

A retrotransposon 412 insertion within an exon of the *Drosophila melanogaster* *vermilion* gene is spliced from the precursor RNA

Robert A. Fridell, Anne-Marie Pret, and Lillie L. Searles¹

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

Three alleles of the *Drosophila melanogaster* *vermilion* (*v*) gene are suppressed by recessive mutations at the *suppressor of sable* [*su(s)*], gene. Previous work has established that these alleles have identical insertions of the 412 retrotransposon in the 5'-untranslated region of the gene. Despite the transposon insertion in an exon, *v* mutants accumulate trace amounts of apparently wild-type-sized transcripts in a *su(s)*⁺ background, and the level of *v* transcript accumulation is increased by *su(s)* mutations. Here, we have characterized transcripts from a suppressible *v* mutant in both *su(s)*⁺ and *su(s)*⁻ backgrounds by S1 nuclease protection experiments and sequence analysis of polymerase chain reaction (PCR) generated cDNA clones. We find that transposon sequences are imprecisely eliminated from *v* mutant transcripts by splicing at donor and acceptor sites located near the ends of the 412 retrotransposon. Four different 5' donor sites are alternatively spliced to a single 3' acceptor site. The implications of this finding are discussed in relation to possible functions of the *su(s)*⁺ gene product.

[Key Words: Suppression; retrotransposon; 412 element; splicing]

Received November 9, 1989; revised version accepted January 31, 1990.

In *Drosophila melanogaster*, the phenotypic effects of certain retrotransposon-induced mutations are reversed by recessive mutations at separate genetic sites, known as suppressor loci (for review, see Kubli 1986). In recent years, an understanding at the molecular level of the relationship between suppressor genes and suppressible alleles has begun to emerge. For example, a mutation of the *Drosophila yellow* (*y*) gene, caused by insertion of a gypsy element upstream of the *y* promoter, is suppressed by mutations at *suppressor of Hairy-wing* [*su(Hw)*] (Parkhurst and Corces 1985). The mutagenic effect of gypsy in this case is believed to result from transcriptional interference between the gypsy element and the *y* promoter (Harrison et al. 1989). Parkhurst et al. (1988) and Spana et al. (1988) have shown that *su(Hw)* encodes a DNA-binding protein that recognizes a specific sequence in the gypsy transposon. This work and transcriptional studies by Mazo et al. (1989) suggest that the *su(Hw)* protein acts as a positive regulator of gypsy transcription. Another suppressor gene, the *suppressor of white-apricot* [*su(w^a)*], suppresses the phenotypic effect produced by insertion of a copia element into an intron of the *white* gene (Bingham et al. 1981). Characterization of *su(w^a)* has revealed that the gene encodes a putative RNA-binding protein that regulates its own expression at the level of splicing (Chou et al. 1987; Zachar et

al. 1987). This suggests that suppression of the *w^a* allele might operate at the level of RNA processing.

We are investigating alleles of the *Drosophila vermilion* (*v*) gene that are suppressed by mutations at *suppressor of sable* [*su(s)*]. The *v* gene encodes tryptophan oxygenase, an enzyme required for brown eye pigment synthesis in *Drosophila* (Walker et al. 1986). Three suppressible *v* alleles, *v¹*, *v²*, and *v^k*, have insertions of the 412 retrotransposon at the same site, in the 5'-untranslated region of the first exon (Fig. 1; Searles and Voelker 1986; Searles et al. 1990). In each mutant the transcriptional orientation of the 412 element is opposite that of *v* (Fig. 1). Northern analysis has shown that the 1.4-kb transcript present in *v⁺* flies is largely absent in *v¹*, *v²*, and *v^k* flies, where only a trace of a wild-type-sized transcript is observed (Searles et al. 1990). However, in flies homozygous for one of these *v* alleles and also mutant at the *su(s)* locus, a wild-type-sized *v* transcript is present at ~10–20% of the *v⁺* level (Searles et al. 1990). Here, we report the results of experiments designed to determine the mechanism by which the normal length transcripts are generated.

Results

Mutant v transcripts include sequences upstream of the 412 insertion site

In a *su(s)*⁻ background, the *v¹*, *v²*, and *v^k* mutants accu-

¹Corresponding author.

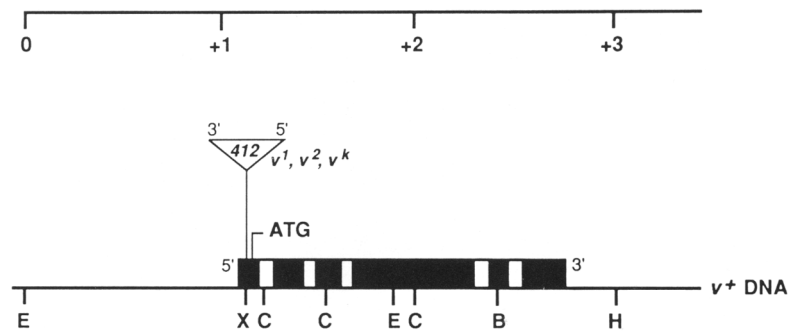


Figure 1. Molecular map of the *v* gene. Solid bars represent exons; open bars represent introns. The open inverted triangle indicates the position of transposon 412 insertion within the *v* transcription unit. The ATG codon indicates the beginning of the long open reading frame. The map coordinates above the restriction map are in kilobase units. The restriction enzyme abbreviations are (B) *Bam*HI; (E) *Eco*RI; (C) *Cl*aI; (H) *H*indIII; (X) *X*hoI.

mulate reduced levels of wild-type-sized transcripts despite the insertion of a 7.5-kb transposon in the first exon (Searles et al. 1990). These transcripts might be generated by one of two mechanisms. If transcription of the mutant gene initiates at the *v* promoter, transposon sequences would have to be eliminated from the mRNA precursor. Alternatively, transcription of *v* in the insertion mutant might initiate at a new site within the 412 element near the *v*-coding region. To distinguish between these possibilities, we set out to determine whether sequences upstream of the 412 insertion site are included in *v* mRNA isolated from a *su(s)* *v* double mutant. The location of the transposon insertion in a small (83-nucleotide) exon only 36 nucleotides downstream of the major transcription start site (Searles et al. 1990) complicated the resolution of this question. However, S1 nuclease protection experiments with cloned wild-type *v* cDNA probes proved to be more informative than we had anticipated. We had isolated several cDNA clones, representing a minor species of the *v* RNA that initiates at a site upstream of the major transcription start site but that is otherwise identical to the major species of *v* mRNA (Searles et al. 1990). Radioactively labeled antisense RNA probes were synthesized from one of these clones and annealed to poly(A)⁺ RNA derived from several different genotypes, and the fragments protected from digestion by S1 nuclease were examined (Fig. 2).

