Characterization of the Transcriptional Properties of Equine Infectious Anemia Virus.

Siyamak Rasty
Louisiana State University and Agricultural & Mechanical College

Follow this and additional works at: https://repository.lsu.edu/gradschool_disstheses

Recommended Citation
https://repository.lsu.edu/gradschool_disstheses/4871

This Dissertation is brought to you for free and open access by the Graduate School at LSU Scholarly Repository. It has been accepted for inclusion in LSU Historical Dissertations and Theses by an authorized administrator of LSU Scholarly Repository. For more information, please contact gradetd@lsu.edu.
INFORMATION TO USERS

The most advanced technology has been used to photograph and reproduce this manuscript from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.
Characterization of the transcriptional properties of equine infectious anemia virus

Rasty, Siyamak, Ph.D.
The Louisiana State University and Agricultural and Mechanical Col., 1989

Copyright ©1990 by Rasty, Siyamak. All rights reserved.
CHARACTERIZATION OF THE TRANSCRIPTIONAL PROPERTIES OF EQUINE INFECTIOUS ANEMIA VIRUS

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

in

The Department of Biochemistry

by

Siyamak Rasty
B.S., Louisiana State University, 1984
December 1989
ACKNOWLEDGMENTS

I would like to extend sincere appreciation to my major professor, Dr. R. C. Montelaro, for his guidance, counsel, encouragement, and supervision during the span of this research and the preparation of this dissertation.

I would also like to express my sincere thanks to Dr. K. E. Rushlow for helpful discussions and suggestions throughout the course of this research, for generous gifts of DNA subclones and various other reagents, and for taking the time to critically review this manuscript.

Special thanks are due to the members of my graduate advisory committee for taking the time to serve on the committee.

I am sincerely grateful to Drs. K. G. Kousoulas and A. Baghian of the LSU Department of Veterinary Microbiology and Parasitology for introducing me to the COS cell expression system and for allowing me to conduct experiments in their laboratory.

I am indebted to M. A. Miller for preparing the computer generated drawings.

I would like to deeply thank my parents, Jamileh and Hossein Ali Rasty, and my brother, Jahangir Rasty, for their endless love, support, understanding, and encouragement. Finally, I am thankful to Faranak Naghavi for her patience, love, and constant words of encouragement throughout the writing of this dissertation.
FOREWORD

## TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>ii</td>
</tr>
<tr>
<td>FOREWORD</td>
<td>iii</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>vi</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>ix</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>MATERIALS AND METHODS</td>
<td>44</td>
</tr>
<tr>
<td>RESULTS</td>
<td>57</td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>123</td>
</tr>
<tr>
<td>LITERATURE CITED</td>
<td>138</td>
</tr>
<tr>
<td>VITA</td>
<td>153</td>
</tr>
</tbody>
</table>
**LIST OF TABLES**

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Taxonomic characteristics common to retroviruses</td>
<td>9</td>
</tr>
<tr>
<td>II</td>
<td>Panel of EIAV proviral DNA restriction fragments used as probes in Northern hybridization analyses</td>
<td>63</td>
</tr>
<tr>
<td>III</td>
<td>Nucleotide sequence of the putative splice donor and acceptor sites located immediately upstream of or within the EIAV env gene</td>
<td>91</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Schematic representation of the different stages of EIA in infected horses</td>
<td>3</td>
</tr>
<tr>
<td>2. A simplified schematic representation of the various stages involved in reverse transcription</td>
<td>18</td>
</tr>
<tr>
<td>3. The molecular events involved in the retroviral life cycle</td>
<td>20</td>
</tr>
<tr>
<td>4. Anatomy of a genomic RNA subunit of a replication-competent retrovirus</td>
<td>23</td>
</tr>
<tr>
<td>5. Anatomy of an LTR</td>
<td>26</td>
</tr>
<tr>
<td>6. The genomic organization of representative lentiviruses</td>
<td>33</td>
</tr>
<tr>
<td>7. The HIV trans-activators</td>
<td>38</td>
</tr>
<tr>
<td>8. Northern hybridization analysis of total RNA from EIAV-infected and uninfected FEK or FDD cells</td>
<td>59</td>
</tr>
<tr>
<td>9. Northern hybridization analysis of EIAV-specific poly(A)^{+} RNA's in FDD cells infected with FDD-adapted EIAV</td>
<td>64</td>
</tr>
<tr>
<td>10. Time-course study on the synthesis of EIAV-specific transcripts</td>
<td>69</td>
</tr>
<tr>
<td>11. Identification of clones containing the splice-junction cDNA of the EIAV 3.5-kb mRNA</td>
<td>70</td>
</tr>
</tbody>
</table>
12. Nucleotide sequence analysis of the 270-bp Smal-BamHI splice-junction cDNA of the EIAV 3.5-kb mRNA ................................. 79
13. Nucleotide sequence of the EIAV provirus near the 5'-terminus of the gag gene ................................................................. 83
14. Northern hybridization analysis of poly(A)+ RNA from EIAV-infected FDD cells using the three oligonucleotide probes SD-1, SD-2, and SD-3 ................................................................. 87
15. Schematic diagram of the EIAV genome ........................................ 90
16. Schematic representation of the probes used in S1 nuclease mapping .................................................................................. 94
17. S1 nuclease mapping of splice junctions in spliced mRNA's of EIAV .................................................................................... 98
18. Predicted amino acid sequences of the three short ORF's of EIAV ....................................................................................... 105
20. Analysis of the 35S-labeled in vitro translational products of the full length and truncated S2-env in vitro transcripts by SDS-PAGE .................................................................................. 111
21. The eukaryotic transient expression vector p91023(B) .............. 115
22. Immunofluorescence photographs of fixed p9SR426-transfected COS-7 cell cultures treated with anti-gp90 monoclonal antibody followed by fluorescein-conjugated goat anti-mouse IgG .......... 120
23. Immunofluorescence photographs of fixed p9SR426-transfected COS-7 cell cultures treated with anti-S2 rabbit polyclonal antibody followed by fluorescein-conjugated goat anti-rabbit IgG

24. Summary of the splicing patterns of the EIAV full-length mRNA
ABSTRACT

The transcriptional properties of equine infectious anemia virus were examined in two distinct equine cell lines in which the virus establishes either a persistent or a cytopathic infection. Northern hybridization analyses were performed to determine the number, sizes, and relative levels of the EIAV transcripts encoded during persistent or cytopathic infections. Three species of viral mRNA were detected in infected cells: an 8.2-kb full-length genomic mRNA, a 3.5-kb single-spliced mRNA, and a low abundance 1.5-kb mRNA, presumably formed by a double-splicing event of the full-length mRNA. Analysis of the levels of EIAV-specific RNA's present during persistent and cytopathic infections has revealed that quantitative differences characterize the transcriptional patterns of EIAV in these two infections. In persistently infected FEK cells the 8.2- and 3.5-kb mRNA’s are the predominant viral transcripts and are detected in approximately equal concentrations, while the 1.5-kb mRNA is detected at very low levels. During the cytopathic infection of FDD cells, however, the 3.5-kb mRNA is the predominant viral transcript, comprising nearly 75% of the total viral mRNA, while the 8.2- and 1.5-kb mRNA’s constitute the remaining 25% of viral transcripts. Moreover, the cytopathic infection is characterized by almost a thirty-fold higher level of viral transcripts than those detected during the persistent infection.

The splicing patterns of the full-length EIAV mRNA during the cytopathic infection were determined by cDNA cloning and sequencing, Northern hybridization analyses using splice donor-specific oligonucleotide probes, and S1
nuclease mapping of RNA from virus-infected cells. The results have identified the splice donor and acceptor sites used to generate the spliced mRNA's of EIAV in infected cells.

The expression of a putative regulatory protein of EIAV from a structural viral gene was investigated by analysis of in vitro and in vivo expression products. In vitro transcription and translation along with in vivo expression in transfected COS-7 cells were used to analyze the expression of the viral env gene. Based on the results of these studies, a potential mechanism for co-expression of two separate proteins from the env mRNA is proposed.
CHAPTER 1
INTRODUCTION

EQUINE INFECTIOUS ANEMIA: THE DISEASE

Persistent retrovirus infections constitute a major challenge in contemporary human and veterinary medicine. With the discoveries of retroviruses associated with human leukemia and acquired immunodeficiency syndrome (AIDS), the urgency to develop preventative or therapeutic measures against these types of viral infections has become increasingly more apparent. To understand the patterns of retrovirus gene expression and to make correlations with viral pathogenesis, it is essential that animal systems be developed to serve as models for human retroviruses, where experimental infections are not possible.

Equine infectious anemia (EIA) is a naturally occurring worldwide disease of horses caused by a nononcogenic retrovirus (Charman et al., 1976; Cheevers et al., 1977; Issel and Coggins, 1979). EIA, or swamp fever as it is known colloquially, has been recognized as a viral disease of horses since the early 1900's when it was shown to be caused by a filterable viral agent (Vallee and Carre, 1904). Each year in the United States nearly 700,000 horses are tested for EIA, and the highest incidence of the disease is reported to be in Louisiana. Although known as a viral disease for over eighty years, significant progress toward understanding the disease at the cellular and molecular level has only been accomplished within the last decade.
Clinical Signs of EIA

One of the features distinguishing EIA from other retrovirus-induced diseases is its episodic nature and variable clinical course (Crawford et al., 1978; Issel and Coggins, 1979; Orrego et al., 1982). Following infection, the disease is characterized by unpredictable bursts of plasma viremia, recurring peaks of fever above normal animal temperatures, and cycles of clinical symptoms which occur in sequential episodes that are separated by several weeks or months. Horses developing the clinical signs of EIA are described to exhibit acute, chronic, or inapparent cases of the disease (Issel and Coggins, 1979), as depicted in Figure 1.

Acute EIA. This stage of the disease is most often associated with the animal's first exposure to the virus and is a result of massive virus replication in infected macrophages. At this stage of the disease the virus is believed to cause a lytic infection of macrophages (Kono, 1969; McGuire et al., 1971) leading to their eventual destruction. Most horses with acute EIA survive the infection and may go unnoticed since the classical signs of the disease such as weight loss, anemia, or edema are not usually evident at this stage.

Chronic EIA. The more classic clinical signs of EIA, such as weight loss, anemia, and ventral edema are exhibited by the infected animal at this stage of the disease. In these animals the characteristic recurrent cycles of illness are observed, the frequency and severity of which usually decline with time. In addition to the above symptoms, infected horses may also exhibit lathargy, leukopenia, depression of central nervous system, fever, and hemorrhages (Orrego et al., 1982).
Figure 1. Schematic representation of the different stages of EIA in infected horses.
**Inapparent EIA.** A high percentage of infected horses exhibit no clinical illness and are inapparent carriers of virus. These animals, although asymptomatic and healthy appearing, harbor the virus for life and virus or viral products can be detected in blood of such animals (Kemen and Coggins, 1972; Issel et al., 1982). Therefore these inapparent carriers pose a serious threat in spreading the disease to normal uninfected horses.

The variable clinical course of EIA depends in part on resistance factors of the host, viral virulence factors, and environmental stresses (Issel and Coggins, 1979). Following the initial infection, the animal may develop chronic EIA with its characteristic recurrent disease episodes, or in some cases the infection may lead to an initial febrile attack, after which the animal may remain asymptomatic for the remainder of its life. Such apparently normal horses may suddenly re-experience an acute febrile episode. Moreover, such inapparently infected animals may be induced to experience clinical EIA by exposure to certain environmental stresses such as hard work, or upon injection with immunosuppressive drugs (Issel and Coggins, 1979).

**Transmission**

Natural transmission of EIA is mainly horizontal and rarely vertical. It usually occurs through the transfer of blood from the infected horse by interrupted blood feeding of hematophagous insects, such as horse flies or deer flies (Issel and Coggins, 1979; Williams et al., 1981; Foil and Issel, 1982; Foil et al., 1983, 1984; Issel et al., 1988). Virus may also be transmitted to an uninfected horse by the transfer of blood or blood products on contaminated instruments or
syringes (Issel and Coggins, 1979). Although the virus can be demonstrated in excretions and secretions of acutely infected horses, there is no evidence of the natural contact transmission of the virus. In all types of transmission, it appears that an infected donor exhibiting the clinical signs of EIA has a greater potential of transmitting the virus when compared to one showing no signs of disease. This observation is apparently correlated with the level of virus in the donor's blood and tissues. The virus can apparently cross the placental barrier and be transmitted to the fetus. Mares with signs of acute EIA during pregnancy seem to have the greatest potential for carrying infected fetuses (Kemen and Coggins, 1972), although more than 75% of the foals born from infected mares appear to be free of EIA (Kemen and Coggins, 1972; Foil and Issel, 1982).

Diagnosis

As a clinical disease, EIA can often be diagnosed on the basis of history and clinical signs of the affected animal. Since clinical signs of EIA are not always definitive, specific laboratory tests are helpful in the diagnosis of the disease. Such tests normally involve the detection of virus or EIA-specific antibody in the animal.

The horse inoculation test (Issel and Coggins, 1979) is by far the most direct method developed to date for demonstration of circulating virus in the blood of an infected horse. It is performed by injection of blood taken from a suspected horse into an uninfected host. Development of clinical disease in the recipient horse conclusively demonstrates the presence of virus in the donor. The main drawback with this procedure, however, is that it is expensive, time
consuming, and the results may be unclear as it is often difficult to interpret the clinical illness in the recipient horse as a definitive sign of EIA in the donor animal.

A variety of serological tests have also been developed and used for diagnosis of EIA. These include the complement fixation test (Kono and Kabayashi, 1966), the serum neutralization test, and the agar gel immunodiffusion (AGID) test, popularly known as the Coggins test (Roth et al., 1971; Coggins et al., 1972). The AGID test is based on the detection of EIA-specific antibody in the serum of the suspected animal and is a reliable and sensitive test for EIA. The correlation between AGID-seropositivity and the presence of infectious virus in the blood of the animal being tested has been documented to approach 100% (Coggins, 1972). However, as with any other serologic test based on antibody detection, the AGID test does not detect the presence of dormant virus in an infected animal with very low levels of virus-specific antibody or the phase of infection prior to antibody production.

Other serological tests, usually more sensitive than the AGID test, including immunofluorescence (Crawford et al., 1971), radioimmunoassay (Coggins et al., 1978), enzyme-linked immunosorbent assay (ELISA) (Shane et al., 1984), and competitive ELISA (C-ELISA) (Hussain et al., 1988) have also been used for detection of EIA-specific antibody.

The more sensitive serological tests for EIA have become important in diagnosis of EIA in cases when the AGID titers are extremely low. However, the guidelines established by the United States Department of Agriculture (USDA)
for diagnosis of EIA-positive horses are currently introducing certain barriers in the use of such tests. The USDA standard for diagnosis of EIA requires that a horse be either: (1) AGID test-positive, or (2) horse inoculation test-positive, or (3) both (Issel et al., 1988). There are inherent problems in diagnosis of EIA-positive horses which do not have a high enough virus titer in their blood to transfer the infection to a recipient horse in a standard horse inoculation test, but are diagnosed as EIA-positive using the more sensitive ELISA tests. However, such antibody-positive "virus-negative" horses are currently being released from USDA restrictions as "false-positive" animals (Issel et al., 1988). Such potential transmitters of EIA pose a great threat to the healthy normal horse population. Therefore, the urgent need for the development of a more accurate set of standards for EIA detection cannot be overemphasized. Accordingly, more definitive methods of diagnosis designed to detect the presence of actively replicating virus in the blood or tissues of a test animal rather than EIA-specific antibody need to be developed. In this regard diagnostic EIA tests based on the polymerase chain reaction (PCR) (Scharf et al., 1986; Mullis and Fallona, 1987; Saiki et al., 1988), which would detect viral genetic material in the suspected animal, would have a great potential in the near future. To date there has been no known treatment or preventative measures developed for EIA. Efforts to control this economically important disease have mainly involved screening for EIA-positive horses and subsequent spatial separation of these carriers from normal healthy horses during housing and transportation.
EQUINE INFECTIOUS ANEMIA VIRUS: THE CAUSATIVE AGENT

Equine infectious anemia virus (EIAV), the causative agent of the disease equine infectious anemia, is a member of the retrovirus subfamily of RNA viruses. Retroviruses are unusual entities in that they propagate themselves by taking advantage of the lifestyles of their host animals in unique and unprecedented ways. They convert their genes from RNA to DNA, incorporate the DNA intermediate stably into chromosomes of somatic or germ cells, and often times mutate or even capture cellular genes. Many retroviruses rarely impair and often potentiate the growth of their host cells, and all retroviruses express their genes by using host-specific mechanisms and machinery under the direction of viral genetic control signals. These viruses have been isolated from a diverse group of vertebrate animals including mammals, birds, and reptiles (Teich, 1982, 1985), and can cause a wide array of oncogenic, nononcogenic, and progressive diseases. The common taxonomic features of retroviruses are summarized in Table I.

Subfamilies of Retroviruses

On the basis of cross-reactivity of the group-specific major antigen within a subfamily, genomic organization, morphological data, and in part, the characteristics of the disease the virus induces in the infected animal, the family of retroviruses are divided into three subfamilies: oncoviruses, spumaviruses, and lentiviruses.
Table I. Taxonomic characteristics common to retroviruses.

Morphology
Spherical enveloped virions (80 to 120 nm in diameter); variable surface projections (8 nm); icosahedral or cubic capsid containing a ribonucleoprotein complex.

Physiochemical properties
Density 1.16 to 1.18 g/ml in sucrose, 1.16 to 1.21 g/ml in cesium chloride; sensitive to lipid solvents, detergents, and heat inactivation; highly resistant to UV and X-ray irradiation.

Nucleic acid
About 1% by weight; linear single-stranded positive-sense RNA (60 to 70 S) composed of two identical subunits (30 to 35 S).

Protein
About 60% by weight; gag-encoded internal structural proteins (3 to 5); pol-encoded reverse transcriptase, protease, and integrase (2 to 3); env-encoded envelope structural proteins (1 to 2).

Lipid
About 35% by weight; derived from the infected cell's membrane.

Carbohydrate
About 4% by weight; associated predominantly with env proteins.

---

*a. From Teich (1982), and Bharat Parekh, Ph.D. dissertation, Louisiana State University, 1983.

*b. The numbers in parantheses indicate the number of corresponding polypeptides.*
The oncovirus subfamily includes the oncogenic and serologically related nononcogenic retroviruses and replication-defective non-infectious oncoviruses. Oncoviruses, recognized for their cancer-causing potential, have been isolated from a wide variety of vertebrates such as fish, chickens, rodents, cats, and primates, and induce various tumors of mesenchymal origin (sarcomas), leukemias, and less often, epithelial malignancies including carcinomas of the breast, kidney, and liver (Weiss et al., 1982, 1985). Representative members of oncoviruses include the avian leukemia and sarcoma viruses, the murine leukemia and sarcoma viruses, bovine and feline leukemia viruses, and Mason Pfizer monkey virus. In addition, Human T-cell leukemia viruses type I and II (HTLV-I and HTLV-II) have also been tentatively placed in this subfamily (Poiesz et al., 1980; Sonigo et al., 1981; Kalyanaraman et al., 1982).

Spumaviruses, commonly referred to as "foamy" viruses, are characterized by induction of foamy degeneration of infected cells in tissue culture. These viruses have been isolated from a number of mammalian species such as cattle, hamsters, rabbits, cats, monkeys, and humans (Gelderblom and Frank, 1987) and establish persistent infections with no evidence of pathogenesis. Representative members of this subfamily include simian foamy virus, bovine syncytial virus, and human syncytium-forming virus.

The lentivirus subfamily derives its name from the slow time course of the infections its members cause in their hosts. Lentiviruses cause chronic infections characterized by prolonged incubation periods of several months to many years, followed by a protracted symptomatic disease phase. The unifying theme among
lentivirus infections is the restricted viral gene expression in the host animal and persistent, latent infections that are not usually apparent for long periods of time (Haase, 1986). Among the chronic diseases which lentiviruses cause in animals are those affecting the lungs, joints, nervous, hematopoietic, and immune systems of humans and animals.

Although the classification of a retrovirus in this subfamily has traditionally been more heavily based on the progressive nature of the induced disease than on group-specific major antigen cross-reactivity, more recently retroviruses have been added to this subfamily based on nucleotide sequence homology and the organization of their genome (Gonda et al., 1985; Sonigo et al., 1985). Members of the lentivirus subfamily include the visna-maedi virus, progressive pneumonia virus (PPV), and caprine arthritis encephalitis virus (CAEV) of sheep and goats, equine infectious anemia virus (EIAV) of horses, feline immunodeficiency virus (FIV) of cats, bovine immunodeficiency virus (BIV) of cows, simian immunodeficiency virus (SIV) of monkeys, and the causative agent of acquired immunodeficiency syndrome (AIDS) in humans, the human immunodeficiency virus type 1 and type 2 (HIV-1 and HIV-2). EIAV was originally classified as a lentivirus due to its morphology and the chronic, persistent nature of its induced disease in horses (Charman et al., 1976; Parekh et al., 1980; Montelaro et al., 1982), although no serological relatedness could be documented between this virus and the prototype lentivirus, visna virus (Stowring et al., 1979). However, recent determination of the EIAV nucleotide sequence and its genomic organization has provided clear evidence for the remarkable relatedness of this
virus to other lentiviruses such as visna, CAEV, and HIV (Chiu et al., 1985; Rushlow et al., 1986; Stephens et al., 1986), thus substantiating its classification as a lentivirus. Moreover, recent serological studies have provided further evidence for the close relationship among the viruses of this subfamily by demonstrating that anti-HIV rabbit serum exhibits cross-reactivity with the BIV p26, CAEV p27, and EIAV p26 core proteins (Casey et al., 1985; Gonda et al., 1987).

