In vitro splicing of mutually exclusive exons from the chicken β-tropomyosin gene: role of the branch point location and very long pyrimidine stretch

Maria Goux-Pelletan, Domenico Libri[1], Yves d'Aubenton-Carafa, Marc Fiszman[1], Edward Brody and Joëlle Marie

Centre de Génétique Moléculaire du CNRS, Laboratoire Propre Associé à l'Université Pierre et Marie Curie, 91190 Gif-sur-Yvette and [1]Molecular Biology Department, Institut Pasteur, 75015 Paris, France

Communicated by E.Brody

The chicken β-tropomyosin gene contains 11 exons, two of which are spliced into mRNA only in skeletal muscle. One pair of alternative exons, 6A and 6B, is found in the middle of the gene; they are spliced in a mutually exclusive manner. The non-muscle splice 6A−7 is by far the predominant *in vitro* reaction in a HeLa cell nuclear extract. A minor product is the 6A−6B splice, which is excluded in all tissues. This minor product results from the use of a branch point located 105 nt upstream of the 3′ end of the intron separating exons 6A and 6B. The region between the branch point sequence and the final AG contains a stretch of ~80 pyrimidines. We have examined the role of the distance of the branchpoint to the 3′ splice site and of the sequences between these two elements. Our results suggest that at least two *cis*-acting elements contribute to the mutual exclusivity of exons 6A and 6B. The intron between exons 6A and 6B is intrinsically poorly 'spliceable' both because the branch point is too far upstream of the 3′ end of the intron to give efficient splicing and because of the particular sequence lying between this branch point and the 3′ splice site.
Key words: *cis*-acting elements/tissue-specific splicing/ tropomyosin pre-mRNA

Introduction

Splicing of nuclear precursor RNA to give mRNA is a central feature of eukaryotic gene expression. In constitutive splicing, introns are removed from a precursor to yield the same mRNA in all tissues and stages of development. This process takes place in a large ribonucleoprotein particle, the spliceosome (Brody and Abelson, 1985; Frendewey and Keller, 1985; Grabowski *et al.*, 1985), and requires precise recognition of correct pairs of splice sites and ligation of contiguous exons. The accuracy of this process is provided by the interaction of snRNPs, some of the hnRNP proteins, and other factors, with required nucleotide sequences in the precursor RNA (for reviews see Green, 1986; Padgett *et al.*, 1986). These sequences include (in metazoans) the 5′ and 3′ splice sites, the pyrimidine tract upstream of the 3′AG and the branch point. The distance between these various elements is crucial for splicing (Fu and Manley, 1987). For some precursors, the terms exon and intron are ambiguous, since the same segment of RNA can function as one or the other. This can then generate different mRNAs from the

same primary transcript, containing combinatory sets of RNA segments. This process, alternative splicing, can be either stochastic or regulated according to the cell environment or the stage of development (tissue-specific splicing). A large number of examples of tissue-specific splicing are known (see Leff *et al.*, 1986; Breitbart *et al.*, 1987 for reviews). Although much progress has been made on the mechanism of constitutive splicing, relatively little is known about the molecular basis underlying alternative splicing, particularly tissue-specific splicing. How does the splicing machinery distinguish an exon that has to be included in mRNA in a specific cellular context while it is used as a part of an intron in another context? As in constitutive splicing, the major problem is the recognition and utilization of the proper pairs of 5′ and 3′ splice sites. In most studies, it has been inferred that *trans*-acting factors control the different splicing pathways used in tissue-specific alternative splicing. Although these factors have not yet been identified, *cis*-acting sequences responsible for dictating tissue-specific splicing have been characterized. These signals, which determine whether an RNA segment is seen as an exon or as an intron, reside within and/or around the differentially spliced exons.

In the case of mutually exclusive exon splicing, two distinct problems arise. One, common to all tissue-specific splicing, is the recognition of the alternative exons. The other pertains only to the splicing of mutually exclusive exons. The contiguous 5′ and 3′ splice sites that separate these exons are never spliced together. However, each splice site reacts separately with other upstream or downstream splice sites.

The β-tropomyosin (β-TM) gene is a good model to study the molecular mechanisms involved in the splicing of mutually exclusive exons. This gene gives rise to at least two isoforms of the protein. One of the maturation events generating this diversity occurs in the middle portion of the gene and involves the tissue-specific splicing of two mutually exclusive exons termed 6A and 6B. Exon 6B is found exclusively in skeletal muscle mRNA, while exon 6A is found in mRNA of all other tissues. All of the information necessary for dictating this tissue-specific splicing lies in a 980 bp fragment extending from exons 5 through 7, the two common exons flanking 6A and 6B (Libri *et al.*, 1989b). Therefore this region can be used to search for both the *cis*-acting sequences which prevent the splicing of exon 6A to 6B, and *trans*-acting factors which select the proper exon according to the cellular environment. We present the results of *in vitro* splicing of a β-TM precursor and show that particular features of the intron separating exons 6A and 6B prevent splicing of the two exons together. This intron cannot be spliced because of the unusual location of the branch point (105 nt from the 3′ acceptor site) and of the presence of a long pyrimidine tract downstream from the branch point. We speculate that alternative secondary structures are involved in the specific selection of mutually exclusive exons.

Results

In vitro splicing of a pre-mRNA containing sequences from exon 6A through exon 7 in a HeLa nuclear extract

Recently, Libri et al. (1989b) have shown by transfection experiments that a 980 bp genomic fragment of the β-TM gene, containing the mutually exclusive exons 6A and 6B flanked by exons 5 and 7, has all the necessary information for directing the skeletal muscle-specific splicing of exon 6B. β-TM mRNA extracted from transfected myoblasts contains exons 5, 6A and 7, while that extracted from myotubes contains exon 6B flanked by the two common exons 5 and 7. We have used this fragment, subsequently cloned in the SP65 transcription vector, to make synthetic pre-mRNA for in vitro splicing in a HeLa cell nuclear extract. More than five lariat forms migrate above the pre-mRNA, and others were detected by 2D gel electrophoresis (data not shown). However, the low efficiency of the splicing reaction did not allow us to analyse precisely the multiple intermediates and products of the reaction. No mRNAs were detected.

