Splicing of Retrotransposon Insertions from Transcripts of the *Drosophila melanogaster vermilion* Gene in a Revertant

Anne-Marie Pret and Lillie L. Searles

Department of Biology, University of North Carolina, Chapel Hill, North Carolina 27599-3280

Manuscript received July 1, 1991
Accepted for publication August 25, 1991

ABSTRACT

A mutation of the *Drosophila melanogaster vermilion* (*v*) gene known as v^1 is caused by the insertion of a *412* retrotransposon into the 5′ untranslated region of the first exon. Mutants carrying this insertion accumulate a low level of mRNA from which most of the transposon sequences have been eliminated by splicing at cryptic sites within transposon sequences. Here, we demonstrate that a revertant of the v^1 allele called v^{+37} is caused by the insertion of a second retrotransposon, the *B104/roo* element, into a site near one end of the *412* element. The revertant strain accumulates a higher level of mRNA from which most of both transposons have been removed by splicing at new donor sites introduced by the *B104/roo* insertion and the same acceptor site within *412*. Mutations at *suppressor of sable* [*su(s)*], which increase the accumulation of v^1 transcripts, slightly elevate the level of v^{+37} RNA. In addition, we show that the first *v* intron downstream of the *412* insertion is not efficiently removed in the v^1 mutant, and suppressor and reversion mutations increase the proportion of transcripts that are properly spliced at that downstream intron. Thus, it appears that both the suppressor and reversion mutations exert an effect at the level of pre-mRNA splicing.

SPONTANEOUS mutations in the fruit fly *Drosophila melanogaster* frequently result from the movement of transposable elements. In some cases, the deleterious effects of transposon insertions can be alleviated by suppressor mutations that map to separate genes. We are investigating mutations of the *D. melanogaster vermilion* (*v*) gene that can be suppressed by mutations at the *suppressor of sable* [*su(s)*] gene. The *v* gene encodes an enzyme required for brown eye pigment synthesis. The suppressible mutations v^1, v^2 and v^k (collectively referred to as v^s), are all associated with the insertion of a 7.4-kilobase (kb) retrotransposon known as *412* at the same position, within the 5′ transcribed but untranslated region of the *v* gene (SEARLES and VOELKER 1986; SEARLES *et al.* 1990). In each mutant, the retrotransposon insertion is in a transcriptional orientation opposite that of *v*. Despite the insertion of a transposon into exon sequences, low levels of wild-type sized transcripts accumulate in flies carrying a v^s allele. In $su(s)^- v^s$ double mutants, the level of *v* mRNA accumulation is partially restored (SEARLES *et al.* 1990), which results in complete reversion of the mutant eye color phenotype. We recently demonstrated that most of the transposable element sequences are eliminated from the *v* mutant transcripts by splicing at cryptic sites within the *412* element (FRIDELL, PRET and SEARLES 1990; see Figure 1B). The primary transcripts are spliced from one of four different donor sites located at one end of the transposon to a single acceptor site at the other end. These novel splice sites are not normally used in generating *412* transcripts as they are derived from what is usually the nontranscribed strand of *412*. The cryptic splice donor sites appear to be somewhat atypical in comparison to other invertebrate splice sites whereas the acceptor appears to be a more normal site (FRIDELL, PRET and SEARLES 1990). The mechanism by which *su(s)* mutations increase the level of these novel spliced transcripts is unknown. Analysis of the *su(s)* gene indicates that it encodes a nuclear protein with a region of similarity to several proteins involved in RNA processing (VOELKER *et al.* 1991). We previously proposed that *su(s)* may encode a protein that normally functions to prevent recognition of cryptic splice sites (FRIDELL, PRET and SEARLES 1990).

The effects produced by the insertion of a transposon into a gene can also be reversed by second-site, intragenic mutations (MOUNT, GREEN and RUBIN 1988; GEYER, GREEN and CORCES 1988a,b; WILLIAMS and BELL 1988; HARRISON *et al.* 1989). The molecular analysis of such reversion mutations has, in at least one case, contributed to the understanding of the mechanism of extragenic suppression. For example, a mutant allele of *yellow*, y^2, caused by the insertion of a *gypsy* retrotransposon into the 5′ untranscribed region of the gene (GEYER, SPANA and CORCES 1986), can be suppressed by mutations at the *suppressor of Hairy-wing* [*su(Hw)*] gene (PARKHURST and CORCES 1986). One class of y^2 revertants has secondary transposable element insertions into a specific region of the *gypsy* element at *y* (GEYER, GREEN and CORCES 1988b; HAR-

RISON *et al.* 1989). The secondary transposon insertions are clustered in a region of *gypsy* containing multiple copies of a sequence motif which is similar to a mammalian enhancer. The *su(Hw)* protein binds to these sequences (SPANA, HARRISON and CORCES 1988; HARRISON *et al.* 1989), and the bound protein is thought to interfere with proper *y* promoter function. Thus, mutations that disrupt either the gene encoding the *su(*Hw) protein or the sequence elements in *gypsy* to which the protein binds restore *y* gene activity. Although *su(*Hw) and *su(s)* probably act by different mechanisms, analysis of *v* revertants might also provide information that will eventually be relevant to the mechanism of suppression by *su(s)*. With this idea in mind, we have undertaken the analysis of a revertant of v^1 called v^{+37}.

MATERIALS AND METHODS

Drosophila mutants: The mutant alleles v^1 and v^k are described in SEARLES *et al.* (1990). The v^{+37} allele was isolated by GERMERAAD (1976). The alleles $su(s)^{51c15}$, $su(s)^{X1}$, $su(s)^{f6}$, and $su(s)^{51j}$ are described in CHANG *et al.* (1986). The stocks containing *X* chromosomes with various *su(s)* alleles in combination with the v^{+37} allele were kindly provided by R. VOELKER. A $y^2\ w^{bf}$ stock was used as the wild-type *v* strain.

Analysis of genomic clones: The v^{+37} allele was cloned by ligation of *Mbo*I partial genomic DNA fragments into the *Bam*HI site of lambda EMBL4 as previously described (SEARLES and VOELKER 1986). Restriction fragments were subcloned into pGEM7Zf (Promega), and single-stranded phagemid DNA was sequenced according to the Sequenase sequencing kit protocol (U.S. Biochemical). Sequencing reactions were primed with either universal or specific oligonucleotide primers. To identify the new transposon insertion in v^{+37}, Genetics Computer Group (GCG) programs were used to search GenBank with a portion of the sequence. The search revealed a match to the published sequence of the *B104* long terminal repeat (LTR) (SCHERER *et al.* 1982).

