

Exon sequences at the splice junctions affect splicing fidelity and alternative splicing

Luciana B. Crotti and David S. Horowitz¹

Department of Biochemistry and Molecular Biology, Uniformed Services University of the Health Sciences, 4301 Jones Bridge Rd., Bethesda, MD 20814

Edited by Christine Guthrie, University of California, San Francisco, CA, and approved September 10, 2009 (received for review July 16, 2009)

Identification of splice sites is essential for the expression of most eukaryotic genes, allowing accurate splicing of pre-mRNAs. The splice sites are recognized by the splicing machinery based on sequences within the pre-mRNA. Here, we show that the exon sequences at the splice junctions play a significant, previously unrecognized role in the selection of 3' splice sites during the second step of splicing. The influence of the exon sequences was enhanced by the Prp18 mutant Prp18 Δ CR, and the strength of an exon sequence in Prp18 Δ CR splicing predicted its effect in wild-type splicing. Analysis of the kinetics of splicing in vitro demonstrated that 3' splice sites were chosen competitively during the second step, likely at the same time as exon ligation. In wild-type yeast, splice site selection for two genes studied was altered by point mutations in their exon bases, affecting splicing fidelity and alternative splicing. Finally, we note that the degeneracy of the genetic code allows competing 3' splice sites to be eliminated from coding regions, and we suggest that the evolution of the splicing signals and the genetic code are connected.

pre-mRNA splicing | Prp18 | splice site selection

Identification of splice sites is necessary both for ensuring the fidelity of pre-mRNA splicing and for the regulation of alternative splicing. Mutations that impair the recognition of the splice sites are responsible for a significant fraction of genetic diseases (1). Splice sites are defined by sequence elements in the pre-mRNA that are recognized by many splicing factors (2). The 3' splice site is usually the first AG downstream of the branchpoint and is immediately preceded by a pyrimidine or an A (3, 4). A polypyrimidine tract is often present upstream of the splice site. The influence of each element varies and not all are required (5). Recognition of elements in the 3' splice site region is required early in splicing in metazoans but not in yeast (6).

Identification of the 3' splice site for exon ligation appears to occur in the second step, and these events are well studied in yeast. After the first transesterification, the RNA helicase Prp16 rearranges the spliceosome (7). Prp18, Slu7, and Prp22 join the spliceosome, which then rapidly ligates the exons (8–10). The U2, U5, and U6 snRNAs and the Prp8 protein, which are required for the first step of splicing, also participate directly in the second step (11, 12).

The 3' splice site is identified by various mechanisms. The distance between the branch point and the 3' splice site is important, with distant sites dependent on their polypyrimidine tracts (13, 14). The G's at the 5' and 3' ends of the intron interact (13, 15), and base –3 at the 3' site interacts with base +3 at the 5' site and the U6 snRNA (16). Prp8 is involved in recognition of both the polypyrimidine tract and the AG dinucleotide at the 3' splice site (12, 17). Mutation of Slu7 impairs the use of distant sites (18, 19). Analyses of in vivo splicing suggest that the 3' splice site recognition events occur during the second step, but this conclusion is not well established.

Prp18 stabilizes the interaction of the ends of the exons with loop 1 of the U5 snRNA (20, 21). This function is lost in the Prp18 Δ CR protein that lacks a conserved 24-amino acid region (22). In splicing with Prp18 Δ CR, the second step is slowed, and its rate depends critically on the exon bases at both the 5' and 3' splice junctions. In general, exon sequences that can base-pair more stably to loop 1 of

the U5 snRNA facilitate the second step of Prp18 Δ CR splicing. The exon sequences have a minimal effect on wild-type splicing (20). Prp18 likely acts together with Prp8 to stabilize interaction of the exons with U5 (23, 24).

In this study we assessed the role of exon sequences at the splice junctions in 3' splice site selection. We found that the exon sequences are important in both model and natural substrates, suggesting that their role is generally significant, and that their influence is evident only during the second step of splicing.

Results

Splicing of Model Substrates in Vivo. To assess the effect of exon sequences on 3' splice site selection in yeast, we made *ACT1-CUP1* reporters with two 3' splice sites (Fig. 1A). We placed one 3' splice site 14 bases from the branch site in an *ACT1*-like sequence that is used efficiently (19). In the substrate AS1 the second 3' splice site was placed 19 bases downstream of the first site in a slightly altered *ACT1* context. In this model substrate, we found that the splice sites were chosen at similar levels. Only the mRNA resulting from splicing to the downstream 3' splice site can be translated into Act1-Cup1 fusion protein. Act1-Cup1 protein production is quantitatively measured by copper resistance (25). We used ag|ACA and ag|TTG as examples of strong and weak exon sequences (the | symbol denotes the splice junction), based on their strengths in Prp18 Δ CR splicing (20).

Individually, both 3' splice sites in AS1 were used efficiently in wild-type yeast and remained sensitive to exon sequence in *prp18 Δ CR* yeast (Fig. 1B, 1–8, 24, 25). In WT *ACT1*, splicing was sensitive to exon sequence in *prp18 Δ CR* yeast (Fig. 1B Left, 1–4), as reported in ref. 20. Splicing to the upstream site was assayed using the Sh substrate (Fig. 1A and B, 5–8), and splicing to the downstream site, by blocking the upstream site in AS1 (Fig. 1B, 24–25). In *prp18 Δ CR* yeast, the exon sequences were influential at both sites, but moving the 3' splice site closer to the branch site improved splicing of weaker substrates (comparing Fig. 1B Left, 1–4 with 5–8 and 24–25), consistent with earlier work (26). In wild-type yeast, the upstream and downstream 3' sites were used efficiently independent of the exon sequence (Fig. 1B Right), as expected (20).

