Structure and Transcription of the *Drosophila melanogaster* Vermilion Gene and Several Mutant Alleles

LILLIE L. SEARLES,* RICHARD S. RUTH, ANNE-MARIE PRET, ROBERT A. FRIDELL, AND AHMAD J. ALI

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

The nucleotide sequence and intron-exon structure of the *Drosophila melanogaster* vermilion (*v*) gene have been determined. In addition, the sites of several mutations and the effects of these mutations on transcription have been examined. The major *v* mRNA is generated upon splicing six exons of lengths (5' to 3') 83, 161, 134, 607, 94, and 227 nucleotides (nt). A minor species of *v* mRNA is initiated at an upstream site and has a 5' exon of at least 152 nt which overlaps the region included in the 83-nt exon of the major *v* RNA. The three *v* mutations, v^1, v^2, and v^k, which can be suppressed by mutations at suppressor of sable, *su(s)*, are insertions of transposon 412 at the same position in exon 1, 36 nt downstream of the major transcription initiation site. Despite the 7.5-kilobase insertion in these *v* alleles, a reduced level of wild-type-sized mRNA accumulates in suppressed mutant strains. The structure and transcription of several unsuppressible *v* alleles have also been examined. The v^{36f} mutation is a B104/roo insertion in intron 4 near the splice donor site. A mutant carrying this alteration accumulates a very low level of mRNA that is apparently polyadenylated at a site within the B104/roo transposon. The v^{48a} mutation, which deletes approximately 200 nt of DNA, fuses portions of exons 3 and 4 without disruption of the translational reading frame. A smaller transcript accumulates at a wild-type level, and thus an altered, nonfunctional polypeptide is likely to be synthesized in strains carrying this mutation. The v^{H2a} mutant has a P element insertion in exon 6 within the coding region.

Several mutations at the *Drosophila melanogaster* vermilion (*v*) gene can be suppressed by recessive mutations at suppressor of sable, *su(s)*. The suppressible *v* mutants have insertions of the retrotransposon 412 in the *v* gene region (36, 43) and thus resemble other mutants with suppressible transposon insertion mutations in *D. melanogaster*. For example, suppressor of Hairy-wing, *su(Hw)*, and suppressor of forked, *su(f)*, suppress gypsy element insertion mutations at several genes; suppressor of white apricot, *su(w^a)*, and suppressor of sable, *su(s)*, suppress copia and 412 insertion mutations, respectively (4, 19, 22, 23, 36, 43; reviewed in reference 13). These mobile elements are all retrotransposons, a group of structurally homologous elements that resemble the integrated form of mammalian retroviruses and make up the largest family of transposons in *D. melanogaster* (31). The major transcription products of these elements are nearly full-length, polyadenylated [poly(A)⁺] RNAs, and transcription of the elements is temporally regulated (24, 34). In an effort to understand the basis of suppression, genes with suppressible mutations (15, 22, 23, 27, 36, 48), as well as the suppressor genes (7, 25, 39, 46, 47), are being cloned and studied at the molecular level. The transposon insertions associated with most suppressible mutations are located either in 5' untranscribed sequences or within introns. The mobile-element insertions typically interfere with synthesis of mRNA from a gene at or near the insertion site, and suppressor mutations partially or completely restore transcript accumulation. Studies of suppression by *su(Hw)* and *su(w^a)* indicate that the suppressor genes encode regulators of mobile-element transcription and RNA processing, respectively (25, 39, 46).

The vermilion gene, located on the X chromosome, is required for brown eye pigment synthesis in *D. melanogaster*, the normal eye color being a composite of red and brown pigments. We cloned the gene by P-element transposon

tagging (36), and the same sequences were cloned independently by Walker et al. (43). In using a mammalian tryptophan oxygenase cDNA probe to identify clones of the *D. melanogaster* *v* gene, Walker et al. (43) verified that *v* encodes tryptophan oxygenase, the enzyme that catalyzes the first step in the brown-pigment biosynthetic pathway (reviewed in reference 26). Mutations at *v* are clustered in a 2-kilobase (kb) region, and the *v* transcript is a 1.4-kb mRNA of relatively low abundance (36, 43).

Three *v* mutations, v^1, v^2, and v^k, can be suppressed by *su(s)*. These alleles produce a leaky mutant phenotype and undetectable levels of tryptophan oxygenase (41). In *D. melanogaster* strains with a mutation at *su(s)* and with the 1, 2, or k mutation at *v*, tryptophan oxygenase activity is increased to 10 to 20% of the wild-type level (41), an amount sufficient to restore a wild-type phenotype. After the gene was cloned, the suppressible *v* alleles were shown to be insertions of the 7.5-kb retrotransposon 412 in the vicinity of the vermilion transcription unit (36, 43). Using *v* as a model system, we are investigating the mechanism of suppression by *su(s)*. Thus, we have characterized the structure of the gene in greater detail and localized the positions of all the suppressible and several unsuppressible mutations relative to introns and exons. In addition, we have examined the effects of these mutations on the transcription of *v*.

MATERIALS AND METHODS

Isolation and sequencing of clones. The mutants used in this study have been described previously (16, 36, 37). The cloning of *v* sequences from wild-type and mutant strains v^{H2a}, v^{36f}, v^k, and v^2 was previously reported (36), and genomic clones containing the v^{48a} and v^1 mutations were isolated by a similar approach. For v^1, *Bam*HI rather than *Mbo*I partial fragments were cloned. Restriction maps of representative mutant clones are shown in Fig. 1. The sequence of a 3-kb region of v^+ DNA between the *Eco*RI site (Fig. 1, coordinate 0) and the *Hin*dIII site (coordinate +3.0)
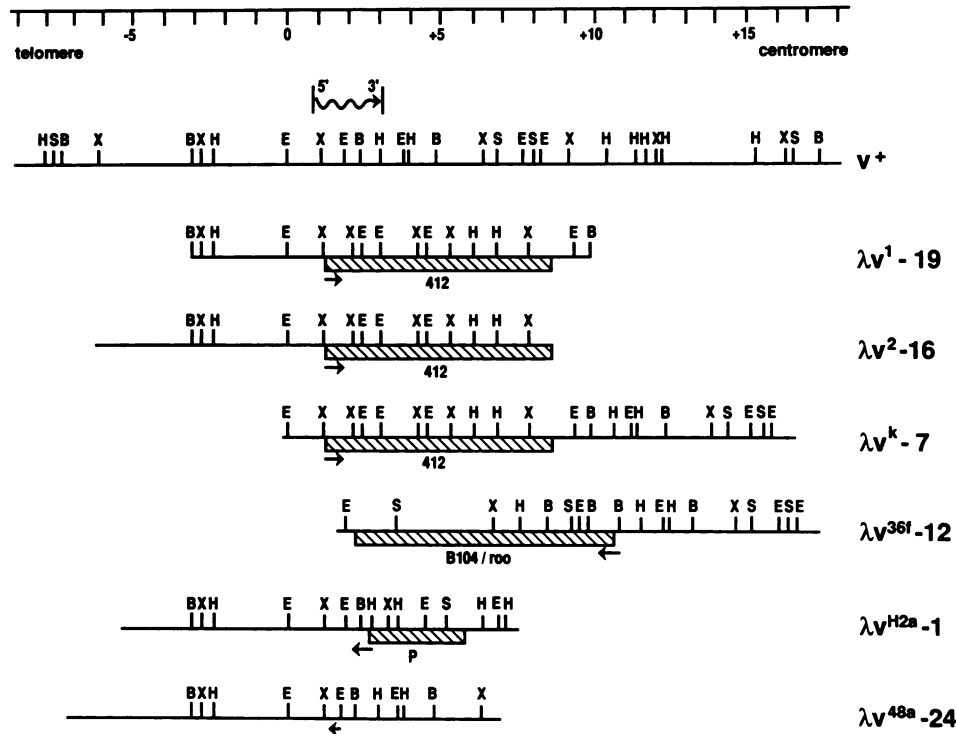
FIG. 1. Restriction maps of mutant v DNA clones. The mutant clones are aligned with the wild-type (v^+) map at coordinate 0. Restriction enzyme abbreviations: E, *Eco*RI; B, *Bam*HI; H, *Hin*dIII; S, *Sal*I; X, *Xho*I. Symbols: ∿, transcribed sequences; →, ←, direction of sequencing; ▧, transposon sequences. The map coordinates are in kilobases.

