MOLECULAR AND CELLULAR BIOLOGY, May 1994, p. 3434–3445
0270-7306/94/$04.00+0

Vol. 14, No. 5

Intron Definition in Splicing of Small *Drosophila* Introns

MELISSA TALERICO AND SUSAN M. BERGET*

Verna and Marrs McClean Department of Biochemistry, Baylor College of Medicine, Houston, Texas 77030

Received 1 October 1993/Returned for modification 10 November 1993/Accepted 8 February 1994

Approximately half of the introns in *Drosophila melanogaster* are too small to function in a vertebrate and often lack the pyrimidine tract associated with vertebrate 3′ splice sites. Here, we report the splicing and spliceosome assembly properties of two such introns: one with a pyrimidine-poor 3′ splice site and one with a pyrimidine-rich 3′ splice site. The pyrimidine-poor intron was absolutely dependent on its small size for in vivo and in vitro splicing and assembly. As such, it had properties reminiscent of those of yeast introns. The pyrimidine-rich intron had properties intermediate between those of yeasts and vertebrates. This 3′ splice site directed assembly of ATP-dependent complexes when present as either an intron or exon and supported low levels of in vivo splicing of a moderate-length intron. We propose that splice sites can be recognized as pairs across either exons or introns, depending on which distance is shorter, and that a pyrimidine-rich region upstream of the 3′ splice site facilitates the exon mode.

Pairing of individual splice sites during early recognition of pre-mRNAs is an attractive model for the orchestration of splicing in genes with multiple introns. Indeed, in *Saccharomyces cerevisiae*, direct support exists for the pairing of splice sites across introns during the first step of spliceosome assembly (4, 8, 30, 31). In vertebrate pre-mRNAs, however, initial interaction between factors recognizing both ends of an intron is problematic because of the large size of many vertebrate introns. We have recently suggested that the exon is the unit of recognition during early spliceosome assembly in vertebrates with the concomitant pairing of splice sites across exons (29, 34). Interactive binding of factors across exons is feasible in vertebrates because internal exons are quite small, rarely exceeding 300 nucleotides (10).

Exon-intron architecture varies widely among eukaryotes. Although vertebrate exons are quite small and cannot be expanded without loss of recognition (29), many genes in lower eukaryotes have large exons (10). Most of these genes with large exons have small introns; however, some have large introns. If splice sites are paired, these genes have an awkward array of large and small distances between pairs of sites, making the mechanism of splice site recognition in these pre-mRNAs unclear. To address this problem, we have turned to the study of splice site recognition in *Drosophila melanogaster*. The majority of *Drosophila* exons are 100 to 180 nucleotides in length; however, 15% are more than 550 nucleotides, which is predicted to be too large to function in a vertebrate (10). This difference suggests that pairing of splice sites across exons will not work for recognition of these large exons.

A comparison of intron sizes between vertebrates and *Drosophila* species also reveals another obvious difference from vertebrates. More than half of *Drosophila* introns have lengths of fewer than 80 nucleotides, with most having lengths in the range of 59 to 67 nucleotides (24). These introns would be too small to function in a vertebrate (35). Thus, there are a number of *Drosophila* genes that have large exons and small introns, which is an inverted exon-intron architecture compared with that found in vertebrate genes. The sequence of *Drosophila* introns also differs from that of vertebrate introns. First, *Drosophila* introns tend not to have a G in the position preceding the branched nucleotide (24). G is the most common nucleotide at this position in mammalian introns. Second, the pyrimidine-rich regions important for vertebrate splicing are frequently missing from the 3′ end of *Drosophila* introns, especially in small introns (24). In addition, *Drosophila* introns have a 17% higher A+T content than their flanking exons; vertebrates do not share this disparity (5, 24).

One of the strongest pieces of evidence supporting the pairing of splice sites across exons in vertebrates is the phenotype of 5′ splice site mutations. Such mutations inactivate the intron in which they reside in all organisms. When local cryptic sites are absent, however, 5′ splice site mutations also suppress the splicing of the proximal upstream intron in vertebrates, resulting in a product RNA missing the exon bordered by the mutation (34). A variety of such mutations in vertebrate genes have now been analyzed (for a review, see references 18 and 34). To date, however, no mutation of this type has been analyzed for RNA phenotype in a gene from a lower eukaryote containing large exons and multiple introns. If the intron is the unit of splice site pairing (or if splice sites are recognized independently), the predicted phenotype is intron inclusion. If the exon is the unit of recognition, the predicted phenotype is exon skipping.

In this report, we present three phenotypes that differentiate the splicing of a typical vertebrate pre-mRNA with small exons from that of a *Drosophila* pre-mRNA with small introns. First, introduction of a 5′ splice site mutation into the second intron of the natural *Drosophila* zeste gene promoted inclusion of the mutated intron rather than exon skipping, when analyzed in vivo in Schneider 2 (S2) cells. Second, expanding the size of two small *Drosophila* introns inhibited wild-type splicing of these introns in vitro and promoted cryptic splice site utilization in vivo, such as to splice small versus large introns. Third, through the development of an in vitro *Drosophila* splicing system, we show that assembly of an initial ATP-dependent splicing complex required both ends of a *Drosophila* intron. Together, these observations suggest mechanistic differences in the process of splice site selection in pre-mRNAs with small introns versus those with small exons. We suggest that splice sites are initially recognized as pairs across the shortest distance between them: either the intron or the exon.

* Corresponding author. Phone: (713) 798-5758. Fax: (713) 796-9438.

MATERIALS AND METHODS

Splicing substrates. The adenovirus substrates used in this study contained information from the human adenovirus late transcription unit based on the MINX transcript described previously by this laboratory (29). adeno 62 (see Fig. 4A and 7B) was a derivative of the wild-type adenovirus duplication construct previously reported. DNA was truncated in the second intron with *Bgl*II to create a two-exon substrate. The 120-nucleotide first intron of the duplication construct was partially deleted by *Bal* 31 exonuclease digestion to produce a 62-nucleotide intron. The deletion was from +8 to +65 of the intron. adeno-X (see Fig. 7B) was also derived from the adenovirus duplication construct. The 323-nucleotide *Bam*HI fragment from the third intron of the human calcitonin/calcitonin gene-related peptide (CGRP) gene was inserted at the *Hin*dIII site in intron 1 to create an expanded intron of 440 nucleotides. To synthesize RNA, substrates were truncated at the *Bgl*II site in the second intron. Thus, adeno 62 and adeno-X were identical substrates, except for the calcitonin/CGRP sequences used to expand the intron.

The zeste in vitro substrates were derived from the natural *Drosophila* zeste gene (kindly provided by V. Pirrotta). zeste 62 (see Fig. 4A, 5B, 6, and 7A) and zeste 62-5′MT (see Fig. 4A) contained the second 62-nucleotide intron of zeste with flanking exon sequences. Twenty-five nucleotides of SP65 polylinker were fused to the last 48 nucleotides of zeste exon 2 to create the first exon of this substrate. The second exon in this construct was composed of the first 62 nucleotides of zeste exon 3. Substrates were truncated at the *Bam*HI site 7 nucleotides downstream of zeste exon 3 sequences. The 5′ splice site mutation is identical to that in the in vivo construct, zeste-5′MT, which is described below.

zeste 62Δ5 (see Fig. 5B) contained the isolated 3′ splice site of the zeste second intron. The construct was made by deleting all of the exon 1 sequences of zeste 62 and the first 7 nucleotides (including the 5′ splice site) of intron 2 by oligonucleotide-mediated PCR. Both zeste 62 and zeste 62Δ5 had identical sequences with the exception of this deletion. zeste-X (see Fig. 7A) was derived from zeste 62 as well and contained identical sequence information, except for the intronic insertion in zeste-X. A 300-nucleotide fragment from intron 7 of the *Drosophila* troponin T gene was obtained by PCR and sequenced. *Drosophila* troponin T was kindly provided by E. Fyrberg. This fragment was cloned into the zeste 62 intron at nucleotide +7 of the intron. The insertion was downstream of the 5′ splice site and upstream of the branch point, pyrimidine tract, and 3′ splice site.

The substrate ZPE (see Fig. 6) was constructed by fusing zeste62Δ5 at the *Bam*HI site downstream of the zeste exon sequences to the first 48 nucleotides of natural zeste exon 2 followed by the first 7 nucleotides of intron 2 containing the 5′ splice site. Twenty-five nucleotides of SP65 polylinker were inserted by the cloning at the fusion site. This construction created a pseudoexon by fusing the 3′ splice site to a downstream 5′ splice site with exon sequences between. RNA was synthesized by truncating 12 nucleotides downstream of the 5′ splice site with *Nco*I.

The substrate mle 59 (see Fig. 4A, 5A, 6, and 7A) contained the first 59-nucleotide intron from the *Drosophila* maleless gene, kindly provided by M. Kuroda, with the last 53 nucleotides of exon 1 and the first 95 nucleotides of exon 2. Thirty-two nucleotides of SP65 polylinker were fused upstream of the maleless exon 1 sequences. RNA was made by truncating at the *Hin*dIII site 5 nucleotides downstream of the exon 2 sequences. mle 59Δ5 (see Fig. 5A) was created by deleting all of the exon 1 sequences and the first 8 nucleotides of the intron using oligonucleotide-mediated PCR. This deleted the 5′ splice site and left intact the 3′ splice site region of the intron.

mle-X (see Fig. 7A) was constructed by inserting the same troponin T intron fragment used in zeste-X in the middle of the mle 59 intron. The fragment was inserted at nucleotide +33 of the intron. Other than the insertion, mle 59 and mle-X had identical sequence information. The pseudoexon substrate MPE (see Fig. 6) contained mle 59Δ5 sequences fused at the *Hin*dIII site downstream of exon 2 to the last 53 nucleotides of mle exon 1 and the first 27 nucleotides of intron 1. This created a structure similar to that of an exon with a 3′ splice site fused to a 5′ splice site downstream. Eleven nucleotides of polylinker were inserted by the cloning at the fusion site.

