JOURNAL OF VIROLOGY, May 1994, p. 3071–3079
0022-538X/94/$04.00+0

Vol. 68, No. 5

# The *tat/rev* Intron of Human Immunodeficiency Virus Type 1 Is Inefficiently Spliced because of Suboptimal Signals in the 3′ Splice Site

ALFREDO STAFFA AND ALAN COCHRANE*

*Department of Microbiology and Immunology, McGill University, Montreal, Quebec, Canada H3A 2B4*

Received 24 August 1993/Accepted 16 February 1994

**Proportional expression of retroviral genes requires that splicing of the viral primary transcript be an inefficient process. Much of our current knowledge about retroviral suboptimal splicing comes from studies with Rous sarcoma virus. In this report, we describe the use of chimeric introns composed of human β-globin and human immunodeficiency virus type 1 (HIV-1) splice sites to establish the basis for inefficient splicing of the intron which comprises most of the HIV-1 *env* coding sequences (referred to as the *tat/rev* intron). S1 RNA analysis of transfected COS-7 cells revealed that the 3′ splice site (3′ ss) of this region was significantly less efficient than the 3′ ss of the first intron of β-globin. Deletion of sequences flanking the *tat/rev* intron 3′ ss demonstrated that the requirements for its inefficiency reside within the region that is expected to comprise the essential signals for splicing (i.e., the branchpoint region, the polypyrimidine tract, and the AG dinucleotide). Introduction of an exact copy of the efficient β-globin branchpoint sequence within a highly conserved region rendered the *tat/rev* intron 3′ ss highly efficient. Improvement of the polypyrimidine tract also increased the splicing efficiency, but to a degree slightly less than that obtained with the branchpoint mutation. Subsequent examination of the *tat/rev* intron 5′ splice site in a heterologous context revealed that it is efficiently utilized. These results indicate that both a poor branchpoint region and a poor polypyrimidine tract are responsible for the low splicing efficiency of the HIV-1 *tat/rev* intron. It is of fundamental interest to establish the basis for inefficient splicing of the HIV-1 *tat/rev* intron since it may provide the key to understanding why nuclear export of mRNAs encoding HIV-1 structural proteins is Rev dependent.**

Primary transcripts of most higher eukaryotic protein-encoding genes transcribed by RNA polymerase II are interrupted by introns. These introns are usually selectively and efficiently removed by splicing, a nuclear process in which the 5′ splice site (5′ ss) and 3′ splice site (3′ ss) are cleaved and joined prior to export of the mRNA to the cytoplasm. The splicing process involves a complex series of protein-protein and protein-RNA interactions between the pre-mRNA, components of the spliceosome (U1, U2, U4/U6, and U5 small nuclear ribonucleoproteins [snRNPs]), and other auxiliary factors (18, 30). Given the complexity of the splicing pathway, there is great potential for elaborate regulation. In fact, splicing of primary transcripts is emerging as an important level of regulation in gene expression. Species ranging from *Drosophila melanogaster* to humans have evolved regulated mechanisms such as alternative splicing and exon skipping to express multiple protein products from a single primary transcript (26, 39). Some viruses of higher eukaryotes have also evolved mechanisms to maximize the coding potential of their genomes. While DNA viruses such as simian virus 40 (SV40) utilize mechanisms of regulated splicing similar to those of their cellular hosts (i.e., alternative splicing) (17), retroviruses have evolved different posttranscriptional regulatory mechanisms. In the case of retroviruses, only a fraction of the primary transcripts are spliced to subgenomic RNAs (8, 46). Proportional expression of retroviral genes requires that splicing of the genome-length primary transcript by the cellular splicing machinery be an inefficient process. Furthermore, unlike most cellular transcripts which are restricted to the nucleus unless spliced, a certain proportion of full-length retroviral transcripts must be exported to the cytoplasm to serve as mRNA for the *gag* and *pol* gene products and as genomic RNA for packaging into virus particles (8, 46).

Although the *Lentiviridae* and *Oncoviridae* retroviral classes share similar pathways of replication, they display striking differences in genome complexity and regulation, particularly in the metabolism of the viral primary transcript (8, 46). Compared with simple retroviruses such as the Rous sarcoma virus (RSV) of the *Oncoviridae* class, the human immunodeficiency virus type 1 (HIV-1) has a more complex RNA-splicing pattern (37, 43). In addition to producing the Gag, Pol, and Env proteins common to all replication-competent retroviruses, the HIV-1 primary transcript also serves to produce six regulatory proteins (10, 14), two of which—Tat and Rev—are essential for HIV-1 replication (2, 11, 15, 31, 41, 42, 44, 47). Moreover, unlike the RNAs of simple retroviruses, which are constitutively transported from the nucleus to the cytoplasm (43), the unspliced and singly spliced HIV-1 RNAs (which encode structural viral proteins) are retained in the nucleus in the absence of the HIV-1 Rev protein; that is, only the doubly spliced, 2-kb class of HIV-1 RNAs is constitutively transported to the cytoplasm and translated independently of the expression of any viral proteins (12, 13, 19, 25). Since all HIV-1 RNAs with an intact *tat/rev* intron are not constitutively transported to the cytoplasm (12, 13, 19, 25), it seems that elements within this intron must contribute to the nuclear entrapment of the incompletely spliced RNAs. Therefore, mechanisms unique to the lentivirus class of retroviruses must exist which (i) allow the majority of the RNA to escape splicing and (ii) cause nuclear retention of incompletely spliced RNAs in the absence of viral regulatory proteins. To date, intronic

* Corresponding author. Mailing address: Department of Microbiology and Immunology, McGill University, 3775 University St., Montreal, Quebec, Canada H3A 2B4. Phone: (514) 398-6653. Fax: (514) 398-7052. Electronic mail address: cza8@musica.mcgill.ca.

sequences crucial for balanced retroviral RNA splicing have been identified in avian sarcoma viruses, namely, a suboptimal 3′ ss (3, 16, 20, 21) and a negative regulator of splicing (NRS) that can function in *cis* to decrease splicing of both RSV and cellular introns (1, 27, 28).

To study the basis for the inefficient splicing observed for the HIV-1 *tat/rev* intron, we have created chimeric introns using splice sites of the human β-globin first intron and the HIV-1 *tat/rev* intron. S1 analysis of total RNA from transfected COS-7 cells reveals that the *tat/rev* intron 3′ ss, like that of RSV, is inefficient. Mutagenesis experiments indicate that the *tat/rev* intron 3′ ss can be rendered efficient either by the introduction of an efficient branchpoint sequence 30 nucleotides (nt) upstream of the AG dinucleotide or by improvement of the polypyrimidine tract. Subsequent experiments carried out with the *tat/rev* intron 5′ ss in a heterologous context reveal that it is efficiently utilized by the splicing machinery. Unlike the NRS-containing intron of RSV (1, 27, 28), the *tat/rev* intron does not appear to contain any *cis*-acting sequences that modulate splicing efficiency. Therefore, it appears that the *tat/rev* intron is inefficiently spliced simply because of the absence of efficient signals within the 3′ ss. Differences in the regulation of splicing between HIV-1 and RSV may have important implications for the fate of inefficiently spliced RNA with respect to nuclear export.

## MATERIALS AND METHODS

**Construction of plasmids.** The parent construct, pSVβ, was constructed as follows. The *Eco*RI-*Bam*HI fragment of pEE14 (a generous gift from Celltech Ltd.) carrying the SV40 early polyadenylation signal was inserted into the respective sites of pBluescript SK(+) and (BlSK; Stratagene). Subsequently, the *Bam*HI and *Xba*I sites were eliminated by simultaneous restriction enzyme digestion, blunting the protruding 5′ ends with Klenow fragment, and religating. The blunt-ended *Sal*I-*Xba*I fragment of Bl-CAT containing the chloramphenicol acetyltransferase (CAT) open reading frame was then inserted into the unique *Eco*RV site upstream of the SV40 early polyadenylation signal. Finally, the SV40 early promoter, contained on a *Xho*I-*Hin*dIII fragment, and a modified human β-globin intron (IVS-1), contained on a 214-bp *Hin*dIII-*Bam*HI fragment, were inserted simultaneously into the *Xho*I and *Bam*HI sites, generating pSVβ.

