Sequences involved in the control of Adenovirus L1 alternative RNA splicing

Jan-Peter Kreivi, Kenn Zerivitz and Göran Akusjärvi*
Department of Microbial Genetics, Medical Nobel Institute, Karolinska Institute, Box 60400, 104 01 Stockholm, Sweden

Received January 25, 1991; Revised and Accepted April 2, 1991

ABSTRACT

During an adenovirus infection the expression of mRNA from late region L1 is temporally regulated at the level of alternative 3′ splice site selection to produce two major mRNAs encoding the 52,55K and IIIa polypeptides. The proximal 3′ splice site (52,55K) is used at all times of the infectious cycle whereas the distal site (IIIa) is used exclusively late after infection. We show that a single A branch nucleotide located at position −23 is used in 52,55K splicing and that two A's located at positions −21 and −22 are used in IIIa splicing. Both 3′ splice sites were active *in vitro* in nuclear extracts prepared from uninfected HeLa cells. However, the efficiency of IIIa splicing was only approximately 10% of 52,55K splicing. This difference in splice site activity correlated with a reduced affinity of the IIIa, relative to the 52,55K, 3′ splice site for polypyrimidine tract binding proteins. Reversing the order of 3′ splice sites on a tandem pre-mRNA resulted in an almost exclusive IIIa splicing indicating that the order of 3′ splice site presentation is important for the outcome of alternative L1 splicing. Based on our results we suggest a *cis* competition model where the two 3′ splice sites compete for a common RNA splicing factor(s). This may represent an important mechanism by which L1 alternative splicing is regulated.

INTRODUCTION

A large number of cellular and viral genes which produce multiple cytoplasmic mRNAs by alternative splicing of a single pre-mRNA have been characterized (reviewed in ref. 1). In several cases the accumulation of alternatively spliced mRNAs have been shown to be regulated either in a temporal or tissue-specific manner. Several patterns of alternative splicing are possible; multiple 5′ splice sites may be paired with a single 3′ splice site, a single 5′ splice site may be paired with alternative 3′ splice sites or both ends of an exon may be bypassed leading to exon skipping. Very little is known how alternative RNA splicing is regulated. Several parameters such as RNA secondary structure (2, 3, 4), relative strength of 5′ splice sites (5, 6), pyrimidine content of 3′ splice sites (7, 8), presence of unique trans-acting factors (9, 10, 11) or *cis* competition for a limiting splicing factor (12, 13, 14) have been shown to influence the choice of alternative splice sites on complex transcripts. However, a coherent picture how alternative splicing is regulated is still lacking.

We are using human adenovirus type 2 (Ad2) late region one (L1) as a model system to study the mechanism of alternative splicing. All adenovirus transcription units encode differentially spliced mRNAs. The accumulation of adenovirus mRNA is subjected to a temporal regulation during the infectious cycle; one set accumulates early after infection whereas a new set is produced at late times (reviewed in ref. 15). L1 belongs to the so called major late transcription unit (Fig. 1A). This unit encodes more than 20 cytoplasmic mRNAs which are grouped into five families, L1 to L5, where each family consists of multiple alternatively spliced mRNAs with coterminal 3′-ends. Through the processing pathway, the vast majority of late mRNAs receive three short leader segments, the tripartite leader, at their 5′-ends (Fig. 1B) (reviewed in ref. 15).

L1 is an alternatively spliced gene where one 5′ splice site is joined to two different 3′ splice sites. During the early phase of the infectious cycle only the proximal 3′ splice site is active, resulting in an almost exclusive production of the 52,55K mRNA. At late times the distal 3′ splice sites becomes active. This leads to the production of approximately equal amounts of two predominant cytoplasmic mRNAs; the 52,55K and the IIIa mRNAs (16, 17, 18). Virus induced or virus-encoded factors have been implicated in the shift from the early to late pattern of L1 pre-mRNA splicing (see ref.19 and refs. therein). Transient transfection studies have also suggested that IIIa splicing may be negatively regulated (20).

To study the mechanism and regulation of L1 pre-mRNA splicing we are using an *in vitro* approach in which SP6 transcripts containing various combinations of the 52,55K and IIIa 3′ splice sites are processed in HeLa cell nuclear extracts. Our results show that the IIIa 3′ splice site is active in the absence of adenovirus regulatory proteins. However, it is much weaker than the 52,55K 3′ splice site. This difference in splice site strength appears to correlate with a reduced capacity of IIIa to associate with 3′ splice site binding factors. Furthermore, we show that the order of splice site presentation is very important for the outcome of alternative L1 splicing. Our results suggest a mechanism where a *cis*

* To whom correspondence should be addressed

competition between the 52,55K and the IIIa 3′ splice sites for RNA splicing factors is, at least in part, responsible for the exclusive synthesis of the 52,55K mRNA early after infection.

MATERIALS AND METHODS

Plasmid constructions

Nucleotide numbers refer to the Ad2 sequence presented by Roberts et al. (21). The structure of the four SP6 templates used for pre-mRNA synthesis is shown in figure 2. pGD52,55K was constructed by cloning the common third leader 5′ splice site (position 9723) and the L1 52,55K 3′ splice site (position 11040) into the HindIII site of pGEM-3. It contains a leader 2, 3 cDNA (22) [ScaI (position 7138 in the second leader) converted to HindIII] fused to genomic sequences encoding the L1 intron and the 5′-part of the 52,55K exon (to HindIII, position 11555). The plasmid was constructed so that the intron was truncated from 1300 to 152 nucleotides by removing genomic Ad2 sequences between positions 9765 (BbvI) and 10937 (NaeI). The end-points of the deletion were fused by HindIII linker addition. Plasmid pGDIIIa is identical to pGD52,55K except that the NaeI-HindIII fragment (positions 10938−11555) encoding the 52,55K 3′ splice site were replaced by a PvuII [position 12178 (converted to HindIII)]-HindIII (position 13636) fragment encoding the IIIa 3′ splice site region. The 3′ duplication plasmid pGD52,3A was created by replacing the ApaI-XbaI fragment in pGD52,55K with the HindIII (converted to ApaI)-AvrII fragment spanning the IIIa 3′ splice site region from pGD3A (Fig. 2). Similarily pGD3A,52 was constructed by cloning the HindIII fragment spanning the 52,55K 3′ splice site region in pGD52,55K (Fig. 2) as an MluI fragment in the unique MluI site in pGD3A.

