

Aberrant Splicing and Transcription Termination caused by *P* Element Insertion into the Intron of a *Drosophila* Gene

Heidi Horowitz and Celeste A. Berg

Department of Genetics, University of Washington, Seattle, Washington 98195

Manuscript received August 1, 1994

Accepted for publication September 30, 1994

ABSTRACT

Insertional mutagenesis screens using the *P*[*lacZ*, *rosy*⁺] (*PZ*) transposable element have provided thousands of mutant lines for analyzing genes of varied function in the fruitfly, *Drosophila melanogaster*. As has been observed with other *P* elements, many of the *PZ*-induced mutations result from insertion of the *P* element into the promoter or 5' untranslated regions of the affected gene. We document here a novel mechanism for mutagenesis by this element. We show that sequences present within the element direct aberrant splicing and termination events that produce a mRNA composed of 5' sequences from the mutated gene (in this case, *pipsqueak*) and 3' sequences from within the *P*[*lacZ*, *rosy*⁺] element. These truncated RNAs could yield proteins with dominant mutant effects.

THE use of *P* transposable elements to create mutations in the fruitfly *Drosophila melanogaster* has been an invaluable tool for *Drosophila* geneticists. In particular, the development of a technique to create single *P* element insertions throughout the genome (COOLEY *et al.* 1988a,b; ROBERTSON *et al.* 1988) and the inclusion of the *lacZ* gene in *P* elements to potentiate their use in "enhancer trap" screens (O'KANE and GEHRING 1987; BELLEN *et al.* 1989; BIER *et al.* 1989; WILSON *et al.* 1989) has greatly facilitated the generation and subsequent characterization of new mutations.

P element insertions may be mutagenic for a variety of reasons, depending on the position of insertion relative to the affected gene. *P* elements show a preference for inserting near transcription start sites (TSUBOTA *et al.* 1985; KELLEY *et al.* 1987; ENGELS 1989). Thus, the insertion may interfere with promoter function, creating a hypomorphic allele by reducing the level of wild-type transcript. Alternatively, insertions into the 5' untranslated regions of genes may affect stability or result in premature termination of the transcript. In addition, insertions into coding regions disrupt open reading frames (ORFs).

Additional alleles of the gene are easily generated by excision of the *P* element after introduction of a source of transposase. Imprecise excisions (ENGELS 1979) can delete genomic DNA flanking the *P* element insertion site, producing (in most cases) a more severe mutant phenotype. The characterization of these deletions often helps in the analysis of the gene, and the existence of a number of alleles with phenotypes of varying severity can often be pivotal in determining function for that gene.

Corresponding author: Celeste Berg, Department of Genetics, SK-50, University of Washington, Seattle, WA 98195.
E-mail: berg@genetics.washington.edu

We describe here the identification of a previously undocumented mechanism of *P* element mutagenesis. Our results show that when the *P*[*lacZ*, *rosy*⁺] element is located within an intron of a gene, specific sequences within the *P* element can direct aberrant splicing and transcription termination events, resulting in a mutant phenotype for that gene. In the example presented, we demonstrate that this aberrant splicing event is entirely responsible for the mutant phenotype observed. Our data suggest that any *P* element bearing the *Hind*III fragment of the *rosy* gene and its flanking DNA could have this effect. In addition, we discuss the ramifications of our findings for simplifying the characterization of certain *P*[*lacZ*, *rosy*⁺] (*PZ*)-induced mutations and suggest features to be considered in the future design of *P* elements to be used for mutagenesis.

MATERIALS AND METHODS

Fly stocks: *PZ* insertional mutations at the *pipsqueak* (SIEGEL *et al.* 1993) gene, lines 2403, 0115, 8109, and 0482, were generated in a large mutagenesis screen conducted in the laboratory of Dr. ALLAN SPRADLING [described in KARPEN and SPRADLING (1992); referred to as *zeppelin* in SPRADLING (1993)]. All four lines contain a *PZ* element [described in MILODZIK and HIROMI (1992)] inserted in an isogenized *cn* chromosome in a *ry*⁵⁰⁶ background. Excision lines were created from the 2403 allele by reintroducing the $\Delta 2-3$ transposase and screening for loss of the *ry*⁺ eye-color marker (see Figure 1). Specifically, *cn psq^{PZ2403(ry⁺)}/CyO; ry⁵⁰⁶* males were mated in bottles to *Sp/SM5; Cy; ry⁵⁰⁶ Sb P[$\Delta 2-3$, *ry*⁺]^{99B/TM6}* virgins. Sons carrying both transposons (*cn psq^{PZ2403(ry⁺)}/CyO; ry⁵⁰⁶ Sb P[$\Delta 2-3$; *ry*⁺]^{99B/ry⁵⁰⁶}) were mated in vials to *Sco/CyO; ry⁵⁰⁶* virgins. Fifty-four matings were established, each with two males and five females. Fifteen days later, the progeny were screened for *ry*⁻ sons that were either *Sco⁺ Cy⁻* or *Sco⁻ Cy⁻*. These progeny bore *PZ* chromosomes in which an excision event had induced loss of the *ry*⁺ marker. All 54 matings provided at least one *ry*⁻ son with the appropriate genotype, most generating more (as many as 12). From*

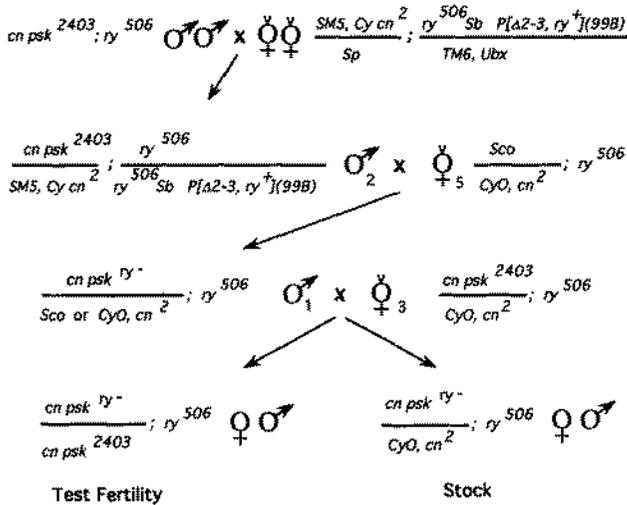


FIGURE 1.—Crosses used to excise the $P[lacZ, ry^+]$ element in line 2403. Stocks were established from single ry^- males. Viability and fertility were examined for flies transheterozygous with the starting allele and also for flies homozygous for the excision allele (see MATERIALS AND METHODS).

each vial, we selected at most two males and mated them individually to $cn PZ^{2403}/CyO; ry^{506}$ virgins. From this cross, females carrying the excision chromosome *in trans* to the starting 2403 allele were tested for fertility. In addition, stocks were founded containing the excision chromosome (ry^-) over CyO . The viability and fertility of flies homozygous for the excision chromosome were then examined.

