Exon Mutations Uncouple 5′ Splice Site Selection from U1 snRNA Pairing

Bertrand Séraphin* and Michael Rosbash*
*Howard Hughes Medical Institute
and Department of Biology
Brandeis University
Waltham, Massachusetts 02254
1Institut Curie Biologie
Bâtiment 110
Campus Universitaire
91405 Orsay Cedex
France

Summary

It has previously been shown that a mutation of yeast 5′ splice junctions at position 5′ (GUAGU) causes aberrant pre-mRNA cleavages near the correct 5′ splice site. We show here that the addition of exon mutations to an aberrant cleavage site region transforms it into a functional 5′ splice site both in vivo and in vitro. The aberrant mRNAs are translated in vivo. The results suggest that the highly conserved G at the 5′ end of introns is necessary for the second step of splicing. Further analyses indicate that the location of the U1 snRNA-pre-mRNA pairing is not affected by the exon mutations and that the precise 5′ splice site is selected independent of this pairing.

Introduction

Early studies indicated that pre-mRNA splicing occurs through an evolutionarily conserved two-step pathway (reviewed by Green, 1986; Padgett et al., 1986). First, cleavage occurs at the 5′ splice site with concomitant joining of the intron 5′ end to the 2′ OH of a residue in the branch site region of the intron, producing the “free” 5′ exon and the lariat intermediate. Second, cleavage occurs at the 3′ splice site with concomitant exon ligation, resulting in the formation of mature mRNA and liberation of the lariat intron. More recent studies have shown that these two reactions occur in a multicomponent complex called the spliceosome. Despite much progress in elucidating the pathway of spliceosome assembly (reviewed by Maniatis and Reed, 1987; Sharp, 1987), we know relatively little about the accuracy of splice site selection. One of the challenges in the study of nuclear pre-mRNA splicing is to understand the rules and mechanisms that underlie splice junction selection and splicing accuracy.

Zhuang and Weiner (1986) verified the predictions of Lerner et al. (1980) and Rogers and Wall (1980) by demonstrating genetically that U1 snRNA base pairs with the 5′ splice junction. This conclusion was reinforced by the finding that the yeast (Saccharomyces cerevisiae) U1 snRNA also base pairs with the 5′ splice junction (Séraphin et al., 1988; Siliciano and Guthrie, 1988) and that this interaction is required for in vitro splicing and spliceosome assembly (Kretzner et al., 1987; Séraphin et al., 1988).

Thus, U1 snRNA is a likely candidate for a molecule involved in 5′ splice site selection.

The effects of 5′ splice site region mutations on 5′ splice site selection are consistent with this view. In mammalian systems, most of the mutations reduced or inhibited splicing at the normal 5′ splice site with, in many cases, activation of cryptic 5′ splice site dozens of nucleotides away (Felber et al., 1982; Treisman et al., 1982; Krainer et al., 1984). This suggests that for most of these mutants the mutated 5′ splice region is no longer recognized by the splicing machinery. Also, studies involving 5′ splice site duplications showed that the 5′ splice site sequence with the greatest similarity to the consensus is frequently selected for splicing (Eperon et al., 1986; Nelson and Green, 1986; Lear et al., 1990). Taken together, these results suggested that splice site region recognition involves cis competition between pre-mRNA sequences; the splice site region selected is the one with the best interaction with U1 snRNA. This idea was strengthened by statistical analyses that showed that, among the sequences of a primary transcript, those with the highest similarity to the 5′ splice site consensus correspond to natural 5′ splice sites (Oshshima and Gotoh, 1987).

However, several mutations at the 5′ extremity of the large rabbit β-globin intron did not result in the activation of cryptic splice sites in vitro. Mutations of the highly conserved GU to GA, GG, or AU were permissive for the first step of splicing in vitro, leading to the accumulation of lariat intermediate. Mutations to UU or CU led to a shift of the 5′ splice site one nucleotide upstream of the normal 5′ splice site (Aebi et al., 1987). In these latter cases, it seemed that the 5′ splice site mutations were not imped-
undertook a new series of experiments to test this hypothesis, i.e., could we effectively uncouple 5'splice site selection? These experiments are the subject of this report.

Figure 1. Sequence of the Wild-Type and Mutant 5' Exon–Intron Junctions

The sequence of the region around the 5' splice site of each construct is depicted. Names of the respective constructs are indicated on the left. WT corresponds to the pH218 plasmid (Teem and Roobash, 1983). Mutated nucleotides (relative to the wild type) are indicated in bold lowercase letters. Large arrows indicate splice site locations for both the WT and 5'I11 constructs, while the small arrow denotes the position of the aberrant cut detected with the 5'I11 mutant (Jacquier et al., 1985). Positions -3 and -4 relative to the normal 5' splice site are indicated.