Physiochemical Properties of EIAV

EIAV has been shown to possess many of the physiochemical properties common to enveloped viruses in general and retroviruses in particular. Mature EIAV particles are enveloped and have a buoyant density of about 1.16 g/ml in cesium chloride (Matheka et al., 1976; Cheevers et al., 1977). Viral infectivity is sensitive to reagents which disrupt the lipid envelope such as detergents or ether, but is resistant to trypsin treatment, ultraviolet, and X-ray irradiation (Nakajima, 1972). Moreover, treatment with certain chemical disinfectants such as sodium hydroxide, sodium hypochlorite, organic phenolic compounds, and chlorohexidine has been shown to inactivate virion infectivity (Shen et al., 1977).

The virus particle contains a 60 to 70S RNA genome composed of two identical single-stranded 34S subunits that are noncovalently bound near their 5'-ends, a virus-encoded RNA-dependent DNA polymerase (reverse-transcriptase) with optimal endogenous activity in the presence of Mg++, and a host-derived lysine-specific tRNA molecule whose 3'-end is hydrogen-bonded to an 18-base sequence (primer binding site, PBS) near the 5'-end of each genomic RNA
subunit (Charman et al., 1976; Archer et al., 1977; Cheevers et al., 1977; Varmus, 1982; Stephens et al., 1986).

Virus Morphology

The morphological properties of retroviruses have been investigated by electron microscopy (Bernhard et al., 1958; Gross, 1970; Vigier, 1970). By virtue of minor differences revealed by electron microscopy, retroviruses can be divided into four morphological groups: type-A, type-B, type-C, and type-D (Bernhard, 1960; Sarkar et al., 1972; Fine and Schochetman, 1978). These differences are related to the presence or absence of visible intracellular or extracellular viral particles, location of the inner core within the viral envelope, and changes in the morphology of the virion particles during the process of budding from the cell membrane.

Morphologically, EIAV exhibits characteristics consistent with that of most retroviruses. Electron microscopy studies have revealed spherical virion particles of 80 to 120 nm in diameter, with a rod- or cube-shaped electron dense nucleoid of about 40 to 60 nm in length (Weiland et al., 1977; Teich, 1982). The virus particle is bounded by a double-layered lipid envelope with spike-and-knob surface projections of about 6 to 8 nm in diameter (Nakajima et al., 1969; Tajima et al., 1969; Weiland et al., 1977). Mature EIAV particles acquire their envelope from the plasma membrane of the host cell during the budding process (McConnell et al., 1977; Gonda et al., 1978).
Structural Components of EIAV

Experiments performed to identify viral structural proteins by guanidine hydrochloride gel filtration (GHCL-GF) and high resolution sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), along with localization studies for determination of the organization of these polypeptides within the virion have provided the basis for understanding the structure of the intact EIAV virion (Montelaro et al., 1982; Parekh et al., 1980). In gel filtration chromatographic analysis of the cell-adapted Wyoming strain of EIAV radiolabeled with radioactive leucine or glucosamine, six distinct peaks of radioactivity, representative of virus-associated proteins, have been observed (Parekh et al., 1980). These proteins exhibit apparent molecular weights of 100,000, 74,000, 26,000, 15,000, 11,000, and 9000. SDS-PAGE analysis of EIAV has revealed four major low molecular weight polypeptides, designated p26, p15, p11, and p9, and six minor components. The minor components have been shown to include two glycosylated proteins of 90,000 and 45,000 molecular weight, designated gp90 and gp45, respectively, and four minor nonglycosylated proteins designated as p70, p61, p30, and p23. The unrelatedness of the four major nonglycosylated proteins (p26, p15, p11, and p9) have been demonstrated by comparative peptide mapping and enzyme-linked immunosorbent assays (Montelaro et al., 1982). Furthermore, biochemical characterization of these four nonglycosylated proteins of EIAV (Montelaro et al., 1982) has revealed that p26 and p9 focus at isoelectric points (pI) of 6.2 and 5.0, respectively, and that they contain no unusual amino acids. The EIAV p15 was shown to display a
heterogeneous isoelectric focusing pattern with a pI ranging from 5.7 to 8.3, which could be attributed to different levels of phosphorylated serine or threonine or both. Finally, p11 was observed to focus at a pI of greater than 10.0, suggesting its high content of basic amino acids. Among the two major glycoproteins of EIAV, gp90 is the more glycosylated polypeptide, while gp45 aggregates in 6 M guanidine hydrochloride, indicative of a hydrophobic nature (Parekh et al., 1980).

Localization studies aimed at assigning the relative location of the virion polypeptides have been performed by treatment of purified virus with the protease bromelain. Such an analysis distinguishes surface proteins from those sequestered by the viral lipid bilayer from enzymatic digestion (Mosser et al., 1975). Application of these studies to EIAV have revealed that gp90 and gp45 constitute the surface proteins of EIAV, whereas p26, p15, p11, and p9 form the internal virion components. Furthermore, treatment of purified EIAV with Nonidet P-40, which releases the viral ribonucleoprotein complex, has shown that the highly basic EIAV p11 is closely associated with the viral RNA genome (Montelaro et al., 1982).

Retroviral Genome: Structure, Integration, and Replication

The single most important feature which separates retroviruses from other animal viruses is their mode of replication. The very name "retrovirus" embodies the biological phenomenon which this family of viruses is most famous for: the ability to copy their single-stranded RNA genome into a double-stranded DNA intermediate in a multistep process coined as "reverse transcription". In fact, it was the pioneering work with these viruses which originally convinced the
scientific community of the "provirus hypothesis" (Temin, 1964, 1976); that transfer of genetic information in biological systems is not limited to the conventional processes of replication (DNA to DNA or RNA to RNA), transcription (DNA to RNA), and translation (RNA to protein).

As mentioned before, the viral genome is wrapped and securely protected in a core of viral protein which, in turn, is surrounded by a lipid envelope derived from the membrane of the previous host cell and containing viral glycoproteins. The infecting virion attaches to its target cell through an interaction with cell surface receptor proteins that are specifically recognized by viral envelope glycoproteins (Tardieu et al., 1982). Entry of the virus particle into the cell is probably mediated by receptor-mediated endocytosis, a mechanism that cells have evolved to ingest beneficial extracellular substances such as growth factors (Mims, 1986).

Although replication of the extracellular virion particle occurs only within cells and depends on cellular functions and machinery, the infecting virus brings with it an organized collection of viral enzymes and RNA to direct its replication process. Following entry into the cell, the quiescent enveloped virion is uncoated and becomes an enzymatically active nucleoprotein complex capable of performing its initial programmed functions with little or no help from the host cell. Each viral genomic RNA molecule is base-paired with a specific, host cell-derived transfer RNA molecule that primes viral DNA synthesis during reverse transcription. Once the genomic RNA is copied into DNA by the virus-coded reverse transcriptase, the nucleoprotein complex migrates into the nucleus and
mediates the covalent integration of the double-stranded DNA copy of the viral genome into the chromosomal DNA of the host cell. A schematic view of this replication cycle is presented in Figure 2 (Varmus, 1983).

Following the integration of its provirus, the retrovirus becomes totally dependent on its host cell for replication of the integrated provirus as a part of the cellular chromosome, for transcription of the proviral DNA by RNA polymerase II, for post-transcriptional processing of viral RNA transcripts by the host capping, polyadenylation, and splicing machinery, and for translation of the resultant messenger RNA's by the host's ribosomes. Recognition signals, one of which may be localized in the leader sequences near the 5'-end of the viral genomic RNA (Shank and Linial, 1980; see below), then draw together the newly synthesized viral genomic RNA and core proteins from the cytoplasm and envelope glycoproteins embedded in the plasma membrane. A virus-encoded protease cleaves the viral polyproteins into the smaller functional components, and mature progeny virions are then released from the plasma membrane of the host cell. Figure 3 shows a schematic representation of the molecular events involved during the replication of the virion particle during the retroviral life cycle (Varmus, 1988).

Retroviral Gene Expression

Within the retrovirus particle, the genome is a complex of two identical RNA strands. Such a state of diploidy and genetic redundancy is a property not seen in other animal viruses and is unique to retroviruses, although its advantages are not well understood (Varmus, 1982, 1983, and 1988). Starting from the 5'-end,
Figure 2. A simplified schematic representation of the various stages involved in reverse transcription. 1. Viral genomic RNA. R5 and R3: 5'- and 3'-direct terminal repeats; U5 and U3: 5'- and 3'-untranslated regions; PB: primer binding site; G,P, and E: *gag*, *pol*, and *env* genes, respectively. 2 and 3. First-strand cDNA and double-stranded DNA synthesis. 4. Circular double-stranded proviral DNA. 5. Integration of proviral DNA into host genome. 6. Transcription of proviral DNA to yield viral mRNA's.
Figure 3. The molecular events involved in the retroviral life cycle. Cap, capped nucleotide at the 5'-end of viral genomic RNA; $A_\text{a}$, poly(A) tail at the 3'-end of viral genomic RNA; R, direct repeat sequence at each end of viral RNA; U3 and U5, unique sequences duplicated during reverse transcription; LTR, long terminal repeat; $gag$, coding region for viral nucleocapsid proteins; $pol$, coding region for RNA-dependant DNA polymerase (reverse transcriptase); $env$, coding region for viral envelope proteins; CJ, circle junction where the ends of linear viral DNA are joined together; $S_\text{a}$ and $S_\text{e}$, splice donor and acceptor sites, respectively; Psi, signal for packaging of viral RNA; P and CHO, post-translational modifications of viral proteins by phosphorylation and glycosylation, respectively. (From Varmus, 1988).
Virus particle

Cap

\[ \beta \text{ gag pol env} \]

\[ \text{R U5} \rightarrow \text{U3 R} \]

LTR

LTR

\[ \text{CJ} \]

Free DNA

Proivirus

Genome + mRNA

Viral proteins

\[ \text{SD} \rightarrow \text{SA} \]

\[ \text{Psi} \]

\[ \text{Cap} \rightarrow \text{env} \]

\[ \text{Splice} \]

\[ \text{CHO} \]
as presented schematically in Figure 4, the arrangement of the noncoding sequences on the RNA molecule has the following features (Varmus, 1983):

A. *The 5'-Cap.* Similar to the 5'-termini of almost all eukaryotic mRNA's, an inverted 7-methylguanyl nucleotide (the "cap" nucleotide) is present at the 5'-end of the viral RNA.

B. *The R Sequence.* In addition to the genetic redundancy of retroviruses in having two identical genomic RNA's, each RNA subunit is itself terminally redundant and carries at each end a direct repeat sequence of about 10-80 nucleotides sequence (Coffin, 1979; see below). This sequence, termed R, plays an important role during the process of reverse transcription, and is present as a component of each long terminal repeat (LTR) sequence at the 5'- and 3'-ends of the proviral DNA (see below).

C. *The U5 Sequence.* Following the R sequence, a sequence of about 80-100 nucleotides, referred to as U5, is present at the 5'-end of the viral RNA. This sequence occurs only once in the viral RNA, but is present twice in the proviral DNA as a segment within the 5'- and 3'-LTR's.

D. *The Primer Binding Site (PBS).* The initiation site for synthesis of the first strand of viral DNA during reverse transcription is the 3'-boundary of U5, which in turn is the site at which the host-derived tRNA primer is hydrogen-bonded to the viral RNA (Taylor, 1977). This sequence, usually 18 nucleotides long, is known as the primer binding site (PBS).

E. *The Leader Region (L).* Following the PBS, and preceding the translational initiation codon of the first structural gene (gag), is a sequence of a
Figure 4. Anatomy of a genomic RNA subunit of a replication-competent retrovirus. The important structural and genetic domains of a typical single-stranded RNA subunit (about 8- to 9-kb) of a replication-competent retrovirus are indicated by the following notations: cap, 5'-cap nucleotide; R, terminal redundancy; U5, sequence unique to the 5'-terminus, present twice in viral DNA; (-)PB, binding site of primer, a host tRNA, for (-)DNA strand synthesis; L, leader region; S_D, splice donor site; S_A, splice acceptor site; gag, pol, and env, coding regions for viral nucleocapsid proteins, reverse transcriptase, and envelope proteins, respectively; NT, nontranslated region between env and U3; (+)P, polypurine tract, primer for (+)DNA strand synthesis; U3, sequence unique to the 3'-terminus, present twice in viral DNA; A_n, poly(A) tract at the 3'-terminus.

For clarity, terminal regions are shown at a 10X expanded scale. Lengths of the various regions are approximate and vary among different retroviruses. (From Varmus, 1983).
few hundred nucleotides referred to as the leader (L) region. In addition to containing splice donor sites for generation of the spliced viral mRNA's in many retroviruses, the L region may also contain the signal which facilitates packaging of viral RNA into virion particles (Shank and Linial, 1980).

**F. The 3'-NT.** The portion of the noncoding region downstream of the *env* gene which does not overlap with the U3 and R regions of the 3'-LTR is termed 3'-NT. This region is variable in length among different retroviruses and lacks substantial open reading frames. A stretch of purine-rich nucleotides, directly preceding the U3 boundary and about 10-20 bases long, is located at the 3'-terminus of the 3'-NT and is a highly conserved feature among retroviruses. This purine-rich tract plays an important role during reverse transcription and appears to serve as primer for the synthesis of plus strand viral DNA.

**G. The U3 Sequence.** Analogous to the U5 sequence at the 5'-end of the viral RNA, the U3 sequence is the unique sequence which appears once at the 3'-end of the RNA molecule and twice in the proviral DNA as a component of each LTR. Within the LTR unit, an inverted repeat is present as a result of the first 5 to 23 nucleotides of U3 being inversely repeated at the 3'-end of U5. Ranging in size from about 170 to 1250 nucleotides among different retroviruses, the U3 sequence also varies greatly in length among different isolates of a single retroviral strain (Ju *et al.*, 1980; Shimotohno and Temin, 1982; Van Beveren *et al.*, 1982). The U3 sequence, although long enough in certain retroviruses to be considered as a potential protein coding domain, is analogous to the 3'-NT region in being usually devoid of open reading frames. The major function of the U3
sequence involves the regulation of synthesis and processing of viral RNA (see below).

**H. The Poly(A) Tail.** Added to the 3'-end of the viral RNA post-transcriptionally, presumably by cellular polynucleotide synthetases, is a poly(A) tail of about 100-200 nucleotides long. Since the poly(A) sequence is added to the RNA after its synthesis, its copy is not present in the proviral DNA template and is not reverse transcribed during the synthesis of viral DNA.

During the intricate process of reverse transcription, sequences present once at the 5'- and 3'-ends of the genomic RNA (U5 and U3, respectively) are duplicated to generate the long terminal repeat (LTR) sequences at the 5'- and 3'-ends of the proviral DNA. The LTR unit, schematically presented in Figure 5, is several hundred bases in length and is a fusion of the sequences characteristic of the 3'-end of viral RNA (U3), the R sequence, and sequences characteristic of the 5'-end of viral RNA (U5), in the order 5'-U3-R-U5-3 (Varmus, 1983). The LTR provides the important regulatory sequence information required for expression of the viral genome, namely, signals for promotion, initiation, and polyadenylation of the viral transcripts. The U3 region of the LTR lies immediately upstream of the initiation site for viral RNA synthesis (Varmus and Swanstrom, 1982); approximately 25-30 nucleotides upstream from the 5'-end of R and within U3 retroviral LTR's invariably possess the consensus "TATAA box" sequence, strongly implicated in initiation of transcription by eukaryotic RNA polymerase II (Breathnach and Chambon, 1981). The 5'-end of R, also known as the cap site, is by definition the 5'-end of the viral RNA and is thus the
Figure 5. Anatomy of an LTR. The figure shows the important features of the U3, R, and U5 sequences of the retroviral LTR and its flanking viral sequence. Features pertaining to a 5'-LTR are shown at the top of the figure, and those relevant to a 3'-LTR are presented at the bottom. The viral sequence at the 5'-boundary of U3 is the sequence of the putative primer for plus-strand DNA synthesis, and the sequence at the 3'-boundary of U5 is the binding site for the tRNA primer for minus-strand DNA synthesis. The LTR terminates with short inverted repeats (IR). The integration sites are usually 2-bp from each boundary. The sequences resembling CCAAT and TATAAA regulate the initiation of viral RNA synthesis from the cap site, while the non-overlapping sequence, usually AATAAA, determines the polyadenylation (A₃) site at the 3'-end of RNA. (From Varmus, 1983).
transcriptional initiation site (Varmus and Swanstrom, 1982; Varmus, 1983). A sequence resembling CCAAT is also usually found about 30-40 nucleotides upstream of the "TATAA box" in the U3 region. The consensus eukaryotic signal for polyadenylation, AATAAA (Proudfoot and Brownlee, 1974; Fitzgerald and Shenk, 1981), is usually located near the 3'-end of U3, about 20 nucleotides upstream from the site of poly(A) addition (Varmus, 1982). In retroviruses with large R sequences the poly(A) addition signal can occur within R and must apparently be ignored during transcription from the 5'-LTR (Varmus, 1983). The dinucleotide CA, usually defining the R-U5 boundary, also serves as the polyadenylation site on the RNA molecule (Bonner et al., 1982; Chen and Barker, 1984).

The coding sequences within the retroviral genome contain at least three genes, located in between the two LTR units, which encode the structural and replicative components of the virus particle. These genes are gag, encoding the viral core proteins or group-specific antiens; pol, encoding the viral RNA-dependent DNA polymerase (reverse transcriptase), integrase, and protease; and env, encoding the viral envelope proteins.

Certain retroviruses possess additional genes which code for proteins with specialized intracellular functions, such as oncogenes or regulatory genes. A group of oncoviruses, typified by Rous sarcoma virus (RSV), carry a viral oncogene whose product is directly involved in the swift induction of tumors in animals and the efficient transformation of cells grown in culture (Bishop and Varmus, 1982). On the other hand, HTLV-I and -II, also recognized as oncoviruses, promote
their oncogenic action through the product of a small open reading frame that lies between the viral env gene and the 3'-LTR. This protein acts as a positive effector of transcription from the viral LTR and from certain cellular promoters, thus promoting tumorigenesis (Sodroski et al., 1984; Greene et al., 1986; Seiki et al., 1986). Moreover, a common feature among lentiviruses, which is absent in oncoviruses, is the presence of several small open reading frames (ORF's) in their genomes in addition to the structural and replicative gag, pol, and env genes (Ratner et al., 1985; Sanchez-Pescador et al., 1985; Sonigo et al., 1985; Wain-Hobson et al., 1985; Rushlow et al., 1986). These short ORF's, usually located between the pol and env genes and near the 3'-end of the env gene, have been intensely studied within the past few years and have invariably been shown to encode regulatory proteins which act as either positive or negative effectors of viral gene expression (see below).

As mobile genetic elements within eukaryotes, retroviruses have evolved into complex entities which have adopted a combination of devices and mechanisms to facilitate their expression in host cells that normally operate in a totally different pattern of genetic organization. In eukaryotic cells, genes are principally arranged as single coding sequences that are interrupted by noncoding regions (introns or intervening sequences). Retroviruses, however, are composed of multicistronic genomes destined to be expressed in eukaryotic cells, which are normally not equipped to translate several independent genes from a single mRNA. To get around this problem, retroviruses have evolved two specific strategies of gene expression: production of spliced subgenomic mRNA's from a
single full-length viral mRNA in order to provide the coding message for proteins encoded at internal positions of the viral genome; and cleavage of the primary viral translational products (polyproteins) to generate multiple functional proteins required for viral replication.

Synthesis and Splicing of the Full-Length Viral mRNA

The primary product of transcription from the proviral DNA template initiates at the 5'-boundary of the R sequence within the 5'-LTR, to which then the 5'-cap nucleotide is added post-transcriptionally (Varmus, 1983). Although this primary transcripts, hereafter referred to as the full-length viral mRNA, is approximately the same length as the subunit of viral genomic RNA, there is evidence to believe that transcription can extend some distance beyond the 3'-end of the R sequence, copied from the 3'-LTR and to which the poly(A) tail is added (Yamamoto et al., 1980; Ucker, 1981; Nevins, 1982).

Since the coding region for the env gene lies in the 3'-half of the genomic RNA, its translation from the full-length viral mRNA requires a splicing event in which the gag-pol intron is removed from this primary transcript, joining 5'-leader sequences to the env coding sequences (Varmus and Swanstrom, 1982). The splice donor site for this splicing event usually resides in the leader region downstream of the transcriptional initiation site and upstream of the gag open reading frame, although there are exceptions to this rule in certain retroviruses, such as RSV, in which the donor site lies within the gag gene (Hacket et al., 1982). The splice acceptor site is usually located just upstream of the env open reading frame, thus
allowing the fusion of the 5'-leader and env sequences through the two splice junction signals.

As mentioned above and presented in detail below, relative to other retroviruses, lentiviruses possess additional smaller regulatory genes near the 3'-end of the env gene or within the pol-env intergenic region. In many cases studied to date (Arya et al., 1985; Arya and Gallo, 1986; Davis et al., 1987; Muesing et al., 1985; Rosenblat et al., 1988), these regulatory genes are encoded by multiple small exons encoded within short open reading frames (ORF's) in the 3'-half of the genome. Generation of mRNA's for these small coding sequence usually occurs through multiple splicing events by which both the gag-pol and the env introns are spliced out of the full-length viral mRNA to generate a small mRNA, usually 1.2-2.0 kb in length and composed of 5'-leader sequences fused to the small exons of the regulatory gene, to allow for the translation of the regulatory mRNA.