In vivo expression constructs. DNA plasmids containing the *Drosophila* zeste gene with its natural polyadenylation site were constructed as diagrammed in Fig. 1 (zeste-WT and zeste-5′ MT). The last 470 nucleotides of zeste exon 1 were fused to the 300-nucleotide *Eco*RI-*Bam*HI fragment from the *Drosophila* actin 5C distal promoter. This fragment contained the first 88 nucleotides of actin exon 1. To create zeste-5′MT, a mutation was introduced at the 5′ splice site beginning intron 2 by PCR-mediated cloning with oligonucleotides altered for 2 bases at this site; the canonical GT was changed to CG.

The following four expression constructs were derived from the natural zeste and maleless genes by utilizing fragments from previously cloned splicing substrates (described above). MXMG and ZXMG (see Fig. 2) were constructed by cloning mle-X and zeste-X into a *Drosophila* in vivo expression vector. mle-X and zeste-X were isolated for cloning by performing PCR with oligonucleotides hybridizing to the 5′ and 3′ ends of the substrates. Both mle-X and zeste-X were fused to the *Eco*RI-*Bam*HI fragment of the *Drosophila* actin 5c distal promoter at nucleotide +88. Downstream, mle-X and zeste-X were fused to the last 1,445 nucleotides of the natural zeste gene (*Nae*I-*Hin*dIII fragment) containing zeste exon 3 sequences and the zeste polyadenylation signals.

M5Z3 and Z5M3 (see Fig. 3) were constructed by switching the 3′ splice sites with flanking exon sequences between MXMG and ZXMG. M5Z3 derived its 5′ half from MXMG and its 3′ half from ZXMG. In ZXMG, there is a unique *Bgl*II site 60 nucleotides upstream of the zeste 3′ splice site. The *Bgl*II-*Hin*dIII fragment from ZXMG containing the zeste 3′ splice site and polyadenylation site replaced the corresponding *Bgl*II-*Hin*dIII fragment in MXMG, which contained the maleless 3′ splice site and zeste polyadenylation site. Thus, M5Z3 had the maleless 5′ splice site and the zeste 3′ splices site bordering an expanded internal intron. Z5M3 was constructed as the reverse of M5Z3. The unique *Bgl*II site in MXIV was 38 nucleotides upstream of the maleless 3′ splice site. The *Bgl*II-*Hin*dIII fragment from MXMG replaced the corresponding fragment in ZXMG to create Z5M3. This construct contained the zeste 5′ splice site and the maleless 3′ splice site bordering an expanded internal intron.

All splicing substrates and in vivo expression constructs cloned by PCR methodology were completely sequenced to confirm that no mutations were introduced during the construction process.

In vivo transfections and reverse transcription (RT)-PCR. Twenty micrograms of each DNA expression plasmid was transfected into *Drosophila* S2 cells with Lipofectin reagent (Gibco BRL) according to the manufacturer's protocol. S2 cells were kindly provided by V. Pirrotta. Total RNA from cells was harvested after 48 h with RNazol B (Biotecx Laboratories, Inc.). A 20-μg amount of total RNA was DNase treated in 50-μl reaction mixtures containing 40 mM Tris-HCl (pH 7.9),

6 mM $MgCl_2$, 2 mM spermidine-$(HCl)_3$, 15 mM dithiothreitol (DTT), 60 U of RNasin RNase inhibitor (Promega), and 2 U of RQ1 RNase-free DNase (Promega). The reaction was performed at 37°C for 20 min and was followed by phenol-chloroform extraction and ethanol precipitation.

The RT-PCR protocol was performed as follows. For annealing of the downstream oligonucleotide primer, an 18.8-μl reaction mixture containing 5 to 10 μg of DNase-treated total RNA, 1 μl of 150 μM 3′ oligonucleotide, and 1.06 mM EDTA was heated for 5 min at 95°C to denature the RNA. After the addition of 1.2 μl of 3M KCl (final concentration, 180 mM), annealing occurred at 37°C for 20 min. To create first-strand cDNAs, RT was performed in a 50-μl reaction mixture containing the annealing reaction mixture (5 to 10 μg of RNA, 1 μl of 150 μM 3′ oligonucleotide, 1 mM EDTA, 180 mM KCl), 40 U of RNasin, 50 mM Tris-HCl (pH 7.5), 10 mM DTT, 3 mM $MgCl_2$, 0.5 mM each dATP, dCTP, dGTP and dTTP, 8.0 μM actinomycin D (Calbiochem), and 200 U of Moloney murine leukemia virus reverse transcriptase (Gibco BRL). The reaction mixture was incubated at 37°C for 1 h. This RT protocol was used to obtain the data presented in Fig. 1. The cDNAs used for Fig. 2 and 3 were obtained by a slightly different protocol using avian myeloblastosis virus reverse transcriptase (Promega).

For these reactions, the KCl was omitted from the annealing reaction mixture and was replaced with water. RT was performed in a 50-μl reaction mixture containing the annealing reaction mixture (same as that described above, without the KCl), 40 U of RNasin, the Promega manufacturer's reaction buffer (50 mM Tris-HCl [pH 8.3], 10 mM DTT, 10 mM $MgCl_2$, 50 mM KCl, 0.5 mM spermidine), 0.5 mM each of dATP, dCTP, dGTP, and dTTP, 8.0 μM actinomycin D (Calbiochem), and 10 U of avian myeloblastosis virus reverse transcriptase (Promega). The reaction mixture was incubated at 53°C for 1 h.

The RNA template was then degraded in the presence of 5 mM EDTA and 40 mM NaOH for 10 min at 100°C. The reaction was neutralized with 40 mM HCl, and the cDNA was precipitated by the addition of 0.27 M NaOAC and 2.5 volumes of ethanol. The final pellet was resuspended in 50 μl of Tris-EDTA.

PCR was performed in multiple ways to increase the specificity of the reaction. For the amplification products shown in Fig. 1, PCR amplification was performed in a 100-μl reaction mixture containing 5 μl of cDNA, 0.25 mM each dATP, dCTP, dGTP, and dTTP, 1 μl of 150 μM 5′ oligonucleotide, 1 μl of 150 μM 3′ oligonucleotide 1.5 mM $MgCl_2$, 50 mM KCl, 10 mM Tris-HCl (pH 8.8), 0.1% Triton X-100, and 2.5 U of *Taq* polymerase. Cycling conditions were 94°C for 5 min, 94°C for 1 min, 55°C for 2 min, and 72°C for 2 min, for 30 cycles. In Fig. 1, the 5′ oligonucleotide hybridized to the actin exon sequences downstream of the start site. The 3′ oligonucleotide hybridized to sequences in zeste exon 3. RT-PCR products were displayed on 10% native acrylamide gels and then were silver stained.

For the amplification of cDNAs from substrates in Fig. 2 and 3, PCR was performed by the hot-start technique. This technique used the Ampliwax PCR Gem 100 (Perkin-Elmer Cetus). The hot-start PCRs were performed according to the manufacturer's protocol with the following variations. The lower reagent mix contained, in a 25-μl volume, 0.25 mM each dATP, dCTP, dGTP, and dTTP, 1 μl of 150 μM 5′ oligonucleotide, 1 μl of 150 μM 3′ oligonucleotide, 2 mM $MgCl_2$, and 1× PCR buffer (50 mM KCl, 10 mM Tris-HCl [pH 8.8], 0.1% Triton X-100). The concentrations given above are final concentrations after combination with the upper reagent mix in a 100-μl final reaction volume. The upper reagent mix contained, in a 75-μl volume, 1× PCR buffer, 2.5 U of *Taq* polymerase, 1 U of Perfect Match DNA polymerase enhancer (Stratagene), and 5 μl of cDNA. Cycling conditions were 94°C for 5 min, 94°C for 1 min, 60°C for 1 min, 72°C for 2 min, for 30 cycles. In Fig. 2 and 3, the 5′ and 3′ oligonucleotides that were used in Fig. 1 were used.

Nuclear extracts and in vitro splicing assays. *Drosophila* S2 cell nuclear extracts were prepared according to procedures described by Dignam et al. (6). RNA precursors were synthesized from plasmid templates with SP6 polymerase and [^{32}P]UTP. *Drosophila* splicing reactions and ribonucleoprotein assays were performed in 25-μl volumes containing 20 mM creatine phosphate or phosphoarginine, 2 mM ATP, 1.6 mM $MgCl_2$, 1.2 mM DTT, 0.6 to 1.5% polyethylene glycol, 80 μg of S2 extract, 20 U of RNasin (Promega), and 2×10^4 to 4×10^4 cpm of RNA. The reaction mixtures were incubated at 23°C. In vitro reaction products were displayed on 5% 7 M urea gels and were then autoradiographed. Reactions were quantitated by scanning the gels in a Betagen Betascope 603 blot analyzer. Counts per minute in each band were converted to femtomoles of product from knowledge of the specific activity of the UTP label, the number of uridines per RNA, and the efficiency of the instrument.

Nucleotide sequence accession number. The sequence described in this paper has been deposited in the GenBank data base under accession number L25807.

RESULTS

Mutation of a 5′ splice site in *Drosophila* zeste promotes intron inclusion in vivo. In vertebrates, mutation of the 5′ splice site terminating an internal exon results in either exon skipping or activation of a nearby cryptic site (1, 2, 34). To study the effect of alteration of a 5′ splice site in *D. melanogaster*, the second intron of the *Drosophila* zeste gene was mutated. zeste is a three-exon gene containing two small introns (121 and 62 nucleotides) and a large internal exon of 542 nucleotides. The 5′ splice site of intron 2 was altered from CCG/**GU**GAGA to CCG/**CG**GAGA to create the zeste mutant construct zeste-5′MT. Both wild-type and mutant zeste genes were transiently transfected into *Drosophila* S2 cells to determine their in vivo RNA splicing phenotypes. Total cellular RNA was analyzed by RT-PCR (Fig. 1).

