MOLECULAR AND CELLULAR BIOLOGY, Feb. 1990, p. 696–704
0270-7306/90/020696-09$02.00/0

Vol. 10, No. 2

Control of Retroviral RNA Splicing through Maintenance of Suboptimal Processing Signals

RICHARD A. KATZ AND ANNA MARIE SKALKA*

Fox Chase Cancer Center, Institute for Cancer Research, Philadelphia, Pennsylvania 19111

Received 24 June 1989/Accepted 6 November 1989

The full-length retroviral transcript serves as genomic RNA for progeny virions, as an mRNA for structural proteins and enzymes, and as a pre-mRNA substrate for splicing that yields subgenomic mRNAs that encode other essential proteins. Thus, RNA splicing to form subgenomic mRNAs must be incomplete or regulated in order to preserve some of the full-length transcripts. We have used the avian sarcoma virus system to delineate the viral functions that are required in the regulation of the splicing event that forms the envelope glycoprotein (*env*) subgenomic mRNA. We observed previously that a specific insertion mutation just 5′ of the *env* splice acceptor site resulted in nearly complete splicing to form *env* mRNA and a concomitant replication defect which is presumably due to a deficit of the full-length transcript. Replication-competent pseudorevertants contained second-site mutations that restored splicing control, and these mapped either just upstream or downstream of the *env* splice acceptor site. In this report, we show that splicing control at this site does not require expression of any known viral replication protein(s), nor does it appear to require the viral splice donor site. From these results and analysis of additional splicing mutations obtained by in vivo selection, we conclude that splicing is controlled through the maintenance of suboptimal *cis*-acting signals in the viral RNA that alter the efficiency of recognition by the cellular splicing machinery.

Pre-mRNA splicing in eucaryotes involves removal of intron sequences from RNA transcripts by specific breaking and joining at splice donor and acceptor sites (for reviews, see references 10, 21, and 28). In most cases, neighboring donor and acceptor sites function as exclusive pairs; introns between neighboring splice sites are efficiently and constitutively removed even in the presence of other distant splice donor and acceptor sites. In other cases, a single splice donor site can be joined to one or another of several distant splice acceptor sites (or vice versa) by alternative splicing.

Many animal viruses use alternative splicing to generate multiple mRNAs from a single large precursor by the splicing of a common leader RNA to multiple splice acceptor sites (reviewed in reference 10). This strategy is not only economical but is also required to bring internal reading frames to a 5′ position from which they can be efficiently translated. In cellular (2) and viral RNAs (10), the relative efficiency of alternative splicing between different pairs of donors and acceptors can be used to regulate the amounts of different mRNAs that are produced from a common precursor or to express combinatorial patterns of coding exons that give rise to variant proteins. Even when splicing is regulated, at least one intron is efficiently removed in the nucleus, and significant amounts of completely unspliced pre-mRNA are not usually found in the cytoplasm. This is not the situation for retroviruses, whose life cycle requires that significant amounts of RNA remain unspliced to perform functions in the cytoplasm.

Retroviruses are enveloped, single-stranded RNA viruses that replicate through a DNA intermediate that is integrated into the host DNA to form a provirus (35). The full-length RNA transcript produced from the integrated provirus encodes precursor proteins for *gag* gene products (structural proteins of the virion core) and *pol* gene products (enzymatic components of the virion core; e.g., reverse transcriptase and integration protein). The full-length RNA also serves as the viral genetic material and is assembled into progeny virions at the cell membrane. A portion of the full-length RNA is also spliced to form *env* mRNA, which encodes a precursor for the virion membrane protein complex which is required for attachment to host cell receptors. Thus, the full-length RNA serves three essential functions: as genomic RNA, as mRNA, and as pre-mRNA for *env* and other subgenomic mRNAs. Therefore, retroviral splicing must be limited or controlled to preserve sufficient amounts of the full-length RNA to perform the first two functions.

There are no obvious features of retroviral gene expression that could account for limited splicing. The full-length retroviral RNA resembles a cellular transcript; it is synthesized by RNA polymerase II from a chromosomal location and is capped and polyadenylated in the nucleus. Mechanisms which might be involved in the preservation of unspliced retroviral RNA have been proposed (3, 31). For example, (i) a portion of the unspliced RNA could be preferentially transported from the nucleus or compartmentalized in order to circumvent the splicing pathway, (ii) a viral protein could serve as a positive or negative regulator of splicing, or (iii) inefficient splicing could be dictated by intrinsic structural or sequence features of the splice sites or RNA molecule. These mechanisms are not mutually exclusive, and control could be exerted through a combination of them.

We have used the avian sarcoma virus (ASV) to determine the relative importance of some of the mechanisms listed above. In our experiments, we focused on the splicing event that is common to all replication-competent retroviruses, in which the primary, full-length transcript (ca. 9.3 kilobases for ASV) undergoes limited splicing to form *env* mRNA (ca. 4.7 kilobases) (Fig. 1). In the ASV strain used in these studies, the steady-state *env* mRNA/unspliced RNA ratio is ca. 0.5 (12). *env* mRNA encodes a protein precursor, $Pr95^{env}$, that contains a signal peptide for membrane insertion and is processed in the rough endoplasmic reticulum and Golgi apparatus to form the surface-transmembrane glyco-

* Corresponding author.

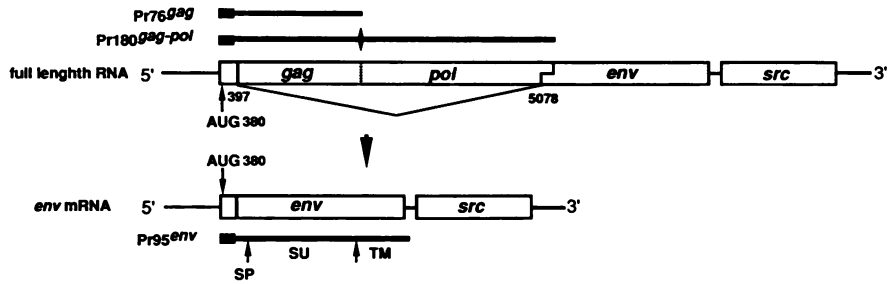

FIG. 1. Maps of ASV full-length and *env* mRNAs. Diagrammatic representations of translation products of full-length ($Pr76^{gag}$, $Pr180^{gag\text{-}pol}$) and *env* mRNA ($Pr95^{env}$) are shown above and below the RNA maps. The *gag-pol* precursor is synthesized by a translational frameshift (|). The RNA splice event that forms *env* mRNA is also indicated (▼). Genes are depicted as boxes, and noncoding regions are depicted as lines. The common AUG initiation codon for *gag*, *gag-pol*, and *env* (position 380); the splice donor site (position 397); and the *env* splice acceptor site (position 5078) are indicated. Numbering of nucleotides follows the convention of Schwartz et al. (27). The *env* leader peptide is released by a signal peptidase (SP). The surface (SU) and transmembrane (TM) components of the *env* glycoprotein complex are produced by proteolytic processing at the position of the vertical arrows. The 6-amino-acid portion of the leader peptide that is shared between the *gag* and *env* precursors is shown as a filled box (5). A second splice event that forms *src* mRNA is not shown. The drawing is not to scale.

protein complex (37). The *env* message is formed by a single splice event in which a 397-nucleotide (nt) leader is joined to the *env* splice acceptor site at position 5078 (Fig. 1). The leader RNA supplies the translational initiation codon (at position 380) and the first five codons for $Pr95^{env}$; thus, the *env* gene is split into two coding exons (5). The ca. 4.6-kilobase intron that is removed includes the *gag* and *pol* genes that are translated from the full-length RNA using the same initiation codon as $Pr95^{env}$ (position 380, Fig. 1).