HIV-1 sequences spanning the *tat/rev* intron 3′ ss (at nt 8377 [the nucleotide numbering system used throughout corresponds to that of the Hxb2 complete genome sequence in GenBank, accession no. K03455]) were derived from the Hxb2 proviral clone. The 321-bp HT fragment from the *Hin*dIII site at nt 8141 to the *Taq*I site at nt 8462 was blunted and inserted into the *Eco*RV site of BlSK in the T7 orientation, generating Bl-HIVSA HT. The CT fragment (nt 8333 to 8462) was subcloned from Bl-HIVSA HT by PCR with the *Cla*I sense primer, 5′-TTTTGCTGTACTATCGATAGTGAATAG AGT-3′ (nt 8317 to 8346), and the T3 primer of BlSK, 5′-CTCGGAATTAACCCTCAC-3′. The 152-bp *Cla*I-*Bam*HI fragment of the PCR-amplified DNA was inserted into BlSK, generating Bl-HIVSA CT.

pSVHTSB and pSVCTSB were generated by inserting the *Sal*I-*Bam*HI fragments of Bl-HIVSA HT and Bl-HIVSA CT, respectively, into the unique *Sal*I and *Bam*HI sites of pSVβ. pSVCBSB was generated by PCR as follows. pSVCTSB was used as template DNA for PCR amplification with the T7 primer of BlSK, 5′-AGTGAATTGTAATACGAC-3′, and the SA-Bam antisense primer, 5′-GGATCCGGGTCTGAAAC GATAATG-3′ (nt 8364 to 8381). The *Xho*I-*Bam*HI fragment of pSVβ was subsequently replaced by the *Xho*I-*Bam*HI fragment of the PCR-amplified DNA, generating pSVCBSB.

The 2.4-kb fragment spanning the *tat/rev* intron 5′ ss (at nt 6045) was derived from Bl-env SX C/X (6a). This construct contains a modified *Sal*I-*Xho*I fragment of Hxb2 (nt 5786 to 8897) in which the *Hin*dIII site at nt 8141 was converted to an *Xba*I site by site-directed mutagenesis (as outlined by Bio-Rad). This *Xba*I site was converted to a *Sal*I site by the addition of a *Sal*I linker. pSVenvβ was constructed by inserting the 2.4-kb *Sal*I fragment into the unique *Sal*I site of pSVβSA, a modified pSVβ in which the β-globin 5′ ss had been deleted. pSVHIVβ was constructed as follows. A 159-bp fragment spanning the *tat/rev* intron 5′ ss was obtained by PCR amplification with the 5′HIVSD sense primer, 5′-CCAAGCTT GAGCTCATCAGAACAGTC-3′ (nt 5999 to 6016) and the 3′HIVSD antisense primer, 5′-GGGAATTCTGATTACTAT GGACCA-3′ (nt 6128 to 6142). The 116-bp *Hin*dIII-*Eco*RI fragment of the PCR-amplified DNA was subcloned into BlSK, generating Bl-envSD. The SV40 early promoter, contained on an *Xho*I-*Hin*dIII fragment, was inserted into the *Xho*I and *Hin*dIII sites of Bl-envSD. The *Eco*RI site was subsequently converted to a *Sal*I site by the addition of a *Sal*I linker, and the *Kpn*I-*Sal*I fragment of the resulting construct was inserted into the *Kpn*I and *Sal*I sites of pSVβ, generating pSVHIVβ.

**Introduction of the branchpoint and polypyrimidine tract mutations into the HT fragment.** For introduction of the β-globin branch point, mutagenesis was carried out entirely in vitro by three sequential PCRs as follows. In the first reaction, a fragment was amplified from BI-HIVSA CT with the mutagenic primer β-BP, 5′-ATAGTGAATACACTGACGCAGG GATATTC-3′ (nt 8333 to 8361), and the T3 primer. The β-BP primer contains four mismatches (underlined) which create an exact copy of the reported β-globin branchpoint sequence, CACTGAC (32). The expected site of the RNA branch formation (shown in boldface type) coincides with an A residue 30 nt upstream of the AG dinucleotide of the HIV-1 *tat/rev* intron 3′ ss. An aliquot of the resulting 229-bp PCR product was used as a pseudoprimer in conjunction with the β-PCR1 primer, 5′-CTTAAGTTGGTGGTGAGG-3′ (a sense primer complementary to sequences 5′ of the β-globin 5′ ss), in a second PCR to amplify a 488-bp fragment from pSVHTSB. To increase the yield of amplified DNA, 1/10 of the second PCR product was further amplified by PCR with the β-PCR1 and T3 primers. The *Sal*I-*Bam*HI fragment of pSVβ was subsequently replaced by the *Sal*I-*Bam*HI fragment of the final PCR product, generating pSVHTSB*.

To introduce mutations into the polypyrimidine tract of the *tat/rev* intron, a strategy identical to that outlined above was employed, the only modification being that the mutagenic primer polypyr, 5′-GGATATTCTCCTTTTTCTTTTCAGAC 3′, was utilized in place of the β-BP primer. Consequently, purine residues at positions −7, −10, −13, and −16 relative to the AG dinucleotide of the 3′ ss (shown in boldface type) were changed to thymidine (underlined). These mutations resulted in the generation of an uninterrupted polypyrimidine tract of 17 pyrimidines. The *Sal*I-*Bam*HI fragment from the final PCR was inserted into pSVβ to generate pSVHTSB$^{ppt}$. To ensure that no other alterations to the sequence had occurred during the amplification cycles, all mutations were confirmed by chain-terminating DNA sequencing (35).

**Cell culture and transfection.** COS-7 cells (ATCC CRL1651) were maintained in Iscove's modified Dulbecco's medium supplemented with 10% fetal bovine serum. Culture dishes (diameter, 100 mm) were seeded with $10^6$ cells, and the cells were transfected 24 h later by the DEAE-dextran method

as previously described (9). Forty-eight hours posttransfection, the cells were washed twice in phosphate-buffered saline, gently scraped off the dishes, and pelleted. The cell pellet was resuspended in 4 M guanidine thiocyanate–25 mM sodium citrate (pH 7.0)–0.5% Sarkosyl–0.1 M β-mercaptoethanol. Total RNA was extracted as previously described (6). After the final precipitation, RNA pellets were washed with 70% ethanol, dried, and redissolved in diethylpyrocarbonate-treated water. RNA was quantitated by determining the $A_{260}$ (an $A_{260}$ value of 1 is 50 μg/ml).