Since exon 6A is never spliced in vivo to exon 6B, we chose to study the molecular mechanism that excludes the usage together of the 5' and 3' splice sites of the intron separating these two exons. Keeping in mind that HeLa cells express the fibroblast isoform of the β-TM gene (MacLeod et al., 1985; Helfman et al., 1986), we asked as well whether the muscle-specific exon 6B is seen as part of an intron or as an exon in this particular non-muscular extract. A simpler construct was used to clarify these questions. A PvuII−HindIII fragment spanning exons 6A−7 was cloned into the SP65 vector (SP65 700) (Figure 1). In vitro transcription of this vector yields a 700 nt long pre-mRNA that was then used for in vitro splicing analysis. Splicing kinetics of this precursor showed four lariat forms which migrate more slowly than the precursor (Figure 2A), lariats 1 and 2 being much more intense than 3 and 4. In addition, the kinetics of appearance suggest that lariats 1 and 3 are intermediate splicing products (intron plus exon) and lariats
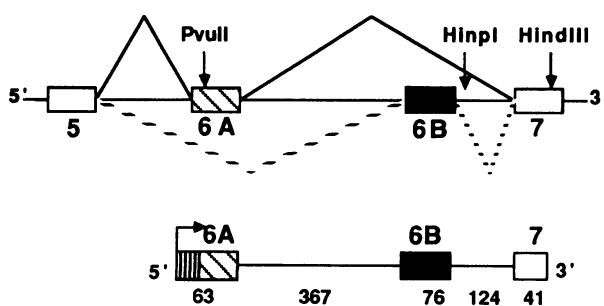
2 and 4 final splicing products (introns only). A fifth smaller lariat, designated i, was revealed by 2D electrophoresis, using 5% acrylamide in the first dimension, and 8% acrylamide in the second (data not shown). The different lariat forms were isolated from large-scale splicing reactions and debranched to determine the length of each lariat (Figure 2B). Lariat 1 (not shown in this figure) and lariat 3 both give a linear RNA of ~620 nt after debranching. Debranched lariat 2 yields a 560 nt linear molecule, while debranched lariat 4 gives rise to a 366 nt linear form.

These results suggest that two different pathways are used: lariats 1 and 2 would correspond to a correct splicing reaction that eliminates exon 6B as part of an intron, (leading thus to the splicing of exons 6A to 7), while lariats 3 and 4 would represent the removal of the intron between the two mutually exclusive exons 6A and 6B, an event completely excluded in vivo. It is to be noted that lariats 1 and 3 give RNA of the same length after debranching (Figure 2B) even though
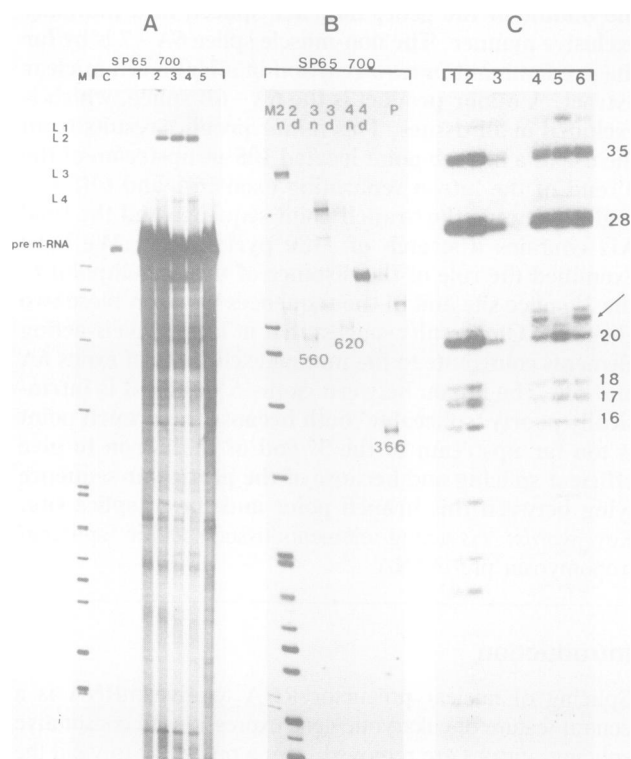

Fig. 2. In vitro splicing reaction kinetics of the 700 nt wild-type pre-mRNA. (A) Pre-mRNA derived from the HindIII linearized SP65 700 vector, was incubated for 30 min (lane 1), 90 min (lane 2), 3 h (lane 3) and 5 h (lane 4) in splicing conditions. Lane 5 presents a mock splicing reaction using a non-hydrolysable ATP analogue. Different lariat forms are designated L₁, L₂, L₃, L₄ and i. Lane C shows the unincubated precursor, and lane M presents size markers (pBR322 DNA digested by restriction enzyme HpaII). (B) The various lariats were subjected to debranching activity. n represent the lariats, d represent the debranched forms. Lanes 2, 3 and 4 designate lariat 2, lariat 3 and lariat 4 respectively. P, indicates the 700 nt precursor. M shows the same size markers as in (A). The various linear lengths obtained are indicated. (C) Each isolated lariat was subjected to RNase T₁ digestion. Lanes 2, 4 and 6 represent respectively pre-mRNA, lariat 3 and lariat 4. Lanes 1, 3 and 5 represent repectively the same samples, all of which had been previously treated with the debranching activity. The arrow indicates the 21mer oligonucleotide containing the branch point linked by a 2'−5' phosphodiester bond to the first guanosine residue of the intron.

Fig. 1. Exon−intron organization of an internal region of the β-TM gene showing the two mutually exclusive exons. Unfilled boxes represent the common exons 5 and 7. The cross-hatched box represents the non-muscular alternative exon (6A). The filled box represents the skeletal muscle-specific exon (6B). Horizontal solid lines represent introns. The tissue-specific alternative pathways used are indicated by dashed diagonal lines (skeletal muscle) or by solid diagonal lines (all other tissues). The PvuII−HindIII fragment cloned into the SP65 vector is shown below. The plasmid was linearized either by HindIII or HinpI. The transcription initiation site is indicated by an arrow. The dashed box represents the SP6 promoter and part of the polylinker (25 nt for SP65 700 and 28 nt for all the other constructs). The size of exons and introns are indicated.

lariat 1 migrates much more slowly than lariat 3 before debranching (Figure 2A). This is to be expected since the circular part of lariat 1 would be larger than the circular part of lariat 3. After debranching, the small lariat i gives a 124 nt linear form corresponding to the size of the intron separating exon 6B from 7. In a non-muscular context this splicing reaction should not take place. In order to confirm our hypotheses about the splicing pathways used *in vitro*, we determined the location of the branch point in each lariat RNA.

For lariats 3 and 4, we took advantage of the following observation: the only sequence resembling the vertebrate branch point consensus, YNCUGAC (Oshima and Gotoh, 1987), is the sequence TCTCAAC, which is 105 nt from the 3' end of the intron separating exons 6A and 6B. This sequence lies in a unique 20 nt long potential digestion product of RNase T_1. If the lariat branch point is in this sequence, RNase T_1 digestion of lariats 3 and 4 should give rise to a 21mer oligonucleotide, due to the addition of the 5' G to the 2' OH group of the branch site adenosine. In contrast, the 21mer oligonucleotide should disappear as an RNase T_1 digestion product of lariats 3 and 4 previously treated by the debranching enzyme. It should be replaced by the unique 20mer oligonucleotide. The results shown in Figure 2C confirm this prediction. RNase T_1 digestion of lariats 3 and 4 (in lanes 4 and 6) yield a novel 21mer oligonucleotide which is not present among the RNase T_1 digestion products of pre-mRNA (lane 1 and 2) or debranched lariats 3 and 4 (lanes 3 and 5). Primer extension analysis carried out on lariat 3 confirms the use of the TCTCAAC consensus, 105 nt upstream of the 3' end of the intron (Figure 3). A 165 nt long extension product is revealed when a 20mer oligonucleotide complementary to exon 6B is used (Figure 3A, lane 1). On the other hand, a 20mer oligonucleotide complementary to exon 7 is extended to give a 320 nt product (Figure 3A, lane 2). No other reverse transcriptase stop was seen between the AG and the consensus branch point sequence, even though numerous A residues exist at 'standard' locations for branch points, 18–40 nt upstream of the AG (Green, 1986). The branch point for lariat i was located on the intermediate form (intron–exon 7) by primer extension using the same 20mer oligonucleotide complementary to the 5' end of exon 7. The extended product shown in Figure 3B is 45 nt long, mapping the branch point to the consensus sequence GCTAAC, 25 nt upstream of the 3' acceptor splice site. Lariats 1 and 2 use the same branch point as lariat i (data not shown).