Previously, we reported that insertion of a specific 24-nucleotide sequence near the ASV *env* splice acceptor site produced a replication-defective virus (mutant pLD6/I) (12). The defect was apparently caused by a splicing imbalance which, surprisingly, heavily favored *env* mRNA. From this original mutant, we isolated a series of replication-competent pseudorevertant viruses, one of which was extensively characterized. In this new mutant, splicing was down regulated to a level that was lower than that observed in the wild type (wt). This phenotypic change was attributed to a single nucleotide substitution (suppressor mutation) at position −20 from the splice acceptor site (12). Thus, passage in vivo selected for a change in splicing efficiency which was accomplished by an alteration in the *env* splice acceptor region. This result suggested that the splice acceptor region can play a significant role in the maintenance of an appropriate splicing balance for retroviruses.

In this report, we present the results of our characterization of other pseudorevertants and an evaluation of the role of viral proteins in ASV splicing regulation. Our results showed that (i) sequences on both sides of the *env* splice acceptor site play a role in determining splicing efficiency, (ii) the *cis*-acting mutations in this region exert characteristic effects on splicing even when RNAs are expressed in heterologous (mammalian) cells, (iii) wt and mutant splicing phenotypes are maintained in the absence of *gag*, *pol*, and *env* protein expression, and (iv) the ASV splice donor site is not required for regulated splicing. These results suggest that the evolutionary maintenance of suboptimal *cis*-acting splicing signals may be the primary mechanism whereby the appropriate balance of spliced and unspliced retroviral RNAs is established in infected cells. In addition, these results demonstrate that biological selection for retroviral splicing mutations is a useful method for identifying specific *cis*-acting sequences that are recognized by the cellular splicing machinery. Moreover, this system allows for selection of informative mutations without a major bias as to the functional importance of any region near the splice site and thus can be used to identify functional regions that might otherwise go undetected.

MATERIALS AND METHODS

Analysis of the *env* splice acceptor regions of pseudorevertants by using the polymerase chain reaction (PCR). Chicken embryo fibroblasts (CEFs) were infected with revertant viruses (12), and low-molecular-weight DNA fractions (11) were prepared approximately 24 h postinfection. A sample of DNA was incubated with primers containing sequences (5'TTGGGTACCCTCTCGAAA3' and 5'CTGCTCGAGTA AGTGAAC3') homologous to sections that flank the *env* splice acceptor region. The primers included a *Kpn*I site (GGTACC) and an *Xho*I site (CTCGAG) (underlined above) located at positions 5000 and 5259 of the ASV genome, respectively (27), flanking the *env* splice acceptor site (5077/5078). The DNA segment was amplified by using *Taq* polymerase, as described by the supplier (Perkin-Elmer Cetus). The amplified DNA was digested with *Kpn*I and *Xho*I restriction endonucleases, and the resultant fragment was cloned in the pMG4 vector (14) by using standard methods (17). The inserts were sequenced by the chemical method (19). For genetic analysis of the mutations contained in the amplified regions, fragments obtained after digestion of pMG4 derivatives with restriction endonucleases, *Kpn*I and either *Xho*I or *Hpa*I (position 5187), were substituted for the analogous fragments in the wt ASV clone, pLD6 (see Fig. 3).

Transfection and RNA analysis. Transfection of CEFs or the QCl-3 quail cell line was carried out as described previously (14). Total RNA was prepared from transiently transfected QCl-3 cells (harvested 60 h posttransfection) or infected CEFs. COS monkey cells were transfected as described previously (4), and RNAs were prepared from cells at 60 h posttransfection. The ratio of unspliced mRNA to *env* mRNA was determined by using an S1 nuclease protection assay (12, 14) (see Fig. 5B). The probes were 5' labeled at an *Xho*I site downstream of the *env* splice acceptor site by using [γ-^{32}P]ATP and polynucleotide kinase. Hybridization of the probe to spliced *env* RNA protects ca. 180 nt, while unspliced RNA protects ca. 280 nt. In all cases, probes contained the same mutations as the viral RNA being assayed (see Fig. 5).

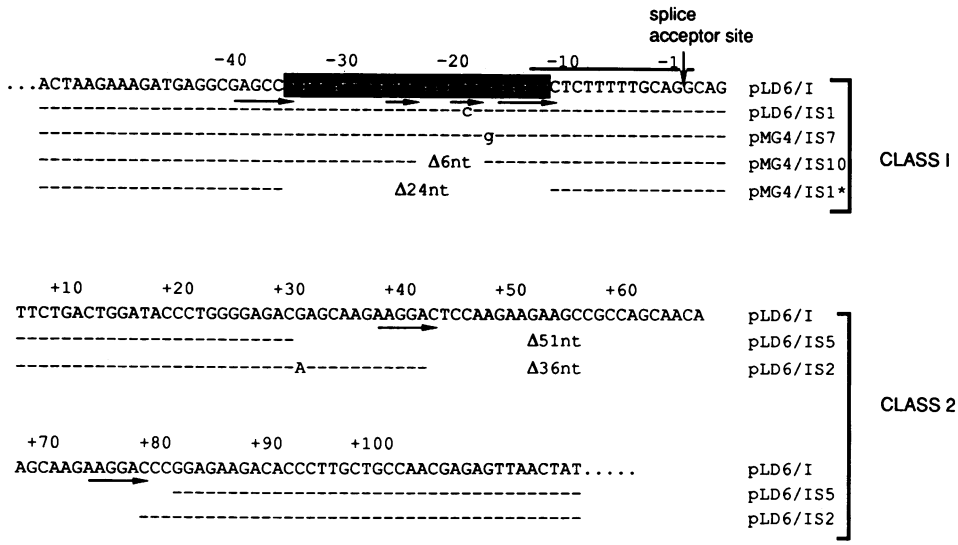

FIG. 2. Nucleotide sequences of *env* splice acceptor regions from mutant and pseudorevertants. *env* splice acceptor regions from pseudorevertants were amplified by PCR as ca. 255- to 280-bp fragments and were cloned into pMG4 (14). The sequence of the *env* splice acceptor region of the pLD6/I splicing mutant is shown at the top. This viral clone differs from the wt (pLD6) by the 24-bp linker insertion identified by lowercase letters in the filled box. The pseudorevertant sequences are shown below; dashes indicate identity with pLD6/I. In the cases of pLD6/IS2 and pLD6/IS5, the entire *Kpn*I-*Xho*I cassette (Fig. 3) was sequenced and only the regions containing differences are shown. The pLD6/ prefix indicates that viral derivatives were constructed for genetic analysis (see Fig. 3). The pMG4 derivatives were not reassembled into viral DNA forms, and the entire sequences of the *Kpn*I-*Xho*I fragments were not determined. Arrows indicate direct repeats that may facilitate deletion events during reverse transcription (20). pMG4/IS1* was identical to the wt (pLD6) in this region, indicating that the virus from which the segment was amplified was likely a true revertant. The splice acceptor consensus sequence (21) is overlined. Numbering is relative to the two G residues that constitute the intron-exon border. The G in the intron (position 5077) is designated –1 and the G in the exon (position 5078) is designated +1.

