MOLECULAR AND CELLULAR BIOLOGY, Feb. 1994, p. 1347–1354
0270-7306/94/$04.00+0

Vol. 14, No. 2

Polypurine Sequences within a Downstream Exon Function as a Splicing Enhancer

KENJI TANAKA,† AKIYA WATAKABE, AND YOSHIRO SHIMURA*

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

Received 12 July 1993/Returned for modification 2 September 1993/Accepted 19 November 1993

We have previously shown that a purine-rich sequence located within exon M2 of the mouse immunoglobulin μ gene functions as a splicing enhancer, as judged by its ability to stimulate splicing of a distant upstream intron. This sequence element has been designated ERS (exon recognition sequence). In this study, we investigated the stimulatory effects of various ERS-like sequences, using the in vitro splicing system with HeLa cell nuclear extracts. Here, we show that purine-rich sequences of several natural exons that have previously been shown to be required for splicing function as a splicing enhancer like the ERS of the immunoglobulin μ gene. Moreover, even synthetic polypurine sequences had stimulatory effects on the upstream splicing. Evaluation of the data obtained from the analyses of both natural and synthetic purine-rich sequences shows that (i) alternating purine sequences can stimulate splicing, while poly(A) or poly(G) sequences cannot, and (ii) the presence of U residues within the polypurine sequence greatly reduces the level of stimulation. Competition experiments strongly suggest that the stimulatory effects of various purine-rich sequences are mediated by the same *trans*-acting factor(s). We conclude from these results that the purine-rich sequences that we examined in this study also represent examples of ERS. Thus, ERS is considered a general splicing element that is present in various exons and plays an important role in splice site selection.

One of the most characteristic features of eucaryotic genes is the presence of intron sequences that interrupt the coding region. Removal of introns by pre-mRNA splicing is, therefore, an essential step for gene expression in eucaryotes. Sequence comparisons and mutational studies have revealed several conserved sequence elements that are required to distinguish exons and introns (for reviews, see references 14 and 22). The best-conserved elements are present at the exon-intron boundaries, namely, the 5′ and 3′ splice sites, whose consensus sequences in mammals are AG/GURAGU and CAG/G, respectively (37). In addition, a polypyrimidine stretch of variable length and a degenerate branchpoint sequence, preceding the 3′ splice site, also play an essential role.

Although these elements are important, they do not seem to be sufficient to account for the accuracy of exon-intron discrimination. Because these elements allow several mismatch deviations from the consensus, potential splice site sequences are present at various sites other than the authentic splice sites. Generally, sequences that show a better match to the consensus are more tightly bound by splicing factors (30, 50) and are more frequently used as the authentic splice sites (2, 31). However, seemingly "good" sequences are not necessarily used as the natural splice sites; conversely, a sequence that is apparently distant from the consensus is sometimes used as an authentic splice site (2, 19). Moreover, it has been shown that the selection of a splice site depends not only on its sequence but on the context where that sequence resides (29). These observations suggest that additional sequence elements may be required for the selection of splice sites. Regarding this point, it was shown that the length of an exon (1a, 9, 13, 34) or the secondary structure of the region around splice sites (10, 11, 25, 38, 39, 46) affects splice site selection. In addition, there are numerous reports that suggest a role for specific exon sequences in splicing (4, 6, 12, 15, 16, 18, 20, 24, 26, 27, 32, 33, 40, 41, 45, 48, 49).

We have previously shown that a purine-rich sequence located within the last exon, M2, of the mouse immunoglobulin μ (IgM) gene plays an essential role in the splicing of this gene (47, 48). We found that this sequence functions as a splicing enhancer: it stimulated splicing of a distant intron that is present in the upstream region. This stimulatory effect was observed even with introns derived from different genes (48). We also found that this sequence promotes the formation of the early splicing complex (48). These results show that the sequence that we identified within IgM exon M2 represents a novel element involved in splice site selection. We designated this sequence element ERS (exon recognition sequence).

Moreover, we have found that several exon sequences whose mutation or deletion affects splice site selection contain purine-rich sequences similar to the ERS of IgM exon M2 (IgM-ERS) (see Table 1 in reference 48). The examples include the avian sarcoma-leukosis virus (ASLV) gene (12, 21), the chicken cardiac troponin T (cTNT) gene (4, 5, 49), the hypoxanthine-guanine phosphoribosyltransferase (hprt) gene (41), the bovine growth hormone (BGH) gene (15), the rat beta-tropomyosin gene (16), the fibronectin gene (27), and the neural cell adhesion molecule (NCAM) gene (42). This observation raised the possibility that the purine-rich sequences of these genes are other examples of ERS and that the effects of exon mutations are ascribed to the loss of the enhancer activity of ERS. In this regard, Xu et al. recently reported that both the purine-rich sequence of the cTNT exon and the synthetic GAAGAGGAGG repeat sequence facilitate splicing of a heterologous intron in in vivo experiments (49). This finding is consistent with our notion

* Corresponding author. Phone: 81-75-753-4231. Fax: 81-75-753-4235. Electronic mail address: g53376@sakura.kudpc.kyoto-u.ac.jp.

† Present address: Molecular Biology Research Laboratories, Tsukuba Research Institute, Banyu Pharmaceutical Co. Ltd., Tsukuba 300-33, Japan.

that purine-rich exon sequences found in various genes serve as splicing enhancer elements.

In this study, we extensively examined the stimulatory effect of natural and synthetic purine-rich sequences on the splicing of a heterologous intron of the *Drosophila melanogaster doublesex* gene. We used the in vitro splicing system with HeLa cell nuclear extracts, in which we can directly examine the stimulatory effect of the purine-rich sequences on the kinetics of the splicing reaction. We found that (i) alternating purine sequences can function as a splicing enhancer, while poly(A) or poly(G) sequences cannot, and (ii) the presence of U residues within the polypurine sequence greatly reduces the level of stimulation. Moreover, we provide evidence that such a purine-rich element is recognized by a common *trans*-acting factor(s). We conclude that ERS is a general splicing element that is present in various genes and plays a vital role in splice site selection of pre-mRNA.

