RNA structural patterns and splicing: Molecular basis for an RNA-based enhancer

D. LIBRI, F. STUTZ, T. McCARTHY, and MICHAEL ROSBASH

1 Department of Biology, Howard Hughes Medical Institute, Brandeis University, Waltham, Massachusetts 02254, USA
2 Institut Pasteur, Unité de Génétique Moléculaire du Développement (C.N.R.S. URA1947), 25 rue du Dr. Roux, 75724 Paris Cedex 15, France

ABSTRACT

Efficient splicing of the 325-nt yeast (Saccharomyces cerevisiae) rp51b intron requires the presence of two short interacting sequences located 200 nt apart. We used the powerful technique of randomization–selection to probe the overall structure of the intron and to investigate its role in pre-mRNA splicing. We identified a number of alternative RNA–RNA interactions in the intron that promote efficient splicing, and we showed that similar base pairings can also improve splicing efficiency in artificially designed introns. Only a very limited amount of structural information is necessary to create or maintain such a mechanism. Our results suggest that the base pairing contributes transiently to the spliceosome assembly process, most likely by complementing interactions between splicing factors. We propose that splicing enhancement by structure represents a general mechanism operating in large yeast introns that evolutionarily preceded the protein-based splicing enhancers of higher eukaryotes.

Keywords: pre-mRNA; splicing; structure; yeast

INTRODUCTION

In nuclear pre-mRNA splicing, RNA–RNA interactions among snRNPs as well as between the snRNPs and the pre-mRNA substrate play a major role in recognition of proper splicing signals and in maturation of the spliceosome (Moore et al., 1993). A still unresolved issue, however, is the extent to which intramolecular RNA–RNA interactions in the pre-mRNA affect the splicing process. In higher eukaryotes, it has been shown that artificial hairpins can hinder splice site access both in vivo and in vitro (Eperon et al., 1986, 1988; Fu & Manley, 1987), but only one example is known of a naturally occurring secondary structure that affects splice site selection in vivo (Clouet-d’Orval et al., 1991; Libri et al., 1991). It has also been shown that secondary structure formation can improve the in vitro splicing efficiency of an adenovirus intron, possibly by reducing the spacing between the branch point and the 3' splice site (Chebli et al., 1989).

In yeast (Saccharomyces cerevisiae), short artificial hairpins that sequester the 5' splice site or the branch point are efficient inhibitors of splicing both in vivo and in vitro (Goguel et al., 1993). There is also a natural case in which a 3' splice site is skipped by virtue of being sequestered in a stem (Deshler & Rossi, 1991). The existence of natural pre-mRNA intramolecular interactions that favor the splicing of yeast introns was initially suggested based on sequence analysis of yeast introns (Parker & Patterson, 1987). These authors pointed out that the yeast introns could be divided into two main size classes: short introns (on average 90 nt long) and large introns (400 nt on average). They suggested that large introns contain short complementary segments whose interaction would have the effect of shortening the distance between the donor site and the branch point to an “operational” length similar to the one found in small introns. Newman (1987) subsequently showed that the interaction of two short regions in the CYH2 intron was essential for optimal splicing efficiency and suggested that the role of the interaction was to prevent one of the interacting segments from somehow inhibiting splicing. More recently, the ribosomal protein rp51b gene intron in yeast has been shown to contain two sequence segments whose interaction contributes to enhance splicing efficiency of this intron and influences splice site choice in competition experiments (Goguel & Rosbash, 1993).
Important questions remained unanswered by all of the previous studies. (1) The position of the inverted repeats (usually located some 200 nt apart) raises the question of the thermodynamic stability of such an interaction and suggests that the stem might be part of a more complex secondary structure. (2) The exact role played by this secondary structure in improving splicing efficiency is unclear. (3) The general applicability of splicing enhancement by secondary structure to other yeast introns is also uncertain.

Madhani and Guthrie (1994) combined in vitro randomization and in vivo selection techniques to investigate RNA–RNA intermolecular interactions in the spliceosome. We have extensively applied this methodology to investigate the role played by the rp51 intron intramolecular base pairing as well as its overall structural context. Our results point to the existence of a general mechanism by which intramolecular RNA–RNA interaction between the two ends of large yeast introns enhances splicing efficiency.