RNA analysis: Poly(A)$^+$ RNA was purified from adult flies as described previously (FRIDELL, PRETT and SEARLES 1990). For Northern blot analysis 20 μg of poly(A)$^+$ RNA from each strain was loaded on a formaldehyde agarose gel, and following electrophoresis, RNA was transferred onto a nylon membrane. A single-stranded probe was synthesized from an M13 template containing the 1.9-kb *Xho*I/*Hin*dIII v^+ fragment (see Figure 1A) as described by O'HARE, LEVIS and RUBIN (1983), except that the Sequenase enzyme was used instead of Klenow fragment and the probes were purified from alkaline agarose gels (MANIATIS, FRITSCH and SAMBROOK 1982) by electroelution. Hybridization and washes were performed as described previously (SEARLES *et al.* 1990). The same filter was washed and reprobed with a clone of the Drosophila ribosomal protein rp49 (O'CONNELL and ROSBASH 1984), labeled by nick translation (MANIATIS, FRITSCH and SAMBROOK 1982). The *rp49* mRNA accumulates to the same level at all stages of development.

S1 nuclease protection experiments were performed as described previously (SEARLES *et al.* 1990) with the following modifications. In order to normalize the amount of RNA in each reaction, the *rp49* probe was added to each reaction with the *v* probe. Single-stranded antisense RNA probes were synthesized from plasmids pv103, a 0.67-kb *Eco*RI *v* cDNA fragment cloned into pGEM7Zf (Promega), pv8.2-2, a 0.32-kb polymerase chain reaction (PCR)-generated *v* genomic fragment cloned into pGEM7Zf, and pGRP49 (O'CONNELL and ROSBASH 1984), a 0.64-kb genomic fragment from the *rp49* gene cloned into pGEM4 (Promega). Plasmid pv103 was linearized with *Cla*I and transcribed with T7 RNA polymerase to generate a 273 nucleotide (nt) probe; pv8.2-2 was linearized with *Xho*I and transcribed with SP6 RNA polymerase to generate a 344-nt probe; and pGRP49 was linearized with *Pvu*II and transcribed with T7 RNA polymerase to synthesize a 71-nt probe. Since the *rp49* gene is expressed at a much higher level than *v*, the specific radioactivity of the *rp49* probe was reduced by the addition of cold CTP to the *in vitro* transcription reaction so that nearly equivalent signals were produced after protection and S1 nuclease digestion. The *v* and *rp49* probes were synthesized in reactions containing 200 μCi of [^{32}P]CTP (800 Ci/mmol) and 10 μCi of [^{32}P]CTP (14 Ci/mmol), respectively. An aliquot of probe, experimentally determined to represent a saturating amount, was included in each hybridization reaction. The hybridization reactions were processed and the protected fragments resolved as described previously (SEARLES *et al.* 1990). Autoradiographic exposures were performed without intensifying screens prior to densitometric analysis.

Isolation of cDNA clones: Partial cDNA clones of v^{+37} and $su(s)^{X1}\ v^{+37}$ mRNA were prepared as previously described (FRIDELL, PRET and SEARLES 1990). Briefly, oligonucleotide 2052, complementary to sequences in the fourth *v* exon, was used to prime first-strand cDNA synthesis on total adult poly(A)$^+$ RNA. Because of the low abundance of the revertant transcripts, the cDNAs were amplified by the PCR (SAIKI *et al.* 1988). Oligo 2052, used in first strand cDNA synthesis, and oligo 1846, complementary to first exon sequences upstream of the transposon insertions, were used as primers in the amplification reactions. *Xho*I/*Cla*I cDNA fragments were cloned into pGEM7Zf and sequenced. The sequence obtained began at the *Xho*I site in the *v* gene, upstream of the *B104/roo* and *412* insertions, and extended into *v* coding sequences.

RESULTS

Molecular structure of v^{+37} DNA: The revertant v^{+37} was originally identified as a putative *v* transformant during early attempts to transform Drosophila by microinjection of genomic DNA (GERMERAAD 1976). Thus, the v^{+37} allele probably arose as a spontaneous reversion of the v^1 mutation in the course of those experiments. On the basis of genomic Southern analysis of *Eco*RI digested v^{+37} DNA, we initially interpreted the reversion as a small deletion near the left end of the *412* element (SEARLES and VOELKER 1986). However, after the revertant allele was cloned, restriction analysis of cloned DNA revealed the presence of a 3.6-kb insertion near the left end of the *412* element (Figure 1C). The insertion introduced a new *Eco*RI site, which was responsible for the restriction pattern alteration previously detected by Southern analysis. Sequence analysis (see MATERIALS AND METHODS) demonstrated that the new insertion occurred 3 bp from the end of the *412* element (Figure 2, B and C). A homology search, performed with a portion of the v^{+37} sequence, indicated that the second insertion is a *B104/roo* element (SCHERER *et al.* 1982; MEYEROWITZ and HOGNESS 1982). The complete *B104/roo* retrotransposon is 9 kb in length, and thus approxi-