was obtained in the following manner. The 1.1-kb *Eco*RI-*Xho*I fragment (coordinates 0 to +1.1) and the 1.9-kb *Xho*I-*Hin*dIII fragment (coordinates +1.1 to +3.0) from λv$^+$8 (36) were cloned into both M13mp18 and M13mp19 as described previously (3). A set of overlapping deletions was generated from each M13 clone by using the procedure described by Dale et al. (10), and a series of deleted clones was sequenced (32) (Fig. 2). The sequences were aligned and compiled by using the programs of Staden (40).

Mutant DNA fragments were subcloned into M13mp18 or M13mp19 and sequenced from appropriate restriction sites (Fig. 1). Sequences of the insertion mutant alleles were compared with published sequences of the transposable elements (21, 33, 44). Since the direction of transcription has been determined for 412 (45), B104/roo (33), and P elements (21), the transcriptional orientation of each element was deduced from the restriction maps.
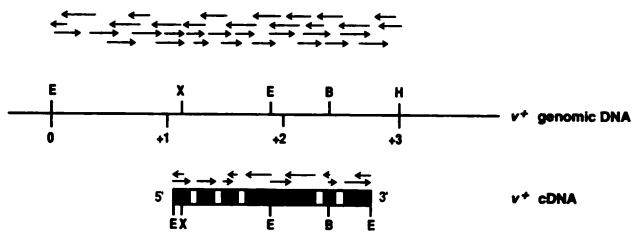
FIG. 2. Strategy for sequencing v^+ genomic and cDNA clones. Symbols: →, ←, direction and extent of sequencing; ■, exons found in the cDNA clones; □, intron sequences absent from cDNA clones. Restriction enzyme abbreviations: E, *Eco*RI; X, *Xho*I; B, *Bam*HI; H, *Hin*dIII. Map coordinates are represented by numbers beneath the genomic DNA map.

Wild-type v cDNA clones were isolated from the late-embryo, larval, and adult cDNA libraries of Poole et al. (28). The libraries were screened as described previously (35) by using the 1.9-kb *Xho*I-*Hin*dIII fragment (coordinates +1.1 to +3.0). The largest cDNAs were purified from the F2 (12- to 24-h embryo) and G2 (first- and second-instar) libraries. Fragments of cDNA were transferred to M13 vectors and sequenced (Fig. 2).

Determining the direction of transcription. The 1.9-kb *Eco*RI A fragment (Fig. 1, coordinates 0 to +1.9) and the 1.9-kb *Eco*RI B fragment (coordinates +1.9 to +3.8) were subcloned from λv$^+$8 into M13mp10. By restriction analysis of double-stranded replicative-form DNA, the orientation of v DNA in the M13 vector was determined. One orientation results in the production of the v transcribed strand as single-stranded DNA, whereas the other orientation produces the v nontranscribed strand in the single-stranded form. Single-stranded M13 DNAs containing transcribed- and nontranscribed-strand sequences were hybridized separately to Northern (RNA) blots containing 20 μg of poly(A)$^+$ RNA from adult flies. Hybrids were detected after a second hybridization with ^{32}P-labeled M13mp10 DNA synthesized with a universal sequencing primer as described by Kuroiwa et al. (14).

Northern analysis. Adult poly(A)$^+$ RNA (20 μg) was fractionated on formaldehyde-agarose gels (9) and transferred to nitrocellulose or nylon membranes (42). The hybridizations and washes were performed as described by O'Hare et al. (20). Plasmid probes were labeled with ^{32}P by nick translation (17).

Exon mapping. Fragments of v genomic DNA and v cDNA F2-2 were subcloned into either pGEM-1 or pGEM-7Zf (Promega Biotec). Antisense RNA probes (riboprobes) were

synthesized from linearized plasmid templates with SP6 or T7 RNA polymerase and 50 to 200 μCi of [^{32}P]CTP (800 Ci/mmol) as described by Melton et al. (18). For the hybridization reaction, 5 to 20 μg of poly(A)$^+$ RNA and 1×10^6 to 5×10^6 cpm of probe were mixed in hybridization buffer (18) and incubated at 85°C for 5 min and then at 42°C overnight. In some experiments hybrids were digested with RNase (18). In other instances hybrids were purified over small oligo(dT)-cellulose columns to reduce background, as described by Chou et al. (8), and then samples were treated with S1 nuclease by the procedure of Quarless and Heinrich (30). Protected fragments were resolved by electrophoresis on an 8 to 15% polyacrylamide–7 M urea sequencing gel. Sequencing reactions of M13mp18 DNA with the universal primer and Sequenase (U.S. Biochemical Corp.) were run in parallel lanes to provide size standards.

Primer extension analysis. Two synthetic oligonucleotides were used in these experiments: 5'-TCCGTTTCCTGCAT AGGG-3' (oligo 1235) and 5'-GCCCACTTCCGTGGTTAA TGGCACCGCCG-3' (oligo 1489). The oligonucleotides were labeled with ^{32}P at the 5' ends by using polynucleotide kinase (17). The labeled oligonucleotides were purified by electrophoresis on a 15% polyacrylamide gel. They were eluted from the gel slice by overnight incubation at 37°C in 0.3 to 0.5 ml of TE (10 mM Tris hydrochloride [pH 8.0], 1 mM EDTA). The oligomer in TE was extracted sequentially with equal volumes of phenol, phenol-chloroform, and chloroform before ethanol precipitation. Primer extension experiments were performed as described by Calzone et al. (6), with the following modifications. The annealing reaction mixture contained 20 μg of poly(A)$^+$ RNA or tRNA (in control reactions) and approximately 0.3 pmol (10^5 to 10^6 cpm) of end-labeled oligonucleotide primer in 30 μl of hybridization solution (6). This mixture was incubated at 85°C for 5 min and then at 42°C for 1 h. To the annealing reaction were added 4 μl of 10× reverse transcriptase buffer (6), 5 μl of a 20 mM deoxynucleoside triphosphate solution (5 mM in each of the four deoxynucleotide triphosphates), 1 μl (40 U) of RNasin, and 0.5 μl (16 U) of avian myeloblastosis virus reverse transcriptase (XL grade; Life Sciences). The reaction mixture was incubated at 42°C for 1 h, and after phenol extraction and ethanol precipitation the reaction products were applied to an 8% acrylamide–7 M urea sequencing gel. The sizes of the extended products were determined by using sequencing reactions of M13mp18 DNA in parallel lanes.