mediated and products; Pikielny and Roobash, 1985) and blocked the second step of splicing (reviewed by Woolford, 1989). Changes at position 5 of the intron produced a new phenotype: aberrant cleavage sites were detected 3 to 5 nucleotides upstream, or 5 nucleotides downstream, of the normal splice site (Parker and Guthrie, 1985; Jacquier et al., 1985; Fouser and Friesen, 1988). Interestingly, products of these aberrant cleavage events were unable to proceed through the second step of the splicing reaction and accumulated as reaction intermediates, i.e., as “free” 5' exons and lariat intermediates. To test if this defect was due to mispairing between the mutant 5' splice site and U1 snRNA, mutant U1 snRNAs with restored complementarity were constructed. Although these mutant snRNAs were functional, they could neither correct the splice site selection defect of the mutant pre-mRNAs nor relieve the block at the lariat intermediate (Seraphin et al., 1988; Siliciano and Guthrie, 1988). As we have been able to show that U1 snRNA–pre-mRNA pairing is required at an early step during in vitro spliceosome assembly (Seraphin et al., 1988) to form a stable complex committed to spliceosome assembly and splicing (Legrain et al., 1988; Seraphin and Roobash, 1989a), these data suggested that U1 snRNA–pre-mRNA pairing is probably involved in an early step of 5' splice site region recognition but not in 5' splice site selection in yeast. We undertook a new series of experiments to test this hypothesis, i.e., could we effectively uncouple 5' splice site selection from U1 snRNA–pre-mRNA pairing? The results of these experiments are the subject of this report.

Results

Strategy

Yeast pre-mRNAs carrying mutations at intron position 5 are affected in the efficiency of splicing and in the specificity of 5' exon–intron cleavage (Parker and Guthrie, 1985; Jacquier et al., 1985; Fouser and Friesen, 1988). In the RPS1A intron, mutations of the 5' splice site from GUA-G1 (wild-type) to GUAU (5'I11 mutant; Jacquier et al., 1985) and to GUAUuU (data not shown) caused the appearance of an aberrant cut 3 nucleotides upstream of the normal splice site. However, RNA cleaved at this aberrant position did not undergo the second splicing step and remained as lariat intermediate. We decided to test whether changing the nucleotides directly flanking this aberrant cleavage site would allow the conversion of this cleavage site into a functional splice site, resulting in the production of aberrantly spliced mature RNA.

The aberrant cleavage site in the 5'I11 mutant is flanked by two A residues (Figure 1). Because a G is always found at the beginning of a yeast intron and because a G is also frequently found at the end of the upstream exon, we mutated simultaneously the two As flanking the aberrant cut of the 5'I11 mutant and changed them toGs (5'I11-GG; Figure 1). As a control, we also constructed each of the single mutants containing a single G either at position -4 (5'I11-4G) or at position -3 (5'I11-3G) relative to the normal 5' splice site. Each of these different exon mutations (GG at -3 and -4, G at -4, or G at -3) was also introduced by site-directed mutagenesis into a wild-type intron background. The resulting constructs (called WT-GG, WT-4G, and WT-3G, respectively; Figure 1) were then inserted into the H218 backbone (Teem and Roobash, 1983). Each of the plasmids contains the RPS1A intron and adjacent exons fused to the β-galactosidase coding sequences. This hybrid gene is galactose inducible, through a GAL-CYC1 composite promoter. Extensive primer extension analyses have identified all of the major RNAs derived from this construct (Teem and Roobash, 1983; Rodriguez et al., 1984; Pikielny and Roobash, 1985).

Analysis of the Splice Sites Used In Vivo

To verify that aberrant cleavages still occurred with the new constructs, we analyzed the location of the cleavage site(s) by primer extension using the RB27 intron primer. Extension products revealed the 5' ends of full-length pre-mRNAs or the 5' ends of the various lariat species (Figure 2A), i.e., the analysis detected all of the frequent cleavage sites used during the first splicing step. Only use of the normal cleavage site (Figure 2B; N) was detected in wild-type intron-containing constructs (Figure 2B, lanes 1, 3, 5, and 7; N). As reported previously (Jacquier et al., 1985), two additional faint cDNA bands were also detected in the 5'I11 mutant (Figure 2B, lane 2; A). These correspond to the aberrant cleavage event 3 nucleotides upstream of the normal 5' splice site (the upper band of the doublet is a reverse transcriptase artifact; Jacquier et al., 1985). With the 5'I11-GG mutant, signals corresponding to the use of the normal 5' splice site were dramatically reduced, while three longer extension products were detected (Fig-
Figure 2. Analysis of the Splice Sites Selected In Vivo

(A) Principle of splicing product analysis using the RB27 primer. The intron primer RB27 hybridizes to pre-mRNA, lariat intermediate, and intron lariat. Extension of the end-labeled RB27 primer hybridized to pre-mRNA elongates to the transcription start sites. Extension products corresponding to the lariat intermediate and lariat intron and at the 5' cleavage site.

(B) Cleavage site analysis. RNAs derived from strains carrying the various constructs (lane 1-6), or carrying the vector alone (lane 0), were analyzed by primer extension using the RB27 primer. The name of each construct is indicated at the bottom of the corresponding lane. Extension products corresponding to the pre-mRNA (P), normal cleavage site (N), or aberrant cleavage site (A) are indicated. The signal corresponding to the aberrant cleavage observed with the 5'11 construct was faint but clearly visible on longer exposures.

(C) Principle of splice site location using RB1 and ddG. Extension of the RB1 primer in the presence of ddG will stop opposite the first C encountered in the RNA. Thus this protocol gives a 40 nucleotide long product corresponding to the pre-mRNA and lariat intermediate, and a 36 nucleotide long product corresponding to the mRNA. Aberrant splice site located X nucleotides upstream of the normal splice site will result in extension products being 36 - X nucleotide long. This is true, however, only if the aberrant cleavage site is located between the normal splice site and the first C upstream. The RNA depicted corresponds to constructs containing the GG mutation in the upstream exon (WT-GG or 511-GG), and the aberrant splice shown occurs 4 nucleotides upstream of the normal 5' splice site. Regions corresponding to exon 1, aberrant exon 1' (Exon 1'), intron, and exon 2 are indicated.

