The role of exon sequences in splice site selection

Akiya Watakabe, Kenji Tanaka, and Yoshiro Shimura¹

Department of Biophysics, Faculty of Science, Kyoto University, Kyoto 606, Japan

Using mouse immunoglobulin μ (IgM) pre-mRNA as the model substrate for in vitro splicing, we have explored the role of exon sequences in splicing. We have found that deletion of the 5' portion of exon M2 of the IgM gene abolishes the splicing of its immediately upstream intron. Splicing was restored when a purine-rich sequence found within the deleted region was reinserted into the deletion construct. This M2 exon sequence was able to stimulate the splicing of a heterologous intron of the *Drosophila doublesex* pre-mRNA that contains a suboptimal 3' splice site sequence. These results show that the IgM M2 exon sequence functions as a splicing enhancer. We found that the assembly of the early splicing complex is stimulated by the M2 exon sequence. In vitro competition experiments show that this stimulatory effect is mediated by the interaction of some *trans*-acting factors. Our results suggest that the U1 snRNP is one such factor. We propose that recognition of an enhancer exon sequence by the components of splicing machinery plays a vital role in the selection of splice sites, not only for the IgM pre-mRNA but for other pre-mRNAs. We designate such a sequence as exon recognition sequence (ERS).

[Key Words: Splice site selection; splicing; exon recognition sequence; spliceosome assembly; U1 snRNP]

Received September 28, 1992; revised version accepted December 28, 1992.

Splicing of eukaryotic pre-mRNAs involves the accurate selection of the correct 5' and 3' splice sites. Previous studies have shown that conserved sequences around the 5' and 3' splice sites, including the site of lariat formation (branchpoint), serve as the major signal sequences in splice site determination (for review, see Krainer and Maniatis 1988; Green 1991). These sequence elements are recognized by splicing factors, which in turn trigger the formation of a multicomponent complex called the spliceosome (Brody and Abelson 1985; Frendewey and Keller 1985; Grabowski et al. 1985). Small nuclear ribonucleoprotein particles (snRNPs) U1, U2, and U4–U6, constitute the framework of the spliceosome. They bind to pre-mRNA in a stepwise manner: U1 and U2 snRNPs bind to the 5' splice site and the branchpoint sequence, respectively, to form an ATP-dependent complex (complex A or pre-spliceosome). Subsequently, U4/U5/U6 snRNPs enter this complex and complete spliceosome (or complex B) formation. Determination of splice sites occurs early during spliceosome formation (Michaud and Reed 1991) and is followed by intron removal and exon ligation.

The consensus for the 5' and 3' splice site sequences in higher eukaryotes has been determined by comparison of known intron sequences (Shapiro and Senapathy 1987). The 5' consensus sequence is AG/GU(A/G)AGU, whereas the 3' consensus sequence contains a polypyrimidine stretch followed by CAG/G at the 3' splice site (YnNC-

AG/G). The branchpoint sequence is also regarded as a part of the 3' consensus, although it is highly degenerate (Krainer and Maniatis 1988; Green 1991). With the exception of the GU and AG at the 5' and 3' splice sites, respectively, splice site sequences contain several mismatch deviations from the consensus. Owing to this low level of conservation, sequences similar to the consensus are often present at various sites within exons and introns. Generally, sequences that show a better match to the consensus are more tightly bound by splicing factors (Nelson and Green 1990; Zamore et al. 1992) and are more frequently used as authentic splice sites (Oshima and Gotoh 1987; Brunak and Engelbrecht 1991). As such, these sites are considered to be "strong", whereas the "weak" sites, or the sites with poor match to the consensus, tend to be inactive or inefficiently used (Fu et al. 1988; Lowery and Van Ness 1988; Peterson and Perry 1989; Hoshijima et al. 1991). Splice site strength is thus an important determinant in splice site selection. However, the consensus sequences are not sufficient to account for the observed high specificity of splice site selection. Seemingly strong sites are not always selected as splice sites, whereas some authentic sites seem to be weak (Brunak and Engelbrecht 1991). Moreover, a synthetic splice site inserted into various regions of a pre-mRNA exhibited variable activity in a manner dependent on its relative location (Nelson and Green 1988). These observations indicate that other sequence elements are also involved in the selection of splice sites. It was shown that the length of an exon (Furdon and Kole

¹Corresponding author.

1986, 1988; Robberson et al. 1990; Black 1991; Dominiski and Kole 1991) or the secondary structure of the region around splice sites (Solnick 1985; Solnick and Lee 1987; Eperon et al. 1988; Watakabe et al. 1989) affects splice site selection.

There has been cumulative evidence suggesting a vital role for exon sequences in splice site selection (Somashekhar and Mertz 1985; Reed and Maniatis 1986; Mardon et al. 1987; Ricketts et al. 1987; Helfman et al. 1988; Cooper and Ordahl 1989; Freyer et al. 1989; Hampson et al. 1989; Kakizuka et al. 1990; Kats and Skalka 1990; Libri et al. 1990; Ligtenberg et al. 1990; Nagoshi and Baker 1990; Fu et al. 1991; Hoshijima et al. 1991; Watakabe et al. 1991; Cooper 1992; Cote et al. 1992; Steingrimsdottir et al. 1992; Wakamatsu et al. 1992). In these studies, mutations of the specific exon sequences abolish the normal splicing pattern. The molecular basis for the effects of these exon mutations remains obscure, although changes in RNA secondary structure are proposed as the cause for such changes in splicing pattern.

We have shown previously through transfection analyses that the sequence within the last exon, M2, of the mouse immunoglobulin μ (IgM) gene affects upstream splicing profoundly (Watakabe et al. 1991). To investigate the role of exon sequences in splice site selection, we have used the IgM pre-mRNA as the model substrate for in vitro analysis. We show here that splicing between exons M1 and M2 of mouse IgM pre-mRNA requires a purine-rich sequence located within the 5' portion of exon M2. This sequence was able to stimulate the assembly of the early splicing complex at the upstream intron. We found that this stimulatory effect is mediated by the interaction of the *trans*-acting factors. Here, we discuss the role of exon sequence recognition in splice site selection.

Results

Splicing between M1 and M2 exons of mouse IgM gene requires a purine-rich sequence located within exon M2

In a previous study, we showed that deletion of the 5' portion of the mouse IgM gene exon M2 affects the splicing of the upstream intron (Watakabe et al. 1991). To elucidate the molecular basis for the effect of the exon deletion, we employed an in vitro splicing system using HeLa cell nuclear extracts and IgM pre-mRNAs containing the region spanning from exon M1 to M2 (Fig. 1A). When we used the pre-mRNA μ M1-2/X, which contains 164 nucleotides of the 5' portion of exon M2, splicing occurred efficiently between M1 and M2 exons, as judged by the accumulation of the final spliced product (Fig. 1B, lanes 1–3). When we deleted the 54 nucleotides that span from residues +38 to +92 with respect to the 3' splice site of exon M2 (μ M40/X), splicing was almost unaffected (Fig. 1B, lanes 4–6). However, when further deletion removed the region from nucleotides +3 to +92 (μ M Δ), splicing was completely abolished (Fig. 1B, lanes 7–9). Splicing did not occur with this substrate, even during 4 hr of incubation (data not shown). Because