Amplification of RNA produced from the wild-type gene yielded a single band with a length (1,115 nucleotides) consistent with correct splicing of both introns. Amplification of mutant RNA produced a single larger band, indicating that mutation altered the splicing phenotype. The produced mutant band had a length intermediate between that produced by amplification of wild-type RNA and the length of the band produced by direct PCR amplification of zeste genomic DNA (1,298 nucleotides). The intermediate length was consistent with an RNA containing one, but not both, introns. Since both introns are similar in length, it was not possible to determine by size analysis alone which of the two introns had been removed. Direct sequencing of the mutant band demonstrated that intron 1 had been correctly removed; however, intron 2 was completely retained in the mutant RNA. A product band of 543 nucleotides, indicative of exon skipping (splicing of exon 1 to exon 3), was never observed for either the wild-type or the mutant construct. Therefore, the in vivo phenotype of a 5′ splice site mutation of the internal exon in *Drosophila* zeste is intron inclusion and not exon skipping, the predominant phenotype for similar mutations in vertebrates (18, 34). These

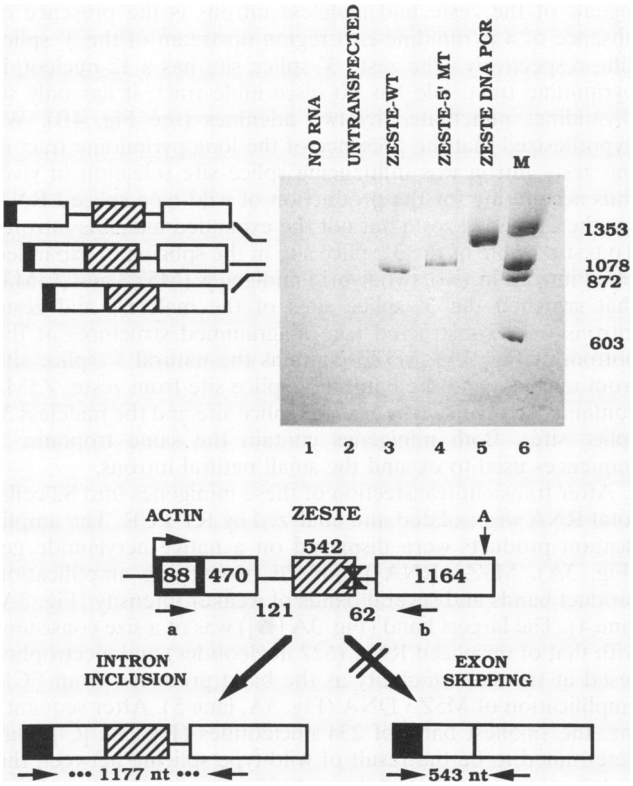

FIG. 1. Mutation of a 5′ splice site in *Drosophila* zeste promotes intron inclusion in vivo. Minigenes of the diagrammed structures were transiently transfected into S2 cells. The natural *D. melanogaster* zeste gene was fused at nucleotide +224 to the *D. melanogaster* actin promoter at nucleotide +88. The wild-type minigene (zeste-5′ WT) was altered to mutate nucleotides +1 and +2 of the 5′ splice site beginning at intron 2 to create the mutant zeste minigene (zeste-5′ MT). Total cell RNA prepared at 48 h posttransfection was analyzed for splicing phenotype by RT-PCR amplification with the diagrammed oligonucleotide primers (arrows a and b). Primer a was complementary to actin sequences derived from the promoter; primer b was complementary to sequences within zeste exon 3. Two possible processing pathways are indicated for the mutant minigene. The sizes of the predicted RT-PCR product for each are indicated. Amplification products were displayed on a native 10% polyacrylamide gel which was subsequently silver stained. The structures of product bands are depicted to the left of the gel. The identities of individual bands were confirmed by diagnostic restriction nuclease digestion and direct sequencing. Lanes: 1, no RNA; 2, RNA from untransfected cells; 3, RNA from cells transfected with the wild-type minigene (zeste-WT); 4, RNA from cells transfected with the mutant zeste minigene (zeste-5′MT); 5, PCR amplification of zeste DNA (marker for unspliced pre-mRNA); 6, *Hae*III-digested ϕX174 replicative form DNA size marker.

data are consistent with an initial pairing of splice sites across the small intron.

Expansion of small introns produces aberrant splicing phenotypes in vivo. If splice sites are selected in pairs across small *Drosophila* introns, then the small size of such introns may be important for their recognition. Thus, expanding these introns might alter the efficiency with which they are recognized. Indeed, in vitro intron size limitations have been observed previously for small *Drosophila* and yeast introns (9, 15). To test this hypothesis in vivo, two small introns from the maleless and zeste genes were expanded by inserting a 300-nucleotide fragment from intron 7 of the *Drosophila* troponin T gene into the middle of each intron. Expansion of the introns increased the sizes of the natural small introns from 59 (maleless) and 62 (zeste) nucleotides to 371 and 373 nucleotides, respectively. Minigenes containing the artificially expanded mle and zeste introns were constructed by fusing these expanded introns to the *Drosophila* actin promoter and the zeste polyadenylation signals to produce MXMG and ZXMG, respectively. Both minigenes were transiently transfected into *Drosophila* S2 cells to determine their in vivo RNA splicing phenotypes. Total cellular RNA was analyzed by RT-PCR (Fig. 2A).

Amplification of the maleless minigene, MXMG, produced two bands (Fig. 2A, lane 4). The larger band (Fig. 2A [●]) had a size consistent with unspliced RNA (640 nucleotides) and electrophoresed with the same mobility as the band produced from PCR amplification of MXMG DNA (Fig. 2A, lane 5). The slightly smaller amplification product band (Fig. 2A [△]) was not of a size predicted for normal splicing. Direct sequencing of this band revealed that the natural mle 5′ splice site was spliced to a cryptic 3′ splice site located 112 nucleotides downstream within the expanded intron (Fig. 2B). No product band of 269 nucleotides, which was indicative of wild-type splicing, was detected. Thus, the preferred in vivo splicing phenotype for the expanded mle intron was activation of a cryptic 3′ splice site within the intron to splice an artificially shortened intron.

Amplification of ZXMG RNA produced two bands of 254 (Fig. 2A [○]) and 310 (Fig. 2A [+]) nucleotides (Fig. 2A, lane 6). The 310-nucleotide amplification band was not of a size predicted for normal splicing, suggesting that it was the product of an aberrant splicing event. After direct sequencing, this band was determined to be the result of activating two cryptic splice sites within the interior of the expanded intron to create a 56-nucleotide pseudoexon (see the structure at the bottom of Fig. 2C). The wild-type zeste 5′ splice site was spliced to a cryptic 3′ splice site 92 nucleotides downstream within the intron, and the wild-type zeste 3′ splice site was spliced to a cryptic 5′ splice site located 225 nucleotides upstream within the intron. Thus, two separate splicing events created a product RNA containing three exons. The smallest product band (Fig. 2A [+]) was of a size (254 nucleotides) consistent with the correct removal of the intron. Direct sequencing of this band confirmed the accurate removal of the intron by using wild-type splice sites (see the structure at the top of Fig. 2C). Thus, expansion of the zeste intron resulted in two phenotypes: (i) double cryptic activation and (ii) wild-type splicing.

Both *Drosophila* minigenes containing artificially expanded introns produced aberrantly spliced RNAs by utilizing cryptic splice sites within the intron. The introns removed by cryptic activation had a size smaller than that of the expanded intron, suggesting that the splicing machinery preferred the removal of small versus large introns, even at the expense of ignoring the wild-type splicing signals in the case of the expanded mle intron. In addition to aberrant splicing, the zeste minigene spliced its expanded intron appropriately by using normal splice sites. It is impossible to determine the relative ratios of these different splicing events, because the conditions for PCR amplification were not quantitative. However, it is apparent that no wild-type splicing was detected for the mle minigene, suggesting an inherent difference between the mle and zeste splicing signals that resulted in different splicing patterns.

A pyrimidine-rich 3′ splice site can overcome aberrant splicing phenotypes produced by the expansion of small introns. The most obvious difference between the splicing