(D) Splice site analysis using RB1 and ddG. Total RNA from strains harboring the indicated constructs (lane 1-8) or the plasmid vector alone (lane 0) was analyzed by primer extension using the RB1 primer and ddG. The name of each construct is indicated at the bottom of the corresponding lane. Extension products are labeled P: pre-mRNA and lariat intermediate; M: mRNA; asterisk: aberrant splice products. Accurate sizing indicates that these latter products are 3 or 4 nucleotides shorter than products derived from normally spliced RNA. The products corresponding to the mRNA show slight variation in their migration behavior due to the differences in sequence in the upstream exon. The background bands observed in lane 0 include an extension product from the related RP51B gene (B), which is 41 nucleotides long (Abovitch and Rosbash, 1984) as well as other by-products. Because the RP51A gene has been deleted from the DB745-ΔA host strain, no extension products derive from the endogenous RP51A RNA.
These three extension products correspond to the overlap of two adjacent doublet bands characteristic of cleavage 3 and 4 nucleotides upstream of the normal 5' splice site. Mutants 5'II-4G and 5'II-3G displayed a relatively normal level of wild-type cleaved molecules and an increased use, compared with the 5'II mutant, of molecules aberrantly cleaved at position -3 (Figure 2B, lanes 6 and 8). Also, traces of molecules cleaved at position -4 were detected in the 5'II-4G mutant (Figure 2B, lane 6). It is important to note that although qualitative change in the use of cleavage sites can be assessed by this analysis, quantitative comparisons are not without reservations. This is due to the fact that turnover and stability of the different molecular species may differ appreciably (Jacquier and Rosbash, 1986).

We pursued this analysis by testing if any of these aberrant cleavage sites would allow the production of mRNA. However, because transcription of this reporter gene starts at numerous sites scattered on a ~60 nucleotide long region (Faye et al., 1981; Teem and Rosbash, 1983) that obscure the use of minor splice sites using a simple primer extension protocol, we used an exon 2 primer in the presence of dideoxy-G (ddG; Figure 2C). In this case, a single extension product is generated from each RNA species. Pre-mRNA and lariat intermediate give rise to 40 nucleotide extension products, while normal mRNA gives rise to a 36 nucleotide extension product. Aberrant mRNA, generated by the use of an aberrant splice site X nucleotides upstream of the normal splice site, gives rise to an extension product X nucleotides shorter than the normal mRNA extension product.

RNA analysis indicates that only the normal 5' splice site was used in wild-type intron-containing constructs (Figure 2D, lanes 1, 3, 5, and 7). The same situation was observed for the 5'II mutant (Figure 2D, lane 2) as reported previously (Jacquier et al., 1985). When the 5'II intron mutation was combined with the A to G changes at positions -3 and -4 in the upstream exon (5'II-GG), there was a dramatic decrease in the level of normal mRNA and the appearance of two new extension products, 3 and 4 nucleotides shorter (Figure 2D, lane 4). This result indicates that, in addition to the normal 5' splice site, two new splice sites located 3 and 4 nucleotides upstream (i.e., in front of each of the two G residues introduced in the upstream exon; Figure 1) are being used. In the case of the 5'II-4G and 5'II-3G constructs, splicing occurred mostly at the normal 5' splice site, with traces of RNA spliced 4 or 3 nucleotides, respectively, upstream of the usual location (Figure 2D, lanes 6 and 8).

Although mutating residues around the aberrant cleavage site allowed aberrant splicing to occur, not all the aberrant cleavage sites gave rise to detectable splicing. For example, most of the aberrant cleavage events detected from the 5'II-4G construct occurred at position -3 (see Figure 2B), although aberrant splicing was detectable only at position -4. This suggests that the sequence requirements for cleavage and splicing are different (see Discussion).

Splicing Efficiency and Branchpoint Selection
Combining the 5'II intron mutation with exon mutations (5'II-GG, 5'II-4G, and 5'II-3G) also had a dramatic effect on splicing efficiency (Figure 2B). This is indicated by the increased level, relative to the intron mutation alone, of pre-mRNA extension products (Figure 2B). For the 5'II-GG and 5'II-3G mutants, lower splicing efficiency is also reflected in lower mRNA levels (Figure 2D, lanes 4 and 8).

To compare the levels of the lariat intermediate species, we analyzed the same RNAs by primer extension using the same exon 2 primer described above (RB1), but in the absence of ddG. Because of the reverse transcriptase block at branched nucleotides, this protocol allowed the detection of lariat intermediate molecules (aberrant and normal) as well as the localization of the branchpoint (Figure 3A; Ruskin et al., 1985). Both pre-mRNA and mRNA levels were also estimated and found to be similar to those obtained in the previous experiments (compare Figures 2B and 2D with Figure 3B). The ratio of precursor RNA levels to lariat intermediate levels reflects mainly the efficiency of the first splicing step (Fouser and Friesen, 1986). Results presented in Figure 3B indicate that this reaction was affected for all the mutants containing the 5'II mutation. Furthermore, the lower ratio of mRNA to lariat intermediate levels observed with the 5'II-GG and 5'II-3G constructs indicate that the second splicing step was also affected in these cases. Note that the exon mutations had a major impact on splicing efficiency only when they were combined with the 5'II mutation (see above). Because the exon and intron mutations had synergistic effects on splicing efficiency, they may affect a common splicing step (see Discussion).