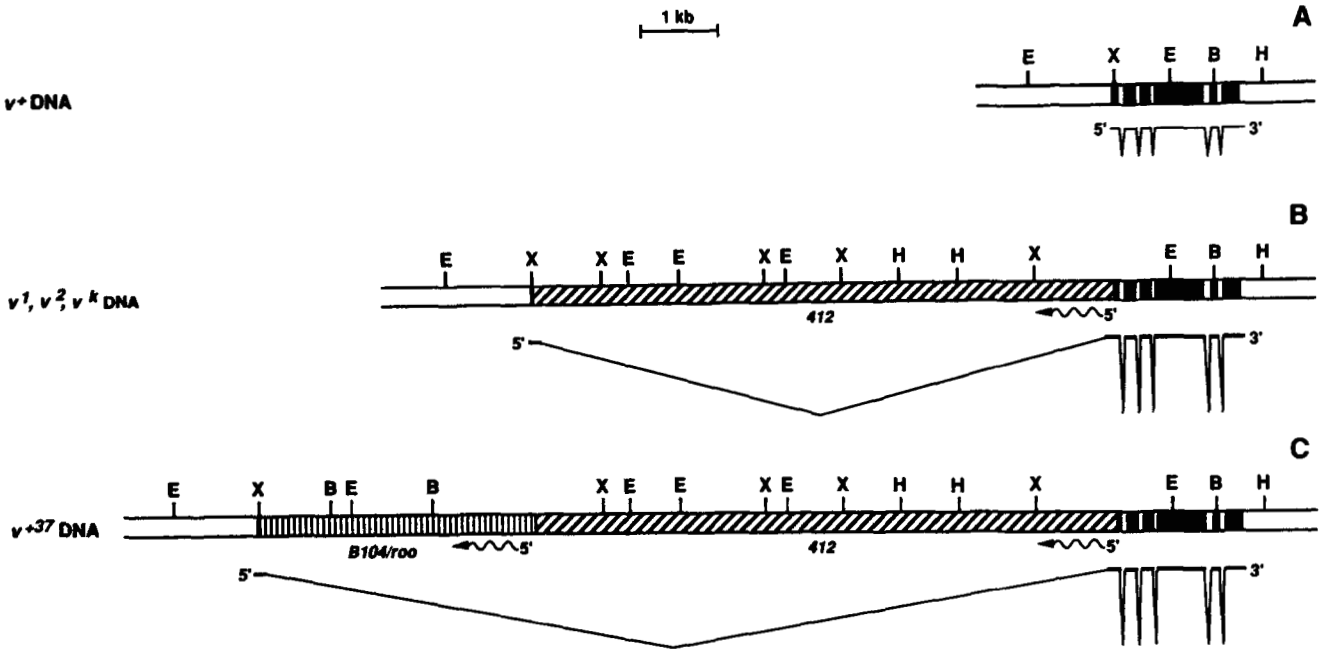

FIGURE 1.—Molecular maps of (A) wild-type, (B) mutant and (C) revertant v alleles. Solid bars represent v exons; bars with diagonal lines represent *412* sequences; bars with vertical lines represent *B104/roo* sequences. The splicing patterns of mRNAs produced from each allele are diagrammed beneath each map. The wavy arrow indicates the direction of transposon transcription. Restriction enzyme abbreviations are B, *Bam*HI; E, *Eco*RI;, H, *Hin*dIII; X, *Xho*I.

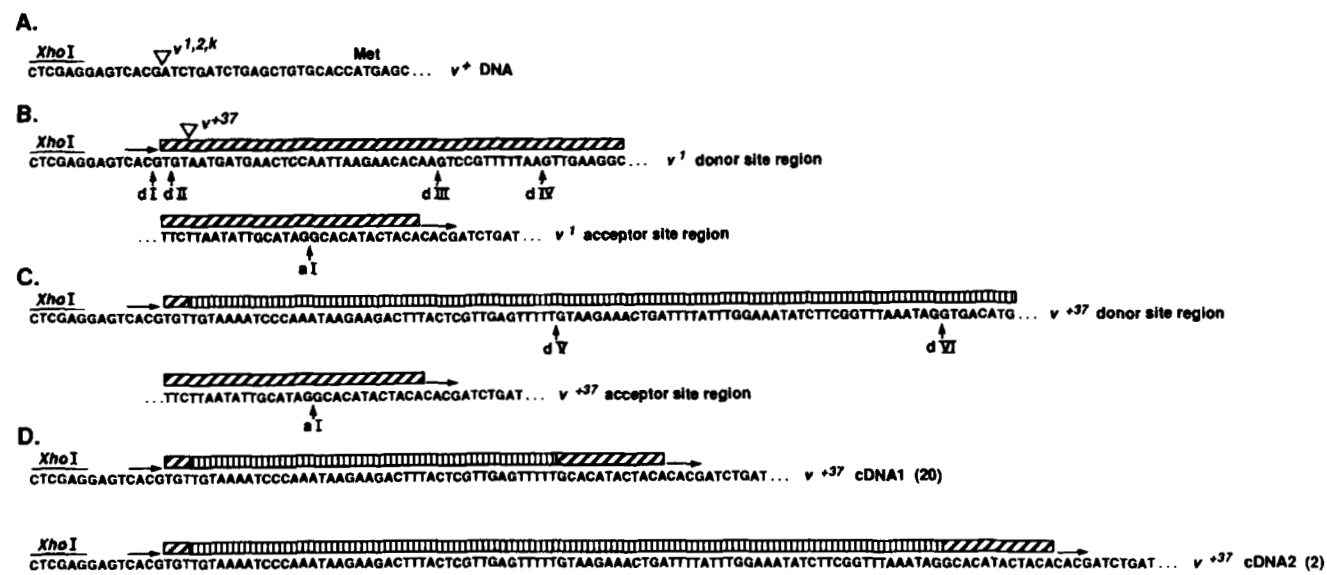

FIGURE 2.—Nucleotide sequences in the vicinity of the cryptic transposon-associated splice sites. (A, B and C) Genomic sequences of v^+, v^1, and v^{+37} DNA, respectively. (D) v^{+37} cDNA sequences. The triangles above the sequence indicate the positions of the transposon insertions. Transposon *412* sequences are indicated by the bars with diagonal lines above the sequence. Transposon *B104/roo* sequences are indicated by the bars with vertical lines above the sequence. The 4-nt direct repeat generated upon *412* insertion is indicated by the arrows above the sequence. The vertical arrows beneath the sequence indicate the position of splice donor (dI–dVI) and acceptor (aI) sites. The numbers included in parentheses in (D) indicate the number of cDNA clones of each type recovered.

mately 5.4 kb of DNA is missing from the element found in v^{+37}. We compared the restriction maps of a full-length *B104/roo* element to the map of the truncated element in v^{+37} DNA, and performed Southern hybridizations (SOUTHERN 1975), probing the cloned truncated element with various fragments of a cloned complete *B104/roo* element from the v^{36f} allele (SEARLES *et al.* 1990). In addition, we determined the sequence at both ends of the *B104/roo* insertion. This analysis indicated that the *B104/roo* insertion in v^{+37} is an internally deleted element with intact LTRs (data not shown). By comparing the sequence of the LTR to the published sequence, we deduced that the transcriptional orientation of the *B104/roo* element is opposite that of the v gene (Figure 1C).