MATERIALS AND METHODS

Plasmid construction. All constructions were made by using standard cloning procedures (36) and confirmed by sequencing.

To construct the template plasmids for the chimeric *dsx* pre-mRNAs, p*dsx*-Sa (48) was cut with *Xba*I and *Hin*dIII and annealed oligonucleotides were inserted into these sites. The template plasmids for competitor RNAs were generated in the same manner except that the pSP72 vector (Promega) was used as the starting plasmid. The oligonucleotides used here are 5′-CTAGAXnA-3′ and 5′-AGCTTXRnT-3′, where Xn and XRn represent various complementary sequences. (See the descriptions of each sequence shown in Fig. 1 and 6 and Tables 1 and 2.)

Pre-mRNA preparation and in vitro splicing. In vitro transcription was carried out with T7 RNA polymerase in the presence of the cap analog (m^7Gp_3G) and [α-^{32}P]GTP. For the preparation of the competitor RNAs, the cap analog and the radiolabeled nucleotides were omitted. HeLa cell nuclear extracts were prepared as described previously (8). The splicing reaction was carried out with 10 μl of a previously described reaction mixture (35). The reaction products were electrophoresed and quantitated with a BA100A Bio-image analyzer (FUJIX).

Competition experiments were carried out as follows. We first incubated a HeLa cell nuclear extract with competitor RNA on ice for 10 min. Radiolabeled pre-mRNA and 10× SP solution (200 mM creatine phosphate, 5 mM ATP, 20 mM $MgCl_2$) were then added to give the standard splicing conditions (35). Incubation was carried out at 30°C for a further 40 min, and the reaction products were analyzed.

RESULTS

Stimulation of splicing by exon sequences from various genes. Several examples of purine-rich exon sequences are shown in Fig. 1B. Deletions or substitutions of these nucleotide sequences were previously shown to affect the splicing of each gene transcript (see the legend for Fig. 1). To investigate the stimulatory effect of these purine-rich exon sequences, we used the variants of the single-intron substrate derived from the *Drosophila doublesex* (*dsx*) gene and carried out in vitro splicing with HeLa cell nuclear extracts. The basic construct used in this study contains the entire third exon, the third intron, and 20 nucleotides (nt) of the 5′ portion of the fourth exon of the *dsx* gene as well as short

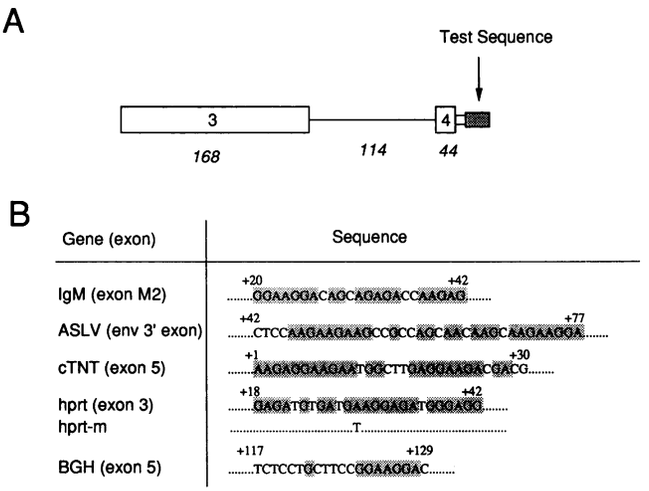

FIG. 1. (A) Schematic representation of *Drosophila doublesex* (*dsx*) chimeric pre-mRNAs. Exon sequences (boxes) and intron sequences (lines) are shown. The linker sequence in the 3′ exon (small box between box 4 and shaded box) that is derived from the pSP72 vector and the insertion site of the test sequence (shaded box) are indicated. The lengths (in nucleotides) of the exons and introns are indicated below the respective regions of the construct. The 3′ exon contains 30 nt of the *dsx* fourth exon and 14 nt of linker sequence derived from pSP72. (B) Natural exon sequences employed for examination of stimulatory effect on splicing. Deletions or substitutions of these sequences are shown to affect splicing of the upstream intron (for IgM [47, 48], ASLV [12, 21], cTNT [4, 49], and BGH [15]) or result in exon skipping (for cTNT [4, 5, 49] and hprt [41]). The purine residues of these sequences are shown by shading. The numbers above the sequences indicate the positions in the original exon. The sequence of hprt-m is exactly the same as that of hprt except for one base substitution (A to T) at the indicated position. env, envelope gene.

leader and trailer sequences derived from the pSP72 vector (Fig. 1A). The polypyrimidine stretch of this third intron was previously shown to be suboptimal (18, 44). Therefore, splicing does not occur with this substrate in a HeLa cell nuclear extract in the absence of any extra exon sequence at the 3′ end of the construct (Fig. 2, lanes 1 to 3). However, when we introduced the ERS of IgM exon M2 to the 3′ side of the downstream exon, splicing was strongly stimulated (Fig. 2, lanes 4 to 6) (48). Using this system, we tested the stimulatory effect of the purine-rich sequences from other genes, such as the ASLV, TNT, hprt, and BGH genes (Fig. 1B). We found that all of these sequences can stimulate splicing, although to different extents (Fig. 2, lanes 7 to 15 and 19 to 21). For example, the levels of stimulation by the hprt and BGH sequences are approximately 1/10 and 1/6, respectively, of that by the IgM sequence, as judged by the accumulation of the final spliced products at 40 min (data not shown). In this case, the difference between such stimulatory effects is due not to the difference in the downstream exon sizes but rather to the sequence contents, because the hprt-m sequence has almost no stimulatory effect on splicing although it is approximately the same size as the IgM sequence. Thus, the purine-rich sequences of natural exons in other genes, like that of the IgM gene, could function as splicing enhancers.