Probe **a** (Fig. 2) includes cDNA sequences from the *Cl*aI site in the third *v* exon to a site in the 5' region that is ~50 nucleotides upstream of the major transcription start site. When this probe is annealed to wild-type poly(A)⁺ RNA, a 320-nucleotide fragment, including cDNA sequences from the major transcription start site to the *Cl*aI site, is the predominant protection product (Fig. 2, lane 3; see Searles et al. 1990). Probe **b** (Fig. 2) has the same 3' boundary as probe **a** but includes 5' *v* sequences up to the *X*hoI site in the first *v* exon. This *X*hoI site is ~10 nucleotides upstream of the 412 insertion site (Searles et al. 1990). A 295-nucleotide fragment is protected from S1 digestion when wild-type RNA is annealed to probe **b** (Fig. 2, lane 4).

The same two cDNA probes were annealed to RNA isolated from a *su(s)*^{51c15} *v*^k mutant background (Fig. 2,

lanes 5 and 6). With each probe we observed two protected fragments: one that is the same size as the fragment protected by wild-type RNA and the other, a smaller 285-nucleotide fragment that represents the protection of sequences between the *Cl*aI site in the third exon and the 412 insertion site. The 320-nucleotide fragment, which includes sequences upstream of the 412 insertion site, predominates when probe **a** is used (Fig. 2, lane 5), whereas the 285-nucleotide fragment, which includes sequences up to the 412 insertion site, is more prevalent when probe **b** is used (Fig. 2, lane 6). We believe the most likely interpretation of these results is as follows. If the mutant *v* RNAs contain sequences upstream and downstream of the 412 insertion site plus additional sequences not found in the wild-type cDNAs, the heteroduplex between wild-type cDNA probes and mutant mRNA would contain an internal loop on the RNA strand. S1 nuclease apparently does not efficiently cleave the site on the strand opposite the loop which, in this case, is the radioactively labeled strand (see Sisodia et al. 1987). When cleavage does not occur, the wild-type-sized (320- or 295-nucleotide) fragment is produced. On the other hand, when cleavage does occur opposite the loop, the smaller (285-nucleotide) fragment is produced. Digestion of sequences upstream of the insertion site might be more efficient in heteroduplexes of probe **b** and mutant mRNA because this probe has only 10 nucleotides upstream of the 412 insertion site (see Fig. 2). Weaker base-pairing upstream of the insertion site would result in those sequences being more vulnerable to cleavage by S1 nuclease. Thus, the appearance of the wild-type-sized protected fragment indicated that sequences upstream of the transposon insertion site were being protected and therefore must be included in the mutant RNA. These results suggested that at least some of the mutant transcripts were generated by a processing event, involving removal of 412 sequences.

Northern analysis previously indicated that the *v* 412 insertion mutant strains accumulate a very low level of wild-type-sized RNA in the absence of a suppressor mutation (Searles et al. 1990). Therefore, we performed the same S1 mapping analysis with RNA from a *su(s)*⁺ *v*^k background. The signals were very faint, but the pattern was the same as that observed with RNA from the

