The EMBO Journal vol.7 no.3 pp.809-817, 1988

The role of the polypyrimidine stretch at the SV40 early pre-mRNA 3′ splice site in alternative splicing

Xin-Yuan Fu, Hui Ge and James L.Manley

Department of Biological Sciences, Columbia University, New York, NY 10027, USA

Communicated by W.Keller

We have studied the role in pre-mRNA splicing of the nucleotide sequence preceding the SV40 early region 3′ splice site. Somewhat surprisingly, neither the pyrimidine at the highly conserved −3 position, nor the polypyrimidine stretch that extends from −5 to −15, relative to the 3′ splice site, were found to be required for efficient splicing. Mutations that delete this region or create polypurine insertions at position −2 had no significant effects on the efficiency of SV40 early pre-mRNA splicing *in vivo* or *in vitro*. Interestingly, however, the pyrimidine content of this region had substantial effects on the alternative splicing pattern of this pre-mRNA *in vivo*. Mutations that increased the number of pyrimidine residues resulted in more efficient utilization of the large T antigen mRNA 5′ splice site relative to the small t 5′ splice site, while mutations that increased the purine content enhanced small t mRNA splicing. A possible molecular mechanism for these findings, as well as a model that proposes a role for the polypyrimidine stretch in alternative splicing, are discussed.

Key words: alternative splicing/mRNA processing/consensus sequences

Introduction

The precise removal of introns during pre-mRNA splicing in higher eukaryotes requires *cis*-acting signal sequences. By comparing sequences at exon/intron junctions, a consensus sequence of AG/GT(A/G)AGT was identified at 5′ splice sites, and $(Py)_{n>11}NPyAG/G$ at 3′ splice sites (Breathnach and Chambon, 1981; Mount, 1982). Additionally, a more loosely defined consensus, PyNPyTPuAPy, was found near the 3′ ends of introns, with the adenosine residue serving as the site of branch formation (Ruskin *et al.*, 1985; Keller and Noon, 1984). The two dinucleotides that define the 5′ and 3′ boundaries of introns are essentially invariant, and in most cases mutations within them either greatly reduce or abolish splicing (Treisman *et al.*, 1982; Wieringa *et al.*, 1983; Aebi *et al.*, 1986).

In contrast, the requirement for the other conserved sequences in pre-mRNA splicing appears to be more flexible. For example, the role of the polypyrimidine stretch at the 3′ site is not completely clear. Several experiments have suggested that this region plays an important role in mammalian pre-mRNA splicing. Wieringa *et al.* (1984) demonstrated that progressive deletions of the polypyrimidine region from −24 to −9 upstream of the 3′ splice site in the rabbit β-globin large intron abolished correct splicing *in vivo* and activated an upstream cryptic 3′ splice site. Similar results were obtained by van Santen and Spritz (1985), who analyzed the effects of such mutations in the human β-globin gene. Ruskin and Green (1985a) showed that a mutation deleting 14 of the original 17 nt polypyrimidine stretch in the human β-globin first intron greatly reduced the efficiency of the first step of splicing *in vitro*; i.e. 5′ splice site cleavage and lariat formation. Recently, a protein associated with a snRNP particle (possibly U5 snRNP) was found to interact preferentially with the 3′ splice site (Chabot *et al.*, 1985; Tazi *et al.*, 1986; Gerk and Steitz, 1986). In one experiment, a mutant pre-mRNA that contained a disrupted polypyrimidine stretch was found not to interact efficiently with this protein, suggesting that the polypyrimidine region may be involved in recognition of the 3′ splice site by a component of the splicing machinery (Tazi *et al.*, 1986).

Although these experiments suggest an important role for the polypyrimidine region in pre-mRNA splicing, several additional observations indicate that the picture may be more complicated. First, unlike at the invariant 5′/GT and AG/3′ dinucleotides, no naturally occurring mutations that affect splicing have been found in the polypyrimidine tract (Treisman *et al.*, 1983; Mitchell *et al.*, 1986). Thus, it may be that many mutations in this region are not deleterious, perhaps because minor changes can be easily tolerated. Second, yeast pre-mRNAs lack pyrimidine-rich regions at 3′ splice sites, yet at least one, a ribosomal protein pre-mRNA, is efficiently spliced in a HeLa cell nuclear extract (Ruskin *et al.*, 1986). Third, several examples in which 3′ splice sites are preceded by purine-rich instead of pyrimidine-rich regions have been observed (Mount, 1982; Bernstein *et al.*, 1986; Eveleth *et al.*, 1986; Morgan *et al.*, 1986). Hypotheses have been proposed in which purine-rich regions immediately upstream of the 3′ splice sites in the *Drosophila* myosin alkali light chain and heavy chain genes act as alternative splicing signals, recognized by developmentally regulated positive or negative *trans*-acting factors (Falkenthal *et al.*, 1985; Rozek and Davison, 1986; Bernstein *et al.*, 1986). Taken together, these observations have raised several intriguing questions about the function(s) of the polypyrimidine region in pre-mRNA splicing.

We have been using the SV40 early gene as a model to investigate the mechanism of one type of alternative splicing. SV40 early pre-mRNA can be spliced *in vivo* to produce large T and small t mRNAs by the utilization of alternative 5′ splice sites and a common 3′ splice site. Our previous studies showed that the alternative splicing pattern of this pre-mRNA is finely controlled by several factors, including the size of the small t intron (Fu and Manley, 1987). In addition, the ratio of small t to large T (t/T) mRNA is significantly higher in human 293 cells than in a variety of other mammalian cell types (Fu and Manley, 1987), suggesting the existence of *trans*-acting factors that influence alternative splicing. This view is supported by the observations that large T, but not small t, mRNA splicing involves the utilization of multiple lariat branch sites, and that the selection of the