For the 5'II-GG constructs, we reproducibly observed two adjacent extension products corresponding to the lariat intermediate (Figure 3B, lane 4). One of these products is identical to the product observed with the wild-type intron, while the second is longer by one nucleotide. This interpretation was supported by analyses using a primer located between the 3' splice site and the branchpoint region, which confirmed the existence of these two species. Furthermore, both of these signals were sensitive to debranching in mammalian cell extracts, indicating that they both result from the branchpoint block to reverse transcriptase rather than from pre-mRNA degradation (data not shown). The results suggest that both of the two adjacent As of the branchpoint sequence (UACUAAC) were used as branchpoint acceptors.

Aberrantly Spliced Messages Are Functional
The experiments described above indicate that some splicing of the 5'II-CC, 5'II-4G, and 5'II-3G constructs occurred a few nucleotides upstream of the normal 5' splice site. However, it was still possible that these aberrantly spliced RNAs were not functional. For example, they might not be released normally from spliceosomes, or they might not be efficiently exported from the nucleus. To test the functionality of these mRNAs, we took advantage of the fact that RNAs spliced 4 nucleotides upstream of
the normal splice site are out of frame for β-galactosidase synthesis. Starting with the WT, WT-GG, and WT-4G constructs and their 5'II derivatives, we constructed six new plasmids (denoted by the suffix FS for frameshifted) in which only the aberrant splicing events would give rise to β-galactosidase (see Figure 4A and Experimental Procedures). Primer extension analysis confirmed that 5' splice site selection in these new constructs was identical to the parental ones (data not shown). The FS plasmids allowed us to estimate the level of functional mRNA produced by aberrant splicing at position -4. β-galactosidase activity produced by the parent plasmids, on the other hand, reflected the use of the aberrant splice site located at position -3 as well as the normal splice site (Figure 4B).

β-galactosidase assays of the parent constructs indicated that exon mutations in a wild-type intron background had no important impact on β-galactosidase expression (less than 1.5-fold). As reported previously, the 5'II mutation alone reduced β-galactosidase expression by a factor of two (Jacquier et al., 1985; note the log scale on the y axis). Combinations of intron and exon mutations had strong synergistic effects on β-galactosidase expression; the 5'II-GG and 5'II-4G constructs lead to reduced β-galactosidase activity relative to wild-type intron controls (WT-GG or WT-4G) and to the wild-type exon control (i.e., the 5'II mutant alone). The most dramatic effect was with the 5'II-GG mutant, which showed a 30-fold decrease in β-galactosidase activity relative to the WT-GG construct and a 17-fold decrease relative to the 5'II mutant. We note that the relative β-galactosidase activity correlated well with the relative mRNA levels observed for each construct (Figures 2 and 3). For the frameshifted plasmids, very low levels of β-galactosidase were detected in strains carrying the WT-FS, 5'II-FS, and WT-4G-FS constructs (Figure 4B). Modest but significant levels of β-galactosidase were detected with the WT-GG-FS and 5'II-4G-FS constructs. These values still correspond to only 0.5%-0.05% of the β-galactosidase produced by translation of correctly spliced mRNA from the same constructs. In contrast, a high level of β-galactosidase was generated from the 5'II-GG-FS construct. In fact, β-galactosidase values produced by the frameshifted and nonframeshifted versions of this gene were very similar (1.7-fold difference). Yet, the total amount of β-galactosidase produced by the 5'II-GG construct (i.e., the sum of the normal reading frame and aberrant reading frame) was only 5% of the β-galactosidase produced by the WT-GG construct. This is consistent with the decreased splicing efficiency of the 5'II-GG construct, resulting in low mRNA levels and high pre-mRNA levels (see above). These results show that the aberrantly spliced mRNAs were fully functional and correlate well with the analysis of splice site selection by primer extension (Figure 2), namely, the 5'II-GG mutant used almost equally the normal splice site and the -4 splice site (Figure 2), and β-galactosidase levels corresponding to the frameshifted and nonframeshifted version of the gene were very similar. (Recall that the frameshifted version reflects the use of the -4 splice site and the nonframeshifted version the use of both the normal and -3 splice sites.) Both analyses
Figure 4. Aberrantly Spliced Messages Are Functional

(A) Construction and sequence of plasmids containing a frameshifted version of the HZ19 gene. The sequences surrounding the intron borders of constructs encoding the normal reading frame and frameshifted version are depicted. The actual sequence shown corresponds to the 5'II-GG construct. The Ddel site used for the linker insertion is indicated. The nucleotides inserted in the frameshifted constructs are shown in bold and underlined. For both constructs, the corresponding mRNA sequence after a normal or aberrant splice event is indicated. Deduced amino acid sequences are indicated below their cognate mRNAs. Three asterisks indicate the presence of a stop codon, whereas the other proteins are in frame with the β-galactosidase coding sequence. Although proteins derived from the normal or frameshifted constructs have slightly different N-terminal amino acid sequences, we assume that they have similar specific activities because it has already been shown that sequence changes in the RP51A part of the fusion protein do not alter its specific activity (Gritz et al., 1995).

(B) Translation of normal and aberrant mRNAs. Histograms depict the β-galactosidase activity for both the normal and frameshifted versions of the various constructs. The name of each construct is indicated below each box. Constructs that have the same exon sequence are shown in a single group. Constructs in different groups with the same intron sequence (WT or 5'II) and the same reading frame are shown in the same tone of gray.