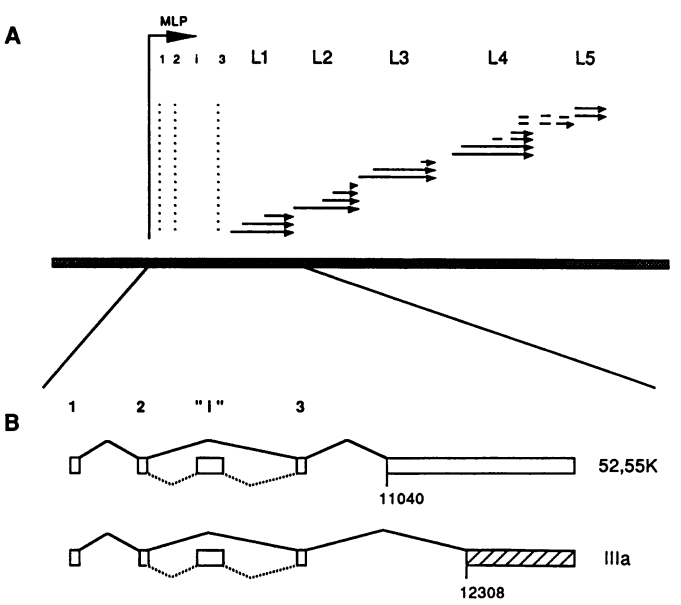

Figure 1. (A) A schematic representation of mRNAs expressed from the Ad2 major late transcription unit late after infection (12−48 hrs). The mRNAs are divided into five families (L1−L5) were each family consist of species with coterminal 3′-ends and a common set of 5′-leaders (1, 2, *i*, 3) spliced onto the RNA body. The arrow indicate direction of transcription. (B) Enlargement illustrating the structure of the two major mRNAs expressed from region L1 Boxes represents exon sequences and thin lines the splicing pathways. Nucleotide positions of the 52,55K and the IIIa 3′ splice sites are indicated. The dotted lines illustrates that a small fraction of late mRNA contain the *i*-leader segment spliced in between leader 2 and 3.

SP6 transcription and *in vitro* splicing

Linearized DNA templates were used for synthesis of capped ^{32}P-labelled SP6 transcripts (23). The reactions were incubated at 40°C for 1 h in a final volume of 50 μl containing 0.5 mM ATP, UTP and m7GpppG (Pharmacia), 0.05 mM CTP and GTP, and 50 μCi of ^{32}p-CTP (800 Ci/mmol, Amersham). Full length transcripts were cut out from a wet 6% polyacrylamide-urea gel and recovered by electroelution (24). Nuclear extracts were prepared from HeLa spinner cells as described by Dignam et al., (25) except that buffer C, contained 0.6 M KCl instead of 0.42 M NaCl (26). Splicing reactions were incubated at 30°C in a total volume of 25 μl containing 40% nuclear extracts, 2.6% polyvinylic alcohol, 12% glycerol, 60 mM HEPES (pH 7.9), 3.2 mM $MgCl_2$, 0.5 mM ATP, 0.5 mM EDTA, 0.3 mM DTT and 10−30 fmol pre-mRNA.

Analysis of processing products

Reactions were terminated by adding 175 μl of a solution containing 1% SDS and 200 μg proteinase-K per ml and further incubating the mixtures at 30°C for 30 min. RNA was then recovered as described (27). Splicing products were distinguished by their mobilities compared to a DNA-marker on a 8% polyacrylamide-urea gel. Splicing intermediates were identified by their disappearance in time course experiments and lariat RNAs by their sensitivity to treatment with debranching extracts (28, 29). Processing products were also verified by primer extension.

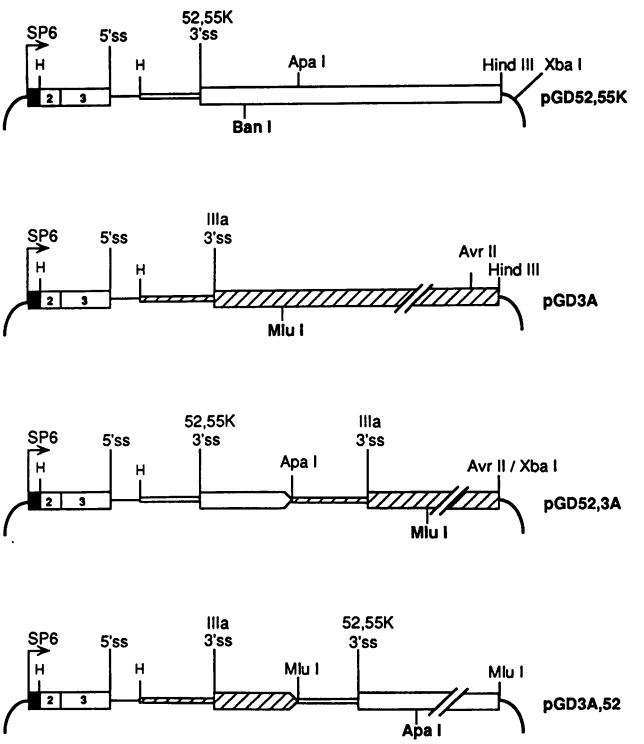

Figure 2. Structure of constructs used for SP6 pre-mRNA synthesis. Boxes denotes exon sequences and thin lines introns sequences (white 52,55K; hatched IIIa). The position of restriction enzyme recognition sites discussed in the text are indicated by their full names (authentic sites) or capital letters (sites constructed by linker addition). The position of restriction sites used for truncation of plasmid DNA for SP6 transcription are shown in bold below each plasmid map. H; Hind III.

Branch point mapping

Exon 2-intron lariat intermediates were located by autoradiography on the wet polyacrylamide-urea gel, recovered by electroelution (24) and ethanol precipitated in the presence of carrier RNA. Approximately one half of the intermediate preparation was subjected to debranching as described by Krämer and Keller (29), except that no $MgCl_2$ was added. Branch points were determined by comparing the primer extension products on debranched exon 2-intron lariat RNAs and untreated lariat intermediates. Two oligodeoxynucleotide primers were used for this analysis. The 52,55K branch point was mapped with a 25-mer (5′-TGCATGTCTGCCGCTGCTCTTGCTC-3′) complementary to nucleotides 11088 to 11112 on the Ad2 genome. The IIIa branch point was mapped with a 24-mer (5′-CCGGGT-CCGTTGCGTCTTGCATCA-3′) complementary to nucleotides 12309 to 12332. Lariat RNAs or debranched intermediates were coprecipitated together with 5′-end labelled oligonucleotide primers and resuspended in 10μl of a hybridization buffer containing 80% (vol/vol) formamide, 0.4 M NaCl, 50 mM PIPES (pH 6.4) and 1 mM EDTA. After denaturation at 80°C for 10 min., hybridization mixtures were annealed at room temperature over night. Hybrids were recovered by ethanol precipitation and resuspended in 10 μl of a buffer containing 0.1 M Tris-HCl (pH 8.3), 16 mM $MgCl_2$, 60 mM KCl and 2 mM DTT. 10 μl of a buffer containing 4 mM dNTP and 7.5 U AMV polymerase was added and the reactions incubated for 1 hr at 42°C. Reactions were terminated by the addition of EDTA to a final concentration of 25 mM. This was followed by phenol extraction and recovery of the RNA by ethanol precipitation. To degrade the radiolabelled RNA, pellets were dissolved in 50μl 0.1 M NaOH and incubated at 65°C for 1 hr. The solutions were neutralized with HCl, precipitated and the products analyzed on a denaturing polyacrylamide gel. A sequence ladder was obtained by dideoxy sequencing of the parental plasmid as described by the manufacturer (Promega).