Transcription of v^{+37}: With the two insertions, v^{+37} has an extra 11 kb of DNA in the first exon, and yet the gene is expressed at a level sufficient to give a

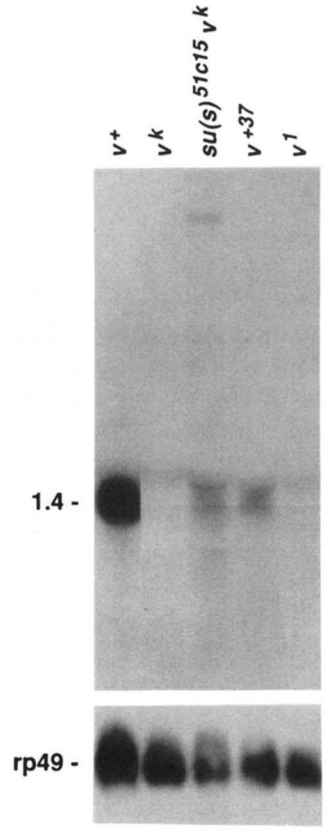

FIGURE 3.—Northern blot analysis of wild-type and mutant *v* RNAs. Each lane contains 20 μg of adult poly(A)$^+$ RNA, isolated from the strains indicated above the lanes. The blot was first hybridized with a single-stranded probe of the 1.9-kb *Xho*I-*Hin*dIII v^+ fragment (see Figure 1A), and subsequently the blot was rehybridized with an *rp49* probe. The positions of the 1.4-kb *v* transcript and the *rp49* transcript are indicated.

wild-type phenotype. Northern blot analysis of poly(A)$^+$ RNA (Figure 3) indicated that the v^{+37} revertant accumulates a transcript that is similar in size to the 1.4-kb v^+ and the $su(s)^{51c15}$ v^k transcripts. The level of v^{+37} transcript accumulation is comparable to that found in a $su(s)^{51c15}$ v^k genotype, and is significantly reduced relative to the wild-type *v* mRNA. A band which migrates slightly slower than the *v* transcript was consistently detected in RNA from all genotypes on Northern blots probed with single-stranded *v* DNA probes (Figure 3). This band is most likely an artifact produced by the 18S ribosomal RNA (1.9 kb) which contaminates the poly(A)$^+$ RNA. We do not know the origin of the large band observed in $su(s)^{51c15}$ v^k mRNA. Although it is approximately the size expected for the *v* mRNA precursor which would include the entire *412* element, this band did not hybridize to the appropriate single-stranded *412* DNA probe (data not shown).

We demonstrated previously that in v^s mutants, carrying the *412* insertion, transposon sequences are imprecisely eliminated from the pre-mRNA by splicing at one of four closely linked donor sites and at a single acceptor site located near the ends of the *412* element (Figures 1B and 2B). Donors dII (CGT/GTAATG) and dIII (CAA/GTCCGT) appear to be used much more frequently than dI (CAG/GTGTAA) and dIV (TAA/GTTGAA) (Figure 2B, see FRIDELL, PRET and SEARLES 1990). The insertion of the *B104/roo* element disrupts donor sites I and II and moves the other two donor sites 3.6 kb downstream (see Figures 1B and 2B).

Since the v^{+37} transcripts are similar in size to wild type, we suspected that transposable element sequences might similarly be removed from these transcripts by splicing. S1 nuclease mapping experiments indicated that sequences upstream of the insertion site were included in the revertant *v* transcripts (data not shown). We verified this result by isolating and sequencing v^{+37} cDNA clones. Poly(A)$^+$ RNA was primed for cDNA synthesis with a *v*-specific oligonucleotide primer and cDNAs were amplified by the PCR (see MATERIALS AND METHODS). Sequence analysis of the PCR-amplified v^{+37} cDNA clones revealed that sequences of both insertions are removed from the revertant transcripts by splicing from new donor sites in *B104/roo* to the same acceptor site in *412* that is used in splicing v^1 transcripts (Figures 1C and 2, C and D). We recovered 22 cDNA clones from 3 different cDNA synthesis and PCR-amplification experiments. Twenty of the cDNAs analyzed were spliced at donor site V (TTT/GTAAGA), located 41 bp from the left end of *B104/roo* (Figure 2, C and D). The other two clones were spliced at donor site VI (TAG/GTGACA), located 85 bp from the same end of the transposon (Figure 2, C and D). Thus, donor site V appears to be the predominant splice donor site utilized in the removal of transposon sequences from revertant transcripts. As is the case for the splice sites in the *412* element, the *B104* associated splice donor sites are not authentic splice sites because they are derived from the nontranscribed strand of the *B104/roo* transposon. To compare these cryptic *B104/roo* splice donor sites to other known splice sites, we analyzed the sequences by the method of SHAPIRO and SENAPATHY (1987). High scores indicate a high degree of similarity to known splice sites with a score of 100 indicating a perfect match to the invertebrate consensus (NAG/GTAAGT). We previously determined that the scores for known Drosophila splice donor sites ranged from 69 to 91, the average being approximately 80 (see FRIDELL, PRET and SEARLES 1990). The scores calculated for the *B104/roo* donor sites V and VI, 72 and 70, respectively, are equivalent to scores at the low end of the range determined for known splice sites. In comparison, the *412* donor site scores (43 to 65) are outside the range of known Drosophila splice sites (see FRIDELL, PRET and SEARLES 1990). Thus, reversion may be the result of more efficient splicing at the *B104/roo* donor sites.