Also revealed a low level of splicing at position -4 for the 5'II-4G construct. Finally, the sensitivity of the β-galactosidase assay allowed the detection of a low level of splicing at position -4 for the WT-GG construct. mRNAs corresponding to this splicing event are probably too rare to permit their detection by primer extension.

Ul snRNA-Pre-mRNA Pairing and In Vitro Splicing

The results presented above indicate that splicing can occur 3 or 4 nucleotides upstream of the normal splice site and produce functional mRNA. There were two likely hypotheses concerning the role of Ul snRNA in this phenomenon. Either use of the aberrant splice sites resulted from a new pairing of the 5' end of Ul snRNA with the mutated exons sequences, or another region of Ul snRNA (and/or other factors) recognized the mutated exon sequences as splice sites and the location of the canonical Ul snRNA-pre-mRNA pairing was unchanged. Although the first hypothesis seemed unlikely (because the newly created 5' splice sites did not show significant complementarity to the 5' end of the Ul snRNA and because in every case there was an equivalent or better pairing between Ul snRNA and the pre-mRNA at the normal 5' splice site; Figure 5a), we decided to distinguish between these two possibilities by altering the Ul-5' splice site pairing. If Ul snRNA pairs independently with the aberrant splice sites and with the normal splice site (e.g., in the case of 5'II-GG with the cleavage sites at positions -4, -3, and the normal location), introduction of the Ul-4U snRNA, which shows lowered complementarity to the aberrant splice sites (Figure 5a; Séraphin et al., 1988), should increase the use of the normal splice site at the expense of the aberrant sites. Alternatively, if Ul snRNA-pre-mRNA pairing still occurred only at the normal 5' splice site, all the splicing events carried out by constructs bearing the 5'II mutation, notably the aberrant splicing events at position -4, should be enhanced in the presence of the complementary Ul-4U snRNA mutant (Figure 5a; Séraphin et al., 1988). Because in vivo effects of Ul mutants are rather weak and difficult to detect (Séraphin et al., 1988; Siliciano and Guthrie, 1988; Séraphin and Rosbash, 1989b), we chose to analyze the effect of the Ul-

Figure 4. Aberrantly Spliced Messages Are Functional

(A) Construction and sequence of plasmids containing a frameshifted version of the HZ19 gene. The sequences surrounding the intron borders of constructs encoding the normal reading frame and frameshifted version are depicted. The actual sequence shown corresponds to the 5'II-GG construct. The Ddel site used for the linker insertion is indicated. The nucleotides inserted in the frameshifted constructs are shown in bold and underlined. For both constructs, the corresponding mRNA sequence after a normal or aberrant splice event is indicated. Deduced amino acid sequences are indicated below their cognate mRNAs. Three asterisks indicate the presence of a stop codon, whereas the other proteins are in frame with the β-galactosidase coding sequence. Although proteins derived from the normal or frameshifted constructs have slightly different N-terminal amino acid sequences, we assume that they have similar specific activities because it has already been shown that sequence changes in the RP51A part of the fusion protein do not alter its specific activity (Gritz et al., 1995).

(B) Translation of normal and aberrant mRNAs. Histograms depict the β-galactosidase activity for both the normal and frameshifted versions of the various constructs. The name of each construct is indicated below each box. Constructs that have the same exon sequence are shown in a single group. Constructs in different groups with the same intron sequence (WT or 5'II) and the same reading frame are shown in the same tone of gray.
Ul snRNA and 5′ Splice Site Selection

(a) Potential pairing of wild-type U1 snRNA and U1-4U at different splice sites. Pairing of nucleotides 3-9 of both wild-type U1 snRNA and U1-4U, at the normal splice site and at the aberrant splice site located 4 nucleotides upstream, is depicted. The sequence around the 5′ splice site corresponds to the 5′I-GG construct. Splice junctions used by this construct are indicated by asterisks. GC, AU, and GU pairs are symbolized by thick, thin, and dotted lines, respectively. Note that the depicted pairing represents only the pairing that conserves the −4 splice site in front of the CC of U1 snRNA, but not necessarily the best potential pairing between U1 snRNA and exon sequences. All other pairing schemes, however, imply that 5′ splice site selection is independent of the canonical U1 snRNA pairing.

(b) In vitro splicing. Products of in vitro splicing of the WT (lanes 1 and 5), WT-GG (lanes 2 and 6), 5′I (lanes 3 and 7), and 5′I-GG (lanes 4 and 8) constructs are shown. Pre-mRNA was incubated in a wild-type extract (lanes 1-4) or a U1-4U-containing extract (lanes 5-8). Bands corresponding to the lariat intermediate (LI), lariat intron (L), pre-mRNA (P), mRNA (M), and free exon 1 (E1) are indicated. Products resulting from an aberrant event are indicated by an asterisk. The lengths of the normal and aberrant free exon 1 were determined by using several labeled oligonucleotides as size markers. Slight differences in the mobility of exon 1 result from the presence of the exon mutations. The different size and heterogeneity of the pre-mRNA and mRNA bands are due to poly(A) addition, which varies between extracts (N. Abovich and B. S., unpublished data).

4U snRNA on splice site selection in vitro where there is a strong assembly phenotype (Séraphin et al., 1988). Also, this approach enabled us to compare splice site selection in vivo and in vitro.

For this analysis we concentrated on the 5′I-GG mutant because it showed the highest level of aberrant splicing in vivo. The WT, WT-GG, and 5′I constructs were selected as controls and the relevant mutations were transferred to T7 promoter–containing plasmids. RNA was analyzed after incubation in splicing extracts derived either from wild-type or U1-4U-containing yeast strains (Séraphin et al., 1988).

