

The Length of the Downstream Exon and the Substitution of Specific Sequences Affect Pre-mRNA Splicing In Vitro

PAUL J. FURDON AND RYSZARD KOLE*

Lineberger Cancer Research Center and Department of Pharmacology, University of North Carolina, Chapel Hill, North Carolina 27514

Received 20 August 1987/Accepted 23 November 1987

We have shown previously that truncation of the human β -globin pre-mRNA in the second exon, 14 nucleotides downstream from the 3' splice site, leads to inhibition of splicing but not cleavage at the 5' splice site. We now show that several nonglobin sequences substituted at this site can restore splicing and that the efficiency of splicing depends on the length of the second (downstream) exon and not a specific sequence. Deletions in the first exon have no effect on the efficiency of in vitro splicing. Surprisingly, an intron fragment from the 5' region of the human or rabbit β -globin intron 2, when placed 14 nucleotides downstream from the 3' splice site, inhibited all the steps in splicing beginning with cleavage at the 5' splice site. This result suggests that the intron 2 fragment carries a "poison" sequence that can inhibit the splicing of an upstream intron.

All pre-mRNAs that contain introns have a number of specific sequence elements that are essential for accurate and efficient splicing. These elements appear to be contained within the intron and include the 5' and 3' splice sites, the branch point near the 3' splice site, and the polypyrimidine region. The importance of these sequences has been shown by (i) their high degree of sequence conservation between pre-mRNAs (18), (ii) the occurrence of mutations in each of the elements which severely alter or completely inactivate normal splicing (for recent reviews see references 9, 21, and 27), and (iii) the observation that these sequences specifically bind ribonucleoprotein particles (RNPs) shown to be involved in pre-mRNA splicing (reviewed in reference 16).

It is likely that there are additional sequence elements in pre-mRNA involved in the regulation of splicing that have not yet been characterized. We have shown previously that efficient removal of intron 1 of the human β -globin pre-mRNA can be dramatically reduced in vitro by truncating the transcript so that only 14 nucleotides of the second exon remain. This transcript is able to undergo the first step in splicing, i.e., cleavage at the 5' splice site and lariat formation, but is unable to efficiently generate the final spliced product or the released lariat intron (7). Similar observations have also been reported by Ruskin and Green (24). Reed and Maniatis (23) reported that different exon sequences can affect the efficiency of splice site selection when tested in an in vitro *cis*-competition assay. Also, examples of mutations affecting splicing at a distance are the skipping of exon 5 in the DHFR gene in Chinese hamster ovary cells (17) or skipping of exon 2 when transcripts with an altered 5' splice site in intron 2 of the rabbit β -globin gene are spliced in vitro (1). Since alternatively spliced mRNAs are frequently generated by exon skipping, it seems plausible that this important process is controlled by sequences distant from a given intron, possibly via the interaction of a specific *trans*-acting factor (for a review of alternative splicing, see reference 15). All these results suggest that splicing of an unaltered intron can be affected by sequence elements around it.

Our previous observations (7) suggest that exon sequences immediately downstream from an intron can affect the removal of the intron during splicing. Therefore, we inves-

tigated whether changing the length of a downstream exon is sufficient to reduce the efficiency of splicing or whether there are certain sequence elements necessary for splicing that are being inadvertently deleted from truncated transcripts. We also considered that the efficiency of splicing may depend on the length of the exon upstream from the intron and on interactions that are needed between the two exons.

To address these questions, we have constructed clones that contain various sequences, 14 nucleotides downstream from the 3' splice site, that replace most of exon 2 of the human β -globin gene. Additional clones were constructed that contain deletions in exon 1. These clones were truncated at various restriction endonuclease sites and used as templates to produce pre-mRNA that could be tested for splicing in an in vitro splicing assay. We found that all of the sequences used had no effect on splicing efficiency except substituted sequences from the 5' end of the large intron of the human or rabbit β -globin gene. The presence of these substituted intron sequences in exon 2 inhibits all the steps in splicing, suggesting that certain regions of the intron may act as a "poison" sequence under certain conditions. We also found that the minimum length requirement seen in exon 2 is independent of the sequence substituted in this region of the transcript. The same requirement for a minimum length was not found for exon 1 because transcripts with only 21 nucleotides of exon 1 remaining were spliced as efficiently as those with the entire exon 1 sequence.

MATERIALS AND METHODS

Plasmid construction. All plasmid constructs were made from pSP64HBA6 (12). The ends of some restriction fragments were filled in as indicated with DNA polymerase, Klenow fragment, under the conditions recommended by the manufacturer (New England Biolabs). pNS constructs were made by ligating a *Hae*III digest of pBR322 to a blunt-ended *Acc*I digest of pSP64HBA6. p96PL was made by recirculating a partial *Acc*I digest of pSP64HBA6, and pPL was made by recirculating a blunt-ended complete *Acc*I digest of pSP64HBA6. pNcoI was constructed by digesting pSP64HBA6 with *Hind*III and *Eco*RI, blunt-ending the two sites, and recirculating the vector fragment. pHaeIII was made by ligating the *Hae*III restriction fragment that begins in exon 1 and ends in exon 2 of β -globin to the *Hind*III-

* Corresponding author.

