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.

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 HindIII 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 insertion 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 CARPEN and SPRADLING (1993); referred to as zeppelin in SPRADLING (1993)]. All four lines contain a P element [described in MLODZIK and HIROMI (1992)] inserted in an isogenized cn chromosome in a ry506 background. Excision lines were created from the 2403 allele by reintroducing the 42–3 transposase and screening for loss of the ry+ eye-color marker (see Figure 1). Specifically, cn+ pipsqueak+/CyO; ry506 males were mated to females from Sb P[A2-3; 99B/TM6] germline stocks. Sons carrying both transposons (cn+ pipsqueak+/CyO; ry506 Sb P[Δ2-3, ry+]) were screened for loss of the PZ element. These truncated RNAs could yield proteins with dominant mutant effects.
and VOCEISTEIN were founded containing the excision chromosome using the random hexamer primer technique of FEINBERG individually to flanking the as in LEE digested with the appropriate restriction endonuclease (results). Flies as described (JOWETT described here. This sequence is located within the HAIZEL-RIGG) and was generously provided by VIVIAN SIEGEL.

prising all of the 3' prime region of the gene (H. HOROWITZ and C. BERG, unpublished results).

13 175 pS 3'UTR, Flies as described (JOWETT 1986). Poly(A) + RNA was reverse transcribed according to the manufacturer's instructions in a total volume of 50 µl using random hexamers (2 pmol/µl) as primers and 200 units of M-MLV Reverse Transcriptase (GIBCO BRL). Five microtubes of the reverse-transcription reaction was then used as a template for a PCR (100 µl total volume) using buffers provided by the manufacturer of the Taq polymerase (Boehringer Mannheim). The primers used (see Figure 4A) were: 18: 5'-GCAAGTCGACGCACGGTTTCAA-3' derived from the 5' UTR of psg [221–230 nucleotides upstream from the predicted start of translation (HOROWITZ and BERG, in preparation)] and 17: 5'-TACAGTCGACGCAGGTTTCC-3', derived from the putative 5' UTR of the l(3)S12 gene (DUTTON and CHOVNICK 1991). This primer extends from 2955 to 2980 of the published ros gene sequence (numbering as in LEE et al. 1987; DUTTON and CHOVNICK 1991) and has a Sall site (GTCGAC) engineered in by the inclusion of the six 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; 45°, 1 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 microtubes of the PCR reaction products was diluted into 100 µl and subjected to a second round of PCR amplification using the primers 16: 5'-ATTTCGCGACGTCCTGACGAG-3', located in the 5' UTR of psg 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 psg sequence, 47 bp 5' to the translation start codon) and Sall (engineered into primer 17, at a position 24 bp beyond the putative start codon of the l(3)S12 gene (DUTTON and CHOVNICK 1991) and ligated into the plasmid KS+ vector (Stratagene) digested with the enzymes EcoRI and Sall. The inserts from two independent clones were sequenced in their entirety by the dideoxy chain termination 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 isolated using the Quick Oligo-tEX mRNA kit. Approximately 0.5 µg of poly(A) + mRNA was loaded per lane on a 1% agarose-formaldehyde gel in 1X MOPS buffer [50 µM MOPS (pH 7.0), 8 µM sodium acetate, 1 mM EDTA) (SAMBOUK 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 X SSC overnight. RNA on filters was subjected to UV cross-linking using a Stratagene Stratagene and hybridized as described above.