Gel shift analysis

10 μl splicing reactions containing the indicated concentrations of polyU (Fig. 7) were mixed. The ^{32}P-labelled RNA was added last and the tubes were incubated at 30°C for 30 min. Complex formation (30) was terminated by addition of 1 μl of heparin (10 mg/ml) followed by freezing in liquid nitrogen. After thawing at room temperature, 3.5 μl aliquots were loaded onto a 4% (80:1) native polyacrylamide gel, cast and run in a buffer consisting of 50 mM Tris-Glycine (30). The gel was pre-electrophoresed at 200V for 1h.

RESULTS

Spliced structure of the IIIa mRNA

Surprisingly little is known about the structure of the IIIa mRNA. By electron microscopy the body of the mRNA has been mapped between coordinates 33.9 and 39 on the Ad2 genome (31) and by S1 mapping a potential 3′ splice site region has been defined (20). To design an experimental approach for the analysis of alternative L1 splicing we first had to verify, at the nucleotide level, the position of the IIIa 3′ splice site. By a combination of S1 cleavage and primer extension we determined the structure of the IIIa mRNA in late (24 hours post infection) cytoplasmic RNA isolated from Ad2 infected HeLa cells (data not shown). The analysis showed that the majority of IIIa mRNAs carry the common 203 nucleotide tripartite leader sequence (32, 33). A minor fraction of the IIIa mRNAs contained an extended 5′-non translated region consisting of the tripartite leader with the 440

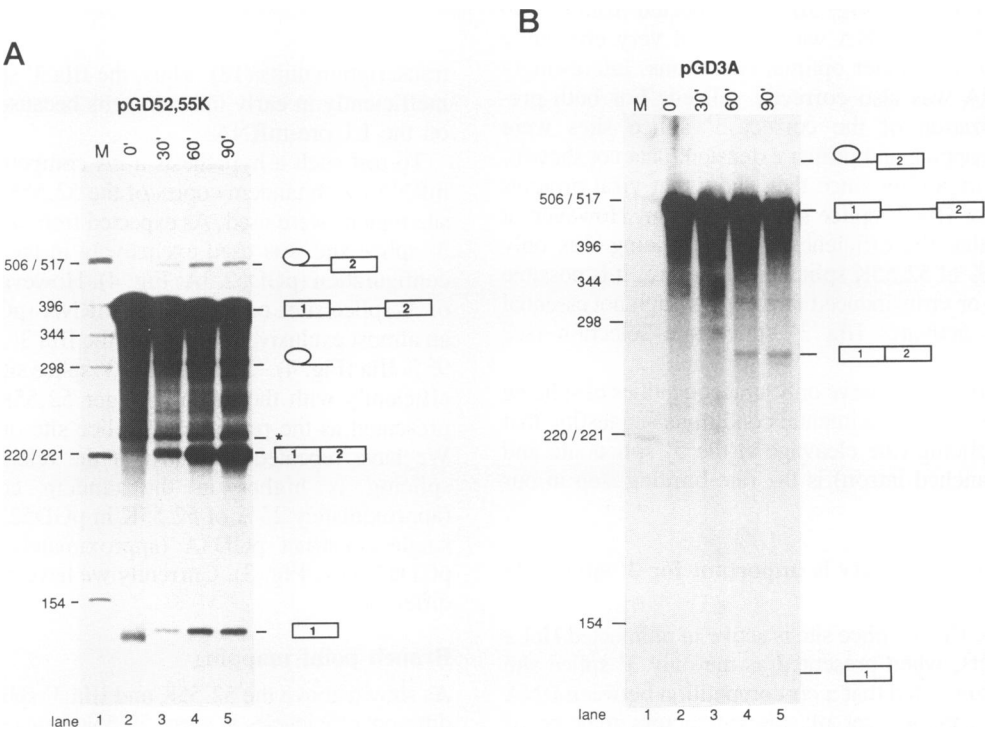

Figure 3. Time course (min.) of formation of spliced products from L1 pre-mRNAs pGD52,55K (panel A) and pGD3A (panel B). The position of reaction products and intermediates are indicated. Boxes represent exon sequences and thin lines intron sequences. The size of ^{32}P-labelled pBR marker fragments are indicated at the left. * denotes an endogenous band due to unspecific cleavage of the pGD52,55K pre-mRNA.

nucleotide long 'i'-leader segment spliced in between the second and third leader segments (34, 35). Sequence analysis of primer extension products demonstrated that both types of mRNAs used a 3′ splice site, located at position 12308 in IIIa splicing.

L1 gene constructs used for pre-mRNA synthesis

Based on the established anatomy of region L1 four SP6 DNA templates were constructed (Fig. 2). A cDNA of the tripartite leader was used in these constructs because previous *in vivo* studies have indicated that major late pre-mRNA splicing may proceed through a two step reaction (36, 37, 38). In the first reaction the tripartite leader is assembled and in the second reaction the common leader is joined to one of the many mRNA bodies within the major late transcription unit (Fig. 1A). To mimic the natural situation as much as possible the third leader 5′ splice site was fused to either the 52,55K or the IIIa 3′ splice site regions, in separate, tandem or reversed tandem constructs (Fig. 2).

In vitro splicing of 52,55K and IIIa pre-mRNAs

The lack of IIIa splicing during the early phase of infection might be explained by several mechanisms. For example, an early viral protein may inhibit the use of the IIIa 3′ splice site. Alternatively, a late viral protein, or virus-induced factor may be required to activate the IIIa 3′ splice site at late times. Since only 52,55K mRNA and no detectable IIIa mRNA is made in transient expression assays (20; unpublished results) it seems unlikely that the lack of IIIa splicing in early infected cells is due to negative regulation of IIIa 3′ splice site utilization by an early viral protein.

