Equine Infectious Anemia Virus Gene Expression: Characterization of the RNA Splicing Pattern and the Protein Products Encoded by Open Reading Frames S1 and S2†

R. LOUIS SCHILTZ,¹ DING S. SHIH,^{1*} SIYAMAK RASTY,² RONALD C. MONTELARO,²
AND KEITH E. RUSHLOW²

Department of Biochemistry and Louisiana State University Agricultural Center, Louisiana State University, Baton Rouge, Louisiana 70803,¹ and Department of Molecular Genetics and Biochemistry, School of Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania 15261²

Received 25 November 1991/Accepted 10 March 1992

The utilization of predicted splice donor and acceptor sites in generating equine infectious anemia virus (EIAV) transcripts in fetal donkey dermal cells (FDD) was examined. A single splice donor site identified immediately upstream of the *gag* coding region joins the viral leader sequence to all downstream exons of spliced EIAV transcripts. The predominant 3.5-kb transcript synthesized in EIAV-infected FDD cells appears to be generated by a single splicing event which links the leader sequence to the first of two functional splice acceptor sites near the 5' end of the S1 open reading frame (ORF). The translation products encoded by the 3.5-kb transcript were examined by producing *in vitro* transcripts from a cDNA corresponding to this RNA followed by *in vitro* translation in wheat germ extracts. These transcripts directed the synthesis of three proteins: the virus *trans*-activator protein (EIAV Tat) encoded by ORF S1, a protein of unknown function encoded by ORF S2, and the virus envelope glycoprotein. When transfected into FDD cells, this cDNA also directed expression of EIAV Tat. Amino-terminal sequence analysis of the *in vitro*-synthesized S1 protein supports the suggestion that translation of EIAV Tat is initiated at a CUG codon within the virus leader region. Both *in vitro*-synthesized S2 protein and synthetic peptides corresponding to S2 are shown to react positively with sera obtained from EIAV-infected horses, providing the first direct evidence of expression of this protein in infected animals.

Equine infectious anemia virus (EIAV) is a member of the lentivirus subfamily of retroviruses, a group that includes the ungulate (visna-maedi, caprine arthritis-encephalitis, and bovine immunodeficiency-like viruses) and the primate (human immunodeficiency virus types 1 and 2 [HIV-1 and HIV-2] and simian immunodeficiency virus) retroviruses (18). The genetic organization of EIAV is relatively simple compared with that of other lentiviruses (6) in that the gene contains only three short open reading frames (ORFs), S1, S2, and S3, in addition to the *gag*, *pol*, and *env* genes common to all replication-competent retroviruses. ORF S1 is located within the *pol-env* intergenic region, while ORF S2 begins between ORF S1 and *env* and overlaps the amino terminus of the *env* gene in a different reading frame. ORF S3 is positioned near the 3' end of the genome and is encoded within an alternate reading frame overlapping the gp45-coding sequence of the *env* gene. The S1 gene product, which shares both structural and functional homology with the Tat protein of HIV-1, has been identified as the EIAV *trans*-activator protein (8, 20, 30). Although ORF S1 contains no AUG codon, it has been suggested that translation of EIAV Tat may begin at a CUG codon within the viral leader sequence, which is joined in-frame with S1 as a result of an RNA splicing event (8, 30). On the basis of sequence homology with HIV-1 Rev and nonsense mutations within S3, which seem to impart a Rev-defective phenotype to EIAV-infected canine cells (30), ORF S3 is believed to encode the EIAV Rev protein. The predicted protein en-

coded by ORF S2 bears no significant homology to any of the lentiviral ancillary proteins described to date and has yet to be assigned a function. Thus far, no transcript or protein product that corresponds to expression of ORF S2 has been identified.

EIAV gene expression has been primarily studied in canine and feline fibroblast cultures which are permissive for persistent EIAV infections. Analyses of RNA transcription patterns in such cells yielded variable observations in the temporal appearance and abundance of EIAV-specific transcripts (8, 21, 30). The simplest and most consistent transcription patterns were observed in EIAV-infected primary horse macrophage cultures and equine fibroblasts (fetal equine kidney [FEK] and fetal donkey dermal [FDD] cells), which predominantly synthesized transcripts corresponding to the 8.2-kb genomic RNA and the presumably singly spliced 3.5-kb envelope message (25, 30). Small multiply spliced transcripts of 1 to 2 kb were mainly undetectable in EIAV-infected equine cells, a surprising observation, since viral *trans*-activation factors are expressed in EIAV-infected equine fibroblasts (25). In the present work, we have systematically examined the splicing patterns of EIAV in FDD cells by using cDNA cloning, Northern (RNA) blot hybridization, and nuclease S1 protection assays.

Previous studies in our laboratory showed that EIAV infection of cultured FDD cells resulted in a cytopathic infection, whereas FEK cells became persistently infected *in vitro* (25). Interestingly, in cytopathically infected FDD cells, the proportion of 3.5-kb mRNA to full-length 8.2-kb transcript was substantially greater than in persistently infected FEK cells, which synthesized the two RNAs in approximately equal abundance at all stages of infection.

* Corresponding author.

† Approved by the Director of the Louisiana Agricultural Experiment Station for publication as manuscript number 91-12-551.

Examination of the properties of the predominant 3.5-kb EIAV transcript of FDD cells might provide some insight into the mechanisms of cytopathicity observed in this cell type. To address this question as part of the study described in this report, *in vitro* transcripts were generated from a cDNA derived from the 3.5-kb RNA and used to prime translation in wheat germ extracts. These transcripts were found to produce the S1, S2, and envelope proteins of EIAV as confirmed by radioimmunoprecipitation assays and amino-terminal sequence analyses of S1 and S2. Sequence analysis of the S1 EIAV Tat protein was consistent with the suggestion of translational initiation at a CUG codon within the leader region. The S2 protein was specifically immunoprecipitated by sera from EIAV-infected horses, demonstrating its production during EIAV infection of host animals. The 3.5-kb RNA-derived cDNA was also tested for encoding *trans*-activation factors in cultured FDD cells by cotransfection of a simian virus 40 (SV40) replacement vector containing the cDNA with a long terminal repeat (LTR)-driven reporter plasmid. The results of these studies are presented in this report.

MATERIALS AND METHODS

Virus strains and cell cultures. Primary cultures of FDD cells were prepared and maintained as previously described (24, 25). An FDD-adapted stock of prototype EIAV was utilized in these studies. This stock was generated by propagation of the Wyoming cell-adapted strain of EIAV (15) in primary cultures of FEK cells to produce prototype virus followed by serial passage in FDD cells. Confluent monolayers of FDD cells were infected with FDD-adapted virus at a multiplicity of infection of 1. Cells were harvested for RNA isolation as previously described (25).

Isolation and purification of RNA. Total cellular RNA was isolated from EIAV-infected or uninfected FDD cells by a modified guanidinium thiocyanate extraction method (4, 16) as previously described (25). Poly(A)⁺ RNA was purified by two cycles of oligo(dT)-cellulose chromatography of total cellular RNA (16) or by direct oligo(dT)-cellulose chromatography of cellular lysates by the commercially available Fast Track mRNA Isolation kit (Invitrogen Corp.) according to the manufacturer's specifications.

cDNA cloning. Double-stranded cDNA was produced from poly(A)⁺ RNA from EIAV-infected cells by using a commercially available cDNA synthesis kit (Bethesda Research Laboratories). First-strand synthesis was primed with a 30-nucleotide (nt) primer complementary to a region of the envelope gene immediately downstream of the *Sma*I site (nt 5693 to 5722). The double-stranded cDNA products were tailed with oligo(dC) by using terminal deoxynucleotidyl transferase and ligated to *Pst*I-digested pUC9 tailed with oligo(dG). The recombinant plasmids were transformed into DH5 α *Escherichia coli* cells and plated on 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside (X-Gal) selective media. White colonies were selected and screened for the presence of EIAV-specific cDNA inserts by colony hybridization (16), using as a probe an α -³²P-labeled *Bam*HI-*Taq*I fragment corresponding to a portion of ORF S2 located immediately upstream of the priming site for first-strand cDNA synthesis. A total of 16 positive clones were further analyzed by Southern blot hybridization. Minilysate DNA from these clones was digested with *Pst*I to release the cDNA inserts and fractionated on 1% agarose gels. Following transfer to nitrocellulose, the membrane was probed with an α -³²P-labeled 226-bp *Mlu*I-*Bam*HI restriction fragment (nt 156 to

386) corresponding to a portion of the viral LTR and leader sequence. Two clones, designated pSR-1 and pSR-2, contained cDNA inserts which hybridized with this probe. These were subcloned as 270-bp *Sma*I-*Bam*HI restriction fragments into M13mp18 or M13mp19 for sequence analysis by the dideoxy chain termination method (27).