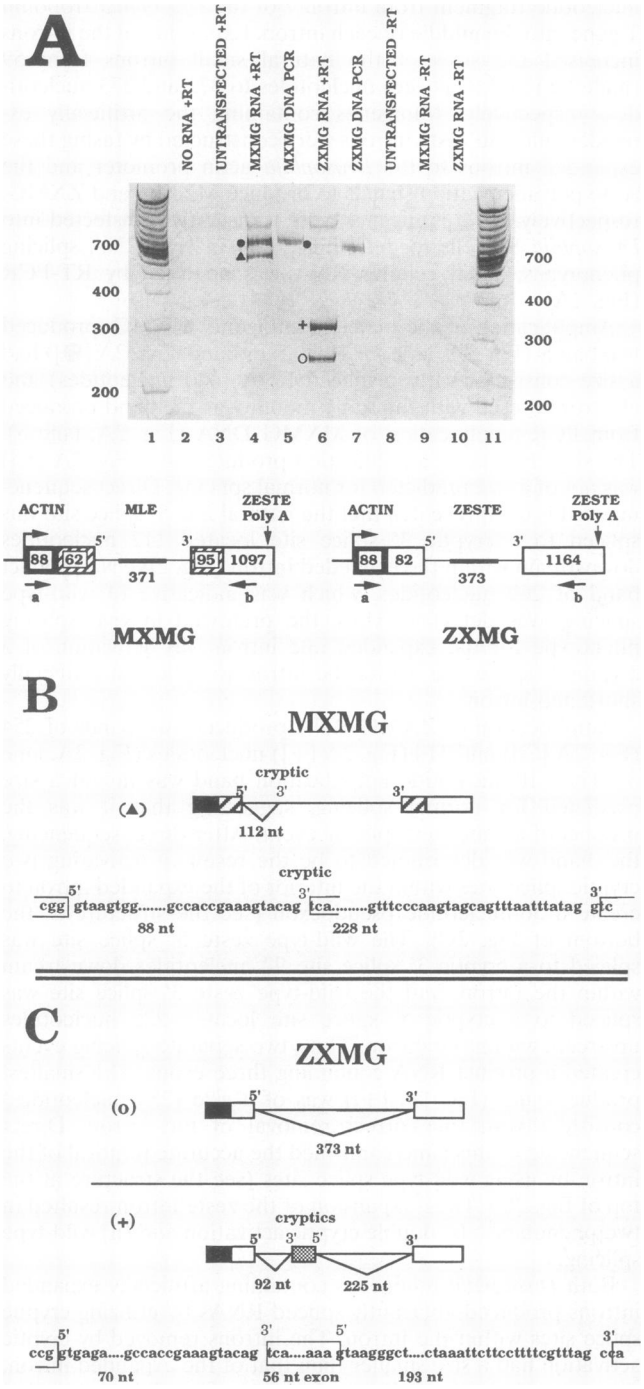

FIG. 2. Expansion of small introns produces aberrant splicing phenotypes in vivo. (A) Minigenes of the structures diagrammed contained the artificially expanded maleless and zeste introns. These sequences were fused to the *Drosophila* actin promoter followed by the first 88 nucleotides of actin exon 1 indicated by black boxes. In MXMG, the hatched boxes represent maleless exon sequences flanking the maleless expanded intron. The maleless 3′ splice site and downstream exon sequences were fused to the last 1,097 nucleotides of zeste exon 3 containing the zeste polyadenylation signals. zeste sequences are represented by white boxes. ZXMG contained the last 82 nucleotides of zeste exon 2 fused to the actin promoter and the complete sequences of zeste exon 3 flanking the zeste expanded intron. The thin black boxes in the introns represent *Drosophila* troponin T sequences used for expansion. The minigenes were transiently transfected into *Drosophila* S2 cells. Total cellular RNA prepared at 48 h posttransfection was analyzed for splicing phenotype by RT-PCR amplification with the diagrammed oligonucleotide primers (arrows a and b). Primer a was complementary to actin sequences derived from the promoter; primer b was complementary to sequences within zeste exon 3. Amplification products were displayed on a native 10% polyacrylamide gel which was subsequently silver stained. The identities of individual bands were confirmed by direct sequencing. Amplification products were obtained by performing RT-PCR in the presence (lanes 2, 3, 4, and 6) or absence (lanes 8, 9, and 10) of reverse transcriptase. Thus, the product bands in lanes 4 and 6 were derived from RNA because in the absence of reverse transcriptase (lanes 8 to 10) these products were not detected. ● in lane 4 and 6, unspliced RNA. (B) The splicing event to produce the band in lane 4 of panel A marked by △, along with the sequences of the 3′ cryptic splice site utilized and the natural maleless 5′ and 3′ splice sites in MXMG. (C) The splicing events responsible for the production of the bands in lane 6 of panel A marked by ○ and +. The sequences of the wild-type and cryptic splice sites utilized in the splicing reactions for ZXMG are shown. Lanes: 1, 100-bp DNA ladder (Gibco BRL); 2, no RNA; 3, RNA from untransfected cells; 4, RNA from cells transfected with the MXMG minigene; 5, PCR amplification of MXMG DNA (marker for unspliced pre-mRNA); 6, RNA from cells transfected with the ZXMG minigene; 7, PCR amplification of ZXMG DNA (marker for unspliced pre-mRNA); 8, RNA from untransfected cells; 9, RNA from cells transfected with the MXMG minigene; 10, RNA from cells transfected with the ZXMG minigene; 11, 100-bp DNA ladder (Gibco BRL).

signals of the zeste and maleless introns is the presence or absence of a pyrimidine-rich region upstream of the 3′ splice site, respectively. The zeste 3′ splice site has a 12-nucleotide pyrimidine tract. mle has no discernible tract; it has only six thymidines punctuated by two adenines (see Fig. 4B). We hypothesized that the presence of the long pyrimidine tract in the zeste intron was influencing splice site selection in vivo, thus accounting for the production of wild-type spliced RNA for the expanded zeste but not the expanded maleless introns. To test the role of the 3′ splice site in the splicing of expanded small introns in vivo, two hybrid minigenes (M5Z3 and Z5M3) that switched the 3′ splice sites of the maleless and zeste introns were constructed (see diagrammed structures at the bottom of Fig. 3A). M5Z3 contains the natural 5′ splice site from maleless and the natural 3′ splice site from zeste. Z5M3 contains the reverse (the zeste 5′ splice site and the maleless 3′ splice site). Both minigenes contain the same troponin T sequences used to expand the small natural introns.

After transient transfection of these minigenes into S2 cells, total RNA was isolated and analyzed by RT-PCR. The amplification products were displayed on a native acrylamide gel (Fig. 3A). M5Z3 RNA produced two strong amplification product bands and several bands of weaker intensity (Fig. 3A, lane 4). The largest band (Fig. 3A [●]) was of a size consistent with that of unspliced RNA (627 nucleotides) and electrophoresed at the same mobility as the band produced from PCR amplification of M5Z3 DNA (Fig. 3A, lane 5). After sequencing, the smallest band of 234 nucleotides (Fig. 3A [○]) was determined to be the result of wild-type splicing between the maleless 5′ splice site and the zeste 3′ splice site (see top of Fig. 3B). The slightly larger band (290 nucleotides) of weaker intensity (Fig. 3A [+]) was also sequenced. This product band was the result of activation of the same two internal cryptic splice sites used in the aberrant splicing event of the expanded zeste minigene (see the bottom of Fig. 2C). Concomitantly, this M5Z3 RNA product contains three exons due to the removal of two introns (see the bottom of Fig. 3B).

RT-PCR performed on RNA from the converse construct, Z5M3, produced two amplification products (Fig. 3A, lane 6).

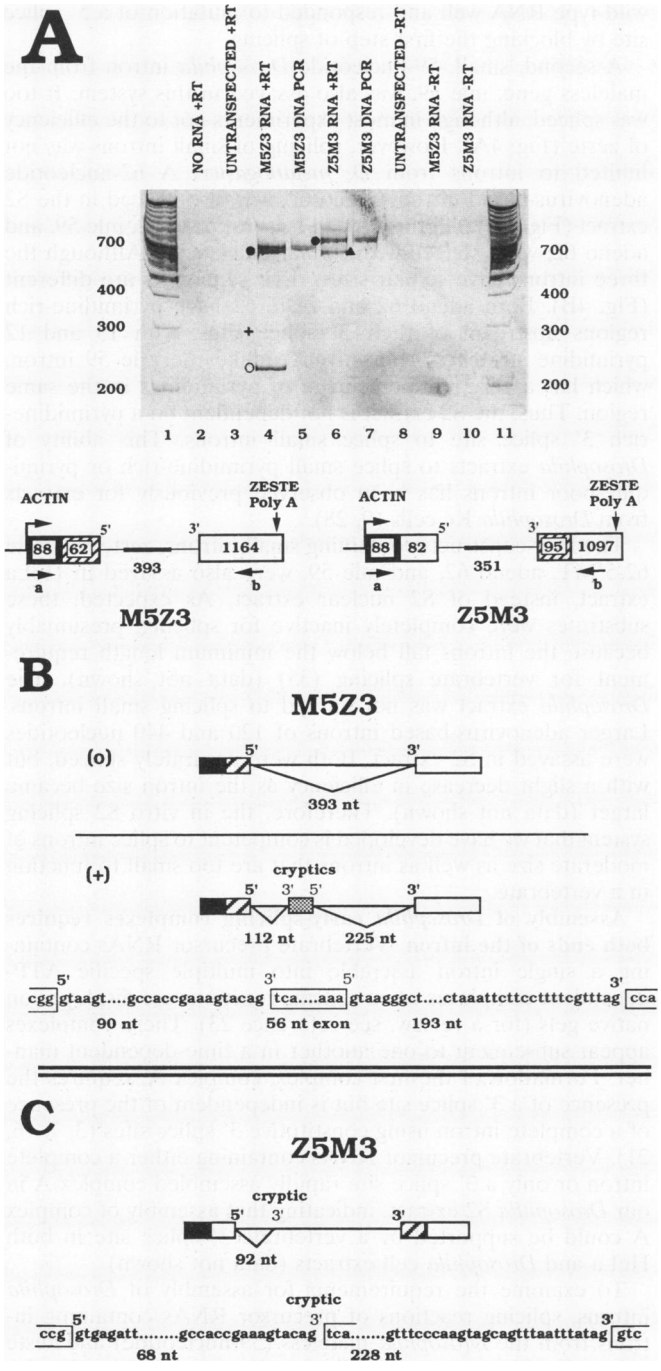

FIG. 3. A pyrimidine-rich 3′ splice site is required for wild-type splicing of an expanded small intron. (A) Two hybrid minigenes were constructed to examine the effect of the pyrimidine content of the 3′ splice site on splicing of an expanded intron in vivo. M5Z3 contained the last 62 nucleotides of maleless exon 1 (hatched box) fused to the *Drosophila* actin promoter with the first 88 nucleotides of actin exon 1. These sequences and the entire last exon of zeste (white box), with its polyadenylation signals, flanked an expanded intron of 393 nucleotides containing troponin T sequences (thin black box). Thus, M5Z3 derived its 5′ splice site from maleless and its 3′ splice site from zeste. Z5M3 contained the last 82 nucleotides of zeste exon 2 fused to the *Drosophila* actin promoter with its exon 1 sequences. The first 95 nucleotides of maleless exon 2 were fused to the last 1,097 nucleotides of zeste exon 3. Thus, Z5M3 had the zeste 5′ splice site and the maleless 3′ splice site flanking an expanded 351-nucleotide intron. After transient transfection of these minigenes into S2 cells, total RNA was analyzed by RT-PCR amplification with the diagrammed oligonucleotide primers (arrows a and b) to determine splicing phenotypes. Primer a was complementary to actin sequences derived from the promoter; primer b was complementary to sequences within zeste exon 3. Amplification products were displayed on a native 10% polyacrylamide gel and then silver stained. The identities of individual bands were confirmed by direct sequencing. Amplification products were obtained by performing RT-PCR in the presence (lanes 2, 3, 4, and 6) or absence (lanes 8, 9, and 10) of reverse transcriptase. Thus, the product bands in lanes 4 and 6 were derived from RNA because in the absence of reverse transcriptase (lanes 8 to 10), these products were not detected. ●, unspliced RNA. (B) the splicing events responsible for the production of the bands in lane 4 of panel A marked by ○ and +. The sequences of the wild-type maleless 5′, zeste 3′, and cryptic splice sites utilized in the splicing reactions for M5Z3 are shown. (C) the splicing event producing the band in lane 6 of panel A marked by △ along with the sequences of the 3′ cryptic splice site utilized and the natural zeste 5′ and maleless 3′ splice sites in Z5M3. Lanes 1, 100-bp DNA ladder (Gibco BRL); 2, no RNA; 3, RNA from untransfected cells; 4, RNA from cells transfected with the M5Z3 minigene; 5, PCR amplification of M5Z3 DNA (marker for unspliced pre-mRNA); 6, RNA from cells transfected with the Z5M3 minigene; 7, PCR amplification of Z5M3 DNA (marker for unspliced pre-mRNA); 8, RNA from untransfected cells; 9, RNA from cells transfected with the M5Z3 minigene; 10, RNA from cells transfected with the Z5M3 minigene; 11, 100-bp DNA ladder (Gibco BRL).

