The EMBO Journal Vol.16 No.4 pp.779–792, 1997

The role of branchpoint-3′ splice site spacing and interaction between intron terminal nucleotides in 3′ splice site selection in *Saccharomyces cerevisiae*

B.G.Mattias Luukkonen[1] and Bertrand Séraphin[2]

European Molecular Biology Laboratory, Meyerhofstrasse 1, D-69117 Heidelberg, Germany

[1]On leave from the Microbiology and Tumorbiology Center, Karolinska Institute, Stockholm, Sweden

[2]Corresponding author

A conserved 3′ splice site YAG is essential for the second step of pre-mRNA splicing but no *trans*-acting factor recognizing this sequence has been found. A direct, non-Watson–Crick interaction between the intron terminal nucleotides was suggested to affect YAG selection. The mechanism of YAG recognition was proposed to involve 5′ to 3′ scanning originating from the branchpoint or the polypyrimidine tract. We have constructed a yeast intron harbouring two closely spaced 3′ splice sites. Preferential selection of a wild-type site over mutant ones indicated that the two sites are competing. For two identical sequences, the proximal site is selected. As previously observed, an A at the first intron nucleotide spliced most efficiently with a 3′ splice site UAC. In this context, UAA or UAU were also more efficient 3′ splice sites than UAG and competed more efficiently than the wild-type sequence with a 3′ splice site UAC. We observed that a U at the first intron nucleotide is used for splicing in combination with 3′ splice sites UAG, UAA or UAU. Our data indicate that the 3′ splice site is not primarily selected through an interaction with the first intron nucleotide. Selection of the 3′ splice site depends critically on its distance from the branchpoint but does not occur by a simple leaky scanning mechanism.

Keywords: pre-mRNA processing/snRNPs/splice site selection/spliceosome/yeast

Introduction

Nuclear pre-mRNA splicing is a process conserved in all eukaryotes from yeast to man. Although differences in intron size and abundance, as well as in splicing regulation, are apparent between species, the fundamental components and chemistry of the splicing reaction remain unaltered. Splicing requires five small nuclear ribonucleoproteins (snRNPs), U1, U2, U4, U5 and U6, consisting of a small nuclear RNA (snRNA) and several protein components. In addition, numerous non-snRNP proteins are also required for splicing. The snRNPs and protein cofactors assemble onto the pre-mRNA in a stepwise, ATP-dependent manner to form the spliceosome, in which the chemical splicing reaction will take place (Green, 1991; Guthrie, 1991; Moore *et al.*, 1993). Removal of the introns occurs through two sequential transesterification reactions. The first step is initiated by a nucleophilic attack from the 2′ OH of the branchpoint adenosine residue on the phosphodiester bond at the 5′ splice site. This results in exon 1 carrying a 3′ terminal OH group, and a lariat–exon 2 intermediate where the 5′ end of the intron is linked to the branch nucleotide through a 5′–2′ phosphodiester bond. In the second transesterification step, the 3′ OH of exon 1 attacks the phosphodiester bond at the 3′ splice site, leading to exon ligation and release of the intron lariat.

Accurate removal of introns from pre-mRNAs is mediated through recognition of short, conserved sequence elements at the intron boundaries. During the formation of the spliceosome, the 5′ end of U1 snRNA is base paired with the 5′ splice site (Zhuang and Weiner, 1986; Séraphin *et al.*, 1988; Siliciano and Guthrie, 1988). This interaction commits the pre-mRNA to the splicing pathway (Séraphin and Rosbash, 1989), but does not specify the phosphodiester bond to be hydrolysed (Séraphin and Rosbash, 1990). U2 snRNA associates with the branchpoint sequence (Parker *et al.*, 1987; Wu and Manley, 1989; Zhuang and Weiner, 1989) forming an imperfect helix from which the branchpoint adenosine is probably bulged (Query *et al.*, 1994). Before the first chemical step of splicing, U1 snRNA base pairing with the 5′ splice site is displaced and substituted by interactions with U5 and U6 snRNAs. Genetic and biochemical analyses suggest that the invariant U5 snRNA loop interacts with exon 1 sequences just upstream of the cleavage site both before and after the first step (Newman and Norman, 1991, 1992; Wyatt *et al.*, 1992; Sontheimer and Steitz, 1993). This interaction is not essential for the first splicing step, but might be involved in keeping the exon 1 intermediate near the active site for step 2 (O'Keefe *et al.*, 1996). UV cross-linking experiments suggested that U6 snRNA might interact with the 5′ end of the intron (Sawa and Abelson, 1992; Sawa and Shimura, 1992; Wassarman and Steitz, 1992). Genetic analysis in *Saccharomyces cerevisiae* demonstrated that U6 snRNA base pairs with intron position +4, +5 and +6, juxtaposing the phosphodiester bond to be hydrolysed with the branchpoint adenosine (Kandels-Lewis and Séraphin, 1993; Lesser and Guthrie, 1993).

While many of the interactions specifying 5′ splice site cleavage have been elucidated, very little is still known about how the 3′ splice site (consensus sequence YAG) is recognized (Umen and Guthrie, 1995c). In *Schizosaccharomyces pombe*, the 3′ splice site YAG is required prior to the first step of splicing, while *S.cerevisiae* introns do not require a 3′ splice site to proceed through step 1 (Rymond and Rosbash, 1985; Parker and Patterson, 1987). Both YAG-dependent and YAG-independent introns are found in mammalian genes (Reed, 1989). In *S.pombe*, the 5′ end of U1 snRNA can base pair with the 3′ AG as well as with the 5′ splice site, and mutations in the AG

dinucleotide can be suppressed by the complementary U1 snRNA mutation. The suppression is partial as only the first step block conferred by the 3′ splice site mutation is rescued, but splicing does not proceed through step 2 (Reich *et al.*, 1992). Thus, the U1 snRNA–3′ splice site pairing is likely to be involved only in an early step in spliceosome assembly, when the intron boundaries are being defined. The U1 snRNA–3′ splice site interaction is not essential in *S.cerevisiae*, and the corresponding conserved U1 snRNA residues base pair with the 3′ nucleotides of the 5′ exon (Séraphin and Kandels-Lewis, 1993). From *in vitro* studies using a substrate with multiple 3′ splice sites, it has been suggested that the YAG is recognized through a scanning mechanism originating from the branchpoint (Smith *et al.*, 1989, 1993). However, in *S.cerevisiae* insertion of a pyrimidine-rich stretch between two 3′ splice sites can promote splicing to the distal YAG, at the expense of the proximal site (Patterson and Guthrie, 1991).

Genetic screens in *S.cerevisiae* have identified a number of protein factors specifically involved in the second splicing step. The SLU7 protein was cloned in a synthetic lethality screen with a U5 snRNA mutation, and appears to be essential for utilizing splice sites distal to the branchpoint (Frank and Guthrie, 1992). The 220 kDa U5-associated protein PRP8 can be cross-linked to both the 5′ splice site and the 3′ splice site (Whittaker and Beggs, 1991; Teigelkamp *et al.*, 1995a,b; Umen and Guthrie, 1995b), and has been shown to be involved in polypyrimidine tract recognition (Umen and Guthrie, 1995a). Recently, an allele of PRP8 was identified that can suppress mutations in the YAG sequence (Umen and Guthrie, 1996). Other factors that affect the second splicing step include PRP16, PRP17 (SLU4), PRP18 and SSF1 (Couto *et al.*, 1987; Vijayraghavan *et al.*, 1989; Burgess *et al.*, 1990; Vijayraghavan and Abelson, 1990; Ansari and Schwer, 1995).

A number of mutations in the U2, U5 and U6 snRNAs that affect the second step have been described. Akin to its role at the 5′ splice site, the U5 snRNA loop interacts with exon 2 sequences just downstream of the 3′ splice site (Newman and Norman, 1992; Sontheimer and Steitz, 1993). It is likely that these interactions play a role in juxtaposing the two exons for ligation. Two mutations have been found that suppress mutations in the YAG sequence to some extent: U2–A25G and U6–G52U (Lesser and Guthrie, 1993; Madhani and Guthrie, 1994). The U2–A25G transition can unspecifically suppress all single mutations in the AG dinucleotide (Madhani and Guthrie, 1994). Other substitutions at, or a deletion of, U2 position 25 do not affect the second step. U6 position 52 is part of the highly conserved ACAGAGA heptanucleotide, and the G–U mutation can unspecifically suppress all changes at the last position of the intron, as well as an A–C change at the penultimate position (Lesser and Guthrie, 1993). Other changes in U6 snRNA at position 52, and all changes at position 51, lead to a complete second step block *in vivo* and *in vitro* (Madhani *et al.*, 1990; Vankan *et al.*, 1992; Datta and Weiner, 1993; Yu *et al.*, 1993; Wolff *et al.*, 1994). Interestingly, Madhani and Guthrie (1994) found genetic evidence for a tertiary interaction between U6–G52 and U2–A25 required for the second step.

A non-Watson–Crick interaction between intron terminal nucleotides has been suggested to play a role in 3′ splice site selection (Parker and Siliciano, 1993; Chanfreau *et al.*, 1994; Scadden and Smith, 1995). A G to A mutation in the first base of the *S.cerevisiae* ACT1 intron leads to a strong second step block, which can be partially suppressed by a G to C change at the last position (Parker and Siliciano, 1993). The same 5′ splice site substitution provokes activation of cryptic 3′ splice sites with the sequence UAC or UAA in the RP51 intron (Chanfreau *et al.*, 1994), and it was found as a suppressor of a G to A substitution in the last nucleotide of the fourth intron of the DHFR gene (Carothers *et al.*, 1993). Suppression of this mutant has also been observed *in vitro* in a HeLa cell extract using a mammalian intron (Scadden and Smith, 1995).

Our laboratory has previously demonstrated that the U1 snRNA was not implicated in 3′ splice recognition in yeast (Séraphin and Kandels-Lewis, 1993). These results suggested that the conserved YAG sequence was recognized by another mechanism. To address this question, we analysed the splicing behaviour of a sensitive reporter intron harbouring two closely spaced 3′ splice sites in direct competition. Our results, presented below, led us to conclude that a direct interaction between intron terminal nucleotides is not the major determinant in 3′ splice site selection. Instead, we show here that the distance from the branchpoint nucleotide is a critical parameter for 3′ splice site activation. Our results also allowed us to exclude a simple leaky scanning mechanism of 3′ splice site selection.