RESULTS
To investigate the role of pre-mRNA secondary structure in splicing of large yeast introns, we employed a reporter gene whose expression is dependent on splicing. The rp51b intron was inserted in the coding region of the CUP1 gene whose product confers copper resistance to yeast (Fig. 1A). Under these conditions, splicing is the limiting factor in the expression of the CUP1 product, and copper resistance of the strain correlates well with the excision efficiency of the intron (Stutz & Rosbash, 1994; our unpubl. data). The stem involved in splicing of the rp51 intron is shown in Figure 1B. We designated UB1 (upstream box 1) and DB1 (downstream box 1) the upstream and downstream constituents of the stem.

To investigate the sensitivity of splicing efficiency to minor alterations in the stem, changes of 3 nt (3mUB1), 5 nt (5mUB1), 4 nt (4mUB1), 6 nt (6mUB1), and 8 nt (8mUB1) were introduced in UB1; the sequence in DB1 was mutated in 3mDB1 and 5mDB1, which carry the compensatory mutations of 3mUB1 and 5mUB1, respectively (Fig. 2A). All the single mutants listed above are expected to weaken or disrupt the secondary structure, whereas the double mutants (3mUB1/3mDB1 and 5mUB1/5mDB1) should restore the stem. This set of mutants was tested both for copper sensitivity and for the pre-mRNA/mRNA ratio in primer extension experiments (Fig. 2C and data not shown).

Mutation of three or five base pairing nucleotides on both sides of the stem had an inhibitory effect on splicing of the intron. The inhibition was partially relieved in the double mutant 5mUB1/5mDB1, although not to levels sufficient to support copper-resistant growth as well as the wild-type construction (5mUB1/5mDB1 grows only as 5mUB1 although better than 5mDB1; Fig. 2C). However, no compensation was observed for the double mutant 3mUB1/3mDB1. Even more surprising was mutant 8mUB1, which caused no significant drop in splicing efficiency despite containing an extension of the same mutations introduced in 3mUB1 and 5mUB1.

One possible explanation for the unexpected splicing phenotypes is that these mutations cause significant changes in pre-mRNA structure, which affects splicing efficiency either positively or negatively. For example, the mutations 3mDB1 and 5mDB1 might induce the formation of an alternative inhibitory stem near the 3' splice site (Fig. 2B), which would be favored over the reestablishment of the original base pairing. The presence of an appropriately located reverse transcriptase strong stop with RNA templates containing the DB1 mutations (Fig. 2B and data not shown) supports this hypothesis. On the other hand, the mutant 8mUB1 might exhibit an unexpected efficient splicing phenotype because the mutated bases can find a different and permissive base pairing partner as confirmed below.

Multiple and alternative intramolecular RNA interactions affect splicing efficiency
If multiple foldings of the intron are compatible with efficient splicing, their analysis might be informative about the mechanism that relates them to splicing ef-
RNA secondary structures and splicing enhancement

To this end, the same 8 nt mutated in 8mUB1 were randomized, and the pool obtained was introduced into yeast. A similar experiment was done by randomizing the 5 nt mutated in 5mUB1 (Fig. 3). The pools were selected by yeast growth in copper-containing liquid media or by growth on copper plates, and their evolution was monitored by sequence analysis at intermediate time points of the liquid assay (data not shown).

The pool containing five randomized bases gave an essentially homogeneous population containing mainly the wild-type sequence (Fig. 3). Analysis of the clones selected on copper plates confirmed this result. The pool containing eight randomized bases, however, gave rise to “winning” sequences that were assigned to four families (named selected box 1–4, SB1–SB4), present to different extents in the total pool.

One of the families, SB1, contained sequences that are variants of the wild-type UB1. A second family (SB3) contained sequences identical to that of 8mUB1. Analysis of the two remaining families and inspection of the sequence of the rp51b intron revealed the existence of potential base pairing regions for every family. The proposed stems are shown in Figure 3 and are compatible with the nucleotide variations observed within the families. (Only SB3, with one family member, has no variation.) Note that family SB4 is predicted to interact with a sequence immediately downstream of the UACUAAC box.

In Figure 4A, SB1–SB4 are represented by different shaped boxes, whereas the putative complementary regions are named DB1–DB4 and drawn as matching boxes, according to their interacting counterparts. These interactions have been tested by mutating every downstream box, either in the presence of the matching upstream selected box or in the presence of the wild-type sequence UB1. The plasmids were tested for their ability to confer a copper-resistant growth phenotype to yeast as well as by primer extension (Fig. 4B and data not shown).

