Volume 16 Number 15 1988 Nucleic Acids Research

Selection of cryptic 5' splice sites by group II intron RNAs *in vitro*

Manfred W.Müller, Rudolf J.Schweyen[1] and Carlo Schmelzer

Institut für Genetik und Mikrobiologie, Universität München, Maria-Ward-Str. 1A, 8000 München 19, FRG and [1]Institute für Genetik und Mikrobiologie, Universität Wien, Althanstr. 14, 1090 Wien, Austria

Received May 16, 1988; Revised and Accepted July 4, 1988

ABSTRACT

Recognition of 5' splice points by group I and group II self-splicing introns involves the interaction of exon sequences - directly preceding the 5' splice site - with intronic sequence elements. We show here that the exon binding sequences (EBS) of group II intron al5c can accept various substitutes of the authentic intron binding sites (IBS) provided *in cis* or *in trans*. The efficiency of cleavages at these cryptic 5' splice sites was enhanced by deletion of the authentic IBS2 element. All cryptic 5' cleavage sites studied here were preceded by an IBS1 like sequence; indicating that the IBS1/EBS1 pairing alone is sufficient for proper 5' splice site selection by the intronic EBS element. The results are discussed in terms of minimal requirements for 5' cleavages and position effects of IBS sites relative to the intron.

INTRODUCTION

RNA splicing has been shown to be catalyzed by some introns *in vitro* in the absence of proteins (self-splicing). In these cases, splice site selection, intron excision and exon ligation rely on elements of the RNA structure only (1-7). Self-splicing introns belong to one or to the other of two classes - group I and group II -, which are discriminated on the basis of conserved sequence elements and the capacity to form particular sets of helices, including several distant sequences (8,9). Initially, genetic studies on mutations and second site suppressors provided evidence for the functional significance of some of these helical structures *in vivo* (10-14); these were extended later on by a variety of studies involving site directed mutagenesis and in vitro splicing assays. Among the helices studied in detail are those which involve intron and exon sequences; they are likely to participate in splice site recognition and in determining its fidelity (12,15-18). In group I introns the internal guide sequence (IGS) (9) has been shown to bind up to six nucleotides of the exon next to the 5' splice site (15,16). This interaction is essential for 5' cleavage (17) and in the absence of the authentic 5' exon sequences aberrant cleaveage at a cryptic site has been observed (18).

Group II introns similarily are provided with intronic sequences which can base pair with 5' exon sequences (19-21). Two exon binding sites (EBS1,EBS2), located at a short distance inside domain I of the core structure, can base pair with two adjacent sequences of the 5' exon - the so-called intron binding sites (IBS1 and IBS2) - next to the 5' splice site (21). The study of mutations which affects the EBS/IBS pairing have revealed that this interaction is of critical importance for 5 'cleavage (21,22).

To further characterize the specificity of EBS/IBS pairings in determining 5' splice sites, we have deleted the authentic IBS2 element in the exon (A5c) upstream of intron al5c. As a result we have observed cleavages at four cryptic 5' splice sites, provided *in cis* or *in trans*; IBS like sequence elements are located 3' to these aberrant splice sites, which can interact with one or both of the exon binding sites (EBS) of intron

al5c by base pairing. These data indicate that surrogate IBS sequences consitute crytpic 5' cleavage sites, irrespective of their location relative to the intron and that IBS1 is the minimal requirement for 5' splice site selection.

MATERIALS AND METHODS

Materials

Restriction enzymes, SP6 RNA Polymerase, Klenow Polymerase, T4 DNA Ligase and Polynucleotide Kinase were purchased from Boehringer Mannheim. Reverse Transcriptase and RNA ligase were from PL Biochemicals and T3 RNA Polymerase from Genofit. ^{32}pCp, alpha ^{35}S-UTP and y-^{32}P-ATP (2000-3000 Ci/mM) were obtained from New England Nuclear.

Plasmid construction

Constructions of the initial plasmids pSP6/bI1S+ (=SP6/bI1) and pSP64/al5c are described in (7). BS/al5c was constructed by recloning of the HindIII-EcoRI insert of pSP64/al5c into the HindIII-EcoRI sites of the vector BlueScript KS+. For introduction of partial 5' exon A5 deletions, BS/al5c was digested with HindIII and KpnI and subjected to exonuclease III digestion as described in (23). Deletions were analyzed by restriction digests and DNA sequencing.

In vitro transcription

Transcripts - uniformely labelled by ^{35}S-UTP - were synthesized from pSP6 and BlueScript vectors by *in vitro* transcription with SP6 RNA-polymerase and T3 RNA polymerase, respectively, after digestion with EcoRI, NsiI or HinfI. *In vitro* transcriptions (either with SP6- or T3 RNA polymerase) were carried out in a 100 ul reaction containing 20 - 50 ug of template DNA, 50 units of the enzyme, 40 mM Tris-HCl (pH 7.5), 6 mM $MgCl_2$, 10 mM Dithiothreitol, 4 mM spermidine, 500 uM of each ribonucleotide triphosphate and 20 uCi ^{35}S-UTP for 30 min at 40^{0}C. Full-length unspliced RNA molecules ("preRNA") were separated on 5% polyacrylamide - 8M urea gels, autoradiographed, gel extracted and purified as described by Frendewey and Keller (24).

In vitro splicing reactions

Splicing reactions of purified preRNA molecules were performed as described in Schmelzer and Müller (25) with the following variation: - 20 ul incubation buffer (220 mM Tris-HCl pH 7.5, 55 mM $MgCl_2$, 20 mM Spermidine and 1.25 M NH_4Cl) at 45^{0}C. The reaction was stopped by ethanol precipitation. The resulting pellet was washed with 70% ethanol before being dried under vacuum.

