. 3 and is compared with the sequences of the HIV-1 (1) and simian immunodeficiency virus from macaques (7) Tat proteins. Only within EIAV exon 2 (ORF S1) is there substantial homology with the other Tat proteins. Highly related regions include the EIAV sequence Y-H-C-Q-L-C-F-L-R-S-L-G-I-D-Y, which is found in very similar forms in all lentiviruses except visna virus (55) and the feline immunodeficiency virus (57), and the EIAV sequence R-K-K-N-K-Q-R-L-K-A which, by analogy with the HIV-1 Tat protein, is probably a nuclear localization signal (48). There are also differences between EIAV and the others. For example, the only cysteine residues in EIAV Tat are those shown above, whereas the primate lentivirus Tat proteins have a highly conserved stretch of five more, immediately upstream. In addition, upstream exons in the other lentivirus proteins do not contribute coding sequence to Tat, while in EIAV, exon 1 has considerable potential coding capacity. Unlike the primate viruses, EIAV has no downstream (beyond exon 2) Tat-coding capacity.

To determine whether the sequence shown in Fig. 3 results in a functional Tat protein, we tested the abilities of several of the cDNAs to transactivate the EIAV LTR. Two features of the sequence made this test particularly important. (i) There are no ATG triplets in either exon 1 or 2, and (ii) the sequence of cDNA 9 predicts a different Tat protein. In that clone, the splice acceptor at the 5′ end of exon 2 is 4 bases downstream from that in the other clones. In consequence, the *tat* ORF of exon 2 is fused with a different upstream ORF in exon 1; there are no ATG triplets in that ORF either.

To test the cDNAs for transactivating activity, the entire insert was removed from the library vector and inserted into expression vectors pRSPA-K (sense orientation) and pRSPA-S (antisense orientation). These vectors use the Rous sarcoma virus LTR as a transcriptional initiator and

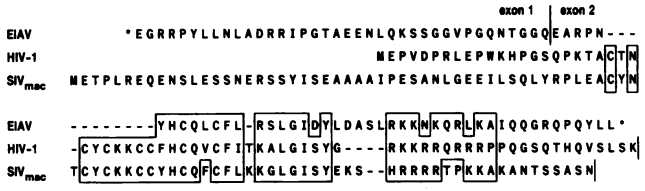

FIG. 3. The translated EIAV *tat* ORF, from the sequence of cDNA 11. The boundary between exons 1 and 2 of EIAV Tat is indicated. The sequence is compared with that of coding exon 1 of the HIV-1 (1) and simian immunodeficiency virus from macaques (7) Tat proteins. The highly conserved cysteine and basic domains are boxed. *, Termination codon.

the simian virus 40 early-region polyadenylation signal (14). In the sense construct, all reading frames upstream of *tat* have termination codons. The constructs were cotransfected into canine (D17) cells with plasmid pEIcat (15), which contains the gene for CAT linked to the EIAV LTR. Without Tat, EIAV LTR-directed CAT activity is very low (13). The result was that both cDNAs 3 and 11 were able to transactivate the EIAV LTR in transient assays. CAT activity was 49- to 75-fold higher in cells that received the cDNA expression plasmid than in those that received the vector alone (Fig. 4). This indicates that cDNAs 3 and 11 do encode Tat and that translation of an active Tat protein does not require an AUG initiation codon.

To determine where translation initiates, the transactivat-

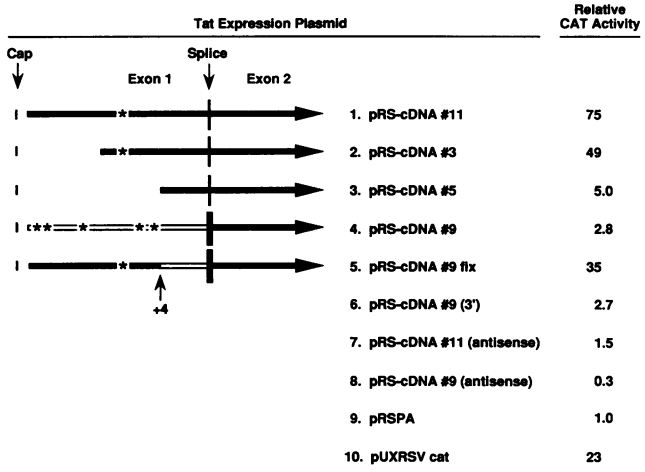

FIG. 4. Transactivation of the EIAV LTR by EIAV cDNAs. For experiments 1 to 9, D17 cells were cotransfected with 5 μg of pEIcat DNA (consisting of the EIAV LTR linked to the gene for CAT [15]) and 1 μg of the indicated Tat expression plasmids. For experiment 10, D17 cells were transfected with 5 μg of pUXRSVcat, which contains the gene for CAT linked to the Rous sarcoma virus LTR (14). The CAT activity in cells transfected with 5 μg of pUXcat (a promoterless CAT expression plasmid]15]) was subtracted from all values. The data are averages of two transfections and are expressed relative to the CAT activity seen with pEIcat plus pRSPA (the expression vector with no insert). The solid lines represent the *tat* ORF of cDNAs 3, 5, and 11. Since the splice acceptor in cDNA 9 is 4 bases downstream of that in the other cDNAs, its exon 1 reading frame is different (open bar). pRS-cDNA 9fix has a four-base insertion (upward-pointing arrow) relative to pRS-cDNA 9. pRS-cDNA 9(3′) has an exchanged cassette with the genomic sequence in exon 4. pRS-cDNA 11 (antisense) and pRS-cDNA 9 (antisense) have the cDNA inserts in the reverse orientation in the vector. *, Termination codon.

ing abilities of cDNAs 5 and 9 were tested. cDNA 5 is shorter than the others and extends in the 5′ direction only to base 190 (cap site, base 1). It was found to be inactive in the transactivation assay. Therefore, initiation must occur upstream of base 190. Since there is a termination codon at positions 134 to 136, initiation must occur between 136 and 190. cDNA 9, in which the *tat* ORF of exon 2 is fused to a different reading frame in exon 1, was also inactive in the transactivation assay, corroborating the importance of the exon 1 sequence.