Northern blot analysis. Poly(A)⁺ RNA (2.5 μ g) isolated from EIAV-infected FDD cells was fractionated on 1.4% agarose-formaldehyde denaturing gels and transferred to nitrocellulose membranes. The membrane was probed with one of three ³²P-labeled synthetic oligonucleotides designed to be complementary to regions of the EIAV leader sequence or amino-terminal portions of the *gag* coding region. These probes were designed to be specific for the three putative splice donors in this region (see Fig. 2A). Probe 1 (SD-1) is a 28-mer complementary to nt 428 to 455 within the EIAV leader immediately upstream of the *Gag* polyprotein AUG initiation codon. Probe 2 (SD-2) is a 28-mer complementary to nt 468 to 495 covering the amino-terminal region of the p15^{gag} protein. Probe 3 (SD-3) is a 27-mer complementary to nt 518 to 544 within the p15 coding sequence.

Nuclease S1 protection assays. Total cellular RNA isolated from EIAV-infected or uninfected FDD cells was annealed to one of four different end-labeled probes (see Fig. 3), and the resulting complex was subjected to S1 nuclease digestion to identify the splice sites of EIAV-specific mRNAs. Probe 1 (P1) was generated by subcloning a 308-bp *Taq*I-*Pvu*II EIAV proviral DNA restriction fragment encompassing the three putative splice donor sequences near the amino terminus of the *gag* coding region into *Acc*I-*Sma*I-digested M13mp18 and then digesting it with *Nar*I. The resulting probe contained a 294-bp *Nar*I-*Pvu*II EIAV-specific sequence (nt 326 to 620) followed by a 250-bp sequence derived from the *Sma*I to *Nar*I restriction sites of M13mp18. This probe was 3' end labeled with [α -³²P]dCTP and the Klenow fragment of *E. coli* DNA polymerase I. Probe 2 (P2) was constructed by subcloning a 450-bp *Nco*I-*Bam*HI EIAV proviral restriction fragment encompassing the 3' end of the *pol* gene and extending through the 5' end of the *env* gene (nt 4889 to 5337) into *Bam*HI-*Sma*I-digested M13mp19 following treatment of the *Nco*I site with the Klenow fragment of *E. coli* DNA polymerase I. The probe DNA was excised by digestion with *Bam*HI and *Bgl*II to yield a 611-bp restriction fragment containing 450 bp of EIAV sequence and 161 bp of M13mp19 sequence. This restriction fragment was dephosphorylated with calf intestine alkaline phosphatase and 5' end labeled with T4 polynucleotide kinase and [γ -³²P]ATP. Probe 3 (P3) corresponds to a 617-bp *Pvu*II-*Hind*III proviral restriction fragment extending from within ORF S1 to the 5' end of the *env* gene (nt 5161 to 5778). This restriction fragment was 3' end labeled with [α -³²P]ATP and [α -³²P]CTP by replacement synthesis with T4 DNA polymerase. Probe 4 (P4) was generated by digested of proviral DNA with *Dra*I and *Sca*I and isolation of a 385-bp restriction fragment corresponding to a region of the *env* gene, which also includes the 5' end of ORF S3 (nt 7026 to 7412). This restriction fragment was 5' end labeled with [γ -³²P]ATP and T4 polynucleotide kinase following dephosphorylation with calf intestine alkaline phosphatase.

The ³²P-labeled probes (10⁶ cpm) were combined with total cellular RNA (40 μ g) isolated from EIAV-infected or uninfected FDD cells and ethanol precipitated. The mixture was suspended in 30 μ l of S1 nuclease hybridization buffer {80% deionized formamide, 40 mM PIPES [piperazine-N,N'-bis(2-ethanesulfonic acid); pH 6.4], 400 mM NaCl, 1 mM EDTA} and placed into an 85°C water bath for 15 min to

denature the probe. Hybridization of the denatured probes to the RNAs was accomplished by overnight incubation at 50°C for probes P1 and P3 or at 55°C for probes P2 and P4. S1 nuclease digestion was accomplished by addition of the following: 150 µl of 2× S1 nuclease buffer (0.56 M NaCl, 0.1 M sodium acetate [pH 4.5], 9 mM ZnSO₄), 3 µl of 2-mg/ml single-stranded calf thymus DNA, 147 µl of diethyl pyrocarbonate (DEPC)-treated doubly distilled H₂O, 300 U of S1 nuclease [Bethesda Research Laboratories]. The mixture was incubated at 30°C for 60 min, after which the reaction was stopped by the addition of 80 µl of S1 stop buffer (4 M ammonium acetate, 20 mM EDTA [pH 8.0], 40 mg of yeast tRNA per ml). The products were concentrated by ethanol precipitation and fractionated by gel electrophoresis on a 6% acrylamide-urea sequencing gel. *Msp*I-digested pBR322 plasmid DNA labeled with [α -³²P]dCTP by the Klenow fragment of *E. coli* DNA polymerase I was utilized as a molecular weight marker.

In vitro transcription and eucaryotic expression plasmid constructs. In vitro transcription and eucaryotic transfection plasmids were constructed by subcloning restriction fragments of the cDNA clone pSR-1 into pSP65 (Promega Corp.) or into an SV40 replacement vector, pSV2Δcat, derived from pSV2cat (11). The pSV2Δcat vector was produced by digestion of pSV2cat with *Hind*III and *Hpa*I to remove the chloramphenicol acetyltransferase (CAT) gene followed by insertion of a synthetic linker that regenerates each of these sites and provides a unique *Xho*I site. In the pSR-1 cDNA clone, ORF S1, ORF S2, and *env* are spliced to the viral leader region through a splice donor site (*sd*1) localized immediately upstream of the *gag* gene (nt 459). pSR-1 was digested with either *Stu*I (nt 256) or *Sma*I (nt 303) and then digested with *Msc*I (nt 5655) to generate a 720-bp (*Stu*I-*Msc*I) or 585-bp (*Sma*I-*Msc*I) restriction fragment. These blunt-end DNAs were subcloned into *Sma*I-digested pSP65, generating pSP720 and pSP585, respectively, or into *Stu*I-digested pSV2Δcat, generating pSV720 and pSV585, respectively (Fig. 5A). The resulting recombinant plasmids were screened for proper orientation by using appropriate restriction enzymes.

The EIAV LTR CAT plasmid, pLTRcat, used in *trans*-activation assays has been previously described (7, 25). The pSV2cat plasmid (11), which contains the CAT gene under the control of the SV40 immediate-early promoter, was used as a positive control in CAT assay transfections. Transfection plasmids were purified by two rounds of cesium chloride density centrifugation as previously described (10).