Figure 1. In vitro splicing of IgM pre-mRNAs. (A) Schematic representation of the IgM pre-mRNAs that contain a region spanning from exon M1 to M2. The boxes represent exon sequences, and the lines between them show intron sequences. The 5' exon contains a short leader sequence derived from pSP65. The lengths (in nucleotides) of the 5' exon and the intron are indicated below the μ M1-2/X pre-mRNA. The region bounded by the broken line represents the deleted sequence (+37 to +92 with respect to the 3' splice site for μ M40 and +3 to +92 for μ M Δ). The 5' portion of exon M2 that is required for the splicing between exons M1 and M2 is indicated by the shaded box. The lengths of the 3' exon are shown to the right of each pre-mRNA. These pre-mRNAs were in vitro-transcribed by SP6 polymerase using the template plasmids linearized either by *Xba*I (for μ M1-2/X, μ M40/X, and μ M Δ), *Spe*I (for μ M1-2/S) or *Sal*I (for μ M40/S). (S) *Spe*I; (X) *Xba*I. (B) In vitro splicing of the IgM pre-mRNAs in a HeLa cell nuclear extract. The pre-mRNAs (10 fmole each) were incubated in a HeLa nuclear extract at 30°C for the times (in min) indicated at the top of each lane. Electrophoresis was carried out using a 5% polyacrylamide gel containing 8 M urea. The bands for the RNA products are shown schematically at the right. The band for the final spliced product of each pre-mRNA is indicated by arrowheads. (Lane M) *Hpa*II digests of pBR322 as size marker; (lanes 1–3) μ M1-2/X; (lanes 4–6) μ M40/X; (lanes 7–9) μ M Δ ; (lanes 10–12) μ M1-2/S; (lanes 13–15) μ M40/S.

we did not change the sequence any further than the first 2 nucleotides of exon M2 in this deletion mutant, the abolition of splicing is not the result of the alteration of the splice site consensus sequences. Neither is the effect caused by shortening of the 3' exon, because pre-mRNAs of similar (μ M1-2/S) or even shorter (μ M40/S) 3' exon length were spliced efficiently (Fig. 1B, lanes 10–15). These results indicate that splicing between M1 and M2

exons (M1–M2 splicing) requires some specific sequence present within the 5' portion of exon M2.

To confirm the requirement of the exon sequences in the M1–M2 splicing and to investigate what sequences are required, we divided the 5' portion of exon M2 into three segments (Fig. 2A; Sa, Sb, and Sc) and inserted three copies of each segment back into the deletion construct ($\mu\text{MSn}\times 3$). As we have shown previously, deletion of the 5' portion of exon M2 abolished splicing (Fig. 2B, lanes 4–6). Splicing still did not occur when three Sa sequences were reinserted (Fig. 2B, lanes 7–9); however, splicing was efficiently restored when Sb sequences were reinserted (Fig. 2B, lanes 10–12). Splicing was not re-

stored when Sb sequences were reinserted in the reverse orientation (Fig. 2B, lanes 16–18). Splicing was also restored, although at a lower efficiency, when Sc sequences were inserted (Fig. 2B, lanes 13–15). These results clearly demonstrate that specific sequences are required within the downstream exon for M1–M2 splicing. Comparison of Sa, Sb, and Sc sequences suggests that the purine-rich sequence present in both Sb and Sc but not in Sa is required for M1–M2 splicing (see Discussion). In addition, we found that three copies of Sb sequences had a greater effect on upstream splicing than just one copy (data not shown). Thus, it is most likely that a sequence encompassing both Sb and Sc comprises the sequence essential for M1–M2 splicing.

Figure 2. Restoration of splicing potential of the deletion constructs by subsequences of M2 exon. (A) The nucleotide sequence of the 5' portion (from -2 to +42) of exon M2 is shown below the schematic representation of $\mu\text{M1-2/X}$ pre-mRNA. This region was arbitrarily divided into three segments: Sa, Sb, and Sc. Three copies of each segment (represented by arrows) were ligated in parallel with a connecting linker sequence and inserted into the deletion region (shown by broken lines) of $\mu\text{M}\Delta$. In $\text{SnR}\times 3$ pre-mRNA, the insert is in the reverse orientation. The boxes and the lines are as described in Fig. 1. The lowercase letters represent intron sequences; the uppercase letters represent exon sequences. The 3' splice junction is indicated by the slash. (B) In vitro splicing of the $\text{Sn}\times 3$ pre-mRNAs. After standard reaction for the times (in min) indicated at the top of each lane, electrophoresis was carried out on a 5% polyacrylamide gel containing 8 M urea. The bands for the RNA products are shown schematically at right. (Lanes 1–3) $\mu\text{M1-2/X}$; (lanes 4–6) $\mu\text{M}\Delta$; (lanes 7–9) $\mu\text{MSa}\times 3$; (lanes 10–12) $\mu\text{MSb}\times 3$; (lanes 13–15) $\mu\text{MSc}\times 3$; (lanes 16–18) $\mu\text{MSbR}\times 3$.

Sb sequence of M2 exon can stimulate the splicing of a heterologous intron of doublesex pre-mRNA

The results described above suggest that the function of the M2 exon sequence is to stimulate splicing of the upstream intron. To test this possibility directly, we constructed a chimeric pre-mRNA in which the IgM M2 exon sequence was connected downstream of the female-specific intron (the intron between exons 3 and 4) of the *Drosophila doublesex* (*dsx*) gene. It was shown previously that splicing of this intron does not usually occur, because its 3' splice site sequences contain a suboptimal polypyrimidine stretch (Hoshijima et al. 1991; Tian and Maniatis 1992). In agreement with these studies, *dsx-S0* pre-mRNA, which contains portions of exons 3 and 4 and the female-specific intron between them, was not spliced in a HeLa nuclear extract (Fig. 3B, lanes 1–3). In contrast, when we inserted three copies of Sb sequences into the 3' exon of *dsx-S0* pre-mRNA (*dsx-Sb*), splicing of the *dsx* intron was strongly stimulated (Fig. 3B, lanes 7–9). Splicing was barely detectable when Sa sequences were inserted (Fig. 3B, lanes 4–6). These results demonstrate that the Sb sequence of exon M2 can stimulate upstream splicing.

The M2 exon sequence stimulates the assembly of the early splicing complex

To elucidate the molecular basis for the stimulatory effect of the M2 exon sequence, we examined its role in spliceosome assembly. The assembly of splicing-specific complexes on pre-mRNA has been investigated previously using a native gel electrophoresis system (Konarska and Sharp 1986, 1987). These studies revealed a stepwise assembly of splicing complexes, which is characterized by the binding of snRNPs U1, U2, and U4–U6, in accordance with various other factors: U1 snRNP first binds to the 5' splice site in an ATP-independent manner; U2 snRNP then binds to the branchpoint sequence and forms the first ATP-dependent complex, often referred to as the pre-spliceosome (or complex A); U4/U5/U6 snRNPs subsequently enter the complex and form the spliceosome (or complex B). We conducted native gel experiments using two IgM pre-mRNAs that differ solely by the presence ($\mu\text{M1-2/X}$) or absence ($\mu\text{M}\Delta$) of the 5'

Figure 3. In vitro splicing of *dsx*-IgM chimeric pre-mRNA. (A) Schematic representation of the *Drosophila dsx*-IgM chimeric pre-mRNAs. The boxes and lines are as described in Fig. 1, except that the linker sequence in the 3' exon that is derived from the pSP72 vector is shown by the narrow box. The 5' exon contains a short leader sequence derived from pSP72. The lengths (in nucleotides) of the exons and introns are indicated above the respective regions of the construct. The 3' exon of *dsx*-S0 pre-mRNA contains 30 nucleotides of the *dsx* fourth exon and 20 nucleotides of linker sequence derived from pSP72. In *dsx*-Sa and *dsx*-Sb, three copies of Sa and Sb sequences are connected to the 3' ends. (B) In vitro splicing of the *dsx* pre-mRNAs. After the standard reaction for the times indicated at the top of each lane, electrophoresis was carried out on a 5% polyacrylamide gel containing 8 M urea. The bands for the RNA products are shown schematically at right. (Lane M) The *Hpa*II digests of pBR322 as size marker; (lanes 1–3), *dsx*-S0; (lanes 4–6) *dsx*-Sa; (lanes 7–9) *dsx*-Sb.

portion of exon M2. These IgM pre-mRNAs were incubated with a HeLa cell nuclear extract, and resultant assembled complexes were resolved by native gel electrophoresis.