Only RNA from the WT and WT-GG constructs was spliced in an extract containing the wild-type U1 snRNA; both products (lariat and mature RNA) and both intermediates (the free 5′ exon and lariat intermediate) were visible (Figure 5b). In an extract containing U1-4U snRNA, splicing intermediates and products were detected from all four pre-mRNAs. Splicing of the WT pre-mRNA in the U1-4U-containing extract gave rise to the same RNAs as those observed in a wild-type extract. As reported previously, in vitro splicing of the 5′I mutant is dependent on the presence of U1-4U snRNA (Séraphin et al., 1988). No aberrant splicing of the 5′I mutant was detected, identical to what was observed in vivo. However, slower migrating lariat intermediates could be detected, consistent with the in vivo observations that some cleavage events occurred at positions upstream of the normal 5′ splice site in the 5′I mutant (retardation of aberrant lariat intermediates is due to the larger “eye” of the lariat). Traces of lariat intermediates corresponding to cleavage events in the upstream exon were also detected with the WT-GG construct after incubation in the U1-4U-containing extract. The corresponding aberrantly spliced mRNA was not detected, probably because of its low abundance.

Most interesting was the aberrant, uniform migration pattern of the products and intermediates corresponding to the 5′I-GG mutant. The free exon 1 was shorter by 4
nucleotides, whereas the mobility of the lariat and lariat intermediate was uniform and retarded. This indicated that the main cleavage site and 5' splice site selected in vitro corresponded to position −4. The small amount of mRNA and slowly migrating lariat intron relative to the large amount of slowly migrating lariat intermediate indicated that, like in vivo, the second splicing step was inefficient. Products and intermediates corresponding to the use of the normal 5' splice site could be detected after long exposures (data not shown), indicating that the normal splice site was also used in vitro, although rarely. Trace amounts of cleavage events and splicing at position −3 were not detected but might well have been obscured by the product of splicing at position −4 and by degradation products. Similar results for all of these constructs were obtained in three wild-type extracts and two U1-4U-containing extracts.

In vitro analyses of spliceosome assembly indicated that detection of all three splicing complexes (Séraphin and Rosbash, 1989a) with 5'II and 5'II-GG constructs required the presence of U1-4U. Otherwise, no effect on spliceosome assembly of the exon GG mutation (either WT-GG or 5'II-GG) was detected (data not shown).

In summary, U1-4U was required for in vitro spliceosome assembly and splicing of the 5'II and 5'II-GG constructs, indicating that U1-snRNA pairing occurred at the location of the normal 5' splice site. It is also interesting to note that the same 5' splice sites were selected in vitro and in vivo, although not always with equal efficiencies.

The above conclusions were supported by in vivo analyses. Plasmids containing either the normal or frameshifted series were introduced into a pair of isogenic yeast strains containing either the wild-type U1 snRNA or U1-4U snRNA (Séraphin et al., 1988). β-galactosidase activities were compared on X-Gal-containing plates (data not shown). For all of the constructs carrying the 5'II mutation (including 5'II-GG-FS and 5'II-4G-FS) we observed a proportionally higher β-galactosidase activity in the U1-4U-containing strain compared with a strain carrying the wild-type U1. These results indicate that the use of the −4 splice site was enhanced (rather than decreased) in the presence of the U1-4U snRNA. We conclude that, in vivo as well as in vitro, aberrant splicing occurs in spite of, and probably requires, U1 snRNA pairing at the normal splice site; no evidence for a potential aberrant pairing was obtained.

Discussion

We have shown that splicing may occur, in vivo and in vitro, at sites located 3 or 4 nucleotides upstream of the normal 5' splice site. Splicing at these sites is significantly enhanced by a mutation at position 5 of the normal 5' splice site. This aberrant splicing occurs at a position with poor complementarity to U1 snRNA, and splicing assays in the presence of the U1-4U mutant indicate that U1 snRNA-pre-mRNA pairing still takes place at the expected position. This indicates that there are other interactions, probably involving the Gs at intron positions 1 and 5, that are required for accurate 5' splice site selection.

These interactions may involve (another region of) U1 snRNA, U1 snRNP proteins, other snRNPs, other splicing factors, or another region of the intron. The involvement of factors other than U1 snRNP in 5' splice site selection is also suggested by a consideration of trypanosome trans splicing. In this organism, U1 snRNP has not been found, yet 5' splice site cleavage occurs at a 5' splice site sequence that resembles that of mammals and yeast (Bruzik et al., 1988; Mottram et al., 1989). Also, recent experiments in mammalian cell extracts indicate that a chimeric substrate containing a trypanosomal 5' leader sequence is accurately spliced in the absence of functional U1 snRNP (Bruzik and Steitz, 1990).

We suggest that, in canonical cis splicing, 5' splice site selection occurs by the successive recognition of the 5' splice site region by several splicing factors during spliceosome assembly. This view is consistent with the results of a chemical modification/interference study that demonstrated an increasing requirement for a consensus 5' splice junction sequence as the pre-mRNA proceeds through the spliceosome assembly and splicing pathway (Rymond and Rosbash, 1988). U1 snRNA is involved in 5' splice site region recognition through the U1 snRNA-pre-mRNA interaction, which occurs at an early phase of spliceosome assembly (Séraphin et al., 1988; Ruby and Abelson, 1988; Séraphin and Rosbash, 1989a). Our results suggest, however, that this interaction is insufficient to define precisely the bond that will be cleaved but only defines a window within which the 5' splice site will be selected (5' splice site region recognition). Accurate 5' splice site selection is achieved by subsequent interactions of the 5' splice site sequence with other splicing factors, resulting in a spliceosome with a precise three-dimensional structure. Repetitive checking of the 5' splice site sequence would account for the high fidelity of splice site selection, despite the fact that 5' splice site sequences are short (6–9 nucleotides) and only partially conserved. Similar kinetic proofreading mechanisms have already been proposed for other highly accurate processes like nucleotide selection during DNA replication or tRNA selection during protein synthesis (Ninio, 1975).