Plasmid DNA transfection and CAT assays. CsCl density gradient-purified DNA (11 µg) was used to transfect 60-mm petri dishes of FDD cells by the calcium phosphate coprecipitation method (10). pLTRcat (5 µg) and pSVβ-gal (1 µg) were cotransfected with pSV720, pSV585, or pSV2Δcat (5 µg). pSVβ-gal plasmid (Promega Biotec) was added to each transfection to control for transfection efficiency. Cell lysates were prepared 48 h posttransfection as previously described (10), and the β-galactosidase activities of the cell lysates from each transfection were measured according to the supplier's recommendations (Promega Corp.). pSV2cat (5 µg) was used as a positive control DNA for CAT activity and was cotransfected with pSV2Δcat (5 µg) and pSVβ-gal (1 µg). Lysate volumes containing equivalent levels of β-galactosidase activity were assayed for CAT enzyme activity by the kinetic diffusion method of Neumann et al. (19) with [¹⁴C]butyryl coenzyme A (New England Nuclear, Dupont) and chloramphenicol as substrates for CAT.

Peptide synthesis and ELISAs. Synthetic peptides were

prepared either by using a SAM-II automated peptide synthesizer (BioSearch) or by using manual methods and the RaMPS system (Dupont). Reaction conditions used for synthesis were according to the manufacturers' specifications. Peptides were initially purified by gel filtration on Sephadex G-25 followed by reverse-phase high-pressure liquid chromatography and further characterized by plasma desorption mass spectrometry to confirm their purity and sequences. Each peptide was reacted against a panel of horse sera in an enzyme-linked immunosorbent assay (ELISA) optimized for use with synthetic peptides antigens (3). Negative ELISA values were established by examining a panel of eight samples of normal horse serum and averaging the reactivity values against each peptide.

In vitro transcription and translation. Capped mRNAs were synthesized from linearized DNA templates by using SP6 RNA polymerase (Promega) and the mCAP mRNA Capping Kit (Stratagene Cloning Systems) according to the manufacturers' recommendations. The capped mRNAs were translated in wheat germ extracts (Promega) in the presence of [³H]leucine, [³H]glycine, or [³H]arginine (Amersham Corp.). The translation products were either directly analyzed on 20% low-molecular-weight polypeptide gels (Hoefer Scientific Instruments) or first subjected to radioimmunoprecipitation with the indicated antisera. Autoradiographic image enhancement was achieved by the use of Autofluor (National Diagnostics) according to the manufacturer's recommendations. The dried gels were exposed to Kodak XAR film at -70°C overnight.

Antiserum preparation and radioimmunoprecipitation assays. Synthetic peptides corresponding to portions of ORFs S1 and S2 were linked to keyhole limpet hemacyanin and injected into New Zealand White rabbits in Freund's complete adjuvant. The rabbits were boosted with the same antigen in Freund's incomplete adjuvant after 3 weeks to produce hyperimmune sera, which were harvested 2 weeks after the rabbits were boosted. The immunoglobulin G (IgG) fraction was partially purified from the sera by ammonium sulfate precipitation and DEAE-Sephadex column chromatography as previously described (5). The appropriate antiserum was reacted with radiolabeled in vitro translation products for radioimmunoprecipitation as previously described (2). The precipitated proteins were analyzed on discontinuous sodium dodecyl sulfate (SDS)-polyacrylamide gels as described above.

Automated protein sequencing. The EIAV ORF S1 protein was synthesized in wheat germ extracts programmed with pSP720 RNA in the presence of [³H]arginine, as described above. The pSP720 transcript encodes both S1 and S2, which are difficult to completely resolve on SDS gels. Because of the similarity in the gel mobilities of the S1 and S2 proteins, the S1 protein was immunoprecipitated from the translation mixture prior to gel electrophoresis. The ORF S2 protein was synthesized in the presence of [³H]glycine from pSP585 RNA, which lacks sequences required for S1 expression. The S1 and S2 proteins were fractionated on 15% discontinuous gels (14), which were placed directly into Autofluor for 30 min. The dried gels were exposed to Kodak XAR film and placed at -70°C overnight. With the autoradiograph as a template, the appropriate protein band was cut from the dried gel and rehydrated in a minimal volume of water. The gel slice was crushed, and the labeled proteins were eluted in 2 to 3 ml of Tris saline buffer (50 mM Tris-HCl [pH 7.5], 100 mM NaCl) at 37°C for 3 h with constant mixing. The gel fragments were removed by centrifugation, and the supernatant was lyophilized and resuspended in 100 µl of

1-mg/ml bovine serum albumin to act as a carrier protein. The protein mixture was desalted by passing it through a G-10 Sephadex column. Radioactive fractions were pooled, lyophilized, and resuspended in 50 to 100 μ l of doubly distilled deionized water. These samples were subjected to automated amino-terminal sequence analysis on an Applied Biosystems 470A Protein Sequencer. Radioactivity from each cycle of the sequencer was determined by directly adding the cycle eluate to 3 ml of Liquescent scintillation cocktail (National Diagnostics) and counting in a Beckman LS 6000IC liquid scintillation counter.

RESULTS

Nucleotide sequence analysis of cDNA clones generated from the 3.5-kb EIAV RNA. The splice site of the EIAV 3.5-kb RNA, which predominates in cytolytically infected FDD cells, was identified by cDNA cloning employing a 30-nt synthetic oligonucleotide primer complementary to a portion of the envelope RNA sequence 400 nt downstream of the putative envelope gene initiator codon (Fig. 1A). The position of this primer downstream of all near-consensus splice donor sequences (17, 23) should preclude synthesis of cDNAs corresponding to multiply spliced mRNAs and thus should allow synthesis only of cDNAs corresponding to the 8.2-kb full-length genomic RNA and the singly spliced 3.5-kb envelope RNA. Two cDNA clones, pSR-1 and pSR-2, which hybridized to both a 114-bp *Bam*HI-*Taq*I probe (nt 5337 to 5456) derived from the *env* gene and a 213-bp *Mlu*I-*Bam*HI probe (nt 156 to 386) derived from LTR and leader sequences, were isolated, indicating that these cDNAs had extended to the 5' terminus of the viral RNA. These two clones, found to be identical by dideoxy nucleotide sequencing (Fig. 1B), were produced from a spliced RNA, which utilized a splice donor at nt 459 (sd1) and a splice acceptor at nt 5135 (sa1). These cDNA clones correspond to an RNA that contains a majority of ORF S1, all of ORF S2, and the *env* gene (Fig. 1C). The splicing event removed four codons from the 5' end of ORF S1 as it was originally defined (26), that is, as the 150 nt of EIAV sequence from the termination codon of the *pol* gene to the next in-frame termination codon. However, S1 is extended an additional 38 codons in the amino-terminal direction through the viral leader region as a result of the splicing. These 38 codons combined with the 46 codons maintained from the original ORF S1 provide 84 codons of polypeptide coding potential. Interestingly, no AUG codon is found within this 84-codon sequence.

Northern blot analyses using splice donor-specific oligonucleotide probes. Although we have isolated two identical cDNA clones corresponding to a singly spliced subgenomic RNA, the possibility that other RNAs with similar mobilities in denaturing agarose gels comigrate with this identified 3.5-kb message still remains. We were particularly interested in examining whether two additional splice donors, sd2 (nt 512) and sd3 (nt 546), located near the 5' end of the *gag* gene, were functional in FDD cells. The utilization of sd2 could result in the first 16 codons of the *gag* gene, including the AUG initiator codon, being spliced to the acceptors at nt 5135 (sa1) or nt 7243 (sa3) to provide translational initiation signals for ORF S1 and ORF S3, respectively. RNA splicing from sd3 to these same acceptors would result in a splicing event that would place ORF S1 or ORF S3 out of frame with the *gag* coding sequence and would supply only two additional codons, neither of which is an AUG, to these ORFs. Northern blot analysis employing splice donor-specific oligonucleotide probes and nuclease S1 protection assays were

FIG. 1. Nucleotide sequence analysis of the splice junction of the predominant 3.5-kb RNA in EIAV-infected FDD cells. (A) Schematic representation of the EIAV proviral genome, indicating the position of potential splice donor (SD) and acceptor (SA) sites. The sequences surrounding the 5' and 3' splice sites of the 3.5-kb mRNA are shown. The position of the 30-nt primer used for first-strand cDNA synthesis is shown with a heavy line marked P. (B) Dideoxynucleotide sequencing gel of the cDNA clone pSR-1, derived from the singly spliced 3.5-kb transcript. Nucleotide sequences corresponding to the two exons are shown (E1 and E2), and the position of the splice site (SS) is indicated by an arrow. (C) Schematic representation of 3.5-kb singly spliced mRNA. This mRNA species contains the viral leader region spliced to ORF S1, which is followed by ORF S2 and the viral envelope gene. The cap structure is indicated by FmeG, and the poly(A) tail is indicated by (A)_n.

performed in order to determine the role of these putative splicing signals in EIAV gene expression.