3' end-labelling of RNA

To label 3' ends of RNAs, precursors were incubated in the presence of 10 uCi of ^{32}pCp with 10 units RNA ligase according to England and Uhlenbeck (25).

Primer extension analysis

RNAs isolated from polyacrylamide gels were co-precipitated with the 5'-^{32}P labelled oligonucleotides. cDNA synthesis with reverse transcriptase was performed in the presence (for sequence reactions) or absence (for continuous cDNA synthesis) of dideoxy nucleotides as described by Tabak et al. (27).

RESULTS

Effect of IBS2 deletion on original and cryptic 5' splice site selection.

The insert sequences of plasmid SP64/al5c (7) harbouring intron al5c with adjacent exon sequences were

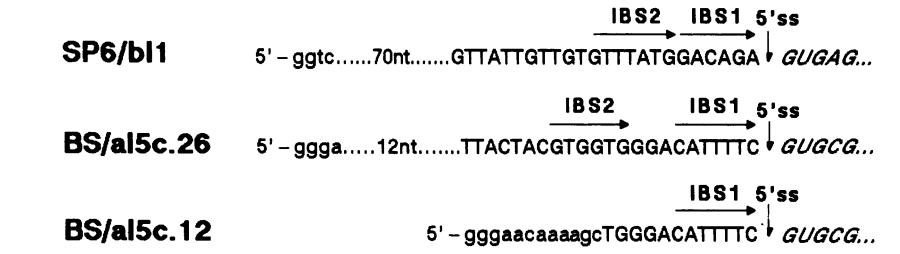


Figure 1. Schematic representation of SP6 and BlueScript (BS) template DNAs.
SP6/bl1: 3' part of exon B1 plus polylinker sequences. BS/al5c.26 and BS/al5c.12: deletions generated by exonuclease III digestion of the al5c insert described in Schmelzer and Schweyen (1986). BS/al5c.26 retains the full size exon A5c (25 nt, yeast strain 777-3A) whereas BS/al5c.12 retains only 12 nt of the 3' end of this exon. Vector sequences are in lower case letters, intron sequences in italics. IBS1,IBS2: intron binding sites; 5'ss: 5' splice site.

reclosed into the BlueScript KS+ vector. We shortened the original 5' exon sequences of BS/al5c by exonuclease III digestion to 26 nt and 12 nt generating plasmids BS/al5c26 and BS/al5c12, respectively.

Transcripts of plasmid BS/al5c26 still contain the complete 5' exon A5c of strain 777-3A with both IBS1 and IBS2 (Fig.1) as defined by Jacquier and Michel (21). They lack upstream sequences present in the longer exon A5 of strain D273-10B which made up the 5' part of the al5c harbouring preRNA studied by these authors as well as by Van der Veen et al. (20).

PreRNAs were synthesized by transcription of EcoRI-digested plasmids *in vitro* in the presence of ^{35}S-UTP; upon incubation under self-splicing conditions they were separated on polyacrylamide gels (see Methods). Major reaction products are displayed in Figures 2A and 2B. BS/al5c26 pre-RNAs (harbouring IBS1 and IBS2) show the typical splicing pattern of group II intron RNAs: the excised intron (887 nt) in lariat form (L), and in linear or branched-linear form (bL) as well as ligated exons (228 nt) and free full length exons (5': 39 nt; 3': 189 nt). In addition, lariat intermediates can be seen which still carry the 3' exon (L-3'E) as revealed from the fact that they retain radioactive label when ^{32}pCp 3' end-labelled preRNA was incubated under self-splicing conditions (not shown).

BS/al5c12 preRNAs (harbouring IBS1, but not IBS2) show the same processing efficiency as preRNAs with both IBS sequence elements. The ratios of excised introns (L-3' E, L, bL), ligated exons (213 nt) and free 5'exon (24 nt) to preRNAs BS/al5c26 and BS/al5c12 are similar, suggesting that the deletion of IBS2 does not affect the efficiency of 5' and 3' splice point recognition and exon ligation to a significant extent. As to the effect of the IBS2 element on *cis*-splicing, Jacquier and Michel (21) came to about the same conclusion when studying al5-containing transcripts with partially deleted IBS2 sequences; however, they note a certain accumulation of the intron-3'exon lariat.

We observe distinct differences, which were not mentioned before, in the fate of the free 3' exon when we compare the splicing reaction of BS/al5c12 and BS/al5c26 transcripts. First, there is significantly less free 3' exon (189 nt) in the BS/al5c12 assay. Second, products shorter than the free 3' exon, especially one in the order of 160 nt, are significantly more prominent (Fig.2A, Fig.2B). The 160 nt RNA (3'E*) species retains the ^{32}pCp 3' end-label added to the preRNA prior the incubation under self-splicing conditions (not shown),