To test the hypothesis of initiation in the long ORF of cDNA 11 exon 1, we mutagenized cDNA 9 by creating a four-base insertion within exon 1 (cDNA 9fix). In consequence, the exon 2 *tat* ORF was fused with a hybrid upstream ORF; the first 15 amino acids are the same as those in cDNA 11, but the next 23 residues are those of inactive cDNA 9 (Fig. 4). The result of this mutation was to restore transactivating activity to cDNA 9. This strongly suggests that initiation occurs within the first 15 codons (bases 136 to 180) of the cDNA 11 exon 1 ORF. It also shows that large variation in the actual amino acid sequence of exon 1 Tat can be tolerated, since substitution of 23 residues affected activity only slightly (about twofold relative to that of cDNA 11).

In summary, five of six cDNAs showed the same splicing pattern between exons 1 and 2, leading to a predicted structure for EIAV *tat*. The longest two of these five cDNAs were able to transactivate the EIAV LTR in a transient assay. The one cDNA with a different splicing pattern, and hence a different predicted structure, was negative in this assay. There are no ATG triplets in the predicted *tat* sequence. Although the experiments reported here did not establish where Tat is initiated, they do suggest that it is within the first 15 codons of the exon 1 *tat* ORF.

Structure of the cDNAs in the exons 3 and 4. Exon 2 ends immediately following the *tat* termination codon. Exon 3 consists of 101 bases located just downstream from the 5′ end of *env*, and splicing between exons 3 and 4 fuses the *env* reading frame in exon 3 with the ORF S3 of exon 4 to form the putative *rev* gene. If initiation occurs at AUG 1 of exon 3, the resulting protein would contain 30 *env*-derived residues. Remarkably, 17 of these are charged.

The sequence of the cDNAs in exon 4 is consistent with the Rev-defective phenotype in the cells from which the RNA was isolated. In five of the cDNAs, there was a two-base insertion relative to the genomic sequence near the 5′ end of exon 4. This shifts the reading frame from ORF S3 back to *env* (Fig. 5C). In addition, a single-base change converted an *env* UGG codon to UAG, so that termination would occur shortly after the frameshift (predicted size of the protein, about 6 kilodaltons [kDa]). In cDNA 6, there was a four-base deletion relative to the prototype genomic sequence, which also shifts the reading frame back to *env* (Fig. 5D). In this case, translation could proceed to the end of the *env* ORF (predicted size of the protein, about 28 kDa). Thus, the sequences of all of the cDNAs show that a frameshift out of the *rev* ORF occurs soon after the junction between exons 3 and 4. The predicted proteins would contain only one-sixth and one-third, respectively, of the total *rev* ORF sequence and, hence, can be expected to lack Rev activity.

Immunoprecipitation of Rev proteins. These abortive *rev-env*-encoded hybrid proteins should be detectable in the cells from which the original mRNA was isolated. We looked for them in immune precipitation experiments using antisera raised against the *rev*- and *env*-encoded peptides diagrammed in Fig. 5A and B. The 28-kDa frameshift protein

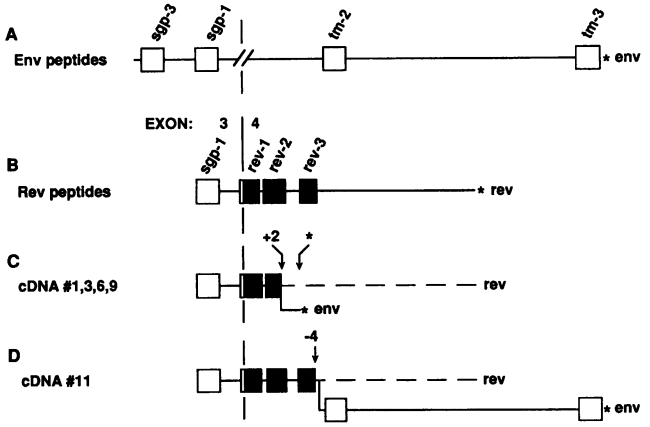

FIG. 5. Translated products predicted by the cDNAs sequences. (A) Diagram of a portion of the wild-type *env* polyprotein. The boxes represent peptides to which antisera were raised. Peptides sgp-3 and sgp-1 begin 7 and 41 residues, respectively, from the N terminus of gp90. Peptides tm-2 and tm-3 are in the long C-terminal extension of the transmembrane protein. *, Termination codon. (B) Diagram of the predicted wild-type *rev* ORF of exon 4 fused in frame with a section of *env* in exon 3. The predicted protein contains the *env* peptide sgp-1, as well as *rev* ORF-encoded peptides (dark boxes) rev-1, rev-2, and rev-3. The rev-1 peptide spans the splice junction. (C) Translation of exons 3 and 4 in cDNAs 1, 3, 5, 6, and 9. A two-base insertion relative to the prototype genomic sequence leads to a frameshift, and a single-base change results in a premature stop codon in the *env* frame. The boxes refer to the peptides defined in diagrams A and B. (D) Translation of exons 3 and 4 in cDNA 11. A four-base deletion relative to the prototype leads to a frameshift into the *env* reading frame. The boxes refer to the peptides defined in diagrams A and B.

should be recognized by six of these sera (anti-sgp-1, anti-Rev-1, anti-Rev-2, anti-Rev-3, anti-tm-2, and anti-tm-3) but not by anti-sgp-3. The 6-kDa frameshift product should be recognized by at least two of the sera (anti-sgp-1 and anti-Rev-1). Immune precipitation of ^{35}S-labeled lysates of clone 6 cells revealed 6- and 28-kDa proteins with exactly these reactivities (Fig. 6A). The 28-kDa protein was also detected in cells stably transfected with cDNA 11 (Fig 6B). Thus, as predicted, cDNA 11 encodes two proteins: Tat (on the basis of the results shown in Fig. 4) and the 28-kDa protein.