By employing three synthetic oligonucleotide probes (SD-1, SD-2, and SD-3) that are complementary to the RNA sequence immediately upstream of each splice donor, we were able to ascertain whether these two alternative donors within the *gag* gene were utilized to generate spliced mRNAs (Fig. 2A). The SD-1 oligonucleotide should hybridize to spliced messages utilizing any of the three putative splice donors. The SD-2 oligonucleotide would be unable to hybridize to spliced mRNAs utilizing sd1 but would hybridize to those utilizing sd2 and sd3. Similarly, the SD-3 oligonucleotide could hybridize only to spliced mRNAs utilizing sd3 and not to those utilizing sd1 or sd2. All three oligonucleotide probes should hybridize to the full-length 8.2-kb genomic RNA. This fact was exploited as an internal control for hybridization efficiency of each of the three splice donor probes.

The results of Northern blot analyses of poly(A)⁺ RNA

FIG. 2. Northern hybridization analysis of EIAV-infected FDD cells using site-specific splice donor oligonucleotide probes. (A) Nucleotide sequence of the EIAV leader RNA coding strand from the *Sma*I site through the 5' end of the *gag* gene, illustrating the positions of the three potential splice donors (arrows). Nucleotides in each of the potential donor sites which conform to the consensus are underlined. The sequences of all of the synthetic splice donor-specific oligonucleotide probes (SD-1, SD-2, and SD-3) are displayed in boxes above their complementary sequences. The amino acid sequence of the p15 coding region of the *gag* gene is shown below the nucleotide codons. (B) Northern blot hybridization of poly(A)⁺ mRNA isolated from EIAV-infected FDD cells with each of the three splice donor probes. Lanes 1 to 3 correspond to RNA probed with radiolabeled oligonucleotides SD-1, SD-2, and SD-3, respectively.

isolated from EIAV-infected FDD cells probed with each of the three ³²P-labeled splice donor probes demonstrate that the 3.5-kb mRNA species can be detected only by the SD-1 probe (Fig. 2B, lane 1), which suggests that sd1 is the only splice donor utilized in this cell type to generate the singly spliced envelope message. The 8.2-kb genomic RNA hybridizes nearly equally well to each probe, suggesting consistency in the techniques employed and efficient labeling of the probes. The extremely low abundance of multiply spliced transcripts in this cell type (25) made it difficult to determine the potential usage of these alternative spliced donors in the generation of this class of RNAs. In an attempt to detect hybridization of the various probes to low-molecular-weight transcripts, the autoradiograph shown in Fig. 2 has been intentionally overexposed. This did not result in the detection of any clear signal corresponding to EIAV-specific transcripts of the 1.5-kb size class but rather resulted in the appearance of a diffuse nonspecific smear in this size range. In order to more clearly define the role of these three putative splice donors, we used the more-sensitive technique of nuclease S1 protection.

FIG. 3. Schematic representation of probes used in nuclease S1 mapping of transcripts. (A) Schematic diagram of the EIAV proviral genome, demonstrating the locations of the four DNA restriction endonuclease fragments (P1, P2, P3, and P4) used as probes for S1 nuclease mapping of potential splice donor (solid triangles) and acceptor (open triangles) sites. Solid bars in each probe represent EIAV sequences, while hatched bars indicate M13 sequences used as tags to distinguish input probe from probe that annealed to unspliced genomic EIAV RNA. Gene constructs and preparation of the probes are detailed in Materials and Methods. Restriction endonuclease abbreviations: N, *Nar*I; P, *Pvu*II; G, *Bgl*II; C, *Nco*I; B, *Bam*HI; H, *Hind*III; D, *Dra*I; S, *Sca*I. (B) Schematic representations of S1 nuclease mapping restriction endonuclease fragment probes P1 to P4. The locations of putative splice donor (SD) and acceptor (SA) sites are indicated. The length of the input probe (IP) and those portions of the probe that are expected to hybridize to unspliced (UP) and spliced (SP) EIAV transcripts are indicated. The asterisk represents the radiolabeled end of each input probe.

Nuclease S1 protection analysis of spliced EIAV mRNAs. The utilization of the various near-consensus splice donor and acceptor sites (17, 23) identified within the EIAV genomic RNA by nucleotide sequence analysis was examined by nuclease S1 protection assays of total cellular RNA isolated from EIAV-infected FDD cells. ³²P-labeled restriction fragments of EIAV proviral DNA were used as probes in these assays (Fig. 3).

Analysis of the splice donors at the 5' end of the genome was accomplished by using a 5'-end-labeled 308-bp *Taq*I-*Pvu*II proviral restriction fragment (nt 309 to 620) with an M13 vector tag sequence at its 3' end (Fig. 3B, P1). When convenient, the probes were constructed with extraneous vector sequences to allow for distinction between input probe and probe annealed to full-length genomic RNA. Hybridization of P1 to spliced mRNA species utilizing sd1, sd2, or sd3 would yield protected fragments of 130, 183, or 217 nt, respectively. The results indicate that only a 300- and a 130-nt probe fragment were protected from S1 nuclease

FIG. 4. Analysis of digestion products from nuclease S1 protection studies. Gels A to D correspond to results of S1 nuclease protection assays of total RNA isolated from EIAV-infected (lanes 1) or uninfected (lanes 2) FDD cells with probes P1 to P4, respectively. Molecular weight markers (lanes M) are 32 P-labeled DNA restriction endonuclease fragments from an *MspI* digest of pBR322.

digestion (Fig. 4A, lane 1). These fragments correspond to probe annealed to full-length EIAV genomic RNA and spliced RNA utilizing sd1, respectively. No protected fragments corresponding to sd2 and sd3 were detected, suggesting that these sequences are not utilized as splice donor sites in EIAV-infected FDD cells. This result is in agreement with that of the Northern blot experiment employing the differential splice donor oligonucleotide probes described above. That is, in FDD cells, apparently only the first splice donor in the leader region is utilized to generate singly or multiply spliced mRNAs.

The splice acceptor sites in the *pol-env* intergenic region were mapped with a 3'-end-labeled 450-bp *NcoI-BamHI* EIAV restriction fragment (nt 4889 to 5337) with 161 bp of M13mp19 sequence tagged to its 5' end (Fig. 3B, P2). This probe was designed to investigate the utilization of the two near-consensus splice acceptor sites located near the 5' ends of ORF S1, sa1 (nt 5135), and sa2 (nt 5183). We have already shown by cDNA cloning and nucleotide sequence analysis of cDNA clones pSR-1 and pSR-2 that sa1 is functional. Probe P2 will allow us to determine whether the second splice acceptor is utilized and what the relative distribution of spliced transcripts employing either acceptor is. Protection

of P2 from S1 nuclease digestion by hybridization to full-length genomic DNA is expected to protect a 450-nt portion of the input probe, while hybridization to spliced mRNAs utilizing sa1 or sa2 would be expected to yield protected fragments of 206 and 157 nt, respectively (Fig. 3B, P2).