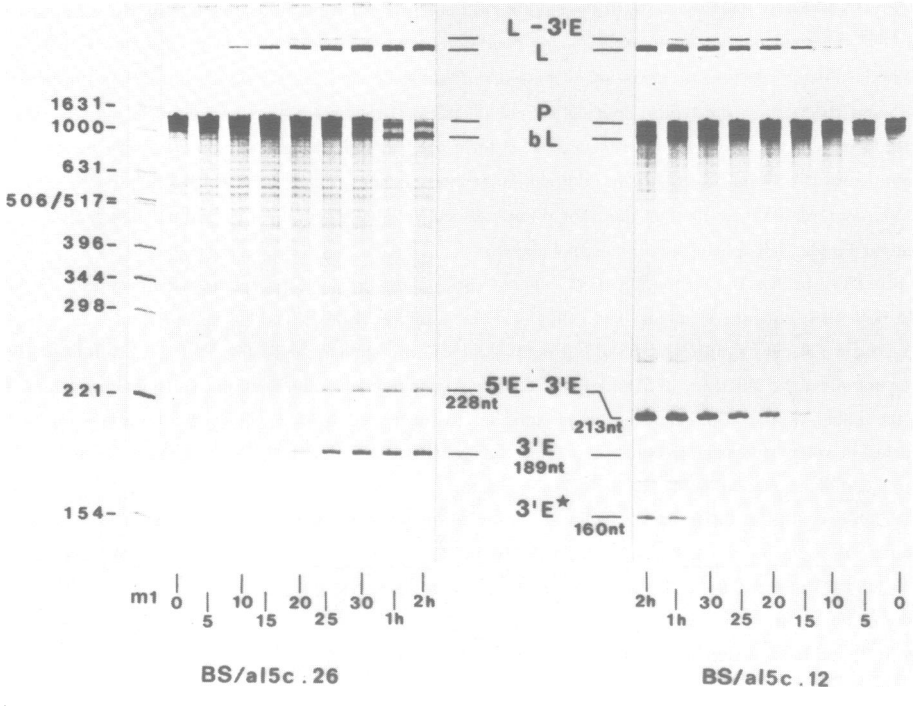
indicating that it represents the 3' part cleaved off of the 3' exon (A6). In addition a 29 nt RNA which represents the corresponding 5' part is detected.

In order to identify the putative cleavage site within the A6 exon sequence, RNA of the 3'E* band originating from BS/al5c12 was examined by primer extension analysis with reverse transcriptase. Ligated exons and free 3' exons were studied in parallel. The 3' extension of the 3'E* cDNA terminates precisely 3' of a **TATTTTC** motif (Fig.3 and Fig.4A). Aberrant cleavage at this site recently has also been noted by Jarrell et al. (28).

This sequence element **TATTTTC** is almost identical with the IBS1 motif in the 5' exon of A5c (**CATTTTC**). Furthermore, this IBS1-analogous sequence is preceded by **TGGGTG** which resembles the IBS2 motif of the 5' exon (**GTGGTG**). These findings indicate that the 3' exon A6 provides sequences similar to those preceding the 5' splice site in 5' exon A5c. The alignement of the authentic IBS1/IBS2 and the analogous sequences IBS1*/IBS2* within the 3' exon with their corresponding exon binding sites (EBS1/EBS2) is shown in Fig.4B.

As long as the authentic IBS1 and IBS2 sequences of the 5' exon are present (BS/al5c26) the 160 nt 3'E* RNA remains a very minor product only. Deletion of the authentic IBS2 sequences, as in the case of BS/al5c12 preRNA, enhances the amount of the 160 nt 3'E* RNA, indicating that the alternative pairing of the intronic EBS sequences with the IBS* sequence elements in the 3' exon is favoured, thus creating a cryptic 5' splice site recognition element.

A



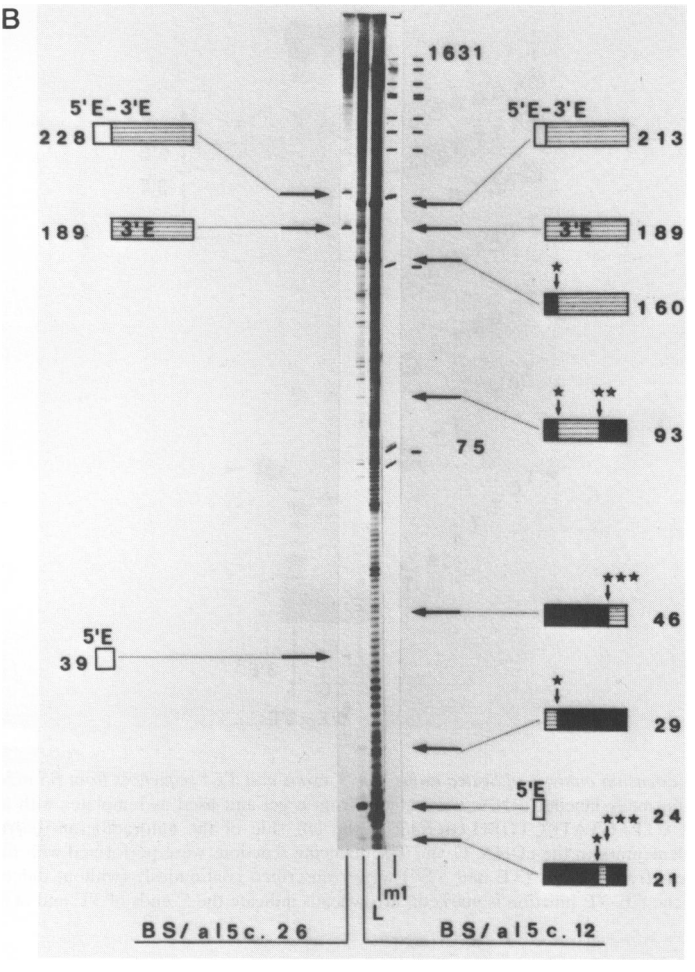


Figure 2. Effect of an IBS2 deletion on original and cryptic 5' splice site selection of al5c.

A: Time course of self-splicing reactions of BS/al5c.26 and BS/al5c.12 (IBS2 deletion). ^{35}S-UTP labelled precursors were gel-extracted and incubated under *in vitro* splicing conditions for different times as indicated. Samples were separated on 5% polyacrylamide-8 M urea gels. m1: pBR322 DNA, digested with HinfI and EcoRI. L, intron lariat; L-3'E, intron lariat-3'exon; P, precursor; bL, broken lariats and linear intron molecules; 3'E, 5'E, and 5'E-3'E are free 3' exons, 5' exons and ligated exons, respectively; 3'E*: 3' exon generated by aberrant cleavage (cf. Text).

