MOLECULAR AND CELLULAR BIOLOGY, Aug. 1994, p. 5392–5401
0270-7306/94/$04.00+0

Vol. 14, No. 8

The *Drosophila* tom Retrotransposon Encodes an Envelope Protein

SOICHI TANDA,† JOSÉ L. MULLOR, AND VICTOR G. CORCES*
Department of Biology, The Johns Hopkins University, Baltimore, Maryland 21218

Received 7 February 1994/Returned for modification 24 March 1994/Accepted 21 April 1994

The tom transposable element of *Drosophila ananassae* is mobilized with high frequency in the germ line of females from the *ca; px* strain, and its insertion results in mutations that show almost exclusively dominant eye phenotypes. tom is a long terminal repeat-containing retrotransposon that encodes three different open reading frames (ORFs). It is expressed in the nurse cells during oogenesis, in the central and peripheral nervous systems during embryonic development, and in the imaginal discs of the larva. tom RNA accumulates in the germarium of ovaries from *ca; px* females but not in the parental inactive strain, suggesting that this altered pattern of tom expression might be the cause of the high rate of mobilization of this retrotransposon. The specificity of tom-induced eye phenotypes can be explained by the presence of regulatory sequences responsible for expression of tom in the eye imaginal discs of third-instar larvae. These sequences might cause overexpression of adjacent genes affected by tom-induced mutations, resulting in the death of undifferentiated cells located anterior to the morphogenetic furrow. In addition to the full-length RNA, tom is also transcribed into a spliced subgenomic transcript that encodes a protein resulting from the fusion between the amino-terminal region of the first (*gag*) and the third ORFs. The protein encoded by this RNA shows structural characteristics such as a signal peptide, glycosylation sites, endopeptidase cleavage site, and fusion peptide that are typical of the envelope proteins of retroviruses. Antibodies against tom ORF3 recognize two different proteins present in female ovaries, suggesting that tom might be able to form infective viral particles that could play a role in the horizontal transmission of this retrotransposon.

Transposable elements constitute about 10% of the *Drosophila* genome, where they form the moderately repetitive class of DNA sequences (6). The structure of particular transposable elements plays an essential role in determining both the mechanisms controlling their mobilization and the basis for their mutagenic effects on adjacent genes. Retrotransposons encode reverse transcriptase involved in the synthesis of DNA intermediates that serve as substrates during the integration process (for reviews, see references 4 and 33). Most retrotransposons move infrequently and unpredictably in the genome of the host, although several cases in which long terminal repeat (LTR)-containing retrotransposons were mobilized at a high frequency in specific *Drosophila* strains have been described (7, 15, 17). Nevertheless, many of these stocks become inactive after a few generations. One exception is that of the tom element of *Drosophila ananassae*.

The tom transposable element is responsible for a syndrome of genetic instability in a specific strain of *D. ananassae* carrying the mutations *claret* (*ca*) and *plexus* (*px*). Mobilization of tom in this strain causes a high incidence of mutations that affect almost exclusively eye morphogenesis (9). These mutations show dominant nonpleiotropic *Optic morphology* (*Om*) phenotypes and arise exclusively in oocytes of *ca; px* females and map to at least 25 different euchromatic loci (9, 10). *Om* alleles have been shown to be associated with the insertion of the tom element at or in close proximity to the cytogenetic location of the mutation (26, 29, 30). The specificity in the type of effects caused by the tom transposable element constitutes an interesting paradigm to study both eye development and the mechanisms by which transposable elements cause mutations. Two simple mechanisms can be put forward to account for the almost exclusive effect on eye morphogenesis that results from the insertion of the tom retrotransposon. One possibility is that the location of de novo insertions of the tom element is highly selective, such that this transposon moves only into regulatory regions that control the expression of a family of genes involved in eye development. A second alternative is that the tom element inserts randomly and the specificity of its mutagenic action could be determined by selective effects of sequences present in the element on genes located nearby. For example, transcriptional regulatory sequences, such as tissue-specific transcriptional enhancers, present in tom could increase the expression of adjacent genes in certain tissues such as the eye imaginal discs, thus accounting for both the dominant characteristics and the specificity of the phenotype. If insertion of the tom element is random, a mutant phenotype would be observed only in those cases in which overexpression of the adjacent gene affects proper differentiation of the cells in the eye imaginal disc. In support of this latter possibility, tom-induced mutations in the *Om(1D)* locus of *D. ananassae* cause a sevenfold accumulation of *Om(1D)* transcripts in the eye imaginal disc of mutant with respect to that in wild-type flies (29).

Partial sequence analysis of the tom element has revealed homologies to the protease and reverse transcriptase activities of retroviruses and the presence of two direct 475-bp flanking repeats, suggesting that tom belongs to the family of LTR-containing retrotransposons (31). Like the proviral form of vertebrate retroviruses, retrotransposons contain signals necessary for the initiation and termination of transcription in their LTRs and adjacent tRNA primer binding sites and purine-rich sequences necessary for the initiation of DNA synthesis in retroviral systems (4). In addition, the genome of a typical retrovirus encodes three different genes termed *gag*,

* Corresponding author. Mailing address: Department of Biology, The Johns Hopkins University, 34th and Charles Streets, Baltimore, MD 21218. Phone: (410) 516-8749. Fax: (410) 516-5456.
† Present address: Department of Zoology, University of Maryland, College Park, MD 20742.

pol, and *env* that are required for viral replication and infectivity (reviewed in reference 33). The products of the *gag* and *pol* genes are translated from a full-length RNA that initiates in the 5′ LTR and terminates in the 3′ LTR. The *gag* region encodes a protein that is cleaved to give rise to several small polypeptides found in the core of the virus particle. The *pol* gene is expressed as a *gag-pol* polyprotein which is the precursor of the mature form of reverse transcriptase. The amino-terminal region of this protein product contains the DNA polymerase and RNase H activities of reverse transcriptase as well as a *gag*-specific protease required for the cleavage of the *gag* polyprotein, whereas the DNA endonuclease (integrase) activity required for provirus integration is located in the carboxy-terminal region. The third open reading frame (ORF) located at the 3′ end of the retroviral genome encodes the components of the viral envelope. This ORF3 is encoded by a subgenomic spliced transcript, and the protein is cleaved by a cellular endopeptidase to give rise to the surface protein or glycoprotein and the transmembrane protein of infective retroviruses (33). All *Drosophila* retrotransposons contain ORFs homologous to *gag* and *pol*, but only a few elements such as gypsy, 297, and 17.6 contain an additional ORF located in a position equivalent to the *env* gene of retroviruses (4). Because the envelope proteins of retroviruses are poorly conserved, the functional significance of the ORF3 of retrotransposons and whether it actually encodes envelope proteins are not currently known.

In this paper we report the complete sequence of tom, showing that this element indeed belongs to the class of LTR-containing retrotransposons. As is the case for the gypsy, 297, and 17.6 elements of *Drosophila melanogaster*, tom encodes two different ORFs that show homology to the *gag* and *pol* genes of retroviruses, but it also contains a third ORF located in a position analogous to the retroviral *env* gene. Analysis of the pattern of tom expression indicates a correlation between tom mobilization and the presence of high transcript levels in the germarium of female ovaries during oogenesis. tom is also expressed in the eye imaginal disc, where it causes cell death by affecting the expression of *Om* genes in the undifferentiated cells located anterior to the morphogenetic furrow. In addition to the full-length transcript, tom encodes an *envelope*-specific RNA that is translated into two proteins detectable in the ovaries of females. These results suggest that the tom retrotransposon has conserved the same strategies as retroviruses in the production of *env*-specific transcripts and therefore may be able to form infectious viruses that could play a role in its mobilization and transmission.