Translation of the Viral mRNA's

The full-length viral mRNA serves as the message for synthesis of a gag-pol precursor polyprotein which is then post-translationally modified and processed to yield the mature viral core proteins and reverse transcriptase (Vogt and Eisenman, 1973; Vogt et al., 1975). The mechanism used for the synthesis of the gag-pol fusion protein illustrates the unique and intricate strategies evolved by retroviruses to utilize the eukaryotic cellular machinery. The gag and pol genes are separated by a nonsense termination codon at the 3'-end of gag in MLV; they read in different and briefly overlapping frames in RSV and all lentiviruses such
as HIV, visna, and EIAV; and are separated by a third open reading frame (pro, which stands for the viral protease) in MMTV, BLV, and HTLV-II (Varmus, 1988). Rather than directing the host cell's splicing machinery to splice out the gag stop codon or to create monocistronic messages by splicing out pol sequences from the full-length viral mRNA, retroviruses instead take advantage of the potential of eukaryotic ribosomes to occasionally insert an amino acid in place of the termination codon at the 3'-end of the MLV gag (Yoshinaka et al., 1985) to suppress translational termination, or to utilize the occasional frameshifting of eukaryotic ribosomes at defined sites and frequencies in order to bypass the gag termination codon in RSV, HIV, MMTV, BLV, and HTLV-II (Jacks and Varmus, 1985; Jacks et al., 1987; Moore et al., 1987, Jacks et al., 1988). To date, these phenomena have not been observed in translation of cellular mRNA's and would indeed be deleterious if occurring frequently and without any control mechanism. However, their occurrence have clear advantages for retroviruses in that structural (gag) proteins can be synthesized in large quantities required for viral replication, while catalytic (pol and pro) proteins not required to be made in large amounts can be synthesized in relatively small quantities. With this translational mechanism, pol products destined to be packaged into progeny virion particles can also be directly incorporated into the viral core through attached gag components.
The Lentivirus Genomic Organization: *Trans*-acting Factors and *Cis*-acting Sequences

As mentioned before, a common feature among lentiviruses, which is absent in oncoviruses, is the presence of several small open reading frames (ORFs) in their genomes in addition to the structural and replicative *gag*, *pol*, and *env* coding sequences. In general, these short ORFs encode *trans*-acting regulatory proteins which act as either positive or negative effectors of viral gene expression. The complex regulation of lentivirus gene expression, brought about by the fine tuning of these regulatory genes, is believed to enable these viruses to establish latency within the infected host, then respond rapidly to various signals to synthesize high levels of viral proteins leading to active viral replication (Peterlin *et al.*, 1988).

The genomic organization of several representative lentiviruses, including EIAV, are presented in Figure 6. Among lentiviruses, the most intensely studied virus within the past five years has been HIV-1, and much has been learned about the molecular biology and genetic complexity of this virus. HIV-1 and HIV-2 are very similar and the two viruses share a 50-60% similarity in their primary proviral DNA sequence (Clavell *et al.*, 1986; Ho *et al.*, 1987; Fauci, 1988).

The HIV-1 genome encodes at least three genes, *tat*, *rev*, and *nef*, whose products regulate viral gene expression. The coding sequences for *tat*, or *trans*-activator of transcription, are within two separate exons of the HIV genome (Figure 6) which are brought together a double-splicing event of the full-length HIV-1 mRNA (Arya *et al.*, 1985; Sodroski *et al.*, 1985). The *tat* gene product of
Figure 6. The genomic organization of representative lentiviruses. In addition to the \textit{gag}, \textit{pol}, and \textit{env} genes, lentiviruses possess additional short open reading frames in between \textit{pol} and \textit{env} and near the 3'-terminus of \textit{env}. These additional ORF's in HIV-1 code for virion infectivity factor (\textit{vif}), viral protein R (\textit{vpr}), \textit{trans}-
activator of transcription (\textit{tat}), regulator of expression of virion proteins (\textit{rev}),
viral protein U (\textit{vpu}), and negative factor (\textit{nef}). The HIV-2 and SIV genomes similarly encode \textit{vif} (\textit{vip} or virion infectivity protein in the case of SIV), \textit{vpr}, \textit{tat},
\textit{rev}, and \textit{nef}, but do not code for \textit{vpu}. Instead, HIV-2 and SIV contain an ORF for viral protein X (\textit{vpx}). Of the three short ORF's of EIAV, \textit{S1}, \textit{S2}, and \textit{S3}, only \textit{S1} is currently known to encode a regulatory protein (see text).
HIV-1

HIV-2

SIV

EIAV
HIV-1 is a 14-kda protein containing a stretch of cysteine residues followed by a stretch of basic amino acids, both of which are required for its activity (Arya et al., 1985; Sodroski et al., 1985; Arya et al., 1987). Moreover, the tat protein forms a non-covalently linked dimer which binds divalent cations in solution (Frankel et al., 1988) and is localized in the nucleus (Hauber et al., 1987). Mutations and deletions in the tat gene have been shown to abolish viral infectivity, thus tat is essential for HIV-1 replication (Dayton et al., 1986; Fisher et al., 1986; Luciw et al., 1987). The cis-acting element on the HIV-1 genome required for tat activity is a trans-acting response sequence (TAR), originally defined as the region from positions -17 to +80 within the 5'-LTR (Rosen et al., 1985). Additional mutational analysis of TAR has revealed that only sequences from +19 to +42 are required for responsiveness to tat (Jacobovits et al., 1988; Peterlin et al., 1988). Although not precisely defined, the mechanism of trans-activation by tat has been suggested to involve transcriptional, post-transcriptional, and translational events. The regulatory function(s) of tat have been related to an increase in steady-state levels of RNA species containing the TAR element at their 5'-ends (Cullen, 1986; Peterlin et al., 1986; Wright et al., 1986; Kao et al., 1987; Muesing et al., 1987), a bimodal mechanism involving two sites of action, one leading to increased levels of TAR-containing RNA's and the other to even higher levels of their encoded protein products (Cullen, 1986; Wright et al., 1986), and an increase in the rates of initiation (Jacobovits et al., 1988; Lasphia et al., 1989) and elongation (Kao et al., 1987; Lasphia et al., 1989) of transcription from the 5'-LTR. More recently, tat has been shown to interact with the RNA
sequence of TAR, rather than its DNA sequence (Berkhout et al., 1989), giving rise to speculations that it might function as an RNA enhancer (Sharp and Marciniak, 1989). In general, the interaction of tat with TAR-containing RNA is believed to facilitate transcriptional elongation, increase mRNA stability, and enhance translation. The predominance of each effect might then depend on quantitative differences in tat and qualitative differences in host cellular factors (Peterlin and Luciw, 1988).

rev, or the regulator of expression of viral proteins, previously known as art/trs (the anti-repression trans-activator/trans-regulator of splicing), is another trans-acting HIV gene product and its coding sequences overlap but are in a different reading frame from those of tat (Figure 6) (Sodroski et al., 1986; Feinberg et al., 1986). Similar to tat, mRNA coding sequences for rev are within two separate exons on the HIV genome which are brought together by RNA splicing. The rev gene product is a 19-kda protein which, like tat, contains basic residues and is localized in the nucleus (Cullen et al., 1988). The cis-acting repression sequences (crs) in the full-length and single-spliced HIV-1 mRNA's, encoding gag and env structural proteins, respectively, have been shown to render these long mRNA's very unstable (Rosen et al., 1988). A separate cis-acting anti-repression sequence (car), located within env and non-overlapping with crs, is the target for the trans-acting rev protein (Felber et al., 1989; Hammarskjold et al., 1989; Malim et al., 1989; Rosen et al., 1988). By interacting with car, also known as the rev-response element or RRE (Malim et al., 1989), the rev protein has been shown to directly effect the post-transcriptional processing of HIV-1 transcripts by
activating the transport of full-length and single-spliced HIV-1 mRNA's to the cytoplasm and away from the splicing machinery in the nucleus (Felber et al., 1989; Hammerskjold et al., 1989; Malim et al., 1989). By so doing, at the expense of viral regulatory proteins being expressed from the double-spliced viral mRNA's, including rev itself, rev rescues and exports incompletely spliced HIV-1 gag and env mRNA's away from the nucleus for translation of viral structural proteins in the cytoplasm.

nef, or the negative factor, previously known as 3'-orf, orfB, E', and F, encodes a 25- to 27-kda protein believed to negatively regulate HIV replication and expression (Ahmad and Venkatesan, 1988; Luciw et al., 1987). Cloned HIV proviruses designed to contain mutations or deletions in the nef gene have been shown to give rise to viruses which replicate to higher titers in tissue culture than the wild-type virus (Luciw et al., 1987). Moreover, the nef gene product has been proposed to downregulate HIV-1 transcription by activating trans-acting factors which in turn interact with the negative regulatory element (NRE) in the upstream U3 region of the 5'-LTR (Ahmad and Venkatesan, 1988).

Figure 7 presents a summary of the regulatory functions of the HIV trans-activators. The cis-acting sequences in the viral genome (TAR and crs) inhibit or downregulate viral gene expression, allowing only low levels of double-spliced viral transcripts to be synthesized early after infection. As the gene products of tat and rev continue to accumulate, a threshold of these positive trans-activators is reached, after which they greatly increase the expression of TAR-containing RNA and incompletely spliced viral transcripts, respectively. In the meantime,
Figure 7. The HIV trans-activators. tat, rev, and nef are translated from double-spliced viral mRNA, denoted as "early transcripts", containing coding sequences flanking the env gene. Transcription initiates from the 5'-LTR. (a) tat acts upon TAR, which is present just 3' to the initiation site of transcription and which becomes a part of the 5'-terminal sequences in HIV full-length RNA, to result in large increases in TAR-containing RNA species. (b) rev acts on car (cis-acting anti-repression sequence) to increase the expression of full-length and single-spliced HIV RNA, denoted as "late transcripts". In the absence of rev, double-spliced "early transcripts" predominate. Full-length and single-spliced viral RNA are quickly processed or degraded since they contain crs elements (cis-acting repression sequence) in gag and env coding regions. (c) nef is encoded by an ORF located 3' to the env gene and negatively regulates viral gene expression, either by decreasing cellular signaling, or by acting on the negative regulatory element (NRE) in the upstream U3 region of the 5'-LTR. The open boxes at the 3'-end of viral transcripts represent U3 and R sequences. (From Peterlin and Luciw, 1988).
Early transcripts (doubly spliced)

Late transcripts (singly spliced and full-length)

(a) Transcription starts

(b) (+) signaling

(c) (-) signaling

TAR RNA

crs

car

crs

car

5'LTR

NRE

gag pol env

3'LTR

DNA

pol env

Signaling

Transcription ends
accumulated levels of *nef* decrease cellular signaling and viral transcription. The balance between the function of these viral regulators and their effects within the infected cell is never completely achieved, and as a result, the infected cell dies while enough virions infect other cells to insure survival of the virus (Franza *et al.*, 1988; Peterlin *et al.*, 1988; Peterlin and Luciw, 1989).

Much less is known about the function of the other small genes of HIV-1 and -2. *vif*, previously known as *sor*, *orfA*, *P*, or *Q*, encodes a 23-kda protein (Lee *et al.*, 1986; Kan *et al.*, 1986). Because cloned proviral genomes containing mutations in the *vif* gene yield low levels of infectious viral particles, the *vif* gene product has been suggested to be required for viral infectivity (Sodroski *et al.*, 1986; Strebel *et al.*, 1987). Precise functions for *vpr*, *vpu*, and *vpx* have yet to be defined.

Much has been learned in the past five years in understanding the molecular biology of lentiviruses. Relative to HIV-1, however, progress in discovering the patterns of gene expression of other lentiviruses has been lagging behind. In the case of visna virus, a triple-spliced viral mRNA, whose four exons contain sequences from the 5'-end of the viral genome, the 3'-end of *pol*, a short ORF immediately upstream of *env* (S), and sequences at the 3'-end of *env*, respectively, has been shown to encode viral *trans*-activating functions (Mazarin *et al.*, 1988; Davis and Clements, 1989). A similar double-spliced viral mRNA, whose three exons contain sequences from the 5'-end of the viral genome, *SI*, and *S3* ORFS of EIAV (Figure 6) has been reported to code for an EIAV *trans*-activator (Derse *et al.*, 1989). It is believed that by dissecting the functions of the
structural and regulatory genes of lentiviruses in general, and the HIV's in particular, measures can then be taken to interfere or block the action of such genes and their encoded products, thus providing therapeutic or preventative protocols against lentivirus-induced diseases, including AIDS.

The patterns of transcription of lentiviruses in infected cells have proven to be quite complex, involving multiple splicing events required for the generation of virus-encoded regulatory trans-acting proteins which, in turn, act on cis-acting sequences on the viral genome to enhance viral transcription or influence the ratio of spliced to unspliced RNA's. The objectives of the current study have been to answer some of the unanswered, yet fundamental, questions on the molecular biology of equine infectious anemia virus (EIAV).

RESEARCH OBJECTIVES

The main objective of the body of research presented here has been to characterize the transcriptional properties of EIAV in virus-infected cells. Questions regarding different aspects of this objective are listed below, along with their significance.

1. How many virus-specific transcripts are encoded by EIAV in virus-infected cells? What are the relative levels, sizes, and kinetics of synthesis of EIAV-specific transcripts? Moreover, are there any differences in patterns of EIAV transcription when comparing persistent to cytopathic EIAV infections? Answers to these questions are necessary in order to carefully examine the genetic properties of EIAV and the possibility of differential expression of the viral genome during persistent and cytopathic infections.
(2) What are the splicing patterns utilized in generation of the spliced EIAV mRNA's in infected cells? Can the location of the splice donor and acceptor sites used for generating the spliced viral mRNA's be mapped on the viral genome? Characterization of the viral mRNA splicing patterns is a prerequisite to defining the exact nucleotide sequence of these mRNA's, which in turn is required for analysis of the potential of these mRNA's to encode viral structural, replicative, and regulatory proteins.

(3) Do the short ORF's on the EIAV genome code for any proteins, such as the trans-acting proteins encoded by HIV-1 or visna virus, and if so what are the transcriptional and translational mechanisms used for their synthesis? To better understand the molecular biology of lentiviruses it is important to search for these regulatory proteins and to understand their function.

(4) Finally, the unique nature of the persistent disease induced by EIAV, in which the infected horse eventually brings the disease under control, provides an exciting venue for answering some of the complex questions on the biology of other important lentiviruses such as HIV-1. Do the determined patterns of EIAV gene expression correlate with the complex patterns observed in other lentivirus systems or are they in any way different or unique? Such information is vital to understanding patterns of lentiviral gene expression and designing diagnostic, therapeutic, and preventative measures for combating lentiviral diseases.

The first objective has been accomplished by analysis of RNA from EIAV-infected cells in Northern blot hybridizations using a battery of subcloned, genespecific proviral DNA probes obtained from the cloned EIAV provirus. The
splicing patterns of EIAV, outlined in objective (2), have been analyzed by cloning and sequencing of cDNA obtained from poly(A)$^+$ RNA of EIAV-infected cells, by Northern hybridization analysis using short oligonucleotide DNA probes complementary to sequences upstream of putative splice donor sites near the 5'-end of the EIAV gag gene, and also by S1-nuclease mapping studies of RNA from virus-infected cells. Characterization of splicing patterns and mapping of splice donor and acceptor sites on the viral genome has facilitated the determination of coding sequences within the spliced EIAV mRNA's, which in turn has allowed the design of experimental approaches for addressing objective (3). Subcloning of specific EIAV cDNA fragments encoding these mRNA's into plasmid vectors for in vitro transcription/translation or expression in eukaryotic cells has provided a means for preliminary analysis of the expressed protein products. Future detailed characterization of the expressed proteins, including an analysis of their reactivity with appropriate antisera, would then allow their assignment to specific EIAV genes.
CHAPTER 2

MATERIALS AND METHODS

Cell cultures and virus strains

Primary cultures of fetal equine kidney (FEK) and fetal donkey dermal (FDD) cells were prepared and maintained as described previously (Orrego et al., 1982). A prototype stock of EIAV, obtained by propagation of the Wyoming cell-adapted strain of EIAV (Malmquist et al., 1973) in FEK cells, was used to carry out infection of confluent monolayers of FEK cells at a multiplicity of infection of 1.0. In addition, an FDD-adapted stock of EIAV, prepared by propagation of prototype EIAV in FDD cells, was used to infect confluent monolayers of FDD cells at a multiplicity of infection of 1.0.

Isolation and purification of total and poly(A)^+ selected cellular RNA

To inactivate contaminating RNases, all solutions used for preparation of RNA were prepared using diethylpyrocarbonate (DEPC)-treated double-distilled water (ddH_2O). In addition, all solutions were subjected to a second round of DEPC treatment following their preparation, with the exception of Tris buffers. For DEPC treatment, 2 ml of DEPC was added to 100 ml of water or solution and shaken vigorously. After an overnight incubation at room temperature, the solution was autoclaved to remove remaining traces of DEPC. In addition, all glassware used for preparation of such solutions was sterilized by baking at 200°C for 8 hours.
Total cellular RNA was isolated from EIAV-infected or uninfected cells by a modified guanidinium thiocyanate extraction method (Chirgwin et al. 1979, Maniatis et al., 1982). Briefly, using repeated washes with phosphate buffered saline (PBS) containing 4mM EDTA (pH 7.0), infected FEK cells were harvested from roller bottles at 3-4 weeks post-infection, and infected FDD cells were recovered from flasks at the peak of the cytopathic effect (10-12 days after infection with FDD-adapted EIAV). The cells were pelleted in 50 ml polypropylene conical centrifuge tubes at 2,000 rpm for five minutes using an IEC HN-SII table-top clinical centrifuge, and immediately resuspended in 5-10 ml of guanidinium thiocyanate lysis buffer [4.0 M guanidinium thiocyanate, 0.5% sodium-N-lauroyl sarcosine, 5.0 mM sodium citrate (pH 7.0), 0.1 M 2-mercaptoethanol]. The cells were lysed by repeated passage through a 23-gauge needle, which also resulted in shearing of the chromosomal DNA. Four grams of solid CsCl was dissolved in each 10.0 ml of homogenate (0.4 g/ml final) and the homogenate was gently layered over a 3.25 ml cushion of sterile 5.7 M CsCl in sterilized 13.5 ml Sarstedt polyallomer ultracentrifugation tubes. Following ultracentrifugation for 18 hours at 28,500 rpm at 20°C in a Beckman SW40Ti rotor, the cellular homogenate and the CsCl cushion were carefully aspirated out and the inside walls of the tube were wiped dry by a sterile gauze. The light brown pellet of RNA was resuspended in 300 ul of 10.0 mM Tris.Cl (pH 7.5), 5.0 mM EDTA (pH 7.0), 1.0% SDS, and extracted twice with phenol:chloroform:isoamyl alcohol (25:24:1). The aqueous supernatant was brought up to 0.3 M in sodium acetate (pH 5.2) and precipitated by addition of three volumes of ethanol.
Poly(A)$^+$ RNA was purified by two cycles of oligo(dT)-cellulose chromatography of total cellular RNA (Maniatis et al., 1982), or by direct oligo(dT)-cellulose chromatography of cellular lysates using the Fast Track mRNA Isolation Kit from Invitrogen, according to the manufacturer's protocol.

For time-course studies on EIAV RNA synthesis, various time-points after infection were used for recovery of infected FEK and FDD cells, as described in Chapter 3.

Subcloning of EIAV proviral DNA fragments

EIAV proviral DNA fragments were obtained by restriction enzyme digestions of an EIAV proviral DNA clone, lambda 12 (Derse et al., 1987; Rushlow et al., 1986; Stephens et al., 1986). Restriction digestions were performed according to standard protocols (Maniatis et al., 1982) using the 10X digestion buffer provided by the enzyme's manufacturer and under the manufacturer's recommended reaction conditions. Typically, 5 to 10 ug of DNA was digested by 10 to 20 units of enzyme for at least 5 hours at the appropriate reaction temperature. Restriction fragments of EIAV proviral DNA were subjected to electrophoresis through 0.8-1.2% low melting-point agarose, normal agarose, or 6.0-8.0% polyacrylamide gels (Maniatis et al., 1982), following which the gels were stained with 0.5 ug/ml ethidium bromide for visualization of the DNA bands by UV illumination. After excision of appropriate DNA bands from gels, DNA was purified by chromatography through a column of NACS-52 resin (BRL) or by the use of a GeneClean Kit (Bio101). Purified viral DNA fragments were typically ligated to appropriately linearized plasmid vectors at a molar ratio of 20:1.
insert:vector under standard ligation reaction conditions (Maniatis et al., 1982).

Aliquots of the ligation reactions were used to transform competent *E. coli* JM101 or DH5α cells to antibiotic resistance (Hanahan, 1985). Recombinant clones were identified by restriction enzyme mapping of minilysate plasmid DNA prepared from 2 ml overnight cultures of antibiotic resistant clones by the alkaline lysis method (Birnboim and Doly, 1979; Maniatis et al., 1982).

Recombinant plasmids were then prepared on a large scale from 2 liter cultures of appropriate clones and purified by two cycles of equilibrium density ultracentrifugation through CsCl-ethidium bromide gradients (Maniatis et al., 1982).

**Labeling of probes for Northern hybridization analysis**

Recombinant plasmids containing subcloned EIAV proviral DNA fragments, which were to be used as probes for Northern hybridizations, were digested with appropriate restriction enzymes, and the EIAV-specific insert DNA fragments were purified following gel electrophoresis as described above. 50 to 100 ng of each fragment was then labeled with [α-32P]dCTP using a random primer labeling kit from BRL (Feinberg and Vogelstein, 1983). Oligonucleotide probes (see below) were prepared by 5'-end labeling of 200 ng of each oligonucleotide with [γ-32P]ATP and polynucleotide kinase (Maniatis et al., 1982). Unincorporated 32P-labeled ribo- or deoxyribonucleotides were removed from the labeling reactions by chromatography through Sephadex G-50 spun columns, as described (Maniatis et al., 1982). Typically, specific activities greater than 5X10⁸
cpm/ug for restriction DNA fragments and about $1 \times 10^9$ cpm/ug for oligonucleotides were obtained.