In *prp18 Δ CR* yeast, the exon sequences at the 3' splice junctions had a very large effect on alternative splice site choice (Fig. 1B 9–21 and Fig. S1). The example substrates (9) ag|TTG–ag|TTG, (10) ag|TTG–ag|ACA, (13) ag|ACA–ag|TTG, and (14) ag|ACA–ag|ACA conferred copper resistances of 0.1, 1.3, 0.04, and 0.4 mM. With either upstream site, changing the downstream site from weak to strong increased copper resistance by \approx 10-fold. With either downstream site, changing the upstream site from weak to strong reduced resistance by 3- to 4-fold, giving an overall maximum change of 30-fold in copper resistance (comparing 10 and 13).

Author contributions: L.B.C. and D.S.H. designed research; L.B.C. performed research; L.B.C. and D.S.H. analyzed data; and D.S.H. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

¹To whom correspondence should be addressed. E-mail: dhorowitz@usuhs.mil.

This article contains supporting information online at www.pnas.org/cgi/content/full/0907948106/DCSupplemental.

Overall, the results show that the exon sequences have a significant effect when there is a choice between 3' splice sites.

The effects of several exon junction sequences were examined. In both wild-type and *prp18ΔCR* yeast the ordering of strengths of the upstream splice sites, $|ACA > |AGG > |CAA \geq |TTG$ (comparing 9–21 in Fig. 1B), was in accord with that found for single-site splicing in *prp18ΔCR* yeast (20). Primer extension results (Fig. 2B, lanes 12–15) were consistent with the copper-resistance measurements. At the downstream site, the ordering of strengths, $|ACA = |AAA > |AGG = |GAA \geq |CAA > |TTG$ (Fig. 1B and Fig. S1), was generally in accord with previous work, with the exception that $|CAA$ was stronger than expected at the downstream site. The overall consistency of our results with expectations based on single-site strengths argues strongly that the changes in splice site choice we observe are dictated by the strength of the adjacent exon sequences. The observation that the exon sequences shift splice site choice in wild-type yeast in the same direction as in *prp18ΔCR* yeast argues that the mechanistic basis for the shifting is the same in wild-type and *prp18ΔCR* yeast.

We previously showed that superior 5' splice sites could compensate for weak 3' splice sites (20). We replaced the exon sequence at the 5' splice site with $AAAA|gt$ in the AS1 substrates $ag|TTG-ag|TTG$ and $ag|TTG-ag|ACA$ (Fig. 1B, 22 and 23). $AAAA|gt$ increased upstream splicing (and reduced downstream splicing) in $ag|TTG-ag|ACA$ (comparing 23 with 10 in Fig. 1B, and lanes 17 and 6 in Fig. 2A and B) by improving splicing to the upstream $|TTG$ site. With $AAAA|gt$, the difference between the substrates 22 and 23 is lessened, compared with 9 and 10 (Fig. 1B). The results are consistent with the notion that the relative strengths of the exon sequences determines which site is chosen. The results show that the 5' splice site can affect the choice of 3' splice sites.

We tested alternative splicing in the AS2 substrate (Fig. 1A), in which the downstream 3' splice site was placed at the wild-type distance from the branch point in an *ACT1* context. The effects of exon sequence in the AS2 substrate (Fig. 1B, 26–29) were similar to, but slightly smaller than, those in AS1 (see Fig. S1).

Alternative Splicing in Vitro. We assayed the splicing of representative AS1 pre-mRNAs in vitro. In *Prp18ΔCR* splicing (Fig. 3A, lanes 1–5), strengthening the downstream site from $|TTG$ to $|AAA$ resulted in more downstream splicing with either $|TTG$ or $|ACA$ as the upstream site (comparing lanes 2 and 3, or 4 and 5). The 4-fold shift in the ratio of proximal to distal splicing in lanes 2 and 3 was similar to the 6-fold shift seen in vivo (Fig. 2A lanes 5 and 8). Strengthening the upstream sites from $|TTG$ to $|ACA$ shifted splicing sharply to the upstream site (comparing lanes 2 and 4, or 3 and 5). Exchanging the strong and weak exon sequences changed the ratio of proximal to distal splicing by at least 50-fold (lanes 3 and 4). In wild-type splicing (Fig. 3A, lanes 6–10), the effects were smaller; strengthening the downstream site from $|TTG$ to $|ACA$ increased downstream splicing approximately twofold (lanes 7 and 8). Placing $|ACA$ at the upstream site shifted essentially all of the splicing to the that site (lanes 9 and 10). Overall, the in vitro results were similar to the in vivo results. The proximal to distal ratios were higher in vitro. This difference might result from slow (relative to splicing in vivo) assembly of Slu7, Prp18, and/or Prp22 onto the substrate, which is expected to favor the upstream site (9, 19, 26), or from the relative instability of the (untranslatable) proximal mRNA in vivo.

