Acta Biochimica Polonica

Vol. 43 No. 1/1996

161–174

QUARTERLY

Dedicated to Professor David Shugar on the occasion of his 80th birthday

Effects of exon sequences on splicing of model pre-mRNA substrates *in vitro**

Zbigniew Dominski** and Ryszard Kole

Department of Pharmacology and Lineberger Comprehensive Cancer Center, University of North Carolina, Chapel Hill, North Carolina 27599, U.S.A.

Key words: model pre-mRNAs, splice site selection, exon skipping, exon recognition sequences

We used several related pre-mRNA substrates consisting of two introns and three exons to study effects of exon sequences on *in vitro* splicing. By varying the sequence of the internal exon and measuring the frequency of its skipping we confirmed that 26-nucleotide exon element naturally existing in β-globin gene and previously analysed *in vivo*, has a strong stimulatory effect on splicing. Sequence analysis of this element suggests that it belongs to a family of purine-rich splicing elements found in exons of several alternatively spliced pre-mRNAs. The 26-nucleotide element can efficiently function in enhancing inclusion of internal exons regardless of their size and sequence composition, suggesting that it plays a role of a general exon recognition element. The purine-rich element is dispensable in exons flanked by strong splice sites, which promote efficient inclusion of otherwise poorly recognized exons. A row of six cytidines inserted into the internal exon (GC2 mutation) initially considered to stimulate exon inclusion to a similar extent as the purine-rich element (Dominski & Kole, 1994, *J. Biol. Chem.* 269, 23590–23596), appears not to affect splice site selection *in vitro*, and *in vivo* it is likely to act by stabilizing mRNA that includes the internal exon against rapid cytoplasmic degradation.

Splicing of mRNA precursors (pre-mRNAs) occurs in multicomponent complexes called spliceosomes, composed of five small nuclear (sn) RNPs (U1, U2, U4, U5 and U6) and yet unknown number of proteins. The splicing reaction proceeds through two transesterfication steps and depends on accurate recognition of the 5′ and 3′ splice sites (see [1] and [2] for recent reviews). In higher eukaryotes, the consensus sequence at the 5′ splice site is CAG/GUAAGU and $(Py)_n$ NPyAG/G at the 3′ splice site, where / separates exon from intron, Py denotes pyrimidine and N — any nucleotide [3, 4]. Two underlined dinucleotides, GU and AG, located at the beginning and at the end of the introns, respectively, are invariant and found in virtually all splice sites whereas nucleotides in other positions are less conserved.

Although mechanism of splicing clearly depends on the recognition of both splice sites, there is a growing amount of data indicating that internal exon sequences (i.e. other than those being a part of consensus sequence) can significantly contribute to the splicing effi-

*This work was supported in part by a Public Health Service grant (GM32994) from the National Institutes of Health to R.K.

Abbreviations: CMV, human cytomegalovirus; dsx, doublesex; ERS, exon recognition sequence; hnRNP, heterogeneous nuclear ribonucloprotein; snRNP, small nuclear ribonucleoprotein; SR proteins, serine and arginine rich splicing factors; tra, transformer; tra-2, transformer 2;

**Present address: Institute of Biochemistry and Biophysics, Polish Academy of Sciences, A. Pawińskiego 5a, 02-106 Warszawa, Poland.

ciency ([5] and references therein). These so called exon recognition sequences (ERS, [5]) include 13-nucleotide repeats located in the female specific exon of *Drosophila* doublesex (*dsx*) gene [6–8]. Binding of two specialized proteins, *tra* and *tra2*, to these repeats results in inclusion of the female specific exon into mature *dsx* mRNA and determines correct sexual differentiation in *Drosophila* [7, 9–11]. In vertebrates, examples of exon recognition elements include purine rich sequences identified in exons of some alternatively spliced pre-mRNAs [5, 12, 13]. These sequences are likely to be sites of interaction with the family of general splicing factors-SR proteins [13–15]. Recently, various purine rich sequences which are recognized by SR proteins and significantly enhance efficiency of *in vitro* splicing were identified using molecular selection/amplification procedures [16–17].

Are purine rich sequences universal and the only splicing elements existing in exons of higher eukaryotes? We have undertaken a series of experiments to investigate the role of exon sequences in splice site selection. Our previous *in vivo* studies identified two distinct sequences, one composed of six cytidines and the other — containing mostly purines, which, when placed within the exon, greatly stimulated its inclusion into the spliced mRNA [18]. Here, in order to verify the *in vivo* results and to provide a valuable system for studies of exon recognition, we tested these sequences under *in vitro* splicing conditions, using HeLa cells nuclear extract and exogenously synthesized transcripts.

MATERIALS AND METHODS

Construction of plasmids

DNA templates were constructed by subcloning of appropriate restriction fragments from CMV vectors [18] into SP64 vector. All mutations within Ex2-184 pre-mRNA are listed in Table 1. The Ex2 constructs described here are

Table 1
List of mutations within 184-nucleotide internal exon used in this work

Mutation of Ex2-184	Existing sequence	Mutated to	Position in 184-nt exon
GC1	UAAAGGCAUU	**CGGGGGCGCC**	38–47
GC2	UUUUUU	**CCCCCC**	63–68
GC3	UUUUUU	**GGGGGG**	63–68
GC4	AGAAAAAUUU	**GGGGGGGCCC**	82–90
GC4a	AUACAGUCAAAU	**CCACGGUCAAGC**	104–115
GC5	AAUAGCAGAA	**GGCGGCGGGG**	139–150
19	AAUGGC	**GUGGUG**	166–171
32	GGCAGACAUUACG AAUGGC	**GUGGAUGAAGUUG** **GUGGUG**	153–171
45	GAAUAGCAGAAUG GGCAGACAUUACG AAUGGC	**GGGCAAGGUGAAC** **GUGGAUGAAGUUG** **GUGGUG**	140–171
32*	GAAUAGCAGAAU GGCAGACAUUACG AAUGGC	**GGGCAAGGUGAAC** GGCAGACAUUACG **GUGGUG**	140–152 and 166–171
45/C	AGGUGAAC AAGUUGGU	AG**GC**GAAC AA**GC**UGGU	148 163
184/AG	GAAUAGCAGAAUG GGCAGACAUUACG	**GAAGAGGAGAAGA** **GGGAGAGAUUACG**	140–165

All sequences written in 5′ to 3′ orientation. Individual nucleotides within each cluster changed due to point mutations are shown in bold letters. Invariant dinucleotides in 5′ cryptic splice sites are underlined. [1]All numbers indicate position within 184-nucleotide exon in Ex2-184 pre-mRNA except for 184/45/C where numbers refer to positions within internal exon in Ex2-184/45 pre-mRNA.

SP6 equivalents of CMV DUP constructs described previously [18].

In vitro transcription and splicing

^{32}P-Labeled transcripts were synthesized by SP6 RNA polymerase (Promega) from DNA templates cleaved at the *Bam*HI site near the 3' end of the third exon [19]. *In vitro* splicing was carried out for 2 h in the nuclear extract from HeLa cells [20], as described previously [21]. Products of the reaction were purified by phenol extraction followed by ethanol precipitation and analyzed by electrophoresis on 8% polyacrylamide/7 M urea sequencing gel. Resulting bands were detected by autoradiography on X-ray films.