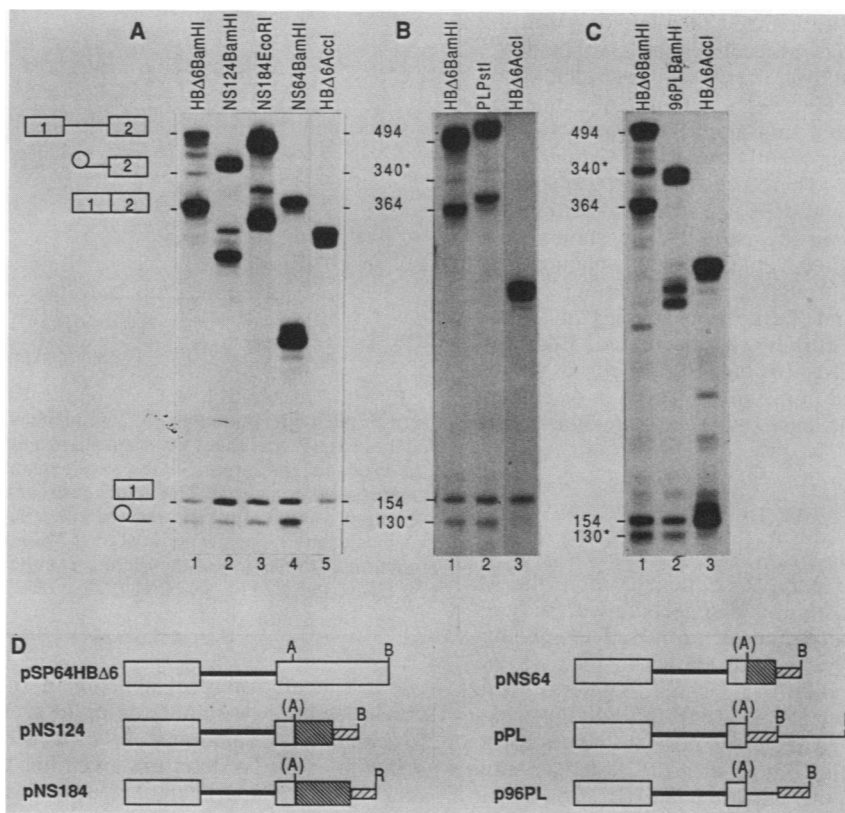


FIG. 1. Effects of nonglobin sequences in exon 2 on the efficiency of in vitro splicing. (A) Splicing of transcripts from the templates pNS64, pNS124, and pNS184. Transcripts (see panel D for structure) were spliced in a nuclear extract, and products were separated on a 5% polyacrylamide sequencing gel (see Materials and Methods). The name of each runoff transcript, shown at the top of the lane, includes the restriction site used to create that transcript as seen in panel D. Lane numbers are shown at the bottom. The structures of the input, intermediates, and products of splicing of the control (HBA6BamHI, lane 1) are given on the left, and the sizes and positions of the corresponding bands are given on the right. (The * indicates aberrant mobility due to lariat structures.) Lanes 2 through 4 show splicing of the NS124, NS184, and NS64 truncated transcripts, respectively. Lane 5 shows splicing of the HBA6AccI truncated transcript, which is not efficiently spliced but undergoes cleavage at the 5' splice site and lariat formation (7). The presence or absence of the 130* released lariat splicing product in each lane is a good indicator for efficient splicing since the size or structure of this product is not altered in any of the constructed templates. (B) Splicing of the PLPstI transcript. Transcript from pPLPstI was spliced in vitro and analyzed in lane 2 as for panel A. The HBA6BamHI and HBA6AccI controls are in lanes 1 and 3, respectively. (C) Splicing of the 96PLBamHI transcript. Transcript from p96PLBamHI was spliced in vitro and analyzed in lane 2 as for panel A. The HBA6BamHI and HBA6AccI controls are in lanes 1 and 3, respectively, with the positions and sizes of the HBA6BamHI input, intermediates, and products of splicing to the left. (D) Structure of the templates. The structures of the templates used to generate the transcripts in panels A, B, and C are shown with the names to the left. The open boxes are exons, the lightly hatched boxes are polylinker sequence, and the thick hatched boxes are substituted sequences. Thick lines represent introns, and thin lines represent flanking, nonglobin sequences. A, *AccI*; B, *BamHI*; P, *PstI*; R, *EcoRI*. Restriction sites destroyed during cloning are shown in parentheses.

BamHI vector fragment (both ends blunt ended) from pSP64 HBA6. p886HBI2 and p466RBI2 were made by ligating the *AvaII-EcoRI* fragment (*AvaII* site blunt ended) from intron 2 of the human β -globin gene (pSP64HBA6) or the *HincII-EcoRI* fragment from intron 2 of the rabbit β -globin gene (pRB1+2, a gift from Michael Green, Harvard University), respectively, to the *AccI-EcoRI* vector fragment (*AccI* site blunt ended) from pSP64HBA6.

Transcription and in vitro splicing. Labeled transcripts were made from templates linearized with various restriction enzymes, using the SP6 system and [³²P]GTP as previously described (12). Transcripts were capped in the transcription reaction (11) and spliced in a 25- μ l reaction for various times as described (12). The products of the reaction were separated on a 5% polyacrylamide sequencing gel (6) unless otherwise indicated.

Quantitation of the efficiency of in vitro splicing. Transcripts were spliced in the standard reactions (see above) in

a time course (tested at 15 min and 30 to 100 min at 10-min intervals), and the products of the reaction were separated on a 5% polyacrylamide sequencing gel. The bands corresponding to the input, the intermediates, and products of splicing were excised, and their radioactivity was determined by Cerenkov counting. The percent splicing for each time point was determined by dividing the amount of input RNA spliced by the total RNA remaining. The amount of spliced input RNA was calculated by measuring the radioactivity (counts per minute) in the released lariat intron band and multiplying it by the ratio of the number of G nucleotides in the total transcript versus the number of G nucleotides in the intron. The released lariat intron was used as a measure of the amount of spliced input RNA instead of the ligated exons because it is the only product that remained unaltered in terms of size and structure in all the transcripts tested. The total RNA remaining was calculated by adding the radioactivity in counts per minute in all the bands remaining.