Results

Construction of an intron with two 3′ splice sites in direct competition

Competing splice sites have been extensively used to study the splicing process, since small effects of mutations in *cis*-acting sequences and *trans*-acting factors can be readily detected using such sensitive assays. Because the sequence of the non-conserved nucleotides flanking a 3′ splice site can influence its usage (Patterson and Guthrie, 1991; Newman and Norman, 1992), we constructed two consensus sites in similar sequence context only five nucleotides apart (Figure 1A; Materials and methods). Henceforth, the first site will be referred to as ‘proximal’ and the second site as ‘distal’, based on their position relative to the branchpoint. The RP51-derived intron has a consensus 5′ splice site and branchpoint, followed by a uracil-rich polypyrimidine tract and the two 3′ splice sites (for details, see Materials and methods). The intron is fused to a LacZ reporter gene, so that the pre-mRNA and mRNA resulting from splicing to the proximal site are out of frame for β-galactosidase translation, while the mRNA resulting from splicing to the distal site is in frame for translation. As premature termination codons can influence RNA splicing and stability (Maquat, 1995; Aoufouchi *et al.*, 1996), several constructs were made with the LacZ ORF fused to the proximal 3′ splice site. With both sets of constructs we observed similar splicing patterns (data not shown). In every case, β-galactosidase assays were performed in parallel to the RNA analyses. These enzymatic assays supported the results of the

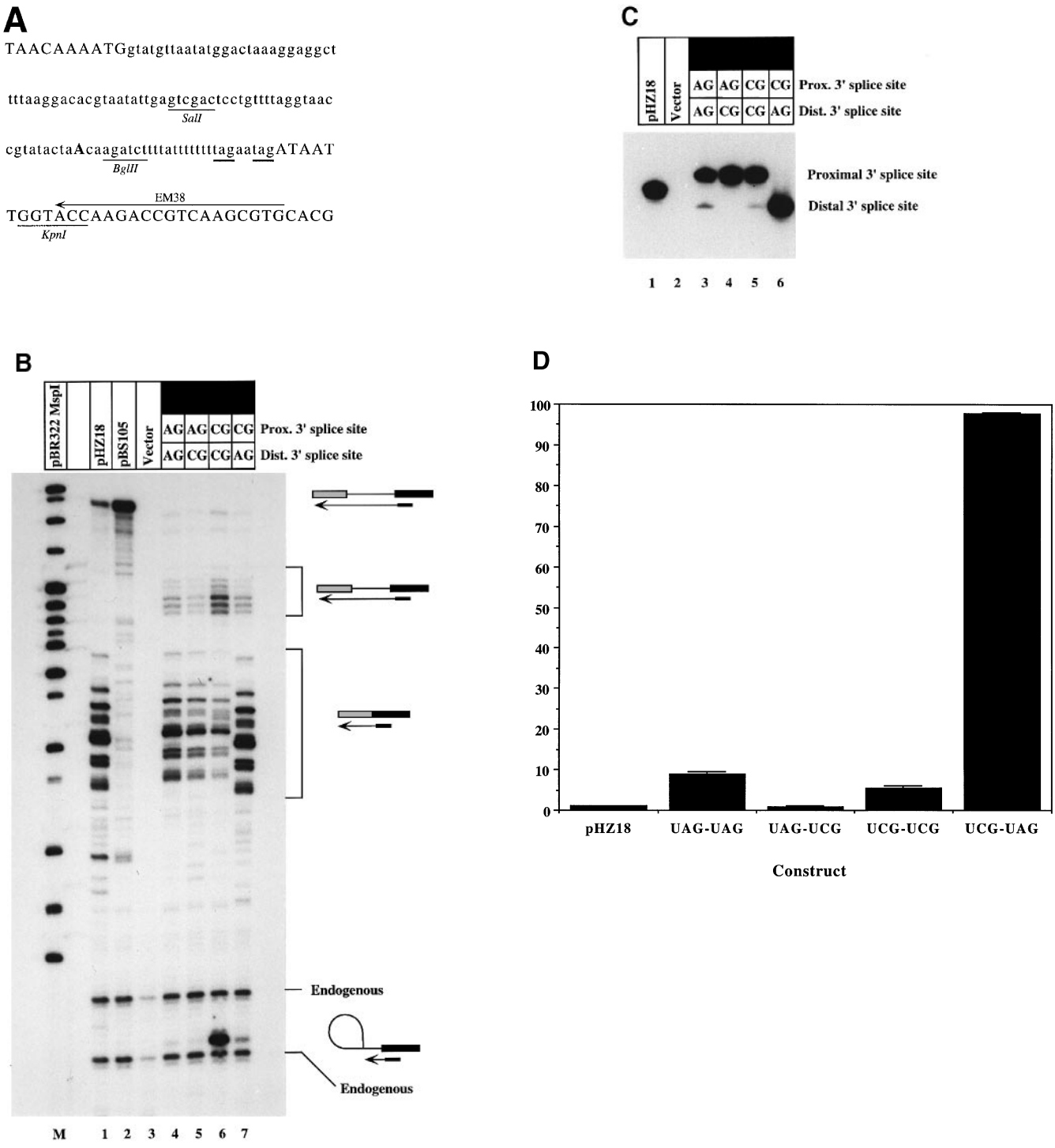

Fig. 1. Construction of an RP51-derived intron with two synthetic 3′ splice sites in direct competition. (**A**) Complete sequence of the RP51-derived *S.cerevisiae* intron used in this study (for details on construction, see Materials and methods). Exon sequences are shown in upper-case letters, intron sequences in lower-case letters. The adenosine residue used for branching is shown in upper-case bold face and the two 3′ splice sites are underlined. Restriction sites used for cloning are indicated, as is the binding site for the exon 2-specific EM38 oligonucleotide used for primer extension analysis. (**B**) Primer extension analysis of introns harbouring a point mutation changing the penultimate adenosine nucleotide to cytosine. Primer extension analysis was performed as previously described (Pikielny and Rosbash, 1985), using a ^{32}P end-labelled oligonucleotide EM38 priming in exon 2 (see Materials and methods). Owing to multiple transcription initiation sites of the inducible GAL-CYC1 promoter, pre-mRNA and mRNAs appear as multiple bands. The lower levels of mRNA in lane 5 resulted from lower induction in this experiment. (Note further that the ratio of mRNA to pre-mRNA was similar to that observed in lane 4.) The different extension products are labelled on the right side of the figure. The product resulting from extension of wild-type RP51 pre-mRNA is longer than that of the reporter constructs, therefore two pre-mRNAs of different sizes are indicated. Endogenous = endogenous rp51 mRNA. (**C**) RT–PCR analysis of RNA samples shown in (B), using exon 1- and exon 2-specific primers (see Materials and methods). Amplification of pre-mRNA and plasmid DNA using this set of primers yields identical products, and only amplified mRNAs are shown. Splicing to the proximal 3′ splice site gives rise to an amplified product of 97 nucleotides, while splicing to the distal site gives rise to a product of 92 nucleotides. Amplification of the wild-type RP51 mRNA yields a product of 95 nucleotides. (**D**) Quantitation of the RT–PCR gel shown in (C), using a PhosphoImager (Molecular Dynamics). The graph displays per cent splicing to the distal site for the different constructs. The mean values and standard deviations from three different transformants are shown.

direct analyses of the pre-mRNA and mRNA levels (data not shown).

We first investigated how efficiently our reporter intron was spliced by primer extension using the exon 2-specific primer EM38 (Figure 1A). Plasmids pHZ18 (Teem and Rosbash, 1983) and pBS105 (Séraphin and Rosbash, 1990) harbouring, respectively, the wild-type RP51 intron and a mutant derivative thereof were used as representatives of efficiently and poorly spliced introns, while RNA derived from a strain transformed with the vector alone was used to ascertain that the signals detected originated from the reporter gene. The results depicted in Figure 1B demonstrate that the reporter intron was spliced with an efficiency similar to the wild-type RP51 intron (Figure 1B, compare lanes 1 and 4). The mRNA and pre-mRNA extension products appear as multiple bands upon separation on polyacrylamide gels. These correspond to the multiple transcriptional initiation sites of the galactose-inducible GAL-CYC1 promoter driving our reporter construct. However, by comparing the size of the major extension product with those observed for the wild-type RP51 intron, we could conclude that the proximal site was preferentially used.

Next, we tested whether both 3′ splice sites could be used. For this purpose, we introduced single A to C mutations in the penultimate position of one or both splice sites (changing the sequence from UAG to UCG), and assayed splicing by primer extension using the exon 2-specific primer. Mutations in either splice site alone did not result in any significant loss of splicing efficiency (Figure 1B, lanes 5 and 7). When the proximal site was mutated, the resulting mRNAs were significantly shorter, consistent with the activation of the distal site (Figure 1B, lane 7). When both sites were mutated, a second step block was clearly apparent from the accumulation of lariat intermediate, accompanied by a slight accumulation of pre-mRNA (Figure 1B, lane 6). Nevertheless, a substantial amount of mRNA could be detected (Figure 1B, lane 6). We conclude that the modified RP51 intron is efficiently spliced when a wild-type 3′ splice site is present, with the absence of an AG resulting in inhibition of splicing mainly because of a second step block.

The primer extension analysis presented above allowed us to measure the levels of pre-mRNA, mRNA and lariat intermediates. Since the GAL-CYC1 promoter has multiple initiation sites resulting in heterogeneous mRNA products, it is difficult to assess to what extent each 3′ acceptor site is used because of the overlapping signals. Therefore, we established a sensitive RT–PCR assay to determine accurately and quantitatively which of the different acceptors was used. The RNAs were subjected to cDNA synthesis using the exon 2-specific EM38 primer, followed by PCR amplification using ^{32}P-labelled EM38 primer and the exon 1-specific oligonucleotide EM84. These primers also amplify plasmid DNA, giving a band of identical size to that of amplified pre-mRNA, and precluding the use of this assay for quantitation of pre-mRNAs. Consequently, only amplification products corresponding to the mRNA species are shown. Control experiments demonstrated that amplification of mRNAs was in the linear range of the PCR reaction (data not shown). With the PCR assay, we detected splicing to both sites when both were wild type or both were mutated (Figure 1C, lanes 1 and 3). In those cases, splicing occurred preferentially to the proximal site. Splicing occurred exclusively to the wild-type site if in competition with a UCG (Figure 1C, lanes 2 and 4). Quantitation of the RT–PCR gel using a PhosphoImager (Molecular Dynamics) revealed that 8.9% splicing occurred to the distal site when both sites were wild type, and 5.4% when both sites were mutated (Figure 1D). In contrast, no splicing above background levels was detected to a UCG site when in competition with a UAG (Figure 1D, lanes 1, 3 and 5). We conclude that the two 3′ splice sites are in direct competition and that no outside cryptic splice sites were activated. When the two sites have the same sequence, the proximal site is preferentially chosen, in concordance with previous observations (Smith *et al.*, 1989, 1993; Scadden and Smith, 1995). However, these data do not support a leaky scanning model for 3′ splice site selection (see Discussion).

