Structural requirements for selection of 5'- and 3' splice sites of group II introns

Christian Wallasch, Mario Mörl, Isabell Niemer and Carlo Schmelzer*

Institut für Genetik und Mikrobiologie der Universität München, Maria-Ward-Str. 1a,
D-8000 München 19, FRG

Received March 12, 1991; Revised and Accepted May 21, 1991

ABSTRACT

The group II intron b11 in the gene for apocytochrome b in yeast mitochondrial DNA (COB) is self-splicing *in vitro*. It could recently be shown that self-splicing of this intron is fully reversible *in vitro*. In addition, intron integration is not restricted to parental exons, since the intron can also integrate into a foreign RNA. The position of insertion seems to be immediately 3' to a cryptic intron binding site 1 (IBS1). We confirmed and extended these results by sequencing 26 individual RNAs with transposed introns after reverse transcription and PCR amplification. Results show that intron integration into authentic exons is generally correct, but that integration into a foreign RNA is often inaccurate, i.e. insertion is one nt downstream or upstream of the 3' end of IBS1. This leads to the generation of 5' splice junctions of the new intron-harboring 'preRNAs' with addition (or deletion) of a single A residue at the 3' end of IBS1. To investigate which structures help to define the position of 5'- and 3' cleavage, preRNAs of i) these clones with aberrant 5' splice junctions and ii) preRNAs with artificial hairpins between domains 5 and 6 of the intron were spliced under different reaction conditions. Results obtained let us conclude that i) branchpoint dependent 5' cleavage is directed by the 5' terminal G residue of the intron and, ii) the first nucleotide(s) of the 3' exon play an important role in defining the 3' splice site.

INTRODUCTION

At least some members of both, group I and group II introns exhibit autocatalytic activities *in vitro*. Among group II introns these are introns aI5c in the mitochondrial subunit I cytochrome oxidase gene of *S.cerevisiae*, and b11, the first intron in the apocytochrome b gene of the same species. Recently, two additional self-splicing introns could be identified, LSUrRNA I2, in the mitochondrial large subunit rRNA gene of the green alga *Scenedesmus obliquus*, and intron COI I1, the first, 'mobile' intron of the mitochondrial gene coding for cytochrome oxidase subunit I of the ascomycete *Podospira anserina* (1–6).

In both intron groups, self-splicing proceeds by two transesterification reactions. The initial step is the attack of the 2' OH group of the branch adenosine (for group II introns) or the 3' OH group of an external guanosine (for group I) at the 5' splice site which generates the free 5' exon and the intron-3' exon intermediate. In the second reaction step, the 3' OH group of this freed exon attacks the 3' splice site generating ligated exons and the excised intron.

For group I introns, base pairing interactions between sequences at the 3' terminus of the 5' exon and a short sequence located near the 5' end of the intron, the so-called internal guide sequence (IGS) have been proven to play a key role for both reaction steps and may help to define the 5' splice site (7–10). Similarly, in the case of group II introns the mutual recognition of two sequence elements at the 3' end of the upstream exon (intron-binding sites 1 and 2; IBS1,2) with complementary exon-binding sites 1 and 2 (EBS1,2) located in domain 1 of the intron is functionally important for the 5' cleavage step and for stabilizing the intermediate complex (5' exon + intron-3' exon lariat) (11). As shown in a recent work, the IBS1-EBS1 interaction is assisted by a long-range base pairing, the ϵ/ϵ' pairing between the third and fourth nucleotides of the intron with two consecutive nucleotides in a loop of domain 1 (12). Whereas some of the elements involved in the choice of the 3' splice site for group I introns could be meanwhile characterized (see, for review 13), the situation is less clear in the case of group II introns. Several conserved structures could theoretically define the splice site, e.g. (i) the structure of domain 6 and/or its location within the catalytic center, (ii) the fixed distance of the 3' splice site from the branch point and, (iii) the γ'/γ' base pairing in which the last intron nucleotide of the conserved AY motif at the 3' end of the intron forms a solitary base pair with a nucleotide in a sequence stretch connecting domains 2 and 3 (11).

Whereas several lines of evidence indicate that the aforementioned long range interactions between exon-intron sequences and intron-intron sequences are important for splicing, so far no experimental evidence has been offered proving that they are also involved in defining the positions of 5'- and 3' cleavage. For example, disruption of the EBS:IBS-, γ'/γ' -, or ϵ/ϵ' - pairing leads to reduction of the splicing efficiency, however

* To whom correspondence should be addressed

selection of aberrant splice sites has as yet not been observed (11,12).

