Excision of an Intact Intron as a Novel Lariat Structure during pre-mRNA Splicing In Vitro

Barbara Ruskin, Adrian R. Krainer, Tom Maniatis, and Michael R. Green

Harvard University
Department of Biochemistry and Molecular Biology
7 Divinity Avenue
Cambridge Massachusetts 02138

Summary

To study the mechanisms of RNA splicing we have analyzed the products generated by in vitro processing of a truncated 32P-labeled human β-globin RNA precursor that contains the first two exons and the first intervening sequence (IVS1). Six major RNA products were detected and characterized. The first detectable RNA processing event is cleavage at the 5' GT of IVS1. Subsequently, accurately spliced RNA and the excised, intact IVS1 are simultaneously observed. The IVS1-containing RNA processing products have several unusual properties, which include: anomalous electrophoretic mobilities on polyacrylamide gels; a block to reverse transcription near the 3' end of IVS1; the presence of a nuclease-resistant component within IVS1. The block to reverse transcription and the nuclease-resistant component map to the same site near the 3' end of IVS1. The nuclease-resistant component appears to be a modified adenosine residue that contains an RNA branch. Based upon these and other structural studies we propose that the 5' end of IVS1 is joined by a 2'→5' phosphodiester linkage to the A residue in the RNAase T1 oligonucleotide ACTCTCTCTG located 28–37 nucleotides upstream from the IVS1 3' end. The IVS1 is therefore in the form of a lariat. These results imply that sequences within IVS1 actively participate in splicing.

Introduction

The primary transcripts of most eucaryotic structural genes contain intervening sequences that are removed by RNA splicing. Although the mechanism of RNA splicing is not understood, it is known that accurate and efficient splicing in vivo requires highly conserved sequences located at the intron/exon junctions (Breathnach et al., 1978; Efstathiades et al., 1980; Mount, 1982). This conclusion is based on the analyses of both naturally occurring splicing mutations (for review see Treisman, Orkin, and Maniatis, 1983a) and in vitro-generated mutations (Wieringa et al., 1983). In contrast to higher eucaryotes, a conserved sequence element has been found near the 3' end of all yeast pre-mRNA introns. Deletions or certain single-base substitutions within this sequence abolish splicing (Langford and Gallwitz, 1983; Pikielny et al., 1983; Langford et al., 1984).

A direct approach to the study of pre-mRNA splicing mechanisms was made possible by the development of in vitro splicing systems (Kole and Weissman, 1982; Goldenberg and Hauser, 1983; Padgett, Hardy, and Sharp, 1983; Hernandez and Keller, 1983; Hardy et al., 1984; Krainer et al., 1984). To study splicing of β-globin pre-mRNAs in vitro, human β-globin RNA precursors were synthesized using plasmids in which a promoter from the bacteriophage SP6 is fused to the human β-globin structural gene (Green, Maniatis, and Melton, 1983; Krainer et al., 1984). The fact that these RNA precursors are accurately spliced following microinjection into frog oocytes (Green et al., 1983), provided the opportunity to develop a highly efficient in vitro splicing system (Krainer et al., 1984). Using this system, we observed in addition to accurately spliced RNA a number of RNA processing products, one of which had the characteristics expected of a normal splicing intermediate (Krainer et al., 1984). The existence of another RNA processing product was evidenced by a primer-extension product whose 3' end mapped to a discrete site within IVS1. Surprisingly, the expected RNA cleavage site was not detected by S1 nuclease analysis. These results implied that termination of reverse transcription was due to the presence of a structure that blocks the enzyme, rather than RNA cleavage (Krainer et al., 1984). We suggested that this structure might be an RNA branch such as those detected in nuclear but not cytoplasmic RNA (Wallace and Edmonds, 1983).

To facilitate the further characterization of all the products generated by in vitro processing, we synthesized a high specific activity 32P-labeled SP6/β-globin RNA precursor that contains the first two exons and IVS1. Following in vitro processing, the reaction products were fractionated by gel electrophoresis and isolated for further analysis. Using this approach we detected an IVS1-containing RNA species, likely to be a normal splicing intermediate, and two other RNA species that contain the excised IVS1. Structural analyses of two of these IVS1-containing RNA species provide direct evidence for the presence of an RNA branch within the intron. Based upon these and other results, we propose a model in which the 5' end of IVS1 is joined to a specific site within IVS1 resulting in the formation of a lariat structure. By mapping the site of the branch, we have identified a specific sequence within IVS1 that is involved in pre-mRNA splicing.

Results

In Vitro Processing of a Truncated SP6/β-Globin RNA Precursor

In a previous study we examined the in vitro splicing of a human β-globin pre-mRNA that contains all three exons...
Figure 1. Time Course of In Vitro Splicing

A 32P-labeled RNA precursor was synthesized, processed in vitro, and the 32P-RNA products fractionated by electrophoresis on a denaturing 5% polyacrylamide gel and detected by autoradiography. The reactions were terminated at the times indicated at the top of each lane. M: 32P-labeled markers of Msp I-digested pBR322 DNA. The size of each RNA species, relative to DNA markers, is indicated on the right. Also indicated on the right is the composition of each of the RNA species as determined by RNA mapping procedures. Boxes: exons. Lines: introns. 1 and 2: exons 1 and 2, respectively. The star indicates those RNA species with unusual structural features (see below). The different position of the star in the case of the 130 RNA species indicates that although this RNA has some unusual structural features (see below), it has not yet been characterized in sufficient detail to know if it has an analogous structure to the 143 and 360 RNA species. Below the autoradiogram is shown the SP6/β-globin transcription unit. Thick black line: vector sequences. SP6: the SP6 promoter. Hatched region: vector and β-globin 5' flanking sequences. B: Bam HI site. The arrow indicates the site of transcription initiation within SP6 sequences. The lengths of the exons and introns are indicated.

and both intervening sequences (Krainer et al., 1984). The reaction products were analyzed by S1 nuclease and primer-extension assays. To simplify the analysis and to visualize directly all of the splicing products and intermediates, in the present study we synthesized a truncated, 32P-labeled RNA precursor that contains the first two exons and IVS1 of the human β-globin gene. The plasmid pSP64-H3Δ6 (Krainer et al., 1984) was linearized at the Bam HI site and transcribed with SP6 RNA polymerase, generating a 497 nucleotide 32P-RNA (Figure 1). This truncated β-globin RNA precursor was capped in vitro with vaccinia virus guanylyltransferase (Green et al., 1983), and processed in a HeLa cell nuclear extract as previously described (Krainer et al., 1984). The time course of the in vitro processing reaction is shown in Figure 1. For convenience, we will refer to the various RNA products based upon their electrophoretic mobilities relative to DNA markers on a denaturing 5% polyacrylamide gel. In some cases the assigned number does not correspond to the actual size of the RNA species (see below). The structure of each RNA species shown schematically in Figure 1 was determined by S1 nuclease and primer-extension analysis of the individual gel-purified RNA species.

Three RNA processing products are detected at 30 min, the earliest time examined. The 380 RNA contains IVS1 and exon 2 and the 155 RNA contains exon 1. These two RNA species, which result from cleavage at the 5' GT of IVS1, are likely to be normal splicing intermediates. This possibility is supported by the observations that these two RNA species: appear early during in vitro processing and do not accumulate (Figure 1); are produced by enzymatic activities that require magnesium and ATP, which are also required for accurate splicing (data not shown; see Krainer et al., 1984); and are not observed when the RNA precursor contains a single-base mutation at the 5' GT of IVS1 (see below). Although this evidence suggests that the 380 and 155 RNA species are splicing intermediates, this precursor-product relationship has not been formally demonstrated.