Data processing

All autoradiograms were captured by DAGE MTI CCD72 video camera and the images were processed using NIH Image 1.43 and McDraw Pro 1.0 software. Relative amounts of spliced products either containing or lacking the internal exon were quantitated using NIH Image 1.43 software. The results were adjusted to account for different number of radioactive cytidine nucleotides in each product and expre- s sed as a percent of the inclusion product to the sum of both products (inclusion *plus* skipping).

RESULTS

Replacement of the majority of 175-nucleotide exon with unrelated sequences leads to its predominant skipping

Figure 1 presents *in vitro* splicing of two related pre-mRNA substrates, Ex2-175 and Ex2-184, consisting of three exons interrupted by two identical introns. Ex2-175 pre-mRNA containing the 175-nucleotide internal exon composed of parts of exon 1 and 2 of the human β-globin gene (Fig. 2B) is spliced *in vitro* exclusively by a correct pathway leading to inclusion of all three exons into the final product (Fig. 1, lane 1, see also [5]). When the majority of the internal exon was replaced by a heterologous sequence (*Nco*I-*Bsm*I restriction fragment of URA3 gene of yeast *Saccharomyces cerevisiae* inserted in direct orientation, [18]) splicing of resultant Ex2-184 pre-mRNA yielded at approximately 70% frequency a product composed of only two flanking exons (Ex2-184, Fig. 1, lane 2). Accumulation of the skipping product lacking the modified, 184-nucleotide exon (Fig. 2A), is accompanied by the appearance of

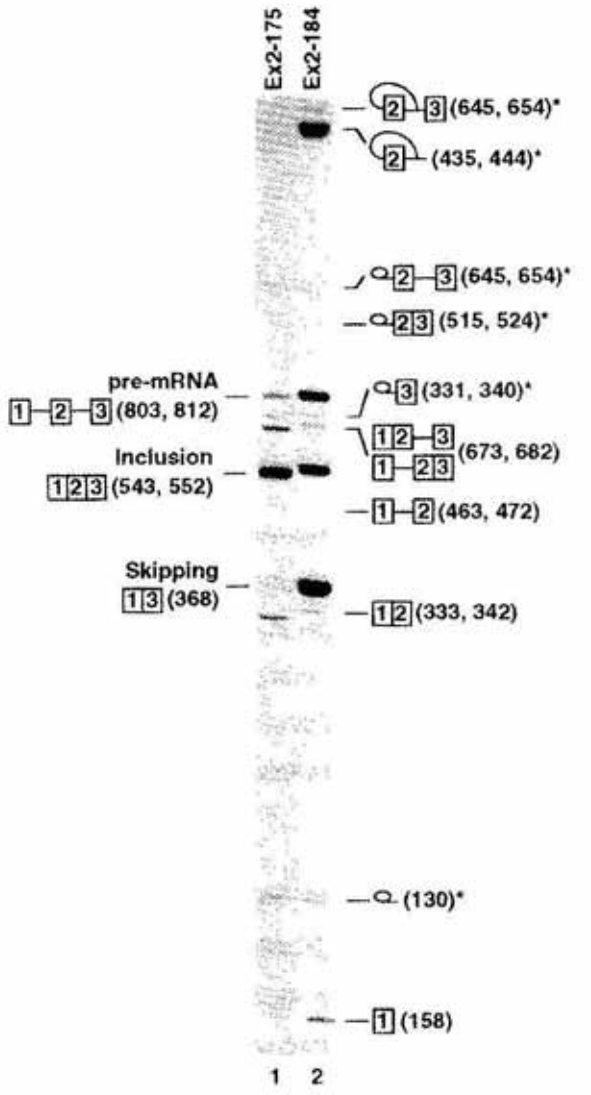
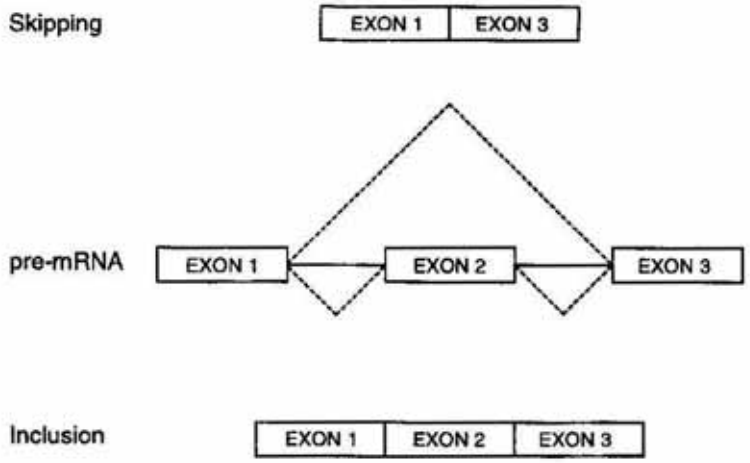


Fig. 1. Internal 184-nucleotide exon is predominantly skipped during in vitro splicing of Ex2-184 pre-mRNA.
Two identical, except for the sequence of the internal exon, pre-mRNA substrates, Ex2-175 (lane 1) and Ex2-184 (lane 2), were compared for *in vitro* splicing patterns. Transcripts were spliced in HeLa nuclear extract for 2 h and reaction products were analyzed on 8% polyacrylamide sequencing gel and visualized by autoradiography (see Materials and Methods). Splicing of Ex2-175 proceeds exclusively *via* inclusion of the internal, 175-nucleotide exon, whereas splicing of Ex2-184 pre-mRNA yields approximately 70% of skipping product, lacking the internal, 184-nucleotide exon. Structures of splicing intermediates and final products are shown and their sizes for Ex2-175 (on the left numbers) and Ex2-184 pre-mRNA (on the right numbers) are given in nucleotides. Boxes represent exon sequences, lines — intron sequences. RNA species exhibiting unusual mobility due to the lariat structure are indicated by asterisks. The same designations are used in all subsequent figures.

A



B

175-nucleotide internal exon:

GCUGCUGGUG (10) GUCUACCCUU (20) GGACAGCUUG (30) CUUACAUUUG
(40) CUUCUGACAC (50) AACUGUGUUC (60) ACUAGCAACC (70)
UCAAACAGAC (80) ACCAUGGUGC (90) ACCUGACUCC (100) UGAGGAGAAG
(110) UCUGCCGUUA (120) CUGCCCUGUG (130) GGGCAAGGUG (140)
AACGUGGAUG (150) AAGUUGGUGG (160) UGAGGCCCUG (170) GGCAG (175)

184-nucleotide internal exon

GCUGCUGGUG (10) GUGCCAUGGA (20) GGGCACAGUU (30) AAGCCGCUAA
(40) AGGCAUUAUC (50) CGCCAAGUAC (60) AAUUUUUUAC (70) UCUUCGAAGA
(80) CAGAAAAUUU (90) GCUGACAUUG (100) GUAAUACAGU (110)
CAAAUUGCAG (120) UACUCUGCGG (130) GUGUAUACAG (140) AAUAGCAGAA
(150) UGGGCAGACA 160) UUACGAAUGG (170) CAGGCCCUGG (180)GCAG (184)

Fig. 2. A, Structure of pre-mRNA substrate consisting of three exons interrupted by two introns, and possible pathways of its splicing leading to either exon inclusion or exon skipping. B, Nucleotide sequence of internal exons in Ex2-175 and Ex-184 pre-mRNAs containing 175 and 184 nucleotides, respectively. The 175-nucleotide exon is composed of sequences derived from exons 1 and 2 of human β-globin gene.
The 184-nucleotide exon was constructed by replacing the majority of the 175-nucleotide exon with a sequence from yeast URA 3 gene. Sequences of globin origin left in the 184-nucleotide exon and located on 5′ (12 nucleotides) and 3′ ends (13 nucleotides) of the exon are underlined.

large lariats, containing both introns and excluded exon (seen at the top of lane 2, Fig. 1). Splicing of Ex2-184 pre-mRNA with the frequency of only 30% followed the correct pathway, resulting in inclusion of all exons into the final product.