We analyzed the kinetics of splicing of pre-mRNAs with $ag|TTG$ as the upstream site and $ag|TTG$, $ag|AAA$, or $cc|TTG$ as the downstream site (Fig. 3B and C). In *Prp18ΔCR* splicing (Fig. 3B), splicing of $ag|TTG-cc|TTG$ (lanes 13–18) was slowest, accumulated the most intermediates and yielded the most proximal mRNA. The $ag|TTG-ag|AAA$ pre-mRNA (lanes 7–12) was spliced fastest accumulating the least intermediates and giving the least proximal mRNA. We calculated second-step rate constants (8, 20, 27) using one rate constant to describe the second step for each product (18)

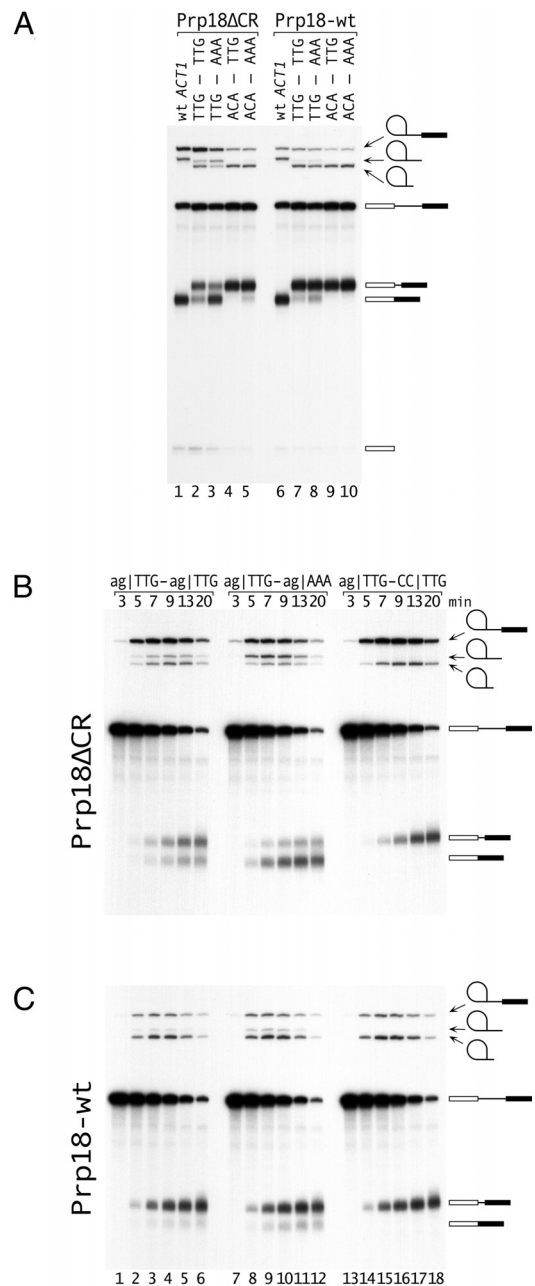


Fig. 3. Analysis of alternative 3' splicing in vitro. (A) Splicing of four substrates with alternative 3' splice sites and wt *ACT1* was assayed using extract from a *PRP18*-knockout strain supplemented with either *Prp18ΔCR* or *Prp18-wt* protein. The sequences of the three exon bases at each 3' splice site are shown at the top; the substrates were based on AS1 (Fig. 1A). The positions of the pre-mRNA, lariat-intron and exon1 intermediates, released lariats (long and short), and alternatively spliced mRNAs are indicated. Splicing was carried out for 15'. (B and C). Kinetics of splicing of three substrates with (B) *Prp18ΔCR* and (C) *Prp18-wt* proteins. The 3' splice sites of the AS1-based substrates are shown at the top of B. In the substrate in lanes 13–18, the downstream AG was mutated to CC. The positions of the RNAs are indicated.

(Fig. S3). In *Prp18ΔCR* splicing (Fig. 3B), upstream splicing had the same rate for all three substrates whereas the downstream rates changed depending on the exon sequence (Table 1). This result implies that the 3' splice site was chosen competitively during the second step; the reduction in the amount of proximal product was accounted for by more rapid splicing to the distal site. The ratio of the two mRNAs was approximately constant over the reaction

Table 1. Second-step rate constants

Prp18	Splice site	<i>ACT1</i>	ag TTG- cc TTG	ag TTG- ag TTG	ag TTG- ag ACA
Wild-type	prox.	1	1.0	1.1	1.0
	dist.	–	0	0.18	0.34
Prp18ΔCR	prox.	0.23	0.13	0.13	0.12
	dist.	–	0	0.06	0.22

Rate constants for the second step of splicing. The rate constant is the ratio of the rate of formation of mRNA to the amount of intermediate (Fig. S3) (20, 27). Values were calculated from replicates of the experiments in Figures 3B and 3C (20) and were normalized to *ACT1* wild-type splicing. Measurements were reproducible to within 10%. prox., proximal; dist., distal.

course, consistent with a single kinetic step. The wild-type splicing results (Fig. 3C) support the idea of competition between splice sites, although the numbers were less compelling because of the dominance of proximal splicing.