The third RNA product observed at 30 min is the 252 RNA species, which results from an aberrant cleavage event at an AG sequence within IVS1. This cleavage event, which is greatly diminished by capping the RNA prior to in vitro processing, has been previously described (Krainer et al., 1984).

At 60 min, two additional RNA processing products are observed: correctly spliced RNA (370 RNA), and the excised IVS1 (143 RNA). The amount of the 370 and 143 RNA species increases during the 4 hr reaction. The relative molar amounts of these two RNA species are approximately equal throughout the time course, suggesting that they are generated by the same processing event(s). At 2–4 hr an RNA species with an electrophoretic mobility of 130 nucleotides is observed (130 RNA). S1 nuclease mapping and RNA fingerprint analysis indicates that this RNA species is another form of the excised IVS1 (data not shown).

IVS1-Containing RNA Processing Products Have Abnormal Electrophoretic Mobilities

The sizes of 380 and 143 RNA species, established by S1 nuclease and primer-extension experiments, do not correspond to the lengths estimated by gel electrophoresis; the 380 and 143 RNA species are actually 339 and 130 nucleotides, respectively. These discrepancies, although small on 5% polyacrylamide gels, are substantial when these RNA species are analyzed on higher percentage polyacrylamide gels. Variations in the relative electrophoretic mobilities of these IVS1-containing RNA species in
5' and 10% polyacrylamide gels provide a basis for identifying and purifying these RNAs. In Figure 2 the products of an in vitro processing reaction were fractionated by two-dimensional denaturing gel electrophoresis, in which the first and second dimensions are 5% and 10% polyacrylamide gels, respectively. Normal RNAs are aligned on a diagonal because they have the same relative electrophoretic mobilities in the two gels. For example, the unspliced 497 RNA, and the exon-containing 370 and 155 RNA species appear on the diagonal. In contrast, the IVS1-containing 380, 143, and 130 RNA species appear above the diagonal, indicating that these RNA species have a decreased electrophoretic mobility in the second dimension gel. The IVS1-containing RNA processing products therefore appear to contain an unusual structural component.

IVS1-Containing RNA Processing Products Contain a Block to Reverse Transcription

Primer extension experiments were carried out to define the structure of IVS1 in various RNA processing products (Figure 3). The 5' end of IVS1 was mapped using an 18 nucleotide 32P-DNA primer that is homologous to the 3' end of IVS1 (Figure 3A). Primer extension of the 497 full-length RNA precursor produces a discrete 215 nucleotide 32P-cDNA, which represents the distance from the primer to the 5' terminus of the SP6 β-globin transcript (Figure 3A). As expected, a primer-extension product is not observed with the 380 RNA species since they lack IVS1 sequences (Figure 3A). Primer extension of the IVS1-containing 380, 143, and 130 RNA species generates a 57 nucleotide cDNA, indicating that the 5' end of these RNA species was generated by cleavage at the 5' GT of IVS1. This conclusion was confirmed by S1 nuclease analyses (data not shown).

When a DNA fragment from exon 2 was used as a primer for reverse transcriptase, a cDNA product of 242 nucleotides was observed with the 380 RNA (Figure 2B). This product maps a reverse transcriptase stop site near the 3' end of IVS1, corresponding to the block to reverse transcription previously observed with the in vitro-processed β-globin pre-mRNA (Krainer et al., 1984). A 339 nucleotide primer-extension product was not observed with the 380 RNA, indicating that the reverse transcriptase block is present in all of the 380 RNA molecules. As expected, a primer-extension product is not observed with the 155, 143, and 130 RNA species, which lack sequences homologous to the exon 2 primer (Figure 3B). Moreover, primer-extension products of 252 and 367 nucleotides are observed with the 252 and 370 RNA species, indicating that neither of these RNAs contains a reverse transcriptase block. The small amount of 242 nucleotide cDNA observed with the 370 RNA is due to contamination of 370 RNA with the 380 RNA species.

Both the 380 and 143 RNA species display aberrant electrophoretic mobilities (Figure 2), and primer-extension experiments indicate that the two RNAs have the same 5' ends (Figure 3A). It is therefore possible that the 143 RNA contains a block to reverse transcription. To test this possibility, a primer-extension experiment was performed using an 18 nucleotide 32P-DNA primer that is complementary to the 3' end of IVS1 (Figure 3C). A 36 nucleotide primer-extension product was observed with the 380 and 143 RNA species, indicating that both RNAs contain a structure that blocks reverse transcriptase, and that the block is located at the same site within IVS1.

IVS1-Containing RNA Processing Products Contain a Nuclease-Resistant Component

To examine further the nature of the block to reverse transcriptase, we directly analyzed the structure of the in vitro RNA processing products by nuclease digestion and RNA fingerprinting methods. Among many possible RNA modifications that might block reverse transcriptase, one is an RNA branch (Wallace and Edmonds, 1983). As described by Wallace and Edmonds (1983), the 2'-5' phosphodiester linkage that joins the branch to nuclear RNA renders the adjacent 3'-5' phosphodiester bond resistant to digestion by nuclease P1 and T2 (see Table 1). We therefore examined the in vitro RNA processing products for the presence of nuclease-resistant components. Purified RNA processing products, labeled with all four 32P-ribonucleoside triphosphates, were digested to completion with nuclease P1 and fractionated by two-dimensional, thin-layer chromatography (Figure 4; Silberklang, Gillum, and RajBhandary, 1979). The four 5' ribonucleoside monophosphates were observed with the 497, 370, and 155 RNA species. However, in the case of the 380 and 143 RNA species, an additional nuclease P1 digestion product is observed (spot X in Figure 4).
Figure 3. Primer-Extension Analysis of In Vitro RNA Processing Products

The RNA processing products of a standard in vitro processing reaction were purified on polyacrylamide gels and subjected to primer-extension analysis. In all cases the RNA species used for the analysis is indicated at the top of each lane. M, 32P-labeled markers of Msp I-digested pBR322 DNA.

(A) The Primer was an end-labeled synthetic oligonucleotide that is homologous to an 18 nucleotide sequence located at position +182 to +199 of the human β-globin gene (Lawn et al., 1986). The sizes of the primer-extension products are indicated to the left of the autoradiogram, and a diagram of the primer-extension products is shown below the autoradiogram. Box: exon 2. Line: IVS1. The position of the 32P-end label is indicated by a star.

(B) The primer was a 76 nucleotide Bam HI-Hae III fragment (positions +405 to +481 of the β-globin gene), which was 5'-end labeled with 32P at the Bam HI site. The sizes of the primer-extension products are indicated to the right of the autoradiogram. A diagram of the primer-extension products is shown below the autoradiogram.

(C) The primer was an end-labeled 18 nucleotide synthetic oligonucleotide, homologous to the 3' end of IVS1 (positions +255 to +272). The sizes of the primer-extension products are indicated to the left of the autoradiogram, and a diagram of the primer-extension product is shown below the autoradiogram.

380 and 143 RNA species also contain a nuclease T2- and RNAase A-resistant component (see below).