MATERIALS AND METHODS

RNA isolation and Northern (RNA) analysis. RNA was extracted from about 1 g of flies by the SDS-phenol technique (27). Samples were homogenized in 10 ml of 10 mM Tris-HCl (pH 7.4)–100 mM NaCl–1 mM EDTA–0.5% sodium dodecyl sulfate (SDS), and the homogenate was extracted several times with phenol-chloroform and then chloroform extracted. Total RNAs were recovered by ethanol precipitation and resuspended in 1 ml of water. Poly(A)$^+$ RNA was then isolated by chromatography on oligo(dT)-cellulose, separated on 1.2% formaldehyde (2.2 M)–agarose gels (7 μg of RNA per lane) in MOPS buffer (20 mM Na-morpholinepropanesulfonic acid [MOPS; pH 8.5]–5 mM sodium acetate–1 mM EDTA), and transferred to Nytran membranes (Schleicher & Schuell) in 10× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate). After being baked at 80°C for 1 h, filters were incubated with ^{32}P-labeled probes in hybridization solution (5× SSCP [1× SSCP is 0.15 M NaCl, 0.015 M sodium citrate, and 0.02 M sodium phosphate, pH 6.8]–5× Denhardt's solution–50% formamide–1% sarcosyl–100 μg of carrier DNA per ml–10% dextran sulfate) at 42°C overnight. The filters were washed at 50°C in 0.1× SSC–0.5% SDS for 30 min and exposed to X-ray film at −80°C. DNA fragments used as probes to detect tom-encoded transcripts were obtained by digestion of the clones pstomBS and pstom3 with *Bam*HI and *Sst*I for probe A or *Eco*RI and *Pst*I for probe B (31). The probe for the *ras2* transcript was obtained by digestion of clone pUC8-HB-1.2 kb with *Bam*HI and *Hin*dIII (3).

Primer extension and cDNA analysis. A primer complementary to the coding strand of tom between nucleotides 470 and 494 was constructed by using a 381A DNA synthesizer (Applied Biosystems). Synthesized oligonucleotides were purified through SEP-PAK cartridges (Millipore). A 10-μg aliquot of poly(A)$^+$ RNA purified from late pupae of the *Om(1D)9e* stock was hybridized overnight at 23°C to 10 ng of the primer end labeled at the 5′ end with [γ-^{32}P]dATP in 40 mM piperazine-*N*,*N*′-bis(2-ethanesulfonic acid) (PIPES; pH 6.4)–1 mM EDTA–0.4 M NaCl–80% formamide (23). The product was then extended with 40 U of reverse transcriptase (Life Sciences, Ltd.) in 50 mM Tris-HCl (pH 8.3)–75 mM KCl–10 mM $MgCl_2$–5 mM spermidine–10 mM dithiothreitol–1 mM each deoxynucleoside triphosphate (dNTP)–40 U of RNasin (Promega)–4 mM sodium PP_i. Extended single-stranded fragments were separated on a 6% denaturing polyacrylamide gel along with the dideoxy sequence reaction of clone pstom5 (29) as a size marker.

A cDNA library was made from poly(A)$^+$ RNA isolated from adult heads of the *ca; px* stock. The library was screened with probe B, shown in Fig. 1A. Two cDNA clones were isolated from 100,000 plaques screened. Cloned fragments were removed from the lambda gt10 vector by cutting with *Eco*RI and then subcloned into the *Eco*RI site of pUC18. Nucleotide sequences were determined by the method of Sanger (24).

Isolation of cDNAs by reverse transcription-PCR. To isolate cDNA clones of the spliced mRNA for the tom ORF3, a set of primers located upstream (from nucleotide 479 to 500) and downstream (from nucleotide 5236 to 5252) from the putative splice sites were synthesized and purified as described above. A 10-μg aliquot of poly(A)$^+$ RNA purified from pupae of the *Om(1D)9e* stock was hybridized to the downstream primer in hybridization buffer (see above). Ethanol-precipitated RNA-primer hybrids were resuspended in reverse transcription mix and then subjected to reverse transcription. After phenol-chloroform extraction and ethanol precipitation, the reverse-transcribed single-stranded DNAs were amplified 30 times by PCR (22) in 10 mM Tris-HCl (pH 8.3)–50 mM KCl–1.5 mM $MgCl_2$–0.01% gelatin, with 150 mM of each dNTP, 100 ng of the upstream and downstream primers, and 2.5 U of *Taq* polymerase (Perkin-Elmer Cetus) by using a DNA thermal cycler (Perkin-Elmer Cetus). Amplified products were separated by electrophoresis on a 1% agarose gel, isolated by electroelution, and subcloned into the *Sma*I site of pUC18. Plasmids carrying inserts that hybridized to probe B (Fig. 1A) were selected by colony hybridization and sequenced.

In situ hybridization and cell death analysis. Three- to 5-day-old females of the *ca; px* and *ca* strains were dissected in 1× phosphate-buffered saline (PBS) and fixed by the method of McKearin and Spradling (16). Probe B (Fig. 1A) containing the 3′ region of the tom element was labeled with digoxigenin-dATP, and hybridization to whole mount ovaries was carried out by the method of Tautz and Pfeiffle (32). After dehydration, the stained ovaries were mounted in Permount and