Nucleic acid isolation. Plasmid and lambda DNAs were purified by standard procedures (17). Single-stranded M13 DNA was isolated as described in the Amersham Corp. sequencing handbook. RNA was, in most cases, prepared as described previously (3) by homogenization in sodium dodecyl sulfate-urea buffer, repeated phenol extractions, and purification on oligo(dT)-cellulose. For use in the primer extension experiments, RNA was isolated by homogenization in guanidine thiocyanate followed by lithium chloride precipitation (2) and purification on oligo(dT)-cellulose. Wild-type (v^+) RNA was isolated from either an Oregon R or $y^2 w^{bf}$ strain.

RESULTS

Localization of the v transcription unit. We previously reported (36) that the 1.4-kb v transcript hybridized to two adjacent 1.9-kb *Eco*RI fragments (Fig. 1, coordinates 0 to +1.9 and +1.9 to +3.8). From Northern analysis with smaller fragments as probes, we subsequently determined

that most if not all v transcribed sequences were found between the *Xho*I and *Hin*dIII sites (coordinates +1.1 to +3.0). The 1.1-kb *Eco*RI-*Xho*I fragment (coordinates 0 to +1.1) did not produce a detectable signal when it was hybridized to wild-type RNA (data not shown). To determine the direction of transcription, we probed Northern blots of poly(A)$^+$ RNA with single-stranded M13 clones of v fragments from this region (see Materials and Methods). These experiments demonstrated that the direction of transcription is from left to right (Fig. 1). This conclusion was subsequently verified by analysis of cDNA clones (see below).

Structure of the v transcription unit. The intron-exon structure of the v gene was determined by comparing the sequences of genomic and cDNA clones and by RNA protection experiments. We sequenced a 3-kb region of v^+ genomic DNA extending from the *Eco*RI site at coordinate 0 to the *Hin*dIII site at coordinate +3.0. In addition, we isolated and sequenced v cDNA clones from the libraries of Poole et al. (28). Approximately 20 v cDNA clones were isolated, and since *Eco*RI linkers were added to the cDNAs during the cloning process (Fig. 2), the sizes of the cDNAs were estimated by examining the *Eco*RI digestion pattern. One of the largest cDNA clones, G2-4, isolated from a first- and second-instar library, was sequenced from convenient restriction sites (Fig. 2), and the sequence was compared with the genomic sequence (Fig. 3). The cDNA sequences align with genomic sequences between nucleotides (nt) 1046 and 2775. Within this region are six exons of the v gene, a 5' exon of at least 152 nt, and five additional exons (ordered 5' to 3') of 161, 134, 607, 94, and 227 nt. As will be discussed below, the first exon in this cDNA extends further 5' than most v mRNAs. The intervening sequences are quite small, ranging from 54 to 91 nt. An open reading frame begins in the first exon at nt 1172. The sequence GCACCATG (nt 1167 to 1174) at the presumptive initiation site matches the consensus for a eucaryotic translation start site, CC(A/G)CCATG (12), at seven of eight positions. Upon splicing, this open reading frame extends to nt 2665 and would generate a 380-amino-acid polypeptide of predicted molecular mass 44.4 kilodaltons. This approximately corresponds to the expected monomer molecular mass of tryptophan oxygenase (1, 36). The polyadenylation signal AATAAA (29) is found beginning at nt 2742, approximately 30 nt upstream of the 3' end of the cDNA.

Mapping the 5' end of the v RNA. S1 nuclease protection experiments were performed to define the 5' exon of v. A 1.9-kb *Xho*I-*Hin*dIII fragment (Fig. 4, probe a) protects most v exons except sequences upstream of the *Xho*I site (Fig. 4A, lane 2), and the sizes of the protected fragments observed are consistent with the cDNA sequence analysis. A probe which covers the entire 5' half of v transcribed sequences (Fig. 4, probe b) generated all of the expected protected fragments except the one expected to be larger than 152 nt. Instead, we observed an 83-nt exon (Fig. 4B, lane 5). Probe c, which includes 5' sequences extending to the *Xho*I site, protects 60 of the 83 nt (Fig. 4B, lane 6). Furthermore, no protected fragments were detected with probes which included sequences between the *Xho*I site in the 5' region to the upstream *Eco*RI site (data not shown). Thus, we concluded that the 5' exon is 83 nt and spans the *Xho*I site, with 60 nt downstream and 23 nt upstream of the *Xho*I site. This analysis places the transcription start site at nt 1115 (Fig. 3).

