MOLECULAR AND CELLULAR BIOLOGY, Feb. 1993, p. 1104–1118
0270-7306/93/021104-15$02.00/0

Vol. 13, No. 2

Species-Specific Signals for the Splicing of a Short *Drosophila* Intron In Vitro

MING GUO, PATRICK C. H. LO, AND STEPHEN M. MOUNT*

Department of Biological Sciences, Columbia University, New York, New York 10027

Received 3 August 1992/Returned for modification 17 September 1992/Accepted 25 November 1992

The effects of branchpoint sequence, the pyrimidine stretch, and intron size on the splicing efficiency of the *Drosophila white* gene second intron were examined in nuclear extracts from *Drosophila* and human cells. This 74-nucleotide intron is typical of many *Drosophila* introns in that it lacks a significant pyrimidine stretch and is below the minimum size required for splicing in human nuclear extracts. Alteration of sequences adjacent to the 3′ splice site to create a pyrimidine stretch was necessary for splicing in human, but not *Drosophila*, extracts. Increasing the size of this intron with insertions between the 5′ splice site and the branchpoint greatly reduced the efficiency of splicing of introns longer than 79 nucleotides in *Drosophila* extracts but had an opposite effect in human extracts, in which introns longer than 78 nucleotides were spliced with much greater efficiency. The *white-apricot* copia insertion is immediately adjacent to the branchpoint normally used in the splicing of this intron, and a copia long terminal repeat insertion prevents splicing in *Drosophila*, but not human, extracts. However, a consensus branchpoint does not restore the splicing of introns containing the copia long terminal repeat, and alteration of the wild-type branchpoint sequence alone does not eliminate splicing. These results demonstrate species specificity of splicing signals, particularly pyrimidine stretch and size requirements, and raise the possibility that variant mechanisms not found in mammals may operate in the splicing of small introns in *Drosophila* and possibly other species.

The splicing of eukaryotic mRNA precursors in mammalian cells has been studied extensively (for reviews, see references 18 and 79), and the signals that govern the identification of splice sites are generally known. A 5′ splice site that conforms to the consensus sequence MAG|GURAGU (M = C or A; R = A or G) and includes the underlined GU dinucleotide is required. 3′ splice sites conform to the consensus sequence YAG|G (Y = C or U) and are typically found at the site of the first AG dinucleotide downstream of the branchpoint. Branchpoints fit the consensus sequence UNCURAC (in which branch formation occurs at the underlined A) and usually reside between 18 and 38 nucleotides upstream of the 3′ splice site. Between the branchpoint and the 3′ splice site is a pyrimidine-rich region. The way in which sequences at the 5′ splice site, the branchpoint, the pyrimidine-rich stretch, and the 3′ splice site act together in mammalian splicing to specify intron boundaries has been investigated in detail, and much is known of the factors that recognize these sites. For example, the 5′ splice site is recognized by the U1 small nuclear ribonucleoprotein (snRNP) via base pairing in both mammals and in yeasts (8, 34, 47, 76, 78, 90), and the branchpoint is similarly recognized by the U2 snRNP (50, 58, 84, 89, 91). Binding of the U2 snRNP to the branchpoint requires a number of factors, including the U1 snRNP (2, 67, 75) and U2AF, a factor that binds to the pyrimidine-rich stretch (70).

Despite the relatively large amount that is known about the sequence requirements for splicing, it is still not possible to accurately predict the positions of introns from sequence information alone, and the basis of alternative, or regulated, splicing is still being elucidated. The genetics of *Drosophila melanogaster* has allowed regulatory factors to be identified for several examples of alternative splicing (4, 10, 11, 41, 49, 85), and it appears that this organism will prove useful for the study of alternative splicing. However, the basic information on splicing signals which exists in yeast and mammalian systems has no counterpart in *D. melanogaster*, and splicing signals do vary between species. For example, the relative A+T richness of plant introns is critical to their proper recognition (17), and animal introns are not properly recognized in transfected plant cells (16, 82). Similarly, most mammalian introns are not recognized by the yeast *Saccharomyces cerevisiae* (3, 32). This is due, at least in part, to the fact that yeast introns almost always use the precise sequence UACUAAC as a branchpoint, and this sequence is the primary determinant of yeast 3′ splice site selection (25, 51, 57). In contrast, the branchpoint sequence of mammalian introns has greater flexibility (26, 28, 50, 65, 69, 88), and the pyrimidine-rich stretch is more important (13, 64, 68).

A thorough analysis of *Drosophila* introns in GenBank (45) revealed that although *Drosophila* splice sites are like those in mammalian genes, *Drosophila* introns as a whole differ from mammalian introns in several significant ways. First, *Drosophila* introns tend not to have a G in the position preceding the branched nucleotide, although G is the most common nucleotide at that position in mammalian introns. Second, many *Drosophila* introns are shorter than the smallest mammalian introns. Third, *Drosophila* introns differ from mammalian introns in base composition, with a 17% greater A+T content in introns than in flanking exons and a much less extreme preference for pyrimidines in the region between the branchpoint and the 3′ splice site.

In this study, we used in vitro splicing in both *Drosophila* and human cell nuclear extracts to explore the signals required for the splicing of a small *Drosophila* intron. The second intron of the *Drosophila white* gene was chosen for this study because it is characteristic of small pyrimidine-poor *Drosophila* introns that lack sequence features required for mammalian splicing. In addition, the splicing of this 74-bp intron is altered in the *white-apricot* (w^a) allele by an insertion of the transposable element copia in the same

* Corresponding author.

transcriptional orientation (5, 15, 56, 60). In w^a, the level of normally spliced mRNA is greatly reduced, and a number of aberrant RNAs that are polyadenylated within copia are observed (36, 46, 61, 86). Interest in w^a derives from the existence of mutations in unlinked genes which alter its expression, resulting in increased or decreased eye pigmentation (6, 7, 11, 38, 59, 62, 81, 85). Correlation of the structure of a number of derivatives of w^a with their phenotypes, the RNAs that they produce, and their response to genetic modifiers has led to the conclusion that there is competition between polyadenylation within copia and the splicing of this intron (31). Furthermore, it appears that the copia insertion in w^a has not only provided a polyadenylation site but has also interfered with splicing in some way. For example, derivatives of w^a that show little or no polyadenylation within copia (including those in which the copia element has been replaced by a single long terminal repeat [sLTR]) are expressed at less than fully wild type levels (31, 46, 86). In addition, because a number of studies have shown that polyadenylation sites within introns can be spliced out (1, 33, 37), even minimal use of the copia polyadenylation site suggests a disruption of splicing. In this report, we show that the w^a copia insertion has altered the branchpoint normally used in the splicing of this intron. Additional results indicate that the increased size of the intron may also contribute to the splicing defect in w^a. A series of introns with alterations between the 5′ splice site and the branchpoint revealed little or no splicing of introns longer than 79 nucleotides in *Drosophila* cell nuclear extract but an opposite effect in human cell extracts, in which introns shorter than 78 nucleotides were not spliced. Furthermore, a pyrimidine stretch adjacent to the 3′ splice site was found to be essential in human extracts and unnecessary (but stimulatory) for splicing in *Drosophila* extracts. These in vitro differences between nuclear extracts from *Drosophila* and human cells in their responses to variation in the sequence of RNA substrates confirm differences suggested by differences between these species in the sequences of their introns.

MATERIALS AND METHODS

Constructions. Plasmids pMG1 and pMG2 (Fig. 1) were constructed by inserting the *Pvu*II (*white* nucleotide 11078)-*Sal*I (position 11867) DNA fragments from plasmids pm12.5 (35) and pC1LTR (46), respectively, between the *Sma*I and *Sal*I sites of the vector pIBI24 (International Biotechnologies, Inc.), downstream of the promoter for T7 RNA polymerase. The resulting clones contained portions of the second and third exons of the wild-type *white* gene with either the wild-type second intron (pMG1) or the same intron carrying an sLTR insertion (pMG2).

The linker insertion mutation pMG3 was constructed by using the polymerase chain reaction (71). pMG1 was used as a template for amplification in two reactions. A 120-bp fragment containing exon 2 and 29 bp of the second intron (11078 to 11182) was amplified by using oligonucleotide MG2 (5′-GGATCCATCGATATCAGATCAGCCGACTGCGA-3′) and reverse sequencing primer 1201 (New England Biolabs). A 670-bp fragment containing the remainder of intron 2 and all of exon 3 (11182 to 11867) was amplified by using sequencing primer 1211 (New England Biolabs) and oligonucleotide MG1 (5′-GATATCGATGGATCCTGTGTGAAA TCTTAAT-3′). Oligonucleotides MG1 and MG2 carried *Cla*I sites within a nonannealing region at their 5′ ends. pMG3 was then constructed by ligating the 120-bp amplified fragment cut with *Eco*RI and *Cla*I, the 670-bp amplified fragment cut with *Cla*I and *Hin*dIII, and pIBI24 cut with *Eco*RI and *Hin*dIII. The resulting construct, pMG3, contains a 16-bp polylinker with sites for *Eco*RV, *Cla*I, and *Bam*HI 29 bp downstream of 5′ splice site. Sequences were confirmed by DNA sequencing (73).

The linker substitution mutations pMG7 and pMG37 were also constructed by using the polymerase chain reaction (71). A fragment of pMG1/31 containing exon 2 and 13 bp of the second intron (11078 to 11169) was amplified by using oligonucleotide MG9 (5′-GGATCCATCGATATCAATAGA AACTCACCGTTC-3′) and reverse sequencing primer 1201 (New England Biolabs). Oligonucleotide MG9 carried a *Cla*I site within a nonannealing region at its 5′ end. pMG7 and pMG37 were then constructed by ligating the amplified fragment cut with *Eco*RI and *Cla*I and pMG3/33 cut with *Eco*RI and *Cla*I. The resulting construct, pMG7/37, contains a 16-bp polylinker substitution between the 5′ splice site and branchpoint, 13 bp downstream of 5′ splice site. Sequences were confirmed by DNA sequencing.

Mutants with increased pyrimidine content (Fig. 1D) or branchpoint alterations (Fig. 3A and C) were generated by oligonucleotide-directed mutagenesis by using oligonucleotides MG3 (for making mutants with increased pyrimidine content; 5′-TTACCAATTTTTTCCTCAGTTTGC-3′), DCB1 (for making mutants pMG4/34 from pMG2/32; 5′-GTAATT GGACCCTTTATTAGTAATTTATAATTTA-3′), and MG6 (for making mutants pMG5/35 from pMG1/31; 5′-CTGTGTG AAAACAACATAAAGGGTCC-3′). The template for mutagenesis was generated by subcloning an *Eco*RI-*Hin*dIII fragment of pMG1 and pMG2 into M13mp18 or M13mp19. Mutagenesis was performed essentially as described by Kunkel et al. (30), and mutants were identified by DNA sequencing.