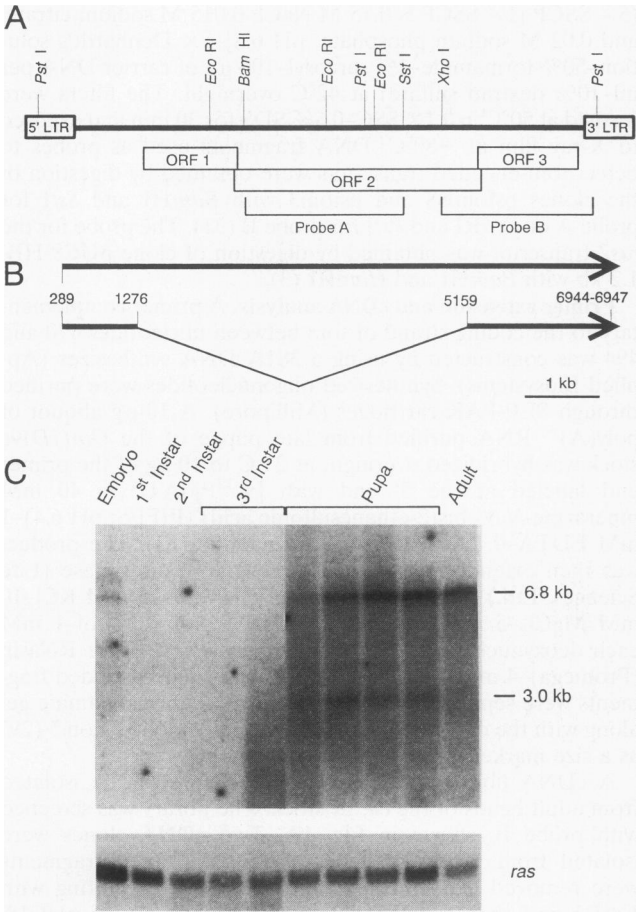

FIG. 1. Structure and developmental expression of the tom element. (A) Organization of the tom element with the locations of the different ORFs. DNA segments used as hybridization probes for Northern analyses are indicated by brackets labeled Probe A and Probe B. Boxes shown below the restriction map indicate the locations of the three ORFs of tom. Boxes at different levels are in different reading frames. ORF1 is in the 0 reading frame, ORF2 is in the −1 frame, and ORF3 is in the 0 frame. (B) Structure of genomic and subgenomic transcripts of the tom element. The schematic drawing shows the unspliced genomic RNA as well as the spliced mRNA. Numbers refer to the nucleotides at the transcription initiation site, the splicing donor and acceptor sites, and the transcription termination site. (C) Northern blot of poly(A)$^+$ RNA isolated from synchronized cultures of the *Om(1D)9e* stock probed with the 3′ end of the tom element (probe B). The sizes of the two major transcripts observed are shown (in kilobases) on the right-hand margin. The lower part of the panel shows the level of *ras2* transcript observed when the same blot was hybridized with the *Drosophila Dras2* gene as a control for the amount of RNA loaded in each lane (21). Stages from which RNA was isolated are indicated on the upper part of the panel. Two third-instar samples were loaded on the gel, one from early- and one from late-third-instar stages. Four lanes contain pupal RNA corresponding to prepupa, early-pupa, mid-pupa, and late-pupa stages of development (left to right, respectively).

observed with Hoffman optics. Probe B was made by PCR amplification of the lambda s9.15 clone (26), using primers located at nucleotides 5310 to 5326 and nucleotides 5589 to 5605. This fragment recognizes the full-length and spliced transcripts of the tom element.

Acridine orange was used to detect cell death in the eye-antenna imaginal discs of late-third-instar larvae. The larvae were dissected in 1× PBS and stained with acridine orange by the method of Spreij (28). Heat-shocked animals were dissected 30 min after incubation at 37°C for 1 h. The discs were observed with a fluorescence microscope immediately after staining.

Antibody production and Western blot (immunoblot) analysis. An 823-bp *Nco*I-*Hae*III fragment from the ORF3 region of the tom element was cloned into the pMAL-c vector (New England Biolabs). The fusion protein was induced in DH5α *Escherichia coli* cells with 3 mM isopropyl-β-D-thiogalactopyranoside. The fusion protein was extracted from an SDS-acrylamide gel after electrophoresis with an Elutrap (Schleicher & Schuell) and injected in rats. An aliquot of 200 μg of the fusion protein was intradermally injected into each rat with complete Freund's adjuvant on day 1. Rats were intradermally injected with 100 μg of the fusion protein with incomplete Freund's adjuvant on day 28. On day 42, an aliquot of 50 μg of the fusion protein was subcutaneously injected with incomplete Freund's adjuvant, and the first serum sample was collected on day 56. The rats were boosted with 50 μg of the fusion protein on day 60. Whole blood was taken on day 84, and the serum was used for the experiments described.

For Western protein analysis, ovaries from two 5- to 10-day-old females were homogenized in 50 μl of homogenization buffer (14). After being boiled for 5 min, the samples were spun for 10 min at 4°C and the supernatants were loaded on a 7.5% acrylamide-SDS gel. After electrophoresis, proteins were electroblotted for 5 h at 300 mA to a nitrocellulose membrane. The membrane was blocked in 2% nonfat milk–1× PBS and then hybridized with primary antibodies (100× dilution) in 1% nonfat milk–1× PBS overnight at 4°C. The membrane was washed in 1× PBS three times at room temperature and treated with secondary alkaline phosphatase-conjugated goat anti-rat antibodies (Cappel; 2,000× dilution) for 1 h at room temperature. The membrane was then washed in 1× PBS three times and stained in 10 mM Tris (pH 9.5)–0.1 M NaCl–50 mM $MgCl_2$–340 μg of nitroblue tetrazolium per ml–175 μg of 5-bromo-4-chloro-3-indolyl phosphate per ml.

Nucleotide sequence accession number. The complete sequence of the tom element has been submitted to the EMBL and GenBank data libraries under accession number Z24451.

RESULTS

Developmental expression of the tom element. A sequence of the LTRs and part of the *pol* gene, including the protease and reverse transcriptase domains, of the tom element has been previously published (31). In order to understand the mechanisms of tom mobilization and mutagenesis, we have determined the complete sequence of this element (data not shown). The tom transposable element of *D. ananassae* is 7,060 bp in length and has three ORFs similar to the *gag*, *pol*, and possibly *env* genes of retroviruses (see diagram in Fig. 1A). The *pol* gene encodes a putative protein with domains homologous to the protease, reverse transcriptase, RNase H, and integrase functions present in other *Drosophila* retrotransposons and vertebrate retroviruses. The order of these different functional domains on the *pol* gene suggests that tom belongs to the gypsy family of LTR-containing retrotransposons, and it is different from members of the Ty-copia family in which the order of *pol* functional domains is protease-integrase-reverse transcriptase-RNase H (4). To gain insights into the basis of high-frequency mobilization and specificity of the mutagenic effect, we examined the pattern of developmental transcription of the tom retrotransposon. Northern blot analysis of poly(A)$^+$ RNA shows that tom encodes a major 6.8-kb RNA and a minor

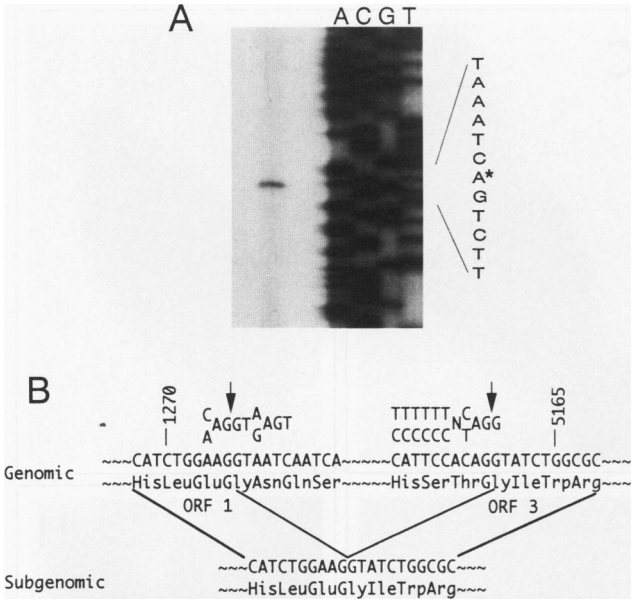

FIG. 2. Transcription initiation, termination, and splice sites of the tom element. (A) The reverse-transcribed product of poly(A)$^+$ RNA primed with a 25-nucleotide extension primer is shown next to dideoxy sequencing reactions from the same primer used as size markers. The nucleotide sequence around the major transcription initiation site is shown to the right. The shaded heptanucleotide is identical to that of the *Drosophila* consensus for transcription initiation by RNA polymerase II (11). Nucleotide A labeled with an asterisk indicates the transcription initiation site of the tom element. (B) Nucleotide sequence from genomic and cDNA clones of the spliced mRNA. Consensus sequences of exon-intron borders of eukaryotic genes are shown above the splicing donor and acceptor sites of the tom element. Vertical arrows in the genomic sequence indicate the proposed splice sites.