Construction of intron deletion and splice site substitution clones. pLD6, pLD6/I, and pLD6/IS1 viral DNA clones (12) were digested with *Sst*II and *Kpn*I restriction endonucleases. Removal of this fragment results in deletion of *gag* and *pol* sequences from ca. positions 547 to 5000 (27) (see Fig. 6B). The DNAs were treated with *Escherichia coli* polymerase I to remove protruding 3′ ends and then were ligated to recircularize the plasmids. The resulting deleted intron consisted of a splice donor site at position 397/398 and intron sequences through to position 547 joined to intron sequences from position 5000 to the *env* splice acceptor site (position 5077/5078). A DNA segment which included this deletion was transferred to an *env* expression vector that contained a simian virus 40 origin of replication (see Fig. 6B).

pCIN was constructed as follows. The expression vector pBC12BI (4) was cleaved with *Nhe*I and repaired with Klenow fragment of DNA polymerase I. The DNA was then cleaved with *Bst*EII to remove the bulk of rat preproinsulin-coding sequences. *Nhe*I cleaves within intron 1, and *Bst*EII cleaves within exon 3. An *Asp*I-*Bst*EII fragment of the ASV genome (from pLD6) spanning positions 5000 through the 3′ long terminal repeat (LTR) was inserted after repair of the *Asp*I site. Ligation of the repaired *Asp*I site to the repaired *Nhe*I site regenerated the *Asp*I site. The *env* splice acceptor site mutations were substituted into pCIN by using the regenerated *Asp*I(*Kpn*I) site and *Xho*I site flanking the *env* splice acceptor region (see Fig. 3). The *Kpn*I-*Xho*I fragments derived from pLD6I and pLD6IS1 were used to construct pCIN/I and pCIN/IS1, respectively.

RESULTS

Cloning and nucleotide sequences of pseudorevertant viruses containing putative splicing mutations. The mutation in pLD6/I consists of a 24-base-pair (bp) linker that was inserted at position −12 from the *env* splice acceptor site in pLD6, an infectious DNA clone of SR-B ASV (15) (Fig. 2). Normally, the *env* mRNA/full-length RNA ratio is ca. 0.5. The insertion produces a 10-fold increase in the ratio of *env* mRNA to unspliced RNA, and we have shown that the replication defect is likely caused by a deficit of full-length RNA. However, this defect is probably not absolute, since some replication must be required for the spontaneous revertants to arise.

One pseudorevertant, pLD6/IS1, was analyzed previously (12). The polymerase chain reaction (PCR) has since been used to amplify and clone the *env* splice acceptor site regions from a number of additional revertant viruses. Culture supernatants that contained a predominant species of revertant virus (12) were used to infect fresh cultures of CEFs. After 24 h, a fraction enriched in the unintegrated viral DNA was used as a template to amplify a region of ca. 283 bp containing the *env* splice acceptor site (Fig. 3). The ends of the PCR primers included a *Kpn*I site and an *Xho*I site which flanked the splice acceptor site. The PCR reaction produced DNA fragments from each of the revertant viruses; each viral DNA gave rise to predominantly one size of fragment, but the collection of revertants yielded fragments which varied in size from ca. 225 to 280 bp (data not shown). After digestion with *Kpn*I and *Xho*I, the amplified DNA was cloned in a plasmid vector and a partial or complete nucleotide sequence of the insert was obtained.

As expected from RNA-mapping studies of these revertants (12), all of the amplified segments showed sequence differences from the original mutant pLD6/I (Fig. 2). These putative suppressor mutations fell into two classes. Class 1 viruses had alterations within the original insert, including an A-to-G change at position −18 and a 6-nt deletion at posi-

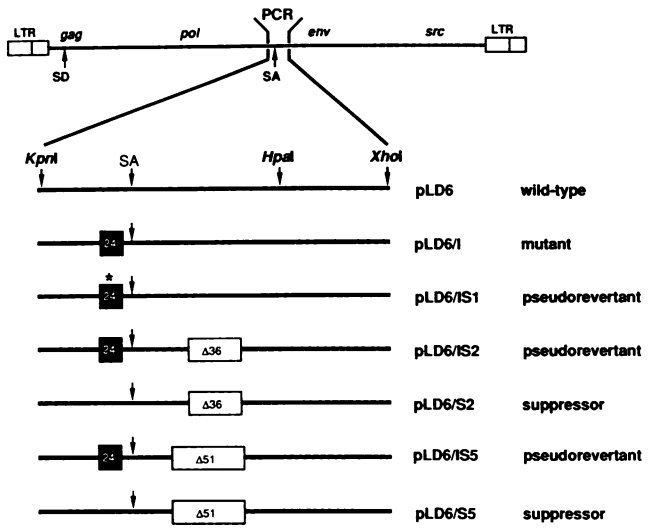

FIG. 3. Maps of reconstructed pseudorevertants in the region of the *env* splice acceptor site. The *env* splice acceptor region from pseudorevertant viral DNA was amplified by PCR (region is indicated). *Kpn*I-*Xho*I cassettes of ca. 225 to 280 bp that included this region (shown as lines) were used to create new viruses that contained defined mutations. In some cases, a *Hpa*I site was used. Symbols: ■, original 24-bp insertion mutation in pLD6/I (Fig. 2); □, deletions of 36 or 51 nt present in the respective pseudorevertant viruses; ⋆, T→C at position −20 in pLD6/IS1 (Fig. 2). In pLD6/S2 and pLD6/S5, the insertion mutation was removed precisely by cleavage with *Ban*II (12) followed by religation, leaving only the suppressor mutation.