To test the hypothesis that IIIa splicing is dependent on a late viral protein or virus-induced host factor, nuclear extracts were prepared from uninfected HeLa cells and used to process SP6 transcripts encoding either a 52,55K (pGD52,55K; Fig. 3A) or a IIIa pre-mRNA (pGD3A; Fig. 3B). As expected from *in vivo* results, the 52,55K pre-mRNA was processed very efficiently when incubated *in vitro* under optimal conditions. Interestingly the IIIa pre-mRNA was also correctly spliced. For both pre-mRNAs the utilization of the correct 3′ splice sites were confirmed by S1 mapping and primer extension (data not shown). These results are important since they show that viral proteins are not essential for IIIa 3′ splice site recognition. However, it should be noted that the efficiency of IIIa splicing was only approximately 10% of 52,55K splicing. Therefore, it is possible that virus encoded or virus-induced factors, although not essential for IIIa splicing, activates IIIa 3′ splice site selection (see discussion).

It is noteworthy that we observe only small quantities of splicing intermediates under our experimental conditions suggesting that the first step in splicing (the cleavage at the 5′ splice site and formation of a branched intron) is the rate limiting step in our reactions.

5′ and 3′ splice site proximity is important for 3′ splice site choice

The finding that the IIIa 3′ splice site is active in uninfected HeLa cell nuclear extracts, when presented as the only 3′ splice site on a pre-mRNA, suggested that a *cis* competition between RNA splice sites for a common set of splicing factors may be of importance for the lack of IIIa splicing in early infected cells. Previous studies have, indeed, shown that RNA splice site proximity is important for 3′ splice site selection on complex transcription units (12). Thus, the IIIa 3′ splice site may be used inefficiently in early infected cells because of its distal location on the L1 pre-mRNA.

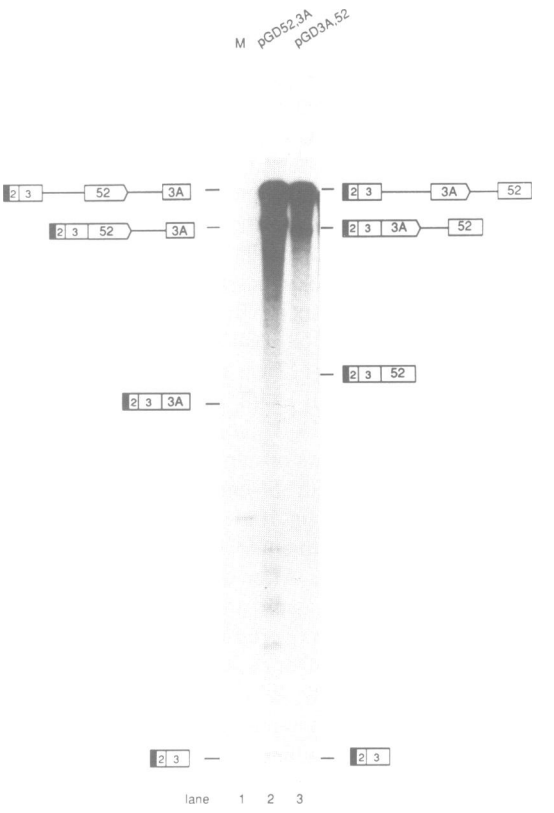

Figure 4. Splice site selection in pre-mRNAs containing tandem duplications of the 52,55K and IIIa 3′ splice sites. Uniformely ^{32}P-labelled pGD52,3A and pGD3A,52 pre-mRNAs were spliced for 2 hrs at 30°C. The resulting products were resolved on a 4% denaturing polyacrylamide gel. The position of pre-mRNAs and expected reaction products are indicated. Boxes represents exon sequences and thin lines intron sequences.

To test such a hypothesis a *cis* competition assay using pre-mRNAs with tandem copies of the 52,55K and the IIIa 3′ splice site regions were used. As expected from *in vivo* data, the 52,55K 3′ splice site was used exclusively in the natural 3′ splice site configuration (pGD52,3A; Fig. 4). However, reversing the order of 3′ splice sites on the L1 pre-mRNA (pGD3A,52) resulted in an almost exclusive selection of the IIIa 3′ splice site; more than 95% IIIa (Fig. 4). Thus, the IIIa 3′ splice site was able to compete efficiently with the much stronger 52,55K 3′ splice site when presented as the proximal 3′ splice site on the L1 pre-mRNA. We have repeatedly found that the relative efficiency of IIIa splicing is higher in the tandem construct pGD3A,52 (approximately 25% of 52,55K in pGD52,3A) compared to the single construct pGD3A (approximately 10% of 52,55K in pGD52,55K; Fig. 3). Currently we have no explanation for this difference.

Branch point mapping

As shown above the 52,55K and IIIa 3′ splice sites are used with different efficiencies in vitro; 52,55K being approximately 10-fold more active compared to IIIa in single constructs (Fig. 3 and data not shown). To begin to better understand the differences in splice site efficiency, we established the precise architecture