The experimental results clearly demonstrate protected fragments representative of all three classes of mRNAs, that is, unspliced genomic and spliced RNAs employing either sa1 or sa2 (Fig. 4B, lane 1). Although equivalent amounts of total cellular RNA were used in each experiment, differences in the autoradiograph signals of the protected fragments for different probes may vary depending on the efficiency of probe labeling and hybridization and on autoradiograph exposure time. However, within a given experiment, it is possible to draw quantitative conclusions about the relative abundance of transcripts employing the individual splice sites. The intensity of the probe fragment protected by mRNAs utilizing sa1 is much greater than it is with mRNAs employing sa2, suggesting that mRNAs spliced at sa2 represent a minor population in EIAV-infected FDD cells. Furthermore, the sa1-protected probe fragment is considerably more intense than that protected by the full-length genomic RNA, which is in agreement with our previously reported result that the 3.5-kb mRNA species is the predominant viral RNA in EIAV-infected FDD cells (25).

The third probe (Fig. 3B, P3) was designed to determine the role of potential splice donor sites within the *pol-env* intergenic region in the formation of multiply spliced mRNA transcripts. This probe was generated from a 617-bp 5'-end-labeled *PvuII-HindIII* EIAV proviral restriction fragment (nt 5162 to 5775). There are four near-consensus splice donors localized from the 3' end of ORF S1 to the 3' end of ORF S2 at nt 5276, 5298, 5306, and 5435 (Fig. 3A). RNA species utilizing these donor sites would be expected to yield protected fragments of 114, 136, 144, and 273 nt, respectively. In fact, only the 144-nt protected probe fragment could be detected (Fig. 4C, lane 1), suggesting that only the splice donor sequence localized to nt 5306 is utilized in FDD cells for the generation of multiply spliced mRNAs. This finding was somewhat surprising, in that multiply spliced ORF S1-ORF S3 RNAs isolated from canine fibroblasts were shown to involve splicing from the donor site immediately downstream of ORF S1 at nt 5276 (30). We were unable to detect a protected fragment 114 nt in length corresponding to a similar splicing event in FDD cells. It should be noted that the design of probe P3 from the *PvuII* site prevents detection of multiply spliced mRNA species, which are produced by splicing events involving sa2, since the labeled portion of the probe would be unable to hybridize to such messages and would therefore be digested by nuclease S1. Therefore, we cannot conclude that the other splice donor sequences within this region are not utilized to generate multiply spliced messages. Further study is required to determine the role of sa2 in both singly and multiply spliced RNAs.

The final S1 protection probe (P4) was designed to investigate the potential usage of a splice acceptor site at the 5' end of ORF S3 (sa3, nt 7234), which may result in the provision of an AUG translational initiation codon for the protein product encoded by this ORF. Probe P4 corresponds to a 3'-end-labeled-385 bp *DraI-ScaI* proviral restriction fragment (nt 7026 to 7412). This probe is expected to yield a 385-nt fragment when annealed to full-length genomic RNA and a 178-nt fragment when hybridized to spliced mRNA utilizing sa3. The S1 mapping results show the presence of the 178-nt protected fragment, indicative of usage of sa3.

The nuclease S1 protection experiments described herein

demonstrate the utilization of a single splice donor site that provides the EIAV leader sequence to all spliced RNAs. Both splice acceptor sites within the 5' end of ORF S1 are functional; however, there is a strong bias towards splicing at the first site in FDD cells. Of the four splice donors located in the *pol-env* intergenic region, only the third (nt 5306) was used in FDD cells in these experiments. Finally, the splice acceptor site near the 5' border of ORF S3 is used to produce presumptive multiply spliced RNAs.

Assays of *trans*-activation activity. As previously stated, the pSR-1 cDNA clone we have characterized, which corresponds to the singly spliced 3.5-kb EIAV mRNA, contains the ORF S1, ORF S2, and *env* genes (Fig. 1C). Since ORF S1 encodes the viral *trans* activator EIAV Tat (8, 20), we were interested to ascertain whether the pSR-1 cDNA clone could direct the synthesis of the Tat protein in cultured FDD cells. Towards this end, restriction fragments of the pSR-1 cDNA sequence were subcloned into a eucaryotic expression plasmid, pSV2Δcat. Two subclones, pSV720 and pSV585, which differ in that pSV585 lacks cDNA sequences between the *Stu*I (nt 256) restriction site located in the terminal redundant (R) region and the *Sma*I (nt 393) restriction site located in the leader region, were constructed (Fig. 5A). The CUG codon, which has been proposed to initiate EIAV Tat synthesis, lies between these two restriction endonuclease cleavage sites at nt 373 (8, 30). The pSV720 and pSV585 expression clones are 3' coterminal, extending to the *Msc*I (nt 5655) restriction site within the *env* gene. If indeed EIAV Tat synthesis is initiated at the suggested CUG codon, then pSV720 would contain all the necessary genetic information for EIAV Tat expression, while pSV585 would lack the translation initiation codon for this protein. FDD cell monolayers were cotransfected with either pSV720 or pSV585 and a plasmid containing the CAT reporter gene under the control of the EIAV LTR, pLTRcat (25). Cotransfection of the pSV720 expression plasmid with pLTRcat clearly results in *trans* activation of the EIAV LTR, as evidenced by an increase in CAT enzyme activity of more than 60-fold compared with that resulting from transfection with the pLTRcat plasmid alone (Fig. 5B). Cotransfection of pSV585 resulted in no significant *trans* activation of the LTR, suggesting that sequences residing between the *Stu*I site within the EIAV R region of the LTR and the *Sma*I site within the leader, which include the suggested CUG translational initiation codon, are absolutely required for expression of EIAV Tat from this cDNA clone in cultured FDD cells.

In vitro expression of ORF S1, ORF S2, and *env* genes from the 3.5-kb EIAV mRNA. The translational coding potential of the pSR-1 cDNA clone derived from the EIAV 3.5-kb transcript was investigated by employing in vitro transcription and translation techniques. The presence of three ORFs in this cDNA suggested the possibility that the parent 3.5-kb mRNA is tricistronic in nature, i.e., capable of encoding EIAV Tat, the ORF S2 protein, and the gp135 envelope glycoprotein. The SP6-based in vitro transcription plasmids pSP720 and pSP585 are analogous to the SV40-based eucaryotic expression plasmids pSV720 and pSV585 used in the *trans*-activation assays described above. That is, these in vitro transcription plasmids differ only in that pSP720 contains pSR-1 cDNA sequences starting from the *Stu*I (nt 356) restriction site in the R region of the LTR, while pSP585 contains cDNA sequences from the *Sma*I (nt 393) restriction site in the viral leader region. The cDNA sequences of both plasmids are 3' coterminal, ending at the *Msc*I (nt 5655) restriction site located within the gp90 coding region of the *env* gene.

FIG. 5. Analysis of *trans*-activated EIAV LTR-driven CAT enzyme activity in transfected FDD cells. (A) Schematic representation of cDNA from pSR-1 corresponding to the singly spliced 3.5-kb EIAV mRNA. The positions of relevant restriction endonuclease sites for subcloning and the presumptive CUG translational initiation codon of the S1 EIAV Tat protein are indicated. Eucaryotic expression plasmids pSV720 and pSV585 were generated by subcloning either a 720-bp *Stu*I-*Msc*I or a 585-bp *Sma*I-*Msc*I DNA restriction fragment into *Stu*I-digested pSV2Δcat, respectively. This expression vector provides the SV40 immediate-early promoter (SV40 Pr) and polyadenylation signal (Poly A) for expression of cloned inserts in eucaryotic cells. (B) Graphical representation of the conversion of [¹⁴C]butyryl coenzyme A to [¹⁴C]butyryl chloramphenicol by incubation with transfected FDD cell lysates as a function of time. A control plasmid, pSVβ-gal, was included in all transfections so that cell lysates could be normalized to β-galactosidase activity to control for transfection efficiency. Solid squares, pLTRcat alone; solid triangles, pSV2cat; solid circles, pLTRcat cotransfected with pSV720; open circles, pLTRcat cotransfected with pSV585.