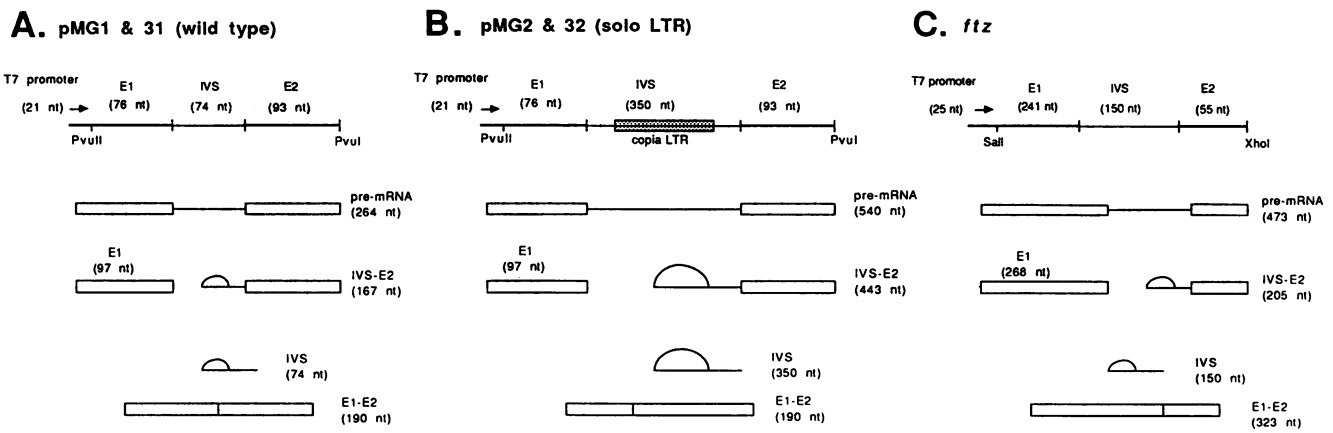
Deletion mutants derived from pMG3 and pMG33 (Fig. 5C) were made by nuclease Bal 31 digestion of pMG3 and pMG33 DNA linearized at the *Eco*RV site. Deletion mutagenesis was performed essentially as described by Sambrook et al. (72), and mutants were identified by DNA sequencing.

Nuclear extracts. *Drosophila* Kc cells were grown in D22 medium. *Drosophila* Kc cell and human HeLa or 293 cell nuclear extracts were prepared by a modification of the protocol of Dignam et al. (12) in which 42 mM $(NH_4)_2SO_4$ was substituted for 0.1 M KCl in the final dialysis step (53).

Precursor preparation and in vitro splicing. Capped precursor RNAs were produced by runoff transcription with T7 RNA polymerase (Promega) of template linearized at the *Pvu*I site in the third exon of the *white* gene (Fig. 1A and B) or at the *Xho*I site of the *fushi tarazu* (*ftz*) gene (Fig. 1C). ^{32}P labeling was provided by inclusion of [α-^{32}P]GTP at a final specific activity of 12 Ci/mmol. Pre-mRNA (50,000 cpm) was incubated in 25-μl reaction mixtures containing 10 μl of nuclear extract, 2.5% polyvinyl alcohol, 20 mM *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid (HEPES)-KOH (pH 7.6), 5 mM creatine phosphate, 1 mM $MgCl_2$, and 3 mM ATP as described previously (66). Reactions were terminated after 3 to 3.5 h of incubation at 20°C, and RNA was extracted after proteinase K digestion (29) and analyzed on denaturing polyacrylamide gels. K^+ ion concentrations between 0 and 100 mM and Mg^{2+} ion concentrations between 0 and 4 mM were tested, and reactions were performed under optimal conditions (described above).

Debranching of lariat RNAs was carried out in 25-μl reaction mixtures containing 20 mM HEPES (pH 7.9), 20 mM KCl, 10 mM EDTA, 20% glycerol, 1 mM dithiothreitol, and 10 μl of HeLa S100 extract (12, 68) at 30°C for 60 min.

Analysis of polyadenylation activity. A nonspecific (AAU



D.

```
                                                    *
consensus:                                        CUAAU          YYYYYYYYYYYYYYYYYNYAG/G
pMG1      CG/GTGAGTTTCTATTCGCAGTCGGCTGATCTGTGTGAAATCTTAATAAAGGGTCCAATTACCAATTTGAAACTCAG/TT
pMG31     CG/GTGAGTTTCTATTCGCAGTCGGCTGATCTGTGTGAAATCTTAATAAAGGGTCCAATTACCAATTTTTTCCTCAG/TT

pMG2      TTTATTTATTTATTAAGAAAGGAAATATAAATTATAAATTACAACATAAAGGGTCCAATTACCAATTTGAAACTCAG/TT
pMG32     TTTATTTATTTATTAAGAAAGGAAATATAAATTATAAATTACAACATAAAGGGTCCAATTACCAATTTTTTCCTCAG/TT
                                                  copia | white  intron
```

FIG. 1. Structures and partial nucleotide sequences of constructs. (A to C) Structures and expected RNA molecules produced during splicing of the pre-mRNAs. Structures of the wild-type and mutant *white* introns are indicated at the top of each panel. (A) Plasmids containing the second intron of the wild-type *white* gene (pMG1) or the same intron with a consensus pyrimidine stretch (pMG31). Synthetic pre-mRNA synthesized by using T7 RNA polymerase from templates truncated at the *Pvu*I site yields a 264-nucleotide precursor RNA. The first step in the splicing reaction should yield a 5′ exon fragment (E1) of 97 nucleotides and an intron-3′ exon (IVS-E2) lariat RNA fragment of 167 nucleotides. The second step should yield a 190-nucleotide mRNA (E1-E2) and the excised lariat intron (IVS) of 74 nucleotides. (B) Plasmids containing a copia LTR insertion (pMG2 and pMG32). Synthetic pre-mRNA synthesized by using T7 RNA polymerase from templates truncated at the *Pvu*I site yields a 540-nucleotide precursor RNA. The first step in the splicing reaction should yield a 5′ exon fragment (E1) of 97 nucleotides and an intron-3′ exon (IVS-E2) lariat RNA fragment of 443 nucleotides. The second step should yield a 190-nucleotide mRNA (E1-E2) and the excised lariat intron (IVS) of 350 nucleotides. (C) Plasmid containing the *ftz* intron (pGEM2 V61 S/B). The plasmid and sizes are as described by Rio (66). (D) Partial nucleotide sequences of constructs pMG1, pMG31, pMG2, and pMG32. The 5′ splice site and 3′ splice site are denoted by slashes. Consensus sequences for the branchpoint (27, 45) and 3′ splice site (45, 74) are shown on the top line. The site of the branch nucleotide within that consensus is indicated by an asterisk. Plasmids pMG31 and pMG32 differ from pMG1 and pMG2 by a substitution mutation that increases pyrimidine content in the −20 to −5 region from 50 to 75%. The consensus branchpoint in wild-type introns is underlined. The sequences are aligned by their 3′ splice sites, and only portions of the large introns from pMG2 and pMG32 are shown. Sizes are indicated in nucleotides.

AAA-independent) polyadenylation activity in our nuclear extract was partially characterized (data not shown). Slowly migrating bands seen in Fig. 2B, 2C, and 3A are ATP dependent but disappear when reactions are carried out in the presence of an inhibitor of polyadenylation, 3′-dATP. Splicing reactions shown in all figures were performed without 3′-dATP. Although concentrations of 3′-dATP above 1 mM inhibit splicing (data not shown), the addition of 0.2 mM 3′-dATP eliminates the slowly migrating bands without significant effects on splicing efficiency, and analysis of the splicing of transcripts from pMG1, pMG31, pMG3, pMG33, pMG5, pMG35, pMG2, pMG32, pMG4, pMG34, pMG7, and pMG37 in combination with a titration of 3′-dATP between 0 and 0.5 mM confirms the major conclusions of this study (data not shown).

RNase T_1 digestion and primer extension experiments. Excised lariat intron (intervening sequence [IVS]) and the intron-exon 2 intermediate (E2-IVS) were gel purified following splicing of RNA from the wild-type intron pMG1 and pMG31 and analyzed with and without treatment with S100 extract (which contains debranching activity). RNase T_1 digestion was performed as described previously (63). RNase T_1 oligonucleotide products containing the branched and debranched nucleotide before and after S100 extract treatment were then analyzed on a 20% denaturing polyacrylamide gel.

Primer extension was performed by the method of Inoue and Cech (24). Oligonucleotide primers purified from 20% denaturing polyacrylamide gels were labeled with [γ-^{32}P] ATP. The primer used for extension on the lariat is a 15-mer (5′-AAATTGGTAATTGGA-3′) complementary to the region near the 3′ splice site of the intron. The primer used for extension on the lariat-E2 intermediate is a 32-mer (5′-GCA GGGTCGTCTTTCCGGCACCGGAACTGCCC-3′) complementary to a region of the 3′ exon 40 nucleotides downstream of the 3′ splice site. Twenty-five to 50% of each splicing reaction product and 10 ng of primer were used for each reaction.

RESULTS

The wild-type second intron of the *Drosophila white* gene is accurately removed in Kc cell nuclear extracts. As a first step toward the investigation of species specificity in splicing signals, an in vitro splicing system was established by using *Drosophila* Kc cell nuclear extracts (12, 66; see Materials and Methods). The *Drosophila ftz* intron was efficiently spliced in these extracts (e.g., Fig. 2C), as previously reported (66). Intron-containing transcripts with *white* second-intron sequences were made from the constructs pMG1, which contains the 74-nucleotide wild-type *white* second intron, and pMG2, which contains the same intron with a 276-nucleotide copia sLTR at the position of the w^a copia insertion. The structures, splicing pathways, and partial sequences of these constructs are shown in Fig. 1. All 18 *white* intron derivatives described in this study are flanked by the same 97-nucleotide 5′ exon (with 20 nucleotides of 5′ plasmid sequence) and 93-nucleotide 3′ exon. Therefore, they would all be expected to generate the same mRNA product (190 nucleotides) but different lariat introns (between 72 and 350 nucleotides). ^{32}P-labeled synthetic precursor RNAs corresponding to the wild-type *white* second intron (pMG1) were accurately spliced in this Kc cell nuclear extract, yielding the expected products in an ATP-dependent reaction (Fig. 2A, lanes 1 and 2). The products and intermediates expected from an accurate splicing of the *white* second intron are designated in Fig. 2A: the 5′ exon (E1), the intron-3′ exon (E2-IVS), the lariat intron (IVS), and the mRNA. In addition, the splicing products were incubated with a HeLa cell cytoplasmic S100 extract which contains a 2′-5′ phosphodiesterase activity that debranches the lariat to generate linear RNA (68). After treatment with debranching activity, the E2-IVS and IVS migrate as linear RNAs at the expected sizes of 167 and 74 nucleotides, respectively (Fig. 2B, lane 1). In the absence of debranching activity, the E2-IVS and IVS retain their lariat structure and migrate anomalously during gel electrophoresis (lane 2). The heterogeneous bands around 90 bp in lane 1 may be due to an exonucleolytic activity that has been shown to remove the intron lariat tail in vitro (68).

Together with these expected products, an unexpected ATP-dependent product of approximately 180 nucleotides was observed. Characterization by T_1 digestion and primer extension (data not shown) revealed that this RNA lacks sequences upstream of a site in the first exon near the 5′ splice site. Consideration of prior results in human cell extracts indicates that such an RNA is likely to be due to the activity of an exonuclease endogenous to the nuclear extract (52, 54). Accordingly, we have designated this band EPP, for exonuclease protection product.