Confirming the importance of the pairing interactions, mutations in DB2, DB3, and DB4 impaired splicing in the presence of SB2, SB3, or SB4, respectively; the same modifications in a wild-type (UB1) context did not reduce splicing efficiency (as expected because the original base pairing [UB1/DB1] should not be affected by the modifications). In some experiments (Fig. 4B), mutation of DB4 seemed to affect splicing efficiency also in the presence of the wild-type sequence in the upstream box (UB1), although less markedly than in an SB4 context. Given the proximity of DB4 to the UACUAAC box, it was not unexpected that some sequences in this position might have a general negative effect on splicing efficiency. To ensure that the observed effect of mutating DB4 in the presence of SB4 was indeed a result of disrupting their base pairing, we constructed two new pools of sequences in which randomization of eight bases in the DB4 region (dB4N) was combined either with the wild-type sequence in UB1 (WT/dB4N) or with SB4 (SB4/dB4N).

The comparison of the growth curves, the plating efficiency of the two pools in the presence of copper, and

FIGURE 2. A: Positions of the mutations introduced in UB1 and DB1. Wild-type sequence is indicated in capital letters. Mutant 3mUB1 introduces four changes instead of three due to a mistake in the original published sequence (Abovich & Rosbash, 1984). However, we decided to maintain the name to avoid confusions with 4mUB1. Note that an A to G change in position 33 does not affect splicing efficiency (see Fig. 3). B: Proposed inhibitory secondary structure for mutant 5mDB1. Mutations introduced are shown in lowercase letters; the branch point (BP) is indicated and the AG acceptor is underlined. Arrow indicates the position of a strong reverse transcriptase stop in primer extension experiments (data not shown). C: Copper growth assay for the mutants of the UB1/DB1 stem. Left part of the figure is a schematic drawing showing the positions of the different mutants. CuSO₄ concentration is indicated under each panel. The higher copper-resistant phenotype displayed by 5mUB1 (and 5mUB1/5mDB1) as opposed to the other supposedly less severe mutants of the stem is most likely due to imperfect base pairing of the 5mUB1 with DB3 (data not shown). The latter mutants can be fully distinguished from the Wt and 8mUB1 at 1.6 mM Cu²⁺ (see Fig. 6B).
the copper sensitivity of randomly chosen individual clones confirmed the SB4/DB4 proposed interaction (data not shown). Analysis of individual sequences from the two pools after copper selection demonstrated the absence of a significant evolution of the Wt/db4N pool but revealed the existence of strong selection pressure on the DB4 region in the presence of SB4. Of the three “winning” families in the SB4/db4N pool, one resembles closely the starting sequence (wild-type DB4), a second contains sequences that can base pair with an alternative site in the intron, and the third contains highly U-rich sequences (see Discussion).

How complex is the pre-mRNA structure?

The middle of the intron (the roughly 200 nt between UB1 and DB1) might also be structured and contribute significantly to the free energy of intron folding. For example, the sequence in UB1, when submitted to randomization-selection, might only “find” base pairing partners in the proximity of DB1 because of additional structural features that juxtapose the two ends of the intron. The putative presence of a more complex intron structure predicts that other regions of the intron will be sensitive to mutations. Because only three mutations in the stem defined by UB1 and DB1 induced a significant drop in splicing efficiency (mutants 3mUB1 and 3mDB1) and even a single nucleotide change could modify the copper sensitivity in some experiments (data not shown), we reasoned that light mutagenesis of the whole intron should be sufficient to perturb optimal folding.

Therefore, we constructed a “doped” pool in which the sequence of the wild-type intron was mutagenized at the level of about 3–6% per position in two independent experiments. Under these conditions, about 75%, 40%, and 15% of the sequences have at least one, two, or three changes, respectively, in any stem of length comparable to the UB1/DB1 interaction. The splice sites and the UACUAAC box were excluded from the mut-
FIGURE 4. A: Schematic drawing showing the sequences selected in UB1 (SB1–SB4) and their interacting partners (DB1–DB4) in the downstream portion of the intron. Every family of selected clones is indicated by a differently shaped box and the corresponding downstream elements are indicated as matching boxes. The effect of every downstream box mutation, in the presence either of the matching upstream box or the wild-type UB1, is shown in the summary of the copper sensitivity of the different clones. Mutated sequences are indicated by a rectangle and lowercase characters (e.g., Wt/db2 contains a mutation in the DB2 box, and is otherwise wild type). B: Primer extension analysis of RNAs from the selected clones and their mutant derivatives. Last lane (labeled Wt/db4) comes from a different experiment and shows a somewhat different pre-mRNA/mRNA ratio (see text).
regions at different rates) or as a consequence of random variation (as obtained in a computer simulation of the experiment, data not shown).

