Equine Infectious Anemia Virus *tat*: Insights into the Structure, Function, and Evolution of Lentivirus *trans*-Activator Proteins

PATRICIA DORN,^{1†} LUIS DASILVA,² LUIS MARTARANO,¹ AND DAVID DERSE^{3*}

Biological Carcinogenesis and Development Program, Program Resources Inc., Frederick Cancer Research Facility, Frederick, Maryland 21701¹; Instituto Nacional do Cancer, Rio de Janeiro, Brazil²; and Laboratory of Viral Carcinogenesis, National Cancer Institute, Frederick, Maryland 21701-1013³

Received 29 September 1989/Accepted 5 December 1989

Equine infectious anemia virus (EIAV) contains a *tat* gene which is closely related to the *trans*-activator genes of the human and simian immunodeficiency viruses. Nucleotide sequence analysis of EIAV cDNA clones revealed that the *tat* mRNA is composed of three exons; the first two encode Tat and the third may encode a Rev protein. Interestingly, EIAV Tat translation is initiated at a non-AUG codon in exon 1 of the mRNA, perhaps allowing an additional level of gene regulation. The deduced amino acid sequence of EIAV *tat*, combined with functional analyses of *tat* cDNAs in transfected cells, has provided some unique insights into the domain structure of Tat. EIAV Tat has a C-terminal basic domain and a highly conserved 16-amino-acid core domain, but not the cysteine-rich region, that are present in the primate immunodeficiency virus Tat proteins. Thus, EIAV encodes a relatively simple version of this kind of *trans* activator.

Equine infectious anemia virus (EIAV) is the etiologic agent of a disease of horses that is characterized by severe anemia, fever, and viremia in a periodically cycling manner (5). The virus infects cells of the monocyte/macrophage lineage and persists for the life of the animal; stress or immunosuppression may precipitate a resumption of clinical disease (25, 29). EIAV is a retrovirus of the lentivirus subgroup whose other family members include the human and simian immunodeficiency viruses (HIV and SIV), visna virus, caprine arthritis-encephalitis virus, and the recently described lentiviruses feline immunodeficiency virus (FIV) (35) and bovine immunodeficiency virus (15). Most of these viruses are known to encode several genes in addition to those required for replication and virion assembly; in some of these viruses, these gene products function as regulators of virus expression.

The control of gene expression has been extensively examined for the primate immunodeficiency viruses and visna virus. These agents appear to control their expression through a complex interaction of viral regulatory proteins and *cis*-acting response elements. The first of several regulatory proteins to be described for these viruses was the HIV transcriptional activator, Tat, which acts in concert with sequence elements in the long terminal repeats (LTRs) to increase the abundance of viral transcripts (2, 40, 43, 44). The HIV and SIV Tat proteins share several highly conserved amino acid domains and act on Tat-response (TAR) elements located downstream of the RNA start site (36, 45). In contrast, the sequence of the visna virus Tat protein is quite different from those of the primate virus Tat proteins and appears to act on sequence elements located in a region upstream from the transcriptional promoter (8, 22). It is apparent that the various lentiviruses have evolved distinct strategies, incorporating discrete *cis*- and *trans*-acting components, for the regulation of their expression.

We previously suggested that EIAV gene expression is controlled in a manner similar to the primate lentiviruses,

since the EIAV TAR elements were present in the region immediately downstream of the RNA start site (9, 11). On the basis of the mechanistic similarities, we hypothesized that EIAV might encode a Tat protein related to the HIV and SIV proteins. A putative *tat* gene has been mapped to the central region of the EIAV genome both by deletion analyses of the provirus (42) and by cloning subgenomic fragments into mammalian expression vectors (11). Because there were several open reading frames (ORFs) and potential splice sites within the region tested, an unambiguous definition of the essential *tat* exon was not possible. In the work described in the present communication, we examined the sequence and expression of the EIAV *tat* gene and analyzed the structure and activity of the Tat protein in transfected cells. The predicted EIAV Tat protein has an amino acid sequence and domain structure closely related to the HIV and SIV Tat proteins. The sequence comparisons, when combined with functional analyses, helped to delineate domains critical for Tat function. A novel aspect of EIAV *tat* expression is the use of a non-AUG codon for translation initiation, thus demonstrating an additional strategy for fine tuning gene expression in the lentiviruses.

MATERIALS AND METHODS

cDNA cloning. cDNA was synthesized by using 5 µg of poly(A)⁺ RNA from EIAV-infected FEA cells, an oligo (dT)₁₂₋₁₈ primer, and reagents and conditions recommended by the supplier (Bethesda Research Laboratories, Inc.). The ends of the double-stranded DNA were modified by the addition of *Eco*RI linkers and ligated to *Eco*RI-digested arms of the bacteriophage lambda ZAP vector (Stratagene Inc.). Approximately 40,000 plaques were recovered, and 50 of these that hybridized to the EIAV LTR were further purified by two rounds of plating and screening. The plasmid portions of the recombinant phage (containing the cDNA insert) were excised by superinfection with helper phage, R408. Plasmids were subsequently analyzed by restriction mapping, Southern blot hybridization to subgenomic probes representing the LTR and the central region of the virus genome, and dideoxy-nucleotide sequencing (20, 39).

Plasmid construction. pElcat contains the EIAV LTR

* Corresponding author.