The pSP720 and pSP585 plasmid DNAs were employed as transcription templates to produce synthetic capped RNAs, which were translated in wheat germ extracts in the presence of [³H]leucine. In vitro translation of pSP720 RNA resulted in the synthesis of three major polypeptide products with relative molecular sizes of 15, 8.5, and 7 kDa (Fig. 6A, lanes 1 and 8). These products correspond closely to the predicted sizes of *Msc*I-truncated gp90 envelope protein, the EIAV Tat protein, and the putative ORF S2 protein, respectively. The pSP585 RNA supported translation of only two of these products, the 15- and 7-kDa polypeptides (Fig. 6A, lane 5). The 15-kDa product from both pSP585 and pSP720 RNA-programmed translation reactions was specifically immunoprecipitated with a monoclonal antibody directed against an epitope of the gp90 envelope protein mapping to a region

FIG. 6. In vitro translation and radioimmunoprecipitation analysis of pSP720 and pSP585 RNAs. (A) Fluorograph of an SDS-PAGE analysis of wheat germ extracts programmed with pSP720 (lanes 1 and 8) and pSP585 (lane 5) synthetic transcripts in the presence of [3 H]leucine. Radioimmunoprecipitation of labeled translation products was done with a monoclonal antibody directed against the gp90 envelope glycoprotein (lanes 2 and 6) and rabbit polyclonal antiserum specific to the S1 (lane 3) or S2 (lanes 4, 7, and 9) protein products. (B) Immunoprecipitation of the S2 protein with EIAV-infected horse sera. In vitro transcripts produced from pSP585 were translated in wheat germ extracts in the presence of [35 S]methionine (lane 1) and immunoprecipitated with either 4 or 20 μ g (lanes 2 and 3, respectively) of partially purified IgG from a pony naturally infected with EIAV. Molecular size markers (lanes M) are BRL 14 C-labeled low-molecular-mass standards.

approximately 50 amino acids from its amino terminus (3) (Fig. 6A, lanes 2 and 6). The 8.5-kDa band specific for pSP720 RNA translation could be immunoprecipitated by a rabbit polyclonal antiserum produced against a synthetic peptide corresponding to a predicted immunoreactive domain of the EIAV Tat protein (Fig. 6A, lane 3). The 7-kDa protein common to both pSP720 and pSP585 RNA translation reactions was immunoprecipitated by a combination of low-titer rabbit polyclonal antisera produced against four different overlapping synthetic peptides that span the entire ORF S2 amino acid sequence (Fig. 6A, lanes 4 and 7). This combination of ORF S2 antisera cross-reacted with the EIAV Tat protein product (Fig. 6A, lane 4). However, a high-titer ORF S2 antiserum obtained more recently demonstrated no cross-reactivity with EIAV Tat (Fig. 6A, lane 9).

This ORF S2 antiserum also specifically precipitates two minor products with mobilities of 5.5 and 4 kDa, which are presumed to be either S2-related peptides produced by translational initiation at the internal AUG codons of S2 or degradation products of S2. These data indicate that a single cDNA species corresponding to the 3.5-kb singly spliced mRNA of EIAV is capable of directing the synthesis of the EIAV Tat, the S2 protein, and the envelope protein in vitro. Furthermore, since no EIAV Tat protein product is detected upon translation of pSP585 RNA, sequences located between the *Stu*I site in the R region and the *Sma*I site in the leader, which include the predicted CUG initiation codon of EIAV Tat, are required for in vitro expression of the EIAV Tat protein.

Humoral immune response to the ORF S2 protein in EIAV-infected horses. The predicted protein product of ORF S2 bears no significant sequence homology to any known lentiviral protein, and to date, no protein product has been shown to be encoded by ORF S2 in vivo. In an effort to establish whether an ORF S2-encoded polypeptide is produced during productive EIAV infection, we have examined sera from EIAV-infected horses for the presence of antibodies to the putative ORF S2 gene product. We have accomplished this goal by examining the abilities of these sera to recognize the in vitro translated product in an immunoprecipitation reaction and by testing the reactivity of a panel of horse sera to ORF S2-specific synthetic peptides in ELISAs.

The ORF S2 polypeptide and a portion of the gp90 envelope protein were synthesized in wheat germ extract from a synthetic RNA similar to that produced by the pSP585 plasmid described above. The transcription template in this experiment differed from pSP585 in that it included sequences from the *Msc*I (nt 5655) to the first *Hind*III (nt 5775) restriction endonuclease digestion sites in the *env* gene. This additional envelope sequence results in the synthesis of a truncated gp90 protein with an apparent molecular size of approximately 20 kDa on SDS-polyacrylamide gels. The 35 S-labeled in vitro translation products were immunoprecipitated with sera obtained from an EIAV-infected horse followed by SDS-polyacrylamide gel electrophoresis (PAGE) analysis. Analysis of the in vitro translation products shows the synthesis of the 20-kDa truncated envelope protein and the 7-kDa ORF S2 polypeptide (Fig. 6B, lane 1). These labeled products were immunoprecipitated with 4 or 20 μ g of partially purified IgG obtained from the serum of an EIAV-infected horse (Fig. 6B, lanes 2 and 3, respectively). As one would anticipate, the truncated gp90 envelope protein reacts strongly with the EIAV-infected horse IgG fraction, particularly at the higher concentration (Fig. 6B, lane 3). The ORF S2 protein was clearly immunoprecipitated by the EIAV-infected horse IgG at the higher concentration (lane 3) and faintly precipitated at the lower concentration (Fig. 6B, lane 2). Neither the truncated gp90 nor the ORF S2 polypeptide demonstrated any reactivity with control sera from normal uninfected horses (data not shown). Furthermore, ELISAs of a standard panel of horse immune sera with ORF S2-specific synthetic peptides showed that 60 to 75% of the horses produced antibodies that reacted with the peptides (data not shown). These data indicate that the ORF S2 protein is expressed during productive EIAV infection, since EIAV-infected horses are capable of mounting humoral immune responses directed against this protein.

Amino-terminal sequencing of the in vitro-synthesized EIAV Tat and ORF S2 polypeptides. In an attempt to gain a better understanding of the expression of the ORF S1, ORF S2, and envelope proteins from the tricistronic 3.5-kb RNA,

FIG. 7. Amino-terminal sequence analysis of the S2 and S1 proteins. In vitro-translated S2 protein produced in the presence of [³H]glycine in wheat germ extracts programmed with pSP585 RNA was purified from an SDS-polyacrylamide gel and subjected to amino-terminal sequencing. Amino acid fractions in 100% acetonitrile were directly placed in Liquiscint liquid scintillation solvent and assayed for radioactivity. The S1 protein was prepared in a similar manner by translation of pSP720 in vitro transcripts in the presence of [³H]arginine. The amino acid sequences of the predicted S2 and S1 polypeptides are aligned beneath the peaks of radioactivity assayed from the automated Sequenator.

we determined the presumptive translation initiation sites of the in vitro-synthesized ORF S1 and ORF S2 proteins by amino-terminal sequencing of radiolabeled polypeptides. The radioactive amino acid of choice for sequencing of the ORF S2 protein was [³H]glycine, since the first methionine codon in ORF S2 is followed immediately by a glycine codon and since three of the first seven codons correspond to glycine residues. The ORF S1 protein was labeled with [³H]arginine, since translational initiation at the proposed CUG codon (8, 30) would place two contiguous arginine residues near the amino terminus.