B: Size determination of products from the BS/al5c.26 and BS/al5c.12 self-splicing reactions.
RNAs were separated on a 7.5% polyacrylamide gel (60 cm) in parallel with an RNA ladder (L). Open boxes: 5' exons, hatched boxes: 3' exons. Asterisks denote cryptic cleavage sites. Black areas represent sequences cleaved off by the aberrant reactions.

Yet turnover of the BS/al5c12 pre-RNA appears to be as efficient as of BS/al5c26, indicating that cuts at the authentic 5' and 3' splice sites also occur efficiently. These results are consistent with the findings of Jacquier and Michel (21) who could show that deletions affecting the IBS2/EBS2 pairing do not significantly

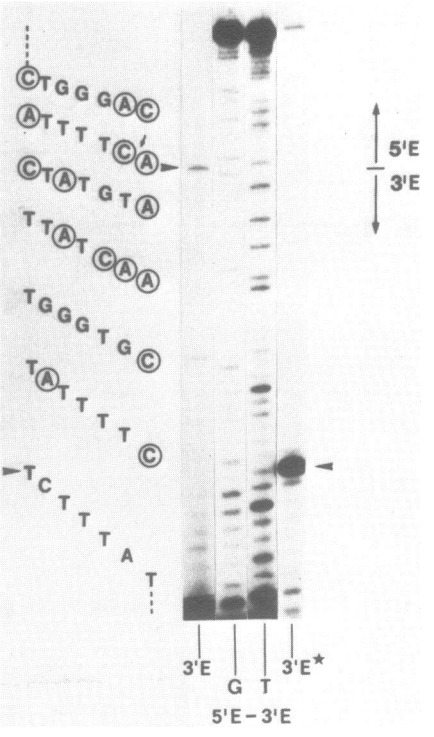


Figure 3. Primer extension analysis of ligated exons, free 3' exons and 3'E sequences from BS/a15c.12 assays.*
35S-UTP uniformelly labelled RNAs were eluted from a gel and used as templates with a 3'exon-primer (GACTTCAATAGTAGTATCCTG). Lettering on the left side of the autoradiogram corresponds to the sequence complementary to the cDNA. G and T sequencing reactions were performed with the ligated exons (5'E-3'E) whereas free 3' exons (3'E and 3'E*) were transcribed continuously (without dideoxynucleotides). The position of the 5'E-3'E junction is marked; arrowheads indicate the 5' ends of 3'E and of 3'E*.

reduce *cis*-splicing activity. Thus, it appears that the IBS1/EBS1 pairing alone is sufficient for proper recognition of the 5' exon by the intron sequence *in cis*, but that this pairing is no longer the preferred one if sequences other than the 5' exon provide both IBS1 and IBS2 motifs.

If the cut at the cryptic splice site in the 3' exon is equivalent to cleavages of authentic exon/intron junctions, the products should undergo the following ligations: 1) the 5' terminus of the 160 nt RNA, being analogous to a 5' intron end, should be transfered to the branch adenosine in stem 6 of the intron, and 2) the 3' terminus of the 29 nt RNA, being analogous to the 3' end of a 5' exon, should be spliced to the authentic 3' exon. Ligation 1) should yield Y-shaped molecules with structures similar to broken lariats but with a lower electrophoretic mobility than these. If 3' endlabelled preRNAs are incubated, these molecules should carry the label. RNAs with molecular weights as expected for such molecules have been detected after prolonged exposure of autoradiograms (not shown) but were not further characterized. Splicing products according to reaction 2) might be 29 nt long circles in the case of a *cis* reaction or 3' exons extended by 29 nt in a