† Present address: Department of Microbiology and Immunology, Stanford University, Stanford, CA 94305.

cloned into pUXcat (11). pC15cat contains an HIV LTR fragment cloned into pSV0cat (40). pUXSV2cat and pUXRSVcat respectively contain the simian virus 40 (SV40) early promoter from pSV2cat and the Rous sarcoma virus LTR from pRSVcat cloned into pUXcat. Fragments from the central region of the EIAV genome were cloned into the expression plasmid, pS, containing the SV40 early promoter and polyadenylation signals at the extreme ends of the multiple cloning site of the Bluescribe vector (11). pS1,2 contains the 800-base-pair (bp) *NcoI-SmaI* fragment (genome positions 4870 to 5670) cloned into the *HincII* site of pS. pS1 contains the 448-bp *NcoI-BamHI* fragment from this region in pS. pS2 contains the 530-bp *PvuII-SmaI* fragment in the *HincII* site of pS. pS1,2δRsa contains the 608-bp *RsaI-SmaI* fragment in pS. Translation termination codons were introduced into ORF S1 or S2 of plasmid pS1,2 to form plasmids pS1,2 and pS1,2δ, respectively. Single nucleotides within the *PvuII* and *BamHI* restriction sites were changed to create stop codons by oligonucleotide-directed mutagenesis of single-stranded DNA. Altered plasmids were identified by the loss of the appropriate restriction site. A slightly different vector system was used to express HIV Tat and EIAV cDNA fragments. pRSPA-K and pRSPA-S contain the Rous sarcoma virus promoter and SV40 polyadenylation sequences cloned into the distal ends of the multiple cloning site of the Bluescript-KS⁺ vector (Stratagene); the two vectors differ in the orientation of the transcription signals with respect to the vector cloning sites. pRS-Htat contains the HIV *tat* gene from the HXB2 provirus clone (41); a 343-bp *SalI-Sau96-1* fragment (genome positions 5785 to 6128) was inserted into the *SalI-to-EcoRV* sites of pRSPA-K. The inserts from EIAV *tat* cDNA clones 442 and 492 were released as *XhoI-XbaI* fragments and cloned directly into the same sites of pRSPA-S to give pRS442 and pRS492, respectively. A 279-bp *HaeII-ApaI* fragment from cDNA 492 was cloned, after T4 DNA polymerase digestion of the *HaeII* end, into the *SmaI-to-ApaI* sites of pRSPA-S to give pRS-Etat-HA. pRS-Etat-HAX was made by ligating the oligonucleotide CTAGTCTAGACTAG to the *SmaI* site, within the *tat* gene, of pRS-Etat-HA. pRS-Etat-S contains the 219-bp *SmaI-ApaI* fragment of cDNA 492 cloned into the same sites of pRSPA-S. pRSPA-S was modified by the insertion of the oligonucleotide CCACCATGG into the *EcoRV* site to give pRSPA-Met; digestion of this plasmid with *NcoI* and end filling with DNA polymerase provides an efficient methionine initiation codon for inserted genes. The *SmaI-ApaI* fragment from pRS-Etat-S was ligated to the polymerase-filled *NcoI-* and *ApaI*-digested sites of pRSPA-Met to give pRS-Etat-M, thus placing the *tat* coding sequences in frame with the methionine initiation codon. A 177-bp fragment was released from cDNA 492 by digestion with *BstNI* and mung bean nuclease followed by *ApaI*; it was then ligated to pRSPA-S (*SmaI* to *ApaI*) and pRSPA-Met (*NcoI* to *ApaI*) to give pRS-Etat-B and pRS-Etat-MB, respectively. The plasmid pRS-Etat-MP was made in a similar way by insertion of the 126-bp *PvuII-ApaI* fragment from cDNA 492 into pRSPA-Met. The plasmids pRS-Etat-M23, pRS-Etat-M24, and pRS-Etat-M25 were constructed by exonuclease III and mung bean nuclease digestion of the 5' end of cDNA 492 followed by *ApaI* digestion and ligation to pRSPA-Met plasmids. All plasmids were banded twice by cesium chloride gradient centrifugation and verified by nucleotide sequence analysis (20, 39).

Cells and transfections. EIAV-infected, feline FEA (E-FEA), canine osteogenic sarcoma (D17), African green monkey kidney (COS), and human (HeLa) cells were main-

tained in Dulbecco modified Eagle medium supplemented with 10% fetal calf serum (GIBCO Laboratories). For chloramphenicol acetyltransferase (CAT) assays, 3×10^5 cells in 5 ml of medium were plated in 3.5-cm-diameter wells of six-well cluster dishes the day before transfection. Plasmid DNAs were transfected as calcium phosphate coprecipitates as previously described (11). After 4 h for D17 and COS or 12 h for HeLa, the cells were washed with phosphate-buffered saline and fresh medium was added. Two days after transfection, cells were harvested by scraping, suspended in 100 μ l of 0.1 M Tris hydrochloride (pH 7.8), and lysed by three cycles of freezing and thawing. The resulting supernatants were then assayed for CAT activity.

CAT assays. CAT activity was assayed by the solvent partition method of Neumann et al. (31). A 200- μ l reaction mixture containing 100 mM Tris hydrochloride (pH 7.8), 1.0 mM chloramphenicol, 0.1 μ Ci of [¹⁴C]acetyl coenzyme A (4 mCi/mmol) (Du Pont, NEN Research Products), and cell extract was overlaid with 3 ml of scintillation fluid (Econofluor; Du Pont, NEN). Reactions were performed at ambient temperature. Production of the radioactively labeled acetyl-chloramphenicol was monitored by counting in a liquid scintillation counter at hourly intervals.

RNA purification and Northern (RNA) blot analysis. Total RNA from E-FEA cells was prepared by lysis in guanidine thiocyanate and ultracentrifugation through a cesium chloride cushion (6). Poly(A)⁺ RNA was prepared by oligo(dT)-cellulose column chromatography, and a 2.5- μ g portion was fractionated on a 1% agarose-0.66 M formaldehyde gel and transferred to a nylon membrane for hybridization. Membranes were hybridized to a ³²P-labeled EIAV LTR DNA fragment as previously described (11).