T7 transcripts of pMG2, containing a 276-nucleotide copia sLTR insertion at the position of the w^a copia insertion, were tested in parallel with pMG1 transcripts in Kc nuclear extract. No splicing intermediates or products were detected from reactions carried out and analyzed in the same way as for pMG1 transcripts (Fig. 2C, lanes 1 and 2). Thus, we conclude that a copia sLTR insertion in the second intron of the *Drosophila white* gene eliminates in vitro splicing in Kc cell nuclear extracts.

A consensus pyrimidine stretch enhances, but is not essential for, the removal of the *white* second intron in *Drosophila* extracts. Like many short *Drosophila* introns (reference 45 and references therein), the wild-type *white* second intron does not have a consensus pyrimidine stretch. To investigate whether a conventional pyrimidine stretch would nevertheless enhance the splicing of this intron, mutations were made in each of these constructs to increase the pyrimidine content. A change of four consecutive purines to four consecutive pyrimidines at positions −9 through −6 increased the fraction of pyrimidines from 50 to 75% in the critical region between −5 to −20 relative to the 3′ splice site (Fig. 1D). In Kc cell nuclear extracts, this change enhanced wild-type second-intron splicing to a variable extent (Fig. 2A and B, pMG1 versus pMG31) but did not allow splicing of the sLTR-containing intron (Fig. 2C, lanes 1 to 4, pMG2 and pMG32).

All four of these constructs were also tested in extracts from two human cell lines. HeLa and 293 nuclear extracts gave identical results with *white*-derived substrates (Fig. 2C and data not shown), but results with these human cell extracts differ from those obtained for the *Drosophila* nuclear extract. Neither of the 74-nucleotide introns (pMG1 or pMG31) was excised (data not shown, but see Fig. 5), but the 350-nucleotide sLTR-containing intron with a consensus pyrimidine tract (pMG32) was efficiently spliced, and the lariat forms E2-IVS and IVS migrated at the correct linear size after treatment with debranching activity (Fig. 2C, lanes 9 and 10). In these human extracts, a pyrimidine stretch appears to be absolutely required; no splicing of pre-mRNA from pMG2 was observed (lanes 7 and 8). Although these results are in sharp contrast to those obtained with the Kc cell nuclear extract, they are in good agreement with previous results from human cell extracts, which have indicated that both a minimum intron size of between 66 and 80 nucleotides and a good pyrimidine tract are required for efficient splicing (14, 79). They also show that the inability of our *Drosophila* extracts to process this intron was not due to a general defect in the RNA preparation. Transcripts containing the *ftz* intron (66) were used as a positive control and were spliced with similar efficiencies in *Drosophila* Kc and human HeLa or 293 cell nuclear extracts (Fig. 2C, lanes 5, 6, 11, and 12; also data not shown).

The branchpoint of the wild-type *white* second intron is at nucleotide −32, immediately adjacent to the site of the w^a copia insertion. To understand the molecular basis of the inefficient splicing of the sLTR-containing introns (pMG2 and pMG32) in *Drosophila* Kc cell nuclear extracts, we sought to determine whether the insertion of the copia element had disrupted any of the normal splicing signals required for removal of the wild-type second *white* intron. The insertion is relatively far from both splice sites (48 nucleotides from the 5′ splice site and 31 nucleotides from the 3′ splice site). However, the sequence UUAAU, which is 32 nucleotides upstream of the 3′ splice site and is disrupted by the copia insertion, is an excellent candidate for the branchpoint sequence (27, 45, 50). Thus, alteration of the natural branchpoint sequence appeared to be a likely explanation for the inefficient expression of w^a in vivo and the lack of splicing of transcripts from the sLTR-containing introns in vitro. We therefore determined the branchpoint used by the second intron of the wild-type *white* allele in vitro.

The branchpoint of the second *white* intron was first localized to the 14-nucleotide RNase T_1 product AAAUC UUAAUAAAG of pMG1 lariat intron, which contains the sequence UUAAU (data not shown). Identical results were obtained with the corresponding products from pMG31 (data not shown). Primer extension experiments (24) were then used to precisely localize the branchpoint. A 15-mer complementary to the region near the 3′ splice site of the intron and a 32-mer complementary to a region of the 3′ exon 40 nucleotides downstream of the 3′ splice site gave identical results. Figure 2D shows results obtained with the 32-mer on

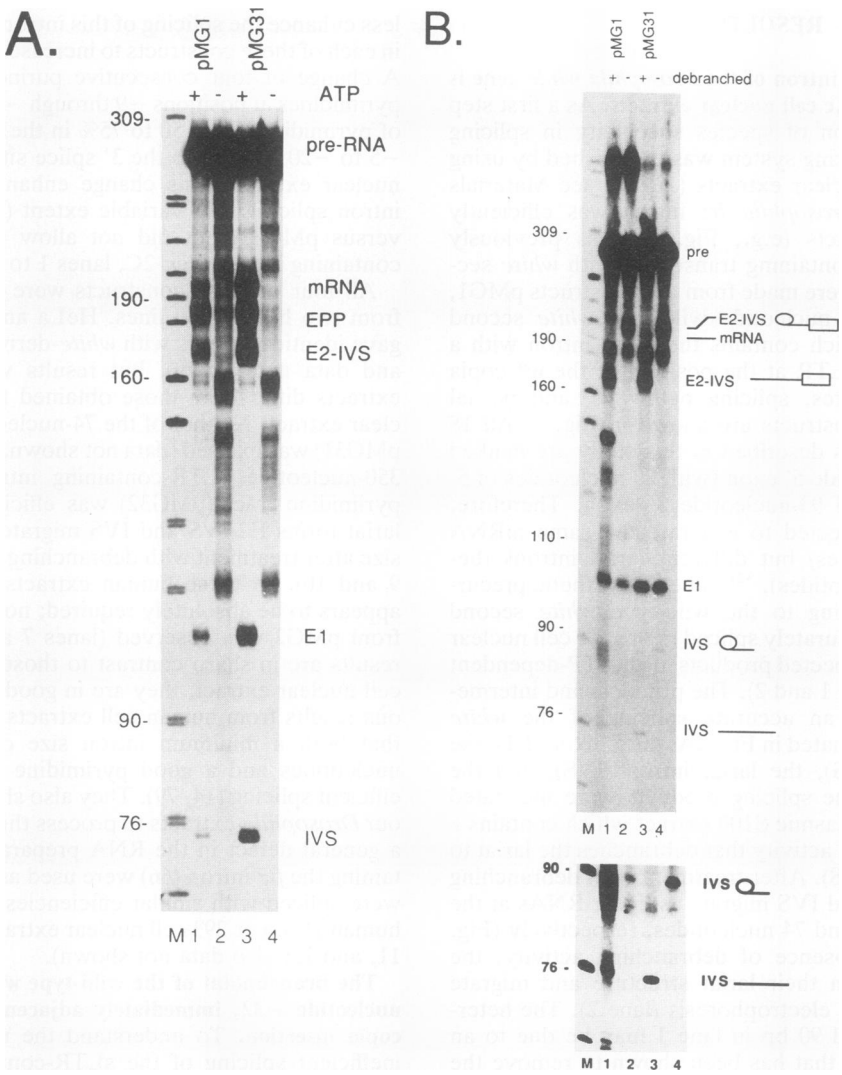
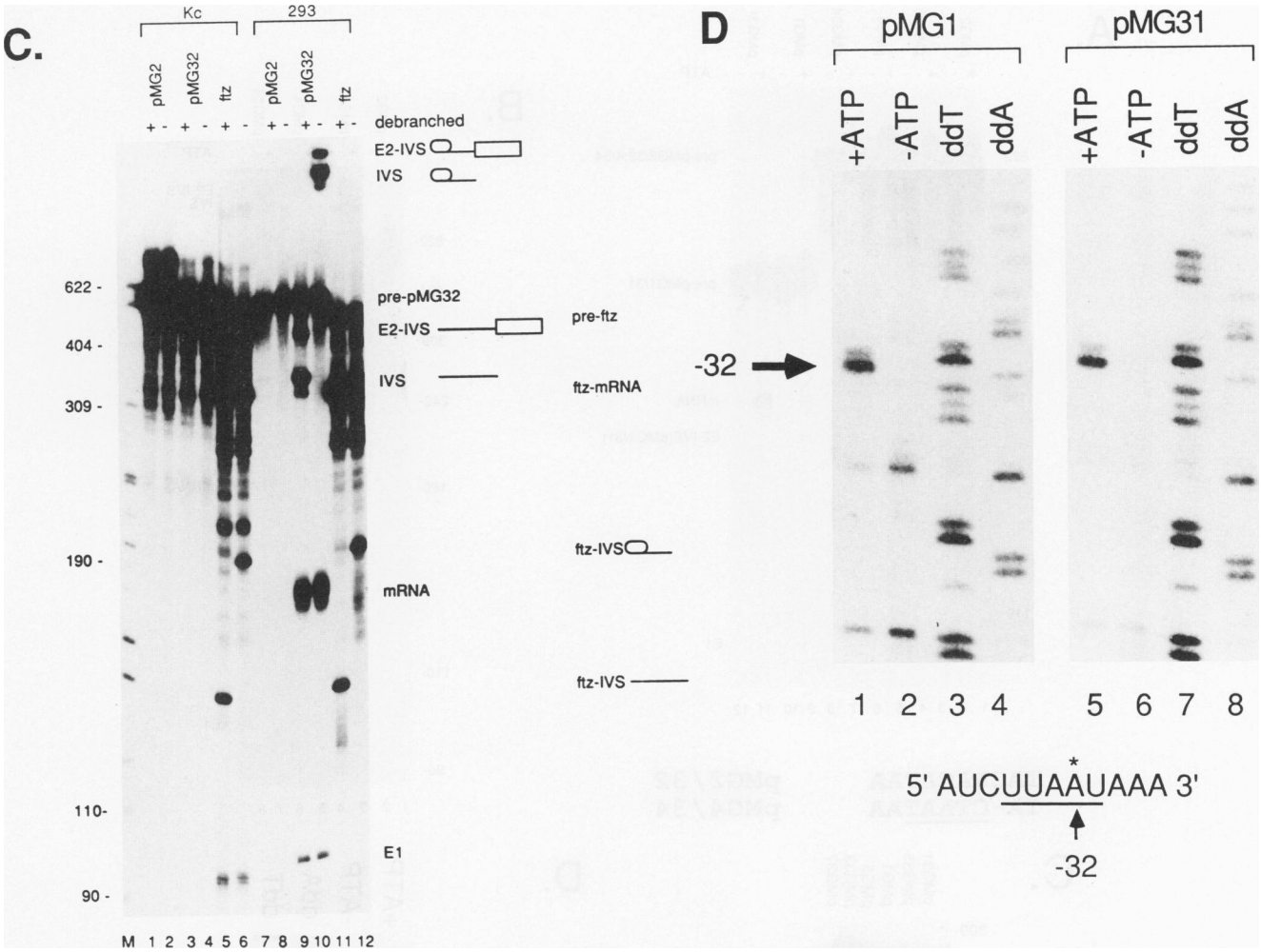