S1 mapping experiments with cDNA probes confirmed the results obtained with genomic probes. The 295 nt of v cDNA

```
    1  GAATTCCAAGCACATTGCAAGAATCCCAAATCAAAAAATCGCATGAAATTGCCCCCGTACCTTTTGCGTTTTACTCCCAGATGTAACTCA    90
   91  ATTTTTTCTATGCAAAAGTAGTTGAAAATTATATAATAAAAACCGATTAGAAAAACAAAACATACATATATATATATATAAAATATATATA   180
  181  TATATTTAGACACACATCGACAGTATCCTATTCAATTGATTTCTTTGAGAACTTTGATTTTGCGATTTTGGATATGCAGCAAGAAAAGTA   270
  271  AAACCAACAACAGAAAATGTGTAAGAAATAGTATAAATAAGGTCGGATATTAATGCCCCGACATTACGCTATATGTATGTGCTGGGGCAA   360
  361  ACAGCAACAAATCCAATAAAACAAAGTAATTAACAACAAACAAAACGGATGAAACAACCAAGATTATGTAAGCGATGGATCGAAGTAAGA   450
  451  ATTGATGCGAACGGCACAAGTATATAACAAATTTCAACAAGTATATGACTGAGCCAATGACTCAAAAATACATTTTAAATAAAAGGGAAA   540
  541  ACCAGAAATATATGAAAAATATAAAAACGATAAGCAAGTGAATGAAAGCTCTTTTTCTTTTTTGGGTTTTGGGAACTGCAAATTATTGAT   630
  631  TGTATGGAAATGTTTGTTTACCTATTTTTGCATATGGTGCGATTGTATCAAAACAAGTTTTGAATTATCAAAATTGGTTCCATTTATTTT   720
  721  TATACAACCTTGACCTATTTTCAAGGACCAATAAGATTGGACCCCACATTAACTTAGAAAACAATACTTGCCATGTTCAATTTTATTCCT   810
  811  ACGCAGGGTTTATTTATTATTATACTATGTTAATCAAAAAATTAAAATGTTAATTTCTCAGTTATTTAACTACACCTTAGGTAACTCTGA   900
  901  TTTGGCATTTCTCACTGAACTGTACTACTGTAGACTACCTTCCATTCAGGAAAATATTTGTGTGCGCCGCACTTTCACCTCAAGTGATTG   990
                                                 ↓G2-4                ↓F2-2
  991  ATAATTCCCAGCCTATCTGGCAGTGCCCATCGCCCAGATCACCGACTGTGCAATCAGTCGGAACTGGAGCTCTCTCGCTCTGTTATCGGT  1080
                    +1  ↓G2-7                                      ▽1,2,k
 1081  TCGCTGGGGTCTCATCTCCGGTCCGCTGGCGGAGATCAGTTCGCCAGCATCCGCCGCTCGAGGAGTCACGATCTGATCTGAGCTGTGCAC  1170

 1171  CATGAGCTGTCCCTATGCAGGAAACGGGTGAGCACCAGCACGTGCTGTCCAGGAATGCCAATCGATCTTCAGTTCTGCGATTCAATTCAA  1260
        MetSerCysProTyrAlaGlyAsnGl

 1261  ACCCATACAGAAACGATCACGATGATTCGGCGGTGCCATTAACCACGGAAGTGGGCAAAATCTATGGAGAGTATCTGATGCTGGACAAAC  1350
                yAsnAspHisAspAspSerAlaValProLeuThrThrGluValGlyLysIleTyrGlyGluTyrLeuMetLeuAspLysL

 1351  TGCTGGATGCCCAGTGTATGCTGTCCGAGGAGGACAAGCGACCCGTGCACGATGAGCATCTGTTCATCATCACGCACCAGGGTGAGTAGG  1440
        euLeuAspAlaGlnCysMetLeuSerGluGluAspLysArgProValHisAspGluHisLeuPheIleIleThrHisGlnA

 1441  TTTACAACTTTGATGACAACACTCAATGGCATTTAAGTACCTTCGCCACAGCCTACGAGCTTTGGTTCAAGCAGATCATCTTTGAGTTCG  1530
                                laTyrGluLeuTrpPheLysGlnIleIlePheGluPheA
                                                                     [48a
 1531  ACTCCATACGAGACATGTTGGATGCAGAGGTCATCGATGAAACCAAGACGCTGGAGATTGTCAAGCGACTGAACCGAGTGGTTCTGATTC  1620
        spSerIleArgAspMetLeuAspAlaGluValIleAspGluThrLysThrLeuGluIleValLysArgLeuAsnArgValValLeuIleL

 1621  TAAAAGTGAGTGCTTTCTGAATCTCTTACCAAAATCCGTTTATAACTTCCTTTGTACAGCTCCTGGTGGACCAAGTGCCCATTCTGGAGA  1710
        euLys                                                      LeuLeuValAspGlnValProIleLeuGluT

 1711  CCATGACCCCGCTAGACTTCATGGACTTCCGCAAGTACCTGGCACCCGCATCTGGTTTTCAGTCGCTGCAGTTCCGTTTGATCGAGAACA  1800
        hrMetThrProLeuAspPheMetAspPheArgLysTyrLeuAlaProAlaSerGlyPheGlnSerLeuGlnPheArgLeuIleGluAsnL
        48a]
 1801  AGCTGGGAGTTCTGACAGAGCAGCGGGTGAGATACAACCAGAAGTACTCGGATGTCTTTAGCGACGAGGAGGCGCGGAATTCGATTCGCA  1890
        ysLeuGlyValLeuThrGluGlnArgValArgTyrAsnGlnLysTyrSerAspValPheSerAspGluGluAlaArgAsnSerIleArgA

 1891  ACTCGGAGAAAGATCCCTCGCTACTGGAGCTAGTGCAGCGATGGCTGGAGAGGACGCCCGGACTGGAGGAGAGTGGCTTCAACTTCTGGG  1980
        snSerGluLysAspProSerLeuLeuGluLeuValGlnArgTrpLeuGluArgThrProGlyLeuGluGluSerGlyPheAsnPheTrpA

 1981  CCAAGTTTCAGGAGAGCGTCGATCGATTCCTGGAGGCGCAGGTACAGAGCGCCATGGAGGAGCCCGTGGAGAAGGCGAAAAACTACCGCC  2070
        laLysPheGlnGluSerValAspArgPheLeuGluAlaGlnValGlnSerAlaMetGluGluProValGluLysAlaLysAsnTyrArgL

 2071  TCATGGACATTGAGAAGCGACGCGAGGTGTATCGCTCCATCTTTGATCCGGCAGTGCACGATGCACTGGTGCGTCGTGGGGATCGCCGGT  2160
        euMetAspIleGluLysArgArgGluValTyrArgSerIlePheAspProAlaValHisAspAlaLeuValArgArgGlyAspArgArgP

 2161  TTAGCCATCGTGCCCTTCAGGGAGCCATCATGATCACCTTCTATAGGGATGAACCCAGGTTCAGCCAACCACACCAGTTGCTCACCCTGC  2250
        heSerHisArgAlaLeuGlnGlyAlaIleMetIleThrPheTyrArgAspGluProArgPheSerGlnProHisGlnLeuLeuThrLeuL
                                                         ▽36f
 2251  TCATGGACATCGACTCGTTAATAACCAAGTGGAGATGTAAGTATTGCATTCTTTGATACTCTTTTATAAATATATCTTATGTTTAAGACT  2340
        euMetAspIleAspSerLeuIleThrLysTrpArgT

 2341  GGTTTTCCTAACCAAATACTTTCTATTCCCGCCGCAGACAATCACGTGATCATGGTGCAACGCATGATTGGATCCCAACAGTTGGGCACT  2430
                                       yrAsnHisValIleMetValGlnArgMetIleGlySerGlnGlnLeuGlyThr

 2431  GGTGGCTCGTCTGGATATCAATATCTGCGCTCCACTCTCAGGTGATCATCGCAGATGTGATTATATCGGGGATCAATGAACTCAAACTGT  2520
        GlyGlySerSerGlyTyrGlnTyrLeuArgSerThrLeuSe

 2521  TCTCCCTTTGTTTTTTTTTGGTTTCAGTGATCGGTACAAGGTGTTTCTGGATCTGTTCAATCTGTCCACTTTTCTGATTCCCCGCGAGGC  2610
                            rAspArgTyrLysValPheLeuAspLeuPheAsnLeuSerThrPheLeuIleProArgGluAl
        ▽H2a
 2611  GATTCCACCGCTGGACGAGACCATTCGCAAGAAACTGATCAACAAAAGTGTCTGACAATCGGCAGGGTATCCAATTGGTCAATGTTTGGC  2700
        aIleProProLeuAspGluThrIleArgLysLysLeuIleAsnLysSerValEnd
                                                               ↓
 2701  TATGCGTTGTTTGTTCTGCCTACTGTTTTGTCGTTTTGGTGTAATAAAATTACTTGTTTAGTCTTTGTTATCACATTTGATGTGTTCCTT  2790
 2791  TTCTTTATGTCTGACATATAATACATATAACATAACAAAATAAATATTCATATTTCAGACATAAACAAATTCTATGGGAATGTGTGAGTC  2880
 2881  AGCAGCCTGAAAGTAGACCATATATATTCTGGTTGTCTTTCTCGCTCGTTTCTATTAGTTCGTTAGCAAATTAAATTCCATAATATTGTG  2970
 2971  TGGCAATACTTGTCAAAATAATAATGGTATAAGTGAATTTTAATTACAAAATACCGATTTAAACAAAAAGCTT                    3043
```