Since sequence modification involved only central part of the internal exon leaving unchanged 12 and 13 nucleotides from the 5′ and 3′ ends of the exon, respectively (Fig. 2 B), skipping of 184-nucleotide exon could not be explained by direct disruption or weakening of the flanking splice sites. Rather it appeared that poor recognition of 184-nucleotide exon resulted from suboptimal character of sequences provided within the yeast insert or due to removal of exon recognition sequences residing in the 175-nucleotide globin exon.

Strong splice sites facilitate efficient inclusion of suboptimal exon

The experiments described below demonstrate that more efficient inclusion of 184-nucleotide exon can be achieved by relatively minor

changes within the flanking splice sites, which improve their matching to the consensus sequences.

Creation of a long uninterrupted polypyrimidine (polyY) tract by substitutions of individual purines with pyrimidines [18, 22] upstream from the internal 184-nucleotide exon leads to substantial increase in the frequency of exon inclusion. The effect depends on the number of substitutions made, as illustrated by more efficient inclusion observed for Ex2-184/Y3 pre-mRNA (3 substitutions, and 57% of inclusion, Fig. 3, lane 3) than for Ex2-184/Y2 clone (two substitutions and only 48%, Fig. 3, lane 2). Comparable stimulation of exon inclusion (51% of total splicing) is promoted by a mutation which improves the upstream branch point sequence to match the yeast consensus UAACUAC sequence (Ex2-184/BP, Fig. 3, lane 4). The effect of individual modifications is additive. This perfect branch point sequence in conjunction with Y3 mutations results in almost complete (85%) *in vitro* inclusion of 184-nucleotide exon (Ex2-184/BP/Y3, Fig. 3, lane 5).

184-Nucleotide exon seems to lack negative sequence elements

To test the possibility that the 184-nucleotide exon contains potential sequences interfering with the correct splicing pathway, we extensively mutated the yeast portion of the exon. The purpose of this strategy was to disrupt negatively acting element(s), that might have been provided during exon modification, and in consequence to restore efficient inclusion pathway. Similar analysis was previously applied to study the mechanism of alternative splicing in pre-mRNA coding for tissue-specific human leukocyte common antigen and suggested the existence of such elements in several skipped exons [23, 24]. We introduced clusters of mutations designated GC into several regions of 184-nucleotide exon, in majority of cases by replacing thymidines with cytidines and adenosines with guanosines. Furthermore, to increase the overall extent of sequence modification, individual mutation clusters were combined in 184-nucleotide exon generating a series of double and triple mutants (Table 1).

Analysis of results in Fig. 4 revealed that none of GC2, GC3, GC4 and GC5 clusters, tested separately or in combination with each other, led to enhanced exon inclusion. In addition to

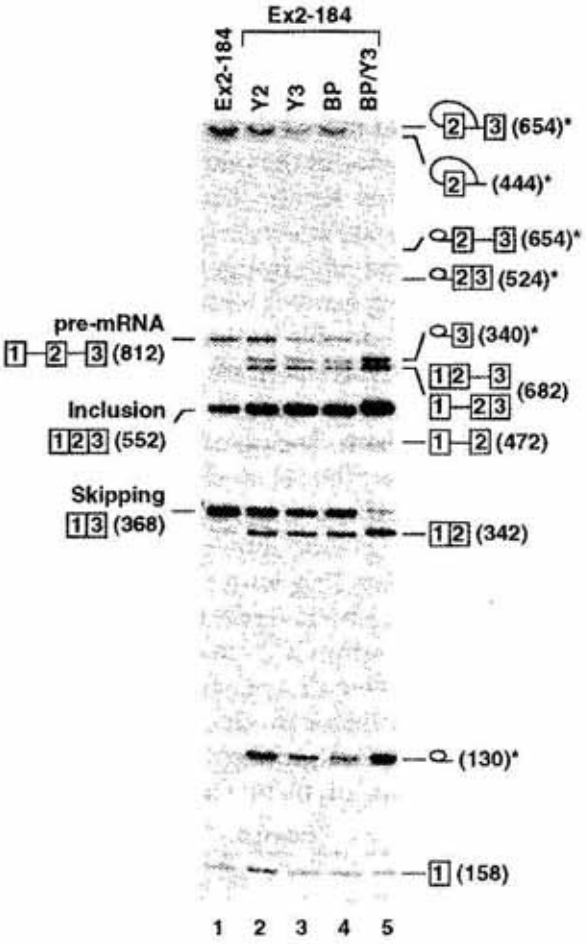


Fig. 3. Improvement of splicing elements flanking the 184-nucleotide exon results in more efficient inclusion.
In vitro splicing of the parental Ex2-184 pre-mRNA (lane 1) and its derivatives with improved splicing elements in the upstream intron (lanes 2–5) as indicated at the top of each lane. Transversion of individual purines within the polypyrimidine tract (Ex2-184/Y2 and Y3) or creation of the consensus branch point sequence (Ex2-184/BP) in front of the internal exon switch the splicing pathway from predominant skipping to predominant inclusion. Mutations within both splicing elements, when combined, act collectively and almost completely prevent exon skippping (Ex2-184/BP/Y3).

4 single and 3 double mutants presented in this Figure, two other single and a number of triple mutants (see Table 1) were analysed and again showed no stimulation of exon inclusion. In all cases tested, with the exception of GC5 and GC3/5 mutants (see below), the overall efficiency of inclusion was maintained at approximately 30%. The most surprising was a neutral behavior of GC2 mutation, replacing six uridines with six cytidines in the middle portion of the internal exon (nucleotides 63–68). While *in vivo* GC2 mutation resulted in approximately 7-fold increase in exon inclusion,

in vitro this alteration did not significantly improve the recognition of the internal exon, maintaining inclusion at normal level. This result suggests that more efficient accumulation of mRNA that includes the internal exon under *in vivo* conditions results from mechanisms other than splice site selection (see Discussion). Note that the relatively stronger intensity of the band representing exon observed in the case of GC2 and GC2/4 double mutants (Fig. 4, lanes 1 and 5, respectively) is misleading and does not result from improved recognition of the 184-nucleotide exon. The quantitative analysis indicates that this effect is mostly due to higher frequency of cytidines in the inclusion product and thus higher specific activity achieved during pre-mRNA labeling with radioactive CTP. Interestingly, GC5 both separately (Fig. 4, lane 4) and together with GC3 cluster (lane 7), resulted in a complete disappearance of 552 nucleotides long inclusion product. Instead, a new spliced product, comigrating with 472-nucleotide intermediate of inclusion pathway was generated. Based on the analysis of products and intermediates, this unusual splicing pattern can be almost certainly attributed to activation of the cryptic 5′ splice site located within the internal exon. Perhaps the cluster of G and C residues created by GC5 mutation spanning nucleotides 139–150 of the 184-nucleotide exon (see Table 1 and Fig. 2B) exerts strong inhibitory effect on the correct 5′ splice site thereby activating the cryptic element located 84 nucleotides upstream.

