Interaction between the first and last nucleotides of pre-mRNA introns is a determinant of 3' splice site selection in S. cerevisiae

Guillaume Chanfreau, Pierre Legrain, Bernard Dujon and Alain Jacquier*
Unité de Génétique Moléculaire des Levures (URA 1149 du CNRS), Département de Biologie Moléculaire, Institut Pasteur, 25 rue du Dr Roux, F-75724 Paris CEDEX 15, France

Received April 8, 1994; Accepted May 5, 1994

ABSTRACT

The splicing of group II and nuclear pre-mRNAs introns occurs via a similar splicing pathway and some of the RNA–RNA interactions involved in these splicing reactions show structural similarities. Recently, genetic analyses performed in a group II intron and the yeast nuclear actin gene suggested that non Watson-Crick interactions between intron boundaries are important for the second splicing step efficiency in both classes of introns. We here show that, in the yeast nuclear rp51A intron, a G to A mutation at the first position activates cryptic 3' splice sites with the sequences UAC or UAAA Moreover, the natural 3' splice site could be reactivated by a G to C substitution of the last intron nucleotide. These results demonstrate that the interaction between the first and last intron nucleotides is a conserved feature of nuclear pre-mRNA splicing in yeast and is involved in the mechanism of 3' splice site selection.

INTRODUCTION

Nuclear pre-mRNAs introns are excised from precursor RNAs by a two step transesterification pathway which is similar to the splicing pathway of group II introns found in organelles of fungi and plants (1) and in bacteria (2). The first step products are the 5' exon, and a lariat intron-3' exon intermediate in which the first nucleotide of the intron (G1) is bound to an internal adenosine by a 5'-2' phosphodiester bond. The second step consists in the ligation of the two exons and liberation of the lariat intron. Group II introns can self-splice in vitro (3; 4) due to their conserved high order structure (1). In contrast, nuclear pre-mRNAs introns do not show strongly conserved secondary structures and their excision requires the assembly of a multimolecular ribonucleoprotein complex called the spliceosome (5; 6). Despite these differences, functional similarities have been found between intermolecular RNA-RNA interactions in the spliceosome and conserved structural features of group II introns (reviewed in 7).

The consensus sequences of nuclear pre-mRNA introns at the 5' splice site are /GUAUGU in yeast and /GURAGU in higher eucaryotes, while the 3' splice site is conserved as YAG/ in all eucaryotes. In addition, two kinds of introns have been defined in yeast (8): the first class, called 3' long (3'L) has a branchpoint away from the 3' splice site (>20nt), with a polypyrimidine tract. The second class, called 3' short (3'S) is characterized by a short distance between the branchpoint and the 3' splice site (<20 nt) with no polypyrimidine tract. The 5' consensus sequences are reminiscent of the consensus sequences found in group II introns (/GUGYG; 1). In contrast, the only similarity between group II and nuclear pre-mRNAs introns at the 3' splice site concerns the penultimate nucleotide (A). The role of these conserved sequences in nuclear introns has been elucidated in part by the description of base-pairing interactions between these sequences and small nuclear RNA (snRNA) (reviewed in 6; see also 9; 10; 11). In addition to these intermolecular interactions, an example of long distance intramolecular interaction has been reported recently in the yeast actin intron (12). By genetic analysis, the first and last nucleotides of this intron were shown to be involved in a non Watson-Crick interaction essential to the second splicing step. Interestingly, boundaries of group II introns are also involved in an interaction which occurs after the first step of the splicing reaction (13). In this case, the first and the penultimate nucleotides of the intron interact. The fact that the first nucleotide is conserved as a G and the penultimate as an A in both group II and nuclear pre-mRNA introns suggested that the first to penultimate nucleotide interaction described for group II introns could also occur in nuclear introns, in addition to the interaction between the first and last nucleotides.

To test this hypothesis, we have undertaken a mutational analysis of the first, penultimate and last nucleotides of the intron of the S. cerevisiae rp51A nuclear gene. No compensatory effect can be found between mutations at the first and penultimate positions. In contrast, compensatory effects between mutations at the first and last positions are observed, showing that the interaction between these nucleotides is not restricted to the actin intron. Moreover, a G1 to A mutation at the 5' splice site induces use of cryptic 3' splice sites according to the compensatory combinations described for the actin intron (12). These results show that the interaction between the first and last intron

*To whom correspondence should be addressed
**A**

**RP51A Intron**

<table>
<thead>
<tr>
<th>Exon 1</th>
<th>Exon 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>GUAUG</td>
<td>AAUAG</td>
</tr>
</tbody>
</table>