Selection of the 3′ splice site in a G1A mutant intron is position but not sequence specific

Following the original observation by Parker and Siliciano (1993) that a G to C mutation in the last nucleotide of the *S.cerevisiae* ACT1 intron could partially, but specifically, suppress a G to A change at the 5′ splice site, it has been suggested that an interaction between intron boundaries is an important determinant in 3′ splice site selection (Chanfreau *et al.*, 1994; Umen and Guthrie, 1995c). If this presumption is true, a mutation at the 5′ splice site would affect selection between two competing 3′ splice sites with different sequences. To test this hypothesis, we made a series of mutants in the proximal and distal 3′ splice site, in combination with mutations at the 5′ splice site. The proximal site was mutated to UAC, UAA and UAU, while the competing distal site was either UAG or UAC. All combinations of mutations at the 3′ splice site were inserted downstream of either the wild-type 5′ splice site or the G1A mutant 5′ splice site (Figure 2A). The constructs were introduced in yeast, and splicing analysed by primer extension (Figure 2B). The wild-type 5′ splice site spliced efficiently to a wild-type 3′ splice site when present, and very low levels of pre-mRNA and exon 2–lariat intermediate accumulated (Figure 2B, lanes 3–7). As described above, the proximal site of a duplicated pair was preferentially selected. In contrast to the results obtained in previous studies (Parker and Siliciano, 1993; Chanfreau *et al.*, 1994), in the absence of a wild-type 3′ splice site a substantial amount of mRNA was still detected (Figure 2B, lanes 8–10). Low levels of pre-mRNA accumulated, and the predominant block was at the second step, as shown by the accumulation of high levels of lariat intermediate (Figure 2B, lanes 8–10). The G1A mutation also led to a strong second step block and a weak first step block (Figure 2B, lanes 11–18). Again, in contrast to previous studies, high levels of mRNA were detected, but a second step block was readily apparent because high levels of lariat intermediate accumulated. Indeed, the levels of mRNA produced by the G1A mutants were similar to those produced by constructs with the wild-type 5′ splice site in the absence of a UAG (Figure 2B, compare lanes 8–10 with lanes 11–18). In contrast to previous reports (Parker and Siliciano, 1993; Chanfreau *et al.*, 1994), no dramatic increase in mRNA levels was seen with a UAC at the 3′ splice site (Figure 2B, lanes 11–18; see also below).

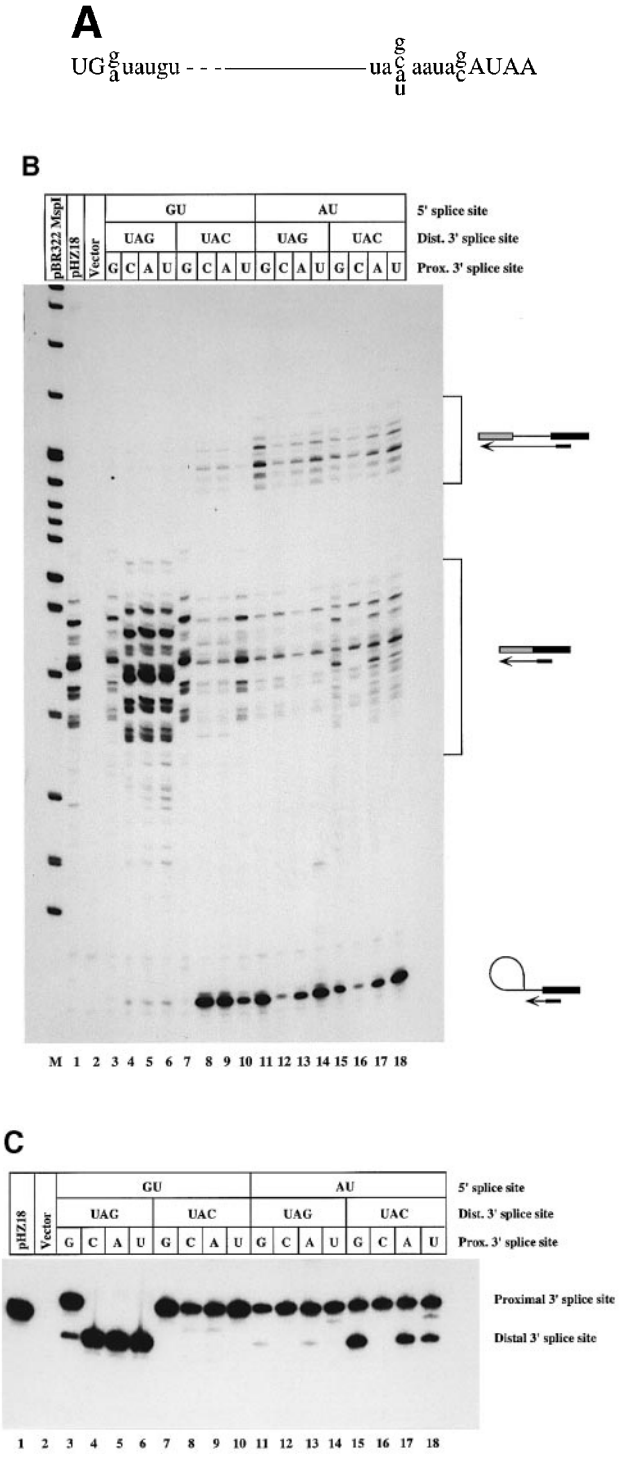

Fig. 2. Splicing of constructs harbouring point mutations at the first and last intron nucleotides. (**A**) Schematic illustration of mutations at the 5′ and 3′ splice sites. (**B**) Primer extension analysis of constructs shown in (A), using the exon 2-specific EM38 primer. The different extension products are labelled on the right. (**C**) RT–PCR analysis of the RNAs shown in (B), using exon 1- and exon 2-specific primers.

However, the levels of exon 2–lariat intermediate are somewhat lower when the proximal site is UAC, suggesting that the second step is slightly more efficient with this 3′ splice site sequence, but only when it is in a branchpoint-proximal position (Figure 2B, lanes 12 and 16).

Several splice acceptors were being used with varying efficiency, and to separate the different mRNA species we next performed an RT–PCR analysis of the RNAs (Figure 2C). The PCR amplification demonstrated that with a wild-type 5′ splice site, a wild-type 3′ splice site was exclusively chosen. In the absence of a wild-type 3′ splice site, splicing to a UAC, UAA or UAU proximal 3′ splice site was detected (Figure 2C, lanes 3–10). The levels of mRNA observed (Figure 2B and C, lanes 7–10) suggested that in this situation UAG is by far the most preferred 3′ splice site, followed by UAU, while UAA and UAC were of equally low strength. When the G1A mutant introns were analysed, the G1A 5′ splice site also spliced to UAG, UAC, UAA and UAU 3′ splice sites in the proximal position (Figure 2C, lanes 11–18). With a duplicated pair of UAC sites, the proximal one was preferentially selected (Figure 2C, lane 16). Low levels of splicing to the distal UAG were detected in combination with UAG and UAA at the proximal 3′ splice site (Figure 2C, lanes 11 and 13). UAC in a distal position was a stronger competitor than the wild-type UAG (Figure 2C, lanes 11–18). Surprisingly, a distal UAC could not entirely out-compete UAA, UAU and even the wild-type UAG at the proximal position. With the UAU–UAG and UAU–UAC combination of 3′ splice sites, the PCR analysis also revealed a spliced product migrating between the proximal and distal mRNAs, which corresponded in size to splicing to a cryptic UAA two nucleotides downstream of the proximal site (Figure 2C, lanes 14 and 18). From this analysis, it appears that with the G1A mutant, UAC was the preferred 3′ splice site, but followed closely by UAU and UAA, with UAG being the weakest site.

In summary, when a wild-type 5′ splice site was present, selection of a 3′ splice site was highly dependent on its sequence (UAG >>> UAU > UAA and UAC), but not on its position. In contrast, with the G1A 5′ splice site, splicing occurred mostly at the proximal site while the sequence of that site had only a minor effect on the selection process. There is apparently no strong sequence specificity for 3′ splice site selection when A is the first intron nucleotide as not only UAC, but also UAA and UAU, are more efficiently selected than the wild-type UAG sequence.

A U at the first position of the intron is used for splicing with various 3′ splice sites

Splicing can occur to most nucleotides at the 3′ splice site using a wild-type or G1A 5′ splice site. Splicing of G1A mutated introns proceeded with little specificity regarding the identity of the last nucleotide of the intron. It was then enticing to ask whether any nucleotide at the 5′ splice site could be used for splicing, in combination with any nucleotide at the 3′ splice site. In previous studies, G1U and G1C were shown to be used for the first but not the second step of splicing (Parker and Siliciano, 1993; Chanfreau *et al.*, 1994). We introduced the 5′ splice site mutants G1U and G1C in our intron reporter, and tested for splicing to a UAG, UAC, UAA or UAU at the proximal position, in competition with a UAG at the distal position.

To investigate what phosphodiester bond at the 5′ splice site was cleaved with our construct, the RNAs were analysed by primer extension using an intron-specific primer (Figure 3A). The extension reaction yielded two products: a longer product from extension of the unspliced pre-mRNA, and a shorter product ending at the 5′ cleavage site originating from the exon 2–lariat intermediate and intron lariat. The latter product appeared as a doublet as the reverse tran-