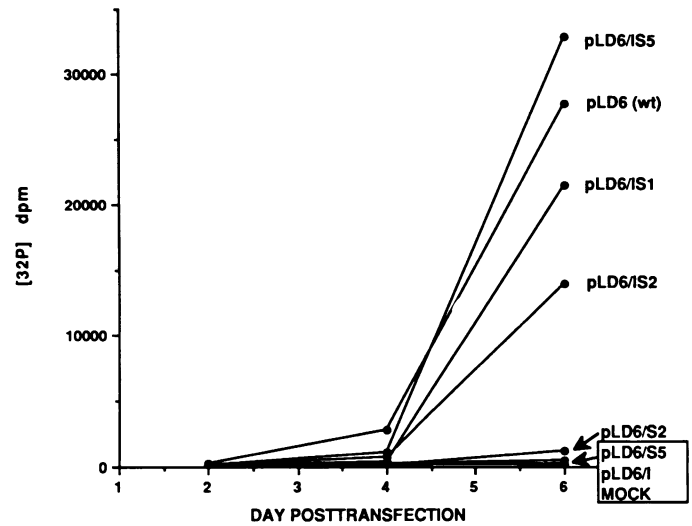

FIG. 4. Replication assay for wt and mutant viruses. CEF cultures were transfected with comparable amounts of the indicated viral DNA clones (Fig. 3). Culture supernatants were assayed for virion-associated reverse transcriptase activity by using a poly(C)-oligo(dG) template/primer system. Incorporation of [α-^{32}P]dGTP into DNA was measured by retention on DEAE paper as described previously (9). Spots were cut out and counted.

tions −19 to −24 from the splice acceptor site. One member of this class (pMG4/IS1*) was apparently a true revertant; the original insert was lost precisely. Class 2 viruses had deletions of 36 or 51 nt starting at positions +42 and +30 downstream of the splice acceptor site (exon sequences), respectively. These deletions occurred within the region coding for the 64-amino-acid *env* leader peptide (Fig. 1) but created in-frame interruptions and therefore should allow *env* translation. Deletions of classes 1 and 2 were frequently flanked by direct repeats (noted in Fig. 2), suggesting that they were produced by slippage during reverse transcription (20). The virus with the 36-nt deletion also contained an A-to-G change at position +30. From these data we conclude that these putative suppressor-type mutations may exert their effects at two sites: within the original linker mutation and immediately downstream of the splice acceptor site.

The class 2 exon mutations are suppressors. In order to determine whether the class 2 mutations alone accounted for the revertant phenotype, we performed the following experiments. The *env* splice acceptor site regions from the two class 2 revertant viruses were substituted into the wild-type DNA clone as a ca. 280-bp fragment (a *Kpn*I-*Xho*I fragment) to create pLD6/IS2 and pLD6/IS5 (Fig. 3). These two viral DNA clones were identical to the original mutant, pLD6/I, except for the deletions of 36 (plus the A-to-G change at +30) or 51 bp. We reasoned that since the revertants were selected to compensate for excessive splicing, these class 2 deletions may interrupt positive regulatory sequences that are required for normal splicing. In order to determine whether the deletion mutations could act independently of the original insertion mutation, derivatives of these two clones were created by removing the original mutation, the 24-bp linker with *Ban*II (5′GAGCCC3′) ends (12). These two derivatives, pLD6/S2 and pLD6/S5, are identical to the wt (pLD6) except for the class 2 mutations.

To determine the effects of the class 2 mutations on viral replication, susceptible CEFs were transfected with each of the clones described above (Fig. 3), and the amount of virus in culture supernatants was estimated by assay of virion-associated reverse transcriptase (Fig. 4). In these experiments, virus was detected only if significant reinfection occurred because cells that initially take up DNA and subsequently release virus are a small percentage of the population. Thus, the appearance of virus is a reflection of the rate at which multiple rounds of infection occurs. As had been observed previously (12), the mutant (pLD6/I) did not give rise to detectable virus by day 6 posttransfection. The pLD6/IS5 clone produced a virus which replicated like the wt (pLD6); culture supernatants contained a high level of reverse transcriptase activity by day 6 posttransfection (Fig. 4), confirming our hypothesis that the 51-nt deletion is a suppressor mutation. pLD6/IS2, which contained the linker plus the 36-nt deletion, gave rise to a virus that grew at a rate reproducibly lower than that of the wt. Thus, the 36-nt deletion in pLD6/IS2 is also a suppressor mutation, but it can only partially suppress the original mutant phenotype. pLD6/S2 and pLD6/S5, which contain only the deletions and not the original linker mutation, produced viruses which spread slowly and did not reach significant titers until 11 or 20 days after transfection, respectively (Fig. 4; data not shown for later times). Thus, the suppressor mutations each displayed a partially defective phenotype in the absence of the original insertion mutation.

Effect of suppressor mutations on *env* mRNA splicing. The excessive splicing phenotype exhibited by the pLD6/I mutant can be suppressed by a T-to-C change at position −20 from the *env* splice acceptor site (12) (Fig. 2). To determine whether the class 2 suppressor mutations also down regulated splicing, we examined the ratio of unspliced RNA to *env* mRNA in cells that had been infected or transfected with

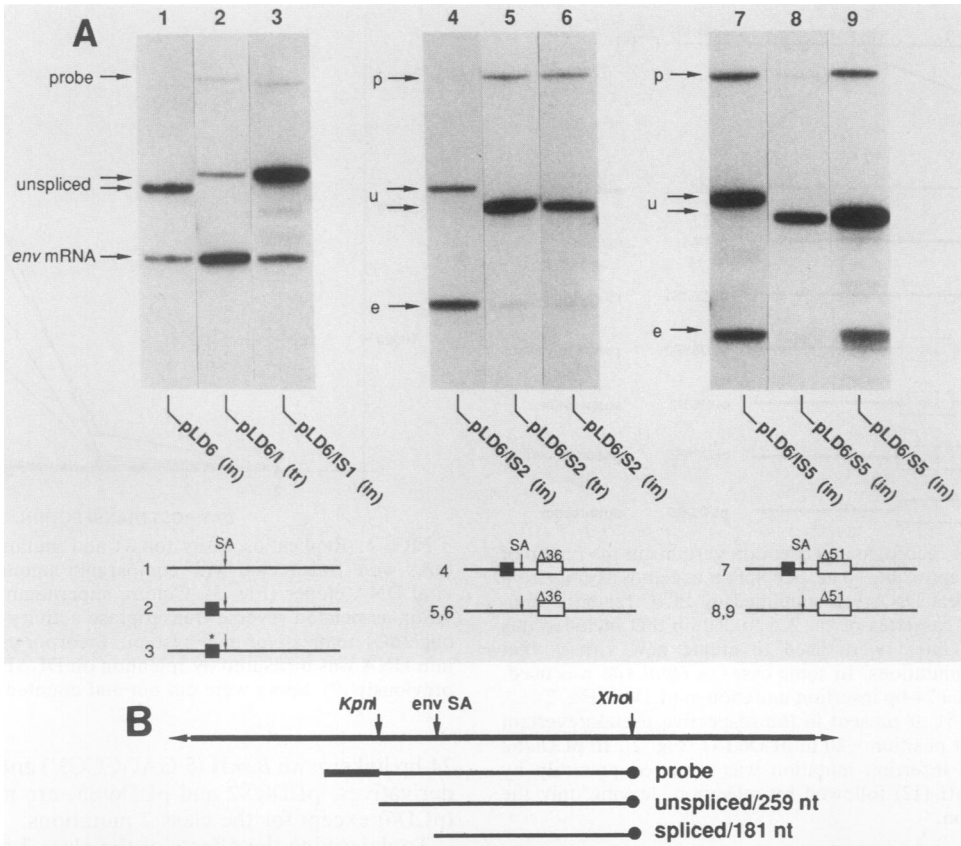

FIG. 5. Determination of the ratio of unspliced RNA to *env* mRNA by S1 nuclease protection assay. (A) Total cell RNA was prepared from QCl-3 quail cells that were transiently transfected (tr) or CEFs that were chronically infected (in) with the indicated viruses or viral DNA clones. Maps below lanes (with corresponding lane numbers) correspond to *env* splice acceptor site segments shown in Fig. 3. Lane 9 is a longer exposure of lane 8. (B) Probe strategy for determining ratio of unspliced RNA to *env* mRNA (12). Circles indicate ^{32}P 5′-end labels; thick line indicates pBR322 sequences that distinguish reannealed probe from probe that is protected by hybridization to unspliced RNA. Lengths are indicated for the wt probe (derived from pLD6). Probes for mutant viral RNAs contained the homologous insertions and/or deletions, as indicated at the bottom of each set in panel A.