**Northern, Southern, and colony hybridization analyses**

For Northern hybridization analysis, 2.5 ug of poly (A)$^+$ RNA or 15.0 ug of total RNA from infected or uninfected cells was subjected to electrophoresis in 20X20 cm 1.0% or 1.4% agarose-formaldehyde gels in 50 mM MOPS (pH 7.0), 1 mM EDTA buffer (Gerard and Miller, 1986). Electrophoresis was carried out at 100 volts for 1 hour without buffer circulation and at 30 volts for 14 hours with buffer circulation. To check for the integrity of RNA samples, the gels were stained by soaking in a 5.0 ug/ml solution of ethidium bromide in water for 5 minutes in the dark and destained for 3 to 8 hours in water in the dark. The stained RNA bands were then visualized by short-wave UV illumination of the gel. For transfer of RNA, the staining step was omitted, and the gel was washed immediately following electrophoresis for 1 hour in 10X SSC (1X SSC is 0.15 M NaCl, 0.015 M sodium citrate) to remove the formaldehyde. The gel was then transferred overnight to GeneScreen or GeneScreenPlus nylon membranes (New England Nuclear-DuPont) in 10X SSC using the capillary Southern transfer method (Maniatis et al., 1982). Following the transfer, the membrane was baked for 2 hours at 80°C in a vacuum oven and prehybridized at 42°C for 2-4 hours in a buffer containing 50% formamide, 5X Denhardt (Maniatis et al., 1982), 6X SSC, 1% SDS, 250 ug/ml denatured salmon sperm DNA, and 25 mM sodium phosphate (pH 6.5). Hybridization was carried out at 42°C for 18-24 hours using $1 \times 10^6$ cpm of $^{32}$P-labeled probe per ml of hybridization solution, which contained
the same composition as the prehybridization buffer except that 1X Denhardt was used. Following hybridization, the membranes were washed for 30 minutes at room temperature in 2X SSC, 0.2% SDS, and for 30 minutes at 65°C in 0.2X SSC, 0.2% SDS. The membranes were then subjected to autoradiography at -70°C for 12-36 hours.

For Southern hybridization analysis, DNA fragments were fractionated by electrophoresis through 1.0% agarose gels. Denaturation, neutralization, and transfer from the gel onto a nitrocellulose membrane was performed according to standard protocols (Maniatis et al., 1982). Baking, prehybridization, hybridization, and autoradiography was exactly as described for Northern hybridizations above.

For colony hybridization analysis, individual *E. coli* colonies were picked by sterile toothpicks and streaked as short lines (2-3 mm long) on sterile circular nitrocellulose membranes placed on LB agar plates containing the appropriate antibiotic. A grid pattern taped to the back of the plate was used to identify the location of individual colonies on the membrane. The plate was inverted and incubated at 37°C until colonies had grown to 0.5-1.0 mm in width (6-8 hours). Lysis of the bacterial colonies on the membrane were performed as described (Maniatis et al., 1982). Baking, prehybridization, hybridization, and autoradiography of the membrane was exactly as described for Northern hybridizations above.

**Chemical synthesis and purification of oligonucleotides**

The four oligonucleotides used in this study were the generous gift of Dr. K. E. Rushlow, Battelle Memorial Institute, Columbus, Ohio. The first
oligonucleotide, used to prime first-strand cDNA synthesis, was a 30-mer and had the following sequence:

\[5'-ACTGAAGTAATTTTGAAATGGTCACACCC-3'\]

The sequence of the oligonucleotide is complementary to a 30 nucleotide-long sequence immediately downstream of the \textit{SmaI} site within the EIAV \textit{env} gene (positions 1642-1672 of Rushlow \textit{et al.}, 1986; Figure 11A). The other three oligonucleotides (SD-1, SD-2, and SD-3) were used as probes in Northern hybridization analyses and their primary sequence is shown in Figure 13.

The oligonucleotides were synthesized on an Applied Biosystems 380A DNA Synthesizer. Following synthesis, the deprotected oligonucleotides were purified by chromatography through oligonucleotide purification cartridges (Applied Biosystems) according to the manufacturer's protocol.

\textbf{cDNA synthesis and cloning}

Double-stranded cDNA was synthesized from 13.0 \mu g of poly(A)$^+$ RNA, obtained from EIAV-infected FDD cells, using a cDNA Synthesis Kit from BRL. The manufacturer's protocol was exactly followed, except that 3.0 \mu g of the synthetic 30-mer oligonucleotide described above was used to prime first-strand cDNA synthesis instead of the oligo(dT) primer provided in the kit. Following the synthesis of second-strand cDNA, the double-stranded cDNA products were precipitated with ethanol.

For cloning of cDNA, the double-stranded cDNA products were tailed with oligo(dC) and annealed to oligo(dG)-tailed, \textit{PstI} digested pUC9 DNA. Following this strategy a \textit{PstI} site is created at the 5'- and 3'-ends of the cloned
cDNA. The ethanol precipitated cDNA products were pelleted, dried, and resuspended in 10 ul sterile ddH₂O. The oligo(dC)-tailing reaction was set up by addition of 5 ul of 10 mM dCTP (1 mM final concentration) and 10 ul of the BRL 5X tailing buffer (1X final concentration) to the double-stranded cDNA. The volume of the mixture was brought up to 48.5 ul with sterile ddH₂O, after which the contents were mixed and the mixture was incubated in a 37°C water bath for 5 minutes. The tailing reaction was initiated by addition of 20 units (1.5 ul) of BRL Terminal Deoxynucleotidyl Transferase (50 ul final reaction volume) and the mixture was incubated at 37°C. Five ul aliquots were removed from the reaction at 30 second intervals (10 aliquots), and to inactivate the enzyme each aliquot was transferred to individual microfuge tubes pre-incubated in a 70°C water bath for 10 minutes. The 20 ul annealing reaction for each aliquot was set up by addition of 2 ul of 10X annealing buffer [100 mM Tris.Cl (pH 7.5), 1,000 mM NaCl, 1 mM EDTA (pH 7.5)], 11 ul of sterile ddH₂O, and 2 ul of 11 ng/ul oligo(dG)-tailed, PstI digested pUC9 (Pharmacia). Annealing was carried out at 58°C for 3 hours. After a brief centrifugation step for collection of residual water on the inside walls of the tubes, the annealing reaction was continued by incubation at room temperature overnight.

A 5 ul aliquot from each 20 ul annealing reaction was used to transform BRL library efficiency competent DH5α E. coli cells (Hanahan, 1985). Transformants were plated on LB plates (Maniatis et al.) containing 100 ug/ml ampicillin, 80 ug/ml X-gal, and 10 ug/ml IPTG. Plates were incubated for 12
hours at 37°C, and colonies exhibiting the white phenotype were randomly picked for colony hybridization analysis, as described above.

**Nucleotide sequencing**

Double-stranded DNA fragments to be sequenced were isolated and purified on 1.0-1.2% agarose gels, as described above, and subcloned into the replicating form (RF) DNA of M13mp18 or M13mp19 vectors. Single-stranded DNA from the M13 phage particles was sequenced by the dideoxy chain termination method (Sanger *et al.*, 1977), following which the sequencing reaction mixtures were subjected to electrophoresis on 10% polyacrylamide-8% urea sequencing gels (BRL).

**S1 nuclease mapping**

The four probes used for S1 nuclease mapping of RNA from EIAV-infected and uninfected FDD cells are described in detail in Chapter 3 and schematically diagrammed in Figure 16. To create the M13 "tag" sequence at the intronic end of probe p1, a 308-bp *TaqI-PvuII* EIAV proviral DNA restriction fragment was subcloned into M13mp18 RF DNA digested with *AccI* and *SmaI*. The cloned DNA was then excised out as a 550-bp *NarI* fragment containing a 300-bp *NarI-PvuII* proviral DNA fragment and a 250-bp M13mp18 *SmaI-NarI* tag attached from its *SmaI* end to the *PvuII* site (Figure 16B).

Similarly, probe p2 was constructed by cloning a 450-bp *NcoI-BamHI* EIAV proviral DNA restriction fragment, which had been previously blunt-ended at the *NcoI* site by treatment with T4 DNA polymerase (Maniatis *et al.*, 1982), into M13mp19 RF DNA digested with *BamHI* and *SmaI*. The cloned DNA was then
excised out as a 611-bp *BamHI-BglII* containing a 450-bp *BamHI-NcoI* proviral DNA fragment and a 161-bp M13mp19 *SmaI-BglII* tag attached from its *SmaI* end to the blunt-ended *NcoI* site (Figure 16C).

Probes p3 and p4 did not contain any tag sequences and were directly used as 617-bp *PvuII-HindIII* and 385-bp *DraI-Scal* proviral DNA restriction fragments.

Probe p1 was prepared by labeling the 550-bp *NarI* fragment at its 3'-ends in a fill-in reaction with [α-32P]-dCTP and the Klenow fragment of *E. coli* DNA polymerase I (Maniatis et al., 1982). Probe p2 was prepared by treatment of the 611-bp *BamHI-BglII* fragment with calf intestinal phosphatase (Maniatis et al., 1982), and labeling of the dephosphorylated 5'-ends with [γ-32P]ATP and polynucleotide kinase (Maniatis et al., 1982). Probe p3 was prepared by labeling the 617-bp *PvuII-HindIII* fragment at its 3'-ends with [α-32P]dATP and [α-32P]dCTP by replacement synthesis with T4 DNA polymerase (O'Farrell, 1981). Finally, probe p4 was prepared by treatment of the 385-bp *DraI-Scal* fragment with calf intestinal phosphatase (Maniatis et al., 1982), and labeling of the dephosphorylated 5'-ends with [γ-32P]ATP and polynucleotide kinase (Maniatis et al., 1982).

For S1 nuclease mapping, a modification of procedures by Nash et al. (1984) and Ausubel et al. (1987) was used. About 1X10^6 Cerenkov counts of each probe were added to 40 μg of total cellular RNA from EIAV-infected or uninfected FDD cells, and the mixture was precipitated with ethanol. The precipitate was pelleted, washed with 70% ethanol-30% DEPC-treated ddH₂O, and air-dried. The pellets were thoroughly dissolved in 30μl of S1 hybridization
buffer [80% deionized formamide, 40 mM PIPES (pH 6.4), 400 mM NaCl, 1 mM EDTA (pH 8.0)], and the resuspended RNA-probe mixture was immersed in an 85°C water bath for 15 minutes. The hybridization mixture was then immediately transferred to water baths adjusted to 50°C (in the case of probes p1 and p3) or 55°C (in the case of probes p2 and p4) and incubated overnight. The following mix was then added to the hybridization reaction: 150 ul 2X S1 nuclease buffer [0.56 M NaCl, 0.1 M NaOAc (pH 4.5), 9 mM ZnSO₄], 3 ul 2.0 mg/ml single-stranded calf thymus DNA, 147 ul DEPC-treated ddH₂O, and 300 units S1 nuclease (BRL). The contents were mixed and incubated at 30°C for 60 minutes, after which the S1 digestion was stopped by addition of 80 ul of S1 stop buffer [4.0 M ammonium acetate, 20 mM EDTA (pH 8.0), 40 μg/ml yeast tRNA]. The digest was extracted twice with phenol:chloroform:isoamyl alcohol (25:24:1), and the aqueous supernatant was precipitated with ethanol. To remove traces of salt, the ethanol precipitated samples were pelleted, resuspended in sterile ddH₂O, and the ethanol precipitation was repeated two more times. The final pellet was washed with 70% ethanol, dried in a Speedvac evaporator, and resuspended in 3 ul of TE buffer (Maniatis et al., 1982). Four ul of formamide loading dye [95% deionized formamide, 20 mM EDTA (pH 7.5), 0.05% bromophenol blue, 0.05% xylene cyanol] was added, the sample was boiled for 4 minutes, and quenched immediately on ice. The entire sample was then subjected to electrophoresis through a 6.0% polyacrylamide-8.0 M urea sequencing gel. Electrophoresis was typically stopped just as the bromophenol blue reached the bottom of the gel.
MspI-digested pBR322 DNA, labeled with [α-32P]dCTP by filling-in with Klenow (Maniatis et al., 1982), was run in parallel lanes as size marker.

**In vitro transcription and translation**

Construction of the pSP65 plasmid containing the 2980-bp BamHI-PstI S2-env cDNA insert is diagrammed in Figure 19. Briefly, a ligation reaction containing the 270-bp SmaI-BamHI splice-junction cDNA, a 2710-bp BamHI-PstI cDNA fragment from a separate cDNA clone (#3006), and pUC19 DNA digested with SmaI and PstI was carried out [The PstI end of the second cDNA fragment had been originally created through cloning of that cDNA into pUC9 by oligo(dC)-(dG) tailing]. As a result, a 2980-bp SmaI-PstI hybrid cDNA insert was cloned into pUC19. Appropriate recombinants were identified by restriction enzyme mapping and the 2980-bp insert was excised out of pUC19 by digestions with SacI and PstI. The 2980-bp SacI-PstI fragment was then cloned into pSP65 DNA which had been digested by SacI and PstI. Appropriate recombinants were again identified by restriction enzyme mapping.

To prepare templates for *in vitro* transcription, the pSP65 plasmid containing the 2980-bp SacI-PstI cDNA insert was then linearized by either HindIII or MluI (Figure 19). Capped transcripts were synthesized in 50 ul reactions, containing the cap analog GpppG (Pharmacia), using an *in vitro* transcription kit (Promega Biotec) and according to the manufacturer’s protocol, based on previously published procedures (Krieg and Melton, 1984). *In vitro* translation reactions of the synthetic transcripts were performed in wheat germ extracts in the presence of 35S-methionine, as described previously (Bartlett et al.,
Incorporation of $^{35}$S-methionine was determined by assaying an aliquot of the translation reaction for acid-insoluble radioactivity using Whatman 3mm filter paper discs.

**Construction of the p9SR426 expression plasmid, transfections, and indirect immunofluorescence assays**

A 426-bp *SspI*-*SmaI* restriction fragment of the EIAV proviral DNA, containing the entire *S2* coding sequence and the first 379 nucleotides of the *env* gene, was cloned into the p91023(B) expression vector (Figure 21). In addition to its unique *EcoRI* cloning site, the p91023(B) vector also contains a unique *BglII* cloning site immediately upstream of the *EcoRI* site. The 426-bp *SspI*-*SmaI* fragment was initially cloned into the *SmaI* site of plasmid TrpSF9, and the orientation of the insert in the recombinant clones was identified by restriction enzyme mapping. The insert DNA was then excised as a *BglII*-*EcoRI* fragment and ligated to p91023(B) DNA digested with *BglII* and *EcoRI*. The recombinant plasmid, termed p9SR426, was then grown on a large scale and subjected to two cycles of equilibrium density ultracentrifugation through CsCl-ethidium bromide gradients (Maniatis *et al.*, 1982). COS-7 cells were grown on sterile tissue culture chamber slides to about 50% confluency. Five ug of the plasmid was transfected onto COS-7 cells as a calcium phosphate co-precipitate followed by a 15% glycerol shock, as described previously (Fordis and Howard, 1987). Indirect immunofluorescence assays were performed at 48 hours post-transfection, using monoclonal antibody #86 or rabbit polyclonal *S2b* antibody, according to previously published procedures (Pereira *et al.*, 1980).
CHAPTER 3
RESULTS

I. IDENTIFICATION OF EIAV-SPECIFIC TRANSCRIPTS IN INFECTED CELLS

The cell-adapted Wyoming (prototype) strain of EIAV establishes a persistent infection in fetal equine kidney (FEK) cells grown in tissue-culture, leading to continuous virus production for up to 4-5 weeks after the initial infection, during which no cytopathic effects are exerted by the replicating virus on the infected cells. Infection of fetal donkey dermal (FDD) cells by FDD-adapted EIAV, a strain of the virus obtained by propagation of prototype EIAV in FDD cells, however, results in a distinct and observable cytopathic effect characterized by rounding and detachment of infected cells at 10-12 days post-infection.

By comparison to other retroviruses, transcription of the EIAV provirus initiates at the 5'-terminus of the R region in the 5'-LTR. Since the R-U5-leader sequences are present in both full-length genomic and spliced viral mRNA's, a proviral DNA fragment containing nucleotide sequences from the R-U5 region of the LTR would share sequence homology with the 5'-ends of all viral mRNA's. Such a DNA fragment, hereafter referred to as the "5'-LTR probe", should hybridize to all virus-specific transcripts and once radiolabeled can be used as a general probe for detection of all viral mRNA's actively transcribed in infected cells.
A 226-bp restriction fragment of the EIAV provirus, beginning at 50 nucleotides upstream of the R region in the 5'-LTR (within the U3 sequences) and extending 63 nucleotides downstream of the U5 region, was radiolabeled with \(^{32}\)P and used as the 5'-LTR probe in Northern blot hybridization analysis of EIAV-specific transcripts during persistent and cytopathic infection of FEK and FDD cells, respectively. Total cellular RNA was isolated from FEK cells at three weeks after infection with prototype EIAV, from FDD cells at the peak of cytopathic effects at twelve days after infection with FDD-adapted EIAV, and from uninfected FEK and FDD cells as negative controls. Fifteen micrograms of total RNA from each source was subjected to electrophoresis in a 1.0% denaturing agarose-formaldehyde gel, transferred to the hybridization membrane, and hybridized with the 5'-LTR probe. The left panel in Figure 8 shows the ethidium bromide-stained agarose-formaldehyde gel of total RNA samples, while an autoradiogram of the Northern blot of the same gel hybridized to the 5'-LTR probe is presented in the right panel of Figure 8. In both infected cell-lines, two major species of hybridizing EIAV-specific transcripts are detected: an 8.2-kb genome-length RNA and a 3.5-kb RNA. Also detected in RNA from both infected cell lines is a low abundance 1.5-kb RNA species which migrates just below the 18S rRNA band on the gel. No EIAV-specific transcripts are detected in RNA from uninfected cells, indicating that there is no endogenous expression of EIAV-related RNA in these cells and that the hybridizing RNA species present in RNA from infected cells are a result of EIAV infection.
Figure 8. Northern hybridization analysis of total RNA from EIAV-infected and uninfected FEK or FDD cells. Left. 1.0% Agarose-formaldehyde gel of total cellular RNA from FEK cells infected with prototype EIAV (lane 1), FDD cells infected with FDD-adapted EIAV (lane 2), uninfected FEK cells (lane 3), and uninfected FDD cells (lane 4). Lane M contains molecular weight markers (0.3- to 9.5-kb RNA ladder of BRL) whose mobilities are indicated on the left. Right. Northern blot of the gel on the left panel, hybridized to the 226-bp *MluI-BamHI* 5'-LTR probe. The position of 28S and 18S rRNA bands on both the gel and the blot are indicated. The numbers on the right correspond to the molecular weight values for the three observed species of EIAV mRNA's.
Densitometric analysis of the Northern blot in Figure 8 reveals that there is almost a 10-fold higher level of viral RNA's in infected FDD cells at the peak of the cytopathic effect relative to that in persistently infected FEK cells. In addition, the results indicate that the 8.2- and 3.5-kb RNA are present at nearly equal concentrations in infected FEK cells, while the concentration of the 3.5-kb RNA is about three-fold greater than that of the full-length transcript.

Immunofluorescence studies have shown the EIAV infection of FDD cells to be asynchronous, resulting in infected cells comprising about 30% of the total cell population, while during infection of FEK cells nearly 100% of the cell population is infected by the virus (Issel and Montelaro, unpublished data). Taking into account the asynchronous nature of viral infection in FDD cells, it appears that the levels of EIAV-specific RNA's are nearly 30-fold higher during the cytopathic infection of FDD cells relative to the persistent infection of FEK cells. Moreover, the relatively high levels of the 3.5-kb mRNA in infected FDD cells suggests enhanced levels of 8.2-kb genome-length viral RNA splicing during this cytopathic EIAV infection.

The presence of abundant quantities of 28S and 18S rRNA in samples of total cellular RNA, as clearly seen in Figure 8A, introduces inherent problems in identification of specific RNA's present on a hybridization membrane which share sequence homology with the hybridization probe. In the absence of the rRNA species, the probe would hybridize to a background "smear" of degraded homologous RNA on the membrane, in addition to the specific non-degraded RNA species. However, when the abundant non-hybridizing rRNA species are
present on the membrane, they physically block the hybridization of the probe to
the background smear of degraded homologous RNA at the position of each
rRNA species on the membrane, resulting in the re-appearance of the
background immediately below the lower boundary of each rRNA which then
appears as a faint hybridizing band on the Northern blot. This band is present
just below the position of 28S and 18S rRNA in Figure 8B, and has been
observed in repeated experiments involving Northern hybridization of total RNA
samples.

To clarify this inherent problem and also to better detect any EIAV-
specific low molecular transcripts which migrate slightly faster than 18S rRNA
during denaturing gel electrophoresis, poly(A)^+ RNA, which contains little or no
rRNA, was isolated and purified from uninfected and EIAV-infected FDD cells
and used as substrate for Northern hybridization analysis. Furthermore, to identify
the gene-specific content of the EIAV-specific mRNA’s in infected cells, a battery
of subgenomic EIAV DNA fragments (Table II), in addition to the 5'-LTR 226-
bp fragment, were used as probes in Northern hybridization analyses. Using
unique restriction sites on the EIAV proviral DNA, the DNA fragments were
individually subcloned so that each, with the exception of the 5'-LTR probe,
represents sequences from each of the seven genes (including the short ORF’s) of
EIAV. By designing the gene-specific probes spanning across the EIAV genome,
separate Northern hybridizations were carried out using each individual probe,
and the identity of viral mRNA’s containing sequences complementary to each
viral gene were identified. Figure 9 shows a composite of the location of the
Table II. Panel of EIAV proviral DNA restriction fragments used as probes in Northern hybridization analyses.