**S1 nuclease protection analysis.** DNA fragments used as probes were from either the *Sal*I or the *Hin*dIII site upstream of the 3′ ss to the *Pvu*II site within the CAT open reading frame. To distinguish reannealed probe from protected probe resulting from unspliced RNA, DNA fragments contained heterologous, BlSK-derived DNA sequences upstream of the 3′ ss. The probe for the β-galactosidase control was contiguous with a 3′ portion of the *lacZ* gene—from the *Pvu*II site to the *Eco*RI site of pCH110 (Pharmacia; nt 2929 to 3286)—and also contained heterologous sequences to distinguish reannealed probe from protected probe resulting from hybridization to RNA. Each DNA fragment was radiolabelled at the 5′ end with [γ-$^{32}$P]ATP by using T4 polynucleotide kinase (34). End-labelled DNA probe (approximately $2 \times 10^3$ cpm; specific activity, approximately $0.5 \times 10^6$ cpm/μg) was added to total RNA (5 μg) in 30 μl of hybridization buffer {80% formamide; 40 mM PIPES [piperazine-*N*,*N*′-bis(2-ethanesulfonic acid)], pH 6.8; 400 mM NaCl; 1 mM EDTA}. The nucleic acid was denatured at 75°C for 10 min and allowed to hybridize for 16 h at 50°C. Following hybridization, samples were rapidly cooled to 4°C. Single-stranded nucleic acid was digested by adding 300 μl of S1 digestion mix (50 mM sodium acetate [pH 4.5], 280 mM NaCl, 4.5 mM $ZnSO_4$, 20 μg of denatured, sonicated salmon sperm DNA per ml, 1,000 U of S1 nuclease per ml) and incubating at 25°C for 2 h. The reaction was terminated by the addition of 5 μl of 500 mM EDTA and 20 μg of *Escherichia coli* tRNA. Proteins were removed by phenol-chloroform (1:1) extraction, and the aqueous phase was ethanol precipitated. The dried nucleic acid pellet was resuspended in 80% formamide–10 mM EDTA, heated at 85°C for 5 min, and subjected to electrophoresis on a 5% polyacrylamide–8 M urea gel. The protected DNA fragments were visualized either by autoradiography on Kodak XAR film or by using a phosphorimager (Molecular Dynamics). Quantitation of phosphorimager-derived data was carried out by using Image Quant software as described by the manufacturer (Molecular Dynamics).

## RESULTS

**Construction of pSVβ for examination of the efficiency of HIV-1 splice sites.** In order to facilitate the study of heterologous splice site efficiency, the pSVβ plasmid, which contains a modified human β-globin intron (IVS-1) upstream of the CAT open reading frame, was constructed (Fig. 1A). The modified β-globin intron contains several unique restriction enzyme sites into which heterologous splice sites can be inserted to analyze their efficiency of utilization. In addition, the β-globin intron contains several AUG codons (data not shown) such that translation of the RNA into functional CAT protein is dependent upon the removal of the intron. This feature provides a readily quantitated marker for gene expression (Table 1). S1 analysis revealed that the β-globin intron is efficiently excised from pre-mRNA in transfected COS-7 cells (Fig. 1B, lane 8; Table 1). Analysis was carried out with a probe spanning the 3′ ss, yielding protected products of 302 and 206 nt corresponding to the unspliced and spliced forms of the RNA, respectively. Furthermore, experiments using a probe spanning the entire CAT open reading frame of pSVβ indicated that no cryptic splice sites within the CAT sequences are utilized (data not shown). This efficient splicing results in the accumulation of a considerable amount of spliced pSVβ RNA (Fig. 1B, lane 8) and, in turn, high-level expression of CAT enzyme activity (Table 1). Thus, the pSVβ construct provides a convenient system to examine the basis for the inefficient splicing of the *tat/rev* intron of HIV-1.

**The *tat/rev* intron 3′ ss is intrinsically inefficient.** In order to study the efficiency of the *tat/rev* intron 3′ ss, the SB series was constructed (Fig. 1A). In this series, the β-globin 3′ ss of pSVβ was replaced by *env* sequences spanning the *tat/rev* 3′ ss by insertion into the *Sal*I and *Bam*HI sites. The largest fragment tested (designated HT) contains 236 nt 5′ and 85 nt 3′ of the *tat/rev* 3′ ss. To test whether flanking sequences modulate the efficiency of the 3′ ss, either 5′ sequences or both 5′ and 3′ sequences were deleted (designated CT and CB fragments, respectively). Analysis of the RNA generated from pSVHTSB revealed that the *tat/rev* intron 3′ ss is utilized (Fig. 1B, lane 9), but at an extremely low level relative to utilization of the β-globin 3′ ss of pSVβ (Fig. 1B, lane 8). The level of splicing observed with the pSVHTSB chimeric construct is comparable to the ratio of unspliced and spliced viral RNAs derived from other HIV *env* expression vectors (13, 19, 25). These results indicate that the requirements for suboptimal splicing are located within the 321-bp HT fragment. Examination of the pSVCTSB and pSVCBSB constructs revealed that deletion of flanking sequences does not increase the efficiency of the *tat/rev* intron 3′ ss (Fig. 1B, lanes 10 and 11). In fact, comparison of the ratio of spliced to unspliced RNA indicates that splicing of pSVCTSB and pSVCBSB transcripts is less efficient (or undetectable in the case of pSVCBSB) than that of pSVHTSB transcripts. Therefore, the results obtained with the SB series indicate that the *tat/rev* intron 3′ ss is functional, albeit at a very poor efficiency. Moreover, requirements for suboptimal splicing appear to be confined to the region that is expected to contain the signals for splicing (i.e., the branch point, the polypyrimidine tract, and the AG dinucleotide). Consistent with the hypothesis that the *tat/rev* intron 3′ ss is inefficiently recognized by the host splicing apparatus, we have also determined that the *tat/rev* intron 3′ ss cannot compete in *cis* with efficient splice sites such as that of β-globin (42a).

**The *tat/rev* intron 3′ ss is rendered efficient by the introduction of the β-globin branch point or mutation of the polypyrimidine tract.** To examine the *tat/rev* intron 3′ ss in more detail, the sequences comprising the 3′ ss from several different HIV-1 proviral clones were compared (Fig. 2). The striking observation is the presence of an extremely conserved region between 20 and 39 nt upstream of the AG dinucleotide of the *tat/rev* intron 3′ ss. It is within this region that the A residue involved in RNA branch formation is expected to be located (18). Although several potential A residues are found within this region, none are contained within a context bearing similarity to the weakly conserved 5′-UNCURAC-3′ branchpoint consensus of higher eukaryotes (18). In addition to the conservation of sequence in the anticipated branchpoint region, there is also conservation of interruptions within the polypyrimidine tract. These findings suggest (i) that inefficient utilization of the *tat/rev* intron 3′ ss may be due to a poor branchpoint region and/or polypyrimidine tract and (ii) that there is an evolutionary constraint on the virus to maintain these suboptimal splicing signals.

To test the hypothesis that inefficient splicing of the *tat/rev* intron 3′ ss is due to either a poor branchpoint region or a poor