3.0-kb RNA whose expression is developmentally regulated (Fig. 1C). While a low level of tom RNA is observed during embryogenesis, tom is poorly transcribed during early fly development. However, the level of tom expression begins to increase at late third instar, and the highest accumulation is observed during pupal development, continuing after eclosion. The number of transcripts and pattern of developmental expression of tom are the same in the active *ca; px* strain, in which tom mobilization takes place, and in inactive stocks. Quantitative differences in the levels of tom-encoded transcripts can be observed between these two types of strains, but the amount of tom RNA is related only to the number of euchromatic copies of tom and seems to play no role in its mobilization (data not shown).

The transcription initiation and termination sites for the 6.8-kb transcript were determined by primer extension and sequence analysis of cDNA clones (Fig. 2A). This transcript begins at nucleotide 289, at the second A in the heptanucleotide ATCAGTC, which matches perfectly the *Drosophila* consensus sequence for RNAs transcribed by RNA polymerase II (11) (Fig. 2A). DNA sequences of two independently isolated cDNA clones indicate that transcription terminates four to seven nucleotides downstream from the proximal polyadenylation signal located at nucleotide 6944. These results indicate that tom LTRs consist of a 288-bp U3 sequence, a 71- to 74-bp R sequence, and a 113- to 116-bp U5 sequence. The 6.8-kb transcript must therefore correspond to the full-length RNA encoded by the tom element, since its size matches well with that predicted from the transcription initiation and termination sites (Fig. 1B).

Tissue-specific expression of the tom element. The pattern of spatial expression of tom was determined by in situ hybridization with a DNA fragment from the 3′ region of the element that recognizes the 6.8- and 3.0-kb transcripts (see Materials and Methods). This pattern was first examined in the inactive *ca* strain from which the *ca; px* active strain arose. tom is expressed in adult females of this strain from very early stages of oogenesis (Fig. 3A). tom RNA accumulates in the nurse cells (Fig. 3B), and it is presumably deposited in the oocyte at later stages, since preblastoderm embryos show a homogeneous pattern of distribution of tom RNA. During embryogenesis, tom is specifically expressed in the central nervous system and the gonads (Fig. 3E). During larval development, tom RNA accumulates at low levels in all imaginal discs but is highly expressed in the eye-antenna imaginal disc; no other larval tissues show tom expression (Fig. 3F). tom is transcribed in cells located on both sides of the morphogenetic furrow, and it is present in the clusters of differentiating photoreceptors posterior to the furrow (Fig. 3G).

Mobilization of the tom element correlates with high levels of expression very early in oogenesis. Northern analysis indicates that the full-length and subgenomic tom-encoded transcripts are expressed in both active and inactive strains, although the levels of these RNAs are higher in the active *ca; px* than in the parental *ca* inactive strain. Other inactive strains accumulate tom RNAs in amounts similar to those in *ca; px* strains (data not shown), suggesting that increased RNA levels are due to the higher number of copies of the tom element present in the genome but are not responsible for tom mobilization. Mobilization of the tom element takes place in the germ line of females of the *ca; px* strain. tom-induced mutants arise from these females not in clusters but as single events, suggesting that tom mobilization takes place during or after meiosis in oogenesis. To examine the molecular basis of tom mobilization in the *ca; px* strain, we compared the expression of this element in ovaries of females of this stock versus the parental inactive *ca* strain. tom is expressed in egg chambers of *ca; px* females (Fig. 3C) in a pattern different from that of females from the parental stock (Fig. 3A). Figures 3C and D are understained with respect to Fig. 3A and B to allow the visualization of particular details in the staining pattern of the germarium. In the inactive *ca* strain, tom is expressed uniformly throughout the germarium, with the exception of region 1, where accumulation of tom RNA is not detectable (Fig. 3A). This region contains stem cells and cystoblasts. To the contrary, the active *ca; px* strain accumulates high levels of tom RNA in this region of the germarium, as well as in region 2b, which contains individual cysts completely surrounded by follicle cells (Fig. 3C). The active strain also shows accumulation of tom RNA in the oocyte nucleus (Fig. 3D), whereas tom expression is not detected in the nucleus of the inactive strain, even in egg chambers that have been overstained (Fig. 3B). The differential pattern of expression of the tom element between the active *ca; px* strain and the inactive parental strain might be responsible for the mobilization of this retrotransposon.

tom-induced phenotypes are caused by increased cell lethality in the region anterior to the morphogenetic furrow. The pattern of tom expression in the eye disc is interesting in view of the specificity in the type of phenotypes induced by tom insertions which seem to affect almost exclusively eye morphogenesis (9). Insertion of the tom element in the 3′ region of the *Om(1D)* gene of *D. ananassae* (which is the homolog of *Bar* in *D. melanogaster*) causes a dominant eye phenotype similar to

