Proc. Natl. Acad. Sci. USA
Vol. 94, pp. 12467–12472, November 1997
Genetics

Intron self-complementarity enforces exon inclusion in a yeast pre-mRNA

KENNETH JAMES HOWE AND MANUEL ARES, JR.*

Center for the Molecular Biology of RNA, Biology Department, University of California, Santa Cruz, CA 95064

Edited by John Abelson, California Institute of Technology, Pasadena, CA, and approved September 15, 1997 (received for review July 18, 1997)

ABSTRACT **Skipping of internal exons during removal of introns from pre-mRNA must be avoided for proper expression of most eukaryotic genes. Despite significant understanding of the mechanics of intron removal, mechanisms that ensure inclusion of internal exons in multi-intron pre-mRNAs remain mysterious. Using a natural two-intron yeast gene, we have identified distinct RNA–RNA complementarities within each intron that prevent exon skipping and ensure inclusion of internal exons. We show that these complementarities are positioned to act as intron identity elements, bringing together only the appropriate 5′ splice sites and branchpoints. Destroying either intron self-complementarity allows exon skipping to occur, and restoring the complementarity using compensatory mutations rescues exon inclusion, indicating that the elements act through formation of RNA secondary structure. Introducing new pairing potential between regions near the 5′ splice site of intron 1 and the branchpoint of intron 2 dramatically enhances exon skipping. Similar elements identified in single intron yeast genes contribute to splicing efficiency. Our results illustrate how intron secondary structure serves to coordinate splice site pairing and enforce exon inclusion. We suggest that similar elements in vertebrate genes could assist in the splicing of very large introns and in the evolution of alternative splicing.**

A key event in the decoding of genetic information in eukaryotes is the removal of intervening sequences or introns by nuclear pre-mRNA splicing (reviewed in refs. 1 and 2). The process of splicing requires the concerted activities of a large number of trans-acting RNA and protein factors that recognize and assemble onto splicing signals located near the branchpoint and splice sites in the pre-mRNA (1, 2). An early step in spliceosome assembly is the formation of a splicing complex within which the reactive sites of the intron destined to be removed have largely been determined and the limits of the intron defined (3–5). Distinguishing the correct splice sites from the many similar pre-mRNA sequences is a problem (1, 2), compounded by the fact that vertebrate genes frequently contain long introns (>5 kb) and short internal exons (<300 bp, ref. 6). In mammals, failure to identify splice sites often results in exon skipping, causing aberrant gene expression (7) as well as a number of genetic diseases in humans (8).

Simply finding correct splice sites is insufficient to determine correct splicing. Experiments using chimeric introns indicate that nearly any 5′ splice site can be joined to nearly any 3′ splice site (1, 2, 9). Because most splice sites are compatible, the splicing machinery must pair them in a fashion that prevents joining of compatible splice sites from different introns. The relative sizes of adjacent exons and introns appear to influence the mechanism by which this is determined (10). This interpretation is based on the general observation that 5′ splice site mutations in a large intron following a small exon causes skipping of that exon (failure of "exon definition"), whereas the same mutation in a small intron causes intron retention rather than exon skipping (failure of "intron definition;" refs. 11–14). The key molecular events that distinguish the intron definition and exon definition pathways of splice site pairing have not been identified. Because the efficiency of splice site recognition clearly plays a role in correct splice site pairing, protein factors that affect splice site recognition through their interaction with splicing enhancer sequences in pre-mRNA are very likely to be involved (for review, see ref. 15). Even given the ability of specific RNA–protein and protein–protein interactions to stimulate spliceosome assembly at authentic splice sites, it is not clear how splice site pairing is coordinated in transcripts with multiple authentic splice sites so that all exons are included.

The exon inclusion problem is intrinsic to the eukaryotic gene expression pathway, and thus mechanistic solutions must also be present in the budding yeast. There are four known multiply interrupted *Saccharomyces cerevisiae* genes (16–19) with different intron-exon architectures (see Fig. 1*A*). *MAT***a***1*, required for determining the diploid state (16), and *SLC1*, a homolog of the dynein light chain 1 gene (17), have short (<100 bp) introns but differ in the sizes of their internal exons (246 nt and 23 nt, respectively). *YL8A* and *YL8B* are related genes for ribosomal protein L6. Each has two large (>400 bp) introns separated by a short (94 nt) exon (18, 19). The pre-mRNAs from all of these genes are spliced with little or no exon-skipping (ref. 16; also Fig. 1*C* and data not shown), suggesting that correct exon inclusion can be managed by the yeast gene expression machinery.

In this report we describe the identification of elements which ensure inclusion of the constitutive internal exon in *YL8A*. Inclusion of the exon is enforced by intron-defining RNA secondary structures that coordinate the pairing of the splice sites of each intron independently, similar to those shown previously to enhance splicing efficiency in single intron yeast pre-mRNAs (20–23). Rearranging complementarities can promote exon skipping or rescue exon inclusion, demonstrating the ability of pre-mRNA folding to contribute to accurate splice site pairing. Extension of our observations to vertebrate genomes suggests that novel and alternative splicing patterns could arise during evolution by the introduction of mobile repeat elements into introns.

