Published online 5 September 2011 *Nucleic Acids Research, 2011, Vol. 39, No. 22* ***9759–9767***
doi:10.1093/nar/gkr662

Secondary structure is required for 3′ splice site recognition in yeast

Ondřej Gahura[1], Christian Hammann[2], Anna Valentová[1], František Půta[1] and Petr Folk[1,*]

[1]Department of Cell Biology, Faculty of Science, Charles University in Prague, Prague, Czech Republic and [2]Heisenberg Research Group Ribogenetics, Faculty of Biology, Technical University Darmstadt, Darmstadt, Germany

Received May 23, 2011; Revised July 6, 2011; Accepted July 27, 2011

ABSTRACT

Higher order RNA structures can mask splicing signals, loop out exons, or constitute riboswitches all of which contributes to the complexity of splicing regulation. We identified a G to A substitution between branch point (BP) and 3′ splice site (3′ss) of *Saccharomyces cerevisiae COF1* intron, which dramatically impaired its splicing. RNA structure prediction and in-line probing showed that this mutation disrupted a stem in the BP-3′ss region. Analyses of various *COF1* intron modifications revealed that the secondary structure brought about the reduction of BP to 3′ss distance and masked potential 3′ss. We demonstrated the same structural requisite for the splicing of *UBC13* intron. Moreover, RNAfold predicted stable structures for almost all distant BP introns in *S. cerevisiae* and for selected examples in several other *Saccharomycotina* species. The employment of intramolecular structure to localize 3′ss for the second splicing step suggests the existence of pre-mRNA structure-based mechanism of 3′ss recognition.

INTRODUCTION

Ribonucleic acid is considered to be the earliest as well as the most versatile information polymer that carries sequential information and forms higher order structures. Precursor mRNA replicas of protein-coding genes are processed by the spliceosome to remove introns; this maturation phase occurs, at least with some transcripts, in all eukaryotic cells studied so far. Spliceosomal introns are marked by four splicing signals on the level of nucleotide sequence: the 5′ splice site (5′ss), branch point (BP), polypyrimidine tract (pY-tract), and 3′ splice site (3′ss). These signals by themselves, however, are not sufficient to predict a splicing event in the Metazoa. Additional inputs, including the propensity of pre-mRNA to attain a thermodynamically stable fold, are required (1). The most context dependent is the recognition of 3′ss, distinguished only by the sequence AG, whereas the positions of 5′ss and BP are demarcated by seven and five-nucleotide sequences, respectively. 3′ss is also the last signal to be recognized in the splicing cycle.

The ability of certain pre-mRNAs to form intramolecular secondary structures that affect the outcome of splicing was recognized more than 25 years ago (2,3). Various types of such structures have since then been shown to impact both the constitutive and alternative splicing in many species (4,5).

(A) The linear sequences can be base paired to complementary regions in stems/helices whereby splicing signals or enhancer/silencer motifs are blocked from recognition by snRNAs or RNA-binding proteins. In the human *SMN2* gene, a secondary structure involving 5′ss of exon 7 hinders the interaction with U1 snRNA and leads to exon exclusion (6).

(B) Distant splicing signals of long introns, which are on a threshold for recognition, can be brought to proximity and thus made available for the spliceosome. In the two tandem introns of the *YL8A* gene, both of which form stems between 5′ss and BP, the swapping of complementary sequences between introns causes exon skipping (7).

(C) Interactions over yet a longer range may loop out whole exons and induce complex patterns of alternative splicing. A conserved stem structure was responsible for alternative exclusion of exon 5 in the *Drosophila Nmnat* gene (8).

(D) Higher order structure-epitopes may bind regulatory proteins or small metabolites such as riboswitches (9). *Saccharomyces cerevisiae RPL30* transcript folds in a structure that binds the gene's product L30. The L30 protein then blocks spliceosomal

*To whom correspondence should be addressed. Tel: +420221951765; Fax: +420221951758; Email: folk@natur.cuni.cz

rearrangements required for U2 snRNP mediated BP-region recognition and hence inhibits the splicing of its own transcript (10).

The yeasts of the *Saccharomyces* genus 'sensu stricto' (11) belong to intron-poor organisms with splicing limited to only ~5% of their genes (12). The introns, mostly one per gene, reach up to ~1000 nt in length. Some introns with long BP to 5′ss distance, e.g. *RPS17B*, require a secondary structure within pre-mRNA for efficient BP recognition and first step-spliceosome assembly (13–15). The recognition of 3′ss also depends on the number of nucleotides separating BP and 3′ss, but this is not well understood at present. Artificially extending the BP to 3′ss distance of *ACT1* intron in *S. cerevisiae* to ~120 nt completely abolishes splicing (16). However, there are 22 introns in *S. cerevisiae* that have BP to 3′ss distance longer than 60 nt (Saccharomyces Genome Database) in which the spliceosome has to rely on additional mechanisms for 3′ss recognition.

Here, we present the comparison of splicing efficiencies of the wild-type and manipulated *COF1* and *UBC13* introns which have a long BP to 3′ss distance. Our data suggest that a stable stem-loop forms between BP and 3′ss in these introns. We show that the secondary structure is essential for the recognition of the proper 3′ss by shortening the structural distance between BP and 3′ss and by masking BP-proximal cryptic 3′ss. As RNA structure analysis tools also predict structures in other long *Saccharomycotina* introns as well, we reason that these and perhaps also other organisms use a pre-mRNA structure-based mechanism of 3′ss recognition.

MATERIALS AND METHODS

Yeast strains, media and growth conditions

Primer extension experiments were performed using *S. cerevisiae* strain EGY48 (*MATα his3 trp1 ura3 LexAop(x6)-LEU2*) (17). Strain 46ΔCup (*MATa ade2 cup1Δ::ura3 his3 leu2 lys2 trp1 ura3, GAL+*) (18) was employed for an *in vivo* copper sensitivity splicing assay. Cells were grown in YPD plus adenine or in synthetic complete drop-out media supplemented with the required amino acids at 30°C. For testing of Cu^{2+} resistance, cells expressing *CUP1* fusion reporter were cultivated to OD_{600} approximately 0.4, concentrated to OD_{600} 4, spotted in 8-fold dilution series on plates with media containing the indicated concentration of $CuSO_4$, and cultivated for 3 days.