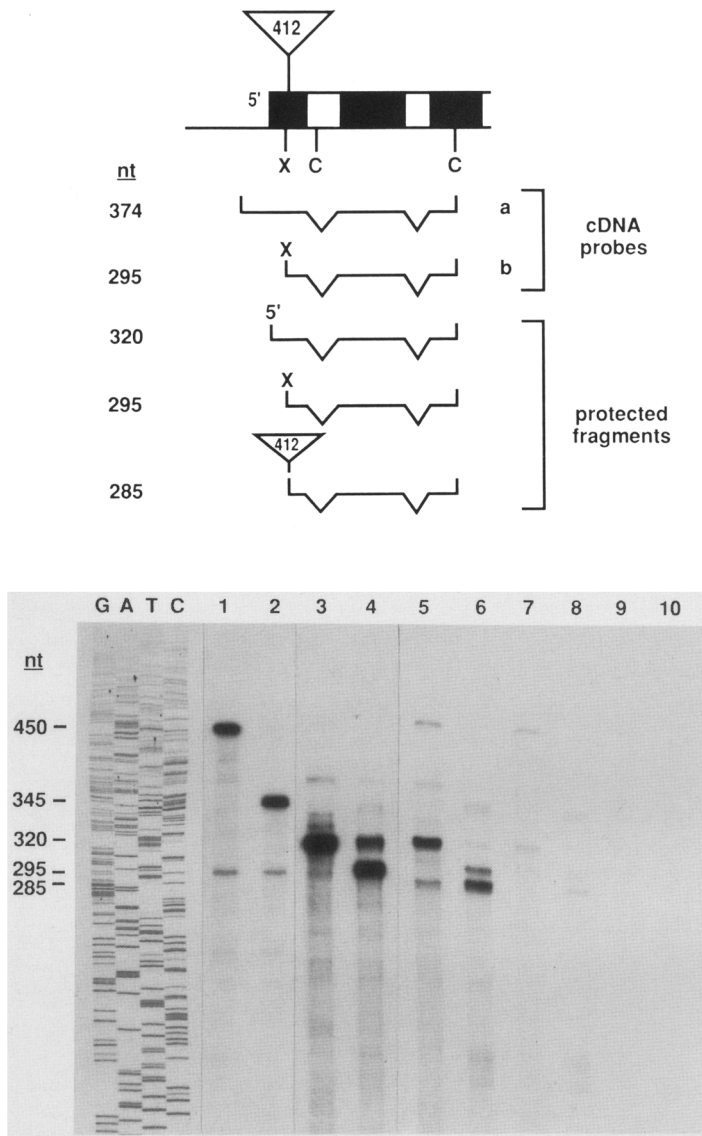


Figure 2. S1 nuclease protection analysis of wild-type and mutant *v* transcripts. Antisense RNA probes **a** and **b**, derived from a wild-type *v* cDNA clone, were annealed to poly(A)⁺ RNA and subsequently treated with S1 nuclease. The reactions contained 10 μ g of *v*⁺ RNA isolated from second instar larvae, 20 μ g of *su(s)*⁻ *v*^{*} RNA from early third-instar larvae, and 20 μ g of *v*^{*} RNA from early third instar larvae. Sequencing reactions of M13mp18 DNA were run in parallel lanes to provide size standards. The approximate sizes of the protected fragments are shown in nucleotide units. The *v*⁺ sequences included in probes **a** and **b** and the regions included in protected fragments are indicated by the horizontal lines beneath the restriction map. The symbols used in the *v* molecular map are described in Fig. 1. The 450-nucleotide probe **a** contains 374 nucleotides of *v* cDNA sequences plus 76 additional nucleotides either generated during the cDNA library construction or derived from the SP6 transcription vector. The 345-nucleotide probe **b** consists of 295 nucleotides of *v* cDNA sequences plus 50 nucleotides of SP6 transcribed vector sequences. (Lane 1) Probe **a** alone, no S1 nuclease; (lane 2) probe **b** alone, no S1 nuclease; (lane 3) *v*⁺ RNA plus probe **a**; (lane 4) *v*⁺ RNA plus probe **b**; (lane 5) *su(s)*⁻ *v*^{*} RNA plus probe **a**; (lane 6) *su(s)*⁻ *v*^{*} RNA plus probe **b**; (lane 7) *v*^{*} RNA plus probe **a**; (lane 8) *v*^{*} RNA plus probe **b**; (lane 9) no RNA control, probe **a**; (lane 10) no RNA control, probe **b**. The 320-nucleotide protected fragment in the reaction containing *v*⁺ RNA and probe **b** (lane 4) is probably caused by slight contamination with probe **a**. The clone used to generate these probes has two *Xho*I sites: one in *v* sequences and the other in the polylinker adjacent to the 5' end of the clone. The enzyme *Xho*I is used to linearize a template for probe **b** synthesis, and if this digestion does not go to completion, probe **a** is synthesized in addition to probe **b**. A longer exposure of this experiment indicates that there is a small amount of probe **a** in the lane containing probe **b** (data not shown).

su(s)⁻ *v*^{*} mutant (Fig. 2, lanes 7 and 8). This suggested that unsuppressed mutant transcripts might also result from a processing event.

Mutant *v* RNAs are spliced at sites within the 412 retrotransposon

To expand on our S1 mapping results, we compared the mature *v* transcript produced in a *su(s)*⁻ *v*^{*} background with the *v*⁺ mRNA by isolating and sequencing mutant *v* cDNA clones. Because these transcripts are of such low abundance, we used polymerase chain reaction (PCR) to amplify cDNA fragments. A synthetic oligonucleotide that includes the *Eco*RI site in the fourth *v* exon (Fig. 1) was used to direct specific first-strand cDNA synthesis on poly(A)⁺ RNA isolated from *su(s)*^{51c15} *v*^{*} late embryo and larval stages. The cDNAs were amplified in

PCR reactions primed with an oligonucleotide covering the *Xho*I site in the first *v* exon (Fig. 1) and the same oligonucleotide used in the first-strand cDNA synthesis. The PCR products were subsequently cloned into a plasmid vector and partially sequenced.

The results of the sequence analysis are summarized in Figure 3. Each cDNA clone contained *v* sequences upstream and downstream of the insertion site and the 4-nucleotide target site duplication, CACG, generated upon 412 insertion. Eleven to 54 nucleotides of additional sequences, derived from the 412 element, were found inserted between the 4-nucleotide duplication. We compared the sequences in these cDNAs to the published sequence of the 481-nucleotide long terminal direct repeat (LTR) of 412 (Will et al. 1981) and deduced that sequences from both ends of the 412 element were spliced together at sites that resemble pre-mRNA splice

Fridell et al.