METHODS AND MATERIALS

Recombinant Techniques. A 1.2-kbp *Bam*HI–*Acc*65I PCR fragment of *YL8A* (ref. 18; GenBank accession no. X62627), including 41 bp of the 5′ untranslated region and 117 bp of exon 3 was cloned into pGEM7zf+ for oligonucleotide-directed mutagenesis (25). The "305" and "692" single intron chimeras were made by deleting a 715-bp *Spe*I (S) or a 328-bp

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

This paper was submitted directly (Track II) to the *Proceedings* office.
*To whom reprint requests should be addressed. e-mail: ares@darwin.ucsc.edu.

*Bst*BI (B) fragment, respectively, from the internal portion of the gene. The "1020" intron was made by inverting the *Bst*BI fragment. The "miniE2" construct was made by deleting a 72-bp *Sma*I–*Sna*BI fragment from exon 2 (*Sma*I and *Sna*BI sites were introduced by oligonucleotide-directed mutagenesis; the altered residues were removed with the deletion). Double and triple secondary structure mutants were made by subcloning fragments containing mutations X, Y, and ¥. Mutant *YL8A* DNA segments were inserted into plasmid pGAC14Uf [a *URA3* reporter derived from pGAC14 (26)] for expression. PCR primers (restriction sites underlined) are as follows: YL8A left, dATTGTCTTCGGATCCTATAAATCCAAATAACC; and YL8A right, dGTAGGTACCAGCAGCCTTGGC; and mutagenic oligonucleotides (mutations underlined) are as follows: WK5′, dGCAAAAGTGGTACGAACAGCCTTACG; X-5′, dGTTTAGAATTCAGCCTTACGAGC; E2SmaI, dCTTGACCCGGGAATCTCAGTTGAAG; E2SnaBI, dGCCTTACGTACAGCTCTTTCAGCAGC; mutation X, dTGCTCCTTTTTCTGTTATACTGCAGATCGAACTAGTATACTTCTC; mutation Y, dACTTTTTTCTACGATCTGCAGTATAACAGAAAACACCACTTTTC; and mutation ¥, dGATGGAAAACACAGTCGATCTGCAGTATAACTGTTACACGGCATT.

Yeast transformation, Reporter Expression, and RNA Preparation. Yeast were transformed by the LiOAC/DMSO method (27) and selected on minimal medium lacking uracil. The yeast strain used in this work is HFY870 (*MAT***a***1 ade2-1 his3-11, 15 leu2-3, 112 trp1-1 ura3-1 can1-100 upf1*::*HIS3 CUP1 YL8A*); the upf1 deficiency does not cause a detectable splicing defect (28). Yeast were grown in synthetic complete media with dextrose lacking uracil at 30°C to mid-log phase. RNA was isolated as previously described (29).

Reverse Transcription and cDNA Analysis. Reverse transcription reactions were performed essentially as described (29). A total of 2–3 μg total RNA/reaction was used for the experiments shown in Figs. 2*B* and 3*B*; 4–6 μg total RNA/reaction was used for the experiment shown in Fig. 4*C*. Primer annealing was carried out at 65°C for 5 min, then 42°C for 25 min. Reverse transcription extensions were performed at 42°C for 30 min. cDNAs were resolved on 6% denaturing polyacrylamide gels. The CUP1 primer was as follows: dGCACTCATGACCTTCATTTTG (8.6×10^6 cpm/pmol was used for Figs. 2*B* and 3*B*; 1×10^7 cpm/pmol was used for Fig. 4*C*). *SCR1* RNA (30) was used as an internal control to equilibrate sample loading. scr1-108 primer (dGGCGTGCAATCCGTGTCT) was used at 4.6×10^6 cpm/pmol for Figs. 2*B* and 3*B*; 4.2×10^6 cpm/pmol for Fig. 4*C*. Signal from endogenous *CUP1* mRNA is absent from the portions of the gels shown. Gels were exposed to PhosphorImager screens, then scanned and quantitated using the Molecular Dynamics IMAGEQUANT software. Because of uncertain effects on stability of RNAs carrying different mutations, quantitation was limited to the stable exon-included and exon-skipped products. Exon-included and exon-skipped cDNAs were amplified using YL8A left and CUP1 oligonucleotides and were subcloned into pGEM7zf+ for sequencing (data not shown).

Pre-mRNA Secondary Structure Prediction. *YL8A* (ref. 18; GenBank accession no. X62627), *YL8B* (ref. 19; GenBank accession no. D25232), *MAT***a***1* (ref. 16; GenBank accession no. J01334), and *SLC1* (ref. 17; GenBank accession no. L13282) sequences were folded using the RNA folding program MFOLD version 2.0 (31).