These results confirm the splicing pattern revealed by the cDNA sequences. In particular, anti-sgp-1 reactivity shows that a small region of *env* is fused in frame with the *rev* ORF. Since anti-sgp-3 did not recognize the proteins, the *env* region does not include the N terminus of the *env* polyprotein. In addition, the existence of the 6-kDa protein and the reactivity of the 28-kDa protein with anti-tm-2 and anti-tm-3 confirm the frameshifts seen in the cDNAs and provide a plausible explanation for the Rev-defective phenotype of the cells.

The remaining question is whether the cDNAs accurately predict the structure of a wild-type protein. Such a protein would weigh about 20 kDa and would show reactivity with anti-sgp-1 and with all three anti-Rev sera but not with anti-sgp-3 or anti-tm serum (Fig. 5B). A protein with exactly these characteristics was readily detectable in immune precipitates of virus-producing equine fetal kidney cells but was not present in uninfected cells (Fig. 7). We conclude that the splicing pattern seen in the cDNAs, which fuses a 30-amino-acid segment of *env* with the *rev* ORF, is not limited to Rev-defective clone 6 cells but occurs in virus-producing

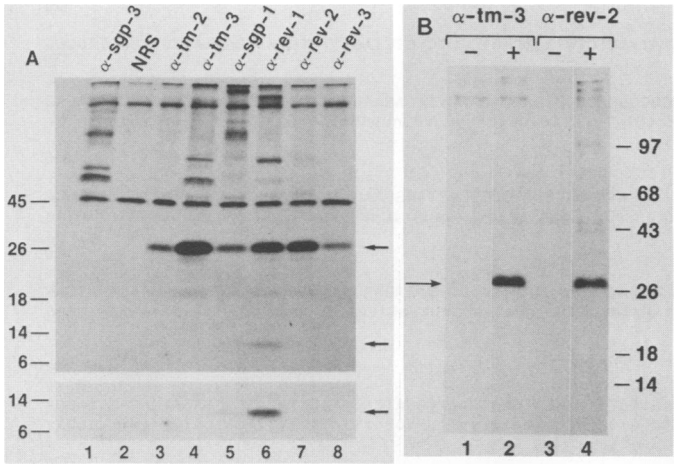

FIG. 6. (A) Immune precipitation of proteins from cell clone 6. Cells were grown in [^{35}S]methionine and [^{35}S]cysteine, and lysates were incubated with the indicated antisera (as defined in Fig. 5A and B). Precipitates were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis on a 10 to 20% gel. The arrows point to the 28- and 6-kDa proteins. A difference between the numbers of methionine-plus-cysteine residues in the 28-kDa protein (total, 11) and the 6-kDa protein (total, 3) may contribute to the difference between their band intensities. The lower gel was exposed for 4 days, and the upper gel was exposed for 16 h. The molecular weights (10^3) of marker proteins are indicated on the left. NRS, Normal rabbit serum. (B) Immune precipitation of proteins from canine cells stably transfected with cDNA 11. Cells were grown in ^{35}S-labeled amino acids, processed as described for panel A, and precipitated with the indicated antisera. Lanes: −, cells transfected with pUC18; +, cells transfected with pRS-cDNA 11. The arrow points to the 28-kDa protein precipitated by both α-rev-2 and α-tm-3. The numbers to the right indicate the molecular weights (10^3) of marker proteins.

cells as well. The predicted sequence of the wild-type cDNA, without the known frameshifting defects found in the clone 6 cells, is shown in Fig. 8.

DISCUSSION

The objective of this study was to determine the structure of EIAV mRNAs that encode the viral regulatory proteins. The source of the mRNA was a cloned cell line derived from a productively infected culture of canine thymus cells. The cloned cells did not produce virus, and most of their EIAV mRNA was of the small size expected for regulatory protein messages. With HIV, such a phenotype is associated with defectiveness of the Rev protein, which, in its active state, promotes cytoplasmic accumulation of unspliced mRNAs (16, 17, 22, 32). Thus, the uniformity of the cDNAs we observed may result from the nature of the source cells, and productively infected cells may yield additional regulatory species. Indeed, the only other EIAV cDNA isolated was derived from virus-producing feline cells and differs from ours in lacking exon 3 (14).

The cDNAs we characterized had four exons, all of which appeared to contain coding sequences. The four exons were (i) the 5′ end of the genome, (ii) the S1 (*tat*) ORF that immediately follows the *pol* gene, (iii) a small segment of *env* beginning 126 bases downstream from the *env* initiation codon, and (iv) the S3 (*rev*) ORF at the 3′ end of the genome. Sequencing of the cDNAs showed that the *tat* ORF of exon 2 extends upstream into exon 1 and that the *rev* ORF of exon 4 extends upstream into the *env* sequence in exon 3. Thus, the cDNAs appear to be capable of encoding two distinct products. We confirmed this by showing that cDNA 11 encodes both Tat activity and a 28-kDa *rev-env* frameshift protein.