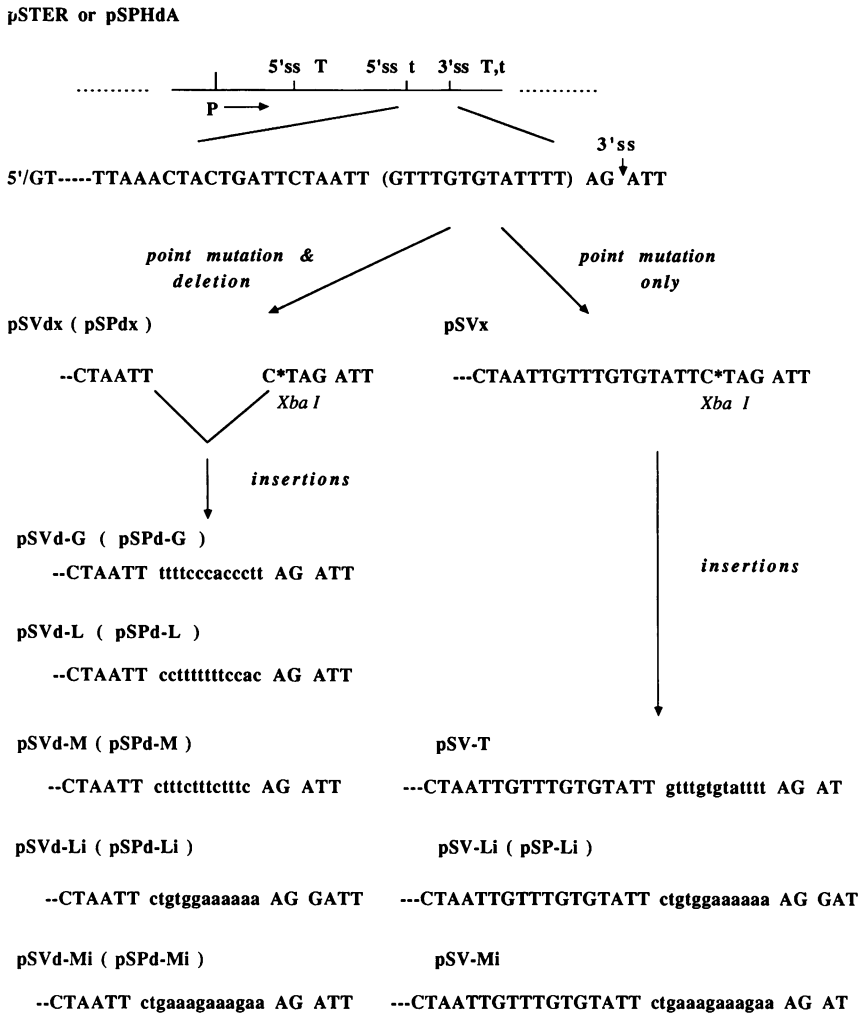

Fig. 1. Nucleotide sequences of SV40 early polypyrimidine region mutants. pSTER and pSPHdA are the wild-type plasmids used for *in vivo* and *in vitro* analyses, respectively. The splice sites utilized for large T and small t splicing are indicated, as is the direction of transcription. The sizes of the wild-type large T and small t introns are 346 and 66 nt, respectively. The 3′ cleavage site in the wild-type is designated by an arrow, and the nucleotides removed in the deletion and substitution mutants are indicated by parentheses. The base substitution (T to C) creating an *Xba*I site is indicated by an asterisk, and the deletion in the pSVd series by a folded line. Substitutions and insertions are indicated by small letters.

predominant branch site varies in a cell-specific manner correlated with the relative utilization of 5′ splice sites (Noble *et al.*, 1987). In order to explore further the interactions of this pre-mRNA with the splicing machinery, and particularly the role(s) of the polypyrimidine stretch in these interactions, we have constructed a series of deletion and substitution mutations in the polypyrimidine region of the SV40 early gene. Our results suggest that the sequences preceding the 3′ splice site have an important but unexpectedly flexible role in SV40 early pre-mRNA splicing.

Results

To study the function of the polypyrimidine stretch preceding the SV40 early mRNA 3′ splice site, we constructed a series of mutations in this region and analyzed their effects in both *in vivo* and *in vitro* assays. The plasmid pSTER, which was used for *in vivo* transient expression experiments, contains the entire SV40 early region (Lewis and Manley, 1985). For *in vitro* analyses, a *Hin*dIII fragment extending from SV40 nt 5171 to 4002 was inserted into the vector pSP64 (Melton *et al.*, 1984) to create pSPHdA, which served as a template for the production of precursor RNAs by bacteriophage SP6 RNA polymerase. The relevant sequences of all mutants described here are shown in Figure 1.

Deletion of the SV40 early polypyrimidine region does not affect large T mRNA splicing

The pyrimidine content of the SV40 early polypyrimidine region is 9 out of 13 nt from position −5 to −17, relative to the 3′ splice site (see Figure 1). To investigate whether this pyrimidine-rich region is required for splicing of SV40 early pre-mRNA, a site-specific mutation was made that deleted 11 nt, from position −5 to −15, to create the plasmid pSVdx. The effects of this mutation on SV40 early splicing *in vivo* were assayed following transfection of HeLa and 293 cells. Cytoplasmic RNA was isolated and the amount of large T and small t mRNAs present was quantitated by S1 nuclease mapping. Small t splicing was completely abolished in pSVdx-transfected HeLa cells (Figure 2A) and 293 cells (Figure 2B). This finding was not unexpected, because previous results in several systems have shown that splicing could be reduced or abolished by deletions that decrease the size of an intron to below a minimum length (Wieringa *et*

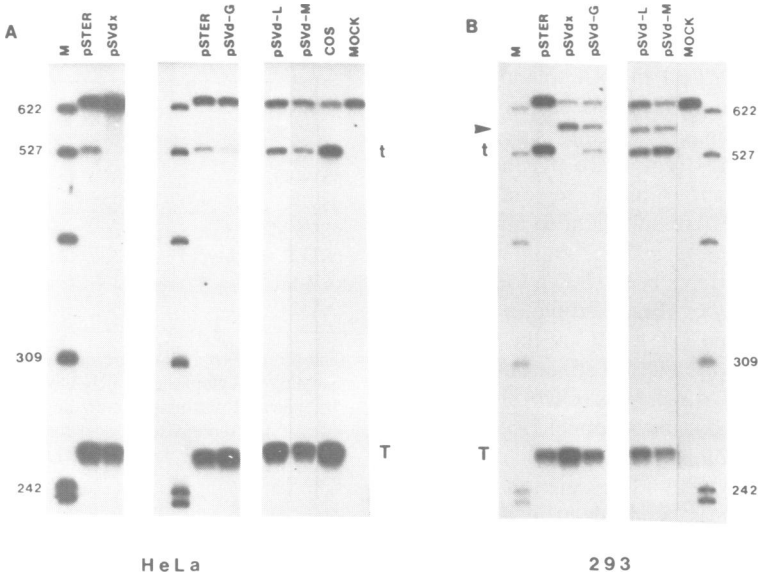

Fig. 2. S1 nuclease analysis of small t and large T mRNA produced *in vivo* by polypyrimidine region mutants. HeLa and 293 cells were transfected with pSTER (wild-type) or mutant plasmids and total cytoplasmic RNA was purified 48 h later. A 3′ end-labelled DNA probe (SV40 nt 5171−4528) was utilized to map the 5′ splice sites of large T (254 nt protected fragment) and small t (534 nt) mRNAs. Following S1 nuclease digestion, protected DNA fragments were resolved by electrophoresis in a denaturing 5% polyacrylamide−urea gel. The plasmids used for transfections are indicated at the top of the figure. **Lane M** contained a ^{32}P-labelled *Hpa*II digest of pBR322; sizes of the DNA fragments are indicated in nucleotides. COS, total cytoplasmic RNA from an SV40 transformed monkey cell line, COS-7 (Gluzman, 1981). MOCK, RNA from HeLa or 293 cells that had not been transfected. The band just above the largest size marker represents undigested DNA probe.