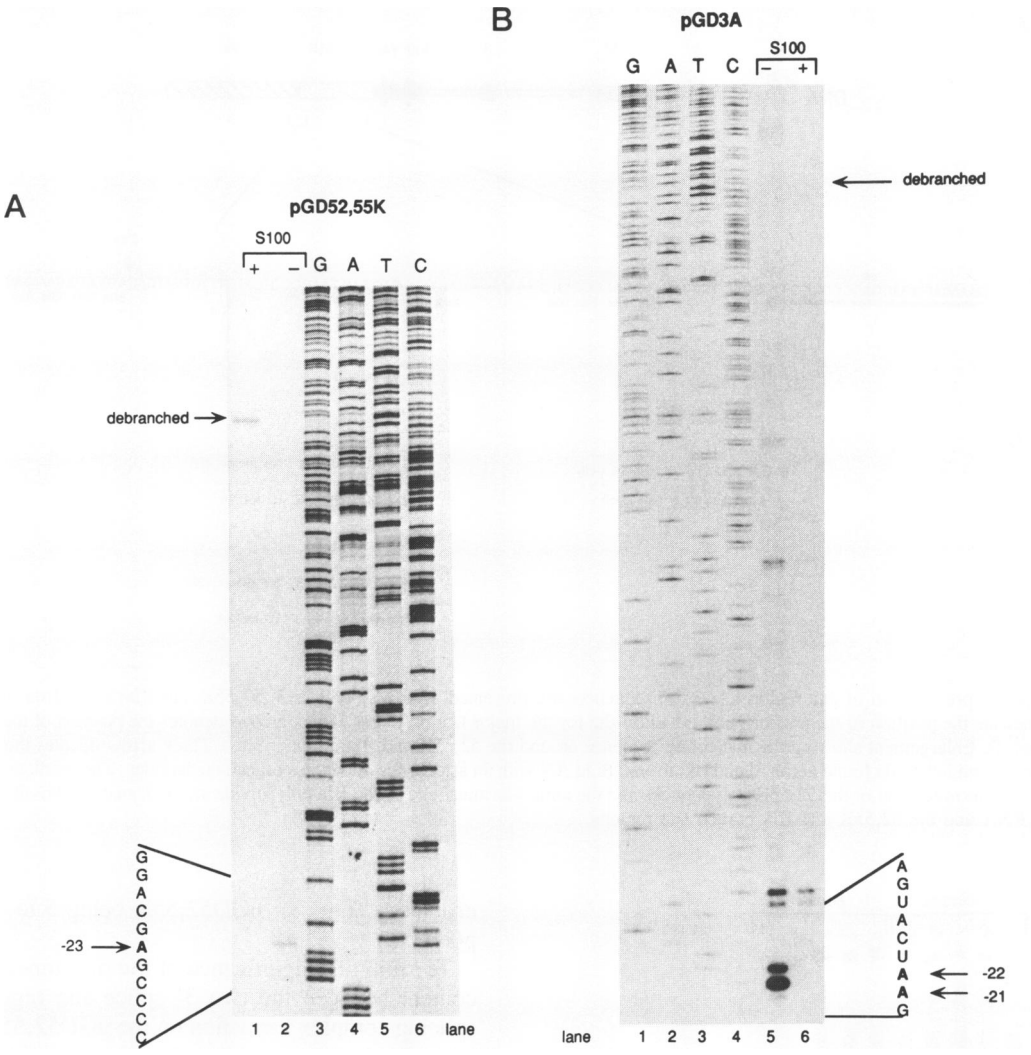

Figure 5. Branch point mapping. Exon-2 intron intermediates were isolated from preparative splicing reactions using the pGD52,55K (panel A) and pGD3A (panel B) pre-mRNAs. Primer extension was performed on lariat (−S100) or debranched intermediates (+S100). The primer extension products were separated on an 8% denaturing polyacrylamide gel in parallel with the corresponding dideoxy sequencing ladder from the parental DNA.

of the 52,55K and IIIa 3′ splice site regions. For this analysis preparative amounts of exon 2-lariat intermediates were isolated from pGD52,55K and pGD3A pre-mRNA splicing reactions. The branch points were mapped by comparing cDNA extension products of untreated or debranched exon 2-lariat intermediate RNAs. The position of the S100 sensitive reverse transcription stop sites were mapped against the appropriate DNA sequence ladder.

Primer extension analysis on pGD52,55K exon 2-intron lariat RNAs identified a single strong stop site (Fig. 5A lane 2). This band was not detected in unprocessed pre-mRNA (data not shown) or in debranched exon-2 intron lariat RNA (Fig. 5A, lane 1). In the debranched sample a single species with the expected mobility of the linear 52,55K intron-exon 2 RNA was detected. Based on this type of analysis we conclude that a single A residue located 23 nucleotides upstream of the 52,55K 3′ splice site (at position 11017; 21) is used for lariat formation (Fig. 6).

Primer extension on pGD3A exon 2-intron lariat RNAs identified two strong stop signals located 22 and 23 nucleotides upstream of the IIIa 3′ splice site (Fig. 5B, lane 5); in a ratio of approximately 2 to 1, respectively. These bands were not present in debranched samples (Fig. 5B, lane 6) or in unprocessed IIIa pre-mRNA (data not shown). The reverse transcriptase stop signals mapped the IIIa branch site to two A residues located at position −22 and −23 (position 12286 and 12287; 21) relative to the IIIa 3′ splice site.

Difference in pre-splicing complex formation between the 52,55K and the IIIa 3′ splice sites

One of the earliest specific interactions taking place during spliceosome assembly is the definition of the 3′ splice site by the formation of a stable complex between U2 snRNP and the branch point region (reviewed in ref. 39). This interaction may, in fact, be the rate limiting step in spliceosome assembly (40). Therefore the difference in 52,55K and IIIa splicing activity (Fig. 3) could be due to the differences in 3′ splice site composition (Fig. 6B). Efficient U2 snRNP binding to the branch point region requires at least one soluble factor, U2AF (41), which binds to the polypyrimidine tract 3′ splice site region of the pre-mRNA.

