Volume 16 Number 2 1988 Nucleic Acids Research

The dependence of splicing efficiency on the length of 3' exon

A.D.Turnbull-Ross, A.J.Else* and I.C.Eperon

Department of Biochemistry, University of Leicester, University Road, Leicester LE1 7RH, UK

Received November 12, 1987; Revised and Accepted December 17, 1987

ABSTRACT

Oligonucleotide-limited transcription has been used to prepare a series of transcripts which allowed the positions of termination by T7 RNA polymerase to be characterized. The same technique was used to prepare a set of transcripts from a rabbit β-globin gene that extend in intervals of two nucleotides from the 3' splice site of IVS-1 into the second exon. Splicing efficiency in a HeLa cell nuclear extract decreased with decreasing length of the 3' exon, although both steps of the splicing reaction could still be detected with as few as four nucleotides in this exon. No evidence was found for a lower limit to the length of the 3' exon below which splicing would not take place. With longer substrates, the rate of the second step of splicing was increased substantially.

INTRODUCTION

In recent years, cell extracts have been developed that are able to support splicing of exogenous pre-messenger RNAs with high efficiency (1,2,3,4). Coupled with the advent of methods for producing abundant, defined transcripts *in vitro*, these systems have shown that splicing involves at least two steps, based on the appearance and characterization of intermediates (5,6,7). The first step involves the formation of a 2'-5' phosphodiester bond between the branch site adenosine, upstream of the 3' splice site, and the first nucleotide of the intron, the first exon being released; the second step involves joining of the two exons with release of the intron as a lariat. This scheme is supported by evidence that the same mechanism takes place *in vivo* (8,9,10).

Analyses of the role of the known conserved splicing signals in each step of the reaction have shown complex interdependence. Mutations in the 5' splice site consensus region (11,12) of human β-globin IVS-1 resulted in cryptic splice site usage *in vitro* and *in vivo* (1). In contrast, mutations in the highly conserved GT of the 5' site of rabbit β-globin IVS-2 gave rise to an accumulation of lariat intermediates which had, predominantly, reacted at the mutated site - i.e., step 2, rather than step

1, had been blocked (13,14). Mutations introduced into the branch site of the two introns behaved likewise: cryptic sites were used (15), or step 2 was found to be affected more severely than the first step of the reaction (16). Experiments with the 3' splice site have shown that the characteristic polypyrimidine tract (11,12) is essential for both steps of the reaction (17,18) and for formation of the spliceosome complex within which splicing reactions take place (19); mutations of the absolutely conserved AG sequence permitted step 1, but not step 2, in human β-globin IVS-1 (17), and in rabbit β-globin IVS-2 such mutations also severely affected step 1 (13). Although each site participates directly in the reactions of covalent bonds in only one of the two steps, it appears that all the signals are involved in both steps of splicing.

In contrast, experiments have shown that the downstream exon is required for step 2 alone, and that truncated exons prevent only the second step (18,19,20,21). The requirement for a minimum length of the 3' exon is surprising: no consensus sequences have been detected downstream of the 3' splice site, suggesting that there is no specific splicing signal there. A minimum length might be required if the 3' exon were released from the pre-mRNA during the splicing reactions and would need to be retained within the spliceosome, but at present there is no evidence that this is the case. Furthermore, studies with different genes have given different values for the minimum necessary length of the 3' exon: in human β-globin IVS-1, step 2 is blocked with a 3' exon length of 14 nucleotides, detectable with 24 nucleotides and fully efficient at 53 nucleotides (20), whereas in a modified adenovirus major late transcription unit step 1 can be detected in the absence of the exon, step 2 is blocked with an exon length of 4 nucleotides, and a 3' exon length of 12 nucleotides (19) permits step 2.

Without a more accurate determination of the length required for splicing, it is not possible to formulate any models for the role of the 3' exon during the reaction; it is not clear whether there is a precise threshold value for the length, or whether the loss of step 2 during progressive shortening of the exon reflects reductions in efficiency or even just the increasing probability of losing 3' splice site signals from intermediates due to 3' exonuclease activity. For these reasons, we have used oligonucleotide-limited transcription (22) to prepare a series of transcripts with 3' exon lengths increasing in intervals of two nucleotides, and we have used these as substrates in splicing reactions.

MATERIALS AND METHODS

Preparation of transcription templates

The plasmid mICE10[IVS-1] was constructed by insertion of a PvuII-TaqI fragment of a rabbit β-globin gene into mICE10 (22) cut with Sma1 and Acc1. Single-stranded DNA templates were prepared using filtration to remove residual host cells from bacteriophage-containing culture supernatants (23). Deoxyoligonucleotides were synthesised on paper discs (24), and purified on polyacrylamide gels (25). The complementary strand was extended from the oligonucleotide past the T7 promoter region in 5 µl reactions comprising 1x TMS (40 mM Tris-HCl pH 7.5, 6 mM $MgCl_2$, 2 mM spermidine), 0.5 mM of each dNTP, 0.5 µg mICE11 (22) or mICE10[IVS-1] template, 0.15 pmol oligonucleotide and 2 units DNA polymerase 1 (Klenow fragment), for 5 to 15 min at 37°C.