We conclude that mutational scanning of the 184-nucleotide exon did not reveal any major negative signals responsible for its efficient skipping *in vitro*. However, since only selected parts of the exon were surveyed by introducing clusters of mutations, the possibility could not be excluded that some negative elements escaped our detection. To address this possibility, we analysed *in vitro* splicing of Ex2-184/Rev pre-mRNA in which yeast portion of the internal exon was inserted in the reversed orientation. This manipulation was expected to generate yet another internal exon, with nucleotide sequence different from that in Ex2-184 pre-mRNA. Fig. 5, lane 2 demonstrates that splicing of Ex2-184/Rev pre-mRNA leads to even less efficient inclusion (14%) than splicing of Ex2-184 pre-mRNA (lane 1). These experiments provided additional indication that skipping of the internal exon in Ex2-184 and Ex2-184/Rev pre-mRNAs must result from the absence of a positive splicing signal normally residing in the globin exon rather than from the presence of a negative signal(s) within the yeast sequence.

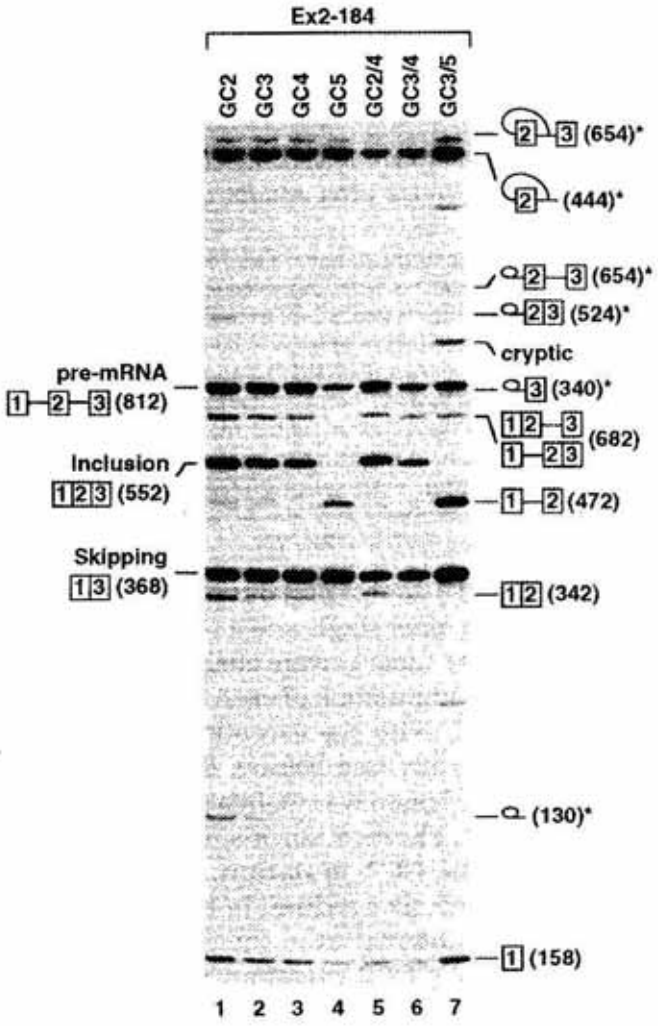


Fig. 4. Mutational scanning of the 184-nucleotide exon does not reveal any negatively acting sequence elements.

In order to identify sequences responsible for exon skipping several block mutations were introduced in various regions of the internal exon and analysed both individually (lanes 1–4) and in combination (lanes 5–7). *In vitro* splicing shows no increase in exon inclusion for any mutant pre-mRNAs tested. Utilization of a cryptic splice site upon GC5 mutation (lanes 4 and 7) is indicated by disappearance of the correct inclusion product and accumulation of a new product, which comigrates with 472-nucleotide intermediate consisting of exons 1 and 2 and the upstream intron.

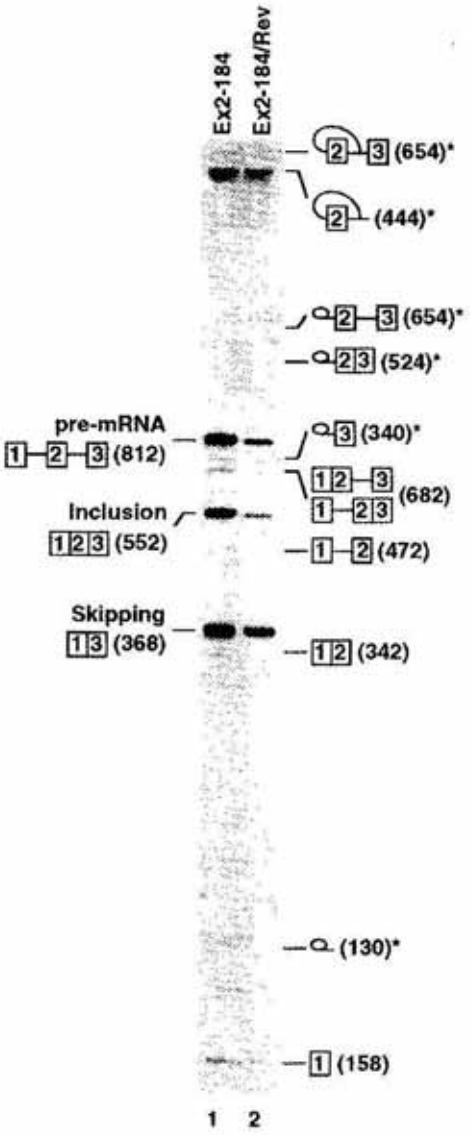


Fig. 5. Insertion of the yeast sequence into the internal exon in reversed orientation does not prevent exon skipping.
In vitro splicing of Ex2-184/Rev pre-mRNA (lane 2) containing the yeast sequence in reversed orientation yields only 14% of inclusion product, approximately 2-fold less than splicing of Ex2-184 pre-mRNA, containing yeast sequence in direct orientation (lane 1).

Deletion analysis of 175-nucleotide exon reveals a region containing a positive splicing element

We began the search for positive splicing elements within the fully globin-derived 175-nucleotide exon with the sequence analysis of its truncated version in Ex2-51 pre-mRNA. The 51-nucleotide internal exon in this pre-mRNA, in spite of the removal of 124 nucleotides from the 175-nucleotide exon, retains full information required to undergo efficient inclusion both *in vivo* and *in vitro* [22]. Sequence analysis revealed that Ex2-51 construct was created by a deletion, which started near the center of the exon and terminated 6 nucleotides from its 5′ and 45 nucleotides from its 3′ end. This observation allowed initial mapping of an apparent positive signal to 32 nucleotides in the 3′ region of 175-nucleotide exon (45 nucleotides *minus* the 13 nucleotides of the most downstream globin sequence already present in 184-nucleotide exon, Fig. 2B).

If the 32-nucleotide globin sequence contains a positive splicing element(s) allowing efficient recognition of internal exons, deletion of this sequence from the 175-nucleotide exon should result in its predominant skipping. Indeed, two deletion mutants of Ex2-175, i.e. Ex2-54 and Ex2-101, which are missing either a large portion or the entire 32-nucleotide sequence in 54- and 101-nucleotide middle exon, respectively (Fig. 6A), are spliced almost exclusively by exon skipping (Fig. 6B, lane 1 and C, lane 1). Note that their counterparts, Ex2-51 and Ex2-107, retaining the sequence and maintaining the internal exon within the same size limit (Fig. 6A), were used in parallel and show very little (Ex2-51) or no skipping (Ex2-107) (Fig. 6B, lane 2 and C, lane 2). Side-by-side analysis of pre-mRNAs containing internal exon of similar length with or without the splicing element allowed to exclude possible size effects [22, 25–27] and facilitated characterization of splicing intermediates and final products.