According to these results, the 700 nt long pre-mRNA is predominantly spliced to give rise to the non-muscular form 6A–7, using 6B as part of an intron. It has been shown by Schmitt *et al.* (1987) that the ionic composition of the splicing buffer could influence the choice of alternative splice sites. Lowering the concentration of $MgCl_2$ and KCl (2 and 40 mM respectively instead of 3 and 66 mM in the standard reaction) does not modify the splicing pattern (data not shown). We can conclude that the *in vitro* splicing of the 700 nt precursor containing exons 6A–7 mimics the tissue-specific splicing of the β-TM gene in non-muscular cells.

Cis-acting elements regulate splicing of exons 6A–6B

The minor *in vitro* splicing pathway of the 700 nt pre-mRNA eliminates the intron separating the two mutually exclusive exons 6A and 6B. The adenosine branch point used in this

splicing reaction is located 105 nt from the 3' acceptor site, far upstream of the 'standard' position (usually 18–40 nt). Examination of the intron sequence reveals a very long stretch of pyrimidines downstream from the branch point. This feature is also found in several other introns separating mutually exclusive exons, e.g. rat α-TM and rat β-TM (Libri *et al.*, 1989a; Smith and Nadal-Ginard, 1989).

Since exon 6A is never spliced to exon 6B in mRNA of any tissue, it is possible that *cis*-acting elements play a role in hindering the removal of the intron separating them. If this hypothesis is correct a modification of these elements would allow the splicing of exon 6A to exon 6B. To verify
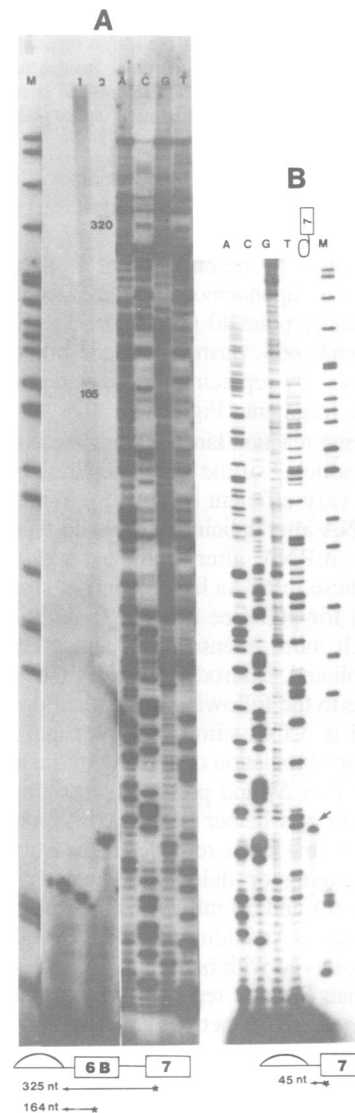
Fig. 3. Location of the branch point by primer extension analysis. (A) Lariat 3 was subjected to primer extension analysis using in **lane 1** a 22mer oligonucleotide complementary to exon 6B and in **lane 2** a 20mer oligonucleotide complementary to exon 7. The length of the extended primers is indicated. Labelled size markers (pBR322 cleaved with *Hpa*II) are shown on the left. RNA sequence using the primer complementary to 6B is shown. (B) The branch point of i was located on the intermediate form (intron–exon 7) using the 20mer oligonucleotide complementary to exon 7. The arrow indicates the extended product. M is same marker as in (A). RNA sequence was performed using the same 20mer primer. A schematic representation of the extension analysis is shown below.

SP65 700

ctcaActtg....ttccctgtgctccttccgctcccaccttcttcccttcatcaccccctcccctgcctggtggcccctgcttgggatgtagTAAATGTG

SP65 Δ1

ctcaActtg....ttcccgggatgtagTAAATGTG

SP65 p63

ctcaActtg....ttcccctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttaggggatgtagTAAATGTG

SP65 p54

ctcaActtg....ttcccctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaacccctgcttgggatgtagTAAATGTG

SP65 p54-1AG

ctcaActtg....ttcccctccggttcccaacgatcggatcgagttacatgatcccccatgttgtgcaaaaacccctgcttgggatgtagTAAATGTG

SP65 p54-2AG

ctcaActtg....ttcccctccggttcccaacgatcggatcggattacatgatcccccatgttgtgcaaaaacccctgcttgggatgtagTAAATGTG

Fig. 4. Sequence of the 3' region of the intron separating 6A and 6B of the wild-type and mutant constructs. The branch point consensus sequence is underlined. The star designates the adenosine residue used as branch site. The first 8 nt of exon 6B are in capital letters. The triangle points to the *Sma*I site created by the deletion of the pyrimidine stretch. The pBR322 fragments inserted into the *Sma*I site are dotted underlined. The arrow indicates the location of the AAG used as 3' splice site.

this, two mutants were constructed to discriminate between the distance parameter and the function of the polypyrimidine region.

Δ_1 was obtained by deleting 65 nt from the stretch of pyrimidines using oligonucleotide-directed mutagenesis. This positions the branch point 40 nt from the 3' end of the intron. The other mutant, p63, pushes back the branch point to its original distance, by replacing the deleted fragment with a 63 nt pBR322 fragment (Figure 4).

In vitro splicing reaction kinetics using these two constructs show that the splicing of the intron separating 6A from 6B now becomes very efficient (Figure 5). Whereas we could not detect mRNA after splicing of the wild-type pre-mRNA, we do detect mRNA after both Δ_1 and p63 splicing. Surprisingly these mRNAs have different sizes: 208 nt for Δ_1 and 288 nt for p63 (see below). In addition, the small lariat i is much more intense in Δ_1 and p63 splicing than in wild-type splicing. Microdensitometry (see Materials and methods) leads to the following quantification: i production in Δ_1 and p63 is 3-fold stimulated over that seen in wild-type. The strong stimulation of splicing of the intron between exons 6B and 7 in Δ_1 and p63 is a surprising result. This stimulation is strong whether or not the intron between exons 6A and 6B has first been removed. The stimulation must mean that changing or deleting sequences in the intron between exons 6A and 6B influences the splicing of the next intron downstream. An additional lariat, designated as lariat 5 or 5', was observed with both mutants. We will show (see below) that lariats 5 and 5' represent a new intermediate form in which the small intron between exon 6B and 7 has already been eliminated.

