Functionally antagonistic sequences are required for normal autoregulation of *Drosophila tra-2* pre-mRNA splicing

Dawn S. Chandler, M. Elaine McGuffin and William Mattox*

Department of Molecular Genetics, University of Texas, MD Anderson Cancer Center, 1515 Holcombe Boulevard, Box 45, Houston, TX 77030-4009, USA

ABSTRACT

Expression of functional TRA-2 protein in the male germline of *Drosophila* is regulated through a negative feedback mechanism in which a specific TRA-2 isoform represses splicing of the M1 intron in the TRA-2 pre-mRNA. We have previously shown that the mechanism of M1 splicing repression is conserved between distantly related *Drosophila* species. Using transgenic fly strains, we have examined the effects on regulation of mutations in two conserved features of the M1 intron. Our results show that TRA-2-dependent repression of M1 splicing depends on the presence of a suboptimal non-consensus 3′ splice site. Substitution of this 3′ splice site with a strong splice site resulted in TRA-2 independent splicing, while substitution with an unrelated weak 3′ splice site was compatible with repression, implying that reduced basal splicing efficiency is important for regulation. A second conserved element internal to the intron was found to be essential for efficient M1 splicing in the soma where the intron is not normally retained. We show that the role of this element is to enhance splicing and overcome the reduction in efficiency caused by the intron's suboptimal 3′ splice site. Our results indicate that antagonistic elements in the M1 intron act together to establish a context that is permissive for repression of splicing by TRA-2 while allowing efficient splicing in the absence of a repressor.

INTRODUCTION

A number of RNA binding proteins have now been identified that regulate the cell-specific splicing of specific pre-mRNAs (1,2). Several of these factors also autoregulate their own expression by directing alternative splicing of the pre-mRNAs that encode them (3–7). For example, the splicing regulator TRA-2 controls the splicing of pre-mRNAs of genes such as *doublesex* (*dsx*) and *fruitless* (*fru*) (8–14) as well as that from the *tra-2* gene itself (4). Autoregulation of splicing in this case serves to limit the expression of a specific isoform of TRA-2

(TRA-2^{226}) that is required in male germ cells for normal spermatogenesis (15). Negative feedback has an important biological role because expression of excessive levels of TRA-2^{226} in the germline results in formation of abnormal spermatids and sterility in males (16).

The focal point of *tra-2* autoregulation is the M1 intron which splits the initiation codon of the TRA-2^{226} open reading frame. M1 splicing is repressed by the TRA-2^{226} protein resulting in the formation of an alternative mRNA encoding TRA-2^{179}, a protein isoform initiating translation at a downstream start codon (15,17,18). Unlike TRA-2^{226}, expression of TRA-2^{179} is neither sufficient nor required in the germline for spermatogenesis. In flies homozygous for point mutations blocking expression of functional TRA-2^{226}, the M1 intron is efficiently spliced and M1-containing RNAs are not detected in the germline. In contrast, in wild-type males expressing this protein the M1 intron is retained in >50% of mature germline *tra-2* mRNA. Thus repression of M1 splicing serves to limit the amount of TRA-2^{226}-encoding mRNA that is produced (16).

In addition to germ cells, TRA-2 mRNAs are produced in a wide variety of somatic tissues in both sexes. Although a low level of M1 retention is observed in the soma, accumulation of somatic M1-containing RNA is not dependent on functional TRA-2 protein (4). The basis of the germline specificity of TRA-2 autoregulation is unknown, but it is unlikely that TRA2^{226} is a limiting factor as significant levels of mRNAs encoding this isoform have been shown to be present both in the soma and germline.

The mechanism responsible for TRA-2-dependent repression of M1 splicing in male germ cells is likely to differ significantly from the well characterized mechanism by which TRA-2 activates female-specific splicing of *dsx* RNA (19–25). The latter activity occurs only in the soma rather than the germline, requires a cofactor (TRA) that is not involved in repression of M1 splicing and can be functionally separated from the repression of M1 splicing by specific mutations in the *tra-2* gene (26). Moreover, in *dsx* splicing, TRA-2^{226} and other TRA-2 isoforms associate with the *dsx* pre-mRNA and enhance splice site recognition by recruiting general splicing factors, but in the case of the M1 intron, TRA-2 protein acts to repress splicing.

We have taken an evolutionary approach to identify *cis*-elements affecting M1 splicing. In a previous study we showed that the negative feedback strategy used by the

*To whom correspondence should be addressed. Tel: +1 713 792 2538; Fax +1 713 794 4394; Email: wmattox@mdanderson.org

Drosophila melanogaster tra-2 gene is conserved in *Drosophila virilis*, a species that diverged >60 million years ago (27). Although the intron sequences of these two species are highly divergent, two short conserved sequence features were identified within M1. In this study we have investigated the possible regulatory roles of these sequences. Our results show that these sequences have opposing activities that are necessary for the correct tissue-specific repression of M1 splicing.

MATERIALS AND METHODS

Construction of transgenes

P[*tra2/ftz*], in which the *ftz* intron is substituted for the M1 intron of *tra-2*, was introduced into the *Drosophila* genome using the plasmid pCSP*tra2/ftz*IVS. To generate this plasmid, primer-directed mutagenesis of sequences flanking both introns was performed to introduce *Cla*I sites. After *Cla*I digestion, the complete *ftz* intron was then ligated in place of the M1 intron in a 1.4 kb Apa-H3 genomic fragment of the *tra-2* gene contained in the plasmid pSK Apa-H3 DAClа to produce the plasmid pSK Apa-H3-*tra2/ftz*. A fragment containing the substitution and flanking regions was then inserted into a 3.8 kb genomic *tra-2* fragment in the plasmid pTZ3.9BΔSac. The entire insert fragment was then transferred to the P element transformation vector pCaSpeR4 (28) to generate pCSP*tra2/ftz*IVS. Ligation junctions and mutations were confirmed by DNA sequencing. The 5′ splice site of the hybrid gene is <u>ATCGATG</u>/gtaggcat and at the 3′ splice site is cuuacag/<u>ATCGAT</u> where the lower case bases represent *ftz* intron sequences and underlined bases encode the *Cla*I restriction site.