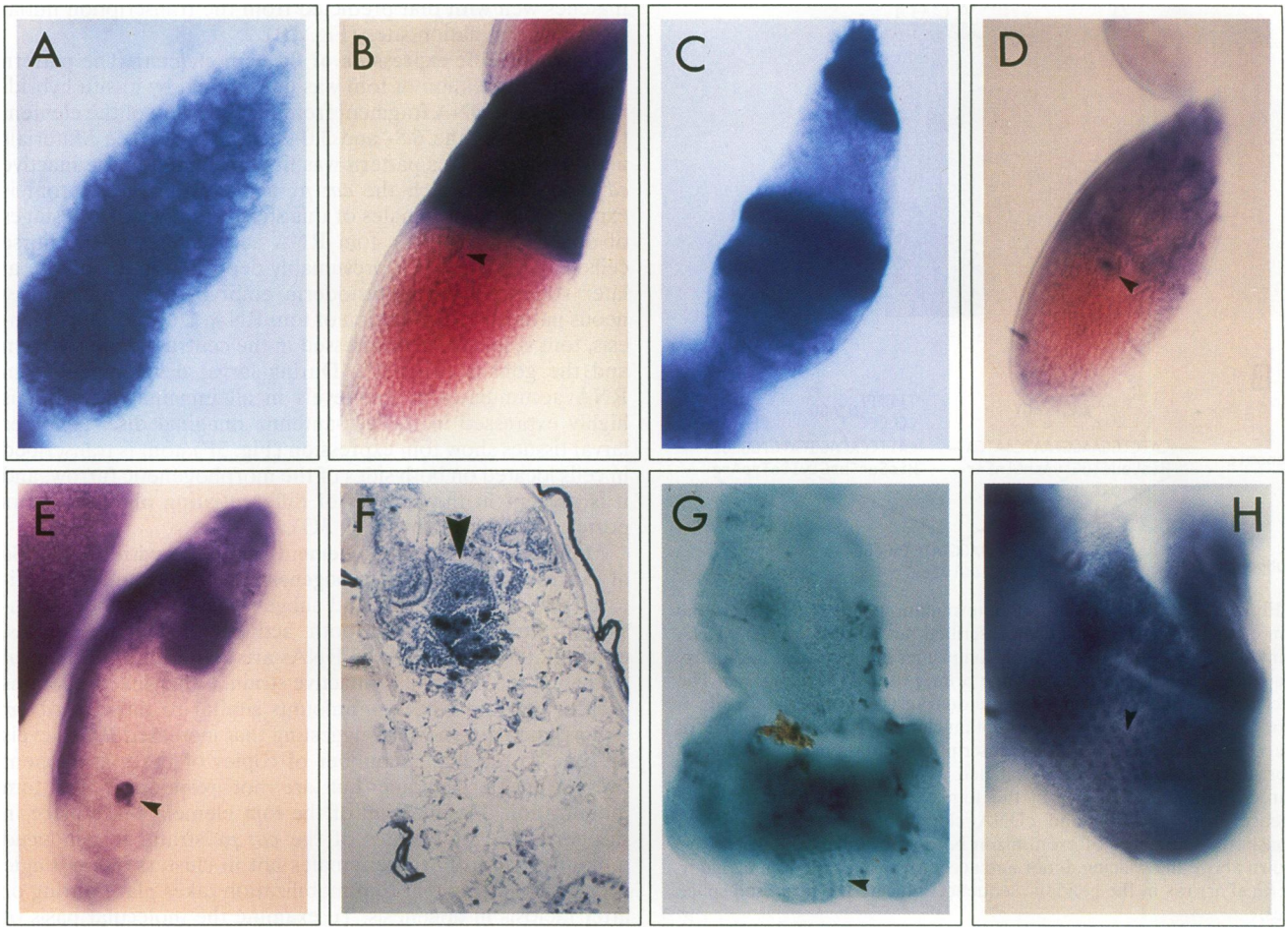

FIG. 3. Tissue-specific expression of the tom element. Probe B (Fig. 1A) containing the 3′ region of the tom element was labeled with digoxigenin-UTP and hybridized to different tissue samples. Anterior is at the top and dorsal is to the right. (A) Germarium from inactive *ca* females. (B) Stage 10 egg chamber from inactive strain females; arrowhead points to the location of the oocyte nucleus. (C) Germarium from females of the active *ca; px* strain. (D) Stage 10 of oogenesis from *ca; px* females; arrowhead points to the oocyte nucleus. (E) Embryo of the *ca; px* strain; arrowhead points to the gonads. (F) Section through a third-instar larva; arrowhead points to the eye imaginal disc. (G) Eye-antenna imaginal disc from a third-instar larva probed with the tom element; clusters of photoreceptors are indicated by arrowhead. (H) Eye-antenna imaginal disc from a third-instar larva probed with sequences encoding the *Om(1D)* gene (29); arrowhead points to clusters of differentiating photoreceptors.

that of *Bar* (8, 13, 29). The *Om(1D)* gene is expressed in the eye imaginal disc with a pattern identical to that of tom. In wild-type flies, *Om(1D)* RNA is present in cells anterior to the morphogenetic furrow and also in the clusters of differentiating photoreceptors (Fig. 3H). In flies carrying a mutation in the *Om(1D)* gene induced by tom insertion, the transcription of *Om(1D)* in the eye imaginal disc as determined by quantitative reverse transcription-PCR is increased sevenfold with respect to that in the wild type (29). In situ hybridization experiments indicate that, within the quantitative limits imposed by this technique, this increased expression of *Om(1D)* is not limited to specific parts of the disc but rather seems to take place on both sides of the morphogenetic furrow (data not shown). These results suggest that the tom-induced eye phenotype might be a consequence of the overexpression of the *Om(1D)* gene throughout the eye imaginal disc. This overexpression could be due to the effect of regulatory sequences present in the adjacent tom element, since tom is transcribed in the same cells of the imaginal disc as the *Om(1D)* gene.

Most tom-induced mutations are characterized by a decreased number of ommatidia, suggesting that increased cell death during eye morphogenesis might be responsible for the mutant defect (9). To test this assumption, we have analyzed the pattern of cell death in the eye imaginal discs of *Om(1D)* mutants. Figure 4 shows a correlation between cell death observed in the eye discs of third-instar larvae and the appearance of the adult eyes that develop from each disc. Cell death can be seen in wild-type eye discs in a scattered pattern on both sides of the morphogenetic furrow (Fig. 4) (35). In discs from *Om(1D)* mutants, the amount of cell death increases dramatically in a stripe of cells running parallel and anterior to the morphogenetic furrow (Fig. 4). The resulting adult eye phenotype is presumably a consequence of the increased cell death in this region of the eye disc. The confinement of tom-induced cell death to a stripe anterior to the morphogenetic furrow could be due to a restriction of *Om(1D)* overexpression to this region or to a special sensitivity of the cells located in this region to high levels of the *Om(1D)* protein. To discern between these two possibilities, we examined the pattern of cell death in the eye imaginal discs of flies transformed with a DNA