For group II introns, the branch point dependent 5' cleavage can be bypassed *in vitro* by incubation of the preRNA in KCl, that favours 5' cleavage by hydrolysis. This reaction mode leads to excision of introns in linear form and to reopening of ligated exons at the ligation junction in a reaction *in trans* yielding free 5'- and 3' exons (14). This *trans* reaction mode is obviously not restricted to authentic ligated exons, since it could be verified that unrelated RNA species can also be cleaved *in trans* (15). The only requirement for this reaction is the presence of a cryptic IBS1 element on the RNA provided *in trans*. The cleavage position is immediately downstream of this cryptic IBS1. Furthermore, we could show in a previous work that bI1 is able to integrate into the parental ligated exons or a foreign RNA (part of yeast *CBS2* transcript) by reversal of the self-splicing reaction downstream of a sequence that matches the IBS1 motif GAC-AGA (16,17). The reaction is initiated by binding of the intron lariat via EBS1 to the IBS1 element of the substrate RNA (Fig.1A). These observations implied that the site for the first step of integration (which corresponds mechanistically to the reverse 3' cleavage) might simply be defined by the end of the IBS1:EBS1 pairing.

Results presented in this work let us now suggest that the EBS1:IBS1 pairing, albeit necessary for splice site recognition and activation, does not unequivocally define the position for 5' cleavage in the forward reaction and the site for intron transposition. For accurate recognition of the 5' cleavage site the 5' terminal guanosine of the intron constitutes an important component. In addition, our data confirm that 3' exon sequences might help in determining the 3' splice site.

MATERIALS AND METHODS

Materials

Restriction enzymes, Klenow polymerase, Taq DNA polymerase, T4 DNA ligase, AMV reverse transcriptase, T7- and T3 RNA polymerase were purchased from Pharmacia. α - 35 S-UTP and γ - 32 P-ATP (2000–3000 Ci/mMol) were obtained from Amersham.

Preparation of RNA

Transcripts—uniformly labelled with 35 S-UTP—were synthesized by *in vitro* transcription with T3 or T7 RNA polymerase. Transcription assays were carried out in a 20–200 μ l reaction containing 5–50 μ g of template DNA, 50–500 units of the enzyme, 40 mM Tris-HCl (pH 7.5), 6 mM MgCl₂, 10 mM dithiothreitol, 4 mM spermidine, 500 μ M of each ribonucleoside triphosphate and 20 μ Ci 35 S-UTP for 2h at 37°C. Following transcription, preRNAs were electrophoresed on 5% polyacrylamide 8M-urea gels, autoradiographed, extracted from the gel and purified. As templates for synthesis of full-length preRNA, plasmids, digested with various restriction enzymes as shown in Fig.2, were used.

In vitro splicing

In vitro splicing was performed in 20 μ l buffer S (40 mM Tris-SO₄ pH 7.0, 60 mM MgSO₄, 2 mM spermidine and 500 mM (NH₄)₂SO₄) or buffer C (40 mM Tris-HCl pH 7.0, 60 mM MgCl₂, 2 mM spermidine and 500 mM KCl) at 45°C for 30 min. The reaction was stopped by ethanol precipitation. The resulting pellet was washed with 70% ethanol, dried under vacuum and products were electrophoresed on denaturing 5% polyacrylamide gels.

Figure 1. A. Proposed mechanism of group II intron transposition. Transposition is initiated by binding of the EBS1 element of the intron lariat (L) to a complementary IBS1 element of a foreign RNA (a). The 3' OH group at the 3' terminus of the intron attacks the phosphodiester bond downstream of IBS1, the 3' part of the foreign RNA is ligated to the 3' end of the intron (b). The 3' OH group of the 5' part of this RNA that is still attached to the intron via the IBS1:EBS1 pairing, attacks the 5'-2' phosphodiester bond of the lariat; this second trans-esterification reaction results in integration of the intron (c). (Left) The secondary structure of helix D3 harbouring the EBS1 sequence in the terminal loop. B. Determination of the insertion site of intron bI1 within *CBS2* RNA. The products of intron transposition were extracted from the gel, reverse transcribed with a 3' primer (GCAGAGGTTATTCTC) complementary to the 3' part of the *CBS2* RNA and PCR amplified in presence of the same 3'- and the 5' primer (GGTCTCCTTAAGTG). After cloning and transformation, DNAs of individual clones were sequenced with a bI1 intron primer (GGTATATGTTATATATAAAC, located 169–188nt from the 5' end of the intron) and the *CBS2* 3' Primer. Only sequences of the 5' junctions are shown. Examples of correct integration (lane 0, CBS2i5'0), +1 shifts (lane +1, CBS2i5'+1) and -1 shifts (lane -1, CBS2i5'-1) are shown. Letters below the gels refer to the dideoxynucleotides used in each sequence reaction, letters on the left side correspond to the RNA sequence. The arrow marks the 5' junction, the cryptic IBS1 motif GACAGA is indicated by the dotted line. *CBS2* sequences are in capital letters, sequences of the 5' end of intron bI1 are in lower case letters.