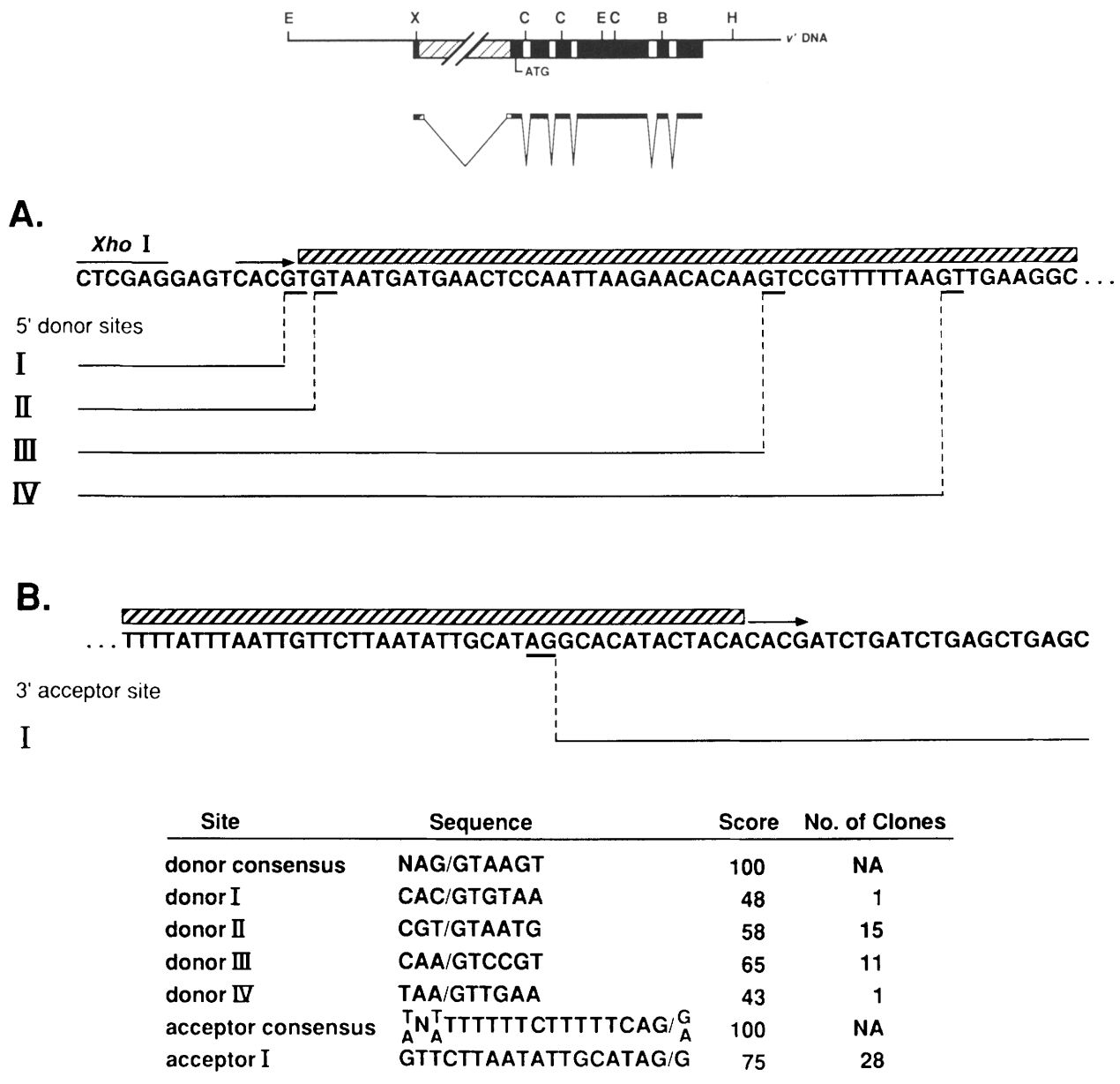


Figure 3. Analysis of PCR-amplified v^k cDNA clones derived from a $su(s)^-$ background. The molecular map of v^k DNA is shown with the inserted 412 transposon represented by the hatched bar. The other symbols are described in Fig. 1. The splicing pattern of mutant transcripts is schematically diagrammed beneath the molecular map. *A* and *B* illustrate the sequence of v^k DNA in the vicinity of the 412 splice sites. Transposon 412 sequences are indicated by the bar with diagonal lines above the sequence. The 4-nucleotide direct repeat generated upon 412 insertion is indicated by the arrows. The splice sites are indicated by the dotted vertical lines. The horizontal lines beneath the sequence indicate regions included in the cDNAs. (*A*) Positions of 5' splice donor sites. The obligatory GT dinucleotide at the 5' splice junction is underlined. (*B*) Position of the 3' acceptor site. The obligatory AG dinucleotide at the 3' splice junction is underlined. A summary list of the splice junction sequences and the invertebrate consensus splice sequences is given (Shapiro and Senapathy 1987). The scoring method has been described by Shapiro and Senapathy (1987) and indicates the extent of similarity between these splice sites and other known invertebrate splice sites (see Discussion).

junctions. The splice donor and acceptor sites are located near the ends of the element within the LTRs (Fig. 3). The 412 element at v is oriented such that the 5' LTR of 412 is adjacent to the v -coding region and the 3' LTR is near the v promoter region (see Fig. 1). Thus, the 412 sequences included in the v precursor RNA are complementary to the normal 412 transcript.

In the 28 $su(s)^-$ v^k cDNAs that were sequenced, four different 5' donor sites were alternatively spliced to a single 3' acceptor site (Fig. 3). One donor site was created by the insertion, and the splice occurs -1 nucleotide upstream of the 412 insertion site. The other splice junctions are located $+1$, $+30$, and $+42$ nucleotides from the left end of the element (in the 3' LTR). The

majority of clones were spliced at donor site II (CGT/GTAATG) or III (CAA/GTCCGT) (Fig. 3). Donor sites I (CAC/GTGTA) and IV (TAA/GTTGAA) were represented in only one clone (Fig. 3). All 28 clones were spliced at the same 3' acceptor site (GTTCTTAAT-TATTGCATAG/G), located 12 nucleotides from the right end of the 412 insertion (in the 5' LTR) (Fig. 3). Three of the 28 cDNA clones contained the first *v* intron downstream of the 412 insertion.

The S1 nuclease mapping experiments described earlier indicated that the very low level of *v* mRNA produced in a *su(s)⁺ v^k* background is also processed to eliminate 412 sequences. To determine whether this transcript is produced by similar splicing events, we used the same PCR amplification approach to isolate *v* cDNA clones from *v^l* and *v^k* mutants in a *su(s)⁺* background. Six cDNA clones have been recovered, and sequence analysis demonstrated that these clones represented mRNAs that were spliced at the 412 donor and acceptor sites identified previously. The single *v^l* cDNA clone was spliced at donor site III, and the five *v^k* cDNA clones were spliced at donor site II. Three of the *v^k* cDNAs retained the first *v* intron downstream of the 412 insertion. Additional analysis will be needed to determine whether the 412 insertion affects splicing of the downstream intron and whether the apparent increased splicing of this intron in an *su(s)⁻* background is significant. Otherwise, the same transposon-related splicing events occur in *su(s)⁺* and *su(s)⁻* backgrounds.