RESULTS

Splice Site Mutations in Intron 2 Do Not Induce Skipping of the Internal Exon. The phenotype of 5′ splice site mutations distinguishes exon from intron definition mechanisms of exon inclusion (11–14). To test whether the efficient inclusion of the small internal *YL8A* exon is enforced by an exon definition mechanism, we altered the 5′ splice site of intron 2, either weakening it (5′-GTtcGT, WK5′) to match that of the naturally weak *MER2* intron (32) or destroying it (5′-GaATtc, X-5′) (Fig. 1*B*). The mutations were tested in the context of a fusion of *YL8A* to *CUP1*, driven by a strong promoter on a high copy plasmid. Total cellular RNA was isolated and reverse transcribed using labeled CUP1-specific primer, and cDNA products were resolved by denaturing gel electrophoresis. Because the exon-skipped mRNA contains in-frame premature stop codons, we used a yeast strain disrupted at the *UPF1* locus. Upf1p is a component in the pathway for cytoplasmic degradation of mRNAs carrying in-frame nonsense codons (28). The upf1-deficiency does not affect splicing (28).

In contrast to the effects of 5′ splice site mutations in many vertebrate genes with small internal exons (11–14), the splice site mutations we tested do not induce exon skipping in *YL8A* (Fig. 1*C*). Weakening the splice site reduces the amount of exon 2-containing mRNA, and increases levels of a higher molecular weight product whose size is consistent with reten-

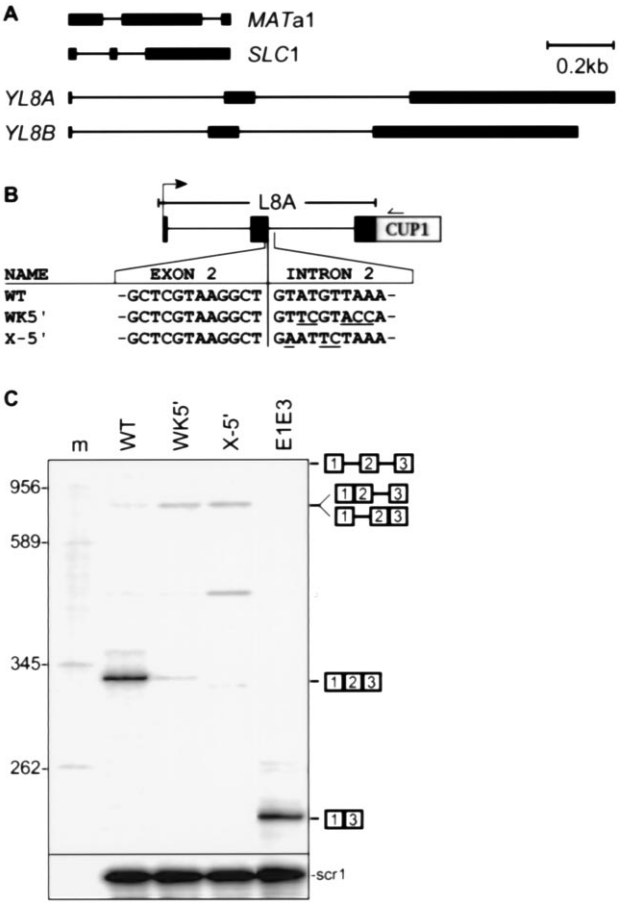

FIG. 1. Splicing phenotype of *YL8A* mutants suggests an intron definition mechanism. (*A*) Intron–exon structures of the known multiply interrupted genes in *S. cerevisiae*. (*B*) Structure of L8A-CUP1 expression constructs used to test *cis*-acting factors in *YL8A* splicing. The sequences shown are at the junction of the second exon and intron. Wild-type (WT) and mutant (WK5′; X-5′) 5′ splice sites were tested (mutant residues are underlined). (*C*) The 5′ splice site mutations do not induce exon skipping *in vivo*. Splicing was analyzed by reverse transcription of total cell RNA using a 5′-labeled (^{32}P) primer complementary to *CUP1* sequences. Lane m, DNA size markers. E1E3 is a marker for exon skipping expressed from a construct in which exons 1 and 3 are directly fused. Expected products are diagrammed at the right. scr1 is a small cytoplasmic RNA used as an internal control for total RNA amount. A strong RT stop (○) correlates with the 5′ end of snR39, a small nucleolar RNA encoded in intron 2 (24) and could be generated by processing events related to snR39 biosynthesis.

tion of intron 2 (Fig. 1*C*, WK5′). Destroying the 5′ splice site also impairs intron removal and yields higher levels of the larger product, but still does not induce exon skipping (Fig. 1*C*, X-5′). A cryptic site (5′-GTAaGg) 8 nt upstream is weakly activated in this mutant, resulting in an mRNA containing a shortened internal exon (Fig. 1*C*, X-5′). Thus, in contrast to the effects of 5′ splice site mutations in vertebrate pre-mRNAs with small internal exons, the *YL8A* exon is not skipped, rather intron 2 is retained. This indicates that *YL8A* exon 2 inclusion is not mediated by an exon-definition mechanism, and suggests instead that the introns are defined (13). A reduction in total L8A-CUP1 RNA yield relative to *SCR1* RNA is also observed, suggesting that mutant pre-mRNAs may be more rapidly degraded.