the four mutant DNAs or viruses by using S1 analyses with probes that spanned the *env* splice acceptor site (Fig. 5). Two types of experiments were performed. First, a quail cell line, QCl-3, was transfected with the viral DNA clones under conditions in which virus spread does not occur (14), and the transiently expressed viral RNA was analyzed. QCl-3 cells are chronically infected with an *env* deletion mutant, Bryan high-titer Rous sarcoma virus, which provides pre-existing *gag* and *pol* proteins, thus mimicking the infected-cell environment. Since we used an *env* probe, Bryan high-titer Rous sarcoma virus RNA was not detected in the S1 analysis (14). Second, to determine whether the mutant phenotypes were stable, we analyzed RNAs from cultures of transfected, susceptible CEFs in which the virus had spread after multiple rounds of infection. For each of these analyses, a probe that was homologous to each mutant was used (Fig. 5B). The pLD6/IS5 clone gave rise to a virus that displayed balanced *env* splicing after multiple rounds of infection (Fig. 5, lane 7). The *env* mRNA/unspliced RNA ratio, estimated by densitometric scanning of the autoradiogram, was ca. 0.2. When transiently expressed, pLD6/IS5 produced RNA that was spliced to the same extent (data not shown). These values were similar to the ratio observed with the original pseudorevertant construct, pLD6/IS1 (Fig. 5, lane 3). Thus, pLD6/IS1 and pLD6/IS5, one with a base substitution upstream and the other with a deletion downstream of the splice acceptor site, respectively, both displayed reduced splicing compared with the original mutant (pLD6/I) and the wild type (*env* mRNA/unspliced RNA ratio, 0.5) (Fig. 5, lanes 1 and 2). We conclude that both types of suppressor mutations can compensate for the oversplicing defect of pLD6/I.

The more slowly growing pseudorevertant, pLD6/IS2, showed a different splicing phenotype. In this case, the ratio of *env* mRNA to unspliced RNA in the transient assay was ca. 2.5 (data not shown). It was reduced compared with that in the original mutant (ca. 5.0), but *env* mRNA was still in excess over unspliced RNA. Therefore, this pseudorevertant showed only a partial reversion as measured by splicing efficiency. This phenotype was stable after multiple rounds of infection (Fig. 5, lane 4). It is noteworthy that although pLD6IS2 and pLD6IS5 differ by only 15 nt (Fig. 2), they show considerable differences in splicing ratios (Fig. 5, compare lanes 4 and 7).

In the absence of the original linker mutation, we observed a dramatic decrease in splicing efficiency for the class 2 deletion mutants (Fig. 5, lanes 5, 6, 8, and 9) compared with the wt (lane 1). Both phenotypes were stable after multiple rounds of infection (shown for pLD6/S2 in Fig. 5, by comparison of lanes 5 and 6). It was difficult to assess the splicing ratios by densitometry in these cases because the intensity of the band representing the spliced form was very weak. In all of the clones described in Fig. 5, the wt *env*

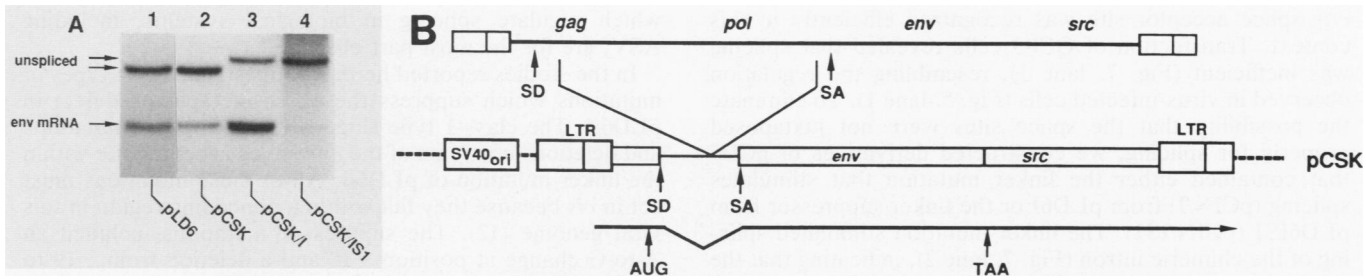

FIG. 6. Ratio of unspliced RNA to *env* mRNA in COS monkey cells. (A) Cells were harvested ca. 60 h posttransfection. S1 nuclease protection assay was as described for Fig. 5. (B) Map and derivation of pCSK. The pLD6 parental genome is illustrated above. The bulk of the intron for *env* mRNA was deleted as indicated. A map of pCSK is shown below. The transcript is illustrated below the map; the splice event is represented. The intron contains ca. 150 nt from the 5′ intron border and 77 nt from the 3′ intron border. Splice donor (SD) and acceptor sites (SA) are indicated. Initiation and termination codons for the *env* precursor are also indicated. Derivatives of pCSK, pCSK/I, and pCSK/IS1 contain the *env* splice acceptor cassette (*Kpn*I/*Xho*I) fragments shown in Fig. 3 from pLD6/I or pLD6/IS1, respectively. - - - -, Vector sequences: SV40 ori, simian virus 40 origin; LTR, long terminal repeat. The map is not drawn to scale.

splice acceptor site still appears to be used, as determined with probes that had been cleaved with restriction enzymes as gel markers (data not shown). This result is consistent with the requirement for accurate splicing to fuse the two *env*-coding exons (Fig. 1), as was expected since all of these replication-competent revertant viruses must express *env* proteins.