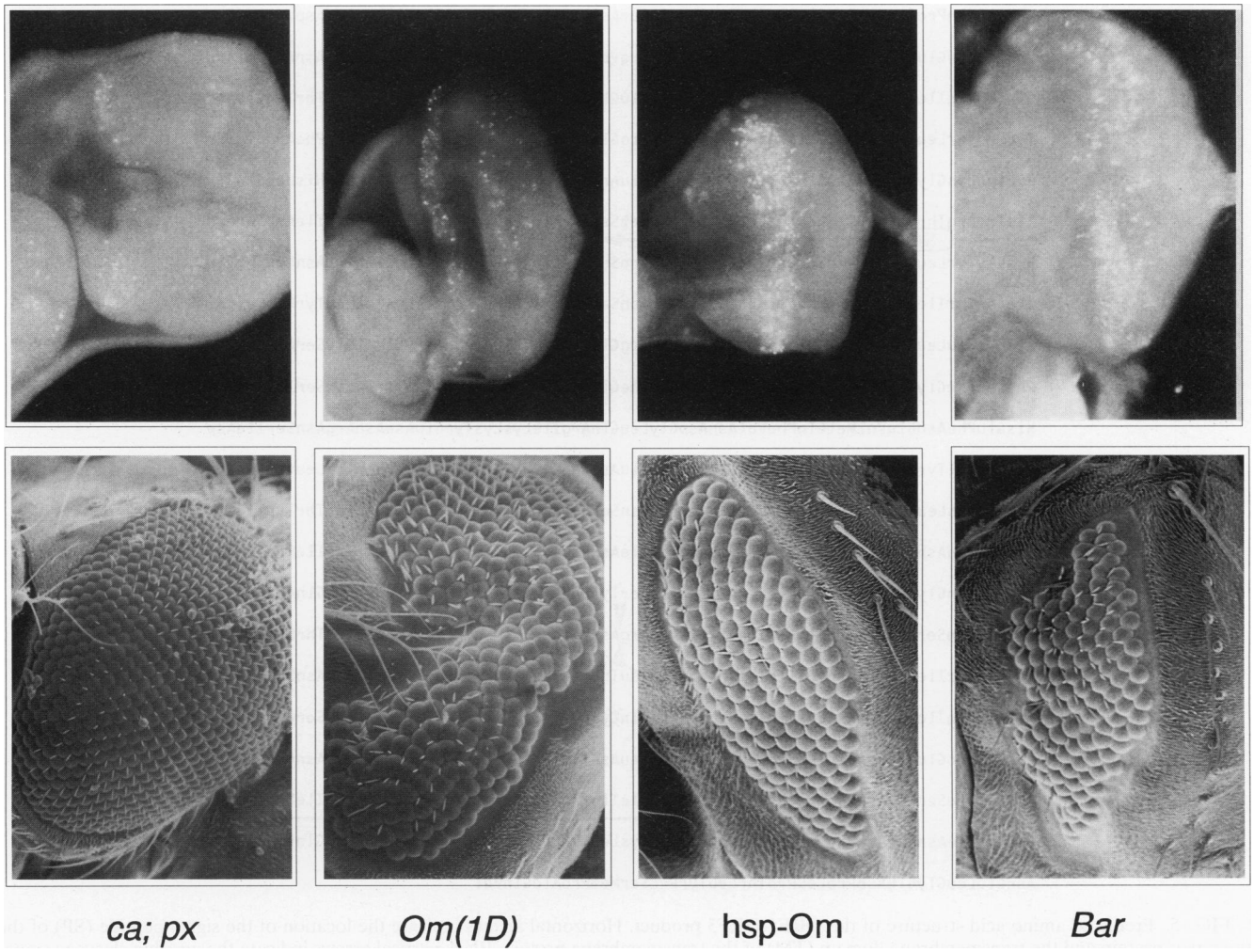

FIG. 4. Cell death in the eye imaginal disc induced by tom insertion. Dorsal is at the top and anterior is to the left. The upper row of photographs shows eye-antenna imaginal discs from third-instar larvae of the strains indicated below stained with acridine orange to visualize cell death. The lower row shows scanning electron micrographs of adult eyes from the corresponding imaginal discs.

fragment containing the coding region of the *Om(1D)* gene under the control of a heat shock promoter (29). Induction of *Om(1D)* gene expression by high temperature causes the uniform expression of the *Om(1D)* protein throughout the eye disc (2, 5), but cell death is induced only in the region located immediately anterior to the morphogenetic furrow (Fig. 4). This result suggests that only the undifferentiated cells in this region of the disc are sensitive to high levels of *Om(1D)* protein.

The tom element encodes a spliced subgenomic transcript. Most *Drosophila* LTR-containing retrotransposons encode a full-length RNA that extends between the two flanking LTRs. Many of these elements also encode smaller RNAs whose structures have not been determined, except in the case of copia (18). In addition to the full-length transcript, copia encodes a *gag*-specific RNA in which most of the *pol* region, with the exception of the protease domain, is removed by splicing. In the case of the tom element, two different RNAs, 6.8 and 3.0 kb in length, were observed when the 3′ region of the element (probe B in Fig. 1A) was used as hybridization probe (Fig. 1C), whereas only the 6.8-kb transcript was present when the central region of the element was utilized (probe A in Fig. 1A). The 3.0-kb transcript hybridizes to sequences containing the third ORF but not the central region of the element, suggesting that this transcript may correspond to a spliced mRNA for the ORF3 gene product. To determine the structure of the small transcript, we isolated cDNA clones by the PCR technique (22). The DNA sequences of three independently isolated clones were identical and showed evidence for a spliced message capable of encoding an ORF1-ORF3 fusion product (Fig. 1B). The pentanucleotide AGGTA was found at the donor splice site in ORF1 and at the acceptor splice site in ORF3 (Fig. 2B). Since 3′ ends of exons end with the dinucleotide AG and 5′ ends of exons begin with the dinucleotide GT in most eukaryotic genes, it is likely that splicing of the 3.0-kb transcript takes place between the two Gs present in the pentanucleotide. If splicing occurs at this step, then the nucleotide sequences of the exon-intron borders agree with the consensus in eukaryotic genes (20) (Fig. 2B). Thus, we conclude that the donor splice site is located at position 1276, 472 bp downstream from the beginning of ORF1, and the acceptor splice site is located at nucleotide 5159, 16 bp downstream from the beginning of ORF3. ORF1 and ORF3 are in frame in the spliced message, suggesting that

FIG. 5. Predicted amino acid structure of the ORF1-ORF3 product. Horizontal arrows indicate the location of the signal peptide (SP) of the precursor protein and the transmembrane domain (TM) of the transmembrane protein. Black vertical arrows indicate the possible cleavage site(s) that eliminates the signal peptide; stippled arrows indicate the cellular endopeptidase cleavage site that cleaves the precursor envelope protein into the surface and transmembrane polypeptides. The Asn residues in putative glycosylation sites conforming to the consensus sequence Asn-Xaa-Ser/Thr are shaded. Cys residues are cross-hatched.

the first Met in ORF1 is used as a translation initiation site to make the ORF3 gene product. The initiation and termination sites of this transcript are the same as those of the full-length RNA, and its structure is diagrammed in Fig. 1B. Its predicted size, not taking into account the poly(A) tail, is 2.8 kb. This size is in good agreement with that observed for the 3.0-kb transcript in the Northern blot shown in Fig. 1C.

The tom element encodes envelope proteins. The ORF1-ORF3 protein encoded by the spliced RNA has a predicted molecular weight of 70,740 and shows several structural characteristics of retroviral envelope proteins (Fig. 5). The putative tom *env* product contains a signal peptide for targeting to the endoplasmic reticulum and secretion to the membrane. Two potential cleavage sites that conform well to the (−3, −1) rule (34) follow the signal peptide and are indicated by arrows in Fig. 5. The predicted size of the processed protein would have a molecular weight of 52,460. The size of the leader peptide eliminated by processing at the signal sequence is 157 amino acids. This peptide is quite large compared with other retroviruses, although a long leader sequence of 62 residues has been previously described for Rous sarcoma virus (12). The tom putative envelope protein also contains a predicted protease cleavage site rich in basic residues that conforms well to the consensus cleavage site of retroviral envelope proteins (Fig. 5). Upon cleavage, the tom protein would give rise to two polypeptides with molecular weights of 32,514 and 19,964 that could correspond to the surface and transmembrane proteins of retroviruses, respectively. The putative transmembrane protein contains a hydrophobic transmembrane domain and hydrophobic amino acids in the amino-terminal end characteristic of the fusion peptide which could mediate penetration of the host cell membrane after receptor binding (Fig. 5) (33). In addition, both proteins contain several putative glycosylation sites for addition of N-linked carbohydrate side chains and Cys residues that could mediate attachment between the surface and transmembrane proteins via disulfide bonds, as is the case in vertebrate retroviruses (Fig. 5). These results point to strong structural homologies between the putative protein products of the ORF3 and retroviral envelope proteins. In addition, the ORF3 products are encoded by an ORF1-ORF3 spliced transcript by a strategy similar to that developed by retroviruses to express the envelope proteins. These two analogies suggest that ORF3 might correspond to the *env* gene of retroviruses.

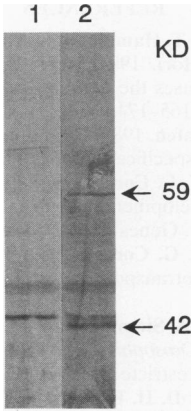

FIG. 6. Western analysis of ORF3-encoded proteins. Polyclonal antibodies were prepared against a maltose-binding protein–ORF3 bacterial fusion protein and were used for Western analysis of protein extracts from ovaries of the *ca; px* strain. Lane 1 was probed with the preimmune serum, and lane 2 was screened with anti-ORF3 antibodies. The sizes of two proteins detected exclusively by the immune serum are indicated on the right.