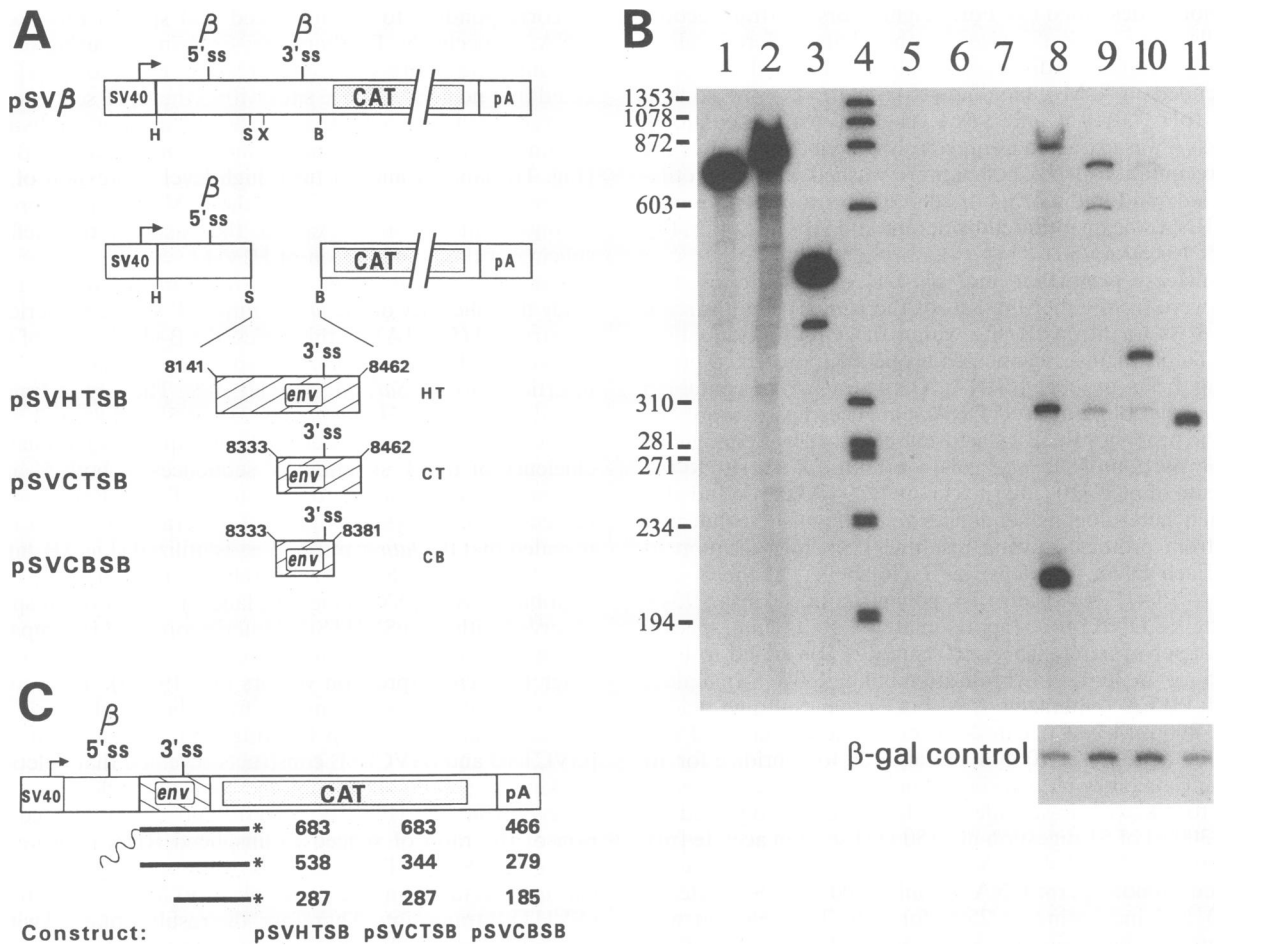


FIG. 1. (A) Schematic representation of pSVβ and its SB derivatives. All constructs contain the SV40 early promoter and early polyadenylation signal (pA). The CAT open reading frame is depicted by a dotted box. Numbers above HIV-1 sequences (HT, CT, and CB fragments [hatched boxes]) spanning the *tat/rev* intron 3′ ss are from the Hxb2 numbering system in GenBank accession no. K03455. Splice sites are indicated by vertical lines. β 5′ ss, β-globin 5′ ss; β 3′ ss, β-globin 3′ ss. The *tat/rev* intron 3′ ss is labelled 3′ ss within the *env* sequence. Important restriction enzyme sites are also shown. H, *Hin*dIII; S, *Sal*I; X, *Xba*I; B, *Bam*HI. The SB constructs were derived by insertion of HIV-1 sequences into the unique *Sal*I and *Bam*HI sites of pSVβ. (B) S1 nuclease protection analysis of total RNA from COS-7 cells transfected with the SB constructs. Lane 1 contains 1/10 the amount of probe used for S1 protection analysis in lanes 5, 9, and 10; lane 2 contains 1/10 the amount of probe used for S1 protection analysis in lanes 6 and 8; lane 3 contains 1/10 the amount of probe used for S1 protection analysis in lanes 7 and 11. Lane 4 contains radiolabelled markers. A 5-μg aliquot of total RNA from COS-7 cells transfected with pSVβ (lane 8), pSVHTSB (lane 9), pSVCTSB (lane 10), and pSVCBSB (lane 11) was used in the S1 protection reactions. As a negative control, total RNA from mock-transfected COS-7 cells (lanes 5, 6, and 7) was incubated with probes and digested with S1 nuclease. The β-galactosidase (β-gal) control S1 protection reactions carried out in parallel using a β-galactosidase-specific probe are also shown below the large gel. Numbers on the left indicate molecular sizes (in base pairs). (C) Schematic representation of the expected protection pattern for the SB constructs outlined in panel A, indicating the sizes of the probes and protected fragments corresponding to the unspliced and spliced RNAs derived from the constructs tested. The wavy line corresponds to heterologous sequence used to discriminate between unspliced RNA and reannealed probe.

polypyrimidine tract, several mutations were generated. In the first instance, the reported β-globin branchpoint sequence CACTGAC (32) was introduced into pSVHTSB by mutagenesis to generate pSVHTSB* (Fig. 3A). This construct is identical to pSVHTSB except that the *env* sequences contain altered bases at positions −29, −31, −33, and −35 relative to the AG dinucleotide of the 3′ ss. The sequence 5′-GAGT TAG-3′ at positions −35 to −29 within the highly conserved region was converted to the β-globin branchpoint sequence shown above. This particular location within the conserved region was chosen for several reasons: (i) only four point mutations (underlined) were required to create an exact copy of the β-globin branchpoint sequence, (ii) the sequence resides within the −18 to −38 distance from the AG dinucleotide commonly found for RNA branch formation, (iii) no A residues within the highly conserved region are altered, and (iv) the expected branchpoint A residue (shown above in boldface) coincides with a naturally occurring A residue 30 nt upstream of the AG dinucleotide of the *tat/rev* intron 3′ ss. S1 analysis (Fig. 3B, lane 4) revealed that this minor alteration of the primary nucleotide sequence, which introduces an efficient branch point, dramatically increases the level of spliced RNA observed. Comparison between pSVHTSB* and pSVβ of the percentage of total RNA that is of the spliced form indicates that splicing of the pSVHTSB* transcript is as efficient as that of the pSVβ transcript (Table 1). The higher ratio of spliced to unspliced RNA observed for pSVHTSB* than for pSVβ is due both to a higher level of spliced product (94.17 versus 81.2%)

TABLE 1. Analysis of CAT activity and RNA species generated by pSVβ and its HTSB derivatives[a]

| Construct[b] | CAT/β-galactosidase activity[c] | RNA analysis[d] | |
|---|---|---|---|
| | | Spliced/unspliced ratio | % Spliced |
| pSVβ | 21.44 ± 15.96 | 4.35 ± 0.53 | 81.20 ± 1.78 |
| pSVHTSB | 0.47 ± 0.18 | 0.56 ± 0.09 | 40.57 ± 6.00 |
| pSVHTSB* | 9.73 ± 1.27 | 16.30 ± 1.89 | 94.17 ± 0.68 |
| pSVHTSB$^{ppt}$ | 6.27 ± 0.88 | 6.92 ± 0.30 | 87.37 ± 0.51 |

[a] Results are averages for three independent sets of transfections. Standard deviations between the three experiments are also indicated.

[b] Each construct was cotransfected with the β-galactosidase expression plasmid pCH110 (Pharmacia).

[c] Parallel assays for CAT and β-galactosidase activities were performed as previously described (34). Numbers represent % conversion of CAT substrate normalized to the $A_{420}$ obtained for the respective β-galactosidase assay.

[d] Quantitation of S1 RNA protection experiments such as that shown in Fig. 3B was carried out by using a phosphorimager as described by the manufacturer (Molecular Dynamics).