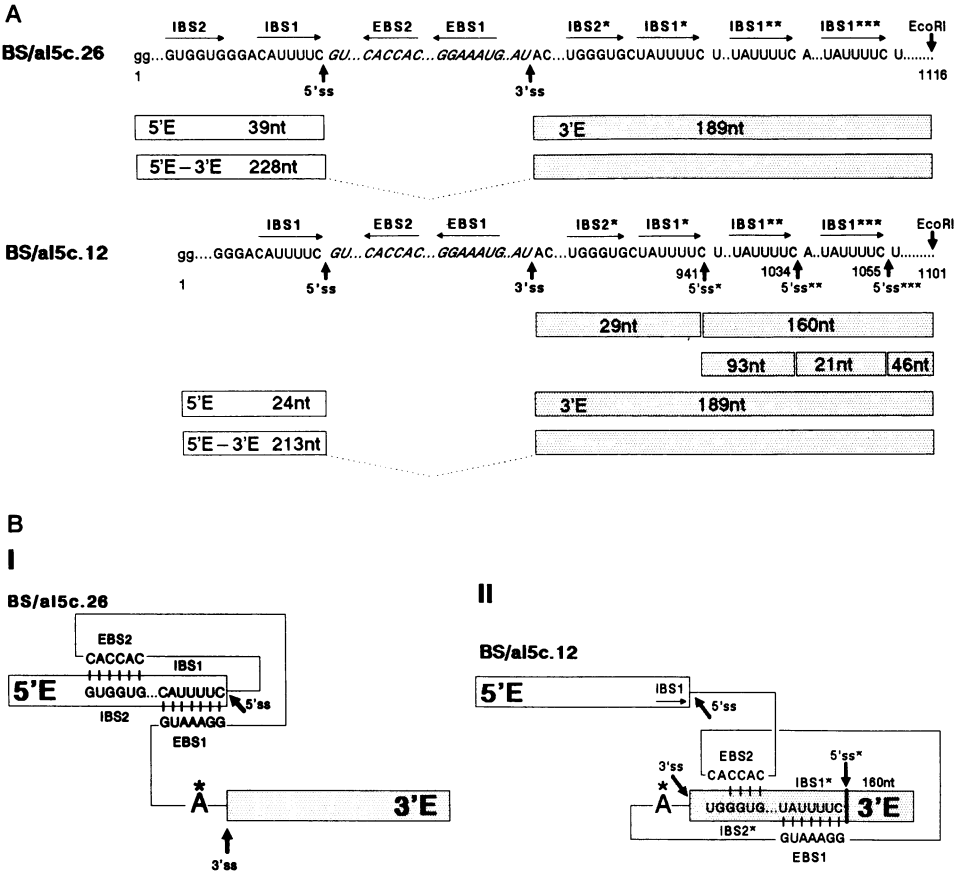


Figure 4. *Schematic representation of products generated by aberrant 5' cleavages within BS/al5c.26 and of BS/al5c.12.*

A: Binding sites and splice sites detected in BS/al5c.26 and of BS/al5c.12 RNA and products generated by cleavage and ligation

IBS1, IBS2, EBS1, EBS2: authentic binding sites; IBS1*, IBS2*, IBS1**,IBS1***: cryptic cleavage sites. Bars represent products derived from 5' exons (open) or 3' exons (dotted) with the respective sizes indicated. 5'E-3'E represents the spliced exons as indicated by a dotted line.

B: Authentic EBS/IBS pairings (I) and pairing of the authentic EBS1-EBS2 sequences with the cryptic IBS1*-IBS2* sequences (II). Arrowheads mark original cleavage sites (5'ss, 3'ss) and the cryptic 5'ss* cleavage site.

trans-reaction. Both products have been observed upon cryptic 5' cleavage at the 3' exon of a group I intron (18) but neither of them has been detected in our studies.

The self-splicing reaction of BS/al5c12 leads to the emergence of a variety of RNA products with low molecular weights, besides those described above. As this suggests the existence of additional cryptic cleavage sites, we inspected the complete BS/al5c12 sequence for further alternative intron binding sites

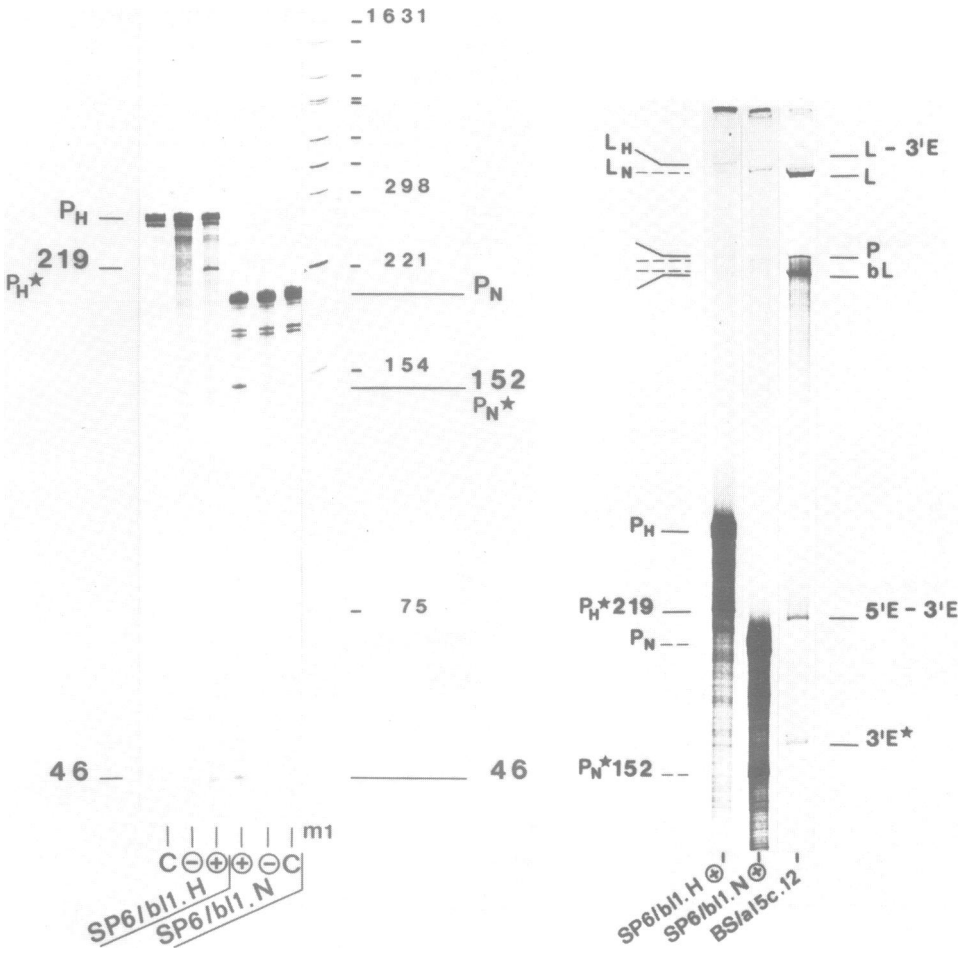


Figure 5. Cleavage of a cryptic 5' splice site by al5c preRNA in a trans reaction.

Unlabelled BS/al5c.12 preRNA was incubated under self-splicing conditions with ^{35}S-UTP-labelled SP6/bl1 transcripts terminated at a HinfI site (SP6/bl1.H) or a NsiI site (SP6/bl1.N) in the 5' portion of the bl1 intron. Reaction products were separated on a 5% denaturing polyacrylamide gel. Note that all products visible in the autoradiographs will contain sequences of the ^{35}S-UTP labelled "substrate" transcripts SP6/bl1.H (= P_{11}) or SP6/bl1.N (= P_N).