Synthesis of RNA substrates

Transcripts were synthesised by additions to the template reactions to give 1x TMS, 5 mM DTT, 0.25 mM ATP, UTP, GTP, 2.5 M CTP, 1.25 M [α^{32}P]CTP (800 Ci/mmol) (Amersham), 1 mM GpppG (Sigma) with 15 units RNAguard (Pharmacia) and 10 units T7 RNA polymerase (Boehringer), in a final volume of 10 µl. Reactions were at 37°C for 45 min. Transcripts for pCp labelling were produced in similar reactions but with omission of [α^{32}P]CTP and GpppG, and the CTP concentration was increased to 0.25 mM. [^{32}P] pCp was synthesised in 15 µl reactions containing 1x TMS, 1.5 mM 3'-cytidine monophosphate with 60 µCi [γ^{32}P] ATP (3000 Ci/mmol) (Amersham) and 2 units T4 polynucleotide kinase (New England Biolabs), with incubation at 37°C for 15 minutes. The kinase was inactivated by heating to 70°C for 10 minutes and 5 µl of this solution was added to the unlabelled transcript in 37.5 mM Hepes pH 7.5, 15 mM $MgCl_2$, 2.25 mM DTT and incubated for 30 minutes at 37°C with 2 units T4 RNA ligase (Pharmacia).

All transcripts were fractionated on denaturing polyacrylamide gels (7M urea, 20% formamide); material was eluted from gel slices overnight at 4°C in 500 mM sodium acetate pH 4.6, 1 mM EDTA, 0.2% SDS, and then precipitated with ethanol at ambient temperature.

Extract preparation and *in vitro* splicing

Nuclear extracts were prepared from HeLa cells grown in spinner culture, according to reference 26, with modifications (27,28). Splicing reactions were carried out as described in reference 1, except the volume of extract used was reduced from 15 µl to 10 µl and it was supplemented with 5 l S100 fraction. The RNA products were subjected to electrophoresis on 6%

polyacrylamide, 7M urea, 20% formamide gels. Autoradiograms were subjected to densitometric scanning (LKB) with the intensity of a peak expressed as a fraction of the total for all the peaks in each track, and these relative intensities were recalculated, based on the number of C residues in each molecule, to express the results in relative numbers of molecules of each RNA species within each reaction.

Analysis of pCp 3'-end labelled transcripts with ribonuclease T2

The major transcript bands were eluted from polyacrylamide gels as described above and ethanol precipitated. Transcripts were then digested in 50 mM sodium acetate pH 5.2 at 37°C for 75 minutes with 1 unit ribonuclease T2 (Sigma). The products were analysed on polyethyleneimine (P.E.I.)-cellulose plates as described in reference 29, using 2 µg of E.coli $tRNA^{phe}$ digested with ribonuclease T2 as a reference for nucleotide migration.

RESULTS

Transcription and termination

In order to produce a series of transcripts with 3' termini staggered at two nucleotide intervals, the strategy shown in Figure 1 was adopted.

If an oligonucleotide is used to prime DNA synthesis on a single-stranded template across a gene cloned into a mICE vector (22), transcription can be initiated in the opposite sense once the T7 RNA polymerase promoter has been made double-stranded, and this transcription should terminate in the nucleotide complementary to the 5' end of the oligonucleotide. In principle, a series of oligonucleotides with 5' termini staggered at intervals of two nucleotides should give rise to a series of transcripts staggered by the corresponding intervals at their 3' end. It is very important for this work that the 3' end of the transcript should be well-defined.

Definition of the 3' end was achieved by using a series of short transcripts, which could be separated to single nucleotide resolution on a gel and the 3' terminal nucleotide determined for each product. Homogeneous primers (4,5,6,8 and 10 of ref.25), which overlapped with intervals of one or two nucleotides, were used for extension on a mICE11 template, and the transcripts analysed by gel electrophoresis. Figure 2(A) shows that the major transcripts were related by the correct intervals in all but the case of primer 10. In order to assign the point of termination, corresponding transcriptions were performed with unlabelled nucleotide triphosphates, and

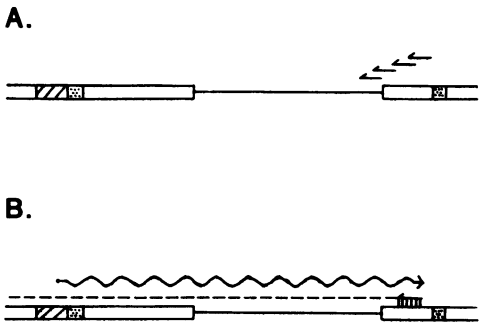

Figure 1. The use of oligonucleotides to specify the 3' terminus of transcription products.
A. Oligonucleotides are synthesized such that the 5' terminus of each oligonucleotide specifies the desired 3' end of a transcript. Each oligonucleotide is annealed to the viral DNA template; shown here is the rabbit β-globin IVS-1 (thin line) and flanking exon sequences cloned into the polylinker sequences (stippled) of an M13 derived vector (mICE 10) with a T7 RNA polymerase promoter (striped).
B. The oligonucleotide is used as a primer for extension (dashed line) by DNA polymerase 1, Klenow fragment; the addition of T7 RNA polymerase and NTPs initiates transcription which, in the case of mICE vectors (22), is dependent on the double-stranded region and in the same sense as the viral DNA. Transcription terminates when the 5' end of the oligonucleotide is reached by the RNA polymerase.