Construction of splicing reporters

All *CUP1*-based reporters were expressed from replicative p423GPD vector. Plasmid constructs used in this study are listed in Supplementary Data (Supplementary Table S2). *COF1-CUP1* reporter was constructed as follows: 221 bp fragment of *COF1* gene (exon 1 including 13 nucleotides upstream of the translation start codon, intron and 15 nucleotides of exon 2) and complete coding sequence of *CUP1* gene was amplified from genomic DNA using polymerase chain reaction (PCR) and primer pairs OG44/OG45 and OG46/OG47, respectively (Supplementary Data, Supplementary Table S3) and TOPO-TA cloned into pCR®II-TOPO® vector (Invitrogen). BamHI/EcoRI fragment of *COF1* and EcoRI/SalI fragment of *CUP1* were inserted into p423GPD vector, resulting in 416 bp *COF1-CUP1* fusion expressed from *TDH3* promoter. *COF1-CUP1* reporters with single-nucleotide substitutions were generated in p423GPD vector by site-directed PCR-based mutagenesis using QuikChange® II Site-Directed Mutagenesis Kit (Stratagene) and primers listed in Supplementary Table S3. DNA fragments encoding *COF1-CUP1* reporters containing internal intron deletions and all *UBC13-CUP1* reporters were synthesized commercially by GeneArt (Germany) and inserted into BamHI/SalI-digested p423GPD. *UBC13-CUP1* reporter contained *UBC13* fragment (exon 1 including 19 nucleotides upstream of the translation start codon, intron and 23 nucleotides of exon 2) and complete coding sequence of *CUP1*.

Primer extension analysis

Cells harboring reporter plasmid were cultivated to OD_{600} approximately 0.5–0.8 and harvested. Total RNA was isolated by MasterPure™ Yeast RNA Purification Kit (Epicentre Biotechnologies). Primer extension reactions were performed with the RevertAid™ M-MuLV Reverse Transcriptase (Fermentas) on 3–4 μg of total RNA. The reactions were primed using the oligonucleotide YAC6, annealing to the 5′-end of *CUP1* ORF, and YU14, complementary to U14 snoRNA. Primers were radiolabeled on 5′-ends by phosphorylation using T4 Polynucleotide Kinase (Fermentas) and [γ-^{32}P]ATP (3000 Ci/mmol; MP Biomedicals). The products were separated on 8% polyacrylamide/7 M urea gels and visualized by phosphorimager. The identities of selected bands (see 'Results and Discussion' section) were confirmed using 5′-RACE System for Rapid Amplification of cDNA Ends (Invitrogen).

RNA in-line probing

Information on the generation of DNA templates for RNA *in vitro* transcription, preparation of RNA, RNA end-labeling and the generation of RNA-ladders is provided in Supplementary Data. In-line probing was carried out essentially as described in (19). To monitor the stability of the various intronic sequences, RNAs were incubated for 45 h at temperatures of 10°C, 20°C, 30°C or 37°C. The incubations were terminated by the addition of an excess of gel loading buffer. The products of spontaneous RNA degradation were separated, together with non-treated RNA and the products of the RNase T1 digest, on denaturing polyacrylamide gels containing 7 M urea. Gels were run at a limiting current of 25 mA for at least 8 h. Visualization was carried out by phosphorimaging.

RNA structure predictions

Secondary structures of introns were predicted by RNAfold (20) and RNAshapes (21) algorithms. Free energy of secondary structures was calculated using

RNAfold with default settings, except that the temperature was set to 30°C. Sequences of analyzed introns from *S. cerevisiae* were downloaded from the Saccharomyces Genome Database (http://www.yeastgenome.org/).

RESULTS AND DISCUSSION

Efficient splicing of COF1 intron requires the formation of a stem between BP and 3′ splice site

Screening UV-mutagenized *S. cerevisiae* cells for splicing-defective mutations, we identified a G to A transition 31 nt upstream of 3′ss in *COF1* intron (referred to as G149A; Figure 1A). The mutation caused an approximately 6-fold increase of pre-mRNA and 2-fold decrease of mRNA levels as compared to wild-type cells (data not shown), suggesting a defect in splicing. *COF1* belongs to the subset of genes in budding yeast with an exceptionally long distance between BP and 3′ss. Intriguingly, according to the literature at present, to be efficiently identified as acceptor site in *S. cerevisiae*, 3′ss ought to be placed no further than 55 nt from BP (16). We thus analyzed the sequence of *COF1* intron (*COF1i*) using available algorithms for RNA structure prediction.

RNAfold- (20) and RNAshapes- (21) based models predicted the formation of a long stable stem structure between BP and 3′ss in wild-type *COF1i*. When G149A substitution was introduced, an additional internal loop appeared within the stem, which resulted in the destabilization of the predicted structure (Figure 1B). Other secondary structure predicting tools, including Mfold (22), which uses different physical parameters, and knowledge-based MC-Fold (23) showed qualitatively similar results. All the algorithms applied to the wild-type intron sequence predicted the existence of two double stranded regions, which hereafter will be referred to as the 'inner' and 'outer' stem (highlighted in Figure 1B). Importantly, the same stem formation was predicted independently of the length of the flanking sequences on 5′- and/or 3′-end. We decided to test the prediction that the stem exists between 75 and 153 nt of *COF1i* (Figure 1B) and to study the splicing of wild-type and mutant *COF1i* versions in more detail.

To test splicing efficiency, we performed a primer extension analysis of *COF1-CUP1* fusion reporters expressed in *S. cerevisiae* EGY48 strain. Unmodified *COF1i* containing pre-mRNA was spliced efficiently, whereas G149A mutation caused a severe splicing defect resulting in barely detectable quantities of spliced mRNA (Figure 2A, lanes 1 and 2). As G149A destabilized pairing in the inner stem of the predicted structure (Figure 1B), we asked whether G149A impairment could be suppressed by the substitution of predicted-complementary nucleotide (C80U). As expected, the G149A+C80U double mutant attained wild-type stability and was spliced efficiently (Figure 2A, lane 3). C80U single mutation affected neither the RNAfold predicted stability nor splicing (Figure 2A, lane 4) hypothetically because the G-U pair in the stem would be stable enough to maintain wild-type properties. In an effort to disrupt the structure by independent mutation, we manipulated a nucleotide adjacent to G149. A148U (Figure 2A, lane 5) and A148C (not shown) substitutions had negative effect on both predicted structure stability and splicing, similarly to G149A, whereas A148G, which allows alternative G-U pairing, had no effect (Figure 2A, lane 6). In all the cases tested, calculated structure stability correlated with experimentally tested splicing efficiency. To test whether the secondary structure or the sequence of the predicted inner stem-region is crucial for efficient *COF1i* splicing, we randomized all 10 base pairs of the inner stem such that the stability of the modeled structure remained the same (this variant is referred to as *COF1*(hel); Figure 2B, left panel). Although the sequence was extensively altered, the splicing of this intron was not impaired, as documented by primer extension (Figure 2B, middle panel). Efficient splicing was demonstrated also by the resistance of *cup1-Δ* cells expressing *COF1*(hel)-*CUP1* reporter to increased Cu^{2+} concentration (Figure 2B, right panel).