The primer extension analysis strongly supports the idea that the 3' splice sites are chosen in a competition that is affected by the exon sequences. In all of the sets of substrates shown in Fig. 2, the amount of product at the upstream site is affected by exon bases at the downstream site, and vice versa. Only a competition can explain the changes in splicing to one site when the other site is altered (14). Previous work showed that the efficiency of splicing in *prp18ΔCR* yeast is affected by exon bases at the splice junctions (20). However, models in which splice site selection occurs independent of exon sequence, followed by a second step whose efficiency depends on exon sequence are not consistent with the results presented here. In any model consistent with our results, the exon sequences must be involved in the selection of 3' splice sites, not just in the efficiency of the second step. The in vitro results are explained by a model with a single kinetic step, but more complicated models cannot be excluded.

Effects of Exon Base Changes in Natural pre-mRNAs. We studied the effects of exon bases on splicing of two natural substrates, *LSM7* and *REC102*. In *LSM7* (Fig. 4A) the correct 3' splice site at +1, aag|AAA, has a strong exon sequence but has an uncommon "a" at base –3. The downstream site at +8 has a strong exon sequence, and 1 of 9 pre-mRNAs in a large-scale cDNA sequencing project was spliced at this site (28). A second potential downstream site at +25 has the weak exon sequence |CAA. We weakened the authentic 3' splice site by changing base +1 from A to C (Fig. 4A, mutants 1 and 3) and strengthened the +25 site by changing base +25 from C to A (mutants 2 and 3), thereby exchanging the strong |AAA and weak |CAA exon sequences.

We analyzed the splicing of wild-type and mutant *LSM7* in wild-type yeast (Fig. 4B). Primer extension (Fig. 4B, lane b) and RT-PCR/sequencing analysis showed splicing only to the authentic +1 site. Mutation of the A at +1 to C (Fig. 4A, mutant 1) resulted in some splicing to the +8 site and a trace of splicing to the +25 site (Fig. 4B, lane c). Strengthening the +25 site from |CAA to |AAA did not have any effect by itself (Fig. 4A and B, mutant 2), but when combined with weakening of the +1 site, this mutation resulted in considerable splicing to the +25 site (mutant 3). In mutant 3, 40% of the splicing was to the +8 and +25 sites, which were used approximately equally (Fig. 4B, lane e). The second-step rates, estimated from the ratios of mRNA to intermediate (Fig. 4B), were consistent with notion that the 3' splice sites were in competition.

In *prp18ΔCR* yeast, the second step of splicing of *LSM7* was slow (Figs. 4B and C, lane h). Weakening the +1 site greatly reduced wild-type splicing and resulted in very limited splicing to the +8 site (Figs. 4B and C, lane i). Strengthening the +25 site (lane k) resulted in a trace of splicing to that site, detectable by RT-PCR (Fig. S4), but activation of this site was less than in wild-type yeast. Levels of all mRNAs are depressed in *prp18ΔCR* yeast (20, 21); our model

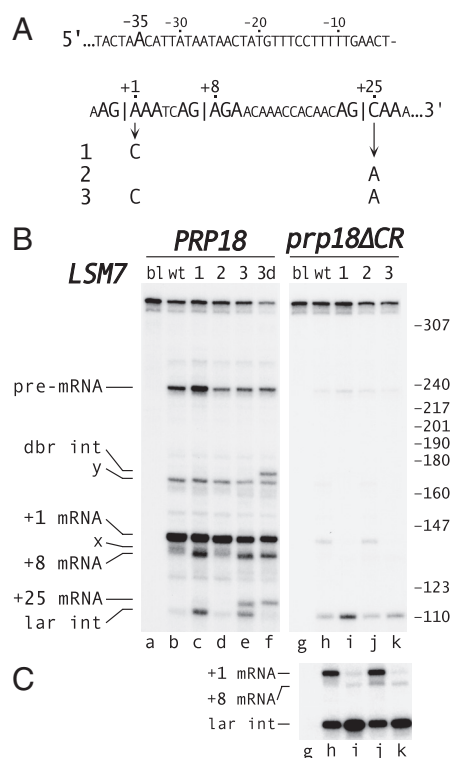


Fig. 4. Effect of exon bases at the splice junctions on fidelity of *LSM7* splicing. (A) Sequence of the 3' splice site region of *LSM7*, emphasizing the branch site at –35 and two downstream AG dinucleotides. Three *LSM7* mutants are shown. (B and C). Primer extension analysis of splicing of wt *LSM7* and the three mutants in *PRP18 ΔUPF1* and *prp18ΔCR ΔUPF1* yeast. Alleles of *LSM7* are indicated at the top. B shows a single, uniformly adjusted exposure, and C shows a longer exposure of the *prp18ΔCR* samples from a separate gel. The positions of the pre-mRNA and mRNAs (numbers denote the first base of exon 2 in the spliced product) are shown at the Left. The positions of the lariat intermediate and debranched intermediate are indicated. To identify the primer extension products, RT-PCR products (using the original primer and a primer at the 5' end of the RNA) were purified from a polyacrylamide gel and sequenced. The lariat intermediate was identified by treating the RNA from mutant 3 with Dbr1 protein before primer extension, leading to a loss of lariat intermediate and the appearance of a new band of the expected size (lane f). The lengths of the primer extension products agree with their calculated values. The identities of the bands "x" and "y" are unknown; RT-PCR results argue that "x" is distinct from +8 spliced product.

substrates show that we can interpret splicing of pre-mRNAs independent of this reduction.