Other transgenes were produced by primer-directed mutagenesis of the plasmid P[*tra2⁺1*] and then insertion of the altered fragments into pCaSpeR (16). All P[*tra2⁺1*] derived constructs contain a 7 nt insertion in exon 4 that produces a translation frameshift, blocking production of functional TRA-2 protein. The 7 nt insertion also makes it possible to distinguish transcripts derived from the wild-type and transgene RNAs in RT–PCR experiments. Sequences of oligonucleotides used in primer-directed mutagenesis are as follows: tra2-VCsub, 5′-TAAGAACCCACCACCTGGATA-TCACCTATTCTACGTCCA-3′; tra2-VCdel, 5′-TGGACGT-AGAATTCGGTTCTTACA-3′; tra2-MHCE3′, 5′-GTCGA-CCTTGTTTGCAAGGGGATAAGTTCAAATTATCTTGCG-3′; tra2-CAG3′, 5′-GTCAGACCTGGATTCGATG-3′ and tra2-ftz3′, 5′-GTAGTCAGACCTGTAAGCATAAGCAAAA-GAAAAAAAATAAGATTATGTTG-3′.

All constructs were injected directly into *Drosophila* embryos of the genotype *w¹¹¹⁸/BˢY; tra-2ᴮ/CyO* in combination with the plasmid pΔ2-3, which provided transposase. Individuals bearing stable P insertions were identified in the G1 generation by changes in eye pigmentation due to the *mini-w⁺* gene contained in pCaSpeR.

Preparation of *D.melanogaster* RNA

In the range of 50–100 flies were homogenized using a Tekmar Tissuemizer in 2 ml phenol combined with 2 ml 2× NETS (200 mM NaCl, 20 mM EDTA, 20 mM Tris pH 8.0, 0.2%

SDS). Phases were separated by centrifugation and the aqueous phase was collected. After 2–3 additional phenol extractions, the RNA was precipitated with 2 vol ethanol. The RNA was resuspended in RNase-free water and quantified by absorbance. Following treatment with DNase at 37°C for 1 h samples were heat-inactivated for 10 min at 70°C, phenol extracted and precipitated with 2 vol ethanol. Poly(A) RNA was isolated by binding to oligo(dT)-cellulose.

Detection of alternatively spliced RNAs

In most experiments the dependence of M1 splicing on *tra-2* was determined by RT–PCR performed in parallel on RNA from adult flies with a functional *tra-2⁺* allele (*tra-2ᴮ/CyO*) and homozygous *tra-2* mutants (*tra-2ᴮ/tra-2ᴮ*). First strand cDNA synthesis was performed on poly(A) RNA or total RNA using the SuperScript Preamplification system (Gibco BRL). Aliquots of 0.8 μg poly(A) RNA or 5 μg total RNA with 0.5 μg Oligo(dT)$_{12-18}$ (New England Biolabs) were denatured at 70°C for 10 min in a final volume of 12 μl and incubated on ice for 1 min. The subsequent annealing, cDNA synthesis, reaction termination, and removal of RNA steps were performed according to manufacturers instructions. An aliquot of 7 μl first strand cDNA was amplified directly using *Taq* polymerase and the following primers: X3S, 5′-CTCAGCCGATTCAGCTG-GTGCTCTTG-3′; X5/X6, 5′-CGCTGTGTTTGTGCGTCAA-TCA-3′ and X2S, 5′-GACAACTTCATCCGGACATAAG-3′.

PCR reactions were carried out in a final volume of 100 μl with 10 mM Tris–HCl pH 8.3, 50 mM potassium chloride, 1.5 mM magnesium chloride, by cycling 20 times at 94°C for 1 min, followed by 60°C for 30 s for primer pair X3S and X5/X6 or 55°C for primer pair X2S and X5/X6, and 72°C for 30 s. Cycling at these temperatures was followed by a single final incubation at 72°C for 5 min. In all experiments the primers were positioned so that PCR products derived from M1-containing RNAs could be readily distinguished from products derived from contaminating genomic DNA based on gel mobility. The above PCR conditions produced ratios of endogenous spliced and unspliced M1 intron consistent with those determined in nuclease protection experiments and were therefore judged to be representative of the relative ratios of transgenic products as well.

PCR products were separated on a 1.2% agarose gel and blotted overnight in 1.5 M sodium chloride, 0.1 M sodium citrate. The *tra-2* transcripts were detected after blotting with a ³²P-end-labeled oligonucleotide specific for either endogenous *tra-2* (Nae-WT, 5′-TGGTGCCGGCGACTG-3′) or transgene-derived RNAs (Nae+1, 5′-GACCGCTGGTGCCGATCCCCG-3′). These oligonucleotides are complementary to sequences at the position of the 7 nt insertion placed in exon 4. Blots were hybridized with the Nae-WT probe at 42°C in 1 M sodium chloride, 0.1 M PIPES pH 7.0, 0.2% BSA, 0.2% Ficoll, 0.2% polyvinylpyrrolidone, 10% dextran sulfate for 12–16 h and washed at 58°C with 6× SSC, 0.1% SDS. In experiments with the Nae +1 probe, incubations with probe were carried out at 60°C, for 12–16 h and washed at 68°C in 6× SSC, 0.1% SDS. Northern blots were performed as previously described (27) using 2 μg adult poly(A) RNA.

RESULTS

### Sequences in the M1 intron are required for splicing repression

To determine if sequences from the M1 intron are necessary for splicing repression we substituted the entire M1 intron with the 150 nt intron from the *Drosophila ftz* gene which is not subject to *tra-2*-dependent regulation in its normal context. A restriction fragment containing the entire *tra-2/ftz* hybrid gene as well as the *tra-2* promoter and its other flanking sequences was then introduced into the P element transformation vector pCaSper4 to produce the plasmid pCSP*tra-2/ftz*IVS (Fig. 1A). In these and all subsequent constructs we included a 7 nt insertion in exon 4 that served both to block translation of TRA-2[226] protein from the transgene and as a means of distinguishing the transgene-derived RNAs from those of the endogenous *tra-2* gene (Fig. 1A). Transcripts from the transgene were specifically detected by RT–PCR using an oligonucleotide (Nae+1) that includes sequences homologous to the 7 nt insertions in exon 4.

Fly strains containing random insertions of the *P[tra-2/ftz]* transgene were produced by microinjection of *tra-2[B]/CyO* embryos. The *tra-2[B]* allele is a nonsense mutation in the region encoding the TRA-2 RNA binding domain that eliminates all autoregulatory activity but produces normal amounts of mutant *tra-2* mRNA. Both homozygous *tra-2[B]* mutants and normal heterozygous *tra-2[B]/CyO* individuals were sorted from transgenic strains and used to isolate RNA.