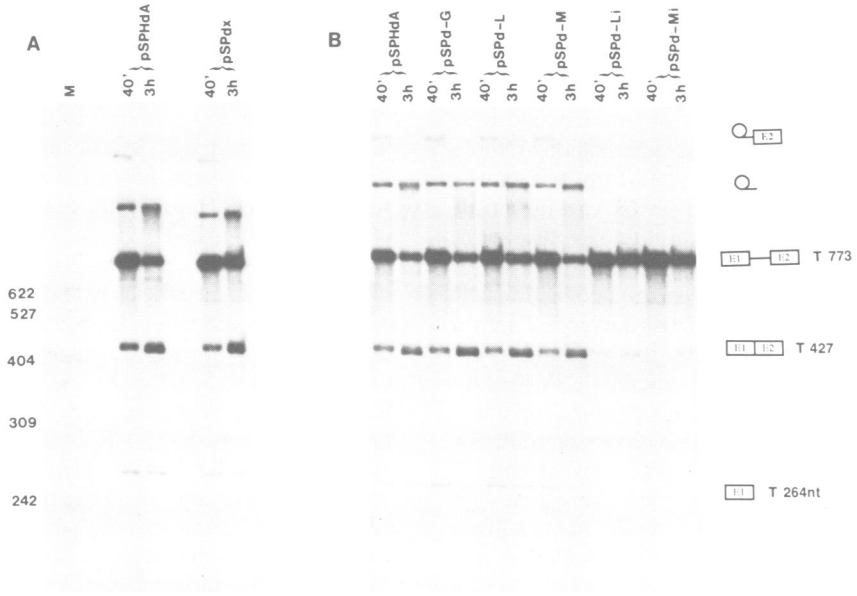

Fig. 3. Effects of polypyrimidine region mutations on *in vitro* splicing. **(A)** *In vitro* processing of the polypyrimidine deletion mutant transcript. ^{32}P-Labelled precursor RNAs synthesized from the wild-type (pSPHdA) and the polypyrimidine deletion mutant (pSPdx) templates were processed in HeLa nuclear extracts for 40 min or 3 h. RNAs were purified and splicing products and intermediates were resolved by electrophoresis in a denaturing 6% polyacrylamide−urea gel. **(B)** *In vitro* processing of precursors containing substitution mutations in the polypyrimidine region. The indicated mutant precursors were processed in HeLa nuclear extracts and products resolved by PAGE. The deduced structures and sizes of the RNA species are schematically diagrammed on the right. Boxes indicate exons, lines denote introns. Sizes of the DNA markers (M) are indicated on the left.

al., 1984; Ruskin *et al.*, 1985; Ulfendahl *et al.*, 1985), and the small t intron in pSVdx is well below this size. Surprisingly, this deletion, which removed almost all of the polypyrimidine region, had little effect on the accumulation of large T mRNA in both HeLa and 293 cells. The slight differences in the amounts of large T mRNA indicated in Figure 2 were due to variations in transfection efficiencies (see Materials and methods). The band indicated by the arrow in panel B most likely resulted from protection of the probe by unspliced pre-mRNA that had leaked into the cytoplasm from the nucleus during RNA preparation. Examination of the sequence immediately preceding the 3′ splice site in pSVdx reveals that a tentative branch point is located at a position 21 nt upstream of the 3′ splice site. The se-

quence TGTTAAA matches the branch site consensus sequence at 5 out of 6 nt, and has been shown to be utilized as a branch site in wild-type large T mRNA splicing *in vitro* (Noble *et al.*, 1987). In addition, the sequence from −5 to −19 (from the 3′ splice site) in pSVdx remains pyrimidine-rich, as 10 out of 15 nt in this region are pyrimidines. It may be that this substitute polypyrimidine region, together with a functional branch site, are sufficient to allow large T mRNA splicing.

Some splice site mutations have been shown to have different effects *in vivo* and *in vitro*. For example, a mutation changing the highly conserved GT to GC in the rabbit β-globin large intron 5′ splice site had no effect on the level of spliced RNA detected *in vivo*, whereas *in vitro* it reduced correct splicing and promoted exon skipping (Aebi *et al.*, 1986). In the light of such findings, we decided to determine the effects of the above polypyrimidine deletion mutation *in vitro*. Capped pre-RNA was synthesized *in vitro* from mutant and wild-type templates (pSP series; see Figure 1) and processed in a HeLa cell nuclear extract as described previously (Noble *et al.*, 1986, 1987). Incubations were terminated after 40 min or 3 h and splicing products were analyzed directly by PAGE (see Materials and methods). Two splicing products, spliced large T RNA and large T lariat intron, and two intermediates, 5′ exon and lariat large T intron-3′ exon, were observed with both wild-type and mutant pre-RNAs (Figure 3, panel A). Identification of these products and intermediates was based on: (i) their relative mobilities compared with size markers; (ii) debranching analysis (Ruskin and Green, 1985b) of the lariat large T intron and large T intron-3′ exon (data not shown); (iii) changes in amounts of each product or intermediate with time; and (iv) additional previous characterizations (Noble *et al.*, 1986, 1987). Pre-RNA transcribed from pSPdx was correctly spliced to produce large T mRNA, with an efficiency similar to that of pSPHdA (Figure 3A). The lariat large T intron and the intron with 3′ exon from pSPdx pre-RNA migrated slightly faster than the corresponding products from pSPHdA, due to the smaller size of the large T intron in pSPdx. These results indicate that the authentic polypyrimidine region is not required for splicing of SV40 early pre-RNA *in vitro*. No small t splicing products or intermediates were found among the *in vitro* processed pSPHdA transcripts, even though the splicing efficiency of large T RNA was high (>50%). These results are consistent with our previous findings that the small t 5′ splice site is utilized inefficiently *in vitro* (Noble *et al.*, 1986; Manley *et al.*, 1987).

Effects of polypyrimidine substitution mutations on SV40 early splicing

The above results suggest that a substitute pyrimidine-rich region can suffice for large T mRNA splicing. To test whether different polypyrimidine regions can influence SV40 early splicing, we constructed three substitution mutants in which the nucleotides from −3 to −15 relative to the 3′ splice site were replaced with the corresponding nucleotides from the human β-globin first intron (pSVd-G), from the adenovirus 2 late mRNA leader first intron (pSVd-L), or by a 'perfect' pyrimidine stretch (pSVd-M). The altered sequences in these mutants, which did not change the size of the intron, are shown in Figure 1. Note that all of these mutations significantly increased the pyrimidine content of the 3′ splice site.

Table I. Effects of polypyrimidine stretch insertion and substitution mutations on the alternative splicing pattern of SV40 early pre-mRNA

Plasmid	Py content[a]	t/T mRNA ratio (HeLa)	t/T mRNA ratio (293)
pSVd-G	1/13	1:25	1:5
pSVd-L	1/13	1:12	1:2
pSVd-M	0/13	1:10	1:1
pSTER (wild-type)	4/13	1:6	2:1
pSV-Mi	11/13	2:1	20:1
pSV-Li	9/13	4:1	24:1

The t/T mRNA ratios were derived from quantitative S1 nuclease mapping of cytoplasmic mRNA obtained from cells transfected with the indicated plasmids. Autoradiograms produced without the use of intensifying screens were used for densitometry scanning.
[a]From position −3 to −16, excluding the nonconserved nucleotide at position −4.