Roles of *gag*, *pol*, and *env* proteins in expression of wt and mutant splicing phenotypes. The mutations we identified may act in *cis* as altered signals for the normal splicing reaction, or alternatively the mutations might have identified targets for viral protein(s), or complexes modified by viral proteins, which exert a regulatory effect on splicing. To distinguish among these possibilities, we first removed the bulk of the viral *env* intron sequences by deleting ca. 4 kilobases (Fig. 6B) between the *Sac*II and *Kpn*I sites. This deletion eliminated most of the *gag* and *pol* coding regions and also any possible *cis*-acting contributions of this portion of the intron to splicing regulation. The intron sequences that remained included 150 nt downstream of the splice donor site and 77 nt upstream from the *env* splice acceptor site. This construct, pCSK, was used to transfect COS monkey cells, and subsequent S1 analysis revealed that the unspliced RNA/*env* mRNA ratio was ca. 0.5, similar to that seen in infected avian cells (Fig. 6A, lanes 1 and 2). Two mutant clones were constructed in which the splice acceptor site segments from pLD6/I and pLD6/IS1 were used to replace the wt splice acceptor region. S1 analysis of the RNA produced after transfection with these clones showed that pCSK/I spliced excessively and that this phenotype was reversed in pCSK/IS1 (Fig. 6A, lanes 3 and 4). Thus, all three splice acceptor sites behaved as they did in the context of the viral genome. These phenotypes are therefore not restricted to avian cells and do not require *cis* or *trans* functions encoded in the major portion of the intron. It seems likely, therefore, that the observed phenotypes reflect a general property of splicing reactions which is conserved in both avian and mammalian cells.

To test the possibility that *env* products are involved in splicing control, translation was prevented by changing the *env* initiation codon (position 380, Fig. 1 and 6B) to a termination codon in all three intron deletion clones. After transfection of COS cells, the same splicing ratios were obtained with the *env* mutant derivatives as had been observed with the original constructs (data not shown). Thus, expression of *env* does not appear to be required for splicing regulation. Since the termination codon also eliminates expression of residual *gag* and *pol* sequences (Fig. 1), we conclude that none of the major viral gene products are involved in this splicing regulation.

Regulated *env* splicing does not require the ASV splice donor site. The results presented above implicate the *env* splice acceptor site as a focal point of regulation. To investigate whether this regulation requires cooperation of the ASV splice donor site, we replaced a ca. 500-bp ASV segment containing the splice donor site with a ca. 130-bp segment containing the rat pre-proinsulin exon 1 splice donor site, as shown in Fig. 7B. The resulting clone (pCIN) contained a chimeric intron with 91 bp derived from insulin and 77 bp derived from ASV. We investigated whether the

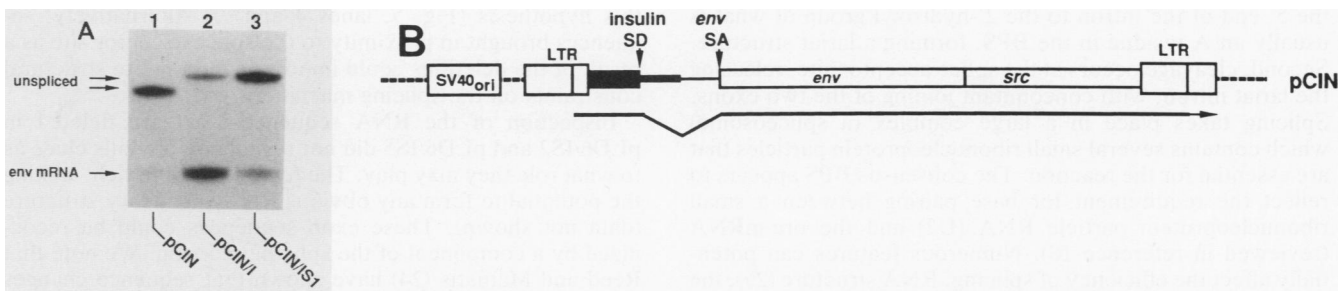

FIG. 7. Effect of replacing the ASV splice donor site with the rat preproinsulin splice donor site. (A) QCl-3 cells were transfected with the indicated plasmid DNAs, and RNAs were analyzed as described for Fig. 5. (B) Map of pCIN plasmid. A segment containing the rat preproinsulin exon 1-intron 1 border (shown as a thick line and a filled box) was substituted for a segment of the ASV leader. The expected transcript and *env* splice event is illustrated below the map (not drawn to scale).

env splice acceptor site was recognized efficiently in this context. Transfection of QCl-3 cells revealed that splicing was inefficient (Fig. 7, lane 1), resembling the regulation observed in virus-infected cells (Fig. 5, lane 1). To eliminate the possibility that the splice sites were not juxtaposed properly for splicing, we constructed derivatives of pCIN that contained either the linker mutation that stimulates splicing (pCIN/I) from pLD6I or the linker/suppressor from pLD6IS1 (pCIN/IS1). The linker mutation stimulated splicing of the chimeric intron (Fig. 7, lane 2), indicating that the splice sites could be used efficiently. This effect could also be reversed by the suppressor mutation (lane 3), as in the context of the virus. These results reinforce the notion that the wt *env* splice acceptor region is a major contributor to regulation.

DISCUSSION

Our studies used a unique ASV mutant (PLD6/I) in which splicing to produce *env* mRNA is nearly as efficient as for most cellular pre-mRNAs. This is detrimental to viral replication, since insufficient genomic RNA and *gag/pol* mRNA are available to support efficient replication. Characterization of a series of pseudorevertants showed that the splice-enhancing effect of the linker insertion, which is upstream of the splice acceptor site, can be suppressed either by changes in the insert (class 1, Fig. 2) or by changes downstream of the splice acceptor site (class 2, Fig. 2). Since the relative efficiencies of the wt and mutant splicing sites were unchanged in the absence of all the known viral replication proteins, we conclude that the mutations probably affect one or more steps in the splicing reaction carried out solely by normal cellular components. Our preliminary experiments with an in vitro system (X. Fu, R. Katz, T. Maniatis, and A. M. Skalka, unpublished observations) also support this conclusion. The incomplete splicing reactions which produce ASV *src* mRNA (33) and influenza virus NS2 mRNAs (23) may also act independently of viral proteins.

What steps in the splicing reaction could be affected by the mutations described here? Avian retrovirus RNA is likely spliced in a manner analogous to that of mammalian pre-mRNAs (31), which is a complex process involving *cis*-acting RNA sequences and *trans*-acting nuclear components (21, 28). The known *cis*-acting RNA signals include the splice donor site (consensus, AG/GUAAGU), the splice acceptor site [consensus, $(Py)_{11}$NCAG/A, Fig. 2], and a weakly conserved branch point sequence (BPS) (consensus, PyNCUGAC) located in the intron, 20 to 50 nt upstream of the splice acceptor site (28). The intron is removed in two steps. First, cleavage occurs at the splice donor site with a concomitant joining of the 5′ phosphate of the G residue at the 5′ end of the intron to the 2′-hydroxyl group of what is usually an A residue in the BPS, forming a lariat structure. Second, cleavage occurs at the splice acceptor site, releasing the lariat intron, with concomitant joining of the two exons. Splicing takes place in a large complex (a spliceosome) which contains several small ribonucleoprotein particles that are essential for the reaction. The consensus BPS appears to reflect the requirement for base pairing between a small ribonucleoprotein particle RNA (U2) and the pre-mRNA (reviewed in reference 18). Numerous features can potentially affect the efficiency of splicing: RNA structure (29), the branch point sequence (22, 25, 26, 38), splice site region consensus sequences (6), exon sequences (8, 24, 30), intron size (7, 36), *trans*-acting factors (2), and the splice sites themselves (10, 21). However, the precise mechanisms which regulate splicing in biological systems, including ASV, are for the most part obscure.