Because of the extreme sensitivity of the PCR technique, an argument could be made that the cDNA clones we recovered represent a minor component of the *v* mRNA from *v^k* mutants. The S1 mapping experiment (Fig. 2) indicates that this is not the case. As mentioned earlier, two bands are protected from S1 nuclease digestion when either wild-type cDNA probe **a** or **b** is annealed to poly(A)⁺ RNA from *v^k* mutants (Fig. 2, lanes 5 and 6). If our interpretation of the S1 experiment is correct, the protection pattern observed with probe **a** suggests that most mutant RNAs include sequences upstream of the insertion site. Yet, the pattern observed with probe **b** indicates that most *v^k* RNAs deviate from wild type at the 412 insertion site.

Another possible explanation for the 285-nucleotide fragment band observed in the S1 nuclease protection experiment is that it represents a separate mRNA species that initiates within the 412 element and does not contain *v* sequences upstream of the 412 insertion site. If this were the case, transcription would initiate on what is normally the nontranscribed strand of the 412 retrotransposon. To test this possibility, we attempted to amplify transcripts that might initiate within the 412 element. A synthetic oligonucleotide, complementary to the nontranscribed strand of 412 at the 5' end of the element, and the oligonucleotide derived from the fourth *v* exon (see above) were used as primers in PCR reactions of mutant first-strand cDNAs as described above. When the PCR products were examined on ethidium bromide-stained agarose gels and by Southern analysis, no amplified fragments were detected (data not

shown). Thus, on the basis of this result and the S1 mapping analysis, we conclude that probably all of the mature *v^k* mutant transcripts are spliced like the cDNAs that we have analyzed.

Discussion

We have demonstrated that the 412 element insertion at the *Drosophila v* gene is spliced from pre-mRNA. Both the 5' donor and 3' acceptor sites are located within the transposon on what is usually the nontranscribed strand of 412. Thus, these novel splice sites are not associated with a transposon gene. Four different donor sites are alternatively spliced to the same acceptor site. The splicing events leave ~10–50 nucleotides of 412 sequences in the mRNA, but because the insertion is in the untranslated region, a normal protein can be synthesized.

The 5' splice sites in the 412 transposon are generally poor matches to the invertebrate donor site consensus AG/GTAAGT (Shapiro and Senapathy 1987). The most frequently used donor sites (II and III; Fig. 3) match the consensus at four of eight and five of eight positions, respectively. The other two donor sites match at only three of eight positions. Shapiro and Senapathy (1987) analyzed all splice junction sequences in GenBank and devised a scoring method that indicates the similarity of a given sequence to other known splice sites. Scores range from 1 to 100 and are based on the observed frequency of a particular nucleotide occurring at each position within the conserved octamer. We used this scoring system to compare the donor sites in the 412 element and in a number of *Drosophila* genes to other invertebrate donor sites. We calculated scores for 45 splice donor sites from seven *Drosophila* genes, including *sevenless* (Basler and Hafen 1988), *bicoid* (Berleth et al. 1988), *su(Hw)* (Parkhurst et al. 1988), *transformer* (Boggs et al. 1987), α -*tubulin* (Theurkauf et al. 1986), *su(w^a)* (Chou et al. 1987), *v* (Searles et al. 1990), and *ovarian tumor* (Steinhauer et al. 1989). These donor site scores ranged from 68 to 91, and the average score was 80 (SD = 5.7). In comparison, the 412 donor sites scores are low, ranging from 43 to 65 (Fig. 3).

The sequences associated with 3' splice acceptor sites are more variable than 5' donor sites (Padgett et al. 1986). A pyrimidine-rich region typically precedes four highly conserved nucleotides at the acceptor site junction. Eight of the 15 nucleotides in this region of the 412 acceptor site are pyrimidines (Fig. 3). We also computed 3' splice acceptor site scores for the 412 and the genes above, except for *su(Hw)*, where full acceptor sites were not reported. The 38 acceptor sites in these *Drosophila* genes have scores ranging from 60 to 98, with an average score of 81 (SD = 9.3). The 412 acceptor site score is 75 (Fig. 3), a value within the range of scores observed for known splice acceptor sites in *Drosophila*. Thus, on the basis of the sequence alone, the 412 donor sites appear to be somewhat atypical, whereas the 412 splice acceptor appears to be a more normal site. During splicing,

Fridell et al.

branch formation occurs at a site within the intron located 20–50 nucleotides upstream of the 3' acceptor site. In higher eukaryotes, the branch site is not highly conserved (Padgett et al. 1986). However, a consensus sequence, CTAAT, for the branch site of *Drosophila* introns has been reported (Keller and Noon 1984). The sequence TTAAT occurs at two positions, beginning 23 and 41 nucleotides upstream of the 3' acceptor site, and could possibly function as the branch site.

The level of *v* mutant mRNA is extremely low in a *su(s)*⁺ background. In a *su(s)* mutant background, the mRNA level increases to 10–20% of the *v*⁺ level. Recessive mutations at the *su(s)* gene could increase the level of *v* mRNA accumulation by one of several possible mechanisms. First, suppressor mutations might affect transcription of the *v* gene. An increase in the rate of transcription initiation or a decrease in the frequency of random transcription termination within the element (e.g., see Dorsett et al. 1989) could result in elevated precursor and mRNA levels. Second, suppressor mutations might decrease the rate of RNA turnover and thus increase the stability of precursor and/or mature RNA. Third, if the precursor RNA that includes transposon sequences is relatively unstable, suppressor mutations might elevate mRNA levels by increasing the rate of splicing at the cryptic sites within the 412 element.