Primer extension analysis

For primer extension analysis, RNAs isolated from polyacrylamide gels were co-precipitated with the 5'-³²P labelled oligonucleotides. cDNA synthesis with AMV reverse transcriptase was performed in the presence (sequence reactions) or absence (continuous cDNA synthesis) of dideoxynucleotides.

PCR

RNAs with integrated bI1 intron were eluted from polyacrylamide gels and reverse transcribed with AMV reverse transcriptase in the presence of primers corresponding to the 3' *CBS2* portion. cDNAs were incubated in a 100 μ l PCR reaction mixture containing 10 mM Tris-HCl (pH 8.3), 3.0 mM MgCl₂, 0.001 % gelatin, 200 μ M of each deoxynucleoside triphosphate, 20 pmol of each primer, and 2.5 units of Taq DNA polymerase. Before addition of the enzyme and deoxynucleotides, the mixture was denatured for 7 min at 92°C. The cycling conditions were as follows: 92°C for 1 min, 40°C for 1 min, and 72°C for 2 min for 30 cycles. PCRs were performed in a Perkin-Elmer/Cetus DNA thermocycler and reaction products were electrophoresed in 1.5 % agarose. Desired DNA fragments were eluted from the gel, treated with Mungbean nuclease, blunt-end ligated into a *Sma*I-digested BlueScript KS+ vector (Stratagene) and sequenced from both ends.

RESULTS

Integration of group II intron RNA into a foreign RNA is frequently inaccurate

We set out to investigate the accuracy of transposition of intron bI1 into the yeast *CBS2*-transcript now in more detail. *CBS2*-RNAs (i.e. 'preRNAs' with the intron integrated between the 5'- and 3' parts of the *CBS2* transcript) were eluted from polyacrylamide gels, reverse transcribed with a primer complementary to the 3' part of *CBS2*, PCR amplified with an additional 5' *CBS2* primer and sequenced. The results showed

Figure 2. Map of DNA templates used for *in vitro* transcription. T7 RNA polymerase was used for transcription of BS/CBS2i5' -1, for all other templates T3 RNA polymerase was used. The sizes of the exons (black boxes), vector sequences (dashed lines) and the introns (solid line) are indicated below the drawings. The positions of restriction enzymes used for termination of transcription are indicated (B, BamHI; E, EcoRI). ▴, ▴ IBS1 and IBS2; ▴, ▴ EBS1 and EBS2. d6*: Insertion of a hairpin structure between domains 5 and 6. The sequences and secondary structures of domains 5, 6 and 6* of BS/S5HA6, BS/S5HB6 and BS/S5HC6 are shown in Fig.5. Intron lengths: BS/S5HA6, 780nt; BS/S5HB6, 784nt; BS/S5HC6, 790nt. Sizes of 5' exons for BS/CBS2i5'0, BS/CBS2i5' +1 and BS/CBS2i5' -1 were 127, 128, and 104 nt, respectively; sizes of 3' exons were 37, 36, and 46nt.

an unexpected high frequency of inaccurate transposition. Eight out of thirteen clones tested exhibited correct intron insertion, i.e. at the 3' nucleotide immediately following IBS1 (the 5' junction is ..GACAGA|gugagac..., the 3' junction ..aaU|AAAA-G.; Fig.1B, lane 0), however we found numerous examples where intron insertion has shifted by one nucleotide upstream (3 clones) or downstream (2 clones) of IBS1 leading to preRNAs with deletion of the 3' nucleotide of IBS1 (Fig.1B, lane -1) and the addition of an A residue (Fig.1B, lane +1), respectively. From these observations the conclusion may be drawn that the presence of a cryptic IBS1 motif on the *CBS2*-RNA is obviously not sufficient to ensure precise intron integration, (i.e. the reverse 3' cleavage).

We also tested the precision of intron integration into the original ligated exons (bE1ΔIBS2-bE2, obtained by the autocatalytic reaction of BS/bI1Δ+6 preRNA with deletion of

Figure 3. Time-course of the self-splicing reactions of CBS2i-preRNAs. ³⁵S-UTP labelled transcripts from templates BS/CBS2i5'0, BS/CBS2i5'+1 and BS/CBS2i5'-1 were incubated for 0,1,2 or 4 h (lanes 0-4, respectively) at 45°C. Reaction products were separated on 8 M-urea/5 % polyacrylamide gels. A, Incubation in reaction buffer C (with 500 mM KCl). B, Incubation in reaction buffer S (with 500 mM (NH₄)₂SO₄). Intron lariat (L), preRNA (P), the linear form of the intron and/or the broken lariat (IVS) and ligated exons (5' E-3' E). As a size marker, characterized reaction products of preRNA from BS/bI1 Δ+24 were coelectrophoresed (M).