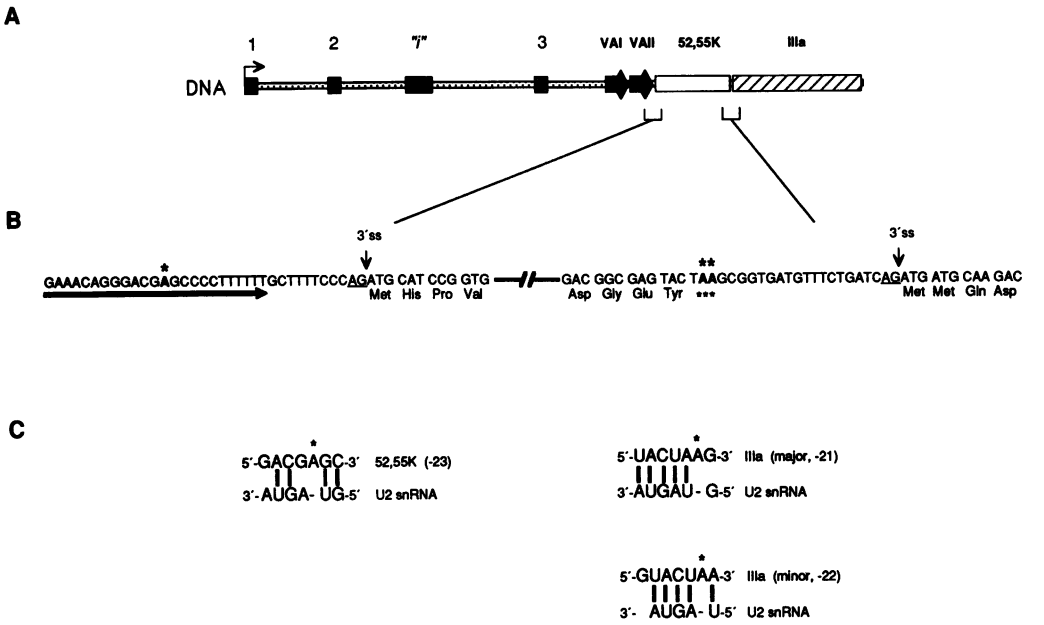

Figure 6. A. A graphic representation of late region L1. Exon sequences are presented as boxes (1, 2, *i*, 3, 52,55K and IIIa) and intron sequences as dotted thin boxes. Thin arrow indicate the position of the transcription initiation site for the major late promoter. Thick arrows denotes the position of the adenovirus VA RNAI and VA RNAII genes. B. Enlargement showing the nucleotide sequence around the 52,55K and IIIa 3′ splice sites. Thick arrow denotes the 3′ end of VA RNAII. *AG* denotes the conserved dinucleotide found at the 3′ end of introns. Bold A's with an asterix above denotes branch nucleotides. The predicted amino-acid sequences at the amino and carboxy terminal end of the 52,55K polypeptides and the amino-terminal end of the IIIa polypeptide are indicated. C. Possible base pairing between the 3′ end of U2 snRNA and the 52,55K and IIIa branch site regions.

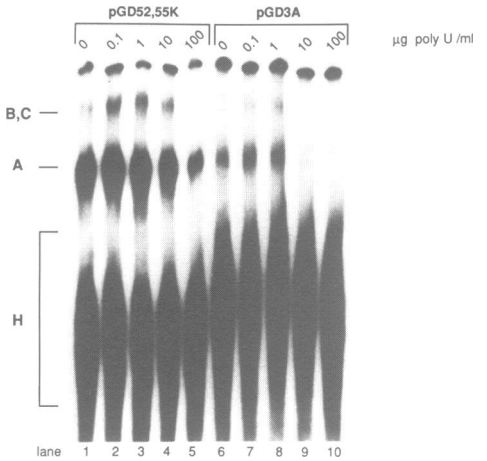

Figure 7. Stable splicing complex formation on 52,55K and IIIa 3′ splice sites. Uniformly ^{32}P-labelled pGD52,55K and pGD3A pre-mRNAs were incubated for 30′ at 30°C under *in vitro* splicing conditions with various amount of polyU (concentration µg polyU per ml noted above each lane). Splicing complexes were resolved on a 4% native polyacrylamide gel.

Since the IIIa 3′ splice site lacks an extended polypyrimidine tract, the relative inefficiency of IIIa splicing (Fig. 3B) could be manifested at the level of stable U2 snRNP/branch point complex formation. To examine this possibility, spliceosome complex formation with ^{32}P-labelled pGD52,55K and pGD3A pre-mRNAs was analyzed by native gel electrophoresis (30). As shown in Fig. 7 (lanes 1 and 6) the difference in stable complex A formation on the two pre-mRNAs paralleled the splicing efficiency (Fig. 3); pGD52,55K being 5 to 10-fold better than pGD3A.

To study the significance of the difference in polypyrimidine content between the two 3′ splice site regions (Fig. 6), pre-splicing complex formation on the pGD52,55K and the pGD3A pre-mRNAs were competed with increasing amounts of cold polyU. As shown in Fig. 7 complex A formation on pGD3A (lanes 6 to 10) was abolished already at 10 µg polyU per ml whereas a 10-fold higher concentration of polyU was necessary to cause a significant reduction in complex A formation on pGD52,55K (lanes 1 to 5). Taken together these results suggest that the relative inefficiency of IIIa splicing could be due to an inability of the IIIa 3′ splice site to efficiently form stable complexes with U2 snRNP. This deficiency appears to be due to a reduced affinity of the IIIa 3′ splice site for polypyrimidine tract binding proteins.

DISCUSSION

With the characterization of the spliced structure of the IIIa mRNA and the mapping of the branch points for both the 52,55K and the IIIa 3′ splice site regions, we now have a more detailed understanding of the structural organization of sequences involved in the control of alternative L1 pre-mRNA splicing. The branch nucleotides are located at about the same distance from the 3′ splice site in both 52,55K (−23) and IIIa (−22 and −23); only A's were used as branch nucleotides. The 52,55K 3′ splice site region is characterized by an unusually long polypyrimidine tract (18 pyrimidines out of 19 nucleotides; Fig. 6B) and a branch site region which shows a poor homology to the proposed branch

point consensus sequence (42, 43, 44). In contrast the IIIa 3′ splice site region show an almost perfect homology with the branch site consensus but lacks an extended polypyrimidine tract (8 pyrimidines out of 11 nucleotides; Fig. 6B).

As has been noted previously the information content of the Ad2 genome is extremely high with few if any surplus nucleotides present (for a review see ref. 15). This is also exemplified in the organization of region L1 where the same nucleotide sequence is shown to serve several regulatory functions (Fig. 6). Adenovirus encodes for two low molecular weight RNA polymerase III products, the so called VA RNAI and VA RNAII (45). The VA RNA genes are encoded within the last intron of region L1, just upstream of the 52,55K 3′ splice site. As shown here the branch nucleotide used in 52,55K splicing (Fig. 5A) is located within the VA RNAII structural gene (46). Also, the T cluster functioning as the VA RNAII transcription termination signal is part of the polypyrimidine tract used for 52,55K splicing. Furthermore, the two A residues used as branch sites in IIIa splicing (Fig 5B) are part of the UAA stop codon terminating 52,55K protein synthesis.