Left panel: ^{35}S-UTP-labelled RNAs were separated without prior incubation (C), upon incubation under standard self-splicing conditions without (-) and with the addition of unlabelled BS/al5c.12 preRNA (+). m1: pBr322 DNA, digested with HinfI and EcoRI. Labelled P_N*152, P_{11}*219 and the 46 nt RNA result form cleavage of the P_{11} and P_N RNA by al5c RNA *in trans* (cf. Text).

Right panel: Upper part of the autoradiograph upon prolonged exposure time, showing labelled products migrating slower than the labelled precursor RNAs (L_N, L_H). Products of a standard self-splicing assay with (labelled) BS/al5c.12 pre-RNA were co-electrophoresed for comparison.

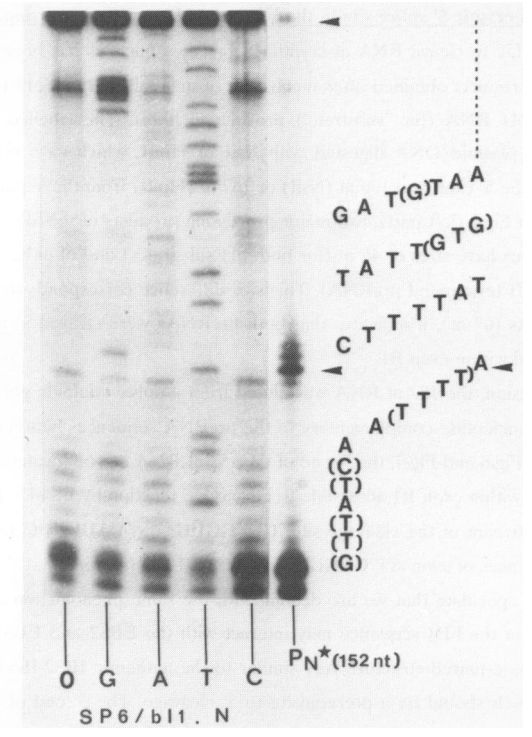


Figure 6. *Primer extension analysis of PN*152.*

The putative 3' part cleaved off the SP6/bI1 substrate, PN*152, was purified from the gel and analysed by reverse transcription with a 5' exon-primer (AATAACCTAAAAAAGCTGTAG, pos. 60-81 in exon B1). The relevant part of the sequence of the SP6/bI1.N transcript - obtained by reverse transcription with the same primer - is shown in parallel together with a control reaction without dideoxynucleotides (0). Arrows mark the 5' end of PN* and a signal representing the end of the SP6/bI1.N RNA which was present as a contaminant in this reaction.

(IBS*). Interestingly, at positions 1029-1034 and 1050-1055 within A6, two sequences with the exact **YATTTTC** aI5c IBS1 motif are located (Fig.4A). The question thus arises if these sequences could constitute cryptic 5' cleavage sites despite the lack of an IBS2 sequence located upstream.

As outlined in Fig.4A, simultaneous cleavages at positions 941, 1034 and 1055 within the 3' exon A6 would yield short RNA molecules with theoretical lengths of 29 nt, 93 nt, 21 nt and 46 nt. RNAs with these lengths can be detected among the products of BS/aI5c12 self-splicing assays (Fig.2B). In addition, slower migrating RNAs are detected which might have arisen from cleavages of various splicing products at one or the other cryptic site (i.e. within the ligated exons). These small RNAs have not been further characterized. Yet, as their lengths precisely match to those expected, it is tempting to conclude that the IBS1** and IBS1*** sequences, as well as the IBS2*/ IBS1* elements, have served as cryptic 5' cleavage sites.

al5c-preRNA cleaves at a cryptic 5' splice site in the 5' exon of bI1 in a *trans*-reaction.

The ability of intron al5c to cleave RNA at cryptic IBS*(al5c) sequences has been further documented by the analysis of reaction products obtained after incubation of unlabelled BS/al5c12 transcript (the 'enzyme') with ^{35}S-UTP labelled bI1 RNA (the 'substrate') provided *in trans*. The labelled RNA was obtained by transcription of SP6/bI1 plasmid DNA digested with NsiI or HinfI, which gave rise to run-off transcripts composed of 98 nt from the 5' exon and 100 nt (NsiI) or 167 nt (HinfI) from the 5' part of the intron.

As can be infered from Fig.5 (left part), two major processing products of the SP6/bI1 transcript are visible in the autoradiograph; they have sizes of 46 nt (for both bI1 substrates) and of either 152 nt (NsiI terminated preRNA) or 219 nt (HinfI terminated preRNA). This size difference corresponds to the length difference of the 3' truncated preRNAs (67 nt); it indicates that both preRNAs were cleaved at identical positions, 46 nt from their common 5' end within exon B1.

To confirm this conclusion, the 152 nt RNA was eluted from a polyacrylamide gel and analyzed by primer extension using an oligonucleotide complementary to the preRNA sequences between positions 60 and 81 in the 5' exon. As shown in Fig.6 and Fig.7, the 5' end of the 152 nt RNA can be located at position 47.

What renders this site within exon B1 accessible to a cleavage reaction by BS/al5c pre-RNA? Sequences in exon B1 immediately upstream of the cleavage site, GUAGGUG$_{UUAU}$UAUUUUC reveal striking similarities with the IBS2-IBS1 sequences of exon A5, GUGGUG$_{GGA}$CAUUUUC (cf. Fig.7).