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

**PCR reaction analysis:** The truncated transcript produced in homoeogamous 2403 adult female flies was analyzed as follows: 1 µg of poly(A) + mRNA was reverse transcribed according to the manufacturer's instructions in a total volume of 50 µl using random hexamers (2 pmol/µl) as primers and 200 units of M-MLV Reverse Transcriptase (GIBCO BRL). Five microtubes of the reverse-transcription reaction was then used as a template for a PCR (100 µl total volume) using buffers provided by the manufacturer of the Taq polymerase (Boehringer Mannheim). The primers used (see Figure 4A) were: 18: 5'-GCAAGTCGACGCACGGTTTCAA-3' derived from the 5' UTR of psg [221–230 nucleotides upstream from the predicted start of translation (HOROWITZ and BERG, in preparation)] and 17: 5'-TACAGTCGACGCAGGTTTCC-3', derived from the putative 5' UTR of the l(3)S12 gene (DUTTON and CHOVNICK 1991). This primer extends from 2955 to 2980 of the published ros gene sequence (numbering as in LEE et al. 1987; DUTTON and CHOVNICK 1991) and has a Sall site (GTCGAC) engineered in by the inclusion of the six 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; 45°, 1 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 microtubes of the PCR reaction products was diluted into 100 µl and subjected to a second round of PCR amplification using the primers 16: 5'-ATTTCGCGACGTCCTGACGAG-3', located in the 5' UTR of psg 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 psg sequence, 47 bp 5' to the translation start codon) and Sall (engineered into primer 17, at a position 24 bp beyond the putative start codon of the l(3)S12 gene (DUTTON and CHOVNICK 1991) and ligated into the plasmid KS+ vector (Stratagene) digested with the enzymes EcoRI and Sall. The inserts from two independent clones were sequenced in their entirety by the dideoxy chain termination method of CHURCH and GILBERT (1984).
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 Karpfen 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 effects \( psq \) in this line is \( \geq 35 \) kb distal to the 2403, 8109, and 0115 \( PZ \) insertion sites.

An additional \( psq \) allele, \( fs1 \), consists of a \( P[lacZ, white^+ \] \) element inserted \( \geq 40 \) 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 HindIII fragment of DNA flanking the \( fs1 \) 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 Chovnick 1986; Dutton and Chovnick 1991). Genomic DNA sequence analysis identified a potential ORF for this gene with a proposed intron of 70 nucleotides (Dutton and Chovnick 1991). The HindIII 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 \( psq \) 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 \( PM \) 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 ligation between the \( psq \) and \( l(3)S12 \) genes could be directly determined by sequence analysis. mRNA isolated from \( cn psq^{259}; ob \) 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 rosa 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 Chovnick 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 Chovnick 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. Conceptual 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. 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), tramtrak (Harrison and Travers 1990; Read and Manley 1992), hunch (Xue and

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 1988; Dutton and Chovnick 1991) suggests, however, that the PCR product we obtained is likely to be bona fide.
Aberrant Splicing and P Elements

Figure 3.—PZ mutations produce a truncated transcript. (A) Northern blot probed with a psg-specific probe from the 5' UTR. RNA samples are from adult females unless otherwise indicated. Genotypes of all lines (in a cn, ry<sup>16</sup> background) are indicated. The 5.1-kb wild-type psg 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 psg probe and reprobed with l(3) Sl2 gene sequence. The 1.6-kb truncated psg transcript and the endogenous l(3) Sl2 transcript are indicated. An additional l(3) Sl2-hybridizing transcript (see text) is indicated with a tailed arrow.

Cooley 1998), and bica-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 psg 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 PZ2403 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 PstI and NruI sites within the l(3) Sl2 sequence (see Figure 4B). Accordingly, the putative l(3) Sl2 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 psg 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 psg message are produced in RF9, despite the presence of the l(3) Sl2 3' splice acceptor site, may be explained by the exon definition model proposed by Robberson 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
(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 (Roberson et al. 1990). For the last exon of a message, the presence of the polya site (rather than a 5' splice site) is presumed to help define the exon (Niwa et al. 1990; Niwa and Berg 1991). Indeed, mutation of the AAUAAA polyadenylation consensus sequence has been shown to inhibit splicing in vitro of an upstream intron (Niwa and Berg 1991).