The only two RNA modifications known to confer resistance to nuclease P1 digestion are a 2' phosphomonoester (Konarska et al., 1981) and an RNA branch that is joined by a 2'→5' phosphodiester linkage (Wallace and Edmonds, 1983). If the nuclease P1 resistance is due to the presence of a 2' phosphomonoester, phosphatase treatment will render the RNA susceptible to nuclease P1 digestion (Konarska et al., 1981). In contrast, the 2' phosphate in an RNA branch would not be removed by phosphatase treatment and therefore the RNA would still be nuclease P1-resistant. In fact, we find that the 380 and 143 RNA species contain a nuclease P1-resistant component following phosphatase treatment (data not shown), suggesting that the modification in these RNAs is a 2'→5' phosphodiester linkage.

Localization of a Modified Nucleotide within the Excised IVS1

To localize the nuclease-resistant component within the excised IVS1, the 143 RNA species was digested with RNAase T1, and the products were fractionated by two-dimensional homochromatography (Brownlee, 1972; Vlalkaert, Min Jou, and Fiers, 1976). A 32P-labeled in vitro transcript containing the 130 nucleotide IVS1 and a small amount of flanking β-globin exon sequences was fingerprinted for comparison. This control fingerprint contains all of the unmodified T1 oligonucleotides of IVS1, as well as several additional T1 oligonucleotides that are derived from the β-globin exon and SP6 sequences (Figure 5). A comparison of the T1 fingerprints of the control RNA and the 143 RNA species reveals two differences. First, a unique 10 nucleotide oligonucleotide, ACTCTCTCTG (T1 fragment 12), located 28–37 nucleotides upstream from the
Table 1. Nucleotide Composition of the Nuclease-Resistant Component in the 143 RNA Species

<table>
<thead>
<tr>
<th></th>
<th>P1</th>
<th>T2</th>
<th>P1+T2</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>C</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>G</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>U</td>
<td>-</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>

3' end of IVS1 is shifted down and to the right in the 143 RNA species compared to the same T1 fragment from the control RNA. Second, the 3'-terminal T1 oligonucleotide of IVS1, TCTATTTTCCCACCCTTAG (T1 fragment 13), is shifted up and to the left compared to the same T1 oligonucleotide in the control RNA (Figure 5).

To localize the nuclease P1-resistant sequence within the 143 RNA species, all of the T1 oligonucleotides were recovered from the T1 fingerprint of Figure 5, digested with nuclease P1, and analyzed by thin-layer chromatography. Only T1 fragment 12 contained a nuclease P1-resistant component. Similarly, only T1 fragment 12 contained an RNAase A-resistant component (data not shown and see below). The nuclease-resistant structure within T1 fragment 12 was further localized by RNAase A digestion analysis of this T1 fragment derived from 143 RNA labeled with 32P-UTP or 32P-CTP. A comparison of the RNAase A digestion products of T1 fragment 12 obtained from control and 143 RNA is shown in Figure 6. When the control transcript is labeled with 32P-UTP, the RNAase A digestion products are C, and A-G, as expected from the sequence. Digestion of the 32P-UTP-labeled T1 fragment 12 derived from the 143 RNA species produces C, but not A-G. The A-G digestion product is not observed with the 143 RNA species and is replaced by a new product. These results suggest that either the A or the adjacent C residue contains a 2' modification. If the C residue contains the 2' modification, a nuclease P1-resistant product should be observed with the 32P-UTP-labeled 143 RNA species. However, no such product is observed (see Table 1). The A residue of the ACTCTCTCTG oligonucleotide must therefore contain the 2' modification. We note that 80% of the branched nucleotides in nuclear RNA are adenosines (Wallace and Edmonds, 1983). The location of the modified A residue within the β-globin IVS1 corresponds to the site of the block to reverse transcrip-
Figure 5. RNAase T1 Fingerprint Analysis of the 143 RNA Species

(A) The sequence of the human β-globin IVS1 (Lawn et al., 1980) is shown above the autoradiograms. The expected RNAase T1 cleavage products are bracketed by slashes. The numbers above the predicted T1 oligonucleotides correspond to the numbered spots in the fingerprints.

(B) An SP6 in vitro transcript of pSP64-IVS1, linearized at the Acc I site, and the 143 RNA species, both labeled with 32P-UTP, were digested with RNAase T1 and fingerprinted as described in Experimental Procedures. The directions of the first-dimension, high-voltage paper electrophoresis at pH 3.5, and second-dimension homochromatography on PEI plates, are indicated by arrows. Oligonucleotides in the SP64-IVS1 transcript that are derived from SP6 or β-globin exon sequences are indicated by an S. The T1 oligonucleotide that contains a nuclease-resistant component (see below) is indicated by an arrow on the fingerprint and a star in the sequence.

The T1 fragment 13 in the fingerprint of the 143 RNA has an altered mobility compared to the control RNA but it does not contain a nuclease P1- or RNAase A-resistant component. Although this T1 fragment migrates more slowly than the corresponding fragment from the control RNA on a denaturing 20% polyacrylamide gel, this mobility difference is eliminated by phosphatase treatment of the control T1 fragment 13 (data not shown). Thus T1 fragment 13 derived from the 143 RNA species must contain a 3' hydroxyl group whereas the corresponding T1 fragment from the control RNA contains a 3' phosphate. These results indicate that the 3' terminus of the 143 RNA species contains a 3' hydroxyl group.

IVS1-Containing RNA Processing Products Contain a Circular Component at Their 5' but Not 3' Ends

The experiments described above indicate that the 380 and 143 RNA species contain an RNA branch at a specific site within IVS1. This branch could be an exogenous RNA joined to the linear IVS1 or the 5' end of IVS1 could be joined to the A residue 37 nucleotides upstream of the IVS1 3' end (see diagram in Figure 7). We refer to this novel structure as a lariat. To distinguish between a branched-linear and a lariat structure, IVS1-specific synthetic oligonucleotides and RNAase H were used to cleave the branched RNAs in a site-specific manner (Donis-Keller, 1979). The experimental strategy is outlined in Figure 7. The synthetic oligonucleotide is annealed to the branched RNAs and the DNA–RNA hybrids digested with RNAase.
Thus the 143 RNA species is circular in the region that is of a 39 nucleotide RNA cleavage product rules out the intact 143 nucleotide RNA (Figure 7A, lane 5). The absence of a resistant component is designated by an X. Composition of the RNAase A digestion products is indicated. The RNAase product that migrates more slowly than the acrylamide gel electrophoresis, a single RNA cleavage cleotide, digested with RNAase H, and analyzed by poly-

Figure 6. RNAase A Digestion of the Modified T1 Oligonucleotide from the 143 RNA Species

The 143 RNA species labeled with 32P-UTP or 32P-CTP was digested with RNAase T1 and fingerprinted. Spot 12 was purified, digested with RNAase A, and the products were fractionated by two-dimensional thin-layer chromatography as described in Experimental Procedures. Control refers to oligonucleotide 12 from the pSP64-HIV1 transcript labeled with the corresponding 32P-ribonucleoside triphosphate. The two panels on the left are from transcripts labeled with 32P-UTP. The two panels on the right are from transcripts labeled with 32P-CTP. The labeled phosphates in the sequence above the fingerprints are indicated by stars. The sequence composition of the RNAase A digestion products is indicated. The RNAase A-resistant component is designated by an X.

H. Digestion of a linear-branched RNA species will produce two RNA fragments. In contrast, RNAase H digestion of RNA in a lariat structure will produce a single linear-branched RNA fragment.