Skipping of Exon 2 Is Prevented by Intron Sequences Near Exon 2. To exclude the possibility that the 5′ splice site of *YL8A* intron 1 is incompatible with the 3′ splice site of intron 2, we deleted the internal exon and flanking portions of each intron (Fig. 2*A*). Splicing of exon 1 to exon 3 in these chimeric constructs shows that the splice sites are compatible (Fig. 2*B*, 305 and 692). The efficiency of splicing (and total transcript yield) is reduced as the distance between the sites is increased, in accord with previously observed distance effects on splicing of single intron yeast substrates (33). Restoring the wild-type

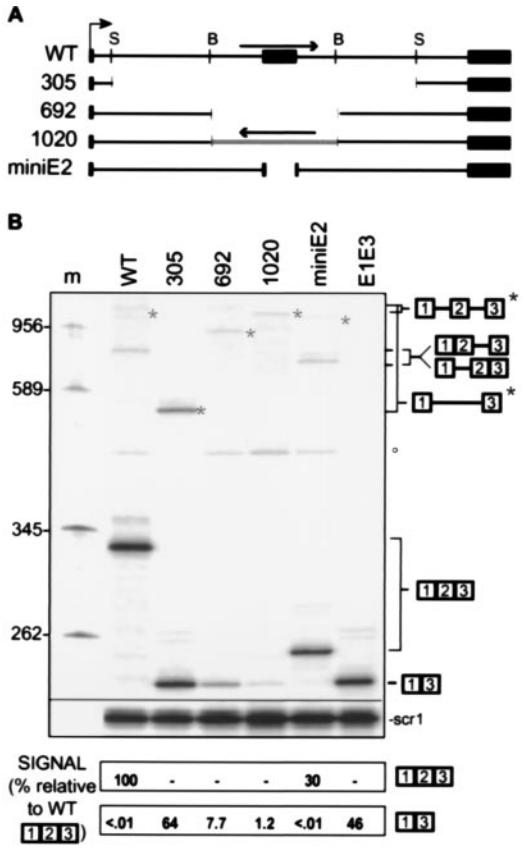

FIG. 2. Compatibility of *YL8A* splice sites and influence of internal exon sequences on exon inclusion. (*A*) Structure of mutant substrates. Single intron substrates with chimeric introns of 305 and 692 nucleotides (deleted regions are depicted as gaps) and 1,020 nucleotides (the inverted region spans exon 2) are shown. The two-intron miniE2 substrate lacks 72 of 94 nucleotides from the internal exon. (*B*) Splicing phenotypes of the mutant constructs. Splicing was analyzed by reverse transcription of total cell RNA using a 5′-labeled (^{32}P) primer complementary to *CUP1* sequences. Lane m, DNA size markers. E1E3 is a marker for exon skipping expressed from a construct in which exons 1 and 3 are directly fused. Expected products are diagrammed at the right. The different unspliced pre-mRNAs are indicated by an asterisk (*). Spliced products were measured relative to scr1; the amount of exon included mRNA in wild type was taken as 100%.

distance of 1,020 nt (by inverting a fragment spanning the internal exon) also allows splicing, although at a reduced level (Fig. 2*B*, 1020). Thus, we conclude that exon skipping is prevented in part by sequences within the fragment spanning the internal exon. Deleting 72 nt of the 94-nt internal exon does not induce exon skipping (Fig. 2*B*, miniE2), arguing that sequences required for exon inclusion are located within the introns.

Intron Self-Complementarities Enforce Inclusion of the Internal Exon. We inspected the sequences near the internal exon and identified complementarity between the regions near the 5′ splice site and branchpoint of each *YL8A* intron (Fig. 3*A*). Complementarities of this type were originally identified in a survey of yeast intron sequences (34). Experiments support a role for such complementarities in splicing efficiency of certain pre-mRNAs (20–23) and in splice site pairing in artificial alternative splicing substrates in yeast (21). If these intron self-complementarities act to promote the pairing of specific pairs of authentic splice sites to the exclusion of incorrect pairs of authentic splice sites as suggested (21), then they would enforce exon inclusion in a natural two-intron pre-mRNA.

To test this hypothesis, we reduced complementarity between sequences near the 5′ splice site and branchpoint in each *YL8A* intron (Fig. 3 *A* and *B*). The sequence downstream of the intron 1 5′ splice site (sequence A) and its complement upstream of the intron 1 branchpoint (sequence B) were separately mutated (to sequences X and Y, respectively, Fig. 3*A*; mutants XBCD and AYCD, respectively, Fig. 3 *A* and *B*). The sequence of the region near the intron 2 branchpoint (sequence D) was mutated (to sequence ¥) to destroy base-pairing potential in intron 2 (Fig. 3 *A* and *B*, mutant ABC¥). The precise sequences of the mutations were selected based on an estimate of their ability to disrupt the natural base-pairing potential shown in Fig. 3*A* without creating new base-pairing potential with other sequences (see below). Each of the three sets of mutations was individually tested for its effect on exon inclusion (Fig. 3*C*).