FIG. 2. Comparison of the nucleotide sequences comprising the *tat/rev* intron 3′ ss of various HIV-1 proviral clones. Sequences conserved in all 10 proviral clones are boxed. At the extreme 3′ end of each sequence is the AG dinucleotide that defines the 3′ end of the *tat/rev* intron. Immediately upstream is the somewhat conserved polypyrimidine tract preceded by a large conserved region between 20 and 39 nt upstream of the AG dinucleotide. It is within this area that the critical A residue involved in formation of the RNA branch point is expected to reside. However, no obvious homology to the 5′-UNCURAC-3′ branchpoint consensus is detected within this region. The amino acid composition of the portion of the *env* gene product encoded by the nucleotide sequence is also shown. All HIV-1 proviral sequences were obtained from the GenBank data base.

and to a slightly lower absolute level of pSVHTSB* unspliced RNA (Table 1 and Fig. 3B). Therefore, four simple point mutations within an evolutionarily conserved region encompassing the putative site of RNA branch formation are sufficient to render the otherwise inefficient *tat/rev* intron 3′ ss highly efficient.

In order to evaluate the contribution of the polypyrimidine tract to the overall splicing inefficiency observed, the four purine residues interrupting the polypyrimidine tract were converted to T residues by site-directed mutagenesis, resulting in the generation of an uninterrupted polypyrimidine tract 17 nt long (Fig. 3A). Analysis of this construct also revealed an increase in utilization of the *tat/rev* intron 3′ ss (Fig. 3B, lane 5; Table 1). Therefore, it can be concluded that the four purines that interrupt the polypyrimidine tract of the *tat/rev* intron 3′ ss contribute to the overall inefficiency of this site. The observation that the steady-state levels of unspliced RNA do not differ between the pSVHTSB construct and its derivatives indicates that the differences in levels of spliced RNA cannot be attributed to differences in the stability of the unspliced RNA. Rather, it would appear that if the RNA is unable to be efficiently spliced, it is degraded in the nucleus.

**The *tat/rev* intron 5′ ss is efficiently utilized in a heterologous context.** The demonstration that the *tat/rev* intron 3′ ss is inefficient is similar to previous findings with the RSV system (3, 20, 21). Moreover, the branchpoint region has also been implicated in splicing inefficiency in the RSV system (16). Further work with RSV has identified an additional *cis*-acting element, NRS, that functions to downregulate splicing efficiency within the context of both RSV and heterologous introns (1, 27, 28). To test for the presence of analogous sequences within the *env* region of HIV-1, the pSVβ construct was used to examine the efficiency of the *tat/rev* intron 5′ ss alone or in the context of most of the *env* coding region (Fig. 4A). The construct with the *tat/rev* intron 5′ ss, pSVHIVβ, contains 17 and 99 bp of HIV-1 sequences upstream and downstream of the 5′ ss, respectively. To address the possibility that the efficiency of the *tat/rev* intron 5′ ss is downregulated by distinct, *cis*-acting elements outside the 116-bp fragment, a larger portion of the HIV-1 genome spanning the 5′ ss was also tested. pSVenvβ contains a 2.4-kb fragment of the HIV-1 genome which includes 258 bp upstream of the 5′ ss and most of the *tat/rev* intron. S1 analysis of the RNA from cells transfected with these constructs (Fig. 4B) demonstrated that the *tat/rev* intron 5′ ss is at least as efficient as the β-globin 5′ ss, whether alone (Fig. 4B, lane 4) or in the context of the *tat/rev* intron (Fig. 4B, lane 5). Therefore, unlike the inefficient splicing in the RSV system, inefficient splicing of the *tat/rev* intron cannot be ascribed to the action of *cis*-acting repressive sequences within the intron.

## DISCUSSION

Previous studies with simple retroviruses such as RSV have demonstrated that a suboptimal 3′ ss and the presence of an NRS which downregulates the utilization of the 5′ ss contribute to the inefficient excision of introns within the viral RNA (1, 3, 16, 20, 21, 27, 28). That is, in order for their primary transcripts to evade the cellular splicing machinery most of the time, these simple retroviruses have evolved an intronic organization which comprises an intrinsically inefficient 3′ ss and a distinct, *cis*-acting intron element which downregulates the efficiency of an otherwise efficient 5′ ss by a currently unknown mechanism.

In an effort to determine the basis for the inefficient splicing of the *tat/rev* intron, chimeric introns composed of β-globin and *tat/rev* intron splice sites were examined. The results obtained indicate that if *cis*-acting sequences which modulate splice site efficiency are present within the *tat/rev* intron, they must be context dependent. Deletion mapping demonstrated that, in the context of the constructs examined, the requirements for suboptimal splicing are intrinsic to the region expected to comprise all of the signals of the *tat/rev* intron 3′ ss, since removal of both 5′ and 3′ flanking sequences did not result in an increase in spliced RNA. In fact, analysis of the SB constructs suggests that sequences 3′ of the AG dinucleotide of the 3′ ss act to augment splicing, given that deletion of these sequences in pSVCBSB dramatically reduces the level of spliced product relative to that obtained with either pSVHTSB or pSVCTSB (Fig. 1). This role of exon sequences in splicing is similar to that observed in other systems (45), including RSV (16). Furthermore, the HIV-1 *env* sequences implicated here (i.e., nt 8382 to 8462) encompass a purine-rich region (data not shown) similar to other exon sequences that have been reported to affect splicing efficiency (45). This observation suggests that these downstream "splicing enhancer" sequences