The larger band (Fig. 3A [●]) was consistent with the predicted size for unspliced RNA (640 nucleotides) and electrophoresed at the same mobility as the band produced from PCR amplification of Z5M3 DNA (Fig. 3A, lane 7). Direct sequencing of the slightly smaller band (Fig. 3A [△]), which was not of a size predicted for wild-type splicing, demonstrated aberrant splicing. The zeste 5′ splice site was spliced to a cryptic 3′ splice site 92 nucleotides downstream in the intron (Fig. 3C). This is the same 3′ splice site that was utilized in the aberrant spliced product obtained from the expanded mle intron (Fig. 2B). No evidence for splicing of the natural zeste 5′ splice site to the maleless 3′ splice site (289 nucleotides) in the Z5M3 minigene was observed.

Thus, the splicing phenotype observed was dictated by the 3′ splice site. The relatively pyrimidine-poor mle 3′ splice site supported no splicing of a large intron. Neither the intact 350- to 400-nucleotide-expanded intron nor the second 203-nucleotide intron that could have been created by recognition of the embedded pseudoexon was spliced. The largest intron removed among introns containing the mle 3′ splice site was 112 nucleotides, suggesting an intron length maximum for the splicing of *Drosophila* introns with pyrimidine-poor 3′ splice sites. In contrast, the zeste 3′ splice site containing a 12-nucleotide uninterrupted pyrimidine tract was capable of directing a low level of splicing of 200- to 300-nucleotide introns. The zeste 3′ splice site, however, was not sufficient to direct 100% splicing of a large intron, because RNA products resulting from activation of cryptic splice sites within the expansion unit were observed.

The cryptic 3′ and 5′ splice sites located within the expansion unit are not, to our knowledge, activated in the *Drosophila* minigene from which they originated. Inspection of the sequences in the cryptic 3′ splice sites did not reveal strong homologies to the *Drosophila* 3′ splice site consensus. The 5′ splice site did have a reasonable fit to consensus. It should be noted that expanded introns containing the mle 3′ splice site utilized the cryptic 3′ splice site but ignored both the cryptic 5′ splice site and its own 3′ splice site. Presumably, this preference reflects a difficulty in removing the 203-nucleotide second

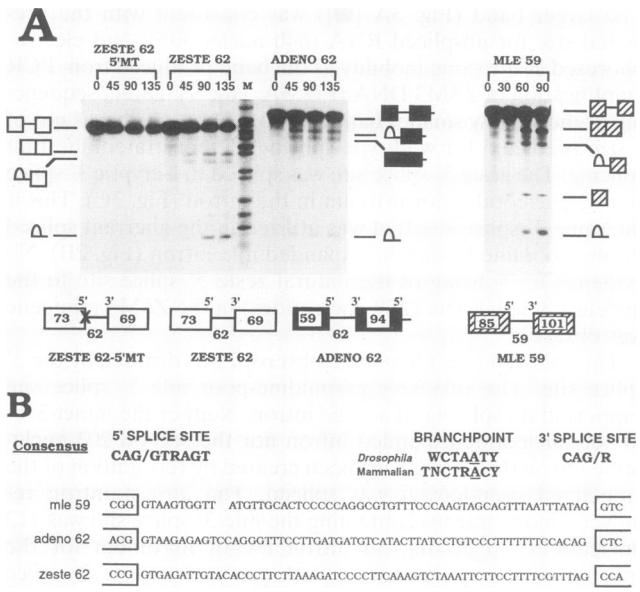

FIG. 4. In vitro splicing of small *Drosophila* introns. (A) Splicing precursor RNAs of the diagrammed structures were incubated in a nuclear extract prepared from S2 cells (Materials and Methods) under splicing conditions for the indicated times and displayed on a 5% denaturing polyacrylamide gel. The zeste 62 precursor RNAs include nucleotides 1308 to 1479, encompassing the second intron of the zeste gene; the mle 59 precursor RNA includes nucleotides 102 to 308, encompassing the first intron of the maleless gene; and the adeno 62 precursor RNA is an internally deleted version of the standard adenovirus MINX precursor RNA with an intron of 62 nucleotides. zeste 62-5′MT contained the same 5′ splice site mutation as the minigene, zeste-5′MT, described in the legend to Fig. 1. Splicing intermediates and products are indicated. These were identified by predicted sizes and changes in electrophoretic mobility on gels with differing percentages of acrylamide. Markers are *Msp*I-digested pBR322 DNA. (B) The sequences of the maleless, zeste, and adeno introns assayed above. Note the presence of a long pyrimidine tract at the 3′ end of both the adeno and the zeste introns and the lack of any discernible tract in the intron from maleless. Splicing signals, both 5′ and 3′, are a fairly strong match to the consensus splice site sequences shown above (24).

intron in a double splicing event and underscores the inability of the pyrimidine-poor mle 3′ splice site to direct the splicing of an expanded intron.

In vitro splicing of small *Drosophila* introns. The contrasting in vivo splicing phenotypes of 5′ splice site mutants in vertebrates and in *Drosophila* species and the observation of a size restriction for normal splicing of small *Drosophila* introns suggested a basic difference in the mechanism of recognition of splice sites in the two types of organisms. To begin to address this possibility, we developed an in vitro splicing system using a nuclear extract from *Drosophila* S2 cells. Initial experiments were performed to test the capabilities of the system. Figure 4A illustrates a time course of a standard splicing assay with several different precursor RNAs. zeste 62 and zeste 62-5′MT precursor RNAs contain, respectively, the wild-type or mutant 62-nucleotide second intron of zeste with its flanking exon sequences. These constructs are derived from the zeste minigenes that are analyzed for in vivo splicing in Fig. 1 to 3. In the in vitro *Drosophila* system, zeste 62 precursor RNA was spliced efficiently, such that ligated product RNA was visible by 45 min of reaction. zeste 62-5′MT precursor RNA was inhibited for any splicing activity (Fig. 4A). Thus, the S2 extract spliced wild-type RNA well and responded to mutation of a 5′ splice site by blocking the first step of splicing.

A second, small, 59-nucleotide *Drosophila* intron from the maleless gene, mle 59, was also assayed in this system. It too was spliced, although in most experiments not to the efficiency of zeste (Fig. 4A). However, splicing of small introns was not limited to introns from *D. melanogaster*. A 62-nucleotide adenovirus-based intron, adeno 62, was also spliced in the S2 extract (Fig. 4A). All three small introns, zeste 62, mle 59, and adeno 62, were spliced with similar efficiencies. Although the three introns have similar sizes, their sequences are different (Fig. 4B). Both adeno 62 and zeste 62 have pyrimidine-rich regions upstream of their 3′ splice sites, with 13 and 12 pyrimidine stretches, respectively, unlike the mle 59 intron, which has a very low percentage of pyrimidines in the same region. Thus, the S2 extract is not dependent on a pyrimidine-rich 3′ splice site to splice small introns. This ability of *Drosophila* extracts to splice small pyrimidine-rich or pyrimidine-poor introns has been observed previously for extracts from *Drosophila* Kc cells (9, 28).

The four constructs containing small introns, zeste 62, zeste 62-5′MT, adeno 62, and mle 59, were also assayed in HeLa extract, instead of S2 nuclear extract. As expected, these substrates were completely inactive for splicing, presumably because the introns fall below the minimum length requirement for vertebrate splicing (35) (data not shown). The *Drosophila* extract was not limited to splicing small introns. Larger adenovirus-based introns of 120 and 440 nucleotides were assayed in S2 extract. Both were accurately spliced, but with a slight decrease in efficiency as the intron size became larger (data not shown). Therefore, the in vitro S2 splicing system that we have developed is competent to splice introns of moderate size as well as introns that are too small to function in a vertebrate.

Assembly of *Drosophila* early-splicing complexes requires both ends of the intron. Vertebrate precursor RNAs containing a single intron assemble into multiple specific ATP-dependent complexes when splicing reactions are displayed on native gels (for a review, see reference 23). These complexes appear subsequent to one another in a time-dependent manner. Formation of the first complex, complex A, requires the presence of a 3′ splice site but is independent of the presence of a complete intron using constitutive 3′ splice sites (3, 7, 16, 21). Vertebrate precursor RNAs containing either a complete intron or only a 3′ splice site rapidly assembled complex A in our *Drosophila* S2 extract, indicating that assembly of complex A could be supported by a vertebrate 3′ splice site in both HeLa and *Drosophila* cell extracts (data not shown).

To examine the requirements for assembly of *Drosophila* introns, splicing reactions of precursor RNAs containing introns from the *Drosophila* maleless (59-nucleotide) and zeste (62-nucleotide) genes were displayed on native gels (Fig. 5A and B, respectively). Wild-type mle 59 and zeste 62 both directed assembly of an ATP-dependent complex by 2 min of reaction. Presumably, this complex is the *Drosophila* analog of the vertebrate A complex. Recent evidence by Spikes and Bingham (32) demonstrated that the complex A that is formed in *Drosophila* embryo extract contains U2 small nuclear RNA and unspliced precursor RNA, similarly to complex A in vertebrates. Indeed, in our assay the *Drosophila* complex showed migration on gels similar to that of the A complex assembled on an adenovirus precursor RNA in *Drosophila* S2 extract (data not shown).