**β-galactosidase activities**

<table>
<thead>
<tr>
<th>Position</th>
<th>G</th>
<th>a</th>
<th>c</th>
<th>u</th>
</tr>
</thead>
<tbody>
<tr>
<td>299</td>
<td>0.66</td>
<td>2.9</td>
<td>0.47</td>
<td></td>
</tr>
<tr>
<td>0.24</td>
<td>3.22</td>
<td>0.97</td>
<td>1.17</td>
<td></td>
</tr>
<tr>
<td>0.08</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>0.39</td>
<td>0.33</td>
<td>0.07</td>
<td>0.01</td>
<td></td>
</tr>
</tbody>
</table>

**B**

**Gacu**

**Position 1**

<table>
<thead>
<tr>
<th>A</th>
<th>G</th>
<th>a</th>
<th>c</th>
<th>u</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

**Position 397**

<table>
<thead>
<tr>
<th>A</th>
<th>G</th>
<th>a</th>
<th>c</th>
<th>u</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

**Figure 1.** Splicing of mutants at the first and penultimate positions of the RP51A gene intron. A: β-galactosidase activities resulting from splicing of mutant RNAs at the first (1) and penultimate (397) intron positions. Wild-type sequences are bold, mutant sequences are in lower case. The activities are defined as in (20). The activities are mean values of at least two independent experiments, in which the activity of at least two independent transformants was measured in duplicate. The background level is 0.01. B: Primer extension analysis; mutant sequences are in lower case. The multiple band pattern of mRNAs results from the multiple transcription starts of the CYC1 promoter used in the reporter gene (see text). Proximal cryptic sites can be detected in mutants U1 (lane 5) and U397 (lane 8) by a slight shift in size in the band pattern. The arrow points to the minor proximal cryptic site activated by the A1/C397 mutation, and to a lesser extent, in other double mutants containing the A1 mutation.

**MATERIALS AND METHODS**

**Plasmids and strains**

All plasmids used are mutation derivatives of pHZ18 (14). This plasmid is a derivative of pLGSD5 (15) which contains a rp51A intron-β-galactosidase hybrid gene under the control of the galactose inducible CYC1 promoter (14). Two potential cryptic 3' splice sites (UAG sequences) are located in exon 2. Although activation of these sites by mutations at the 3' splice site was not tested, we used a derivative of pHZ18 in which these two UAG sequences were substituted to UAA to avoid these events. These mutations did not affect the splicing of the wild-type intron (data not shown). In vitro mutagenesis was performed after subcloning into pBluescript® KS- vector (Stratagene Co., La Jolla, CA) and using the dut-', ung' procedure (16). Recombinant pHZ18 plasmids were sequenced as described (17) to verify constructions. Yeast cells (strain MGD353-46D (MATα ura3-52 trpl-289 leu2-3, 112 his3-delta1), (18)) were transformed using the LiCl method as described (19).

**RNA analysis and β-galactosidase assays**

Yeast cells were grown at 30°C in glycerol-lactate medium lacking uracil, with 2% galactose in order to induce the reporter gene. At O.D. 600 0.5-1.2, cells were harvested to perform β-galactosidase assay (20) or to extract total RNAs as described (21) with minor modifications. Primer extension analyses were done as follows using primer PL117 (5'CGCTTGACGGTCTTGG). 5 μg of total RNAs were hybridized 1 hour at 42°C with 0.25 pmol of gel-purified 5' end labeled PL117 primers in 6 μl of reverse transcriptase buffer (50mM Tris·Cl pH 8.3, 40mM KCl, 1mM DTT, 6mM MgCl2). A volume of 4 μl of reverse transcription mix (50mM Tris·Cl pH 8.3, 40 mM KCl, 1mM DTT, 6mM MgCl2, 125 μM dNTPs, 0.5 μg of actinomycin D, 4 units of Stratagene AMV reverse transcriptase) was added and reaction tubes were incubated at 42°C for 30 minutes. Reactions were stopped by adding 10 μl of formamide-NaOH loading buffer (deionised formamide: 98%; 10mM NaOH; 1mM EDTA pH 8.0; 0.025% xylene cyanol; 0.025% bromophenol blue). Aliquot volumes (typically 8μl) were loaded on 5% sequencing gels. Dried gels were exposed using intensifying screens.