Initially we used an 18 nucleotide synthetic DNA fragment whose 3' end is 39 nucleotides from the 5' end of IVS1. Because this oligonucleotide is, by coincidence, homologous to sequences between +182 to +199 in exon 2, RNAase H cleavage of the 497 RNA generates four rather than two products. In addition to the 282 and 197 nucleotide fragments resulting from cleavage within IVS1, the 282 nucleotide fragment is further cleaved to generate fragments of 146 and 128 nucleotides (Figure 7A, lane 2). When the 143 RNA species is hybridized to this oligonucleotide, digested with RNAase H, and analyzed by polyacrylamide gel electrophoresis, a single RNA cleavage product is observed that migrates more slowly than the intact 143 nucleotide RNA (Figur 7A, lane 5). The absence of a 39 nucleotide RNA cleavage product rules out the possibility that the 143 RNA is a branched-linear molecule. Thus the 143 RNA species is circular in the region that is homologous to this oligonucleotide. The 3-ended molecule is probably less compact than the uncleaved circle, which could account for the slower electrophoretic mobility of the RNAase H cleavage product. Model studies comparing the electrophoretic mobilities of circular RNAs to RNAs with multiple ends are consistent with this explanation (Branch and Robertson, personal communication).

A circular component was also detected in the 380 RNA species using the same synthetic oligonucleotide described above to direct site-specific cleavage with RNAase H. In this case the major RNAase H cleavage product migrates more slowly than the intact 380 RNA (Figure 7B, lane 5) and a 39 nucleotide RNA cleavage product is not detected even after long autoradiographic exposures. Thus the 380 RNA species, like the 143 RNA, is not a branched-linear molecule. The band migrating slightly faster than the intact 380 RNA species, and a faint 128 nucleotide fragment in lane 5 of Figure 7, result from homology between the oligonucleotide and a sequence within exon 2 (see above).

In conclusion, the experiments described above indicate that both the 380 and 143 RNA species contain a circular component. This fact, and the presence of an RNA branch at the same site within IVS1 in both of these RNA species indicates that they are in the form of a lariat as shown in Figures 7A and 7B.

To determine whether the structure of IVS1 is the same in the 380 and 143 RNA species, we used a synthetic oligonucleotide homologous to the 3' end of IVS1 to direct site-specific cleavage of both RNAs with RNAase H. As shown in Figure 7C, if both RNA species contain the same circular component, cleavage at the 3' IVS1 junction will generate lariat structures with identical electrophoretic mobilities. In the case of the 143 RNA, a single cleavage product was observed with a faster mobility than the intact RNA species (Figure 7C, lane 4). Significantly, the electrophoretic mobility of this cleavage product is still much less than expected for its size, consistent with the presence of a circular component. A similar RNAase H experiment with the 380 RNA species generates two cleavage products (Figure 7C, lane 6). One of the RNAase H cleavage products from the 380 RNA comigrates with the exon 2 fragment from the 497 RNA and therefore is the linear, unmodified exon 2. The second cleavage product from the 380 RNA comigrates with the single cleavage product of the 143 RNA species. The identical, anomalous electrophoretic mobilities of these cleavage products indicate that both RNA species contain the same lariat structure.

To determine whether the lariat is formed by covalent linkage of the 5' end of IVS1 to a site within IVS1, we used an oligonucleotide that is homologous to the 5' end of IVS1 to site-direct cleavage of the 380 and 143 RNA species. If this proposed structure is correct, cleavage at the 5' end of IVS1 will convert the lariat to a linear RNA, which will display a significantly increased mobility. As shown in Figure 7D (lanes 5 and 7), RNAase H digestion of the 380 and 143 RNA species generates single cleavage products with electrophoretic mobilities that are dramatically increased relative to the intact RNA species. These electrophoretic mobilities are close to those expected if RNAase H cleavage results in the linearization of IVS1.
Figure 7. Site-Specific Cleavage of the IVS1-Containing RNA Species by RNAase H

The ^32P-labeled purified RNA processing products were cleaved in a site-specific manner using various synthetic oligonucleotides and RNAase H as described in Experimental Procedures. The RNA species analyzed is indicated above each lane. RNAase H digestions were performed with (+) or without (−) hybridization to the synthetic oligonucleotide. The predicted sizes of the full-length input RNA (497) and its RNAase H cleavage products are indicated to the left of each autoradiogram. The apparent sizes relative to DNA markers (lanes M, ^32P labeled markers of Msp I digested pBR322 DNA) differ slightly from the predicted sizes. These differences are due to both the nonlinearity of large RNA fragments on the 10% polyacrylamide gel, and incomplete removal of the RNA sequences homologous to the DNA oligonucleotide by RNAase H.

(A) Site-specific cleavage of the 143 RNA species with a synthetic oligonucleotide homologous to sequences from positions +162 to +199 of the human β-globin gene. Cleavage of the 497 RNA species without (lane 1) or with (lane 2) hybridization to the synthetic oligonucleotide. Due to cross-hybridization of the oligonucleotide to +347 to +352 in exon 2, the 282 nucleotide cleavage product is further cleaved to generate two additional cleavage products with expected electrophoretic mobilities of 146 and 128 nucleotides. Cleavage of the 143 RNA species without (lane 4) or with (lane 5) hybridization to the synthetic oligonucleotide. The diagrams below (A) show the predicted RNAase H cleavage products for a linear or circular-branched structure of the 143 RNA species.

(B) Site-specific cleavage of the 380 RNA species with a synthetic oligonucleotide homologous to sequences from positions +182 to +199. Cleavage of the 497 RNA species without (lane 1) or with (lane 2) hybridization to the synthetic oligonucleotide. Cleavage of the 380 RNA species without (lane 4) or with (lane 5) hybridization to the synthetic oligonucleotide. Cleavage of the 143 RNA species without (lane 4) or with (lane 5) hybridization to the synthetic oligonucleotide. The diagrams below (B) show the predicted RNAase H cleavage products for a linear or circular-branched structure of the 380 RNA species.

(C) Site-specific cleavage of the 143 and 380 RNA species with a synthetic oligonucleotide homologous to sequences from positions +255 to +272. Cleavage of the 497 RNA species without (lane 1) or with (lane 2) hybridization to the synthetic oligonucleotide. Cleavage of the 380 RNA species without (lane 6) or with (lane 7) hybridization to the synthetic oligonucleotide. The diagrams below (C) show the predicted structures of the RNAase H cleavage products.

(D) Site-specific cleavage of the 143 and 380 RNA species with a synthetic oligonucleotide homologous to sequences from positions +143 to +160. Cleavage of the 497 RNA species without (lane 1) or with (lane 2) hybridization to the synthetic oligonucleotide. Cleavage of the 143 RNA species without (lane 4) or with (lane 5) hybridization to the synthetic oligonucleotide. Cleavage of the 380 RNA species without (lane 6) or with (lane 7) hybridization to the synthetic oligonucleotide. The diagrams below (D) show the predicted structures of the RNAase H cleavage products.

and removal of 18 nucleotides from the 380 and 143 RNA species (321 and 112 nucleotides, respectively; see Figure 7D). The observed electrophoretic mobilities (330 and 123 nucleotides, respectively) correspond to RNA molecules slightly larger than expected. This difference may be due to incomplete removal of the RNA branch by RNAase H.