The internal, 54-nucleotide exon retained 35 nucleotides from the 5′ end and 19 nucleotides from the 3′ end of the 175-nucleotide exon. Based on this observation, the region likely to contain the positively acting element was narrowed down to 26 nucleotides, encompassing positions 131–156 in 175-nucleotide exon.

The 26-nucleotide globin sequence acts as a positive splicing element in the background of 184-nucleotide exon

In order to determine whether a positive splicing element identified in 175-nucleotide exon can act in a different sequence context, we started a process of stepwise replacement of the corresponding sequences in 184-nucleotide exon with sequences derived from the 3′ region of the globin exon.

In the first step a GTGGTG sequence of 175-nucleotide exon was used to substitute for AAUGGC sequence of 184-nucleotide exon.

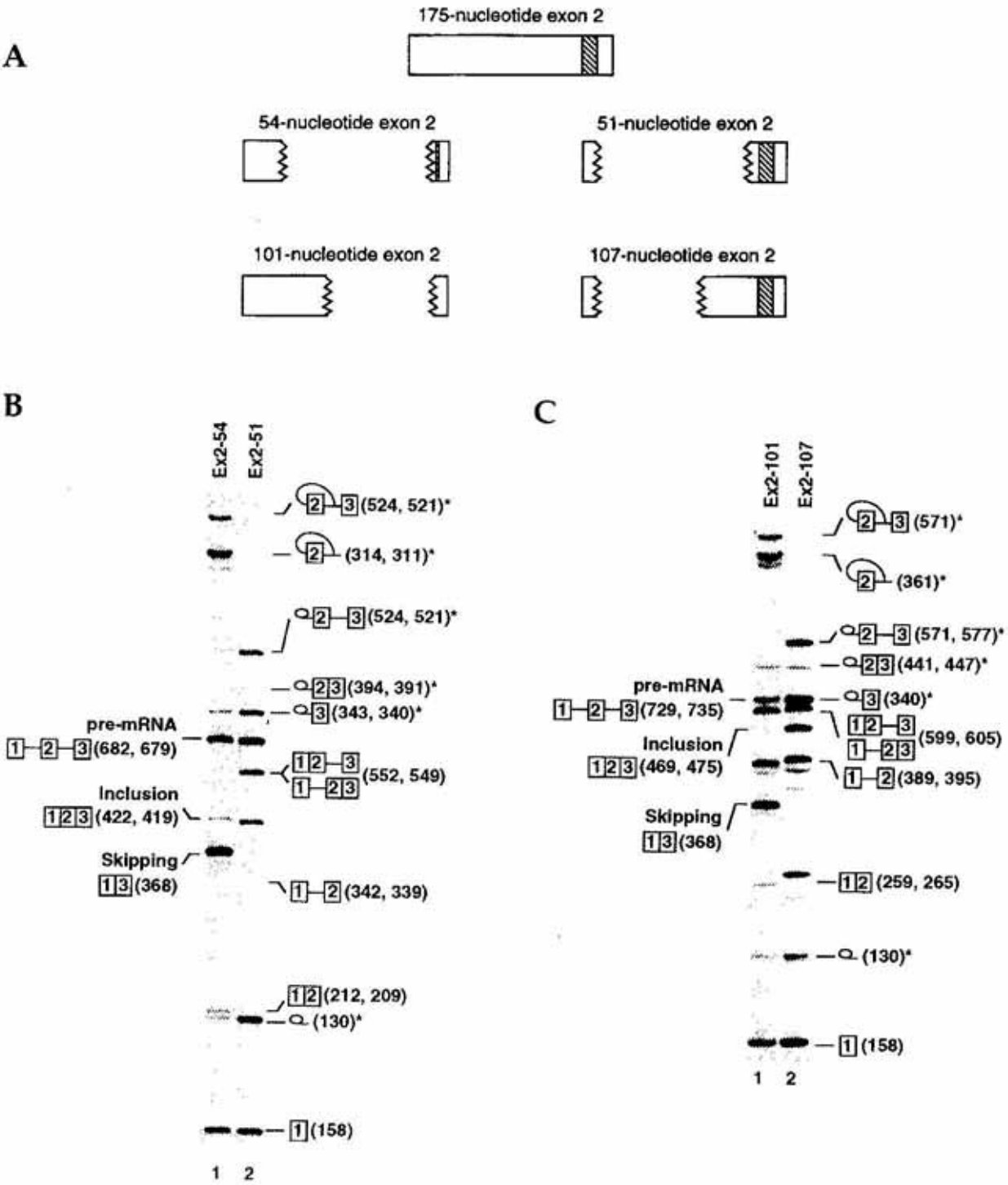


Fig. 6. A 26-nucleotide element is responsible for efficient recognition and inclusion of the 175-nucleotide globin exon.

A, Two pairs of deletion mutants of Ex2-175, lacking various portions of the internal 175-nucleotide exon were constructed and analysed for *in vitro* splicing. The first pair includes Ex2-54 (54-nucleotide internal exon, deletion of nucleotides 36 to 156) and Ex2-51 (51-nucleotide internal exon, deletion of nucleotides 7 to 130), the second pair — Ex2-101 (101-nucleotide internal exon, deletion of nucleotides 88 to 161) and Ex2-107 (107-nucleotide exon, deletion of nucleotides 14–81). In each pair, internal exons have similar length and contain or do not contain the 32-nucleotide sequence represented by hatched area. Jagged lines indicate end points of the deletions in the 175-nucleotide exon. B, *In vitro* splicing of Ex2-54 and Ex2-51 pre-mRNAs. C, *In vitro* splicing of Ex2-101 and Ex2-107 pre-mRNAs. Accumulation of large amounts of the skipping product and the presence of slowly migrating lariats at the top of the gel during *in vitro* splicing of Ex2-54 and Ex2-101 pre-mRNAs strongly indicates the importance of the 26-nucleotide element in exon recognition.

This alteration extended of the already existing globin sequence from 13 nucleotides to the total of 19 nucleotides. As expected from *in vitro* splicing of the Ex2-54 pre-mRNA, containing the same number of globin nucleotides at the 3' end, splicing of the resultant Ex2-184/19 pre-mRNA shows no indication of improved exon inclusion (Fig. 7, lane 2). Extension of the globin sequence to the total of 32 nucleotides in Ex2-184/32 pre-mRNA, achieved by replacement of additional 13-nucleotide region (i.e. half of the 26 nucleotides likely to contain the positive element) leads to a strong shift in splicing pathway increasing exon inclusion to almost 50% (Fig. 7, lane 3). Further stimulation of exon inclusion, to approximately 70%, was observed when the remaining 13-nucleotide region from 175-nucleotide exon was substituted for corresponding yeast sequence (Ex2-184/45, Fig. 7, lane 4). This modification extended the globin sequence in yeast-globin exon to 45 nucleotides and completed the restoration of the downstream portion of 51-nucleotide exon. Interestingly, when the upstream 13-nucleotide half was inserted in Ex2-184/19 mutant to create Ex2-184/32*, leaving a 13-nucleotide spacer of the yeast sequence, the effect was even stronger than that observed for Ex2-184/32 (Fig. 7, lane 5). Clearly, each of the two 13-nucleotide globin fragments contains positively acting elements leading to efficient stimulation of exon inclusion. Introduction of the GC2 mutation into Ex2-184/45 background does not increase exon inclusion (Fig. 7, lane 6). This further indicates the essential role for the globin element derived from 175-nucleotide exon and confirms lack of the effect of 6 consecutive C's on exon inclusion *in vitro*.