Destroying the complementarity in either intron reduces the amount of correctly spliced product, and allows a modest amount of exon skipping to occur (Fig. 3*C*, XBCD, AYCD, ABC¥). The appearance of exon 2-skipped product supports our previous conclusion that the 5′ splice site of intron 1 is compatible with the 3′ splice site of intron 2 (Fig. 2), and demonstrates that skipping can occur despite the presence of intact authentic splice sites elsewhere in the *YL8A* pre-mRNA. A 2–4 fold reduction in levels of correctly spliced transcript is also observed, extending the role for intron complementarities in splicing efficiency of single intron pre-mRNAs (20–23) to both introns of *YL8A*. In addition to their role in splicing efficiency, the intron complementarities act to prevent skipping and enforce inclusion of the internal exon by determining intron identity.

To show that the splicing defects we observed were caused by disruption of base-pairing interactions within intron 1, we combined mutations X and Y, which together restore complementarity within intron 1 (Fig. 3, mutant XYCD). The exon skipping defect observed in the XBCD and AYCD mutants is completely suppressed in the XYCD double mutant (Fig. 3*C*), consistent with the idea that the complementarities act to form base-pairing interactions between sequences located near the splice sites. As expected, destroying both intron complementarities (mutant AYC¥) without intentionally creating any new base-pairing potential does not restore exon inclusion (Fig. 3*C*). These results provide genetic evidence that base-pairing interactions between the regions near the correct splice sites function to ensure exon inclusion in *YL8A*, by stimulating the pairing of the appropriate authentic splice sites. Thus, intron self-complementarities function as internal identity elements