The $P[lacZ, w^+]$ line *fs1* was generated by BIER *et al.* (1989) and was generously provided by VIVIAN SEGEL.

DNA probes: *pipsqueak* probe: A 6.7-kb *HindIII* fragment flanking the *fs1* P element (provided by V. SEGEL) was used as a probe to an ovary cDNA library (provided by Dr. T. HAZELRIGG). Two cDNA clones, 8'A and 11'A, were obtained. Northern analysis revealed that these clones contained cDNAs from multiple unrelated mRNAs ligated into the same vector. The library was rescreened using a portion of the 11'A clone, and one of the clones obtained, 13-13, contained a 1-kb insert that by Northern analysis was entirely derived from the *psq* gene. A 300 bp *EcoRI* DNA fragment representing the most 5' region of this clone was used in the Northern hybridization described here. This sequence is located within the 5' UTR of the *psq* gene (H. HOROWITZ and C. BERG, unpublished results).

l(3)S12 probe: A 1376 bp *Msd-HindIII* fragment extending from -1755 to -3131 of the *rosy* gene [numbering as in LEE *et al.* (1987); DUTTON and CHOVNICK (1991)] was isolated from the plasmid pCar20 (RUBIN and SPRADLING 1983). This fragment contains all of the putative ORFs identified for the *l(3)S12* gene (DUTTON and CHOVNICK 1991), as well as 1155 bp of additional 3' sequence.

3' P element probe: A 245-bp *HindIII-Msd* fragment comprising all of the 3' P element sequence was isolated from the plasmid pCar20 (RUBIN and SPRADLING 1983).

Nucleic acid hybridizations: DNA was prepared from adult flies as described (JOWETT 1986). DNA from five flies was digested with the appropriate restriction endonuclease(s), subjected to agarose gel electrophoresis, and blotted to nylon membranes as previously described (SOUTHERN 1975; REED and MANN 1985). [^{32}P]-labeled DNA probes were prepared using the random hexamer primer technique of FEINBERG and VOGELSTEIN (1983, 1984). Hybridizations were per-

formed according to the method of CHURCH and GILBERT (1984).

Total RNA was isolated from adult female flies or from hand-dissected ovaries using the hot phenol method (JOWETT 1986). Poly(A)⁺ RNA was selected using the Qiagen Oligotex-dT mRNA kit. Approximately 0.5 μ g of poly(A)⁺ mRNA was loaded per lane on a 1% agarose-formaldehyde gel in 1× MOPS buffer [20 mM MOPS (pH 7.0), 8 mM sodium acetate, 1 mM EDTA] (SAMBROOK *et al.* 1989). RNA size standards (Bethesda Research Laboratories) were used for molecular weight determinations. After electrophoresis, gels were transferred to ZetaProbe nylon filters (BioRad) in 20× SSC overnight. RNA on filters was subjected to UV cross-linking using a Stratagene Stratalinker and hybridized as described below.

Northern blots were prehybridized for 3 h at 42° in 50% formamide, 5× SSC, 1× Denhardt's solution, 1% SDS, 20 mM NaPO₄ (pH 7.0), 100 μ g/ml single-stranded salmon sperm DNA, and then hybridized overnight in the same solution to which [^{32}P]-labeled DNA probe was added. Blots were washed twice in 1× SSC, 1% SDS at 42° for 15 min and once in 0.1× SSC, 1% SDS at 50° for 15 min and subjected to autoradiography.

PCR reaction analysis: The truncated transcript produced in homozygous 2403 adult female flies was analyzed as follows: 1 μ g of poly(A)⁺ mRNA was reverse transcribed according to the manufacturer's instructions in a total volume of 50 μ l using random hexamers (2 pmoles/ μ l) as primers and 200 units of M-MLV Reverse Transcriptase (GIBCO BRL). Five microliters of the reverse-transcription reaction was then used as a template for a PCR (100 μ l total volume) using buffers provided by the manufacturer of the Taq1 polymerase (Boehringer Mannheim). The primers used (see Figure 4A) were 18: 5'-GCAACGTAACGGGAAGCTCTTTG-3', derived from the 5' UTR of *psq* [221-200 nucleotides upstream from the predicted start of translation (HOROWITZ and BERG, in preparation)], and 17: 5'-TACAGTCGACGCACGGTTTCAA-3', derived from the putative 3' UTR of the *l(3)S12* gene (DUTTON and CHOVNICK 1991). This primer extends from -2885 to -2900 of the published *rosy* gene sequence [numbering as in LEE *et al.* (1987); DUTTON and CHOVNICK (1991)], CGACGCACGGTTTCAA, and has a *SalI* site (GTCGAC) engineered in by the inclusion of the six most 5' bases of the primer, TACAGT.

PCR thermocycler (Eppendorf) conditions were as follows: 10 cycles of 94°, 1 min; 45°, 1.5 min; 72°, 1.5 min, followed by 20 cycles of 94°, 1 min; 60°, 1.5 min; 72°, 1 min, and ending with a final extension at 72° for 8 min. Southern blot analysis of the PCR products using a *l(3)S12* probe indicated the presence of the predicted product of ~750 bp. Two microliters of the PCR reaction products was diluted into 100 μ l and subjected to a second round of PCR amplification using the primers 16: 5'-ATTTCGCGACCGTGTGAGCAC-3', located in the 5' UTR of *psq* from 100-89 nucleotides 5' to the predicted start codon of the gene, and 17 (at the 3' end of *l(3)S12*, see above). PCR thermocycler conditions were 30 cycles of 94°, 30 sec; 63°, 45 sec; 72°, 45 sec, followed by a 5-min final extension at 72°. Reaction products were subjected to preparative agarose gel electrophoresis and the PCR product of ~630 bp in length was isolated. The purified DNA fragment was digested with the restriction enzymes *EcoRI* (located within the *psq* sequence, 47 bp 5' to the translation start codon) and *SalI* [engineered into primer 17, at a position 24 bp beyond the putative stop codon of the *l(3)S12* gene (DUTTON and CHOVNICK 1991)] and cloned into the Bluescript KS⁻ vector (Stratagene) digested with the enzymes *EcoRI* and *SalI*. The inserts from two independent clones were sequenced in their entirety by the dideoxy chain termination

method using reagents supplied by US Biochemical. Both clones were of identical sequence.