the products were 3' end-labelled with [α-^{32}P]pCp before fractionation. A similar pattern of bands was produced; the major band and the heavy band above it were excised, subjected to digestion with ribonuclease T2, and the products analysed by chromatography on P.E.I.-cellulose. From Figure 2(B), it can be seen that the major transcripts produced using primers 4,5,6 and 8 end in the expected nucleotide; the band above this represents template-independent addition of an extra nucleotide to the 3' terminus, with an approximate order of preference: A>C>G=U. Knowing the identity of the major band in these four cases, it is apparent that the major band derived from use of primer 10 is one nucleotide longer than expected, and the P.E.I.-cellulose analysis confirms that both this band and the faint band above are produced by template-independent nucleotide addition; the proper full-length product is the band below the major band. This interpretation fits the observed pattern of bands in Figure 2(A), even though the transcriptions were performed under slightly different conditions: the concentration of CTP was higher for those in Figure 2(B). The preference of transcripts terminated by primer 10 for

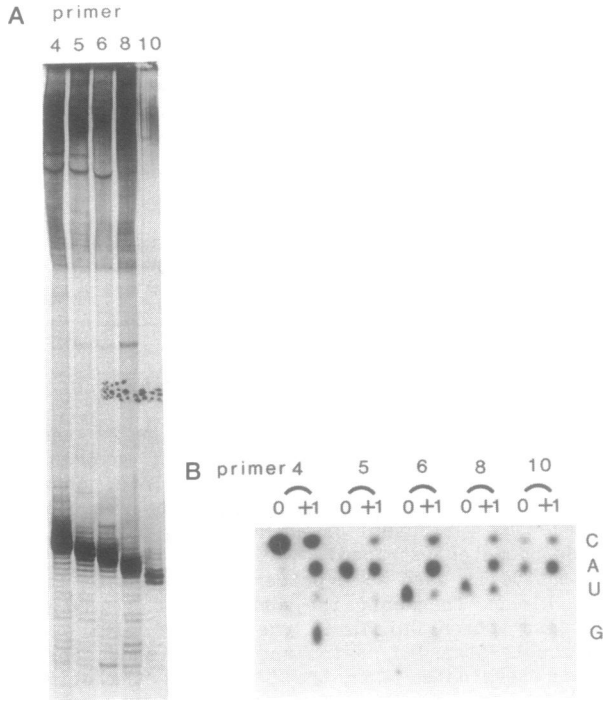

Figure 2. Determination of the 3' terminus of transcripts produced by the strategy shown in Fig. 1.
A. Fractionation of transcripts on an 8% polyacrylamide, 7M urea, 20% formamide sequencing gel. Primer 4 is a "universal" M13 primer; the 5' terminus is 17 nucleotides 5' to the first nucleotide of the M13 polylinker. For primers 5,6,8 and 10 this distance is 16,15,13 and 11 nucleotides respectively.

The sequences of the primers are as follows: primer 4, 5'-GTAAAACGACGGCCAGT; primer 5, 5'-TAAAACGACGGCCAGTG; primer 6, 5'-AAAACGACGGCCAGTGA; primer 8, 5'-AACGACGGCCAGTGAAT; primer 10, 5'-CGACGGCCAGTGAATTC. Transcription terminating with the nucleotide complementary to the 5' nucleotide of the primer would give rise to transcripts of the following lengths, with the expected 3' terminal nucleotide in brackets: primer 4, 73 nucleotides (C); primer 5, 72 nucleotides (A); primer 6, 71 nucleotides (U); primer 8, 69 nucleotides (U); primer 10, 67 nucleotides (G).
B. Chromatography on P.E.I.-cellulose of ribonuclease T2 digestion products. Transcripts produced using the above primers were 3' end-labelled with [α-^{32}P]pCp and fractionated on a gel as in A. The major band ("0") and the one above it ("+1") were excised from each track and digested with ribonuclease T2.

template-independent addition is a characteristic of this primer. Other experiments (data not shown) show that transcription terminating at the site of cleavage by a restriction enzyme gives rise to a similar distribution of 3' termini.

Knowing the pattern of termination, transcriptions were performed with a template derived from cloning a portion of rabbit β-globin gene (from -9 relative to the normal transcription start to +310, i.e., all of exon 1 and IVS-1 and 39 nucleotides of exon 2) into mICE10. Oligonucleotides whose 5' termini lie 0,2,4,6,8 and 10 nucleotides into exon 2 were used to prime DNA synthesis. The transcripts were fractionated on a denaturing polyacrylamide gel; the major products were eluted and analysed by electrophoresis on a second gel. Figure 3 shows that the relative lengths are as expected (DNA sequencing reactions run alongside are too faint to see on this exposure), i.e. 301-311 nucleotides. On the original autoradiograph it can be seen that the products are predominantly in the form of doublets, which, based on the above experiments, should comprise the expected transcript and the product of template-independent nucleotide addition.