The percent splicing was normalized for degradation using the 15-min time point as zero percent splicing. Although the endogenous nucleases could interfere with the calculations by degrading the final products, the amount of remaining RNA after the initial 15-min time point remained constant throughout the reaction, apparently due to the formation of the splicing complex (21). Therefore, the percent splicing calculated for each time point does not appear to be affected by endogenous nucleases in the extract. The values were plotted as a function of time versus percent splicing. The equation for the best line through the points was determined using linear regression, and the percent splicing at 50 min (the linear region of the curve) was calculated. For each transcript, this value was used to construct the curves shown in Fig. 3 and 5. Thus each point on the curve is calculated based on the data from at least seven experimental points (lanes from the gel).

RESULTS

Human β -globin gene constructs. We have made a series of constructs which replace exon 2 sequences after the 14 nucleotides downstream from the 3' splice site with various sequences to see how substitutions in exon 2 affect splicing. The structure of these constructs is shown in Fig. 1D and 5D. They include three different *Hae*III fragments from pBR322 (pNS64, pNS124, and pNS184), the polylinker sequence from pSP64 (pPL), a sequence normally found in the 3'-flanking region of the β -globin gene (p96PL), and a region from the large intron of the human (p886HBI2) or rabbit (p466RBI2) β -globin gene. An additional construct (pRR102, a gift from Robin Reed and Tom Maniatis, Harvard University, Cambridge, Mass.), similar in structure to p886HBI2, which contains a portion of the 5' region of the large intron from the human β -globin gene was also tested. pRR102 and p886HBI2 are somewhat different in that the inserted intron 2 sequence in pRR102 is missing the first 3 nucleotides of the intron and the inserted intron 2 sequence in p886HBI2 is missing the first 11 nucleotides of the intron (see below).

Effects of sequence substitutions in exon 2 on splicing. The splicing of pre-mRNAs with sequence substitutions in exon 2 is shown in Fig. 1A, B, and C. As a positive control, splicing of the β -globin pre-mRNA truncated at the *Bam*HI site is shown with the intermediates and products of the reaction depicted on the left (lane 1 in Fig. 1A, B, and C). As a negative control, processing of the β -globin pre-mRNA truncated at the *Acc*I site, 14 nucleotides downstream from the 3' splice site, is also shown (Fig. 1A, lane 5; B, lane 3; C, lane 3). This transcript generates exon 1 and the remaining 3' half of the molecule in the lariat form, which comigrate in this gel system, but does not give a band corresponding to the spliced product or the released lariat intron (7). In all the lanes containing pre-mRNA with substitutions in exon 2, the presence of the released lariat intron can be seen (band 130*), indicating that the transcripts are spliced. The identity of the lariat intron has been determined by its aberrant mobility on a two-dimensional gel and by its change in mobility after treatment with debranching activity (25), present in the cytoplasmic S100 fraction (data not shown). We use this released lariat intron as an indicator for splicing since it is the one splicing product that does not change in size or structure in any of the constructed transcripts. Additional bands that correspond to the spliced products and intermediates as determined by their mobility on the gel can also be seen in each lane, as well as breakdown products produced by the endogenous nucleases in the splicing ex-

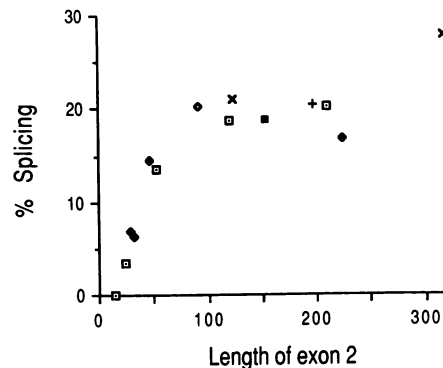


FIG. 2. Quantitative analysis of the efficiency of splicing of transcripts with altered exon 2 sequences. The transcripts with altered exon 2 sequences were spliced in a time-course reaction, and the products and intermediates were separated on a 5% polyacrylamide sequencing gel. The bands were excised, their Cerenkov radiation was determined, and the efficiency of splicing at 50 min was calculated (see Materials and Methods) and plotted as a function of the length of the second exon. Symbols for transcripts: □, HBA6; +, NS124; ◇, NS184; ■, NS64; ◆, PL; ×, 96PL.

tract. The efficiency of splicing of the transcripts shown above seems to be slightly different in some cases (compare, for example, the amounts of the released lariat in lane 1 and 4 in Fig. 1A). To determine whether these differences are real or due to experimental error, we carried out a quantitative analysis for each of the transcripts (see below). However, we conclude from these results that three different pBR322 sequences (Fig. 1A, lanes 2 through 4), the polylinker sequence from pSP64 (Fig. 1B, lane 2), and a region of the human β -globin 3'-flanking sequence (Fig. 1C, lane 2) can all replace exon 2, 14 nucleotides downstream from the 3' splice site, and restore splicing.