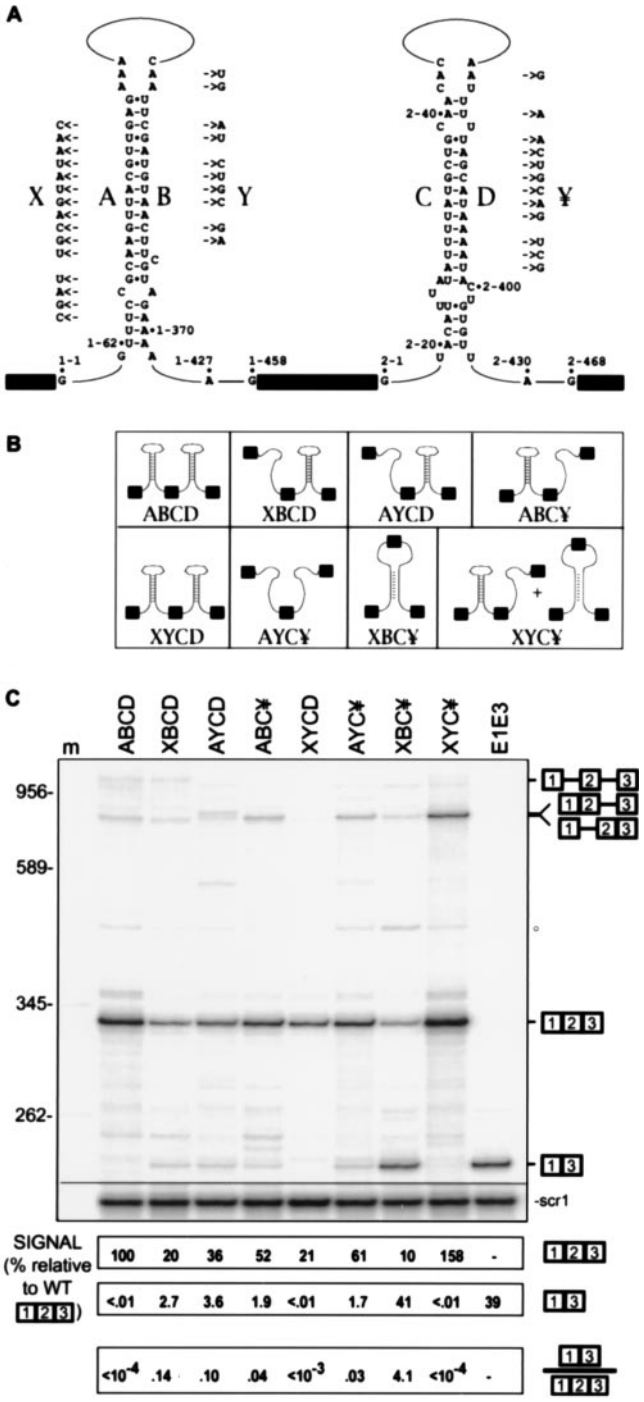

FIG. 3. Intron self-complementarities ensure exon-inclusion. (*A*) Sequences of intron complementarities and mutant derivatives. Intron complementarities between regions downstream of the 5′ splice site and upstream of the branchpoint in introns 1 and 2 (designated A/B and C/D, respectively), and their predicted secondary structures. Additional secondary structure is predicted for the regions internal to each intron (data not shown). Mutations (X, Y, and ¥) that replace intron complementarities are shown next to the affected element. Complementarity exists between X/Y and between X/¥. For each intron, 5′ and 3′ splice site guanosine and branchpoint adenosine residues are indicated. Intron sequences are numbered 1–1 through 1–458 for intron 1, and 2–1 through 2–468 for intron 2. (*B*) Predicted effect of the mutations on substrate secondary structure. Presence or absence of pairing is shown schematically for each construct. (*C*) Splicing phenotypes of the mutant constructs. Splicing was analyzed by reverse transcription of total cell RNA using a 5′-labeled (^{32}P) primer complementary to *CUP1* sequences. Lane m, DNA size markers. E1E3 is a marker for exon skipping expressed from a construct in which exons 1 and 3 are directly fused. Expected products are diagrammed at the right. The different unspliced pre-mRNAs are indicated by an asterisk (*). Spliced products were measured relative to scr1; the amount of exon included mRNA in wild type was taken as 100%.

to help define introns in a multi-intron pre-mRNA by base pairing.

New Complementarities Can Stimulate the Inappropriate Pairing of Authentic Splice Sites. To test the ability of intron complementarities to stimulate the use of alternative pairs of authentic splice sites, we introduced new base-pairing potential between the regions near the intron 1 5′ splice site and the intron 2 branchpoint (Fig. 3 *A* and *B*, mutant XBC¥). Exon skipping is significantly enhanced in this construct (Fig. 3*C*), clearly demonstrating that new complementarity can stimulate pairing of distal splice sites separated by large distances. Sequencing of double stranded cDNA after amplification indicates that the intron 1 5′ splice site is correctly joined to the intron 2 3′ splice site in the skipped product (data not shown). In this case the long-range splice site pairing supported by complementarities is more efficient than the local pairing events not supported by complementarities (Fig. 3*C*; see also ref. 21).

To test whether restoring the intron 1 complementarity in the efficiently skipping XBC¥ construct could suppress exon skipping and restore exon inclusion, the Y sequence was introduced upstream of the intron 1 branchpoint (Fig. 3 *A* and *B*, mutant XYC¥). This should restore identity to intron 1, while maintaining the pairing that promotes exon skipping. When mutant XYC¥ is expressed in yeast, exon skipping is completely suppressed, suggesting that base pairing between the more distant X¥ complementarity is no longer able to promote exon skipping. The complete absence of exon skipping is somewhat surprising given that intron 2 remains compromised in mutant XYC¥ (as it is in mutant ABC¥). The weaker complementarity in the natural AB sequences as compared with the XY complementarity is consistent with the hypothesis that the potency of a complementarity to influence splice site pairing is in part a function of its stability (23). In any case, our results demonstrate that exon skipping caused by introduction of a novel complementarity can be overcome by increasing the intron identity afforded by intron secondary structure.

DISCUSSION

Natural intron secondary structure plays a positive role in ensuring inclusion of an internal exon in a multi-intron pre-mRNA (Fig. 3). The ability of complementarity between sequences adjacent to the 5′ splice site and branchpoint to enhance splice site pairing has been demonstrated previously using artificial splice site competition substrates (21). In the case of *YL8A*, two distinct pairs of naturally complementary sequences act positively to stimulate correct pairing of authentic splice sites. Because the two pairs of elements differ in sequence, base pairing between intron 1 and 2 is inefficient, and the incorrect pairing of splice sites that leads to skipping of exon 2 is discouraged (Fig. 4*A*). The difference in sequence thus confers identity upon the introns, organizing them independently and possibly sequestering them from exon sequences in the folded pre-mRNA. We show that intron identity can be blurred by introducing new complementarity between normally unpaired authentic splice sites (Fig. 4*B*). Strong and independent intron identity enforces exon inclusion by ensuring that each intron is separately defined. This is an example of an intron definition mechanism that uses intrinsic intron RNA secondary structure as the defining feature.