We used in-line probing analysis (19) to compare relative stabilities of 5′-3′ phosphodiester bonds of wild-type and G149A RNAs. The method allows monitoring of secondary structures in RNA molecules: base paired regions cannot adopt a conformation that allows for spontaneous RNA degradation, while single stranded regions can. For wild-type *COF1i*, degradation started to appear at 30°C and increased slightly at 37°C (wild-type; Figure 2C). G149A-mutated structure deviated from the wild-type in several aspects. First, there were strong additional in-line cuts in the regions of A147-A151 and C80 (G149A; Figure 2C), which were predicted to form the complementary arms of the inner stem (Figure 1B). Second, the G149A intron was considerably less stable than wild-type at 37°C; in some experiments, we observed its almost complete degradation (data not shown). When the compensatory C80U mutation was introduced into the G149A intron, the in-line probing pattern was reversed to that of the wild-type (G149A+C80U; Figure 2C), which indicated that the secondary structure was re-stabilized. Taken together, we demonstrated that a secondary structure which reduces the BP to 3′ss distance, rather than any particular sequence motif between the two splicing signals, is critical for efficient *COF1i* splicing.

Secondary structure within *COF1* intron masks potential 3′ splice sites

We generated a set of internal *COF1i* deletions (Figure 3A) and tested their splicing efficiency using *CUP1* fusion reporters. Deletion of nucleotides 91 to 140, which are predicted to be involved in outer stem formation, did not detectably affect splicing efficiency (*cof1*(Δ91-140); Figure 3B, lanes 1 and 2). However, concomitant destabilization of the inner stem led to dramatic decrease of mRNA signal and the appearance of an additional product [*cof1*(Δ91-140, G149A); Figure 3B, lane 3]. Using 5′-RACE technique, we found that this product corresponds to mRNA spliced to AAG located 27 nt upstream of the regular 3′ss. As expected, adding the G149A-complementary C80U mutation, which should stabilize the stem, partially restored the use of the annotated 3′ss (*cof1*(Δ91-140, G149A+C80U); Figure 3B, lane 4). We then deleted the whole stem, obtaining an intron with

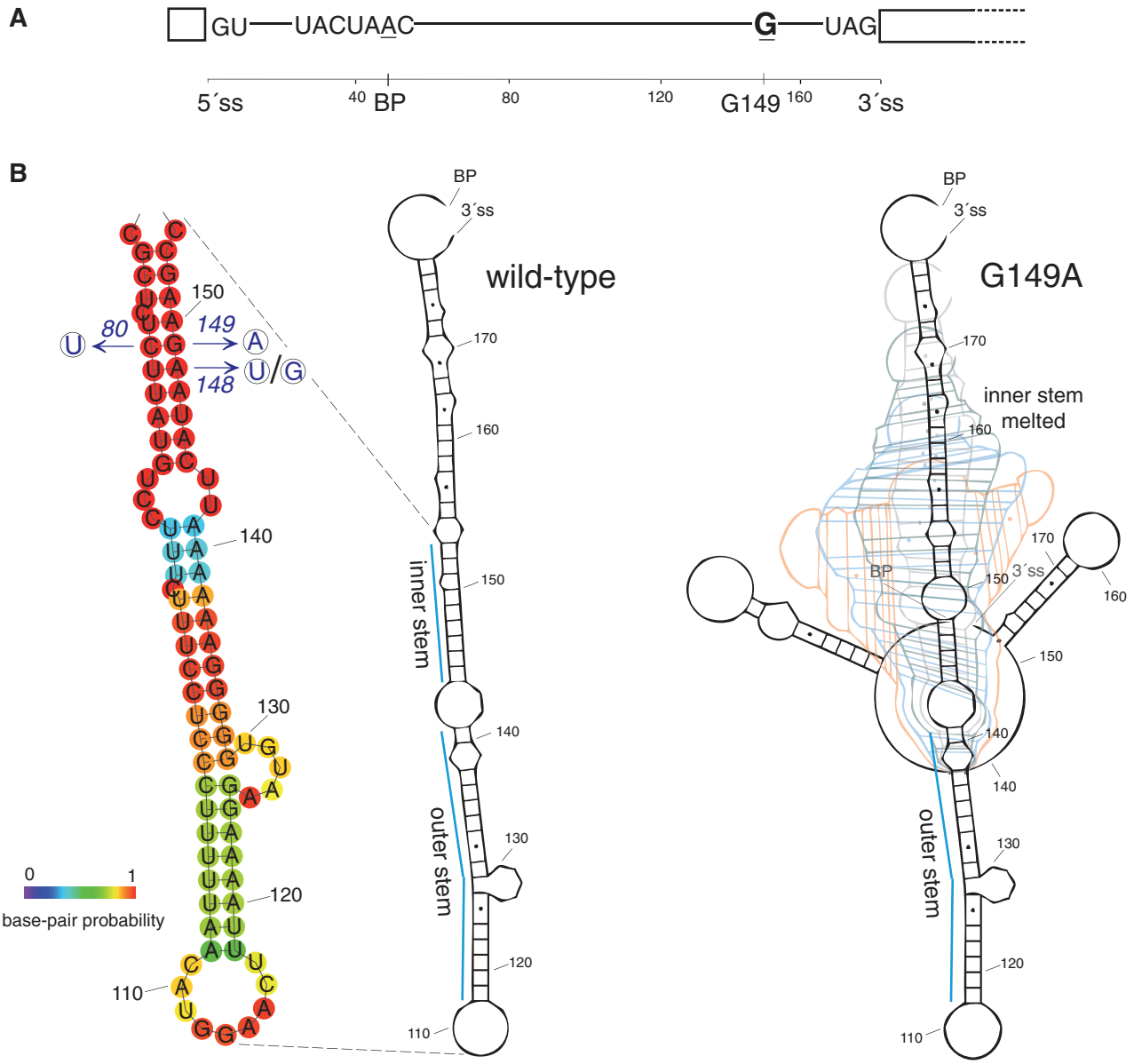


Figure 1. G149A substitution destabilizes predicted secondary structure between branch point and 3′ splice site of *COF1* intron. (**A**) Schematic representation of *COF1* intron. The numbering starts at the first intron nucleotide; all substitutions in this study are numbered according to the nucleotide position in non-manipulated intron. 5′ss, branch point region, 3′ss and G149 nucleotide are depicted. (**B**) Secondary structure predicted between BP and 3′ss of wild-type (left panel) and G149A (right panel) *COF1i*. Models were generated by RNAshapes program. Dynamic structure of G149A variant is depicted as a set of six overlaid still images extracted from an animated output of the program. Structure representation of the 'inner' and 'outer' stem containing part of wild-type *COF1i* including the sequence and base pair probabilities was generated by RNAfold. Mutations of the inner stem used in this study are shown in the left panel.