The appearance of the maleless and zeste complex A was dependent on the 5′ splice site beginning the intron. Deletion of the 5′ splice site beginning the intron in mle 59Δ5 (Fig. 5A)

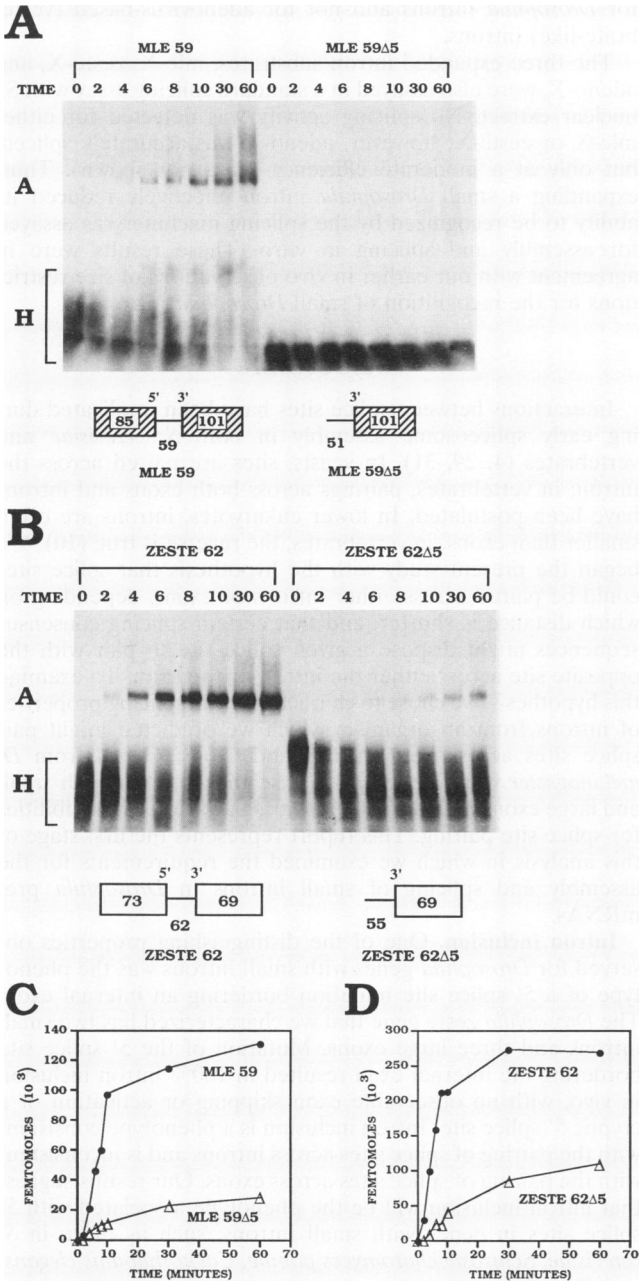

FIG. 5. Assembly of *Drosophila* early splicing complexes requires both ends of the intron. Equal amounts of the diagrammed precursor RNAs derived from maleless (A) and zeste (B) containing either all or the 3′ half of a functional intron were incubated in a *Drosophila* extract splicing assay for the indicated times. The assembled complexes were resolved on a native ribonucleoprotein gel. Well-resolved higher-order complexes were never detected in these assays. The numbers to the left of the 3′ splice sites in mle 59Δ5 and zeste 62Δ5 represent the number of nucleotides from the natural intron remaining after deletion of the 5′ splice site. (C and D) The amounts of complex A in panels A and B, respectively, quantitated and plotted versus incubation time.

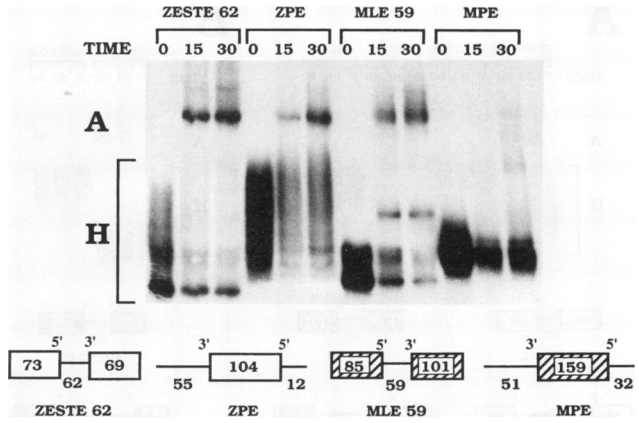

FIG. 6. Assembly of an isolated exon requires a pyrimidine-rich 3′ splice site. Equal amounts of precursor RNAs of the diagrammed structures were assayed under standard splicing conditions with *Drosophila* S2 extract. The assembled complexes were visualized on a native ribonucleoprotein gel. The isolated exons ZPE and MPE were constructed with splice sites and flanking intron and exon sequences from zeste 62 and mle 59, respectively. The abilities to form complex A of the isolated exons, ZPE and MPE, and their corresponding intact introns, zeste 62 and mle 59, on the same gel were compared. For ZPE and MPE, the numbers to the left of the 3′ splice site (55 and 51, respectively) indicate the numbers of nucleotides of natural intron sequence upstream.

or zeste 62Δ5 (Fig. 5B) depressed assembly of the complex. Quantitation of complex formation for the mle and zeste substrates (Fig. 5C and D, respectively) indicated that the 5′ splice site deletion caused a five- to ninefold repression of initial assembly. Thus, assembly of the 3′ splice sites from the *Drosophila* zeste and maleless small introns were more dependent on the presence of a complete intron than comparable vertebrate 3′ splice sites. Furthermore, the assembly properties of these 3′ splice sites were similar to those observed with *S. cerevisiae* splice sites, in which intact introns are required to observe ATP-dependent complex formation (4, 31).

Assembly of an isolated exon requires a pyrimidine-rich 3′ splice site. Because the presence of pyrimidines in a 3′ splice site permitted splicing of an expanded intron in vivo, we were interested in determining whether a pyrimidine tract would permit a *Drosophila* 3′ splice site to assemble an ATP-dependent complex when presented to the extract as an exon, rather than as an intron or an isolated 3′ splice site. In Fig. 5, we demonstrated that isolated 3′ splice sites from maleless and zeste were severely depressed for assembly of complex A compared with intact introns containing both 5′ and 3′ splice sites. We then asked if placing a 5′ splice site downstream of the 3′ splice site (creating a pseudoexon) would have a similar effect. More importantly, we asked if there is a difference in directing spliceosome assembly between the maleless and zeste 3′ splice sites.

To address these questions, two in vitro substrates, ZPE and MPE, were constructed. These substrates contained either the zeste (ZPE) or maleless (MPE) 3′ splice site with flanking exon sequences fused to its cognate 5′ splice site, normally upstream across the intron, now placed downstream to create a pseudoexon structure (Fig. 6). The constructs were assayed for their ability to assemble complex A in a standard splicing reaction with S2 extract. The reaction products were displayed on a native agarose-acrylamide gel. Considerable complex A was observed for the zeste exon but not for the mle exon. The mle exon assembled no visible complex A, even after 30 min of reaction. The zeste exon was competent to assemble at virtually the same efficiency as the zeste intron. Therefore, the zeste 3′ splice site containing a pyrimidine-rich region was able to direct assembly of complex A when the zeste 5′ splice site was

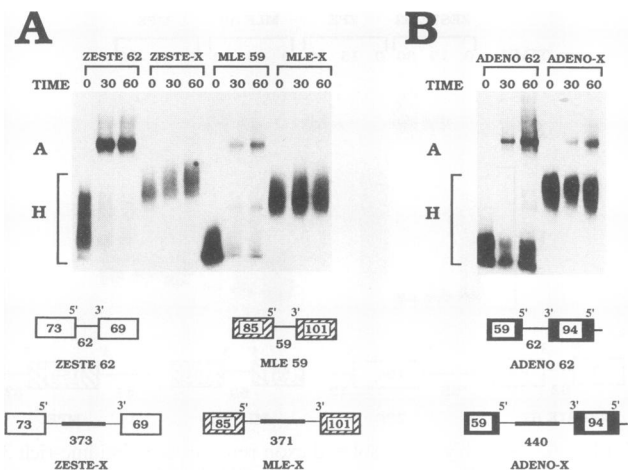

FIG. 7. Expansion of small introns inhibits in vitro assembly. Standard splicing reactions in S2 extract were performed with precursor RNAs of the diagrammed structures. Assembly complexes were displayed on native ribonucleoprotein gels. (A) Complex formations between the small *Drosophila* introns, zeste 62 and mle 59, and the same introns expanded with *Drosophila* troponin T intron sequences, zeste-X and mle-X, respectively, were compared. The thin black boxes in the middle of the zeste-X and mle-X introns represent the troponin T sequences. The numbers below the black boxes, 373 and 371, indicate the total lengths in nucleotides of the expanded introns. (B) Vertebrate-like substrates derived from adenovirus sequences were assayed in S2 extract to compare the efficiencies of assembly between a small versus an expanded vertebrate intron. Sequences from intron 3 of the human calcitonin/CGRP gene were used to expand the intron of adeno 62 to create adeno-X. These calcitonin/CGRP sequences are represented by the thin black box in the intron of adeno-X. The total size of the expanded intron is 440 nucleotides, as indicated.

placed either upstream across an intron or downstream across an exon. In contrast, the maleless 3′ splice site without a pyrimidine stretch was competent for assembly only when paired with a 5′ splice site across an intron.

Assembly of splicing complexes is depressed for expanded small introns. Expansion of the small zeste and maleless introns promoted aberrant splicing phenotypes in vivo, as seen in Fig. 2. These data suggested that the splicing machinery preferred to select splice sites in pairs across the shortest distance. To examine this at an early stage of assembly, two expanded intron substrates for in vitro analysis were constructed by using the maleless and zeste introns to create mle-X and zeste-X, respectively. The exon and intron sequences in these constructs were identical to those in the expanded intron constructs used in vivo. mle-X and zeste-X were first assayed for their abilities to assemble into splicing complexes in S2 nuclear extract along with their naturally small parent constructs, mle 59 and zeste 62 (Fig. 7A). Both mle-X and zeste-X were depressed for the formation of complex A compared with mle 59 and zeste 62. Quantitation of this assay revealed that complex A assembly was depressed 15-fold for zeste-X and to nondetectable levels for mle-X.