The overall efficiencies of pre-mRNA splicing *in vivo* in all three mutant transcripts were similar to the level of the wild-type (pSTER) transcript, in both HeLa and 293 cells (Figure 2). However, the ratios of small t mRNA to large T mRNA were significantly changed. We showed previously that the t/T mRNA ratio is unusually high in 293 cells, between 2:1 and 3:1 (Fu and Manley, 1987). This ratio was also seen with pSTER in the experiment shown in Figure 2, panel B. However, in pSVd-G transfected cells, the t/T mRNA ratio was 1:5, representing a greater than 10-fold decrease compared with wild-type. In HeLa cells, which display the 'normal' t/T mRNA ratio, also ~1:5, the pSVd-G mutation again resulted in increased large T splicing, although the effect was less dramatic than in 293 cells, as the ratio was decreased by only about a factor of four. With the other two mutants, pSVd-L and pSVd-M, the t/T mRNA ratios were reduced as well, but somewhat less than with pSVd-G (see Table I).

To determine whether these polypyrimidine substitution mutations affect the efficiency and kinetics of *in vitro* splicing of SV40 early pre-RNA, the mutated fragments were introduced into SP64 for subsequent analysis in HeLa cell nuclear extracts (Figure 3, panel B). As with pSPHdA, no small t RNA splicing products or intermediates were detected. Mutant and wild-type large T RNA splicing patterns were very similar, except that large T spliced RNA was formed somewhat more efficiently from the pSPd-G pre-RNA, perhaps reflecting its altered splicing pattern *in vivo*. These results together indicate that polypyrimidine regions from different sources can be interchanged without greatly affecting the overall efficiency of splicing, although they appear to influence the choice of alternative 5′ splice sites *in vivo*.

Substitution of the authentic polypyrimidine stretch with purine-rich fragments impairs pre-mRNA splicing

The above results have shown that deletion of the authentic polypyrimidine stretch of the SV40 early 3′ splice site did not severely affect the accumulation of the spliced products. To test whether a pyrimidine-rich region of any sort is required for SV40 early pre-mRNA splicing, we analyzed two polypurine substitution mutants, pSVd-Mi and pSVd-Li, in which the authentic polypyrimidine region was replaced with purine-rich fragments (see Figure 1, and Materials and methods). As shown in Figure 4, the *in vivo* splicing efficiencies of both pSVd-Li and pSVd-Mi pre-mRNAs were greatly reduced, and substantial amounts of unspliced precur-

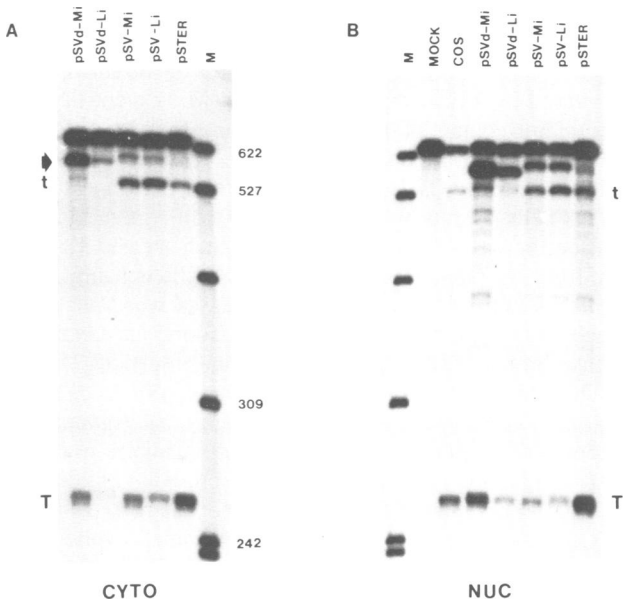

Fig. 4. Effects of polypurine substitution and insertion mutations on splicing *in vivo*. HeLa cells were transfected with wild-type (pSTER) or the indicated mutant DNAs, and RNA extracted from the cytoplasm (CYTO) or nucleus (NUC) was analyzed by S1 nuclease mapping using a 3′ end-labelled probe prepared from wild-type DNA. t and T represent small t and large T mRNA, respectively. The band indicated by the arrow in each lane that contained RNA from mutant transfected cells corresponds to unspliced precursor RNA. The probe was cleaved at the point at which the DNA and RNA sequences diverged.

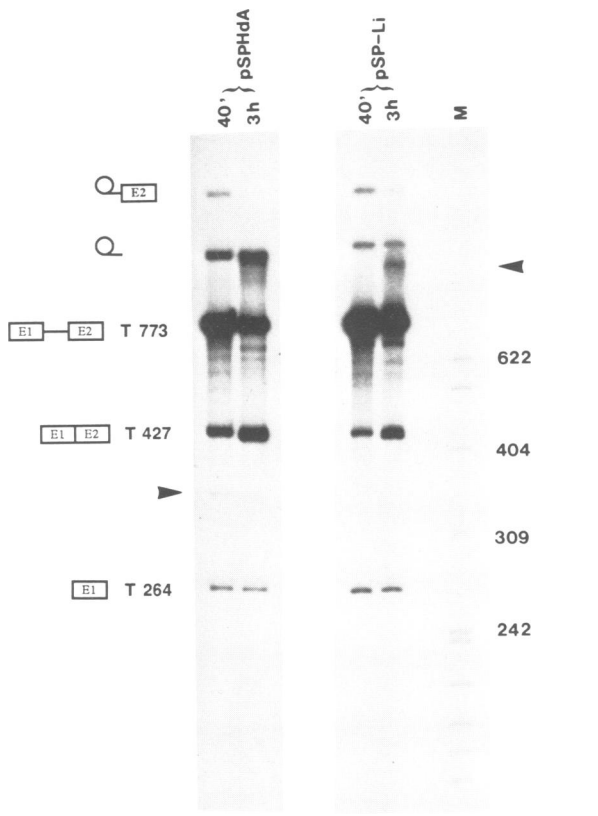

Fig. 5. *In vitro* splicing of the polypurine insertion mutant pSP-Li. Wild-type and mutant pre-RNAs were prepared and processed in HeLa nuclear extracts for 40 min or 3 h. The structure of each species is indicated by the diagram on the left. The two arrows indicate bands that represent linear large T intron (in pSPHdA) and a lariat large T intron with a shortened 3′ tail, presumably generated by a 3′ exonuclease.

sors accumulated in both cytoplasmic and nuclear fractions of transfected HeLa cells. (Similar results were observed in 293 cells; data not shown.) The observed inhibition was greater with pSVd-Li than with pSVd-Mi. Approximately 5% of the wild-type amount of spliced mRNA accumulated in pSVd-Li transfected cells. Notably, the t/T mRNA ratio in pSVd-Li transfected cells (0.8:1 in the nuclear fraction) was higher than that of pSTER (1:6), suggesting that this mutation also influenced 5′ splice site selection. S1 nuclease mapping of the 3′ splice site with a 5′ end-labelled DNA probe demonstrated that only the wild-type 3′ splice site was used (data not shown; see also below).