Our analysis allows us also to establish an upper limit for the error rate of 5' splice site selection during in vivo splicing. The β-galactosidase produced by the WT-FS construct reflects the use of aberrant splice sites located in frame (N-1) relative to the normal splice site (i.e., position −1, −4, +2 . . . ). Because this amount is about 10^−4 of the level of β-galactosidase produced by the normal reading frame (WT construct), this gives an approximate indication of splicing accuracy with this construct. However, this value should be considered as a tentative maximum, because much of the activity detected with the WT-FS construct might be due to translational frameshifting or other translation "artifacts" that occur on accurately spliced mRNA.

Several facts are consistent with the view that the aberrant splice site selection and poor splicing efficiency are a consequence of an abnormal structure of the mature spliceosome. First, the effects of exon and intron mutations are synergistic, suggesting that they impact on a
common structure or step. Second, the simultaneous activation of aberrant 5' splice sites and an aberrant branch point by the 5'll-GG mutant suggests that the structural defects are not localized to the 5' splice site region but are of a more general nature. Third, the low splicing efficiency of the mutants probably results from the formation of abnormal spliceosomes with low activity. This interpretation suggests that the phenotypes observed with these constructs result from the mispositioning of a splicing factor(s) due to the exon and intron mutations. Because the substantial pre-mRNA levels of 5'll mutants are not reversed in a U1-4U strain and because in vitro spliceosomes assembly proceeds identically in the presence or absence of exon mutations (data not shown), it is likely that the pre-mRNA accumulation results from a rather late block during in vivo spliceosome assembly. However, we do not know at which step of spliceosome assembly the putative mispositioning takes place. In any case, our study uncovers an unsuspected role of yeast exon sequences in splice site selection and splicing efficiency. A similar conclusion about the importance of exon sequences was previously reached by Reed and Maniatis (1986) for mammalian splicing.

Our results show that splicing can occur, to varying extents, at some positions with noncanonical sequences. In other cases, 5' splice site mutations or the use of aberrant cleavage sites lead to the accumulation of lariat intermediate species that cannot proceed through the second splicing step. A comparison of these two classes of yeast sequences (Figure 6) indicates that splicing continues after the first step only if the first nucleotide of the intron is a G. Because no other nucleotide is absolutely required, we suggest that a G attached to the 2' OH of the branchpoint is the only part of the 5' splice junction required for the second step to take place. Given the resemblance between yeast and mammalian pre-mRNA splicing, it is likely that a G at this position is also necessary for an efficient second splicing step in higher eukaryotes (Aebi et al., 1987). It is interesting to note that substituting the branchpoint A with a G also blocks the second splicing step, resulting in the accumulation of lariat intermediate molecules (Fouser and Friesen, 1986). Thus, the 5' splice site and branchpoint regions are not only recognized and "checked" during spliceosome assembly, but are also re scrutinized after the first splicing reaction.

Our results may also shed some light on the mechanisms of intron evolution. By comparing the structure of a given gene in a variety of species, it has often been observed that homologous introns may not be present at precisely the same position but may be located a few nucleotides apart in the same region of the molecule. Similar cases of intron sliding have been reported for sets of duplicated coding segments in a single organism (Craik et al., 1983; Gilbert et al., 1986; Rogers, 1986). It was difficult to understand how intron sliding has occurred, because it requires the simultaneous inactivation of the old 5' and 3' splice sites and the creation of a new pair of sites (Rogers, 1986). The observation that the 5'll-GG, 5'll-4G, 5'll-3G, and WT-GG constructs use a set of closely spaced 5' splice sites suggests that intron sliding may occur in several steps. First, intron and exon mutations close to the intron 5' end create a new 5' splice site without completely abolishing the activity of the previous one. Second, mutations within or close to the 3' end of the intron would shift the 3' splice site by a few nucleotides. If the newly created
3' and 5' splice sites allow the production of functional protein, mutations destroying the old 5' splice site and improving the new one could then fix the intron in its new location. The observation that splicing can occur outside of the canonical U1 snRNA pairing region suggests that other factors contribute to accurate 5' splice site selection. Using the above constructs, we are now in the position of screening for yeast mutants in genes affecting splicing fidelity, by selecting strains for enhanced or suppressed use of the aberrant 5' splice sites.

**Experimental Procedures**

**Strains and Media**

E. coli strain TG1 was used for plasmid propagation. Yeast strain DB743-Δ (MATα ade1-100 leu2-d3,112 ura3-52 rps5a::LEU) was used for RNA analysis and β-galactosidase assays. Because the chromosomal copy of the RPS1A gene has been disrupted in this strain (Abovich and Rosbash, 1984), there is no background due to the endogeneous intron. Wild-type extracts were derived either from strain EJ101 (MATα trpl-Δ 25 trpl-Δ 112 pep3-Δ 3 trcl-Δ 80) or from strain DB746 (MATα trpl-289 ura3-52 arg4 leu2-d3,112 ade2 ssrt1::LEU2 pBS33 [SNR1 TRP1 ARS1 CEN3]). Extracts containing U1-4U were prepared from strain BS-Y46 (isogenic to BS-Y46 but containing U1-4U; Séraphin et al., 1986). Strains BS-Y46 and BS-Y48 were used to analyze the effects of Δ 114U on vivo. Strains were grown, transformed, and manipulated according to standard protocols (Maniatis et al., 1982; Sherman et al., 1985).