RESULTS

Identification of the essential *tat* exon. As a first step in identifying the EIAV *tat* gene, genomic fragments derived from the central region of the virus genome were expressed in mammalian cells (11). EIAV *tat* coding sequences were thus mapped to an 800-bp fragment that contained the 3' end of the *pol* gene, two small ORFs designated S1 and S2, and the 5' end of the *env* gene (Fig. 1). EIAV LTR-directed transcription was activated in cells transfected with the expression plasmid pS1,2, which contains this 800-bp *NcoI-SmaI* fragment controlled by the SV40 early promoter (11). To precisely identify the ORF required for the expression of Tat activity, we have introduced several deletions and point mutations into pS1,2 (Fig. 1). The resulting plasmids were then transfected into canine D17 cells with pElcat, a plasmid containing the bacterial *cat* gene controlled by the EIAV LTR. CAT activity expressed in the transfected cells reflects Tat activation of the EIAV LTR sequences (Table 1).

Levels of CAT activity directed by pElcat were approximately 10 times higher in cells cotransfected with pS1,2 than with pS, the noncoding expression vector (Table 1). Plasmid pS1 is deleted of sequences 3' of the *BamHI* site; it contains all of ORF S1 and only the first 18 codons of ORF S2 (Fig. 1). pS1 expressed Tat activity in transfected cells, yielding levels of CAT activity higher than produced by pS1,2 (Table 1). We previously reported that a pS1 plasmid did not express Tat activity (11); we later discovered that this erroneous result was due to secondary alterations in that plasmid. Deletion of sequences upstream of the *PvuII* site gave rise to pS2, which contains all 66 codons of ORF S2 but is lacking the first 12 codons of ORF S1. pS2 did not express Tat activity in the transfected cells, suggesting that ORF S1

FIG. 1. Structures of expression plasmids containing subgenomic fragments from the central region of the EIAV genome. At the top is shown the 800-bp *NcoI*-*SmaI* fragment containing the 3' end of *pol*, ORFs S1 and S2, and the 5' end of *env*. Numbers refer to nucleotide positions in the genome. Translation initiation codons (AUG) and restriction sites are indicated. This 800-bp fragment was inserted into the expression plasmid, pS, to give pS1,2, from which the other plasmids were derived. Termination codons were introduced into ORFs S1 and S2, altering the *PvuII* and *BamHI* sites, by site-directed mutagenesis.

but not ORF S2 is required for Tat expression. In a complementary approach, translation stop codons were introduced into ORF S1 or S2 by site-directed mutagenesis, consequently altering the *PvuII* or *BamHI* site, respectively. Plasmid pS1,2, which has a stop codon in S1, did not express Tat activity, whereas plasmid pS1,σ2, which has a stop codon in ORF S2, retained Tat activity. These experiments revealed that ORF S1 but not S2 is an essential coding exon of the *tat* gene.

It was not clear how translation of ORF S1 was initiated, since it lacks an AUG codon. To determine whether an AUG from the *pol* region was being joined to S1 by splicing, we removed sequences upstream of the *RsaI* site, yielding the plasmid pS1,2Δ*Rsa*. Although pS1,2Δ*Rsa* does not possess any AUG codons preceding S1, it still expressed Tat activity

TABLE 1. *trans* activation of the EIAV LTR by subgenomic expression plasmids

CAT plasmid	Tat expression plasmid	Relative CAT activity ^a
pElcat	pS	0.29
pElcat	pS1,2	3.4
pElcat	pS1	7.5
pElcat	pS2	0.30
pElcat	pS1,2	1.0
pElcat	pS1,σ2	4.1
pElcat	pS1,2Δ <i>Rsa</i>	3.3

^a Data are the mean of 4 to 12 separate transfections. Background CAT activity obtained with a promoterless plasmid, pUXcat, was subtracted from all values, which are expressed relative to the values obtained with pUXSV2cat, which contains the SV40 early promoter. pElcat (3 μg) plus 3 μg of expression plasmid were transfected onto D17 cells as calcium phosphate precipitates. Two days later CAT activities were assayed by the solvent partition method (31).

(Table 1). It should be noted that the genomic fragment in pS1,2Δ*Rsa*, as well as the other pS1,2 plasmids, contain two UUG codons preceding ORF S1, which may act as alternative start codons. Thus, expression of ORF S1, even in the absence of an AUG initiation codon, results in EIAV-specific transactivation.

Characterization of *tat* cDNA clones. The transcription pattern of EIAV is depicted in the Northern blot analysis of poly(A)⁺ RNA prepared from EIAV-infected FEA cells, E-FEA (Fig. 2A). The blot was hybridized with an LTR probe which revealed the approximately 8-kilobase (kb) genomic mRNA, an approximately 4-kb singly spliced *env* transcript, and what appears to be a population of small RNAs of approximately 1.5 kb. As in the other lentiviruses, these small RNAs probably represent small, multiply spliced transcripts encoding regulatory proteins.

To identify and characterize the *tat* mRNA, we constructed a cDNA library from poly(A)⁺ RNA from the E-FEA cell line. It was known from earlier transfection experiments that this cell line produces significant Tat activity (9, 11). Of approximately 40,000 recombinant plaques, 5 clones, whose inserts ranged in size from 0.4 to 4.0 kb and showed positive hybridization to a probe containing ORFs S1 and S2, were examined by nucleotide sequencing. Of these five clones, two appeared to represent truncated copies of unspliced RNAs and the other three, designated 492, 272, and 442, were examined in detail. Their basic features are summarized in Fig. 2B. Clone 492 appears to represent a nearly complete copy of a multiply spliced RNA, since 5' and 3' LTR sequences were present at the termini. Clone 272 is identical to 492 except for the addition of some unknown sequences at the extreme 5' end of the cDNA. Clone 442 differs from the other clones in several respects; first, cDNA synthesis, primed by oligo(dT), was initiated within an adenosine-rich region immediately downstream of ORF S2. Second, the splicing pattern at the 3' end of clone 442 differs from the other clones; the splice donor following ORF S1 was joined to a splice acceptor within ORF S2. Thus, clone 442 lacks the *env* and ORF S3 sequences present in the other clones.