The effect of *su(s)* mutations on accumulation of v^{+37} transcripts: To get an indication of whether the v^{+37} reversion corrects the defect in v^1 that allows a response to *su(s)*, we compared the effect of *su(s)* mutations on v^{+37} and v^1 expression. RNA levels were measured by quantitative S1 protection analysis, which is preferable to quantitation by Northern blot analysis in this case because of the apparently nonspecific RNA signal mentioned earlier that approximately comigrates with *v* transcripts. For these experiments, two radioactively labeled antisense RNA probes, one synthesized from a v^+ cDNA clone and the other from an *rp49* clone, were annealed simultaneously to poly(A)$^+$ RNA isolated from flies of different genotypes, and subsequently the RNA/RNA hybrids were digested with S1 nuclease. The 273-nt *v* probe includes cDNA sequences from a *Cla*I site in exon three to the *Eco*RI site in exon four (see SEARLES *et al.* 1990). Thus, only sequences downstream of the transposon insertion sites are included. When this probe is annealed to poly(A)$^+$ RNA, a 264-nt *v* fragment is protected (Figure 4, A and B). The 71-nt *rp49* probe includes *rp49* second exon sequences extending from a *Pvu*II site to a *Hin*dIII site (O'CONNELL and ROSBASH 1984) and generates a 61-nt protected fragment when annealed to poly(A)$^+$ RNA (Figure 4, A and B). Since the level of *rp49* mRNA is the same in all *v* mutant strains, this probe provides a standard against which the *v* signal can be normalized to correct for RNA loading and recovery differences. To quantitate the level of v^{+37} transcript relative to wild type, densitometric analysis was performed on autoradiographs from four different S1 nuclease protection experiments like the ones shown in Figure 4, and the average relative amounts determined. The level of transcript accumulation in the revertant was estimated to be 24% (SD = 4) of the wild-type level. The $su(s)^{X1}$ v^1 mutant accumulates approximately the same amount of *v* RNA as v^{+37} (Figure 4A, compare lanes 5 and 7). The signal obtained from v^1 mutant RNA (Figure 4A, lane 3; Figure 4B, lane 3) was clearly detectable in longer exposures but was too faint for proper densitometric analysis. Therefore, the level of v^1 RNA accumulation was determined by comparing the autoradiographic exposures required to obtain signals of equal intensity from $su(s)^{X1}$ v^1 and v^1 mutant RNAs. Using this approach, we estimate that the v^1 RNA level is approximately 5-fold lower than the level found in a suppressed v^1 mutant or about 5% of the wild-type level.

To determine how *su(s)* mutations affect expression of the v^{+37} allele, we obtained (from R. VOELKER) strains that carry both the v^{+37} allele and one of three *su(s)* alleles on the same chromosome. The $su(s)^{X1}$ mutation was X-ray induced but the defect is not detectable by Southern analysis (R. VOELKER, personal communication). The $su(s)^{r6}$ allele was ethyl methanesulfonate (EMS)-induced and is, therefore, likely to be a point mutation. The $su(s)^{51j}$ allele has a *gypsy* element inserted into the first intron of the *su(s)* gene (VOELKER *et al.* 1990). Both $su(s)^{X1}$ and $su(s)^{r6}$ affect expression of v^{+37}. The $su(s)^{X1}$ allele increases *v* RNA accumulation in the v^{+37} strain from 24% to 50% (SD = 6) of wild type (Figure 4A, compare lanes 5 and 6; Figure 4B, compare lanes 4 and 7). Similarly, the $su(s)^{r6}$ allele increases the amount of v^{+37} RNA to 59% (SD = 11) of the wild-type level (Figure 4B, compare lanes 4 and 6). However, the level of RNA accumulation in the $su(s)^{51j}$ v^{+37} strain, 34% (SD = 11) of wild type, is not significantly higher than the v^{+37} strain

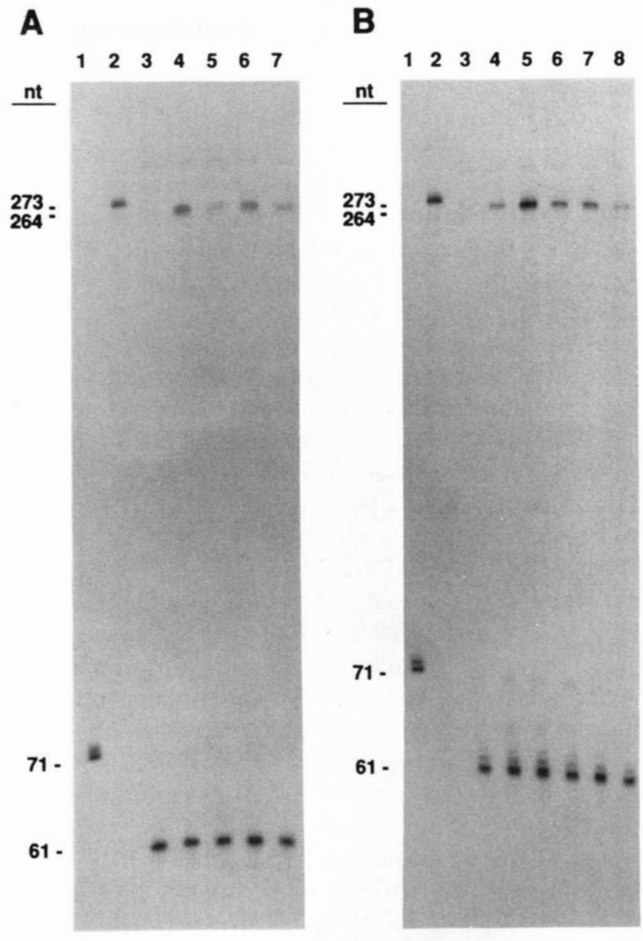

FIGURE 4.—Quantitative S1 protection analysis of wild-type and mutant *v* mRNA. A 273-nt antisense v^+ cDNA probe and a 71-nt antisense *rp49* probe were simultaneously annealed to 20 μg of poly(A)$^+$ RNA, and the reactions were subsequently treated with S1 nuclease. Sequencing reactions were run in parallel lanes to provide size standards (not shown). The approximate sizes of the protected fragments are shown in nt units. Panels A and B represent separate experiments. (Panels A and B, lane 1) *rp49* probe alone, no S1 treatment; (panels A and B, lane 2) *v* cDNA probe alone, no S1 treatment. The remaining lanes contain samples hybridized with both probes and treated with S1 nuclease but differ with respect to the source of RNA. The RNA samples were isolated from the following genotypes. (A) (lane 3) v^1; (lane 4) v^+; (lane 5) v^{+37}; (lane 6) $su(s)^{X1}$ v^{+37}; (lane 7) $su(s)^{X1}$ v^1. (B) (lane 3) v^1; (lane 4) v^{+37}; (lane 5) v^+; (lane 6) $su(s)^{r6}$ v^{+37}; (lane 7) $su(s)^{X1}$ v^{+37}; (lane 8) $su(s)^{51j}$ v^{+37}.

(Figure 4B, compare lanes 4 and 8). These results are consistent with studies that examined the cold-sensitive male sterility associated with *su(s)* mutants (VOELKER *et al.* 1989). The $su(s)^{X1}$ and $su(s)^{e6}$ alleles impair male fertility to a greater extent than $su(s)^{51j}$. Thus, whereas *su(s)* mutations increase the level of v^1 RNA by at least a factor of five, the effect of *su(s)* mutations on expression of the revertant is no more than a twofold increase. Sequence analysis of seven *v* cDNA clones prepared from $su(s)^{X1}$ v^{+37} RNA showed that the same donor (dV) and acceptor (aI) sites are used to splice the revertant transcripts made in a *su(s)* mutant background (data not shown).