The effects of the mutations in pSVd-Li and pSVd-Mi were tested *in vitro* by processing pre-RNAs from pSPd-Li and pSPd-Mi in HeLa cell nuclear extracts. As shown in Figure 3, panel B, the purine-rich substitution mutations greatly reduced the *in vitro* splicing efficiency, although detectable amounts of spliced large T RNA were produced from both pSPd-Li and pSPd-Mi pre-RNAs after 3 h incubation. Quantitation of this and other autoradiograms revealed that 5−10% of each mutant pre-RNA was spliced to large T RNA, representing a 5- to 10-fold reduction in splicing efficiency compared with wild-type. The faint band representing pSPd-Mi spliced large T RNA appeared to migrate somewhat slower than the large T product. This is consistent with utilization of the most upstream AG as the major 3′ splice site, resulting in a spliced large T RNA with an additional 8 nt (Figure 1; see also below). At least one intermediate, large T 5′ exon, was clearly detected after a 40 min incubation. Notably, the ratio of this intermediate to product, large T RNA, was 0.8:1 with both pSPd-Li and pSPd-Mi, whereas the ratio was 1:6 with wild-type pSPHdA (Figure 3). This finding suggests that these mutations significantly inhibited the second, as well as the first, step of pre-mRNA splicing.

The nature of the mutation in pSVd-Li was somewhat different from the other substitutions described above. Rather than precisely replacing sequences within the intron, the net effect of this mutation was to insert a G immediately downstream of the splice acceptor and to reduce the size of the intron by 1 nt (see Figure 1). It is unlikely that these additional changes contributed to the observed phenotype of the mutation, for several reasons. First, the G inserted at the +1 position would be expected to, if anything, enhance splicing efficiency, because it matches more precisely the consensus sequence (Mount, 1982). Likewise, deletion of 1 nt from the intron would not be expected to influence large T mRNA splicing (see, for example, pSVdx above), and, if anything, would decrease rather than increase the relative efficiency of size-limited small t mRNA splicing (Fu and Manley, 1987; manuscript in preparation). Thus, the effects resulting solely from the polypurine insertion in pSVd-Li may actually have been greater than observed.

Insertions of purine-rich fragments between the 3′ AG splice acceptor and the polypyrimidine region do not affect the efficiency of pre-mRNA splicing, but alter the t/T mRNA ratio

We next wished to ask what would result if, instead of replacing the polypyrimidine region with a polypurine stretch, we inserted purine-rich fragments between the 3′ splice site

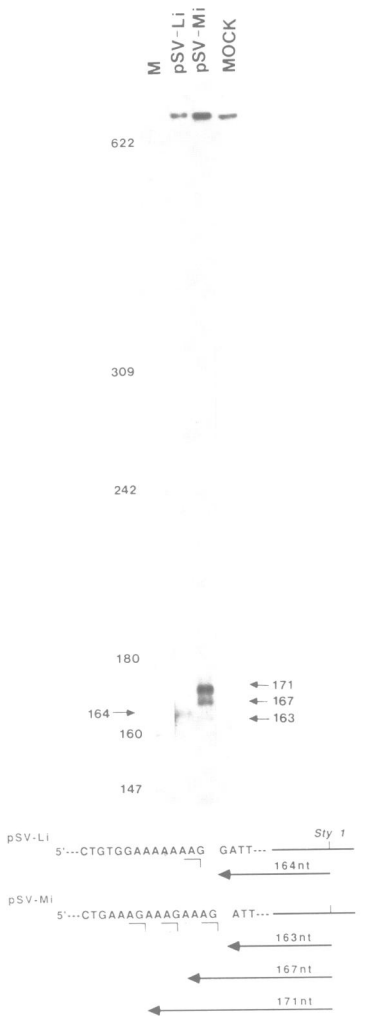

Fig. 6. Determination of the 3′ splice sites used *in vivo* by polypurine insertion mutants. The 3′ splice sites utilized in splicing of pSV-Li and pSV-Mi pre-RNAs were mapped by S1 nuclease analysis. Homologous 5′ end-labelled probes prepared from *Sty*I DNA fragments (SV40 nt 4409−5187) and hybridized with cytoplasmic RNA extracted from pSV-Li- or pSV-Mi-transfected HeLa cells. Following digestion with S1 nuclease, the products were resolved by PAGE. The sizes in nucleotides of the protected fragments are indicated. The diagram of these fragments and corresponding RNA sequences is shown at the bottom of the figure.

and pyrimidine-rich region. The insertion mutants pSV-Li and pSV-Mi were therefore constructed by inserting the same purine-rich fragments described above between the 3′ AG and the original polypyrimidine stretch (see Figure 1, and Materials and methods). These mutations increased the sizes of both the small t and large T introns by 10 nt in pSV-Li and 11 nt in pSV-Mi.

In contrast to the related mutants pSVd-Li and pSVd-Mi, the *in vivo* splicing efficiencies of pSV-Li and pSV-Mi pre-mRNAs following transfection into HeLa cells were comparable with those observed with pSTER (Figure 4). However, the ratio of small t to large T mRNA was increased by these mutations at least 20-fold, to ~4:1 in both cytoplasmic and nuclear fractions with pSV-Li and 2.5:1 with pSV-Mi transfected HeLa cells. In 293 cells, the t/T mRNA ratios were 24:1 and 20:1 with mutants pSV-Li and pSV-Mi, respectively, whereas wild-type was 2.5:1 (data not shown). That the elevated ratios of small t to large T mRNAs were not due simply to the increased size of the small t intron in these two mutants was shown by analysis of an additional mutant, pSV-T. In this mutant, the size of the intron was increased by 11 nt, exactly as in pSV-Mi, except that the pyrimidine content of the region was not altered (see Figure 1). The effects of this mutation were to increase the t/T mRNA ratios to 0.7 in transfected HeLa cells and to 5.0 in 293 cells (results not shown). Thus, while increasing the distance between the small t mRNA branch site and 3′ splice site from 18 (Noble *et al.*, 1987) to 30 nt did enhance small t mRNA splicing, the observed increases were relatively small and not sufficient to explain the more substantial increase brought about by the polypurine insertions.

The effects of the purine-rich fragment in pSV-Li on splicing *in vitro* were also determined (Figure 5). Again, only large T RNA splicing products and intermediates were detected following processing of pSP-Li pre-RNA, indicating that, although this mutation significantly increased the t/T mRNA ratio *in vivo*, it failed to activate small t splicing *in vitro*. The efficiency of *in vitro* splicing of pSP-Li pre-RNA was reduced, but only slightly, to ~50% that of pSPHdA pre-RNA, as determined by the ratio of spliced products to pre-RNA. After 40 min, the ratio of intermediates (5′ exon and lariat intron-3′ exon) to products (large T RNA and lariat intron) was significantly higher with the mutant pre-RNA (1:1.5) than with wild-type pre-RNA (1:6), indicating that the mutation may have somewhat slowed the second step in the splicing pathway.