Splicing reactions

Transcripts were prepared as above and purified by gel electrophoresis. Samples were then added to a nuclear extract (1,27,28), and incubated at 30°C for three hours. The products were analysed by gel electrophoresis and the result is shown in Figure 4. The mobilities of the 5' exon intermediates and the lariat product are independent of the length of the 3' exon; the lariat-3' exon intermediate and the spliced exons shift with increasing 3' exon length. The bands were assigned on this basis, and the assignments were consistent with their observed mobilities, relative to markers, on gels with different concentrations of acrylamide (not shown).

In order to determine the length of 3' exon involved in each step of the splicing reaction, the products of that step should be sufficiently abundant to be attributable only to reactions of the major species of substrate, of known length ; the lengths of scarce products are not necessarily a reliable indication of the length of the reactant RNA, because longer, spurious transcription products might have undergone the reaction and have been subjected to 3' exonuclease activity subsequently.

The difficulty is exemplified in the case of step 1 with RNA 0. The lariat-3' exon bands are too faint to ascertain whether it is plausible that the reaction could be attributed to the 0 or +1 transcripts, the expected major products of transcription. Step 1 has clearly taken place with much higher efficiency with longer 3' exons, and the abundance of intermediates indicates that the major transcripts are responsible for this. Furthermore, the distributions of lengths in the lariat-3' exon intermediates correlate

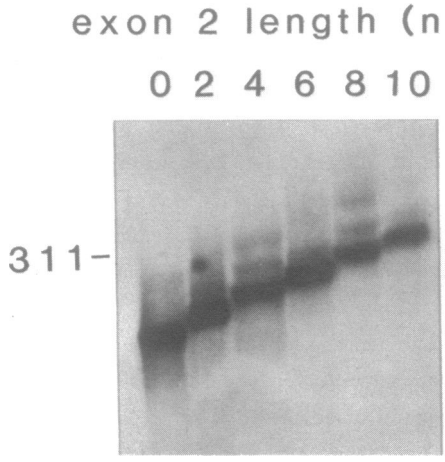

Figure 3. The products of oligonucleotide-limited transcription of β-globin IVS-1.
Transcription reactions were fractionated on a 6% polyacrylamide gel in the presence of 7M urea and 20% formamide. The full length products were eluted for use in splicing reactions, and an aliquot was subjected to the second round of electrophoresis shown here.

very well with the expected lengths of the transcripts. In RNA 2, the intermediate shows a faint band at a length equivalent to lariat +1 (i.e., one nucleotide longer than the intron product after cleavage at the 3' splice site), and strong bands at lariat +2 and +3; this agrees well with the expectation from the transcription data above that an oligonucleotide with a 5' terminus complementary to the second nucleotide in the exon would give rise to major products with 3' exon lengths of two and three nucleotides. Similarly, the major intermediate of RNA 4 corresponds to lariat +4, with a fainter band at +3. However, with RNA 6 the same bands are found, together with a band at +6 and a fainter band at +7. These results are also consistent with the expected pattern of transcription; the nucleotide mix used for transcription contained a low concentration of CTP, and it is possible that the T7 RNA polymerase paused, prior to addition of C at +5, giving rise to the +4 band in lane 6 (although no sign of such a pause can be seen in figure 3). We conclude that the lengths of the lariat-3'exon intermediates seen in Figure 4 do reflect the lengths of the transcripts used as substrate for step 1.

The lariat-3' exon intermediate is a precursor to step 2, and in these

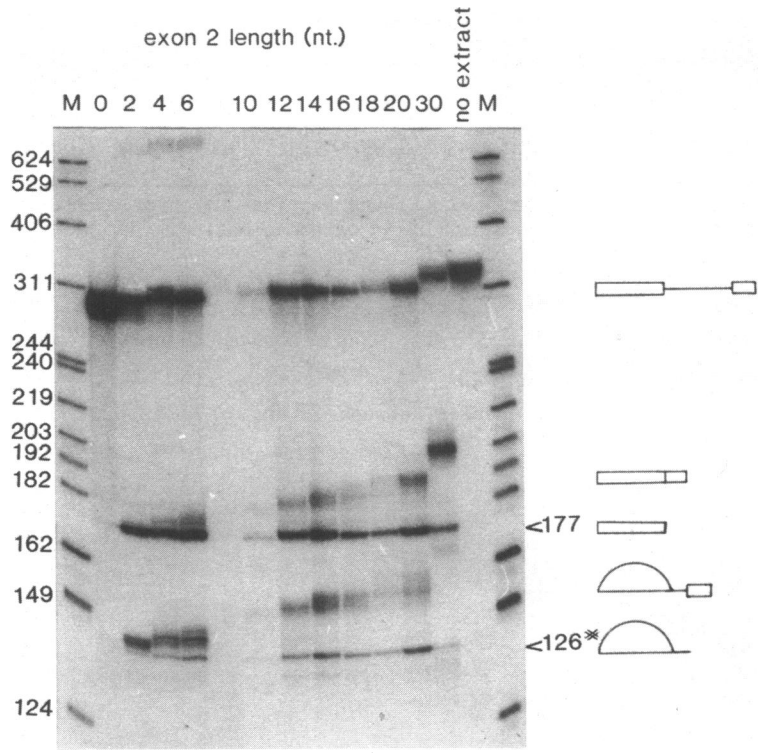

Figure 4. Analysis of splicing reactions performed in vitro with β-globin IVS-1 transcripts with varying lengths of 3' exon. The length of 3' exon (exon 2) indicated is that of the major product expected from use of a primer with a 5' terminus at the corresponding position. M, markers produced by incubating Hpa11 digestion products of pBR322 in the presenceof [α - ^{32}P]dCTP and dGTP and DNA polymerase 1. The structures of the intermediates and products are shown alongside the appropriate band (arrowhead) or bands; boxes indicate exons, and lines, IVS-1.