RESULTS

***P* element alleles:** Four noncomplementing female-sterile alleles that map to chromosomal position 47A10–12 were obtained from a mutagenesis screen using the *P*[*lacZ, rosy*⁺] (*PZ*) element [see KARPEN and SPRADLING (1992); referred to as *zeppelin* in SPRADLING (1993)]. These insertions affect a gene, *pipsqueak*, that is required in oogenesis (SIEGEL *et al.* 1993). Mutant alleles display a range of phenotypes, including early degeneration of egg chambers, dorsalization of eggshells, and deletion of posterior segments in embryos. Three of the lines (2403, 8109, and 0115) contain insertions of a single *P* element at sites clustered within a 1-kb region of the genomic DNA (Figure 2A). All three elements are oriented in the same direction with respect to the genomic DNA. The 8109 insertion contains an altered *P* element in which the *rosy* portion of the element is duplicated and present in tandem array (H. LIN, personal communication). The fourth *PZ* line, 0482, contains at least two *P* elements. This line fails to complement all *pipsqueak* alleles and also fails to complement any of a group of nine other *PZ* lines belonging to a distinct and neighboring lethal complementation group. The *P* element that affects *psq* in this line is \approx 35 kb distal to the 2403, 8109, and 0115 *PZ* insertion sites.

An additional *psq* allele, *fsI*, consists of a *P*[*lacZ, white*⁺] element inserted \approx 40 kb distal to the triplet of *PZ* insertion sites.

***Pipsqueak* transcripts:** Initial attempts to identify a *psq* transcript(s) using genomic DNA flanking the *PZ*2403 insertion site failed to detect any message by Northern blot analysis (data not shown). A 6.7-kb *Hind*III fragment of DNA flanking the *fsI* insertion site, however, did identify a 5.1-kb message in wild-type adult females. This fragment was used as a probe to isolate a partial cDNA for *psq* (see MATERIALS AND METHODS). Using a 300-bp fragment from the 5' end of this cDNA as a probe to Northern blots, a 5.1-kb transcript was again revealed in RNA from wild-type ovaries and adult female flies (Figure 3A). This transcript is reduced in amount in females from all the *PZ* lines, and a new abundant 1.6-kb transcript is apparent (Figure 3A).

Because the four *PZ* insertions produce the same new transcript despite their differing locations in the genomic DNA, it seemed possible that these elements lay within an intron of the *psq* gene and that the *psq* transcript was being spliced aberrantly using sequences located within the *P* element. To test this possibility, we took advantage of several excision lines we had generated that deleted portions of the *P* element and/or flanking DNA. One of these lines, RV34a, deleted most of the *P* element, retaining \sim 1 kb of sequence from the 3' end of the transposon. This remaining sequence

contained DNA upstream of the *rosy* gene as well as all of the 3' *P* element end. Females of this line also produced the shorter 1.6-kb transcript found in the line carrying the entire transposon (data not shown), suggesting that the sequences responsible for aberrant splicing were located at this end of the *P* element.

A literature search revealed the existence of another gene upstream of *rosy*, *l(3)S12*, that encodes a putative ribosomal protein (CLARK and CHOVIK 1986; DUTTON and CHOVIK 1991). Genomic DNA sequence analysis identified a potential ORF for this gene with a proposed intron of 70 nucleotides (DUTTON and CHOVIK 1991). The *Hind*III fragment of *rosy* that was originally used to construct the *PZ* element contains a portion of this putative *l(3)S12* intron, the splice acceptor site, and the remaining coding and transcription termination sequences for the *l(3)S12* gene (see Figure 2A).

To test the possibility that the shorter transcript observed in the *PZ* lines resulted from aberrant splicing of the *psq* transcript into *l(3)S12* sequence, the Northern blot shown in Figure 3A was stripped of *psq* probe and reprobed with a fragment of DNA containing *l(3)S12* sequences (Figure 3B). It is clear that the 1.6-kb message observed in the *PZ* lines with a *psq* 5' UTR probe (Figure 3A) also contains *l(3)S12*-hybridizing sequence. An additional less abundant RNA \sim 0.9 kb in length is also apparent. This species may represent truncation of a message initiated at an alternative site or a splicing intermediate of the 1.6 kb *PZ* message. Also apparent with this probe is the endogenous *l(3)S12* message of 0.7 kb in length (which serves as a control for quantitating amounts of RNA in each sample). Reprobing the blot with 3' *P* element sequences did not reveal any transcripts (data not shown) as expected for the splicing event proposed. In summary, the "truncated" message we observe in the *PZ* lines appears to result from normal transcription initiation from the *psq* promoter, followed by aberrant splicing and termination of transcription using *l(3)S12* signals (see Figure 2B).

Nucleotide sequence of the truncated *PZ* RNA: We used the method of RT-PCR to clone out a copy of the truncated transcript so that the point of fusion between the *psq* and *l(3)S12* genes could be directly determined by sequence analysis. mRNA isolated from *cn psq*^{PZ2403}; *ry*⁵⁰⁶ female flies was reverse transcribed and subjected to two rounds of PCR amplification, using partially nested primers (see MATERIALS AND METHODS). The positions of primers used for amplification and restriction sites used for cloning are indicated relative to a map of the truncated transcript presented in Figure 4A. The nucleotide sequence and conceptual translation of the cloned region of the truncated message is presented in Figure 4B.

The nucleotide sequence of the PCR product demonstrates that *psq* sequences have indeed been joined to those of *l(3)S12* by an aberrant splicing event. This aberrant splicing event uses a splice donor from *psq* that