The plot in Figure 5A shows the mutation frequency of the doped pool after selection on copper (24 sequences). As expected, two "cold" spots were present at UB1 and DB1. Analysis of the few mutations found in these regions strengthened the wild-type base pairing model (Fig. 2A; data not shown). A few additional cold spots were identified whose functional relevance was tested by site-directed mutagenesis in four cases (Fig. 5A). Random sequences were introduced at every site, and a few arbitrarily chosen clones were tested for survival on copper plates. All the clones displayed

![Graph showing mutation frequency](image-url)
normal copper-resistant growth (data not shown). Moreover, the percentage survival of the pools on copper plates was indistinguishable from that of wild type. Because similar cold spots were also observed in other experiments (Fig. 5b and data not shown), we conclude that they were biases generated in the initial pool by an uneven distribution of introduced mutations. Together with experiments presented below, the results argue against the involvement of a more complex pre-mRNA structure in the splicing of the rp51b intron.

What is the function of the UB1/DB1 interaction?

The interaction UB1/DB1 might function to neutralize a negative element, such as the inhibitory binding of a protein or an alternate folding that might limit splice site accessibility. When this interaction is disrupted, as in mutant 5mUB1 for example, the “poison” sequence (perhaps in collaboration with a protein) would actively block splicing. Restoration of efficient processing should thus be possible by mutation of the poison sequence as well as by reconstitution of the original (or an equivalent) pairing. To examine this possibility, we performed a selection experiment on a doped pool containing the mutation 5mUB1, which disrupts the interaction UB1/DB1. Winners of the selection are expected to display high mutation frequencies in the regions of UB1 and DB1 (due to the reestablishment of the pairing), and additionally, “hot” spots might be expected where poison sequences are located. About 50 winner clones were fully sequenced, and the distribution of mutations is plotted in Figure 5B.

Two major peaks fall in the UB1 region, and a third is in the DB1 region. Closer inspection of the sequences revealed that all selected clones could be divided into two mutually exclusive families: in the first (around 70% of the total), all the sequences contain an A to G mutation at position 25 of the intron; in the second (30% of the total) a CT doublet (positions 30–31) belonging to the original mutation in 5mUB1 reverted to the wild type GA. Both the “G” and the “GA” classes still displayed a highly mutagenic peak in the region of DB1 when the distribution of mutations was plotted separately (data not shown). Moreover, the G or GA mutations alone are not sufficient to restore splicing in 5mUB1; the clustered mutations in the DB1 region are also required (data not shown; note also that the poor growing mutant 3mUB1 has a GA in the same position as the GA clones). The analysis of the covariances fully supports the reformation of the base pairing interactions between the regions of UB1 and DB1 (shown in Fig. 5C for the GA clones).

The almost complete coincidence of the hot spots with the two regions involved in base pairing argues against the existence of a negative cis-acting element whose action is somehow counteracted by the interaction of UB1 with DB1.

In vivo splicing of artificial substrates and their evolution

The absence of an ordered inhibitory state for the structure mutants (i.e., the apparent absence of sequence information which would actively inhibit splicing) suggests that splicing of a randomly chosen long intron sequence should be inefficient (i.e., the default state is inefficient splicing). To test this hypothesis, two arbitrarily chosen sequences were inserted between the UACUUAAC box and the 5' splice site of the wild-type intron. The constructions b/bl/b and b/Tm/b are two chimeric introns containing an inversion of the whole central portion of the rp51b intron and a fragment of the chicken β-tropomyosin gene in inverted orientation, respectively. Consistent with expectation, neither construct conferred substantial copper-resistant growth to yeast, and this phenotype was due to poor splicing efficiency as verified by primer extension experiments. A third construct containing another tropomyosin fragment gave essentially similar results (data not shown).

To test whether poor splicing was due to the lack of some limited structural information, we performed an in vivo selection experiment similar to the one already performed on the wild-type and the 5mUB1 mutant intron. Two doped pools were constructed from the starting chimeric intron sequences. Two clones were isolated from the evolution of b/Tm/b (b/Tm/b.S1, six times, and b/Tm/b.S6, once), and five different clones were chosen among the “winners” of the b/bl/b evolution (survival rate 10⁻³–10⁻⁴). The improvement in splicing efficiency after evolution varies from 2.5-fold for the b/bl/b based clones to almost 15-fold for b/Tm/b.S6 (data not shown; Fig. 6A,B).