Splicing of the *REC102* pre-mRNA uses alternative 3' splice sites (29, 30). Fig. 5A shows the 4 AG dinucleotides that could be 3' splice sites. Only the +30 site allows translation of the *REC102* mRNA. We strengthened the weak upstream (+1) site, |CTT, and weakened the relatively strong downstream (+30) site, |ATT (Fig. 5A).

We assayed splicing of wild-type and mutant *REC102* by primer extension in both *PRP18-wt* and *prp18ΔCR* yeast (Fig. 5B). All 4 potential 3' splice sites were used in wild-type yeast (Fig. 5B, lane b) with the +1 and +30 sites used approximately equally, and the +19 and +25 sites, ≈5-fold less. Strengthening the +1 site (mutants 1 and 5) increased its utilization at the expense of the other sites (Fig. 5B, lanes c and g). Weakening the +30 site (mutants 2 and 4) reduced its usage somewhat (lanes d and f). Combining the mutations (mutants 3 and 6, lanes e and h) gave the largest shifts in splicing, increasing the ratio of +1 to +30 site usage by threefold. The effects of exon sequences were larger in *prp18ΔCR* yeast than in wild-type yeast. The +1 site was used preferentially with wild-type *REC102* (lane j), and strengthening it increased its usage two- to threefold (lanes k and o). Weakening the +30 site reduced its

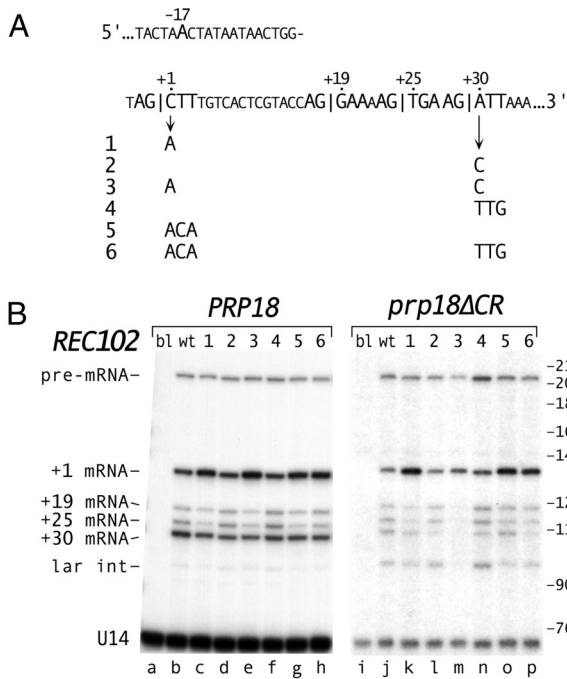


Fig. 5. Effect of exon bases at the splice junctions on *REC102* splice site choice. (A) Sequence of *REC102* including its 3' splice sites and branch site. Six mutations at the exon junctions are shown. (B) Primer extension analysis of splicing of *REC102*. The mutants of *REC102* (blank, wt and 1–6) and the *PRP18* Δ UPF1 (lanes a–h) and *prp18* Δ CR Δ UPF1 (lanes i–p) yeast are indicated at the top. The positions of the pre-mRNA, 4 mRNAs (numbers denote the first base of exon 2 in the spliced product), and lariat intermediate are indicated. Primer extension products were annealed to an oligo within exon1 and cut with HpaI before electrophoresis to facilitate separation (Fig. S5). To identify the primer extension products, RT-PCR products (using the original primer and a primer in exon 1) were purified from denaturing polyacrylamide gels and sequenced. The lariat intermediate was identified by treating the RNA from *prp18* Δ CR yeast with Dbr1 before primer extension (Fig. S5). The lengths of the primer extension products agree with their calculated values. B was assembled from two separate primer extension experiments. In direct comparisons, the amounts of spliced products in *prp18* Δ CR yeast were two- to fourfold less than in wild-type yeast.

usage up to threefold (lanes l and n). Combining the mutations (mutants 3 and 6) shifted splicing fivefold (lane p) to >10-fold (lane m) to the +1 site.

Discussion

Our experiments show that the selection of the 3' splice site is significantly influenced by exon bases at the splice junctions. The choice of splice sites is made competitively during the second step, and the exon sequences play an important role in selection in natural substrates. Previous work showed that exon sequences at both splice junctions significantly affect the rate of the second step of splicing with the mutant Prp18 Δ CR protein but have almost no effect on the rate of wild-type splicing (20). Here, in model pre-mRNAs with two 3' splice sites, we found that the choice of the 3' splice site in *prp18* Δ CR splicing was governed by the exon bases at the 3' splice junction, with nearly quantitative switches in splice site choice with the strongest and weakest exon sequences. The strengths of exon sequences in alternative splicing and in single-site *prp18* Δ CR splicing were well-correlated, consistent with a common mechanism for both processes. Surprisingly, the exon sequences had a significant effect on 3' splice site selection by wild-type spliceosomes, changing the ratio of alternatively chosen sites in model substrates severalfold. Analysis of splicing kinetics in vitro showed that the 3' splice site was chosen in competition during the second

step. The dependence of the choice on Prp18 strongly suggests that the choice is coincident with exon ligation. In wild-type yeast, the exon sequences affected splice site choice in natural pre-mRNAs. Mutation of the exon bases at the splice junctions altered the splicing of the pre-mRNAs of *LSM7* and *REC102*, suggesting a general role for these bases in splice site selection.