The nucleotide sequence of *tat* cDNA clone 492 (Fig. 3) (GenBank accession no. M30138) revealed that it is 1,325 bp long and begins 14 bases downstream of the previously determined RNA start site (9). Exon 1 (bases numbered 1 to 236 in Fig. 3) contains the R and U5 regions of the 5' LTR and 130 nucleotides of the region preceding the *gag* gene. This is joined to exon 2 by a splice donor immediately before the *gag* initiation codon and a splice acceptor in ORF S1 (genome positions 440 and 5116, respectively) (24, 37). Exon 2, containing ORF S1, is 142 bases long, ending at nucleotide 378 of the cDNA sequence. It is joined to exon 3 by a splice donor following the S1 stop codon and a splice acceptor within the *env* gene (genome positions 5257 and 7216, respectively). Exon 3 is 948 bases long and contains the 3' end of the *env* gene, ORF S3, and the U3 and R regions of the 3' LTR. There are several nucleotide changes in this cDNA compared with reported genomic sequences; however, none of these are located within the *tat* coding region (see below).

Inferred amino acid sequence of the Tat protein. The deduced translation of the cDNA in clone 492 revealed two extended ORFs. The first, containing 83 codons, starts at nucleotide 126 and ends at nucleotide 377 of the cDNA sequence (Fig. 3); the 37 amino-terminal codons derived from the pre-*gag* region of exon 1 are joined to 46 codons of ORF S1, supplied by exon 2. Since this region contains ORF

FIG. 2. (A) Northern blot analysis of poly(A)⁺ RNA from EIAV-infected FEA cells hybridized with an EIAV LTR probe. Numbers to the left refer to positions of RNA molecular size markers in kilobases. The probe identifies the full-length genomic RNA of approximately 8 kb, the singly spliced RNA of approximately 4 kb, and a population of small multiply spliced mRNAs in the 1.5-kb size range. (B) Genetic organization of EIAV and structures of three EIAV cDNA clones. The major ORFs of EIAV, including the *gag*, *pol*, and *env* genes and ORFs S1 (*tat*), S2, and S3, are shown. The restriction enzyme sites shown at the top are *Nco*I (N), *Pvu*II (P), *Bam*HI (B), *Sma*I (Sm), and *Hpa*I (Hp). Sequences in the three cDNA clones (492, 272, and 442) are aligned with genomic sequences to show the exon (—) and intron (—) boundaries. The three clones used the same splice donor and acceptor for the upstream splice which joins the 5' end to the central region. Clones 492 and 272 use the same splice sites to join the central region to the 3' end of the genome. In contrast, clone 442 used a different 3' splice acceptor, located within ORF S2; in addition, cDNA synthesis was apparently initiated in an A-rich region in *env*.

S1, it is presumed to represent the *tat* mRNA. The second extended ORF, of 160 codons, follows the stop codon at the end of ORF S1. Thus, exon 3 is read for 105 codons in ORF S3; then the reading frame shifts relative to the genomic sequence (due to a single base deletion at position 836 of the cDNA) and continues for an additional 55 codons in the *env* ORF. By analogy to the primate lentiviruses, it is possible that this ORF encodes a *rev* gene.

The inferred amino acid sequence of EIAV Tat is compared with that of Tat proteins of other lentiviruses in Fig. 4. Two regions are conserved among the primate viruses and EIAV. The first is a region near the carboxy terminus, containing a string of basic amino acids, that may be involved in nuclear localization of the HIV-1 protein (21). The second, or core domain, is a 16-amino-acid sequence containing 12 residues that are very highly conserved among this group of proteins. This degree of sequence conservation suggests a pivotal role of this domain in Tat function. Interestingly, EIAV Tat lacks the cysteine-rich region that is conserved among the HIV and SIV Tat proteins, thereby defining the boundary of the core and upstream domains. The N-terminal portions of the various Tat proteins vary significantly in size and sequence. The EIAV, HIV, and SIV Tat proteins form a group quite distinct from the visna virus Tat and the putative Tat of FIV; these two show little similarity to any of the other proteins.

Expression and activity of EIAV *tat* cDNAs. Expression plasmids derived from EIAV *tat* cDNA clones were transfected into mammalian cells to evaluate *tat* expression and to examine the effects of amino acid deletions on Tat activity. The cDNA in clone 492 was inserted into the mammalian expression vector pRSPA, containing a Rous sarcoma virus

promoter and SV40 polyadenylation signals, to give pRS492. Transfection of D17 cells with pElcat in combination with pRS492 revealed that Tat activity was expressed at high levels (Table 2), suggesting that cDNA clone 492 represents an authentic *tat* mRNA. Expression plasmids containing cDNA from clones 442 and 272 yielded levels of Tat activity comparable to those of pRS492 (data not shown). A series of plasmids was then constructed in which fragments from cDNA clone 492 were inserted into the expression plasmid pRSPA to define the region containing the translation initiation codon (Fig. 5A). The first plasmid in this series, pRS-Etat-HA, contains the two *tat* coding exons located between the *Hae*II and *Apa*I sites (Fig. 3, positions 111 to 390). A translation termination codon was introduced at the *Sma*I site of pRS-Etat-HA to create the plasmid pRS-Etat-HAX. Expression of Tat activity in cells transfected with pRS-Etat-HA and pRS-Etat-HAX was compared over a range of DNA concentrations (Fig. 6; Table 2). pRS-Etat-HA activated EIAV LTR-controlled CAT expression even at the lowest concentration tested (1 ng), whereas pRS-Etat-HAX was relatively inactive. Plasmid pRS-Etat-S is deleted of sequences upstream from the *Sma*I site (Fig. 5A) and was also inactive in the cotransfection experiments (Table 2). Together, these data suggested that *tat* translation is initiated between the stop codon at position 123 and the *Sma*I site at position 170 of the cDNA. The cDNA fragment from the plasmid pRS-Etat-S was fused to a methionine initiation codon in a related vector to give pRS-Etat-M; this plasmid expressed Tat activity, indicating that sequences upstream of the *Sma*I site are required for translation initiation but are not essential for Tat function. Within this region of exon 1 are several triplets that have been shown to