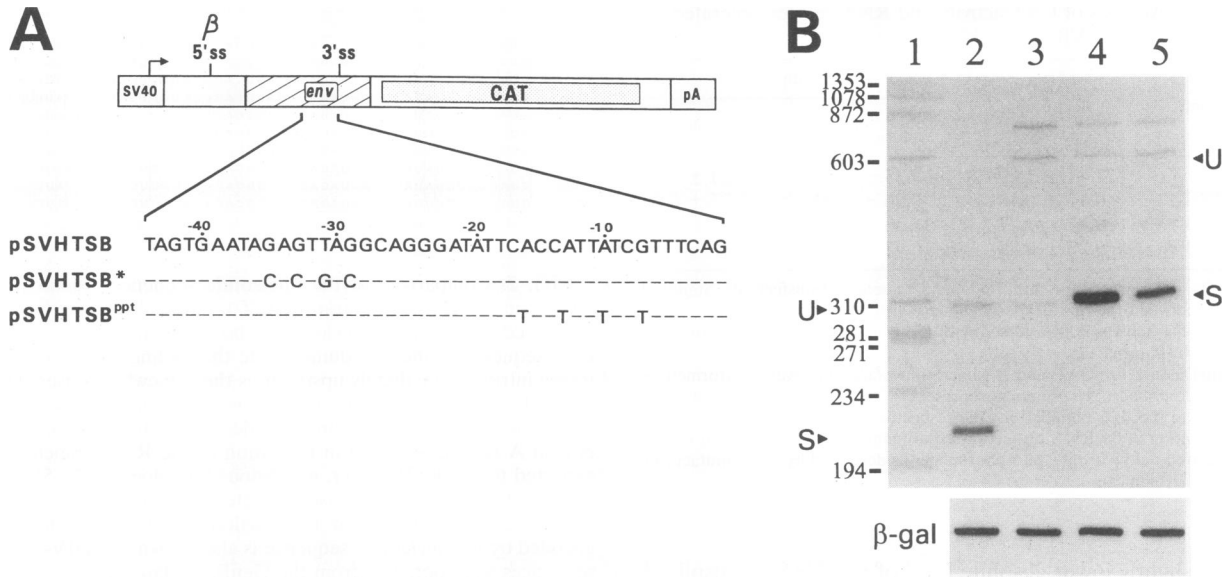


FIG. 3. Analysis of *tat/rev* intron 3′ ss containing either branchpoint or polypyrimidine tract mutations. (A) Schematic representation of the constructs. The pSVHTSB construct (Fig. 1A) was mutagenized by an in vitro PCR-based method as described in Materials and Methods to generate pSVHTSB* or pSVHTSB$^{ppt}$. The pSVHTSB* construct is identical to pSVHTSB except for the four point mutations in the expanded region (dashes denote unaltered bases). The AG dinucleotide that defines the 3′ end of the intron is shown at the extreme 3′ end of the sequence. The three G-to-C mutations and the T-to-G mutation at positions −29, −33, −35, and −31, respectively, relative to the AG create an exact copy of the reported β-globin branchpoint sequence, 5′-CACTGAC-3′. Consequently, the A residue at position −30 is the expected site of RNA branch formation in the case of pSVHTSB*. The exact site of RNA branch formation for the authentic *tat/rev* intron 3′ ss in the context of either pSVHTSB or the HIV-1 genome is unknown. To assess whether the polypyrimidine tract could also play a role in determining the extent of splicing of the *tat/rev* intron, the tract was mutagenized as indicated to create a continuous stretch of 17 pyrimidines, generating pSVHTSB$^{ppt}$. β, β-globin. (B) S1 nuclease protection analysis. A 5-μg aliquot of total RNA from COS-7 cells transfected with pSVβ (lane 2), pSVHTSB (lane 3), pSVHTSB* (lane 4), and pSVHTSB$^{ppt}$ (lane 5) was used in the S1 protection reactions. The probe used for RNA analysis of cells transfected with pSVβ is identical to that used in previous experiments (Fig. 1B), the probe used for pSVHTSB is identical to that used previously for pSVHTSB (Fig. 1), and the probes used for pSVHTSB* and pSVHTSB$^{ppt}$ are similar to that used for pSVHTSB except for the point mutations described above. The protection patterns for pSVHTSB, pSVHTSB*, and pSVHTSB$^{ppt}$ are as depicted in Fig. 1C. Unspliced (U) and spliced (S) forms of the RNAs are indicated, those on the left referring to RNA derived from pSVβ and those on the right referring to RNAs derived from pSVHTSB and its derivatives. At the bottom of the figure are the results obtained upon parallel incubation with probe to β-galactosidase (β-gal) mRNA. Numbers on the left indicate molecular sizes (in base pairs).

from diverse transcripts may have similar mechanisms of action.

Subsequent analysis of the region comprising the *tat/rev* intron 3′ ss revealed a striking conservation of sequences between 20 and 39 nt 5′ of the AG dinucleotide of the 3′ ss among several HIV-1 proviruses (Fig. 2). Although this region encodes a portion of the *env* protein, given the high mutation rate of retroviruses and the redundancy of the genetic code, one would expect some variation within the third base of some codons within this region. The observed 100% conservation suggests another function for this region encompassing the putative site of RNA branch formation. Base pairing between the branchpoint sequence on the pre-mRNA and the U2 snRNA is important for binding of the U2 snRNP to the pre-mRNA and subsequent biochemical reactions (32, 48, 49) and may be a critical step in the splicing process (40). Because the *tat/rev* intron 3′ ss does not contain an obvious branchpoint sequence, efficient base pairing with the RNA component of the U2 snRNP may be a rate-limiting step in splicing of the *tat/rev* intron, resulting in inefficient splicing. Consistent with this hypothesis, introduction of an exact copy of the β-globin branchpoint sequence within the highly conserved, putative branchpoint region converted the intrinsically inefficient *tat/rev* intron 3′ ss into a highly efficient splice site. This demonstrates that the inefficiency of the *tat/rev* intron 3′ ss is due, at least in part, to the absence of a strong branchpoint consensus sequence and that the observed sequence conservation in this region of the HIV-1 genome may be due to an evolutionary constraint to maintain a suboptimal branch point, as in RSV (16). Thus, maintenance of a suboptimal branchpoint region may be a common theme among retroviruses. Subsequent examination of the polypyrimidine tract of the *tat/rev* intron 3′ ss also revealed a conservation in the disruption of the tract by purine residues (Fig. 2). Mutations which converted these purines to thymidine resulted in a dramatic increase in the efficiency of splicing (Fig. 3). However, the effect of the polypyrimidine tract mutation was marginally less than that obtained with the introduction of the β-globin branch point (the branchpoint mutant yielded 94% spliced RNA, as opposed to 87% spliced RNA observed with the polypyrimidine tract mutant). The introduction of an efficient branch point would appear to provide a signal that can function despite the suboptimal nature of the polypyrimidine tract. Similarly, improvement of the polypyrimidine tract increases splicing efficiency despite the absence of a consensus branchpoint sequence. The increase in splicing efficiency observed with the pSVHTSB* and pSVHTSB$^{ppt}$ mutants, in addition to ruling out the possibility that heterologous 3′ ss are not functional in pSVβ, demonstrates the absence of any other distinct *cis*-acting negative sequences within this region. Therefore, it appears that a poor branchpoint region and a suboptimal

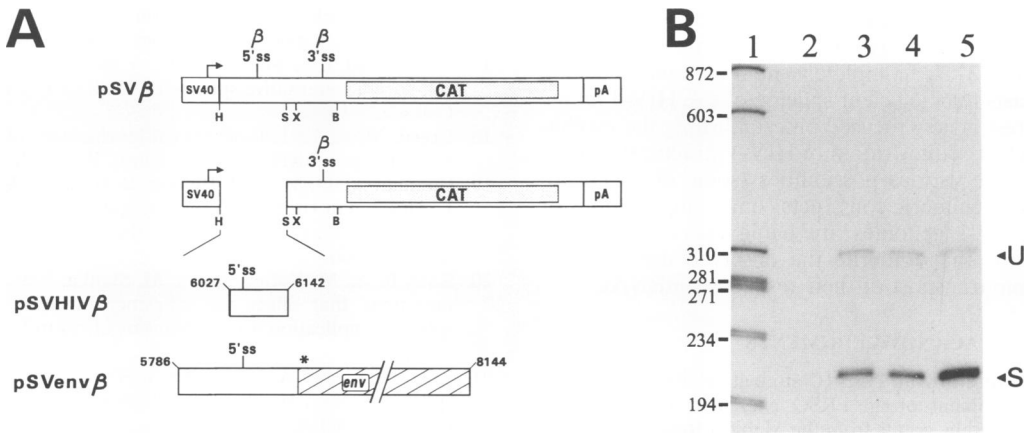


FIG. 4. Analysis of the *tat/rev* intron 5′ ss. (A) Schematic representation of the constructs. pSVHIVβ and pSVenvβ were derived by replacing the β-globin (β) 5′ ss of pSVβ with portions of the HIV-1 genome spanning the *tat/rev* intron 5′ ss. The initiation codon of the *env* open reading frame (hatched box) is indicated by an asterisk. (B) S1 nuclease protection analysis. The end-labelled DNA probe used for analysis of the constructs spanned the 3′ ss and was identical to that used for pSVβ. Lane 1 contains radiolabelled markers. A 5-μg aliquot of total RNA from COS-7 cells transfected with pSVβ (lane 3), pSVHIVβ (lane 4), and pSVenvβ (lane 5) was used in the S1 protection reactions. As a negative control, total RNA from mock-transfected COS-7 cells (lane 2) was incubated with probe and digested with S1 as normal. Bands arising from S1-resistant hybrids corresponding to unspliced (U) and spliced (S) RNA are indicated by arrows. Numbers on the left indicate molecular sizes (in base pairs).

polypyrimidine tract both contribute to the inefficient splicing of introns containing the *tat/rev* intron 3′ ss.