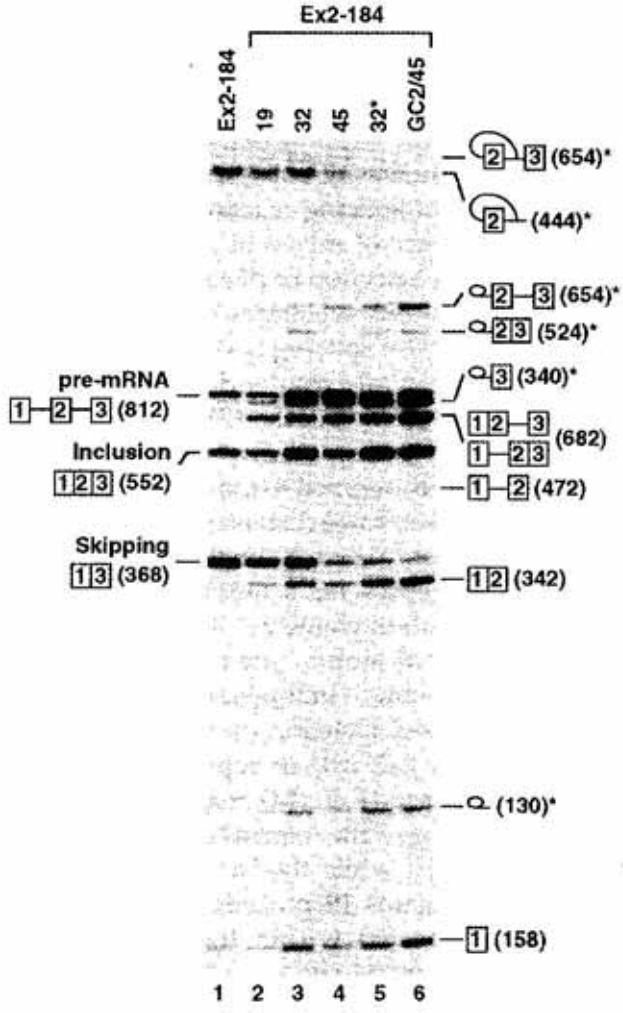


Fig. 7. Step-wise reconstitution of the 175-nucleotide exon in the background of the 184-nucleotide exon confirms a positive role of the globin element in exon recognition.
In vitro splicing of the parental Ex2-184 pre-mRNA (lane 1) and its derivatives containing various substitutions in the downstream portion of the 184-nucleotide exon (lanes 2–5). Sequence in the downstream region of 184-nucleotide exon was altered by progressive substitution with sequence from corresponding region of 175-nucleotide exon. Since both internal exons share identical sequence of 13 nucleotides at the most downstream region, replacement of immediately adjacent 6 nucleotides creates the Ex2-184/19 mutant (lane 2). Replacement of additional 13 or 26 consecutive nucleotides result in Ex2-184/32 (lane 3) and Ex2-184/45 (lane 4) mutants, respectively. Ex2-184/32* (lane 5) represents a mutant in which replacement of the most downstream 6 nucleotides was combined with replacement of the upstream 13 nucleotides, leaving the 13-nucleotide yeast spacer. Lane 6, *in vitro* splicing of Ex2-184/GC2/45 pre-mRNA containing GC2 mutation (6 consecutive C's in positions 63–68 of the 184-nucleotide exon) in Ex2-184/45 background.

26-Nucleotide globin element acts as a purine-rich exon recognition sequence

What sequences of 26-nucleotide element function as a positive splicing signal? Since cryptic splice sites were often implicated in splice site selection [28–30] we considered a possibility that sequences AGGUGAAC and AAGUUGGU resembling 5' splice sites (invariant nucleotides are underlined, see Table 1 and Fig. 2B) are required for activity of the upstream and downstream halves of the globin exon recognition element, respectively. However, re-

placement of invariant U's with C's in Ex2-184/45 background did not decrease its stimulatory effect on exon inclusion and in fact increased inclusion to over 90% (Ex2-184/45/C, Fig. 8, lane 2), consistent with some reports on antagonistic effects of juxtaposed 5' splice sites [31, 32].

Recently, a number of independent studies demonstrated the importance of purine-rich exon sequences in splicing of several natural [5, 12, 13, 27, 33] and artificial pre-mRNAs [16, 17]. To test the possibility that these sequences can indeed efficiently facilitate inclusion of the internal exon in our model pre-mRNA substrates, we created a polypurine tract encompassing positions 143–160 of the 184-nucleotide exon in Ex2-184 pre-mRNA. This tract was thus placed in the same distance from the 3' end of 184-nucleotide exon as was the 26-nucleotide sequence in the 175-nucleotide exon. Specific positions and types of mutations (mostly pyrimidine to purine transversions) were selected in such a way that mutated region, in addition to overall high purine content contained several short adenosine and guanosine stretches found in many natural exon recognition elements [27, 33]. As seen in Fig. 8, lane 3 (Ex2-184/AG pre-mRNA), this modification resulted in over 80% of exon inclusion indicating that purine rich sequences are functional in the context of 184-nucleotide yeast-globin exon.

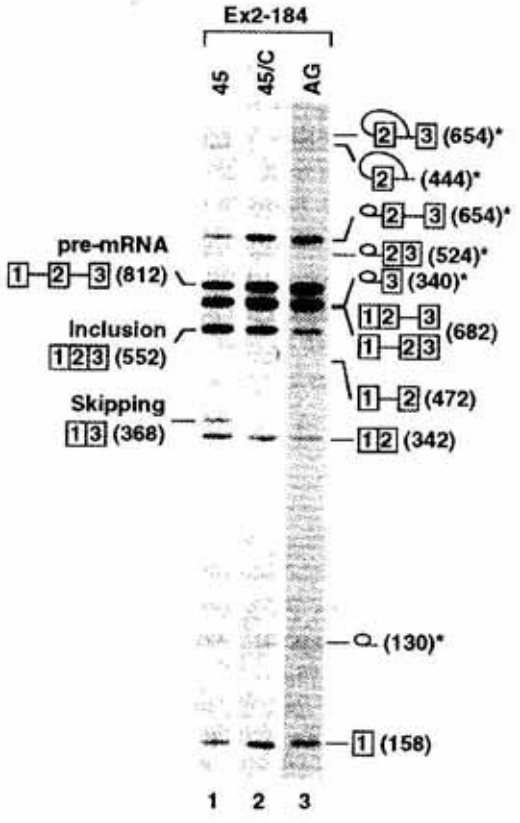


Fig. 8. The 26-nucleotide sequence seems to represent a purine-rich exon recognition element.
In vitro splicing of Ex2-184/45/C pre-mRNA (lane 2) in which both cryptic 5' splice sites located in the 26-nucleotide globin region were disrupted by converting GU's into GC's. *In vitro* splicing of the parental Ex2-184/45 pre-mRNA is shown for comparison (lane 1). Disruption of cryptic 5' splice sites do not inactivate the 26-nucleotide element. Lane 3, *in vitro* splicing of Ex2-184/AG pre-mRNA containing a long polypurine tract in the downstream portion of the 184-nucleotide exon. Creation of an artificial polypurine sequence in the 184-nucleotide exon, spanning nucleotides 140–163, results in a very efficient inclusion of otherwise poorly recognized yeast-globin exon.