FIG. 3. Nucleotide sequence and deduced translation of EIAV cDNA clone 492. Bases that differ from published proviral sequences are shown above the cDNA sequence; plus and minus signs indicate the addition or deletion, respectively, of bases in the cDNA compared with genomic sequence. The splice sites are underlined. The one-letter amino acid code is used below the cDNA sequence, and asterisks denote stop codons. The sequence has been entered into the GenBank data base as accession no. M30138.

act as alternative initiation codons in other systems (34); these include the CUG at position 141, UUG at 144, CUG at 150, and AUC at 165 of the cDNA sequence (Fig. 3). The CUG at 150, followed by the AUC at 165, would be predicted to be the most efficiently utilized start codons based on the Kozak consensus sequence context (26). The precise identification of the *tat* start codon is currently being addressed by site-directed mutagenesis.

The preceding experiments indicated that the region upstream of the *Sma*I site supplies the translation initiation codon but does not contain amino acids essential for Tat function. To define the essential N-terminal amino acids, cDNA fragments with progressive deletions of 5' sequences were ligated to expression plasmids that supply a translation initiation codon (Fig. 5B). Plasmid pRS-Etat-B was deleted of sequences upstream from the *Bst*NI site (position 216) and did not express Tat activity. When this fragment was ligated to a vector that provides a methionine initiation codon, pRS-Etat-MB, Tat activity was expressed at high levels in

transfected cells (Table 2). pRS-Etat-M25 lacks the entire exon 1 and 4 residues of exon 2 of *tat* and still expressed Tat activity in transfected cells. pRS-Etat-M24 is missing an asparagine and a tyrosine residue at the start of the core domain; this deletion reduced Tat activity by approximately 60%. Removal of the tyrosine and histidine at the N terminus of the core domain in plasmid pRS-Etat-M23 reduced Tat activity by 97%. Plasmid pRS-Etat-MP, which lacks the first 4 residues of the core region, did not express Tat activity. Thus, a non-AUG codon present in exon 1 is used for initiation of EIAV *tat* translation; however, other amino acids in exon 1 are not required for Tat function. Tat activity was reduced when the first two residues of the core domain were removed and abolished when the first four residues were deleted. These results support the domain topography inferred from the pattern of conserved Tat amino acids shown in Fig. 4.

Cell and virus specificity of EIAV Tat function. The *trans*-activation of EIAV and HIV-1 LTRs in canine D17 cells,

FIG. 4. Comparison of the amino acid sequences of lentivirus Tat proteins. The one-letter amino acid code is used, asterisks refer to stop codons, and dashes represent gaps introduced to align sequences. The Tat amino acid sequences are shown in two parts: the upper part shows the N termini, and the lower part shows the central and C-terminal regions of the proteins. The deduced amino acid sequences for the first coding exons of *tat* genes are shown for HIV-1 (2), HIV-2 (18), SIVmac (4), and SIVagm (14). Visna virus Tat is based on the sequence of Davis and Clements (8). For FIV, the deduced amino acid sequence is that of ORF 2, analogous to visna virus ORF S (*tat*), from FIV-14 (32). The sequence for EIAV Tat assumes that translation is initiated at the CUG codon at position 150 of cDNA 492. The boxed sequences, from left to right, correspond to the cysteine-rich, core, and basic domains; these are contiguous regions that are separated in the figure for emphasis.

primate COS cells, and human HeLa cells transfected with plasmids that expressed EIAV or HIV-1 Tat was compared (Table 3). The HIV-1 *tat* gene was cloned into the expression plasmid pRSPA to give pRS-Htat; plasmid pC15cat contains an HIV-1 LTR-controlled *cat* gene (40). Two different EIAV Tat expression plasmids were tested with pElcat; pRS-Etat-HA has the native non-AUG start codon, whereas pRS-Etat-M has a synthetic methionine initiation codon (Fig. 5A). The plasmids yielded nearly identical levels of Tat activity in all cell lines examined, indicating that the non-AUG start codon is utilized in cells other than D17. The ability of either HIV-1 or EIAV Tat to *trans*-activate the

heterologous LTR was examined in transfections with pElcat plus pRS-Htat or pC15cat plus pRS-Etat-M, respectively. In both cases a slight activation of the heterologous LTR was observed in most cell lines; however, this level of activity was 1 to 2 orders of magnitude lower than the activities obtained with the homologous components. Inspection of the CAT activities obtained in the various cell lines transfected with the HIV-1 components suggests that these function best in HeLa cells. The pattern of activities seen in the same cell lines transfected with the EIAV components indicated that these worked poorly in HeLa cells but gave the highest levels of activity in D17 cells. It is possible that these differences in activity result either from different affinities or associations of the Tat proteins with cellular cofactors or from cell-type-specific interactions of transcription factors with promoter control elements.

DISCUSSION

EIAV encodes a *tat* gene whose active exon, like those of the other lentiviruses, is located in the intergenic region between *pol* and *env*. The EIAV *tat* mRNA has three exons; the first two encode *tat* and the third may encode *rev*. The EIAV Tat protein is similar to but simpler than the primate lentivirus Tat proteins; although it shares C-terminal basic and highly conserved core domains with these other proteins, it lacks the N-terminal cysteine-rich domain. Surprisingly, the EIAV Tat protein is quite different from the visna virus Tat or the predicted Tat of FIV. The relationships among the various lentivirus Tat proteins parallel the similarities in the structure and position of the TAR elements. EIAV, HIV, and SIV have TAR elements that are located 3' from the RNA start site and form similar stem-loop structures (1, 12, 23, 30; P. L. Dorn and D. Derse, unpublished data), whereas visna virus TAR elements are located upstream of the promoter (22). Although EIAV Tat is related to the primate lentivirus Tat proteins, it has a distinct TAR specificity.