The sequence of the *su(s)* gene indicates that it encodes a polypeptide with a region of similarity to vertebrate and *Drosophila* proteins involved in RNA processing (Voelker et al., in prep.). Furthermore, a reversion of the *v*¹ mutation is associated with the introduction of a new 5' donor site into the 412 element (A.-M. Pret and L.L. Searles, in prep.). These results suggest that suppression occurs at the RNA level, that is, splicing or RNA stability is likely to be affected. We speculate that the *su(s)* protein may be involved, directly or indirectly, in pre-mRNA splicing. Splicing of mRNA precursors involves the assembly of the pre-mRNA into spliceosome complexes composed of small nuclear ribonucleoprotein particles (snRNPs) (for review, see Maniatis and Reed 1987; Sharp 1987). Recent studies have identified proteins that bind to the 3' splice site region and facilitate U2 snRNP binding and spliceosome complex formation (Ruskin et al. 1988; Garcia-Blanco et al. 1989). Because base-pairing between U1 snRNA and the conserved 5' splice site is not sufficient for splice donor site recognition, similar proteins are likely to act at the donor site region. The *su(s)* gene may encode one of these proteins. A normal function for the *su(s)* protein in 5' splice site recognition may be to prevent recognition of, or destabilize complex formation on, cryptic splice donor sites. In *su(s)* mutants, the weak splice donor sites may be recognized with a higher efficiency. The consequences of *su(s)* mutations may be to increase the efficiency of splicing the "412" intron, resulting in higher mRNA levels. Alternatively, more efficient recognition of the 412 donor sites may increase the rate of association of the precursor into a splicing complex, thereby protecting the precursor from degradation without necessarily affecting the rate of splicing. Future

experiments will test this and other models of suppression.

To our knowledge, this is the first example of the splicing of a retrovirus-like transposable element from RNA. Analysis of P-element insertions into the 5'-untranslated region of the *Drosophila y* gene suggests that P elements can also be spliced from RNA (Geyer et al. 1988). These events may occur more commonly but may not be detected because the removal of transposon sequences is imprecise. For example, the splicing of transposons inserted into coding regions could generate defective mRNAs by introducing changes in translational reading frames. Furthermore, the low level of spliced mutant transcript would probably not provide sufficient gene activity to be phenotypically detectable in many instances.

The splicing of transposable elements from mRNA precursors was first demonstrated in maize. Both the activator/dissociation (Ac/Ds) and the suppressor/mutator (Spm) elements of maize can be spliced from RNA (for review, see Wessler 1988). In the examples that have been reported, either the splice donor or the acceptor site is located within the transposon, and the other site is found in the resident gene. Some spliced insertions of maize elements are in coding sequences, and splicing maintains the reading frame. These mutants accumulate a reduced level of altered protein. All spliced transposons, including the ones in *Drosophila*, are inserted in a transcriptional orientation opposite that of the resident gene. When transposons insert within a gene in the same transcriptional orientation, most mRNAs are polyadenylated at sites within the transposon. Thus, the observed orientation of spliced transposons may be necessary for the inclusion of sequences downstream of the transposon in the RNA. However, it is somewhat surprising that diverse families of transposable elements have potential splice junctions near their boundaries on the nontranscribed strand. These insertions are, in effect, introducing new introns into genes (Wessler 1988; Raboy et al. 1989). In fact, one model of intron evolution suggests that introns arose from transposon insertions (Cavalier-Smith 1985; Sharp 1985). As more examples of transposon splicing are discovered, the possibility that transposons may be a source of new introns becomes more intriguing.

Methods

RNA isolation

Mutant *D. melanogaster* alleles used in this study have been described previously (Shepard 1960; Lindsley and Grell 1968). *Drosophila* strains *y*² *w*^{bf} (*v*⁺), *y*² *su(s)*^{51c15} *ras v*^k [*su(s)*⁻ *v*^k], and *ras v*^k [*v*^k] were used as sources of RNA. The 1.4-kb *v* transcript accumulates from midembryogenesis until pupation and reappears in young adults (unpubl.). Thus, RNA was isolated from 12- to 24-hr embryo and early larval stages. Total RNA was prepared by homogenization in guanidine thiocyanate (Bermingham and Scott 1988), and poly(A)⁺ RNA was purified by oligo(dT)-cellulose chromatography.

S1 mapping experiments

Single-stranded antisense RNA probes were synthesized from plasmid pV110, a 0.4-kb *EcoRI*–*Clai* *v* fragment from the embryonic cDNA F2-2 (Searles et al. 1990) cloned into pGEM-7Zf (Promega). Plasmid DNA was linearized by digesting with either *EcoRI* or with *XhoI*. (An *EcoRI* site was introduced at the cDNA 5' end during construction of the library.) Radioactively labeled probes were made using SP6 RNA polymerase as described by Melton et al. (1984), except that 200 μ Ci of [³²P]CTP (800 Ci/mmol) was included in the reaction to increase the probe specific activity. The hybridization reaction (Melton et al. 1984) contained 10–20 μ g of poly(A)⁺ RNA (or tRNA in control reactions) and 5 \times 10⁶ cpm of probe. Hybrids were purified over small oligo(dT)-cellulose columns to reduce background, as described by Chou et al. (1987), and subsequently treated with S1 nuclease (Quarless and Heinrich 1986). Protected fragments were resolved by electrophoresis on an 8% polyacrylamide/7 M urea sequencing gel. M13mp18 DNA was sequenced in reactions containing ³⁵S-labeled dATP, the universal primer, and Sequenase (U.S. Biochemical), and the sequencing reactions were run in parallel lanes to provide size standards.