Nucleotide Composition of the Nuclease-Resistant Structure in IVS1

We have shown that the 380 and 143 RNA species contain a circular component that is formed by covalent linkage of the 5' end of IVS1 to a site within IVS1. Moreover, our data suggests that the A residue located 37 nucleotides up-
stream from the 3' end of IVS1 is involved in a 2'-5' phosphodiester linkage. This A may therefore be linked to the 5'-terminal G of IVS1 by a 2'-5' phosphodiester bond. To obtain additional information regarding the structure of the branched nucleotide and to determine the source of the phosphate in the proposed 2'-5' phosphodiester bond, we have examined the nucleotide composition of the nuclease-resistant structure within the 143 RNA species. This was accomplished by labeling the RNA precursor with each of the four nucleoside triphosphates separately, and then determining whether the nuclease-resistant component can be detected in the 143 RNA species. The results of the nuclease-digestion experiments are shown in Table 1, and a diagram of the proposed nuclease-resistant structure is shown at the bottom of the table. As predicted, the nuclease P1-resistant structure is labeled by A, G, and C, but not U, and the nuclease T2-resistant structure is labeled by G, C, and U, but not A. The results of the combined nuclease P1 and T2 digestion experiments are also consistent with the proposed structure. Significantly, these data indicate that the origin of the phosphate in the 2'-5' phosphodiester bond is the 5' phosphate of the 5'-terminal G of IVS1. We note, however, that definitive proof of the chemical structure of the nuclease-resistant component will require further studies.

The fact that the 3'-terminal nucleotide of the excised IVS1 contains a 3' hydroxyl group (Figure 5), and the observation that the 5'-terminal phosphate of IVS1 is conserved (Table 1), indicates that the pre-mRNA splicing endonuclease(s), unlike most ribonucleases, may generate 5' phosphate and 3' hydroxyl termini. However, these results were obtained with a crude extract and must be confirmed by using more purified components. We note that T4 RNA ligase joins RNAs that, like the splicing naturally occurring &thalassemia splicing mutations were processed similarly in vivo and in vitro (Krainer et al., 1984). In particular, we found that a single-base mutation in the first residue of the GT dinucleotide at the 5' end of IVS1, which abolishes normal splicing in vivo (Trunschman et al., 1983b), has the same effect in vitro (Krainer et al., 1984). The loss of normal splicing leads to the activation of three cryptic 5' splice sites but a common 3' splice site. It was therefore predicted, the nuclease-resistant component will require further studies.

Analysis of In Vitro RNA Processing Products Generated from Mutant RNA Precursors

In our previous study on in vitro processing of SP6/β-globin RNA precursors we demonstrated that pre-mRNAs bearing naturally occurring β-thalassemia splicing mutations were processed similarly in vivo and in vitro (Krainer et al., 1984). In particular, we found that a single-base mutation in the first residue of the GT dinucleotide at the 5' end of IVS1, which abolishes normal splicing in vivo (Trunschman et al., 1983b), has the same effect in vitro (Krainer et al., 1984). The loss of normal splicing leads to the activation of three cryptic 5' splice sites both in vivo and in vitro. The faithful in vitro splicing of the mutant RNA precursor provides the opportunity to investigate the structure of IVS1 containing RNA processing products generated by cryptic splicing events. Figure 8 shows the time course of in vitro processing of an SP6/β-globin RNA precursor that contains the G to A substitution at position 1 of IVS1 (Orkin et al., 1982). The major splicing event observed in vitro involves the joining of a 5' cryptic splice site, located 16 nucleotides upstream from the normal IVS1 splice junction, to the normal 3' splice site (Krainer et al., 1984; see diagram below Figure 8A). The structures of these RNA processing products are diagrammatically represented and compared to the corresponding RNA products generated by the normal SP6/β-globin RNA precursor in Figure 8. For each RNA product that is generated by the normal substrate there is an analogous product that is produced by the mutant substrate. For example, correctly spliced RNA and the exon 1 RNA fragment are 16 nucleotides shorter than the corresponding products generated by the normal RNA precursor. In addition, two major and one minor IVS1-containing RNA processing products are detected, which are larger than the corresponding products generated by the normal RNA precursor (Figure 8A).

The normal and cryptic splicing events involve different 5' splice sites but a common 3' splice site. It was therefore of interest to determine whether a branch is formed at the same site within IVS1 of the normal and mutant RNA processing products. The mutant 448 RNA species, which contains IVS1 and exon 2 (equivalent to the 3R0 RNA species from the normal RNA precursor), was analyzed by primer-extension analysis. Using a primer homologous to the middle of IVS1, we detect 73 and 57 nucleotide primer-extension products with the mutant 448 and normal 380 RNA species, respectively. The 73 nucleotide primer-extension product corresponds to the predominant cryptic splice site activated by this thalassemia mutation in vitro (Krainer et al., 1984). To determine whether the 448 RNA contains an RNA branch at the same site as the 380 RNA, an exon 2 UPA primer was annealed to both RNA species and extended with reverse transcriptase. As shown in Figure 8B, a 242 nucleotide primer extension product is observed with both the 380 and 448 RNA species, indicating that both RNAs contain an RNA branch that is located at the same position upstream from the 3' end of IVS1. The same block to reverse transcription was previously observed with the other splicing mutants in IVS1, which activate the same set of cryptic 5' splice sites (Krainer et al., 1984). We conclude that although the 5' ends of the normal and mutant IVS1-containing RNA processing products differ, in each case the 5' end is covalently joined to the same site within IVS1.

Discussion

By characterizing the in vitro processing products of a simple, uniformly labeled human β-globin RNA precursor we have identified a novel lariat structure in which the 5' end of IVS1 is covalently joined to a specific site within IVS1. These conclusions are based on the fact that the IVS1 containing RNA processing products have the following properties:

— they display anomalous electrophoretic mobilities on polyacrylamide gels, suggesting the presence of an unusual structural component.

— they contain a block to reverse transcription near the 3' end of IVS1.
Figure 8. In Vitro Splicing of a Mutant \( \beta \)-Globin RNA Precursor

(A) \(^{32}P\)-labeled SP6/\( \beta \)-globin RNA precursor containing a G to A substitution at position 1 of IVS1 (Orkin et al., 1982; Treisman et al., 1983; Krainer et al., 1984) was synthesized, processed in vitro, and the \(^{32}P\)-RNA products fractionated by electrophoresis on a denaturing 5% polyacrylamide gel and detected by autoradiography. The reaction was terminated at the times indicated at the top of each lane. M, \(^{32}P\)-labeled markers of Msp I-digested pBR322 DNA. The sizes of the RNA products relative to DNA markers in this gel are indicated to the right of the autoradiogram. For comparison, the RNA processing products resulting from in vitro processing of a normal SP6/\( \beta \)-globin RNA precursor are shown. The correspondence between the normal and mutant RNA products is indicated by the arrows. On the right, the composition of each of the RNA species as determined by RNA mapping procedures is shown. Boxes; exons. Lines; introns. 1 and 2; exons 1 and 2, respectively. The star indicates those RNA species with unusual structural features. The diagram below the figure shows the splicing of the normal and mutant SP6/\( \beta \)-globin RNA precursors. Boxes; exons. Lines; introns. Primer-extension analysis of normal and mutant IVS1-containing RNA processing products.