FIG. 2. In vitro splicing and characterization of *white* second introns. (A) In vitro splicing of the wild-type intron in *Drosophila* Kc cell nuclear extracts. ^{32}P-labeled pMG1 and pMG31 precursors were synthesized and incubated in the presence (+) or absence (−) of ATP. Splicing products were analyzed on 6% polyacrylamide gel. Lane M, ^{32}P-labeled *Msp*I digest of pBR322. Molecular sizes (in nucleotides) in all panels are indicated on the left. The positions of substrate, intermediates, and products are shown on the right. (Abbreviations are as described in the legend to Fig. 1 except for EPP [exonuclease protection product]. See text for explanation.) (B) Debranching assay. Splicing products were treated with (+) or without (−) S100 debranching activity and analyzed on 10% polyacrylamide gel. The structures of substrate, intermediates, and products are shown on the right. Lariat E2-IVS (lanes 2 and 4) runs close to (just above) mRNA (190 nucleotides) on this 10% polyacrylamide gel. The unshifted RNA between mRNA and lariat E2-IVS is the exonuclease protection product (see text). After treatment with debranching activity (lanes 1 and 3), the linear E2-IVS runs at the expected size of 167 nucleotides. Slowly migrating bands visible in this gel are ATP dependent (see Fig. 3A and B) and are apparently due to nonspecific polyadenylation activity present in the *Drosophila* extracts (see Materials and Methods). The inset at the bottom shows a longer exposure of the lower portion of the gel. (C) In vitro splicing of sLTR-containing introns. pMG2 and pMG32 precursors were synthesized, incubated in Kc (lanes 1 to 4) or 293 (lanes 7 to 10) nuclear extracts, and debranched in parallel with pMG1 and pMG31 (B) and *ftz* positive controls (lanes 5, 6, 11, and 12). The structures of substrate, intermediates, and products of pMG32 and *ftz* are shown on the right, inside and outside, respectively. (D) Primer extension mapping of branchpoints. RNAs from in vitro splicing reactions in the presence (+) or absence (−) of ATP were analyzed by primer extension using a primer complementary to the 3′ exon, 40 nucleotides downstream of the 3′ splice site (see Materials and Methods). Primer extension products corresponding to the branched nucleotide are indicated by the arrow. Lanes 3, 4, 7, and 8, markers from dideoxy sequencing reactions performed using the same primer with precursor RNA as the template. A partial sequence of the *white* second intron is shown at the bottom. The position of the branchpoint, which is at nucleotide −32, is indicated (*). Sequence with similarity to the *Drosophila* branch site consensus (27, 45) is underlined.

the E2-IVS intermediates from both pMG1 and pMG31. The precise location of the branchpoint was found to be position −32 in both cases. Thus, our suspicion that the copia insertion disrupted the natural branchpoint sequence was confirmed.

Nevertheless, human nuclear extracts were able to splice the sLTR-containing transcript from pMG32. This observation could be explained either by use of an alternate branchpoint in these extracts or by an insensitivity to alteration of the branchpoint sequence. Analysis of the branchpoint used

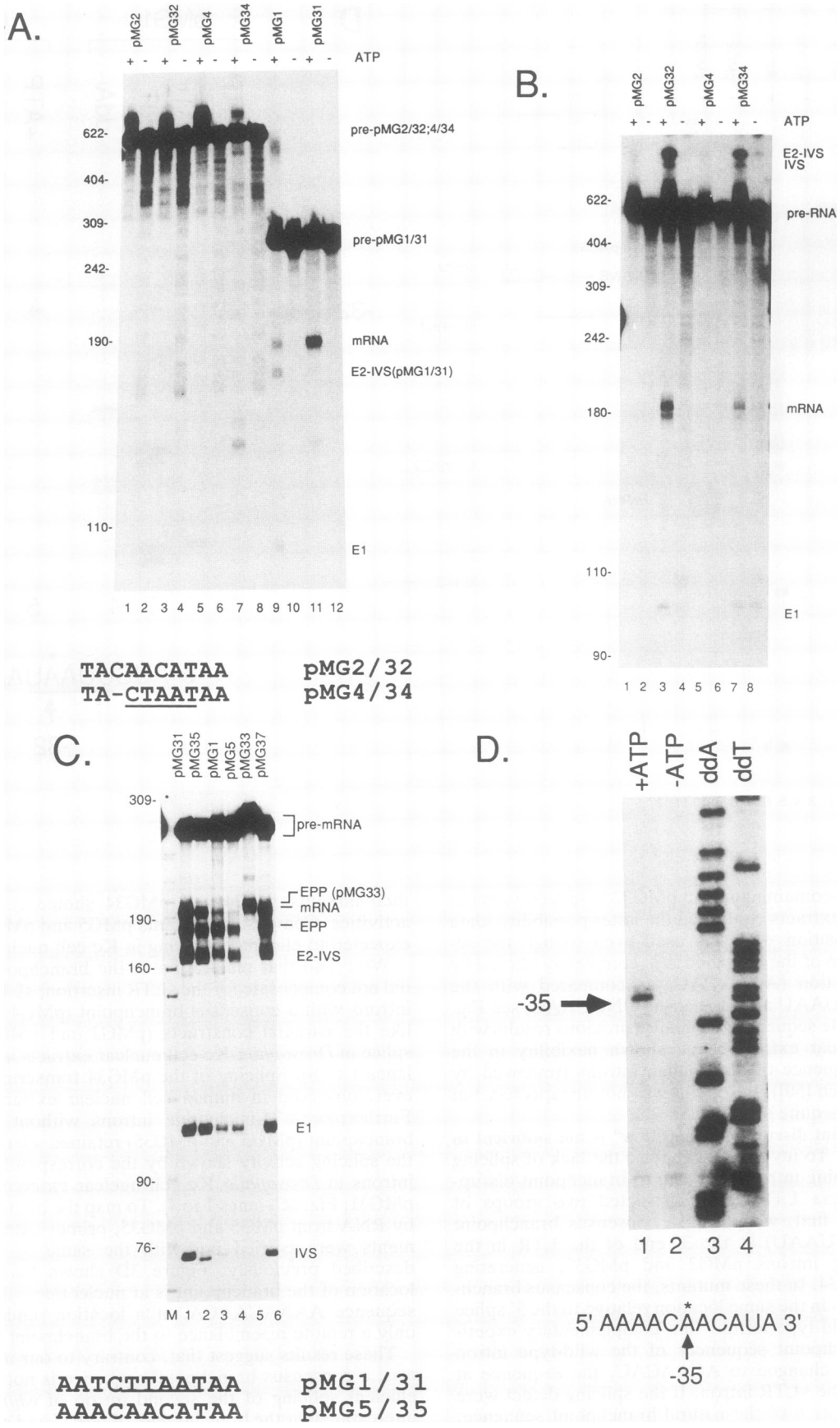


when the sLTR-containing intron pMG32 is spliced in mammalian 293 cell extracts confirmed the latter possibility (data not shown); position −32 was used even though the sequence upstream of the branched A is completely altered by the sLTR insertion (ACAACAU, as compared with the wild-type UCUUAAU and consensus UNCURAC; see Fig. 1 for the complete sequences). Because previous results with HeLa cell nuclear extracts have shown flexibility in the branchpoint sequence of mammalian introns (reviewed by Nelson and Green [50]), this result was not unexpected, but it is nevertheless quite striking.

The branchpoint disruption found in *w*^{a} is not sufficient to prevent splicing. To investigate whether the lack of splicing in sLTR-containing introns was due to branchpoint disruption by the copia LTR, we constructed two groups of mutants. In the first, we placed a consensus branchpoint sequence (UACUAAU) at the 3′ end of the LTR in the sLTR-containing introns pMG2 and pMG32, generating pMG4 and pMG34. In these mutants, the consensus branchpoint sequence is in the same location relative to the 3′ splice site as in the wild-type intron. In a complementary experiment, the branchpoint sequences of the wild-type intron constructs were changed to ACAACAU, the sequence at this position in the sLTR intron. If the splicing defect were due to the disruption of the natural branchpoint sequence, then mutants pMG4 and pMG34 should restore splicing activities. Conversely, mutants pMG5 and pMG35 would be expected to eliminate splicing in Kc cell nuclear extracts.

We found that restoration of the branchpoint consensus did not compensate for the sLTR insertion; sLTR-containing introns with a consensus branchpoint (pMG4 and pMG34), like the parental constructs (pMG2 and pMG32), did not splice in *Drosophila* Kc cell nuclear extracts at all (Fig. 3A, lanes 1 to 8). Splicing of the pMG34 transcripts was, however, observed in human cell nuclear extracts (Fig. 3B). Furthermore, 74-nucleotide introns without a consensus branchpoint (pMG5 and pMG35) retained a large fraction of the splicing activity shown by the corresponding wild-type introns in *Drosophila* Kc cell nuclear extracts (pMG1 and pMG31; Fig. 3C, lanes 1 to 4). To map the branchpoints used by RNA from pMG5 and pMG35, primer extension experiments were carried out with the same 3′ exon 32-mer described previously. Figure 3D shows that the precise location of the branchpoint is at nucleotide −35, within the sequence AAAACAACA, at a location (underlined) with only a remote resemblance to the branchpoint consensus.

These results suggest that, contrary to our initial hypothesis, a consensus branchpoint sequence is not essential for efficient splicing of the second intron of *white*. It is also interesting that the branchpoints of pMG5 and pMG35 are at


A.
pMG2
pMG32
pMG4
pMG34
pMG1
pMG31
ATP
622-
404-
309-
242-
190-
110-
pre-pMG2/32;4/34
pre-pMG1/31
mRNA
E2-IVS(pMG1/31)
E1
1 2 3 4 5 6 7 8 9 10 11 12
TACAACATAA pMG2/32
TA-CTAATAA pMG4/34
B.
pMG2
pMG32
pMG4
pMG34
ATP
E2-IVS
IVS
622-
404-
309-
242-
180-
110-
90-
pre-RNA
mRNA
E1
1 2 3 4 5 6 7 8
C.
pMG31
pMG35
pMG1
pMG5
pMG33
pMG37
309-
190-
160-
90-
76-
pre-mRNA
EPP (pMG33)
mRNA
EPP
E2-IVS
E1
IVS
M 1 2 3 4 5 6
AATCTTAATAA pMG1/31
AAACAACATAA pMG5/35
D.
+ATP
-ATP
ddA
ddT
-35
1 2 3 4
5' AAAACAACAUA 3'
-35

a location different from those observed when the sLTR-containing intron (pMG32) is spliced in mammalian extracts (AACAU; data not shown).

Increasing the size of the *white* second intron eliminates splicing in *Drosophila* Kc cell nuclear extracts but improves splicing in mammalian HeLa cell nuclear extracts. Because the size of the wild-type second *white* intron (74 nucleotides) is close to the minimum size which can be spliced in HeLa cell nuclear extracts, we were originally concerned that the small intron size might also prevent splicing in our *Drosophila* in vitro system. Therefore, we made two additional mutants, pMG3 and pMG33, by inserting a 16-bp polylinker into pMG1 and pMG31, respectively, at a point between the 5′ splice site and the branchpoint, 29 bp downstream of the 5′ splice site, resulting in 90-nucleotide introns (Fig. 4A). To our great surprise, transcripts from pMG3 and pMG33 were not spliced in Kc cell nuclear extracts, although wild-type introns from pMG1 and pMG31 spliced well in parallel reactions (Fig. 4B, lanes 1, 3, 4, and 6). In contrast, the pMG33 intron was a good substrate for human extracts (data not shown and Fig. 5B).