Effects of the length of exon 2 on the efficiency of splicing. To determine the efficiency of splicing of the transcripts with exon 2 substitutions, templates were generated by linearizing the constructs with different restriction endonucleases. The transcripts generated from these templates were spliced in a time-course reaction, and the efficiency of splicing was calculated as described in detail in Materials and Methods. The results (Fig. 2) were plotted as a function of the percent splicing at a 50-min time point versus the length of exon 2. Included in this experiment were RNAs with various lengths of unmodified β -globin exon 2 transcribed from the pSP64 HBA6 template (HBA6). The efficiency of splicing of HBA6 increased as the length of exon 2 increased, up to 100 nucleotides where it began to plateau. In spite of the small scattering of the points, this general pattern was seen for all the truncated transcripts, regardless of the sequence present in exon 2. For example, transcript HBA6 truncated at the *Sau*3A1 site and transcript PL truncated at the *Eco*RI site each have a 53- and 47-nucleotide-long exon 2, respectively, after the 3' splice site. Both of these transcripts spliced at approximately the same efficiency (15% at 50 min in the splicing reaction) even though their exon 2 sequences are different after the initial 14 nucleotides. These results again show that it is not the specific sequence of the second exon but the length of the exon that is important for splicing.

Effect of changes in the length of exon 1 on splicing efficiency. Constructs were made with 50-nucleotide (p5' NcoI) and 133-nucleotide (p5' HaeIII) deletions in exon 1 of the human β -globin gene (Fig. 3, bottom) to determine

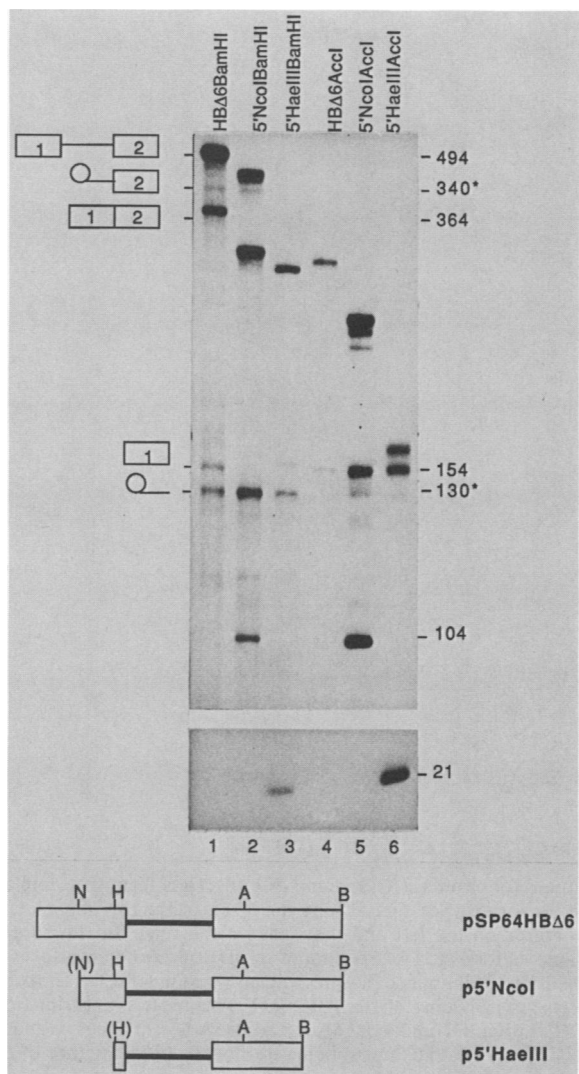


FIG. 3. Effects of deletions in exon 1 on the efficiency of splicing. The transcripts (from the templates diagrammed at the bottom) were spliced and analyzed as for Fig. 1. The 21-nucleotide band shown in the center of the figure was resolved on a 20% polyacrylamide sequencing gel. The sizes (in nucleotides) on the right refer to the positions and sizes of the HBA6BamHI input, intermediates, and products of splicing as diagrammed on the left (see Fig. 1) as well as the 104- and 21-nucleotide exon 1 intermediates from p5'NcoI and p5'HaeIII, respectively. Names of the transcripts, including restriction enzymes used for template truncation, are shown at the top of each lane. N, *NcoI*; H, *HaeIII*; other symbols are described in the legend of Fig. 1. p5'HaeIII has a partial deletion of exon 2 as a result of cloning.

whether transcripts with exon 1 truncations were able to splice efficiently. RNAs transcribed from these templates were spliced as judged by the release of the free lariat intron (Fig. 3, lanes 2 and 3, band 130*). Additional bands in these lanes displayed mobilities expected for the other intermediates and products of splicing, e.g., the 21- and 104-nucleotide released exon 1 splicing intermediates for p5'HaeIII and p5'NcoI, respectively. Small amounts of the released lariat intron were also detected when p5'NcoI and p5'HaeIII RNAs that extended to the *AccI* site in exon 2 were spliced in vitro (Fig. 3, lanes 5 and 6). Since β -globin transcripts with normal exon 1 sequences that were truncated at the *AccI* site

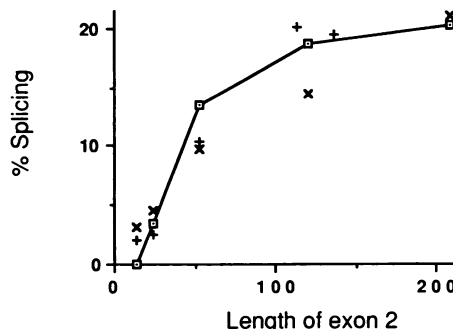


FIG. 4. Quantitative analysis of the efficiency of splicing of transcripts with deletions in exon 1. The efficiency of splicing of transcripts shown in Fig. 3 was quantitated and plotted as for Fig. 2. The solid line shows the efficiency of splicing of the transcripts from pSP64HBA6 with the full-length exon 1. Symbols for templates: □, SP64HBA6; x, *NcoI*; +, *HaeIII*.

were not efficiently spliced (Fig. 3, lane 4; see also Fig. 1A, lane 5, Fig. 1B and C, lane 3, and reference 7), the splicing of these transcripts was unexpected. However, although the spliced products from these exon 1 deletion transcripts could be seen, splicing was still very inefficient (see quantitations in Fig. 4 below).