56 nt between BP and 3′ss [*cof1*(Δ76-152)]. This variant was spliced to CAG positioned 23 nt upstream of the regular 3′ss (Figure 3B, lane 5). Thus, both *cof1*(Δ91-140, G149A) and *cof1*(Δ76-152), which are predicted to lack stable structures (RNAfold), were spliced to the first acceptor AG downstream of BP. Crucially, a variant with the BP–3′ss distance of 31 nt (similar to *S. cerevisiae* median) was spliced as efficiently as full-length *COF1i*, generating wild-type mRNA [*cof1*(Δ76-176); Figure 3B, lane 6].

In summary, we demonstrated that neither the complementary sequences nor the stem–loop structure *per se* are needed by the spliceosome. Rather, the secondary structure formed between BP and 3′ss of *COF1* intron masks sequences that might otherwise serve as acceptor sites and thereby ensures proper 3′ss choice. Similar conclusions were reached when distant branch point (dBP) intron of *ACT* gene of *Kluyveromyces lactis* was analyzed as a heterologous construct in *S. cerevisiae* (24).

Secondary structure within *UBC13* intron aids splicing in a temperature-dependent manner

We modeled the structures of all *S. cerevisiae* introns with BP to 3′ss distance longer than 50 nt (RNAfold). The vast