To distinguish the influence of intron size from the possibility that the 16-bp polylinker insertion in pMG3 and pMG33 prevented splicing by disrupting a previously unrecognized sequence element, two substitution mutants were made in which a 16-bp sequence between the 5′ splice site and the branchpoint was replaced by the same 16-bp polylinker, resulting in altered 74-nucleotide substitution introns (pMG7 and pMG37; Fig. 4A). If the linker insertion altered a previously unrecognized splicing signal, then these two mutants would be expected to be defective for splicing, but if the linker insertion prevented splicing by lengthening the intron, then transcripts from pMG7 and pMG37 should splice with efficiencies similar to those of the parental 74-nucleotide introns pMG1 and pMG31. Figure 4B shows that these substituted transcripts are indeed spliced in *Drosophila* Kc cell nuclear extracts with efficiencies comparable to those of the corresponding wild-type introns (lanes 1, 2, 4, and 5). Like the wild-type introns, these substituted introns are not spliced in HeLa cell nuclear extracts (data not shown).

Thus, both 90-nucleotide introns containing an innocuous linker insertion (pMG3/33) and 350-nucleotide introns bearing an insertion of the copia LTR (pMG2/32) were defective for splicing in *Drosophila* Kc cell nuclear extracts. In each case, the sequence change per se did not eliminate splicing in control constructs of wild-type length (pMG7/37 and pMG5/35). In each case, the elongated intron, if provided with a pyrimidine-rich region, was an effective substrate for splicing in human cell nuclear extracts (pMG33 and pMG32). Consideration of these results led us to the hypothesis that intron size is a critical factor in the splicing of these introns in our *Drosophila* nuclear extracts.

To test this idea, a series of deletion mutations around the polylinker insertion region of pMG3 and pMG33 was constructed (Fig. 5C). Each of the resulting mutants contains a different distance between the 5′ splice site and branchpoint. If size is the critical factor, mutants with smaller introns should be able to restore splicing activity in Kc cell nuclear extracts.

The results obtained with three deletions from each of pMG3 and pMG33 are shown in Fig. 5A. As expected, splicing efficiency increases with decreasing intron size in Kc cell nuclear extracts. The 78- and 79-nucleotide introns, pMG3Δ14, pMG3Δ28, pMG33Δ14, and pMG33Δ28 (lanes 3, 4, 8, and 9), showed splicing activity close to that of the corresponding wild-type introns pMG1 and pMG31 (lane 5 and 10). Note that although the exonuclease protection product generated from some substrates of intermediate length obscures the mRNA band, the exon 1 intermediate and intron products are better isolated on the gel and serve as excellent, and equally valid, indicators of splicing efficiency. The 84-nucleotide introns from pMG3Δ6 and pMG33Δ23 showed very little splicing activity (lanes 2 and 7) but were better substrates than their 90-nucleotide parental introns, which showed almost no splicing activity at all (lanes 1 and 6). In agreement with earlier results, a consensus pyrimidine stretch confers greater splicing efficiency in all cases (lanes 6 to 10 versus lanes 1 to 5). All of these 10 constructs were also tested in HeLa cell nuclear extracts (Fig. 5B). In agreement with previous studies using mammalian cell nuclear extracts, but in contrast to our results for *Drosophila* Kc cell nuclear extracts, we found that splicing efficiency decreases with decreasing intron size. Derivatives of pMG3 lack a consensus pyrimidine stretch and did not show any splicing activity in HeLa cell extracts (lanes 1 to 5). As observed earlier, the 90-nucleotide pMG33 intron was spliced in HeLa nuclear extracts. Deletion of the intron in construct pMG33 to less than 84 nucleotides resulted in decreased splicing efficiencies in HeLa nuclear extracts (Fig. 5B, lanes 6 to 10). These results suggest that removal of the *white* second intron is exquisitely sensitive to intron length in both *Drosophila* cell nuclear extracts and human cell nuclear extracts but that the relationship between length and efficiency is species specific.

DISCUSSION

Species-specific splicing signals. We have examined the effects on in vitro splicing of sequence alterations in the 74-nucleotide second intron of the *Drosophila white* gene. This intron differs from mammalian introns in two ways that

FIG. 3. Evidence that mutations in the branchpoint consensus do not control *white* second intron splicing. (A) Splicing of sLTR-containing introns with and without a consensus branchpoint in *Drosophila* Kc nuclear extracts. Constructs pMG4 and pMG34 encode sLTR-containing introns with a consensus branchpoint sequence and were made from pMG2 and pMG32. The sequences of the branchpoint regions are shown underneath. Splicing reactions were performed and analyzed as described in the legend to Fig. 2A. The positions of precursors and splicing products of pMG1 and pMG31 are shown on the right. (B) Splicing of pMG4 and pMG34 in HeLa nuclear extracts. (C) Splicing of introns with and without a consensus branchpoint in *Drosophila* Kc nuclear extracts. Constructs pMG5 and pMG35 carry 74-nucleotide introns lacking a consensus branchpoint sequence and were made from pMG1 and pMG31. The sequences of the branchpoint regions are shown underneath. Splicing reactions were performed and analyzed as described in the legend to Fig. 2A. The exonuclease protection product of pMG33 runs with or slightly slower than mRNA (190 nucleotides), and the best estimate of splicing efficiency is obtained from noting the quantity of IVS product or intermediates (E1 and E2-IVS). (D) Primer extension mapping of the branchpoint in pMG35 RNA. Primer extension experiments were performed as described in the legend to Fig. 2D. Primer extension products corresponding to the branched nucleotide are indicated by the arrow. Lanes 3 and 4, markers. Partial sequence of the intron is shown underneath. The position of the branchpoint, which is at nucleotide −35, is indicated (*). Sizes are indicated in nucleotides.

are characteristic of a large subset of *Drosophila* introns (45). It is shorter than the length of approximately 80 nucleotides required for efficient splicing in mammalian cells (68, 83), and it lacks a tract of pyrimidines adjacent to the 3′ splice site. In results summarized in Fig. 6, we have confirmed that this *Drosophila* intron is not spliced by extracts from human cells, but becomes a substrate for splicing in human extracts if these two sequence features are modified to correspond to the requirements of human cells. Thus, all introns tested that were longer than 80 nucleotides and had a good pyrimidine stretch (12 pyrimidines among 15 nucleotides in positions −5 to −19) were efficiently spliced in human cell extracts. This set of introns includes the LTR-containing introns pMG32 (Fig. 2C, lanes 9 and 10, and Fig. 3B, lane 3, show the pMG32 intron spliced in extracts from 293 cells and HeLa cells, respectively) and pMG34 (Fig. 3B, lane 7), as well as introns enlarged to 90 or 84 nucleotides by the insertion of a linker (Fig. 5B, lanes 6 and 7, pMG33 and pMG33Δ23). Although not present in the wild-type *Drosophila* intron, a pyrimidine tract is absolutely required for these introns to be recognized by human extracts; no intron with the wild-type *Drosophila* sequence (GAAA) at −9 to −6 relative to the 3′ splice site showed detectable splicing (Fig. 2C, lanes 7 and 8 versus 9 and 10; Fig. 3B, lanes 1 versus 3 and 5 versus 7; Fig. 5B, lanes 1 to 5 versus 6 to 10). Likewise, introns shorter than 80 nucleotides (pMG31, pMG33Δ14, and pMG33Δ28 [Fig. 5B] and pMG37 [data not shown]) are not spliced well in HeLa cell nuclear extracts. Thus, our results confirm previous reports on the sequence features required in nuclear extracts derived from human cells and indicate that the ability of this intron to be properly spliced in *Drosophila* extracts is indeed due to species specificity in the recognition of splicing signals rather than to some unrecognized feature of this intron.

Indeed, extracts derived from *Drosophila* Kc cells behave very differently from those derived from human cells (summarized in Fig. 6). Most significantly, the wild-type intron is spliced efficiently in *Drosophila* extracts, despite its short length and the absence of a pyrimidine tract. To explore this species specificity, which was consistently observed in five independent preparations of Kc cell nuclear extract, and to characterize the sequence requirements for the splicing of this intron in *Drosophila* nuclear extracts, mutations in the pyrimidine stretch, in the branchpoint, and in the overall length were analyzed. Each of these three sequence features makes a significant contribution to splicing efficiency in *Drosophila* extracts, but *Drosophila* extracts differ from human extracts in that the effect of the pyrimidine stretch is quantitative rather than absolute. When the ratio of pMG31 to pMG1 splicing produced by five different extract preparations was measured by quantitation of the excised intron lariat, measurements varied between 7.73 and 18.04, with a mean of 12.2 and a standard deviation of 3.6 (data not shown). Examples are visible in Fig. 2A, 2B, 3C (lanes 1 versus 3 and 2 versus 4), 4B (lanes 1 to 3 versus 4 to 6), and 5A (lanes 1 to 5 versus 6 to 10). A strong preference for small, rather than large, introns was also observed in *Drosophila* extracts. This critical dependence of splicing efficiency on length is made clear by data from the deletion