intermediates the sizes of the shorter 3' exons can be seen clearly. However, although RNA 2 gives rise to a lariat, the spliced exons cannot be seen; in their absence the origin of the apparent product is indeterminate. In contrast, RNA 4, with intermediates with 3' exon lengths of 4 and, to a lesser extent, 3 nucleotides, gives rise to a major spliced product with a 3' exon length of probably 4 nucleotides. For RNA 6, too, the distribution of spliced products matches that of lariat-3' exon intermediates.

The levels of spliced products increase with increasing length of 3' exon, with a corresponding decline in the observable levels of 5' exon intermediate. Table 1 shows the ratio of lariat product to 5' exon

Table 1. The ratio of lariat product to 5' exon intermediate: analysis of data from Figure 4.

RNA:	2	4	6	10	12	14	16	18	20	30
Ratio:	0.11	0.22	0.23	0.63	0.31	0.32	0.43	0.47	0.83	0.40

Laser densitometry was performed on Figure 4. The number of each RNA refers to the length of 3' exon specified by an oligonucleotide.

intermediate in each case. This ratio was chosen because the levels of those species including the 3' exon could not be quantified accurately because of the heterogeneity of 3' end of this exon. The ratio increases as the 3' exon lengthens, and doubles between 18 and 20 nucleotides in the 3'

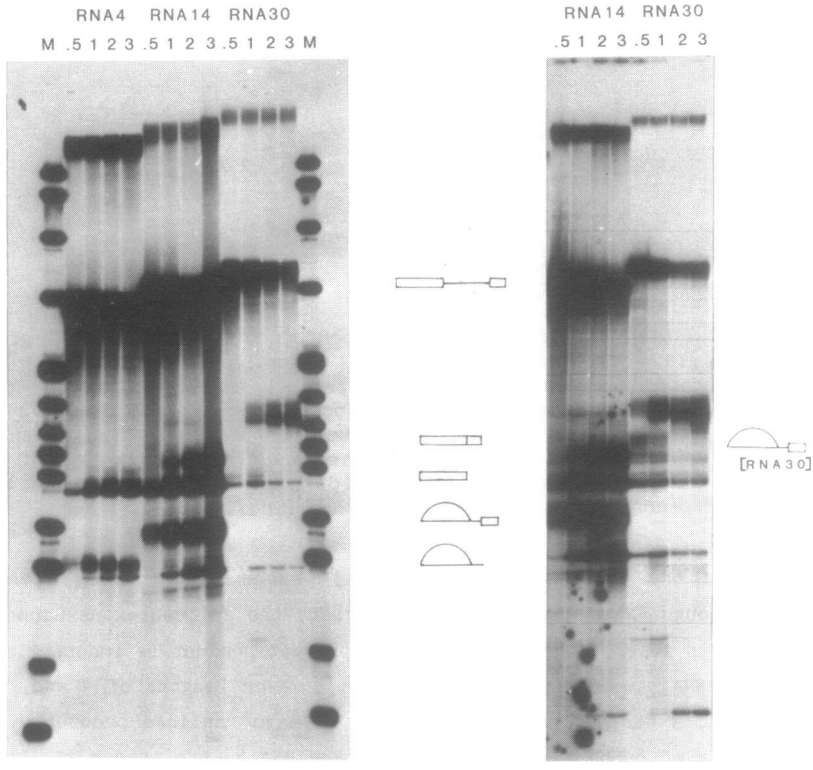

Figure 5. Time courses of splicing in vitro with three transcripts. Times of reaction are shown in hours above each track. Markers and the structures of the reaction products are shown as before. The nature of the bands running near the top of the gel is unknown; note that the transcripts were gel-purified before use (Figure 3).

exon. There are two anomalous points: the bands for RNA 10 are very faint, leading to high errors in measurement, and the lariat product of RNA 30 has been degraded (lariat degradation is significant in this reaction only; see later). If it is assumed that the reactions proceed with a constant or steady-state level of intermediates, and with an unaltered rate of step 1, then the ratio used measures the rate constant for step 2. However, if there is no steady state, the very different ratios could reflect different stages reached in each reaction. Even without more extensive quantitative data, the results shown in Figure 4 and Table 1 strongly suggest that, as 3' exon length increases, the rate of step 2 (V2) increases with respect to that of step 1 (V1). Despite the degradation of lariat products, the very high ratio of intensities of spliced product to 5' exon intermediate in RNA 30 compared with RNA 20 suggests that the increase in V2 relative to V1 is very considerable between RNA 20 and RNA 30. It is not possible to demonstrate whether the rate of step 1 is altered as well as that of step 2.