The structures of the cDNAs are compatible with the synthesis of both proteins. The combined *tat* ORF lacks an AUG codon and is therefore likely to be translated inefficiently. There is only one AUG in any frame upstream of exon 3, and it is in a poor context for initiation (TxxAUGC) (29, 30). Therefore, scanning ribosomes may be expected to reach exon 3 at rather high frequency. Once they are there, initiation should be efficient, for AUGs 1 and 2 (which are in the *env* reading frame) are in a favorable context (AxxAUGG and GxxAUGA). The scanning model thus predicts inefficient synthesis of Tat and efficient synthesis of the presumptive Rev protein, a prediction that correlates well with our finding that the exon 3 and 4 product was readily detectable in the cloned cells. This difference in efficiency of translation may reflect (i) a difference in the relative levels of the two proteins necessary for function, (ii) a difference in stability, (iii) synthesis of Tat from multiple mRNAs (for example, the exon 1-2-4 mRNA found by Dorn et al. [14]), or (iv) some combination of these possibilities. Alternatively, the lack of a strong initiator for Tat may reflect the attenuated nature of the tissue culture-adapted virus, and Tat encoded by a pathogenic virus may be different. We are testing this possibility. It is also possible that, as in HIV-1 (1, 50), simian immunodeficiency virus from macaques (10), and visna virus (35), a monocistronic EIAV *rev* mRNA will be found.

tat. The strongest evidence that the cDNAs described in this study represent authentic EIAV *Tat* mRNA is that they were able to transactivate the EIAV LTR when expressed in transfected cells. A cDNA isolated by Dorn et al. (14) from virus-producing feline cells had a very similar *tat* structure and also had transactivating activity. Our results complement theirs in at least five ways. (i) The *tat* cDNA found by Dorn et al. lacked an initiator AUG. It was therefore of interest to examine other infected cell types to learn whether

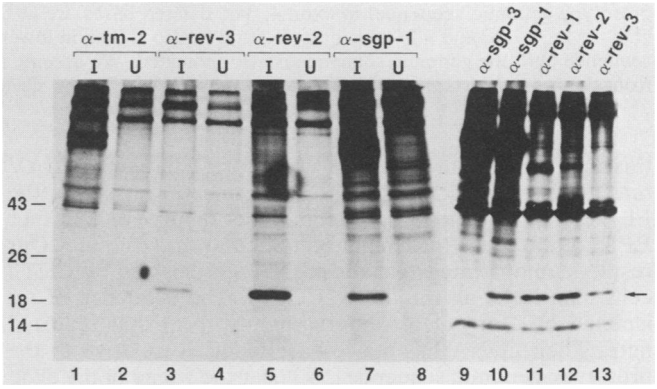

FIG. 7. Immune precipitation of proteins from virus-producing equine cells. Cells were grown in ^{35}S-labeled amino acids, and lysates were immune precipitated with the indicated antisera. The arrow points to the 20-kDa protein precipitated with anti-sgp-1 and with all three anti-Rev sera. Lanes 1 to 8: U, uninfected equine dermal cells; I, EIAV-producing equine fetal kidney cells. To reduce the background, the lysate was boiled in 1% sodium dodecyl sulfate before immune precipitation. Lanes 9 to 13: EIAV-producing equine fetal kidney cells. The lysate was not boiled. The numbers to the left indicate the molecular weights (10^3) of marker proteins.