The inhibition of complex A formation in mle-X and zeste-X is not due to an inability of the S2 extract to recognize large introns, because expansion of a small adenovirus-based intron from 62 (adeno 62) to 440 (adeno-X) nucleotides in length produced only a twofold reduction of spliceosome assembly in this extract (Fig. 7B). Therefore, the inhibition of complex assembly due to increasing the size of a small intron is specific for *Drosophila* introns and not for adenovirus-based (vertebrate-like) introns.

The three expanded intron substrates, mle-X, zeste-X, and adeno-X, were also assayed in a standard splicing assay with S2 nuclear extract. No splicing activity was detected for either mle-X or zeste-X; however, adeno-X was accurately spliced, but only at a moderate efficiency (data not shown). Thus, expanding a small *Drosophila* intron effectively reduced its ability to be recognized by the splicing machinery as assayed for assembly and splicing in vitro. These results were in agreement with our earlier in vivo observations of size restrictions for the recognition of small *Drosophila* introns.

DISCUSSION

Interactions between splice sites have been implicated during early spliceosome assembly in both *S. cerevisiae* and vertebrates (4, 29, 31). In yeasts, sites are paired across the intron; in vertebrates, pairings across both exons and introns have been postulated. In lower eukaryotes, introns are often smaller than exons; in vertebrates, the reverse is true (10). We began the present study with the hypothesis that splice sites could be paired across either introns or exons, depending on which distance is shorter, and that certain splicing consensus sequences might dispose a given splice site to pair with the opposite site across either the intron or the exon. To examine this hypothesis, we chose to characterize the splicing properties of introns from an organism which we predicted might pair splice sites across both introns and exons. Genes from *D. melanogaster* were analyzed because they contain both small and large exons and introns, presenting a variety of possibilities for splice site pairing. This report represents the first stage of this analysis in which we examined the requirements for the assembly and splicing of small introns in *Drosophila* pre-mRNAs.

Intron inclusion. One of the distinguishing properties observed for *Drosophila* genes with small introns was the phenotype of a 5′ splice site mutation bordering an internal exon. The *Drosophila* zeste gene that we characterized has two small introns and three large exons. Mutation of the 5′ splice site bordering the internal exon resulted in 100% intron inclusion in vivo, with no observable exon skipping or activation of a cryptic 5′ splice site. Intron inclusion is a phenotype consistent with the pairing of splice sites across introns and is inconsistent with the pairing of splice sites across exons. Our results suggest that intron inclusion will be the phenotype associated with 5′ splice sites in genes with small introns, such as those in *S. cerevisiae*, *Schizosaccharomyces pombe*, *Caenorhabditis elegans*, and *D. melanogaster*.

While this work was in progress, we learned that Mount et al. (25a) have isolated a 5′ splice site mutation in the 74-nucleotide second intron of the white gene of *D. melanogaster* that results in intron inclusion. Four of the five introns in the white gene are small; thus, for the most part, the white gene has the intron-exon architecture of the zeste gene used in the present study.

At the same time, another *Drosophila* gene with a 5′ splice site mutation was brought to our attention by Bob Schulz (29a). He has isolated a mutant allele of the *Drosophila ets*-like gene D-elg, containing a mutation in the 5′ splice site of intron 2, which is 61 nucleotides in length. D-elg is a five-exon gene with four introns of fewer than 80 nucleotides. Thus, it has an exon-intron architecture similar to that of zeste. We performed RT-PCR analysis on RNAs from flies carrying the mutant allele to determine its splicing phenotype. We observed 100% intron inclusion (data not shown). Therefore, three different

genes containing large exons and small introns, zeste, white, and D-elg, all respond to a 5′ splice site mutation similarly by retaining the mutant intron in the final product RNA, and all three thereby have nonvertebrate splicing phenotypes.

Two other splice site mutants in *Drosophila* genes with multiple introns have been reported. One of these is a 3′ splice site mutant in the shaker locus which leads to intron inclusion (22); the second is a 5′ splice site mutation in the myosin heavy chain region that leads to aberrant products containing multiple introns (19). The intron inclusion phenotype produced by mutating the 3′ splice site in the 220-nucleotide intron 19 of shaker is in agreement with our hypothesis that small introns pair splice sites across the intron. However, the myosin heavy chain mutation resides in a region of the pre-mRNA that is differentially spliced; therefore, it is difficult to sort out alterations in constitutive versus differential processes.

Until recently, no case of intron inclusion in a vertebrate had been identified. However, we have observed in vivo inclusion of a 166-nucleotide intron residing between exons that are 30 and 7 nucleotides long from chicken fast skeletal troponin I when the 5′ splice site of the intron had been mutated (33). This event occurs in a region of the pre-mRNA in which the exon-intron-exon distance is less than 300 nucleotides, a length compatible with the size limits of vertebrate internal exons. In addition, several other examples of intron inclusion of small vertebrate introns have been reported elsewhere (17, 26, 27). Therefore, it seems probable that intron inclusion phenotypes in vertebrates can be observed when introns and neighboring exons are small.

Intron size limitation. Recently, Guo et al. have observed that increasing the size of the white intron from 74 to 90 nucleotides severely depressed the efficiency of splicing activity in in vitro *Drosophila* extracts (9). We also observed that expanding the zeste and mle introns from approximately 60 to 400 nucleotides inhibited in vitro *Drosophila* splicing. Such expansion also inhibited in vivo splicing and caused the activation of cryptic splice sites within the intron. The expanded maleless intron utilized a single cryptic 3′ splice site within the intron to remove an artificial 112-nucleotide intron. By sequence analysis, this activated site is a poorer match to the *Drosophila* 3′ splice site consensus sequence than is the natural mle 3′ splice site ignored in splicing of the expanded intron. Thus, the in vivo splicing machinery seemed to prefer a small intron with a poor 3′ splice site to a larger intron with the natural 3′ splice site.

Expansion of the pyrimidine-rich zeste intron with the same troponin T sequences also caused activation of intron-located splice sites. Interestingly, some level of wild-type splicing was also observed, despite the size of the expanded intron. Thus, both the expanded zeste intron and the expanded chimeric intron with the zeste 3′ splice site permitted the splicing of 373- or 393-nucleotide introns, respectively. In contrast, expansion of introns with the pyrimidine-poor mle 3′ splice site supported no wild-type splicing of the expanded intron. This difference suggests that pyrimidine tracts are necessary for splicing of large introns. Analysis of the frequency of pyrimidine content in *Drosophila* introns supports this premise, in that large introns have a higher frequency of pyrimidines near the 3′ splice site than do small introns (24).

However, correlation of the ability to splice large introns in *D. melanogaster* and the presence of 3′ pyrimidine tracts is not absolute. Although expansion of the zeste intron permitted some wild-type splicing, it also strongly activated internal cryptic splice sites to remove two small introns instead of the larger intron. Therefore, small introns were also spliced, despite the presence of a 12-nucleotide uninterrupted pyrimidine tract within the zeste 3′ splice site. The converse is also true. Expansion of the pyrimidine-poor white intron with a copia element permits some wild-type splicing in vivo; however, the insertion does depress overall splicing of this intron (20, 25, 36). More extensive mutational analysis of *Drosophila* splicing signals from small and large introns will be required to determine which elements are required for efficient processing of large introns.

SUBSTRATE STRUCTURES	3′ SPLICE SITE: PY POOR, MLE 59 NT	3′ SPLICE SITE: PY RICH, ZESTE 62 NT	3′ SPLICE SITE: PY RICH, FTZ 150 NT
Isolated 3′ SS	-	-	+
Intact Intron	+	+	+
Intact Exon	-	+	?
Expanded Intron (In Vitro)	-	-	?
Expanded Intron (In Vivo)	-	+	+

FIG. 8. Splicing and assembly properties of three *Drosophila* introns with pyrimidine (PY)-poor versus pyrimidine-rich 3′ splice sites (SS). The requirements for splicing and assembly of two small introns from zeste and maleless presented in this report are displayed, along with previously published information on the larger ftz intron. Each intron is labeled either pyrimidine poor of pyrimidine rich on the basis of the frequency of pyrimidines present upstream of the 3′ splice site. The natural length of each intron is also indicated. +, intron capable of wild-type splicing and/or assembly depending on the substrate assayed; −, intron not capable of splicing and/or assembly. ?, no information for the substrates diagrammed available. References for previously published data are as follows: isolated 3′ splice sites (32), intact intron (9, 28, 32), and expand intron (in vivo) (12).

Spliceosome assembly. In support of mechanistic differences between the recognition of splice sites in pre-mRNAs with small introns versus large introns, we observed that in vitro assembly of *Drosophila* spliceosome complex A required sequences at both ends of the zeste and mle introns, in contrast to vertebrate assembly of complex A, which required only sequences at the 3′ end of the intron with a minimal amount of flanking exon sequence (3, 7, 16, 21). The zeste and mle assembly requirements for intact introns are similar to those observed for *S. cerevisiae* (4, 31).

Surprisingly, assembly of complex A on the small maleless and zeste introns required sequences different from those necessary for complex formation on the larger, 150-nucleotide intron from the *Drosophila* fushi tarazu (ftz) gene (32). The ftz 3′ splice site contains an 11-nucleotide uninterrupted pyrimidine tract. The isolated 3′ splice site from ftz assembles complex A, similarly to a vertebrate 3′ splice site. Moreover, the efficiencies of complex formation on the ftz isolated 3′ splice site and intact intron are nearly identical. Thus, the ftz intron exhibits properties similar to those of a vertebrate intron. In vivo, the ftz intron tolerates considerable expansion (greater than 2 kb), again reminiscent of vertebrate pyrimidine-rich introns (12, 32). Inspection of ftz 3′ splice site sequences does not reveal an obvious difference between these sequences and that of the 3′ splice site of zeste, in that both introns have uninterupted pyrimidine tracts. Mutational analysis of both 3′ splice sites should yield interesting clues as to the features of a 3′ splice site necessary for the splicing of large introns and/or exon definition.