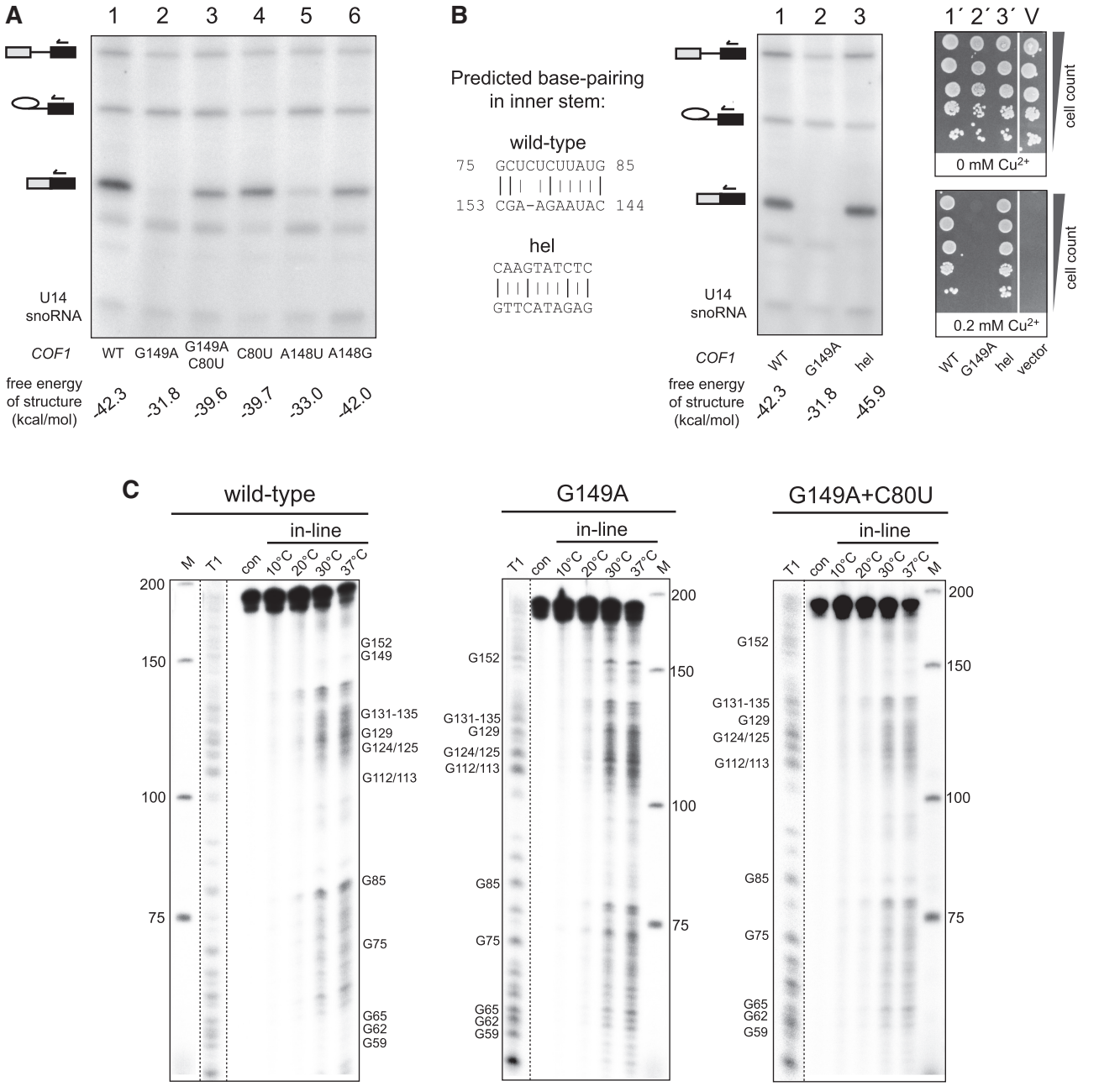


Figure 2. Intramolecular structure between BP and 3′ss is required for efficient splicing of *COF1* intron. (**A**) The stability of secondary structure predicted between BP and 3′ss of *COF1i* correlates with splicing efficiency. RNA from cells expressing the indicated *COF1-CUP1* constructs was subjected to primer extension analysis as described in Materials and Methods. The products corresponding to pre-mRNA, mRNA and lariat-exon 2 intermediate are indicated by their icons. U14 snoRNA was assayed as a loading control. Energy values (kcal/mol) for the minimal free energy structures calculated by RNAfold are shown below each lane. Mutation G149A caused severe inhibition of *COF1-CUP1* reporter gene splicing. The defect was suppressed by restoring complementarity in the predicted stem (G149A+C80U; lane 3). Destabilization of the predicted structure at a neighboring position likewise decreased splicing efficiency (A148U; lane 5). Substitutions which were predicted to have negligible impact on the stem between BP and 3′ss did not affect splicing (C80U and A148G; lane 4 and 6, respectively). (**B**) Intramolecular stem is indispensable for efficient splicing of *COF1i*. We completely altered the sequence of the inner stem of *COF1i* (left panel; see text and Figure 1B for explanation) such that the structure's predicted stability remained the same. Minimal free energies are indicated below lanes. The variant hel was spliced with the same efficiency as wild-type as shown by primer extension (lane 1 and 3) and by Cu^{2+}-resistance of cells expressing the reporter construct (lane 1′ and 3′). (**C**) RNA in-line probing confirms the existence of secondary structure between BP and 3′ss of *COF1i*. RNAs encompassing wild-type, G149A and G149A+C80U mutant *COF1i* were subjected to in-line probing analysis as described in Materials and Methods and in Supplementary Data. RNAs were incubated for 45 h at the indicated temperatures and the extent of spontaneous RNA degradation was analyzed by denaturing polyacrylamide gel electrophoresis. The bands were annotated based on guanosine ladders generated by T1 RNase digestion (T1). Untreated RNA and the GeneRuler Ultra Low Range DNA Ladder (Fermentas) were run in the 'con' and 'M' lanes, respectively. Strong in-line cuts in the regions of A147-A151 and C80 of G149A intron (middle panel), but not of wild-type (left panel) or G149A+C80U mutant (right panel) are visible. They confirm the RNAfold prediction that the stem that forms between G75-G85 and C153-C145 sequences in wild-type or G149A+C80U intron is disrupted in G149 mutant (see Figure 1B).

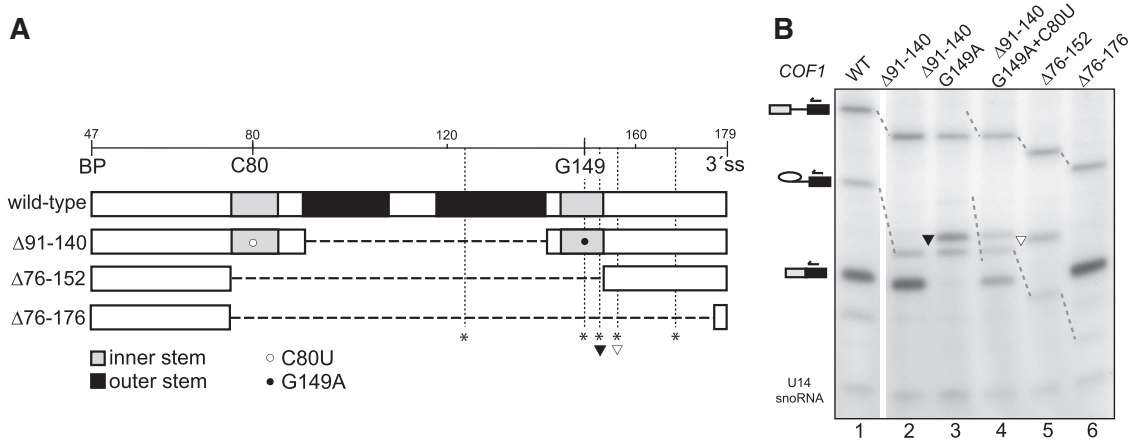


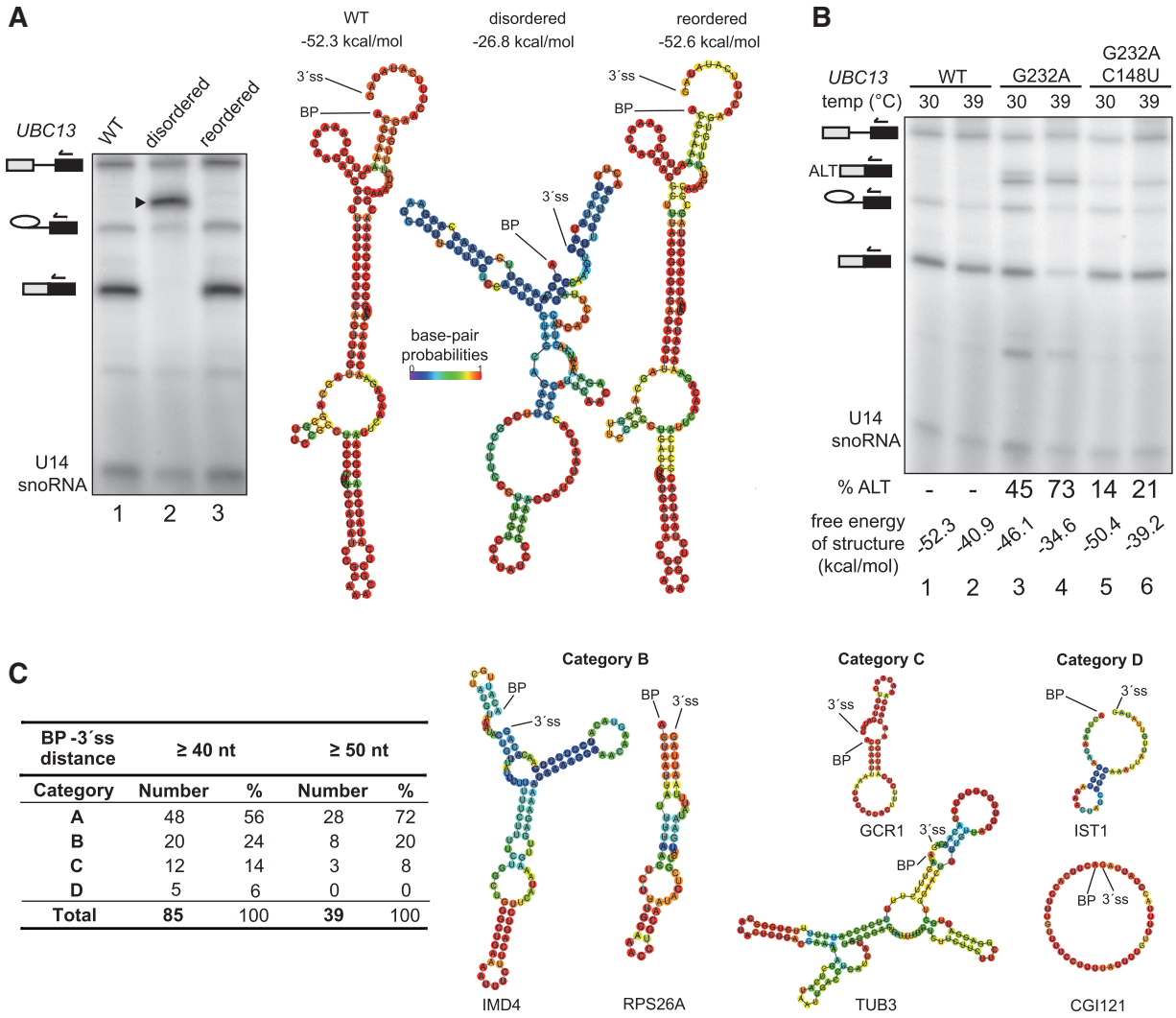
Figure 3. Destabilization of intramolecular stem within *COF1* intron unmasks potential 3′ splice sites. (**A**) Summary of *COF1i* constructs used. Sequences forming inner and outer stem are shaded; deleted regions are represented by dashed lines. Positions of G149 and C80 are marked by empty and filled circle, respectively. Potential 3′ss (all A/C/UAG trinucleotides) are marked by asterisks; acceptor sites used in the mutants, AAG152 and CAG156, are indicated by filled and empty arrowhead, respectively. (**B**) Destabilization or removal of stem within *COF1i* proves the structure's role in presenting the appropriate portion of the pre-mRNA molecule to spliceosome for 3′ss recognition. Deletion of the outer-stem nucleotides 91–140 did not affect splicing efficiency of *COF1i* (lane 2). Destabilization of the inner stem in this variant (Δ91-140, G149A) resulted in splicing to AAG152 (lane 3). Mutation of the nucleotide which is supposed to base pair with A149 partially suppressed the phenotype (Δ91-140, G149A+C80U; lane 4). Pre-mRNA with the whole stem-region deleted spliced to the first AG proximal to BP (CAG156; Δ76-152; lane 5). More extensive deletion, which shortened the BP–3′ss distance to 31 nt, resulted in efficient use of the annotated 3′ss (Δ76-176; lane 6). Filled and empty arrowheads indicate bands corresponding to the cryptic splice sites marked in (A). Notably, the cryptic sites of the *COF1i* mutants were only used when they were not blocked within the secondary structure or when they were brought closer to BP through deletion.