Transcripts produced from the *P[tra-2/ftz]* transgene were efficiently spliced and no *ftz* intron-containing RNA was detected in either *tra-2[B]/CyO* males (Fig. 1B, lanes 7 and 8) or their *tra-2[B]* homozygous mutant siblings (Fig. 1B, lanes 5 and 6). Control experiments confirmed that the M1 intron containing RNA from the endogenous *tra-2* gene accumulated normally (Fig. 1C). M1-containing *tra-2* transcripts were readily detected in transgenic males with a wild-type endogenous *tra-2* allele (Fig. 1C, lane 4) but not in flies from *tra-2[B]* homozygous males (Fig. 1C, lane 3). Together these results show that *tra-2/ftz*IVS transcripts are not subject to splicing repression by products from the endogenous *tra-2* gene. We conclude from this that sequences within the M1 intron are necessary for splicing repression.

A suboptimal 3′ splice site is essential for repression of M1 splicing *in vivo*

Examination of the sequence of the M1 3′ splice site region revealed that it differs significantly in several respects from the *Drosophila* 3′ splice site consensus. An adenosine is present at the –3 position rather than the consensus pyrimidine that is found in ~96% of all *Drosophila* introns (29). In addition, the 3′ splice site was not preceded by the usual long polypyrimidine tract, but instead only by several short runs of pyrimidines each <5 nt in length (Fig. 2A). Finally, there was no good match to the *Drosophila* consensus branchpoint in the distal region of the intron (not shown). We previously found that the 3′ splice site region of the M1 intron of the *D.virilis* tra-2 gene shares these characteristics and that this intron is also repressed by TRA-2 (27). This led us to hypothesize that these atypical sequences are suboptimally recognized by the general splicing
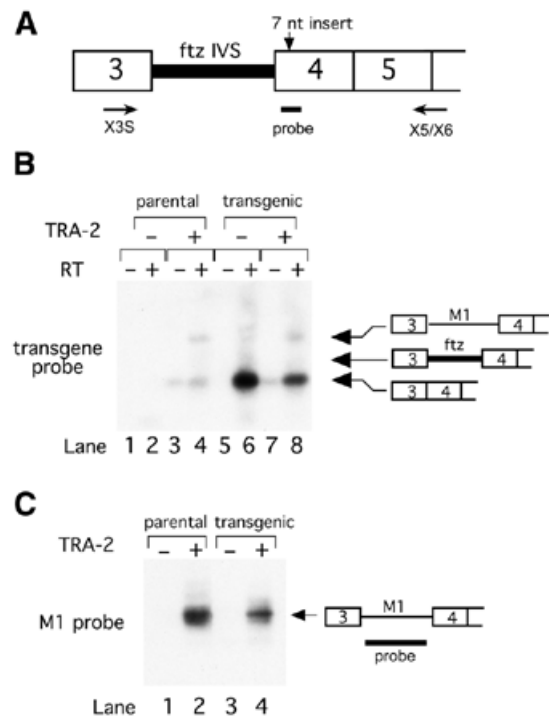
Figure 1. Replacement of the *tra-2* M1 intron with the intron from the *ftz* gene. (**A**) A schematic showing the predicted 5′ end of male germline *tra-2* pre-mRNA in which M1 intron of *tra-2* has been replaced with the constituitively spliced intron of the *ftz* gene (*ftz* IVS). Also shown is the relative location of the 7 nt insertion downstream of the 3′ splice site, the transgene probe and the positions of the PCR primers (X3S and X5/X6). Male germline specific transcripts initiate within exon 3 of the *tra-2* gene. (**B**) A Southern blot of low-cycle RT–PCR products from male *tra-2/ftz*IVS transgenic flies hybridized with a ^{32}P-end-labeled transgene-specific oligonucleotide containing the 7 nt insertion is shown (lanes 2, 4, 6, 8). Positions of products corresponding to different spliced forms are indicated on the right. Products from PCR with material for mock reverse transcriptase reactions (RT) in which enzyme was omitted are also shown (lanes 1, 3, 5, 7). Samples denoted '–' for TRA-2 are derived from males homozygous for a strong loss of function allele (*tra-2[B]*) that contains a nonsense mutation located within the TRA-2 RNA binding domain. Samples denoted '+' for TRA-2 are derived from sibling heterozygous mutant males (*tra-2[B]/+*). The lanes marked 'parental' derive from the strain used for microinjection (*w[1118]/B[s]Y; tra-2[B]/CyO*) which contains no transgene. (**C**) RT–PCR products from the same RNA samples hybridized with a probe from the M1 intron to confirm that endogenous transcripts undergo *tra-2*-dependent M1 splicing repression.

machinery and that poor recognition of the 3′ splice site/polypyrimidine tract facilitates TRA-2-dependent repression of M1 splicing. To test this idea, we produced transgenes with two mutations, in which sequences from the 3′ splice site region of M1 were replaced with sequences more closely matching the *Drosophila* consensus (29). In *P[ftz 3′]* the M1 3′ splice site and polypyrimidine tract were replaced with the analagous sequences from the *ftz* intron (Fig. 2A). In the other construct, *P[CAG 3′]*, the non-consensus A residue at position –3 was changed to C as in the *Drosophila* consensus. Transgenic fly strains were generated using both of these mutations and assayed for the ability of transgene-derived RNA to undergo M1 repression in the presence and absence of functional endogenous TRA-2. Splicing of the 'improved' M1 intron in transcripts from *P[ftz 3′]* was not repressed in the presence of TRA-2 protein (Fig. 2B, lanes 5–8). This indicates
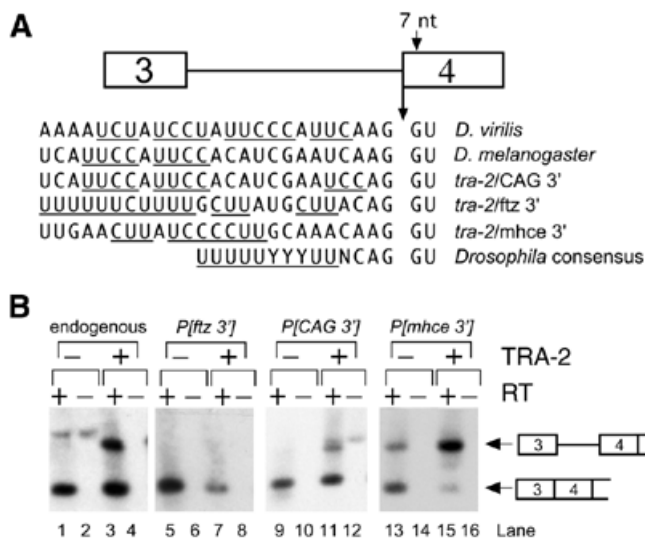
Figure 2. Substitutions of sequences near the M1 3′ splice site. (**A**) Schematic showing the M1 intron (line) and sequences of relevant 3′ splice sites. The 3′ intron sequences present in the *ftz3′*, CAG3′ and *mhce* 3′ transgenes are shown. Runs of three or more consecutive pyrimidines are underlined. (**B**) Southern blots of low-cycle RT–PCR products of RNAs from males of various fly strains. The primers used and the labeling of lanes are as described in Figure 1. Products in lanes 1–4 are derived from RNA of *tra-2^B/tra-2^B* and *tra-2^B/+* of the non-transgenic parental strain and were hybridized with an oligonucleotide probe specific for the endogenous *tra-2* gene lacking the 7 nt insertion in exon 4. Products in lanes 5−16 are derived from the various transgenic strains indicated and were hybridized with a transgene-specific oligonucleotide. Expected mobilities of products from spliced and unspliced RNA are indicated by arrows on the right. The faint lower mobility band appearing in lanes 1, 2, 11 and 12 is not dependent on reverse transcriptase and corresponds in size to the products expected from contaminating genomic DNA.