(B) A synthetic oligonucleotide homologous to the sequence between +182 to +199 of the human \( \beta \)-globin gene was used to prime reverse transcription on the normal (NL) 497 RNA species, the normal (NL) 380 RNA species, and the equivalent mutant (MUT) 448 RNA species. The sizes of the primer-extension products are indicated on the left of the autoradiogram. The diagram below the autoradiogram shows the observed primer-extension products for both the normal and mutant RNA species.

(C) A Bam H I to Hae III restriction fragment from exon 2 was used to prime reverse transcription of the same RNA species as in B above. The diagram below the autoradiogram shows the observed primer-extension products. B, Bam H I; H, Hae III.

—they contain a nuclease-resistant component located at the site at which reverse transcriptase is blocked. The nuclease-resistant component includes the 5' phosphate from a G residue, most likely the 5' terminal G of IVS1.

—they contain a circular component at their 5' but not 3' ends.

These observations indicate that IVS1-containing in vitro RNA processing products are in the form of a lariat. However, the nature of the chemical bonds that join the 5' end of IVS1 to the site within IVS1 remains to be proven. The results of our nuclease-digestion experiments, and the nucleotide composition of the nuclease-resistant component within IVS1, are consistent with the structure shown in Table 1, based on the original proposal of Wallace and Edmonds (1983). Their model was based on the resistance of the branch structure to enzymatic and chemical cleavage, and the charge of the resultant products. Formal proof for the structure detected in polyadenylated nuclear RNA (Wallace and Edmonds, 1983) and the analogous structure we have detected in IVS1-containing RNA processing products will require further studies. For example, we cannot rule out the possibility that these structures contain chemical bonds that have not been previously described in RNA molecules.

It is unlikely that the lariat is an artifact of in vitro processing. First, Zeitlin and Efstratiadis (submitted) have recently detected the excised introns of the adult rabbit \( \beta \)-globin in RNA prepared from rabbit fetal liver cells. Structural analyses of the excised introns suggest that they are in the form of a lariat. Moreover, the electrophoretic mobilities of the excised introns detected in vivo (Zeitlin and Efstratiadis, submitted) are the same as those observed with in vitro processed rabbit \( \beta \)-globin pre-mRNA (R. Reed, unpublished observations). Second, there is evidence that
excised yeast introns detected in vivo display certain features that are consistent with a lariat structure (J. Abelson, personal communication; C. Guthrie, personal communication). Finally, in vitro processing of an adenovirus major late RNA precursor generated a 3' RNA fragment containing RNA species that display several of the unusual properties reported here for β-globin RNA (P. Sharp, personal communication).

In contrast to previous reports (Kinniburgh and Ross, 1979; Grosveld, Koster, and Flavell, 1981; Avvedimento et al., 1980), pre-mRNA introns appear to be excised as single RNA fragments in vitro (this report) and in vivo (Zetlin and Efstratiadis, submitted). Likewise, the introns of tRNA precursors (Knapp et al., 1979) and ribosomal RNA precursors (Zaug and Cech, 1980; Green et al., 1981) are excised as single RNA fragments. Although a lariat structure is not known to be associated with the splicing of transfer RNA or ribosomal RNA precursors, excised intron sequences of tRNA precursors (Grabowski et al., 1981) and yeast mitochondrial ribosomal RNA precursors (Halbreich et al., 1980; Arnborg et al., 1980) can be detected as circles.

The Role of Intron Sequences in Splicing

A surprising implication of the presence of an RNA branch at a specific site in IVS1 is that sequences within the intron are involved in the pre-mRNA splicing mechanism. Although splice junction consensus sequences are clearly necessary for accurate and efficient splicing in higher eukaryotes, large deletions within introns have only a negligible effect on splicing, suggesting that sequences within introns are not essential for splicing (Khoury et al., 1979; Gruss and Khoury, 1980; Wieringa et al., 1983). The apparent contradiction between these observations could be explained if, in the absence of the normal branchpoint sequence, alternative sequences could function as branchpoints. To determine whether a consensus branchpoint sequence could be identified within intron sequences, and whether multiple copies of the sequence are present, we examined a number of different β-globin IVS1 sequences for the presence of a sequence similar to the human β-globin IVS1 branchpoint sequence.

The following criteria were established for identifying putative branchpoint sequences. First, the branched nucleotide is an A. This criterion was based on the fact that the branched nucleotide in IVS1 is A, and the observation that the predominant branched nucleotide in polyadenylated nuclear RNA is A (Wallace and Edmonds, 1983). Second, the A must be followed by a pyrimidine. This criterion was based on the fact that a pyrimidine-rich region follows the A in the human β-globin IVS1. Moreover, Wallace and Edmonds (1983) found that the composition of the two nucleotides linked to the branched nucleotide in total nuclear RNA is 50% G and 40% pyrimidine. Assuming the branched structures in total nuclear RNA are analogous to the structure shown in Table 1, the base linked to A through a 2'-5' phosphodiester bond is G. Third, after establishing sequence alignments within the human β-globin IVS1, we learned of a computer-derived consensus sequence located near the 3' splice sites of eucaryotic introns (E. B. Keller and W. A. Noon, personal communication). This consensus, based upon its homology with the yeast TACTAAC box (see below), is CT(A/G)A(C/T) in the case of mammalian introns. Remarkably, this sequence fit many of our independently predicted putative branchpoint sequences. In particular, the known branchpoint sequence in IVS1 of the human β-globin gene precisely matches this consensus sequence. Fourth, the branchpoint must be the first sequence upstream from the 3' splice junction consensus sequence that fulfills the three criteria described above. This criterion is based on the fact that a sequence that fulfills the first three criteria, but is located upstream from the actual branchpoint, does not function as a branchpoint in the human β-globin IVS1.

The results of our analysis of mammalian globin genes is presented in Table 2. The consensus sequence is located between 19 and 37 nucleotides (with an average of 31 nucleotides) upstream from the 3' splice site in all of the IVS1 sequences examined. In all of the β-globin IVS1 sequences except two, the three nucleotides preceding the A are CTG. Putative branchpoint sequences can also be identified in β-globin IVS2 sequences. In the case of the human γ-globin gene a consensus with only one purine can be found 29 nucleotides upstream from the 3' splice site. Moreover, we find a putative branchpoint consensus sequence in both introns of the human α-globin genes; one of these sequences contains only a single purine (Table 2). Considering only globin genes introns, we derived the consensus branchpoint sequence shown in Table 2. This consensus is similar to but not identical with consensus sequences derived by computer analysis of sea urchin, Drosophila, and mammalian introns by Keller and Noon (personal communication). We note that the globin consensus Py-X-Py-T-Pu-A-Py matches the yeast TACTAAC sequence in the arrangement of pyrimidine and purine boxes. The same conclusion was reached independently by Zeitlin and Efstratiadis (submitted).