<table>
<thead>
<tr>
<th>Probe</th>
<th>Restriction Fragment</th>
<th>Specificity within the EIAV genome</th>
<th>EIAV mRNA Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>226-bp MluI-BamHI</td>
<td>5' - LTR</td>
<td>All mRNA's</td>
</tr>
<tr>
<td>B</td>
<td>393-bp PvuII-KpnI</td>
<td>gag</td>
<td>Full-length</td>
</tr>
<tr>
<td>C</td>
<td>511-bp SmaI-SsrI</td>
<td>pol</td>
<td>Full-length</td>
</tr>
<tr>
<td>D</td>
<td>102-bp PvuII-SspI</td>
<td>S2</td>
<td>Full-length, S1, env*</td>
</tr>
<tr>
<td>E</td>
<td>114-bp BamHI-TaqI</td>
<td>S2</td>
<td>Full-length, S2, env</td>
</tr>
<tr>
<td>F</td>
<td>417-bp SphI-XbaI</td>
<td>env</td>
<td>Full-length, env</td>
</tr>
<tr>
<td>G</td>
<td>320-bp AvaII</td>
<td>S3</td>
<td>Full-length, env, S3</td>
</tr>
</tbody>
</table>

* The hybridization of this probe to the env mRNA would be dependent on the env splice acceptor site being located upstream of the SspI site.
Figure 9. Northern hybridization analysis of EIAV-specific poly(A)$^+$ RNA's in FDD cells infected with FDD-adapted EIAV. A diagram of the EIAV proviral DNA genome and the location of the six open reading frames is represented on top of the figure. Based on DNA sequence data (Rushlow et al., 1986; Stephens et al., 1986) solid triangles represent the location of consensus splice donor sites and open triangles represent the location of consensus splice acceptor sites. The stippled bars, labeled A through G, show the location and size of the subgenomic probes relative to the EIAV genome. The bottom portion of the figure shows the results of hybridization of 2.5-ug of poly(A)$^+$ RNA from FDD cells infected with FDD-adapted EIAV, isolated at the peak of the cytopathic effect, to each of the subgenomic probes A-G. Panel H, RNA from uninfected FDD cells. The positions of the 8.2-, 3.5-, and 1.5-kb viral mRNA's are indicated. Panel A1 represents a Northern blot of a 1.5% agarose-formaldehyde gel, while the remaining panels, including A2, are blots of a 1.0% gel. The molecular weight markers used in this experiment were the 0.3- to 9.5-kb RNA Ladder (BRL).
various probes on the EIAV genome and their pattern of hybridization to Northern blots of poly(A)$^+$ RNA from FDD cells infected with FDD-adapted EIAV. As shown in Figure 9, panels A1 and A2, three species of viral RNA were identified in poly(A)$^+$ RNA of infected FDD cells following hybridization to the 5'-LTR probe: an 8.2-kb genome-length mRNA, a 3.5-kb mRNA, and a low-abundance 1.5-kb mRNA. Panel A1 is a Northern blot of a 1.4% agarose-formaldehyde gel, with the higher percentage gel resulting in a sharpened 1.5-kb hybridizing RNA band, while panel A2 is a blot of a 1.0% agarose-formaldehyde gel. As demonstrated before, no viral-specific RNA species were detected in RNA from uninfected cells (Figure 9, panel H).

Northern blots of poly(A)$^+$ RNA from infected FDD cells, run on a 1.0% agarose-formaldehyde gel, were also hybridized to a panel of probes spanning the EIAV genome (probes B through G). As shown in Figure 9, the gag- and the pol-specific probes (B and C) hybridize only to the full-length 8.2-kb viral mRNA, indicating that this transcript represents both the gag-pol message as well as the genomic RNA. The env-specific probe (F) hybridizes to the full-length viral mRNA, in addition to the 3.5-kb mRNA, suggesting that the 3.5-kb transcript may be the single-spliced mRNA encoding the EIAV env gene. Moreover, although not clearly observed in panel F due to the hybridization of the probe to a smear of degraded RNA below the 3.5-kb mRNA band, the results of separate experiments have revealed that probe F doesn't hybridize to the 1.5-kb mRNA (data not shown). This indicates that probe F is complementary to an intron within the env gene which is not present in the 1.5-kb mRNA.
Each of the subgenomic probes representing sequences from the three short open reading frames (ORF's) S1, S2, and S3 (Dorn & Derse, 1988; Rushlow et al., 1986) (probes D, E, and G) hybridizes to all three virus-specific transcripts, suggesting that the 1.5-kb mRNA is presumably a double-spliced message containing sequences from the 5'-end of the genome and each of the ORF's S1, S2, and S3. It should be noted that a low molecular weight hybridizing RNA band, migrating at about 0.8- to 1.0-kb, is also detected by the 5'-LTR probe and each of the probes representing sequences from S1, S2, and S3 (probe A in panel A2, and probes D, E, and G). Because this band is not readily detected in Northern blots of 1.4% gels (e.g., probe A in panel A1), it is uncertain whether it actually represents a viral mRNA species or is a degradation product of the larger mRNA's. Thus in this cytopathic EIAV infection three distinct virus-specific RNA species are detected in infected cells. Moreover, the levels of the 3.5-kb viral transcript appears to be much more abundant than that of the 8.2-kb or the 1.5-kb RNA's, as observed previously by Northern blot analysis of total cellular RNA from infected FDD cells (Figure 8).

The differential steady-state levels of the viral RNA's detected late in this cytopathic EIAV infection could be attributed to a temporally regulated, differential splicing of the genome-length RNA, resulting in relatively different concentrations of the various viral RNA species late during the infection. To examine this possibility, a time-course study was performed during which total cellular RNA isolated from FEK cells at 3-, 5-, 7-, and 9-days after infection with prototype EIAV and from FDD cells at 2-, 4-, 6-, 8-, 10-, and 12-days after
infection with FDD-adapted EIAV, was subjected to Northern blot analysis using a mixture of probes A and G. As shown in Figure 10A, no viral transcripts are detected in infected FEK cells until 9-days after infection, and the earliest detectable RNA species, present at almost equal concentrations, are the full-length 8.2-kb and the spliced 3.5-kb transcripts. In infected FDD cells, however, viral transcripts are detectable as early as 2-days after infection, and the increase in their concentrations reaches a plateau at the sixth day of infection (Figure 10B). It is evident that overproduction of the 3.5-kb spliced RNA occurs early during the cytopathic infection of FDD cells and does not appear to be temporally regulated. Moreover, it appears that the synthesis of the 1.5-kb mRNA may be temporally regulated as the levels of this transcript increase during the course of the cytopathic infection, with the highest level being detected at 12-days after infection.

II. CHARACTERIZATION OF THE SPlicing PATTERNS OF EIAV-SPECIFIC mRNA'S IN INFECTED CELLS

cDNA Cloning and Sequencing

Nucleotide sequence analysis of the EIAV proviral DNA (Rushlow et al., 1986; Stephens et al., 1986) has revealed the presence of putative splice donor and acceptor sites located throughout the viral genome. A schematic diagram of the location of the various putative splice donor and acceptor sites in relation to the six open reading frames of EIAV is presented in Figure 11A.
Figure 10. Time-course study on the synthesis of EIAV-specific transcripts. The numbers above each lane in panels A and B correspond to the time-points at which total RNA was isolated from infected cells (days post-infection). A. Northern blot of total RNA isolated from FEK cells infected with prototype EIAV. B. Northern blot of total RNA isolated from FDD cells infected with FDD-adapted EIAV. In both panels 15.0 ug of total RNA, isolated at each time-point, was loaded on the appropriate lane of the gel. The positions of the 8.2-kb and the 3.5-kb viral RNA’s and of the 28S and 18S rRNA’s are indicated.
Figure 11. Identification of clones containing the splice-junction cDNA of EIAV 3.5-kb mRNA.

A. Schematic diagram of the EIAV genome. Solid and open triangles represent the location of putative splice donor and acceptor sites, respectively. The location of the region within the env gene, to which the 30-mer oligonucleotide primer used for priming of first-strand cDNA synthesis is complementary, is also indicated. The small bar below ORFS2, denoted as "B-T", represents the 114-bp BamHI-TaqI fragment of ORFS2 used for the initial screening of the cDNA library via colony hybridization. The small bar below the 5'-LTR, denoted as "M-B", represents the 226-bp MluI-BamHI 5'-LTR probe used for the secondary screening of the 16 EIAV-specific cDNA clones via Southern blot analysis.

B. Colony hybridization analysis of 213 random clones from the cDNA library which exhibited white phenotype on media containing X-gal. The probe used for this experiment was the $^{32}$P-labeled 114-bp BamHI-TaqI fragment of ORFS2. Closed arrows point to positive controls for colony hybridization to the probe, which were E. coli DH5α colonies transformed with a recombinant pUC13 plasmid containing the 114-bp BamHI-TaqI fragment of ORFS2 as insert and exhibiting a white phenotype on media containing X-gal. These positive controls also served as markers on the autoradiogram for determining its correct orientation relative to the location of the clones taken from the cDNA library. Open arrows point to negative controls for colony hybridization to the probe,
which were *E. coli* DH5α colonies transformed with pUC19 plasmid and exhibiting a blue phenotype on media containing X-gal.

C. Ethidium bromide-stained 1.0% agarose gel of *PstI*-digested minilysate plasmid DNA isolated from the 16 EIAV-specific cDNA clones (lanes 1-16) which were identified in panel B by colony hybridization. Lane C contains a positive control for Southern blot hybridization of the gel to the 226-bp *MluI*-BamHI probe, which was an *MluI*/BamHI double-digested recombinant pIBI76 plasmid containing the 226-bp *MluI*-BamHI fragment as insert. Arrows above the gel point to cDNA clones 4 and 13, whose inserts were not cut out by *PstI*.

D. Southern blot of the gel in panel C, hybridized to the 32P-labeled 226-bp *MluI*-BamHI probe. Arrows above the blot point to cDNA clones 4 and 13, whose plasmid DNA strongly hybridized with the probe.
30-mer primer for 1st-strand cDNA synthesis
With the availability of the Northern hybridization data, it was clear that the single-spliced 3.5-kb env mRNA is the predominant viral transcript during cytopathic infection of FDD cells. Moreover, since the 102-bp \textit{PvuII-SspI} fragment of \textit{ORFS1} was shown to hybridize to the 3.5-kb mRNA (Figure 9, panel D), it was clear that the splice acceptor site for this mRNA is located upstream of the \textit{SspI} site and thus, as expected, 5' to the translational initiation codon of the env gene. Thus the exact splicing pattern of this mRNA was investigated by identification and sequencing of cDNA clones in a cDNA library constructed from poly(A)$^+$ RNA of EIAV-infected FDD cells.

The cDNA library was constructed by utilizing a synthetic oligonucleotide, complementary to a 30-nucleotide-long sequence 400 nucleotides downstream of the initiation codon of the env gene, as the primer for first-strand cDNA synthesis (Figure 11A). The primer was designed to hybridize to RNA sequences within the env reading frame. Thus it was expected to hybridize to both the unspliced, 8.2-kb full-length mRNA, and the spliced, 3.5-kb env mRNA. Moreover, the 5'-to-3' direction of first-strand cDNA synthesis, initiated from the primer towards the 5' end of its complementary viral mRNA, was expected to generate cDNA copies of both the 8.2- and the 3.5-kb mRNA's. Since the 3.5-kb mRNA splice acceptor site was known to be situated upstream of the env gene initiation codon, therefore cDNA copies of this mRNA were expected to span the splice junction of this transcript.

Following the synthesis of both first- and second-strand cDNA products, the double-stranded cDNA products were cloned into the plasmid vector pUC9.
and propagated in *E. coli* DH5α. Previous cDNA cloning experiments from poly(A)⁺ RNA of EIAV-infected cells had revealed that the first-strand cDNA synthesis reaction, under the conditions employed, results in a majority of the cDNA products being synthesized from cellular rather than viral mRNA's (data not shown), most possibly due to non-specific hybridization of the oligonucleotide primer to cellular mRNA's. To identify recombinant clones containing viral cDNA inserts, 213 independent clones of the library, exhibiting a white phenotype on media containing X-gal, were randomly selected and initially screened by colony hybridization using the ³²P-labeled 114-bp *Bam*HI-*Taq*I DNA fragment of the *env* gene as the hybridization probe (Figure 11A). This fragment contains DNA sequences immediately upstream of the priming site for first-strand cDNA synthesis, and the use of it as probe in the initial screening of the library permitted the identification of those clones which contained cDNA inserts generated from the priming site on EIAV-specific poly(A)⁺ RNA's. The result of colony hybridization to the 114-bp *Bam*HI-*Taq*I probe is presented in Figure 11B, showing DNA from 16 of the 213 bacterial clones hybridizing to the probe.

Because each clone was individually picked from the cDNA library and streaked on the nitrocellulose membrane used for colony hybridization analysis, differential intensities of the hybridization signal from the various clones are due to quantitative differences in sizes of the various colonies after their growth on the membrane, resulting from differences in either the growth rate of the individual clones or the number of viable bacteria transferred onto the membrane.
A splicing event joins the R-U5-leader sequences from the 5'-end of the full-length genomic RNA to the splice acceptor site upstream of the \textit{env} gene. Therefore it was apparent that among the 16 EIAV-specific cDNA clones, shown by colony hybridization to contain cDNA sequences located between the \textit{BamH}I and \textit{TaqI} sites of the \textit{env} gene, those possessing sequences spanning the splice junction of the 3.5-kb mRNA must also contain sequences from the 5'-region of the EIAV genome upstream of the \textit{gag} gene. Thus the 16 EIAV-specific cDNA clones were subjected to a second round of screening using the radiolabeled 226-bp \textit{MluI-BamH}I fragment as the hybridization probe (Figure 11A). The \textit{BamH}I-end of the \textit{MluI-BamH}I probe is situated upstream of the three putative splice donor sites near the 5'-end of the \textit{gag} gene (Figure 11A), therefore it was clear that those clones hybridizing to this probe would contain cDNA sequences extending upstream of the \textit{BamH}I site and thus would also span the splice junction of the 3.5-kb mRNA.

The cloning strategy used for the generation of the cDNA library was based on oligo(dC)-tailling of the double-stranded cDNA products, in the presence of deoxycytidine and terminal deoxynucleotidyl transferase, followed by annealing of the oligo(dC)-tailed cDNA to \textit{PstI}-linearized, oligo(dG)-tailed pUC9. In such a manner, a \textit{PstI} site is created at each end of the cDNA insert within the pUC9 vector, allowing recovery of the cDNA insert by a \textit{PstI} digestion of the plasmid DNA. Thus minilysate plasmid DNA, prepared from the 16 positive clones, was digested with \textit{PstI} to release the cDNA inserts, fractionated by electrophoresis through a 1.0% agarose gel, and Southern blotted using the 226-bp \textit{MluI-BamH}I
probe. Figure 11C shows the ethidium bromide-stained 1.0% agarose gel of the 
PstI-digested minilysate plasmid DNA from the 16 EIAV-specific cDNA clones 
(lanes 1-16). As a positive control for the hybridization of the Southern blot of 
the same gel to the 226-bp *MluI-BamHI* probe, the pIBI76 plasmid DNA 
containing the 226-bp *MluI-BamHI* fragment was digested with *MluI* and *BamHI* 
and loaded on lane C of the gel. Digestion of this plasmid with *MluI* and *BamHI* 
yields two DNA fragments: a 4200-bp linearized pIBI76 vector DNA and a 226-
bp *MluI-BamHI* fragment (Figure 11C, lane C). Digestion of the minilysate 
plasmid DNA from the 16 EIAV-specific cDNA clones with *PstI* yields a 2700-bp 
linearized pUC9 vector DNA, a 1500- to 1700-bp undigested plasmid DNA, and 
the cDNA inserts of the individual clones, ranging from 400- to 1000-bp in size 
(Figure 11C, lanes 1-16). Interestingly, the cDNA inserts in plasmid DNA of 
clones 4 and 13 (Figure 11C, lanes 4 and 13, marked by arrows) are not excised 
from the vector following the *PstI* digestion. Additional restriction enzyme 
mapping of plasmid DNA from these two clones indicated that they were 
identical and revealed a deletion of one of the two *PstI* sites present at each end 
of the cDNA insert, thus explaining the inability of the *PstI* digestion to release 
their cDNA inserts.

The Southern blot of the same gel, hybridized to the 226-bp *MluI-BamHI* 
probe, is presented in Figure 11D. Out of the 16 EIAV-specific cDNA clones, 
only plasmid DNA from clones 4 and 13, which contain a deleted *PstI* site, 
hybridizes with this probe (Figure 11D, lanes 4 and 13, marked by arrows). As 
can be seen, the probe hybridizes with two specific DNA bands in lanes 4 and 13.
of Figure 11D. The slower migrating band is the PstI-linearized plasmid DNA of each of the two clones, while the faster migrating band is undigested plasmid DNA. Since the cDNA insert is not excised by PstI due to a deleted PstI site in plasmid DNA of these two clones, the cDNA sequence is therefore present in both the PstI-linearized and the undigested plasmid DNA bands. Thus the probe, recognizing the cDNA sequence present in both DNA bands, strongly hybridizes with both. Moreover, the particular preparation of the 226-bp MluI-BamHI probe used in this experiment also contained low amounts of contaminating vector DNA, as evidenced by the weak hybridization of the probe to the pIBI vector DNA band in addition to the 226-bp MluI-BamHI fragment in the control sample (Figure 11D, lane C). Since the labeled contaminating pIBI vector DNA present in the probe preparation also contains a pBR322-derived sequence present in pUC9 DNA, it also recognizes the linearized and undigested pUC9 DNA fragments of clones 1-3, 5-12, and 14-16 (Figure 11D, lanes 1-3, 5-12, 14-16), thus explaining the weak hybridization of these fragments to the probe. The hybridization of plasmid DNA from these latter 14 clones only to the 114-bp BamHI-TaqI probe (Figure 11B), and not to the 226-bp MluI-BamHI probe (Figure 11D), indicated that their cDNA inserts had primed from the 8.2-kb full-length EIAV mRNA and thus had extended into the pol gene instead of spanning the splice junction of 3.5-kb mRNA.

The hybridization of plasmid DNA from clones 4 and 13 to the 226-bp MluI-BamHI probe indicated that they contained cDNA inserts extending upstream of the gag gene and thus spanning the splice junction of the 3.5-kb
mRNA. This finding was further confirmed by restriction enzyme mapping of plasmid DNA from these two clones using restriction enzymes SmaI and BamHI. The distance between the SmaI site upstream of the gag gene and the BamHI site near the 5'-end of the env gene on the EIAV proviral DNA is 4945-bp. However, a SmaI/BamHI double-digestion of plasmid DNA from both clones had yielded a 270-bp fragment (data not shown), clearly indicating that a splicing event between the SmaI site, upstream of the gag gene, and the BamHI site, near the 5'-end of the env gene, had reduced the distance separating the two restriction sites from 4945-bp on the provirus to 270-bp on the cDNA.

The 270-bp SmaI-BamHI fragment from both cDNA clones was subcloned into M13mp18 and M13mp19 RF DNA, allowing nucleotide sequencing of both strands of each clone by the dideoxy chain termination method (Sanger et al., 1977). The two clones were found to contain identical splice sites, and the nucleotide sequence of the splice junction present in both clones is presented in Figure 12A. The sequencing data identifies the splice donor and acceptor sites, in the full-length EIAV genomic mRNA, used to generate the spliced mRNA copy of the two cDNA clones. Of the three potential splice donor sites near the 5'-terminus of the gag gene (Figure 12B; see below), in this spliced mRNA the 5'-most splice donor site is spliced to the splice acceptor site near the 5'-end of ORF51, 178 nucleotides upstream of the env gene translational initiation ATG codon. Also shown in Figure 12B are the nucleotide sequences of the identified splice donor and acceptor sites and the extent of their homology to the consensus splice donor and acceptor sequences (Mount, 1982; Ohshima and Gotoh, 1987).
Figure 12. Nucleotide sequence analysis of the 270-bp Smal-BamHI splice-junction cDNA of the EIAV 3.5-kb mRNA.

A. Autoradiogram of a sequencing gel showing the region of nucleotides immediately 5' and 3' to the splice junction site, which is pointed out by the arrow.

B. Schematic diagram of the EIAV genome, showing the location of putative splice donor sites (solid triangles) and acceptor sites (open triangles) on the viral genome. The bold-face nucleotide sequences within the left and right lower brackets are those of the identified splice donor and acceptor site, respectively, with underlined bases showing perfect homology to the consensus donor and acceptor site sequences (Mount, 1982; Ohshima and Gotoh, 1987). The arrows show the site of splicing.
Northern Hybridization Analysis Using Splice Donor-Specific Oligonucleotide Probes

As mentioned previously, three potential splice donor sites are located near the 5'-terminus of the EIAV gag gene (Figure 12B) based on matches to the consensus sequence for splice donor sites (Mount, 1982; Ohshima and Gotoh, 1987). Based on the utilization of one or more of the three donor sites for generation of the spliced viral transcripts, a number of differentially spliced mRNA's can be generated in EIAV-infected cells with varying coding potentials. The nucleotide sequence of the three short ORF's in the EIAV genome has revealed that only S2 contains a translational initiation methionine codon (Rushlow et al., 1986). Thus for functional expression of S1 or S3, an initiation codon from an upstream gene needs to be spliced, in frame, to the coding sequences of the two ORF's.

By comparison to other lentiviruses, which utilize multiple splicing of their full-length genomic mRNA to generate the small mRNA's needed to encode their short ORF's (Arya et al., 1985; Arya and Gallo, 1986; Davis and Clements, 1989; Mazarin et al., 1988; Muesing et al., 1985), the initiation codon for the EIAV S1 or S3 ORF's would be assumed to be located within the first exon of such multi-spliced mRNA. This exon would contain sequences from the 5'-region of the viral genome, and the only translational initiation ATG codon in this region is that of the gag gene.
The nucleotide sequence of the EIAV provirus near the 5'-terminus of the 
gag gene is presented in Figure 13 (Stephens et al., 1986). The sequence starts 
from the Smal site, 71 nucleotides upstream of the gag gene's ATG translational 
initiation codon, and continues for 120 nucleotides (40 codons) into gag. 
Moreover, the location of each of the three putative splice donor sites in relation 
to the gag ATG codon is shown with underlined bases in each donor site 
representing a perfect match to the consensus donor site sequence (Mount, 1982; 
Ohshima and Gotoh, 1987). As can be seen, the 5'-most donor site is an excellent 
match with the consensus sequence except in the last position; the middle donor 
site is a very good match with the consensus sequence with the exception of the 
last two nucleotides; and the 3'-most donor site is a good match with the 
consensus sequence except in the second and the eighth positions.