The accumulation of IIIa mRNA is subjected to a temporal regulation during the infectious cycle; it is made only at late times of infection. Here we show that the sequences surrounding both the 52,55K and the IIIa 3′ splice sites are sufficient for accurate splicing. However, in similar constructs (Fig. 2) 52,55K splicing was approximately 10-fold better than IIIa splicing (Fig. 3). The relative inefficiency of IIIa splicing was manifested at the level of stable U2 snRNP binding (Fig. 7). IIIa was reduced 5 to 10-fold in complex A formation compared to 52,55K although the IIIa branch site can theoretically form a much better base pairing with U2 snRNA (Fig. 6C). Previous studies have shown that U2 snRNP binding to a 3′ splice site, in part, is defined by RNA-RNA base pairing between the 3′ end of U2 snRNA and the branch point sequence (47, 48, 49). Efficient U2 snRNP binding also requires at least one protein factor, U2AF which recognizes the 3′ splice site polypyrimidine tract (41). Most likely U2AF binds first to the polypyrimidine tract and directs the association of U2 snRNP to the highest affinity branch site in the vicinity of the 3′ splice site (49). Efficient U2AF binding appears to require an extended polypyrimidine tract preferably not interrupted by purines (41, 50). The 52,55K 3′ splice site contains a very long polypyrimidine tract (Fig. 6B) and would be expected to efficiently bind U2AF. In contrast the IIIa 3′ splice site contains a much shorter polypyrimidine tract which also is interrupted by three purines. Thus, the relative inefficiency of IIIa spliceosome assembly (Fig. 7, lane 6) and splicing (Fig. 3B) could be due to a reduced affinity of polypyrimidine tract binding factors for the IIIa 3′ splice site. This conclusion is supported by the observation that stable complex A formation on the IIIa pre-mRNA was outcompeted at significantly lower concentrations of polyU compared to the 52,55K pre-mRNA (Fig. 7). Although we have only considered U2AF above, it should be emphasized that several factors have been shown to recognize the polypyrimidine/3′ splice site region. For example, a second polypyrimidine tract binding protein, pPTB, which is also required for U2 snRNP/pre-mRNA complex formation and splicing has been identified (51, 52). Also the hnRNP C protein which binds preferentially to sequences typical of polypyrimidine tracts (53) or one of the several factors which require the AG dinucleotide at the 3′ splice site for binding (53, 54, 55) may contribute to the unequal recognition of the 52,55K and the IIIa 3′ splice sites.

RNA splice site proximity and RNA splice site strength (defined as the ability of a splice site to associate with splicing factors and participate in spliceosome assembly) have been shown to be major determinants in the selection of alternative 5′ and 3′ splice sites on complex pre-mRNAs. This has been demonstrated in several model experiments where tandem duplications of 5′ and 3′ splice sites have been used (12, 44, 56, 57). Although the IIIa 3′ splice site appears to be much weaker than the 52,55K 3′ splice site when assayed as a simple transcription unit (Fig. 3), it competes efficiently, both *in vitro* (Fig. 4) and *in vivo* (data not shown), with the 52,55K 3′ splice site when presented as the proximal site in a tandem construct. This result indicates that the lack of IIIa splicing in early infected cells, at least in part, may be due to the distal location of the IIIa 3′ splice site on the L1 pre-mRNA.

How then is the shift from 52,55K to IIIa 3′ splice site usage regulated during the infectious cycle? Our results are compatible with a model where all 3′ splice sites compete for a common set of splicing factors which are in excess in uninfected or early infected cells. This would probably lead to stable U2 snRNP binding to both the proximal and distal 3′ splice sites on the L1 pre-mRNA (58). Under such conditions the 52,55K 3′ splice site would be expected to be used preferentially in spliceosome assembly and splicing because of its closeness to the third leader 5′ splice site (12). Late in infection when large numbers of late transcripts are synthesized, these factor(s) may become limiting and thus allow U2 snRNP complexes formed on the IIIa 3′ splice site to compete more effectively for spliceosome assembly and splicing. Precedence for such a *cis* competition model comes from experiments where dilution of nuclear extracts (12) or addition of high concentrations of purified splicing factors (13, 14) have been shown to alter *in vitro* 3′ and 5′ splice site selection. In support for such a mechanism we have also shown that in wild type Ad2 infected cells, IIIa splicing can be activated by inducing synthesis of large amounts of L1 transcripts in the absence of late protein synthesis (Larsson et al. in preparation).

In one report SV40 large T antigen was suggested be able to activate IIIa splicing in transient expression experiments (20). The significance of this observation is unclear. The plasmids used in that study contained the SV40 origin and were therefore competent for DNA replication. The possibility exists that the observed shift to IIIa splicing by providing SV40 large T in trans was due to a *cis* competition mechanism via a replication dependent increase in L1 transcripts rather than a direct involvement of large T in alternative L1 3′ splice site selection.

The 52,55K and the IIIa mRNAs accumulate to about a 1:1 ratio in late infected cells. Since the IIIa 3′ splice site is intrinsically much weaker than the 52,55K 3′ splice site (Fig. 3), a *cis* competition model alone cannot satisfactorily explain the abundance of the individual L1 mRNAs late during the infectious cycle. Therefore it does not seem unreasonable to postulate that several mechanisms are simultaneously in operation. Several studies have provided indirect evidence that newly synthesized or activated factor(s) may contribute to the specificity of L1 processing (19, 59, 60). However, the search for viral factor(s) which directly participate in alternative RNA splicing has so far been unsuccessful. It should also be mentioned that it cannot be excluded, at the present time, that accumulation of alternatively spliced mRNAs is regulated at levels other than RNA splice site choice. For example, at the level of cytoplasmic mRNA stability as has been shown in the case of adenovirus early region 1B (61, 62).

ACKNOWLEDGEMENTS

We thank Elvira Metéreus for excellent secretarial help. This work was supported by grants from the Swedish Cancer Society and the Swedish Board for Technical Development.