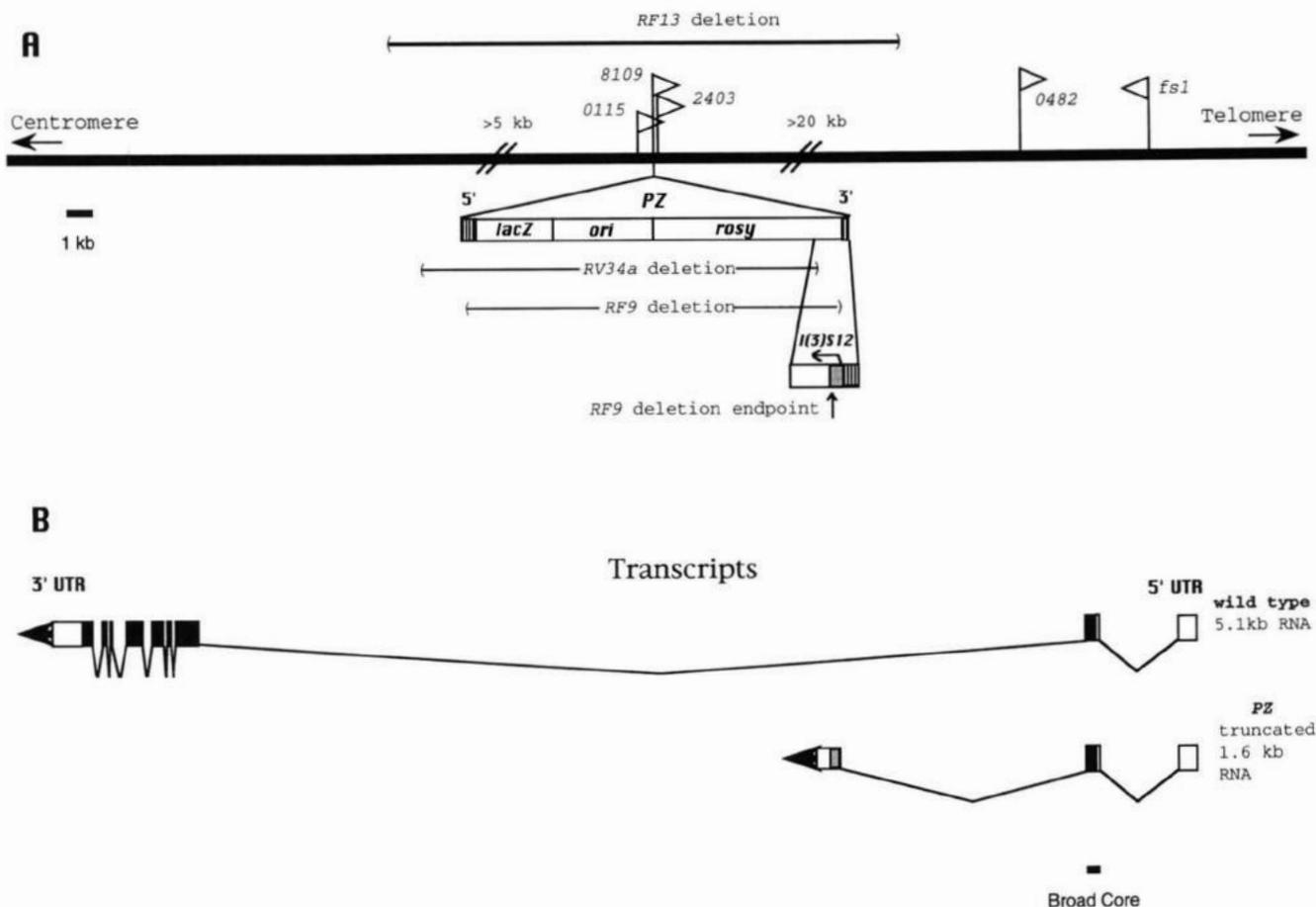


FIGURE 2.—Genomic map of the *psq* gene. (A) P element insertion sites for the various alleles are designated by flags. Below the map is shown the structure of the PZ element. Transcription of the *rosy* and *l(3)S12* genes is from right to left. The extent of the deletions of P element and genomic DNA in the excision lines, RF13, RV34a, and RF9, are indicated. (B) Structure of transcripts observed in wild-type and PZ lines is shown. Open boxes indicate untranslated sequence; shaded boxes indicate translated sequence. The *Broad Core*-homologous region of *psq* is indicated below the transcripts. The region of conceptual translation of *l(3)S12* (DUTTON and CHOVIK 1991) is indicated by stippled shading.

normally functions to splice out the large intron of *psq* when joined to the acceptor sequence many kilobases downstream in the *psq* pre-mRNA. In the PZ mutants, however, this donor sequence is joined to a sequence in *l(3)S12* previously identified as a probable splice acceptor on the basis of computer analysis of the genomic DNA sequence for that gene (DUTTON and CHOVIK 1991). Our analysis shows that the assignment of an intron and splice acceptor to this region of *l(3)S12* is indeed correct. The position of fusion between *psq* and *l(3)S12* sequences is indicated by an arrow in the sequence presented in Figure 4B and in the sequencing gel shown in Figure 4C.

The truncated message in PZ mutants encodes a fusion protein: We have previously identified the ATG indicated in Figure 4B as the most likely candidate for the start codon for translation of the 5.1-kb *psq* mRNA (H. HOROWITZ and C. BERG, unpublished results) based on homology to other proteins (see below) and on the finding that multiple stop codons occur in all three frames in the sequence preceding this codon. Concep-

tual translation of the truncated (mutant) mRNA indicates that a fusion protein containing amino acids from both genes should be produced (Figure 4, B and C). The predicted fusion protein contains 105 amino acids derived from the N-terminus of *psq* fused to 63 amino acids from *l(3)S12*.¹ The region of *psq* contained in the fusion protein bears homology to several other *Drosophila* proteins, including those encoded by the *Broad Complex* (DiBELLO *et al.* 1991), *tramtrack* (HARRISON and TRAVERS 1990; READ and MANLEY 1992), *kelch* (XUE and

¹ This structure is based upon the RT-PCR sequence data. Because the intron/exon structure for *l(3)S12* has not been determined, it is possible that the *l(3)S12* primer (17) used in the PCR reaction might actually reside in an intron for that gene. If so, no PCR product would be generated except by aberrant means (*e.g.*, by using the cDNA from an incompletely processed PZ pre-mRNA as a template). The production of such a species would not affect our assignment of the site of fusion between *psq* and *l(3)S12* but would preclude assignment of the C-terminus of the hybrid protein. The observation that both codon usage and conservation of sequence is poor in the region downstream of the putative *l(3)S12* termination codon (RILEY 1989; DUTTON and CHOVIK 1991) suggests, however, that the PCR product we obtained is likely to be *bona fide*.