Furthermore, as shown in Figure 13, the 5'-most splice donor site is 
situated immediately upstream of the gag ATG codon, and its utilization in a 
splicing event would remove the gag initiation codon, as part of the intron, from 
the spliced mRNA. Since the other two putative splice donor sites are located 
downstream of the gag ATG codon, potential splicing events utilizing either one 
of these two donor sites, however, would result in the ATG codon of gag being 
retained in the spliced mRNA.

Although the use of the 5'-most splice donor site for generation of a 
spliced EIAV mRNA was already determined by cDNA cloning and sequencing, 
this finding does not rule out the potential utilization of either of the other two 
putative splice donor sites to generate additional alternatively spliced
Figure 13. Nucleotide sequence of the EIAV provirus near the 5'-terminus of the \textit{gag} gene (Stephens \textit{et al.}, 1986). The sequence starts from the \textit{SmaI} site, 71 nucleotides upstream of the \textit{gag} gene's ATG translational initiation codon, and continues for 120 nucleotides (40 codons) into \textit{gag}. The sequences within each of the three putative splice donor sites are in bold-face, with underlined nucleotides in each donor site representing a perfect match to the consensus donor site sequence (Mount, 1982; Ohshima and Gotoh, 1987), which is shown at the bottom of the figure. Solid triangles above each donor site show the expected site of splicing. Brackets above the EIAV DNA sequence and immediately upstream of each donor site sequence contain the sequences the three oligonucleotides, SD-1, SD-2, and SD-3, which were used as probes in Northern hybridization analysis. One-letter amino acid codes below each the \textit{gag} gene codons, beginning with the ATG methionine codon, denote the amino-terminal sequence of the \textit{gag} polyprotein.
**SD-1**

*SmaI*

$3'\text{-} \text{CTCCACAAGGACCCTTGT}

CCAAGCAGAGGAGAACCTGATCCTGTGAAGGTGCCTGGCCAGAACAC

**SD-2**

$3'\text{-} \text{CTCTGGGAACTGTACCCTCGTGCGC}

AGGAGGA\text{CAGTTAGAGTGGAGAGCCTTGACATGAGCGGCTCAAGAAGTTAGAGMGDP}L TW SK ALK K L E$

**SD-3**

$3'\text{-} \text{CCATGTTCCAGAGTTTAAATTGATG}$

\text{AGGCTACGGTACAAGGTTCTCAGAAATTAACTACCTCCTACTGTAATTTGGCCGCTAAGTC}KVTVQGSQKLTGNCCNWALS

CTAGTAGCTTTATTT....

\text{LVDLP}

Consensus splice donor sequence: \(\text{CAGGT}\text{AGT}\)
EIAV mRNA's in infected cells. For instance, the 3.5-kb mRNA could represent a heterogeneous population of mRNA molecules generated by splicing of each of the three donor sites to the same acceptor site upstream of the env gene, or by splicing of one donor site to different, but proximally located, acceptor sites upstream of the env gene. As a direct method for addressing the first possibility, that is differential splicing of the three donor sites to the same acceptor site, oligonucleotides complementary to sequences immediately upstream of each of the three splice donor sites were radiolabeled with $^{32}$P and used as probes in Northern hybridization analysis of poly(A)$^+$ RNA from infected FDD cells. The sequence of the three oligonucleotides, termed SD-1, SD-2, and SD-3, along with the location of their complementary region upstream of each splice donor site on the viral genome is also indicated in Figure 13.

Since all three probes contain complementary sequences to the full-length genomic mRNA, they would all be expected to hybridize to this viral transcript. Moreover, because SD-1 is complementary to a region upstream of all three putative splice donor site, it is expected to also hybridize to all spliced viral mRNA's, namely the 3.5- and 1.5-kb mRNA's, whose production would utilize either one, two, or all three of the donor sites downstream of SD-1. Furthermore, the hybridization of SD-2 to either the 3.5- or the 1.5-kb mRNA, or to both, would indicate that the middle splice donor site, in addition to the 5'-most donor site, is utilized to generate the corresponding spliced viral mRNA(s). Similarly, the hybridization of SD-3 to either of the two spliced mRNA's, or to both, would reveal the usage of the 3'-most donor site, in addition to the 5'-most donor site,
for generation of the corresponding viral mRNA(s). However, the hybridization of SD-2 and SD-3 to only the full-length 8.2-kb mRNA, and not to any spliced mRNA, would indicate that only the 5'-most splice donor site is used to generate the 3.5- and 1.5-kb mRNA's of EIAV.

The results of the three Northern hybridizations are presented in Figure 14. As expected, SD-1 hybridizes to the three species of EIAV-specific mRNA's. However, SD-2 and SD-3 only hybridize to the 8.2-kb full-length mRNA and do not detect either the 3.5- or the 1.5-kb spliced mRNA's recognized by SD-1. These results indicate that the 5'-most splice donor site, already shown to be used for generating a spliced EIAV mRNA, is the only donor site utilized in EIAV-infected FDD cells for splicing of the full-length viral genomic mRNA. Thus the R-U5-leader sequences initiating from the 5'-boundary of the R region and terminating at the 5'-most splice donor site constitute the first exon of all spliced EIAV mRNA's. The consequences of this finding are that spliced EIAV mRNA's containing sequences from ORFS1 or ORFS3 lack the gag gene AUG codon, and if these short ORF's are indeed expressed in EIAV-infected cells, translation of their corresponding mRNA's would presumably have to involve the use of non-AUG initiation codons.
Figure 14. Northern hybridization analysis of poly(A)$^+$ RNA from EIAV-infected FDD cells using the three oligonucleotide probes SD-1, SD-2, and SD-3 (lanes 1, 2, and 3, respectively). The electrophoretic mobilities of the three species of EIAV-specific mRNA's are indicated on the left.
S1 Nuclease Mapping of Viral mRNA's

The results of Northern hybridizations with the splice donor-specific oligonucleotide probes indicated that all spliced EIAV mRNA's utilize the 5'-most splice donor site immediately preceeding the gag gene. Moreover, the results obtained from cDNA cloning and sequencing had revealed the splicing of this donor site to the acceptor site 178-nucleotides upstream of the 5'-boundary of the env gene to generate the single-spliced 3.5-kb env mRNA.

By analogy to other lentiviruses, it is presumed that the majority of sequences within the EIAV single-spliced mRNA, lying outside of those within the short ORF's, would be removed as an intron in a second splicing event to produce the double-spliced viral mRNA (Arya et al., 1985; Arya and Gallo, 1986; Davis and Clements, 1989; Mazarin et al., 1988; Muesing et al., 1985). The double-spliced message, containing coding sequences of the short ORF's, would then encode the regulatory proteins of the virus such as the viral trans-activator(s). Indeed a 1.4-kb cDNA copy of one such double-spliced mRNA, containing sequences from ORFS1 and ORFS3, has already been isolated and shown to encode a viral trans-activator (Derse et al., 1989; see below). Moreover, in agreement with our finding that only the 5'-most splice donor site on the EIAV genome, upstream of the gag gene AUG codon, is utilized to generate the first exon of all virus-specific spliced mRNA's in infected cells, the translation of the aforementioned double-spliced viral mRNA is proposed to be initiated from a non-AUG codon (Derse et al., 1989).
To map the location of the splice donor site(s) within the EIAV env gene that are used to produce the double-spliced mRNA species, and also to determine whether any additional splice acceptor sites, located upstream of and within the env gene, are used to generate the spliced EIAV mRNA's, S1-nuclease mapping studies were performed using RNA from FDD cells infected with FDD-adapted EIAV. The studies were designed to also include the 5'-region of the gag gene to test the validity of the results obtained previously using the splice donor-specific oligonucleotide probes.

Figure 15 shows the relative location of putative splice donor and acceptor sites, represented by closed and open triangles, respectively, on the EIAV genome. In addition to the acceptor site located 178-nucleotides upstream of the env gene initiation codon and identified by cDNA cloning and sequencing, a second putative acceptor site is present at 129-nucleotides upstream of the env initiation codon. Following the 3'-terminus of ORFS1 and extending into the coding sequences of ORFS2 are four putative donor sites, any one of which could potentially be utilized in the second splicing event of the single-spliced 3.5-kb mRNA to generate a double-spliced 1.5-kb mRNA. Moreover, a putative splice acceptor site is situated immediately at the 5'-boundary of the ORFS3 reading frame.

The sequence of each putative donor and acceptor site, with the exception of the three putative donor sites near the 5'-terminus of the gag gene whose sequence was presented earlier in Figure 13, is presented in Table III. In each case, underlined bases correspond to a perfect match with the consensus donor
Figure 15. Schematic diagram of the EIAV genome. The figure shows the location of putative splice donor and acceptor sites, represented by solid and open triangles, respectively, on the viral genome.
Table III. Nucleotide sequence of the putative splice donor and acceptor sites located immediately upstream of or within the EIAV env gene.

<table>
<thead>
<tr>
<th>Nucleotide sequence of putative splice sites</th>
<th>Nature of splice site</th>
<th>Position on the EIAV genome</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAGCTTTGAG</td>
<td>Donor</td>
<td>1226</td>
</tr>
<tr>
<td>TTGTAAG</td>
<td>Donor</td>
<td>1248</td>
</tr>
<tr>
<td>GGGTAACA</td>
<td>Donor</td>
<td>1256</td>
</tr>
<tr>
<td>ACAGTAAGA</td>
<td>Donor</td>
<td>1385</td>
</tr>
<tr>
<td>TTGATTTGTTCAG</td>
<td>Acceptor</td>
<td>1084</td>
</tr>
<tr>
<td>TTCTTGAGTCAGTCTAGG</td>
<td>Acceptor</td>
<td>1133</td>
</tr>
<tr>
<td>AACCTTCCGCTAAGA</td>
<td>Acceptor</td>
<td>3184</td>
</tr>
</tbody>
</table>

a. The splice junction site within the putative donor and acceptor sequences is denoted by closed and open triangles, respectively. Underlined nucleotides are a perfect match with the consensus splice junction sequences (Mount, 1982; Ohshima and Gotoh, 1987).

b. Based on the numbering system of Rushlow et al. (1986).

c. The utilization of this acceptor site in prototype EIAV-infected FDD cells has already been confirmed by cDNA cloning and sequencing (Figure 11).
and acceptor site sequences (Mount, 1982; Ohshima and Gotoh, 1987). Both acceptor sites located upstream of the *env* gene show an excellent match with the consensus acceptor site sequence. Moreover, the four putative donors which follow the 3'-end of *ORFS1* and extend into *ORFS2* exhibit good matches with the consensus donor sequence. Finally, the acceptor site located immediately at the 5'-boundary of the *ORFS3* reading frame is a good match with the consensus sequence but contains an A instead of a G in the +1 position and an A instead of a pyrimidine in the -3 position.

For S1 nuclease mapping studies of the splice donor and acceptor sites in spliced viral mRNA's, four specific EIAV proviral DNA restriction fragments, spanning the various putative splice donor and acceptor sites on the viral genome, were used as probes. Since the entire nucleotide sequence within each of the four probes comes from the EIAV proviral DNA, the full-length nonsense strand of each probe is complementary to the sense strand sequence of the unspliced, full-length 8.2-kb EIAV genomic mRNA. Moreover, by virtue of spanning the putative splice donor or acceptor sites on the full-length genomic mRNA, only a segment of the nucleotide sequence within the nonsense strand of each probe is complementary to the corresponding exon sequence of the full-length viral mRNA and its spliced mRNA product. The other segment of the nonsense strand of the respective probe, however, is complementary to the corresponding intron sequence of the full-length viral mRNA which is not present in its spliced product. Such differential complementarity of each splice junction probe to the
unspliced and spliced viral mRNA's provides the basis for mapping the exact location of splice donor and acceptor sites on the viral genome.

Following the hybridization of each denatured probe to RNA from EIAV-infected cells, a protection assay is performed in which the RNA-DNA hybrid is treated with the single-strand-specific nuclease S1. Single-stranded probe DNA or viral RNA is degraded by the nuclease into mononucleotides while double-stranded DNA or RNA-DNA hybrids are left intact (Berk and Sharp, 1977). Thus in this assay the full-length probe fragment is expected to be protected from S1 nuclease digestion by the full-length viral genomic mRNA, while smaller probe fragments would be rescued from S1 nuclease digestion by the spliced viral mRNA's. The size of the smaller probe fragments, determined by denaturing polyacrylamide gel electrophoresis, would then be equal to the actual distance between the splice site on the spliced mRNA relative to the radiolabeled end of the probe.

Figure 16A presents a schematic diagram of the EIAV genome, showing the location of the four probes used for S1 nuclease mapping studies relative to the putative splice donor and acceptor sites on the viral genome. Because the probes are radiolabeled double-stranded DNA fragments, in each hybridization reaction a portion of the denatured nonsense strand input probe molecules rehybridize back to their complementary sense DNA strand instead of hybridizing to viral mRNA's. Such full-length nonsense strand probe molecules cannot be distinguished from those that are protected from S1 nuclease digestion by the full-length mRNA. To allow for distinction of the full-length input probe from the
Figure 16. Schematic representation of the probes used in S1 nuclease mapping.

A. Schematic diagram of the EIAV genome, showing the location of the four probes used for S1 nuclease mapping studies (p1, p2, p3, p4) relative to the location of putative splice donor (solid triangle) and acceptor sites (open triangle) on the viral genome. Probe p1 consists of a 300-bp NarI-PvuII proviral DNA fragment and a 250-bp M13mp18 PvuII-NarI tag fragment at the PvuII site. Probe p2 consists of a 450-bp NcoI-BamHI proviral DNA fragment and a 161-bp M13mp19 NcoI-BglII tag fragment at the NcoI site. Probes p3 and p4 do not contain tag sequences and were directly used in S1-nuclease protection analysis as 617-bp PvuII-HindIII and 385-bp DraI-ScaI proviral DNA restriction fragments, respectively. Abbreviations: N, NarI; P, PvuII; G, BglII; C, NcoI; B, BamHI; P, PvuII; H, HindIII; D, DraI; S, ScaI; SD, putative splice donor site(s); SA, putative splice acceptor site(s).

B.-E. Schematic representations of probes p1-p4, respectively. Also shown in each panel are fragments of the nonsense strand of each respective probe, and their lengths in nucleotides (nt), which could potentially be protected from S1 nuclease digestion by the unspliced, single-spliced, and double-spliced EIAV mRNA’s. The asterisk represents the $^{32}$P-labeled end of the nonsense strand of each probe.
probe fragment rescued by the full-length mRNA, two of the probes were specifically designed to contain "tag" sequences taken from the double-stranded replicating form (RF) DNA of either M13mp18 or M13mp19 and attached to their intronic end. As shown in Figure 16B, probe 1 (p1) consisted of a 300-bp NarI-PvuII proviral DNA fragment and a 250-bp M13mp18 SmaI-NarI tag at the PvuII site. Thus following S1-nuclease protection analysis using this probe, the input probe and the probe fragment rescued by the full-length mRNA would be 550- and 300-nucleotides (nt) long, respectively. Moreover, smaller probe fragments, whose size equals the distance between the splice donor site and the NarI-end of the probe, would be protected from S1-nuclease digestion by the spliced mRNA’s. This distance is 130-nt for the 5’-most donor site, 183-nt for the middle donor site, and 217-nt for the 3’-most donor site.

Similarly, as shown in Figure 16C, probe 2 (p2) consisted of a 450-bp NcoI-BamHI proviral DNA fragment and a 161-bp M13mp19 SmaI-BglII tag at the repaired NcoI site. Thus following S1-nuclease protection analysis using this probe, the expected length of the input probe would be 611-nt while the probe fragment protected by the full-length mRNA would be 450-nt long. In addition, the distance between each of the two putative splice acceptor sites in this region and the BamHI-end of the probe is 157- and 206-nt long, respectively.

The other two probes, probes p3 and p4, did not contain tag sequences and were directly used in S1-nuclease protection analysis as 617-bp PvuII-HindIII and 385-bp DraI-Scal proviral DNA restriction fragments, respectively. As shown
in Figure 16D, the distance between each of the four putative splice donor sites located within the sequences of probe p3 and the PvuII-end of this probe is 114-, 136-, 144-, and 273-nt long, respectively. Moreover, Figure 16E shows the the distance between the putative acceptor site located within the sequences of probe p4 and the ScaI-end of the probe, which is 178-nt long.

For S1 nuclease protection assays, RNA-DNA hybrids, obtained from hybridization of each end-labeled probe with total RNA isolated from FDD cells infected with FDD-adapted EIAV, were digested with S1 nuclease and the digestion products were resolved on 6% polyacrylamide-8M urea sequencing gels. As negative control, RNA from uninfected FDD cells, lacking EIAV-specific unspliced and spliced transcripts, was used in parallel experiments. The results of these studies are presented in Figure 17.

Panels A, B, C, and D in Figure 17 correspond to individual S1-nuclease protection assays using probes p1, p2, p3, and p4, respectively, which were described above. In each panel, lane M contains end-labeled MspI fragments of plasmid pBR322, while lane 1 or lane 2 contain the corresponding probe fragments protected from S1 nuclease digestion by RNA from EIAV-infected FDD cells or uninfected FDD cells, respectively. In panel A, three distinct fragments of probe p1 (550-bp NarI; Figure 16A and 16B), migrating at about 550-nt, 300-nt, and 130-nt are protected following the S1-nuclease protection assay utilizing RNA from infected cells (lane 1), while only the 550-nt full-length probe fragment is present following the assay in which control RNA from uninfected cells is utilized (lane 2). The 550-nt probe fragment in lanes 1 and 2 corresponds
Figure 17. S1 nuclease mapping of splice junctions in spliced mRNA's of EIAV. Panels A, B, C, and D correspond to S1 nuclease protection assays of probes p1, p2, p3, and p4, respectively, using total cellular RNA from FDD cells infected with FDD-adapted EIAV (lane 1), or uninfected FDD cells (lane 2), Lane M in each panel contains $^{32}$P-labeled $MspI$-digested pBR322 as size marker.
to the portion of full-length input probe molecules, containing the M13 tag sequence, which have escaped degradation by S1 nuclease due to re-hybridization with their complementary DNA strand. Moreover, the additional probe fragments present in lane 1, but not in lane 2, are those which have been protected from S1-nuclease degradation by EIAV-specific mRNA's. In lane 1, the 300-nt fragment of the probe is that which is protected by the full-length 8.2-kb EIAV mRNA (Figure 16B), while the 130-nt probe fragment is that which is protected by a spliced viral mRNA whose splice junction site is 130-nt downstream of the NarI-end of the probe within the EIAV genome. This splice site corresponds exactly to the 5'-most donor site among the three putative donor sites near the 5'-end of the EIAV gag gene (Figure 16A). Furthermore, as mentioned above, the protected probe fragments which would correspond to spliced mRNA's utilizing the middle or the 3'-most putative donor sites in this region would be expected to be 183- or 217-nt long (Figure 16A and 16B). Since these additional bands are not present in lane 1 of panel A, it appears that the 5'-most splice donor site is indeed the only donor site near the 5'-end of the gag gene that is utilized in EIAV-infected FDD cells for splicing of the full-length viral mRNA, confirming the earlier results obtained using the splice-donor-specific oligonucleotide probes.

In Figure 17, panel B, four distinct fragments of probe p2 (611-bp BamHI-BglII; Figure 16A and 16C), migrating at about 611-nt, 450-nt, 206-nt, and 155-nt are protected by RNA from infected cells (lane 1). As expected, the 611-nt full-length probe fragment, containing the M13 tag sequence, is the only protected probe fragment in the assay utilizing RNA from uninfected cells (lane 2).
1 of panel B, the 450-nt fragment of the probe is that which is protected by the full-length 8.2-kb EIAV mRNA, while the probe fragments migrating at about 206-nt and 155-nt are those protected by spliced viral mRNA's whose splice junction sites are therefore 206-nt and 155-nt, respectively, upstream of the \textit{BamHI}-end of the probe. The higher intensity of the 206-nt probe fragment corresponds to the higher concentration of the viral mRNA which has protected this fragment relative to the other viral mRNA's present in infected FDD cells. The splice junction site located at 206-nt upstream of the \textit{BamHI} site corresponds exactly to the acceptor site, already identified by cDNA cloning and sequencing, which is located at 178-nt upstream of the \textit{env} gene initiation codon (position 1084 on the nucleotide sequence of Rushlow \textit{et al.}, 1986; Table III). Interestingly, the second splice junction site identified by this assay and located at 155-nt upstream of the \textit{BamHI}-end of the probe corresponds to the acceptor site present at 129-nt upstream of the \textit{env} gene initiation codon (position 1133 on the nucleotide sequence of Rushlow \textit{et al.}, 1986; Table III). The length of the probe fragment protected by a spliced mRNA utilizing this acceptor site was actually expected to be 157-nt (Figure 16C), although the apparent mobility of this fragment is closer to 154- or 155-nt during denaturing polyacrylamide gel electrophoresis (Figure 17, panel B, lane 2).

Panel C in Figure 17 shows the results of S1 nuclease protection analysis of probe p3 (617-bp \textit{PvuII-HindIII}; Figure 16A and 16D). As expected, the full-length 617-bp probe fragment is protected by RNA from either infected or uninfected FDD cells (lanes 1 and 2). Moreover, a protected probe fragment,
migrating at about 140- to 144-nt, is protected by RNA from infected cells (lane 1) but not by RNA from uninfected cells (lane 2). This fragment of the probe is therefore protected by a spliced EIAV mRNA whose splice junction site is situated at about 140- to 144-nt downstream of the PvuII site on the viral genome. As mentioned previously, the four putative splice donor sites near the end of ORFS1 and within the sequences of the PvuII-HindIII probe (p3) are located at 114-, 136-, 144-, and 273-nt from the PvuII-end of the probe. Since the mobility of the protected probe fragment identified in this assay corresponds to a fragment of about 140- to 144-nt, the donor site located at 144-nt from the PvuII-end of the probe (position 1256 on the nucleotide sequence of Rushlow et al., 1986; Table III) appears to be the donor site used to generate a spliced EIAV mRNA which has protected this fragment of probe p3.