that splicing of an intron containing a consensus 3′ splice site cannot be repressed by TRA-2. In contrast, transcripts from *P[tra-2/CAG]* were repressed by endogenous TRA-2 protein (Fig. 2B, lanes 9–12) to a similar extent to endogenous *tra-2* RNAs (Fig. 2B, lanes 1–4). This shows that a non-consensus nucleotide at the –3 position is not essential for *tra-2*-dependent repression of M1 splicing.

A possible explanation for the above findings is that non-consensus sequences near the 3′ splice site are necessary to lower the basal rate of M1 splicing to a level that permits repression by TRA-2[226]. A second possibility is that the substituted sequences are directly and specifically recognized by TRA-2 or another regulatory factor involved in repression. To distinguish these possibilities we used sequences from the *Drosophila* myosin heavy chain intron E (mhce) (30), which also has a non-consensus 3′ splice site, to substitute for those of the M1 3′ splice site/polypyrimidine tract region. The mhce 3′ splice site sequences differ from those of the native M1 intron (Fig. 2A) and thus, would not be expected to bind similar regulatory factors. Transcripts produced from a transgene with this substitution *P[mhce 3′]* are shown in Figure 2B (lanes 13 and 15). Unlike endogenous *tra-2* transcripts, a large fraction of transgene-derived RNAs retained the M1 intron in the absence of functional TRA-2 protein (lane 13). This suggests that recognition of the 3′ splice site in *P[mhce 3′]* by the basal splicing machinery is less efficient than splicing of transcripts from the endogenous *tra-2* gene. Significantly, in the presence

of a functional endogenous TRA-2 protein, splicing of the M1 intron from *P[mhce 3′]* was further reduced (lane 15). These results indicate that substitution of mhce sequences near the 3′ end of the intron are compatible with TRA-2-dependent splicing repression. Because repression occurs with either the mhce or the native M1 sequences, we conclude that a critical feature of this region necessary for splicing repression is a suboptimal (but functional) polypyrimidine tract.

Mutation of a conserved sequence internal to the M1 intron affects splicing in somatic tissues

Comparison of the *D.virilis* and *D.melanogaster tra-2* genes also revealed a sequence element (VC) internal to the M1 intron of both genes that is conserved at 15 out of 18 positions (27). To determine if the VC element plays an important role in M1 repression we generated two transgenes with different mutations disrupting the element in the *D.melanogaster* M1 intron (Fig. 3A). In the transgene *P[VC-sub]* we replaced the VC element with random nucleotides at 14 out of 18 positions. In the transgene *P[VCdel]* we deleted 14 nt from the element. Both transgenes produced M1-containing RNAs in a TRA-2 independent manner, suggesting that the VC element is necessary for efficient splicing of the intron (Fig. 3B). Northern blot analysis of M1-containing RNAs from a transgenic strain carrying *P[VC-sub]* revealed the appearance of a novel M1-containing RNA that is independent of the presence of functional TRA-2 protein and is larger than the normal male germline M1-containing *tra-2* transcript (Fig. 3C). The size of the novel RNA suggests that it might be derived from the soma where an upstream transcription start site is utilized (Fig. 4A) and where the M1 intron is normally not retained at significant levels (31). RNAase protection assays confirmed that these flies expressed novel transcripts with the somatic form of exon 3 and retained the entire M1 intron (data not shown). To further determine that these novel transcripts accumulate in the soma, we utilized primers from exons 2 and 4 in RT–PCR experiments with RNA extracted from gonad-free carcasses of adult male flies. M1-containing sequences made up a large fraction of the products deriving from both the *P[VC-sub]* and *P[VC-del]* transgenes but were only a minor fraction of products from the endogenous *tra-2* gene (Fig. 4B). As expected based on previous observations (4,31), neither endogenous nor transgenic M1 containing RNAs found in the soma were dependent on the activity of the TRA-2 protein. We conclude that somatic splicing of the M1 intron is reduced in the absence of the VC element.

The VC element is necessary to overcome suboptimal splicing signals near the 3′ end of the M1 intron