majority of these introns was predicted to fold into a structure resembling the stem–loop characterized in *COF1i* (RNAfold; Figure 4A, C and Supplementary Table S1). To further support the evidence that *S. cerevisiae* dBP introns depend on secondary structure between BP and 3′ss for splicing, we designed a *CUP1*-based splicing reporter for the *UBC13* gene, which has the second longest BP–3′ss sequence in *S. cerevisiae* (155 nt). We introduced multiple substitutions in one arm of the presumed stem of *UBC13* intron (*UBC13i*), which destabilized the predicted structure ('disordered' in Figure 4A and Supplementary Figure S1). The disordered *UBC13i* did not produce wild-type mRNA but was instead spliced to CAG located 38 nt downstream of BP (5′-RACE confirmed); this site was apparently masked by the stem structure in wild-type intron (Figure 4A, lane 1 and 2). When the base pairing (but not the original sequence) in the predicted stem was restored through a set of complementary mutations (Figure 4A and Supplementary Figure S1), wild-type splicing pattern was observed (Figure 4A, lane 3). Clearly, the requirement of secondary structure to overcome long BP–3′ss distance and to mask BP proximal sequences is not limited to *COF1i*.

Point mutation G232A, which caused only a mild decrease in the stability of the predicted structure at 30°C (data not shown), resulted in splicing that used several AGs, including the annotated 3′ss (Figure 4B lane 3). However, at 39°C, aberrant 3′ss were preferentially employed (Figure 4B, lane 4). The compensating C148U mutation suppressed, in a temperature-dependent manner, the inclusion of additional 3′ss (Figure 4B, lanes 5 and 6). These findings further support the hypothesis that the stem structure is responsible for proper 3′ss selection, as the stability of folded RNA is temperature dependent.

Long BP–3′ss sequences encode secondary structures in several *Saccharomycotina* species

We demonstrated that higher order structures are required for splicing of dBP introns in *S. cerevisiae*. We also noticed the occurrence of such structures in other intron-poor *Saccharomycotina* species (hemiascomycetes) (11). RNAfold predicted stem-loops downstream of BP in *COF1* introns in five species of the *Saccharomyces* 'sensu stricto' genus (*S. cerevisiae*, *S. paradoxus*, *S. kudriavzevii*, *S. mikatae* and *S. bayanus*; Supplementary Figure S2B). A multiple alignment of intron sequences revealed a high conservancy in the regions predicted to base pair (Supplementary Figure S2A). For most of the nucleotides that are not conserved, base pairing is preserved. A change in one strand either preserves base pairing with the nucleotide of the other strand (e.g. A-U pair is changed to G-U), or is matched by the co-evolution of the opposite strand (e.g. G-C pair is replaced by A-U). Thus, there seems to be a selection against mutations destabilizing the secondary structure. Notably, the AGs present between BP and the physiological 3′ss do not seem to be immediately usable (they do not give rise to translatable mRNAs). We found the same conservancy on the level of both primary and secondary structure also for *UBC13, YDR381C-A* and *UBC12* introns (data not shown). In fact, every other dBP intron we examined within the genus was similarly conserved. Outside of the genus, *COF1i* was conserved in position and secondary structure in *Candida glabrata* and *Kluyveromyces lactis*. In *C. glabrata*, all dBP introns tested, e.g. *RPS4A*, were



BP -3´ss distance	≥ 40 nt		≥ 50 nt	
Category	Number	%	Number	%
A	48	56	28	72
B	20	24	8	20
C	12	14	3	8
D	5	6	0	0
Total	85	100	39	100