Effects of exon 1 deletions on the efficiency of splicing with different exon 2 truncations. The efficiency of splicing of transcripts containing exon 1 deletions with different amounts of exon 2 was quantitated and plotted in the same manner as described for Fig. 2. The results are shown (Fig. 4) superimposed on the curve of the splicing efficiency of the HBA6 transcripts containing the full-length exon 1 with different amounts of exon 2. Although small deviations from the curve can be seen, the efficiency of splicing of the exon 1 deletion transcripts followed the same general pattern found for the HBA6 truncated transcripts. Therefore, we conclude that deletions in exon 1 do not affect the efficiency of splicing or significantly alter the requirement for a minimum exon 2 length seen for transcripts truncated at various positions in exon 2. In addition, the overall size of the transcript, within the limits tested, does not appear to be important for splicing. A transcript with a shortened exon 1 and a full-length exon 2 spliced with a much higher efficiency than a transcript with a full-length exon 1 and a truncated exon 2, even though both transcripts were the same size (see Fig. 3, lane 3 versus lane 4, respectively; see quantitation in Fig. 4).

Inhibition of splicing by substituting fragments from the large β -globin intron 2 for exon 2 sequences. The results presented above establish that transcripts with truncated exon 2 sequences are not efficiently spliced but that splicing can be restored if a certain length of exon 2 is reconstructed with nonglobin fragments, regardless of the sequence. Surprisingly, we have encountered a few substituted fragments that will not rescue splicing, one of which is a *HinI-DraI* fragment from the large intron (IVS2) of the human β -globin gene termed plasmid pRR102. When the transcript from the pRR102 template, truncated at an *RsaI* site (leaving 100 nucleotides of IVS2 sequence), was spliced in vitro, none of the intermediates or products were generated (Fig. 5A, lane 3). The same result was obtained when a transcript carrying 500 nucleotides of IVS2 was generated from the pRR102 template (data not shown). These results are different from those obtained with the HBA6 transcript truncated at the *AccI* site, which did not efficiently generate the final splicing

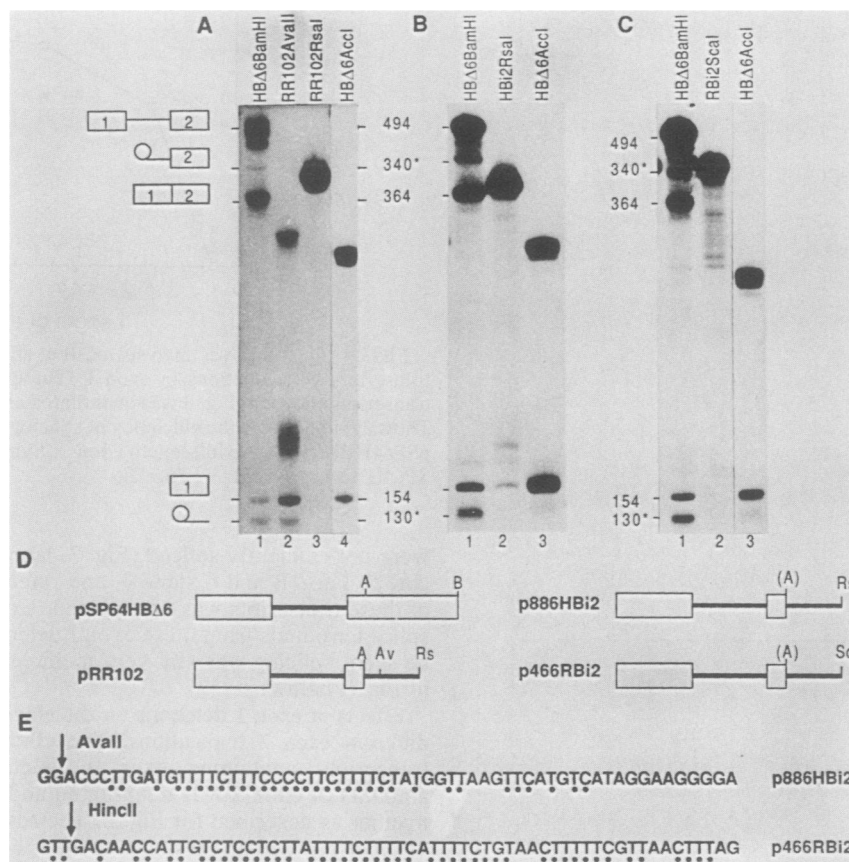


FIG. 5. Analysis of the splicing of transcripts with intron sequences substituted for exon 2. (A) Splicing of transcripts from the template pRR102. Transcript was spliced in a nuclear extract and analyzed as described for Fig. 1. (See Fig. 5D and the labels at the top of each lane for the structure and a description of each transcript, respectively.) The structures on the left and the sizes on the right for the control HBA6BamHI (lane 1) input, intermediates, and products of splicing are the same as for Fig. 1A. Splicing of the HBA6AccI transcript is in lane 4 (see Fig. 1A). (B) Splicing of the HBI2RsaI transcript. Transcript from p886HBI2 was spliced in vitro and analyzed in lane 2 as for panel A. The HBA6BamHI and HBA6AccI controls are in lanes 1 and 3, respectively. (C) Splicing of the RBi2ScaI transcript. Transcript from p466RBI2 was spliced in vitro and analyzed in lane 2 as for panel A. The HBA6BamHI and HBA6AccI controls are in lanes 1 and 3, respectively, with the positions and sizes of the HBA6BamHI input, intermediates, and products of splicing to the left. (D) Structure of the templates. The structures of the templates used to generate the transcripts in panels A, B, and C are shown with the names to the left. Av, *AvaII*; Rs, *RsaI*; Sc, *ScaI*; other symbols and restriction sites are described in the legend of Fig. 1. (E) Nucleotide sequence of the poison sequence found in p886HBI2 and p466RBI2. A portion of the nucleotide sequence from a fragment of intron 2 of the human (GenBank entry HUMHBB, nucleotides 723 to 784) and rabbit (GenBank entry RABHBB1A1, nucleotides 1085 to 1148) β -globin genes, used to construct p886HBI2 and p466RBI2, respectively, is shown. The positions of the *AvaII* and the *HincII* sites used to construct the clones (see Materials and Methods for the construction protocol) are indicated by an arrow. The pyrimidines found in this sequence are indicated with a dot.