The low levels of M1-containing transcripts present in the soma suggest that the intron is normally efficiently spliced in these tissues despite the presence of suboptimal 3′ signals. This led us to hypothesize that a function of the VC element is to overcome these suboptimal splicing signals. If so, then an intron with consensus 3′ splicing signals would not be expected to require the VC element in the soma for efficient splicing. We therefore produced transgenic flies that expressed transcripts carrying a combination of mutations in both VC and the 3′ splice site region *P[tra-2/VCsub, ftz 3′]*. This transgene contains a combination of the substitutions used in *P[VC-sub]* and *P[tra-2/ftz 3′]*. As predicted by our hypothesis, the M1
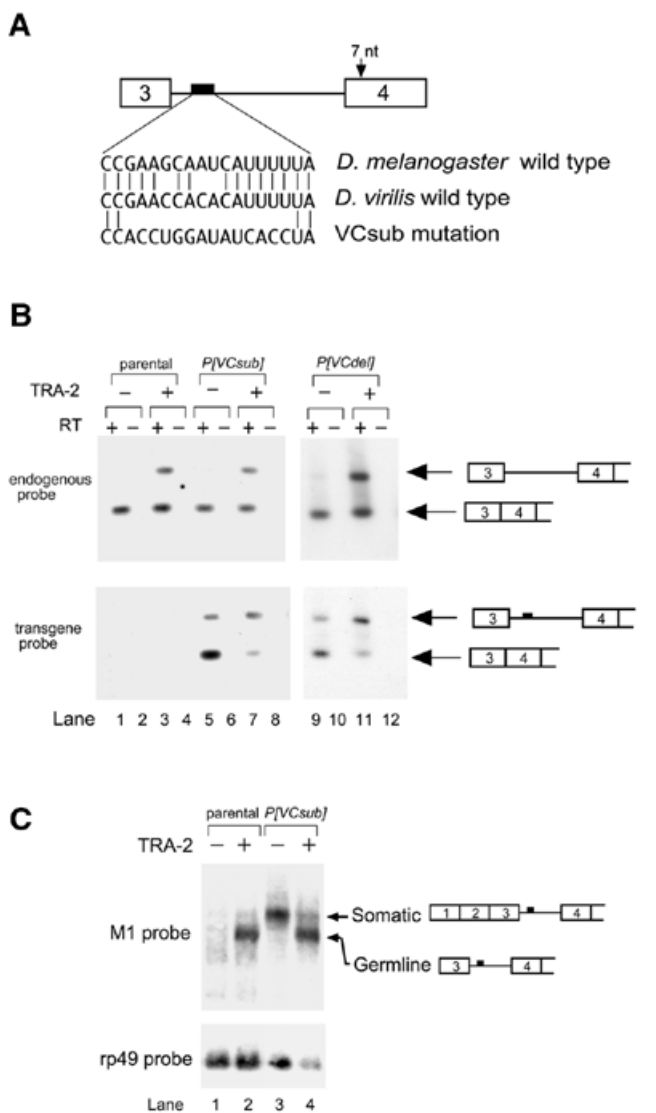
Figure 3. Mutation of a conserved intron sequence decreases M1 splicing in the soma. (**A**) Schematic showing the sequence and location of the conserved VC element (black box) in the M1 intron (line). The sequence of the VC element and the substitution mutation (VCsub) is shown. (**B**) Low cycle RT–PCR was performed as in Figure 1B. RT–PCR products were hybridized with oligonucleotides specific for either endogenous or transgene-derived RNA. Expected mobilities and the structure of spliced and unspliced products are shown to the right of the panels. After hybridization with the probe for endogenous transcripts the same blot was stripped and used to detect products from the transgene. Endogenous M1 containing RNA is observed only in wild-type males (upper panel, lanes 7 and 11), but M1-containing RNA derived from the transgenes is observed in both wild-type (lower panel, lanes 7 and 11) and *tra-2B* mutant males (lower panel, lanes 5 and 9). (**C**) Northern blot analysis of poly(A)+ RNA from flies expressing the VCsub mutation. M1 intron containing transcripts were detected using a ^{32}P-labeled *in vitro* transcribed antisense M1 RNA. The smaller, TRA-2-dependent RNA corresponds to the normal germline M1 containing RNA and is observed in both the parental (lanes 1 and 2) and transgenic flies (lanes 3 and 4). The larger RNA which is independent of TRA-2 function, is observed at high levels in transgenic flies and at much lower levels in the parental strain. This RNA corresponds in size to transcripts initiating at the somatic transcription initiation site as indicated on the right of the figure. As a loading control the blot was rehybridized with a probe for transcripts from the *ribosomal protein 49* (*rp49*) gene. Note that lane 4 is underloaded.
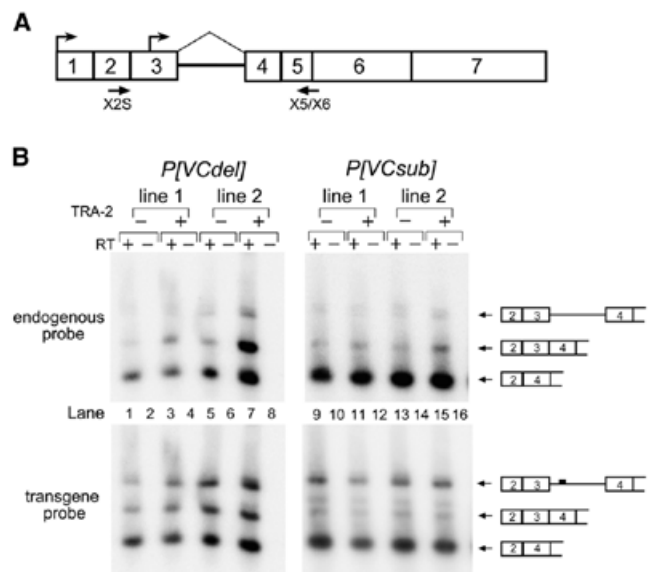
Figure 4. The VC element is necessary for the efficient removal of the M1 intron in the soma. (**A**) Schematic diagram of an M1-containing somatic *tra-2* RNA is shown. Boxes indicate exons and arrows above indicate both the somatic transcription start site at the beginning of exon 1 and the germline transcription start site within exon 3. PCR primers designed to specifically recognize cDNA generated from the longer somatic *tra-2* RNAs are indicated below the RNA. Primer X2S is upstream of the germline-specific transcription start site and primer X5/X6 spans the splice junction between exons 5 and 6. (**B**) Low cycle RT–PCR analysis of somatic RNA extracted from male fly carcasses from which the testes had been removed by dissection are shown. The flies examined expressed RNA from either the *P[VCdel]* or the *P[VCsub]* *tra-2* transgenes. Endogenous and transgene RT–PCR products were distinguished with specific oligonucleotide probes as in previous experiments. Note that in all samples the band corresponding to somatic M1-containing RNA comprises a significantly higher fraction of the transgene transcripts compared to endogenous gene transcripts. As observed in previous studies, somatic M1 splicing is independent of the endogenous *tra-2* genotype.

intron was efficiently spliced from transcripts containing this transgene despite the absence of a functional VC element (Fig. 5). These results indicate that the VC element is required for M1 splicing only in the presence of the suboptimal, native 3′ splicing signals.

DISCUSSION

The alternative splicing of *tra-2* appears to be conserved during evolution. TRA-2 protein variants that are generated by alternative splicing in humans are strikingly similar to those produced in the fly (26,32). Although conservation of the regulatory mechanisms between human and insects appears unlikely, examination of the splicing of *tra-2* RNA from *D.virilis*, a species that diverged from *D.melanogaster* over 60 million years ago, revealed that it undergoes similar regulation in these two species (27). Testes-specific splice variants in *D.virilis* retain an intron analogous to M1 that interrupts the first translation initiation codon. When introduced into *D.melanogaster*, the *D.virilis tra-2* gene acted *in trans* to regulate splicing of the endogenous *tra-2* pre-mRNA. Using a defective *D.virilis* transgene it was found that the endogenous *D.melanogaster* TRA-2 protein could likewise repress splicing

of *D.virilis tra-2* pre-mRNA. Thus, the *cis*-regulatory sequences affecting *tra-2* splicing are likely to be very similar in the two species.