Clones b/Tm/b.S1 and b/Tm/b.S6 share the same G to A change at position +7 of the intron (i.e., the first position outside the conserved 6-nt sequence for the 5' splice site), which likely improves base pairing with U1 (Rosbash & Seraphin, 1991). When introduced into an otherwise unmodified b/Tm/b intron, this change improved splicing efficiency, but not to levels sufficient to achieve the same copper-resistant growth phenotype as b/Tm/b.S1 and b/Tm/b.S6. For all the selected clones, pairings similar to the ones previously described for the rp51b intron can be drawn; the one for clone b/Tm/b.S1 is shown in Figure 6C. The results indicate that even very limited in vivo evolution of an inefficiently spliced artificial intron can improve splicing efficiency up to 15-fold. Taken together, the data indicate that structural modifications of the pre-mRNA account for this improvement.

Intramolecular interactions can be designed to improve splicing in chimeric introns

To investigate the possibility of designing de novo a structural element required to achieve efficient splicing
in the two artificial introns, we introduced in the 5' end of both b/bi/b and b/Tm/b introns a 15-nt mutation that would be complementary to a region upstream of the UACUAAC box; the position of these inverted repeats in the artificial introns is identical to the position of UB1 and DB1 in the wild-type intron with respect to the 5' splice site and the UACUAAC box. The construction derived from b/bi/b (b/bi/b.St) displayed a copper-resistance growth phenotype and a splicing efficiency comparable to the clones derived by in vivo selection of the same intron (Fig. 6B). In contrast, the modified intron derived from b/Tm/b (b/Tm/b.St) was unable to confer copper-resistant growth to yeast (Fig. 6B), suggesting perhaps that this stem is not allowed to form in vivo or is not fully compatible with the splicing enhancement mechanism.

**DISCUSSION**

**The wild-type pairing**

The presence of an intramolecular interaction within the rp51b intron was first proposed by Parker and Patterson (1987) and more recently demonstrated by Goguel and Rosbash (1993) in this laboratory. The open questions of the overall structural context of this interaction and of its mode of action have been addressed in this report using the sensitive and highly informative technique of in vivo selection.

The results of two different experiments allowed us to confirm the previously reported interaction of UB1 and DB1 (Goguel & Rosbash, 1993). First, the molecular evolution of 8 nt contained in UB1 yielded four fam-
ilies of sequences, one of which included the wild-type sequence. All the nucleotide variations contained in
this family either maintained or improved the strength of the original interaction. For example, a bulged U
present in the original stem disappeared in most of the selected sequences (Fig. 3 and data not shown). Sec-
ond, evolution of the doped wild-type intron gave a very low frequency of mutations in UB1 or DB1 (about
half the expected value; Fig. 5A and data not shown), almost all of which are compatible with the proposed
stem. Only two positions were modified in UB1 (a total of 5 clones out of the 25 sequenced): U_{25} to A (three
clones), which most likely allows stronger stacking inter-
actions with the neighboring purines, and G_{39} to A
two clones), which gives rise to a more canonical A–U
interaction. This same base pair is affected by a muta-
tion in DB1 (U_{251} to C, two clones), which allows for-
mation of a more stable G–C. Only one nonconservative
change is observed in these two regions (U_{247} to A,
one clone), and it is located in a region that most likely
does not provide a major contribution to the overall
stem stability.

A stronger interaction might not be necessarily
favored for splicing enhancement

The molecular evolution of UB1 and some site-directed
mutagenesis experiments revealed the existence of at
least three additional base pairing possibilities for se-
quencies in UB1. One intriguing observation stems
from the analysis of these interactions and their vari-
ation. Because most of the winners of the UB1 pool
evolution have been competitively selected in copper-
containing medium before separation on nonselective
plates, their relative abundance should reflect relative
growth advantage, due in principle to stronger inter-
actions between the two regions of the intron. Indeed,
sequences with lower base pairing potentials are in
general underrepresented within a given family. There
are, however, some exceptions.