Two lines of evidence suggest that it is the polypurine sequences that are important for splicing stimulation. First, a purine-to-pyrimidine mutation that was reported to affect hprt splicing in culture cells (41) also decreased the stimula-

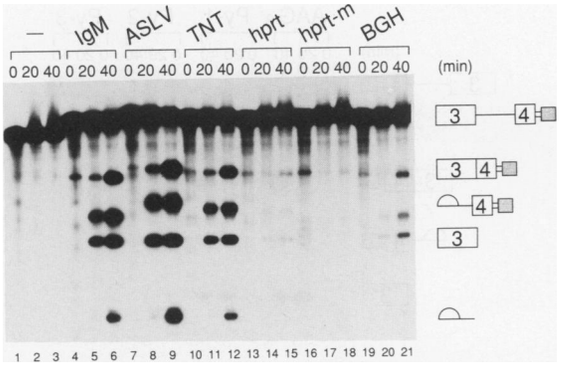

FIG. 2. Stimulation of upstream splicing by various exon sequences. In vitro splicing of chimeric *dsx* pre-mRNAs in a HeLa cell nuclear extract is shown. The substrates used in this experiment contain the sequences listed in Fig. 1B at the site indicated in Fig. 1A. The pre-mRNAs (10 fmol each) were incubated in a HeLa cell nuclear extract at 30°C for the times indicated at the top of each lane. Electrophoresis was carried out with a 5% polyacrylamide gel containing 8 M urea. The bands for the RNA products are shown schematically on the right (with sequences represented as in Fig. 1A; semicircle, lariat RNA molecule). Lanes 1 to 3, *dsx* pre-mRNA with no exon sequence (−); lanes 4 to 6, *dsx*-IgM; lanes 7 to 9, *dsx*-ASLV; lanes 10 to 12, *dsx*-TNT; lanes 13 to 15, *dsx*-hprt; lanes 16 to 18, *dsx*-hprt-m; lanes 19 to 21, *dsx*-BGH.

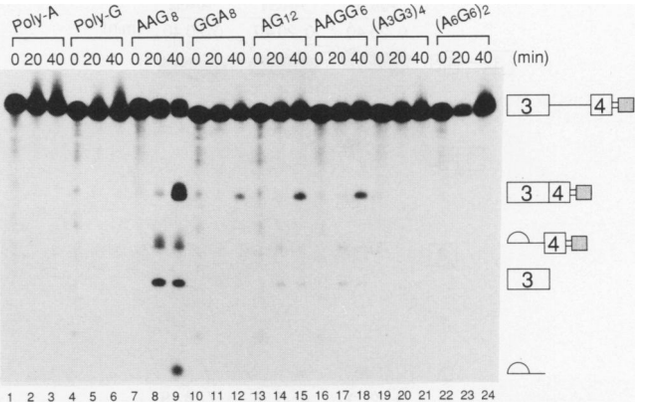

FIG. 3. Stimulation of upstream splicing by synthetic polypurine sequences. The synthetic polypurine sequences listed in Table 1 were connected to *dsx* pre-mRNA, and their stimulatory effects on splicing were tested. After the standard reaction for the times indicated at the top of each lane, RNAs were extracted and electrophoresed on a 5% polyacrylamide gel containing 8 M urea. The bands for the RNA products are shown schematically on the right (with sequences represented as for Fig. 1A; semicircle, lariat RNA molecule). Lanes 1 to 3, *dsx*-poly(A); lanes 4 to 6, *dsx*-poly(G); lanes 7 to 9, *dsx*-AAG_8; lanes 10 to 12, *dsx*-GGA_8; lanes 13 to 15, *dsx*-AG_{12}; lanes 16 to 18, *dsx*-$AAGG_6$; lanes 19 to 21, *dsx*-$(A_3G_3)_4$; lanes 22 to 24, *dsx*-$(A_6G_6)_2$.

tory effect of the hprt exon sequence in our system (Fig. 2, lanes 13 to 18). This decrease is rather small and may not be necessarily clear in the figure, especially since the band present near the spliced product band is degraded. However, while faint bands corresponding to the 5′ exon and the lariat intermediate were observed with the hprt sequence, no bands corresponding to these products were observed with the hprt-m sequence. This difference became more apparent in another film of longer exposure (data not shown). Second, deletion of the pyrimidine-rich portion of the BGH sequence did not affect splicing stimulation (data not shown). Thus, the polypurine segments of the exons listed in Fig. 1 seem to represent important sequences for splicing stimulation.

Splicing stimulation by synthetic polypurine sequences. To confirm the importance of the polypurine sequence and to investigate the sequence requirements for splicing stimulation in more detail, we connected various synthetic polypurine sequences to *dsx* pre-mRNA and tested their stimulatory effects on splicing. In this experiment, 24 nt of various polypurine sequences was used (Table 1). When the poly(A) or poly(G) sequence was connected, splicing was not stimulated at all (Fig. 3, lanes 1 to 6). In contrast, eight consecutive repeats of AAG had a strong stimulatory effect on the upstream splicing (Fig. 3, lanes 7 to 9). The extent of stimulation was approximately the same as that by the IgM sequence (data not shown). Other sequences (such as GGA, AG, and AAGG repeats) also stimulated splicing, although at lower efficiencies (Fig. 3, lanes 10 to 18). The levels of stimulation by these sequences were approximately 4.5-, 3-, and 3.5-fold less, respectively, than that by the AAG repeat. A_3G_3 repeats and A_6G_6 repeats failed to stimulate splicing (Fig. 3, lanes 19 to 24). These results demonstrate that alternating purine sequences can stimulate the splicing of the upstream intron, even without any pyrimidine residues in the sequences. Moreover, our results show that the stimulatory effects are variable, depending on the given polypurine sequences.