Structure and expression of EIAV *tat* cDNA. The arrangement of the coding exons in the EIAV *tat* mRNA differs from the other lentivirus *tat* mRNAs in two respects. First, EIAV contains *tat* in the first two, rather than the last two, exons; EIAV *tat* translation is initiated at a non-AUG codon in exon 1 rather than at a methionine codon at the beginning of exon 2. Second, the putative EIAV *rev* would be contained entirely within exon 3 rather than being split between the last two exons. Until more EIAV isolates are sequenced, we cannot exclude the possibility that the absence of an AUG initiation codon for *tat* is peculiar to certain strains of the virus. However, the use of a non-AUG initiation codon may represent a novel mechanism for balancing protein synthesis from polycistronic mRNAs. The first *tat* exon is present in the full-length mRNA which contains *gag* and *pol* genes; furthermore, the first two *tat* exons may be present in the *env* and *rev* mRNAs. Therefore, the efficient expression of the distal genes would require that the *tat* initiation codon be poorly utilized and frequently passed over by the ribosome. A regulatory strategy used by human T-cell lymphotropic virus and bovine leukemia virus to coordinately express proteins from polycistronic mRNAs makes use of the effects of AUG context on translation efficiency; synthesis of the downstream genes is apparently aided by the relatively poor sequence context of the upstream initiation codons (D. Derse, unpublished data). Certain codons that differ from AUG at a single base have been demonstrated to function as translation initiation codons, albeit at reduced efficiency, in

TABLE 2. *trans* activation of the EIAV LTR by *tat* cDNA expression plasmids

CAT plasmid	Tat expression plasmid	CAT activity (10 ³ cpm/h) ^a
pUXRSVcat	pRSPA	10.0
pElcat	pRSPA	1.2
pElcat	pRS492	16.6
pElcat	pRS-Etat-HA	20.0
pElcat	pRS-Etat-HAX	1.9
pElcat	pRS-Etat-S	1.4
pElcat	pRS-Etat-M	21.7
pElcat	pRS-Etat-B	1.3
pElcat	pRS-Etat-MB	18.6
pElcat	pRS-Etat-M25	19.0
pElcat	pRS-Etat-M24	7.8
pElcat	pRS-Etat-M23	2.0
pElcat	pRS-Etat-MP	1.4

^a Data are the mean of at least three transfections and are presented as the net cpm of product formed per hour under the conditions described in Materials and Methods. D17 cells were transfected with 5 µg of CAT plasmid and 0.01 µg of the indicated Tat expression plasmids. pUXRSVcat contains the Rous sarcoma virus promoter.

FIG. 5. (A) Structures of cDNA expression plasmids. At the top is shown the basic arrangement of the *tat* coding sequences in the EIAV *tat* cDNA. Numbers at the top refer to nucleotide positions in the cDNA (Fig. 3). Asterisks denote stop codons in the cDNA. The splice site is shown, indicating the sequences derived from exons 1 and 2; the two conserved domains are represented as boxes. Potential alternative initiation codons are indicated by CUG and AUC. Restriction enzyme sites are *Hae*II (H), *Sma*I (S), *Bst*NI (B), *Pvu*II (P), and *Apa*I (A). Restriction fragments from cDNA 492 were cloned into the expression plasmid, pRSPA, to give the series of plasmids starting with pRS-Etat-HA. Plasmid pRS-Etat-HAX has an oligonucleotide containing a stop codon inserted at the *Sma*I site. The fragments preceded by (Met) were cloned into vectors that supply an in-frame translation initiation codon. (B) Predicted N-terminal amino acid sequences of Tat proteins expressed from plasmids containing cDNA fragments fused to a methionine initiation codon. The two Tat-encoding exons are indicated. The core domain is boxed, and the C terminus of Tat is not shown. In pRS-Etat-M25, the vector donates Met and Pro residues; in pRS-Etat-M24, the vector supplies the N-terminal Met, Asp, Pro, and Pro residues; in pRS-Etat-M23, only the N-terminal Met is derived from the vector.

transfected cells and in cell-free systems (34). Furthermore, several eucaryotic and viral proteins, including human basic fibroblast growth factor (13), a second product of the human *c-myc* gene (19), a mouse mutant *dhfr* gene (33), adeno-associated virus capsid proteins (3), and Sendai virus C' protein (7), have recently been found to begin at non-AUG codons. The use of a non-AUG codon by EIAV may therefore represent yet another level of control of lentivirus gene expression.

Structure and function of Tat proteins. The importance of

specific protein domains with respect to Tat function can be deduced both from Tat amino acid sequence comparisons and from experimental manipulation of Tat proteins. EIAV encodes a comparatively simple Tat protein, related to but distinct from those of the primate lentiviruses. The presence of a C-terminal basic domain and a highly conserved core domain in the EIAV, HIV, and SIV Tat proteins suggests that these are functionally critical regions. The N-terminal boundary of the core domain and the boundaries of the cysteine-rich domain of HIV and SIV Tat are defined by the

FIG. 6. *trans*-activation of the EIAV LTR by various concentrations of EIAV Tat expression plasmids. D17 cells were transfected with 5 μ g of pElcat and the indicated amounts of pRS-Etat-HA (□) or pRS-Etat-HAX (♦). Two days later cells were assayed for CAT activity by the solvent partition method (see Table 1). CAT activity is expressed as cpm of [^{14} C]acetyl coenzyme A converted to product per hour multiplied by 10^{-3} .

