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

PATRICIA DORN,† LUIS DASILVA,2 LUIS MARTARANO,1 AND DAVID DERSE3

Biological Carcinogenesis and Development Program, Program Resources Inc., Frederick Cancer Research Facility, Frederick, Maryland 21701; Instituto Nacional do Cancer, Rio de Janeiro, Brazil2; and Laboratory of Viral Carcinogenesis, National Cancer Institute, Frederick, Maryland 21701-1013

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 EcoRI linkers and ligated to EcoRI-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, andideoxy-nucleotide sequencing (20, 39).

Plasmid construction. pElcat contains the EIAV LTR
cloned into pUXcat (11). pC15cat contains an HIV LTR fragment cloned into pSV0cat (40). pUXS2Vcat 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) Ncol-Smal fragment (genome positions 4870 to 5670) cloned into the HindII site of pS. pS1 contains the 448-bp Ncol-BamHI fragment from this region in pS. pS2 contains the 530-bp PvuII-Smal fragment in the HindII site of pS. pS1,2,Rsa contains the 608-bp Rsal-Smal fragment in pS. Translation termination codons were introduced into ORF S1 or S2 of plasmid pS1,2 to form plasmids pS1,2,R and pS1,a2, 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. pRS-PKA and pRS-PSA 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 start site with respect to the EIAV genome. pRS-Hoe contains the HIV tat gene from the HXB2 provirus clone (41); a 343-bp Sall/Sau96I fragment (genome positions 5785 to 6128) was inserted into the Sall-to-EcorV sites of pRS-PKA. The inserts from EIAV tat cDNA clones 442 and 492 were released as Xhol/XbaI fragments and cloned directly into the same sites of pRS-PKA to give pRS442 and pRS492, respectively. A 279-bp HaelI-Apal fragment from cDNA 492 was cloned, after T4 DNA polymerase digestion of the HaeII end, into the SmaI-to-Apal sites of pRS-PKA to give pRS-PS-Etat-HA. pRS-PS-Etat-HAX was made by ligating the oligonucleotide CTAAGTCAGA to the SmaI site, within the tat gene, of pRS-PS-Etat-HA. pRS-PS-Etat-S contains the HaeII fragment of the EIAV-infected cells DNA cloned pS1-Hoe; the same sites of pRS-PKA. pRS-PK-S was modified by the insertion of the oligonucleotide CCACCATGG into the EcorV site to give pRS-PK-Met; digestion of this plasmid with Ncol and end filling with DNA polymerase provides an efficient methionine initiation codon for inserted genes. The Smal-Apal fragment from pRS-PS-Etat-S was ligated to the polymerase-filled Ncol- and Apal-digested sites of pRS-PKA to give pRS-PS-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 BsrNI and mung bean nuclease followed by Apal; it was then ligated to pRS-PKA (SmaI to Apal) and pRS-PK-Met (Ncol to Apal) to give pRS-PS-Etat-B and pRS-PS-Etat-MB, respectively. The plasmid pRS-PS-Etat-MP was made in a similar way by insertion of the 126-bp PvuII-Apal fragment from cDNA 492 into pRS-PK-Met. The plasmids pRS-PS-Etat-M23, pRS-PS-Etat-M24, and pRS-PS-Etat-M25 were constructed by exonuclease III and mung bean nuclease digestion of the 5′ end of cDNA 492 followed by Apal digestion and ligation to pRS-PK-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 maintained in Dulbecco modified Eagle medium supplemented with 10% fetal calf serum (GIBCO Laboratories). For chloramphenicol acetyltransferase (CAT) assays, 3 × 10⁵ 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 precipitates 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 μl 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-μl reaction mixture containing 100 mM Tris hydrochloride (pH 7.8), 1.0 mM chloramphenicol, 0.1 μCi of [3H]acetate, and 4 mM NADH (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 acetylchloramphenicol 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-μg 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 tat gene, two small ORFs designated pS1 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-Smal 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
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 pSor1,2, which has a stop codon in S1, did not express Tat activity, whereas plasmid pS1,2r, 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 pS1,25Rsa. Although pS1,25Rsa does not possess any AUG codons preceding S1, it still expressed Tat activity (Table 1). It should be noted that the genomic fragment in pS1,25Rsa, 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 E-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 coding 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 (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 precursor region of exon 1 are joined to 46 codons of ORF S1, supplied by exon 2. Since this region contains ORF

---

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

<table>
<thead>
<tr>
<th>CAT plasmid</th>
<th>Tat expression plasmid</th>
<th>Relative CAT activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>pElcat</td>
<td>pS</td>
<td>0.29</td>
</tr>
<tr>
<td>pElcat</td>
<td>pS1,2</td>
<td>3.4</td>
</tr>
<tr>
<td>pElcat</td>
<td>pS1</td>
<td>7.5</td>
</tr>
<tr>
<td>pElcat</td>
<td>pS2</td>
<td>0.30</td>
</tr>
<tr>
<td>pElcat</td>
<td>pSor1,2</td>
<td>1.0</td>
</tr>
<tr>
<td>pElcat</td>
<td>pSor1,2r</td>
<td>4.1</td>
</tr>
<tr>
<td>pElcat</td>
<td>pS1,25Rsa</td>
<td>3.3</td>
</tr>
</tbody>
</table>

* Data are the mean of 4 to 12 separate transfections. Background CAT activity obtained with a promoterless plasmid, pUXSV2cat, 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).
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 HaeII and Apal sites (Fig. 3, positions 111 to 390). A translation termination codon was introduced at the Smal 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-HAX 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 Smal 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 Smal 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 Smal 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
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 SmaI 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 BstN1 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 topology inferred from the pattern of conserved Tat amino acids shown in Fig. 4.

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

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

<table>
<thead>
<tr>
<th>CAT plasmid</th>
<th>Tat expression plasmid</th>
<th>CAT activity (10^5 cpm/b)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pUXRSVcat</td>
<td>pRS492</td>
<td>10.0</td>
</tr>
<tr>
<td>pElcat</td>
<td>pRS492</td>
<td>1.2</td>
</tr>
<tr>
<td>pElcat</td>
<td>pRS-Etat-HA</td>
<td>16.6</td>
</tr>
<tr>
<td>pElcat</td>
<td>pRS-Etat-HAX</td>
<td>20.0</td>
</tr>
<tr>
<td>pElcat</td>
<td>pRS-Etat-M</td>
<td>1.9</td>
</tr>
<tr>
<td>pElcat</td>
<td>pRS-Etat-S</td>
<td>21.7</td>
</tr>
<tr>
<td>pElcat</td>
<td>pRS-Etat-B</td>
<td>1.4</td>
</tr>
<tr>
<td>pElcat</td>
<td>pRS-Etat-MB</td>
<td>1.3</td>
</tr>
<tr>
<td>pElcat</td>
<td>pRS-Etat-M23</td>
<td>18.6</td>
</tr>
<tr>
<td>pElcat</td>
<td>pRS-Etat-M24</td>
<td>19.0</td>
</tr>
<tr>
<td>pElcat</td>
<td>pRS-Etat-M24</td>
<td>7.8</td>
</tr>
<tr>
<td>pElcat</td>
<td>pRS-Etat-M23</td>
<td>2.0</td>
</tr>
<tr>
<td>pElcat</td>
<td>pRS-Etat-EP</td>
<td>1.4</td>
</tr>
</tbody>
</table>

* 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.
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), adenovirus-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 μg of pElcat and the indicated amounts of pRS-Etat-HAX 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 [14C]acetate 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 proteaselle 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 the 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