Determination of the 3′ splice sites used in mutants pSV-Li and pSV-Mi

Previous experiments by Wieringa *et al.* (1984) showed that deletions in the polypyrimidine region of the β-globin large intron activated a cryptic 3′ splice site, in addition to reducing accumulation of correctly spliced products. In the SV40 early gene, several sequences similar to 3′ splice sites are present downstream of the authentic splice site. Could these cryptic 3′ splice sites be used in pSV-Li or pSV-Mi? Additionally, in the mutant pSV-Mi the inserted purine-rich fragment contains two AG dinucleotides preceding the original AG (see Figure 1). Which one(s) constitute(s) the 3′ cleavage sites?

To address these questions, 5′ end-labelled DNA probes were produced from pSV-Li and pSV-Mi and used to determine the 3′ splice sites by S1 nuclease mapping (Figure 6). With RNA from pSV-Li-transfected HeLa cells, only a single protected fragment of 164 nt was detected, corresponding to the expected 3′ splice site. A shorter time of electrophoresis provided no indication that downstream cryptic 3′ splice sites were used (data not shown). When RNA from pSV-Mi-transfected cells was analyzed, three protected fragments of 171, 167 and 163 nt were detected, suggesting that all three AGs were used as 3′ splice sites. However, the predominant cleavage site appeared to be the most upstream AG, as the 171 nt fragment was the major species. If this upstream AG was used, the intron would not only be smaller, but would contain only five, instead of ten, purine residues at the conserved positions preceding the AG. These two changes may explain both the decreased t/T ratio observed in pSV-Mi relative to pSV-Li-transfected cells and also the greater inhibition of splicing observed with pSVd-Li compared with pSV-Mi (Figure 3). No cryptic 3′ splice sites were activated by the mutation in pSVd-Mi, as determined

from a gel electrophoresed for a shorter time (data not shown). These results suggest that when more than one AG dinucleotide is present at a 3′ splice site, the upstream most AG is preferred, but need not be utilized exclusively.

Discussion

The results presented here lead to the somewhat unexpected conclusion that the primary role of the pyrimidine-rich sequence immediately preceding the 3′ splice site in the SV40 early region is not so much to enhance the efficiency of splicing, but rather to influence the selection of alternative 5′ splice sites. Thus, deletion of the entire polypyrimidine region, substitution of the pyrimidine-rich region with heterologous polypyrimidine tracts, and the insertion of purine-rich fragments between the 3′ AG splice acceptor and the polypyrimidine stretch all had only minor effects on the overall efficiency of splicing, both *in vivo* and *in vitro*. These results also argue that the highly conserved pyrimidine at position −3 is not required for efficient SV40 early splicing. Only when the polypyrimidine stretch was deleted and replaced with polypurine fragments was splicing significantly inhibited, and even these mutations did not totally prevent splicing.

Both the highly conserved nature of the polypyrimidine region in mammalian introns and also the results of analyses of deletion mutations that affect this region in other systems suggested an extremely important role for the polypyrimidine region in pre-mRNA splicing. In addition to preventing splicing *in vivo* (Wieringa *et al.*, 1984; van Santen and Spritz, 1985), it has been shown that deletions of these sequences block 5′ splice site cleavage *in vitro* (Ruskin *et al.*, 1985), prevent binding of a protein, probably associated with U5-snRNP, to the 3′ splice site (Tazi *et al.*, 1986; Gerk and Steitz, 1986), and prevent incorporation of the pre-RNA into spliceosomes (Frendewey and Keller, 1985).

Why does the SV40 early pre-RNA appear to be so different in its requirement for a polypyrimidine-rich region? We suggest two possible explanations. First, the types of mutations that we analyzed were different from those studied by others. While the mutations analyzed previously were all deletions, the mutations in our studies that provided the strongest evidence that an intact polypyrimidine tract is not essential for SV40 early splicing were the polypurine insertion mutations, especially pSV-Li. Here, although the T at the −3 position was replaced by an A, the remainder of the polypyrimidine tract was not deleted, but rather displaced by the 12 bp polypurine insertion. This in effect separates the 3′ AG splice acceptor from the polypyrimidine tract. One interpretation of the results obtained with this mutant is that such separations need not have significant effects on splicing. However, if this is a general property of 3′ splice sites, then it is somewhat difficult to understand why the structure of 3′ splice sites is so highly conserved.

A second possibility, which we favor, is that the role of the polypyrimidine tract in SV40 early pre-mRNA splicing is unusual compared with other pre-mRNAs studied to date. Our previous experiments have suggested that the branch site region of this precursor may have an unusually high affinity for (a) splicing factor(s), presumably, at least, U2 snRNP. These studies include the demonstration that large T mRNA splicing utilizes at least five different adenosines and one cytosine (Noble *et al.*, 1987; manuscript in preparation) as lariat branch acceptors, due at least in part to the presence of two perfect and a third near perfect consensus branch site sequences (Ruskin *et al.*, 1985; Keller and Noon, 1984); an unusually high degree of protection of this site in an endogenous RNase protection assay (Noble *et al.*, 1986); and the ability of the branch site region to allow SV40 early splicing in microinjected *Xenopus laevis* oocytes to continue under conditions that reduce the concentrations of functional snRNPs to levels insufficient to allow splicing of other pre-RNAs (Fradin *et al.*, 1984; Pan *et al.*, manuscript in preparation). We propose that the presence of this unusual branch site region relaxes the requirement for a polypyrimidine stretch in SV40 early pre-mRNA splicing.

Two different types of observations provide some support for this hypothesis. First, in yeast, the lariat branch site sequence (TACTAAC) is highly conserved (Langford and Gallwitz, 1983; Pikielny *et al.*, 1983), and mutations within it can significantly inhibit splicing. Additionally, yeast introns lack polypyrimidine stretches. While these properties contrast with those of metazoan introns, they are consistent with the idea that a strong lariat branch site might lessen the importance of the polypyrimidine stretch. Second, Chabot and Steitz (1987) studied the interactions of snRNPs with a human β-globin pre-RNA *in vitro*. These authors found that the interaction of U2 snRNP with the lariat branch site required an intact 3′ splice site, including the polypyrimidine stretch. In addition, following formation of the RNA lariat, the interaction of U2 snRNP with the RNA was extended downstream to include much of the polypyrimidine region. These findings support the idea that the polypyrimidine and lariat branch site regions are functionally interdependent.