Figure 4. Determination of 5' splice sites by reverse transcription of excised introns of BS/CBS2i5'0, BS/CBS2i5'-1 and BS/CBS2i5'+1. 5' ends of linear introns (obtained by incubation in buffer C) or lariats (obtained by incubation of preRNAs in buffer S) were analysed by primer extension with the intron primer (cf. Fig. 1). To obtain a size marker, the same primer was used to sequence BS/CBS2i5'0 DNA by the chain termination reaction. 5' ends of the introns are marked by an arrow. Lanes are labelled according to the dideoxynucleotides used in the sequence reactions. A: intron lariats (L) from BS/CBS2i5'0 (0) and BS/CBS2i5'-1 (-1), B: linear introns (IVS), C: intron lariat and linear intron from BS/CBS2i5'+1 (+1). 5' junction sequences of the different preRNAs with the IBS1:EBS1 pairings and the positions of 5' cleavages are shown below the figure. Filled triangles mark sites of cleavage by hydrolysis, open triangles mark sites of cleavage by transesterification. Intron sequences are in lower case letters, the IBS1 sequence of the 5' exon is in capital letters. The sequence of EBS1 is marked by the cross-hatched box.

IBS2, cf. Fig. 2 and ref. 17). Strikingly, among 13 clones tested we found only 1 example of incorrect integration with a -1 shift, whereas 5 out of 13 intron transposition events were inaccurate. Both RNAs, Δ bE1 IBS2 and CBS2, differ by the sequences surrounding the common IBS1 motif. Since we have thus far no indication that could point to the importance of sequences upstream of IBS1 for splice site selection, we conclude that these observations reflect the difference in the sequences downstream of IBS1 which formally correspond to the 3' exons (5' terminus of the cryptic 3' exon of CBS2 is AAAAGU... instead of UGUC-AC... of the authentic 3' exon). Thus, it is tempting to speculate that the inaccuracy of intron integration into CBS2-RNA is largely due to the absence of authentic exon sequences.

Effects of 5' splice site mutations on self-splicing efficiency *in vitro*

In order to gain more insight into the role of the EBS1:IBS1 base pairing for splice site selection we took advantage of the constructs with deletion (BS/CBS2i5'-1) and addition (BS/CBS2i5'+1) of the adenosine at the 3' end of IBS1, respectively. These clones contain the complete sequence of bI1 with adjacent 5'- and 3' parts of CBS2 (Fig. 2). To examine the effect of each 5' splice site alteration on the *in vitro* self-splicing forward reaction transcripts from these clones were incubated under different self-splicing conditions. Incubations were carried out either in buffer C, which favours the hydrolytic 5' cleavage mode, or in buffer S, which allows 5' processing by both transesterification and hydrolysis.

Fig. 3 shows the time course of self-splicing assays of preRNAs of BS/CBS2i5'0, BS/CBS2i5'+1 and BS/CBS2i5'-1 in buffer C (Fig. 3A), and buffer S, respectively (Fig. 3B). BS/CBS2i5'0 preRNA reacted very efficiently in buffer C giving rise to a gel pattern similar to those observed with bE1-bI1-bE2 preRNA; that

is, the excised intron lariat, minor amounts of the linear form of the intron and/or y-shaped broken lariats and ligated exons. The reaction intermediate, the intron-3' exon lariat, is not detectable. At first sight this observation is surprising, since it could be shown (for the related intron aI5c) that mutations that impair the IBS2:EBS2 pairing led to the accumulation of intron-3' exon lariats. This indicates that the IBS2:EBS2 interaction is required to prevent premature dissociation of the 5' exon from the intermediate splicing complex after the first splicing step (11). We surmise that our result simply reflects the higher stability of the IBS1:EBS1 base pairing of bI1 ($\Delta G = 11.0$ kcal/mol) compared to $\Delta G = 6.4$ kcal/mol for aI5c). I.e., for bI1 the IBS1:EBS1 pairing alone seems to be sufficient to ensure the stable association of the 5' exon with the intron.