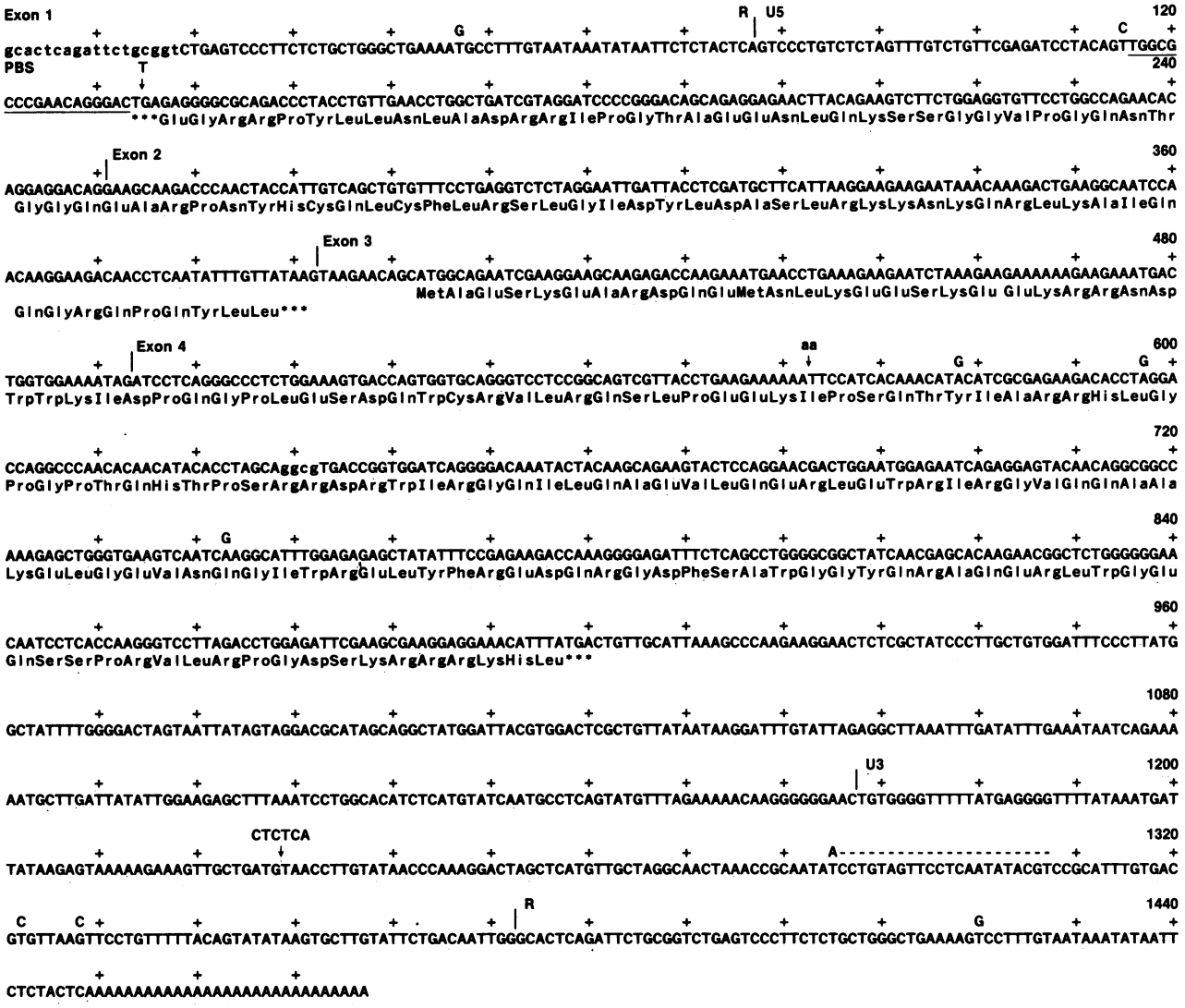

FIG. 8. Sequence of the predicted wild-type cDNA. The sequence is that of cDNA 11 without its four-base deletion (relative to the prototype genomic sequence) in exon 4. The deleted bases are at 629 to 632 and are shown in lowercase. The two-base insertion found in cDNAs 1, 3, 5, 6, and 9 follows base 562 and is also shown in lowercase. Where the cDNA 11 sequence differs from the published genomic sequence (49), the genomic sequence is also indicated. We assume that the RNA cap site is the authentic 5′ end of the mRNA; the 18 bases from the presumed (13) cap site to the 5′ end of cDNA 11 are shown in lowercase. PBS, Primer-binding site; ***, termination triplet.

this unconventional structure was representative of EIAV *tat*. This was particularly important, since in our study the FEA cell line was a poor virus producer relative to Cf2Th or EFK, a difference that might be traced to *tat* structure. Our results demonstrated that *tat* mRNA synthesized by Cf2Th cells also lacks an initiator AUG. (ii) The sequence of *tat* isolated by Dorn et al. differs slightly from that reported here. Their cDNA has a two-base deletion (relative to the prototype genomic sequence [56]) near the 5′ end of the exon 1 ORF. The result is that the first four amino acids of their translated exon 1 ORF differ from the first five amino acids of ours. The fact that both cDNAs have comparable transactivating activities shows that this sequence difference is unimportant. (iii) The effects of various exon 1 deletions on transactivating activity allowed Dorn et al. to conclude that Tat initiation occurs within the first 15 codons of the exon 1 ORF. The context rules of Kozak (29, 30) suggested that the most likely candidate initiator was codon 10 of the exon 1 *tat* ORF, AxxCUGG. Our results also indicated initiation within the first 15 residues. Since the extreme 5′ codons differ in our cDNAs (see ii, above), initiation can be further localized to codons 7 to 15. It should be pointed out that this is the only known example among the lentiviruses of translation initiation at a non-AUG codon. However, there is a growing list of such examples in eucaryotic cells, including murine dihydrofolate reductase (40), human basic fibroblast growth factor (18, 42), human c-*myc* (24), adeno-associated virus capsid protein (4), and Sendai virus C protein (11). These precedents suggest that other lentiviral ORFs without AUG codons should be carefully scrutinized for potential coding capacity. (iv) Dorn et al. showed that exon 1 *tat* sequences in feline cell cDNA could be deleted entirely without loss of function if a translation initiation codon for exon 2 was supplied by the expression vector. Thus, the only apparent role of exon 1 is to supply an initiation codon. Insensitivity to exon 1 structure was shown in a different way in our

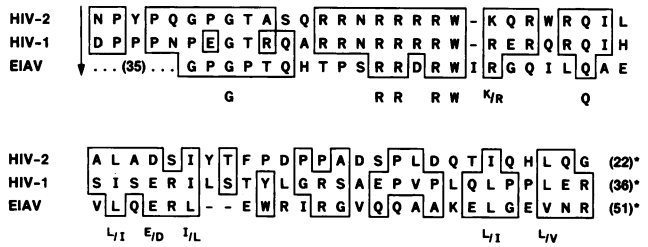

FIG. 9. Sequence homology between the presumptive Rev protein of EIAV and the Rev proteins of HIV-1 (44) and HIV-2 (21). The sequences shown are from coding exon 2 of each protein and begin at the splice site (arrow). The number of residues between the EIAV splice acceptor and the beginning of the sequence above is indicated in parentheses, as is the number of residues following the sequence above in each protein. Identities and conserved residues are boxed. Conservative groupings are P, S, A, T, and G; I, L, V, and M; K, R, and H; D, E, N, and Q; and F, Y, and W. *, Termination codon.

experiments. We found that much of exon 1 could be substituted with an unrelated sequence without loss of function. Thus, it appears likely that Tat exon 2 constitutes a structurally and functionally distinct domain(s) that is unaffected by the presence or absence of the exon 1 sequence. (v) The fact that the cDNA of Dorn et al. differs in exon structure from ours (exons 1-2-4 versus 1-2-3-4) shows that there are at least two Tat-encoding mRNAs in EIAV-infected cells. Both are able to express functional Tat, as measured by the CAT assay.