Exon Sequences and Splice Site Choice. The exon bases at the splice junctions are an important determinant of 3' splice site choice, affecting both splicing fidelity and alternative splicing. The activation of the downstream splice sites in *LSM7* by the mutations in the exon bases demonstrates the importance of the exon sequences in ensuring the fidelity of *LSM7* splicing. Many transcripts will have a potential 3' splice site—a YAG or AAG—within 25 bases of the correct 3' splice site. Splicing fidelity is likely ensured by a combination of 3' splice site determinants, including a pyrimidine tract, the –3 base, and the distance from the branch site (4, 13, 14, 16). For *LSM7*, point mutations in the exon bases shifted splicing to a site that lacks a pyrimidine tract and is distant (60 bases) from the branch site, suggesting that the exon bases play an important and general role in ensuring the fidelity of splice site choice.

The *REC102* results show that exon bases are important in determining alternative 3' splice site choice. Alternative splicing is uncommon in *S. cerevisiae* but is common in higher eukaryotes. Despite differences in yeast and metazoan splicing, yeast Prp18 can replace human Prp18 in splicing in vitro (31), suggesting that 3' splice site identification mechanisms have been conserved. In metazoans, the 3' splice site region is identified before the first step of splicing, but the 3' splice site for exon joining is chosen after the first step (2, 5). Exon sequences could affect this second-step selection of 3' splice sites. Our results could apply to cases where alternative 3' splice sites are chosen using the same branch site (32), including the closely spaced 3' splice sites that occur often (33).

Mechanism of Choosing the 3' Splice Site. Our results show that the 3' splice site is chosen competitively during the second step of splicing. The sensitivity of the outcome to Prp18 argues that the choice is made during the Prp18-dependent, ATP-independent stage of the second step during which the second transesterification occurs (8). After the first step of splicing, ATP hydrolysis by Prp16 leads to protection of the 3' splice site (7). We think that the ATP hydrolysis activates the spliceosome, permitting binding of Slu7, Prp18, and Prp22, and that this second-step spliceosome selects the 3' splice site and carries out the reaction (similar to the scheme of (9, 10)). This model has a single kinetic step for identifying the 3' splice site and carrying out the reaction. Our results are inconsistent with mechanisms in which the spliceosome selects a 3' splice site before binding of Prp18, with Prp18 needed only to facilitate the transesterification. Previous work on 3' splice site selection in yeast argued against directional scanning mechanisms and in favor of competition (13, 14), in agreement with our conclusions.

Our work shows that Prp18 can influence splice site selection. Wild-type Prp18 mutes the response to different exon sequences, which can cause a 100-fold shift in 3' splice site choice in *prp18* Δ CR yeast. The second-step proteins Slu7 and Prp8 also affect 3' splice site selection in that mutants are sensitive to the branchpoint-3' splice site distance or the presence of a polypyrimidine tract (17, 18, 34). Slu7, Prp8, Prp22, and Prp18 act in concert during the transesterification stage of the second step (9, 19, 23, 26), and this stage appears to be critical for defining the 3' splice site.

Competing 3' Splice Sites and the Genetic Code. We suggest that the competition between splice sites during the second step may have been a force in evolution. Any AG near the correct splice site can be a competitor because no other bases are required for a 3' splice site (although GAG is strongly disfavored (3, 4, 16)). Problems would be expected if the encoded protein's function demanded an amino acid encoded by AGN or NAG near a 3' splice site. However,

all such AG-containing codons have substitutes that lack an AG dinucleotide. For NAG, NAA is always a synonymous codon, and this third position degeneracy is typical in the genetic code. For AGN, it is unusual that there are substitute codons; AGY is Ser, which is also encoded by UCN, and AGR is Arg, which is also encoded by CGN. (An AG dinucleotide is required in the rare case in which a selenocysteine UGA codon is followed by a GNN codon, resulting in a weak GAG 3' splice site.) Thus, an AG dinucleotide is never required for coding, and we suggest that splicing and/or the genetic code may have evolved to avoid potential difficulties with 3' splice site selection using this degeneracy of the genetic code. Only one other amino acid, Leu, is encoded by different initial bases (CUN and UUR), making CU another avoidable dinucleotide.

The genetic code is thought to have evolved to minimize replication and translation errors (35), and it continues to evolve slowly (36). It is not known when pre-mRNA splicing developed relative to the fixing of the genetic code, but several arguments favor its presence early in evolution (37, 38). We can redirect a modern spliceosome to an incorrect site. We surmise that an ancient spliceosome could be more easily mis-directed, and that this difficulty could have provided sufficient selective pressure to affect the evolution of the genetic code. Alternatively, splicing could have evolved to use AG. The spliceosome can, for example, use AC as the 3' splice site (15), and the dominant use of AG might reflect evolution to nonconflicting signals.