In order to analyse more precisely the splicing pathways of Δ_1 and p63, purified lariat forms were characterized by debranching, RNase T_1 digestion and by primer extension assays. Table I summarizes the results obtained. Lariats 1 and 2 correspond to the splicing pathway utilizing 6B as an intron, as they did for the wild-type pre-mRNA splicing. Whereas this was by far the predominant splicing pathway for the wild-type precursor, it is now minor for Δ_1 and p63. Splicing of the Δ_1 precursor shows a huge stimulation of lariats 3, 4 and 5. Lariat 3 represents an intermediate form in the reaction, which splices exon 6A to 6B, as it did in the case of the wild-type pre-mRNA lariat 3. The small intron between exons 6B and 7 has not been spliced out of
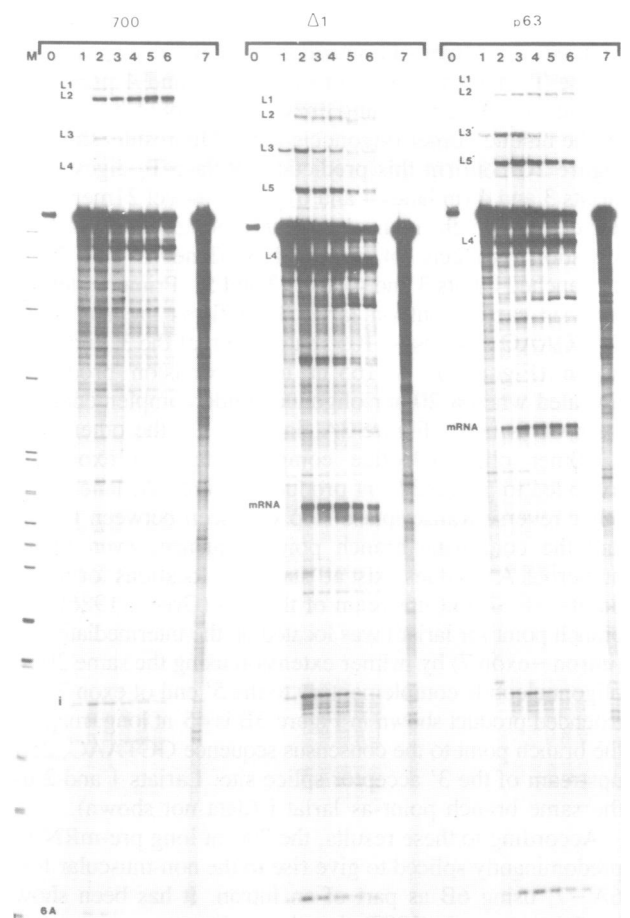

Fig. 5. *In vitro* splicing kinetics of pre-mRNAs 700, Δ_1 and p63. The precursors were incubated for 30 min (lane 1), 60 min (lane 2), 90 min (lane 3), 2 h (lane 4), 2 h and 30 min (lane 5) and 3 h (lane 6). Mock splicing reactions are shown in lane 7. Lanes O represent the unincubated precursors. Lane M designates ^{32}P-labelled *Hpa*II fragments of pBR322. The different lariat forms are designated by numbers following the letter L. Δ_1 and p63 mRNA are indicated. The letter i represents the lariat intron separating exons 6B from 7. 6A represents free exon 6A, the splicing intermediate generated by the first splicing reaction. The primes indicate that lariat forms of p63 use a 3' splice site coded in the pBR322 sequence (see the text). The splicing products were analysed on a 5% acrylamide gel. This explains why the final lariats of the Δ_1 and p63 splicing reactions migrate faster than the precursor. They migrate above the precursor in a 6% acrylamide gel (see Figure 6).

Table I. A summary of the results obtained after analysis of lariats generated during the splicing of Δ_1 and p63

Δ1

RNA species	Size (nt)	Branch pt position	Structure
L1	543	−25	
L2	502	−25	
L3	543	−40	
L4	302	−40	
L5	419	−40	

p63

RNA species	Size (nt)	Branch pt position	Structure
L1	606	−25	
L2	565	−25	
L3'	606	−52	
L4'	314	−52	
L5'	482	−52	

The size of each debranched lariat is indicated. The distance between the branch point and the 3' acceptor site used is reported. Schematic representation of each lariat is shown. Filled boxes represent exon 7 while empty boxes represent the skeletal muscle specific exon 6B. Solid lines represent introns. The arrows indicate the site in the pBR322 insert which is used as a 3' splice acceptor in p63 and p54.

lariat 3. Lariat 5, on the other hand, refers to another pathway where the small intron has already been spliced out before the removal of the intron separating 6A for 6B. Lariat 4 is the final splicing product of both intermediates 3 and 5. Lariats 3', 4' and 5' seen after p63 splicing represent analogous products. All these lariat forms (3 → 5 and 3' → 5'), which correspond to the splicing of the large intron separating 6A from 6B, use the same branch point as the one determined for the wild-type precursor (see Table I).

After debranching the lariat 4 of Δ_1, we find the expected intron length of 302 nt. Lariat 4' of p63 splicing, however, yields a 314 nt linear molecule after debranching. This is shorter than the expected length of 365 nt, which would have been observed if p63 had used the normal 3' end of this intron for splicing. The inserted pBR322 sequence in p63 contains two PuAG trinucleotides, 52 and 57 nt downstream from the branch point. One of these, most probably the AAG, which is closest to the branch point, is used in p63 as the 3' acceptor site; the original 3' splice site is not used at all. This explains why the mRNA found after p63 splicing is longer than that found after Δ_1 splicing.

We then asked if the p63 construct might have inactivated the original 3' splice site, since only nine of the original nucleotides were left at the 3' end of the intron. It could be that 9 nt are not sufficient to constitute a functional 3' splice site. We constructed another mutant, p54, derived from p63, by oligonucleotide-directed mutagenesis. The last 9 nt at the 3' end of the pBR322 fragment were replaced by the 9 nt from the wild-type sequence (Figure 4). We thus obtained an acceptor site with 18 original nucleotides at the 3' end of the intron. We asked whether the original 3' splice site is now able to compete *in cis* with the PuAG from the pBR322 insert. Figure 6 shows that it cannot: p54 splicing is indistinguishable from that of p63. Therefore, the PuAG sequences introduced by the inserted pBR322 fragment completely eliminate utilization of the wild-type 3' splice site. Since the normal 3' splice site is 103 nt away from the branch point in p54, competitor 3' splice sites lying nearer are preferentially used (the discrepancy between 103 and 105 nt is explained by the replacement of 65 nt from the wild-type sequence by 63 pBR322 nucleotides in the p63 construct).
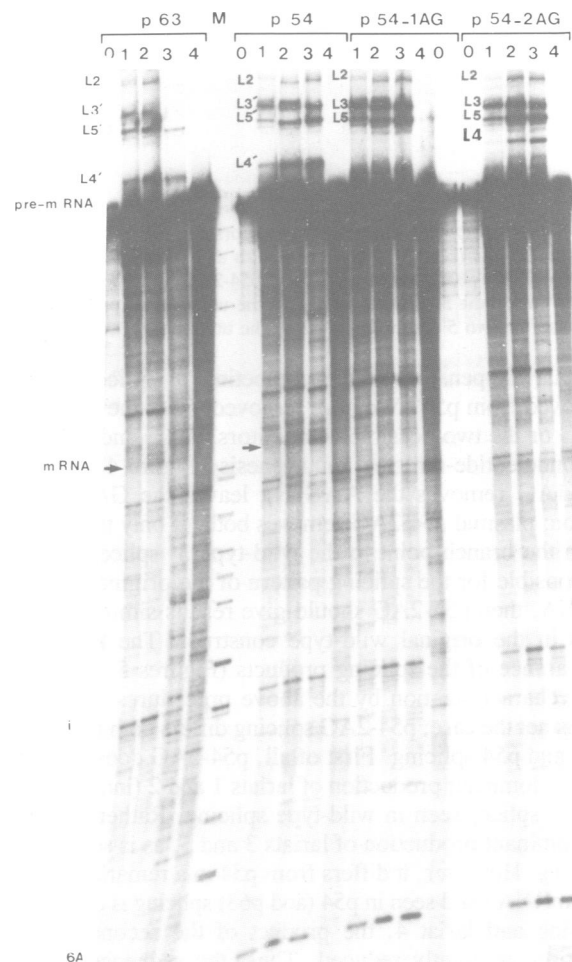