The putative Rev protein. We have no direct evidence that the protein encoded by exons 3 and 4 is the homolog of HIV Rev. However, there are several pieces of indirect evidence. (i) The exon 4 ORF and HIV Rev are derived from the same region of their respective genomes, overlapping the 3′ end of the *env* gene. (ii) Limited but significant sequence homology between the translated EIAV exon 4 and Rev of HIV-1 and HIV-2 can be seen (Fig. 9). Over a stretch of 41 EIAV residues, 27 are related to those of HIV-1 and/or HIV-2. Twelve residues are very highly conserved among all three viruses, including five from the basic domain presumed to direct nuclear localization. (iii) The cells from which the cDNA library was derived exhibited a Rev-defective phenotype. This correlates with the frameshifting mutations within exon 4 that were observed in all of the cDNAs. It is therefore our working hypothesis that the EIAV exon 3 and 4 product will show Rev activity similar to that of HIV. We are attempting to demonstrate this activity directly.

However, there is reason to be cautious about this interpretation. The presumptive EIAV Rev and HIV-1 Rev differ in the origins of their upstream coding exons. In HIV, Rev exon 1 is between *pol* and *env* and overlaps exon 1 of Tat. In both EIAV and visna virus (12, 20, 35), exon 1 of the putative Rev protein is derived from the *env* gene. A newly described HIV-1 protein demonstrates that coding exon 1 of HIV-1 Rev can be varied somewhat and still retain function. Tev cDNA contains (i) exon 1 of *tat*, (ii) 114 bases of *env*, and (iii) exon 2 of *rev*. An ORF extends throughout these three regions, and a tripartite protein with both Tat and Rev activities was detected in infected cells (5). Thus, segments of Tat plus *env* can substitute for the normal exon 1 of Rev. An additional protein found in HIV-infected cells demonstrates that there are limits on the variability of Rev exon 1. $p18^{Drev}$ initiates within the same *env* segment found in Tev and also contains exon 2 of *rev*, a structure much like that of the predicted EIAV and visna virus Rev proteins. However, $p18^{Drev}$ did not show Rev activity (5).

It is worth emphasizing that although many common features link the members of the lentivirus group, the individual viruses exhibit unique characteristics. For example, all lentiviruses encode a transactivating protein, yet that of visna virus appears to be different from that of the primate viruses and EIAV (14). Similarly, all of the *tat* mRNAs are produced by multiple splicing events, yet the Tat protein of EIAV is unique in being encoded by exons 1 and 2 and in initiating without an AUG codon. A *rev* ORF is identifiable in all lentiviruses, but the upstream exon differs: in the primate viruses, it overlaps the *tat* ORF, while in EIAV and visna virus (12, 20) it is derived from *env*. Thus, one must be cautious in extrapolating from one lentivirus to another. More positively, knowledge of the range of strategies and mechanisms found in the group as a whole may be helpful in unraveling particular pathways in individual viruses.

ACKNOWLEDGMENTS

We are grateful to Terry Copeland for synthesis of peptide rev-3, to Beth Cassell and Pat Wesdock for antiserum production, to Tom Wood for pNeoHiC, and to Marilyn Powers for synthesis of oligonucleotides.

This research was sponsored in part by the National Cancer Institute under contract NO1-CO-74101 with ABL.

LITERATURE CITED

1. **Arya, S. K., C. Guo, S. F. Josephs, and F. Wong-Staal.** 1985. Transactivator gene of human T-lymphotropic virus type III (HTLV-III). Science **229**:69–73.
2. **Aviv, H., and P. Leder.** Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. Proc. Natl. Acad. Sci. USA **69**:1408–1412.
3. **Ball, J. M., S. L. Payne, C. J. Issel, and R. C. Montelaro.** 1988. EIAV genomic organization: further characterization by sequencing of purified glycoproteins and cDNA. Virology **165**: 601–605.
4. **Becerra, S. P., J. A. Rose, M. Hardy, B. M. Baroudy, and C. W. Anderson.** 1985. Direct mapping of adeno-associated virus capsid proteins B and C: a possible ACG initiation codon. Proc. Natl. Acad. Sci. USA **82**:7919–7923.
5. **Benko, D. M., S. Schwartz, G. N. Pavlakis, and B. K. Felber.** 1990. A novel human immunodeficiency virus type 1 protein, *tev*, shares sequences with *tat*, *env*, and *rev* proteins. J. Virol. **64**:2505–2518.
6. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–1523.
7. **Chakrabarti, L., M. Guyader, M. Alizon, M. D. Daniel, R. C. Desrosiers, P. Tiollais, and P. Sonigo.** 1987. Sequence of simian immunodeficiency virus from macaques and its relationship to other human and simian retroviruses. Nature (London) **328**: 543–547.
8. **Chirgwin, J. M., A. E. Przybyla, R. J. MacDonald, and W. J. Rutter.** 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry **18**:5294–5299.
9. **Chiu, I., A. Yaniv, J. F. Dahlberg, A. Gazit, S. F. Skuntz, S. R. Tronick, and S. A. Aaronson.** 1985. Nucleotide sequence evidence for relationship of AIDS retrovirus to lentiviruses. Nature (London) **317**:366–368.
10. **Colombini, S., S. K. Arya, M. S. Reitz, L. Jagodzinski, B. Beaver, and F. Wong-Staal.** 1989. Structure of simian immunodeficiency virus regulatory genes. Proc. Natl. Acad. Sci. USA **86**:4813–4817.
11. **Curran, J., and D. Kolakofsky.** 1988. Ribosomal initiation from an ACG codon in the Sendai virus P/C mRNA. EMBO J. **7**:245–251.
12. **Davis, J. L., and J. E. Clements.** 1989. Characterization of a cDNA clone encoding the visna virus transactivating protein.