While our results suggest that disruption of the SV40 early region polypyrimidine stretch need not significantly affect the overall efficiency of splicing, both *in vivo* and *in vitro*, they also show that alterations in this region can have a marked effect on the pattern of alternative splicing *in vivo*; i.e. on the relative utilization of the large T and small t 5′ splice sites. Examination of the t/T mRNA ratios that resulted from the different pyrimidine stretch mutations reveals an interesting pattern (Table I). Mutations that increased the pyrimidine content of this region resulted in relatively more efficient large T splicing, while those that decreased the pyrimidine content brought about enhanced small t mRNA splicing. We suggest an explanation for this, based on the size of the small t intron (66 nt). Our previous studies have shown that small t intron size is a limiting factor in small t splicing *in vivo* (Fu and Manley, 1987). Indeed, the size appears to be at or near a minimum, because deletions that removed only 2 nt from the region separating the 5′ splice site and branch site sequence virtually abolished small t splicing (Fu *et al.*, manuscript in preparation). These size constraints most likely reflect steric hindrance that prevents efficient, or productive, binding of factors (snRNPs) required for splicing. We propose that the greater the pyrimidine content at the 3′ splice site, the tighter the binding of the cognate factor (U5 snRNP?). Such tight binding might then interfere with binding and/or productive interaction of other snRNPs with the size-limited small t intron, thereby favoring large T splicing. Conversely, looser factor binding to a polypyrimidine stretch disrupted with purines might allow more efficient binding of U1 snRNP to the small t 5′ splice site, U2 snRNP to the lariat branch site, and/or subsequent interaction between these particles, or others, i.e. U4/U6

snRNP (Berget and Robberson, 1986; Black and Steitz, 1986). This would enhance small t splicing at the expense of large T splicing. It is noteworthy that pSVd-G resulted in the greatest relative enhancement of large T splicing *in vivo*, and also increased most significantly the efficiency of large T splicing *in vitro*. These findings suggest that the polypyrimidine region in this mutant, which matches that of the human β-globin first intron, may be the most efficient in factor binding. Why this pyrimidine stretch is better than those in pSVd-L and pSVd-M is not clear, although we note that the former contains more cytosine residues than any of the others.

In the model just described, intron size is an important element in the ability of the pyrimidine stretch to influence alternative splicing. However, it is possible that the nature of interactions at pyrimidine tracts might be a more general way by which alternative splicing can be modulated, perhaps, for example, by influencing the rate of formation or stability of spliceosomes, factors which could play roles in intron or exon 'skipping'. Interestingly, we have noticed that in several genes that undergo this type of alternative splicing, relevant 3′ splice sites contain purine-rich sequences immediately adjacent to the 3′ splice site. These include the fifth intron in *Drosophila* myosin alkali light chain gene (Falkenthal *et al.*, 1985), and the intron prior to exon B in the myosin heavy chain gene (Rozek *et al.*, 1986), the first intron in the *Drosophila* dopa decarboxylase gene (Eveleth *et al.*, 1986), the fifth intron in the human CSF-1 gene (Ladner *et al.*, 1987), the intron preceding exon III 9 in the rat fibronectin gene (Schwarzbauer *et al.*, 1987), the intron before the B exon in the rat calcitonin gene (Amara *et al.*, 1982) and the fifth intron in the rat preprotachykinin gene (Nawa *et al.*, 1984). The purine content of the region immediately upstream of the 3′ splice sites (from −5 to −15) in these eight genes averages almost 60%, and is as high as 90%.

In conclusion, the results of experiments described here and previously indicate that the relatively simple pattern of alternative splicing that occurs in the SV40 early region is finely tuned by a surprising number of *cis*-acting sequences. These include at least the size of the small t intron (Fu and Manley, 1987), the relative 'strengths' of the small t and large T 5′ splice sites (Zhuang *et al.*, 1987), the existence of multiple alternative branch sites (Noble *et al.*, 1987), and finally the pyrimidine content of the 3′ splice site.

Materials and methods

Site specific mutagenesis and plasmid construction

All plasmids used in this study were derived from pSTER (Lewis and Manley, 1985), which contains the entire SV40 early region from nt 294 to 2533 (Tooze, 1981) inserted between the *Eco*RI and *Bam*HI sites of pBR322. Oligonucleotide-directed mutagenesis using a single-stranded M13 derivative containing SV40 DNA sequences from nt 4002 to 5171 was performed by a modification (Fu and Manley, 1987) of a procedure described previously (Norris *et al.*, 1983; Zoller and Smith, 1983). All oligonucleotides were synthesized with an Applied Biosystems Model 380A DNA synthesizer.

A unique *Xba*I restriction site was created by changing a T residue at nt 4575 (−4 from the 3′ splice site; see Figure 1) into a C, to generate the mutant pSVx. pSVdx, which contains a deletion of the polypyrimidine sequence from −5 to −15, in addition to a T to C transition at −4, was constructed in a similar manner. To replace the polypyrimidine sequence, pSVdx was cleaved by *Xba*I, blunt-ended with S1 nuclease, and then ligated with synthetic double-stranded 15mers, which correspond to the sequences from −15 to −1 of the 3′ splice sites in the first intron of human β-globin gene and in the adenovirus 2 late mRNA leader sequence, and also to a 'perfect' polypyrimidine sequence, CTTTCTTTCTTTCAG. Insertion of the 15mers used in the construction of pSVd-L and pSVd-M in the opposite orientation resulted in plasmids pSVd-Li and pSVd-Mi. To generate the insertion mutants pSV-Li and pSV-Mi, pSVx was cleaved with *Xba*I, blunt-ended with S1 nuclease and then ligated with the appropriate synthetic 15mers. *Hin*dIII DNA fragments (SV40 nt 4002−5171) from the above mutants were cloned into the vector SP64 (Melton *et al.*, 1984) to create the pSP-series of plasmids. The sequences of all mutants described here were verified by DNA sequence analysis (Maxam and Gilbert, 1980), and are indicated in Figure 1.

Cell transfection and RNA analysis

Transfection of human HeLa and 293 cells was performed as previously described (Fu and Manley, 1987). Based on the results of multiple experiments, transfection efficiencies were estimated to vary by at most 2-fold. RNA was extracted from the cytoplasm and the nucleus and analyzed by S1 nuclease mapping using 5′ or 3′ end-labelled DNA probes (Weaver and Weissmann, 1979) as described previously (Fu and Manley, 1987). The protected products were analyzed under denaturing conditions on 5% acrylamide−8 M urea sequencing-type gels. Autoradiograms produced without the use of intensifying screens were scanned with a Gilford model 250 densitometer.

In vitro splicing and analysis of RNA products

HeLa nuclear extracts were routinely prepared by the procedure of Dignam *et al.* (1983), with minor modifications (Noble *et al.*, 1987). Capped precursor RNAs were produced from the pSP series of plasmids using SP6 RNA polymerase (Melton *et al.*, 1984). Run-off transcripts extended to the *Sty*I site (SV40 nt 4409). *In vitro* splicing reactions were carried out at 30°C for 40 min or 3 h in 25 μl volume using conditions previously described (Noble *et al.*, 1986, 1987). Following RNA extraction, splicing products and intermediates were analyzed directly by electrophoresis through 6% acrylamide−8 M urea denaturing gels.