In the studies reported here, we investigated two types of mutations which suppress the excessive splicing defect in pLD6/I. The class 1 type (Fig. 2) includes point mutations and deletions upstream of the *env* splice acceptor site within the linker mutation of pLD6/I. All of these mutations must act in *cis* because they fall within a noncoding region in this viral genome (12). The suppressor mutations included an A-to-G change at position −18 and a deletion from −19 to −24 with respect to the splice acceptor site. Assuming that these mutations act in a manner similar to that of the T-to-C change at −20 in the characterized pLD6/IS1 (12), they are expected to compensate for the pLD6/I defect by down regulating splicing. What might be the mechanism of down regulation by the class 1 suppressor mutations? Neither the sequence of the pLD6/I mutation nor its cognate class 1 suppressor mutations (Fig. 2) produced a change in the splice acceptor consensus sequence (overlined in Fig. 2). They do, however, impinge on the region in which the branch point is expected to lie (−20 to −50). Thus, it is possible that the linker provides a better BPS and that the class 1 suppressor mutations make it less favorable. This hypothesis is consistent with the fact that wt and mutant splicing phenotypes were observed when a heterologous splice donor site was used (Fig. 7). However, inspection of the linker sequence does not reveal any strong BPS homologies. An alternative notion, that the linker could act by altering the spacing between the splice site and the branch point, seems unlikely, since point mutations are sufficient to reverse the phenotype. Also, the same reasoning makes it seem unlikely that a negative control element was interrupted by the linker sequence.

The class 2 pseudorevertants contain deletions in exon sequences downstream of the *env* splice acceptor site. In pLD6/IS2 and pLD6/IS5, 31 and 56 overlapping nucleotides were deleted, respectively (Fig. 2). Both deletions were suppressor mutations that reduced the efficiency of splicing and thus compensated for the excessive-splicing phenotype induced by the linker in pLD6/I (Fig. 4 and 5). Both of these deletions affected the *env* leader peptide-coding region (Fig. 1); however, the interruptions are in translational frame and therefore preserve the *env* reading frame. The virus can evidently tolerate small deletions in this region (13). Our results showed that the two class 2 deletions reduce the efficiency of splicing and act independently of the linker mutation (Fig. 5). Thus, it seems possible that the deleted sequences identify a region that is required in *cis* as part of the normal splicing reaction and therefore may reveal a previously unrecognized splicing signal. The larger deletion had a stronger negative affect on splicing, consistent with this hypothesis (Fig. 5, lanes 4 and 7). Alternatively, sequences brought in proximity to the splice acceptor site as a result of the deletions could impose inappropriate structural constraints on the splicing machinery.

Inspection of the RNA sequences that are deleted in pLD6/IS2 and pLD6/IS5 did not reveal any obvious clues as to what role they may play. The region is purine rich, lacking the potential to form any obvious RNA secondary structure (data not shown). These exon sequences could be recognized by a component of the splicing reaction. We note that Reed and Maniatis (24) have shown that sequence changes downstream of the splice acceptor sites in the globin exons can also have a considerable effect on splicing.

We observed regulated splicing even when ca. 95% of the *gag-pol* intron region was removed (Fig. 6). Thus, it appears

that 120 nt from the 5′ intron border and 77 nt from the 3′ intron border are sufficient to interact with the cellular splicing machinery in a manner that is similar to that of the full-size intron. However, two separate studies by others have indicated that deletions in the *gag-pol* region can decrease the ratio of unspliced RNA to *env* or *src* ASV mRNAs (1, 32). Some of the deletions appear to affect RNA stability, while others appear to affect splicing per se. We observed no similar effect in monkey cells (Fig. 6), but the specific deletions and virus strains differed.

The human retroviruses have evolved unique and intricate mechanisms to control RNA availability that exploit elements which potentially govern transcription, RNA stability, splicing, and transport from the nucleus (34). They include both *cis*-acting sequences in the viral RNA and *trans*-acting viral proteins that recognize these sequences. One such protein is required for efficient expression of viral RNA (e.g., the Tat protein of human immunodeficiency virus) and a second is required for cytoplasmic expression of full-length and *env* mRNA (e.g., the Rev protein of human immunodeficiency virus). These controlling elements appear to allow these viruses to establish a latent phase in which viral structural proteins are not expressed; subsequently, high-level expression of structural proteins can be triggered through modulation of these controls to give rise to fulminant release of virions. The ASV life cycle does not include such a latent phase, and gene expression, as well the RNA splicing pattern, appears to be constitutive. Thus, the fact that *trans*-acting viral proteins do not appear to be involved in splicing control (see Results) (1, 32) or transcriptional activation is not totally unexpected. However, a balance between unspliced and *env* mRNAs must be maintained by all retroviruses. Consequently, it would not be surprising if such a balance is also maintained in human immunodeficiency virus by *cis*-acting features of the *env* splice signals and that the regulation provided by *rev* is superimposed upon this control.

Recent studies which used a system which measured pre-mRNA expression in yeast cells have indicated that blocking an early step in splicing is sufficient to allow unspliced RNA to be expressed in the cytoplasm (16). These authors speculate that such an early block might allow some intron-containing RNA molecules to escape commitment to a splicing pathway and thus be more accessible for some other transport mechanism, for example, that which allows RNAs without introns to be introduced into the cytoplasm. It seems possible that retroviral RNAs with suboptimal signals may be poor substrates for spliceosome assembly. In the case of ASV, an inefficient splice acceptor site seems to be sufficient. In other retroviruses, *cis*-acting control could potentially occur at the splice donor site. However, as illustrated in other studies (discussed in reference 16), inhibition of pre-mRNA splicing by mutation of splice sites does not necessarily result in the appearance of stable, unspliced RNA in the cytoplasm. Thus, it is possible that additional RNA sequences which affect the stability and/or transport of unspliced RNAs may also be important for balanced ASV RNA expression.

The biological importance of *cis*-acting sequences in ASV splicing control is supported by the fact that our down regulating splicing mutations were isolated by selection in vivo. It seems possible that such suboptimal splicing signals may also have been selected for during evolution of the virus. We are currently evaluating the ASV *env* splice event in vitro to determine which steps in the splicing reaction are regulated and how the mutations we have described affect these steps. These studies should provide additional information regarding general mechanisms which govern splicing efficiency.

ACKNOWLEDGMENTS

We gratefully acknowledge George Merkel for excellent technical assistance. We also thank John Taylor and Martha Peterson for critical review of this manuscript.

This work was supported by Public Health Service grants CA-06927, CA-48703, and RR-05539 from the National Institutes of Health, a grant from the Pew Charitable Trust, and an appropriation from the Commonwealth of Pennsylvania. Partial support for R.A.K. was provided by the W. W. Smith Charitable Trust.