Comparison of the *D.virilis* and *D.melanogaster tra-2* pre-mRNA sequences revealed conserved features within the M1 intron. In both species the 3′ splice site of the M1 intron is similarly divergent from the *Drosophila* 3′ splice site consensus, with both introns containing a non-consensus A at position –3, as well as interrupted polypyrimidine tracts. Sequences near the M1 5′ splice site (AT/GTAAGA) are also conserved in the M1 intron. However, in contrast to the 3′ splice site, this conservation is likely to be due only to the canonical requirement for these sequences. The conserved 5′ splice site closely matches the *Drosophila* 5′ splice site consensus (AG/GTRAGT) both in its extent and sequence and is thus similar to 5′ splice sites found in many other introns (29). A final conserved feature of M1 identified in the comparison between *Drosophila* species is an intron-internal sequence (the VC element) that is identical at 15 out of 18 positions but does not correspond to any known canonical splicing signals.

To determine if these conserved sequences are needed for regulation of M1 splicing we generated transgenic fly strains expressing altered *tra-2* pre-mRNAs. Mutations of the M1 3′ splice site region suggest that the non-consensus sequences at this position are critical for normal splicing regulation. Substitution of the wild-type M1 3′ splice site region with sequences from a different non-consensus site supported TRA-2-dependent splicing repression but substitutions that more closely match the *Drosophila* consensus did not. One possibility is that a poor 3′ splice site limits the rate of splicing allowing regulatory factors an opportunity to intervene before splicing is initiated. This idea is supported by *in vitro* splicing experiments showing that substitution of the *ftz* 3′ splice site signals into M1 dramatically increases the rate of splicing relative to RNAs with the wild-type 3′ splice site (D.S.Chandler and W.Mattox, manuscript in preparation).

Although the non-consensus sequences in the wild-type 3′ splice site appear to confer a relatively slow rate of splicing on M1, the intron is removed from all germline *tra-2* mRNA produced in the absence of functional TRA-2[226] protein (4,16). Thus, the inherent 'strength' of 3′ splicing signals in the native M1 intron appears to be precisely tuned to a level that permits intron retention in the presence of repressor and complete splicing in the absence of repressor.

Non-consensus 3′ splice sites appear to be a common, but not universal, feature of introns where splicing is repressed. The 3′ splice sites of several other *Drosophila* RNAs with similarly regulated introns are shown in Figure 6 (3,33–36). Four of the five introns deviate significantly from the *Drosophila* 3′ splice site consensus (29), and in the fifth case (the *Drosophila* P element) splicing is known to be repressed through interactions of regulators near the 5′ splice site. We speculate that a reduced rate of basal splicing caused by non-consensus sequences near the 3′ splice sites in many such introns is essential for regulated intron retention *in vivo*.

Analysis of the VC element, the most highly conserved sequence within the M1 introns of *D.melanogaster* and *D.virilis*, showed that this sequence acts to facilitate M1 splicing. This effect appears to be specific to the somatic tissues of adults, where TRA-2-dependent repression of M1 splicing does not occur. When the VC element was deleted or
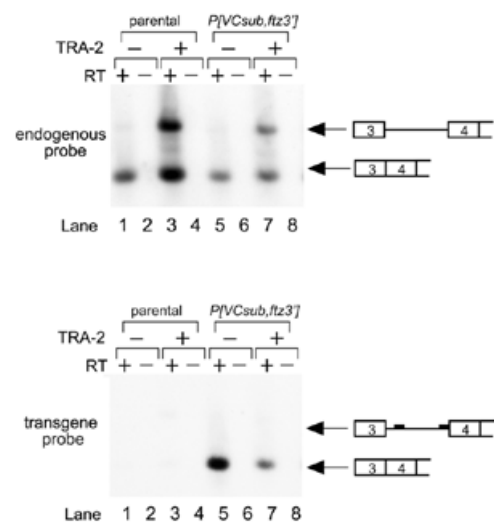
Figure 5. Substitution of a strong polypyrimidine tract/3′ splice site into the M1 intron alleviates the somatic requirement for the VC element. M1 splicing in flies expressing *P[VCsub,ftz3′]* transgene is analyzed by low cycle RT–PCR and Southern blotting. In this transgene the VCsub mutation is combined with a substitution of the strong *ftz* 3′ splice site. Products from endogenous and transgene transcripts were detected using specific oligonucleotide probes as in other experiments. Expected mobilities of spliced and unspliced products are shown schematically on the right of the panels. Intron-containing transcripts were detected from the endogenous gene (upper panel, lanes 3 and 7) but not from the transgene (lanes 5 and 7) in the presence or absence of functional TRA-2 protein.
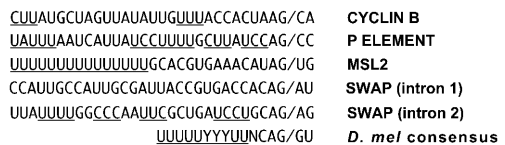
Figure 6. Comparison of 3′ splice site regions of introns that undergo splicing repression. The 3′ splice site regions of five introns which are retained in mature mRNA under specific developmental conditions or in response to known regulatory factors are shown in comparison to the *D.melanogaster* consensus sequences. The splice junctions are indicated by a forward slash and runs of three or more pyrimidines are underlined. Three of the five sequences (*cyclin B* and two introns from the *SWAP* gene) lack a polypyrimidine tract of >4 nt near the 3′ splice site. A fourth sequence (*msl-2*) has a polypyrimidine tract starting 16 nt upstream of the splice site and this spacing has been shown to reduce recognition of the splice site by U2AF in a manner important for splicing repression (46). Splicing of the intron from P element pre-mRNA is known to be repressed by complexes formed near the 5′ splice site of that intron (35).

substituted, adult flies accumulated high levels of somatic RNAs containing M1. Thus, the VC sequence appears to be necessary for efficient splicing of M1 in the soma.

The above results raise the issue of why the *tra-2* pre-mRNA contains sequences that both positively and negatively affect M1 splicing. We hypothesized that positive elements, such as VC, might be required to overcome the weak 3′ splice site and allow efficient M1 splicing in somatic tissues. In support of this idea we found that substitution of the M1 3′ splice site with a strong 3′ site alleviated the need for the VC element in the soma. Thus, the VC element allows for efficient M1 splicing in the presence of the intron's weak 3′ splice site, which is itself a necessary feature of germline splicing repression.

Neither of the sequences examined here appears to be a direct target of TRA-2^{226} or a co-repressor of M1 splicing. The VC element clearly has a positive rather than repressive role and alterations in it do not affect repression of M1 splicing in the germline. The weak M1 3′ splice site is required for germline splicing repression, but the ability of another weak 3′ splice site to functionally replace it suggests that it does not contain sequences recognized specifically by factors involved in M1 splicing repression. Therefore, it seems most likely that these conserved sequences affect only the basal level of M1 splicing in the soma and germline. The evolutionary conservation of such 'context elements' and the observation that alteration of these elements leads to qualitative changes in M1 splicing indicates that the basal splicing efficiency is critical to normal regulation. Increased rates of basal M1 splicing overcome the ability of TRA-2^{226} to repress it in the germline, while a decrease in basal splicing in the soma leads to accumulation of M1 containing transcripts where they are normally not found.