TABLE 1. Synthetic polypurine sequences tested for splicing stimulation

Element	Sequence of:		
	*Xba*I	Insert	*Hin*dIII
Poly(A)	CTAGA	AAAAAAAAAAAAAAAAAAAAAAAA	AAGCT
Poly(G)	CTAGA	GGGGGGGGGGGGGGGGGGGGGGGG	AAGCT
AAG_8	CTAGA	AAGAAGAAGAAGAAGAAGAAGAAG	AAGCT
GGA_8	CTAGA	GGAGGAGGAGGAGGAGGAGGAGGA	AAGCT
AG_{12}	CTAGA	AGAGAGAGAGAGAGAGAGAGAGAG	AAGCT
$AAGG_6$	CTAGA	AAGGAAGGAAGGAAGGAAGGAAGG	AAGCT
$(A_3G_3)_6$	CTAGA	AAAGGGAAAGGGAAAGGGAAAGGG	AAGCT
$(A_6G_6)_2$	CTAGA	AAAAAAGGGGGGAAAAAAGGGGGG	AAGCT
AAG_4	CTAGA	AAGAAGAAGAAG	AAGCT
AAG_2	CTAGA	AAGAAG	AAGCT

Next, we examined whether the length of polypurine sequence affects the efficiency of splicing stimulation. In addition to the *dsx* pre-mRNA containing eight repeats of AAG (*dsx*-AAG_8), we constructed substrates containing four (*dsx*-AAG_4) and two (*dsx*-AAG_2) repeats of AAG (Table 1). Because of the presence of purine residues adjacent to the connection site, these pre-mRNAs contain 30, 18, and 12 nt of consecutive polypurine stretches, respectively. As shown in Fig. 3, the pre-mRNA with eight AAG repeats was spliced very efficiently (Fig. 4, lanes 1 to 3). The pre-mRNA with four AAG repeats was spliced slightly less efficiently (Fig. 4, lanes 4 to 6). The pre-mRNA with two AAG repeats was spliced but with much lower efficiency, approximately one-fifth, than the pre-mRNA with eight AAG repeats (Fig. 4, lanes 7 to 9). The results show that the stimulatory effect of the polypurine sequence depends on the length of the stretch.

Effect of pyrimidine residues on splicing stimulation. The natural exon sequences that we tested for splicing stimulation contain several pyrimidine residues within the purine-

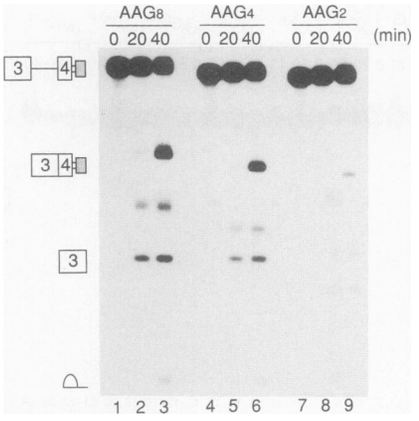

FIG. 4. Effect of the length of the polypurine sequence on splicing stimulation. Polypurine sequences containing eight, four, and two repeats of AAG (Table 1) were connected to *dsx* pre-mRNA, and their stimulatory effects on splicing were tested. After the standard reaction for the times indicated at the top of each lane, electrophoresis was carried out on a 5% polyacrylamide gel containing 8 M urea. Lanes 1 to 3, *dsx*-AAG_8; lanes 4 to 6, *dsx*-AAG_4; lanes 7 to 9, *dsx*-AAG_2. The bands for the RNA products are shown schematically on the left (with sequences represented as for Fig. 1A; semicircle, lariat RNA molecule).

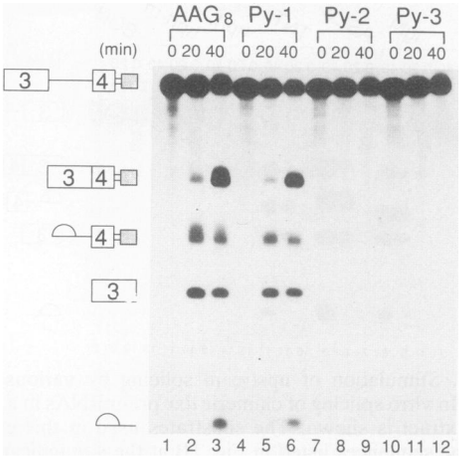

FIG. 5. Effect of pyrimidine residues on splicing stimulation. Synthetic sequences that contain several pyrimidine residues within the polypurine stretch (Table 2) were connected to *dsx* pre-mRNA, and their stimulatory effects on splicing were tested. In *dsx*-Py-1, *dsx*-Py-2, and *dsx*-Py-3 pre-mRNAs, a polypurine stretch is divided into two, three, and five parts, respectively, by pyrimidine residues. After the standard reaction for the times indicated at the top of each lane, RNAs were electrophoresed on a 5% polyacrylamide gel containing 8 M urea. Lanes 1 to 3, *dsx*-AAG_8; lanes 4 to 6, *dsx*-Py-1; lanes 7 to 9, *dsx*-Py-2; lanes 10 to 12, *dsx*-Py-3. The bands for the RNA products are shown schematically on the left (with sequences represented as for Fig. 1A; semicircle, lariat RNA molecule).

rich region (Fig. 1B). To examine the effect of the pyrimidine residues on splicing stimulation, we replaced eight purine residues of *dsx*-AAG_8 pre-mRNA with pyrimidine residues and divided the polypurine stretch of this pre-mRNA in three different ways (Table 2). In *dsx*-Py-1 pre-mRNA, the polypurine stretch was divided into two parts by alternating C and U(T). In *dsx*-Py-2 and *dsx*-Py-3 pre-mRNAs, the polypurine stretches are divided into three and five parts, respectively. As shown in Fig. 5, *dsx*-Py-1 pre-mRNA was spliced almost as efficiently as *dsx*-AAG_8 (Fig. 5, lanes 1 to 6). In contrast, splicing of *dsx*-Py-2 and *dsx*-Py-3 pre-mRNAs was almost undetectable (Fig. 5, lanes 7 to 12). These results suggest that the length of the consecutive purine residues is critical for splicing stimulation.