Preparation of partial cDNA clones

Poly(A)⁺ RNA isolated from *v*⁺ and *su(s)^{51c15} v*⁺ embryo and early larval stages was specifically primed for synthesis of *v* cDNAs. Oligonucleotide 2052 (5'-CCGAGTTGCGAATCGAATTCGCGCCTCC-3'), which is derived from a region of the sense strand spanning the *EcoRI* site in the fourth exon of the *v* gene (Searles et al. 1990; see Fig. 1) was used to prime first-strand cDNA synthesis (Calzone et al. 1987). Poly(A)⁺ RNA (10–20 μ g) was mixed with 0.2 μ g of oligonucleotide in 20 μ l of hybridization solution [0.4 M NaCl, 40 mM PIPES (pH 6.5), 1.0 mM EDTA] and incubated for 5 min at 85°C and then for 1 hr at 42°C. The annealed oligomer was extended by adding 3 μ l of 10 \times reverse transcriptase buffer [0.5 M Tris-HCl (pH 8.5), 0.1 M MgCl₂, 0.4 M KCl, 10 mM DTT], 5 μ l of a 20 mM dNTP solution (5 mM each for dATP, dGTP, dTTP, and dCTP) 1 μ l (40 units) of RNasin, and 0.5 μ l (16 units) of AMV reverse transcriptase (Life Sciences, XL grade).

The products of the cDNA synthesis were amplified by PCR (Saiki et al. 1988). Oligonucleotide 1846 (5'-CGCCAGCATCCGCGCGAGGAG-3'), which is complementary to the sense strand of the first exon in the region that includes the *XhoI* site (Searles et al. 1990; see Fig. 1), and oligonucleotide 2052 were used as PCR primers. One to 3 μ l of the first-strand cDNA synthesis reaction mixture was mixed with 0.2 μ g of each oligonucleotide and 0.5 μ l (2.5 units) of *Taq* polymerase (Perkin Elmer–Cetus) in 100 μ l of PCR reaction buffer [50 mM KCl, 1.5 mM MgCl₂, 10 mM Tris-HCl (pH 8.5), 0.25 mM with respect to each dNTP]. The reaction mixture was overlaid with mineral oil and subjected to 30 temperature cycles (94°C denaturing, 1 min; 58°C or 65°C annealing, 2 min; 72°C extension, 3 min).

Fragments of *v* cDNA were cloned into a plasmid vector through the use of either the *XhoI* and *EcoRI* sites contained within the primers or the *XhoI* site within the upstream primer and an internal *Clai* site located in the third *v* exon (see Fig. 1) as follows. After amplification, the PCR reaction mixture was extracted twice with equal volumes of phenol/chloroform and once with chloroform. Following ethanol precipitation, DNA was digested with the appropriate enzymes, the fragments were ligated into pGEM-7Zf and transformed into *Escherichia coli* strain JM105 by standard procedures (Maniatis et al. 1982).

Plasmid DNA was isolated from 5-ml cultures derived from individual colonies by the alkaline lysis procedure (Maniatis et al. 1982) and analyzed by restriction enzyme digestion prior to sequencing. The same double-stranded plasmid DNA obtained from a small-scale alkaline lysis was prepared for sequencing by extracting twice with phenol/chloroform and once with chloroform, followed by ethanol precipitation. Sequencing was performed according to the Sequenase sequencing kit protocol (U.S. Biochemical). The sequence was determined beginning at the *XhoI* site immediately upstream of the 412 insertion site and extending into the *v*-coding region.

Acknowledgments

We thank Dr. R. Voelker for kindly providing *Drosophila* stocks. We also thank Drs. P. Bedinger, J. Lucchesi, T. Petes, and P. Pukkila for critically reading the manuscript, S. Whitfield for the art work, and S. Lennon for typing. This work was supported by grants DCB-8602354 and DMB-8904471 from the National Science Foundation and by Basil O'Connor Starter Scholar Research Award 5-642.

References

- Basler, K. and E. Hafen. 1988. Control of photoreceptor cell fate by the *sevenless* protein requires a functional tyrosine kinase domain. *Cell* **54**: 299–311.
- Berleth, T., M. Burri, G. Thoma, D. Bopp, S. Richstein, G. Frigerio, M. Noll, and C. Nüsslein-Volhard. 1988. The role of localization of bicoid RNA in organizing the anterior pattern of the *Drosophila* embryo. *EMBO J.* **7**: 1749–1756.
- Bermingham, J.R., Jr. and M.P. Scott. 1988. Developmentally regulated alternative splicing of transcripts from the *Drosophila* homeotic gene *Antennapedia* can produce four different proteins. *EMBO J.* **7**: 3211–3222.
- Bingham, P.M., R. Levis, and G.M. Rubin. 1981. Cloning of DNA sequences from the *white* locus of *D. melanogaster* by a novel and general method. *Cell* **25**: 693–704.
- Boggs, R.T., P. Gregor, S. Idriss, J.M. 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.
- Calzone, F.J., R.J. Britten, and E.H. Davidson. 1987. Mapping gene transcripts by nuclease protection and cDNA primer extension. *Methods Enzymol.* **152**: 611–632.
- Cavalier-Smith, T. 1985. Selfish DNA and the origin of introns. *Nature* **315**: 283–284.
- Chou, T.-B., Z. Zachar, and P.M. Bingham. 1987. Developmental expression of a regulatory gene is programmed at the level of splicing. *EMBO J.* **6**: 4095–4104.
- Dorsett, D., G.A. Viglianti, B.J. Rutledge, and M.M. Meselson. 1989. Alteration of *hsp82* gene expression by the gypsy transposon and suppressor genes in *Drosophila melanogaster*. *Genes Dev.* **3**: 454–468.
- Garcia-Blanco, M.A., S.F. Jamison, and P.A. Sharp. 1989. Identification and purification of a 62,000-dalton protein that binds specifically to the polypyrimidine tract of introns. *Genes Dev.* **3**: 1874–1886.
- Geyer, P.K., K.L. Richardson, V.G. Corces, and M.M. Green. 1988. Genetic instability in *Drosophila melanogaster*: P-element mutagenesis by gene conversion. *Proc. Natl. Acad. Sci.* **85**: 6455–6459.
- 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.

Fridell et al.