Thus it is tempting to speculate that we are dealing with the same phenomenon as described above. The cryptic IBS2*/IBS1* within the bE1 sequence may interact with the EBS2 and EBS1 sequences within al5c preRNA to establish a base-paired structure very similar to the authentic IBS2-IBS1/EBS2-EBS1 sequences of the al5c transcript which should be a prerequisite to 5' cleavage. The 5' end of the cleaved 152nt RNA corresponds to the first unpaired nucleotide 3' of IBS2*(al5c)-IBS1*(al5c), which is consistent with observations made for authentic IBS sequences (21).

Substitution of BS/al5c12 RNA (IBS2 deletion) by BS/al5c26 (IBS2 + IBS1) leads to a reduction of the recognition of the cryptic splice site in exon B1 (not shown). This indicates that the presence of the authentic intron binding sites within the same molecule reduces the intermolecular base pairing reaction of the cryptic intron binding sites within bE1 with the exon binding sites of the intron al5c.

According to the group II self-splicing scheme (6) we might expect that the branch point of intron al5c exerts a nucleophilic attack on the exon B1 sequence, serving as a surrogate for the authentic 5' splice junction A5c/al5c. If this reaction is at least partially coupled with covalent linkage of the pseudo 5' intron terminus (in B1) with the 2'OH group of the branch nucleotide (in al5c), we would expect labelled Y-shaped molecules with distinct lengths, depending on whether the "trans-branched" al5c-intron still carried its 5' or 3' exon or both. Actually, minor RNA bands can be detected on the autoradiographs which might represent such molecules (Fig.5, right panel); their low concentration, however, did not allow them to be further characterized.

Unexpected was the finding that labelled RNAs occur which have electrophoretic mobilities typical of lariats (L_{II}, L_N), but which migrate slower than the authentic al5c lariat (L). This decrease of electrophoretic mobility correlates with the length difference at the 3' end of the substrate RNAs. In addition, these lariats carried the label when 3' endlabelled RNAs were used as substrates (not shown). Both results indicate that the 3' portion of the cleaved substrate molecules is transferred to this lariat structure. Occupation of the

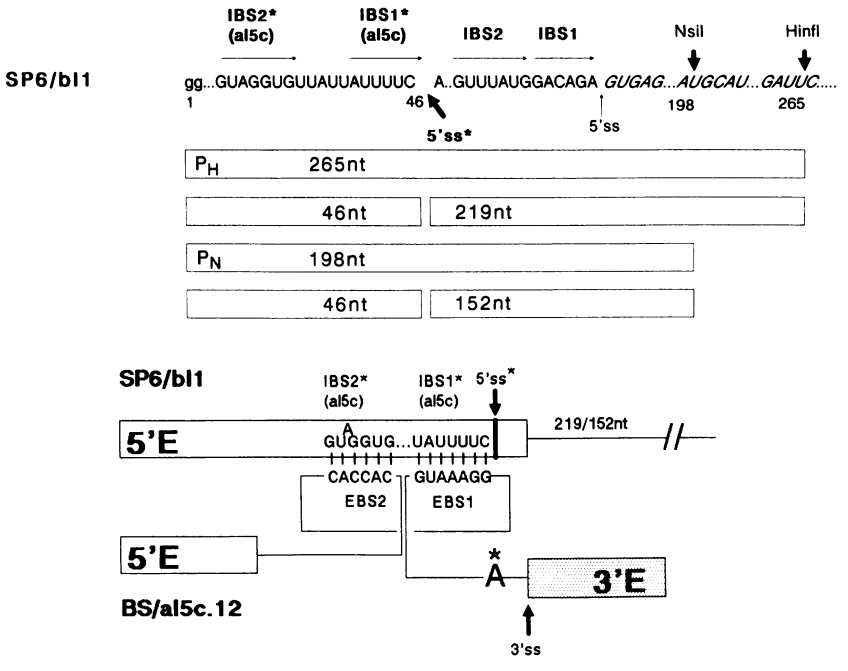


Figure 7. Schematic representation of transcript sequences and their putative interaction in trans.

Upper part: Positions of binding sites in the 5' exon of SP6/bl1 RNA recognized by intron bl1 (authentic IBS2-IBS1) or by intron al5c (cryptic IBS2*-IBS1*$_{al5c}$). NsiI and HinfI mark restriction sites at positions 198 and 265 in SP6/bl1 RNA where run-off transcripts P_N and P_{II} were terminated. Open bars represent the P_N and P_{II} transcripts of 198 nt and 265 nt, respectively, and products generated by their cleavage at the cryptic 5' splice site (5'ss*) recognized by intron al5c *in trans*.

Lower part: Putative parings *in trans* of authentic EBS2-EBS1 sequences in intron al5c and cryptic IBS2*-IBS1*$_{(al5c)}$ sequences in the 5' exon preceding intron bl1.

original branch point of al5c by a transfer of the 3' portion of the cleaved, labelled RNA would result in formation of Y-shaped molecules; thus, the question arises how this lariat structure can be formed. One possible explanation would be that these labelled lariats have two branches, one formed by intramolecular interaction of al5c, the other by B1 sequences in a 5'-2' linkage of P_{II}^* or P_N^* to the al5c sequence.

DISCUSSION

We have shown here that group II intron al5c cleaves surrogate 5' junctions irrespective of whether they are provided *in cis* (upstream or downstream of the intron) or *in trans*. The minimal requirement for recognition of a surrogate 5' junction apparently is the base pairing of the intronic EBS1 with a complementary (IBS1*) sequence, forming a short helix. We have observed cleavages at all sites fulfilling this requirement (a total of 4 surrogate sites), except possibly one IBS* motif within the intron which we have not tested for its ability to cause an aberrant 5' cleavage. However, as this sequence forms part of a stable intronic helical region, it may not be accessible to an alternative binding with an EBS sequence.