Moreover, the comparison of *dsx*-Py-1 and *dsx*-AAG_2 pre-mRNAs is informative. The *dsx*-Py-1 pre-mRNA that contains two stretches of 11-nt purine residues was spliced much more efficiently than the *dsx*-AAG_2 pre-mRNA that contains a single stretch of 12-nt purine residues (Fig. 4, lanes 7 to 9; Fig. 5, lanes 4 to 6). This suggests that multiple polypurine stretches have an accumulative effect on splicing even when they are separated by pyrimidine residues.

Next, we asked whether C and U residues have any different effects on splicing stimulation. In this connection, the IgM sequence that strongly stimulates splicing contains four C residues within the polypurine stretch, while the hprt sequence that weakly stimulates splicing contains four T(U) residues (Fig. 6A). To investigate the differential effect of the two pyrimidine residues, all the C's of the IgM sequence were replaced with T(U)s and all the T(U)s of the hprt sequence were changed to C's. When the C's in the IgM sequence were changed to T(U)s, the stimulatory effect of the IgM sequence was completely lost (Fig. 6B, lanes 1 to 6). In contrast, when the T(U)s in the hprt gene were changed to C's, the stimulatory effect of the hprt sequence was enhanced approximately 10-fold (Fig. 6B, lanes 7 to 12). These results demonstrate that the presence of U residues in the polypurine stretch greatly reduces the level of splicing stimulation by the polypurine sequence, while C residues seem to be neutral.

TABLE 2. Synthetic polypurine sequences containing pyrimidine interruptions

Element	Sequence of:		
	*Xba*I	Insert	*Hin*dIII
AAG_8	CTAGA	AAGAAGAAGAAGAAGAAGAAGAAG	AAGCT
Py-1	CTAGA	AAGAAGAACTCTCTCTGAAGAAGA	AAGCT
Py-2	CTAGA	AAGACTCTAGAAGAAGCTCTAAGA	AAGCT
Py-3	CTAGA	ACTAGAAGCTAAGACTAGAAGCTA	AAGCT

Involvement of common *trans*-acting factors in the stimulatory effects of polypurine sequences. The results described above indicate that polypurine stretches with several structural features have a function similar to, if not the same as, that of the ERS of the IgM gene. We have previously shown in the competition experiment that the stimulatory effect of the IgM-ERS is mediated by a *trans*-acting factor(s) (48). To see whether the same factor(s) is involved in splicing stimulation by other polypurine sequences, we carried out competition experiments, using different competitors and different substrates.

In the first series of experiments, the nuclear extracts were preincubated on ice with various competitor RNAs and then the chimeric *dsx* pre-mRNA containing the IgM-ERS was added to the reaction mixture. The splicing reaction was carried out at 30°C for a further 40 min. We first used two competitors containing different regions of the IgM gene. Cont40 RNA contains 40 nt of IgM exon M2 that does not stimulate splicing, while IgM40 contains 40 nt of the region that covers the IgM-ERS (48). Consistent with the previous results, Cont40 RNA did not affect splicing (Fig. 7A, lanes 1

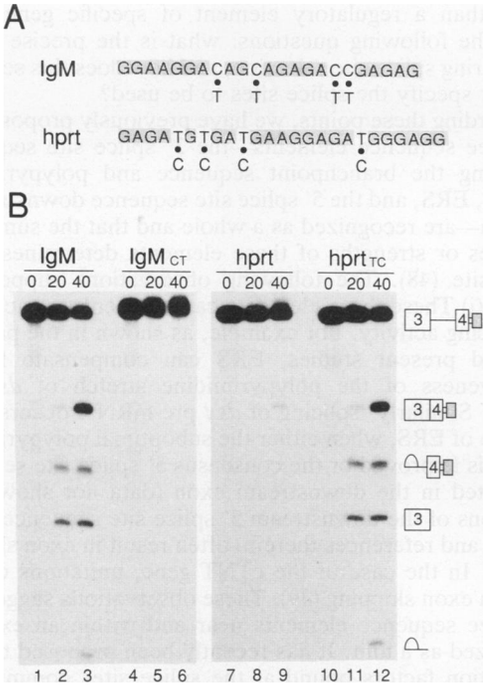

FIG. 6. Effect of U residues on splicing stimulation. (A) The purine-rich sequences of the IgM and the hprt exons. Pyrimidine residues are indicated by the filled circles. (B) In vitro splicing of the *dsx*-IgM and *dsx*-hprt derivative pre-mRNAs. In *dsx*-IgM$_{CT}$ and *dsx*-hprt$_{TC}$ pre-mRNAs, the C and T residues indicated by the filled circles (in panel A) are mutated to T and C residues, respectively. After the standard reaction for the times indicated at the top of each lane, RNAs were extracted and electrophoresed on a 5% polyacrylamide gel containing 8 M urea. Lanes 1 to 3, *dsx*-IgM; lanes 4 to 6, *dsx*-IgM$_{CT}$; lanes 7 to 9, *dsx*-hprt; lanes 10 to 12, *dsx*-hprt$_{TC}$. The bands for the RNA products are shown schematically on the right (with sequences represented as for Fig. 1A; semicircle, lariat RNA molecule).