Materials and Methods

Yeast Strains. The copper-reporter strain YLC10 (*prp18::KAN^r*) was described in ref. 20. *UPF1*-knockout strains were made by recovering the *upf1::KAN^r* fragment from BY4741 (*YMR080C::KAN^r*) (Research Genetics) by PCR, transforming it either into DB120 (*PRP18/prp18::URA3*) (derived from W303 (*al α leu2-3,112 his3-11,15 trp1-1 can1-100 ade2-1 ura3-1*)) and dissecting to obtain DB127 (*MAT α upf1::KAN^r prp18::URA3*), or into W303a to obtain DB129 (*upf1::KAN^r*).

- Lopez-Bigas N, Audit B, Ouzounis C, Parra G, Guigo R (2005) Are splicing mutations the most frequent cause of hereditary disease? *FEBS Lett* 579:1900–1903.
- Wahl MC, Will CL, Lührmann R (2009) The spliceosome: Design principles of a dynamic RNP machine. *Cell* 136:710–718.
- Sheth N, et al. (2006) Comprehensive splice-site analysis using comparative genomics. *Nucl Acids Res* 34:3955–3967.
- Spingola M, Grate L, Haussler D, M Ares J (1999) Genome-wide bioinformatic and molecular analysis of introns in *Saccharomyces cerevisiae*. *RNA* 5:221–234.
- Moore MJ (2000) Intron recognition comes of AGE. *Nat Struct Biol* 7:14–16.
- Rymond B, Rosbash M (1985) Cleavage of 5' splice site and lariat formation are independent of 3' splice site in yeast mRNA splicing. *Nature* 317:735–737.
- Schwer B, Guthrie C (1992) A conformational rearrangement in the spliceosome is dependent on PRP16 and ATP hydrolysis. *EMBO J* 11:5033–5039.
- Horowitz DS, Abelson J (1993) Stages in the second reaction of pre-mRNA splicing: The final step is ATP independent. *Genes Dev* 7:320–329.
- Schwer B, Gross CH (1998) Prp22, a DEXH RNA helicase, plays two distinct roles in yeast pre-mRNA splicing. *EMBO J* 17:2086–2094.
- James SA, Turner W, Schwer B (2002) How Slu7 and Prp18 cooperate in the second step of yeast pre-mRNA splicing. *RNA* 8:1068–1077.
- Umen JG, Guthrie C (1995) The second catalytic step of pre-mRNA splicing. *RNA* 1:869–885.
- Grainger RJ, Beggs JD (2005) Prp8 protein: At the heart of the spliceosome. *RNA* 11:533–557.
- Luukkonen BGM, Séraphin B (1997) The role of the branchpoint-3' splice site spacing and interaction between intron terminal nucleotides in 3' splice site selection in *Saccharomyces cerevisiae*. *EMBO J* 16:779–792.
- Patterson B, Guthrie C (1991) A U-rich tract enhances usage of an alternative 3' splice site in yeast. *Cell* 64:181–187.
- Parker R, Siliciano PG (1993) Evidence for an essential non-Watson-Crick interaction between the first and last nucleotides of a nuclear pre-mRNA intron. *Nature* 361:660–662.
- Collins CA, Guthrie C (2001) Genetic interactions between the 5' and 3' splice site consensus sequences and U6 snRNA during the second catalytic step of pre-mRNA splicing. *RNA* 7:1845–1854.
- Umen JG, Guthrie C (1995) A novel role for a U5 snRNP protein in 3' splice site selection. *Genes Dev* 9:855–868.
- Frank D, Guthrie C (1992) An essential splicing factor, SLU7, mediates 3' splice site choice in yeast. *Genes Dev* 6:2112–2124.
- Brys A, Schwer B (1996) Requirement for SLU7 in yeast pre-mRNA splicing is dictated by the distance between the branchpoint and the 3' splice site. *RNA* 2:707–717.
- Crotti LB, Bačíková D, Horowitz DS (2007) The Prp18 protein stabilizes the interaction of both exons with the U5 snRNA during the second step of pre-mRNA splicing. *Genes Dev* 21:1204–1216.
- Bačíková D, Horowitz DS (2005) Genetic and functional interaction of evolutionarily conserved regions of the Prp18 protein and the U5 snRNA. *Mol Cell Biol* 25:2107–2116.
- Bačíková D, Horowitz DS (2002) Mutational analysis identifies two separable roles of the *Saccharomyces cerevisiae* splicing factor Prp18. *RNA* 8:1280–1293.

Plasmids. Plasmids bearing *act1-CUP1* genes with two 3' splice sites were made in pGAC14 (25) by replacing the ClaI-KpnI fragment that spans the region with inserts made from oligonucleotides. *LSM7* and *REC102* were recovered by PCR and cloned into the single-copy vector pYX132 (Novagen) in which transcription is driven by the *TPI1* promoter (39). Two additional mutations were made in *LSM7*, A165G and A166C (numbered relative to the first coding base in the gene), and in *REC102*, A301C and A302C. These mutations are translationally silent and did not detectably affect splicing of the wild-type substrates. These bases were the 3' termini of the primers to prevent the wild-type mRNAs from interfering in the extension. Mutations in the *LSM7* and *REC102* were made by QuikChange Mutagenesis (Stratagene).