Our results are genetic, and the evidence for RNA folding in *YL8A* pre-mRNA is inferred from the suppression afforded

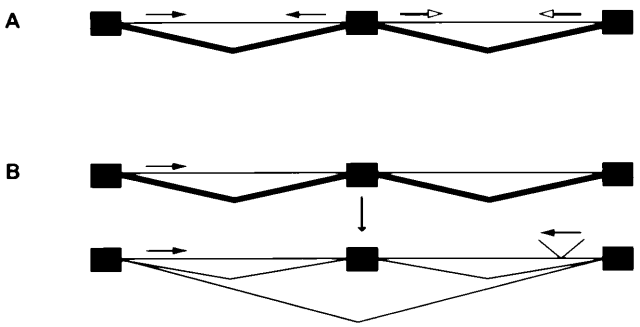

FIG. 4. Intron complementarities and splicing. (*A*) Intron complementarities (inverted arrows) coordinate splice site pairing in yeast *YL8A*. Open arrows represent one complementarity and filled arrows represent a distinct complementarity. Pairing between complementarities defines introns and enforces exon inclusion. (*B*) New complementarity within a splicing substrate induces a new pattern of splicing. In *YL8A* a new engineered complementarity (arrows) causes exon skipping as well as allowing residual exon inclusion (Fig. 3). In higher organisms, the introduction of new complementarity by the transposition of a mobile repeat element (arrows) is proposed to induce an alternative splicing pathway and increase the coding potential of the genome (see text).

by compensatory mutations. RNA structure probing experiments using synthetic single intron transcripts *in vitro* indicate that intron folding is stimulated by such elements (23, 35), and this property is correlated with improved splicing complex formation (20, 22, 35). Presumably the organization of the pre-mRNA into folded intron and exon domains assists the correct pairing of splice sites, but the true impact of these elements on folding and splicing *in vivo* will be challenging to determine. Of the four known multi-intron yeast genes, only *YL8A* and its sister gene *YL8B* have obvious intron complementarities (data not shown). The absence of such elements in *MAT***a***1* and *SLC1* suggests that other mechanisms enforce inclusion of internal exons in transcripts from those genes.

Could long-range intron complementarities act positively to coordinate splice site pairing and ensure exon inclusion in higher eukaryotes? Local secondary structure that includes splice sites can negatively influence splice site selection and cause splice site or exon skipping (for a review, see ref. 36). A large number of experiments support a positive role in exon inclusion for interactions between splicing enhancer sequences in pre-mRNA and protein factors that direct the binding of constitutive components of the splicing machinery to nearby splice sites (for a review, see ref. 15). Despite the predominant role of these RNA-protein interactions in metazoan splicing, complementarity introduced between introns activates exon skipping in artificial constructs transfected into HeLa cells (37, 38), suggesting that complementarities like those in *YL8A* could play a role in regulating splice site pairing in humans. Circumstantial evidence also suggests that natural intron complementarities can influence exon inclusion in mammals. For example, it has been proposed that complementarities contribute to skipping of an optional exon in the mouse neural cell adhesion molecule pre-mRNA (39). In the mouse Sry transcript, inverse splicing occurs in the adult but not fetal testis (40), and this is correlated with the presence of large complementarities flanking the exon (41).

Abundant potential for formation of RNA duplex exists in heterogeneous nuclear RNA (42). Much of this complementarity is derived from *Alu* and other mobile repeat element family members, which are common in higher eukaryotic genomes (43). Because their transposition into coding sequences is disruptive, these elements are often found in large introns, occasionally in inverse orientation with respect to each other (44). Although no systematic studies have been done to date, it seems possible that extended intron complementarities formed by multiple copies of *Alu* or other repeat family sequences could help define and increase the splicing efficiency of very large metazoan introns by a mechanism that parallels the one observed in yeast (refs. 20–23 and this work). Furthermore, the transposition of a new copy of such a sequence could promote a new pathway of alternative splicing in pre-existing pre-mRNAs (Fig. 4*B*). Such events would have the advantage of increasing the combinatorial potential of splicing and provide a direct role for so-called "selfish" DNAs in generating protein diversity. A complementary evolutionary role based on the effects of these elements on transcription regulation has recently been proposed (45).

We thank Amy Kistler and Christine Guthrie for valuable discussions, and Michael Rosbash, Alan Zahler, Sherif Abouelela, Rhonda Perriman, and Kevin Wilson for helpful comments on the manuscript. Bryn Weiser produced the secondary structures from which the *YL8A* and *YL8B* intron complementarities were identified. Keiko Mizuta provided pYL8A and pYL8B, Christine Guthrie provided plasmid pGAC14, and Allan Jacobson provided yeast strain HFY870. This work was supported by a Graduate Assistance in Areas of National Need (GAANN) Fellowship from the Department of Education (to K.J.H.) and Grant GM40478 from the National Institutes of Health (to M.A.).