FIG. 3. Sequence of the *v* gene. The genomic sequence from the *Eco*RI site (coordinate 0 in Fig. 2) to the *Hin*dIII site (coordinate +3.0 in Fig. 2) is shown. The following restriction sites are underlined: *Eco*RI, nt 1 to 6 and nt 1877 to 1882; *Xho*I, nt 1137 to 1142; *Bam*HI, nt 2310 to 2315; and *Hin*dIII, nt 3038 to 3043. Symbols: ↓, 5′ ends of the longest cDNA clones; ♦, 3′ end of one cDNA, ▽, transposon insertion sites. The boundaries of the deletion mutation are indicated by brackets above the relevant sequences. The major transcription start site (+1) was deduced by S1 and primer extension analyses. Splice sites are indicated by interruptions of the translational reading frame.
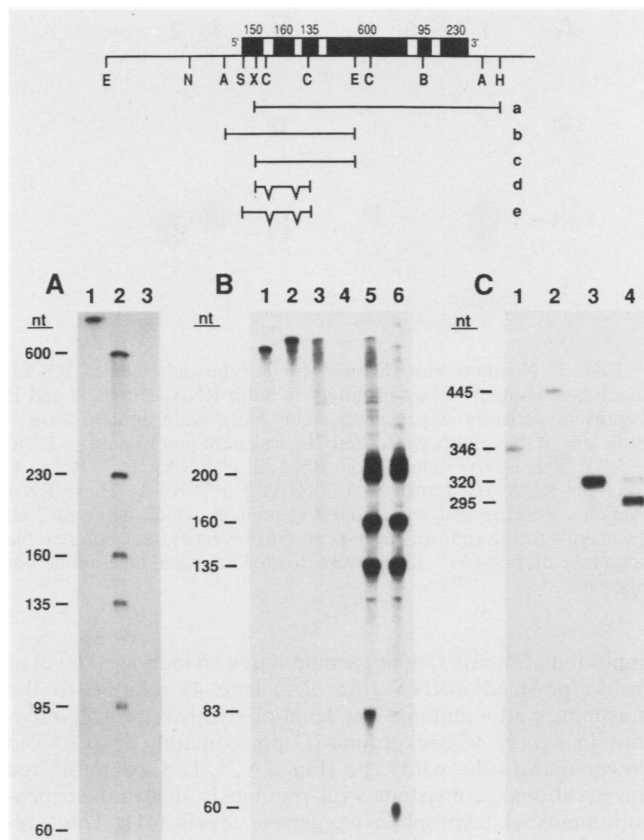
FIG. 4. S1 protection analysis of v^+ exons. The exon lengths indicated above the solid bars in the map are approximations based on the cDNA sequence. Probes a, b, and c are derived from genomic clones, whereas probes d and e are derived from cDNA clones. Restriction enzyme abbreviations: A, *Acc*I; B, *Bam*HI; C, *Cla*I, E, *Eco*RI; H, *Hin*dIII; N, *Nde*I; S, *Sac*I; X, *Xho*I. In lanes containing probe alone there was no S1 treatment. In the experimental lanes (probe plus RNA), wild-type poly(A)$^+$ RNA from second-instar larvae was annealed to radioactively labeled antisense RNA probes and subsequently treated with S1 nuclease. In the no-RNA controls, probe was incubated with tRNA and subsequently treated with S1 nuclease. Panels A to C represent separate experiments. (A) Lanes: 1, probe a; 2, probe a plus RNA; 3, probe a, no-RNA control. (B) Lanes: 1, probe c; 2, probe b; 3, probe b, no-RNA control; 4, probe c, no-RNA control; 5, probe b plus RNA; 6, probe c plus RNA. (C) Lanes: 1, probe d; 2, probe e; 3, probe e plus RNA; 4, probe d plus RNA. No bands were detected in the no-RNA controls for the experiment in panel C (data not shown). The faint band at 320-nt in lane 4C was caused by slight contamination with probe e.

sequences between the *Cla*I and *Xho*I sites were completely protected when annealed to wild-type RNA (Fig. 4C, lane 4). However, only 25 additional nucleotides of cDNA sequences upstream of the *Xho*I site were protected (Fig. 4C, lane 3). We did not readily detect the 374-nt fragment, which would be observed if the entire v cDNA fragment had been protected. A faint band of this size was seen in longer exposures; however, we do not know whether it is a relevant product. The same S1 protection pattern was observed regardless of the source (i.e., developmental stage or *D. melanogaster* strain) of RNA (data not shown).

We independently mapped the transcription start site by primer extension experiments. End-labeled synthetic oligonucleotides, complementary to sequences near the 5' end of v RNA, were used to prime cDNA synthesis. Oligo 1235, an
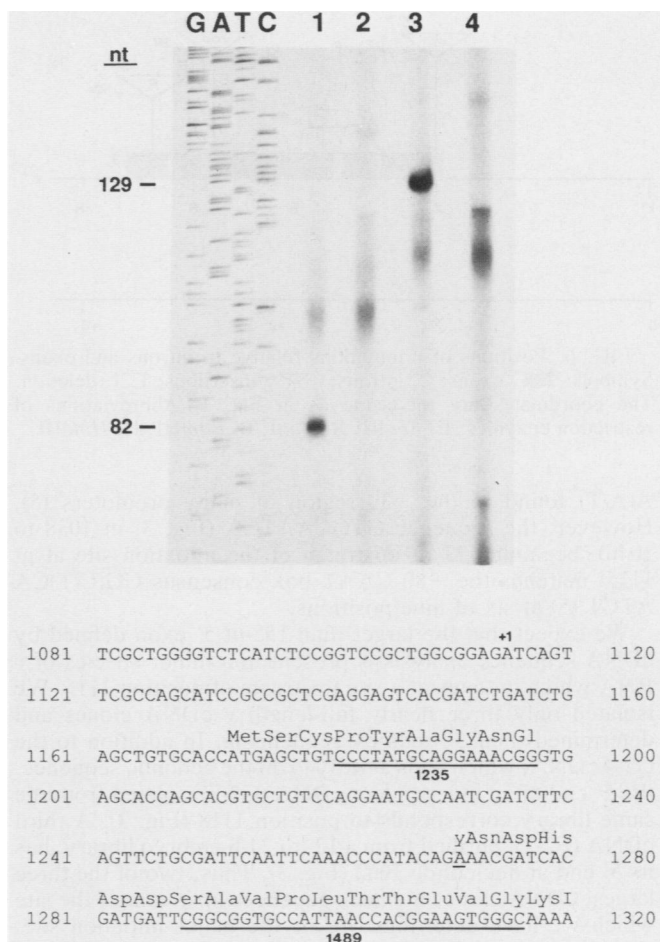

FIG. 5. Mapping the 5' end the v by primer extension. End-labeled synthetic oligos 1235 and 1489, complementary to the underlined sequences, were used to prime cDNA synthesis on wild-type RNA. Lanes: 1, oligo 1235 plus RNA; 2, oligo 1235 plus tRNA; 3, oligo 1489 plus RNA; 4, oligo 1489 plus tRNA. The lengths of the extension products are indicated. M13mp18 DNA was sequenced by using the universal primer and used as the size standards. The major transcription start site (+1) is indicated on the DNA sequence.

18-mer which anneals to 17 nt in exon 1 and 1 nt in exon 2 (Fig. 5, nt 1181 to 1197 and nt 1271, respectively), primed the synthesis of an 82-nt extension product (Fig. 5, lane 1), which places the 5' end of the RNA at the C residue at position 1117 (Fig. 3 and 5). Oligo 1489, a 29-mer which anneals to exon 2 sequences (Fig. 5, nt 1288 to 1316), generated a 129-nt extension product (Fig. 5, lane 3), placing the 5' end at nt 1115 (Fig. 3 and 5).