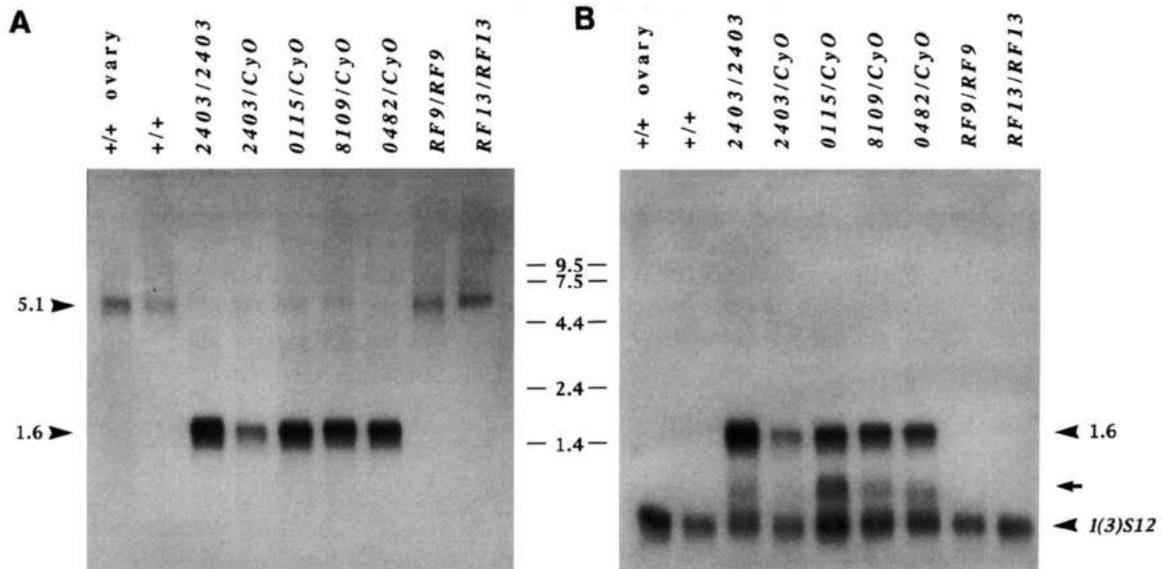


FIGURE 3.—*PZ* mutations produce a truncated transcript. (A) Northern blot probed with a *psq*-specific probe from the 5' UTR. RNA samples are from adult females unless otherwise indicated. Genotypes of all lines (in a *cn, ry⁵⁰⁶* background) are indicated. The 5.1-kb wild-type *psq* and 1.6-kb truncated *PZ* transcripts are indicated. Positions of migration of RNA size markers are also shown. (B) The same blot as in A after being stripped of *psq* probe and reprobed with *l(3)SI2* gene sequence. The 1.6-kb truncated *psq* transcript and the endogenous *l(3)SI2* transcript are indicated. An additional *l(3)SI2*-hybridizing transcript (see text) is indicated with a tailed arrow.

COOLEY 1993), and *bric-a-brac* (F. LASKI, personal communication) genes.

The mutant phenotypes of the *PZ* insertions are due to an aberrant splicing event: The experiments described above document the production of a truncated *psq* message in the *PZ* lines. We could not be certain, however, that the truncation was, in fact, the true cause of the mutant phenotypes observed in these lines. It was possible that these *P* elements were affecting another as yet unidentified transcript and that the reduction of the 5.1-kb message and concomitant production of the truncated message were irrelevant to the phenotypes observed.

To examine this possibility, we characterized the structure of fertile excision lines generated from the *PZ*2403 allele. In general, it is expected that the majority of fertile revertants one might expect to recover would represent precise excisions of the *P* element, restoring the genomic DNA to its previous unmutated state. We reasoned, however, that if the mutant effect of the *PZ* lines were due solely to an aberrant splicing event, then any alteration of the DNA interfering with this splicing event would restore fertility. Upon molecular characterization of seven fertile revertants of 2403, we found that five lines had undergone precise excision of the *P* element. Two lines, however, had different and informative structures.

In the revertant line RF9, the majority of the *P* element had been deleted, but a small portion near the 3' end remained (see Figure 2A). Southern blot analysis placed the right-hand deletion endpoint between the *Pst*I and *Nru*I sites within the *l(3)SI2* sequence (see Figure 4B).

Accordingly, the putative *l(3)SI2* splice acceptor site was retained in this line, but a portion of the predicted coding sequence and the 3' end of the message had been deleted.

Analysis of the RF13 revertant showed that a large deletion had occurred that removed not only all of the *P* element but ≥ 40 kb of flanking genomic DNA as well (see Figure 2A). The finding that female flies from this line are viable as homozygotes and are fertile in spite of the presence of the large deletion indicates that, in fact, the original *P* element insertion was not affecting any other essential transcripts in this intron and that the mutant phenotype is best explained by the splicing event that occurs in these *PZ* lines. If the alterations that restore fertility in these two excision lines prevent the aberrant splicing event, then the wild-type 5.1-kb *psq* transcript should be restored. Northern blots of RNA obtained from these lines reveal that normal levels of the 5.1-kb message are restored and none of the truncated transcripts are produced (Figure 3A).

The finding that normal levels of the 5.1-kb *psq* message are produced in RF9, despite the presence of the *l(3)SI2* 3' splice acceptor site, may be explained by the exon definition model proposed by ROBBERTSON *et al.* (1990) to explain splicing of vertebrate messages. In this model, an internal exon must be defined by the presence of appropriate 3' and 5' splice sites at its ends for it to be incorporated into a spliced message. Mutation of the 5' splice site of an exon precludes proper exon definition, leading either to activation of normally silent cryptic sites or to inhibition of splicing of the upstream intron