- Keller, E.B. and W.A. Noon. 1984. Intron splicing: A conserved internal signal in introns of animal pre-mRNAs. *Proc. Natl. Acad. Sci.* **81**: 7417–7420.
- Kubli, E. 1986. Molecular mechanisms of suppression in *Drosophila*. *Trends Genet.* **2**: 204–209.
- Lindsley, D.L. and E.H. Grell. 1968. Genetic Variations of *Drosophila melanogaster*. *Carnegie Inst. Washington Publ.* 627.
- Maniatis, T., E.F. Fritsch, and J. Sambrook. 1982. *Molecular cloning: A laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor New York.
- Maniatis, T. and R. Reed. 1987. The role of small nuclear ribonucleoprotein particles in pre-mRNA splicing. *Nature* **325**: 673–678.
- Mazo, A.M., L.J. Mizrokhi, A.A. Karavanov, Y.A. Sedkov, A.A. Krichevskaja, and Y.V. Ilyin. 1989. Suppression in *Drosophila*: *su(Hw)* and *su(f)* gene products interact with a region of gypsy (*mdg4*) regulating its transcriptional activity. *EMBO J.* **8**: 903–911.
- Melton, D.A., P.A. Krieg, M.R. Rebagliati, T. Maniatis, K. Zinn, and M.R. Green. 1984. Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Res.* **12**: 7035–7056.
- Padgett, R.A., P.J. Grabowski, M.M. Konarska, S. Seiler, and P.A. Sharp. 1986. Splicing of messenger RNA precursors. *Annu. Rev. Biochem.* **55**: 1119–1150.
- Parkhurst, S.M. and V.G. Corces. 1985. 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.
- Parkhurst, S.M., D.A. Harrison, M.P. Remington, C. Spana, R.L. Kelly, R.S. Coyne, and V.G. Corces. 1988. The *Drosophila su(Hw)* gene, which controls the phenotypic effect of the gypsy transposable element, encodes a putative DNA-binding protein. *Genes Dev.* **2**: 1205–1215.
- Quarless, S.A. and G. Heinrich. 1986. The use of complementary RNA and S1 nuclease for the detection and quantitation of low abundance mRNA transcripts. *Biotechniques* **4**: 434–438.
- Raboy, V., H.-Y. Kim, J.W. Schiefelbein, and O.E. Nelson, Jr. 1989. Deletions in a dSpm insert in a maize *bronze-1* allele alter RNA processing and gene expression. *Genetics* **122**: 695–703.
- 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.
- 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.
- Searles, L.L. and R.A. Voelker. 1986. Molecular characterization of the *Drosophila* vermilion locus and its suppressible alleles. *Proc. Natl. Acad. Sci.* **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.* (in press.)
- Shapiro, M.B. and P. Senapathy. 1987. RNA splice junctions of different classes of eukaryotic sequence statistics and functional implications in gene expression. *Nucleic Acids Res.* **15**: 7155–7174.
- Sharp, P.A. 1985. On the origin of RNA splicing and introns. *Cell* **42**: 397–400.
- . 1987. Splicing of messenger RNA precursors. *Science* **235**: 766–771.
- Shepard, P.B. 1960. A physiological study of the *vermilion* eye color mutants of *Drosophila melanogaster*. *Genetics* **45**: 359–376.
- Sisodia, S.S., D.W. Cleveland, and B. Sollner-Webb. 1987. A combination of RNAase H and S1 nuclease circumvents an artefact inherent to conventional S1 analysis of RNA splicing. *Nucleic Acids Res.* **15**: 1995–2011.
- 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.
- Steinhauer, W.R., R.C. Walsh, A. Ousley, and L.J. Kalfayan. 1989. Sequence and structure of the *Drosophila melanogaster* ovarian tumor gene and generation of an antibody specific for the ovarian tumor protein. *Mol. Cell. Biol.* **9**: 5726–5732.
- Theurkauf, W.E., H. Baum, J. Bo, and P. Wensink. 1986. Tissue-specific and constitutive alpha-tubulin genes of *Drosophila melanogaster* code for structurally distinct proteins. *Proc. Natl. Acad. Sci.* **83**: 8477–8481.
- Walker, A.R., A.J. Howells, and R.G. Tearle. 1986. Cloning and characterization of the vermilion gene of *Drosophila melanogaster*. *Mol. Gen. Genet.* **202**: 102–107.
- Wessler, S.R. 1988. Phenotypic diversity mediated by the maize transposable elements Ac and Spm. *Science* **242**: 399–405.
- Will, B.M., A.A. Bayev, and D.J. Finnegan. 1981. Nucleotide sequence of terminal repeats of 412 transposable elements of *Drosophila melanogaster*. **153**: 897–915.
- Zachar, A., T.-B. Chou, and P.M. Bingham. 1987. Evidence that a regulatory gene autoregulated splicing of its transcript. *EMBO J.* **6**: 4105–4111.



A retrotransposon 412 insertion within an exon of the *Drosophila melanogaster* vermilion gene is spliced from the precursor RNA.

R A Fridell, A M Pret and L L Searles

Genes Dev. 1990, 4:

Access the most recent version at doi:[10.1101/gad.4.4.559](https://doi.org/10.1101/gad.4.4.559)

References

This article cites 36 articles, 14 of which can be accessed free at:
<http://genesdev.cshlp.org/content/4/4/559.full.html#ref-list-1>

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

A dark green banner advertisement for Horizon. On the left is the "horizon" logo in white, with "INSPIRED CELL SOLUTIONS" underneath. To the right of the logo is the text "Inspired Custom Oligo Synthesis Solutions" in white, followed by "Limitless modifications, greater yields, rapid delivery" in a smaller white font. On the far right is a white rounded rectangle containing the text "Request a quote" in dark green. The background of the banner features a faint, light-colored molecular structure graphic.

horizon
INSPIRED CELL SOLUTIONS

Inspired Custom Oligo
Synthesis Solutions
Limitless modifications, greater yields, rapid delivery

Request
a quote