The self-splicing reaction of CBS2i5'+1 preRNA exhibits similar kinetics, however a clear shift of the lariat/linear intron ratio towards linear intron molecules is detectable. Only faint bands of lariats can be seen suggesting that the A insertion interferes with branch point dependent 5' cleavage, however does not hamper the hydrolytic 5' cleavage. In addition, the disappearance of spliced exons with concomitant accumulation of free 5'- and 3' exons (not visible on the autoradiograph) after one hour of incubation is indicative of an increased reopening of spliced exons. Splicing phenotypes of BS/CBS2i5'0 and BS/CBS2i5'+1 in buffer S are comparable to those observed in buffer C. However, as expected, the increased accumulation of excised introns in lariat form confirms that in presence of $(\text{NH}_4)_2\text{SO}_4$ 5' cleavage by transesterification is the preferred reaction mode. In addition, exon reopening does not occur under these reaction conditions.

In contrast to the results of the self-splicing reactions obtained with BS/CBS2i5'0 and BS/CBS2i5'+1, the preRNA of BS/CBS2i5'-1 is almost unreactive under both assay conditions.

Figure 5. Characterization of cryptic 3' cleavage sites of mutant RNAs with d6* insertions. RNA bands corresponding in size to (i) free 3' exons resulting from cryptic 3' cleavages and subsequent exon reopening and, (ii) ligated exons, also resulting from aberrant 3' cleavages, were analysed by reverse transcription with the 3' primer in the absence (for determination of the 5' termini) or presence (for sequencing) of dideoxynucleotides. The left autoradiograph in the upper part of the figure shows the results of reverse transcription of free 3' exons. lane a, BS/S5HC5; lane b, BS/S5HB6, lanes c1 and c2, BS/S5HA6. The 5' ends are indicated by arrows, UG's downstream of cryptic cleavages are boxed. As a size marker, the sequence of BS/S5HB6 DNA obtained by the chain termination reaction with the same 3' primer was coelectrophoresed. The right autoradiograph shows the junction sequence of ligated exons, the arrow indicates the 5' E-3' E* junction. The lower part of the figure shows the secondary structure models of domains 5, 6* and 6 of the wildtype and mutants. The arrowhead marks the normal 3' splice site, cryptic 3' cleavages are marked by arrows.

Virtually no bands corresponding to ligated exons are appearing. Besides products generated by unspecific hydrolytic degradation, only trace amounts of lariats and, in buffer C, linear intron molecules can be detected. The simplest explanation for this observation could be that for this mutant RNA the IBS1:EBS1 pairing is no longer stable enough to allow initiation of the splicing reaction ($\Delta G = -8.8$ kcal/mol versus $\Delta G = -11.0$ kcal/mol for an undisturbed IBS1:EBS1 pairing). This observation would be in agreement with the recent findings of Jacquier and Michel (11). They could show for an intron $\alpha 5c$ harbouring preRNA, that the combination of an IBS2 deletion and an IBS1 double mutation, which partially disrupted the IBS1:EBS1 pairing led to a complete loss of reactivity of the preRNA. On the other hand, we cannot exclude the possibility that so far unknown interactions constrain the last IBS1 nucleotide being an A residue (in the case of b11).

Branchpoint-dependent 5' cleavage assists accurate splice site selection

As shown above, the reactivity of BS/CBS2i5' - 1 preRNA was very poor, whereas the overall efficiency of the BS/CBS2i5' + 1 self-splicing reaction was almost indistinguishable from BS/CBS2i5'O, despite the additional A residue separating the end of the IBS1:EBS1 pairing from the 5' guanosine of the intron. Whereas the A insertion obviously had no influence on the efficiency of hydrolytic 5' cleavage, the question arises as to where cleavage occurred in this mutant, i.e. was it directed by the 3' end of the normal IBS1 sequence GACAGA or by the 5' terminus of the intron. Cleavage positions were determined by assaying the 5' ends of excised linear introns (obtained upon incubation of preRNAs in buffer C = hydrolytic cleavage mode) and lariats (incubation was in buffer S = branch point dependent

cleavage by transesterification) of BS/CBS2i5' + 1 and the other mutant-RNAs.

Results of primer extensions are shown in Fig.4. Both linear introns and lariats of BS/CBS2i5'O and BS/CBS25' - 1 preRNA show the normal 5' ends, i.e. hydrolytic cleavage and branch point dependent cleavage occurred immediately upstream of the G residue of the intron (Fig.4A,B). Results obtained with BS/CBS2i5' + 1 preRNA are different. Whereas cleavage by transesterification generates the correct intron 5' end, the hydrolytic reaction mode leads to ambiguous cleavages. About 70% of 5' cuts follow the normal IBS1 sequence, however the other 30% are 5' of the intron guanosine (Fig.4C). These findings let us suggest that for the branch point dependent 5' cleavage mode of group II introns, which should reflect the physiologically relevant pathway, the highly conserved G residue of the 5' intron end plays a crucial role in defining the splice site.