Inefficient splicing of the first *v* intron downstream of the *412* insertion in v^1 mutant derivatives: While sequencing *v* cDNA clones derived from the *412* insertion mutants, we observed that a significant proportion of the cDNAs retained the first *v* intron downstream of the *412* insertion. These observations suggested that the *412* insertion interferes with splicing of that intron. Failure to remove the first *v* intron would cause translation to terminate prematurely at a stop codon in the intron. Thus, transcripts of this type would produce a nonfunctional polypeptide. Furthermore, from the analysis of cDNA clones it appeared that the intron was retained in a larger proportion of clones derived from a $su(s)^+$ background than from *su(s)* mutants (see FRIDELL, PRET and SEARLES 1990). However, since some of the PCR-generated cDNAs may not have been independently derived, the analysis of cDNAs might have been somewhat misleading. Therefore, we employed S1 nuclease mapping to quantitate the proportion of properly spliced *vs.* unspliced mRNA in various mutant backgrounds (Figure 5). A 344-nt v^+ riboprobe was generated that includes the 60 nt of first exon sequences downstream of the *Xho*I site, the 73-nt first intron, and the entire 161-nt second exon. If wild-type transcripts are spliced, protected exon fragments of 161 nt and 60 nt would be produced. On the other hand, unspliced wild-type transcripts would generate a 294-nt protected fragment (Figure 5), whereas unspliced mutant transcripts should generate a 281-nt protected fragment, representing protection of sequences extending 5′ to the *412* insertion site (Figure 5). However, sometimes the 294-nt fragment is also observed when *v* mutant RNA with residual *412* sequences is annealed to the wild-type *v* probe, due to the failure of S1 nuclease to cleave efficiently the site opposite a mismatched heteroduplex loop (for a detailed explanation, see FRIDELL, PRET and SEARLES 1990).

We compared the relative amounts of spliced 161-nt exon to that of unspliced 294-nt and/or 281-nt fragment to get an indication of the efficiency of splicing in this region of the gene. Figure 5 illustrates the results of this analysis with poly(A)$^+$ RNA from

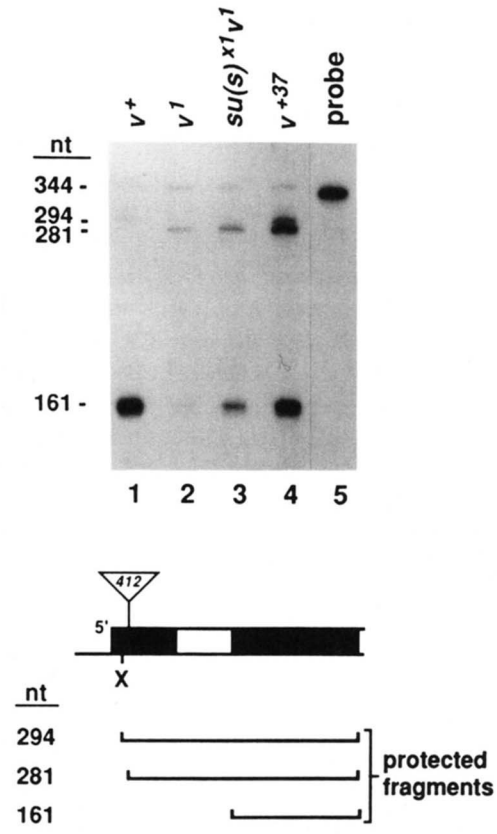

FIGURE 5.—S1 protection analysis of *v* first intron splicing in wild-type and mutant strains. A 344-nt antisense RNA probe, derived from a wild-type genomic clone and containing 294 nt of *v* genomic sequences, was annealed to 20 μg of adult poly(A)$^+$ RNA and treated with S1 nuclease. Lanes 1–4 contain RNA samples isolated from the strains indicated above the lanes. Lane 5 contains a sample of the probe alone with no S1 nuclease treatment. Sequencing reactions were run in parallel lanes to provide size standards (not shown). The lengths of the protected fragments are indicated in nt units. The map beneath the autoradiogram illustrates the region included in each protected fragment. Solid bars on the molecular map represent the first two *v* exons, and the open bar represents the first *v* intron. The open triangle indicates the position of the *412* insertion immediately downstream of the *Xho*I (X) site. The 294- and 281-nt protected fragments are derived from unspliced transcripts whereas the 161-nt fragment is derived from spliced transcripts.

different genetic backgrounds. The region of the gel containing the 60-nt protected fragment is not shown. Whereas the vast majority of v^+ transcripts are properly spliced (Figure 5, lane 1), all of the v^1 mutant derivatives accumulate a significant proportion of transcripts which retain the 73-nt *v* intron (Figure 5, lanes 2–4). We performed densitometric analysis on the gel shown in Figure 5 and determined the ratio of spliced to unspliced protected fragments, taking into account the fact that the larger protected fragment would be expected to contain about 1.8 times as much radioactivity as the smaller 161-nt fragment. The ratio of spliced/unspliced RNA in v^1 (Figure 5, lane 2) was determined to be 0.7. However, for $su(s)^{X1}$ v^1 and v^{+37} RNAs (Figure 5, lanes 3 and 4) the spliced/

unspliced ratios are 3 and 2.5, respectively. These results were consistently observed in three different experiments involving different preparations of RNA. Other experiments showed that the remaining *v* introns are spliced completely in these mutants. Furthermore, data obtained with v^k RNA were similar to the v^l mutant results (data not shown). Thus, it appears that the *412* insertion interferes with splicing of the first *v* intron downstream of the insertion site. Therefore, not all of the *v* RNAs that accumulate in mutant strains can produce functional protein. In addition, this analysis suggests that both suppressor and reversion mutations affect the splicing of this small intron.