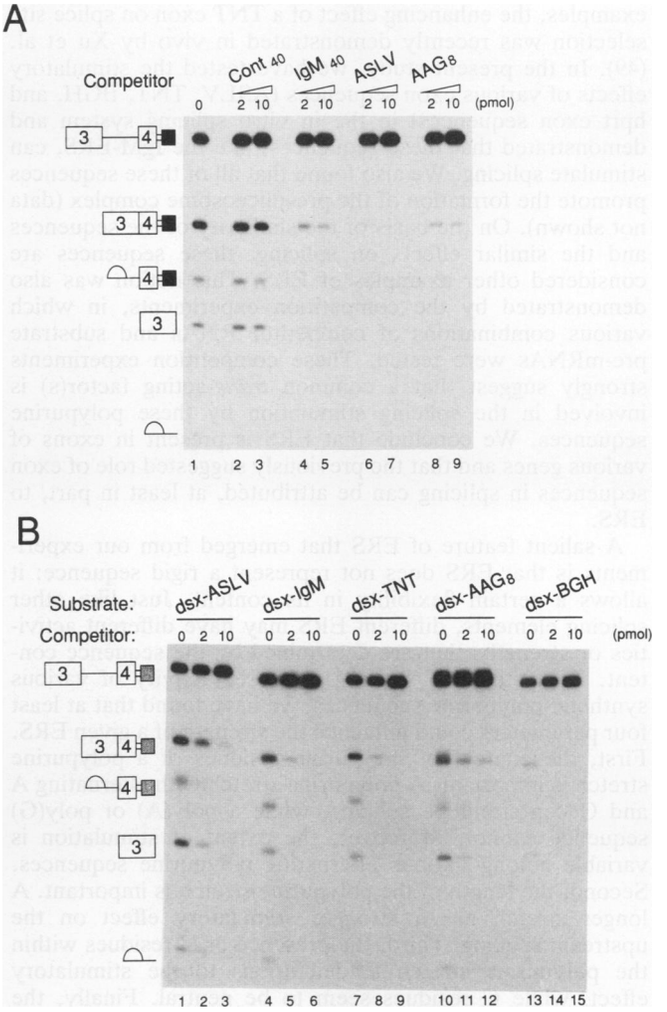

FIG. 7. Competition experiments with RNAs containing polypurine sequences. (A) In vitro splicing carried out with *dsx*-IgM pre-mRNA as the substrate in the presence of various competitor RNAs. HeLa cell nuclear extracts were first incubated on ice with the competitor RNAs indicated above the gel. The *dsx*-IgM pre-mRNA was then added to the reaction mixture, and splicing was carried out at 30°C for a further 40 min. The amounts of the competitors used in the experiment are shown on the top of each lane. Lane 1, no competitor RNA; lanes 2 and 3, Cont40 RNA containing a portion of IgM exon M2 that does not stimulate splicing (48); lanes 4 and 5, IgM40 RNA containing the first 40 nt of IgM exon M2 (48); lanes 6 and 7, ASLV RNA that contains the ASLV sequence (Fig. 1); lanes 8 and 9, AAG$_8$ RNA containing eight repeats of AAG (Table 1). (B) In vitro splicing carried out with various substrates in the presence of ASLV RNA. HeLa cell nuclear extracts were first incubated on ice with various amounts of ASLV RNA (shown at the top of each lane). The *dsx* pre-mRNAs containing various polypurine sequences were then added, and the reaction was continued for a further 40 min. Lanes 1 to 3, *dsx*-ASLV; lanes 4 to 6, *dsx*-IgM; lanes 7 to 9, *dsx*-TNT; lanes 10 to 12, *dsx*-AAG$_8$; lanes 13 to 15, *dsx*-BGH. RNA products are shown schematically on the left (with sequences represented as for Fig. 1A; semicircle, lariat RNA molecule).

to 3), while IgM40 RNA inhibited splicing (Fig. 7A, lanes 4 and 5). Moreover, other polypurine stretches such as the ASLV exon sequence and the synthetic AAG repeat sequence also inhibited splicing of *dsx*-IgM pre-mRNA effectively (Fig. 7A, lanes 6 to 9).

Next, we carried out the in vitro splicing, using the chimeric *dsx* pre-mRNAs with various exon sequences as the substrates and ASLV RNA as the competitor. As shown in Fig. 7B, ASLV RNA interfered with the splicing activities of all the tested *dsx* pre-mRNAs, namely, *dsx*-ASLV (lanes 1 to 3), *dsx*-IgM (lanes 4 to 6), *dsx*-TNT (lanes 7 to 9), *dsx*-AAG$_8$ (lanes 10 to 12), and *dsx*-BGH (lanes 13 to 15) pre-mRNAs. Together with the results in Fig. 7A, these results show that the stimulatory effects of various polypurine sequences are mediated, at least in part, by a common *trans*-acting factor(s) that specifically interacts with such sequences.

DISCUSSION

We previously identified a novel splicing element within exon M2 of the mouse IgM gene (47, 48) that was required for splicing of the upstream introns of exon M2 and that functioned as a splicing enhancer. We designated this sequence element ERS (48). Here, we have demonstrated that the same type of splicing element is present in at least several exons of other genes. Sequence comparison showed that purine-rich sequences similar to the IgM-ERS are often found within the exon sequences that have been reported to play an important role in splice site selection. Among those

examples, the enhancing effect of a TNT exon on splice site selection was recently demonstrated in vivo by Xu et al. (49). In the present study, we have tested the stimulatory effects of various exon sequences (ASLV, TNT, BGH, and hprt exon sequences) in the in vitro splicing system and demonstrated that these sequences, like the IgM-ERS, can stimulate splicing. We also found that all of these sequences promote the formation of the prespliceosome complex (data not shown). On the basis of the similarity of the sequences and the similar effects on splicing, these sequences are considered other examples of ERS. This notion was also demonstrated by the competition experiments, in which various combinations of competitor RNAs and substrate pre-mRNAs were tested. These competition experiments strongly suggest that a common *trans*-acting factor(s) is involved in the splicing stimulation by these polypurine sequences. We conclude that ERS is present in exons of various genes and that the previously suggested role of exon sequences in splicing can be attributed, at least in part, to ERS.