Authentic ends of group II introns consistently contain the well conserved GUGYG motif directly 3' to the 5' splice site, but none of the five surrogate IBS* sites is followed by a similar sequence. They even lack the 5' G residue of this motif; instead an A or a U is next to the cleavage site. This suggests that neither the entire GUGYG motif nor its first base are essential parts of the substrate which is to be recognized and cleaved. The only information required for a substrate to precisely direct a nucleophile to the 5' cleavage site is thus provided by the IBS1 sequence of the 5' exon sequences. In addition to promoting the lariat formation, the correct GUGYG sequence could be important for exon ligation. However, results of Jaquier and Rosbash (19) who have shown that even a solitary 5' exon can be ligated to the 3' exon of an intron lariat-3' exon, argue against an important role of the GUGYG motif for ligation.

The rate of cleavage at surrogate 5' junctions depends on two major factors. First, it is highest at sites bearing both IBS1* and IBS2* sequences in the proper order. However, the authentic junctions with both IBS sequences are highly preferred over surrogate junctions which also provide both IBS* sequences. Second, deletion of the authentic IBS2 sequence strongly enhances cleavage at all surrogate splice sites.

In line with previous studies we regard EBS1/IBS1 pairings as being sufficient for efficient and precise recognition of the authentic 5' junctions *in cis*, but not for such recognition of surrogate 5' junctions (this work) or of authentic 5' exons provided *in trans* (19).

The IBS2/EBS2 pairing, thus, may not serve to ensure proper recognition of authentic 5' junctions (in addition to IBS1/EBS1), but rather to prevent the intron sequences from accepting surrogate 5' junctions, which in the long transcripts of the yeast mitochondrial mosaic genes *COB* and *OXI3 in vivo* (8kb and 13 kb, respectively) are abundant.

We may ask, then, what gives the authentic 5' junction this preference over others. Unique is its position at the 5' end of the intron and it is conceivable that the three-dimensional structure of the intron helps to bring the authentic IBS sites into contact with the EBS sequences of the intron. *In vivo* physiological ionic conditions and accessory factors like RNA binding proteins or RNPs might help to ensure the proper IBS/EBS interaction.

ACKNOWLEDGMENTS

We thank Benjamin Feldman and Klaus Wolf for critical reading of the manuscript, Barbara Gelhaus for her expert technical assistance and Thomas Jansen for synthesis of the oligonucleotide primers. This work was supported by a grant from the Deutsche Forschungsgemeinschaft.

REFERENCES

1. Cech,T.R., Zaug,A.J., and Grabowski,P.J. (1981). Cell **27**, 487-496.
2. Garriga,G., and Lambowitz,A.M. (1984). Cell **39**, 631-641.
3. Van der Horst,G., and Tabak,H.F. (1985). Cell **40**, 759-766.
4. Chu,F.K., Maley,G.F., West,D.K., Belfort,M., and Maley,F. (1986) Cell **45**, 157-166.
5. Peebles,C.L., Perlman,P.S., Mecklenburg,K.L., Petrillo,M.L., Tabor,J.H., Jarrell,K.A., and Cheng,H.-L. (1986). Cell **44**, 213-223.
6. Van der Veen,R., Arnberg,A.C., Van der Horst,G., Bonen,L., Tabak,H.F., and Grivell,L.A. (1986). Cell **44**, 225-234.
7. Schmelzer,C., and Schweyen,R.J. (1986). Cell **46**, 557-565.
8. Michel,F., and Dujon,B. (1983). EMBO J. **2**, 33-38.
9. Waring,R.B., Scazzocchio,C., Brown,T.A. and Davies,R.W. (1983) J.Mol. Biol. **167**, 595-605.

10. Weiss-Brummer,B. Rödel,G., Schweyen,R.J. and Kaudewitz, J. (1983). Cell, **33**, 195-202.
11. Schmelzer,C., Schmidt,C. May,K. and Schweyen,R.J. (1983) EMBO J. **2**, 2047-2052.
12. Perera,J. and Jacq,C. (1985) EMBO J. **4**, 3281-3288.
13. Holl,J., Rödel,G. and Schweyen, R.J. (1985) EMBO J. **4**, 2081-2085.
14. Waring,R.B., Ray,J.A., Edwards,S.W., Scazzocchio,C. and Davies,R.W. (1985) Cell **40**, 371-380.
15. Waring,R.B., Towner,P., Minter,S.J., and Davies,R.W. (1986). Nature **321**, 133-139.
16. Been,M.D., and Cech,T.R. (1986). Cell **47**, 207-216.
17. Chandry,P.S. and Belfort,M. (1987) Genes Dev. **1**, 1028-1037.
18. Price,J.V., Engberg,J., and Cech,T.R. (1987). J.Mol.Biol. **196**, 49-60.
19. Jaquier,A., and Rosbash,M. (1986). Science **234**, 1099-1104.
20. Van der Veen,R., Arnberg,A.C., and Grivell,L.A. (1987). EMBO J. **6**, 1079-1084.
21. Jaquier,A., and Michel,F. (1987). Cell **50**, 17-29.
22. Michel,F., and Jaquier,A. (1987). Cold Spring Harbor Symp.Quant.Biol. **52** NY.
23. Henikoff,S. (1984). Gene **28**, 351-359.
24. Frendewey,D., and Keller,W. (1985). Cell **42**, 355-367.
25. Schmelzer,C. and Müller, M. (1987) Cell **51**, 753-762.
26. England,T.E., and Uhlenbeck,O.C. (1978). Nature **275**, 560-561.
27. Tabak,H.F., Van der Horst,G., Osinga,K.A., and Arnberg,A.C. (1984). Cell **39**, 623-629.
28. Jarrell,K.A., Peebles,C.L., Dietrich,R., Romiti,S.L., and Perlman,P.S. (1988). J.Biol.Chem. **263**, 3432-3439.