3' exon sequences contribute to accuracy of 3' cleavage

The role of 3' exon sequences in the forward reaction was investigated by two different approaches.

1.) We designed a modified version of intron bI1 by inserting different hairpins between domains 5 and 6 (Fig.2). Incubation of preRNAs derived from these clones led to a complete loss of intron reactivity under branch-point dependent reaction conditions indicating that a functional domain 6 does not tolerate major deviations from its conserved distance from domain 5 and that domain 6* does not represent a surrogate for authentic domain 6 (data not shown).

However, with incubation in buffer C, these altered intron versions resume their capability for 5' cleavage, since in the presence of Cl^- ions 5' cleavage can occur by hydrolysis and exhibits complete independence from a functional domain 6 (18,19). As already mentioned, this reaction mode leads to excision of the intron in linear form and—to some extent—reopening of ligated exons. Primer extension analyses of ligated exons and 3' exons (resulting from reopening of joined exons) are shown in Fig.5. Interestingly, these intron forms no longer catalyse attacks at the authentic 3' splice sites, but exhibit an amazing flexibility in selecting new, cryptic 3' splice sites. These cleavages proceed now efficiently 3' to the new hairpin d6* (BS/S5HB6, BS/S5HC6) or within it (BS/S5HA6) at various distances from domain 5 in the sequence AG↓UG and CU↓UG, respectively. i.e., cryptic 3' cleavages are nonrandom and only observed 5' to the dinucleotide UG regardless of its base paired status. Since this dinucleotide corresponds to the first two bases of the normal 3' exon following bI1 we conclude that at least the first one or two bases of the 3' exon help to direct the 3' OH group of the 5' exon for proper 3' cleavage.

Not all NNUG motifs in the new sequence could be recognized as cryptic 3' splice sites. One reasonable interpretation could be that within the three-dimensional structure of the ribozyme the spatial orientation of these UG's does not allow the attack by the 5' exon. Interestingly, cleavages at the original 3' splice site were no longer observed. Therefore it is likely that as a prerequisite for recognition, a critical distance from domain 5 has to be preserved.

2.) The impressive efficiency of the forward reaction with BS/CBS2i5'O and BS/CBS2i5' + 1 preRNAs lacking authentic 3' exons suggests that 3' exon sequences do not contribute to the efficiency of the reaction, as already suggested in a previous work (16). To test whether 3' splice site recognition is affected in these constructs (= combination of the wild type version of

Figure 6. Determination of the junction sequence of ligated 5'- and 3' parts of BS/CBS2i5'O. After the self-splicing reaction of BS/CBS2i5'O RNA the RNA band corresponding in size to ligated exons (= 5'- and 3' portions of the original CBS2 RNA) was eluted from the gel and analysed by reverse transcription with the 3' primer in the presence of dideoxynucleotides. The junction sequence is shown on the left, the arrow indicates the ligation junction. The sequence of the 5' end of 3' CBS2 is given in italics.

bI1 with foreign 3' exon sequences) we analysed the sequence of the ligated exons generated by the self-splicing reaction of mutant-RNA CBS2i5'O under standard reaction conditions by RNA sequencing with reverse transcriptase with a 3' exon primer. As shown in Fig.6, ligated 5'- and 3' parts of CBS2 gave the sequence expected for correct joining, the ligation junction reads ...GACAGAAAAAGU..., i.e. if 3' cleavage was occasionally incorrect it would be below the level of detection by this approach. From these observations we conclude that in the normal forward reaction 3' exon sequences are less important for accuracy as long as the additional intronic recognition elements are intact.

DISCUSSION

Selection of the 5' splice site

It could be convincingly shown that the EBS1+2:IBS1+2 interaction of group II introns is necessary for both, 5'- and 3' cleavage (11). Whereas none of the mutations that weaken the EBS2:IBS2 base pairing has major effects on the efficiency of 5' cleavage, those alterations which reduce the stability of the EBS1:IBS1 interaction exhibit a complete abolishment or, at least, a drastic reduction of 5' cleavage.