It is surprising that comparison of the *tra-2* genes from different *Drosophila* species did not identify conserved sequences more directly involved in splicing repression. In a similar comparison, a critical 13 nt repeated element of the splicing enhancer from the *dsx* gene was found to be well conserved in the same two species (25,37). Several of the same 13 nt repeat elements are also found in the regulatory sequences within *fru* pre-mRNA, another target of *tra*-2 (12,14). Although this suggests that the 13 nt repeat element may be common to RNAs where splicing is affected by *tra*-2, no sequences identical to the 13 nt element were found in or near the M1 intron. *In vivo* studies on two related but non-identical and non-conserved repeat sequences in exon 3 upstream of the M1 intron showed that they are not required for M1 repression (D.S.Chandler and W.Mattox, unpublished). It is possible that the absence of a similar regulatory sequence in *tra-2* pre-mRNA reflects differences in the role of the TRA protein, which is required in combination with TRA-2 for enhancement of *dsx* and *fru* splicing but does not function in the male germline where M1 splicing repression occurs. Thus, germline specific targets of *tra-2* including the RNAs from *tra-2*, *exuperantia* and *alternative-testes-transcript* (38,39) may be regulated via different *cis*-regulatory elements than those found in *dsx* and *fru*.

The idea that regulation of alternative splicing depends on sequences that affect the efficiency of splicing is supported by observations that other regulated pre-mRNAs contain poor matches to consensus splicing signals (40–44). In several cases it has been shown that substitution of these sequences with consensus splice sites leads to constitutive use of the splice site and thus correct regulation of splicing depends on the weak splice site. In addition, some splicing enhancers have been shown to be essential for regulated splicing despite not being the direct targets of determinative tissue-specific regulators (45). Our results emphasize the sensitivity of splicing regulation to the basal rate of splicing and the importance of studying splicing regulatory elements within an appropriate context. Outside of a context where basal splicing occurs at normal efficiency, the effects of splicing regulators such as TRA-2 would not be observed.

REFERENCES

1. Chabot,B. (1996) Directing alternative splicing: cast and scenarios. *Trends Genet.*, **12**, 472–478.
2. Graveley,B.R. (2000) Sorting out the complexity of SR protein functions. *RNA*, **6**, 1197–1211.
3. Zachar,Z., Chou,T.-B. and Bingham,P.M. (1987) Evidence that a regulatory gene autoregulates splicing of its transcript. *EMBO J.*, **6**, 4105–4111.
4. Mattox,W. and Baker,B.S. (1991) Autoregulation of the splicing of transcripts from the transformer-2 gene of *Drosophila*. *Genes Dev.*, **5**, 786–796.
5. Bell,L.R., Horabin,J.I., Schedl,P. and Cline,T.W. (1991) Positive autoregulation of sex-lethal by alternative splicing maintains the female determined state in *Drosophila*. *Cell*, **65**, 229–239.
6. Lejeune,F., Cavaloc,Y. and Stevenin,J. (2000) Alternative splicing of intron 3 of the SR protein 9G8 gene: identification of flanking exonic splicing enhancers and involvement of 9G8 as a trans-acting factor. *J. Biol. Chem.*, **276**, 7850–7858.
7. Sureau,A., Gattoni,R., Dooghe,Y., Stevenin,J. and Soret,J. (2001) SC35 autoregulates its expression by promoting splicing events that destabilize its mRNAs. *EMBO J.*, **20**, 1785–1796.
8. Nagoshi,R.N., McKeown,M., Burtis,K.C., Belote,J.M. and Baker,B.S. (1988) The control of alternative splicing at genes regulating sexual differentiation in *D.melanogaster*. *Cell*, **53**, 229–236.
9. Ryner,L.C. and Baker,B.S. (1991) Regulation of *doublesex* pre-mRNA processing occurs by 3′-splice site activation. *Genes Dev.*, **5**, 2071–2085.
10. Tian,M. and Maniatis,T. (1992) Positive control of pre-mRNA splicing *in vitro*. *Science*, **256**, 237–240.
11. Wu,J.Y. and Maniatis,T. (1993) Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell*, **75**, 1061–1070.
12. Ryner,L.C., Goodwin,S.F., Castrillon,D.H., Anand,A., Villella,A., Baker,B.S., Hall,J.C., Taylor,B.J. and Wasserman,S.A. (1996) Control of male sexual behavior and sexual orientation in *Drosophila* by the *fruitless* gene. *Cell*, **87**, 1079–1089.
13. Zuo,P. and Maniatis,T. (1996) The splicing factor U2AF(35) mediates critical protein–protein interactions in constitutive and enhancer-dependent splicing. *Genes Dev.*, **10**, 1356–1368.
14. Heinrichs,V., Ryner,L.C. and Baker,B.S. (1998) Regulation of sex-specific selection of *fruitless* 5′ splice sites by *transformer* and *transformer-2*. *Mol. Cell. Biol.*, **18**, 450–458.
15. Mattox,W., McGuffin,M.E. and Baker,B.S. (1996) A negative feedback mechanism revealed by functional analysis of the alternative isoforms of the *Drosophila* splicing regulator *transformer-2*. *Genetics*, **143**, 303–314.
16. McGuffin,M.E., Chandler,D., Somaiya,D., Dauwalder,B. and Mattox,W. (1998) Autoregulation of *transformer-2* alternative splicing is necessary for normal male fertility in *Drosophila*. *Genetics*, **149**, 1477–1486.
17. Amrein,H., Maniatis,T. and Nothiger,R. (1990) Alternatively spliced transcripts of the sex-determining gene *tra-2* of *Drosophila* encode functional proteins of different size. *EMBO J.*, **9**, 3619–3629.
18. Mattox,W., Ryner,L. and Baker,B.S. (1992) Autoregulation and multifunctionality among trans-acting factors that regulate alternative pre-mRNA processing. *J. Biol. Chem.*, **267**, 19023–19026.
19. Tian,M. and Maniatis,T. (1993) A splicing enhancer complex controls alternative splicing of *doublesex* pre-mRNA. *Cell*, **74**, 105–114.
20. Amrein,H., Hedley,M.L. and Maniatis,T. (1994) The role of specific protein–RNA and protein–protein interactions in positive and negative control of pre-mRNA splicing by Transformer 2. *Cell*, **76**, 735–746.
21. Tian,M. and Maniatis,T. (1994) A splicing enhancer exhibits both constitutive and regulated activities. *Genes Dev.*, **8**, 1703–1712.
22. Lynch,K.W. and Maniatis,T. (1995) Synergistic interactions between two distinct elements of a regulated splicing enhancer. *Genes Dev.*, **9**, 284–293.