Thus, both primer extension and S1 mapping experiments place the 5' end of the v RNA at approximately the same site. Within this region is found the heptanucleotide sequence ATCAGTT (Fig. 3 and 5, nt 1115 to 1121), which perfectly matches the consensus sequence ATCA(G/T)T(C/T) which is found at the transcription initiation sites of a number of *D. melanogaster* genes (11) and which is thought to be important for capping. On the basis of primer extension and S1 analysis, we have designated the A residue at nt 1115 (Fig. 3 and 5) as the major v initiation site. Transcription of other *D. melanogaster* genes is most frequently found to start at that position within the heptanucleotide consensus. Upstream of the initiation site at nt 1115 there are no sequences similar to the TATA box consensus TATA(A/T)
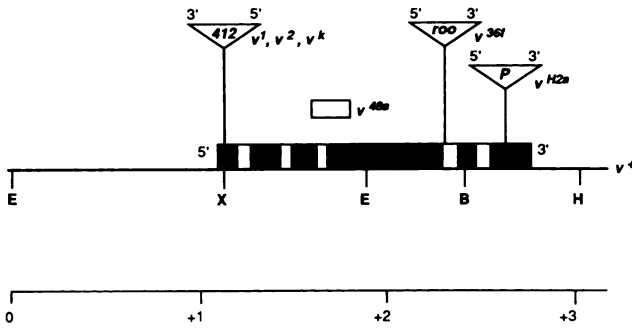
FIG. 6. Positions of ν mutations relative to introns and exons. Symbols: ■■, exons; ⌐, introns; ▽, insertions; □, deletion. The coordinates are the same as in Fig. 1. Abbreviations of restriction enzymes: E, *Eco*RI; X, *Xho*I; B, *Bam*HI; H, *Hin*dIII.
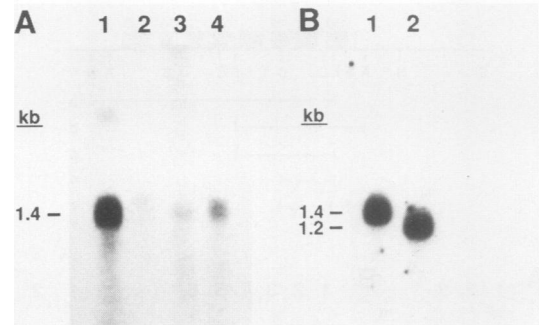

FIG. 7. Northern blot analysis of wild-type and mutant ν RNAs. Each lane contains 20 μg of poly(A)⁺ adult RNA. Panels A and B represent separate experiments. The blots were probed with a subclone of the 1.9-kb *Xho*I-*Hin*dIII ν fragment (coordinates +1.1 to +3.0 in Fig. 1). (A) Lanes: 1, ν⁺ RNA; 2, ν¹ RNA; 3, ν³⁶ᶠ RNA; 4, *su(s)²* ν¹ RNA. (B) Lanes: 1, ν⁺ RNA; 2, ν⁴⁸ᵃ RNA. These RNA samples were probed with cloned sequences which detect a 2-kb transcript to the right of the ν gene (36) to verify that comparable amounts of poly(A)⁺ RNA were loaded in each lane (data not shown).

A(A/T) found in the −30 region of many promoters (5). However, the sequence GTGCAATCA (Fig. 3, nt 1038 to 1046), beginning 77 nt upstream of the initiation site at nt 1115, matches the −80 CAAT box consensus GG(C/T)CA ATCT (5) at six of nine positions.

We expect that the larger-than-152-nt 5' exon defined by cDNA sequence analysis is present in a minor species of ν RNA which initiates at a site upstream of position 1115. We isolated only three nearly full-length ν cDNA clones and determined their 5' ends by sequencing. In addition to the G2-4 cDNA, which ends at nt 1046 on the genomic sequence, the 5' end of an independent cDNA (G2-7) isolated from the same library corresponds to position 1118 (Fig. 3). A third cDNA (F2-2), purified from a 12- to 24-h embryo library, has its 5' end at nucleotide 1062 (Fig. 3). Thus, two of the three largest cDNA clones include sequences upstream of the site which we have determined to be the major initiation site. The sequence ATCAGTC, which also perfectly matches the heptanucleotide consensus sequence, is found in the region where the largest cDNA clone ends (Fig. 3, nt 1043 to 1049). If transcription initiates at the A residue at position 1043, the 5' exon of this minor ν RNA species would be 155 nt. However, since we have not detected these 5' sequences in S1 or primer extension experiments, initiation at this position has not been verified. There are no potential TATA or CAAT boxes upstream of the possible minor transcription start site at position 1043.

Analysis of the 412 insertions. The ν¹, ν², and ν^k mutations differ from other ν alleles in that they are both leaky and suppressible by the suppressor of sable. Each of these mutations is associated with a 412 insertion in the 5' gene region (Fig. 1). We cloned ν DNA from the three mutants (Fig. 1) and localized the positions of mutations by DNA sequencing. The 412 element is located at the same position, between nt 1150 and 1151, in all three mutations (Fig. 3). This insertion site is within exon 1, 36 nt downstream of the major transcription start site and 21 nt upstream of the translation initiation site (Fig. 3 and 6). On the basis of the restriction map of the ν 412 insertion mutants (Fig. 1) and the sequence data, we determined that the 412 element in each mutant is oriented such that it would be transcribed in the opposite direction to ν (Fig. 6). Northern blot analysis showed that, as reported previously (36), the level of ν RNA is drastically reduced in the 412 insertion mutant ν¹. However, a faint, wild-type-sized band is consistently observed (Fig. 7A, lane 2). Similarly, reduced levels of transcript are apparent in ν² and ν^k mutants (data not shown). The *su(s)²* ν¹ mutant, which has a 412 insertion mutation at ν and a

mutation at the *su(s)* gene, accumulates an increased level of wild-type-sized mRNA (Fig. 7A, lane 4) relative to the unsuppressed ν mutant. The level of ν transcript accumulation in a *su(s)²* ν¹ background is approximately 5- to 10-fold lower than in the wild type (Fig. 7A; L.L.S., unpublished observations), consistent with previously determined measurements of tryptophan oxygenase levels (41). Thus, although a 7.5-kb transposable element has inserted into an exon, the mature RNA is the same size as in the wild type. A more detailed analysis of the ν RNA produced in the presence of the suppressor is currently in progress.

Analysis of other v alleles. We had previously determined that the ν⁴⁸ᵃ allele has a 200-nt deletion in the ν gene between the *Xho*I and *Eco*RI sites (Fig. 1, coordinates +1.1 and +1.9). Northern analysis showed that the ν⁴⁸ᵃ mutant accumulates the same level of RNA as a wild-type strain (Fig. 7, lane 2), and, as expected, the RNA is slightly smaller than the wild-type ν transcript. When poly(A)⁺ RNA from the ν⁴⁸ᵃ mutant was hybridized to wild-type sequences between the *Xho*I and *Hin*dIII sites (probe a) and subsequently treated with RNase, the 600- and 135-nt protected ν⁺ fragments were not detected, and new fragments, approximately 480 and 110 nt, were observed instead (Fig. 8A, lane 3). The region of the 600-nt exon affected is the portion which lies upstream rather than downstream of the *Eco*RI site (Fig. 8A, lanes 6 and 9). Sequence analysis of cloned ν⁴⁸ᵃ DNA revealed that sequences between nt 1599 and 1802 (Fig. 3) are deleted in this mutant. Interestingly, the deletion does not alter the translational reading frame (Fig. 3); thus, a mutant polypeptide consisting of 50 fewer amino acids might be synthesized from this mutant gene.