Fig. 6. *In vitro* splicing patterns of pre-mRNAs p63, p54, p54-1AG and p54-2AG. Lane O shows the unincubated precursor, lanes 1−3 represent the following incubation times: 60 min, 90 min, 2 h. Mock splicing reactions are shown in lane 4. Lane M designates ^{32}P-labelled *Hpa*II fragments of pBR322. The arrows indicate mRNA.
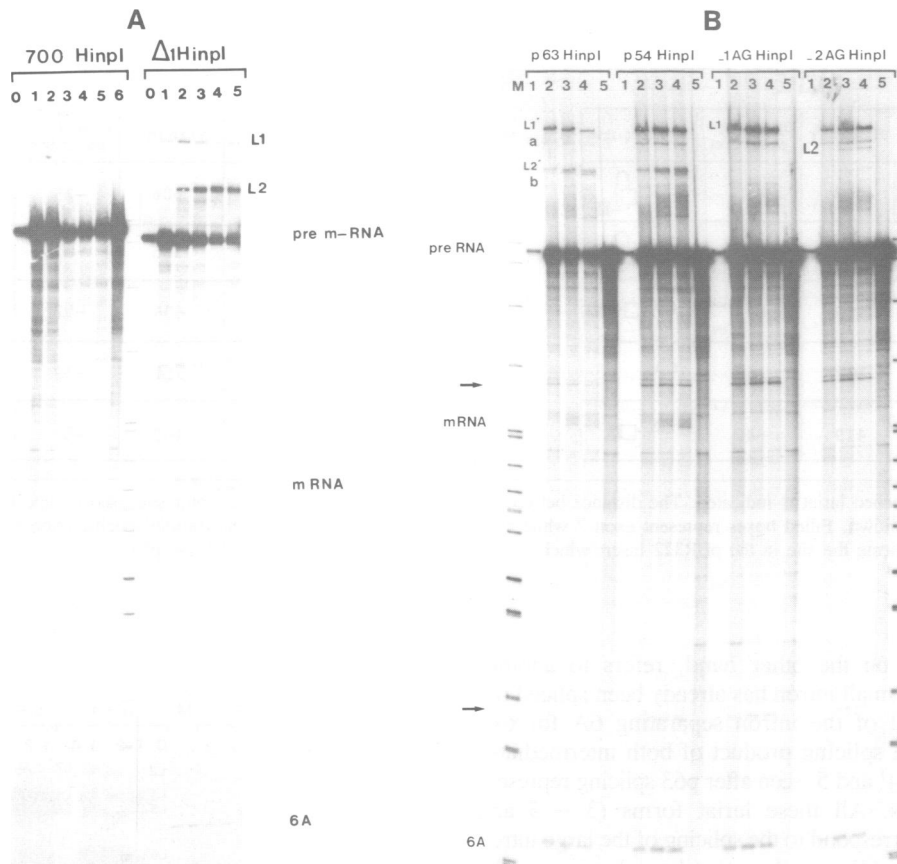
Fig. 7. *In vitro* splicing kinetics of precursors derived from *Hin*pI digested vectors. (**A**) 700 and Δ_1 were spliced for 30 min ((**lane 1**), 60 min (**lane 2**), 90 min (**lane 3**), 2 h (**lane 4**), 2 h and 30 min (**lane 5**). Mock splicing reaction is shown in **lane 6**. **Lane O** represents the unincubated precursors. (**B**) p63, p54, p54-1AG and p54-2AG precursors were spliced for 60 m (**lane 2**), 90 min (**lane 3**), 2 h (**lane 4**). Mock splicing reactions are shown in **lane 5**. **Lane 1** represents the unincubated precursor. **Lane M** designates ^{32}P-labelled *Hpa*II fragments of pBR322. The arrows designate the two 5' exons liberated by the utilization of two cryptic 5' splice sites.

What happens when the competing 3' splice sites are removed from p54? We have removed either the competing AAG or the two possible competitors, AAG and GAG, by oligonucleotide-directed mutagenesis (Figure 4). Plasmid p54-1AG removes the AAG but leaves the GAG in this region; plasmid p54-2AG removes both. If only the distance from the branch point to the wild-type 3' splice site were responsible for the splicing pattern of the original 700 pre-mRNA, then p54-2AG should give results similar to those seen in the original wild-type construct. The kinetics of appearance of the splicing products (Figures 5 and 6) and their characterization by the above procedures shows that this is not the case; p54-2AG splicing differs both from wild-type and p54 splicing. First of all, p54-2AG does not show the predominant production of lariats 1 and 2 (indicating the 6A−7 splice) seen in wild-type splicing. Rather, it yields predominant production of lariats 3 and 5, as is seen in p54 splicing. However, it differs from p54 in a remarkable way. The mRNA band seen in p54 (and p63) splicing is completely missing and lariat 4, the product of the second splicing reaction, is greatly reduced. Thus, the sequence that has replaced the polypyrimidine stretch in p54-2AG neither restores the wild-type splicing pattern nor does it allow the lariat intermediates of 6A−6B splicing to undergo the second splicing reaction efficiently. This is true in spite of the fact that the branch point used in formation of lariats 3 and 5 of p54-2AG is the same as that used in all the other

constructs. The wild-type 3' splice site cannot be used in this context, even though the first step in splicing proceeds normally. It is also apparent in Figure 6 that the p54-1AG precursor produces lariats 3 and 5, but that the second splicing reaction is completely inhibited; both mRNA and lariat 4 are completely missing. The GAG remaining in this construct is apparently not usable as a 3' splice site, which suggest that only the AAG is used as a 3' splice site in p63 and p54 splicing. Moreover, the reproducible difference between p54-1AG and p54-2AG splicing (the production of a small amount of lariat 4 in the latter and its complete absence in the former) may mean that the GAG sequence actually inhibits the utilization of the wild-type 3' splice site, possibly by forming a 'dead-end' complex wth a splicing component.