Time courses were performed with three substrates. Figure 5 shows that spliced products are again found with as few as 4 nucleotides in the 3' exon: with RNA4, 2-4 nucleotides are seen in the precursor to step 2, the lariat-3' exon intermediate. In this preparation of the nuclear extract, a little more degradation of the lariat product is evident, particularly in the reaction with RNA 30, where (with a longer exposure) the ladder of degradation products shows that a shorter lariat accumulates before further 3' exonuclease activity completely removes the tail. It is not clear whether the same activity is responsible for the very diffuse appearance of the ligated exons with the same substrate. The lariat-3' exon intermediate is extremely faint with RNA 30 in one case, and bimodal in the other case, with the upper (correct length) band disappearing with time.

The levels of the intermediates for RNA 30 decline sharply between one and two hours, in contrast to the intermediates for RNAs 4 and 14, the levels of which are roughly constant between times of one and three hours. Together with the abundance of spliced products for RNA 30, this is consistent with a very high ratio of V2 to V1 when compared with the other substrates, despite the scarcity of measurable lariat products.

The rates of the reactions were determined using scanning densitometry of the data in Figure 5. The overall autoradiographic intensity within each track was normalized to one, and the relative proportions of each component calculated, allowing for the number of radioactive nucleotides in each component. This procedure gives rise to skewed data where, as with the

lariat of RNA 30, any component is disproportionately subject to degradation. Plots of this data (not shown) and the fitting of exponential curves suggested that the more rapid approach of the levels of intermediates to their peak levels as the 3' exon was lengthened could be attributed to an increased rate of step 1 or to a decreased lag time, and that it could not be fitted by an increase in the rate of step 2 alone. However, it has proved to be very difficult to establish consistent quantitative values when more detailed time courses have been undertaken.

DISCUSSION

The strategy described in this paper has allowed us to examine whether a clear-cut minimum length is required for the second exon in an *in vitro* splicing reaction, and to test the importance of this aspect of pre-mRNA structure for steps 1 and 2 of the reaction.

The central strategy was to specify the length of the second exon by limiting transcription with oligonucleotides. An essential first step in following this approach is to characterise the products of transcription. Experiments in our model system showed that transcription terminated over a range of positions around the 5' end of the oligonucleotide template, with a very sharp peak at the expected position and after addition of one more (template-independent) nucleotide. A similar phenomenon is also apparent when transcription is limited by restriction endonuclease cleavage of a template. It has been inferred, from experiments where the 3' end of *Tetrahymena* rRNA has been shown to react at the 3' splice site of the upstream intron, that transcription with SP6 RNA polymerase gives rise to a similar product with an extra nucleotide at the 3' end (30). Because of the difficulties inherent in studying the precise termini of transcripts of several hundred nucleotides, this artefact should be borne in mind where short lengths of 3' exon have been studied previously; in particular, it is possible that weak step 1 activity seen with an RNA meant to lack any 3' exon (19) is actually taking place with components of the substrate bearing at least one nucleotide in the 3' exon.

The addition of extra nucleotides during transcription meant that the results of splicing experiments had to be treated with caution. However, there are several reasons why the reactions shown in Figure 4 do not seem to be jeopardised. First, Figure 3 shows that the major transcripts in our reactions seem to be of the correct length or one nucleotide longer. Thus, only with the least efficient reactions is it possible that minor

contaminants could account for the observed splicing patterns. Secondly, these are precisely the reactions where the lengths of lariat-3' exon and ligated product can be assigned most accurately. The lengths of lariat-3' exon forms in such cases match the expected primary products of transcription accurately, again suggesting that step 1 was not followed by nuclease activity. Thirdly, the most important point is whether or not step 2 takes place with small 3' exons, and the length of small 3' exons attached to the lariat-3' exon intermediate, the precursor to step 2, can be seen quite readily. Finally, there is little or no evidence of 3' exonuclease activity in the extract used for the experiment shown in Figure 4, with the exception that the lariat product of the reaction of RNA 30 seems to have degraded.

The results showed that step 1 in the splicing reaction may have taken place at an extremely low level with no 3' exon, but that it takes place efficiently with 3' exon lengths of two or more nucleotides. This agrees with previous data that step 1 takes place with short lengths of 3' exon for which step 2 is inactive (apparently (see above) at 0,3,4,12 and 14 nucleotides (18,19,20,21)). The ability to detect step 2, on the other hand, depends very strongly on length of the 3' exon, and there is a continual increase in the ratio of product to reactant at a fixed time over a range of 3' exon lengths from 0 to 20 nucleotides. This can be seen as a continual increase in the rate of step 2. It is striking that the ratio of lariat product to 5' exon intermediate rises so sharply between RNA 18 and RNA 20; if the ratio of spliced exons to the 5' exon intermediate is considered, an even greater increase in efficiency is seen between RNA 20 and RNA 30. The appearance of step 2 is strongly dependent on the nuclear extract used. The experiments shown in Figures 4 and 5 used extracts prepared in triethanolamine buffer, reported to be preferable to HEPES buffer (28), but, even so, we have found that not all preparations will perform step 2 with 3' exons smaller than 8 or 10 nucleotides. The importance of the extract preparation does not invalidate the observations; our data clearly show that step 2 can take place with 3' exons of about 4 nucleotides, and thus we conclude that there is no reason intrinsic to the splicing reaction mechanism that has given rise to the variety of lengths found to fall below a threshold for step 2 (18,19,20,21). We note that, in recent work with IVS-1 of rabbit β-globin, no evidence was found for step 2 activity with 3' exon lengths of 5 nucleotides; step 2 was found to be

inefficient with 14 nucleotides in the 3' exon and almost fully efficient with 21 nucleotides (34).