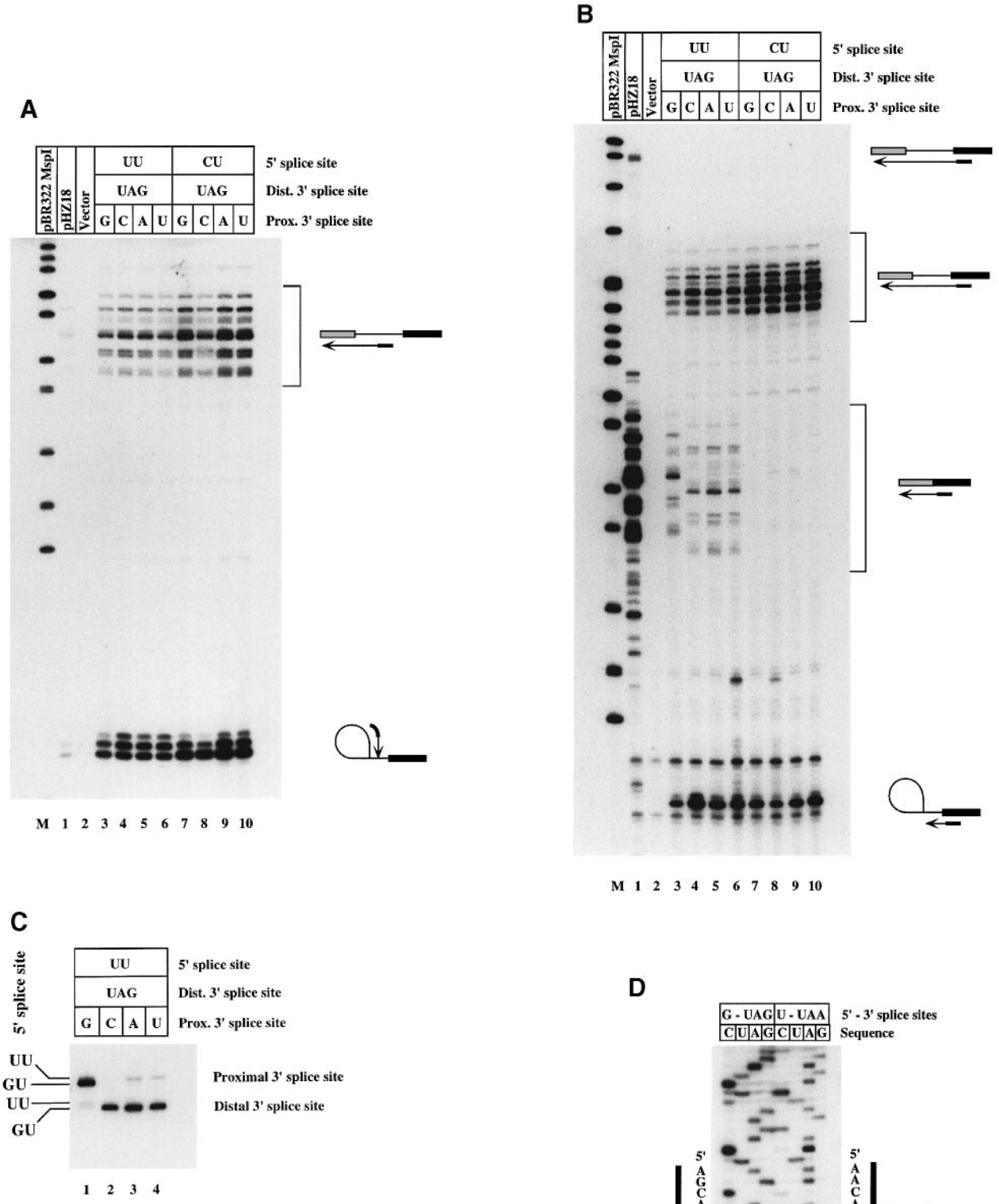

Fig. 3. The G1U mutation is competent for both steps of splicing. (**A**) Primer extension analysis of RNAs harbouring the G1U and G1C mutations, using the intron-specific RB27 primer (Jacquier *et al.*, 1985) (see Materials and methods). The products are identified on the right side of the figure. (**B**) Primer extension analysis of the RNAs shown in (A), using the exon 2-specific EM38 primer. Extension products are labelled on the right side of the figure. (**C**) RT–PCR analysis of the G1U-derived RNAs shown in (A) and (B), using exon 1- and exon 2-specific primers EM38 and EM84. (**D**) Direct cycle sequencing of two mRNAs from G1U introns (proximal 3′ splice site UAA, Figure 3C, lane 3). Right sequence shows the upper band, left sequence the lower. All mRNAs derived from the G1U mutated introns were sequenced, and the data are summarized in Table I.

Table I. Sequence of mRNAs generated with the G1U mutation at the 5′ splice site

Proximal 3′ splice site	Band No.	mRNA sequence	Splice sites used	
			5′	3′
UAG	1	AAAAUG/AAUAGAUAA	+1 UU	Prox. UAG
	2	AAAAU/AAUAGAUAA	–1 GU	Prox. UAG
	3	AAAAUG/AUAA	+1 UU	Dist. UAG
	4	AAAAU/AUAA	–1 GU	Dist. UAG
UAC	1	AAAAU/AUAA	–1 GU	Dist. UAG
UAA	1	AAAAUG/AAUAGAUAA	+1 UU	Prox. UAA
	2	AAAAU/AUAA	–1 GU	Dist. UAG
UAU	1	AAAAUG/AAUAGAUAA	+1 UU	Prox. UAU
	2	AAAAU/AUAA	–1 GU	Dist. UAG

For each construct, increasing numbers correspond to mRNA species of increasing mobility. Slashes indicate the splice site.

scriptase occasionally adds one non-templated random nucleotide to the full extension product (Jacquier *et al.*, 1985; see Figure 3A, lane 1, for the wild-type RP51 intron). The G1U mutated introns displayed three bands corresponding to two overlapping doublets (Figure 3A, lanes 3–6). The lower doublet originated from cleavage at the normal 5′ splice site (UU), while the upper doublet originated from cleavage at the adjacent 5′ GU sequence created by the G to U substitution (Figure 1A). Usage of similar cryptic sites has been previously reported (Aebi *et al.*, 1987; Chanfreau *et al.*, 1994). Introns harbouring the G1C mutant also displayed three bands at the same positions (Figure 3A, lanes 7–12). We conclude that with both G1U and G1C the normal site and a cryptic splice site one nucleotide upstream are used for the first splicing step. Primer extension analysis with an exon 2 primer revealed the accumulation of exon 2–lariat intermediates for both the G1U and G1C mutants (Figure 3B). No mature mRNAs were detected with the G1C mutation, whereas low levels of spliced products were detected in the G1U mutated introns (Figure 3B). Previous studies failed to detect any splicing from a U residue at the 5′ splice site and indicated that all mature RNA species originated from the usage of the cryptic 5′ splice site (Chanfreau *et al.*, 1994). The size of the spliced products correlated with usage of the wild-type 3′ splice site (Figure 3B, lanes 3–6), but did not allow us to determine accurately which 5′ splice site was used. To map precisely which 5′ and 3′ splice sites were used with our reporter, we performed an RT–PCR analysis with the same RNAs (Figure 3C). Consistent with the result of the primer extension analysis above, no mRNAs were reproducibly detected from PCR amplification of G1C mutated RNAs (data not shown). With the G1U mutant and two wild-type 3′ splice sites, four mRNAs were detected as two closely migrating doublets (Figure 3C, lane 1). The upper doublet corresponded to usage of the proximal 3′ splice site, while the lower doublet corresponded to splicing at the distal 3′ splice site. The two bands of each doublet were one nucleotide apart, as judged by migration next to a sequencing ladder (not shown). The size of the lower band corresponded to splice products originating from the aberrant 5′ splice site while, surprisingly, the upper band suggested splicing from the UU at the normal 5′ splice site. To confirm these unanticipated results, PCR products were eluted from the gel, re-amplified and subjected to cycle sequencing. The sequenced confirmed that a UU 5′ splice site could splice to a 3′ UAG acceptor (Table I).

By comparing the constructs carrying the various 3′ splice site combinations, it appeared that when the cryptic 5′ splice site GU was used, a UAG 3′ splice site was preferentially selected. When both competing 3′ splice sites had identical sequence, the upstream one was preferred (Figure 3C, lane 1, see above). This is similar to the results obtained with a GU at the normal 5′ splice site (see above). We observed PCR amplification products corresponding to splicing from the UU present at the normal 5′ splice site to a UAG, UAA and UAU, but not a UAC, at the 3′ splice site (Figure 3C, lanes 1–4). Splicing to these various 3′ splice site sequences occurred with similar efficiency. Again, the sequence of the PCR products determined by cycle sequencing after elution of the bands from the gel and re-amplification confirmed the prediction of the RT–PCR analysis (Figure 3D; Table I).

Our analysis demonstrates that a U can be used as a 5′ splice site. This has not been reported previously. As with the G1A mutant (see above), we observe little sequence specificity in the selection of a 3′ splice site in combination with the UU 5′ splice site.

Branchpoint proximity is an important determinant in 3′ splice site selection

A number of observations in the previous experiments suggested that the location of a 3′ splice site is more important than its sequence for its activation. (i) When two identical sites were present, the proximal one was preferentially used. (ii) With the G1A 5′ splice site, splicing to a proximal UAC was more efficient and less lariat–exon 2 intermediate accumulated than when UAC was in a distal position (Figure 2B, lanes 12 and 16). (iii) Similarly, with the 5′ splice site, G1A, UAA and UAU could partially compete with a UAC site if they were located in a proximal, but not in a distal position (Figure 2B and data not shown). A similar observation was made by Chanfreau *et al.* (1994) in the context of the wild-type RP51 intron, as mutation of the natural 3′ splice site to UAC activated a proximal UAA sequence that is otherwise silent.

In all the cases stated above, the proximal position had the sequence UAN and it is possible that the UA dinucleotide may be partially recognized as a 3′ splice site, and thereby impede recognition of or splicing to downstream sites. To test this possibility, we changed the proximal site to CCC, making this stretch divergent from a consensus 3′ splice site (Figure 4A). An alternative explanation for the poor activation of the downstream site might be that it is located too far from the polypyrimidine tract. Thus, we changed the proximal site to UUU, and the two following residues also

to uracils, extending the polypyrimidine tract to the distal site (Figure 4A). The distal site was either UAG or UAC, and these constructs were appended to either a wild-type or G1A 5′ splice site. Splicing was analysed by primer extension using the exon 2 primer (Figure 4B). With the wild-type 5′ splice site, splicing was again very efficient to a wild-type 3′ splice site in the presence of either the UUU or the CCC substitution, but undetectable to a UAC 3′ splice site. With a UAC at the 3′ splice site, the UUU construct triggered activation of a cryptic CAU two nucleotides downstream of the distal site (Figure 4B, lane 8, see below). All constructs harbouring a G1A 5′ splice site accumulated lariat–exon 2 intermediate to a large extent, even with the extended polypyrimidine tract (Figure 4B, lanes 9–14). Furthermore, the polypyrimidine tract extension or the CCC substitution did not significantly enhance splicing to the distal UAC (Figure 4B, compare lanes 12 and 14). To map precisely which sites were used and to quantify the usage of each site, we next analysed the RNAs by RT–PCR (Figure 4C). With the exception of the aforementioned cryptic CAU downstream from the distal site, the PCR analysis did not reveal any other cryptic 3′ splice site (Figure 4C, lane 8). However, as reported above, a UAG could be activated by the G1A mutation when in a proximal position (Figure 4C, lanes 9 and 10). Low levels of splicing of the 5′ A to the UAG at the downstream site could be detected when the upstream site was UAG or with the UUU substitution (Figure 4C, lanes 9 and 13). Therefore, splicing to the downstream site was not more efficient when it was put close to the polypyrimidine tract, or when the proximal site was destroyed.

We conclude that location of the 3′ splice site has a major effect on its selection. Since a mutant downstream site was not efficiently utilized when the proximal site was destroyed, or even when closely juxtaposed with the polypyrimidine tract, our results suggest that it is the distance between the branchpoint and the 3′ splice site that might be critical for its activation.