The ν³⁶ᶠ mutant was previously shown to contain a B104/roo insertion to the right of the 412 insertion site at ν (Fig. 1). The sequence of mutant DNA revealed that the insertion occurred between nt 2298 and 2299 (Fig. 3) in intron 4 of the ν gene, 12 nt downstream of the splice donor site. The element is oriented such that it would be transcribed in the same direction as ν (Fig. 6). The ν³⁶ᶠ mutation drastically reduces the level of ν transcript accumulation. A faint signal was frequently observed on Northern blots, indicating that at least some ν RNA accumulates which is approximately the same size as in the wild type (Fig. 7A, lane 3). The ν³⁶ᶠ
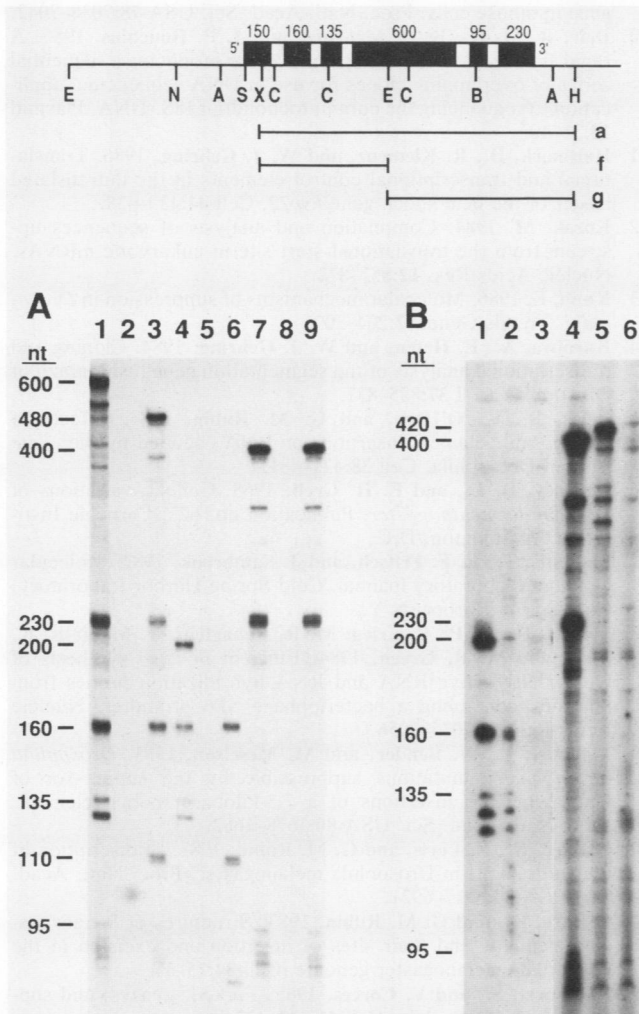
FIG. 8. RNase mapping of v mutant exons. The exon lengths indicated above the solid bars on the map are approximations based on the cDNA sequence. Probes a, f, and g are derived from genomic clones. Restriction enzyme abbreviations: E, EcoRI; N, NdeI; A, AccI; S, SacI; X, XhoI; C, ClaI; B, BamHI; H, HindIII. In these experiments, antisense RNA probes were annealed to RNA from wild-type or mutant adults, and the hybrids were treated with RNase. When RNase was used, several protected fragments appeared as doublets or triplets, owing to variable RNase cleavage at the ends. Panels A and B represent separate experiments. (A) Lanes: 1, probe a and v^+ RNA; 2, probe a and v^1 RNA; 3, probe a and v^{48a} RNA; 4, probe f and v^+ RNA; 5, probe f and v^1 RNA; 6, probe f and v^{48a} RNA; 7, probe g and v^+ RNA; 8, probe g and v^1 RNA; 9, probe g and v^{48a} RNA. (B) Lanes: 1, probe f and v^+ RNA; 2, probe f and v^{36f} RNA; 3, probe f and v^1 RNA; 4, probe g and v^+ RNA; 5, probe g and v^{36f} RNA; 6, probe g and v^1 RNA. The sizes of the protected fragments are indicated.

mutant transcript is more readily detectable in RNase mapping experiments (Fig. 8B). This analysis showed that the mutant RNA contains v sequences upstream of the B104/roo insertion site but lacks v sequences downstream of the insertion. Exons 5 and 6 are not present in v^{36f} RNA, and a few additional nucleotides are included in exon 4 (Fig. 8B, lane 5). This result would be expected if the RNA includes intron sequences upstream of the B104/roo insertion site and terminates at a site within the element. Polyadenylation in the 5' long terminal repeat of the B104/roo element would

generate a mature RNA of approximately the same size as the v^+ RNA. Unlike the v^1, v^2, and v^k alleles, the v^{36f} mutant does not exhibit a leaky phenotype and has no detectable tryptophan oxygenase activity (41). Thus, sequences downstream of the insertion site are probably needed to produce functional enzyme. The suppressor of sable has been reported to slightly elevate v expression in the 36f mutant (41). However, this effect is not detectable by Northern analysis (data not shown).

The P-element insertion mutation, v^{H2a}, that was used to clone v is located to the right of the B104/roo insertion site (Fig. 1). Sequence analysis revealed that the insertion is within v exon 6 between nt 2626 and 2627 (Fig. 3 and 6). The 8-nt sequence adjacent to the insertion site, CGCTGGAC, matches the P-element target site consensus GGCCAGAC (21) at five of eight positions. The P-element insertion at this site within the gene would be expected to alter the carboxy-terminal 12 amino acids of tryptophan oxygenase. An insertion at this position might also affect the termination of transcription or translation. However, we have not examined v transcript accumulation in the v^{H2a} mutant.

DISCUSSION

The v^1, v^2, and v^k mutants have 412 insertions at identical positions, and the elements are in the same orientation. These observations raise the possibility that these mutations have a common origin, although they were isolated independently, i.e., in different laboratories and at different times (16, 37). An alternative explanation would be that specific sequences are recognized as target sites for 412 insertion. Will et al. (44) have shown that a 4-base-pair sequence is duplicated at the insertion site. Our unpublished data indicate that the sequence 5'-CACG-3' located immediately upstream of the 412 insertion site (Fig. 3) is duplicated. The complement of this sequence, 5'-CGTG-3', is similar to the three other sequenced genomic 412 target site duplications, CTTG (two independent events) and CTGG, but is less similar to the duplication in a free 412 long terminal repeat, ATAG, (38, 44). Thus, although the number of target sites sequenced is small, specificity for 412 element insertion sites seems quite probable.

Furthermore, this work has established that the 412 insertions are in the 5' untranslated region of the v gene and that the RNA synthesized in $su(s)$ v double mutants is apparently the same size as that in the wild type. A similar v RNA accumulates at a very low level even in the absence of the suppressor mutation, and this low level of expression may be responsible for the leaky phenotype associated with these mutants. These observations raise two questions. First, how is the wild-type sized RNA generated with a 7.5-kb transposon inserted into exon sequences? Second, how does a suppressor mutation act to elevate the level of v mRNA? One mechanism for generating a wild-type-sized RNA would be the excision of transposable element sequences by splicing, either at sites within the v gene or within the element. An alternative mechanism would be the initiation of transcription at a different site, either in transposon sequences adjacent to the v coding region or in v sequences just downstream of the insertion.

It was important to define v splice junctions in the vicinity of the 412 insertion in order to determine whether wild-type strains have alternative splicing mechanisms that could eliminate the transposable element sequences at v. Although our experiments suggest that more than one transcription start site may be used, we have not detected alternatively spliced

products. Our S1 mapping experiments demonstrated conclusively that all wild-type v mRNAs include sequences from the 412 insertion site. This result eliminates the trivial explanation that the 412 insertion would be in an intron of a minor species of v RNA. In addition, there are no splice donor sites in v sequences upstream of the 412 insertion which could be spliced to a downstream acceptor to eliminate 412 sequences from the RNA. Therefore, these results suggest that if the 412 sequences are removed from the RNA by splicing, this event would involve sequences that are not normally used in generating the v RNA. Since the insertion is in the untranslated region, the removal of transposon sequences could be imprecise and still produce an RNA which could be translated into a wild-type polypeptide.