REFERENCES

1. Smith, C.W.J., Patton, J.G. and Nadal-Ginard, B. (1989) *Annu. Rev. Genet.*, **23**, 527–577.
2. Eperon, L.P., Graham, I.R., Griffiths, A.D. and Eperon, I.C. (1988) *Cell*, **54**, 393–401.
3. Chebli, K., Gattoni, R., Schmitt, P., Hildwein, G. and Stevenin, J. (1989) *Mol. Cell. Biol.*, **9**, 4852–4861.
4. Solnick, D. and Lee, S.I. (1987) *Mol. Cell. Biol.*, **7**, 3194–3198.
5. Fu, X.-Y. and Manley, J.L. (1987) *Mol. Cell. Biol.*, **7**, 738–748.
6. Zhuang, Y., Leung, H. and Weiner, A.M. (1987) *Mol. Cell. Biol.*, **7**, 3018–3020.
7. Fu, X.-Y., Ge, H. and Manley, J.L. (1988) *EMBO J.*, **7**, 809–817.
8. Emeson, R.B., Hedjran, F., Yeakley, J.M., Guise, J.W. and Rosenfeld, M.G. (1989) *Nature*, **341**, 76–80.
9. Bell, L.R., Maine, E.M., Schedl, P. and Cline, T.W. (1988) *Cell*, **55**, 1037–1046.
10. Chou, T.-B., Zachar, Z. and Bingham, P.M. (1987) *EMBO J.*, **6**, 4095–4104.
11. Zachar, Z., Chou, T.-B. and Bingham, P.M. (1987) *EMBO J.*, **6**, 4105–4111.
12. Reed, R. and Maniatis, T. (1986) *Cell*, **46**, 681–690.
13. Ge, H. and Manley, L. (1990) *Cell*, **62**, 25–34.
14. Krainer, A.R., Conway, G.C. and Kozak, D. (1990) *Cell*, **62**, 35–42.
15. Akusjärvi, G., Pettersson, U. and Roberts, R.J. (1986) In Doerfler, W. (ed.), Adenovirus DNA: the viral genome and its expression. Martin Nijhoff Publishing, Boston, Vol.8, pp. 53–95
16. Chow, L.T., Broker, T.R. and Lewis, J.B. (1979) *J Mol Biol*, **134**, 265–303.
17. Akusjärvi, G. and Persson, H. (1981) *Nature*, **292**, 420–426.
18. Nevins, J.R. and Wilson, M.C. (1981) *Nature*, **290**, 113–118.
19. Svensson, C. and Akusjärvi, G. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 4690–4694.
20. Delsert, C., Morin, N. and Klessig, D.F. (1989) *Mol. Cell. Biol.*, **9**, 4364–4371.
21. Roberts, R.J., Akusjärvi, G., AlestrÖm, P., Gelinas, R.E., Gingeras, T.R., Sciaky, D. and Pettersson, U. (1986) In Doerfler, W. (ed.), Adenovirus DNA: the viral genome and its expression. Martin Nijhoff Publishing, Boston, Vol.8, pp. 1–52
22. Svensson, C. and Akusjärvi, G. (1985) *Embo J.*, **4**, 957–964.
23. Melton, D.A., Krieg, P.A., Rebagliati, M.R., Maniatis, T., Zinn, K. and Green, M.R. (1984) *Nucleic Acids Res.*, **12**, 7033–7034.
24. Öfverstedt, L.G., Hammarström, K., Balgobin, N., Hjertén, S., Pettersson, U. and Chattopadhyaya, J. (1984) *Biochemica et Biophysica Acta*, **782**, 120–126.
25. Dignam, J.D., Lebovitz, R.M. and Roeder, R.G. (1983) *Nucleic Acids Res.*, **11**, 1475–1489.
26. Heintz, N. and Roeder, R.G. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 2713–2717.
27. Krainer, A.R., Maniatis, T., Ruskin, B. and Green, M.R. (1984) *Cell*, **36**, 993–1005.
28. Ruskin, B. and Green, M.R. (1985) *Science*, **229**, 135–140.
29. Krämer, A. and Keller, W. (1985) *EMBO J.*, **4**, 3571–3581.
30. Konarska, M.M. and Sharp, P.A. (1987) *Cell*, **49**, 763–774.
31. Chow, L.T. and Broker, T.R. (1978) *Cell*, **15**, 497–510.
32. Akusjärvi, G. and Pettersson, U. (1979) *Cell*, **16**, 841–850.
33. Zain, S., Sambrook, J., Roberts, R.J., Keller, W., Fried, M. and Dunn, A.R. (1979) *Cell*, **16**, 851–861.
34. Virtanen, A., Aleström, P., Persson, H., Katze, M.G. and Pettersson, U. (1982) *Nucleic Acids Res.*, **10**, 2539–2548.
35. Falvey, E. and Ziff, E. (1983) *J. Virol.*, **45**, 185–191.
36. Berget, S.M. and Sharp, P.A. (1979) *J. Mol. Biol.*, **129**, 547–565.
37. Keohavong, P., Gattoni, R., LeMoullec, J.M., Jacob, M. and Stévenin, J. (1981) *Nucleic Acids Res.*, **10**, 1215–1229.
38. Mariman, E.C., van, B., Reinders, R.J. and van Venrooij, W. J. (1983) *J. Mol. Biol.*, **163**, 239–256.
39. Krainer, A.R. and Maniatis, T. (1988) In Hanes, B. D. and Glover, D. M. (ed.), Frontiers in transcription and splicing. IRL Press, Oxford, pp. 131–206
40. Bindereif, A. and Green, M.R. (1987) *EMBO J.*, **6**, 2415–2424.
41. Ruskin, B., Zamore, P.D. and Green, M.R. (1988) *Cell*, **52**, 207–219.
42. Keller, E.B. and Noon, W.A. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 7417–7420.
43. Ruskin, B., Greene, J.M. and Green, M.R. (1985) *Cell*, **41**, 833–844.
44. Reed, R. and Maniatis, T. (1988) *Genes Dev.* **2**, 1268–1276.
45. Söderlund, H., Pettersson, U., Vennström, B., Philipson, L. and Mathews, M. (1976) *Cell*, **7**, 585–593.
46. Akusjärvi, G., Mathews, M.B., Andersson, P., Vennström, B. and Pettersson, U. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 2424–2428.
47. Zhuang, Y. and Weiner, A.M. (1989) *Genes Dev.*, **3**, 1545–1552.
48. Wu, J. and Manley, J.L. (1989) *Genes Dev.*, **3**, 1553–1561.
49. Nelson, K.K. and Green, M.R. (1989) *Genes Dev.*, **3**, 1562–1571.
50. Zamore, P.D. and Green, M.R. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 9243–9247.
51. García-Blanco, M.A., Jamison, S.F. and Sharp, P.A. (1989) *Genes Dev.*, **3**, 1874–1886.
52. Wang, J. and Pederson, T. (1990) *Nucleic Acids Res.*, **18**, 5995–6001.
53. Swansson, M.S. and Dreyfuss, G. (1988) *EMBO J.*, **7**, 3519–3529.
54. Gerke, V. and Steitz, J.A. (1986) *Cell*, **47**, 973–984.
55. Tazi, J., Alibert, C., Temsamani, J., Reveillaud, I., Cathala, G., Brunel, C. and Jeanteur, P. (1986) *Cell*, **47**, 755–766.
56. Lang, K.M. and Spritz, R.A. (1983) *Science*, **220**, 1351–1355.
57. Nelson, K.K. and Green, M.R. (1988) *Genes Dev.*, **2**, 319–329.
58. Christofori, G., Frendewey, D. and Keller, W. (1987) *EMBO J.*, **6**, 1747–1755.
59. Johnston, J.M., Anderson, K.P. and Klessig, D.F. (1985) *J. Virol.*, **56**, 378–385.
60. Pilder, S., Moore, M., Logan, J. and Shenk, T. (1986) *Mol. Cell. Biol.*, **6**, 470–476.
61. Wilson, M.C. and Darnell, J.J. (1981) *J. Mol. Biol.*, **148**, 231–251.
62. Montell, C., Fisher, E.F., Caruthers, M.H. and Berk, A.J. (1984) *Mol. Cell. Biol.*, **4**, 966–972.