LITERATURE CITED

1. **Arrigo, S., and K. Beemon.** 1988. Regulation of Rous sarcoma virus RNA splicing and stability. Mol. Cell. Biol. **8**:4858–4867.
2. **Breitbart, R. E., A. Andreadis, and B. Nadal-Ginard.** 1987. Alternative splicing: a ubiquitous mechanism for the generation of multiple protein isoforms from single genes. Annu. Rev. Biochem. **56**:467–495.
3. **Coffin, J.** 1985. Genome structure, p. 17–73. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), Molecular biology of tumor viruses: RNA tumor viruses, 2nd ed., vol. 2. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
4. **Cullen, B. R.** 1987. Use of eukaryotic expression technology in the functional analysis of cloned genes. Methods Enzymol. **152**:684–704.
5. **Ficht, T. A., L.-J. Chang, and C. M. Stoltzfus.** 1984. Avian sarcoma virus *gag* and *env* gene structural protein precursors contain a common amino-terminal sequence. Proc. Natl. Acad. Sci. USA **81**:362–366.
6. **Fu, X., H. Ge, and J. L. Manley.** 1988. The role of the polypyrimidine stretch at the SV40 early pre-mRNA 3′ splice site in alternative splicing. EMBO J. **7**:809–817.
7. **Fu, X., and J. L. Manley.** 1987. Factors influencing alternate splice site utilization in vivo. Mol. Cell. Biol. **7**:738–748.
8. **Furdon, P. J., and R. Kole.** 1986. Inhibition of splicing but not cleavage at the 5′ splice site by truncating human β-globin pre-mRNA. Proc. Natl. Acad. Sci. USA **83**:927–931.
9. **Goff, S., P. Traktman, and D. Baltimore.** 1981. Isolation and properties of Moloney murine leukemia virus mutants: use of a rapid assay for release of virion reverse transcriptase. J. Virol. **38**:239–248.
10. **Green, M. R.** 1986. Pre-mRNA splicing. Annu. Rev. Genet. **20**:671–708.
11. **Hirt, B.** 1967. Selective extraction of polyoma DNA from infected mouse cell cultures. J. Mol. Biol. **26**:365–369.
12. **Katz, R. A., M. Kotler, and A. M. Skalka.** 1988. *cis*-acting intron mutations that affect the efficiency of avian retroviral RNA splicing: implications for mechanisms of control. J. Virol. **62**:2686–2695.
13. **Katz, R. A., and A. M. Skalka.** 1988. A C-terminal domain in the avian sarcoma-leukosis virus *pol* gene product is not essential for viral replication. J. Virol. **62**:528–533.
14. **Katz, R. A., R. W. Terry, and A. M. Skalka.** 1986. A conserved *cis*-acting sequence in the 5′ leader of avian sarcoma virus RNA is required for packaging. J. Virol. **59**:163–167.
15. **Kopchick, J. J., and D. W. Stacey.** 1984. Differences in intracellular DNA ligation after microinjection and transfection. Mol. Cell. Biol. **4**:240–246.
16. **Legrain, P., and M. Rosbash.** 1989. Some *cis*- and *trans*-acting mutants for splicing target pre-mRNA to the cytoplasm. Cell **57**:573–583.
17. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
18. **Maniatis, T., and R. Reed.** 1987. The role of small nuclear ribonucleoprotein particles in pre-mRNA splicing. Nature (London) **325**:673–678.
19. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol.

65:499–560.
20. **Miller, C. K., J. E. Embretson, and H. M. Temin.** 1988. Transforming viruses spontaneously arise from nontransforming reticuloendotheliosis virus strain T-derived viruses as a result of increased accumulation of spliced viral RNA. J. Virol. **62**:1219–1226.
21. **Padgett, R. A., P. J. Grabowski, M. M. Konarska, S. Seiler, and P. A. Sharp.** 1986. Splicing of messenger RNA precursors. Annu. Rev. Biochem. **55**:1119–1150.
22. **Padgett, R. A., M. M. Konarska, M. Aebi, H. Hornig, C. Weissmann, and P. A. Sharp.** 1985. Nonconsensus branch site sequences in the *in vitro* splicing of transcripts of mutant rabbit β-globin genes. Proc. Natl. Acad. Sci. USA **82**:8349–8353.
23. **Plotch, S. J., and R. M. Krug.** 1986. *In vitro* splicing of influenza viral NS1 mRNA and NS1-β-globin chimeras: possible mechanisms for the control of viral mRNA splicing. Proc. Natl. Acad. Sci. USA **83**:5444–5448.
24. **Reed, R., and T. Maniatis.** 1986. A role for exon sequences and splice-site proximity in splice-site selection. Cell **46**:681–690.
25. **Reed, R., and T. Maniatis.** 1988. The role of the mammalian branchpoint sequences in pre-mRNA splicing. Genes Dev. **2**:1268–1276.
26. **Ruskin, B., J. M. Greene, and M. R. Green.** 1985. Cryptic branch point activation allows accurate in vitro splicing of human β-globin intron mutants. Cell **41**:833–844.
27. **Schwartz, D. B., R. Tizard, and W. Gilbert.** 1983. Nucleotide sequence of Rous sarcoma virus. Cell **32**:853–869.
28. **Sharp, P. A.** 1987. Splicing of messenger RNA precursors. Science **235**:766–771.
29. **Solnick, D.** 1985. Alternative splicing caused by RNA secondary structure. Cell **43**:667–676.
30. **Somasekhar, M. B., and J. E. Mertz.** 1985. Exon mutations that affect the choice of splice sites used in processing the SV40 late transcripts. Nucleic Acids Res. **13**:5591–5609.
31. **Stoltzfus, C. M.** 1989. Synthesis and processing of avian sarcoma retrovirus RNA. Adv. Virus Res. **35**:1–38.
32. **Stoltzfus, C. M., and S. J. Fogarty.** 1989. Multiple regions in the Rous sarcoma virus *src* gene intron act in *cis* to affect the accumulation of unspliced RNA. J. Virol. **63**:1669–1676.
33. **Stoltzfus, C. M., S. K. Lorenzen, and S. L. Berberich.** 1987. Noncoding region between the *env* and *src* genes of Rous sarcoma virus influences splicing efficiency at the *src* gene 3′ splice site. J. Virol. **61**:177–184.
34. **Varmus, H.** 1988. Regulation of HIV and HTLV gene expression. Genes Dev. **2**:1055–1062.
35. **Weiss, R., N. Teich, H. Varmus, and J. Coffin (ed.).** 1982. Molecular biology of tumor viruses: RNA tumor viruses, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
36. **Wieringa, B., E. Hofer, and C. Weissmann.** 1984. A minimal intron length but no specific internal sequence is required for splicing the large rabbit β-globin intron. Cell **37**:915–925.
37. **Wills, J. W., J. M. Hardwick, K. Shaw, and E. Hunter.** 1983. Alterations in the transport and processing of Rous sarcoma virus envelope glycoproteins mutated in the signal and anchor regions. J. Cell. Biochem. **23**:81–94.
38. **Zhuang, Y., A. M. Goldstein, and A. M. Weiner.** UACUAAC is the preferred branch site for mammalian mRNA splicing. Proc. Natl. Acad. Sci. USA **86**:2752–2756.