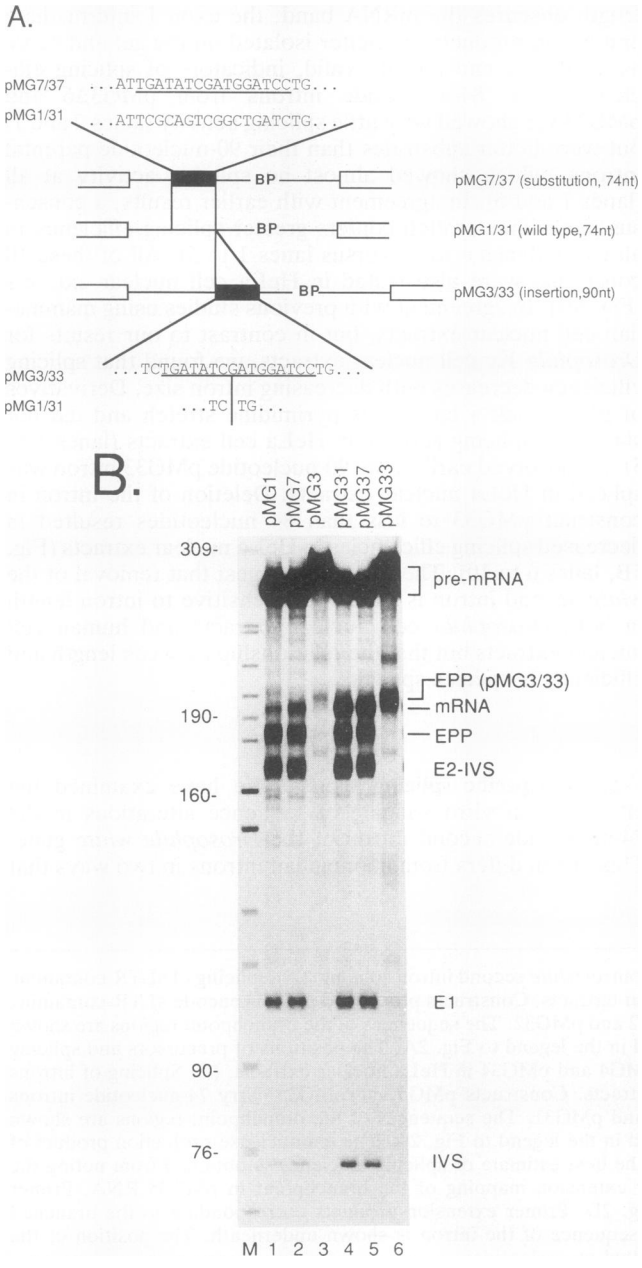
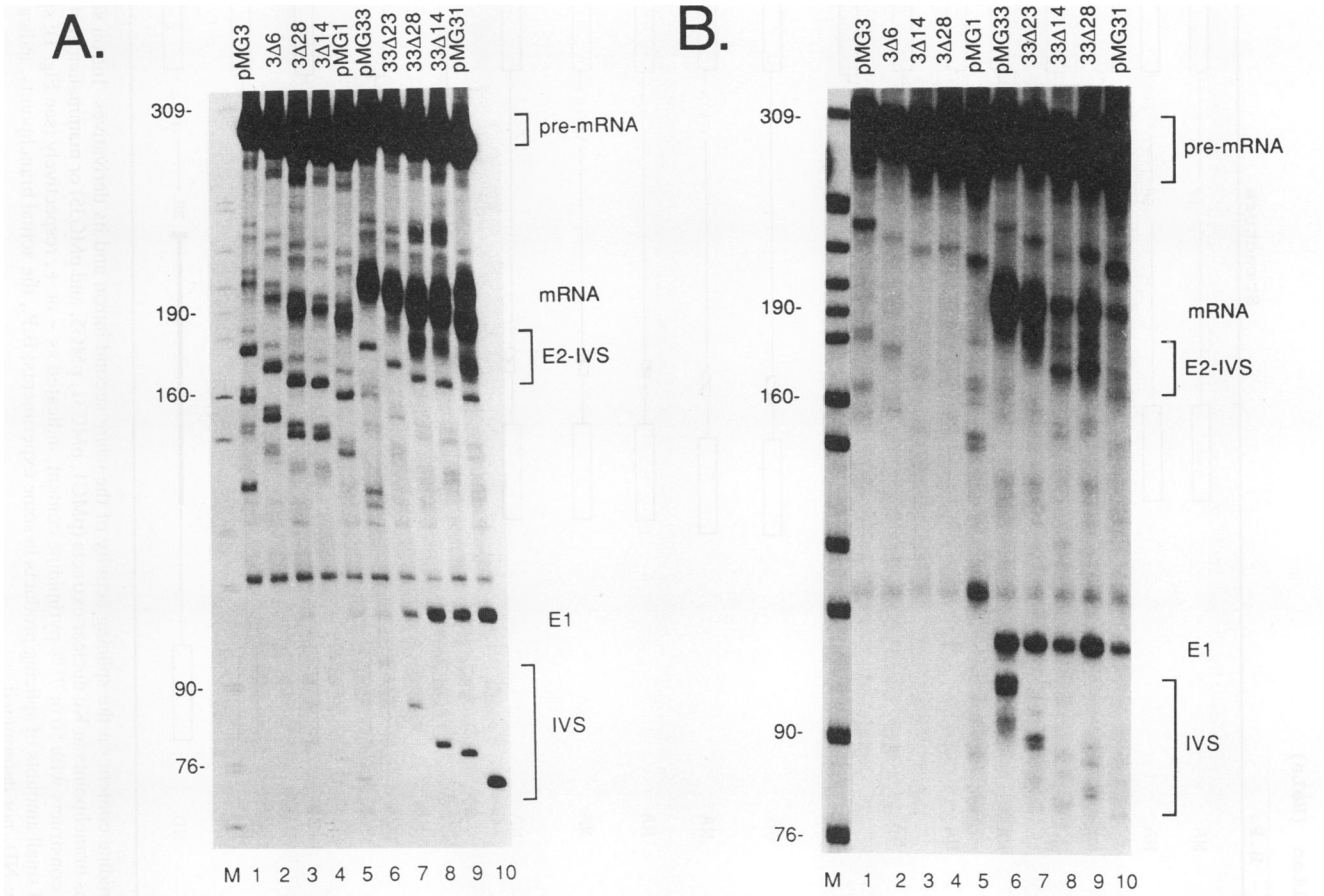


FIG. 4. Splicing of introns carrying a 16-nucleotide insertion in *Drosophila* Kc cell nuclear extracts. (A) Schematic diagram of insertion construct pMG3/33 and the control (substitution) construct pMG7/37. pMG3 and pMG33 contain 90-nucleotide (nt) introns which were made from pMG1 and pMG31 by a 16-nucleotide insertion (shown by a black box) at a point between the 5′ splice site and the branchpoint, 29 nucleotides downstream of the 5′ splice site. Constructs pMG7 and pMG37 contain 74-nucleotide introns which were made from pMG1 and pMG31 by a substitution of the same 16 nucleotides (shown by a black box) for sequences in the region between the 5′ splice site and the branchpoint, starting at 13 nucleotides downstream of the 5′ splice site. Exons are shown by open boxes. The intron is shown by a line. The sequence of the 16-nucleotide insertion/substitution is underlined, and the position of the insertion in pMG1/31 is indicated by the vertical bar. (B) Splicing of pMG3, pMG33, pMG7, and pMG37 transcripts in *Drosophila* Kc nuclear extracts. Splicing reactions were performed and analyzed as described in the legend to Fig. 2A. The exonuclease protection products from pMG3 and pMG33 run with or slightly slower than mRNA (190 nucleotides), and the best estimate of splicing efficiency is obtained from the quantity of IVS or intermediates. Sizes are indicated in nucleotides.



C.

constructs	intron size	sequence
		EcoRV/ClaI/BamH
pMG3/33	90	CG/GTGAGTTTCTATTCGCAGTCGGCTGATC[TGATATCGATGGATCC]TGTGTGAAATCTTAATAA----------TCAG/TT
Δ6/Δ23	84	CG/GTGAGTTTCTATTCGCAGTCGGCTGATC[T------GATGGATCC]TGTGTGAAATCTTAATAA----------TCAG/TT
Δ28/Δ28	79	CG/GTGAGTTTCTATTCGCAGTCGGCTGATC[TGAT----------C]TGTGTGAAATCTTAATAA----------TCAG/TT
Δ14/Δ14	78	CG/GTGAGTTTCTATTCGCAGTC--------[----ATCGATGGATCC]TGTGTGAAATCTTAATAA----------TCAG/TT
pMG1/31	74	CG/GTGAGTTTCTATTCGCAGTCGGCTGATC[----------]TGTGTGAAATCTTAATAA----------TCAG/TT

FIG. 5. Effect of intron size on splicing of *white* second introns in vitro. (A) Splicing of deletion mutants derived from pMG3 and pMG33 in *Drosophila* Kc cell nuclear extracts. Splicing reactions were performed and analyzed as described in the legend to Fig. 2A. (B) Splicing of the same set of deletion mutants in mammalian HeLa nuclear extracts. (C) Intron size (length in nucleotides) and partial nucleotide sequences of the introns in deletion mutants. Introns derived from pMG3 (lanes 1 to 5 in panels A and B) have the wild-type sequence at positions −9 to −6 upstream of the 3′ splice site, while introns derived from pMG33 (lanes 6 to 10 in panel A and B) have the pyrimidines at positions −9 to −6 upstream of the 3′ splice site (see Fig. 1D).

construct	intron size	B.P. seq.	(Py)n	Drosophila (Kc) spl.	Drosophila (Kc) B.P.	mammalian (HeLa) spl.	mammalian (HeLa) B.P.	Structure
pMG1	74	TTAAT	-	+	-32	-	NA	
pMG31	74	TTAAT	+	+	-32	+/-	NA	
pMG5	74	AACAA	-	+	-35	ND	ND	
pMG35	74	AACAA	+	+	-35	ND	ND	
pMG7	74	TTAAT	-	+	ND	-	NA	
pMG37	74	TTAAT	+	+	ND	+/-	NA	
pMG3	90	TTAAT	-	-	NA	-	NA	
pMG33	90	TTAAT	+	-	NA	+	-32	
pMG3Δ6	84	TTAAT	-	-	NA	-	NA	
pMG33Δ23	84	TTAAT	+	+/-	NA	+/-	NA	
pMG3Δ28	79	TTAAT	-	+/-	NA	-	NA	
pMG33Δ28	79	TTAAT	+	+	ND	+/-	NA	
pMG3Δ14	78	TTAAT	-	+/-	NA	-	NA	
pMG33Δ14	78	TTAAT	+	+	ND	+	ND	
pMG2	350	AACAT	-	-	NA	-	NA	
pMG32	350	AACAT	+	-	NA	+	-32	
pMG4	350	CTAAT	-	-	NA	-	NA	
pMG34	350	CTAAT	+	-	NA	+	ND	

FIG. 6. Summary of the effects of intron size, branchpoint sequence, and pyrimidine content on the splicing activity of the *white* second intron and its derivatives. Intron size, number of nucleotides in each intron; B.P. seq., sequences which are actually used as branchpoints in Kc nuclear extracts (pMG1, pMG31, pMG5, and pMG35) or mammalian nuclear extracts (pMG32 and pMG33) or the sequence at the corresponding location; (Py)n, constructs with 50 or 75% pyrimidine content, indicated as − or +, respectively (see Fig. 1); spl., splicing activities indicated by + for splicing or − for not splicing; +/−, detection of small amounts of splicing products in some experiments; B.P., the actual branchpoints, indicated by the distance between the branchpoint and the 3′ splice site; NA, not applicable; ND, not determined.

series derived from pMG3 and pMG33, which indicate a maximum length of between 79 and 84 nucleotides for efficient splicing of the second *white* intron in Kc cell nuclear extracts (Fig. 5).

In summary, the species specificity of splicing signals, particularly pyrimidine stretch and size requirements, implied by sequence differences between *Drosophila* and mammalian introns has now been demonstrated in vitro. We have also discovered an unexpected inability of *Drosophila* extracts to recognize elongated forms of this short intron.

This is not the first report of species-specific splicing in *Drosophila* versus human cell extracts. Previously, it was observed that substrates containing the regulated intron of the *Drosophila* P transposable element were spliced accurately in human extracts but not in *Drosophila* extracts (77). However, the P intron conforms well to the features characteristic of mammalian introns, and the species specificity observed is contrary to expectation (splicing occurs in the heterologous extract). These results are best attributed to the regulated splicing of the intron; the *Drosophila* extracts used were made from somatic cells, which do not normally splice the P-element third intron, and an inhibitory factor can also be isolated from human cells (80).

The molecular basis of the w^a mutation. We have duplicated the splicing defect observed in the w^a allele in vitro with introns containing an sLTR at the site of the copia insertion, and we have mapped the branchpoint of the second *white* intron to a location disrupted by the copia insertion in the w^a allele. However, when the branchpoint of an intron of wild-type length is altered to match the sequence alteration induced by the copia insertion, splicing is not abolished, either in vitro (Fig. 3C) or in vivo (data not shown). Instead, a nonconsensus branchpoint sequence is used (Fig. 3D). Thus, the branchpoint disruption per se cannot be held accountable for the splicing defect in w^a. Similarly, the provision of a consensus branchpoint to LTR-disrupted introns does not restore splicing in vitro (Fig. 3A). Like all introns derived from the second *white* intron and greater than 80 nucleotides in size, these LTR-containing introns yielded no spliced products in *Drosophila* extracts that carry out efficient removal of the wild-type intron in parallel (Fig. 2C and 3A). Thus, our observations are most consistent with the idea that introns carrying the copia LTR are defective for splicing in Kc cell nuclear extracts not because of their altered branchpoint sequence but because of their overall size. This effect is accentuated in the case of the full copia insertion w^a because of copia's polyadenylation activity (31).