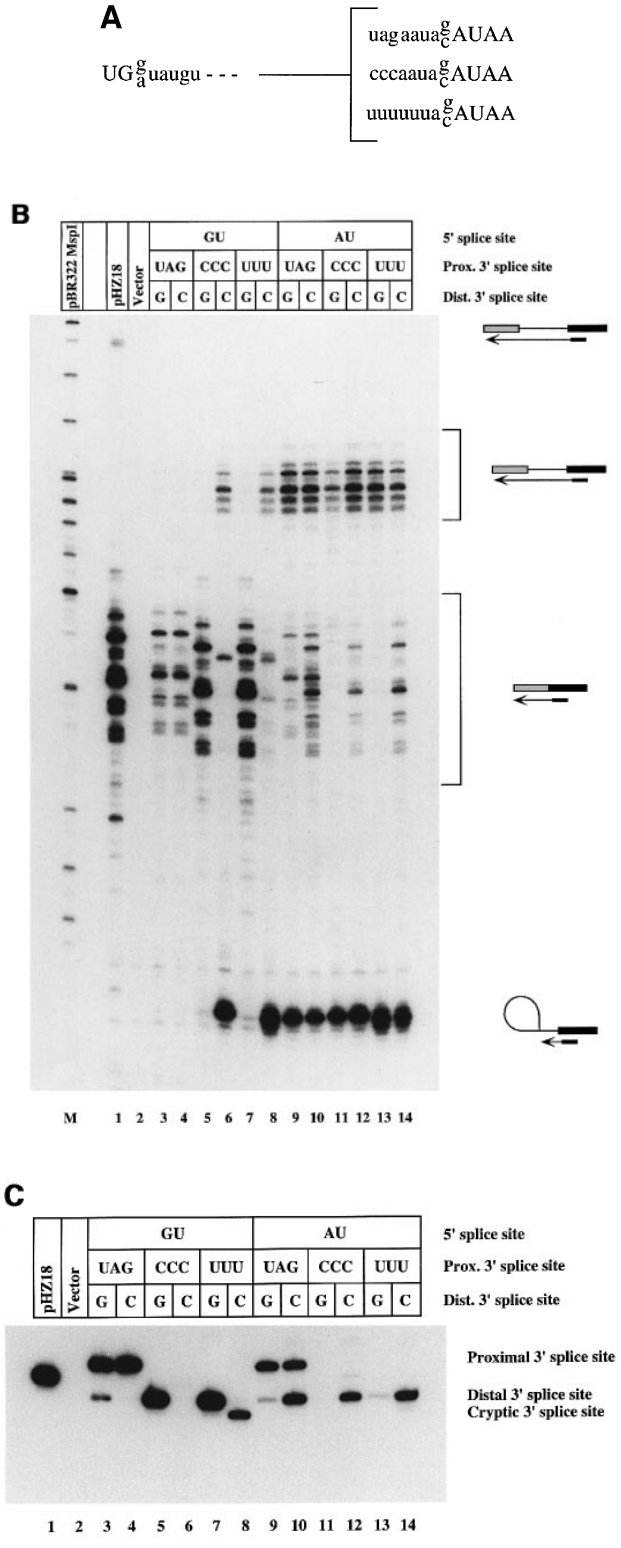

Fig. 4. Selection of a 3′ splice site is dependent on distance from the branchpoint. (**A**) Schematic illustration of mutations at the 5′ and 3′ splice sites. (**B**) Primer extension analysis of introns with a mutated proximal 3′ splice site, using an exon 2-specific primer. Extended products are shown to the right of the figure. In lanes 6 and 8, an extension product appears that migrates slower than the major mRNA species. This product does not display the multiple bands characteristic of a GAL-CYC1 transcript, it only appears when high levels of lariat intermediate accumulate, and it is not amplified in the RT–PCR assay. From these observations, and its apparent size in a denaturing polyacrylamide gel, we deduce that it corresponds to extension of debranched but undegraded lariat intermediate. (**C**) RT–PCR analysis of the RNAs shown in (B), using exon 1- and exon 2-specific primers.

Optimal activation of a wild-type or mutant 3′ splice site occurs between 18 and 22 nucleotides from the branchpoint

As it appeared that the distance between the 3′ splice site and the branchpoint might be important for efficient 3′ splice site selection, we next sought to determine what the optimal distance was. To this end, we constructed plasmids carrying repeats of UAG, AAG or UUG trinucleotides, spaced by UAU triplets, inserted between the branchpoint and distal 3′ splice site (Figure 5A). The UAG, AAG and UUG sequences were selected to cover a wide range of 3′ splice site efficiencies. UAG is a canonical 3′ splice site sequence that is efficiently used, while AAG, which is only rarely found at the end of yeast introns (Kalogeropoulos, 1995), is probably a weaker 3′ splice site sequence. The UUG sequence is not a natural 3′ splice site. The constructs were made in two versions where the second version (b) is frame shifted by three nucleotides with respect to the first one (a) (Figure 5A), making it possible to measure the distance effect to an accuracy of three nucleotides in a window from 10 to 28 nucleotides downstream of the

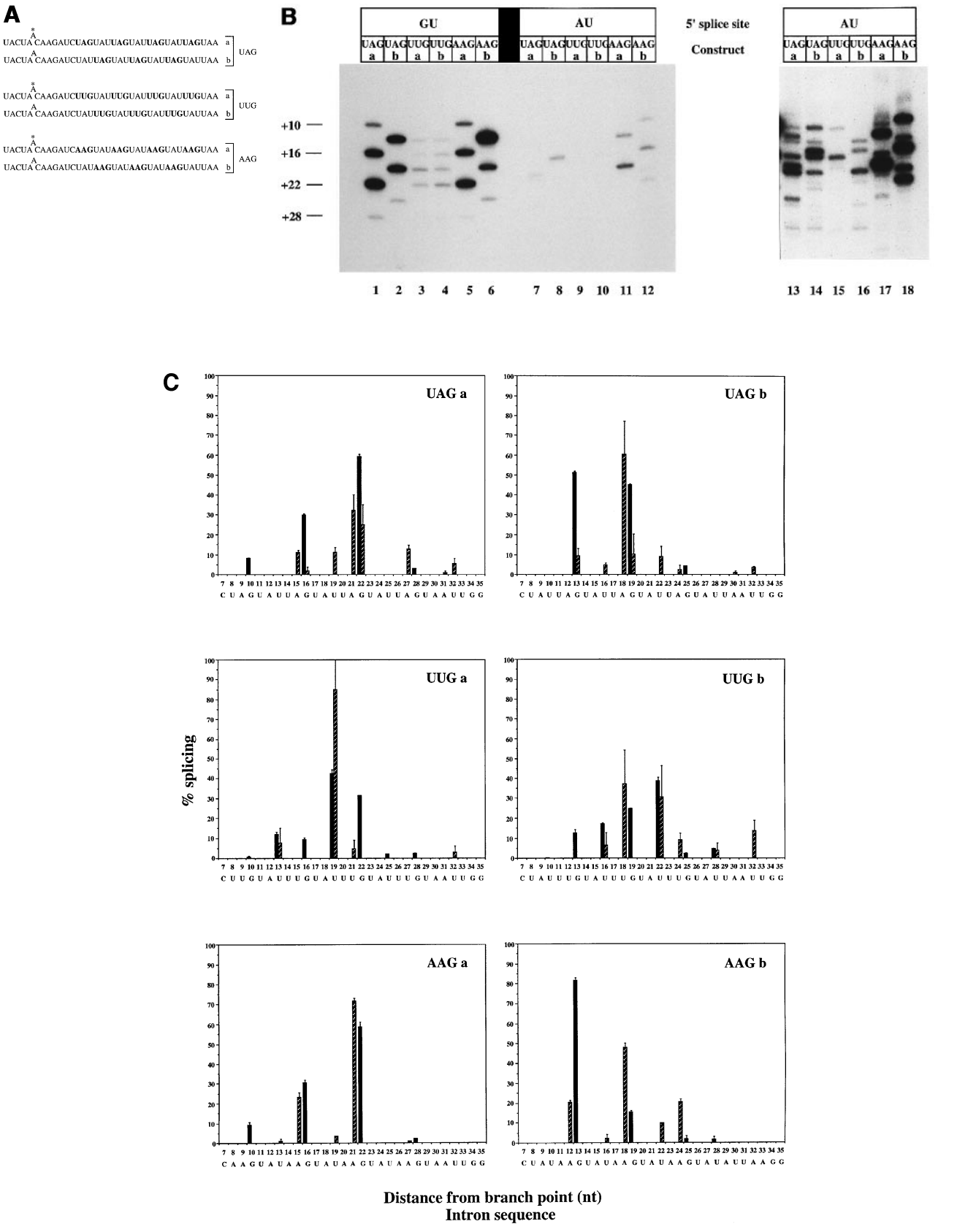

Fig. 5. Distance requirements for selecting a wild-type or cryptic 3′ splice site. (**A**) Schematic illustration of the constructs harbouring multiple 3′ splice sites. An asterisk indicates branchpoint adenosine. (**B**) RT–PCR analysis of splicing pattern of the RNAs shown in (A), using exon 1- and exon 2-specific primers. Lanes 13–18: 72 h exposure of lanes 7–12 (G1A). (**C**) PhosphoImager quantitation of the gels shown in (B). The graphs display per cent splicing as a function of distance from the branchpoint adenosine. Full bars indicate splicing of constructs harbouring a wild-type 3′ splice site, striped bars indicate splicing of constructs harbouring the G1A 5′ splice site. The mean values and standard deviations from two independent transformants are shown.

branchpoint. These constructs have an altered polypyrimidine tract but, consistent with the poor conservation of the polypyrimidine tract in yeast, this does not affect their splicing efficiency (Rymond *et al.*, 1987; our unpublished observations).

The repeated 3′ splice site sequences were inserted behind either a wild-type or G1A 5′ splice site, and splice site selection was monitored by RT–PCR (Figure 5B). With the wild-type 5′ splice site, efficient splicing to a UAG or AAG was detected (Figure 5B, lanes 1, 2, 5 and 6). The 3′ splice site at position +10 (relative to the branchpoint adenosine) was weakly activated. This is consistent with a previous observation in a metazoan system (Smith *et al.*, 1993). In contrast, the AG at position +13 was efficiently used (Figure 5B, compare lanes 1 with 2 and 5 with 6). Both the UAG and AAG constructs utilized at least three different acceptors, demonstrating that the proximal AGs were too close to the branchpoint to be used efficiently. Interestingly, splicing to the AAG was as efficient as to the UAG, suggesting that the pyrimidine of the conserved YAG may be redundant for 3′ splice site selection when this site is at an optimal distance from the branchpoint. This is further supported by the natural occurrence of AAG at the 3′ end of yeast introns (Kalogeropoulos, 1995).

For the UAG and AAG constructs version a, the strongest acceptor was at position 22, with the site at position +16 being only slightly weaker (Figure 5B, lanes 1 and 5), while the optimal region for 3′ acceptor selection in the b version was at position +13, followed by position +19 (Figure 5B, lanes 2 and 6). One would have anticipated that the peak positions for 3′ splice site selection in the two versions would be separated by three nucleotides. However, it is possible that some upstream sites (e.g. site +10 in version a), while being too close to the branchpoint for efficient cleavage, would interfere sterically with a site located immediately downstream (e.g. site +16 in version a) by binding some splicing factor. This would redirect splicing to the next available site (i.e. site +22 in version a), as has been previously observed (Smith *et al.*, 1993). Alternatively, packing of the RNA by hnRNP proteins might favour one site while masking another site of identical sequence.

Splicing was significantly reduced when the 3′ splice site was changed to a UUG (Figure 5B, lanes 3 and 4). Interestingly, either of two 3′ splice site sequences were chosen in this construct: the UUG or the spacer UAU triplet. These sites were selected with almost identical efficiency between positions +13 and +22 (Figure 5B, lanes 3 and 4). This result suggests that the two last nucleotides of the intron are equally important for 3′ splice site selection. In contrast, with the UAG and AAG constructs, the UAU site was not activated suggesting that recognition of the AG is more important for 3′ splice site selection than recognition of the preceding pyrimidine nucleotide. The RT–PCR was quantified, and the data plotted as percentage splicing as a function of distance from the branchpoint (Figure 5C). Quantitation revealed that optimal splicing occurs between 13 and 22 nucleotides from the branchpoint adenosine, irrespectively of whether a consensus or cryptic 3′ splice site is activated. A primer extension reaction using ddGTP performed in parallel demonstrated that the bands observed did not result from mispriming of PCR products on repeated sequences and confirmed the quantitation data obtained using RT–PCR (not shown).