When the 5' portion of exon M2 was present in the pre-mRNA (μ M1-2/X), two ATP-dependent complexes with different mobilities were detected (Fig. 4, lanes 2–6, bands A and B). Considering the formation time course, ATP dependency, and mobility of the complexes, the faster migrating complex probably corresponds to a U2 snRNP-containing complex A and the slower migrating one to a U4/U5/U6 snRNP-containing complex B (Konarska and Sharp 1986, 1987). On the other hand, when the 5' portion of exon M2 was deleted (μ M Δ), complex A

was observed, but to a lesser extent, and complex B was not detectable (Fig. 4, lanes 7–12). These results suggest that the M2 exon sequence stimulates the assembly of the initial ATP-dependent complex (complex A) and subsequent formation of complex B.

To confirm the results of the native gel electrophoresis experiments, we investigated the interactions between snRNPs and pre-mRNAs during spliceosome assembly by a UV cross-linking assay (Sawa and Shimura 1992). In this assay, the reaction mixtures, incubated under the conditions for in vitro splicing, are irradiated by UV light and deproteinized extensively. After the deproteinization, only the snRNAs remain cross-linked to the radio-labeled pre-mRNAs. This cross-linked product can be detected as bands shifted above the pre-mRNAs upon gel electrophoresis. When the pre-mRNA containing the 5' portion of exon M2 (μ M1-2/S) was analyzed by this UV cross-linking experiment, three shifted bands were mainly detected. Among these, the band that migrates just above the pre-mRNA was formed even in the absence of nuclear extracts (data not shown). The other two bands, α and β , were generated upon incubation with a HeLa cell nuclear extract. The slower migrating band, α , appeared within 1 min (Fig. 5A, lane 2) and disappeared with further incubation (Fig. 5A, lanes 3,4). Subsequently, the faster migrating band, β , appeared within 5

Figure 4. Splicing complex formation as analyzed by native gel electrophoresis. The pre-mRNAs that are shown schematically beneath the panel were incubated under splicing conditions for the times indicated at the top of each lane, either in the presence (lanes 2–6, 8–12) or absence (lanes 1,7) of ATP. The reaction mixtures were then treated with 10 mg/ml of heparin and loaded directly onto a 4% native gel in Tris-glycine buffer. The complexes formed on the pre-mRNAs are indicated as H, A, and B (right), according to previous reports (Konarska and Sharp 1986, 1987).

complementary to a portion of U1 (lanes 2, 8) or U6 snRNA (lane 6), respectively. The bands corresponding to the cleavage products are indicated at *right*. The oligonucleotides used are complementary to positions 64–75 of U1 snRNA, 28–42 of U2 snRNA, 1–15 of U4 snRNA, 33–51 of U5 snRNA, and 78–95 of U6 snRNA.

min (Fig. 5A, lane 3). Only band α was observed in the absence of ATP (Fig. 5A, lane 5). To identify which of the small nuclear RNAs (snRNAs) are cross-linked to the pre-mRNA, oligonucleotide-directed RNase H cleavage was performed (Sawa and Shimura 1992). After UV irradiation and deproteinization, the recovered RNAs were annealed to oligonucleotides complementary to snRNAs U1, U2, and U4–U6, respectively, and digested with RNase H. As judged by the disappearance of that band and the appearance of the faster migrating bands, the cross-linked product corresponding to band α was cleaved with an oligonucleotide complementary to U1 snRNA (Fig. 5B, lane 2). Band α was not cleaved with any oligonucleotides complementary to other snRNAs (Fig. 5B, lanes 3–6). These results indicate that band α corresponds to the product cross-linked to U1 snRNA. By the same criteria, band β was shown to be the product of pre-mRNA cross-linked to U6 snRNA (Fig. 5B, lanes 1–6). Thus, U1 snRNP binds to the μ M1–2/S pre-mRNA within 1 min (Fig. 5A, lane 2), and its interaction with the pre-mRNA weakens as time passes (Fig. 5A, lanes 3,4). After 5 min, U6 snRNP interacts with this pre-mRNA (Fig. 5A, lane 3). These results are consistent with the observation that the slower migrating complex, which appeared after a 5-min incubation (Fig. 4B, lane 4), corresponds to complex B and contains U6 snRNP.

We then conducted UV cross-linking experiments using the pre-mRNA in which the 5' portion of exon M2 is deleted (μ M Δ). We observed only one band that appeared in an ATP-independent manner (Fig. 5A, lanes 7–10). This band was first detected within 1 min of incubation in the extract (Fig. 5A, lane 7) and did not disappear with further incubation. By oligonucleotide-directed RNase H cleavage, this band was shown to represent a cross-linked product between U1 snRNA and μ M Δ (Fig. 5B, lanes 7–12). Thus, the band detected with μ M Δ seems to

Figure 5. UV cross-linking experiments with IgM pre-mRNA. (A) 32 P-labeled μ M1–2/S and μ M Δ pre-mRNAs were incubated with a HeLa cell nuclear extract in a 5- μ l reaction mixture under splicing conditions for the times indicated, either in the presence (lanes 1–4, 6–9) or absence (lanes 5, 10) of ATP. Each reaction mixture was then irradiated with UV light, and the RNAs were recovered as described in Materials and methods. Two cross-linked products, α and β , are indicated at *right*. (P) Pre-mRNAs. (B) Incubation was carried out for 5 min, and UV cross-linking was performed. The recovered RNAs were treated with RNase H and oligonucleotides complementary to a specific snRNA (indicated at the top of each lane). The RNA products cross-linked to U1 snRNA (P–U1) and to U6 snRNA (P–U6) were cleaved by RNase H, when annealed to oligonucleotides

correspond to band α detected with μ M1–2/S. These results show that U1 snRNP binds to the pre-mRNA even when the M2 exon sequence is deleted. The M2 exon sequence should therefore be required for the subsequent changes that occur during spliceosome assembly, such as the weakening of the U1 snRNP interaction and the binding of U6 snRNP. This supports the notion that M2 exon sequences stimulate spliceosome assembly.

Stimulation of splicing is mediated by the specific interaction of trans-acting factors with the M2 exon sequence

To determine whether the stimulation of spliceosome assembly is mediated by some *trans*-acting factor that specifically interacts with the M2 exon sequence, we carried out in vitro competition experiments using two kinds of competitor RNAs containing either the 5' (5'P) or other portion (Cont) of exon M2 (Fig. 6). HeLa cell nuclear extracts were preincubated with these competitor RNAs on ice for 10 min and incubated at 30°C for an additional 20 min after the addition of the μ M1–2/X pre-mRNA into the reaction mixture. When we used the RNA containing the 5' portion of exon M2 (5'P), splicing was titrated by increasing the levels of this RNA (Fig. 6, lanes 1–4). On the other hand, similar levels of a control RNA that contained the same length of the other portion of exon M2 (Cont) did not affect splicing (Fig. 6, lanes 5–8). These results strongly suggest that the stimulatory effect of the M2 exon sequence is mediated by some *trans*-acting factor that specifically interacts with the sequence.