In order to test whether the putative tom-encoded envelope proteins are actually produced, we prepared antibodies against a bacterial fusion protein containing tom sequences between nucleotides 5244 and 6067. This region of the tom element encodes the complete putative surface protein, with the exception of the first 11 amino acids, and contains only the first six residues of the putative transmembrane protein. Antibodies against the fusion protein would then be expected to recognize the surface protein and the uncleaved precursor. These antibodies were used to carry out Western analysis of proteins present in the ovaries of females from the *ca; px* active strain. Figure 6 shows the result of this experiment. The antibodies recognize two different proteins present in ovaries that do not react with the preimmune serum. The sizes of these two proteins are 59 and 42 kDa, which are in good agreement with the expected sizes of the precursor envelope protein, after cleavage of the signal peptide, and the surface glycoprotein. The slight increase in the observed with respect to the predicted sizes could be due to glycosylation; indeed, several putative glycosylation sites can be observed in the sequence of the tom envelope protein (Fig. 5). These two proteins are also present in ovaries from the inactive *ca* strain (data not shown). These results indicate that tom ORF3 is expressed and the encoded product might be processed in a fashion similar to that for retroviral envelope proteins.

DISCUSSION

Mobilization of LTR-containing retrotransposons in the genomes of eukaryotic organisms is sporadic, suggesting the existence of tight control mechanisms by the cell machinery of the host. This is necessarily so because germ line mobilization of retrotransposons would lead to deleterious effects caused by the accumulation of mutations. Inheritable mutations caused by retrotransposon insertion require the expression of these elements in germ line cells, in order for the full-length RNA to be reverse transcribed into DNA that will serve as a substrate for integration. To understand the mechanisms that control retrotransposon mobilization, we have studied the pattern of germ line and tissue-specific expression of the tom element of *D. ananassae*. This transposon is particularly interesting because it is consistently mobilized in the *ca; px* strain at a frequency of approximately 5×10^{-4}, and its insertion results almost exclusively in mutations that show defects in eye morphogenesis (9).

The tom element is an LTR-containing retrotransposon that is expressed into a full-length RNA that extends between initiation signals located in the 5′ LTR and termination sites located in the 3′ LTR. In addition, tom encodes a subgenomic transcript that arises by splicing of sequences located between the amino-terminal region of ORF1 (*gag*) and the beginning of ORF3. The analysis of tissue-specific tom expression has been confined to those stages significant in the understanding of the mechanisms of tom mobilization and mutagenesis. Both RNAs encoded by tom are expressed in a stereotyped pattern during *Drosophila* development, accumulating in the nurse cells during oogenesis, in the central and peripheral nervous systems during embryogenesis, and in the eye imaginal discs in third-instar larvae. Particularly interesting is the expression of tom in egg chambers, since the mobilization of this element takes place in the female germ line. Clues for the basis of high frequency of tom mobilization in the *ca; px* active strain were obtained by comparing the expression of tom in this and the parental inactive strain. tom mobilization is probably not related simply to levels of RNA; although both tom-encoded transcripts accumulate at approximately threefold-higher levels in the *ca; px* active strain, this is likely due to the presence of three times as many copies of the element in this strain compared with the inactive stock. Nevertheless, tom mobilization might be related to differences between the two strains in the stages of oogenesis when tom transcription takes place. In the active strain, tom is expressed at very high levels in regions 1 and 2b of the germarium as well as in the oocyte nucleus. This particular pattern of tom expression in the *ca; px* active strain could be due to the presence of a copy of tom inserted near regulatory sequences that drive tom transcription in cells of regions 1 and 2b of the germarium. Alternatively, a mutation present in the active strain could be responsible for this pattern of tissue-specific expression. In either case, since tom is expressed at low levels in these particular cells of the germarium in the inactive strain, this pattern of tom transcription in *ca; px* females might be responsible for its mobilization.

tom is also transcribed in the eye imaginal discs, and this expression might be responsible for the specificity in the phenotypes induced by insertion of this retrotransposon. These phenotypes are dominant and usually result in a decrease in the number of ommatidia present in the adult eye. For example, tom-induced mutations in the *Om(1D)* gene of *D. ananassae* result in a phenotype very similar to *Bar* in *D. melanogaster* (13, 29). *Om(1D)* is the homolog of the *D. melanogaster Bar* gene, and it encodes a homeobox protein expressed in photoreceptor cells during eye development (8). Insertion of the tom element in different *Om(1D)* alleles takes place in the adjacent 3′ region, causing overexpression of the *Om(1D)* gene in the eye disc (29, 30). This overexpression has no effect on the differentiated photoreceptor cells located posterior to the morphogenetic furrow but causes cell death in the undifferentiated cells in the anterior region of the disc. tom-induced *Om* mutations in other genes also seem to result from overexpression of the affected gene in the eye imaginal disc. For example, the *Om(1A)* mutation results from tom insertion into the *cut* locus of *D. ananassae*, causing ectopic expression of this gene in the eye imaginal disc (1). In addition, the mutant phenotype of *Om(2D)* is caused by high levels of expression in the eye imaginal disc of a protein containing His Pro repeats; this protein is normally present only during embryonic development and its accumulation in eye discs in

the *Om(2D)* mutant is due to the presence of the adjacent tom element (36). These results suggest that the specificity in the eye phenotypes caused by tom might not be a result of tom insertion into genes exclusively involved in eye development. tom might insert randomly in the genome, but a phenotype is observed only when overexpression of the adjacent gene has an effect on eye morphogenesis. This model would require that tom insertion has a preference for noncoding regions in order to explain the seldomly observed recessive phenotypes caused by tom mobilization.

In addition to the full-length RNA, tom encodes a subgenomic transcript that seems to be expressed with the same tissue specificity. This transcript encodes an ORF1-ORF3 fusion protein which might offer some clues to the relationship between retrotransposons and retroviruses. Members of the *Drosophila* Ty-copia family of LTR-containing retrotransposons contain ORFs corresponding to retroviral *gag* (ORF1) and *pol* (ORF2) genes but lack an ORF3 that possibly corresponds to retroviral *env* (4). These retrotransposons encode virus-like particles that are located in the nucleus (19, 25), in contrast to retroviral particles in vertebrate cells which are assembled in the cytoplasm. In addition, these virus-like particles are not infectious to *D. melanogaster* tissue culture cells, suggesting that retrotransposons of this family are more closely related to mouse intracisternal type A particles than infectious retrovirus particles (4). This inability to form infectious viruses could be due to the lack of an *env* ORF in the members of this retrotransposon family, which may represent an intermediate step in the evolution of retrotransposons to retroviruses. In contrast, members of the gypsy group of LTR-containing retrotransposons, such as gypsy, 297, 17.6, and tom, encode a third ORF containing a membrane-spanning domain near the C terminus which, if properly expressed, could allow them to form true retroviral particles. The results presented here indicate that the tom element can in fact regulate its expression by splicing of the full-length genomic RNA to give rise to an ORF1-ORF3 subgenomic transcript. This process is necessary in vertebrate retroviruses for expression of a functional *env* protein and viral infectivity. In support of this analogy, the protein encoded by the spliced tom RNA contains many similarities to the retroviral *env* products. These similarities include the presence of a signal peptide for targeting to the endoplasmic reticulum, glycosylation sites, and a protease cleavage site to process the *env* precursor into surface and transmembrane proteins. The putative transmembrane product contains a transmembrane domain and a region of homology to the fusion peptide located in the amino-terminal end and involved in the fusion of the viral and cell membranes during the process of infection. Antibodies against the tom ORF3 region recognize two proteins in extracts from ovaries whose sizes correspond well to the predicted surface protein and its uncleaved precursor after taking into account the possibility of modification by glycosylation. The ability of tom to encode an *env*-specific mRNA and the presence of *env* products in ovaries establish further parallels between retroviruses and retrotransposons and support the possibility that elements such as tom might encode infective virus particles that could mediate the horizontal transmission of this class of retrotransposons.