RT–PCR analysis of the G1A constructs revealed a complex splicing pattern (Figure 5B, lanes 7–12). Compared with the construct harbouring a wild-type 5′ splice site, splicing was substantially decreased (Figure 5B, compare lanes 7–12 with lanes 1–6). Upon prolonged exposure of the film, multiple bands were visible in each lane (Figure 5B, lanes 13–18). Three major splice acceptors were being used: UAA, UUA and UAU, the UAA being most efficient (long exposure, lanes 17 and 18). Quantitation of the gel showed that efficient splice site activation occurred in a more narrow window, between 18 and 22 nucleotides from the branchpoint (Figure 5C). Interestingly, all constructs lacking an AG dinucleotide were spliced significantly less efficiently than when an AG was present, both with a wild-type and G1A 5′ splice site (Figure 5B, lanes 3 and 4, and long exposure lanes 15 and 16). It is possible that recognition of an AG before step 2 affects splicing efficiency, even though it is eventually not used as a splice acceptor.

In summary, the usage of constructs containing tandemly repeated 3′ splice sites demonstrate that those located too close or too far from the branchpoint are not used. With both the wild-type 5′ splice site and G1A mutant, we observe that the most efficient activation of a 3′ splice site occurred between 13 and 22 nucleotides from the branchpoint.

Discussion

Interaction between intron terminal nucleotides and 3′ splice site selection

The analyses of substitution mutants in the 3′ terminal AG of nuclear introns demonstrate that this sequence is the primary signal for selection of the 3′ splice site. These nucleotides must therefore be recognized by some *trans*- and/or *cis*-acting factor(s). It is even likely that, like the nucleotides of the 5′ splice site, the 3′ splice site sequence is recognized multiple times both before and during catalysis of the second step of splicing. Mutants of U2 and U6 snRNA and of PRP8 that suppress mutations of the 3′ AG have been described. However, the suppressions are not sequence specific, suggesting that these factors do not interact directly with the 3′ splice site sequence. As many factors interacting with the 5′ splice site have been identified, it is surprising that no *trans*-acting factors specifically recognizing the 3′ splice site and affecting its selection have yet been described.

It has been proposed that selection of the 3′ splice site is mediated through a direct interaction of the intron 3′ G with the first intron nucleotide. This conclusion came from the observation that a mutation at the 5′ splice site suppressed a substitution at the 3′ splice site. As this effect was neither intron nor species specific, it has been suggested to be a major event in 3′ splice site recognition (Parker and Siliciano, 1993; Chanfreau *et al.*, 1994; Scadden and Smith, 1995). Our data, in part, duplicate the results of these previous analyses, in this case using a sensitive reporter harbouring two competing 3′ splice sites. However, the results presented here examine a more extensive collection of 5′ and 3′ splice site mutants, and the sum of our data leads us to question whether there is a direct interaction between the first and last intron nucleotides. The original proposal for such an interaction was based on the fact that (i) a G at the 3′ splice site was only efficiently used when a G was present

at the 5′ splice site, and (ii) a 3′ splice site mutation from AG to AC was suppressed by a G to A substitution at the 5′ splice site (Parker and Siliciano, 1993; Chanfreau *et al.*, 1994; Scadden and Smith, 1995). Our results do not contradict these original observations. Indeed, we also observe that the G–G combination is optimal and a change in any of the nucleotides results in a severe second step block. Furthermore, we also observe that, with an A at the 5′ splice site, the best 3′ splice site sequence is UAC. However, we also observe that UAA and UAU are better 3′ splice sites than UAG under these conditions. The suppression of an A at the 3′ splice site by an A at the 5′ splice site has also been reported by others (Carothers *et al.*, 1993; Parker and Siliciano, 1993; Chanfreau *et al.*, 1994). With our reporter, we also observed splicing, albeit inefficient, from a U at the 5′ splice site to a 3′ G, A and U (but not C) with similar efficiencies. Splicing from a 5′ U had not been reported in previous studies. Given the minimal, or lack of, specificity in the 3′ splice site sequence used with either A or U at the 5′ splice site, our results do not support the hypothesis that a direct interaction between the first and last intron nucleotides is involved in 3′ splice site selection. Alternative models appear as likely. For example, these results might reflect independent but cooperative interactions of *trans*-acting factors with the first and last intron nucleotides during the second splicing step. Alternatively, these results could be anticipated if some nucleotide combinations lead to steric clashes between the first and last intron nucleotides. Mechanisms facilitating a G to G splicing at the expense of other G–X or X–G (X = A, U or C) combinations could have been selected during evolution as a proofreading step ensuring accurate splicing. However, as described above, such a mechanism would not require a direct interaction between the intron ends.

This issue appears even more complicated as the 3′ splice site sequence is likely to be recognized several times during the splicing reaction. In *S.pombe* and human systems, it has been shown the YAG sequence is required before the first splicing step, and that is recognized by the U1 snRNA, at least in *S.pombe* (Reich *et al.*, 1992). This early recognition appears to be absent in *S.cerevisiae* (Rymond and Rosbash, 1985; Séraphin and Kandels-Lewis, 1993). We observe that two 3′ splice sites located five nucleotides apart 22 nucleotides downstream of the branchpoint (i.e. well within the range for yeast introns; Kalogeropoulos, 1995) are in direct competition. Indeed, with a wild-type intron 5′ end, mutation of one 3′ splice site redirects the splicing machinery to the second site. The situation is somewhat different when the 5′ splice site is mutated to an A or a U. In that case, the proximal site of the competing pair is preferentially selected, in a way that is not strongly dependent on its sequence (see Results). These results would not be expected if a direct interaction between the first and the last intron nucleotide was a major step in 3′ splice site selection. The results do not completely rule out a direct interaction between the intron ends, as this interaction could occur at a late step in 3′ splice site activation (e.g. during catalysis) following an initial recognition of the YAG sequence by another factor. If this is the case, this would still support our conclusion that a 5′–3′ splice site interaction is not the primary determinant in 3′ splice site selection.

A minor class of pre-mRNA introns harbouring the A–C combination (AT–AC introns) at the intron boundaries is found in higher eukaryotes (Jackson, 1991; Hall and Padgett, 1994). These introns require the low-abundance U11 and U12 snRNPs, rather than the U1 and U2 snRNPs (Hall and Padgett, 1996; Tarn and Steitz, 1996). Interestingly, even though these introns are removed in a spliceosome containing at least some diverging components, the combination of intron terminal nucleotides is A–C. Although this observation has also been taken to support an interaction between the first and last nucleotide of the intron (because it is the best mutant–mutant combination in the major type of spliceosome), this does not provide evidence for a direct interaction. Indeed, this could again reflect independent but cooperative interactions of *trans*-acting factors with intron ends. Alternatively, there might be strict steric limitations to the possible variation of the nucleotides near the active site of the spliceosomes.

In summary, from the available evidence, we conclude that a direct interaction between the 5′ and 3′ intron nucleotides is not the primary determinant of 3′ splice site selection. Although we cannot definitively rule out that such a direct interaction occurs after initial recognition of the YAG sequence by another factor, the lack of sequence specificity in 3′ splice site selection with either A or U at the 5′ splice site rather supports an indirect interaction between the first and last intron nucleotides.

Evidence against a leaky scanning model for 3′ splice site selection

Using duplicated 3′ splice sites in direct competition, we found that the branchpoint-proximal site of an identical pair was almost exclusively used (Figure 1C and D). This was previously observed in metazoan systems (Fu *et al.*, 1988; Smith *et al.*, 1989, 1993; Scadden and Smith, 1995). It has been suggested that the low level of splicing detected to the distal site results from a 'leaky scanning' phenomenon (Smith *et al.*, 1993). The leaky scanning model proposes that the splicing machinery scans the pre-mRNA downstream of the branchpoint/polypyrimidine tract and selects the first suitable sequence as the 3′ splice site. However, each site can be overlooked with a given frequency ('leakage'), allowing the splicing machinery to scan to the next site downstream. According to this model, UAG is well recognized in most cases (>90% in our case), preventing the splicing machinery from reaching the downstream site. In contrast, a UCG in a proximal position is poorly recognized and often overlooked. Therefore, this model explains the efficient use of a wild-type splice site located downstream of a mutant site. Even though a mutant site can be only very poorly recognized, it might be efficiently used because of multiple rounds of scanning.

Our data contradict a simple leaky scanning model for 3′ splice site selection (Figure 1C and D). When both splice acceptors harboured a single point mutation at the A residue, the proximal to distal ratio was not changed compared with an intron with two wild-type splice acceptors. For the UCG–UCG construct, the poor recognition of the upstream (and downstream) sites should shift the ratio of proximal to distal 3′ splice usage towards 50:50 (see Materials and methods for a mathematical treatment). The two sites should now be 'randomly' selected. Because the proximal to distal ratio is nearly identical between the UAG–UAG and UCG–UCG constructs, we conclude that a simple leaky scanning mechanism cannot operate for 3′ splice site selection. One comes

to a similar conclusion when the splice site selection between repeated UAG trinucleotides and weaker AAG trinucleotides is compared (Figure 5B). The two splicing patterns are almost indistinguishable and there is no evidence for an increase in leaky scanning. Therefore, our data demonstrate that 3′ splice site selection does not occur through a simple leaky scanning mechanism.

Mechanisms for 3′ splice site selection

Our data rule out a simple leaky scanning mechanism for 3′ splice site selection. At the same time, they demonstrate a strong effect on 3′ splice site selection of both the position of one possible 3′ splice site relative to another one, and on the distance from the branchpoint to the 3′ splice site. How can these observations be reconciled? In a first model, the 3′ splice site sequence would first be selected through a scanning event by the splicing machinery. This first selection would define a region for 3′ splice site cleavage, but not determine the phosphodiester bond to be cleaved. The site originally selected would in most cases also be selected for cleavage in a second recognition step. However, when a strong site is present nearby, it could compete for binding to the catalytic site. The situation would, therefore, be similar to selection of the 5′ splice site where U1 snRNA first binds to the 5′ splice site sequence. In most cases, the site defined by this early recognition, which is in an optimal position, is subsequently selected by the U6 and U5 snRNAs. However, in some situations, cryptic sequences can compete with the original site for U6 and U5 snRNA binding, and are therefore selected for use in splicing (Séraphin and Rosbash, 1990; Kandels-Lewis and Séraphin, 1993; Lesser and Guthrie, 1993). This model would explain: (i) why a proximal site is preferentially selected; (ii) why a canonical sequence is preferred independently of its position; (iii) why we observe an identical ratio of proximal to distal site usage with either wild-type or mutant sequences. To accommodate our data showing that sites are preferentially selected in a window of 13–22 nucleotides downstream of the branchpoint, and that in this window multiple sites are selected, one would imagine that the start site for scanning is selected in this window through a random mechanism (i.e. random collision of the scanning factor with nucleotides in this window). Our results suggest that the location of the initiation site for the scanning process is determined by its distance from the branchpoint. Previous studies have shown that the sequence and structure of the pre-mRNA downstream of the branchpoint can affect 3′ splice site selection. For example, a stable hairpin loop positioned between the branchpoint and the 3′ splice site was shown to inhibit the second step of splicing *in vitro* (Smith *et al.*, 1989, 1993). Surprisingly, there was a strong position effect of the secondary structure element, so that in a branchpoint-proximal position the loop could be bypassed, while in a more distal position it strongly inhibited splicing (Smith *et al.*, 1993). It has been reported from both yeast and viral systems that AGs sequestered in stable RNA structures can be bypassed in favour of more distal sites, and in both cases the sequence immediately downstream of the stem/loop was pyrimidine rich (Chebli *et al.*, 1989; Deshler and Rossi, 1991). Patterson and Guthrie (1991) demonstrated that insertion of a strong polypyrimidine tract between two competing AGs could shift splicing to the distal site. In this case, it is possible that scanning was initiated at the polypyrimidine tract rather than at the branchpoint. Alternatively, a polypyrimidine tract could favour the first or second 3′ splice site selection step described above.