Intrinsic 'spliceabilty' of the intron between 6A and 6B

In order to examine the 'spliceability' of the intron separating the two mutually exclusive exons 6A and 6B in the absence of any other acceptor and donor splice sites, the different SP65 plasmids were linearized with restriction enzyme *Hin*pI instead of with *Hind*III. The different pre-mRNAs synthesized contain only the two exons 6A and 6B and the intron separating them. The wild-type transcript 580 is barely spliced (Figure 7). Δ_1 *Hin*pI, on the other hand, is spliced very efficiently, yielding a 205 nt mRNA. p63 and p54 *Hin*pI

are spliced as efficiently as Δ_1, still utilizing the 3' splice site provided by the pBR322 insert. When debranched, the lariats 2' from p63 and p54 *Hin*pI splicing give rise to a 314 nt linear form instead of the 365 nt form that would have been seen if the authentic 3' acceptor site was used (data not shown). This is correlated with the appearance of a 257 nt long mRNA. p54-1AG *Hin*pI gives only the lariat product of the first splicing reaction; neither the final intron lariat nor mRNA are produced. This is not because of the use of an abnormal branch point. The branch point determined after RNase T_1 digestion lies in the 20mer oligonucleotide just as it did in p54 splicing (data not shown). p54-2AG produces a small amount of the final intron lariat. The second splicing reaction is nevertheless so reduced that no mRNA is detectable. Thus, the major splicing products of the *Hin*pI transcripts parallel those seen for the intron separating exons 6A and 6B in the *Hin*dIII transcripts. Additional intermediate lariat forms (designated as a and b) are also detected (Figure 7). They are correlated with the appearance of two linear forms of 117 and 287 nt, designated by arrows in Figure 7. These observations indicate that two minor splicing reactions occur using cryptic 5' splice sites. The one corresponding to lariat a could be assigned to the sequence GAG/GTGCAG, 26 nt downstream of the authentic 5' splice site and the other could be mapped either 191 or 198 nt downstream of the normal splice site where there are two overlapping potential 5' splice sites (GCA/GTGACA or CAC/GTGAGG). Lariats a and b use the same branch point as the major lariat forms (data not shown).

Discussion

Mutually exclusive splicing of adjacent exons is one of the patterns seen when a common pre-mRNA is spliced differentially to give two or more mRNAs (Leff *et al.*, 1986; Breitbart *et al.*, 1987; Brody *et al.*, 1989). In the tissue-specific pair studied here, as well as in other tissue-specific pairs of mutually exclusive exons, the 5' and 3' sequences at either end of the intron that separates the exons (in our case exons 6A and 6B) are unremarkable: both can function as splice sites, but never with each other. This sort of tissue-regulated availability of splice sites may be related to alternative RNA secondary or tertiary structures, each of which can be trapped in one form by tissue specific *trans*-acting factors.

Such a mechanism is the basis for attenuation in prokaryotes where ribosome passage modulates switching from one secondary structure to another (Landick and Yanofsky, 1987). A number of experiments suggest that alternative RNA structures could account for alternative splicing (Solnick, 1985; Solnick and Lee, 1987; Eperon *et al.*, 1988; Munroe, 1988). Although tissue-specific factors that hide or activate a particular sequence may bind specifically to segments of RNA, the mechanism of activation could be indirect. Aebi *et al.* (1986) first proposed that splicing decisions might be made quickly with respect to transcription (first come first served) and that such considerations could influence competition between splice sites (Beyer and Osheim, 1988; Eperon *et al.*, 1988).

None the less, *in vitro* splicing in HeLa cell nuclear extracts can mimic certain aspects of tissue-specific splicing as they do in the study presented here. We have found that the intron separating exons 6A and 6B of the β-TM gene

has a sequence that renders it poorly 'spliceable' *in vitro*. The non-utilization of this potential intron as an intron is directly related to the mutually exclusive splicing of exons 6A and 6B (Libri *et al.*, 1989a,b and unpublished results). Thus, the HeLa cell nuclear extract sees exon 6B as part of a large intron extending from exon 6A to exon 7, just as would all tissues except skeletal muscle. There are two minor reaction products that use exon 6B as an exon: in one, the splice joining 6A and 6B is made (generating lariats 3 and 4); in the other, exon 6B is joined to exon 7 (generating lariat i). The branch point and 3' splice site of lariat i and lariat 2 are the same (data not shown). Why is the 5' splice site following exon 6A used preferentially over that following exon 6B? The 5' splice site following 6A, GTACTG, has three deviations from the consensus sequence GTAAGT, and it is selected although it is in competition with the 5' splice site after exon 6B, GTATGA, which has only two deviations from the consensus. Moreover, both 5' splice sites represent rather common deviations from the 5' splice site consensus sequence, both in mammals and in non-mammalian vertebrates (Jacob and Gallinaro, 1989). We think the answer is given by the fact that Δ_1, which eliminates sequences in the intron between exons 6A and 6B, stimulates utilization of the 5' splice site of the downstream intron. The 5' splice site of this intron must be masked by a structure in which the polypyrimidine stretch (which is eliminated in Δ_1) interacts with the downstream purine-rich sequences in such a way as to render its 5' splice site inaccessible to some part of the splicing apparatus. It is, nonetheless, also possible that a component of the HeLa cell extract binds specifically to the very long polypyrimidine stretch in this intron, and, as a consequence, blocks access to the downstream 5' splice site.

The splicing of exons 6A and 6B, which is never found *in vivo*, takes place to a minor extent in the wild-type construct. The branch point found in these minor lariat forms is 105 nt upstream of the 3' end of the intron. The overwhelming majority of branch points used in vertebrate splicing are between 18 and 40 nt upstream of the 3' end of the introns (Green, 1986); moreover, if one eliminates wild-type branchpoint sequences, splicing can often continue at a reduced rate because the splicing apparatus will use A residues at the appropriate distance to make lariats (Padgett *et al.*, 1985; Reed and Maniatis, 1985; Ruskin *et al.*, 1985). The branch point consensus far upstream is used in β-TM splicing even though there are numerous A residues between 18 and 40 nt upstream of exon 6B. Between this branch point and exon 6B is a stretch of 105 nt, 84 of which are pyrimidines. This very long pyrimidine stretch has been found associated with other pairs of mutually exclusive exons (Libri *et al.*, 1989b). This raises the question of whether the distance between the branch point and the 3' acceptor site and/or the stretch of pyrimidines is responsible for the 'non-spliceability' of the intron separating the mutually exclusive exons.