In RF9, a deletion removes the 3' end of the PZ-residing l(3)S12 gene; previous work (Dutton and Chovnick 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 psg 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 psg message would occur. Alternatively, if the l(3)S12 sequences spliced onto psg in the PZ mutants comprise the single terminal exon and if the RF9 deletion removes the l(3)S12 polyadenylation signal, exon definition would

---

**Figure 4.**—Nucleotide and amino acid sequence of the truncated PZ transcript. (A) Schematic of the truncated mRNA generated in mutant PZ lines of psg. 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 psg 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 psg and l(3)S12 sequences is indicated by a vertical arrow. The dashed arrow indicates the position and orientation of the primer used for sequencing the gel presented in C. (C) Sequencing gel showing the fusion of psg 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 psg and l(3)S12 sequences.
again be precluded and the normal \textit{psq} message would be produced.

**DISCUSSION**

We have demonstrated a novel mechanism for \textit{P} element mutagenesis by the \textit{P} [\textit{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 \textit{MuI} transposable element into the first intron of the \textit{Adhi} gene alters processing of the pre-

mRNA of that gene (Ortiz and Strommer 1990). In mouse, the \textit{lpr} mutation results from an insertion of an early transposable element (\textit{En}) in the second intron of the \textit{Fas} antigen gene (Aadachi et al. 1998). 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 \textit{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 \textit{Adhi}) or by use of a cryptic splice site acceptor in the intron (in \textit{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. 1986; Gossler et al. 1989; Keur et al. 1989; Friedrich and Soriano 1991). A splice acceptor is strategically placed in front of a promoterless reporter gene (often \textit{lacZ}) such that integrations of the vector into an intron of a gene in the correct orientation should create \textit{lacZ} fusion transcripts. If the reading frames of the endogenous gene and \textit{lacZ} are the same, an active \textit{\beta}-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 \textit{PZ} element in Drosophila, is highly analogous to the gene trap designed for use in mammalian systems. Like our finding with the \textit{psq} \textit{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 \textit{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 \textit{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 \textit{PZ} insertions into the \textit{psq} gene), a fusion protein between coding sequences of the gene of interest and the putative ribosomal protein \textit{1(3)S12} is predicted to occur. If the \textit{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 \textit{in vitro}) have shown that overexpression of the \textit{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 (Rabbay et al. 1995). An analogous situation could occur if the \textit{PZ} element inserted into an intron of any gene such that it separated the \textit{N}-terminal receptor domain of a protein from the \textit{C}-terminal effector.

In the example of the \textit{PZ} insertions into \textit{psq}, conceptual translation of the truncated message produces a protein in which the \textit{N}-terminal portion of the \textit{psq} gene is fused in frame to \textit{1(3)S12}. This region of \textit{psq} bears homology to the \textit{Broad Core}, a domain identified in a number of Drosophila transcription factors (Broad Complex proteins and \textit{tramtrack}) as well as in a protein with a presumptive structural role (\textit{helch}). Included in the \textit{Broad Core} domain of \textit{psq} is a potential leucine zipper (see Figure 4B) (H. Horkowitz 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 \textit{psq} message may reflect a disruption of normal protein–protein interactions in which pipsqueak is involved and/or a disruption of the interactions of \textit{1(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 that the \textit{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 \textit{psq} \textit{PZ} insertions in heterozygous flies. It is possible, however, that subtle effects have gone undetected. Af-
very large introns, such that cosmid walking to obtain gene function, provide a compelling argument for the mutated genes that have been used to isolate genes identified in gene trap libraries. We also are indebted to Dr. Bob Spalding (1992) mutagenesis screen, the four that introduced into an intron is underscored by the finding that the target of this manuscript and to Joy Sarl for her participation in the excision analysis of the P2'403 line. This work was supported by National Institutes of Health grant GM47548 and March of Dimes grants SF91-0486, SF92-0031, and 1F495-0073.

LITERATURE CITED


Harrison, S. D., and A. A. Travers, 1990 The tramtrack gene encodes a Drosophila finger protein that interacts with the fmd transcriptional regulatory region and shows a novel embryonic expression pattern. EMBO J. 9: 207–216.


McGee, M., and Y. Hirota, 1992 The enhancer trap method in...
Aberrant Splicing and P Elements


Communicating editor: A. Citovick