As yet, there was no experimental proof for the role of the EBS1:IBS1 helix in the accuracy of 5' cleavage. Whereas for group I introns obviously the 5' splice site is defined by the last U residue of the 5' exon, which is almost always base paired with a G residue of the IGS, it was generally believed that for group II introns the end of the EBS1:IBS1 helix defines the point of cleavage (see, for review ref. 13,20). Yet, there are exceptions to this rule. Among 70 group II introns sequenced so far (compiled in ref.20) there are several examples where the EBS1:IBS1 interaction either does not reach up to the 5' splice site or could extend beyond it into the intron, i.e. the end of the IBS1:EBS1 helix does not unambiguously define the 5' junction, e.g.

for bI1 (IBS1)..GACAGA|gugaga..
 (EBS1)..c u g u c u ua..
 or, for Cl.ndh1 (IBS1)..CUUAUU|gugcg..
 (EBS1)..g a g u a a cgu..

(Cl.ndh1 is the first intron of the gene encoding subunit 1 of the NADH-dehydrogenase in *Citrullus lanatus* mitochondria, IBS1 sequences are in capital letters).

Our data confirm that the IBS1:EBS1 base pairing is necessary for activation of the 5' junction but they also indicate that this interaction alone is not sufficient for unequivocal definition of the cleavage site. We conclude that for the branch point dependent 5' cleavage mode the 5' G of the intron, which belongs to the most conserved characteristics of group II introns, helps to position the cleavage site precisely.

It has been argued that in subgroup IIB introns the symmetrical location of the EBS1 element within the hairpin loop of subdomain D3 could prevent the extension of the IBS1 pairing with EBS1 beyond the 3' end of the exon into intronic sequences by a possible steric hindrance (20)(cf. Fig.1A). Thus, for bI1 the IBS1:EBS1 pairing would be restricted to the 6 nt helix G-ACAGA:cugucu and, with this scenario, the last Watson-Crick pairing of the helix would define the 5' junction in the wildtype preRNA. In accordance with this model, hydrolytic and branch point dependent cleavages of preRNAs from mutants BS/CBS2i5'0 and BS/CBS2i5'-1 are directed by the last Watson-Crick base pair at the end of the IBS1:EBS1 helix which is shortened by one base pair for BS/CBS2i5'-1.

Perhaps a more interesting effect is represented by mutant BS/CBS2i5'+1. The presence of an additional adenosine following IBS1 leads to ambiguous hydrolytic cleavages (Fig.4C). Following the idea that hydrolytic 5' cleavages might be directed by the end of the IBS1:EBS1 pairing, this could reflect an instability of the additional A.U base pair. In contrast, branch point dependent cleavage seems to be positioned by the 5' G residue of the intron and not by the ambiguous end of IBS1:EBS1, thus ensuring that the excised intron obtains a proper 5' terminus.

In the case of incorrect transposition of the intron at a suitable IBS1 motif, this reaction mode would help to preserve the integrity of the 5' terminus of the RNA-transposon after reexcision (in contrast to the IBS1:EBS1-directed 5' cleavage). Furthermore, the host-RNA would also benefit from this mechanism, since it would avoid the generation of deletions or insertions of nucleotide(s) at the target site after the disappearance of the intron.

Taken together, we conclude that the 5' G of the intron is an important element for the choice of the splice site but, by no means, the only one. We do not rule out the possibility that additional structural features, like the ϵ/ϵ' interaction mentioned above and the overall architecture of the intron molecule could support proper 5' splice site selection.

3' splice site selection

In contrast to group I introns, little is as yet known about structural elements defining the 3' splice site of group II introns and theoretical criteria which could define this site are listed in the introduction. Our data imply that (i) recognition of the original 3' splice site requires a distinct distance of domain 6 from domain 5, (ii) artificial hairpins inserted between domain 5 and 6 cannot functionally substitute for the original domain 6, but lead to

aberrant cleavages at variable distances downstream from domain 5 (ranging from 10 to 26 nt), (iii) these cleavages are always 5' to a dinucleotide corresponding to the original 3' exon terminus and, (iv) as long as structural elements at the 3' end of the intron are intact 3' exon sequences are dispensable for splice site selection in the normal forward reaction.

Thus, the recognition of the group II intron 3' splice site is likely to be accomplished by at least two structural elements. These are (i) the first 3' exon nucleotides that are not sufficient by themselves but may assist correct positioning and, (ii) the structure of domain 6 and its appropriate distance and/or the appropriate spatial orientation toward the catalytic center of the ribozyme. The latter findings are consistent with previously published data (21).

On the basis of comparative sequence data, Jaquier and Michel suggested for group II introns a 'guide' mechanism similar to group I (11). They noted that for most of subgroup IIA introns the first 3' exon nt can base pair with the nt immediately 5' to EBS1. However, this model cannot be directly applied for bI1 and many other subgroup IIB introns. Furthermore, disruption of this interaction by replacing the first A nucleotide of the 3' exon of aI5c by a C had no measurable effect on the reaction (11).