To answer these questions, we have changed the two variables with one construction, Δ_1. Sixty-five pyrimidines from the pyrimidine stretch were deleted, thus moving the branch point consensus sequence to 40 nt from the 3' end of the intron. Changing these two variables makes the intron between exons 6A and 6B completely accessible to the HeLa cell splicing apparatus; as a result, lariats 3 and 4 represent the major splicing pathway of this precursor (along with a

stimulation of splicing of the downstream intron, as discussed above). Which of the variables is responsible for this? When we fill in Δ_1 with 63 nt from pBR322 (p63), the splicing efficiency does not change. However, the spliced product does not arise from the use of the wild-type 3' splice site, neither for the p63, nor for the p54 precursors. Instead the AG chosen is the AAG present in the pBR322 sequence; PuAGs are found at the end of introns only rarely (Mount, 1982; Oshima and Gotoh, 1987) and may be inherently less strong 3' splice sites than PyAGs. Eliminating the AAG does not restore the wild-type splicing pattern. Pre-mRNA from p54-2AG, in contrast to the wild-type precursor, produces lariats 3 and 5 to the same extent as the p54 precursor. However, these intermediates cannot be efficiently used for the second splicing reaction, probably due to the large distance separating the branch point from the 3' splice site (Δ_1 precursor undergoes both splicing reactions efficiently). The intron between exons 6A and 6B is intrinsically poorly 'spliceable' both because the branch point is too far upstream of the 3' end of the intron to give efficient splicing and because of the particular sequence lying between this branch point and the 3' splice site. A similar location (177 nt upstream of the 3' end of the intron) has been found for a branch point separating a pair of mutually exclusive exons in the rat α-TM gene (Smith and Nadal-Ginard, 1989). In that case, however, the branch point left too little distance between itself and the 5' end of the intron to allow splicing to occur. That intrinsic 'non-spliceability' could be relieved by inserting sequences in between these two elements. Clearly, this mechanism is not responsible for the 'non-spliceability' of the intron between exons 6A and 6B studied here, since 263 nt lie between the end of exon 6A and the branch point, approximately five times longer than the minimal distance required between a 5' splice site and a branch point (Wieringa *et al.*, 1984; Ruskin *et al.*, 1985; Fu and Manley, 1988).

Finally, a third element suggested by our data is that sequences in this region are involved in structures which are mutually interfering. In the wild-type construct, splicing of the intron between exons 6B and 7 takes place at a low level. In Δ_1, the deletion in the intron upstream of exon 6B strongly stimulates splicing of the small intron downstream of exon 6B, suggesting either that spliceosome formation is interactive on adjacent introns or that we have eliminated a structural element that interferes with splicing of the small intron between exons 6B and 7. We prefer the latter interpretation. A very stable potential structure which forms between the pyrimidine-rich region of the upstream intron and the purine-rich regions (exon 6B and the downstream intron) has previously been proposed (Libri *et al.*, 1989a). In addition, Helfman *et al.* (1988) have shown that a similar situation may exist in the rat β-TM gene.

In sum, we have identified different structural elements surrounding mutually exclusive exons that act as signals, presumably to allow *trans*-acting factors to fix alternative RNA structures.

Materials and methods

Vector constructions

The different fragments were cloned into the SP65 transcription vector using standard procedures (Maniatis *et al.*, 1982). Different restriction enzymes were purchased from New England Biolabs, Amersham and Bethesda Research Labs. Calf intestinal phosphatase was obtained from Promega.

All constructions described below derive from a 1.7 kb chicken β-TM genomic clone spanning exons 4–7 (Libri *et al.*, 1989a). Oligonucleotide-directed mutagenesis permitted us to create a *Bam*HI restriction site 15 nt upstream of exon 5. A 980 bp fragment starting at this created *Bam*HI site and terminating at a *Hin*dIII restriction site lying 41 nt downstream of the 5' end of exon 7 was cloned in the pSP65 vector, giving pSP65 980. A *Pvu*II–*Hin*dIII sub-fragment (671 nt) was used to construct the pSP65 700 vector (the *Pvu*II restriction site lies 13 nt downstream from the start of exon 6A). Twenty-five nucleotides of vector sequence are transcribed upstream of exon 6A.

Site-directed mutagenesis was performed using the Amersham handbook procedures. a 65 nt deletion was carried out on the *Pvu*II–*Hin*dIII fragment (cloned into the M13mp18 vector) by using a 43mer oligonucleotide 5'-CCACATTTACTACATCCCGGGAAGGAGGAAAGAAGGAGAG-AGA-3'; this led simultaneously to the creation of a *Sma*I site (CCCGGG). We call this construct Δ_1. A 63 nt fragment provided by an *Alu*I digestion of pBR322 was inserted in the *Sma*I site of Δ_1, giving rise to p63. A 42mer oligonucleotide 5'-CACCACATTTACTACATCCCAAGCAGGGGTTT-TTGCACAACA-3' was used to transform p63 into p54 by mutagenesis. p54 differs from p63 in that the 9 nt at the 3' of the pBR322 sequence have been replaced with the original 9 nt from the wild-type sequence. The p54-1 AG construct derives from p54 and differs by the mutation of the AAG trinucleotide from the pBR322 insert, to GGA. This mutant was obtained by olignucleotide-directed mutagenesis using a 34mer 5'-GGATCATGT-AACTCGATCCGATCGTTGGGAACCG-3'. The elimination of both PuAG trinucleotides (AAG and GAG both to GGA) was also performed on p54 by directed mutagenesis using a 38mer oligonucleotide 5'-TGGGGGATCATGTAATCCGATCCGATCGTTGGGAACCG-3'. This yields the p54-2AG construct.

Transcription and splicing experiments

Uniformly labelled capped SP6 transcripts were synthesized in a 20 μl transcription reaction containing 40 mM Tris (pH 7.5), 6 mM $MgCl_2$, 2 mM spermidine, 10 mm DTT, 0.1 mM ATP and CTP, 0.05 mM GTP, 1.5 mM RNA cap analogue [$m^7G(5')$ ppp(5')G], 0.075 mM UTP, 40 μCi of [α-^{32}P]UTP (800 Ci/mmol; Amersham) 40 U RNasin (Promega), 10–20 U SP6 polymerase (Promega) and 0.5–1.0 μg of *Hin*dIII linearized plasmid, unless otherwise indicated. The reactions were incubated for 2 h at 37°C, phenol/chloroform extracted, and ethanol precipitated. The transcripts were purified by electrophoresis on a 5% acrylamide–7 M urea sequencing gel. The specific activity of the pre-mRNA varied between 10 and 15 × 10^6 c.p.m./pmol. HeLa cell nuclear extracts were prepared as described (Dignam *et al.*, 1983). *In vitro* splicing reactions were performed at 30°C for the indicated times. Standard splicing reactions were carried out in 25 μl containing 15 μl nuclear extract, 5–10 fmol labelled precursor, 3 mM $MgCl_2$, 20 mM creatine phosphate, 0.5 mM ATP, 3% (w/v) polyvinyl alcohol and 0.5 U/μl RNasin (Krainer *et al.*, 1984). In mock splicing reactions, ATP and creatine phosphate were replaced by adenyl (β,γ-methylene)-diphosphonate (AMP-PCP) and water respectively. The splicing reactions were stopped at the indicated times by the addition of SDS and proteinase K followed by phenol/chloroform extraction and ethanol precipitation (Krämer *et al.*, 1987). The splicing products were analysed on a 6% (unless otherwise indicated) denaturing polyacrylamide gel.

Large-scale splicing reactions were performed in 50–100 μl reactions containing 0.5–1 pmol of pre-mRNA.

Analysis of the different lariat forms

Lariat forms migrating more slowly than the pre-mRNA on denaturing urea–polyacrylamide gels were easily identified. Two-dimensional electrophoresis using increasing percentages of acrylamide revealed lariats that migrate faster than the precursor in one-dimensional electrophoresis. The different lariat forms were cut out of 6% acrylamide denaturing, 1 mm thick gels after large-scale reactions. The different RNAs were eluted and ethanol precipitated. One thousand counts per minute (counted as Cerenkoff radiation in a scintillation counter) of each isolated lariat were subjected to debranching in buffer D (Dignam *et al.*, 1983) containing 8 mM EDTA and 20% (v/v) HeLa cell S-100 fraction for 45 min at 30°C (Ruskin and Green, 1985). Proteinase K and SDS were added to stop the reaction. Incubation was continued for 30 min at 30°C before phenol/chloroform extraction and ethanol precipitation.