products but accumulated the intermediates of splicing (Fig. 5A, lane 4). In contrast, when pRR102 was truncated 30 nucleotides after the 3' splice site (with 15 nucleotides of intron 2 sequences remaining), the transcript was efficiently spliced and the intermediates and products were generated at the same rate as those of other transcripts having a similar length sequence after the 3' splice site (Fig. 5A, lane 2; quantitation not shown). These results suggest that the sequence element involved in the splicing inhibition in pRR102 is contained in a region of intron 2 between the *AvaII*-*RsaI* sites (Fig. 5D). To show that the first 15 nucleotides of the inserted sequence are not involved (a possibility stemming from the fact that this sequence has a portion of the proposed U1-RNP binding site from intron 2 remaining [18, 19]), we constructed a clone, p886HBI2, by placing an intron 2 fragment that begins at the *AvaII* site after the first 14 nucleotides of exon 2. Again, transcripts truncated at the *RsaI* site were not efficiently spliced in vitro (although a small amount of cleavage occurred at the 5' splice site; Fig.

5B, lane 2), identifying the inhibiting sequence within the 70 nucleotides of the *AvaII*-*RsaI* region. We also constructed a clone with a structure similar to that of p886HBI2 by substituting a portion of the rabbit β -globin intron 2 at a point 14 nucleotides from the 3' splice site (p466RBI2). A transcript with 117 nucleotides of intron 2 (RBI2ScaI) was not spliced in vitro (Fig. 5C, lane 2). To ascertain that these results were not due to slower kinetics of splicing, we performed time-course experiments for the RR102RsaI, HBI2RsaI, and RBi2ScaI truncated transcripts. No significant accumulation of splicing products could be detected for any of the transcript after 120 min of incubation, although a small amount of 5' splice site cleavage products could be seen for HBI2RsaI (data not shown).

A comparison of the poison sequence from the human versus the rabbit intron 2 is shown in Fig. 5E. The most striking similarity between these sequences is that they have long stretches of pyrimidines (20 long in the human intron and 26 long with only two purines in the rabbit intron).

A possible mechanism for the inhibition of splicing by the poison sequence is that it competes better for binding to certain splicing factors than the normal binding sites in the pre-mRNA. To test this possibility, we carried out splicing of HBA6 transcripts truncated at either the *Bam*HI or *Acc*I site in the presence of a 2.5-fold excess of transcript HBi2RsaI. We did not attempt competitions with higher concentrations of the transcript containing the poison sequence because we have found that high concentrations of the substrate alone inhibit the splicing reaction. Under these conditions HBiRsaI had no effect on the efficiency of splicing of HBA6 transcripts, suggesting that the poison sequence cannot act in *trans* and is not depleting the extract of an essential splicing factor (data not shown).

DISCUSSION

We have previously shown (7; see also reference 24) that a transcript from the human β -globin gene, with only 14 nucleotides of exon 2 sequence at the 3' end, does not efficiently undergo the final step in splicing, i.e., cleavage at the 3' splice site and exon ligation. The transcript is cleaved at the 5' splice site, accumulating the exon 1 and the intron 1-exon 2 lariat intermediates. The inefficient splicing of this transcript suggested that either there is a minimum size requirement of the second exon for splicing, or an important sequence found in this exon has been deleted in the truncated transcript. The likelihood of this latter possibility was raised by the results of Reed and Maniatis (23), who showed that exon sequences affect the efficiency of splice site selection in a *cis*-competition assay. The results shown here suggest that the inability of this transcript to generate spliced products is due to the lack of a minimum size for the second exon. We also show that splicing can be restored with a variety of non- β -globin-related sequences substituted for exon 2 and that these sequences do not affect the exon 2 minimum size requirement. In contrast to truncations in exon 2, deletions in exon 1, the largest of which reduces the size of exon 1 to 21 nucleotides, have no major effect on splicing in this system. Although a very small increase in the splicing efficiency of transcripts with exon 1 deletions truncated at the *Acc*I site was observed (Fig. 4), it does not seem to change the general pattern of the minimum size requirement for exon 2. All of the above results were quantitated by time-course experiments and show that the inability of certain transcripts to splice is not due to slower kinetics in the *in vitro* reaction. Since, within the limits tested, the overall size of the transcript is unrelated to the efficiency of splicing, it appears instead that an important requirement for efficient splicing is that an intron be present no less than 15 nucleotides upstream from the 3' end of the pre-mRNA.