1. Green, M. (1991) *Annu. Rev. Cell Biol.* **7,** 559–599.
2. Moore, M. J., Query, C. C. & Sharp, P. A. (1993) in *The RNA World*, eds. Gesteland, R. F. & Atkins, J. F. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 303–357.
3. Ruby, S. R. & Abelson, J. (1988) *Science* **242,** 1028–1035.
4. Seraphin, B. & Rosbash, M. (1991) *EMBO J.* **10,** 1209–1216.
5. Michaud, S. & Reed, R. (1993) *Genes Dev.* **7,** 1088–1020.
6. Hawkins, J. D. (1988) *Nucleic Acids Res.* **16,** 9893–9908.
7. Nakai, K. & Sakamoto, H. (1994) *Gene* **141,** 171–177.
8. Krawczak, M., Reiss, J. & Cooper, D. N. (1992) *Hum. Genet.* **90,** 41–54.
9. Chu, G. & Sharp, P. (1981) *Nature (London)* **289,** 378–382.
10. Sterner, D. A, Carlo, T. & Berget, S. M. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 15081–15085.
11. Talerico, M. & Berget, S. (1990) *Mol. Cell. Biol.* **10,** 6299–6305.
12. Robberson, B. L., Cote, G. J. & Berget, S. M. (1990) *Mol. Cell. Biol.* **10,** 84–94.
13. Talerico, M. & Berget, S. (1994) *Mol. Cell. Biol.* **14,** 3434–3445.
14. Kuo, H.-C., Nasim, F.-U. H. & Grabowski, P. J. (1991) *Science* **251,** 1045–1051.
15. Black, D. (1995) *RNA* **1,** 763–771.
16. Miller, A. M. (1984) *EMBO J.* **3,** 1061–1065.
17. Dick, T., Surana, U. & Chia, W. (1996) *Mol. Gen. Genet.* **251,** 38–43.
18. Mizuta, K., Hashimoto, T. & Otaka, E. (1992) *Nucleic Acids Res.* **20,** 1011–1016.
19. Mizuta, K., Hashimoto, T. & Otaka, E. (1995) *Curr. Genet.* **28,** 19–25.
20. Newman, A. (1987) *EMBO J.* **6,** 3833–3839.
21. Goguel, V. & Rosbash. M. (1993) *Cell* **72,** 893–901.
22. Libri, D., Stutz, F., McCarthy, T. & Rosbash, M. (1995) *RNA* **1,** 425–436.
23. Charpentier, B. & Rosbash, M. (1996) *RNA* **2,** 509–522.
24. Bachellerie, J.-P., Michot, B., Nicoloso, M., Balakin, A., Ni, J. & Fournier, M. J. (1995) *Trends Biol. Sci.* **20,** 261–264.
25. Kunkel, T. A. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 488–492.
26. Lesser, C. F. & Guthrie, C. (1993) *Genetics* **133,** 851–863.
27. Hill, J., Ian, K. A., Donald, G. & Griffiths, D. E. (1991) *Nucleic Acids Res.* **19,** 5791.
28. He, F., Peltz, S. W., Donahue, J. L., Rosbash, M. & Jacobson, A. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 7034–7038.
29. Ares, M., Jr., & Igel, A. H. (1990) *Genes Dev.* **4,** 2132–2145.
30. Felici, F., Cesareni, G. & Hughes, J. M. X. (1989) *Mol. Cell. Biol.* **9,** 3260–3268.
31. Jaeger, J. A., Turner, D. H. & Zuker, M. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 7706–7710.
32. Engebrecht, J., Voelkel-Meiman, K. & Roeder, G. S. (1991) *Cell* **66,** 1257–1268.
33. Klinz, F.-J. & Gallwitz, D. (1985) *Nucleic Acids Res.* **13,** 3791–3804.
34. Parker, R. & Patterson, B. (1987) in *Molecular Biology of RNA:*

New Perspectives, eds. Dudock, B. & Inouye, M. (Academic, New York), pp. 133–149.
35. Mougin, A., Gregoire, A., Banroques, J., Segault, V., Fournier, R., Brule, F., Chevrier-Miller, M. & Branlant, C. (1996) *RNA* **2,** 1079–1093.
36. Balvay, L., Libri, D. & Fiszman, M. (1993) *BioEssays* **15,** 165–169.
37. Solnick, D. (1985) *Cell* **43,** 667–676.
38. Solnick, D. & Lee, S. I. *Mol. Cell. Biol.* **7,** 3194–3198.
39. Barbas, J. A., Chaix, J. C., Steinmetz, M. & Goridis, C. (1988) *EMBO J.* **7,** 625–632.
40. Capel, B., Swain, A., Nicolis, S., Hacker, A., Walter, M., Koopman, P., Goodfellow, P. & Lovell-Badge, R. (1993) *Cell* **73,** 1019–1030.
41. Hacker, A., Capel, C., Goodfellow, P. & Lovell-Badge, R. (1995) *Development (Cambridge, U.K.)* **121,** 1603–1614.
42. Jelinek, W. & Darnell, J. E. (1972) *Proc. Natl. Acad. Sci. USA* **69,** 2537–2541.
43. Schmid, C. (1996) *Prog. Nucleic Acids Res. Mol. Biol.* **53,** 283–319.
44. McNaughton, J. C., Broom, J. E., Hill, D. F., Jones, W. A., Marshall, C. J., Renwick, N. M., Stockwell, P. A. & Petersen, G. B. (1993) *J. Mol. Biol.* **232,** 314–321.
45. Britten, R. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 9374–9377.