In the first family of sequences (interaction SB1/DB1),
the best base pairing sequence is only half as abundant
as the most frequent family member (Fig. 3). In the
second family (interaction SB2/DB2), the most abun-
dant clone contains a C–A mismatch in the center of
an otherwise perfect 13-nt stem. Clones that contain a
Watson–Crick interaction at this position exist, but they
are not four or five times less abundant and contain at least
one additional noncanonical interaction. Because the
most abundant second family clone accounts for 40%
of the family (and 25% of the whole selection), it is very
unlikely that its high frequency derives from an initial
bias in the starting pool. Indeed, a comparison of the
copper growth curves of the first and third most abun-
dant clones of this family (which have almost identical
sequences with the exception of a G–C pair instead
of the A–C mismatch, and a G–U instead of an A–U
pair; Fig. 3) directly proved the existence of a selective
advantage for the clone containing the less stable stem
(data not shown). This shows that a stronger inter-
action does not necessarily favor higher splicing effi-
ciency. The extent of splicing enhancement is also
likely to depend on the sequence of the stem as well as
its stability (data not shown; Goguel & Rosbash,
1993).

Stem disruption as well as stem formation might also
be limiting processes for splicing efficiency enhance-
ment. This disruption is most likely accomplished be-
fore the first splicing step, at least in the case of the
sequences belonging to the fourth family (interaction
SB4/DB4; Fig. 3) and possibly those of the evolved arti-
ficial intron b/Tm/b.S1 (Fig. 6C). In both cases, one of
the two interacting sequences is located downstream
of the branch point and maintaining this interaction
might be incompatible with function in a branched
structure.

How complex is the intronic secondary structure?

The experiments presented in this report indicate that
a wide range of structural patterns that improve splic-
ing efficiency share the feature of having the 5′ and 3′
end of the intron in close proximity. It may seem sur-
prising that the UB1 or the 5mUB1 doped pool evolu-
tions do not identify shorter range interactions, which
are expected to be favored by thermodynamic criteria.
Particularly striking is the case of the UB1 evolution,
because essentially all possible sequences are expected
to be contained in the starting pool and, taking into ac-
count the contribution of the nucleotides flanking the
randomized bases, a number of alternative interaction
sites scattered throughout the intron can be identified
by inspection of the sequence.

These observations can be explained by two alterna-
tive hypotheses: (1) UB1 pairing is restricted to the DB1
region (the region between DB4 and DB3) because there
is a more complex intronic structure that restricts
pairing to these two regions. Undetermined features
of this complex folding account for the effect on splic-
ing efficiency. (2) There is functional selective pressure
that favors base pairings between the two ends of the
introns, because their close proximity is the feature that
enhances splicing efficiency. A distinction between the
two models is the extent to which the region between
UB1 and DB1 is structured.

The experiment in which a high number of func-
tional intron variants were selected from a doped pool
derived from the wild-type sequence (Fig. 5) argues
against the presence of a complex structure; the few
cold spots identified (besides UB1 and DB1) were in-
sensitive to site-directed mutagenesis (data not shown),
indicating that their occurrence is likely due to an un-
even distribution of mutations in the starting pool.
The finding that a small stem in an artificial intron
(b/bI/b.St, which is very unlikely to share the same structural organization as the wild-type rp51b intron) is sufficient as well as necessary for efficient splicing also argues against the existence of a more complex structure.

The data indicate that the interaction of UB1 and DB1 (or any of the similar possible interactions) is the primary contributor to splicing efficiency enhancement. Its ΔG is mostly independent from other free energy contributions and likely near an optimum value.

**The role of the interaction**

The evolution of the doped pool obtained by light mutagenesis of a non-base pairing mutant (5mUB1) failed to reveal any event other than the formation of a base pairing interaction between the same regions of the intron. Given the frequency of the G and GA clones and the number of sequences analyzed, we would likely have detected events from as frequent as the reestablishment of a pairing (about 1/1,000) to probably 10 times less frequent, i.e., any actively inhibited splicing state that would be sensitive to the simultaneous mutation of 3–4 nt. The results argue against the existence of an “ordered” nonfunctional state characterized by the existence of negative cis-acting elements in the rp51b intron; rather, they suggest a direct and positive role for intramolecular RNA–RNA interactions in large yeast introns. Using these same in vivo evolution techniques, we found very similar structural elements in the otherwise poorly conserved rp51a intron (data not shown; see also Pikielny & Rosbash, 1985); together with other data presented in this report and in the literature (Newman, 1987; Parker & Patterson, 1987), the results suggest that this mechanism is of general significance for large yeast introns.