DISCUSSION

Each suppressible *v* mutant has an identical 7.4-kb insertion of the *412* retrotransposon 36 nt downstream of the major transcription start site in an 83-nt exon (FRIDELL, PRET and SEARLES 1990). The transposon insertion in this position behaves essentially as if it were a large intron in that most of the transposon sequences are removed from the pre-mRNA by splicing at cryptic sites. In a $su(s)^+$ background, the level of spliced mutant transcripts is very low, and we previously showed that recessive *su(s)* mutations increase the accumulation of these transcripts (SEARLES *et al.* 1990; FRIDELL, PRET and SEARLES 1990). In this report we have demonstrated that a particular intragenic reversion event is the result of a secondary insertion of another transposon, the *B104/roo* element, near one end of the *412* element. The second insertion leads to an increase in *v* RNA accumulation even in a $su(s)^+$ background, and the revertant transcripts are spliced from different donor sites in the *B104/roo* element to the same acceptor site in the *412* element. Thus, the majority of both transposons are removed from the 5′ untranslated region v^{+37} mRNA.

It is surprising that splicing occurs at the observed sites within *412* and *B104/roo* because these are not authentic splice sites as they are derived from the nontranscribed strand of both transposons. While the sequences recognized as splice sites in higher eukaryotes can deviate considerably from the consensus, the many sequences that resemble splice sites are not usually recognized by the splicing machinery. The mechanism of splice site recognition is an important unresolved issue in splicing, and some experimental evidence suggests that sequences flanking a splice site modulate splicing activity (REED and MANIATIS 1986; NELSON and GREEN 1988; NAGOSHI and BAKER 1990). One explanation for splicing at these cryptic sites could be that the *412* and *B104* elements at *v* inserted into a context that allows splicing to occur. For example, suppose that *v* sequences within the first *v* exon are important for recognition of the downstream 5′ donor site. If the transposons inserted in such a position that the relevant *v* first exon sequences are now immediately upstream of transposon sequences, then recognition of nearby transposon sequences that resemble splice donor sites might be enhanced. This model would also predict that the insertions would disrupt the context needed for efficient splicing of the first *v* intron downstream of the insertion, and might explain why splicing of the first authentic *v* intron is impaired. However, an alternative explanation for inefficient splicing of the first *v* intron could be that inclusion of the large *412* insertion in the precursor RNA alters the secondary structure of the transcript, making the downstream intron inaccessible to the splicing machinery.

The finding that revertant transcripts are spliced at new donor sites suggests that this change in the splicing pattern is the cause of the higher level of *v* mRNA accumulation in the revertant strain. If the *B104/roo* insertion introduces splice donor sites that are recognized at a higher efficiency, the transposon-containing mRNA precursors might be more readily targeted toward a splicing pathway rather than being degraded. Recognition of the 5′ splice site appears to be an early step in formation of a stable, committed splicing complex in yeast (SERAPHIN and ROSBASH 1989; RUBY and ABELSON 1988). In comparing the cryptic transposon donor sites to other invertebrate donor sites (see RESULTS), we found that the *B104/roo* donor sites appear to be slightly better than the preferentially used *412* donor sites. Whether this difference is sufficient to support the notion that the *B104* sites are more efficiently recognized is not clear. Our analysis of cDNA clones indicates that most transcripts are spliced at the donor site (dV) closest to the 5′ end of the revertant pre-mRNA. Thus, this donor site might be located in a better context for splicing. Alternatively, if splicing is initiated during transcription, donor site V might be preferred because it is synthesized first. It is interesting to note that although the *B104/roo* insertion disrupts *412* donor sites I and II, donor sites III and IV are intact in v^{+37} DNA. Yet, we have no indication that splicing occurs at these donor sites within *412* when the *B104/roo* donor sites are present. Perhaps that is because when moved 3.6 kb downstream by the *B104/roo* insertion, these sites are no longer in a context that is favorable for splicing. Thus, while we cannot rule out the possibility that reversion is due to an effect on transcription initiation or RNA stability, we believe that the mechanism of reversion is more likely to involve splicing since the transcript splicing pattern is altered in v^{+37}.

We have not succeeded in detecting the transposon-containing pre-mRNAs, perhaps because they are unstable. Since we are unable to assess how precursor

levels vary in different backgrounds, it is impossible to determine if suppressor and reversion mutations alter the efficiency of splicing the "transposon-intron." However, we have shown that the transposon insertions impair splicing of the *v* intron downstream of the insertion and have detected both spliced and unspliced forms of the RNA. The observation that suppression and reversion mutations appear to increase the efficiency of splicing this small intron suggests that both mutations exert an effect at the level of splicing. We have also shown that the revertant allele does respond to *su(s)* mutations. The steady state level of revertant *v* transcripts increases by approximately a factor of two in a *su(s)* mutant background. This result might be expected if the reversion introduces a somewhat better splice donor site, the recognition of which can be enhanced slightly by *su(s)* mutations. However, a similar effect might be observed if *su(s)* mutations directly increase stability of the transposon-containing precursor RNA and the reversion allows more efficient elimination of transposon sequences by affecting splicing.

Our experiments suggest that *su(s)* mutations increase the steady state level of v^1 mRNA as well as the proportion of transcripts that are correctly spliced. Thus, although the v^1 mutant accumulates about 5% of the wild-type level of RNA, only about 40% of those transcripts have the 73-nt first *v* intron removed. Therefore, the level of functional *v* mRNA in the mutant is about 2% of the wild-type level. On the other hand, in $su(s)^- v^1$ double mutants, the level of RNA accumulation increases to 24% of the wild-type level and about 80% of the transcripts are correctly spliced. Thus, in this case the level of functional mRNA is about 20% of the wild-type level. Both of these results are consistent with our hypothesis that *su(s)* mutations affect recognition of poor splice sites. Recognition of the cryptic *412* splice sites and/or the *v* splice site(s) immediately downstream of the *412* insertion might be affected by *su(s)* mutations. Enhanced recognition might facilitate assembly of the transcript in a splicing complex to stabilize the precursor RNA and/or to facilitate the rate of intron splicing. Similar effects on RNA accumulation and first *v* intron splicing were observed in the revertant strain. In this case, it is more difficult to imagine how the *B104/roo* insertion might increase the efficiency of splicing the first downstream *v* intron that is so far removed. Perhaps the effect is indirect, *i.e.*, improved splicing of transposon sequences in the revertant might somehow lead to more efficient splicing of the next intron downstream. On the other hand, the *B104/roo* insertion might produce an alteration in the secondary structure of the precursor RNA that results in enhanced splicing of the downstream intron.