Our S1 mapping experiments demonstrated that there are no wild-type transcription initiation sites downstream of the 412 insertion site. Furthermore, it is unlikely that the 412 promoter controls transcription initiation in the insertion mutant because, in the observed orientation, this promoter would direct transcription away from v. Therefore, if the site of transcription initiation eliminates transposon sequences, the initiation site would also be one that is not normally used in v or 412 DNA. Future experiments will determine whether transposon sequences are removed by splicing or by initiation at a new site and define how $su(s)$ mutations alleviate the effect caused by 412 insertion.

LITERATURE CITED

1. **Baillie, D. L., and A. Chovnick.** 1971. Studies on the genetic control of tryptophan pyrrolase in Drosophila melanogaster. Mol. Gen. Genet. **112:**341–353.
2. **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.
3. **Biggs, J., L. L. Searles, and A. L. Greenleaf.** 1985. Structure of the eucaryotic transcription apparatus: features of the gene for the largest subunit of Drosophila RNA polymerase II. Cell **42:**611–621.
4. **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.
5. **Breathnach, R., and P. Chambon.** 1981. Organization and expression of eukaryotic split genes coding for proteins. Annu. Rev. Biochem. **50:**349–383.
6. **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.
7. **Chang, D.-Y., B. Wisely, S.-M. Huang, and R. A. Voelker.** 1986. Molecular cloning of *suppressor of sable*, a Drosophila melanogaster transposon-mediated suppressor. Mol. Cell. Biol. **6:**1520–1528.
8. **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.
9. **Corces, V., A. Pellicer, R. Axel, and M. Meselson.** 1981. Integration, transcription, and control of a Drosophila heat shock

gene in mouse cells. Proc. Natl. Acad. Sci. USA **78:**7038–7042.
10. **Dale, R. M. K., B. A. McClure, and J. P. Houchins.** 1985. A rapid single-stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: application to sequencing the corn mitochondrial 18S rDNA. Plasmid **13:**31–40.
11. **Hultmark, D., R. Klemenz, and W. J. Gehring.** 1986. Translational and transcriptional control elements in the untranslated leader of the heat-shock gene *hsp22*. Cell **44:**429–438.
12. **Kozak, M.** 1984. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucleic Acids Res. **12:**857–872.
13. **Kubli, E.** 1986. Molecular mechanisms of suppression in *Drosophila*. Trends Genet. **2:**204–209.
14. **Kuroiwa, A., E. Hafen, and W. J. Gehring.** 1984. Cloning and transcriptional analysis of the segmentation gene fushi tarazu of Drosophila. Cell **37:**825–831.
15. **Levis, R., K. O'Hare, and G. M. Rubin.** 1984. Effects of transposable element insertions on RNA encoded by the *white* gene of Drosophila. Cell **38:**471–481.
16. **Lindsley, D. L., and E. H. Grell.** 1968. Genetic variations of *Drosophila melanogaster*. Publication no. 627. Carnegie Institution, Washington, D.C.
17. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
18. **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.
19. **Modollel, J., W. Bender, and M. Meselson.** 1983. *Drosophila melanogaster* mutations suppressible by the suppressor of Hairy-wing are insertions of a 7.3-kilobase mobile element. Proc. Natl. Acad. Sci. USA **80:**1678–1682.
20. **O'Hare, K., R. Levis, and G. M. Rubin.** 1983. Transcription of the *white* locus in Drosophila melanogaster. Proc. Natl. Acad. Sci. USA **80:**6917–6921.
21. **O'Hare, K., and G. M. Rubin.** 1983. Structures of P transposable elements and their sites of insertion and excision in the Drosophila melanogaster genome. Cell **34:**25–35.
22. **Parkhurst, S., and V. Corces.** 1985. Forked, gypsys, and suppressors in Drosophila. Cell **41:**429–437.
23. **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.
24. **Parkhurst, S. M., and V. G. Corces.** 1987. Developmental expression of *Drosophila melanogaster* retrovirus-like transposable elements. EMBO J. **6:**419–424.
25. **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:**1202–1215.
26. **Phillips, J. P., and H. S. Forrest.** 1980. Ommochromes and pteridines, p. 542–623. *In* M. Ashburner and T. R. F. Wright (ed.), The genetics and biology of *Drosophila*, vol. 2d. Academic Press, Inc., New York.
27. **Pirrotta, V., and C. Brockl.** 1984. Transcription of the *Drosophila white* locus and some of its mutants. EMBO J. **3:**563–568.
28. **Poole, S. J., L. M. Kauvar, B. Drees, and T. Kornberg.** 1985. The *engrailed* locus of Drosophila: structural analysis of an embryonic transcript. Cell **40:**37–43.
29. **Proudfoot, N. J., and G. G. Brownlee.** 1976. The 3' non-coding sequences in eukaryotic messenger RNA. Nature (London) **263:**211–214.
30. **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.
31. **Rubin, G. M.** 1983. Dispersed repetitive DNAs in *Drosophila*, p. 329–361. *In* J. A. Shapiro (ed.), Mobile genetic elements. Academic Press, Inc., New York.

32. **Sanger, F., A. R. Coulson, B. G. Barrell, A. J. H. Smith, and B. A. Roe.** 1980. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. **143:**161–178.

33. **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.

34. **Schwartz, H. E., T. J. Lockett, and M. W. Young.** 1982. Analysis of transcripts from two families of nomadic DNA. J. Mol. Biol. **157:**49–68.

35. **Searles, L. L., R. S. Jokerst, P. M. Bingham, R. A. Voelker, and A. L. Greenleaf.** 1982. Molecular cloning of sequences from a Drosophila RNA polymerase II locus by P element transposon tagging. Cell **31:**585–592.

36. **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.

37. **Shapard, P. B.** 1960. A physiological study of the *vermilion* eye color mutants of *Drosophila melanogaster.* Genetics **45:**359–376.

38. **Shepard, B. M., and D. J. Finnegan.** 1984. Structure of circular copies of the 412 transposable element present in *Drosophila melanogaster* tissue culture cells, and isolation of a free 412 long terminal repeat. J. Mol. Biol. **180:**21–40.

39. **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.

40. **Staden, R.** 1980. A new computer method for the storage and manipulation of DNA gel reading data. Nucleic Acids Res. **8:**3673–3694.

41. **Tartof, K. D.** 1969. Interacting gene systems. I. The regulation of tryptophan pyrrolase by the vermilion-suppressor of vermilion system in Drosophila. Genetics **62:**781–795.

42. **Thomas, P. S.** 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. USA **77:**5201–5205.

43. **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.

44. **Will, B. M., A. A. Bayev, and D. J. Finnegan.** 1981. Nucleotide sequence of terminal repeats of 412 transposable elements of *Drosophila melanogaster.* J. Mol. Biol. **153:**897–915.

45. **Yuki, S., S. Inouye, S. Ishimaru, and K. Saigo.** 1986. Nucleotide sequence characterization of a *Drosophila* retrotransposon, 412. Eur. J. Biochem. **158:**403–410.

46. **Zachar, A., T.-B. Chou, and P. M. Bingham.** 1987. Evidence that a regulatory gene autoregulates splicing of its transcript. EMBO J. **6:**4105–4111.

47. **Zachar, A., D. Garza, T.-B. Chou, J. Goland, and P. M. Bingham.** 1987. Molecular cloning and genetic analysis of the suppressor-of-white apricot locus from *Drosophila melanogaster.* Mol. Cell. Biol. **7:**2498–2505.

48. **Zachar, Z., D. Davison, D. Garza and P. M. Bingham.** 1985. A detailed developmental and structural study of the transcriptional effects of insertion of the copia transposon into the white locus of *Drosophila melanogaster.* Genetics **111:**495–515.