An alternative model would be that 3′ splice sites are not recognized by scanning, but rather by a diffusion/collision mechanism. According to this model, recognition of the 3′ splice site by components of the active site is a stochastic process, where the most branchpoint-proximal site has the highest probability of recognition. Given the high flexibility of RNA, how could one explain why the proximal site would have such a strong advantage over the downstream site? A possibility is that this is the result of packing of pre-mRNA into hnRNP-type complexes that might restrict the flexibility of the RNA chain. This would strongly favour the proximal sites at the expense of the distal ones. Our experiment with repeated 3′ splice site sequence suggests that, in that case, the preferred landing site for the splicing machinery would be 13–22 nucleotides downstream of the branchpoint. Efficient activation of sites that are located further downstream (as is often found naturally in yeast introns; Kalogeropoulos, 1995) could then require additional factors that would facilitate interaction of this sequence with the branchpoint-bound splicing machinery, either by affecting RNA folding or by pre-recognizing the consensus YAG. In this vein, it is interesting that splicing to distal splice acceptors is also dependent on the U5 snRNP-associated SLU7 protein (Frank and Guthrie, 1992). Strikingly, SLU7 was found to be required for splicing to 3′ acceptors further than 22 nucleotides from the branchpoint (Frank and Guthrie, 1992). It was, therefore, tempting to suggest that activation of distal 3′ splice sites could be triggered by base pairing between the downstream exon sequence and the U5 snRNA loop. We have tested this possibility with our reporter construct by increasing the complementarity of the nucleotide following the distal site to U5 snRNA. This did not enhance the usage of the downstream site and suggests, therefore, that U5 snRNA–exon 2 interaction is not an important factor in 3′ splice site selection.

As more and more data on 3′ splice site selection become available, some of the proposed models for this mechanism, such as leaky scanning and first–last nucleotide interaction, become less certain. Clearly, the mechanism for selecting a splice acceptor is complex and may involve multiple sequential recognition events. Several different mechanisms may be required for distal 3′ splice sites, secondary structure elements or pyrimidine-rich sequences. However, the factors that actually bind to and select the 3′ splice site still remain to be determined.

Materials and methods

Plasmids and strains

The duplicated 3′ splice sites were constructed by replacing the downstream part of the RP51 intron in pBS7 (Séraphin *et al.*, 1988) with a synthetic DNA-generating plasmid pBS533. The sequence of this synthetic 3′ splice site region was designed to fit the *S.cerevisiae* consensus and is depicted in Figure 1A. A *Sal*I–*Sac*I fragment containing the synthetic 3′ splice sites was subcloned to a *Sal*I–*Sac*I-digested pHZ18, generating plasmid pBS546.

Site-directed mutagenesis was performed on single-stranded DNA prepared from plasmid pBS533, using the dut- ung- procedure (Kunkel, 1985). All mutations were confirmed by DNA sequencing using T7 DNA polymerase (Pharmacia). A completely sequenced *Kpn*I–*Sal*I fragment from each mutant was subcloned into the pBS546 backbone to reconstitute the reporter plasmid. The UAG, UUG and AAG constructs harbouring

multiple 3′ splice sites (Figure 5A) were synthesized as oligonucleotides spanning the *Bgl*II–*Kpn*I sequence of pBS546. The oligonucleotides were hybridized and extended using Klenow polymerase (Boehringer Mannheim), and cut with *Sau*3A and *Kpn*I. This fragment was subcloned to the *Bgl*II–*Kpn*I-digested pBS546 reporter, and the *Bgl*II–*Kpn*I spanning region was sequenced using T7 DNA polymerase (Pharmacia). Plasmids harbouring 5′ splice site mutations were a kind gift of G.Chanfreau and A.Jacquier (Chanfreau *et al.*, 1994). All plasmids were propagated in *Escherichia coli* strain MC1066.

All yeast transformations were carried out in strain MGD353-13D (MAT**a**, *trp1-289, ura3-52, arg4, leu2-3, 112, ade2*) (Séraphin *et al.*, 1988) using the LiCl method (Ito *et al.*, 1983).

RNA extraction and analysis

Total RNA was extracted as previously described (Pikielny and Rosbash, 1985), from cells grown in lactate–glycerol-containing URA medium to an OD_{600} of 0.5–1.0, and induced for 2 h with galactose at a final concentration of 2%. Three micrograms of total RNA were analysed by primer extension using the exon 2-specific oligonucleotide EM38 (5′-CACGCTTGACGGTCTTGGT-3′), or the intron-specific RB27 oligonucleotide (Jacquier *et al.*, 1985) as described (Pikielny and Rosbash, 1985). RT–PCR was performed by an initial cDNA synthesis step using primer EM38 (as above), and a subsequent amplification of the cDNA using oligos EM38 (1:20 ^{32}P-labelled:unlabelled) and the exon 1 specific EM84 primer (5′-CACACTAAATTAATAATGACC-3′) (42 s at 95°C, 1 min at 57°C and 2 min at 72°C for 25 cycles) using AmpliTaq DNA polymerase (Perkin-Elmer). Primer extensions and RT–PCRs were separated on denaturing polyacrylamide gels, dried and exposed with intensifying screens (DuPont). Gels were quantified in a PhosphoImager (Molecular Dynamics).

For sequencing of amplified mRNAs, wet RT–PCR gels were exposed overnight, and a gel piece corresponding to the amplified mRNA was cut out and eluted in 300 µl TE buffer overnight. A fraction of the eluate was run on a denaturing polyacrylamide gel to rule out the presence of contaminating bands, while a second fraction was re-amplified by PCR. The re-amplified DNA was purified on a 3% NuSieve agarose gel (FMC) and eluted on QiaQuick columns (Qiagen). A fraction of this eluate corresponding to 10 ng DNA was subjected to cycle sequencing using Vent exo- (New England Biolabs) and a ^{32}P end-labelled exon 2-specific primer (RB1; Teem and Rosbash, 1983), according to the manufacturer's recommendations.

Mathematical model for leaky scanning

This analysis of the leaky scanning mechanism assumes that the reporter pre-mRNA has two 3′ splice sites and that scanning will be reinitiated indefinitely upstream of the proximal site until one of the sites is selected for splicing. If *a* and *b* are the respective frequencies with which the first and second sites are selected at each round of scanning, the final frequency of usage of the first site will be *A*:

$$A = \frac{a}{a + b - ab}$$

while the frequency of usage of the second site is *B*:

$$B = \frac{b - ab}{a + b - ab}$$

This assumes that either *a* or *b* are different from 0, otherwise no splicing would ever occur. From this model, one observes that if the first site is efficiently selected ($a \approx 1$) it is exclusively used independently of the nature of the second site ($A = 1, B = 0$; e.g. with the UAG–UAG construct). The second site will only be efficiently used ($B = 1$) if the first site is poorly selected (*a* tends toward 0; e.g. in the case of the UCG–UAG construct). If the two sites are identical, $a = b \neq 0$, one gets:

$$A = \frac{1}{2 - a} \quad \text{and} \quad B = \frac{1 - a}{2 - a}$$

If the sites are efficiently selected ($a \approx 1$), the first one is used preferentially ($A = 1$, $B = 0$; e.g. with the UAG–UAG construct), while if the sites are poorly selected (*a* tends toward 0), then both sites should tend to be used to the same extent ($A = B = 0.5$). However, this is not what we observe with the UCG–UCG construct.

Acknowledgements

We are grateful to A.Jacquier and G.Chanfreau for providing plasmids used in this study, and to S.Kandels-Lewis for technical assistance and advice. We want to thank I.Mattaj, J.Valcarcel, C.Gamberi, D.Jackson, I.Palacios, J.Salgado-Garrido, S.Vagner and partly E.Izaurralde for helpful comments on the manuscript and for stimulating discussions. The excellent secretarial assistance of C.Kjaer and the support from the EMBL services are gratefully acknowledged. This work was supported by EMBL. B.S. is on leave from CNRS.

References

Aebi,M., Hornig,H. and Weissmann,C. (1987) 5′ cleavage site in eukaryotic pre-mRNA splicing is determined by the overall 5′ splice region, not by the conserved 5′ GU. *Cell*, **50**, 237–246.

Ansari,A. and Schwer,B. (1995) SLU7 and a novel activity, SSF1, act during the PRP16-dependent step of yeast pre-mRNA splicing. *EMBO J.*, **14**, 4001–4009.

Aoufouchi,S., Yelamos,J. and Milstein,C. (1996) Nonsense mutations inhibit RNA splicing in a cell-free system: recognition of mutant codon is independent of protein synthesis. *Cell*, **85**, 415–422.

Burgess,S., Couto,J.R. and Guthrie,C. (1990) A putative ATP binding protein influences the fidelity of branchpoint recognition in yeast splicing. *Cell*, **60**, 705–717.

Carothers,A., Urlaub,G., Grunberger,D. and Chasin,L. (1993) Splicing mutants and their second-site suppressors at the dihydrofolate reductase locus in Chinese hamster ovary cells. *Mol. Cell. Biol.*, **13**, 5085–5098.

Chanfreau,G., Legrain,P., Dujon,B. and Jacquier,A. (1994) Interaction between the first and last nucleotides of pre-mRNA introns is a determinant of 3′ splice site selection in *S. cerevisiae*. *Nucleic Acids Res.*, **22**, 1981–1987.

Chebli,K., Gattoni,R., Schmitt,P., Hildwein,G. and Stevenin,J. (1989) The 216 nucleotide intron of the E1a pre-mRNA contains a hairpin structure that permits utilization of unusually distant branch acceptors. *Mol. Cell. Biol.*, **9**, 4852–4861.

Couto,J.R., Tamm,J., Parker,R. and Guthrie,C. (1987) A trans-acting suppressor restores splicing of a yeast intron with a branch point mutation. *Genes Dev.*, **1**, 445–455.

Datta,B. and Weiner,A.M. (1993) The phylogenetically invariant ACAGAGA and AGC sequences of U6 small nuclear RNA are more tolerant of mutation in human cells than in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **13**, 5377–5382.

Deshler,J.O. and Rossi,J.J. (1991) Unexpected point mutations activate cryptic 3′ splice sites by perturbing a natural secondary structure within a yeast intron. *Genes Dev.*, **5**, 1252–1263.

Frank,D. and Guthrie,C. (1992) An essential splicing factor, SLU7, mediates 3′ splice site choice in yeast. *Genes Dev.*, **6**, 2112–2124.

Fu,X.Y., Ge,H. and Manley,J.L. (1988) The role of the polypyrimidine stretch at the SV40 early pre-mRNA 3′ splice site in alternative splicing. *EMBO J.*, **7**, 809–817.

Green,M.R. (1991) Biochemical mechanisms of constitutive and regulated pre-mRNA splicing. *Annu. Rev. Cell Biol.*, **7**, 559–599.

Guthrie,C. (1991) Messenger RNA splicing in yeast: clues to why the spliceosome is a ribonucleoprotein. *Science*, **253**, 157–163.