DISCUSSION

We have previously reported the use of an artificial Ex2-175 construct, derived from human β-globin gene in studies of various aspects of pre-mRNA splicing [18, 21, 22, 34]. The three-exonic Ex2-175 pre-mRNA undergoes an efficient splicing *in vitro* and *in vivo* resulting in accurate removal of both introns and ligation of all three exons. Skipping of the internal 175-nucleotide exon was not observed under normal conditions of splicing reaction or remarkably, in the presence of excess of hnRNP A1 factor, favoring the selection of distal 3' splice site in other pre-mRNA substrates [35]. In this work, by using *in vitro* splicing system we confirmed our previous *in vivo* results, demonstrating that efficient recognition of the 175-nucleotide exon is mediated by a sequence, residing in this exon. The sequence, mapped to the downstream portion of the exon some distance from the 5' splice site, does not exceed 26 nucleotides in length and likely represents a family of purine-rich exon recognition elements. It contains at least two functional motifs, one in each 13-nucleotide half, capable of independent stimulation of exon inclusion. Deletion of all 26 nucleotides (Ex2-54 and Ex2-101) or replacing them with yeast sequence (Ex2-184) results in predominant skipping of the internal exon.

Surprisingly, while the 26-nucleotide globin element contains 19 purines (73%), the yeast sequence located within the same distance from the 5' splice site in Ex2-184 pre-mRNA contains only one purine less (69%). This com-

parison confirms other observations [13, 33] that purine richness in itself is not a sufficient explanation for the activity of the globin element in splice site selection. Other features of this element may be important, for example presence of not equally deleterious U *versus* C residues, alternation of A and G residues [33], the length of uninterrupted stretches of purines or their specific sequence configurations. Interestingly, among various synthetic polypurine sequences tested for stimulatory effect on *in vitro* splicing, AAG represented once in each of 13-nucleotide halves and not found in corresponding yeast sequence, seems to be the most active [33]. Accordingly, generation of two such trinucleotides in the yeast sequence by individual point mutations was sufficient to convert weakly recognized exon into predominantly included one (Ex2-184/AG).

The activity of 26-nucleotide exon recognition sequence is not restricted to 175- and 184-nucleotide exons. With similar efficiency this element facilitates inclusion of the internal exon, containing yeast sequence in the reversed orientation (Ex2-184/Rev, Wozniak, A. & Kole, R., unpublished results). This, together with the stimulatory effect on 51- and 107-nucleotide exon, indicates the universal role of the element and its ability to cooperate with exons containing various nucleotide sequences and composition.

The fact that exon recognition elements are functional regardless of their specific location in exons and consist of rather weakly defined and short sequences [5, 12, 27] may explain why they escaped earlier detection by comparison of known exon sequences. The same properties of exon recognition elements seem to explain the lack of their major interference with the primary role of exons in translation. It may also be significant that many known elements are located within 3′ flanking exons, hence code for the most carboxy-terminal amino acids or resides outside stop codons and do not overlap with the information for protein synthesis.

The 184-nucleotide internal exon created by extensive sequence modification of 175-nucleotide exon is predominantly ignored by splicing machinery *in vitro*. Experiments with improved splice sites adjacent to this exon presented here and in the previous report [18] indicate that both splice sites of β-globin origin are not sufficiently strong to direct inclusion of internal exons, lacking the positively acting element. Consistently, most of the exon recognition elements identified so far, including those in doublesex [9], cardiac troponin T [27, 36] and avian sarcoma virus *env* pre-mRNAs [37, 38], seem to act by cooperating with functionally weak splice sites. These results provide additional support for exon definition model [39] in which splicing factors recognizing the splice sites communicate with each other through the network of interactions across the exons [14, 40]. The existence of exon recognition elements, likely targets for a family of SR proteins [12, 14–16, 41, 42] would result in stabilization of such interactions and thereby allow efficient definition of exons as primary units of spliceosome assembly [39].

A direct comparison between *in vitro* splicing of Ex2-184 pre-mRNA and previously reported *in vivo* splicing of the same construct [18] indicates 30% and 5% of exon inclusion, respectively. This is in contrast to several earlier observations which consistently demonstrated more efficient recognition of the internal exon *in vivo* [22, 34, 43]. This contradiction can be explained by lower cytoplasmic stability of mRNA containing the 184-nucleotide exon of yeast origin as compared to skipped mRNA which leads to modification of splicing-determined ratio of both products. Moreover, given that amounts of spliced products accumulated during *in vitro* splicing accurately reflect relative frequencies of exon skipping and exon inclusion, we conclude that GC2 mutation does not play a role in splice site selection. We suggest that improved ratio of inclusion to skipping previously observed for this mutation *in vivo* resulted from stabilizing effect of six consecutive C's on mRNA containing the internal exon and its increased half life time during 48 h of transient expression. Experimental support for this interpretation will require *in vivo* expression of corresponding cDNA constructs (which do not involve splicing) and direct determination of the steady-state level of mRNAs that include or exclude the internal exon [27, 36]. Interestingly, cytidine-rich sequences found in α2-globin gene were recently shown to participate in regulation of mRNA stability perhaps *via* a mechanism which involves association with polycytosine binding proteins to prevent rapid degradation at the 3′ untranslated region [44].

Identification of the positively acting purine-rich element in our model pre-mRNA raised the possibility that previously reported skipping of 23- and 33-nucleotide internal exons [22] was caused by deletion of this element rather than by general shortening of the exon. This possibility was addressed by Xu *et al.* [27], who using model constructs identical to ours inserted strong polypurine sequences within several internal mini-exons and measured the frequency of skipping in such constructs *in vivo*. It was found [27] that purine rich sequences, facilitating exon recognition in a variety of longer exons, when placed within mini-exons, worked to a very limited extent or did not work at all. These studies, together with increasing number of recent reports, provided a strong experimental evidence indicating that in conjunction with other features of pre-mRNAs, both exon length and exon sequence play an important role in splice site selection.

We thank Elizabeth Smith for excellent technical assistance.