Although we were able to find a putative branchpoint sequence that fulfills the criteria stated above in all of the mammalian β-globin gene introns examined, the functional significance of these sequences remains to be determined. We are currently testing whether a number of sequences in Table 2 actually function as branchpoints during in vitro processing. In fact, preliminary in vitro splicing studies are consistent with the assignments for the first introns of the human ε, the mouse and rabbit β-, and the human α-globin genes shown in Table 2 (R. Reed, unpublished data). Clearly, it is essential to localize the branchpoints in a variety of introns to establish a functional branchpoint consensus sequence. The assignments in Table 2 were based on the analysis of the human β-globin IVS1, the structure of branched nucleotides in nuclear RNA (Wallace and Edmonds, 1983), comparisons with a variety of other introns, and, finally, on a functional analogy with a yeast...
Table 2. Putative Branchpoint Sequences in Pre-mRNA Introns

<table>
<thead>
<tr>
<th>Intron Sequence</th>
<th>Branchpoint Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>5'- GTACAA -3'</td>
<td>GTNAA</td>
</tr>
<tr>
<td>5'- CTACAA -3'</td>
<td>CTNAA</td>
</tr>
<tr>
<td>5'- TACTAAC -3'</td>
<td>TACTAAC</td>
</tr>
<tr>
<td>5'- TCGAAC -3'</td>
<td>TCGAAC</td>
</tr>
</tbody>
</table>

Shown is an alignment of putative branchpoint sequences near the 3' splice site of a number of mammalian globin genes. The putative branchpoint nucleotide of each intron sequence is set off by a space and underlined. The AG dinucleotide at the 3' splice sites is also underlined. The number above the underlined A residue indicates the number of nucleotides between the putative branchpoint and the AG dinucleotide. The adenine residue underlined in the middle of the H. globin IVS1 sequence fulfills the criteria for a putative branchpoint, but is not utilized in vitro. Preliminary data indicates that the underlined adenine is actually the branchpoint (F. Flood, unpublished data). The mammalian globin gene sequences are abbreviated as follows: Hb, human /-globin, Lawh et al., 1980; H, human /-globin, Spritz et al., 1980; Hb, human /-globin, Slightom et al., 1980; Hc, human /-globin, Bumad et al., 1980; R, rabbit /-globin, Hardison, 1983; Rf3, rabbit /3-globin, Hardison, 1981; Rf1, rabbit /1-globin, Hardison et al., 1979; M, mouse /-globin, Konkel et al., 1978; Hb, human /-globin, Michelson and Orkin, 1990. The TACTAAC sequence that is essential for splicing pre-mRNAs in yeast is also shown (Langford and Gallwitz, 1983). Single-base mutations that inactivate (-) or have no effect (+) on splicing are also shown (Langford et al., 1984). A consensus sequence derived from the putative branchpoint sequences of globin genes is also shown.

The intron sequence that is essential for splicing (see below). These criteria can change as more examples are studied. For example, although A is the predominant branched nucleotide in nuclear RNA, U and C are also found as branched nucleotides (Wallace and Edmonds, 1983). In addition, it is not clear whether a minimum of one or two purines are required for a functional branchpoint. Finally, we do not have direct evidence that the first putative branchpoint sequence upstream from the 3' splice site is always the one that forms a branch.

It is possible that the branchpoint sequence is the functional analog of a highly conserved sequence element in yeast introns. Yeast pre-mRNA introns contain a conserved sequence element located near their 3' ends referred to as the TACTAAC box (Langford and Gallwitz, 1983). It is therefore possible that the two sequences fulfill the same function; the yeast TACTAAC box may function as a branchpoint in yeast pre-mRNA splicing. In fact, primer-extension stops have been mapped to the yeast TACTAAC box (Pikielny et al., 1983; J. Abelson, personal communication; C. Guthrie, personal communication). In addition, the excised intron of yeast pre-mRNAs have anomalous electrophoretic mobilities (J. Abelson, personal communication), suggesting that they contain unusual structures possibly identical with the lariat structures we have described. The yeast pre-mRNA splicing mechanism may therefore be similar to the pre-mRNA splicing mechanism in higher eucaryotes. We speculate that the more stringent requirement for the TACTAAC box in yeast pre-mRNA splicing and the apparent absence of such a requirement for splicing in higher eucaryotes is due to the fact that a functional branchpoint sequence may be present many times in the introns of higher eucaryotes. In fact, as mentioned above, the human /-globin IVS1 contains at least one putative branchpoint sequence in addition to the actual branchpoint. We also find multiple putative branchpoint sequences in other /-globin introns we have examined. Another explanation for the difference between yeast and higher eucaryotes is that the sequence requirements for a functional branchpoint may be more stringent in yeast than in higher eucaryotes. Thus the apparent discrepancy between directed mutagenesis studies showing that most intron sequences in higher eucaryotes can be removed without affecting splicing (Wierenga et al., 1983), and the formation of a branchpoint at a specific sequence within introns, may be explained by the redundancy of putative branchpoint sequences, and the utilization of alternative sequences in the absence of the primary branchpoint. In fact, we have found that removal of the branchpoint sequence in the human /-globin IVS1 by deletion does not prevent splicing, but instead, activates a cryptic branchpoint (unpublished data).

**Significance of the Lariat Structure**

At present, the functional significance of the lariat structure of the /-globin IVS1-containing RNA processing products is not known. We suggest several possibilities. First, branch formation and cleavage at the 5' splice site may be mechanistically coupled. Relevant to this possibility is the fact that we previously detected a small amount of RNA that is cleaved at the 5' splice junction but is unbranched (Krainer et al., 1984). The generation of this RNA species could precede the formation of a lariat, arguing against a coupled mechanism or, alternatively, it could be a breakdown product that results from an activity in crude extracts. In this regard, we have detected an activity in these extracts that can convert gel-purified 380 RNA species to a 339 RNA that no longer contains the block to primer extension (data not shown).

A second possibility for the function of the lariat is that it removes the 5' end of the intron from the splicing
reaction, thus driving the reaction in the forward direction toward exon ligation. Sequencing the 5' end of the intron could also protect the intron from degradation during the splicing reaction, or it could prevent splicing errors by reducing the number of substrates available for ligation in a pre-mRNA containing multiple introns.

Finally, we speculate that lariat formation may play a central role in the interaction between 5' and 3' splice sites. Based on the unproven assumption that the branchpoint occurs near the 3' splice site of all introns, it is possible that the branch is a signal for the splicing endonuclease to cleave at the first downstream AG dinucleotide. Alternatively, the 3' splice site may be involved in determining the location of the branchpoint. The nucleotide at the branch site would then interact with the 5' end of the intron, leading to the formation of the lariat. In both cases, the 3' end of the intron would indirectly interact with the 5' end via the branch site, which may be an important mechanism involved in achieving accurate splicing. A detailed understanding of this mechanism should provide insights into the problem of RNA splice-site selection.

The Pathway of Pre-mRNA Splicing

Based on the data in this report and our previous study on SP6/β-globin pre-mRNA splicing in vitro (Krainer et al., 1984) we propose the following pathway for pre-mRNA splicing (Figure 9). This pathway is based upon the structural characterization of the RNA processing products. We cannot rule out the existence of short-lived intermediates that we have not detected. For example, the pre-mRNA may be covalently modified at the branchpoint or other regions prior to cleavage at the 5' splice site and branch formation. These modifications might occur during the 45–60 min lag period observed in the in vitro splicing reaction (Hernandez and Keller, 1983; Hardy et al., 1984; Krainer et al., 1984). The first detectable processing event is the formation of the pre-mRNA at the 5' splice site. Formation of the lariat may occur concomitantly with cleavage or shortly thereafter. In any case, the cleavage at the 5' splice site generates two splicing intermediates: the linear first exon and the second exon-intron lariat RNA molecule. Presumably, these two intermediates are held in place until the next step, cleavage at the 3' splice site, and ligation of the exons. Since we have never detected an RNA species that is cleaved at the 3' splice site and not ligated to the first exon, these two steps may be coupled reactions, or ligation may rapidly follow cleavage. The products of the splicing reaction are the linear spliced RNA and the excised intron in the form a lariat. The excised lariat is then converted to another RNA species (130 RNA; Figure 1) that is found off the diagonal in the two-dimensional polyacrylamide gel (Figure 2) and is therefore not a simple linear form of the 130 nucleotide IVS1. The structure of this RNA species remains to be established. Using the simple SP6/β-globin RNA precursor and the in vitro splicing assay described here, it should be possible to purify and characterize the individual components involved in pre-mRNA splicing, and thereby establish the validity of this proposed pathway.