Proc. Natl. Acad. Sci. USA **86**:414–418.
13. **Derse, D., P. L. Dorn, L. Levy, R. M. Stephens, N. R. Rice, and J. W. Casey.** 1987. Characterization of equine infectious anemia virus long terminal repeat. J. Virol. **61**:743–747.
14. **Dorn, P. L., L. DaSilva, L. Martarano, and D. Derse.** 1990. Equine infectious anemia virus *tat*: insights into the structure, function, and evolution of lentivirus *trans*-activator proteins. J. Virol. **64**:1616–1624.
15. **Dorn, P. L., and D. Derse.** 1988. *cis*- and *trans*-acting regulation of gene expression of equine infectious anemia virus. J. Virol. **62**:3522–3526.
16. **Feinberg, M. B., R. F. Jarrett, A. Aldovini, R. C. Gallo, and F. Wong-Staal.** 1986. HTLV-III expression and production involve complex regulation at the level of splicing and translation of viral RNA. Cell **46**:807–817.
17. **Felber, B. K., M. Hadzopoulou-Cladaras, C. Cladaras, T. Copeland, and G. N. Pavlakis.** 1989. The rev protein of HIV-1 affects the stability and transport of the viral mRNA. Proc. Natl. Acad. Sci. USA **86**:1495–1499.
18. **Florkiewicz, R. Z., and A. Sommer.** 1989. Human basic fibroblast growth factor gene encodes four polypeptides: three initiate translation from non-AUG codons. Proc. Natl. Acad. Sci. USA **86**:3978–3981.
19. **Gonda, M. A., H. P. Charman, J. L. Walker, and L. Coggins.** 1978. Scanning and transmission electron microscopic study of equine infectious anemia virus. Am. J. Vet. Res. **39**:731–740.
20. **Gourdou, I., V. Mazarin, G. Querat, N. Sauze, and R. Vigne.** 1989. The open reading frame S of visna virus genome is a trans-activating gene. Virology **171**:170–178.
21. **Guyader, M., M. Emerman, P. Sonigo, F. Clavel, L. Montagnier, and M. Alizon.** 1987. Genome organization and transactivation of the human immunodeficiency virus type 2. Nature (London) **326**:662–669.
22. **Hadzopoulou-Cladaras, M., B. K. Felber, C. Cladaras, A. Athanassopoulos, A. Tse, and G. Pavlakis.** 1989. The *rev* (*trs/art*) protein of human immunodeficiency virus type 1 affects viral mRNA and protein expression via a *cis*-acting sequence in the *env* region. J. Virol. **63**:1265–1274.
23. **Hammarskjold, M.-L., J. Heimer, B. Hammarskjold, I. Sangwan, L. Albert, and D. Rekosh.** 1989. Regulation of human immunodeficiency virus *env* expression by the *rev* gene product. J. Virol. **63**:1959–1966.
24. **Hann, S. R., M. W. King, D. L. Bentley, C. W. Anderson, and R. N. Eisenman.** 1988. A non-AUG translational initiation in c-myc exon 1 generates an N-terminally distinct protein whose synthesis is disrupted in Burkitt's lymphomas. Cell **52**:185–195.
25. **Hauber, J., A. Perkins, E. P. Heimer, and B. R. Cullen.** 1987. Trans-activation of human immunodeficiency virus gene expression is mediated by nuclear events. Proc. Natl. Acad. Sci. USA **84**:6364–6468.
26. **Kao, S.-Y., A. F. Calman, P. A. Luciw, and B. M. Peterlin.** 1987. Anti-termination of transcription within the long terminal repeat of HIV-1 by *tat* gene product. Nature (London) **330**:489–493.
27. **Kawakami, T., L. Sherman, J. Dahlberg, A. Gazit, A. Yaniv, S. R. Tronick, and S. A. Aaronson.** 1987. Nucleotide sequence analysis of equine infectious anemia virus proviral DNA. Virology **158**:300–312.
28. **Kobayashi, K., and Y. Kono.** 1967. Serial passages of equine infectious anemia virus in horse leucocyte cultures. Natl. Inst. Anim. Health Q. **7**:1–7.
29. **Kozak, M.** 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. Cell **44**:283–292.
30. **Kozak, M.** 1987. An analysis of the 5'-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res. **15**:8125–8148.
31. **Malim, M. H., S. Bohnlein, J. Hauber, and B. R. Cullen.** 1989. Functional dissection of the HIV-1 Rev trans-activator—derivation of a trans-dominant repressor of Rev function. Cell **58**:205–214.
32. **Malim, M. H., J. Hauber, R. Fenrick, and B. R. Cullen.** 1988. Immunodeficiency virus *rev* trans-activator modulates the expression of the viral regulatory genes. Nature (London) **335**:181–183.
33. **Malim, M. H., J. Hauber, S. Le, J. V. Maizel, and B. R. Cullen.** 1989. The HIV-1 rev transactivator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. Nature (London) **338**:254–257.
34. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
35. **Mazarin, V., I. Gourdou, G. Querat, N. Sauze, and R. Vigne.** 1988. Genetic structure and function of an early transcript of visna virus. J. Virol. **62**:4813–4818.
36. **McGuire, T. C., T. B. Crawford, and J. B. Henson.** 1971. Immunofluorescent localization of equine infectious anemia virus in tissue. Am. J. Pathol. **62**:283–294.
37. **Mount, S. M.** 1982. A catalogue of splice junction sequences. Nucleic Acids Res. **10**:459–472.
38. **Muesing, M. A., D. H. Smith, C. D. Cabradilla, C. V. Benton, L. A. Lasky, and D. J. Capon.** 1985. Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovirus. Nature (London) **313**:450–458.
39. **Muesing, M. A., D. H. Smith, and D. J. Capon.** 1987. Regulation of mRNA accumulation by a human immunodeficiency virus trans-activator protein. Cell **48**:691–701.
40. **Peabody, D. S.** 1989. Translation initiation at non-AUG tripets in mammalian cells. J. Biol. Chem. **264**:5031–5035.
41. **Peterlin, B. M., P. A. Luciw, P. J. Barr, and M. D. Walker.** 1986. Elevated levels of mRNA can account for the transactivation of human immunodeficiency virus. Proc. Natl. Acad. Sci. USA **83**:9734–9738.
42. **Prats, H., M. Kaghad, A. C. Prats, M. Klagsbrun, J. M. Lelias, P. Liauzun, P. Chalon, J. P. Tauber, F. Amalric, J. A. Smith, and D. Caput.** 1989. High molecular mass forms of basic fibroblast growth factor are initiated by alternative CUG codons. Proc. Natl. Acad. Sci. USA **86**:1836–1840.
43. **Rasty, S., B. R. Dhruva, R. L. Schiltz, D. S. Shih, C. J. Issel, and R. C. Montelaro.** 1990. Proviral DNA integration and transcriptional patterns of equine infectious anemia virus during persistent and cytopathic infections. J. Virol. **64**:86–95.
44. **Ratner, L., W. Haseltine, R. Patarca, J. Livak, B. Starcich, S. F. Josephs, E. R. Doran, and J. A. Rafalski.** 1985. Complete nucleotide sequence of the AIDS virus, HTLV-III. Nature (London) **313**:277–283.
45. **Rice, N. R., T. D. Copeland, S. Simek, S. Oroszlan, and R. V. Gilden.** 1986. Detection and characterization of the protein encoded by the v-rel oncogene. Virology **149**:217–229.
45a. **Rice, N. R., L. E. Henderson, R. C. Sowder, T. D. Copeland, S. Oroszlan, and J. F. Edwards.** 1990. Synthesis and processing of the transmembrane envelope protein of equine infectious anemia virus. J. Virol. **64**:3770–3778.
46. **Rimsky, L., J. Hauber, M. Dukovich, M. H. Malim, A. Langlois, B. R. Cullen, and W. C. Greene.** 1988. Functional replacement of the HIV-1 *rev* protein by the HTLV-I *rex* protein. Nature (London) **335**:738–740.
47. **Rosen, C. A., J. G. Sodroski, and W. A. Haseltine.** 1985. The location of cis-acting regulatory sequences in the human T cell lymphotropic virus type III (HTLV-III/LAV) long terminal repeat. Cell **41**:813–823.
48. **Ruben, S., A. Perkins, R. Purcell, K. Joung, R. Sia, R. Burghoff, W. A. Haseltine, and C. A. Rosen.** 1989. Structural and functional characterization of human immunodeficiency virus *tat* protein. J. Virol. **63**:1–8.
49. **Rushlow, K., K. Olsen, G. Stiegler, S. L. Payne, R. C. Montelaro, and C. Issel.** 1986. Lentivirus genomic organization: the complete nucleotide sequence of the *env* gene region of equine infectious anemia virus. Virology **155**:309–321.
50. **Sadaie, M. R., J. Rappaport, T. Benter, S. F. Josephs, R. Willis, and F. Wong-Staal.** 1988. Missense mutations in an infectious human immunodeficiency viral genome: functional mapping of tat and identification of the rev splice acceptor. Proc. Natl. Acad. Sci. USA **85**:9224–9228.
51. **Sanger, O., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.