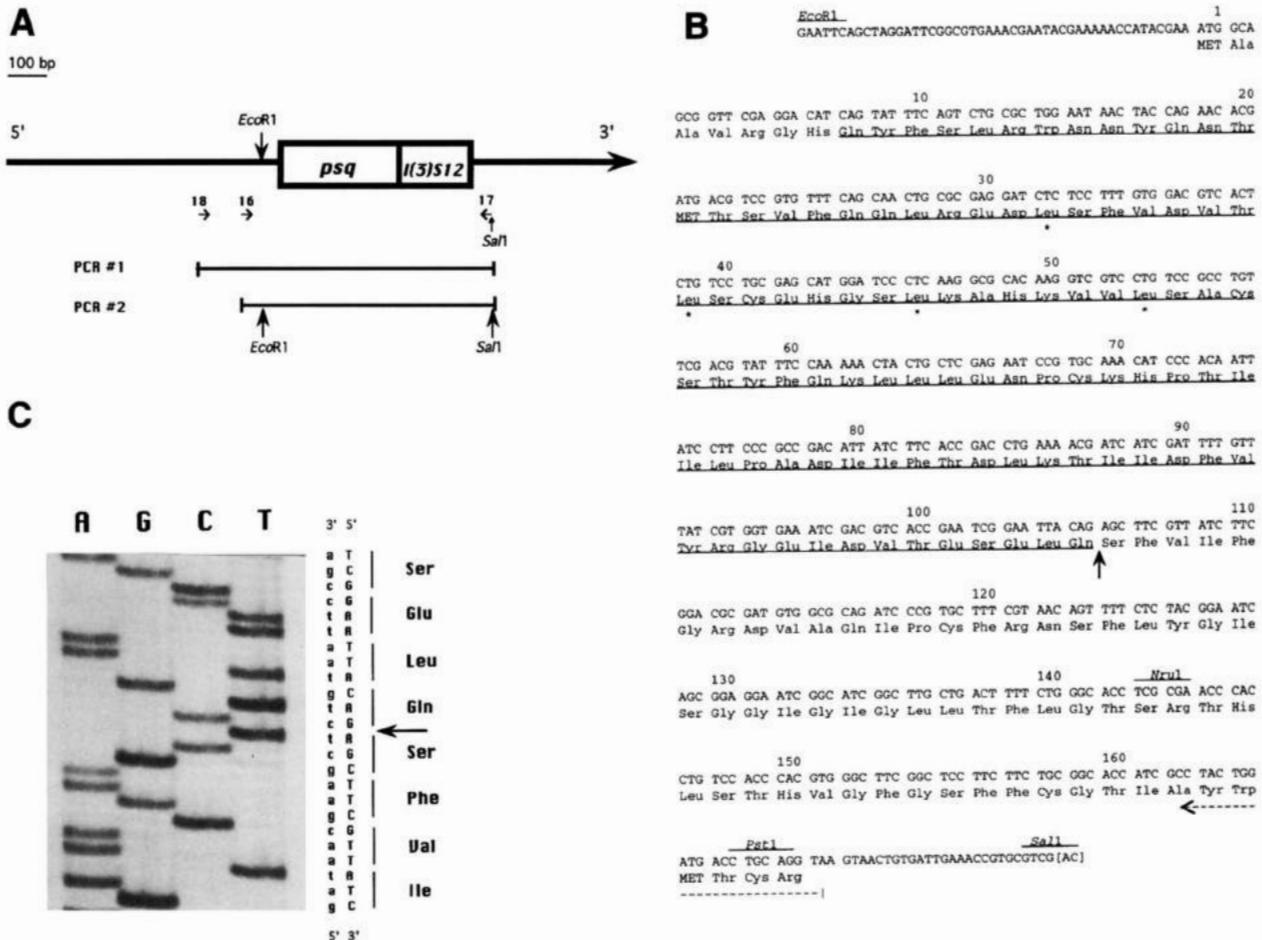


FIGURE 4.—Nucleotide and amino acid sequence of the truncated *PZ* transcript. (A) Schematic of the truncated mRNA generated in mutant *PZ* lines of *psq*. Positions of PCR primers 18, 16, and 17 are indicated by small arrows beneath the map. Restriction sites used for cloning the RT-PCR product are indicated. Open boxes indicate translated sequence; solid lines indicate UTRs. PCR products obtained using primers 18 and 17 (e.g., PCR #1) were reamplified with primers 16 and 17 to produce the PCR product indicated as PCR #2. This fragment was then cloned using the *EcoRI* and *SalI* restriction sites indicated. (B) Nucleotide and (conceptual) amino acid sequence of the hybrid transcript from 2403/2403 female flies. The *EcoRI* and *SalI* sites used for cloning the RT-PCR product are indicated. (Brackets surround bases in the *SalI* site derived from the PCR primer; these are not present in the truncated transcript *in vivo*.) Additional restriction sites referred to in the text are indicated. The region of homology of *psq* to the Broad Complex and other proteins (see text) is underlined. The potential leucine zipper in this region is indicated by asterisks beneath the leucine residues. The position of the fusion between *psq* and *l(3)S12* sequences is indicated by a vertical arrow. The dashed arrow indicates the position and orientation of the primer used in sequencing for the gel presented in C. (C) Sequencing gel showing the fusion of *psq* and *l(3)S12* sequences. The cloned RT-PCR product was sequenced using a primer located within *l(3)S12* sequence (dashed arrow shown in B). Sequencing progresses up the gel (lowercase letters) in a 3' to 5' direction relative to the message. Reading down the gel, uppercase letters indicate the sequence of the truncated message in the 5' to 3' direction, with the corresponding amino acids indicated. The arrow indicates the position of fusion of *psq* and *l(3)S12* sequences.

(exon skipping) despite the presence of wild-type 5' and 3' splice sites flanking that intron. This observation suggested that exon sequences downstream of an intron affect recognition and splicing of that intron (ROBBERTSON *et al.* 1990). For the last exon of a message, the presence of the poly(A) site (rather than a 5' splice site) is presumed to help define the exon (NIWA *et al.* 1990; NIWA and BERGET 1991). Indeed, mutation of the AAUAAA polyadenylation consensus sequence has been shown to inhibit splicing *in vitro* of an upstream intron (NIWA and BERGET 1991).

In RF9, a deletion removes the 3' end of the *PZ*-

residing *l(3)S12* gene; previous work (DUTTON and CHOVIK 1991) has established that genomic sequences ≥ 165 bp downstream of the RF9 deletion endpoint that we mapped are present in the *l(3)S12* message. If this deletion removes a 5' splice site flanking the first *l(3)S12* exon fused to *psq* in the *PZ* lines, it would be predicted that the *l(3)S12* 3' splice acceptor would not be used and normal splicing of the *psq* message would occur. Alternatively, if the *l(3)S12* sequences spliced onto *psq* message in *PZ* mutants comprise the single terminal exon and if the RF9 deletion removes the *l(3)S12* polyadenylation signal, exon definition would

again be precluded and the normal *psq* message would be produced.

DISCUSSION

We have demonstrated a novel mechanism for P element mutagenesis by the *P*[*lacZ*, *rosy*⁺] element. Our findings show that when this element inserts into the intron of a gene in the correct orientation, it can cause a mutant phenotype by inducing aberrant splicing and termination events that truncate the wild-type message.

Examples of aberrant splicing and termination of transcripts induced by transposable elements have been observed in two other organisms. In maize, the insertion of the *Mu1* transposable element into the first intron of the *Adh1* gene alters processing of the pre-mRNA of that gene (ORTIZ and STROMMER 1990). In mouse, the *lpr* mutation results from an insertion of an early transposable element (ETn) in the second intron of the *Fas* antigen gene (ADACHI *et al.* 1993). In both of these examples, hybrid transcripts composed of sequence derived from the affected gene and the transposable element are produced, and premature termination of the message results from signals present in the transposable element. Unlike the process we observe for the *PZ* truncations, however, splicing does not occur into the element; intronic sequences directly upstream of the element are incorporated into the hybrid transcript either by "read-through" from the preceding exon (in *Adh1*) or by use of a cryptic splice site acceptor in the intron (in *lpr*). Furthermore, in neither of these two examples is it expected that the hybrid transcript will produce a fusion protein. Stop codons in the intron sequences included in the hybrid transcripts preclude this possibility.

In mammalian systems, the challenge of identifying genes with interesting expression patterns has led to the development of vectors originally termed "gene traps" (GOSSLER *et al.* 1989). These vectors are designed to generate spliced fusion transcripts between a reporter gene and the endogenous gene present at the site of integration (BRENNER *et al.* 1989; GOSSLER *et al.* 1989; KERR *et al.* 1989; FRIEDRICH and SORIANO 1991). A splice acceptor is strategically placed in front of a promoterless reporter gene (often *lacZ*) such that integrations of the vector into an intron of a gene in the correct orientation should create *lacZ* fusion transcripts. If the reading frames of the endogenous gene and *lacZ* are the same, an active β -galactosidase fusion protein should be produced. Consequently, the process of aberrant splicing and termination, with potential formation of a fusion protein that we observe for the *PZ* element in *Drosophila*, is highly analogous to the gene trap designed for use in mammalian systems. Like our finding with the *psq* *PZ* alleles, recent work has demonstrated that the aberrant splicing into a gene trap vector can be very efficient, resulting in drastic reductions in

the amount of endogenous message and the generation of a mutant phenotype (SKARNES *et al.* 1992).