The observation that the UG dinucleotide following the cryptic 3' cleavages of the d6* constructs is always involved in a stable canonical pairing within the hairpin could also argue against a guide mechanism for bI1, since the long-range pairing of the UG with bases adjacent to EBS1 would require the disruption of the existing local base-interaction. In conclusion, our results support the hypothesis that 3' exon sequences participate in conferring specificity of 3' cleavage, by which kind of mechanism, however, remains to be elucidated.

Although our data suggest that the only structural motif required for intron transposition is a cryptic IBS1 motif, it is obvious that the IBS1:EBS1 base pairing alone is not sufficient to define exactly the position of integration. Results obtained by determination of the insertion site either within original exons or within a foreign RNA let us suggest that 3' exon sequences could also participate in selecting the 3' splice site in the reverse 3' cleavage (=the first step of transposition). One could argue that the inexact integration into CBS2-RNA is not due to the lack of authentic 3' exon nucleotides but induced by the presence of an A immediately downstream of the cryptic IBS1 that could allow extension of the IBS1:EBS1 base pairing (cf. Fig.4). Yet, this interpretation would only explain occasional +1 shifts. It would not be in accordance with the observation of the -1 shifts found for integration into CBS2 (3 examples) but not for integration into the original ligated exons. Nevertheless, we recognize that the importance of 3' exon sequences is not directly shown by our studies and further experiments will be required to elucidate the exact role of sequences downstream of IBS1.

Unfortunately, predictions on the structural elements important for splice site recognition are so far restricted to the two-dimensional model of group II introns and it is obvious that a detailed understanding of the mechanism of catalytic RNAs requires data about the spatial structure. Furthermore, it is possible that precise splice site selection of group I and group II introns is not only achieved by intramolecular RNA-RNA interactions but also by the help of proteins *in vivo*. A major area of future investigations will thus be not only to study the effect of further manipulations of RNA structure on splice site selection but also to investigate the function of protein factors in group II intron splicing *in vivo*.

ACKNOWLEDGEMENTS

This work was supported by a grant from the Deutsche Forschungsgemeinschaft (SFB 184, project A9). We thank Ana Maria Merlos, Gerhard Rödel and Udo Schmidt for helpful discussions and careful reading of the manuscript. Special thanks to Rebecca Hix for looking over the English. We also thank Barbara Gelhaus for her expert technical assistance.

REFERENCES

1. Cech, T.R. (1986). *Sci.Am.* 254, 76–84.
2. 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.
3. Schmelzer, C., and Schweyen, R.J. (1986). *Cell* 46, 557–565.
4. 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.
5. Kück, U., Godehardt, I., and Schmidt, U. (1990). *Nucl.Acids Res.* 18, 2691–2697.
6. Schmidt, U., Riederer, B., Mörl, M., Schmelzer, C., and Stahl, U. (1990). *EMBO J.* 9, 2289–2298.
7. Been, M.D., and Cech, T.R. (1988). *Science* 239, 1412–1416.
8. Waring, R.B., Towner, P., Minter, S.J., Davies, R.W. (1986) *Nature* 321, 133–139.
9. Barford, E.T., Cech, T.R. (1989). *Mol.Cell.Biol.* 9, 3657–3666.
10. Doudna, J.A., Cormack, B.P., and Szostak, J.W. (1989). *Proc.Natl.Acad.Sci.USA* 86, 7402–7406.
11. Jacquier, A., and Michel, F. (1987). *Cell* 50, 17–23.
12. Jacquier, A., and Michel, F. (1990). *J.Mol.Biol.* 213, 437–447.
13. Cech, T.R. (1990). *Annu.Rev.Biochem.* 59, 543–568.
14. Jarrell, K.A., Peebles, C.L., Dietrich, R., Romiti, S.L., and Perlman, P.S. (1988a). *J.Biol.Chem.* 263, 3432–3439.
15. Müller, M.W., Schweyen, R.J., and Schmelzer, C. (1988). *Nucl. Acids Res.* 16, 7383–7395.
16. Mörl, M. and Schmelzer, C. (1990a). *Cell* 60, 629–636.
17. Mörl, M., and Schmelzer, C. (1990b). *Nucl. Acids Res.* 18, 6545–6551.
18. Jarrell, K.A., Dietrich, R., and Perlman, P.S. (1988b). *Mol.Cell.Biol.* 8, 2361–2366.
19. Altura, R., Rymond, B., Seraphin, B., and Rosbash, M. (1989). *Nucl. Acids Res.* 17, 335–354.
20. Michel, F., Umesono, K., and Ozeki, H. (1989). *Gene* 82, 5–30.
21. Schmelzer, C., and Müller, M.W. (1987). *Cell* 51, 753–762.