Acknowledgements

We thank M.Kapczynski for technical assistance and W.Weast for preparing the manuscript. We are also grateful to J.Noble for discussions and comments on the manuscript. This work was supported by NIH grant CA33620.

References

Aebi,M., Hornig,H., Padgett,R.A., Reiser,J. and Weissmann,C. (1986) *Cell*, **47**, 555−565.
Amara,S.G., Jonas,V., Rosenfeld,M.G., Ong,E.S. and Evans,R.M. (1982) *Nature*, **298**, 240−244.
Berget,S.M. and Robberson,B.L. (1986) *Cell*, **46**, 691−696.
Bernstein,S.I., Hansen,C.J., Becker,K.D., Wassenberg D.R.,II, Roche, E.S., Donady,J.J. and Emerson,C.P., Jr (1986) *Mol. Cell. Biol.*, **6**, 2511−2519.
Black,D.L. and Steitz,J.A. (1986) *Cell*, **46**, 697−704.
Breathnach,R. and Chambon,P. (1981) *Annu. Rev. Biochem.*, **50**, 349−383.
Chabot,B. and Steitz,J.A. (1987) *Mol. Cell. Biol.*, **7**, 281−293.
Chabot,B., Black,D.L., LeMaster,D. and Steitz,J.A. (1985) *Science*, **230**, 1344−1349.
Dignam,J.D., Lebovitz,R.M. and Roeder,R.G. (1983) *Nucleic Acids Res.*, **11**, 1475−1489.
Eveleth,D.D., Gietz,R.D., Spencer,C.A., Nargang,F.E., Hodgette,R.B. and Marsh,J.L. (1986) *EMBO J.*, **5**, 2663−2672.
Falkenthal,S., Parker,V.P. and Davidson,N. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 449−453.
Fradin,A., Jove,R., Hemenway,C., Keiser,H.B., Manley,J.L. and Prives,C. (1984) *Cell*, **37**, 927−936.
Frendewey,D. and Keller,W. (1985) *Cell*, **42**, 355−367.
Fu,X.-Y. and Manley,J.L. (1987) *Mol. Cell. Biol.*, **7**, 738−748.
Gerk,V. and Steitz,J.A. (1986) *Cell*, **47**, 973−984.
Gluzman,Y. (1981) *Cell*, **23**, 175−182.
Keller,E.B. and Noon,W.A. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 7417−7420.
Ladner,M.B., Martin,G.A., Noble,J.A., Nikoloff,D.M., Tal,R., Kawasaki, E. and White,T.J. (1987) *EMBO J.*, **6**, 2693−2698.
Langford,C.J. and Gallwitz,D., (1983) *Cell*, **33**, 519−527.
Langford,C.J., Klintz,F.J., Donath,C. and Gallwitz,P. (1984) *Cell*, **36**, 645−653.
Lewis,E.D. and Manley,J.L. (1985) *Nature*, **317**, 172−175.
Manley,J.M., Noble,J.C.S., Chaudhuri,M., Fu,X.-Y., Michaeli,T. and Prives,C. (1986) *Cancer Cells 4*. Cold Spring Harbor Laboratory Press,

New York, pp. 259–265.
Manley,J.L., Fu,X.-Y., Noble,J.C.S. and Ge,H. (1987) In Inouye,M. and Dudock,B. (eds), *Molecular Biology of RNA*. Academic Press, New York, pp. 97–112.
Maxam,A.M. and Gilbert,W. (1980) *Methods Enzymol.*, **65**, 499–560.
Melton,D.A., Krieg,P.A., Rebaghiati,M.R., Maniatis,T., Zinn,K. and Green,M.R. (1984) *Nucleic Acids Res.*, **12**, 7035–7056.
Mitchell,P.J., Urlaub,G. and Chasin,L. (1986) *Mol. Cell. Biol.*, **6**, 1926–1935.
Morgan,B.A., Johnson,W.A. and Hirsh,J. (1986) *EMBO J.*, **5**, 3335–3342.
Mount,S.M. (1982) *Nucleic Acids Res.*, **10**, 459–472.
Nawa,H., Kotani,H., Nakanishi,S. (1984) *Nature*, **312**, 729–734.
Noble,J.C.S., Prives,C. and Manley,J.L. (1986) *Nucleic Acids Res.*, **14**, 1219–1235.
Noble,J.C.S., Pan,Z.-Q., Prives,C. and Manley,J.L. (1987) *Cell*, **50**, 227–236.
Norris,K., Norris,F., Christiansen,L. and Fiil,N. (1983) *Nucleic Acids Res.*, **11**, 5103–5112.
Padgett,R.A., Konarska,M.M., Aebi,M., Hornig,H., Weissmann,C. and Sharp,P.A. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 8349–8353.
Pikielny,C.W., Teem,J.L. and Rosbash,M. (1983) *Cell*, **34**, 395–403.
Rozek,C.E. and Davidson,N. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 3128–3132.
Ruskin,B. and Green,M.R. (1985a) *Nature*, **317**, 732–734.
Ruskin,B. and Green,M.R. (1985b) *Science*, **229**, 135–140.
Ruskin,B., Green,J.M. and Green,M.R. (1985) *Cell*, **41**, 833–844.
Ruskin,B., Pikielny,C.W., Rosbash,M. and Green,M.R. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 2022–2026.
Schwarzbauer,J.E., Patel,R.S. Fonda,D. and Haynes,R.O. (1977) *EMBO J.*, **6**, 2573–2580.
Tazi,J., Alibert,C., Temsamani,J., Reveilland,I., Cathala,G., Brunel,C. and Jeanpeur,P. (1986) *Cell*, **47**, 755–766.
Tooze,J. (1981) *DNA Tumor Virus*, 2nd edn. Cold Spring Harbor Laboratory Press, New York.
Treisman,R., Proudfoot,N.J., Shander,M. and Maniatis,T. (1982) *Cell*, **29**, 903–911.
Treisman,R., Orkin,S.H. and Maniatis,T. (1983) In Stamatoyannopoulos,G. and Nienhuis,A.W. (eds), *Globin Gene Expression and Hematopoietic Differentiation*. Alan R.Liss, New York, pp. 99–121.
Ulfendahl,P.J., Petterson,U. and Akusjarvi,G. (1985) *Nucleic Acids Res.*, **13**, 6299–6315.
van Santen,L.V. and Spritz,R.A. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 2885–2889.
Weaver,R.F. and Weissmann,C. (1979) *Nucleic Acids Res.*, **7**, 1175–1193.
Wieringa,B., Meyer,F., Reiser,J. and Weissmann,C. (1983) *Nature*, **301**, 38–43.
Wieringa,B., Hofer,E. and Weissmann,C. (1984) *Cell*, **37**, 743–751.
Zhuang,Z., Leung,H. and Weiner,A.M. (1987) *Mol. Cell. Biol.*, **7**, 3018–3020.
Zoller,M.J. and Smith,H. (1983) *Methods Enzymol.*, **100**, 468–500.

Received on December 15, 1987