A salient feature of ERS that emerged from our experiments is that ERS does not represent a rigid sequence: it allows a certain flexibility in its content. Just like other splicing elements, different ERS may have different activities or strengths that are determined by the sequence content. By testing the splicing enhancer activity of various synthetic polypurine sequences, we have found that at least four parameters could influence the strength of a given ERS. First, the content of the purine residues of a polypurine stretch is important. A polypurine stretch with alternating A and G can stimulate splicing, while a poly(A) or poly(G) sequence cannot. Moreover, the extent of stimulation is variable among various alternating polypurine sequences. Second, the length of the polypurine stretch is important. A longer stretch has a stronger stimulatory effect on the upstream splicing. Third, the presence of U residues within the polypurine stretch is deleterious to the stimulatory effect, while C residues seem to be neutral. Finally, the presence of multiple stretches of polypurine sequences within an exon seems to have accumulative effects on splicing.

Although we have shown the positive effect of exon sequences (ERS) in this study, the negative effect of exon sequences should also be considered. In vitro experiments using β-globin pre-mRNA have shown that a sequence containing many U residues inhibits splicing when it is present in the downstream exon (13). Moreover, in vivo selection of the mutants in the hprt gene has been described previously (41). This experiment revealed eight exon mutations that result in aberrant splicing. Interestingly, seven of these mutations were A-, G-, or C-to-T substitutions. These results imply that a certain sequence containing U residues could function as an inhibitory element. Good candidates for such element are stop codons (TAA, TAG, and TGA), because there are reports that suggest the negative effect of the stop codons on splicing (7, 28). The dramatic effects of the mutual substitutions of C's and T's seen in Fig. 6 might be caused by the formation or disruption of such an inhibitory element.

In this study, we have shown four examples of ERS present within alternatively spliced exons (the IgM, ASLV, TNT, and BGH genes) and one example present within a constitutively spliced exon (the hprt gene). Considering the relatively loose sequence requirements, ERS appears to be present within a large number of exons besides these examples. Thus, ERS is thought to be a general splicing element rather than a regulatory element of specific genes. This raises the following questions: what is the precise role of ERS during splice site selection, and how does this sequence element specify the splice sites to be used?

Regarding these points, we have previously proposed that the three sequence elements—the 3′ splice site sequences (including the branchpoint sequence and polypyrimidine stretch), ERS, and the 5′ splice site sequence downstream of an exon—are recognized as a whole and that the sum of the activities or strengths of these elements determines the 3′ splice site (48). The following observations support this model. (i) These three elements can compensate each other for splicing activity. For example, as shown in the previous (48) and present studies, ERS can compensate for the defectiveness of the polypyrimidine stretch of *dsx* pre-mRNA. Similarly, splicing of *dsx* pre-mRNA occurs in the absence of ERS, when either the suboptimal polypyrimidine stretch is improved or the consensus 5′ splice site sequence is inserted in the downstream exon (data not shown). (ii) Mutations of the downstream 5′ splice site sequence (reference 43 and references therein) often result in exon skipping in vivo. In the case of the cTNT gene, mutations of ERS result in exon skipping (49). These observations suggest that the three sequence elements near and within an exon are recognized as a unit. It has recently been proposed that the recognition factors bound at the splice sites communicate across an exon to define the exon region prior to intron removal (34). ERS may facilitate the process of such exon recognition. In this connection, it would be important to investigate the effect of ERS on splicing, in the presence of the downstream 5′ splice site. In this respect, in vivo experiments using constructs of multiple introns would be especially informative.

As described previously, Xu et al. showed that the purine-rich sequence of cTNT exon 5 and the synthetic GAAGAG GAGG repeat sequences facilitate splicing in vivo (49). They also showed that insertion of these sequences into a short exon promotes inclusion of that exon in mRNA. These results are quite consistent with our experiments and provide evidence that the purine-rich element (ERS) functions as the determinant of exons and introns in vivo.

In their work, Xu et al. (49) used the term ESE (exon splicing element) to describe the splicing element within an exon. This term is used to designate all kinds of splicing elements within an exon, including gene-specific elements required for the regulation of alternative splicing, general positive-acting elements and putative negative-acting elements. For example, they say that the repeated 13-nt sequence of the *Drosophila doublesex* female exon (3) is the best-characterized ESE. Therefore, ERS is considered a type of ESE and corresponds to what is called a purine-rich ESE.

To clarify the role of ERS in splice site selection, it is also important to identify the factors that are involved in the recognition of ERS and in splicing stimulation. We previously found that the 5′ end of U1 small nuclear RNA (snRNA) is specifically cross-linked to the ERS of the IgM gene (48). This observation raised the possibility that U1 small nuclear ribonucleoprotein (snRNP) may recognize ERS through base pairing at its 5′ end. However, the data presented in this study are not consistent with this possibility. For example, if the base-pairing interaction were the major determinant, a poly(G) sequence that can form three G-C pairs and four G-U pairs with the 5′ end of U1 snRNA (5′-AUACUUACCU) would be able to stimulate splicing. But the poly(G) sequence had no effect on the upstream

splicing. In addition, the strong negative effect of U residues on splicing stimulation cannot be explained simply by the base-pairing interaction with U1 snRNA. Furthermore, we found that single-stranded DNA that contains the same sequence as ASLV RNA does not interfere with splicing (1). Thus, the base-pairing interaction between U1 snRNP and ERS is not likely to be the major determinant of ERS recognition. Presumably, some other factors besides U1 snRNP are responsible for the recognition of ERS and U1 snRNP plays a more indirect role in ERS recognition. Considering the stimulatory effect of U1 snRNP on the upstream splicing (17, 23), U1 snRNP may play a role primarily in splicing stimulation rather than in ERS recognition. The role of U1 snRNP and other factors in ERS recognition and splicing stimulation should be clarified by further investigation.

ACKNOWLEDGMENTS

We thank Kazuyuki Hoshijima, Kunio Inoue, Mutsuhito Ohno, and Tilmann Achsel for critical reading of the manuscript as well as for helpful discussion. This work was supported by grants from the Ministry of Education and Science, Japan, and from the Mitsubishi Foundation.