Finally, in Figure 17, panel D, the full-length probe p4 (385-bp DraI-Scal; Figure 16A and 16E) is protected from SI nuclease digestion by RNA from either infected or uninfected FDD cells (lanes 1 and 2), while an additional fragment of the probe migrating at about 180-nt is only protected by RNA from infected cells (lane 1). Therefore the 180-nt probe fragment is one which is protected from SI nuclease digestion by a spliced EIAV mRNA whose splice junction site is 180-nt upstream of the Scal site on the viral genome. The only splice junction site in this region of the viral genome is a putative splice acceptor site situated at the 5'-boundary of ORFS3 (Figure 16A; position 3184 on the nucleotide sequence of Rushlow et al., Table III). This acceptor site is exactly 178-nt upstream of the Scal-end of the DraI-Scal probe (p4), and a spliced mRNA generated by the
utilization of this acceptor site would be expected to protect a 178-nt fragment of
the probe (Figure 16E). The apparent mobility of the protected probe fragment
during denaturing polyacrylamide gel electrophoresis, although closer to 180-nt
than to 178-nt, suggests that this splice acceptor site is indeed utilized in EIAV-
infected FDD cells to generate a spliced viral mRNA which has protected the
180-nt probe fragment.

Longer exposures of the autoradiograms in panels A, B, C, and D of
Figure 17 did not reveal any additional protected probe fragments (data not
shown), suggesting that the detected EIAV mRNA's constitute the major species
of viral transcripts present in infected FDD cells. The S1 nuclease mapping
studies described above have confirmed the earlier results obtained through
Northern hybridizations using the splice donor-specific oligonucleotide probes,
showing that the 5'-most splice donor site, immediately upstream of the gag gene,
is the common donor site to all spliced EIAV mRNA's. Moreover, in addition to
the splice acceptor site previously identified by sequencing of splice-junction
cDNA obtained from the 3.5-kb mRNA, the S1 nuclease mapping studies have
identified and localized a second splice acceptor site upstream of the env gene, a
splice donor site immediately downstream of ORFS1, and a splice acceptor site at
the 5'-boundary of ORFS3.
III. The 3.5-kb EIAV mRNA: Potential To Encode The ORFS2 Gene Product In Addition To The env Proteins

Based on the nucleotide sequence of the EIAV proviral DNA, the predicted amino acid sequences of the three short ORFs of EIAV is presented in Figure 18 (Rushlow et al., 1986). As with the short ORF's of other lentiviruses, such as HIV-1 or visna virus, whose encoded products have been generally shown to code for trans-acting viral regulatory proteins (Arya et al., 1985; Davis and Clements, 1989; Mazarin et al., 1988; Sodroski et al., 1986), attempts to delineate the function of the three short ORF's of EIAV have mainly focused on assessing the ability of their gene products to trans-activate the viral LTR. Based on the results of early deletion mutagenesis experiments, by which parts of the coding sequence within each ORF were removed and the resulting mutant proviruses were assayed for their capacity to trans-activate the viral LTR, sequences within the pol-env intergenic region on the viral genome, to which both SI and S2 belong, were shown to be required for trans-activation of the viral LTR (Sherman et al., 1988). More recently, Derse et al. (1989) have isolated a cDNA clone of a double-spliced EIAV mRNA, containing sequences from S1 and S3, and have shown that it can trans-activate the viral LTR in transfected cells. Since the coding sequence of S1 is located in the pol-env intergenic region, it is therefore presently believed that S1 encodes a viral trans-activating protein. The function of the other two ORF's, however, is not known and to date no protein products from any of the three ORF's have been detected in EIAV-infected cells.
**ORFS1**

VLLQ\(\vec{\text{E}}\)ARPNYHCQLCFLRSL\(\vec{G}\)IDYLDASLRKKNK
QR\(L\)KAIQ\(Q\)GR\(Q\)PQYLL

**ORFS2**

YMGLFGKGV\(\text{T}\)WSASHSMGGSO\(\text{Q}\)G\(E\)S\(\text{Q}\)PLL\(N\)S\(Q\)K
NLS\(V\)R\(R\)T\(Q\)CF\(N\)LIVIIMTV\(R\)TA\(W\)QN\(R\)R\(K\)Q\(E\)TKK

**ORFS3**

\(\vec{D}\)P\(\text{Q}\)G\(\text{P}\)\(L\)\(E\)SD\(Q\)\(W\)CRVLR\(Q\)SLPE\(E\)\(K\)I\(S\)\(T\)\(C\)I\(A\)RRH
LG\(P\)GP\(T\)Q\(H\)TPSR\(R\)DR\(W\)IR\(G\)\(Q\)\(I\)L\(Q\)AE\(V\)L\(Q\)E\(R\)LE\(W\)R
IR\(G\)V\(Q\)\(Q\)\(A\)A\(K\)E\(L\)GE\(V\)N\(R\)\(G\)I\(W\)REL\(Y\)F\(R\)EDQ\(R\)G\(D\)FS
AW\(G\)\(G\)\(Y\)\(Q\)\(R\)A\(Q\)ER\(L\)W\(G\)E\(Q\)SS\(P\)\(R\)VLR\(P\)G\(D\)\(S\)\(K\)RR\(R\)KH

**Figure 18.** Predicted amino acid sequences of the three short ORFs of EIAV. The sequences are "translated" from the nucleotide sequence of EIAV proviral DNA (Rushlow et al., 1986). The splice acceptor sites in S1 and at the 5'-boundary of S3, identified by sequencing of splice-junction cDNA and S1 nuclease mapping, are indicated by open triangles.
A unique property among these short ORF's is that S2 is the only one which contains a translational initiator ATG codon, located at its second codon position (Figure 18). Thus, because S1 and S3 lack such a codon, their potential expression in EIAV-infected cells would require the in-frame splicing of an initiator codon from an upstream reading frame. The results of our S1 nuclease mapping experiments point to the existence of at least one double-spliced mRNA in EIAV-infected FDD cells whose first exon contains the 5'-leader sequences upstream of the gag gene, and in its second and third exons it contains the coding sequences of S1 and S3, respectively. Almost the entire coding sequence of S2, with the exception of its first seven codons, is removed from this mRNA by splicing, suggesting that potential expression of S2 in virus-infected cells would require the existence of a separately spliced mRNA.

Nucleotide sequencing of the splice-junction cDNA of the 3.5-kb mRNA had revealed that the splice acceptor site for this mRNA is located 178 nucleotides upstream of the env gene ATG initiator codon. Because the S2 ATG codon is 23 nucleotides upstream of the ATG codon of env, and the coding sequences of the two genes are present on the same spliced 3.5-kb mRNA and in different reading frames, we postulated that this mRNA could potentially encode the S2 gene product in addition to the env proteins. To test the hypothesis, a transcript designed to contain the coding sequences of both S2 and env identical to that of the 3.5-kb mRNA was synthesized in vitro and its coding capacity was evaluated by in vitro translation.
The 270-bp Smal-BamHI fragment of the splice junction cDNA, sequenced earlier for identification of the splice donor and acceptor sites of the 3.5-kb mRNA, was ligated to a 2710-bp BamHI-PstI fragment containing the remainder of the env coding sequence and derived from a separate cDNA clone. The hybrid Smal-PstI cDNA fragment, containing the entire coding sequences of both S2 and env genes, was then cloned into the in vitro transcription vector pSP65 downstream of the SP6 promoter (Figure 19). Two separate linear DNA templates of the plasmid DNA containing the S2-env cDNA insert were prepared for in vitro transcription by SP6 RNA polymerase. One of the linearized templates was prepared by digestion of the plasmid with HindIII, an enzyme which cleaves the env gene at two unique restriction sites, one at about 300-bp and the other at approximately 1030-bp downstream of the S2 coding sequence (Figure 19). The second linearized template was prepared by digestion of the plasmid with MluI, an enzyme which cuts at a unique restriction site within the 3'-LTR downstream of the env gene (Figure 19). The design of the two linear templates for in vitro transcription was based on the premise that if translational initiation from this mRNA begins at the ATG codons of both S2 and env, then in vitro translation of the truncated transcript generated from the HindIII-linearized template would yield a full-length S2 protein and a truncated env protein, while translation of the full-length transcript generated from the MluI-linearized template would yield both S2 and env proteins in their full-length forms.
Figure 19. Construction of the S2-env *in vitro* transcription vector. The 270-bp *Smal-BamHI* fragment of the splice junction cDNA was ligated to a 2710-bp *BamHI-PstI* fragment containing the remainder of the *env* coding sequence and derived from a separate cDNA clone. The hybrid *Smal-PstI* cDNA fragment, containing the entire coding sequences of both S2 and env genes in their natural spliced form, was subcloned into pUC19 and taken out as a *SacI-PstI* fragment. The 2980-bp *SacI-PstI S2-env* hybrid cDNA was then cloned into the *in vitro* transcription vector pSP65 downstream of the SP6 promoter. The location of the *HindIII* and *MluI* restriction sites on the EIAV genome are also indicated.
The transcripts derived from the two templates were translated in a cell-free extract of wheat-germ in the presence of $^{35}\text{S}$-methionine, and the $^{35}\text{S}$-labeled translational products were fractionated by SDS-polyacrylamide gel electrophoresis (SDS-PAGE). The results, presented in Figure 20, indicate that a lower molecular weight polypeptide of about 7.5-kda is encoded by both transcripts (lanes 1 and 2). Moreover, the longer transcript, generated from the $MluI$-linearized template, encodes a second protein of about 27.0- to 28.0-kda (lane 1), while the second protein expressed by the truncated transcript, produced from the $\text{HindIII}$-linearized template, migrates at about 16.0- to 17.0-kda (lane 2).

Based on the nucleotide sequence of $S2$ and $env$, the coding potential of these two genes were calculated to be approximately 7.5- and 97.0-kda, respectively, from the long transcript. Moreover, the size of the $env$ protein translated from the truncated transcript was calculated to be around 17.0-kda.

The translation of the smaller protein from both truncated and full-length transcripts and its apparent molecular weight of 7.5-kda, as expected for the $S2$ gene product, suggested that it was expressed from the $S2$ coding sequence in both transcripts. Moreover, in the case of the truncated transcript, the larger protein exhibited the 17.0-kda electrophoretic mobility expected from the truncated $env$ protein (Figure 20, lane 2), whose coding sequence would have had ended at the $\text{HindIII}$ site. This observation suggested that both $S2$ and $env$ could have been translated from this transcript. However, in the case of the full-length transcript, the larger protein exhibited a molecular weight of about 27.0-kda (Figure 20, lane 1), much smaller than the 97.0-kda full-length protein expected...
Figure 20. Analysis of the $^{35}$S-labeled *in vitro* translational products of the full-length (lane 1) and truncated (lane 2) S2-env *in vitro* transcript by SDS-PAGE. Size markers on the left are the $^{14}$C-labeled 3.0- to 43.0-kda molecular weight standards from BRL.
to have been synthesized from the env coding sequences of the full-length transcript.

Because most plasmid clones of the full-length EIAV env gene, once propogated in E. coli, are very unstable and contain nucleotide deletions, insertions, or substitutions (K. Rushlow, unpublished data; see below), it was suspected that the env coding sequences within the cDNA insert of the transcription vector, from which the full-length transcript had been transcribed, had undergone a mutation during the plasmid's propogation in E. coli. Such mutation(s) could have potentially introduced premature termination codon(s) in the env coding sequence of the full-length transcript, resulting in prematurely terminated env translational products synthesized from this RNA.

To examine this possibility, the env gene of the same plasmid preparation used for in vitro synthesis of the truncated and full-length transcripts was subjected to nucleotide sequencing (Sanger et al., 1977). The results indicated that indeed a deletion had occured within the env coding sequence, 718 nucleotides downstream of the env ATG initiation codon (K. Rushlow, personal communication; unpublished data). As a result, a C in a CCT (proline) codon had been deleted, bringing a TGA termination codon located at 748 nucleotides downstream of the env ATG codon (30 nucleotides downstream of the mutation) in frame with the env coding sequence and thus terminating translation. In the wild-type env sequence, this TGA codon is located in the +1 frame relative to the env reading frame, thus it does not function to terminate the translation of the wild-type env mRNA.
The expected molecular weight of the terminated env protein, translated from the mutated full-length env transcript, was then calculated and found to be 27.5-kda, a value matching the observed size of the larger protein translated from the full-length transcript (Figure 20, lane 1). Thus in vitro translation of transcripts containing the coding sequences of both S2 and env gives rise to two distinct proteins having the sizes expected of the gene products of these two genes, suggesting that translational initiation can occur from each gene's AUG codon on the same mRNA. That the two proteins are indeed encoded by env and S2 have been confirmed by immunoprecipitation analysis of the in vitro translational products. Polyclonal or monoclonal antibodies against the EIAV outer envelope glycoprotein, gp90, and a rabbit polyclonal antibody against a synthetic peptide containing amino acid sequences of S2 have been used in immunoprecipitation analyses of the translational products. The results indicate that the individual large and small proteins expressed from the in vitro transcripts described above are recognized by the anti-gp90 and anti-S2 antibodies, respectively (R.L. Schiltz, unpublished data).

To examine whether the co-expression of env and S2 proteins from the same mRNA, which was observed in vitro, can also be shown in vivo, the expression of the two proteins was investigated by utilizing a eukaryotic transient expression vector, p91023(B), and a simian cell line, COS-7. In addition to having the prokaryotic origin of replication of plasmid pBR322, the p91023(B) eukaryotic expression vector (Figure 21) contains the origin of replication derived from the small DNA tumor virus, simian virus 40 (SV40). Moreover, COS-7 cells are
African green monkey cells which have been transformed by an origin of replication-defective mutant of SV40 and constitutively express high levels of the SV40 large T antigen, a viral protein required for initiation of SV40 DNA synthesis from the SV40 origin of replication (Gluzman, 1981; Cullen, 1987). Thus the p91023(B) plasmid, once introduced into COS-7 cells, can replicate to a very high copy number in these cells and efficiently express the protein products of the cloned gene of interest.

As mentioned previously, most plasmid clones of the full-length EIAV env gene, once propagated in E. coli, are very unstable and contain nucleotide deletions, insertions, or substitutions. It is believed that a stretch of nucleotides near the 5'-terminus of the env gene, containing a sequence very similar to that of the prokaryotic ribosome binding site (Shine and Dalgarno, 1975), causes cryptic or "leaky" expression of the env gene in bacteria which harbor plasmids containing the env coding sequences (K. Rushlow, unpublished data). The expressed env proteins, most possibly toxic to the bacteria, result in aberrant cell growth patterns such as extremely slow growth in liquid cultures or small colony morphology on solid media. Those bacterial clones which grow normally or exhibit a wild-type phenotype, however, are usually selected for by mutations in the env gene which block the full expression of env. This phenomenon was documented earlier by the recombinant pSP65 in vitro transcription vector, containing the full-length S2-env sequence, which had expressed a truncated env
Figure 21. The eukaryotic transient expression vector p91023(B) (Kaufman et al., 1985; Wong et al., 1985; Sambrook and Gething, 1988). The vector contains the pBR322 prokaryotic origin of replication and tetracycline resistance gene, but lacks the "poison" sequences of pBR322 which have been claimed to inhibit DNA replication in eukaryotic cells (Luskey and Botchan, 1981). In addition, the vector contains the following eukaryotic regulatory elements: 1) the SV40 origin of replication and enhancer; 2) the adenovirus major late promoter coupled to a cDNA copy of the adenovirus tripartite leader; 3) a hybrid intron consisting of splice donor site from the first exon of the tripartite leader and a splice acceptor site from a mouse immunoglobulin gene; 4) the SV40 polyadenylation signal; 5) the adenovirus VAI and VAIII genes. The vector also contains the mouse dihydrofolate reductase (DHFR) coding sequence immediately downstream of the unique EcoRI cloning site. Both the cloned DNA and the DHFR gene lie within the same transcription unit, allowing the transcription of a dicistronic mRNA and translation of the DHFR coding sequence from the 3'-proximal open reading frame of the dicistronic transcript. In addition, the DHFR gene serves to provide a marker for selection and amplification of the vector DNA in transformed cells (Sambrook and Gething, 1988). The adenovirus tripartite leader and the VA RNA's have been proposed to increase the efficiency with which polycistronic mRNA's are translated (Logan and Shenk, 1984; Kaufman, 1985).
SV40 origin

Xha I  Xho I

intron

adenovirus major late promoter

Eco RI

adenovirus late tripartite leader sequence

dhfr

SV40 poly (A) addition site

adenovirus VA genes

Bam HI

tetracycline resistance

p91023 (B)
protein as a result of a single nucleotide deletion, and has also been observed and documented in a variety of EIAV env subclones (K. Rushlow, personal communication).

To assess the potential of the single-spliced EIAV env mRNA to encode both env and S2 proteins in vivo, the 2980-bp hybrid cDNA, previously cloned into the pSP65 in vitro transcription vector and containing the coding sequences of both env and S2, was cloned into the p91023(B) vector and propagated in E. coli. All attempts for recovering a recombinant clone containing the full-length cDNA insert were unsuccessfull, as all recombinant plasmids contained large deletions, presumably within the env gene. These observations were not surprising, in light of the problems mentioned above regarding plasmids harboring full-length EIAV env sequences. These problems are complicated even further due to the presence of an elaborate series of prokaryotic and eukaryotic regulatory elements within the p91023(B) expression vector (see below) which probably enhance the cryptic expression of the cloned EIAV env sequences when the plasmid is propogated in E. coli.

To obtain a functional expression plasmid lacking major deletions in the env gene, a 426-bp SspI-SmaI fragment of the EIAV provirus, starting at 21 nucleotides upstream of the S2 ATG codon and extending for 379 nucleotides downstream of the env ATG codon, was cloned into the p91023(B) vector. The recombinant plasmid, designated p9SR426 and containing the entire S2 coding sequence followed by the 5'-terminal 379 nucleotides of env, was then expected to encode the 7.5-kda S2 protein and an approximately 14.0-kda amino-
terminal env protein in transfected COS-7 cells, assuming the in vivo expression of both S2 and env from the same mRNA. Restriction enzyme mapping of recombinant p9SR426 plasmid DNA from many individual clones had indicated that the full-length 426-bp insert was present within the plasmid (data not shown). Thus the putative p9SR426 expression plasmid was transfected into COS-7 cells for an assessment of its ability to express both the S2 and env coding sequences.

To date no function has been associated with the putative EIAV S2 protein. Therefore detection of any expressed S2 protein in COS-7 cells transfected with the p9SR426 expression plasmid could not be based on a functional assay. Instead, assays for detection of S2 and the amino-terminal env protein, both of which could potentially be expressed from the expression plasmid, were based on immunological detection methods.

The EIAV env gene, 2578 nucleotides long, codes for the viral envelope and transmembrane glycoproteins (gp90 and gp45, respectively; Rushlow et al., 1986). Starting from the 5'-end of env, the first 1330 nucleotides code for gp90, while its latter 1248 nucleotides code for gp45. The env coding sequences present in p9SR426, as mentioned previously, contain the first 379 nucleotides of env. Thus the expressed env protein product of p9SR426 should be the amino-terminal 126 amino acids of gp90.

As an initial approach for detection of the expressed protein products of p9SR426 in COS-7 cells, indirect immunofluorescence microscopy was used. As primary antibody for detection of the putative env protein, a monoclonal antibody recognizing an epitope in the amino-terminal region of gp90 was used. On the
other hand, the primary antibody used for detection of the putative S2 protein was a rabbit polyclonal antibody raised against a 36-residue synthetic peptide derived from the sequences of S2. Each primary antibody was reacted with a separate preparation of COS-7 cells transfected with p9SR426. Subsequently, the cells were incubated with fluorescein-conjugated goat anti-mouse (in the case of anti-gp90 monoclonal antibody) or goat anti-rabbit IgG (in the case of anti-S2 polyclonal antibody), and examined under UV light. As shown in Figure 22, high level expression of the env coding sequence of p9SR426 can be clearly observed in 4-6% of COS-7 cells transfected with this plasmid, the percentage of cells expected to have incorporated the plasmid DNA. The expression of S2 in p9SR426-transfected cells, however, is not as clearly observed (Figure 23), although faint yellow-green fluorescent areas can be detected within the cell population examined. The high level of observed background fluorescence is due to the polyclonal nature of the anti-S2 antibody, making it difficult to distinguish positive fluorescent cells from the background of non-specific fluorescence.

To better detect the putative expression of S2 from p9SR426 in transfected COS-7 cells, an alternative approach was taken. Following transfection with p9SR426, COS-7 cells were labeled with 35S-methionine and -cysteine. The cellular extract was then immunoprecipitated using the anti-S2 polyclonal antibody described above, the precipitated products were fractionated by SDS-polyacrylamide gel electrophoresis, and they were detected by autoradiography. The results, however, indicated no detectable levels of the S2 protein (data not
Figure 22. Immunofluorescence photographs of fixed p9SR426-transfected COS-7 cell cultures treated with anti-gp90 monoclonal antibody followed by fluorescein-conjugated goat anti-mouse IgG.
Figure 23. Immunofluorescence photographs of fixed p9SR426-transfected COS-7 cell cultures treated with anti-S2 rabbit polyclonal antibody followed by fluorescein-conjugated goat anti-rabbit IgG.
shown), suggesting either a low level or a lack of S2 expression in p9SR426-transfected COS-7 cells.

The indirect immunofluorescence microscopy technique has allowed us to conclude that the env coding sequence of p9SR426 is indeed expressed in COS-7 cells, although no such conclusion can be made regarding the expression of the S2 gene from this plasmid. The potential implications of these findings are presented in Chapter 4.
CHAPTER 4
DISCUSSION

The results of the experiments described above have allowed us to determine the following:

(1) The number, sizes, and relative levels of the transcripts of EIAV encoded during persistent and cytopathic infections of equine cell lines.