EIAV Tat sequence, since it does not have the latter domain. The absence of the cysteine-rich domain in EIAV Tat suggests that it may not be required for *trans*-activation by the other Tat proteins. That this region may play an auxiliary role, perhaps in modulating Tat activity by altering protein conformation, is supported by the experimental observations that HIV-1 Tat peptides lacking the cysteine-rich region possessed Tat activity (17), whereas activity was lost when specific amino acids within this domain were altered (38). By using synthetic peptides, it was revealed that the minimal, active HIV-1 Tat peptide contained the core and basic domains (16, 17), consistent with the conservation of these regions among the divergent viruses. Removal of the four N-terminal amino acids of the EIAV Tat core domain abolished activity, whereas the minimal, active HIV-1 Tat

peptides lack six N-terminal residues from this domain (16). This discrepancy may reflect differences in the Tat proteins, cell lines, or experimental approaches used. The basic domain has been shown to be required for HIV-1 Tat activity and nuclear localization (21). Although it is hypothesized to act in nucleic acid binding (16), the role of this and other domains in *trans* activation is not known. The distinct TAR and cell type specificities of EIAV and HIV Tat proteins, combined with their modular domain structures, suggest that chimeric EIAV-HIV Tat proteins may be useful in analyzing the function of each region in the *trans*-activation process.

Evolution of lentivirus Tat proteins. The similarity of the *cis*- and *trans*-acting components of the EIAV Tat system to the primate lentiviruses but not to visna virus or FIV is paradoxical in light of the lentivirus phylogenies derived from nucleotide sequence comparisons of *pol* genes (10, 28). These phylogenies would generally place EIAV closer to visna virus and FIV than to the primate lentiviruses. Moreover, EIAV, visna virus, and FIV contain a proteaselike gene segment, inserted between the RNase H- and endonuclease-encoding regions of the *pol* gene, that is not present in HIV or SIV (27). It was suggested that the gene segment was captured after the divergence of an EIAV, FIV, and visna virus ancestor from the primate lentivirus progenitor. The alternative possibility, that the ancestor of HIV and SIV lost this gene segment, is supported by the relationship of EIAV to the other lentiviruses based on the comparison of the *trans*-activation components. EIAV *trans* activation is more similar to that of HIV and SIV than to that of the nonprimate lentiviruses. The position of the *tat* (and perhaps *rev*) exons in the genome and the splicing patterns that are used to generate the *tat* mRNA are very similar in all the lentiviruses examined in detail, suggesting that these genes were present before the divergence of EIAV, HIV, SIV, FIV, and visna virus. It is possible that recombination events have led to mixing or that FIV and visna virus independently evolved *trans* activators of a different type from those of EIAV, HIV, and SIV. As more and different lentiviruses are examined, perhaps a clearer picture of the evolution of regulatory function in lentiviruses will emerge.

ACKNOWLEDGMENTS

This project was funded in part by federal funds from the Department of Health and Human Services under contract N01-CO-74102 with Program Resources Inc. L.D.S. was supported by CAPES fellowship 3721/87-2, Ministry of Education, Brazil.

LITERATURE CITED

1. Arya, S. K., and R. C. Gallo. 1988. Human immunodeficiency virus type 2 long terminal repeat: analysis of regulatory elements. *Proc. Natl. Acad. Sci. USA* 85:9753-9757.
2. Arya, S. K., C. Guo, S. F. Josephs, and F. Wong-Staal. 1985. *trans*-activator gene of human T-lymphotrophic virus type III (HTLV-III). *Science* 229:69-73.
3. Bacerra, 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.
4. Chakrabarti, L., M. Guyader, M. Alizon, M. D. Daniel, R. C. Desrosiers, P. Tiollais, and P. Sonigo. 1987. Sequence of simian immunodeficiency virus from macaque and its relationship to other human and simian retroviruses. *Nature (London)* 328:543-547.
5. Cheevers, W. P., and T. C. McGuire. 1985. Equine infectious anemia virus: immunopathogenesis and persistence. *Rev. Infect. Dis.* 7:83-88.
6. Chirgwin, J. M., A. E. Przybyla, R. J. MacDonald, and W. J. Rutter. 1979. Isolation of biologically active ribonucleic acid

TABLE 3. *trans* activation of EIAV and HIV LTRs in various cell lines by EIAV and HIV Tat expression plasmids

CAT plasmid (LTR)	Tat plasmid	CAT activity (10^3 cpm/h) ^a in:		
		HeLa	COS	D17
pUXRSVcat (Rous sarcoma virus)	pRSPA	1.10	10.5	6.71
pElcat (EIAV)	pRSPA	0.01	0.04	0.45
pElcat (EIAV)	pRS-Etat-HA	0.75	0.93	20.3
pElcat (EIAV)	pRS-Etat-M	0.86	0.34	24.1
pElcat (EIAV)	pRS-Htat	0.03	0.02	1.91
pC15cat (HIV-1)	pRSPA	0.09	0.08	1.88
pC15cat (HIV-1)	pRS-Htat	162	48.6	35.7
pC15cat (HIV-1)	pRS-Etat-M	0.59	0.17	3.84

^a The data are the mean of four transfection experiments. CAT activities are presented as the net cpm of product formed per hour. CAT plasmid (5 μ g) plus 1 μ g of the indicated Tat expression plasmid were transfected onto cultures of HeLa, COS, or D17 cells. Two days later, CAT activity was assayed as described in Table 1, footnote a.