We favor the hypothesis that the base pairing contributes to the overall strength of 5′ splice site–branch point interactions and acts cooperatively with splicing factor interactions that bridge these two intronic regions. The increased local concentration of binding sites and factors would contribute to lowering the entropic cost of one or more interactions that take place during spliceosome assembly. For example, interactions between yeast U1 snRNP and the yeast Mud2 protein might be aided by the pre-mRNA intramolecular base pairing. Both factors are known to bridge interactions between the branch site and the 5′ splice site. Mud2p is known to associate functionally with U1 snRNP, to have homology with mammalian U2AF, and to crosslink to pre-mRNA (Abovich et al., 1994). Although the binding site of Mud2p has not yet been identified, like U2AF it might be a U-rich region near the branch site. It is intriguing to note that one of the families emerging from the evolution of the DB4 region (immediately downstream of the branch point) contains an extremely U-rich sequence that might partially overcome the usual requirement for a base-paired stem (see also Patterson & Guthrie, 1991). Members of the family defined by the SB4/DB4 interaction also exhibit a very poor growth phenotype in the absence of the Mud2 protein (data not shown), suggesting a tight link between this particular structural pattern and the binding of Mud2p. It is interesting to note that none of these elements (the U-rich tract, Mud2 or the base pairing interaction) is absolutely required for splicing, but they all contribute to its efficiency.

Splicing enhancement by structure might be considered an evolutionarily “primitive” enhancer mechanism, whose presence may be related to a less prominent role played in yeast by SR or HnRNP proteins (or their as yet unidentified equivalents). In higher eukaryotes, these proteins are believed to be involved in splice site recognition and pairing, by contributing to spliceosomal protein–protein interactions (Wu & Maniatis, 1993; Amrein et al., 1994; Staknis & Reed, 1994) and by helping to neutralize nonproductive pre-mRNA folding patterns (Burd & Dreyfuss, 1994). In this context, it is interesting to note that negative secondary structures of the pre-mRNA seem to play a more prominent role in yeast than in higher eukaryotes (Eperon et al., 1986, 1988; Fu & Manley, 1987), which might also be related to the striking intron-length dependence of splicing in yeast (Klinz & Gallwitz, 1985).

Although more detailed considerations must await in vitro experiments, this view suggests that the base pairing interaction may catalytically or thermodynamically enhance yeast pre-mRNA splicing in a manner that resembles the proposed roles of HnRNP or SR proteins in mammalian systems. It also emphasizes the related roles of RNA–RNA and RNA–protein interactions (e.g., Mohr et al., 1994) and underscores the dual origins of nuclear pre-mRNA splicing: the ancient RNA world and the more recent world of protein catalysis.

**MATERIALS AND METHODS**

**Cloning, mutagenesis, and pool constructions**

The rp51b intron was cloned between a BamHI I and a Sma I site in a reporter construct containing the yeast CUP1 gene (Stutz & Rosbash, 1994) so that the starting ATG is located immediately upstream of the donor site.

Site-directed mutagenesis has been performed either as described in Kunkel et al. (1987) or as follows: a mutagenic oligonucleotide, an oligonucleotide containing one of the cloning sites (e.g., Sma I), and 100 ng of the wild-type template were used to obtain a primary PCR product. One of the strands of this product was subsequently used as primer in a secondary PCR reaction that contained, besides the primary PCR product, a wild-type template and an oligonucleotide corresponding to the second cloning site (e.g., BamHI I). The resulting PCR product was subsequently cloned in the CUP
reporter plasmid. In most cases, Vent polymerase was used to minimize unwanted errors.

The pools of sequences containing random nucleotides in the UB1 region were constructed by cloning a PCR product obtained with a degenerate 5′ oligonucleotide (containing an Nco I site) and a 3′ primer (containing a Sma I site) between the Sma I site and an Nco I site created by mutagenesis in the reporter gene. All the other pools containing completely randomized regions were obtained using the mutagenesis procedure described above. In all cases, the PCR product containing the degenerate sequences was cloned in a vector prepared from a plasmid containing a nonfunctional intron. This effectively ruled out the presence in the cloned pools of functional introns that do not derive from the mutagenesis procedure. The percentage of nonfunctional introns deriving from the vector (uncut or cut only once) was generally lower than 1%.

The doped pools were constructed by mutagenic PCR on the whole intron essentially as described by Cadwell and Joyce (1992) and modified by Bartel and Szostak (1993). The conditions described by Leung et al. (1989), which allow a higher mutagenic rate but with a strong AT to GC bias, were also employed in duplicate experiments. The mutation rate was estimated a priori by monitoring the disappearance of a restriction site in the intermediate PCR products. The mutation rate of the different pools was as follows: 3–6% (wild-type intron); 4–10% (intron containing the mutation 5mUB1); and 5% and 8%, respectively, for the evolution of b/Tm/b and b/bI/b. The splice sites were excluded from mutagenesis and the UACUAAC box was “repaired” by subsequent site-directed mutagenesis on the final doped pool.