The ratio of product lariat to intermediate exon 1 cannot take account of the slight heterogeneity of the 3' termini of each substrate. However, the use of this ratio and examination of the lengths of 3' exon-containing intermediate and product suggest that there is no precise minimum value for the length of 3' exon in order to see step 2; rather, our inability to see definitive products with 3' exons of one or two nucleotides may simply reflect a continuing fall in the efficiency of step 2.

An analysis of the role of step 1 and step 2 in the link between efficiency and 3' exon length was undertaken via time courses. Several points arise from inspection of Figure 5. As in Figure 4, these data represent for each substrate an amalgam of data from a narrow distribution of lengths of 3' exons. The decline in the levels of precursor shows a curious profile. Regardless of the level of unspliced material left, after about two hours very little more precursor enters the reaction path. However, the levels of spliced exons continue to rise at this time. Several explanations are possible. One is that, although step 2 is obviously still proceeding, components essential for step 1 are inactivated with time of incubation; the level of precursor left for each substrate merely reflects the fraction that has undergone step 1 before that step fails. Another explanation is that there is limited recycling of spliceosomes; the proportion of the input RNA incorporated into spliceosomes during the lag phase of splicing may, like the rate of splicing, be related to 3' exon length but, after completion of the reaction, spliceosomes will not reassemble on to fresh substrate molecules . Alternatively, excess substrate molecules may become sequestered, i.e. inaccessible to recycling spliceosomes. We have tried to discriminate between these possibilities by adding a second substrate, with a different 3' exon length, to a splicing reaction which had been proceeding for three hours. With the extract used for the experiments shown in Figure 4, the second substrate was used almost as well when the extract was preincubated (with or without the first substrate) as when there was no preincubation. These results support the suggestion that substrates not bound to spliceosomes upon addition to the reaction become inaccessible.

The susceptibility of the 3' exon of RNA 30 to degradation was evident from the time course experiments, where the lariat-3' exon intermediate was affected as well as the lariat product. The result in Table 1 for RNA 30

has clearly been affected by the substantial degradation of the lariat product. However, the abundance of spliced products, the apparent rise in V2/V1 seen in the appearance of the intermediates and the unlikelihood that V1 has decreased for RNA 30 all suggest that the rate of step 2 is far higher for RNA 30 than RNA 14. It seems that, although the increased length of 3' exon mediates this effect, any association of the 3' exon with reaction components is likely to be transient and that the longer 3' exons are more exposed to 3' exonuclease activity. In contrast to the profile for RNA 30, the time-courses of RNAs 14 and 4 show that the intermediates do not decline, but continue to accumulate, between 0.5 and 1h. This probably arises because the increase in V2 from RNA 4 to RNA 14 is too small to affect the profile. For both these substrates, the levels of intermediates in the time courses suggest that the data in Table 1 could be used to infer the relative rate constants for step 2, at least for the smaller substrates.

We conclude that splicing intermediates form even with very short 3' exons, and that the relative rates of the reactions in step 2, compared with those involved in the lag phase and step 1, are substantially increased as the 3' exon lengthens to 20 and then 30 nucleotides. Furthermore, there is no evidence for a threshold minimum length that is required for step 2. These observations suggest that the 3' exon is involved in non-specific interactions during step 2.

It is clear from the data we have presented that there are two separate processes in splicing which exhibit different requirements for the length of 3' exon sequences. Step 1 is not efficient with RNA 0. Frendewey and Keller (19) have shown that a substrate thought to lack a 3' exon is incorporated with very low efficiency into a 50S spliceosome complex, whereas a substrate with a 3' exon of 4 nucleotides is incorporated efficiently - even though only step 1 of the reaction was seen. Assembly of complete spliceosomes is known to require functional 5' and 3' splice site sequences (19,21,31,32), and assembled spliceosomes show interactions between small nuclear ribonucleoproteins or other components bound at the 5' and 3' splice sites and the branch site (21). Such observations are clearly consistent with the influence of mutations at the 3' splice site on step 1 (13,17). However, assembly requirements cannot account for the relationship between 3' exon length and V2: the range of lengths of 3' exon found to limit the rate of step 2 in our experiments would not be expected to affect spliceosome assembly, and the comparatively efficient rate of step 1 with

our shorter substrates (apart from RNA 0) also suggests that spliceosome assembly is not limiting.