Figure 4. Secondary structure in *UBC13* intron is responsible for proper 3′ss selection. (**A**) Mutations of one arm of the predicted stem in *UBC13i*, which destabilized the structure, resulted in splicing to cryptic CAG located 38 nt downstream of BP (lane 2; band is marked by arrowhead). Introducing mutations to the complementary arm, such that the pairing energy but not the original sequence was restored, repaired the defect (lane 3). Schematic representations of BP to 3′ss regions in wild-type and manipulated *UBC13i* variants based on RNAfold predictions are shown in the left panel. The values of free energy are indicated for each structure. Summary of wild-type and manipulated *UBC13i* sequences is provided in Supplementary Figure S1. (**B**) Use of cryptic 3′ss in a mutant with partially destabilized structure is enhanced at higher temperature. G232A substitution in *UBC13i* mildly destabilized the secondary structure and led to the use of several 3′ss (lane 3; bands corresponding to mRNA spliced to cryptic 3′ss are marked by ALT + mRNA icon). The defect was exacerbated at 39°C (lane 4). Compensating substitution partially suppressed the defect in temperature dependent manner (G232A+C148U; lanes 5 and 6). Densitometric quantification of alternatively spliced RNA (expressed as percent of total spliced RNA; % ALT) and free energy of structure calculated for a given temperature are indicated below each lane. (**C**) Majority of dBP introns of *S. cerevisiae* are predicted to form secondary structure between BP and 3′ss. Sequences between BP and 3′ss were analyzed by RNAfold. Predicted structures were sorted into four categories. A—structures with extensive and stable stems; none to small bulge/internal loops. B—structures with less extensive but still stable stems; bulge/internal loops. C—other type of secondary structure. D—unstable structures or unstructured. The relative proportions of the categories are shown. Two examples of predicted structures in each category are depicted in the right panel. Category A is represented by *UBC13i* and *COF1i* (Figure 1B). Full list of introns together with additional information on sorting the structures is provided in Supplementary Data (Supplementary Table S1).

structured. Long and structured intron was previously found in *K. lactis ACT* gene (24), but it is not conserved in the *Saccharomyces* genus.

Previous analysis of phylogenetic distribution of BP–3′ss distances within *Saccharomycotina* revealed two groups of species (12). A group with constrained BP to 3′ss distance (e.g. *Debaryomyces hansenii*; 7–8 nt) and yeasts with unconstrained BP–3′ss spacing (*Saccharomyces* genus, *C. glabrata* and *K. lactis*; distances reach up to 166, 471 and 185 nt, respectively; http://genome.jouy.inra.fr/genosplicing/index.html). Outside of the *Saccharomyces* 'sensu stricto' genus, the conserved position of an intron within a gene did not imply the conservancy of its BP–3′ss length. Also, complementary regions of dBP introns that we analyzed did not show any homology to transposons (24). Importantly however, in every *Saccharomycotina* intron

we checked, BP–3′ss sequence over 60 nt folded into a stable structure (RNAfold).

Mechanisms of 3′ss recognition

In *S. cerevisiae*, splice site consensus sequences are recognized repeatedly during across-intron spliceosome assembly and subsequent rearrangements through both catalytic steps (25). For lariat formation, pre-mRNA substrate must contain 23 nucleotides downstream of BP in a sequence independent manner (26,27). This type of splicing, which occurs typically in *S. cerevisiae*, where the branch site is strictly conserved, is called AG independent. For the so-called AG dependent splicing, which is typical for some mammalian introns with weak pY-tracts, the acceptor YAG trinucleotide must be present for the first step of splicing to proceed (28). However, the YAG seems to be required for spliceosome assembly rather than for exon ligation. Experiments with 3′ substrates *in trans* clearly showed that these two phases can be separated (29). It seems that for both *S. cerevisiae* and mammalian introns, acceptor YAG must be correctly positioned only before the second step. We clearly observed lariat-exon 2 intermediate accumulation in all cases where the disruption of secondary structure inhibited mRNA formation (Figures 2–4).

The recognition of 5′ss and BP_pY-tract_3′ss regions of long introns in mammalian cells involves the cotranscriptional formation of complexes across flanking short exons (exon definition complex) (30,31). dBP introns, comprising around 0.6% of all human introns, represent an additional problem of overcoming the separation of BP_pY tract from 3′ss. Long sequences between BP and 3′ss (>40 bp) are usually devoid of AG dinucleotides (AG exclusion zone, AGEZ) (32) and are presumed to be scanned by the spliceosome (leaky scanning model) (33). An example is the human serotonin receptor 4 gene (*HTR4*), in which dBP introns 3, 4 and 5 contain AGEZs of 149–291 nt (34). We examined these and other human BP_pY-tract_3′ss regions and found them to be unstructured (RNAfold; data not shown).

In contrast, the recognition of distant 3′ss in *S. cerevisiae* does not obey the scanning model (35–37) and is dependent on the formation of a secondary structure. We reason that the requirement of a secondary structure for splicing of dBP introns as well as the presence of silent proximal AGs within it confirms that *S. cerevisiae* spliceosome cannot use a processive scanning mechanism to locate distant acceptor 3′ ss.

CONCLUSIONS

Splicing signals must be recognized both over distance and among competing sequences. We experimentally demonstrated in *S. cerevisiae* that *COF1* and *UBC13* introns, which both have distant branch points, are spliced with the aid of intramolecular structure within pre-mRNA. This dependence of splicing on intron structure may have evolved during the reductive evolution of hemiascomycetes (12,38). It remains to be seen whether pre-mRNA structure-mediated recognition of 3′ss is confined to intron-poor yeasts or whether it exists also in higher eukaryotes, which employ more complex networks of splicing regulation. The role of nascent RNA secondary structure offers exciting possibilities for the discovery of novel regulatory mechanisms, as the introns with long BP–3′ss distance may have acquired additional functions which proved advantageous for the organisms.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENT

Anne Kalweit is acknowledged for generating the DNA templates for *in vitro* transcription.

FUNDING

Czech Ministry of Education, Youth and Sports grants (MSM0021620858; LC07032); Grant Agency of the Charles University grant (398811); Heisenberg stipend by the Deutsche Forschungsgemeinschaft (HA 3459/5 to C.H.). Funding for open access charge: Czech Ministry of Education, Youth and Sports grant LC07032.

Conflict of interest statement. None declared.