Interaction of U1 snRNP with the M2 exon sequence

To test whether one of the snRNPs is involved in the

Figure 6. In vitro competition experiments with the exon M2 sequence. A schematic representation of the IgM pre-mRNA (μ M1-2/X) used for in vitro competition experiments is shown beneath the panel. The regions contained in the competitor RNAs are indicated by the thick bar below the pre-mRNA. The 5'P RNA contains 40 nucleotides of the 5' portion (from +1 to +40 with respect to the 3' splice site) of exon M2; the Cont RNA contains another portion (+118 to +158). A HeLa cell nuclear extract was preincubated on ice with these competitor RNAs for 10 min. After adding the pre-mRNA into the reaction mixture, it was incubated at 30°C for 20 min. The amount of competitors added in each case is indicated at the top of each lane. The bands for the RNA products are shown schematically at right. (Lanes 1-4) 5'P RNA was used as the competitor; (lanes 5-8) Cont RNA was used as the competitor.

recognition of the M2 exon sequences and in the stimulation of spliceosome assembly, we carried out UV cross-linking experiments using the RNA probes containing either the 5' (5'P) or other portion (Cont) of exon M2, used for the competition experiments. When we used 5'P RNA, several cross-linked RNA products were observed as shifted bands (Fig. 7A, lane 4). To identify these bands, oligonucleotide-directed RNase H cleavage was carried out (Fig. 7A, lanes 5-10). The two closely migrating bands denoted as 5'P-U1 were selectively cleaved with an oligonucleotide complementary to U1 snRNA (Fig. 7A, lane 6). These bands were not detected when we carried out the UV cross-linking experiments with control RNA (Fig. 7A, lanes 1,2), demonstrating that the cross-linking is sequence specific. Another strong band that migrates slower than 5'P-U1 (denoted as X) was not digested with any oligonucleotide complementary to the five snRNAs (Fig. 7A, lanes 5-10). We presume that this

band represents an intramolecular cross-link, because, even in the absence of a HeLa nuclear extract, a faint band was detected (Fig. 7A, lane 3).

To further investigate the specificity of U1 cross-linking, we carried out similar experiments using RNA probes containing three copies of Sa, Sb, or Sc sequences (Sa $\times$ 3, Sb $\times$ 3 and Sc $\times$ 3 probes, respectively). With the Sb $\times$ 3 probe, we detected a band that shifts above the probe (Fig. 7B, lane 7). This band was specifically cleaved by RNase H, when annealed to two oligonucleotides that are complementary to the different regions of U1 snRNA (Fig. 7B, lanes 7-10). The band migrated faster when the RNase H digestion was carried out with the U1:3' oligonucleotide, which is complementary to a region on the 3' side (from +121 to +136, with respect to the 5' end) of U1 snRNA (Fig. 7B, lane 9). The band disappeared, when the U1:5' oligonucleotide, which is complementary to a region near the 5' end (from +12 to +27), was used (Fig. 7B, lane 8). These results indicate that the shifted band represents a cross-linked product between U1 snRNA and the Sb $\times$ 3 probe. Moreover, the disappearance of the band shows that most of the U1 snRNA that is cross-linked to the Sb $\times$ 3 probe was removed by RNase H digestion. This should occur only when the cross-linked product was cleaved near the cross-linking site. Thus, the cross-linking site is thought to reside near the 5' end of U1 snRNA.

We also carried out the UV cross-linking experiment using the Sa $\times$ 3 probe. As in Figures 2 and 3, Sa sequences failed to activate the upstream splicing. Except the band that appeared in the absence of a HeLa nuclear extract (Fig. 7B, lane 1), we could not observe any discrete band with mobility that corresponds to the U1 cross-linked product of the Sb $\times$ 3 probe (Fig. 7B, lane 2). Thus, the cross-linking of U1 snRNA to Sa and Sb probes is in good correlation with the ability of these sequences to activate the upstream splicing. We tested another probe, Sc $\times$ 3, containing the sequence that weakly activated upstream splicing (see Fig. 2B). We detected a shifted band whose mobility is similar to the cross-linked product of U1 snRNA and the Sb $\times$ 3 probe (Fig. 7B, lane 12). Although this band was faint and not necessarily clear in the photograph, it was specifically cleaved by RNase H, when annealed to the U1:5' and U1:3' oligonucleotides (Fig. 7B, lanes 12-15). Thus, U1 snRNA is also cross-linked to the Sc $\times$ 3 probe. These results strongly suggest that although other factors may also be involved in exon recognition, U1 snRNP is at least one of the factors that recognize the M2 exon sequence.

To assess the significance of U1 snRNP interaction with the M2 exon sequence, this sequence was replaced by a 5' splice site consensus sequence. Splicing between exons M1 and M2 was abolished by deleting the Sb and Sc sequences (μ MA Δ 20) (Fig. 8B, lanes 4-6). When we inserted the consensus 5' splice site sequence into the 3' end of this pre-mRNA (μ MA + U1), splicing occurred efficiently with this substrate (Fig. 8B, lanes 7-9). In contrast, when the inserted 5' splice site was mutated (μ MA + U1M), splicing was abolished (Fig. 8B, lanes 10-12). These results suggest that the binding of U1 snRNP

Figure 7. UV cross-linking experiments with the stimulatory sequence within exon M2. 32 P-labeled RNAs were incubated under splicing conditions for 5 (A) or 10 min (B). UV cross-linking and RNase H cleavage were carried out as described in Materials and methods and in the legend to Fig. 3. (A) The 5'P and Cont RNAs were used as probes. (Lanes 1–4) UV cross-linking experiments with Cont and 5'P RNA, either in the absence (lane 1, 3), or presence (lane 2, 4) of a HeLa cell nuclear extract. (Lanes 5–10) Identification of the cross-linked products to 5'P by RNase H and oligonucleotides complementary to specific snRNAs (see the legend to Fig. 3). RNase H digestion was carried out without the oligonucleotides in lane 5. Two closely migrating bands that are selectively cleaved with an oligonucleotide

complementary to U1 snRNA are denoted as 5'P-U1. Their cleavage products are also indicated at right. Band X represents the product that was not cleaved with any oligonucleotides. (B) UV cross-linking and RNase H identification experiments with Sa×3, Sb×3, and Sc×3 RNAs. UV cross-linking was carried out in the absence of a HeLa nuclear extract in lanes 1, 6, and 11. RNase H digestion was carried out without the oligonucleotides in lanes 2, 7, and 12. The oligonucleotides used for the RNase H cleavage are as follows. (U1:5') Complementary to positions 12–27 of U1 snRNA; (U1:3') complementary to positions 121–136 of U1 snRNA; (U2) complementary to positions 28–42 of U2 snRNA. The bands specifically cleaved with the oligonucleotides complementary to U1 snRNA are denoted as U1 cross-link, and the cleavage products are indicated.

to the downstream exon stimulates the splicing of the upstream intron.

Discussion

The role of exon sequences in splice site selection

We have shown here that the splicing between exons M1 and M2 of the mouse IgM gene requires a specific sequence within the 5' portion of exon M2. We found that this exon sequence can stimulate the splicing of the upstream intron. Thus, this sequence is regarded as a splicing enhancer. There are two possible mechanisms by which this M2 exon sequence exerts this stimulatory effect: One is that the M2 exon sequence forms a secondary structure that improves the accessibility of splicing factors to the splice sites; the other is that the M2 exon sequence serves as the target of some *trans*-acting factor that stimulates the splicing. The competition experiments showed that the stimulatory effect of the M2 exon sequence is titrated out by a competitor RNA that contains the 5' portion of exon M2. This result strongly favors the latter mechanism. Moreover, the result that the M2 exon sequence stimulated splicing of a heterologous intron suggests that the formation of a specific secondary structure should not be so important for stimulation. We conclude from these results that the stimulatory effect of the M2 exon sequence is mediated by the interaction of some *trans*-acting factor.

Two lines of evidence suggest that the putative factor that recognizes the M2 exon sequence is not specific to the IgM pre-mRNA. First, we obtained the above results

by using the nuclear extracts of HeLa cells that do not express IgM. It is not likely that such cells express a factor that is required only for IgM splicing. Second, two different sequences, Sb and Sc, are thought to be recognized by the putative factor, because they both restore M1–M2 splicing when reinserted into the deletion constructs. Comparison of these sequences revealed that they are both rich in purine residues and contain consecutive polypurine stretches: a 7-nucleotide stretch for Sb and a 5-nucleotide stretch for Sc (Fig. 2). However, we could not identify any common sequences shared by Sb and Sc. These observations suggest that the sequence recognized by the putative factor may be a weakly defined purine-rich sequence that could be found in the exons of other genes (see below). Consistent with this notion, the Sa sequence, which does not contain consecutive polypurine stretch, failed to stimulate upstream splicing. It is therefore most likely that this putative factor is not a regulatory factor that recognizes a specific sequence of the IgM gene but is one of the general factors involved in the splicing of other genes.