We presume that the reason for this size requirement is the physical need for a certain length of an exon to act as an anchor for splicing factors that are positioned at the 3' splice site. This is supported by a study that used the Miller chromatin spreading technique to identify 25-nm particles that bind to intron-exon junctions and cover a total of approximately 150 bases of RNA (20). Therefore, we reason that transcripts with less than 50 nucleotides (see Fig. 2) of exon 2 are not spliced well because they are unable to bind efficiently to certain factors in the splicing complex. Transcripts generated with our clones have been examined by P. Goswami and C. Goldenberg (personal communication) for UV cross-linking to proteins present in the splicing complex. As a general pattern, Goswami and Goldenberg found that heterogeneous nuclear RNP core proteins, which have been

shown to be essential for splicing (5, 28), bound to these transcripts in increased amounts as the length of exon 2 in these transcripts (and the efficiency of splicing) increased. Since Bindereif and Green (2a) showed that truncation of the second exon does not affect binding of small nuclear RNPs, it appears that the interaction of heterogeneous nuclear RNP core proteins with exon sequences may contribute to the efficiency of splicing.

Another important consideration is whether a minimum exon size requirement could be involved in splicing *in vivo* since short exons, shorter than 50 nucleotides, do occur in pre-mRNA (for example, the third exon of the troponin I gene is only 7 nucleotides long; see reference 2). For the splicing of a multi-intron pre-mRNA, the results above would first argue that an intron with a short downstream exon cannot be spliced out unless the short exon is first joined to the exon further downstream. This would effectively increase the size of the exon and remove any inhibitory sequences that may be found in the adjacent intron. Thus, it seems plausible that this mechanism could describe a method of ordering the removal of introns.

We have shown by quantitating the efficiency of splicing that most but not all sequences can substitute for exon 2 and allow splicing in the human β -globin gene. The advantages of this procedure are that it uses the standard assay, can be applied to any transcript, and corrects for differences in the initial degradation of the input pre-mRNA seen in different experiments. With this method it is possible not only to identify transcripts which can or cannot splice, but also to observe which transcripts splice at efficiencies somewhere in between these extremes. It is unlikely that this method can be used to differentiate the relative strength of particular sequences that show similar efficiencies of *in vitro* splicing. For these results, a competition assay, such as the one used by Kuhne et al. (13), Lang and Spritz (14), and Reed and Maniatis (23), is more useful for seeing these small differences.

One striking result that we observed is the identification of two sequences, one in the 5' region of the large intron of the human β -globin gene and the other located in approximately the same region of the rabbit β -globin gene, that act as poison sequences when inserted close to the 3' splice site in the downstream exon. As shown above, these sequences completely inhibit splicing when positioned 14 and 31 nucleotides downstream from the 3' splice site (transcripts HBi2RsaI or RBi2ScaI and transcript RR102RsaI, respectively). To see whether the poison sequence could inhibit splicing when positioned further away from the 3' splice site, we generated a transcript from the pSP64HBA6 template truncated at the *Rsa*I site in IVS2. In this transcript, the poison sequence begins 236 nucleotides downstream from the 3' splice site of IVS1 (11 nucleotides from the 5' splice site of IVS2). The efficiency of splicing for this transcript appeared to be lower than the efficiency seen with the *Bam*HI-truncated transcript (data not shown). Since the effect was small, it is not clear whether this sequence affects the efficiency of splicing at a distance.

We do not yet know how the poison sequences can inhibit splicing *in vitro*, since they are found normally in IVS2 of both human and rabbit β -globin and do not appear to inhibit splicing *in vivo*. Moreover, the poison sequence can be deleted from the rabbit IVS2 without affecting splicing *in vivo* (30). The poison sequence may be inhibiting splicing in our constructs by altering the binding of specific factors or affecting the secondary structure of the pre-mRNA. We have looked at computer-generated secondary structures (10) of

the poison sequences alone and in a series of transcripts and can see no common significant features that might suggest that the poison sequence may be blocking essential regions of the molecule by base pairing. Therefore it seems more likely that this sequence interacts with factors in the splicing extract and inhibits formation of the active splicing complex. These factors could be the same ones involved in binding at the 3' splice site near the polypyrimidine stretch and the branch point (3, 4, 8, 26, 29) or a new series of factors that have not yet been identified. Alternatively, the poison sequence may prevent essential factors from binding to the transcript. This idea is supported by the observation of Goswami and Goldenberg that, in contrast to normal β -globin pre-mRNA (HBA6BamHI), transcript HBi2RsaI was not able to cross-link any of the heterogeneous nuclear RNP proteins in the splicing extract (personal communication). In addition, preliminary results suggest that transcripts with poison sequence form complexes that sediment slower than active spliceosomes through glycerol gradients (R. Kole, unpublished data). Understanding how these sequences are inhibiting splicing will determine whether they play a role in splicing *in vivo*.

After this work was completed, Parent et al. (22) reported similar results regarding exon requirements in *in vitro* splicing.

ACKNOWLEDGMENTS

We thank Kathleen Rogers and Jill Stutzman for excellent technical assistance. We thank Robin Reed, Tom Maniatis, and Michael Green (Harvard University) for DNA clones and Carlos Goldenberg (University of Miami) for communicating his results before publication.

This work was supported by a Public Health Service grant (GM32994) from the National Institutes of Health. P.F. is a recipient of a Leukemia Society Postdoctoral Fellowship.