The mutant phenotype resulting from a *PZ*-induced splicing and termination event could result entirely from reduction in the amount of full-length wild-type message. However, depending on the location of the intron in which the *PZ* element resides relative to the coding sequences of the affected gene, additional modes of mutagenesis can be envisioned as well. In some situations (as in the case of the *PZ* insertions into the *psq* gene), a fusion protein between coding sequences of the gene of interest and the putative ribosomal protein I(3)S12 is predicted to occur. If the N-terminus of the fusion protein functions in DNA binding, in interactions with other proteins, or in signal transduction, the observed result might be a dominant or semidominant negative effect on the normal gene function. Experiments examining the effect of truncated Notch proteins (engineered *in vitro*) have shown that overexpression of the N-terminal extracellular portion of the Notch receptor protein resulted in dominant negative wing, eye, and bristle phenotypes in the presence of a wild-type Notch background (REBAY *et al.* 1993). An analogous situation could occur if the *PZ* element inserted into an intron of any gene such that it separated the N-terminal receptor domain of a protein from the C-terminal effector.

In the example of the *PZ* insertions into *psq*, conceptual translation of the truncated message produces a protein in which the N-terminal portion of the *psq* gene is fused in frame to I(3)S12. This region of *psq* bears homology to the *Broad Core*, a domain identified in a number of *Drosophila* transcription factors (*Broad Complex* proteins and *tramtrack*) as well as in a protein with a presumptive structural role (*kelch*). Included in the *Broad Core* domain of *psq* is a potential leucine zipper (see Figure 4B) (H. HOROWITZ and C. BERG, unpublished results). The predicted fusion protein thus contains a domain that may be involved in mediating protein-protein interactions between pipsqueak and itself or heterologous proteins. Consequently, the mutant effect of the truncation of *psq* message may reflect a disruption of normal protein-protein interactions in which pipsqueak is involved and/or a disruption of the interactions of I(3)S12 with other ribosomal proteins, as well as the reduction of wild-type full-length pipsqueak protein.

The possibility that the fusion protein itself may have some deleterious effect is suggested by the observation that the *PZ* allele, 8109, displays more severe phenotypes as a homozygote than when present over a deficiency for the region (SIEGEL *et al.* 1993). This finding suggests that the presence of truncated transcript is more deleterious than no transcript at all. We have not observed any obvious dominant negative effect of the *psq* *PZ* insertions in heterozygous flies. It is possible, however, that subtle effects have gone undetected. Al-

ternatively, it might be necessary for the *PZ* alleles to be in a homozygous state to provide sufficient truncated fusion protein to abrogate the effects of any wild-type gene product that is present. Although such a scenario does not conform to the classical definition of a dominant negative interaction, the effect is essentially the same: mutant protein interferes with the activity of wild-type protein, with the added requirement that the ratio of the two species falls within a certain critical range. Indeed, evidence for a potentially similar type of dominant negative effect was observed in the studies of truncated Notch protein (REBAY *et al.* 1993) described above. It was found that when a truncated portion of the *Notch* gene containing the extracellular domain was introduced into wild-type flies, it resulted in a mutant phenotype *only* when the truncated construct was homozygous. We are currently conducting experiments to more fully examine this possibility for the *psq* *PZ* mutants.

The enhanced mutagenic power gained by using a *P* element that can cause disruption even when it inserts into an intron is underscored by the finding that of the 8500 *P* element lines created in the KARPEN and SPRADLING (1992) mutagenesis screen, the four that mutated the *psq* gene resulted from *P* element insertion into an intron. We are aware of at least one other example of *PZ*-induced truncation of a wild-type message (K. RITTENHOUSE and C. BERG, unpublished data).

Finally, an additional and powerful technical advantage of *P* element-induced truncations should be noted. In some cases (as in *psq*), *P* elements insert into very large introns, such that cosmid walking to obtain probes for use in a cDNA screen is very tedious. In such a situation, the 5' end of the gene can be obtained by performing 5' RACE cloning (FROHMAN *et al.* 1988) using reverse transcriptase primed with an oligonucleotide derived from the *l(3)S12* sequence. This technique, which greatly simplifies the isolation of the gene, has been used to isolate genes identified in gene trap experiments in mouse (SKARNES *et al.* 1992).

The combination of these features, broader targets for mutagenesis, facilitated cloning, and generation of unique and potentially negative alleles for analysis of gene function, provide a compelling argument for the inclusion of splicing/termination signals in the design of *P* elements used for mutagenesis.

We thank Dr. HAIFAN LIN for providing clones and maps of the DNA flanking the *PZ* insertions 8109 and 2403, Dr. VIVIAN SIEGEL for providing the *P*[*lacZ, w'*] line *fs1* and DNA flanking the *P* element, and Dr. TULLE HAZELRIGG for providing the ovary cDNA library. We also are indebted to Dr. BOB BRAUN for his critical reading of this manuscript and to JOY SABL for her participation in the excision analysis of the *PZ*2403 line. This work was supported by National Institutes of Health grant GM45248 and March of Dimes grants 5-FY91-0486, 5-FY92-0081, and 1-FY93-0973.

LITERATURE CITED

ADACHI, M., R. WATANABE-FUKUNAGA and S. NAGATA, 1993 Aberrant transcription caused by insertion of an early transposable ele-