Splicing Assays. Protocols for measuring copper-resistance and for primer extension, using the *ACT1-CUP1* and U14 primers from Siatecka et al. (40), were described in ref. 20. *LSM7* and *REC102* were tested in DB129 and in DB127 transformed with pRS315-*prp18 Δ CR*. The *LSM7* results were similar in wild-type and Δ *UPF1* yeast, although the +8 mRNA (Fig. 4A) is not translatable. For *REC102*, there was an increase in the relative amounts of the +1, +19, and +25 mRNAs (Fig. 5A) compared with the in-frame +30 mRNA in the Δ *UPF1* yeast, and we used Δ *UPF1* yeast for analysis of *LSM7* and *REC102* splicing. Primer for *LSM7* complemented bases 165–191, and, for *REC102*, bases 301–322. *REC102* cDNAs were treated with RNase A and RNase H, annealed to oligonucleotide complementary to bases 105–127, and digested with HpaII before electrophoresis (Fig. 5S). Lariat intermediates were identified by treating the RNA from one mutant with Dbr1 protein (from B. Schwer, Weill Cornell Medical College, New York) (41) before primer extension.

For in vitro splicing, *act1* genes were cloned into pDKSa2, and splicing was assayed at 32 °C, as described in ref. 20. We used |AAA as the strong downstream site in vitro because |ACA affected the first step of splicing. For Fig. 3 B and C, aliquots from single reactions were withdrawn at the times indicated at the top of B.

ACKNOWLEDGMENTS. We thank Beate Schwer for helpful comments during this work, Michelle Hastings for comments on the manuscript, and Dagmar Bačíková (Uniformed Services University of the Health Sciences, Bethesda) for construction of *UPF1* yeast strains. This work was supported by National Institutes of Health Grant GM57267 (to D.S.H.), and by Uniformed Services University of the Health Sciences Grant C017HO (to D.S.H.).

- Aronova A, Bačíková D, Crotti LB, Horowitz DS, Schwer B (2007) Functional interactions between Prp8, Prp18, Slu7, and U5 snRNA during the second step of pre-mRNA splicing. *RNA* 13:1437–1444.
- Teigelkamp S, Newman AJ, Beggs JD (1995) Extensive interactions of PRP8 protein with the 5' and 3' splice sites during splicing suggest a role in stabilization of exon alignment by U5 snRNA. *EMBO J* 14:2602–2612.
- Lesser CF, Guthrie C (1993) Mutational analysis of pre-mRNA splicing in *Saccharomyces cerevisiae* using a sensitive new reporter gene, *CUP1*. *Genetics* 133:851–863.
- Zhang X, Schwer B (1997) Functional and physical interaction between the yeast splicing factors Slu7 and Prp18. *Nucl Acids Res* 25:2146–2152.
- Horowitz DS, Abelson J (1993) A U5 small nuclear ribonucleoprotein particle protein involved only in the second step of splicing in *Saccharomyces cerevisiae*. *Mol Cell Biol* 13:2959–2970.
- Miura F, et al. (2006) A large-scale full-length cDNA analysis to explore the budding yeast transcriptome. *Proc Natl Acad Sci* 103:17846–17851.
- Juneau K, Palm C, Miranda M, Davis RW (2007) High-density yeast-tiling array reveals previously undiscovered introns and extensive regulation of meiotic splicing. *Proc Natl Acad Sci* 104:1522–1527.
- Maleki S, Neale MJ, Arora C, Henderson KA, Keeney S (2007) Interactions between Mei4, Rec114, and other proteins required for meiotic DNA double-strand break formation in *Saccharomyces cerevisiae*. *Chromosoma* 116:471–486.
- Horowitz DS, Krainer AR (1997) A human protein required for the second step of pre-mRNA splicing is functionally related to a yeast splicing factor. *Genes Dev* 11:139–151.
- Lallena MJ, Chalmers KJ, Llamazares S, Lamond AI, Valcárcel J (2002) Splicing regulation at the second catalytic step by Sex-lethal involves 3' splice site recognition by SPF45. *Cell* 109:285–296.
- Hiller M, Platzer M (2008) Widespread and subtle: Alternative splicing at short-distance tandem sites. *Trends Genet* 24:246–255.
- Chua K, Reed R (1999) The RNA splicing factor hSlu7 is required for correct 3' splice-site choice. *Nature* 402:207–210.
- Crick FHC (1968) The origin of the genetic code. *J Mol Biol* 38:367–379.
- Santos MA, Moura G, Massey SE, Tuite MF (2004) Driving change: The evolution of alternative genetic codes. *Trends Genet* 20:95–102.
- Rodriguez-Trelles F, Tarrío R, Ayala FJ (2006) Origins and evolution of spliceosomal introns. *Ann Rev Genet* 40:47–76.
- Roy SW, Gilbert W (2006) The evolution of spliceosomal introns: Patterns, puzzles and progress. *Nat Rev Gen* 7:211–221.
- Scott EW, Allison HE, Baker HV (1990) Characterization of TPI gene expression in isogenic wild-type and *gcr1*-deletion mutant strains of *Saccharomyces cerevisiae*. *Nucl Acids Res* 18:7099–7107.
- Siatecka M, Reyes JL, Konarska MM (1999) Functional interactions of Prp8 with both splice sites at the spliceosomal catalytic center. *Genes Dev* 13:1983–1993.
- Khalid MF, Damha MJ, Shuman S, Schwer B (2005) Structure-function analysis of yeast RNA debranching enzyme (Dbr1), a manganese-dependent phosphodiesterase. *Nucl Acids Res* 33:6349–6360.