Is there a specific mechanism for the splicing of small introns in *D. melanogaster*? It is well established that *Drosophila* and vertebrate splicing signals and mechanisms are generally similar. Consensus sequences for *Drosophila* splice sites are extremely similar to those found in mammals, and similar branchpoint sequences can also be found in *Drosophila* gene. Furthermore, identified components of the splicing machinery appear to be conserved. The collection of small RNAs that is known to function in mammalian splicing has been identified in *Drosophila* cells, and these RNAs are highly conserved in sequence (19, 39, 48), as are several proteins involved in splicing that have been identified in both *Drosophila* and human cells (20, 22, 23, 40, 42–44, 87). In addition, there are natural introns that are recognized by the splicing machinery of both species (63, 77) (Fig. 2C).

However, many *Drosophila* introns are very small relative to mammalian introns. Approximately half of all sequenced *Drosophila* introns are less than 80 nucleotides, with a modal length between 60 and 65 nucleotides (21, 45), which is smaller than the size of all but a few mammalian introns (21; reviewed in reference 14). An even more extreme situation exists in *Caenorhabditis elegans*, in which intron lengths of less than 50 nucleotides are common (9). Consistent with these observations, a *C. elegans* intron of 53 nucleotides was efficiently spliced in HeLa cell nuclear extracts only when expanded to 84 nucleotides (55).

The size dependence that we have observed for the naturally small *white* intron not only is species specific but also is specific to small introns. We have obtained similar results with the fifth intron of the myosin heavy-chain gene, in which case expansion from 68 to 84 nucleotides permitted accurate splicing in extracts from HeLa cells but prevented splicing in Kc cell extracts (data not shown). However, the *Drosophila ftz* intron, which is 150 nucleotides, is efficiently spliced in extracts from both *Drosophila* and human cells (Fig. 2C). This finding rules out the possibility that Kc extracts are generally unable to splice introns larger than 84 nucleotides but is consistent with the notion of an as yet unidentified independent signal for small introns. Consideration of all of the data (most of them summarized in Fig. 6) leads us to favor the hypothesis that the splicing of small introns, or a subset of small introns, in *Drosophila* cells differs in mechanistic detail from the splicing of mammalian introns and that *Drosophila* nuclear extracts are sensitive to this difference.

The effect of size may be more significant in vitro than in vivo. For example, although RNA transcripts from pMG2 and pMG32 are not spliced at all in *Drosophila* nuclear extracts, and detectable levels of intron-containing transcripts are observed in flies carrying the corresponding *white* allele with an LTR-containing intron (30a), these same flies do have significant levels of normally spliced RNA (46, 86). Similarly, analysis of the splicing of introns described in this study (pMG1, -31, -3, -33, -5, -35, -7, and -37) in transfected *Drosophila* Schneider cells (38a) indicates that lengthening the intron to 90 nucleotides decreases splicing efficiency roughly twofold, comparable in magnitude to the effects of branchpoint and pyrimidine tract alterations in the same study. Thus, it appears that although the length effect can be observed in vivo, it is less pronounced.

We consider three possibilities for size-specific splicing signals. One is a positive signal (such as a mammalian-style pyrimidine stretch) that operates in the splicing of long introns but is not required for short introns. In small introns, simple direct contact between a U1 snRNP at the 5′ splice and a U2 snRNP at the branchpoint might supersede the requirement for this signal. In this case, increasing the distance between the 5′ splice site and the branchpoint beyond 53 nucleotides (the distance in pMG3Δ6 and pMG33Δ23) might prevent splicing unless the additional signal were present. A positive signal specific for the splicing of small introns might also exist. This second hypothesis (which is not inconsistent with the first) is attractive in that it explains how pMG5, which lacks both a consensus branchpoint and pyrimidine stretch, might be recognized. The observation that pMG5 is spliced better than pMG33 would imply that such a small intron-specific positive signal plays a greater role in *Drosophila* extracts than does either the branchpoint sequence or the pyrimidine stretch. Finally, our results are consistent with a negative signal that inhibits the pairing of one of the two splice sites of a short intron with distant partners.

Having shown that a size-dependent signal acts in the splicing of this intron but not the *ftz* control intron, we are

now in a position to localize such a signal and to identify the factor or factors responsible for its activity. The existence of *trans*-acting genetic modifiers of w^a, at least some of which may act by overcoming the size limitations of this splicing event, should prove useful to this approach. In any case, if there is indeed a distinct mechanism for the splicing of small introns in species such as *D. melanogaster* and *C. elegans*, the elucidation of that mechanism is likely to reveal aspects of splicing common to all species.

ACKNOWLEDGMENTS

We thank Don Rio and Chris Siebel for providing plasmid pGEM2 V61 S/B and for technical advice; Hui Ge for an introduction to preparation of S100 and nuclear extracts and for technical advice; Mathew Wang for HeLa and 293 cells; and Zhenqiang Pan, James Manley, Frank Laski, Daniel Kalderon, and Jym Mohler for valuable suggestions and critical comments on the manuscript.

This work was supported by Public Health Service grant GM 37991 from the National Institute of General Medical Sciences, by a National Science Foundation Presidential Young Investigator award to S.M.M., and by Basil O'Connor Starter Scholar research award 5-630 from the March of Dimes Birth Defects Foundation.

REFERENCES

1. **Adami, G., and J. R. Nevins.** 1988. Splicing site selection dominates over poly(A) choice in RNA production from complex adenovirus transcription units. EMBO J. **7**:2107–2116.
2. **Barabino, S. L., B. J. Blencowe, U. Ryder, B. S. Sproat, and A. I. Lamond.** 1990. Targeted snRNP depletion reveals an additional role for mammalian U1 snRNP in spliceosome assembly. Cell **63**:293–302.
3. **Beggs, J. D., J. V. D. Berg, A. V. Ooyen, and C. Weissman.** 1980. Abnormal expression of a chromosomal rabbit β-globin gene in *Saccharomyces cerevisiae*. Nature (London) **283**:835–840.
4. **Bell, L. R., E. M. Maine, P. Schedl, and T. W. Cline.** 1988. *Sex-lethal*, a *Drosophila* sex determination switch gene, exhibits sex-specific RNA splicing and sequence similarity to RNA binding proteins. Cell **55**:1037–1046.
5. **Bingham, P. M., and B. H. Judd.** 1981. A copy of the copia transposable element is very tightly linked to the w^a allele at the *white* locus of *D. melanogaster*. Cell **25**:705–711.
6. **Birchler, J. A., and J. C. Hiebert.** 1989. Interaction of the *Enhancer-of-white-apricot* with transposable element alleles at the *white* locus in *Drosophila melanogaster*. Genetics **122**:129–138.
7. **Birchler, J. A., J. C. Hiebert, and L. Rabinow.** 1989. Interaction of the *mottler-of-white* with transposable element alleles at the *white* locus in *Drosophila melanogaster*. Genes Dev. **3**:73–84.
8. **Black, D. L., B. Chabot, and J. A. Steitz.** 1985. U2 as well as U1 small nuclear ribonucleoproteins are involved in pre-messenger RNA splicing. Cell **42**:737–750.
9. **Blumenthal, T., and J. Thomas.** 1988. Cis and trans mRNA splicing in C. elegans. Trends Genet. **4**:305–308.
10. **Boggs, R. T., P. Gregor, S. Idriss, J. Belote, and M. McKeown.** 1987. Regulation of sexual differentiation in *D. melanogaster* via alternative splicing of RNA from the *transformer* gene. Cell **50**:739–747.
11. **Chou, T., Z. Zachar, and P. M. Bingham.** 1987. Developmental expression of a regulatory gene is programmed at the level of splicing. EMBO J. **7**:4095–4104.
12. **Dignam, J. D., R. M. Lebovitz, and R. G. Roeder.** 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res. **11**:1475.
13. **Frendeway, D., and W. Keller.** 1985. The stepwise assembly of a pre-mRNA splicing complex requires U-snRNPs and specific intron sequences. Cell **42**:355–367.
14. **Ge, H., J. Noble, J. Colgan, and J. L. Manley.** 1990. Polyoma virus small tumor antigen pre-mRNA splicing requires cooperation between two 3′ splice sites. Proc. Natl. Acad. Sci. USA **87**:3338–3342.
15. **Gehring, W. J., and R. Paro.** 1980. Isolation of a hybrid plasmid with homologous sequences to a transposing element of *Drosophila*. Cell **19**:897–904.
16. **Goodall, G. J., and W. Filopowicz.** 1989. The AU-rich sequences present in the introns of plant nuclear pre-mRNAs are required for splicing. Cell **58**:473–483.
17. **Goodall, G. J., and W. Filopowicz.** 1991. Different effects of intron nucleotide composition and secondary structure on pre-mRNA splicing in monocot and dicot plants. EMBO J. **10**:2635–2644.
18. **Green, M. R.** 1991. Biochemical mechanisms of constitutive and regulated pre-mRNA splicing. Annu. Rev. Cell Biol. **7**:559–600.
19. **Guthrie, C., and B. Patterson.** 1988. Spliceosomal snRNAs. Annu. Rev. Genet. **22**:387–419.
20. **Harper, D. S., L. D. Fresco, and J. D. Keene.** 1992. RNA binding specificity of a *Drosophila* snRNP protein that shares homology with mammalian U1-A and U2-B″ proteins. Nucleic Acids Res. **20**:3645–3650.
21. **Hawkins, J. D.** 1988. A survey on intron and exon lengths. Nucleic Acids Res. **16**:9893–9905.
22. **Haynes, S. R., D. Johnson, G. Raychaudhuri, and A. L. Beyer.** 1991. The *Drosophila* Hrb87F gene encodes a new member of the A and B hnRNP proteins group. Nucleic Acids Res. **19**:25–31.
23. **Haynes, S. R., G. Raychaudhuri, and A. L. Beyer.** 1990. The *Drosophila* Hrb98DE locus encodes four protein isoforms homologous to the A1 protein of mammalian heterogeneous nuclear ribonucleoprotein complexes. Mol. Cell. Biol. **10**:316–323.
24. **Inoue, T., and T. R. Cech.** 1985. Secondary structure of the circular form of the *Tetrahymena* rRNA intervening sequence: a technique for RNA structure analysis using chemical probes and reverse transcriptase. Proc. Natl. Acad. Sci. USA **82**:648–652.
25. **Jacquier, A., J. R. Rodriguez, and M. Rosbash.** 1985. A quantitative analysis of the effects of 5′ junction and TACTAAC box mutants and mutant combinations on yeast mRNA splicing. Cell **43**:423–430.
26. **Keller, E. B., and W. A. Noon.** 1984. Intron splicing: a conserved internal signal in introns of animal pre-mRNAs. Proc. Natl. Acad. Sci. USA **81**:7417–7420.
27. **Keller, E. B., and W. A. Noon.** 1985. Intron splicing: a conserved internal signal in introns of *Drosophila* pre-mRNAs. Nucleic Acids Res. **13**:4971–4981.
28. **Konarska, M. M., P. J. Grabowski, R. A. Padgett, and P. A. Sharp.** 1985. Characterization of the branch site in lariat RNAs produced by splicing of mRNA precursors. Nature (London) **313**:552–557.
29. **Krainer, A. R., T. Maniatis, B. Ruskin, and M. R. Green.** 1984. Normal and mutant human β-globin pre-mRNAs are faithfully and efficiently spliced *in vitro*. Cell **36**:993–1005.
30. **Kunkel, T. A., J. D. Roberts, and R. A. Zakour.** 1987. Rapid and efficient site-specific mutagenesis without phenotypic selection. Methods Enzymol. **154**:367–382.