The ORF S2 product was synthesized from pSP585 RNA, and approximately 7×10^4 cpm of gel-purified protein was subjected to amino-terminal sequence analysis. The sequencer fraction corresponding to the first amino acid residue registered in excess of 7×10^3 cpm, while further peaks of radioactivity were observed for amino acid residues 4 and 6 (Fig. 7A). The anticipated amino acid sequence of the S2 polypeptide initiating at the first methionine codon in this ORF is M G L F G K G. Assuming that the amino peptidase activity of the wheat germ extract has cleaved the initiator methionine residue from this protein, the peaks of radioactivity occur precisely as expected, that is, in fractions 1, 4, and 6. Additional peaks of radioactivity were noted in fractions 16, 17, and 20, which are also predicted sites of glycine residues in the ORF S2 protein. These data confirm that the in vitro-synthesized polypeptide assigned to be encoded by ORF S2 on the basis of its ability to specifically react with ORF S2-specific peptide antisera is indeed the ORF S2 gene product and that its synthesis in vitro is apparently initiated at the first methionine residue in this ORF.

The sequence of the ORF S1 protein, EIAV Tat, was of particular interest, since it has been suggested that this protein may initiate translation at a CUG codon. Of the two CUG codons located within the leader region, the one more downstream (nt 373) lies in the most favorable sequence context for translation initiation (12). Initiation of EIAV Tat at this CUG codon would place two consecutive arginine codons near its amino terminus. The EIAV Tat protein was immunoprecipitated from pSP720 RNA-directed in vitro

translation products labeled with [³H]arginine and subjected to amino-terminal sequencing. Isolation by immunoprecipitation was necessary to purify EIAV Tat from the similarly sized ORF S2 protein as well as from a number of comigrating wheat germ proteins, which become amino terminally modified with labeled arginine by an arginine-terminal transferase activity found in wheat germ extracts (9). Although we were able to purify relatively large amounts of [³H]arginine-labeled EIAV Tat protein, initial amino-terminal sequence analysis attempts resulted in no distinctive sequence information, suggesting modification of the amino-terminal amino acid residue that prevented the Edman degradation reaction. To overcome this problem, we carried out translation for only a short time (30 min) and immediately boiled the translation reaction mixture after the incubation, anticipating that such treatment could yield a fraction of the synthesized S1 protein in an unmodified state. Amino-terminal sequencing of 5×10^4 cpm of EIAV Tat protein prepared in this manner resulted in two clear consecutive peaks of radioactivity, each of approximately 2×10^2 cpm, corresponding to the third and fourth residues of the polypeptide (Fig. 7B). The positions of these peaks are consistent with translational initiation at the previously proposed CUG codon, presuming amino peptidase processing of the initiator methionine residue. On the basis of these findings, we propose that the amino terminus of the in vitro-synthesized EIAV Tat protein is blocked by an unknown functional group and that limited translation time followed by heat treatment allowed us to capture a fraction of the newly synthesized Tat protein in an unblocked form.

DISCUSSION

EIAV RNA splicing in FDD cells. Although a complex pattern of RNA splicing has been observed in cultured canine fibroblasts persistently infected with avirulent EIAV (21, 32), considerably less complex patterns have been observed in cells of equine origin. In primary horse macrophage cultures infected with the horse-virulent Wyoming strain of EIAV, only the 8.2-kb full-length genomic and 3.5-kb singly spliced envelope RNAs were detected (30). We

have previously reported similar results for primary equine fibroblasts (FDD) infected with avirulent cell culture-adapted EIAV (25). In this report, we have analyzed the EIAV splicing pattern in FDD cells and examined the coding potential of the major 3.5-kb EIAV transcript produced in FDD cells.

The 3.5-kb EIAV transcript(s) of FDD cells can be produced by splicing of the 243-nt leader region from a unique splice donor (sd1, nt 459) to either of two functional splice acceptors (sa1 [nt 5135] and sa2 [5184]) localized within ORF S1. The S1 nuclease protection studies described here indicate that the first of these two acceptor sites is preferentially utilized in FDD cells. Nucleotide sequence analysis of cDNA clones clearly demonstrate the involvement of this splicing event in the production of the major 3.5-kb RNA of FDD cells. This splicing event is identical to the first of three splicing reactions in a previously defined ORF S1-ORF S3 RNA, which was shown to produce functional *trans*-activator protein encoded by ORF S1 (22, 30). Since the major 3.5-kb singly spliced RNA in FDD cells contains the same ORF S1 sequences as this previously identified multiply spliced message, it follows that the 3.5-kb transcript may also be able to produce the EIAV Tat *trans*-activator protein encoded by ORF S1. In addition to ORF S1, this 3.5-kb RNA also includes the overlapping ORF S2 and *env* genes.

A second class of 3.5-kb transcripts can be produced by splicing the leader to the second acceptor located within ORF S1 (sa2 [nt 5184]). RNAs of this class would not be expected to produce functional EIAV Tat protein, since sequences shown to be critical for *trans* activation reside between the two alternative splice acceptor sites (8) and therefore would not be present in RNAs spliced at sa2. Such RNAs would, however, contain the entire ORF S2 and *env* genes. It appears highly unlikely that any envelope RNA which does not also contain ORF S2 will be detected, since there are only 26 nt separating the presumptive AUG initiation codons of S2 and the gp135 envelope gene product, and no reasonable splice acceptor sequences lie in this region. Indeed, we were unable to detect in FDD cells any functional splice acceptor sites that map to this region by S1 nuclease protection analysis. Therefore, singly spliced 3.5-kb RNAs involving sa2 may be bicistronic in nature, encoding the S2 and gp135 proteins in a manner similar to that by which the HIV-1 Vpu and gp160 proteins are encoded (28).

The possibility remains that either of the identified splice acceptor sites (sa1 and sa2) could be involved in the production of multiply spliced transcripts. Of the four putative splice donor sequences in the *pol-env* intergenic region, only one, located between the ORF S2 and the *env* AUG initiation codons, was detected in these studies to be functional in FDD cells. The splice acceptor site (sa3) near the 5' end of ORF S3 is also functional in these cells, suggesting the possibility that the ORF S2 AUG is spliced to ORF S3, thus providing a translational initiation for S3 expression. However, direct splicing of the splice donor within ORF S2 (nt 5306) to the ORF S3 acceptor site (nt 7235) would not result in proper alignment of the S2 AUG with the S3 coding frame. Only two additional codons, neither of which is an AUG, would be placed in frame with S3. The third exon of the triply spliced ORF S1-ORF S3 RNA identified in EIAV-infected canine cells consists of a short stretch of sequence from the *env* gene which appears to provide a translational initiation codon to ORF S3 (30). It remains possible that an exon within the *env* coding region provides a stretch of

amino acid-coding potential allowing for proper alignment of S3 with the S2 AUG.

Translational potential of the major 3.5-kb transcript. The major 3.5-kb EIAV transcript of FDD cells is shown here to be tricistronic *in vitro*, encoding the protein products of the ORF S1, ORF S2, and *env* genes in the wheat germ cell-free translation system. Amino-terminal sequence analysis of *in vitro*-synthesized EIAV Tat is consistent with initiation at a CUG codon (nt 373), which lies in an otherwise favorable sequence context for translational initiation, AAC CUG G, in which underlined nucleotides match the consensus sequence for translation initiation (12). We have shown EIAV Tat synthesis from cDNA sequences corresponding to the 5' end of the 3.5-kb transcript *in vivo*, as evidenced by a 60-fold increase in CAT expression directed by the EIAV LTR. The relatively low levels of *trans* activation observed for EIAV compared with those reported for HIV may reflect differences in the amount of *trans*-activator protein produced as a result of initiation at the suboptimal CUG codon.