REFERENCES

1. Madhani, H.D. & Guthrie, C. (1994) Dynamic RNA-RNA interactions in the spliceosome. *Annu. Rev. Genet.* **28**, 1–26.
2. Moore, M.J., Query, C.C. & Sharp, P.A. (1993) Splicing of precursors to mRNAs by the spliceosome; in *RNA World* (Gesteland, R.F. & Atkins, J.F., eds.) pp. 303–358, Cold Spring Harbor Laboratory Press.
3. Penotti, F.E. (1991) Human pre-mRNA splicing signals. *J. Theor. Biol.* **150**, 385–420.
4. Senapathy, P., Shapiro, M.B. & Harris, N.L. (1990) Splice junctions, branch point sites, and exons: sequence statistics, identification, and applications to genome project. *Methods Enzymol.* **183**, 252–278.
5. Watakabe, A., Tanaka, K. & Shimura, Y. (1993) The role of exon sequences in splice site selection. *Genes & Dev.* **7**, 407–418.
6. Burtis, K.C. & Baker, B.C. (1989) *Drosophila* doublesex gene controls somatic sexual differen- tiation by producing alternatively spliced mRNAs encoding related sex-specific polypeptides. *Cell* **56**, 997–1010.
7. Hedley, M.L. & Maniatis, T. (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.
8. Nagoshi, R.N. & Baker, B.S. (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.
9. Hoshijima, K., Inoue, K., Higuchi, I., Sakamoto, H. & Shimura, Y. (1991) Control of doublesex alternative splicing by transformer and transformer-2 in *Drosophila*. *Science* **252**, 833–836.
10. Inoue, K., Hoshijima, K., Higuchi, I., Sakamoto, H. & Shimura, Y. (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. U.S.A.* **89**, 8092–8096.
11. Tian, M. & Maniatis, T. (1992) Positive control of pre-mRNA splicing *in vitro*. *Science* **256**, 237–240.
12. Dirksen, W.P., Hampson, R.K., Sun, Q. & Rottman, F.M. (1994) A purine-rich exon sequence enhances alternative splicing of bovine growth hormone pre-mRNA. *J. Biol. Chem.* **269**, 6431–6436.
13. Lavigueur, A., La Branche, H., Kornblihtt, A.R. & Chabot, B. (1993) A splicing enhancer in the human fibronectin alternate ED1 exon interacts with SR proteins and stimulates U2 snRNP binding. *Genes & Dev.* **7**, 2405–2417.
14. Staknis, D. & Reed, R. (1994) SR proteins promote the first specific recognition of pre-mRNA and are present together with the U1 small nuclear ribonucleoprotein particle in a general splicing enhancer complex. *Mol. Cell. Biol.* **14**, 7670–7682.
15. Sun, Q., Mayeda, A., Hampson, R.H., Krainer, A.R. & Rottman, F.M. (1993) General splicing factor SF2/ASF promotes alternative splicing by binding to exonic splicing enhancer. *Genes & Dev.* **7**, 2598–2608.
16. Tacke, R. & Manley, J.L. (1995) The human splicing factors ASF/SF2 and SC35 possess distinc, functionally significant RNA binding specificities. *EMBO J.* **14**, 3540–3551.
17. Tian, H. & Kole, R. (1995) Selection of novel exon recognition elements from a pool of random sequences. *Mol. Cell. Biol.* **15**, 6291–6298.
18. Dominski, Z. & Kole, R. (1994) Identification of exon sequences involved in splice site selection. *J. Biol. Chem.* **269**, 23590–23596.
19. Konarska, M.M., Padgett, R.A. & Sharp, P.A. (1984) Recognition of cap structure in splicing *in vitro* of mRNA precursors. *Cell* **38**, 731–736.

20. Krainer, A.R., Maniatis, T., Ruskin, B. & Green, M.R. (1984) Normal and mutant human β-globin pre-mRNAs are faithfully and efficiently spliced *in vitro*. *Cell* **36**, 993–1005.
21. Dominski, Z. & Kole, R. (1993) Restoration of correct splicing in thalassemic pre-mRNA by antisense oligonucleotides. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 8673–8677.
22. Dominski, Z. & Kole, R. (1991) Selection of splice sites in pre-mRNAs with short internal exons. *Mol. Cell. Biol.* **11**, 6075–6083.
23. Streuli, M. & Saito, H. (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.
24. Tsai, A.Y.M., Streuli, M. & Saito, H. (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.
25. 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.
26. Sterner, D. & Berget, S.M. (1993) *In vivo* recognition of a vertebrate mini-exon as an exon-intron-exon unit. *Mol. Cell. Biol.* **13**, 2677–2687.
27. Xu, R., Teng, J. & Cooper, T. (1993) The cardiac troponin T alternative exon contains a novel purine-rich positive splicing element. *Mol. Cell. Biol.* **13**, 3660–3674.
28. Ligtenberg, M.J.L., Gennissen, A.M.C., Vos, H.L. & Hilkens, J. (1991) A single nucleotide polymorphism in an exon dicatates allele dependent differential splicing of episialin mRNA. *Nucleic Acids Res.* **19**, 297–301.
29. Steingrimsdottir, H., Rowley, G., Dorado, G., Cole, J. & Lehmann, A.R. (1992) Mutations which alter splicing in the human hypoxanthine-guanosine phosphoribosyltransferase gene. *Nucleic Acids Res.* **20**, 1201–1208.
30. Zhuang, Y. & Weiner, A.M. (1990) The conserved dinucleotide AG of the 3′ splice site may be recognized twice during *in vitro* splicing of mammalian mRNA precursors. *Gene* **90**, 263–269.
31. Nelson, K.K. & Green, M.R. (1988) Splice site selection and ribonucleoprotein complex assembly during *in vitro* pre-mRNA splicing. *Genes & Dev.* **2**, 319–329.
32. Siebel, P.G., Fresco, L.D. & Rio, D.C. (1992) The mechanism of somatic inhibition of *Drosophila* P-element pre-mRNA splicing: multiprotein complexes at an exon pseudo-5′ splice site control U1 snRNP binding. *Genes & Dev.* **2**, 1258–1267.
33. Tanaka, K., Watakabe, A. & Shimura, Y. (1994) Polypurine sequences within a downstream exon function as a splicing enhancer. *Mol. Cell. Biol.* **14**, 1347–1354.
34. Dominski, Z. & Kole, R. (1992) Cooperation of pre-mRNA sequence elements in splice site selection. *Mol. Cell. Biol.* **12**, 2108–2114.
35. Mayeda, A., Helfman, D.M. & Krainer, A.R. (1993) Modulation of exon skipping and inclusion by heterogenous nuclear ribonucleoprotein A1 and pre-mRNA splicing factor SF2/ASF. *Mol. Cell. Biol.* **13**, 2993–3001.
36. Cooper, T.A. & Ordahl, C.P. (1989) Nucleotide substitutions within the cardiac troponin T alternative exon disrupt pre-mRNA alternative splicing. *Nucleic Acids Res.* **17**, 7905–7921.
37. Fu, X.-D., Katz, R.A., Skalka, A.M. & Maniatis, T. (1991) The role of branchpoint and 3′-exon sequences in the control of balanced splicing of avian retrovirus RNA. *Genes & Dev.* **5**, 211–220.
38. Katz, R.A. & Skalka, A.M. (1990) Control of retroviral RNA splicing through maintenance of suboptimal processing signals. *Mol. Cell. Biol.* **10**, 696–704.
39. Robberson, B.L., Cote, G.J. & Berget, S.M. (1990) Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol. Cell. Biol.* **10**, 84–94.
40. Hoffman, B.E. & Grabowski, P.J. (1992) U1 snRNP targets an essential splicing factor, U2AF65, to the 3′ splice site by a network of interactions spanning the exon. *Genes & Dev.* **6**, 2554–2568.
41. Tian, M. & Maniatis, T. (1993) A splicing enhancer complex controls alternative splicing of doublesex pre-mRNA. *Cell* **74**, 105–114.
42. Zahler, A.M., Lane, W.S., Stolk, J.A. & Roth, M.B. (1993) SR proteins: a conserved family of pre-mRNA splicing factors. *Genes & Dev.* **6**, 837–847.
43. Aebi, M., Horning, H., Padgett, R.A., Reiser, J. & Weissmann, C. (1986) Sequence requirement for splicing of higher eukaryotic nuclear pre-mRNA. *Cell* **47**, 555–565.
44. Weiss, I.M. & Liebhaber, S.A. (1995) Erythroid cell-specific mRNA stability elements in the α2-globin 3′ nontranslated region. *Mol. Cell. Biol.* **15**, 2457–2465.