- from sources enriched in ribonuclease. *Biochemistry* **18**:5294–5299.
7. Curran, J., and D. Kolakofsky. 1988. Ribosomal initiation from an ACG codon in the Sendai virus P/C mRNA. *EMBO J.* **7**:245–251.
  8. 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.
  9. 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.
  10. Doolittle, R. F., D. F. Feng, M. S. Johnson, and M. A. McClure. 1989. Origins and evolutionary relationships of retroviruses. *Q. Rev. Biol.* **64**:1–29.
  11. 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.
  12. Feng, S., and E. C. Holland. 1988. HIV-1 tat trans-activation requires the loop sequence within TAR. *Nature (London)* **334**:165–167.
  13. 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.
  14. Fukasawa, M., T. Miura, A. Hasegawa, S. Morikawa, H. Tsujimoto, K. Miki, T. Kitamura, and M. Hayami. 1988. Sequence of simian immunodeficiency virus from African green monkey, a new member of the HIV/SIV group. *Nature (London)* **333**:457–461.
  15. Gonda, M. A., M. J. Braun, J. W. Bess, Jr., L. O. Arthur, and M. J. Van Der Maaten. 1987. Characterization and molecular cloning of a bovine lentivirus related to human immunodeficiency virus. *Nature (London)* **330**:388–391.
  16. Green, M., M. Ishino, and P. M. Loewenstein. 1989. Mutational analysis of HIV-1 Tat minimal domain peptides: identification of trans-dominant mutants that suppress HIV-LTR-driven gene expression. *Cell* **58**:215–223.
  17. Green, M., and P. M. Loewenstein. 1988. Autonomous functional domains of chemically synthesized human immunodeficiency virus tat trans-activator protein. *Cell* **55**:1179–1188.
  18. 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.
  19. 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.
  20. Hattori, M., and Y. Sakaki. 1986. Dideoxy sequencing method using denatured plasmid templates. *Anal. Biochem.* **152**:232–238.
  21. Hauber, J., M. H. Malim, and B. R. Cullen. 1989. Mutational analysis of the conserved basic domain of human immunodeficiency virus tat protein. *J. Virol.* **63**:1181–1187.
  22. Hess, J. L., J. A. Small, and J. E. Clements. 1989. Sequences in the visna virus long terminal repeat that control transcriptional activity and respond to viral trans-activation: involvement of AP-1 sites in basal activity and *trans*-activation. *J. Virol.* **63**:3001–3015.
  23. Jakobovits, A., D. H. Smith, E. B. Jakobovits, and D. J. Capon. 1988. A discrete element 3' of human immunodeficiency virus 1 (HIV-1) and HIV-2 mRNA initiation sites mediates transcriptional activation by an HIV transactivator. *Mol. Cell. Biol.* **8**:2555–2561.
  24. 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.
  25. Kono, Y., K. Hirasawa, Y. Fukunaga, and T. Taniguchi. 1976. Recrudescence of equine infectious anemia by treatment with immunosuppressive drugs. *Natl. Inst. Anim. Health Q.* **16**:8–15.
  26. Kozak, M. 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**:283–292.
  27. McClure, M. A., M. S. Johnson, and R. F. Doolittle. 1987. Relocation of a protease-like gene segment between two retroviruses. *Proc. Natl. Acad. Sci. USA* **84**:2693–2697.
  28. McClure, M. A., M. S. Johnson, D. F. Feng, and R. F. Doolittle. 1988. Sequence comparisons of retroviral proteins: relative rates of change and general phylogeny. *Proc. Natl. Acad. Sci. USA* **85**:2469–2473.
  29. 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.
  30. 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.
  31. Neumann, J. R., C. A. Morency, and K. O. Russian. 1987. A novel rapid assay for chloramphenicol acetyltransferase gene expression. *BioTechniques* **5**:444–447.
  32. Olmsted, R. A., V. M. Hirsch, R. H. Purcell, and P. R. Johnson. 1989. Nucleotide sequence analysis of feline immunodeficiency virus: genome organization and relationship to other lentiviruses. *Proc. Natl. Acad. Sci. USA* **86**:8088–8092.
  33. Peabody, D. S. 1987. Translation initiation at an ACG triplet in mammalian cells. *J. Biol. Chem.* **262**:11847–11851.
  34. Peabody, D. S. 1989. Translation initiation at non-AUG triplets in mammalian cells. *J. Biol. Chem.* **264**:5031–5035.
  35. Pedersen, N. C., E. W. Ho, M. L. Brown, and J. K. Yamamoto. 1987. Isolation of a T-lymphotropic virus from domestic cats with an immunodeficiency-like syndrome. *Science* **235**:790–793.
  36. 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) long terminal repeat. *Cell* **41**:813–823.
  37. Rushlow, K., K. Olsen, G. Stiegler, S. L. Payne, R. C. Montelaro, and C. J. Issel. 1986. Lentivirus genomic organization: the complete nucleotide sequence of the env gene region of equine infectious anemia virus. *Virology* **155**:309–321.
  38. 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.
  39. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463–5467.
  40. Seigel, L. J., L. Ratner, S. F. Josephs, D. Derse, M. B. Feinberg, G. R. Reyes, S. J. O'Brien, and F. Wong-Staal. 1986. Transactivation induced by human T-lymphotropic virus type III (HTLV III) maps to a viral sequence encoding 58 amino acids and lacks tissue specificity. *Virology* **148**:226–231.
  41. Shaw, G. M., B. H. Hahn, S. K. Arya, J. E. Groopman, R. C. Gallo, and F. Wong-Staal. 1984. Molecular characterization of human T-cell (lymphotropic) virus type III in the acquired immune deficiency syndrome. *Science* **226**:1165–1171.
  42. 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.
  43. Sodroski, J., R. Patarca, C. A. Rosen, F. Wong-Staal, and W. A. Haseltine. 1985. Location of the trans-activating region on the genome of human T-cell lymphotropic virus type III. *Science* **229**:74–77.
  44. 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.
  45. Viglianti, G. A., and J. I. Mullins. 1988. Functional comparison of transactivation by simian immunodeficiency virus from Rhesus macaques and human immunodeficiency virus type 1. *J. Virol.* **62**:4523–4532.