Interestingly, although the zeste 3′ splice site was unable to support ATP-dependent complex formation, we observed that

it directed formation of considerable complex A when a 5′ splice was present downstream, such that it created an isolated internal exon. Therefore, this 3′ splice site directed ATP-dependent complex formation when present as either an exon or an intron. Strikingly, the pyrimidine-poor 3′ splice site from maleless was not rescued for assembly by addition of a downstream 5′ splice site; it could assemble only when present within an intron. Thus, the pyrimidine content of the 3′ splice site may be crucial in maintaining the balance between selecting splice sites across introns and/or exons in *Drosophila* genes with multiple introns.

Splice site pairing. We have summarized the data presented here in Fig. 8 by displaying the distinguishing splicing and assembly properties of three *Drosophila* introns. We propose from our observations that splice sites can be paired across either introns or exons. It is unclear if distance between sites is the dominant determinant of the pairing pathway or if splice site consensus sequences contribute. The ability of a pyrimidine-rich 3′ splice site to assemble both an intron and an exon reported here suggests that pyrimidine stretches and the factors that recognize them may influence the pairing pathway. In this regard, it will be interesting to examine the processing of pre-mRNAs that are transcribed from genes with a mixed exon-intron architecture. A few examples of genes of this type in *Drosophila* species include notch (14), abelson (11), and lethal (2) giant larvae (13). Their mixed architecture suggests that both intron-based and exon-based recognition mechanisms could operate in a single pre-mRNA. The mechanism whereby this is possible remains unclear. Mutation of 5′ splice sites within genes such as notch and abelson may help sort out these questions.

ACKNOWLEDGMENTS

We are grateful to P. Miller for help in setting up *Drosophila* tissue culture; G. Cote, T. Cooper, B. Mattox, D. Stolow, and C. Kennedy for stimulating discussions; and B. Moore for technical assistance.

This research was supported by Public Health Service grant GM38526 to S.B. M.T. was supported by the Robert A. Welch Foundation.

REFERENCES

1. **Aebi, M., H. Hornig, R. A. Padgett, J. Reiser, and C. Weissmann.** 1986. Sequence requirements for splicing of higher eukaryotic nuclear pre-mRNAs. **47:**555–565.
2. **Aebi, M., and C. Weismann.** 1987. Precision and orderliness in splicing. Trends Genet. **3:**102–107.
3. **Bindereif, A., and M. R. Green.** 1987. An ordered pathway of snRNP binding during mammalian pre-mRNA splicing complex assembly. EMBO J. **6:**2415–2424.
4. **Cheng, S. C., and J. Abelson.** 1987. Spliceosome assembly in yeast. Genes Dev. **1:**1014–1027.
5. **Csank, C., F. M. Taylor, and D. W. Martindale.** 1990. Nuclear pre-mRNA introns: analysis and comparison of intron sequences from *Tetrahymena thermophila* and other eukaryotes. Nucleic Acids Res. **18:**5133–5141.
6. **Dignam, J. D., R. M. Lebovitz, and R. G. Roeder.** 1983. Accurate and efficient transcription initiation by RNA polymerase II in a soluble extract from mammalian nuclei. Nucleic Acids Res. **11:** 1475–1489.
7. **Frendewey, D., and W. Keller.** 1985. Stepwise assembly of a pre-mRNA splicing complex requires U-snRNPs and specific intron sequences. Cell **42:**355–367.
8. **Goguel, V., and M. Rosbash.** 1993. Splice site choice and splicing efficiency are positively influenced by pre-mRNA intramolecular base pairing in yeast. Cell **72:**893–902.
9. **Guo, M., P. C. H. Lo, and S. M. Mount.** 1993. Species-specific signals for the splicing of a short *Drosophila* intron in vitro. Mol. Cell. Biol. **13:**1104–1118.
10. **Hawkins, J. D.** 1988. A survey on intron and exon lengths. Nucleic Acids Res. **16:**9893–9908.
11. **Henkemyer, M. J., R. L. Bennett, F. B. Gertler, and F. M. Hoffmann.** 1988. DNA sequence, structure, and tyrosine kinase activity of the *Drosophila melanogaster Abelson* proto-oncogene homolog. Mol. Cell. Biol. **8:**843–853.
12. **Horabin, J. I., and P. Schedl.** 1993. Regulated splicing of the *Drosophila sex-lethal* male exon involves a blockage mechanism. Mol. Cell. Biol. **13:**1408–1414.
13. **Jacob, L., M. Opper, B. Metzroth, B. Phannavong, and B. M. Mechlen.** 1987. Structure of the l(2)gl gene of *Drosophila* and delimitation of its tumor suppressor domain. Cell **50:**215–225.
14. **Kidd, S., M. R. Kelley, and M. W. Young.** 1986. Sequence of the *notch* locus of *Drosophila melanogaster*: relationship of the encoded protein to mammalian clotting and growth factors. Mol. Cell. Biol. **6:**3094–3108.
15. **Klinz, F. J., and D. Gallwitz.** 1985. Size and position of intervening sequences are critical for the splicing of pre-mRNA in the yeast *Saccharomyces cerevisiae*. Nucleic Acids Res. **13:**3791–3804.
16. **Konarska, M. M., and P. A. Sharp.** 1986. Electrophoretic separation of complexes involved in pre-mRNA splicing. Cell **46:**845–855.
17. **Kontusaari, S., G. Tromp, H. Kuivaniemi, R. L. Ladda, and D. J. Prokop.** 1990. Inheritance of an RNA splicing mutation ($G^{+1\ IVS\ 20}$) in the type III procollagen gene (COL3A1) in a family having aortic aneurysms and easy bruisability: phenotypic overlap between familial arterial aneurysms and Ehler-Danlos syndrome type IV. Am. J. Hum. Genet. **47:**112–120.
18. **Krawczak, M., J. Reiss, and D. N. Cooper.** 1992. The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. Hum. Genet. **90:**41–54.
19. **Kronert, W. A., K. A. Edwards, E. S. Roche, L. Wells, and S. I. Bernstein.** 1991. Muscle specific accumulation of *Drosophila myosin heavy chains*: a splicing mutation in an alternative exon results in an isoform substitution. EMBO J. **10:**2479–2488.
20. **Kurkulos, M., J. M. Weinberg, M. E. Pepling, and S. M. Mount.** 1991. Polyadenylation in copia requires unusually distant upstream sequences. Proc. Natl. Acad. Sci. USA **88:**3038–3042.
21. **Lammond, A. I., M. M. Konarska, and P. A. Sharp.** 1987. A mutational analysis of spliceosomal assembly: evidence for splice site collaboration during spliceosome formation. Genes Dev. **1:**532–543.
22. **Lichtinghagen, R., M. Stocker, R. Wittka, G. Bonheim, W. Stumer, A. Ferrus, and O. Pongs.** 1990. Molecular basis of altered excitability in *shaker* mutants of *Drosophila melanogaster*. EMBO J. **9:**4399–4407.
23. **Moore, M. J., C. C. Query, and P. A. Sharp.** 1993. Splicing of precursors to messenger RNAs by the spliceosome, p. 303–357. *In* R. Gesteland and R. Atkins (ed.), The RNA world. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
24. **Mount, S. M., C. Burks, G. Hertz, G. D. Stormo, O. White, and C. Fields.** 1992. Splicing signals in *Drosophila*: intron size, information content and consensus sequences. Nucleic Acids Res. **20:** 4255–4262.
25. **Mount, S. M., M. M. Green, and G. M. Rubin.** 1988. Partial revertants of the transposable element-associated suppressible allele *white-apricot* in *Drosophila melanogaster*: structure and responsiveness to genetic modifiers. Genetics **118:**221–234.

25a. **Mount, S. M.** Personal communication.

26. **Ohno, K., and K. Suzuki.** 1988. Multiple abnormal β-hexaminidase a chain mRNAs in a compound-heterozygous Ashkenazi Jewish patient with Tay-Sachs disease. J. Biol. Chem. **263:**18563–18567.
27. **Patterson, D., R. Berger, J. Bleskan, D. Vannais, and J. Davidson.** 1992. A single base change at a splice acceptor site leads to a truncated CAD protein in Urd—a mutant Chinese hamster ovary cells. Somatic Cell Mol. Gen. **18:**65–75.
28. **Rio, D. C.** 1988. Accurate and efficient pre-mRNA splicing in *Drosophila* cell-free extracts. Proc. Natl. Acad. Sci. USA **85:**2904–2909.
29. **Robberson, B. L., G. J. Cote, and S. M. Berget.** 1990. Exon definition may facilitate splice site selection in RNAs with multiple exons. Mol. Cell. Biol. **10:**84–94.

29a. **Schulz, R.** Personal communication.
30. **Seraphin, B., and M. Rosbash.** 1989. Identification of functional U1 snRNA-premRNA complexes committed to spliceosome assembly and splicing. Cell **59:**349–358.
31. **Seraphin, B., and M. Rosbash.** 1991. The yeast branch point sequence is not required for the formation of a stable U1 snRNP pre-mRNA complex and is recognized in the absence of U2 snRNA. EMBO J. **10:**1209–1216.
32. **Spikes, D., and P. M. Bingham.** 1992. Analysis of spliceosome assembly and the structure of a regulated intron in *Drosophila* in vitro splicing extracts. Nucleic Acids Res. **20:**5719–5727.
33. **Sterner, D. A., and S. M. Berget.** 1993. In vivo recognition of a vertebrate mini-exon as an exon-intron-exon unit. Mol. Cell. Biol. **13:**2677–2687.
34. **Talerico, M., and S. M. Berget.** 1990. Effect of 5′ splice site mutations on splicing of the preceding intron. Mol. Cell. Biol. **10:**6299–6305.
35. **Wieringa, B., E. Hofer, and C. Weissmann.** 1984. A minimal intron length but no specific internal sequence is required for splicing the large rabbit β-globin intron. Cell **37:**915–925.
36. **Zachar, Z., D. Davidson, D. Garza, and P. M. Bingham.** 1985. A detailed developmental and structural study of the transcriptional effects of insertion of the *copia* transposon into the *white* locus of *Drosophila melanogaster*. Genetics **111:**495–515.