30a. **Kurkulos, M.** Unpublished data.

31. **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.
32. **Langford, C. J., and D. Gallwitz.** 1983. Evidence for an intron-contained sequence required for the splicing of yeast RNA polymerase II transcripts. Cell **33**:7–19.
33. **Leff, S. E., R. M. Evans, and M. G. Rosenfeld.** 1987. Splice commitment dictates neuron-specific alternative RNA processing in calcitonin/CGRP gene expression. Cell **48**:517–524.
34. **Legrain, P., B. Séraphin, and M. Rosbash.** 1988. Early commitment of yeast pre-mRNA to the spliceosome pathway does not require U2 small nuclear ribonucleoprotein. Mol. Cell. Biol. **8**:3755–3760.
35. **Levis, R., P. M. Bingham, and G. M. Rubin.** 1982. Physical map of the *white* locus of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **79**:564–568.
36. **Levis, R., K. O'Hare, and G. M. Rubin.** 1984. Effects of transposable element insertions on RNA encoded by the *white* gene of *Drosophila melanogaster*. Cell **38**:471–481.

37. **Levitt, N., D. Briggs, A. Gil, and N. J. Proudfoot.** 1989. Definition of an efficient synthetic poly(A) site. Genes Dev. **3:**1019–1025.
38. **Lindsley, D. L., and G. G. Zimm.** 1992. The genome of *Drosophila melanogaster*. Academic Press, New York.
38a. **Lo, P. C. H.** Unpublished data.
39. **Lo, P. C. H., and S. M. Mount.** 1990. *Drosophila melanogaster* genes for U1 snRNA variants and their expression during development. Nucleic Acids Res. **18:**6971–6979.
40. **Mancebo, R., P. C. H. Lo, and S. M. Mount.** 1990. Structure and expression of the *Drosophila melanogaster* gene for the U1 small nuclear ribonucleoprotein particle 70K protein. Mol. Cell. Biol. **10:**2492–2502.
41. **Mattox, W., and B. S. Baker.** 1991. Autoregulation of the splicing of transcripts from the *transformer-2* gene of *Drosophila*. Genes Dev. **5:**786–796.
42. **Matunis, E. L., M. J. Matunis, and G. Dreyfuss.** 1992. Characterization of the major hnRNP proteins from *Drosophila melanogaster*. J. Cell Biol. **116:**257–269.
43. **Matunis, M. J., E. L. Matunis, and G. Dreyfuss.** 1992. Isolation of hnRNP complexes from *Drosophila melanogaster*. J. Cell Biol. **116:**245–255.
44. **Mayeda, A., A. M. Zahler, A. R. Krainer, and M. B. Roth.** 1992. Two members of a conserved family of nuclear phosphoproteins are involved in general and alternative splicing. Proc. Natl. Acad. Sci. USA **89:**1301–1304.
45. **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.
46. **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.
47. **Mount, S. M., I. Petterson, M. Hinterberger, A. Karmas, and J. A. Steitz.** 1983. The U1 small nuclear RNA-protein complex selectively binds a 5′ splice site in vitro. Cell **33:**509–518.
48. **Mount, S. M., and J. A. Steitz.** 1981. Sequence of U1 RNA from Drosophila melanogaster: implications for U1 secondary structure and possible involvement in splicing. Nucleic Acids Res. **9:**6351–6368.
49. **Nagoshi, R. N., M. McKeown, K. C. Burtis, J. M. Belote, and B. S. Baker.** 1988. The control of alternative splicing at genes regulating sexual differentiation in *D. melanogaster*. Cell **53:** 229–236.
50. **Nelson, K. K., and M. R. Green.** 1989. Mammalian U2 snRNP has a sequence-specific RNA-binding activity. Genes Dev. **3:**1562–1571.
51. **Newman, A. J., R.-J. Lin, S.-C. Cheng, and J. Abelson.** 1985. Molecular consequences of specific intron mutations on yeast mRNA splicing in vivo and in vitro. Cell **42:**335–344.
52. **Noble, J. C. S., H. Ge, M. Chaudhuri, and J. L. Manley.** 1989. Factor interactions with the simian virus 40 early pre-mRNA influence branch site selection and alternative splicing. Mol. Cell. Biol. **9:**2007–2017.
53. **Noble, J. C. S., Z. Pan, C. Prives, and J. L. Manley.** 1987. Splicing of SV40 early pre-mRNA to large T and small t mRNA utilizes different patterns of lariat branch sites. Cell **50:**227–236.
54. **Noble, J. C. S., C. Prives, and J. L. Manley.** 1986. In vitro splicing of simian virus 40 early pre-mRNA. Nucleic Acids Res. **14:**1219–1235.
55. **Ogg, S. C., P. Anderson, and M. P. Wickens.** 1990. Splicing of a C. elegans myosin pre-mRNA in a human nuclear extract. Nucleic Acids Res. **18:**143–149.
56. **O'Hare, K., C. Murphy, R. Levis, and G. M. Rubin.** 1984. DNA sequence of the *white* locus of *Drosophila melanogaster*. J. Mol. Biol. **180:**437–455.
57. **Parker, R., and C. Guthrie.** 1985. A point mutation in the conserved hexanucleotide at a yeast 5′ splice junction uncouples recognition, cleavage and ligation. Cell **41:**107–118.
58. **Parker, R., P. G. Siliciano, and C. Guthrie.** 1987. Recognition of the TACTAAC box during mRNA splicing in yeast involves base-pairing to the U2-like snRNA. Cell **49:**229–239.
59. **Peng, X., and S. M. Mount.** 1990. Characterization of *Enhancer-of-white-apricot* in *Drosophila melanogaster*. Genetics **136:** 1061–1069.
60. **Pepling, M. E., and S. M. Mount.** 1990. Sequence of a cDNA from the *Drosophila melanogaster white* gene. Nucleic Acids Res. **18:**1633.
61. **Pirrotta, V., and C. Brockl.** 1984. Transcription of the *Drosophila white* locus and some of its mutants. EMBO J. **3:**563–568.
62. **Rabinow, L., and J. A. Birchler.** 1989. A dosage-sensitive modifier of the retrotransposon-induced alleles of the *Drosophila white* locus. EMBO J. **8:**879–889.
63. **Reed, R., and T. Maniatis.** 1985. Intron sequences involved in lariat formation during pre-mRNA splicing. Cell **41:**95–105.
64. **Reed, R., and T. Maniatis.** 1986. A role for exon sequences and splice site proximity in splice site selection. Cell **46:**681–690.
65. **Reed, R., and T. Maniatis.** 1988. The role of mammalian branchpoint sequences in pre-mRNA splicing. Genes Dev. **2:**1268–1276.
66. **Rio, D. C.** 1988. Accurate and efficient pre-mRNA splicing in *Drosophila* cell-free extracts. Proc. Natl. Acad. Sci. USA **85:**2904–2909.
67. **Ruby, S. W., and J. Abelson.** 1988. An early hierarchic role of U1 small nuclear ribonucleoprotein in spliceosome assembly. Science **242:**79–85.
68. **Ruskin, B., J. M. Greene, and M. R. Green.** 1985. Cryptic branch point activation allows accurate in vitro splicing of human β-globin intron mutants. Cell **52:**207–219.
69. **Ruskin, B., A. R. Krainer, T. Maniatis, and M. R. Green.** 1984. Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. Cell **38:**317–331.
70. **Ruskin, B., P. D. Zamore, and M. R. Green.** 1988. A factor, U2AF, is required for U2 snRNP binding and splicing complex assembly. Cell **52:**207–219.
71. **Saiki, R. K., D. H. Gelfand, S. Stoffel, S. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis, and H. A. Ehrlich.** 1988. Primer-directed enzymatic amplification of DNA with a thermostable polymerase. Science **239:**487–491.
72. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
73. **Sanger, F., S. Nicklen, and A. R. Coulsen.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
74. **Senapathy, P., M. B. Shapiro, and N. L. Harris.** 1990. Splice junctions, branch point sites, and exons: sequence statistics, identification, and applications to the human genome project. Methods Enzymol. **183:**252–278.
75. **Séraphin, B., L. Kretzner, and M. Rosbash.** 1988. A U1 snRNA: pre-mRNA base pairing interaction is required early in yeast spliceosome assembly but does not uniques define the 5′ splice site. EMBO J. **7:**2533–2538.
76. **Séraphin, B., and M. Rosbash.** 1989. Mutational analysis of the interactions between U1 small nuclear RNA and pre-mRNA of yeast. Gene **82:**145–151.
77. **Siebel, C. W., and D. C. Rio.** 1990. Regulated splicing of the Drosophila P transposable element third intron in vitro: somatic repression. Science **248:**1200–1208.
78. **Siliciano, P. G., and C. Guthrie.** 1988. 5′ splice site selection in yeast: genetic alterations in base-pairing with U1 reveal additional requirements. Genes Dev. **2:**1258–1267.
79. **Smith, C. W. J., J. G. Patton, and B. Nadal-Ginard.** 1989. Alternative splicing in the control of gene expression. Annu. Rev. Genet. **23:**527–577.
80. **Tseng, J. C., S. Zollman, A. C. Chain, and F. A. Laski.** 1991. Splicing of the *Drosophila* P element ORF2-ORF3 intron is inhibited in a human cell extract. Mech. Dev. **35:**65–72.
81. **von Halle, E. S.** 1969. Pursuing the *Enhancer-of-white-apricot*. *Drosophila* Inf. Serv. **44:**119.
82. **Weibauer, K., J.-J. Herrero, and W. Filopowicz.** 1988. Nuclear pre-mRNA processing in plants: distinct modes of 3′ splice site selection in plants and animals. Mol. Cell. Biol. **8:**2042–2051.
83. **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.

84. **Wu, J., and J. L. Manley.** 1989. Mammalian pre-mRNA branch site selection by U2 snRNP involves base pairing. Genes Dev. **3**:1553–1561.
85. **Zachar, Z., T. B. Chou, and P. M. Bingham.** 1987. Evidence that a regulatory gene autoregulates splicing of its transcript. EMBO J. **6**:4105–4111.
86. **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.
87. **Zamore, P. D., and M. R. Green.** 1991. Biochemical characterization of U2 snRNP auxiliary factor: an essential pre-mRNA splicing factor with a novel intramolecular distribution. EMBO J. **10**:207–214.
88. **Zeitlin, S., and A. Efstratiatis.** 1984. In vivo splicing products of the rabbit β-globin gene. Cell **39**:589–602.
89. **Zhuang, Y., A. M. Goldstein, and A. M. Weiner.** 1989. UACU AAC is the preferred branch site for mammalian mRNA splicing. Proc. Natl. Acad. Sci. USA **86**:2752–2756.
90. **Zhuang, Y., and A. M. Weiner.** 1986. A compensatory base change in U1 snRNA suppresses a 5′ splice site mutation. Cell **46**:827–835.
91. **Zhuang, Y., and A. M. Weiner.** 1989. A compensatory base change in human U2 snRNA can suppress a branch site mutation. Genes Dev. **3**:1545–1552.