Having identified elements within the *tat/rev* intron 3′ ss that are responsible for the inefficient splicing observed, we then focused attention on other elements of the *tat/rev* intron. In particular, it was of interest to examine the efficiency of the *tat/rev* intron 5′ ss and to test for the presence of negative, *cis*-acting intron elements analogous to the NRS sequences of RSV (1, 27, 28). Upon examination of the HIV-1 *tat/rev* intron 5′ ss in a heterologous context, no such element was detected. Splicing of chimeric introns containing the *tat/rev* intron 5′ ss with either minimal flanking sequences or the 5′ ss in the context of most of the *tat/rev* intron was at least as efficient as that of the β-globin intron (Fig. 4). Moreover, there is no evidence for the existence elsewhere in the *tat/rev* intron of *cis*-acting negative regulatory elements that modulate splicing efficiency. From the data regarding the *tat/rev* intron 5′ ss and 3′ ss and their effects on splicing efficiency, it can be concluded that the only basis for inefficient splicing of the *tat/rev* intron is the inefficient signals present within the 3′ ss. While it appears that HIV-1 and RSV share some requirements for inefficient splicing (i.e., a suboptimal branchpoint region), their respective unspliced RNAs differ dramatically in their capacity to be transported to the cytoplasm (8, 43, 46).

Different mechanisms for the achievement of suboptimal splicing in simple retroviruses (e.g., RSV) and HIV-1 could explain the observed differences in the metabolism of their respective mRNAs. Both spliced and unspliced RSV mRNAs are constitutively transported from the nucleus to the cytoplasm (43). However, in the case of HIV-1, RNAs containing an intact *tat/rev* intron are sequestered within the nucleus in the absence of the viral protein, Rev (12, 13, 19, 25). Experiments using subgenomic constructs of HIV-1 have demonstrated that the *env* region contains all the signals necessary for nuclear sequestration of viral mRNA observed in the absence of Rev (12, 13, 19, 25). There are currently two theories to explain how these HIV-1 RNAs are retained in the nucleus: (i) partial recognition of splice sites by components of the spliceosome, resulting in the nuclear entrapment of pre-mRNA within spliceosome complexes (4, 5, 23), and (ii) the presence of distinct, *cis*-acting repressive sequence (CRS) elements (7, 24, 33, 36, 38). If sequestration of the HIV-1 RNA is the result of entrapment into spliceosome complexes because of the inefficiency of the splicing signals, then there must exist differences between HIV-1 and RSV that permit unspliced RNA of the latter to avoid the entrapping process. The observation that splicing inefficiency in the RSV system is also partly due to an inefficient branch point (16) suggests the presence of either positive RNA export signals in RSV or negative signals in HIV-1. To date, several distinct CRS-like elements have been identified throughout the HIV-1 transcript (7, 24, 33, 36, 38). Furthermore, the absence of an NRS-like element within the *env* region of HIV-1 may also explain the observed difference in RNA metabolism. The NRS of RSV may function to downregulate recognition of the 5′ ss. Failure to recognize the 5′ ss may prevent RSV primary transcripts from becoming committed to the splicing pathway, making them available for transport to the cytoplasm. Support for such a model does exist, since inactivation of the 5′ ss and 3′ ss of an intron has been shown to result in transport of unspliced mRNA into the cytoplasm (4, 5, 22). In the case of the HIV-1 *tat/rev* intron, the results presented provide strong circumstantial evidence that both the branch point and the polypyrimidine tract of the 3′ ss interact poorly with the essential splicing components, U2 snRNP and U2AF, respectively. Poor recognition of the polypyrimidine tract by U2AF implies that the recently described, ATP-independent early (E) splicing commitment complex (29) composed of the pre-mRNA, U1 snRNP, and U2AF may not form efficiently with the *tat/rev* intron of HIV-1. This implication is inconsistent with the splice-site-mediated model for nuclear retention of HIV-1 structural mRNA. Instead, on the basis of the data outlined in this paper, it is tempting to suggest that sequences distinct from the HIV-1 *tat/rev* splicing signals play a role in mediating nuclear retention of structural mRNA. Indeed, recent results indicate that the signals for suboptimal splicing and nuclear retention can be segregated in the case of HIV-1 (42a).

In summary, the data presented in this paper indicate that the balanced splicing of the HIV-1 *tat/rev* intron is due solely to the action of signals within the 3′ ss and the neighboring exon. There is evidence that a purine-rich sequence within the 3′ exon acts to increase recognition of the *tat/rev* intron 3′ ss

despite the suboptimal nature of both the branch point and the polypyrimidine tract; deletion of the purine-rich region renders the *tat/rev* intron 3′ ss completely nonfunctional. Having identified the basis for inefficient splicing of the HIV-1 *tat/rev* intron, our interest is now focused on establishing the mechanism for the nuclear sequestration of HIV-1 structural protein mRNA. To test the various possibilities (some of which are mentioned above), chimeric constructs comprising portions of the RSV and HIV-1 genomes are being tested to identify sequence elements that underlie the observed difference in nucleocytoplasmic transport of their respective mRNAs.

## ACKNOWLEDGMENTS

A. Staffa is supported by an NSERC student scholarship, and A. Cochrane is the recipient of an FRSQ chercheur boursier. This research was supported by grants from the Medical Research Council of Canada.

We thank P. Harakidas and the other members of our laboratory for their help and support throughout the course of this work, P. Liston and D. J. Briedis for kindly providing access to HIV sequences from GenBank, and N. H. Acheson for his advice and suggestions.

## REFERENCES

1. **Arrigo, S., and K. Beemon.** 1988. Regulation of Rous sarcoma virus RNA splicing and stability. Mol. Cell. Biol. **8**:4858–4867.
2. **Arya, S. K., C. Guo, S. F. Josephs, and F. Wong-Staal.** 1985. Trans-activator gene of human T-lymphotrophic virus type III (HTLV-III). Science **229**:69–73.
3. **Berberich, S. L., and C. M. Stoltzfus.** 1991. Mutations in the regions of the Rous sarcoma virus 3′ splice sites: implications for regulation of alternative splicing. J. Virol. **65**:2640–2646.
4. **Chang, D. D., and P. A. Sharp.** 1989. Regulation by HIV Rev depends upon recognition of splice sites. Cell **59**:789–795.
5. **Chang, D. D., and P. A. Sharp.** 1990. Messenger RNA transport and HIV *rev* regulation. Science **249**:614–615.
6. **Chomczynski, P., and N. Sacchi.** 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. **162**:156–159.

6a. **Cochrane, A.** Unpublished data.