**nucleotides is not only essential for efficient 3' cleavage, as previously described, but is also an element of 3' splice site specificity.**
Compensatory effects of mutations at the first and last positions of the rp51A intron on 3' cleavage at the natural site

The activation by the A1 mutation of compensatory cryptic 3' splice sites raised the possibility of a similar compensatory effect at the natural 3' splice site. To investigate this possibility, we have analyzed constructions bearing further alterations of the 3' splice site. A G to C mutation at the last intron nucleotide was combined with the three different substitutions at the first position to test restoration of cleavage at the wild-type 3' splice site as described for the actin intron. In addition, the effect of a double mutation of the 3' splice site (AG to CC) was examined.

Splicing was analyzed by primer extension (figure 3A; Wild-type, A1 and A1-C397 mutants are shown as controls in lanes 2, 3 and 6, respectively). No mRNA could be detected in the C398 and C397-C398 mutants and the transcripts accumulated as precursors and lariat intermediates, showing an impairment of the first and second splicing steps (figure 3A; lanes 4 and 5). These constructs gave low levels of β-galactosidase activity generated by this double mutation (figure 3B). When combined with the A1 mutation the major splicing product from these mutants of the 3' splice site was the cryptic -39 mRNA already present in the A1 single mutant. In addition, the double mutant A1-C398 exhibited a band migrating at the same size as the wild-type mRNA (figure 3A, lane 7). The increase of β-galactosidase activity generated by this double mutation indicated a significant compensatory effect. Sequencing the PCR product from this mRNA (figure 2A) confirmed that it was the wild-type mRNA (data not shown). This compensatory effect confirms the result observed by Parker and Siliciano (12), i.e. a partial but reciprocal suppression of second step defects of mutants at the first and last position by the double mutant combination. The compensation was weaker than in the actin intron most likely because of competition for 3' cleavage by the −39 site, which is used much more efficiently.

The triple mutation A1-C397-C398 induced an increase in the use of the minor 3' cryptic splice site at position −3 already detected in the A1-C397 mutant (compare lanes 6 and 8 in figure 3A). The location of this site was confirmed by sequencing the corresponding PCR product (figure 2A and 2B). This result correlates with an increase of the β-galactosidase levels (compare the β-galactosidase values in figure 1A and 3B; note that the similar values observed in the A1-C398 and A1-C397-C398 combinations do not result of the utilisation of the same 3' splice site). Thus, when the sequence of the penultimate nucleotide is altered by an A to C mutation, the G to C mutation at the last
position fails to restore splicing at the natural site in the A1 mutant. This observation shows that a productive interaction between the first and last nucleotides requires unaltered sequences at the penultimate position.

**DISCUSSION**

We have undertaken a mutational analysis of the 5' and 3' splice sites of the rp51A intron. The effects of mutations of the first nucleotide are consistent with previous studies in other *S. cerevisiae* introns (23; 24). Mutations at the 3' splice site blocked the second splicing step as shown by the large increase of the lariat intermediate amount. In addition, these mutations induced some increase of the amount of unspliced precursor. Such observations were previously reported (26; 27). It suggests that 3' splice site sequences, although non-essential for the first splicing step in *S. cerevisiae* (28), are recognized prior to 5' cleavage and moderately affect the first step (27). However, the


**Figure 3.** Splicing of mutant RNAs at the first, penultimate and last positions. A: Primer extension analysis of mutants at the first (1), penultimate (397) and last (398) positions. Legends are the same as in figure 1B. B: β-galactosidase activity resulting from splicing of mutants at the first, penultimate and last positions. Legends are the same as in figure 1A; ND = not determined. *: this β-galactosidase activity results from a 3' cryptic site cleavage (position -3; see text).
amount of total plasmid borne RNAs was decreased in these mutants. We propose that in 3' splice site mutants the majority of RNAs are blocked after the first splicing step and that the resulting lariat intermediates are unstable. The half life of the intermediate could be controlled by a spliceosomal component like the PRP16 protein (29). The observation that the combination of these mutations with position 1 mutations which strongly block splicing prior to lariat formation, restores the amount of transcripts in the double mutants is consistent with this hypothesis (compare lanes 6—8 to lanes 9—17 in figure 1B).

In order to test an interaction between the first and penultimate nucleotides of a yeast nuclear pre-mRNA intron, as observed in group II introns, we analyzed the splicing of a complete panel of double mutants at the first and penultimate positions. This study did not reveal any compensatory effect. In particular, the equivalent of the compensatory combinations in group II introns (U1-G397 and C1-G397; (13)) are completely inactive for the second step of splicing in the rp51A intron. In contrast, we observed a compensatory effect of a double mutation at the first and last positions, which occurred in the same combination (A1-C at the last position) as reported in the actin intron (12). In addition to this compensatory effect at the wild-type splice site, the activation of cryptic 3' splice sites by a 5' mutation (A1) provides new perspective on the mechanism of 3' splice site selection in yeast. The two alternative sites occur downstream of a nucleotide which differs from the universally conserved G (UAC/ at —39 and UAA/ at —3). The fact that the sequence of the alternative splice sites are in accordance with the compensatory combinations at the natural sites described here (figures 3A and 3B) and in the actin intron (12) indicates that the interaction between the first and last nucleotides is involved in the mechanism of 3' splice site selection.