REFERENCES

1. **Achsel, T.** Unpublished data.
1a. **Black, D. L.** 1991. Does steric interference between splice sites block the splicing of a short c-*src* neuron-specific exon in nonneuronal cells? Genes Dev. **5**:389–402.
2. **Brunak, S., and J. Engelbrecht.** 1991. Prediction of human mRNA donor and acceptor sites from the DNA sequence. J. Mol. Biol. **220**:49–65.
3. **Burtis, K. C., and B. S. Baker.** 1989. Drosophila *doublesex* gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding related sex-specific polypeptides. Cell **56**:997–1010.
4. **Cooper, T. A.** 1992. *In vitro* splicing of cardiac troponin T precursors. J. Biol. Chem. **267**:5330–5338.
5. **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.
6. **Cote, G. J., D. T. Stolow, 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.
7. **Dietz, H. C., D. Valle, C. A. Francomano, R. J. Kendzior, Jr., R. E. Pyeritz, and G. R. Cutting.** 1993. The skipping of constitutive exons in vivo induced by nonsense mutations. Science **259**:680–683.
8. **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.
9. **Dominski, Z., and R. Kole.** 1991. Selection of splice sites in pre-mRNAs with short internal exons. Mol. Cell. Biol. **11**:6075–6083.
10. **Eperon, L. 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.
11. **Estes, P. A., N. E. Cooke, and S. A. Liebhaber.** 1992. A native RNA secondary structure controls alternative splice site selection and generates two human growth hormone isoforms. J. Biol. Chem. **267**:14902–14908.
12. **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.
13. **Furdon, P. J., and R. Kole.** 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.
14. **Green, M. R.** 1991. Biochemical mechanisms of constitutive and regulated pre-mRNA splicing. Annu. Rev. Cell Biol. **7**:559–599.
15. **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.
16. **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.
17. **Hoffmann, B. E., and P. J. Grabowski.** 1992. U1 snRNP targets an essential splicing factor, U2AF, to the 3′ splice site by a network of interactions spanning the exon. Genes Dev. **6**:2554–2568.
18. **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.
19. **Jackson, I. J.** 1991. A reappraisal of non-consensus mRNA splice sites. Nucleic Acids Res. **19**:3795–3798.
20. **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.
21. **Kats, R. A., and A. M. Skalka.** 1990. Control of retroviral RNA splicing through maintenance of suboptimal processing signals. Mol. Cell. Biol. **10**:696–704.
22. **Krainer, A. R., and T. Maniatis.** 1988. RNA splicing, p. 131–206. *In* B. D. Hames and D. M. Glover (ed.), Frontiers in molecular biology: transcription and splicing. IRL Press, Oxford.
23. **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.
24. **Libri, D., M. Goux-Pelletan, E. Brody, and M. Y. Fiszman.** 1990. Exon as well as intron sequences are *cis*-regulating elements for the mutually exclusive alternative splicing of the β tropomyosin gene. Mol. Cell. Biol. **10**:5036–5046.
25. **Libri, D., A. Piseri, and M. Y. Fiszman.** 1991. Tissue-specific splicing in vivo of the β-tropomyosin gene: dependence on an RNA secondary structure. Science **252**:1842–1845.
26. **Ligtenberg, M. J. L., M. C. Gennissen, H. L. Vos, and J. Hilkens.** 1990. A single nucleotide polymorphism in an exon dictates allele dependent differential splicing of episialin mRNA. Nucleic Acids Res. **19**:297–301.
27. **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.
28. **Naeger, L. K., R. V. Schoborg, Q. Zhao, G. E. Tullis, and D. J. Pintel.** 1992. Nonsense mutations inhibit splicing of MVM RNA in *cis* when they interrupt the reading frame of either exon of the final spliced product. Genes Dev. **6**:1107–1119.
29. **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.
30. **Nelson, K. K., and M. R. Green.** 1990. Mechanism for cryptic splice site activation during pre-mRNA splicing. Proc. Natl. Acad. Sci. USA **87**:6253–6257.
31. **Oshima, Y., and Y. Gotoh.** 1987. Signals for the selection of a splice site in pre-mRNA. J. Mol. Biol. **195**:247–259.
32. **Reed, R., and T. Maniatis.** 1986. A role for exon sequences and splice-site proximity in splice site selection. Cell **46**:681–690.
33. **Ricketts, M. H., M. J. Simons, J. Parma, L. Mercken, Q. Dong, and G. Vassart.** 1987. A nonsense mutation causes hereditary goitre in the Afrikander cattle and unmasks alternative splicing of thyroglobulin transcripts. Proc. Natl. Acad. Sci. USA **84**:3181–3184.
34. **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.
35. **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.

36. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
37. **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.
38. **Solnick, D.** 1985. Alternative splicing caused by RNA secondary structure. Cell **43**:667–676.
39. **Solnick, D., and S. I. Lee.** 1987. Amount of RNA secondary structure required to induce an alternative splice. Mol. Cell. Biol. **7**:3194–3198.
40. **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.
41. **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.
42. **Tacke, R., and C. Goridis.** 1991. Alternative splicing in the neural cell adhesion molecule pre-mRNA: regulation of exon 18 skipping depends on the 5′-splice site. Genes Dev. **5**:1416–1429.
43. **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.
44. **Tian, M., and T. Maniatis.** 1992. Positive control of pre-mRNA splicing in vitro. Science **256**:237–240.
45. **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.
46. **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.
47. **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.
48. **Watakabe, A., K. Tanaka, and Y. Shimura.** 1993. The role of exon sequences in splice site selection. Genes Dev. **7**:407–418.
49. **Xu, R., J. Teng, and T. A. Cooper.** 1993. The cardiac troponin T alternative exon contains a novel purine-rich positive splicing element. Mol. Cell. Biol. **13**:3660–3674.
50. **Zamore, P. D., J. G. Patton, and M. R. Green.** 1992. Cloning and domain structure of the mammalian splicing factor U2AF. Nature (London) **355**:609–614.