(2) Quantitative differences in patterns of EIAV transcription during persistent and cytopathic infections.

(3) The splicing patterns of the full-length EIAV mRNA, including the splice donor and acceptor sites, used to generate the spliced EIAV mRNA's in virus-infected cells.

(4) A potential mechanism for expression of the viral env proteins and a putative regulatory protein from the same single-spliced viral transcript.

The results of the present investigations have indicated that EIAV encodes three species of mRNA in infected cells: an 8.2-kb full-length genomic mRNA, a 3.5-kb single-spliced mRNA, and a low abundance 1.5-kb mRNA, presumably formed by a double-splicing event of the full-length mRNA. Analysis of the levels of EIAV-specific RNA's present during persistent and cytopathic infections has revealed that quantitative differences characterize the transcriptional patterns of EIAV in these two infections. In persistently infected FEK cells the 8.2- and 3.5-kb transcripts are the predominant viral transcripts and are detected in approximately equal concentrations, while the 1.5-kb viral mRNA is detected at
very low levels. During the cytopathic infection of FDD cells the 3.5-kb mRNA is the predominant viral transcript, comprising nearly 75% of the total viral mRNA. Moreover, the cytopathic infection is characterized by almost a thirty-fold higher level of viral transcripts than those detected during the persistent infection.

It is not known whether the nearly 30-fold higher concentration of viral transcripts observed in the cytopathic infection of FDD cells, relative to that in persistently infected FEK cells, is a result of more efficient viral transcription or a consequence of greater stability of viral RNA's in infected FDD cells. Moreover, the differential transcriptional pattern of EIAV in the two cell-lines examined in these investigations suggests the involvement of a cell type-specific regulatory mechanism of viral transcription. It is possible that the more efficient splicing of the full-length EIAV genomic mRNA to the 3.5-kb env mRNA in infected FDD cells is a cell-specific phenomenon resulting from potential interactions between certain cellular factor(s) and viral regulatory protein(s). This may in turn affect the transcriptional initiation or the stability of the full-length viral mRNA, or the efficiency of its transport from the nucleus to the cytoplasm relative to that of the 3.5-kb mRNA. In the case of HIV-1, the trans-activator tat protein has already been shown to enhance the rate of viral transcription from the 5'-LTR (Jacobovits et al., 1988; Laspia et al., 1989), and the trans-acting rev protein has already been reported to have a role in transporting the full-length and single-spliced viral mRNA's to the cytoplasm and away from the splicing machinery in the nucleus (Felber et al., 1989; Hammarskjold et al., 1989; Malim et al., 1989). Furthermore, the enhanced levels of env mRNA in EIAV-infected FDD cells may result in
higher levels of viral glycoprotein production in these cells, thus contributing to
the cytopathic effects of the infection. In this regard, high level expression of the
HIV-1 envelope gene has been shown to have a direct role in syncytium
formation and cytopathicity (Sodroski et al., 1986).

The transcriptional patterns of the lentivirus full-length mRNA are
generally believed to be more complex than those of other retroviruses, and
studies on HIV-1 and visna virus point to the presence of a family of double-
spliced viral mRNA's, 1.5- to 2.0-kb long, which in general encode a variety of
trans-acting viral regulatory proteins (Arya et al., 1985; Rabson et al., 1985; Arya
and Gallo, 1986; Davis et al., 1987; Vigne et al., 1987; Mazarin et al., 1988; Davis
and Clements, 1989). Through Northern hybridization analysis of RNA from
virus-infected cells, reports by separate groups of investigators on the patterns of
HIV-1 or visna virus transcription in different cell lines have generally revealed
analogously complex but slightly different findings. For instance, in the continuous
virus producer human T-cell line H9 (Popovic et al., 1984), which is relatively
resistant to the cytopathic effects of HIV-1, Arya et al. (1984) have reported the
presence of four virus-specific RNA's of 9.4-, 4.2-, 2.0-, and 1.8-kb. On the other
hand, using the same infected H9 T-cell line, or phytohemagglutinin (PHA)-
stimulated normal human lymphocytes in which infection by HIV-1 results in
cytopathic effects and cell death, Rabson et al. (1985) have observed five virus-
specific RNA's of 9.1-, 5.5-, 5.0-, 4.3-, and 1.8-kb. Moreover, in the case of
visna virus, Davis et al. (1987) have reported the presence of viral transcripts of
9.4-, 5.0-, 4.3-, 1.8-, and 1.5-kb in virus-infected primary sheep choroid plexus
cells, while a separate report by Vigne et al. (1987) has pointed to the presence of viral RNA's of 9.4-, 4.8-, 4.3-, 3.7-, 1.6-, and 1.2-kb in infected American ovine fetal trachea cells. Additionally, Davis et al. (1987) have reported a temporal regulation of visna virus transcription in which the smallest viral transcript, the 1.5-kb RNA species, is detected 6 hours after the other viral transcripts are observed. On the other hand, studies by Vigne et al. (1987) have suggested a different pattern of temporal regulation of visna virus transcription in which the 1.2- and 1.6-kb viral RNA's are detected early after infection (at 24 hours), while all RNA's, including the 1.2- and 1.6-kb species, are observed late (at 72 hours) after infection. It is apparent that a determination of lentivirus transcriptional patterns by Northern hybridization analysis yields only an approximation of the actual mRNA structure. This approximation is partly due to variations in the length of the electrophoresis gel used by separate investigators to fractionate the viral mRNA's prior to Northern blotting, resulting in differences in resolution of the detected mRNA's on the blot. Other more vigorous analyses, such as determination of mRNA splice junctions by S1 nuclease mapping, or isolation and sequencing of cDNA clones of such spliced mRNA's are required before the exact structure of lentivirus mRNA's are determined. In this regard, the polymerase chain reaction (PCR) (Mullis and Fallona, 1987; Saiki et al., 1988) would be an extremely useful experimental tool for amplification and cloning of low abundance spliced viral mRNA's.

In addition to their complexity, the transcriptional patterns of HIV-1 and visna virus have a second characteristic in common which is the expression of the
family of low molecular weight transcripts of these viruses at concentrations equivalent to those of the full-length and single-spliced viral RNA species (Arya et al., 1985; Davis et al., 1987; Muesing et al., 1985; Rabson et al., 1985; Vigne et al., 1987). The results of the studies on patterns of EIAV transcription in the present studies, however, point to low detectable levels of the double-spliced 1.5-kb viral mRNA in cells of equine origin. Whether this observation is due to specific differences between the genetic organization of EIAV and the other lentiviruses, or is a result of the cell-type specific regulation of viral transcription described above remains to be determined. Future studies on the patterns of EIAV transcription in cell lines different from those utilized in the present studies should provide useful information in this regard.

The results of the S1 nuclease mapping experiments in the present investigations have confirmed the results obtained through Northern hybridization analyses using the splice junction-specific oligonucleotide probes, showing that the 5'-most splice donor site, immediately upstream of the gag gene, is the common donor site to all spliced EIAV mRNA's. Moreover, in addition to the splice acceptor site previously identified by sequencing of splice-junction cDNA obtained from the 3.5-kb mRNA, the S1 nuclease mapping experiments have identified and localized a second splice acceptor site upstream of the env gene, a splice donor site immediately downstream of S1, and a splice acceptor site at the 5'-boundary of S3.

Based on the body of information obtained on the splicing patterns of EIAV mRNA's through the experimental results, the splicing patterns of the
EIAV full-length 8.2-kb mRNA are summarized in Figure 24. The splice donor site located immediately upstream of the *gag* gene, 136-nt from the 5'-boundary of the primer binding site for reverse transcription, is the common donor site used to generate the first exon of all EIAV spliced mRNA's (Figure 24C-F). Following a single splicing event of the full-length mRNA, by which the donor site at position 136 is spliced to the acceptor site at position 1084, the *gag-pol* intron is removed and a 3.5-kb mRNA is produced whose second exon begins at the fourth codon of *SI* (position 1084 of Rushlow *et al.*, 1986), and contains the entire *S2* and *env* coding sequences (Figure 24C). Direct evidence for the existence of this spliced mRNA was already presented by cloning of cDNA primed within the coding sequences of the *env* gene. A second potential single splicing event, different than that presented in Figure 24C, is depicted in Figure 24D and would involve the splicing of the donor site at position 136 to the acceptor site at position 1133. This single-spliced mRNA, while still 3.5-kb long, would have a first exon identical to that of the former 3.5-kb mRNA (Figure 24C), while its second exon would start following the twentieth codon of *SI* and would contain the entire coding sequences of *S2* and *env*. It should be noted that the results of *S1* nuclease mapping studies, as shown in Figure 17B, indicate that the 3.5-kb mRNA utilizing the splice acceptor site at position 1084 (Figure 24C) is more abundant than the one using the acceptor site at position 1133 (Figure 24D).

A second splicing event of either or both 3.5-kb mRNA's, utilizing the donor site at position 1256 and the acceptor site at position 3184, would then
Figure 24. Summary of the splicing patterns of the EIAV full-length mRNA, as determined by cDNA cloning and sequencing, Northern hybridization analysis using splice donor-specific oligonucleotide probes, and S1 nuclease mapping.

A. Schematic diagram of the EIAV genome. B. The full-length, unspliced 8.2-kb EIAV mRNA. Solid and open triangles represent the location of splice donor and acceptor sites on the full-length mRNA, respectively. a. 136 nucleotides downstream of position 1, defined here as the 5'-boundary of the primer binding site (PBS) (Stephens et al., 1986). b. Based on the numbering system of Rushlow et al. (1986). C. The single-spliced 3.5-kb env mRNA. The existence of this spliced EIAV mRNA in virus-infected FDD cells was documented by cloning and sequencing of cDNA primed within the env coding sequences of the viral genome. D. An alternative single-spliced 3.5-kb mRNA generated by a potential splicing event utilizing the acceptor site at position 1133 of env. E and F. Double-spliced 1.4-kb mRNA's, either one or both of which may be present in EIAV-infected FDD cells.
A. □ □ gag □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ ^

B. 136a ▼ 1084b ▼ 1133b ▼ 1256b ▼ 3184b ▼ poly (A)

gag  pol  S1  S2  env  S3  poly (A)

C. 136 ▼ 1084 ▼ poly (A)

D. 136 ▼ 1133 ▼ poly (A)

E. 136 ▼ 1084 ▼ 1256 ▼ 3184 ▼ poly (A)

F. 136 ▼ 1133 ▼ 1256 ▼ 3184 ▼ poly (A)
remove most of the env coding sequences as a second intron and could potentially generate either one or two double-spliced 1.5-kb mRNA's (Figure 24E and F). The first exon of both double-spliced mRNA's would be identical to that of the two single-spliced transcripts. However, their second exon, beginning at position 1084 or 1133 (Figure 24E and F, respectively), would contain different coding capacities. The mRNA utilizing the acceptor site at position 1084 (Figure 24E) would retain all but the initial four codons of SI; however, in the case of the mRNA utilizing the acceptor site at position 1133 (Figure 24F), the first twenty codons of SI would be spliced out and the mRNA would retain only the latter thirty codons of SI. Following the UAA translational termination codon of SI, both mRNA's would contain the AUG initiation codon of S2 and its following six codons spliced onto the third exon of both mRNA's at the acceptor site located at the 5'-boundary of S3.

Although the S1 nuclease mapping data have identified the splice donor and acceptor sites used to generate the spliced mRNA's of EIAV, direct evidence for the actual presence of the latter three spliced mRNA's in EIAV-infected cells, as depicted in Figure 24D-F, can only be provided by obtaining their respective cDNA clones, similar to that of the 3.5-kb mRNA shown in Figure 24C. Moreover, it should be noted that only the most abundant viral mRNA's would be expected to be detected by the S1 nuclease mapping technique. As mentioned previously, other low abundance spliced viral mRNA's would best be detected by amplification techniques such as the PCR.
Recently, a cDNA clone of a double-spliced 1.4-kb EIAV mRNA has been isolated from an EIAV-infected feline embryo fibroblast (FEA) cell line (Derse et al., 1989). The sequence of the 1.4-kb cDNA indicates that its mRNA copy contains three exons. The first exon, containing the R and U5 regions of the 5'-LTR and 130 nucleotides of the region preceding the gag gene, is joined to the second exon using the splice donor site immediately upstream of the gag AUG codon and a splice acceptor site near the 5'-end of SI. The second exon, containing SI, is 142 nucleotides long and is joined to the third exon using a splice donor site immediately following the SI translational termination codon and a splice acceptor site within the env gene at the 5'-boundary of the S3. The third exon is 948 nucleotides long and contains S3 followed by the U3 and R regions of the 3'-LTR. It is interesting to note that this double-spliced mRNA, and the splice donor and acceptor sites used to generate it, are identical to those defined by the results of S1 nuclease mapping experiments in the present studies (Figure 24E), with one exception. The splice donor site at the end of the second exon, as identified by our S1 mapping experiments, is located at 31 nucleotides downstream of the SI termination codon (position 1256 of Rushlow et al., 1986; Figure 24E; Table III). However, the sequencing of the cDNA has identified this donor site as the one located at 1 nucleotide downstream of the SI termination codon (position 1226 of Rushlow et al., 1986; Table III; Derse et al., 1989). It is uncertain whether the identification of two different splice donor sites by the two separate studies is a result of differential cell type-specific mechanisms which control post-transcriptional processing of EIAV mRNA's in the EIAV-infected
FDD cells examined in the present studies, relative to the infected FEA cells used by Derse et al. (1989).

An additional observation regarding the double-spliced 1.4-kb mRNA identified by Derse et al. (1989) is that in this mRNA the S3 sequence, lacking an AUG initiation codon, is immediately preceded by the S1 termination codon and no translational initiation AUG codon is spliced onto the S3 sequence. On the other hand, the double-spliced mRNA identified by our S1 mapping experiments would have the S2 AUG codon near the 3'-end of its second exon spliced onto the S3 sequence (Figure 24E). However, this AUG codon would not be spliced in frame with the S3 coding sequence, thus leaving S3 without an AUG initiation codon. Therefore, in neither double-spliced mRNA does an AUG codon splice in frame onto the S3 coding sequence. Whether the S3 gene of EIAV indeed encodes a protein product remains to be determined, although serological evidence points to the expression of an S3-related protein in EIAV-infected horses (J. M. Ball and R. C. Montelaro, unpublished data).

As shown by the results of the present studies, a feature common to the spliced mRNA's of EIAV is that they all use the splice donor site located immediately upstream of the gag AUG codon to generate their first exon. This splicing pattern of the full-length genomic mRNA, which would generate spliced mRNA's sharing similar 5'-ends, has also been observed in other lentiviruses. S1 nuclease mapping studies of the 5'-region of visna virus (Vigne et al., 1987) and sequencing of cDNA clones obtained from spliced visna virus mRNA's (Mazarin et al., 1988; Davis and Clements, 1989) have indicated that all spliced visna virus
transcripts use a common donor site located at 188 nucleotides upstream of the visna virus gag gene (Sonigo et al., 1985). Moreover, sequencing of cDNA clones obtained from the spliced mRNA's of HIV-1 have revealed that the splice donor site located at 45 nucleotides upstream of the HIV-1 gag gene generates the first exon of these spliced transcripts (Arya et al., 1985; Muesing et al., 1985).

The interesting finding regarding the spliced transcripts of EIAV, however, is that following such splicing events the gag AUG codon is removed from all these mRNA's. Thus, for instance, in the double-spliced mRNA identified either by the present studies or by those of Derse et al. (1989), the SI coding sequence located within the second exon of the mRNA lacks an AUG initiation codon. In the case of HIV-1 or visna virus, however, the coding sequences of the short ORF's encoded by these viruses contain their own AUG codons and their translational initiation is not dependant on the splicing of an AUG codon from an upstream gene (Arya et al., 1985; Muesing et al., 1985; Mazarin et al., 1988; Davis and Clements, 1989). In the case of EIAV, it is proposed that the translation of the SI coding sequence on the double-spliced 1.4-kb mRNA is initiated from a non-AUG codon (Derse et al., 1989). Although the translational mechanism of SI from this mRNA is not fully understood yet, the usage of a non-AUG codon for translational initiation of a viral regulatory gene would potentially provide an additional control mechanism for fine-tuning of EIAV gene regulation not previously observed in other lentiviruses. The first two exons of the double-spliced 1.4-kb mRNA identified by Derse et al. (1989) are identical to the first two exons of the single-spliced 3.5-kb env mRNA identified in the present
studies (Figure 24C). Therefore it is plausible to assume that 51, 52, and env genes could all be expressed from this single-spliced mRNA. If this phenomenon indeed occurs in vivo, the efficient expression of the distal 52 and env genes would require that the proximal 51 coding sequence be poorly expressed. The utilization of a non-AUG codon, which might be poorly utilized and frequently passed over by the ribosomes, would potentially provide such a control mechanism for translation of 51 (Derse et al., 1989).

The results of the present investigations on the co-expression of 52 and env from the same transcript indicate that expression of both proteins from the same transcript can be detected in vitro. Studies on the in vivo expression of the two proteins, however, have not been as conclusive. Although expression of the env gene from a proviral DNA fragment containing the ATG codons of both 52 and env can be clearly demonstrated in transfected COS-7 cells, no such conclusion can be made regarding the expression of 52.

It is possible that the 52 protein is indeed not expressed in these cells. However, this lack of 52 detection, based solely on results of immunological detection methods, could also be due to other factors. It is possible, for instance, that the anti-52 titer of the polyclonal antibody is perhaps insufficient to detect any expressed protein in these experiments. In addition, as mentioned before, the length of the 5'-upstream sequences of the 52 gene in the p9SR426 expression plasmid have been minimized to avoid the problems associated with cloning of the env gene. It is possible that additional 5'-untranslated sequences, not present in p9SR426, are required for efficient translation of the 52 reading frame.
An interesting possibility is that the *in vivo* expression of the EIAV *S2* gene occurs at a much lower frequency relative to that of a structural viral gene such as *env*, making the *S2* protein difficult to detect. In this regard, it should be noted that the nucleotide sequence surrounding the *S2* AUG codon does not constitute a favorable context for translational initiation at this AUG, as defined by Kozak (1986a and 1986b). The sequence of nucleotides around the *S2* AUG codon is GAUAUAUGG (Rushlow *et al.*, 1986), while the AUG sequence context for optimal translational initiation is defined as CCACCAUGG (Kozak, 1986b). On the other hand, the *env* AUG codon, present in a GUAACAUUGG sequence context (Rushlow *et al.*, 1986), is much more favored for translational initiation. It has been suggested that positions -3 and +4 in this sequence context, where the "A" nucleotide of the AUG triplet is defined as the +1 position, are especially critical; the presence of a pyrimidine instead of a purine at these two positions can reduce the frequency of translational initiation by 5- to 10-fold (Kozak, 1986a). It is interesting to note that the *S2* AUG contains a pyrimidine at the -3 position, while the *env* AUG contains a purine. This suggests that if indeed translational initiation does occur from the *S2* AUG *in vivo*, it would happen much less frequently than from the *env* AUG. In the *in vitro* analysis of *S2* and *env* expression in a wheat-germ extract, however, it was observed that higher levels of *S2* were synthesized relative to *env* (Figure 20, lanes 1 and 2). In this regard, it should be noted that in the wheat-germ system translational initiation occurs much more efficiently from the 5'-most AUG codon (D. Shih, personal communication), which in this case is that of *S2*. 
Taken together with the model of infrequent $S_I$ translation from a non-AUG codon, as discussed above, it is plausible to propose that $S_I$, $S_2$, and $env$ could all be potentially expressed from the 3.5-kb $env$ transcript of EIAV. The virus can then be assumed to have evolved the following fine-tuning mechanism to insure optimal expression of the structural $env$ gene located distal to the 5'-end of the 3.5-kb mRNA. The first two genes ($S_I$ and $S_2$) would be inefficiently translated from a non-AUG codon and an AUG codon in a non-optimal sequence context, respectively, while a majority of the ribosomes initiate translation from the $env$ AUG codon, resulting in efficient synthesis of the viral envelope proteins.

The co-expression of more than one viral protein from the same mRNA has been observed in many different animal virus systems, including retroviruses such as HTLV-I (Nagashima et al., 1986), HTLV-II (Shimotohno et al., 1985), BLV (Rice et al., 1987), and HIV-1 (Arya and Gallo, 1986; Sodroski et al., 1986). An ideal means of detecting the potential expression of the $S_2$ gene from the expression plasmid p9SR426 used in the present studies would be based on a functional assay by which the function encoded by $S_2$ could be assayed in transfected cells. Design of such functional assays, however, await the assignment of a specific function to the putative EIAV $S_2$ protein.


### VITA

<table>
<thead>
<tr>
<th>Year</th>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>December 15, 1963</td>
<td>Born, Tehran, Iran.</td>
</tr>
<tr>
<td>1969-1975</td>
<td>Attended Mahboubi Grammar School, Tehran, Iran.</td>
</tr>
<tr>
<td>1975-1979</td>
<td>Attended Kharazmi High School, Tehran, Iran.</td>
</tr>
<tr>
<td>1984</td>
<td>Received B. S. in Biochemistry.</td>
</tr>
<tr>
<td>1989</td>
<td>Candidate for Ph. D. in Biochemistry, Louisiana State University, Baton Rouge, Louisiana.</td>
</tr>
</tbody>
</table>
Candidate: Siyamak Rasty

Major Field: Biochemistry

Title of Dissertation: Characterization of the Transcriptional Properties of Equine Infectious Anemia Virus

Approved:

Ronald C. Montelaro
Major Professor and Chairman

Dean of the Graduate School

EXAMINING COMMITTEE:

Kenneth M. Ford
Earl C. Coon
Sierra H. Cheung
Kathleen M. Veith
O. A. Shih

Date of Examination:
November 27, 1989