ACKNOWLEDGMENTS

We thank Jef Boeke for invaluable discussions and suggestions in the course of these studies. We also thank Karen Wendel and Tibor Roberts for technical assistance.

This work was supported by American Cancer Society grant DB-7E.

REFERENCES

1. **Awasaki, T., N. Juni, T. Hamabata, K. Yoshida, M. Matsuda, Y. N. Tobari, and S. H. Hori.** 1994. Retrotransposon-induced ectopic expression of *cut* causes the *Om(1A)* mutant in *Drosophila ananassae*. Genetics **137**:165–174.
2. **Basler, K., and E. Hafen.** 1989. Ubiquitous expression of *sevenless*: position-dependent specification of cell fate. Science **243**:931–934.
3. **Bishop, J. G., and V. C. Corces.** 1988. Expression of an activated *ras* gene causes developmental abnormalities in transgenic *Drosophila melanogaster*. Genes Dev. **2**:567–577.
4. **Boeke, J. D., and V. G. Corces.** 1989. Transcription and reverse transcription of retrotransposons. Annu. Rev. Microbiol. **43**:403–434.
5. **Bowtell, D. D. L., M. A. Simon, and G. M. Rubin.** 1989. Ommatidia in the developing *Drosophila* eye require and can respond to *sevenless* for only a restricted period. Cell **56**:931–936.
6. **Finnegan, D. J., and D. H. Fawcett.** 1986. Transposable elements in *Drosophila melanogaster*, p. 1–62. *In* N. Maclean (ed.), Oxford surveys in eukaryotic genes, vol. 3. Oxford University Press, Oxford.
7. **Gerasimova, T. I., L. V. Matjunina, L. J. Mizrokhi, and G. P. Georgiev.** 1985. Successive transposition explosions in *Drosophila melanogaster* and reverse transpositions of mobile dispersed genetic elements. EMBO J. **4**:3773–3779.
8. **Higashijima, S., T. Kojima, T. Michiue, S. Ishimaru, Y. Emori, and K. Saigo.** 1992. Dual *Bar* homeo box genes of *Drosophila* required in two photoreceptor cells, R1 and R6, and primary pigment cells for normal eye development. Genes Dev. **6**:50–60.
9. **Hinton, C. W.** 1984. Morphogenetically specific mutability in *Drosophila ananassae*. Genetics **106**:631–653.
10. **Hinton, C. W.** 1988. Formal relations between *Om* mutants and their suppressors in *Drosophila ananassae*. Genetics **120**:1035–1042.
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. **Hunter, E., E. Hill, M. Hardwick, A. Bhown, D. E. Schwartz, and R. Tizard.** 1983. Complete sequence of the Rous sarcoma virus *env* gene: identification of structural and functional regions of its product. J. Virol. **46**:920–936.
13. **Kojima, T., S. Ishimaru, S. Higashijima, E. Takayama, H. Akimaru, M. Sone, Y. Emori, and K. Saigo.** 1991. Identification of a different-type homeobox gene, *BarH1*, possibly causing *Bar* (*B*) and *Om(1D)* mutations in *Drosophila*. Proc. Natl. Acad. Sci. USA **88**:3434–3447.
14. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
15. **Laverty, T. R., and J. K. Lim.** 1982. Site-specific instability in *Drosophila melanogaster*: evidence for transposition of destabilizing element. Genetics **101**:461–476.
16. **McKearin, D. M., and A. C. Spradling.** 1990. *bag-of-marbles*: a *Drosophila* gene required to initiate both male and female gametogenesis. Genes Dev. **4**:2242–2251.
17. **Mevel-Ninio, M., M.-C. Mariol, and M. Gans.** 1989. Mobilization of the gypsy and copia retrotransposons in *Drosophila melanogaster* induces reversion of the ovo^D dominant female-sterile mutations: molecular analysis of revertant alleles. EMBO J. **8**:1549–1558.
18. **Miller, K., J. Rosenbaum, V. Zbrzezna, and A. O. Pogo.** 1989. The nucleotide sequence of *Drosophila melanogaster* copia-specific 2.1 kb mRNA. Nucleic Acids Res. **11**:2134.
19. **Miyake, T., N. Mae, T. Shiba, and S. Kondo.** 1987. Production of virus-like particles by the transposable genetic element copia of *Drosophila melanogaster*. Mol. Gen. Genet. **207**:29–37.
20. **Mount, S.** 1982. A catalogue of splice junction sequences. Nucleic Acids Res. **10**:459–472.
21. **Mozer, B., R. Marlor, S. Parkhurst, and V. Corces.** 1985. Characterization and developmental expression of a *Drosophila ras* oncogene. Mol. Cell. Biol. **5**:885–889.
22. **Saiki, R. K., S. Scharf, F. Faloona, K. B. Mullis, G. T. Horn, H. A. Erlich, and N. Arnheim.** 1985. Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of

sickle cell anemia. Science **230:**1350–1354.
23. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
24. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:**5463–5467.
25. **Shiba, T., and K. Saigo.** 1983. Retrovirus-like particles containing RNA homologous to the transposable element copia in *Drosophila melanogaster*. Nature (London) **302:**119–124.
26. **Shrimpton, A. E., E. A. Montgomery, and C. H. Langley.** 1986. *Om* mutations in *Drosophila ananassae* are linked to insertions of a transposable element. Genetics **114:**125–135.
27. **Spradling, A. C., and A. P. Mahowald.** 1979. Identification and genetic localization of mRNAs from ovarian follicle cells of *Drosophila melanogaster*. Cell **16:**589–598.
28. **Spreij, T. E.** 1971. Cell death during the development of the imaginal discs of *Calliphora erythrocephala*. Neth. J. Zool. **21:**221–264.
29. **Tanda, S., and V. G. Corces.** 1991. Retrotransposon-induced overexpression of a homeobox gene causes defects in eye morphogenesis in *Drosophila*. EMBO J. **10:**407–417.
30. **Tanda, S., A. E. Shrimpton, C. W. Hinton, and C. H. Langley.** 1989. Analysis of the *Om(1D)* locus in *Drosophila ananassae*. Genetics **123:**495–502.
31. **Tanda, S., A. E. Shrimpton, C. Ling-Ling, H. Itayama, H. Matsubayasi, K. Saigo, Y. N. Tobari, and C. H. Langley.** 1988. Retrovirus-like features and site specific insertions of a transposable element, tom, in *Drosophila ananassae*. Mol. Gen. Genet. **214:**405–411.
32. **Tautz, D., and C. Pfeiffle.** 1989. A nonradioactive *in situ* hybridization method for the localization of specific RNAs in *Drosophila* embryos reveals transcriptional control of the segmentation gene *hunchback*. Chromosoma **89:**81–85.
33. **Varmus, H., and P. Brown.** 1989. Retroviruses, p. 53–108. *In* D. E. Berg and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
34. **Von Heijne, G.** 1986. A new method for predicting signal sequence cleavage sites. Nucleic Acids Res. **14:**4683–4690.
35. **Wolf, T., and D. Ready.** 1991. Cell death in normal and rough eye mutants of *Drosophila*. Development **113:**825–839.
36. **Yoshida, K., and S. Hori.** Personal communication.