Previous experiments using human β -globin pre-mRNA showed that most of the 3' exon sequence is not required for splicing (Parent et al. 1987; Furdon and Kole 1988). This information and our results suggest that there are two classes of pre-mRNAs: those that require specific downstream sequences for splicing and those that do not. What is the difference between these two classes of pre-mRNA? The experiments using *dsx* pre-mRNA provide an important clue to this question. It was shown previously that the female-specific 3' splice site of *dsx* pre-mRNA is not normally used, owing to a defect in its polypyrimidine stretch (Hoshijima et al. 1991).

Figure 8. In vitro splicing of a pre-mRNA containing the 5' consensus splice site sequence in the 3' exon. (A) A schematic representation of the pre-mRNAs is as described in Fig. 1. Sa, Sb, and Sc sequences (see Fig. 4) are shown. Sb and Sc sequences are shaded. In μMΔ20, the 3' exon is truncated so that Sb and Sc sequences are deleted. In μMΔ+U1, a short sequence containing the consensus 5' splice site (underlined) is connected to the 3' end of μMΔ20. In μMΔ+U1M, the consensus 5' splice site of μMΔ+U1 is mutated (indicated by × under the sequence). (B) After standard in vitro splicing reactions for the times (in min) indicated at the top of each lane, electrophoresis was carried out on a 5% polyacrylamide gel containing 8 M urea. Representations of the products defined by each band are shown schematically at right. The band corresponding to the final spliced product of each pre-mRNA is indicated by arrowheads. Lane M) HpaII digests of pBR322 as size marker; (lanes 1–3) μM1-2/X; (lanes 4–6) μMΔ20; (lanes 7–9) μMΔ+U1; (lanes 10–12) μMΔ+U1M.

However, even such a defective site was used efficiently for splicing when a specific sequence (Sb sequence of IgM M2 exon) was present in the downstream region. This result suggests that the pre-mRNA with suboptimal consensus sequences requires the stimulation by the downstream sequence. As we have shown, the downstream 5' splice site can also stimulate upstream splicing. Taken together, we propose the model illustrated in Figure 9 to account for these findings.

When the 3' splice site sequences (including branch-point, polypyrimidine tract, and AG/ at the 3' splice site) are strong, or close to the consensus sequence, splicing occurs independently of the downstream region (Fig. 9A).

However, when the 3' splice site sequences are weak, or poorly match the consensus sequence, they are not efficiently recognized by the splicing factors and thus cannot be used efficiently as the splice site. The use of such a weak site requires stimulation by the downstream sequence (Fig. 9B,C). For example, a specific exon sequence can stimulate the use of upstream 3' splice sites as in the case of IgM splicing (Fig. 9B). We designate a sequence that serves such a role as an ERS (exon recognition sequence). Alternatively, a strong 5' splice site sequence can also stimulate the use of the upstream 3' splice site across an exon (Fig. 9C). We propose that the three sequence elements—the 3' splice site, ERS, and the downstream 5' splice site—are recognized as a whole and that the sum of the strength of these elements determines the 3' splice site selection.

This model provides a good explanation for why alterations of the polypyrimidine stretch did not abolish splicing with some pre-mRNAs (Fu et al. 1988; Freyer et al. 1989). It is conceivable that mutations in the polypyrimidine stretch would not affect splicing so much if the downstream exon contained an ERS or a strong 5' splice site. Indeed, the alterations of the polypyrimidine stretch of the adenovirus major late transcript had different effects on splicing depending on the downstream sequence (Freyer et al. 1989). This observation is consistent with our model.

Figure 9. A model for 3' splice site selection. Schematic representation of the *cis*-acting elements that affect the selection of the 3' splice site. (■) Strong signals; (□) weak signals. The arrow represents stimulation of the weak 3' splice site by the interaction of U1 snRNP (and possibly other factors) with ERS (B) or with the downstream 5' splice site (C). (D) ● Indicate regulatory factors that bind to a specific element within an exon (shown by shading) and stimulate upstream splicing (for details, see Discussion). (BPS) Branchpoint sequence; (Py) polypyrimidine tract; (5' SS) 5' splice site; (ERS) exon recognition sequence.

Our model also provides a hint about the regulation of a certain type of alternative splicing. It is possible that specific regulatory factors may stimulate exon incorporation by mimicking the function of the splicing factor that recognizes the ERS (Fig. 9D). It was shown previously that female-specific splicing of the *dsx* gene is mediated by two regulatory factors, *transformer* and *transformer-2*, products that bind to a regulatory element in the female-specific exon (Nagoshi and Baker 1990; Hedley and Maniatis 1991; Hoshijima et al. 1991; Ryner and Baker 1991; Inoue et al. 1992; Tian and Maniatis 1992). We speculate that these factors might induce female-specific splicing by a similar mechanism as in the case of IgM splicing.

The presence of ERS may account for the effect of some exon mutations

Many groups have reported the effects of exon mutations on splicing (Mardon et al. 1987; Ricketts et al. 1987; Helfman et al. 1988; Cooper and Ordahl 1989; Hampson et al. 1989; Kats and Skalka 1990; Fu et al. 1991; Cooper 1992; Cote et al. 1992; Steingrimsdottir et al. 1992; Wakamatsu et al. 1992). However, with the exception of *dsx* splicing, the molecular basis for such effects has remained obscure (Nagoshi and Baker 1990; Hedley and Maniatis 1991; Hoshijima et al. 1991; Ryner and Baker 1991; Inoue et al. 1992; Tian and Maniatis 1992). As we know now that the purine-rich sequence of IgM exon M2 stimulates splicing, we examined whether similar sequences could be found in the mutated exon sequences of other genes. Interestingly, we found that several exon mutations disrupt polypurine stretches similar to the one in the IgM exon M2 (Table 1). In these cases, it is likely that exon mutations affect splicing by disrupting ERS. Recently, we found that some such sequences also have a strong stimulatory effect on upstream splicing (data not shown). This is consistent with our notion that

the weakly defined polypurine stretches serve as ERS. In this connection, it is worth noting that a pyrimidine-rich sequence placed in the downstream exon acts as a poison sequence (Furdon and Kole 1988). There are several other cases that can be explained by a different mechanism. In such cases, exon mutations may disrupt a specific element that is recognized by the regulatory factors of alternative splicing (Streuli and Saito 1989; Tsai et al. 1989; Nagoshi and Baker 1990; Hoshijima et al. 1991) or change the secondary structure of pre-mRNAs that either sequesters or exposes the splice sites (Libri et al. 1990; Ligtenberg et al. 1990; Steingrimsdottir et al. 1992).

The mechanism of splicing activation

We have yet to clarify the precise mechanism by which ERS stimulates the splicing of the upstream intron. The important question is what factor recognizes ERS and how it stimulates splicing. Regarding this point, we have shown in the UV cross-linking experiments that U1 snRNA is specifically cross-linked to the M2 exon sequence. Moreover, binding of U1 snRNP to the 5' splice site present at the downstream exon stimulated the splicing of the preceding intron. These results strongly suggest that U1 snRNP is the factor that recognizes ERS and stimulates the splicing of the upstream intron (Fig. 9B), although other factors may also be involved in this process.