The role of the 3' exon in step 2 is unclear. The lack of a clear threshold length confirms that there are no essential sequences, in agreement with the absence of a simple consensus sequence in the 3' exon. Furthermore, protection experiments based on ribonuclease T1 digestion and immunoprecipitation with anti-Sm or anti-trimethyl guanosine antibodies have shown that 3' exon sequences are not associated with snRNPs after brief incubation of the substrate with nuclear extract on ice (33), and that only a very small proportion of protected fragments at the 3' splice site extend into the exon with longer periods of incubation at 30°C (21). The 3' exon portion adjacent to the 3' splice site is not protected from oligonucleotide-directed RNase H cleavage (J.P. Estibeiro and I.C. Eperon, unpublished work). Thus, if the 3' exon is required for spliceosome assembly, it does not appear to be involved in stable interactions. An attractive possibility is that the 3' exon is involved in a transient association with spliceosome components during step 2 - for example, in facilitating the configurational changes required for nucleophilic attack on the phosphorous at the 3' splice site. Thus, strong, non-specific binding of 3' exon sequences to the spliceosome may be important in activation of the 3' splice site, even if the 3' exon is not released during the reaction.

ACKNOWLEDGEMENTS

We acknowledge gratefully the work of Mr J. Keyte and Mr J. Turner in synthesizing the oligonucleotides used, and financial support from the SERC and MRC. mICE10[IVS-1] was a gift from Mr A.D. Griffiths. A.D.T.-R. is a recipient of an SERC postgradute training award.

*Present address: Laboratoire d'Enzymologie, CNRS, 91190 Gif-Sur-Yvette, France

REFERENCES

1. Krainer, A.R., Maniatis, T., Ruskin, B. and Green, M.R. (1984) Cell 36, 993-1005.
2. Hardy, S.F., Grabowski, P.J., Padgett, R.A. and Sharp, P.A. (1984) Nature 308. 375-377.
3. Hernandez, N. and Keller, W. (1983) Cell 35, 89-99.
4. Goldenberg, C.J. and Hauser, S.D. (1983) Nucleic Acids Res. 11, 1337-1348.
5. Padgett, R.A., Konarska, M.M., Grabowski, P., Hardy, S.F. and Sharp, P.A. (1984) Science 225, 898-903.

6. Ruskin, B., Krainer, A.R., Maniatis, T. and Green, M.R. (1984) Cell 38, 317-331.
7. Konarska, M.M., Grabowski, P.J., Padgett, R.A. and Sharp, P.A. (1985) Nature 313, 552-557.
8. Domdey, H., Apostol, B., Lin, R.-J., Newman, A., Brody, E. and Abelson, J. (1984) Cell 39, 611-621.
9. Rodriguez, J.R., Pikielny, C.W. and Rosbash, M. (1984) Cell 39, 603-610.
10. Zeitlin, S. and Efstratiadis, A. (1984) Cell 39, 589-602.
11. Mount, S. (1982) Nucleic Acids Res. 10, 459-472.
12. Padgett, R.A., Grabowski, P.J., Konarska, M.M., Seiler, S. and Sharp, P.A. (1986) Ann. Rev. Biochem. 55, 1119-1150.
13. Aebi, M., Hornig, H., Padgett, R.A., Reiser, J. and Weissmann, C. (1986) Cell 47, 555-565.
14. Aebi, M., Hornig, H. and Weissmann, C. (1987) Cell 50, 237-246.
15. Ruskin, B., Greene, J.M. and Green, M.R. (1985) Cell 41, 833-844.
16. Hornig, H., Aebi, M. and Weissmann, C. (1986) Nature 324, 589-591.
17. Reed, R. and Maniatis, T. (1985) Cell 41, 95-105.
18. Ruskin, B. and Green, M.R. (1985) Nature 317, 732-734.
19. Frendewey, D. and Keller, W. (1985) Cell 42, 355-367.
20. Furdon, P.J. and Kole, R. (1986) Proc. Natl. Acad. Sci. USA 83, 927-931.
21. Chabot, B. and Steitz, J.A. (1987) Mol. Cell. Biol. 7, 281-293.
22. Eperon, I.C. (1986) Nucleic Acids Res. 14, 2830.
23. Matthes, H.W.D., Zenke, W.M., Grundstrom, T., Staub, A., Wintzerith, M. and Chambon, P. (1984) EMBO J. 3, 801-805.
24. Eperon, I.C. (1986) Analytical Biochem. 156, 406-412.
25. Skinner, J.A. and Eperon, I.C. (1986) Nucleic Acids Res. 14, 6945-6964.
26. Dignam, J.D., Lebovitz, R.M. and Roeder, R.G. (1983) Nucleic Acids Res. 11, 1475-1489.
27. Heintz, N. and Roeder, R.G. (1984) Proc. Natl. Acad. Sci. USA 81, 2713-2717.
28. Tazi, J., Alibert, C., Temsamani, J., Reveillaud, I., Cathala, G., Brunel, C. and Jeanteur, P. (1986) Cell 47, 755-786.
29. Volckaert, G. and Fiers, W. (1977) Analytical Biochem. 83, 222-227.
30. Waring, R.B., Towner, P., Minter, S.J. and Davies, R.W. (1986) Nature 321, 133-139.
31. Perkins, K.K., Furnea, H.M. and Hurwitz, J. (1986) Proc. Natl. Acad. Sci. USA 83, 887-891.
32. Bindereif, A. and Green, M.R. (1987) EMBO J. 6, 2415-2424.
33. Chabot, B., Black, D.L., Le Master, D.M. and Steitz, J.A. (1985) Science 230, 1344-1349.
34. Parent, A., Zeitlin, S. and Efstratiadis, A. (1987) J. Biol. Chem. 262, 11284-11291.