52. **Shapiro, M. B., and P. Senapathy.** 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. Nucleic Acids Res. **15**:7155–7174.
53. **Sherman, L., A. Gazit, A. Yaniv, T. Kawakami, J. E. Dahlberg, and S. R. Tronick.** 1988. Localization of sequences responsible for *trans*-activation of the equine infectious anemia virus long terminal repeat. J. Virol. **62**:120–126.
54. **Sodroski, J., C. A. Rosen, F. Wong-Staal, S. Z. Salahuddin, M. Popovic, S. Arya, R. C. Gallo, and W. A. Haseltine.** 1985. Trans-acting transcriptional regulation of human T-cell leukemia virus type III long terminal repeat. Science **227**:171–173.
55. **Sonigo, P., M. Alizon, K. Staskus, D. Klatzmann, S. Cole, O. Danos, E. Retzel, P. Tiollais, A. Haase, and S. Wain-Hobson.** 1985. Nucleotide sequence of the visna lentivirus: relationship to the AIDS virus. Cell **42**:369–382.
56. **Stephens, R. M., J. W. Casey, and N. R. Rice.** 1986. Equine infectious anemia virus *gag* and *pol* genes: relatedness to visna and to the AIDS virus. Science **231**:589–594.
57. **Talbott, R. L., E. E. Sparger, K. M. Lovelace, W. M. Fitch, N. C. Pederson, P. A. Luciw, and J. H. Elder.** 1989. Nucleotide sequence and genomic organization of feline immunodeficiency virus. Proc. Natl. Acad. Sci. USA **86**:5743–5747.
58. **Wright, C. M., B. K. Felber, H. Paskalis, and G. N. Pavlakis.** 1986. Expression and characterization of the transactivator of HTLV-III/LAV virus. Science **234**:988–992.