23. Heinrichs,V. and Baker,B.S. (1995) The *Drosophila* SR protein RBP1 contributes to the regulation of *doublesex* alternative splicing by recognizing RBP1 RNA target sequences. *EMBO J.*, **14**, 3987–4000.

24. Lynch,K.W. and Maniatis,T. (1996) Assembly of specific sr protein complexes on distinct regulatory elements of the *Drosophila doublesex* splicing enhancer. *Genes Dev.*, **10**, 2089–2101.

25. Hertel,K.J., Lynch,K.W., Hsiao,E.C., Liu,E.H. and Maniatis,T. (1996) Structural and functional conservation of the *Drosophila doublesex* splicing enhancer repeat elements. *RNA*, **2**, 969–981.

26. Dauwalder,B. and Mattox,W. (1998) Analysis of the functional specificity of RS domains *in vivo*. *EMBO J.*, **17**, 6049–6060.

27. Chandler,D., McGuffin,M.E., Piskur,J., Yao,J., Baker,B.S. and Mattox,W. (1997) Evolutionary conservation of regulatory strategies for the sex determination factor transformer-2. *Mol. Cell. Biol.*, **17**, 2908–2919.

28. Pirrotta,V. (1988) In Rodriguez,R.L. and Denhardt,D.T. (eds), *Vectors: A Survey of Molecular Cloning Vectors And Their Uses.* Butterworths, Boston, pp. 437–456.

29. Mount,S. (1993) Messenger RNA splicing signals in *Drosophila* genes. In Maroni,G. (ed), *Atlas of Drosophila Genes.* Oxford University Press, New York, NY, pp. 333–358.

30. Wassenberg,D.R.,II, Kronert,W.A., O'Donnell,P.T. and Bernstein,S.I. (1987) Analysis of the 5′ end of the *Drosophila* muscle myosin heavy chain gene. Alternatively spliced transcripts initiate at a single site and intron locations are conserved compared to myosin genes of other organisms. *J. Biol. Chem.*, **262**, 10741–10747.

31. Mattox,W., Palmer,M.J. and Baker,B.S. (1990) Alternative splicing of the sex determination gene *transformer-2* is sex-specific in the germ line but not in the soma. *Genes Dev.*, **4**, 789–805.

32. Nayler,O., Cap,C. and Stamm,S. (1998) Human transformer-2-beta gene (SFRS10): complete nucleotide sequence, chromosomal localization and generation of a tissue-specific isoform. *Genomics*, **53**, 191–202.

33. Dalby,B. and Glover,D.M. (1992) 3′ non-translated sequences in *Drosophila* cyclin B transcripts direct posterior pole accumulation late in oogenesis and peri-nuclear association in syncytial embryos. *Development*, **115**, 989–997.

34. Siebel,C.W. and Rio,D.C. (1990) Regulated splicing of the *Drosophila* P transposable element third intron *in vitro*: somatic repression. *Science*, **248**, 1200–1208.

35. Siebel,C.W., Fresco,L.D. and Rio,D.C. (1992) The mechanism of somatic inhibition of *Drosophila* P-element pre-mRNA splicing: multiprotein complexes at an exon pseudo-5′ splice site control U1 snRNP binding. *Genes Dev.*, **6**, 1386–1401.

36. Zhou,S., Yang,Y., Scott,M.J., Pannuti,A., Fehr,K.C., Eisen,A., Koonin,E.V., Fouts,D.L., Wrightsman,R., Manning,J.E. *et al.* (1995) *Male-specific lethal* 2, a dosage compensation gene of *Drosophila*, undergoes sex-specific regulation and encodes a protein with a RING finger and a metallothionein-like cysteine cluster. *EMBO J.*, **14**, 2884–2895.

37. Burtis,K.C. and Baker,B.S. (1989) *Drosophila doublesex* gene controls somatic sexual differentiation by producing alternatively spliced mRNAs encoding related sex-specific polypeptides. *Cell*, **56**, 997–1010.

38. Hazelrigg,T. and Tu,C. (1994) Sex-specific processing of the *Drosophila exuperantia* transcript is regulated in male germ cells by the *tra-2* gene. *Proc. Natl Acad. Sci. USA*, **91**, 10752–10756.

39. Madigan,S.J., Edeen,P., Esnayra,J. and McKeown,M. (1996) *Att*, a target for regulation by *Tra2* in the testes of *Drosophila melanogaster*, encodes alternative RNAs and alternative proteins. *Mol. Cell. Biol.*, **16**, 4222–4230.

40. Hoshijima,K., Inoue,K., Higuchi,I., Sakamoto,H. and Shimura,Y. (1991) Control of *doublesex* alternative splicing by transformer and *transformer-2* in *Drosophila*. *Science*, **252**, 833–836.

41. Watakabe,A., Tanaka,K. and Shimura,Y. (1993) The role of exon sequences in splice site selection. *Genes Dev.*, **7**, 407–418.

42. Sun,Q., Mayeda,A., Hampson,R.K., Krainer,A.R. and Rottman,F.M. (1993) General splicing factor SF2/ASF promotes alternative splicing by binding to an exonic splicing enhancer. *Genes Dev.*, **7**, 2598–2608.

43. Humphrey,M.B., Bryan,J., Cooper,T.A. and Berget,S.M. (1995) A 32-nucleotide exon-splicing enhancer regulates usage of competing 5′ splice sites in a differential internal exon. *Mol. Cell. Biol.*, **15**, 3979–3988.

44. Zheng,Z.M., Quintero,J., Reid,E.S., Gocke,C. and Baker,C.C. (2000) Optimization of a weak 3′ splice site counteracts the function of a bovine papillomavirus type 1 exonic splicing suppressor *in vitro* and *in vivo*. *J. Virol.*, **74**, 5902–5910.

45. Xu,R., Teng,J. and Cooper,T.A. (1993) The cardiac troponin T alternative exon contains a novel purine-rich positive splicing element. *Mol. Cell. Biol.*, **13**, 3660–3674.

46. Merendino,L., Guth,S., Bilbao,D., Martinez,C. and Valcarcel,J. (1999) Inhibition of *msl-2* splicing by *Sex-lethal* reveals interaction between U2AF35 and the 3′ splice site AG. *Nature*, **402**, 838–841.