We do not know at present how U1 snRNP recognizes ERS. Rough mapping by RNase H cleavage suggest that the cross-linking is near the 5' end of U1 snRNA. Therefore, one possible mechanism is that the weak base-pairing between the 5' end of U1 snRNA and ERS is involved in recognition: The Sb sequence contains a sequence that matches 5 of 8 nucleotides of the 5' splice site consensus, whereas the Sc sequence contains a sequence that matches 3 of 8 nucleotides of this consensus. Alter-

Table 1. Comparison of exon sequences affecting splicing

Exon	Sequence ^a	Reference
1. Mouse IgM exon M2	...GGAAGGACAGCAGAGACCAAGAG...	this study
2. Human FN gene EDIII exon	...TGAGGATGGAAT TGGTGAAGAAGAC...	Mardon et al. (1987)
3. BGH gene exon 5	...TCTCCTGCTTCCGGAAGGACCTGCATAAGA...	Hampson et al. (1989)
4. ASLV <i>env</i> 3' exon	...CGAGCAAGAAGGACTCCAAGAAGAAGCCGCCAG CAACAAGCAAGAAGGACCCGGA...	Katz and Skalka (1990) Fu et al. (1991)
5. Rat β -TM gene exon 8	...TATTCCACCAAGAGGACAAATACGA	Helfman et al. (1988)
6. cTNT gene exon 5	... / AAGAGGAAGAAATGGCTTGAGGAAGACGACG / ..	Cooper and Ordahl (1989) Cooper (1992)
7. Human <i>hprt</i> gene exon 3	...GAGATGTGATGAAGGAGATGGGAGG...TCAAGGGGGCTATAAA...	Steingrimsdottir et al. (1992)

The purine-rich sequences present within the exons of various genes are shown. It was reported previously that deletions (1–5) or substitutions (6 and 7) of these sequences severely affect the splicing of the upstream intron (1 and 3–6) or the inclusion of the exon into mRNAs (2, 6, and 7). The underscoring of the sequences from the cTNT gene and *hprt* gene sequence shows the substituted residues. The slashes in the cTNT gene sequence indicate the splice sites. The abbreviations used for the names of these genes are as follows: (IgM) immunoglobulin μ ; (FN) fibronectin; (BGH) bovine growth hormone; (ASLV) avian sarcoma–leukosis virus; (TM) tropomyosin; (cTNT) cardiac troponin T; (*hprt*) hypoxanthine–guanine phosphoribosyltransferase.

^aConsecutive purine residues (>5 nucleotides) are boldfaced.

natively, there might be additional factors that recognize ERS and facilitate the binding of U1 snRNP. Bennett et al. demonstrated that pre-mRNAs are differentially bound by a unique set of heterogeneous nuclear RNP (hnRNP) proteins (Bennett et al. 1992). It is possible that the preferential binding of one such hnRNP protein may facilitate the interaction of U1 snRNP to the exon sequence. We have detected a protein that is specifically cross-linked to ERS of the IgM M2 exon upon UV irradiation (data not shown). This protein may be involved in the recognition of ERS.

Several studies including the present one demonstrate that the interaction of U1 snRNP with the 5' splice site downstream of the 3' exon of a pre-mRNA can activate the splicing of the upstream intron (Robberson et al. 1990; Talerico and Berget 1990; Kreivi et al. 1991; Kuo et al. 1991). Robberson et al. suggested that the recognition factors bound at the downstream 5' splice site may interact with the 3' splice site recognition factors across an exon and may stabilize the complex formed on the 3' splice site region (Robberson et al. 1990). Similarly, U1 snRNP bound to ERS may interact with the 3' splice site recognition factors, and the stabilization of this 3' splice site complex may facilitate subsequent spliceosome assembly and splicing reaction.

In conclusion, having identified a novel *cis*-acting element involved in splice site selection, we can now direct our efforts toward a greater understanding of how splice sites are selected. We are currently investigating the ERS motif in more detail and determining whether cellular factors besides U1 snRNP are involved in recognition of this motif. Further work will reveal the precise mechanism by which recognition of ERS leads to the stimulation of upstream splicing.

Materials and methods

Oligonucleotides

Oligonucleotides were synthesized with Applied Biosystems DNA Synthesizer A380 and purified by electrophoresis on a 10% denaturing polyacrylamide gel. The oligonucleotides used for plasmid construction are as follows: pcr-1, 5'-CTGCTGGTCGACCTCTCTGCTGTCCTTCCA-3'; pcr-2, 5'-CTGCTGTGCGACCTTGAACAGGGTGAC-3'; pcr-3, 5'-CTGCTCTAGATGCTGAGAGTCATTTTC-3'; sa, 5'-TCGACACTCTCAGCATGC-3'; saR, 5'-TCGAGCATGCTGAGAGTG-3'; sb, 5'-TCGACGGAAGGACAGCAC-3'; sbR, 5'-TCGAGTGCTGTCCTTCCG-3'; sc, 5'-TCGACAGAGACCAAGAGC-3'; scR, 5'-TCGAGCTCTTGGTCTCTG-3'; compA, 5'-GTGAAATGACTCTCAGCATGGAAGGACAGCAGAGACCAAG-3'; compAR, 5'-CTTGGTCTCTGCTGTCTCTCCATGCTGAGAGTCATTTCAC-3'; compB, 5'-CCTGTGTTGCCCTCCAGCTTTATCTCTGAGATGGTCTTC-3'; compBR, 5'-GAAGACCATCTCAGAGATAAAGCTGGAGGGCAACACAGG-3'; U1, 5'-CTAGACAGGTAAGTACA-3'; U1R, 5'-AGCTTGTAACCTGT-3'; U1M, 5'-CTAGACAGCCAACTACA-3'; U1MR, 5'-AGCTTGTAAGTTGGCTGT-3'.

Plasmid construction

All constructions were made using standard cloning procedures (Sambrook et al. 1989), and most were confirmed by sequencing.

Mouse IgM gene fragments were obtained from plasmid pM μ Δ A (kindly provided by Dr. N. Tsurushita) or its derivatives (Tsurushita et al. 1987; Watakabe et al. 1991). *Drosophila* *dsx* gene fragments were obtained from pSPdsxE34f (Inoue et al. 1992). To construct the templates for μ M1-2/X and μ M1-2/S pre-mRNAs (p μ M1-2), the *BsmI*-*XbaI* fragment spanning portions of exons M1 and M2 and the intron between them were inserted into the *SacI*-*XbaI* site of pSP65 vector (Promega). The deletion plasmids p μ M40, p μ M Δ , and p μ M Δ 20 (the templates for μ M40/X, μ M40/S, μ M Δ , and μ M Δ 20 pre-mRNAs) were constructed by subcloning the PCR-amplified fragments of the IgM gene, using SP6 promoter primer and pcr-1, pcr-2, and pcr-3 (see above) primers as the first and the second primers, respectively. To construct the plasmids p μ MSax3 and pT7-Sax3, the sa and saR oligonucleotides were annealed and ligated by T4 ligase. The fragment containing three of these sequences ligated in parallel were then inserted into p μ M Δ cut with *SalI* or pSP 72 vector cut with *SalI* and *XhoI*, generating p μ M Δ Sax3 and pT7-Sax3, respectively. The orientation of the insert was confirmed by sequencing. The plasmids p μ MSbx3, p μ MSbRx3, p μ MScx3, pT7-Sbx3, and pT7-Scx3 were produced in the same way. To construct the template plasmids for *dsx*-Sa and *dsx*-Sb pre-mRNAs, *BglII*-*HincII* fragment of pSPdsxE34f (Inoue et al. 1992) was ligated into the *BglII*-*SmaI* site of pT7-Sax3 and pT7-Sbx3. The template plasmids for Sax3, pT7-Sax3-d, was generated by self-ligating the pT7-Sax3 cut with *BglII* and *SalI*. pT7-Sbx3-d and pT7-Scx3-d were generated in a similar way. The template plasmids for competitor RNAs, 5'P and Cont, were generated by inserting annealed oligonucleotides (compA and compAR oligonucleotides for 5'P and compB and compBR oligonucleotides for Cont) into pSP72 vector cut with *SmaI*. p μ M Δ 20-U1 was constructed by inserting annealed oligonucleotides (U1 and U1R) into p μ M Δ 20 cut with *XbaI* and *HindIII*. p μ M Δ 20-U1M was constructed in a similar way.

Pre-mRNA preparation and in vitro splicing

In vitro transcription was carried out either with SP6 or T7 RNA polymerase. HeLa cell nuclear extracts were prepared as described previously (Dignam et al. 1983). The splicing reaction was carried out in 10 μ l of the previously described reaction mixture (Sakamoto et al. 1987).