REFERENCES

1. Chen,M. and Manley,J.L. (2009) Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches. *Nat. Rev. Mol. Cell Biol.*, **10**, 741–754.
2. Munroe,S.H. (1984) Secondary structure of splice sites in adenovirus mRNA precursors. *Nucleic Acids Res.*, **12**, 8437–8456.
3. Solnick,D. (1985) Alternative splicing caused by RNA secondary structure. *Cell*, **43**, 667–676.
4. Buratti,E. and Baralle,F.E. (2004) Influence of RNA secondary structure on the pre-mRNA splicing process. *Mol.Cell. Biol.*, **24**, 10505–10514.
5. Warf,M.B. and Berglund,J.A. (2010) Role of RNA structure in regulating pre-mRNA splicing. *Trends Biochem. Sci.*, **35**, 169–178.
6. Singh,N.N., Singh,R.N. and Androphy,E.J. (2007) Modulating role of RNA structure in alternative splicing of a critical exon in the spinal muscular atrophy genes. *Nucleic Acids Res.*, **35**, 371–389.
7. Howe,K.J. and Ares,M. Jr (1997) Intron self-complementarity enforces exon inclusion in a yeast pre-mRNA. *Proc. Natl Acad. Sci. USA*, **94**, 12467–12472.
8. Raker,V.A., Mironov,A.A., Gelfand,M.S. and Pervouchine,D.D. (2009) Modulation of alternative splicing by long-range RNA structures in Drosophila. *Nucleic Acids Res.*, **37**, 4533–4544.
9. Blouin,S., Mulhbacher,J., Penedo,J.C. and Lafontaine,D.A. (2009) Riboswitches: ancient and promising genetic regulators. *Chembiochem.*, **10**, 400–416.
10. Macias,S., Bragulat,M., Tardiff,D.F. and Vilardell,J. (2008) L30 binds the nascent RPL30 transcript to repress U2 snRNP recruitment. *Mol. Cell*, **30**, 732–742.
11. Scannell,D.R., Butler,G. and Wolfe,K.H. (2007) Yeast genome evolution—the origin of the species. *Yeast*, **24**, 929–942.
12. Irimia,M. and Roy,S.W. (2008) Evolutionary convergence on highly-conserved 3′ intron structures in intron-poor eukaryotes and insights into the ancestral eukaryotic genome. *PLoS Genet.*, **4**, e1000148.

13. Libri,D., Stutz,F., McCarthy,T. and Rosbash,M. (1995) RNA structural patterns and splicing: molecular basis for an RNA-based enhancer. *RNA*, **1**, 425–436.
14. Charpentier,B. and Rosbash,M. (1996) Intramolecular structure in yeast introns aids the early steps of in vitro spliceosome assembly. *RNA*, **2**, 509–522.
15. Rogic,S., Montpetit,B., Hoos,H.H., Mackworth,A.K., Ouellette,B.F. and Hieter,P. (2008) Correlation between the secondary structure of pre-mRNA introns and the efficiency of splicing in *Saccharomyces cerevisiae*. *BMC Genomics*, **9**, 355.
16. Cellini,A., Felder,E. and Rossi,J.J. (1986) Yeast pre-messenger RNA splicing efficiency depends on critical spacing requirements between the branch point and 3' splice site. *EMBO J.*, **5**, 1023–1030.
17. Golemis,E.A., Gyuris,J. and Brent,R. (1996) Interaction trap/ two-hybrid system to identify interacting proteins. In Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.D., Smith,J.A. and Struhl,K. (eds), *Current Protocols in Molecular Biology*, Wiley, New York, Unit 20.1.
18. Lesser,C.F. and Guthrie,C. (1993) Mutational analysis of pre-mRNA splicing in Saccharomyces cerevisiae using a sensitive new reporter gene, CUP1. *Genetics*, **133**, 851–863.
19. Regulski,E.E. and Breaker,R.R. (2008) In-line probing analysis of riboswitches. *Methods Mol. Biol.*, **419**, 53–67.
20. Gruber,A.R., Lorenz,R., Bernhart,S.H., Neubock,R. and Hofacker,I.L. (2008) The Vienna RNA websuite. *Nucleic Acids Res.*, **36**, W70–W74.
21. Steffen,P., Voss,B., Rehmsmeier,M., Reeder,J. and Giegerich,R. (2006) RNAshapes: an integrated RNA analysis package based on abstract shapes. *Bioinformatics*, **22**, 500–503.
22. Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.
23. Parisien,M. and Major,F. (2008) The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature*, **452**, 51–55.
24. 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.
25. Wahl,M.C., Will,C.L. and Luhrmann,R. (2009) The spliceosome: design principles of a dynamic RNP machine. *Cell*, **136**, 701–718.
26. 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.
27. Cheng,S.C. (1994) Formation of the yeast splicing complex A1 and association of the splicing factor PRP19 with the pre-mRNA are independent of the 3' region of the intron. *Nucleic Acids Res.*, **22**, 1548–1554.
28. Wu,S., Romfo,C.M., Nilsen,T.W. and Green,M.R. (1999) Functional recognition of the 3' splice site AG by the splicing factor U2AF35. *Nature*, **402**, 832–835.
29. Anderson,K. and Moore,M.J. (2000) Bimolecular exon ligation by the human spliceosome bypasses early 3' splice site AG recognition and requires NTP hydrolysis. *RNA*, **6**, 16–25.
30. Robberson,B.L., Cote,G.J. and Berget,S.M. (1990) Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol. Cell Biol.*, **10**, 84–94.
31. Berget,S.M. (1995) Exon recognition in vertebrate splicing. *J. Biol. Chem.*, **270**, 2411–2414.
32. Gooding,C., Clark,F., Wollerton,M.C., Grellscheid,S.N., Groom,H. and Smith,C.W. (2006) A class of human exons with predicted distant branch points revealed by analysis of AG dinucleotide exclusion zones. *Genome Biol.*, **7**, R1.
33. 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.
34. Hallegger,M., Sobala,A. and Smith,C.W. (2010) Four exons of the serotonin receptor 4 gene are associated with multiple distant branch points. *RNA*, **16**, 839–851.
35. Patterson,B. and Guthrie,C. (1991) A U-rich tract enhances usage of an alternative 3' splice site in yeast. *Cell*, **64**, 181–187.
36. Luukkonen,B.G. and Seraphin,B. (1997) The role of branchpoint-3' splice site spacing and interaction between intron terminal nucleotides in 3' splice site selection in *Saccharomyces cerevisiae*. *EMBO J.*, **16**, 779–792.
37. Liu,Z.R., Laggerbauer,B., Luhrmann,R. and Smith,C.W. (1997) Crosslinking of the U5 snRNP-specific 116-kDa protein to RNA hairpins that block step 2 of splicing. *RNA*, **3**, 1207–1219.
38. Collins,L. and Penny,D. (2005) Complex spliceosomal organization ancestral to extant eukaryotes. *Mol. Biol. Evol.*, **22**, 1053–1066.