LITERATURE CITED

- Aebi, M., H. Hornig, R. A. Padgett, J. Reiser, and C. Weissmann. 1986. Sequence requirements for splicing of higher eucaryotic nuclear pre-mRNA. *Cell* 47:555-565.
- Baldwin, A. S., E. W. Kittlek, and C. P. Emerson. 1985. Structure, evolution, and regulation of a fast skeletal muscle troponin I gene. *Proc. Natl. Acad. Sci. USA* 82:8080-8084.
- Bindereif, A., and M. R. Green. 1987. An ordered pathway of snRNP binding during mammalian pre-mRNA splicing complex assembly. *EMBO J.* 6:2415-2424.
- Black, D. L., B. Chabot, and J. A. Steitz. 1985. U2 as well as U1 small nuclear ribonucleoproteins are involved in pre-mRNA splicing. *Cell* 42:737-750.
- Chabot, B., D. L. Black, D. M. LeMaster, and J. A. Steitz. 1985. The 3' splice site of pre-messenger RNA is recognized by a small nuclear ribonucleoprotein. *Science* 230:1344-1349.
- Choi, Y. D., P. J. Grabowski, P. A. Sharp, and G. Dreyfuss. 1986. Heterogeneous nuclear ribonucleoproteins: role in RNA splicing. *Science* 231:1534-1539.
- Donis-Keller, H. 1979. Site specific enzymatic cleavage of RNA. *Nucleic Acids Res.* 7:179-192.
- Furdon, P. J., and R. Kole. 1986. Inhibition of splicing but not cleavage at the 5' splice site by truncating human β -globin pre-mRNA. *Proc. Natl. Acad. Sci. USA* 83:927-931.
- Gerke, V., and J. A. Steitz. 1986. A protein associated with small nuclear ribonucleoprotein particles recognizes the 3' splice site of pre-messenger RNA. *Cell* 47:973-984.
- Green, M. R. 1986. Pre-mRNA splicing. *Annu. Rev. Genet.* 20: 671-708.
- Jacobson, A. B., L. Good, J. Simonetti, and M. Zuker. 1984. Some simple computational methods to improve the folding of large RNAs. *Nucleic Acids Res.* 12:45-52.
- Konarska, M. M., R. A. Padgett, and P. A. Sharp. 1984. Recognition of cap structure in splicing *in vitro* of mRNA precursors. *Cell* 38:731-736.
- Krainer, A. R., T. Maniatis, B. Ruskin, and M. R. Green. 1984. Normal and mutant human β -globin pre-mRNAs are faithfully and efficiently spliced *in vitro*. *Cell* 36:993-1005.
- Kuhne, T., B. Wieringa, J. Reiser, and C. Weissmann. 1983. Evidence against a scanning model of RNA splicing. *EMBO J.* 2:727-733.
- Lang, K. M., and R. A. Spritz. 1983. RNA splice site selection: evidence for a 5' to 3' scanning model. *Science* 220:1351-1355.
- Leff, S. E., M. G. Rosenfeld, and R. M. Evans. 1986. Complex transcriptional units: diversity in gene expression by alternative RNA processing. *Annu. Rev. Biochem.* 55:1091-1117.
- Maniatis, T., and R. Reed. 1987. The role of small nuclear ribonucleoprotein particles in pre-mRNA splicing. *Nature (London)* 325:673-678.
- Mitchell, P. J., G. Urlaub, and L. Chasin. 1986. Spontaneous splicing mutations at the dihydrofolate reductase locus in Chinese hamster ovary cells. *Mol. Cell. Biol.* 6:1926-1935.
- Mount, S. M. 1982. A catalogue of splice junction sequences. *Nucleic Acids Res.* 10:459-472.
- Mount, S. M., I. Petterson, M. Hinterberger, A. Karmas, and J. A. Steitz. 1983. The U1 small nuclear RNA-protein complex selectively binds a 5' splice site *in vitro*. *Cell* 33:509-518.
- Osheim, Y. N., O. L. Miller, Jr., and A. L. Beyer. 1985. RNP particles at splice junction sequences on *Drosophila* chorion transcripts. *Cell* 43:143-151.
- Padgett, R. A., P. J. Grabowski, M. M. Konarska, S. Seiler, and P. A. Sharp. 1986. Splicing of messenger RNA precursors. *Annu. Rev. Biochem.* 55:1119-1150.
- Parent, A., S. Zeitlin, and A. Efstratiadis. 1987. Minimal exon sequence requirements for efficient *in vitro* splicing of mono-intronic nuclear pre-mRNA. *J. Biol. Chem.* 262:11284-11291.
- Reed, R., and T. Maniatis. 1986. A role for exon sequences and splice-site proximity in splice site selection. *Cell* 46:681-690.
- Ruskin, B., and M. R. Green. 1985. Role of the 3' splice site consensus sequence in mammalian pre-mRNA splicing. *Nature (London)* 317:732-734.
- Ruskin, B., and M. R. Green. 1985. An RNA processing activity that debranches RNA lariats. *Science* 229:135-140.
- Ruskin, B., and M. R. Green. 1986. Specific and stable intron-factor interactions are established early during *in vitro* pre-mRNA splicing. *Cell* 43:131-142.
- Sharp, P. A. 1987. Splicing of messenger RNA precursors. *Science* 235:766-771.
- Sierakowska, H., W. Szer, P. J. Furdon, and R. Kole. 1986. Antibodies to hnRNP core proteins inhibit *in vitro* splicing of human β -globin pre-mRNA. *Nucleic Acids Res.* 14:5241-5254.
- Tazi, J., C. Alibert, J. Temsamani, I. Reveillaud, G. Cathala, C. Brunel, and P. Jeanteur. 1986. A protein that specifically recognizes the 3' splice site of mammalian pre-mRNA introns is associated with a small nuclear ribonucleoprotein. *Cell* 47:755-766.
- Wieringa, B., E. Hofer, and C. Weissman. 1984. A minimal intron length but no specific internal sequence is required for splicing the large rabbit β -globin intron. *Cell* 37:915-925.