**Plasmids**

Plasmids carrying the wild-type or mutant U1 snRNA genes have been described (Séraphin et al., 1986). To mutate nucleotides flanking the 5' splice site, BamHI-SacI fragments from plasmid pHZ16 (WT; Team and Rosbash, 1983) and from the 51 derivative (Jacquier et al., 1985) were subcloned into the pTZ19R backbone (US Biochemical Corp.), giving pBS7 and pBS87, respectively. Mutagenesis was done following the method of Nechayev and Eckstein (1990) using the following oligonucleotides: DT257, ATACAAAGTGTGAT; DT317, ATACAAGATGGTAT; DT318, ATACAAAGTGTGAT. Presence of the expected mutations was confirmed by sequencing using M13 reverse primer and Sequenase (US Biochemical Corp.). The mutant BamHI-SacI fragments were then reinserted in the pHZ16-A2 backbone (Pikielny et al., 1963) and named pBS122, WTdG; PBS-123, WTdG; pBS124, 51-4G; pBS126, WT-3G; pBS127, WT-3G.

To construct the FS series of plasmid, the pBS7 plasmid was partially cleaved with DdeI to give on average one cut per molecule. DNA extremities were blunt ended with Klenow enzyme and full-sized linear plasmid recovered by electrophoresis through low melting agarose. An unphosphorylated BamHI linker (CGGGATCCCG; New England Biolabs) was added by ligation, and high molecular weight DNA separated from unincorporated linkers by gel electrophoresis before recircularization. After transformation, a plasmid, pBS160, carrying the BamHI linker inserted at the DdeI site just following the RP61A intron was selected. The reading frame of the insertion was confirmed by DNA sequencing using a downstream primer. The structure of the resulting construct in the intron region is depicted in Figure 4A. Finally, a SacI-SacI fragment carrying the inserted linker was transferred into each of the plasmids containing the various intron and exon mutations.

To allow RNA preparation for in vitro splicing analyses, HindIII-SalI fragments carrying the 5' splice site region of the WT, 51, WT-GG, or 51-GG construct were ligated with a SacI-SalI fragment carrying the downstream part of a shortened RPS1A intron derived from plasmid pHZ18-Δ2 into the pTZ19R backbone.

**RNA Extraction and Analysis**

RNA extraction and analysis by reverse transcriptase were done essentially according to Pikielny and Rosbash (1985). The R51 primer complementary to exon 2 (CGCTTGACGTCCTG; Team and Rosbash, 1983), the RB27 primer complementary to the upstream part of the intron (ACGTGTCTTAAAAAGCTC; Jacquier et al., 1985), or the RB45 complementary to part of the intron downstream of the branch point (TAAAAAATTGTTGAAATTG; oligo A in Rymond et al., 1997) was used. Debranching was according to the method of Ruskin and Green (1985).

**β-Galactosidase Assay**

β-galactosidase assays were done according to the protocol of Leclair and Rosbash (1989). Analysis of β-galactosidase production on X-Gal-containing plates was according to Rose et al. (1981).

**In Vitro Splicing**

RNAs were transcribed in vitro using T7 RNA polymerase (US Biochemical Corp.) essentially according to the procedure of Melton et al. (1984). For extract preparation and in vitro splicing, we followed the protocol of Lin et al. (1985). Splicing products were analyzed on 15% acrylamide–urea gels (20:1 acrylamide: bisacrylamide ratio).

**Acknowledgments**

We are grateful to our colleagues in both laboratories, especially H. V. Culot, I. Edery, G. Faye, A. Jacquier, P. Legrain, and C. W. Pikielny for careful reading of the manuscript and fruitful discussion. Thanks are due to Jackie Wyatt and Joan A. Steitz for helpful suggestions. B. S. wishes to thank G. Faye for laboratory space and support. T. Tishman provided excellent secretarial assistance. This work was supported by grant number GM23549 to M. R.

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 16 USC Section 1734 solely to indicate this fact.

**Received May 14, 1990; revised August 14, 1990.**

**References**


Aebi, M., Hornig, H., and Weissmann, C. (1987). 5' cleavage site in eu-

karyotic pre-mRNA splicing is determined by the overall 5' splice re-

gion, not by the conserved 5' GU. Cell 50, 237–246.


transcriptional regulation and assembly into ribosomes of a Saccha-

romyces cerevisiae ribosomal protein β-galactosidase fusion. Mol.


are greatly diminished by a mutant yeast branch point. Proc. Natl. Acad. Sci. USA 63, 5835-5839.


Kretzner, L., Raymon, B. C., and Rosbash, M. (1987). S. cerevisiae U1 RNA is large and has limited primary sequence homology to metazoan U1 snRNA. Cell 60, 363-362.


Weber, S., and Aebl, M. (1988). In vitro splicing of mRNA precursors: 5' cleavage site can be predicted from the interaction between the 5' splice region and the 5' terminus of U1 snRNA. Nucl. Acids Res. 16, 471-486.