Separation of splicing complexes in a native gel

In vitro splicing was carried out in 5 μ l of the reaction mixture. After treatment with heparin (10 mg/ml) for 10 min on ice, the reaction mixture was loaded directly onto a 4% polyacrylamide, 50 mM Tris-glycine (pH 8.8) gel (acrylamide/bisacrylamide weight ratio of 80 : 1) as described previously (Konarska and Sharp 1986, 1987).

UV cross-linking analyses

UV cross-linking experiments were performed essentially as described previously (Sawa and Shimura 1992). After incubation of ³²P-labeled RNAs in a HeLa cell nuclear extract, the reaction mixtures were diluted 20-fold with buffer E [12 mM HEPES-NaOH (pH 7.9), 60 mM KCl, 1.5 mM MgCl₂, 0.12 mM EDTA, 12% glycerol] and irradiated with UV light (wavelength 254 nm) in a Stratilinker (Stratagene) at 250,000 μ J/cm² on a microtiter plate on ice at a distance of 10 cm from UV light. The irradiated samples were deproteinized with proteinase K (Merck) and pronase (Calbiochem), phenol extracted, and ethanol precipitated and analyzed by electrophoresis on a 5% denaturing polyacrylamide gel.

RNase H digestion experiments

RNase H digestion experiments were performed essentially as described previously (Sawa and Shimura 1992). RNA preparations were annealed with 10 μ g/ml of oligonucleotides complementary to snRNAs. After annealing, *Escherichia coli* RNase H (Takara Shuzo Co.) and $MgCl_2$ were added to 100 U/ml and 1.5 mM, respectively, and the digestion was carried out at 30°C for 10 min in the presence of 1 mg/ml of yeast tRNAs.

Acknowledgments

We are grateful to Dr. Naoya Tsurushita and Kazuma Tomizuka for the gift of μ gene plasmids and for valuable advice. We thank Kazuyuki Hoshijima, Dr. Hiroshi Sakamoto, and Dr. Kunio Inoue for critical reading of the manuscript, as well as helpful discussion. We thank Dr. Iain Hagan for proofreading. This work was supported by grants from the Ministry of Education and Science and from Mitsubishi Foundation.

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

References

- Bennett, M., S. Piñol-Roma, D. Staknis, G. Dreyfuss, and R. Reed. 1992. Differential binding of heterogeneous nuclear ribonucleoproteins to mRNA precursors prior to spliceosome assembly in vitro. *Mol. Cell. Biol.* **12**: 3165–3175.
- Black, D.L. 1991. Does steric interference between splice sites block the splicing of a short *c-src* neuron-specific exon in non-neuronal cells? *Genes & Dev.* **5**: 389–402.
- Brody, E. and J. Abelson. 1985. The "spliceosome": Yeast pre-messenger RNA associates with a 40S complex in a splicing-dependent reaction. *Science* **228**: 963–967.
- Brunak, S. and J. Engelbrecht. 1991. Prediction of human mRNA donor and acceptor sites from the DNA sequence. *J. Mol. Biol.* **220**: 49–65.
- Cooper, T.A. 1992. *In vitro* splicing of cardiac troponin T precursors. *J. Biol. Chem.* **267**: 5330–5338.
- Cooper, T.A. and C.P. Ordahl. 1989. Nucleotide substitutions within the cardiac troponin T alternative exon disrupt pre-mRNA alternative splicing. *Nucleic Acids Res.* **17**: 7905–7921.
- Cote, G.J., D.T. Stelow, S. Peleg, S.M. Berget, and R.F. Gagel. 1992. Identification of exon sequences and an exon binding protein involved in alternative RNA splicing of calcitonin/CGRP. *Nucleic Acids Res.* **20**: 2361–2366.
- Dignam, J.D., R.M. Lebovitz, and R.G. Roeder. 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* **11**: 1475–1489.
- Dominski, Z. and R. Kole. 1991. Selection of splice sites in pre-mRNAs with short internal exons. *Mol. Cell. Biol.* **11**: 6075–6083.
- Eperon, I.P., I.R. Graham, A.D. Griffiths, and I.C. Eperon. 1988. Effects of RNA secondary structure on alternative splicing of pre-mRNA: Is folding limited to a region behind the transcribing RNA polymerase? *Cell* **54**: 393–401.
- Frederick, D. and W. Keller. 1985. Stepwise assembly of a pre-mRNA splicing complex requires U-snRNPs and specific intron sequences. *Cell* **42**: 355–367.
- Freyer, G.A., J.P. O'Brien, and J. Hurwitz. 1989. Alterations in the polypyrimidine sequence affect the *in vitro* splicing reactions catalyzed by HeLa cell-free preparations. *J. Biol. Chem.* **264**: 14631–14637.
- Fu, X.Y., H. Ge, and J.L. Manley. 1988. The role of the polypyrimidine stretch at the SV40 early pre-mRNA 3' splice site in alternative splicing. *EMBO J.* **7**: 809–817.
- Fu, X.-D., R.A. Kats, A.M. Skalka, and T. Maniatis. 1991. The role of branchpoint and 3'-exon sequences in the control of balanced splicing of avian retrovirus RNA. *Genes & Dev.* **5**: 211–220.
- Furdon, P.J. and R. Kole. 1986. Inhibition of splicing but not cleavage at the 5' splice site by truncating human β -globin pre-mRNA. *Proc. Natl. Acad. Sci.* **83**: 927–931.
- . 1988. The length of the downstream exon and the substitution of specific sequences affect pre-mRNA splicing *in vitro*. *Mol. Cell. Biol.* **8**: 860–866.
- Grabowski, P.J., S.R. Seiler, and P.A. Sharp. 1985. A multicomponent complex is involved in the splicing of messenger RNA precursors. *Cell* **42**: 345–353.
- Green, M.R. 1991. Biochemical mechanisms of constitutive and regulated pre-mRNA splicing. *Annu. Rev. Cell Biol.* **7**: 559–599.
- Hampson, R.K., L. La Follette, and F.M. Rottman. 1989. Alternative processing of bovine growth hormone mRNA is influenced by downstream exon sequences. *Mol. Cell. Biol.* **9**: 1604–1610.
- Hedley, M.L. and T. Maniatis. 1991. Sex-specific splicing and polyadenylation of *dsx* pre-mRNA requires a sequence that binds specifically to tra-2 protein *in vitro*. *Cell* **65**: 579–586.
- Helfman, D.M., W.M. Rich, and L.A. Finn. 1988. Alternative splicing of tropomyosin pre-mRNAs *in vitro* and *in vivo*. *Genes & Dev.* **2**: 1627–1638.
- Hoshijima, K., K. Inoue, I. Higuchi, H. Sakamoto, and Y. Shimura. 1991. Control of *doublesex* alternative splicing by *transformer* and *transformer-2* in *Drosophila*. *Science* **252**: 833–836.
- Inoue, K., K. Hoshijima, I. Higuchi, H. Sakamoto, and Y. Shimura. 1992. Binding of the *Drosophila transformer* and *transformer-2* proteins to the regulatory elements of *doublesex* primary transcript for sex-specific RNA processing. *Proc. Natl. Acad. Sci.* **89**: 8092–8096.
- Kakizuka, A., T. Ingi, T. Murai, and S. Nakanishi. 1990. A set of U1 snRNA-complementary sequences involved in governing alternative RNA splicing of the Kininogen genes. *J. Biol. Chem.* **265**: 10102–10108.
- Katz, R.A. and A.M. Skalka. 1990. Control of retroviral RNA splicing through maintenance of suboptimal processing signals. *Mol. Cell. Biol.* **10**: 696–704.
- Konarska, M.M. and P.A. Sharp. 1986. Electrophoretic separation of complexes in the splicing of precursors to mRNAs. *Cell* **46**: 845–855.
- . 1987. Interactions between small nuclear ribonucleoprotein particles in formation of spliceosomes. *Cell* **49**: 763–774.
- Krainer, A.R. and T. Maniatis. 1988. RNA splicing. *Frontiers in molecular biology: Transcription and splicing* (ed. B.D. Hames and D.M. Glover), pp. 131–206. IRL Press, Oxford/Washington D.C.
- Kreivi, J.-P., K. Zefrivotz, and G. Akusjarvi. 1991. A U1 snRNA binding site improves the efficiency of *in vitro* pre-mRNA splicing. *Nucleic Acids Res.* **19**: 6956.
- Kuo, H.-C., F.H. Nasim, and P.J. Grabowski. 1991. Control of alternative splicing by differential binding of U1 small nuclear ribonucleoprotein particle. *Science* **251**: 1045–1050.
- Libri, D., M. Goux-Pelletan, E. Brody, and M.Y. Fisman. 1990. Exon as well as intron sequences are *cis*-regulating elements for the mutually exclusive alternative splicing of the β tropo-