Hall,S. and Padgett,R. (1994) Conserved sequences in a class of rare eukaryotic nuclear introns with non-consensus splice sites. *J. Mol. Biol.*, **239**, 357–365.

Hall,S. and Padgett,R. (1996) Requirement of U12 snRNA for *in vivo* splicing of a minor class of eukaryotic nuclear pre-mRNA introns. *Science*, **271**, 1716–1718.

Ito,H., Fukada,Y., Murata,K. and Kimura,A. (1983) Transformation of intact yeast cells treated with alkali cations. *J. Bacteriol.*, **153**, 163–168.

Jackson,I.J. (1991) A reappraisal of non-consensus mRNA splice sites. *Nucleic Acids Res.*, **19**, 3795–3798.

Jacquier,A., Rodriguez,J.R. and Rosbash,M. (1985) A quantitative analysis of the effects of 5′ junction and TACTAAC box mutants and mutant combinations on yeast mRNA splicing. *Cell*, **43**, 423–430.

Kalogeropoulos,A. (1995) Automatic intron detection in nuclear DNA sequences of *Saccharomyces cerevisiae*. *Yeast*, **11**, 555–565.

Kandels-Lewis,S. and Séraphin,B. (1993) Involvement of U6 snRNA in 5′ splice site selection. *Science*, **262**, 2035–2039.

Kunkel,T. (1985) Rapid and efficient site-specific mutagenesis without phenotypic selection. *Proc. Natl Acad. Sci. USA*, **82**, 488–492.

Lesser,C.F. and Guthrie,C. (1993) Mutations in U6 snRNA that alter splice site specificity: implications for the active site. *Science*, **262**, 1982–1988.

Madhani,H.D. and Guthrie,C. (1994) Randomization-selection analysis of snRNAs *in vivo*: evidence for a tertiary interaction in the spliceosome. *Genes Dev.*, **8**, 1071–1086.

Madhani,H.D., Bordonne,R. and Guthrie,C. (1990) Multiple roles for U6 snRNA in the splicing pathway. *Genes Dev.*, **4**, 2264–2277.

Maquat,L.E. (1995) When cells stop making sense: effects of nonsense codons on RNA metabolism in vertebrate cells. *RNA*, **1**, 453–465.
Moore,M., Query,C. and Sharp,P. (1993) Splicing of precursors to mRNA by the spliceosome. In Gesteland,R. and Atkins,J. (eds), *The RNA World*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, pp. 303–357.
Newman,A. and Norman,C. (1991) Mutations in yeast U5 snRNA alter the specificity of 5′ splice-site cleavage. *Cell*, **65**, 115–123.
Newman,A.J. and Norman,C. (1992) U5 snRNA interacts with exon sequences at 5′ and 3′ splice sites. *Cell*, **68**, 743–754.
O'Keefe,R., Norman,C. and Newman,A. (1996) The invariant U5 snRNA loop 1 sequence is dispensable for the first catalytic step of pre-mRNA splicing in yeast, **86**, *Cell*, 679–689.
Parker,R. and Patterson,B. (1987) Architecture of fungal introns: implications for spliceosome assembly. In Inouye,M. and Dudock,B. (eds), *Molecular Biology of RNA. New Perspectives*. Academic Press, New York, pp. 133–150.
Parker,R. and Siliciano,P.G. (1993) Evidence for an essential non-Watson-Crick interaction between the first and last nucleotides of a nuclear pre-mRNA intron. *Nature*, **361**, 660–662.
Parker,R., Siliciano,P.G. and Guthrie,C. (1987) Recognition of the TACTAAC box during mRNA splicing in yeast involves base pairing to the U2-like snRNA. *Cell*, **49**, 229–239.
Patterson,B. and Guthrie,C. (1991) A U-rich tract enhances usage of an alternative 3′ splice site in yeast. *Cell*, **64**, 181–187.
Pikielny,C.W. and Rosbash,M. (1985) mRNA splicing efficiency in yeast and the contribution of nonconserved sequences. *Cell*, **41**, 119–126.
Query,C.C., Moore,M.J. and Sharp,P.A. (1994) Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. *Genes Dev.*, **8**, 587–597.
Reed,R. (1989) The organization of 3′ splice-site sequences in mammalian introns. *Genes Dev.*, **3**, 2113–2123.
Reich,C., VanHoy,R., Porter,G. and Wise,J. (1992) Mutations at the 3′ splice site can be suppressed by compensatory base changes in U1 snRNA in fission yeast. *Cell*, **69**, 1159–1169.
Rymond,B.C. and Rosbash,M. (1985) Cleavage of 5′ splice site and lariat formation are independent of 3′ splice site in yeast mRNA splicing. *Nature*, **317**, 735–737.
Rymond,B.C., Torrey,D.D. and Rosbash,M. (1987) A novel role for the 3′ region of introns in pre-mRNA splicing of *Saccharomyces cerevisiae*. *Genes Dev.*, **1**, 238–246.
Sawa,H. and Abelson,J. (1992) Evidence for a base-pairing interaction between U6 small nuclear RNA and 5′ splice site during the splicing reaction in yeast. *Proc. Natl Acad. Sci. USA*, **89**, 11269–11273.
Sawa,H. and Shimura,Y. (1992) Association of U6 snRNA with the 5′ splice site region of pre-mRNA in the spliceosome. *Genes Dev.*, **6**, 244–254.
Scadden,A.D.J. and Smith,C.W. (1995) Interactions between the terminal bases of mammalian introns are retained in inosine-containing pre-mRNAs. *EMBO J.*, **14**, 3236–3246.
Séraphin,B. and Kandels-Lewis,S. (1993) 3′ splice site recognition in *S. cerevisiae* does not require base pairing with U1 snRNA. *Cell*, **73**, 803–812.
Séraphin,B. and Rosbash,M. (1989) Identification of functional U1 snRNA-pre-mRNA complexes committed to spliceosome assembly and splicing. *Cell*, **59**, 349–358.
Séraphin,B. and Rosbash,M. (1990) Exon mutations uncouple 5′ splice site selection from U1 snRNA pairing. *Cell*, **63**, 619–629.
Séraphin,B., Kretzner,L. and Rosbash,M. (1988) A U1 snRNA:pre-mRNA base pairing interaction is required early in yeast spliceosome assembly but does not uniquely define the 5′ cleavage site. *EMBO J.*, **7**, 2533–2538.
Siliciano,P.G. and Guthrie,C. (1988) 5′ splice site selection in yeast: genetic alterations in base-pairing with U1 reveal additional requirements. *Genes Dev.*, **2**, 1258–1267.
Smith,C.W., Porro,E.B., Patton,J.G. and Nadal Ginard,B. (1989) Scanning from an independently specified branch point defines the 3′ splice site of mammalian introns. *Nature*, **342**, 243–247.
Smith,C.W., Chu,T.T. and Nadal Ginard,B. (1993) Scanning and competition between AGs are involved in 3′ splice site selection in mammalian introns. *Mol. Cell. Biol.*, **13**, 4939–4952.
Sontheimer,E.J. and Steitz,J.A. (1993) The U5 and U6 small nuclear RNAs as active site components of the spliceosome. *Science*, **262**, 1989–1996.
Tarn,W.-Y. and Steitz,J. (1996) A novel spliceosome containing U11, U12 and U5 snRNPs excises a minor class (AT-AC) intron *in vitro*. *Cell*, **84**, 801–811.
Teem,J.L. and Rosbash,M. (1983) Expression of a beta-galactosidase gene containing the ribosomal protein 51 intron is sensitive to the rna2 mutation of yeast. *Proc. Natl Acad. Sci. USA*, **80**, 4403–4407.
Teigelkamp,S., Newman,A.J. and Beggs,J.D. (1995a) Extensive interactions of PRP8 protein with the 5′ and 3′ splice sites during splicing suggest a role in stabilization of exon alignment by U5 snRNA. *EMBO J.*, **14**, 2602–2612.
Teigelkamp,S., Whittaker,E. and Beggs,J.D. (1995b) Interaction of the yeast splicing factor PRP8 with substrate RNA during both steps of splicing. *Nucleic Acids Res.*, **23**, 320–326.
Umen,J.G. and Guthrie,C. (1995a) A novel role for a U5 snRNP protein in 3′ splice site selection. *Genes Dev.*, **9**, 855–868.
Umen,J.G. and Guthrie,C. (1995b) Prp16p, Slu7p, and Prp8p interact with the 3′ splice site in two distinct stages during the second catalytic step of pre-mRNA splicing. *RNA*, **1**, 584–597.
Umen,J.G. and Guthrie,C. (1995c) The second catalytic step of pre-mRNA splicing. *RNA*, **1**, 869–885.
Umen,J. and Guthrie,C. (1996) Mutational analysis of PRP8 reveals separate domains governing the specificity and fidelity of 3′ splice site selection. *Genetics*, **143**, 723–739.
Vankan,P., McGuigan,C. and Mattaj,I.W. (1992) Roles of U4 and U6 snRNAs in the assembly of splicing complexes. *EMBO J.*, **11**, 335–343.
Vijayraghavan,U. and Abelson,J. (1990) PRP18, a protein required for the second reaction in pre-mRNA splicing. *Mol. Cell. Biol.*, **10**, 324–332.
Vijayraghavan,U., Company,M. and Abelson,J. (1989) Isolation and characterization of pre-mRNA splicing mutants of *Saccharomyces cerevisiae*. *Genes Dev.*, **3**, 1206–1216.
Wassarman,D.A. and Steitz,J.A. (1992) Interactions of small nuclear RNA's with precursor messenger RNA during *in vitro* splicing. *Science*, **257**, 1918–1925.
Whittaker,E. and Beggs,J.D. (1991) The yeast PRP8 protein interacts directly with pre-mRNA. *Nucleic Acids Res.*, **19**, 5483–5489.
Wolff,T., Menssen,R., Hammel,J. and Bindereif,A. (1994) Splicing function of mammalian U6 small nuclear RNA: conserved positions in central domain and helix I are essential during the first and second step of pre-mRNA splicing. *Proc. Natl Acad. Sci. USA*, **91**, 903–907.
Wu,J. and Manley,J.L. (1989) Mammalian pre-mRNA branch site selection by U2 snRNP involves base pairing. *Genes Dev.*, **3**, 1553–1561.
Wyatt,J., Sontheimer,E. and Steitz,J. (1992) Site-specific cross-linking of mammalian U5 snRNP to the 5′ splice site before the first step of pre-mRNA splicing. *Genes Dev.*, **6**, 2542–2553.
Yu,Y.T., Maroney,P.A. and Nilsen,T.W. (1993) Functional reconstitution of U6 snRNA in nematode cis- and trans-splicing: U6 can serve as both a branch acceptor and a 5′ exon. *Cell*, **75**, 1049–1059.
Zhuang,Y. and Weiner,A.M. (1986) A compensatory base change in U1 snRNA suppresses a 5′ splice site mutation. *Cell*, **46**, 827–835.
Zhuang,Y. and Weiner,A.M. (1989) A compensatory base change in human U2 snRNA can suppress a branch site mutation. *Genes Dev.*, **3**, 1545–1552.

Received on September 6, 1996; revised on October 15, 1996