The minor site at position —3 (UAA/) is activated by the A1 mutation, but only when the wild-type 3'SS is altered by an A to C substitution at the penultimate position (mutants A1-C397 and A1-C397-C398). Interestingly, the wild-type 3'SS is not used in the triple mutant A1-C397-C398, even though it presents the compensatory change G to C at the last position. Indeed, the more the natural 3' splice site is altered, the more the —3 site is used (compare utilisation of the —3 site in mutants A1-C397 and A1-C397-C398 in lanes 6 and 8 of figure 3A). This suggests that other elements recognize the 3' splice site and that the —3 and natural sites are in competition for these elements. In this hypothesis, the mutated (UCG/) natural site would be recognized but not cleaved. In Schizosaccharomyces pombe, the 3' splice site AG is recognized by base pairing with positions 7 and 8 of U1 snRNA (30). However, this interaction does not seem to occur with the corresponding positions of U1 (9 and 10) in S. cerevisiae (31). Moreover, in S. pombe, its role is limited to 5' cleavage. Alternatively, U6 snRNA (10) or some U5 associated proteins (32) are potential candidates for being such factors. Indeed, non Watson—Crick base pairs are important RNA recognition elements for proteins (reviewed in 33), and an interesting hypothesis is that the RNA—RNA interaction discussed here constitutes an element of recognition for proteins essential for the second step.

In addition, one has to emphasize that the —3 splice site is preferred over two other potential UAA/ sites located in exon 2, near the 3' junction. Similarly, in the A1-C398 double mutant, the cryptic site UAC/ at —39 is predominantly used over the UAC/ site wild-type position. Thus, it appears that 3' splice sites located closer to the branchpoint are preferred to more distant ones (note that the fact that the —39 site UAC/ is predominant over the —3 site UAA/ in the A1-C397 and A1-C397-C398 mutants can also be explained by the better compensatory effect of an A-C pairing over an A-A pairing; 12). This branchpoint proximity rule is consistent with previous experiments in yeast (27) or mammalian systems (34). In light of our results, a possible interpretation of these observations is that a short distance between the branched G1 and the 3' splice site favors the formation of the first to last interaction which is implicated in 3' splice site selection; thus 3' splice sites closer to the branchpoint would be more easily positioned near the branched G1. Furthermore, it has been shown that the polypyrimidine tract is essential for 3'SS selection in 3' long introns (27). This sequence could serve to fold the 3' end of the intron in order to bring the branched G1 close to the last intron nucleotide. It is possible that some factor(s) which have been identified as important for 3' splice site selection (32) would interact with the polypyrimidine tract to induce this folding.

To our knowledge, our results constitute the first example of a mutation of the 5' splice site inducing utilization of cryptic 3' splice sites. There have been several reports of cryptic splice sites activation by point mutations, for example in globin gene introns where such mutations are responsible for thalassemia (35 and references therein). However, in previously reported cases, cryptic 3' splice sites were always induced by mutations at the 3' end of the intron. Despite this finding, several points remain unanswered. First, does the interaction between the first and last intron nucleotides also occur in other systems, as in mammalian cells? The analysis of second site suppressors of splicing mutants of the dihydrofolate reductase gene in Chinese hamster ovary cells led to the observation that a G to A mutation at the first position could be compensated by a G to A substitution at the last position (36). This observation suggests the existence of a similar interaction in mammalian introns. Thus, the interaction between the first and last nucleotides may be a general feature of nuclear pre-mRNA splicing. Second, our analysis did not allow us to elucidate the precise hydrogen bond pattern of the interaction. In addition, the molecular elements in the spliceosome which are responsible for the stabilisation and/or the recognition of this interaction remain to be identified.

ACKNOWLEDGEMENTS
Special thanks are due to Bertrand Seraphin for helpful advice on primer extension experiments, communication of unpublished results and gift of yeast strains. We also thank Paul Siliciano for the kind gift of plasmids and for sharing unpublished results. James Haber and Alan Tartakoff are greatly acknowledged for the kind gift of plasmids and for sharing unpublished results.

REFERENCES

Downloaded from https://academic.oup.com/nar/article-abstract/22/11/1981/2400053/Interaction-between-the-first-and-last-nucleotides by guest on 15 September 2017