A reduced mutational bias was observed with the conditions described by Cadwell and Joyce (1992) as opposed to the conditions of Leung et al. (1989). However, A/T to X/Y mutations occurred with a frequency about four times higher than G/C to X/Y. Notably, no G/C to C/G changes were observed. A positional bias was also observed in the distribution of mutations, maybe due to some sequence-dependent variations in the polymerase speed (see text).

The artificial intron b/bl/b was constructed by inverting the intron sequence between positions +12 and +281 with appropriately designed oligonucleotides. The intron b/Tm/b was obtained by cloning a fragment from the chicken tropomyosin gene (sequence positions 7401–7593; Libri et al., 1989) in the antisense orientation between the same intron positions in the reporter gene (UB1 and DB1 are not included in the chimeric introns). The chimeric introns containing the artificially designed inverted repeats were obtained by mutagenizing 15 nt starting at intron position 31 and by introducing changes complementary to a sequence located at position −41 with respect to the UACUAAC box.

Transformations

The pools of sequences were first transferred to bacteria by electroporation and the DNA prepared was then transferred to yeast by lithium acetate/PEG transformation as described (Gietz et al., 1992). The complexity of the pools varied between $10^4$ and $10^5$. The transformants were first plated on synthetic Leu− plates and harvested as soon as colonies were visible. After overnight growth of representative aliquots of the pools, they were submitted to copper selections.

Copper-selection assays

All selections and growth assays on plates were performed at CuSO4 concentrations of 1.2, 1.4, and 1.6 mM. The selective concentrations employed in liquid media were 0.4–0.8 mM. This last competitive assay proved to be extremely sensitive, being able to detect growth differences almost undetectable by other methods.

Sequence analysis

Most sequencing was performed directly on PCR products derived from nucleic acid preparations or cell suspensions of individual clones or pools, with the femtomol sequencing kit (Promega).

The “winners” from the doped pool selections were analyzed as follows: 25 clones were sequenced from two independent experiments (a total of 50); all the sequences obtained were introduced in Microsoft Excel worksheets and the total number of mutations contained in each window of 6 nt starting from every position in the intron has been plotted to obtain the graphics shown in Figure 5A and B.

The secondary structure analysis was performed either by eye or by searching the whole intron sequence for homologies to the degenerate complement of the sequence of interest.

Yeast strain

All the experiments were performed using the copper-sensitive yeast strain Y69ΔCUP1ΔB::URA3 (α, leu2-3, leu2-112, ura3-52, trp1-289, his3-Δ1, cyhR, ΔCUP1, Δrp51B::URA3). This strain was obtained by deleting the tandemly repeated X-CUP1 genes from strain MGD 353-46D (Séraphin et al., 1988) as described earlier (Stutz & Rosbash, 1994). In a second step, the nonessential rp51B gene was completely deleted by homologous recombination using a rp51B genomic construct in which the rp51B sequences between the EcoR V site (position −210) and the Hinc II site in the 3′ flanking region (Abovich & Rosbash, 1984) had been replaced by the URA3 gene.

Various

Primer extension of total RNAs and ligations was performed essentially as suggested by the enzyme manufacturer (Gibco-BRL and USB, respectively). For some experiments, a thermostable reverse transcriptase from Epicentre was employed. Taq and Vent polymerase were purchased from Boehringer Mannheim and New England Biolabs, respectively.

ACKNOWLEDGMENTS

We thank D. Bartel, M. Buvoli, M. Moore, C. Pikielny, and H.V. Colot for critical reading of the manuscript and all the members of the Rosbash lab for fruitful discussions. The technical assistance of D. Johns-Gordon and the expert secretarial help of L.A. Monaghan are kindly acknowledged. D.L. is
supported by the C.N.R.S., France. This work was supported in part by a grant from the National Institutes of Health to M.R. (GM 23549).

Received April 24, 1995; returned for revision May 8, 1995; revised manuscript received May 15, 1995

REFERENCES

Seraphin B, Kretzner L, 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.
Slaknis D, Reed R. 1994. SR proteins promote the first specific recognition of pre-mRNA and are present together with the U1 small nuclear ribonucleoprotein particle in a general splicing enhancer complex. Mol Cell Biol 14:7670–7682.

D. Libri et al.