We thank R. VOELKER for kindly providing Drosophila stocks and R. S. RUTH for cloning the v^{+37} allele. We also thank G. MARONI, P. PUKKILA and R. FRIDELL for critically reading the manuscript, S. WHITFIELD for the art work, and A. BUTCHER for typing. This work was supported by grants DMB-8904471 and DMB-9004708 from the National Science Foundation.

LITERATURE CITED

CHANG, D., WISELY, S. HUANG and R. A. VOELKER, 1986 Molecular cloning of suppressor of sable, a *Drosophila melanogaster* transposon-mediated suppressor. Mol. Cell. Biol. **6:** 1520–1528.

FRIDELL, R. A., A. M. PRET and L. L. SEARLES, 1990 A retrotransposon 412 insertion within an exon of the *Drosophila melanogaster* vermilion gene is spliced from the precursor RNA. Genes Dev. **4:** 559–565.

GERMERAAD, S., 1976 Genetic transformation in *Drosophila* by microinjection of DNA. Nature **262:** 229–231.

GEYER, P. K., M. M. GREEN and V. G. CORCES, 1988a Reversion of a gypsy-induced mutation at the yellow (*y*) locus of *Drosophila melanogaster* is associated with the insertion of a newly defined transposable element. Proc. Natl. Acad. Sci. **85:** 3938–3942.

GEYER, P. K., M. M. Green and V. G. CORCES, 1988b Mutant gene phenotypes mediated by a *Drosophila melanogaster* retrotransposon require sequences homologous to mammalian enhancers. Proc. Natl. Acad. Sci. **85:** 8593–8597.

GEYER, P. K., C. SPANA and V. G. CORCES, 1986 On the molecular mechanism of gypsy-induced mutations at the *yellow* locus of *Drosophila melanogaster*. EMBO J. **5:** 2657–2662.

HARRISON, D. A., P. K. GEYER, C. SPANA and V. G. CORCES, 1989 The gypsy retrotransposon of *Drosophila melanogaster*: mechanisms of mutagenesis and interaction with the *suppressor of Hairy-wing* locus. Dev. Genet. **10:** 239–248.

MANIATIS, T., E. F. Fritsch and J. SAMBROOK, 1982 *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

MEYEROWITZ, E. M., and D. S. HOGNESS, 1982 Molecular organization of a *Drosophila* puff site that responds to ecdysone. Cell **28:** 165–176.

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 **188:** 221–234.

NAGOSHI, R., and B. BAKER, 1990 Regulation of sex-specific RNA splicing at the *Drosophila* doublesex gene: *cis*-acting mutations in exon sequences alter sex-specific RNA splicing patterns. Genes Dev. **4:** 89–97.

NELSON, K., and M. GREEN, 1988 Splice site selection and ribonucleoprotein complex assembly during *in vitro* pre-mRNA splicing. Genes Dev. **2:** 319–329.

O'CONNELL, P., and M. ROSBASH, 1984 Sequence, structure, and codon preference of the *Drosophila* ribosomal protein 49 gene. Nucleic Acids Res. **12:** 5495–5513.

O'HARE, K., R. LEVIS and G. RUBIN, 1983 Transcription of the white locus in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **80:** 6917–6921.

PARKHURST, S. M., AND V. G. CORCES, 1986 Interactions among the gypsy transposable element and the *yellow* and the suppressor of Hairy-wing loci in *Drosophila melanogaster*. Mol. Cell. Biol. **6:** 47–53.

REED, R., and T. MANIATIS, 1986 A role for exon sequences and splice-site proximity in splice-site selection. Cell **46:** 681–690.

RUBY, S. W., and J. ABELSON, 1988 An early role of U1 small nuclear ribonucleoprotein in spliceosome assembly. Science **242:** 1028–1035.

SAIKI, R. K., D. H. Gelfand, S. Stoffel, S. J. Scharf, R. Higuchi, G. T. Horn, K. B. Mullis and H. A. ERLICH, 1988 Primer-

directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239:** 487–491.

SCHERER, G., C. Tschudi, J. Perera, H. Delius and V. PIRROTTA, 1982 B104, a new dispersed repeated gene family in *Drosophila melanogaster* and its analogies with retroviruses. J. Mol. Biol. **157:** 435–451.

SEARLES, L. L., and R. A. VOELKER, 1986 Molecular characterization of the *Drosophila vermilion* locus and its suppressible alleles. Proc. Natl. Acad. Sci. USA **83:** 404–408.

SEARLES, L. L., R. S. RUTH, A. M. PRET, R. A. FRIDELL and A. J. ALI, 1990 Structure and transcription of the *Drosophila melanogaster* vermilion gene and several mutant alleles. Mol. Cell. Biol. **10:** 1423–1431.

SERAPHIN, B., and M. ROSBASH, 1989 Identification of functional U1 snRNA-pre-mRNA complexes committed to spliceosome assembly and splicing. Cell **59:** 349–358.

SHAPIRO, M., and P. SENEPATHY, 1987 Splice junctions of different classes of eukaryotes: Sequence statistics and functional implications in gene expression. Nucleic Acids Res. **15:** 7155–7175.

SOUTHERN, E. M., 1975 Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98:** 503–517.

SPANA, C., D. A. HARRISON and V. G. CORCES, 1988 The *Drosophila melanogaster* suppressor of Hairy-wing protein binds to specific sequences of the gypsy retrotransposon. Genes Dev. **2:** 1414–1423.

VOELKER, R. A., S.-M. HUANG, G. B. WISELY, J. F. STERLING, S. P. BAINBRIDGE and K. HIRAIZUMI, 1989 Molecular and genetic organization of the *suppressor of sable* and *Minute(1)1B* region in *Drosophila melanogaster*. Genetics **122:** 625–642.

VOELKER, R. A., J. GRAVES, W. GIBSON and M. EISENBERG, 1990 Mobile element insertions causing mutations in the *Drosophila* suppressor of sable locus occur in DNase I hypersensitive subregions of 5′ transcribed nontranslated sequences. Genetics **126:** 1071–1082.

VOELKER, R. A., W. Gibson, J. P. Graves, J. F. Sterling and M. EISENBERG, 1991 The *Drosophila* suppressor of sable gene encodes a 150kD polypeptide with regions similar to those of RNA binding proteins. Mol. Cell. Biol. **11:** 894–905.

WILLIAMS, J. A., and J. B. BELL, 1988 Molecular organization of the vestigial region in *Drosophila melanogaster*. EMBO J. **7:** 1355–1363.

Communicating editor: M. J. SIMMONS