- ment in an intron of the Fas antigen gene of *lpr* mice. *Proc. Natl. Acad. Sci. USA* **90**: 1756-1760.
- BELLEN, H. J., C. J. O'KANE, C. WILSON, U. GROSSNIKLAUS, R. K. PEARSON *et al.*, 1989 *P* element-mediated enhancer detection: a versatile method to study development in *Drosophila*. *Genes Dev.* **3**: 1288-1300.
- BIER, E., H. VAESSIN, S. SHEPHERD, K. LEE, K. MCCALL *et al.*, 1989 Searching for pattern and mutation in the *Drosophila* genome with a *P-lacZ* vector. *Genes Dev.* **3**: 1273-1287.
- BRENNER, D. G., S. LIN-CHAO and S. N. COHEN, 1989 Analysis of mammalian cell genetic regulation *in situ* by using retrovirus-derived "portable exons" carrying the *Escherichia coli lacZ* gene. *Proc. Natl. Acad. Sci. USA* **86**: 5517-5521.
- CHURCH, G. M., and W. GILBERT, 1984 Genomic sequencing. *Proc. Natl. Acad. Sci. USA* **81**: 1991-1995.
- CLARK, S.H., and A. CUOVNICK, 1986 Studies of normal and position-affected expression of *rosy* region genes in *Drosophila melanogaster*. *Genetics* **114**: 819-840.
- COOLEY, L., C. BERG and A. SPRADLING, 1988a Controlling *P* element insertional mutagenesis. *Trends Genet.* **4**: 253-258.
- COOLEY, L., R. KELLEY and A. SPRADLING, 1988b Insertional mutagenesis of the *Drosophila* genome with single *P* elements. *Science* **239**: 1121-1128.
- DIBELLO, P. R., D. A. WITHERS, C. A. BAYER, J. W. FRISTROM and G. M. GUILD, 1991 The *Drosophila* Broad-Complex encodes a family of related proteins containing zinc fingers. *Genetics* **129**: 385-397.
- DUTTON, F. L., and A. CHOVIK, 1991 The *l(3)S12* locus of *Drosophila melanogaster*: heterochromatic position effects and stage-specific misexpression of the gene in *P* element transposons. *Genetics* **128**: 103-118.
- ENGELS, W. R., 1979 Extrachromosomal control of mutability in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **76**: 4011-4015.
- ENGELS, W. R., 1989 *P* elements in *Drosophila melanogaster*, pp.437-484 in *Mobile DNA Elements*, edited by D. E. BERG and M. M. HOWE. American Society for Microbiology, Washington, DC.
- FEINBERG, A. P., and B. VOGELSTEIN, 1983 A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132**: 6-13.
- FEINBERG, A. P., and B. VOGELSTEIN, 1984 A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Addendum. *Anal. Biochem.* **137**: 266-267.
- FRIEDRICH, G., and P. SORIANO, 1991 Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes Dev.* **5**: 1513-1523.
- FROHMAN, M. A., M. K. DUSH and G. R. MARTIN, 1988 Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. USA* **85**: 8998-9002.
- GOSSLER, A., A. L. JOYNER, J. ROSSANT and W. C. SKARNES, 1989 Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes. *Nature* **244**: 463-465.
- HARRISON, S. D., and A. A. TRAVERS, 1990 The *tramtrack* gene encodes a *Drosophila* finger protein that interacts with the *ftz* transcriptional regulatory region and shows a novel embryonic expression pattern. *EMBO J.* **9**: 207-216.
- JOWETT, T., 1986 Preparation of nucleic acids, pp. 275-286 in *Drosophila: A Practical Approach*, edited by D. B. ROBERTS. IRL Press, Oxford.
- KARPEN, G. H., and A. C. SPRADLING, 1992 Analysis of subtelomeric heterochromatin in the *Drosophila* mini-chromosome *Dp1187* by single *P* element insertional mutagenesis. *Genetics* **132**: 737-753.
- KELLEY, M. R., S. KIDD, R. L. BERG and M. W. YOUNG, 1987 Restriction of *P* element insertions at the *Notch* locus of *Drosophila melanogaster*. *Mol. Cell. Biol.* **7**: 1545-1548.
- KERR, W. G., G. P. NOLAN, A. T. SERAFINI and L. A. HERZENBERG, 1989 Transcriptionally defective retroviruses containing *lacZ* for the *in situ* detection of endogenous genes and developmentally regulated chromatin. *Cold Spring Harbor Symp. Quant. Biol.* **54**: 767-776.
- LEE, C. S., D. CURTIS, M. MCCARRON, C. LOVE, M. GRAY *et al.*, 1987 Mutations affecting expression of the *rosy* locus in *Drosophila melanogaster*. *Genetics* **116**: 55-66.
- MLODZIK, M., and Y. HIROMI, 1992 The enhancer trap method in

- Drosophila*: its application to neurobiology. *Methods Neurosci.* **9**: 397-414.
- NIWA, M., and S. M. BERGET, 1991 Mutation of the AAUAAA polyadenylation signal depresses *in vitro* splicing of proximal but not distal introns. *Genes Dev.* **5**: 2086-2095.
- NIWA, M., S. D. ROSE and S. M. BERGET, 1990 *In vitro* polyadenylation is stimulated by the presence of an upstream intron. *Genes Dev.* **4**: 1552-1559.
- O'KANE, C. J., and W. J. GEHRING, 1987 Detection *in situ* of genomic regulatory elements in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **84**: 9123-9127.
- ORTIZ, D. F., and J. N. STROMMER, 1990 The *Mu1* maize transposable element induces tissue-specific aberrant splicing and polyadenylation in two *Adh1* mutants. *Mol. Cell. Biol.* **10**: 2090-2095.
- READ, D., and J. L. MANLEY, 1992 Alternatively spliced transcripts of the *Drosophila tramtrack* gene encode zinc finger proteins with distinct DNA binding specificities. *EMBO J.* **11**: 1035-1044.
- REBAY, L., R. G. FEIGON and S. ARTAVANIS-TSAKONAS, 1993 Specific truncations of *Drosophila* Notch define dominant activated and dominant negative forms of the receptor. *Cell* **74**: 319-329.
- REED, K. C., and D. A. MANN, 1985 Rapid transfer of DNA from agarose gels to nylon membranes. *Nucleic Acids Res.* **20**: 7207-7221.
- RILEY, M. A., 1989 Nicotinic sequence of the Xdh region in *Drosophila pseudoobscura* and an analysis of the evolution of synonymous codons. *Mol. Biol. Evol.* **6**: 33-52.
- ROBBERTSON, B. L., G. J. COTE and S. M. BERGET, 1990 Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol. Cell. Biol.* **10**: 84-94.
- ROBERTSON, H. M., C. R. PRESTON, R. W. PHILLIS, D. M. JOHNSON-SCHLITZ, W. K. BENZ *et al.*, 1988 A stable genomic source of P element transposase in *Drosophila melanogaster*. *Genetics* **118**: 461-470.
- RUBIN, G. M., and A. C. SPRADLING, 1983 Vectors for P element-mediated gene transfer in *Drosophila*. *Nucleic Acids Res.* **11**: 6341-6351.
- SAMBROOK, J., E. F. FRITSCH and T. MANIATIS, 1989 *Molecular Cloning, A Laboratory Manual*, Ed. 2. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- SIEGEL, V., T. A. JONGENS, L. Y. JAN and Y. N. JAN, 1993 *pipsqueak*, an early acting member of the posterior group of genes, affects vasa level and germ cell-somatic cell interaction in the developing egg chamber. *Development* **119**: 1187-1202.
- SKARNES, W. C., B. A. AUERBACH and A. L. JOYNER, 1992 A gene trap approach in mouse embryonic stem cells: the *lacZ* reporter is activated by splicing, reflects endogenous gene expression and is mutagenic in mice. *Genes Dev.* **6**: 903-918.
- SOUTHERN, E. M., 1975 Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**: 503-517.
- SPRADLING, A. C., 1993 Developmental genetics of oogenesis, pp. 1-70 in *The Development of Drosophila melanogaster*, edited by M. BATE. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- TSUBOTA, S., M. ASHBURNER and P. SCHEDL, 1985 P element induced control mutations at the r gene of *Drosophila melanogaster*. *Mol. Cell. Biol.* **5**: 2567-2574.
- WILSON, C., R. K. PEARSON, H. J. BELLEN, C. J. O'KANE, U. GROSS-NIKLAUS *et al.*, 1989 P element-mediated enhancer detection: an efficient method for isolating and characterizing developmentally regulated genes in *Drosophila*. *Genes Dev.* **3**: 1301-1313.
- XUE, F., and L. COOLEY, 1993 *kelch* encodes a component of intercellular bridges in *Drosophila* egg chambers. *Cell* **72**: 681-693.

Communicating editor: A. CHOVNICK