Serum samples from EIAV-infected horses are immunoreactive with both *in vitro*-synthesized S2 protein and synthetic peptides corresponding to portions of S2. Characterization of viral envelope variants propagated from horse macrophage cultures infected with horse-virulent EIAV has shown that although a number of mutations arise throughout the gp90-coding sequence and in the major coding exon for the S3 protein, no mutations occur in ORF S2 (1). Taken together, these data strongly indicate a role for the ORF S2 gene product during EIAV infection, although no function for this protein has yet been defined.

Although a number of singly spliced HIV-1 transcripts contain multiple ORFs (28, 29, 31), only the *vpu-env* (28) transcripts are known to act as multicistronic transcripts *in vivo*. The HIV-1 transcript encoding the 72-amino-acid form of the Tat protein, Tat-1, is structurally similar to the major EIAV 3.5-kb transcript described here in that it contains the *tat-1*, *vpu*, and *env* genes in positions nearly identical to those of the EIAV *tat*, ORF S2, and *env* genes (28). This HIV transcript has been shown to be monocistronic *in vivo*, producing only the Tat-1 protein, presumably because of the relatively strong translation initiation codon of the Tat protein (GAA AUG G). Similarly, infrequent translation initiation of EIAV Tat from an inefficient non-AUG codon and of the S2 protein from an AUG codon in a relatively weak sequence context (UAU AUG G) may allow significant leaky ribosomal scanning (13) to the highly favorable gp135 AUG (AAC AUG G) of the major 3.5-kb transcript (sd1-sa1).

During the preparation of this paper, a report which seems to suggest translational initiation of EIAV Tat in transfected canine fibroblasts from an AUC codon (nt 388) upon deletion of sequences including the CUG (nt 373) initiation codon identified here was published (22). This result is somewhat surprising, since this AUC codon (AGG AUC C) lies in a less favorable context for translation initiation than does the CUG codon (AAC CUG G). It is possible that deletion of the more favorable CUG could allow synthesis at the next most favorable codon, the AUC. Clearly, future studies involving site-specific mutagenesis of each codon and its surrounding sequence are required to determine which is the actual translation initiator in EIAV-infected cells.

ACKNOWLEDGMENTS

We thank C. Issel for providing equine fibroblast cells and virus. We thank J. Ball and M. Miller for peptide synthesis, peptide ELISAs, and preparation of some of the antisera used in these studies.

This research was supported in part by funds provided by the Louisiana Agricultural Experiment Station and Public Health Service grant CA49296.

REFERENCES

- Alexandersen, S., and S. Carpenter. 1991. Characterization of variable regions in the envelope and S3 open reading frame of equine infectious anemia virus. *J. Virol.* **65**:4255-4262.
- Anderson, D. J., and G. Blobel. 1983. Immunoprecipitation of proteins from cell-free translations. *Methods Enzymol.* **96**:111-120.
- Ball, J. M. 1990. Ph.D. thesis. Louisiana State University, Baton Rouge.
- Chirgwin, J. M., A. E. Przybyl, R. J. MacDonald, and W. J. Rutter. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* **18**:5249-5299.
- Chua, N.-H., S. G. Bartlett, and M. Weiss. 1982. Preparation and characterization of antibodies to chloroplast proteins, p. 1063-1080. *In* M. Edelman, R. B. Hallick, and N.-H. Chua (ed.), *Methods in chloroplast molecular biology*. Elsevier Biomedical Press, Amsterdam.
- Cullen, B. R. 1991. Human immunodeficiency virus as a prototypic complex retrovirus. *J. Virol.* **65**:1053-1056.
- Derse, D., P. L. Dorn, L. Levy, R. M. Stephens, N. R. Rice, and J. W. Casey. 1987. Characterization of the equine infectious anemia virus long terminal repeat. *J. Virol.* **61**:743-747.
- Dorn, P., L. DaSilva, L. Martrano, 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.
- Elias, S., and A. Ciechanover. 1990. Post-translational addition of an arginine moiety to acidic NH₂ termini of proteins is required for their recognition by ubiquitin-protein ligase. *J. Biol. Chem.* **265**:15511-15517.
- Fordis, M., and B. H. Howard. 1987. Use of the CAT reporter gene for the optimization of gene transfer into eucaryotic cells. *Methods Enzymol.* **151**:382-397.
- Gorman, C. M., L. F. Moffat, and B. H. Howard. 1982. Recombinant genomes which express chloramphenicol acetyl transferase in mammalian cells. *Mol. Cell. Biol.* **2**:1044-1051.
- Kozak, M. 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**:283-292.
- Kozak, M. 1989. The scanning model for translation: an update. *J. Cell Biol.* **108**:229-241.
- Laemmli, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature (London)* **227**:680-685.
- Malmquist, W. A., D. Barnett, and C. S. Becvar. 1973. Production of equine infectious anemia antigen in a persistently infected cell line. *Arch. Virol.* **42**:361-370.
- Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Mount, S. M. 1982. A catalogue of splice junction sequences. *Nucleic Acids Res.* **10**:459-472.
- Narayan, O., and J. E. Clements. 1990. Lentiviruses, p. 1571-1589. *In* B. N. Fields and D. M. Knipe (ed.), *Virology*, 2nd ed. Raven Press, New York.
- Neumann, J. R., C. A. Morency, and K. O. Russian. 1987. A novel rapid assay for chloramphenicol acetyltransferase gene expression. *BioTechniques* **5**:444-447.
- Noiman, S., A. Gazit, O. Tori, L. Sherman, T. Miki, S. R. Tronick, and A. Yaniv. 1990. Identification of sequences encoding the equine infectious anemia virus *tat* gene. *Virology* **176**:280-288.
- Noiman, S., A. Yaniv, L. Sherman, S. R. Tronick, and A. Gazit. 1990. Pattern of transcription of the genome of equine infectious anemia virus. *J. Virol.* **64**:1839-1843.
- Noiman, S., A. Yaniv, T. Tsach, T. Miki, S. Tronick, and A. Gazit. 1991. The Tat protein of equine infectious anemia virus is encoded by at least three types of transcripts. *Virology* **184**:521-530.
- Ohshima, Y., and Y. Gotoh. 1987. Signals for the selection of a splice site in pre-mRNA. Computer analysis of splice junction sequences and like sequences. *J. Mol. Biol.* **195**:247-259.
- Orrego, A., C. J. Issel, R. C. Montelaro, and W. V. Adams, Jr. 1982. Virulence and *in vitro* growth of a cell adapted strain of equine infectious anemia virus after serial passage in ponies. *Am. J. Vet. Res.* **43**:1556-1560.
- 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 lytic infection. *J. Virol.* **64**:86-95.
- Rushlow, K. E., 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.
- Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463-5467.
- Schwartz, S., B. K. Felber, E. Fenyó, and G. N. Pavlakis. 1990. Env and Vpu proteins of human immunodeficiency virus are produced from multiple bicistronic mRNAs. *J. Virol.* **64**:5448-5456.
- Schwartz, S., B. K. Felber, and G. N. Pavlakis. 1991. Expression of human immunodeficiency virus type 1 vif and vpr mRNAs is rev-dependent and regulated by splicing. *Virology* **183**:677-686.
- Stephens, R. M., D. Derse, and N. R. Rice. 1990. Cloning and characterization of cDNAs encoding equine infectious anemia virus Tat and putative Rev proteins. *J. Virol.* **64**:3716-3725.
- Vaishnav, Y. N., and F. Wong-Staal. 1991. The biochemistry of AIDS. *Annu. Rev. Biochem.* **60**:577-630.
- Yaniv, A., L. Sherman, S. Noiman, O. Tori, H. Lichtman-Pleban, T. Miki, S. R. Tronick, and A. Gazit. 1989. Studies on the regulation and patterns of expression of the equine infectious anemia virus genome, p. 59-73. *In* D. Gaudy and W. Hennessen (ed.), *Developments in biological standardization: progress in animal retroviruses*, vol. 72. S. Karger AG, Basel.