Experimental Procedures

Materials

Restriction enzymes were from Promega Biotech or New England Biolabs. SP6 RNA polymerase, RNAase H, polynucleotide kinase, and T4 RNA ligase were from Promega Biotech. Vaccinia virus guanylyltransferase was from Bethesda Research Laboratories. 51 nuclease was from Sigma. Rovosor transcriptase was from Life Sciences, Inc. RNAsin II was from Promega Biotech. RNAsin T1 and nuclease T2 were from Calbiochem. RNase A was from Boehringer Mannheim. Nuclease P1 was from FL Biochemicals. α-32P-NTPs were from ICN and Amersham. γ-32P-ATP was from ICN, PEI, and cellulose plates were EM products.

Plasmids, SP6 Transcriptions, In Vitro Processing Reactions

The SP6/β-globin plasmids, pSP64-HA6 and pSP64-HA5EVS1-1A, were previously described (Krainer et al., 1984; see Figure 1). pSP64-HA5EVS1 was made by inserting the human β-globin sequences from a blunt-ended Bst N1 site at position +136 to the Acc I site at position +266 into the plasmid pSP64 between a blunt-ended Hind III site and the Acc I site. In vitro transcription with SP6 RNA polymerase, the plasmid was linearized by cleavage at the Acc I site. Syntheses of SP6/β-globin 32P-RNA for in vitro processing was a modification of our previously described procedures (Green et al., 1963; Krainer et al., 1984). Twelve-microcarrier transcription reactions contained 2 μg of DNA template, 500 μM of three ribonucleoside triphosphates, and the fourth triphosphate was at 41 μM plus 50 μCi of the same 32P-ribonucleoside triphosphate. Following the transcription reaction, the 32P-RNA was precipitated with 1 μg of RNA carrier and washed with vaccinia virus guanylytransferase as described previously (Green et al., 1963; Krainer et al., 1984). In vitro processing was as described (Krainer et al., 1984) in reactions that contained 50 mM KCl.

S1 Nuclease Mapping and Primer-Extension Analysis

51 nuclease and primer-extension analyses were essentially as described previously (Krainer et al., 1984). For primer-extension analysis with synthetic oligonucleotides, 0.002–1 pmole of 5' 32P-end-labeled oligonucleotide was annealed overnight at 37°C with the RNA sample in 30 μl of 40 mM PIPES (pH 6.7), 0.4 M NaCl, 1 mM EDTA, and 0.2% SDS. The 32P-RNA–DNA hybrids were precipitated with ethanol and extended with reverse transcriptase as previously described (Krainer et al., 1984).

Analysis of RNAs by Two-Dimensional Gel Electrophoresis

32P-RNAs from an in vitro processing reaction were electrophoresed in a 20 cm x 40 cm denaturing 5% polyacrylamide gel. A strip containing the gel lane was excised and placed 4 cm from the bottom of a 35 cm x 40 cm glass plate. Spacers and a second glass plate were added to form a gel mold, followed by pouring a denaturing 10% polyacrylamide gel, which surrounds the strip from the first-dimension gel. The direction of electrophoresis is perpendicular to the first-dimension gel strip, from the bottom to the top of the second-dimension gel.
Site-Specific Cleavage of RNAs with Rnase H

Site-specific cleavage of RNAs with E. coli Rnase H was a modification of the previously described procedure (Donis-Keller, 1979). Unlabelled RNA species purified by two-dimensional gel electrophoresis were incubated with 60 ng/ml of synthetic DNA oligonucleotide in 10 µl of 40 mM Tris-HCl (pH 7.9), 4 mM MgCl₂, 1 mM DTT, and 0.03 µg/ml BSA. The reaction mixture was heated to 50°C for 3 min followed by incubation at 32°C for 30 min, at which time 1 µl of Rnase H was added and the reaction allowed to proceed for 2 hr at 32°C. The RNA cleavage products were fractionated on denaturing 10% polyacrylamide gels.

**Direct RNA Analysis**

RNA fingerprinting of Rnase H T1 digests (Brownlee, 1972) was with homoxynux β using PEI plates (Volckaert et al., 1976). In some cases, Rnase H T1 digests were fractionated by two-dimensional gel electrophoresis (Pedersen and Haseltine, 1980). Rnase H T1 digests containing 10 µg of carrier RNA, 2 µl of Rnase H T1 in 2 µl of 10 mM Tris-HCl (pH 7.5), 1 mM EDTA were at 37°C for 45 min. Rnase A digests containing 10-50 µg of carrier RNA, 10 µg of Rnase H A in 5 µl of 10 mM Tris-HCl (pH 7.5), 1 mM PFA were at 37°C for 1 hr. The Rnase A digestion products were fractionated by two-dimensional thin-layer chromatography on PEI plates (Volckaert and Fiers, 1977). Nuclease P1 digests containing 10 µg of carrier RNA, 2 µg of nuclease P1 in 20 mM sodium acetate (pH 5.6) were at 37°C for 2 hr. Nuclease P1 digestion were fractionated by two-dimensional thin-layer chromatography on cellulose plates (Siberglant et al., 1979). In some cases, the nuclease P1 digestion products were fractionated on cellulose plates using as a solvent 0.1 M sodium phosphate (pH 6.8): ammonium sulfate-n-propanol (100/60/2 v/w/v) (Silberklang et al., 1979). Nuclease T2 digestion products containing 10 µg of carrier RNA, 2 µl of nuclease T2 in 20 mM sodium acetate (pH 5.6) were at 37°C for 2 hr. Nuclease T2 digestion products were fractionated on PEI plates using as a solvent 1 M formic acid adjusted with pyridine to pH 4.3 (Volckaert and Fiers, 1978). For digests with nuclease P1 and T2, the nuclease T2 was added first, and after 2 hr at 37°C nuclease P1 was added, followed by incubation at 37°C for an additional 2 hr. Nuclease P1 plus T2 digestion products were fractionated on cellulose plates using as a solvent 0.1 M sodium phosphate (pH 6.8): ammonium sulfate-n-propanol (100/60/2 v/w/v) (Silberklang et al., 1979).

**Acknowledgments**

We gratefully acknowledge Scott Zettin and Arg Efstratiadis for originally suggesting the idea of the Rnase H experiment of Figure 7, and for their continual open communication throughout this investigation. We also thank R. Reed for discussion and advice, and E. Keller and R. Noon for communicating their work prior to publication. We are grateful to Alison Cowie for technical advice on direct RNA analysis, and R. Myers, R. Treisman, and D. Melton for their comments on the manuscript. B. R. was supported by a National Institute of Health postdoctoral training grant, and A. R. K. by a Harvard University Grant-in-Aid. This work was supported by Departmental start-up funds to M. R. G. and a grant from the National Institutes of Health to T. M.

The costs of publication of this article were defrayed in part by the costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received June 9, 1984

References


and mutant human β-globin pre-mRNAs are faithfully and efficiently spliced in vitro. Cell 36, 993–1005.