7. **Cochrane, A. W., K. S. Jones, S. Beidas, P. J. Dillon, A. M. Skalka, and C. A. Rosen.** 1991. Identification and characterization of intragenic sequences which repress human immunodeficiency virus structural gene expression. J. Virol. **65**:5305–5313.
8. **Coffin, J. M.** 1990. Retroviridae and their replication, p. 1437–1500. *In* B. N. Fields and D. M. Knipe (ed.), Virology. Raven Press, New York.
9. **Cullen, B. R.** 1988. Use of eukaryotic expression technology in the functional analysis of cloned genes. Methods Enzymol. **152**:684–704.
10. **Cullen, B. R.** 1991. Human immunodeficiency virus as a prototypic complex retrovirus. J. Virol. **65**:1053–1056.
11. **Dayton, A. I., J. G. Sodroski, C. A. Rosen, W. C. Goh, and W. A. Haseltine.** 1986. The transactivator gene of the human T-cell lymphotrophic virus type III is required for replication. Cell **44**:941–947.
12. **Emerman, M., R. Vazeux, and K. Peden.** 1989. The *rev* gene product of the human immunodeficiency virus affects envelope-specific RNA localization. Cell **57**:1155–1165.
13. **Felber, B. K., M. Hadzopoulou-Cladaras, C. Cladaras, T. Copeland, and G. N. Pavlakis.** 1989. The *rev* protein of HIV-1 affects the stability and transport of the viral mRNA. Proc. Natl. Acad. Sci. USA **86**:1495–1499.
14. **Felber, B. K., and G. N. Pavlakis.** 1993. Molecular biology of HIV-1: positive and negative regulatory elements important for virus expression. AIDS **7**:S51–S62.
15. **Fisher, A. G., M. B. Feinberg, S. F. Josephs, M. E. Harper, L. M. Marselle, G. Reyes, M. A. Gonda, A. Aldovini, C. Debouk, R. C. Gallo, and F. Wong-Staal.** 1986. The transactivator gene of HTLV-III is essential for virus replication. Nature (London) **320**:367–371.
16. **Fu, X.-D., R. A. Katz, A. M. Skalka, and T. Maniatis.** 1991. The role of branchpoint and 3′-exon sequences in the control of balanced splicing of avian retrovirus RNA. Genes Dev. **5**:211–220.
17. **Ge, H., and J. L. Manley.** 1990. A protein factor, ASF, controls cell-specific alternative splicing of SV40 early pre-mRNA in vitro. Cell **62**:25–34.
18. **Green, M. R.** 1991. Biochemical mechanisms of constitutive and regulated pre-mRNA splicing. Annu. Rev. Cell Biol. **7**:559–599.
19. **Hammarskjöld, M.-L., J. Heimer, B. Hammarskjöld, I. Sangwan, L. Albert, and D. Rekosh.** 1989. Regulation of human immunodeficiency virus *env* expression by the *rev* gene product. J. Virol. **63**:1959–1966.
20. **Katz, R. A., M. Kotler, and A. M. Skalka.** 1988. *cis*-acting intron mutations that affect the efficiency of avian retroviral RNA splicing: implication for mechanisms of control. J. Virol. **62**:2686–2695.
21. **Katz, R. A., and A. M. Skalka.** 1990. Control of retroviral RNA splicing through maintenance of suboptimal splicing signals. Mol. Cell. Biol. **10**:696–704.
22. **Legrain, P., and M. Rosbash.** 1989. Some *cis*- and *trans*-acting mutants for splicing target pre-mRNA to the cytoplasm. Cell **57**:573–583.
23. **Lu, X., J. Heimer, D. Rekosh, and M.-L. Hammarskjöld.** 1990. U1 small nuclear RNA plays a direct role in the formation of a *rev*-regulated human immunodeficiency virus *env* mRNA that remains unspliced. Proc. Natl. Acad. Sci. USA **87**:7598–7602.
24. **Maldarelli, F., M. A. Martin, and K. Strebel.** 1991. Identification of posttranscriptionally active inhibitory sequences in human immunodeficiency virus type 1 RNA: novel level of gene regulation. J. Virol. **65**:5732–5743.
25. **Malim, M. H., J. Hauber, S.-Y. Le, J. V. Maizel, and B. R. Cullen.** 1989. The HIV-1 *rev trans*-activator acts through a structured target sequence to activate nuclear export of unspliced viral mRNA. Nature (London) **338**:254–257.
26. **Maniatis, T.** 1991. Mechanisms of alternative pre-mRNA splicing. Science **251**:33–34.
27. **McNally, M. T., and K. Beemon.** 1992. Intronic sequences and 3′ splice sites control Rous sarcoma virus RNA splicing. J. Virol. **66**:6–11.
28. **McNally, M. T., R. R. Gontarek, and K. Beemon.** 1991. Characterization of Rous sarcoma virus intronic sequences that negatively regulate splicing. Virology **185**:99–108.
29. **Michaud, S., and R. Reed.** 1993. A functional association between the 5′ and 3′ splice sites is established in the earliest prespliceosome complex (E) in mammals. Genes Dev. **7**:1008–1020.
30. **Moore, M. J., C. C. Query, and P. A. Sharp.** 1993. Splicing of precursors to mRNA by the spliceosome, p. 303–357. *In* R. F. Gesteland and J. F. Atkins (ed.), The RNA world. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
31. **Muesing, M. A., D. H. Smith, and D. J. Capon.** 1987. Regulation of mRNA accumulation by a human immunodeficiency virus *trans*-activator protein. Cell **48**:691–701.
32. **Nelson, K. K., and M. R. Green.** 1989. Mammalian U2 snRNP has a sequence-specific RNA-binding activity. Genes Dev. **3**:1562–1571.
33. **Rosen, C. A., E. Terwilliger, A. Dayton, J. G. Sodroski, and W. A. Haseltine.** 1988. Intragenic *cis*-acting *art* gene-responsive sequences of the human immunodeficiency virus. Proc. Natl. Acad. Sci. USA **85**:2071–2075.
34. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
35. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
36. **Schwartz, S., M. Campbell, G. Nasioulas, J. Harrison, B. Felber, and G. Pavlakis.** 1992. Mutational inactivation of an inhibitory sequence in human immunodeficiency virus type 1 results in Rev-independent *gag* expression. J. Virol. **66**:7176–7182.
37. **Schwartz, S., B. K. Felber, D. M. Benko, E.-M. Fenyo, and G. N. Pavlakis.** 1990. Cloning and functional analysis of multiply spliced mRNA species of human immunodeficiency virus type 1. J. Virol. **64**:2519–2529.
38. **Schwartz, S., B. K. Felber, and G. N. Pavlakis.** 1992. Distinct RNA

sequences in the *gag* region of human immunodeficiency virus type 1 decrease RNA stability and inhibit expression in the absence of Rev protein. J. Virol. **66:**150–159.

39. **Smith, C. W. J., J. G. Patton, and B. Nadal-Ginard.** 1989. Alternative splicing in the control of gene expression. Annu. Rev. Genet. **23:**527–577.
40. **Smith, C. W. J., E. B. Porro, J. G. Patton, and B. Nadal-Ginard.** 1989. Scanning from an independently specified branch point defines the 3′ splice site of mammalian introns. Nature (London) **342:**243–247.
41. **Sodroski, J., W. C. Goh, C. Rosen, A. Dayton, E. Terwilliger, and W. A. Haseltine.** 1986. A second post-transcriptional transactivator gene required for the HTLV-III replication. Nature (London) **321:**412–417.
42. **Sodroski, J., C. A. Rosen, and W. A. Haseltine.** 1984. *Trans*-acting transcriptional activation of the long terminal repeat of human T lymphotropic viruses in infected cells. Science **225:**381–385.

42a. **Staffa, A., and A. Cochrane.** Unpublished data.

43. **Stoltzfus, C. M.** 1988. Synthesis and processing of avian sarcoma retrovirus RNA. Adv. Virus Res. **35:**1–38.
44. **Terwilliger, E., R. Burghoff, R. Sia, J. Sodroski, W. Haseltine, and C. Rosen.** 1988. The *art* gene product of the human immunodeficiency virus is required for replication. J. Virol. **62:**655–658.
45. **Watakabe, A., K. Tanaka, and Y. Shimura.** 1993. The role of exon sequences in splice site selection. Genes Dev. **7:**407–418.
46. **Wong-Staal, F.** 1990. Human immunodeficiency viruses and their replication, p. 1529–1543. *In* B. N. Fields and D. M. Knipe (ed.), Fields virology. Raven Press, New York.
47. **Wright, C. M., B. K. Felber, H. Paskalis, and G. N. Pavlakis.** 1986. Expression and characterization of the trans-activator of HTLV-III/LAV virus. Science **234:**988–992.
48. **Wu, J., and J. L. Manley.** 1989. Mammalian pre-mRNA branch site selection by U2 snRNP involves base pairing. Genes Dev. **3:**1553–1561.
49. **Zhuang, Y., and A. M. Weiner.** 1989. A compensatory base change in human U2 snRNA can suppress a branch site mutation. Genes Dev. **3:**1545–1552.