- myosin gene. *Mol. Cell. Biol.* **10**: 5036–5046.
- Ligtenberg, M.J.L., M.C. Gennissen, H.L. Vos, and J. Hilken. 1990. A single nucleotide polymorphism in an exon dictates allele dependent differential splicing of episialin mRNA. *Nucleic Acids Res.* **19**: 297–301.
- Lowery, D.E., and B.G. Van Ness. 1988. Comparison of *in vitro* and *in vivo* splice site selection in κ -immunoglobulin precursor mRNA. *Mol. Cell. Biol.* **8**: 2610–2619.
- Mardon, H.J., G. Sebastio, and F.E. Baralle. 1987. A role for exon sequences in alternative splicing of the human fibronectin gene. *Nucleic Acids Res.* **15**: 7725–7733.
- Michaud, S. and R. Reed. 1991. An ATP-independent complex commits pre-mRNA to the mammalian spliceosome assembly pathway. *Genes & Dev.* **5**: 2534–2546.
- Nagoshi, R.N. and B.S. Baker. 1990. Regulation of sex-specific RNA splicing at the *Drosophila doublesex* gene: *Cis*-acting mutations in exon sequences alter sex-specific RNA splicing patterns. *Genes & Dev.* **4**: 89–97.
- Nelson, K.K. and M.R. Green. 1988. Splice site selection and ribonucleoprotein complex assembly during *in vitro* pre-mRNA splicing. *Genes & Dev.* **2**: 319–329.
- . 1990. Mechanism for cryptic splice site activation during pre-mRNA splicing. *Proc. Natl. Acad. Sci.* **87**: 6253–6257.
- Oshima, Y. and Y. Gotoh. 1987. Signals for the selection of a splice site in pre-mRNA. *J. Mol. Biol.* **195**: 247–259.
- Parent, A., S. Zeitlin, and A. Efstratiadis. 1987. Minimal exon sequence requirements for efficient *in vitro* splicing of mono-intronic nuclear pre-mRNA. *J. Biol. Chem.* **262**: 11284–11291.
- Peterson, M.L. and R.P. Perry. 1989. The regulated production of μ m and μ s mRNA is dependent on the relative efficiencies of μ s poly(A) site usage and the C μ 4-to-M1 splice. *Mol. Cell. Biol.* **9**: 726–738.
- Reed, R. and T. Maniatis. 1986. A role for exon sequences and splice-site proximity in splice site selection. *Cell* **46**: 681–690.
- Ricketts, M.H., M.J. Simons, J. Parma, L. Mercken, Q. Dong, and G. Vassart. 1987. A nonsense mutation causes hereditary goitre in the Afrikaner cattle and unmasks alternative splicing of thyroglobulin transcripts. *Proc. Natl. Acad. Sci.* **84**: 3181–3184.
- Robberson, B.L., G.J. Cote, and S.M. Berget. 1990. Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol. Cell. Biol.* **10**: 84–94.
- Ryner, L.C. and S.B. Baker. 1991. Regulation of *doublesex* pre-mRNA processing occurs by 3'-splice site activation. *Genes & Dev.* **5**: 2071–2085.
- Sakamoto, H., M. Ohno, K. Yasuda, K. Mizumoto, and Y. Shimura. 1987. *In vitro* splicing of a chicken δ -crystallin pre-mRNA in a mammalian extract. *J. Biochem.* **102**: 1289–1301.
- Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: A laboratory manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
- Sawa, H. and Y. Shimura. 1992. Association of U6 snRNA with the 5' splice site region of mRNA precursor in the spliceosome. *Genes & Dev.* **6**: 244–254.
- Shapiro, M.B. and P. Senapathy. 1987. RNA splice junctions of different classes of eukaryotes: Sequence statistics and functional implications in gene expression. *Nucleic Acids Res.* **15**: 7155–7174.
- Solnick, D. 1985. Alternative splicing caused by RNA secondary structure. *Cell* **43**: 667–676.
- Solnick, D. and S.I. Lee. 1987. Amount of RNA secondary structure required to induce an alternative splice. *Mol. Cell. Biol.* **7**: 3194–3198.
- Somasekhar, M.B. and J.E. Mertz. 1985. Exon mutations that affect the choice of splice sites used in processing the SV40 late transcripts. *Nucleic Acids Res.* **13**: 5591–5609.
- Steingrimsdottir, H., G. Rowley, G. Dorado, J. Cole, and A.R. Lehmann. 1992. Mutations which alter splicing in the human hypoxanthine-guanine phosphoribosyltransferase gene. *Nucleic Acids Res.* **20**: 1201–1208.
- Streuli, M. and H. Saito. 1989. Regulation of tissue-specific alternative splicing: Exon-specific *cis*-elements govern the splicing of leukocyte common antigen pre-mRNA. *EMBO J.* **8**: 787–796.
- Talerico, M. and S.M. Berget. 1990. Effect of 5' splice site mutations on splicing of the preceding intron. *Mol. Cell. Biol.* **10**: 6299–6305.
- Tian, M. and T. Maniatis. 1992. Positive control of pre-mRNA splicing *in vitro*. *Science* **256**: 237–240.
- Tsai, A.Y., M. Streuli, and H. Saito. 1989. Integrity of the exon 6 sequence is essential for tissue-specific alternative splicing of human leukocyte common antigen pre-mRNA. *Mol. Cell. Biol.* **9**: 4550–4555.
- Tsurushita, N., N.M. Avdalovic, and L.J. Korn. 1987. Regulation of differential processing of mouse immunoglobulin μ heavy-chain mRNA. *Nucleic Acids Res.* **15**: 4603–4615.
- Wakamatsu, N., H. Kobayashi, T. Miyatake, and S. Tsuji. 1992. A novel exon mutation in the human β -hexosaminidase β subunit gene affects 3' splice site selection. *J. Biol. Chem.* **267**: 2406–2413.
- Watakabe, A., K. Inoue, H. Sakamoto, and Y. Shimura. 1989. A secondary structure at the 3' splice site affects the *in vitro* splicing reaction of mouse immunoglobulin μ chain pre-mRNAs. *Nucleic Acids Res.* **17**: 8159–8169.
- Watakabe, A., H. Sakamoto, and Y. Shimura. 1991. Repositioning of an alternative exon sequence of mouse IgM pre-mRNA activates splicing of the preceding intron. *Gene Expression* **1**: 175–184.
- Zamore, P.D., J.G. Patton, and M.R. Green. 1992. Cloning and domain structure of the mammalian splicing factor U2AF. *Science* **355**: 609–614.

The role of exon sequences in splice site selection.

A Watakabe, K Tanaka and Y Shimura

Genes Dev. 1993, **7**:

Access the most recent version at doi:[10.1101/gad.7.3.407](https://doi.org/10.1101/gad.7.3.407)

References

This article cites 60 articles, 33 of which can be accessed free at:
<http://genesdev.cshlp.org/content/7/3/407.full.html#ref-list-1>

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

Focused on your science.

USC
SCIENTIFIC
SIDE BY SCIENCE