RNase T_1 digestions were performed on isolated lariats or precursor (2000 c.p.m.) using 50 U of RNase T_1 in TE buffer at 37°C for 2 h. The reactions were stopped by lyophilization. The T_1 RNase digestion products were separated on a 20% denaturing polyacrylamide gel.

The branch points were located by primer extension analysis using a 22mer oligonucleotide 5'-CCAGAGACTTCAAGTTGTTGGT-3' complementary

to nt 38−59 of exon 6B and a 20mer oligonucleotide 5′-TTGTCCTCCT-TGGTGGAATA-3′, complementary to the 5′ first 20 nt of exon 7. Lariat RNA (0.2 fmol) was annealed to 0.1 pmol of 5′ labelled primer in 3 μl of hybridization buffer (50 mM Tris, pH 8.3, 60 mM NaCl, 30 mM $MgCl_2$). The mixture was incubated for 5 min at 80°C, then rapidly frozen in a solid CO_2 (dry ice)/ethanol bath. After thawing the samples on ice, primer extension was performed for 45 min at 50°C by adding 1 μl of reverse transcriptase buffer (50 mM Tris, pH 8.3, 60 mM NaCl, 20 mM dithiothreitol), 1 μl of all four dNTPs (0.4 mM each) and 0.1 U of avian myeloblastosis virus (AMV) reverse transcriptase. Labelled RNA was hydrolysed by the addition of NaOH to a final concentration of 0.25 M. After neutralization, the extended product was ethanol precipitated and analysed on an 8% acrylamide denaturing gel. The sequence was performed by four chain termination reactions on cold pre-mRNA in the same conditions as above plus the corresponding ddNTP (50 μM).

The quantification of lariat i was performed using two different methods. First, the autoradiographs were scanned with a microdensitometer (Vernon Phi 5). Lighter exposures than the one shown in Figure 5 were scanned. Second, bands were cut out from the same splicing kinetics gel and counted as Cerenkoff radiation in a scintillation counter. The results were expressed as a ratio between i and the input precursor for each incubation time. A 3-fold increase of i is observed for A_1 and p63 compared to the wild-type.

Acknowledgements

References

Aebi,M., Hornig,H., Padgett,R.A,. Reiser,J. and Weissman,C. (1986) *Cell*, **47**, 555−565.

Beyer,A.L. and Osheim,Y.N. (1988) *Genes Dev.*, **2**, 754−765.

Breitbart,R.E,. Andreadis,A. and Nadal-Ginard,F. (1987) *Annu. Rev. Biochem.*, **56**, 467−495.

Brody,E. and Abelson,J. (1985) *Science*, **228**, 963−967.

Brody,E., Marie,J., Goux-Pelletan,M.S. and Clouet d'Orval,B. (1989) In Grunberg-Manago,M. (ed.), *Evolutionary Tinkering in Gene Expression*. Plenum, New York, pp. 203−213.

Dignam,J.D., Lebovitz,R.M. and Roeder,R.G. (1983) *Nucleic Acids Res.*, **11**, 1475−1489.

Eperon,L.P., Graham,I.R., Griffiths,A.D. and Eperon,I.C. (1988) *Cell*, **54**, 393−401.

Frendewey,D. and Keller,W. (1985) *Cell*, **42**, 355−367.

Fu,X.Y. and Manley,J.L. (1987) *Mol. Cell. Biol.*, **7**, 738−748.

Fu,X.Y. and Manley,J.L. (1988) *Mol. Cell. Biol.*, **8**, 3582−3590.

Grabowski,P.J., Seiler,S.R. and Sharp,P.A. (1985) *Cell*, **42**, 345−353.

Green,M.R. (1986) *Annu. Rev. Genet.*, **20**, 671−708.

Helfman,D.M., Cheley,S., Kuismanen,E., Finn,L.A. and Yamawaki-Kataoka,Y. (1986) *Mol. Cell. Biol.*, **6**, 3582−3595.

Helfman,D.M., Ricci,W.M. and Finn,L.A. (1988) *Genes Dev.*, **2**, 1627−1638.

Jacob,M. and Gallinaro,H. (1989) *Nucleic Acids Res.*, **6**, 2159−2180.

Krainer,A.R., Maniatis,T., Ruskin,B. and Green,M.R. (1984) *Cell*, **36**, 993−1005.

Krämer,A., Frick,M. and Keller,W. (1987) *J. Biol. Chem.*, **262**, 17630−17640.

Landick,R. and Yanofsky,C. (1987) In Neidart,F.C. (ed.), *Escherichia coli and Salmonella typhimurium: Cellular and Molecular Biology*. American Society for Microbiology, Washington, DC, vol.2, pp. 1276−1301.

Leff,S.E., Rosenfeld,M.G. and Evans,R.M. (1986) *Annu. Rev. Biochem.*, **55**, 1091−1117.

Libri,D., Lemonnier,M., Meinnel,T. and Fiszman,M.Y. (1989a) *J. Biol. Chem.*, **264**, 2935−2944.

Libri,D., Marie,J., Brody,E. and Fiszman,M.Y. (1989b) *Nucleic Acids Res.*, **17**, 6449−6462.

MacLeod,A., Houlker,C., Reinach,F., Smillie,L., Talbot,K., Modi,G. and Walsh,F. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 7835−7839.

Maniatis,T., Frisch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Mount,S.M. (1982) *Nucleic Acids Res.*, **10**, 459−472.

Munroe,S.H. (1988) *EMBO J.*, **7**, 2523−2532.

Oshima,Y. and Gotoh,Y. (1987) *J. Mol. Biol.*, **195**, 247−259.

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.

Padgett,R.A., Grabowski,P.J., Konarska,M.M., Seiler,S. and Sharp,P.A. (1986) *Annu. Rev. Biochem.*, **55**, 1119−1150.

Reed,R. and Maniatis,T. (1985) *Cell*, **41**, 95−105.

Ruskin,B. Greene,J.M. and Green,M.R. (1985) *Cell*, **41**, 833−844.

Schmitt,P., Gattoni,R., Keohavong,P. and Stevenin,J. (1987) *Cell*, **50**, 31−39.

Smith,C.W. and Nadal-Ginard,B. (1989) *Cell*, **56**, 749−758.

Solnick,D. (1985) *Cell*, **43**, 667−676.

Solnick,D. and Lee,S.I. (1987) *Mol. Cell. Biol.*, **7**, 3194−3198.

Wieringa,B., Hofer,E. and Weissmann,C. (1984) *Cell*, **37**, 915−925.

Note added in proof

D.M.Helfman and W.M.Ricci (1989) reported a similar unusual branch point location for the analogous mutually-exclusive exons in the rat β-tropomyosin gene (*Nucleic Acids Res.*, **14**, 5633−5645).