THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 280, No. 13, Issue of April 1, pp. 12397–12404, 2005

Printed in U.S.A.

DTL, the *Drosophila* Homolog of PIMT/Tgs1 Nuclear Receptor Coactivator-interacting Protein/RNA Methyltransferase, Has an Essential Role in Development*

Received for publication, August 12, 2004, and in revised form, January 31, 2005
Published, JBC Papers in Press, January 31, 2005, DOI 10.1074/jbc.M409251200

Orbán Komonyi‡, Gábor Pápai§, Izzet Enunlu§, Selen Muratoglu§¶, Tibor Pankotai‡, Darija Kopitova‖, Péter Maróy‡, Andor Udvardy§, and Imre Boros‡§**

From the ‡Department of Genetics and Molecular Biology, University of Szeged, Közép fasor 52, H-6726 Szeged, Hungary, §Institute of Biochemistry, Biological Research Center, Temesvari körut 62, H-6726 Szeged, Hungary, and ‖Institute of Gene Biology, Russian Academy of Sciences, Vavilov Street 34/5, Moscow 117 334, Russia

We describe a novel *Drosophila* gene, *dtl* (*Drosophila Tat-like*), which encodes a 60-kDa protein with RNA binding activity and a methyltransferase (MTase) domain. *Dtl* has an essential role in *Drosophila* development. The homologs of DTL recently described include PIMT (peroxisome proliferator-activated receptor-interacting protein with a methyltransferase domain), an RNA-binding protein that interacts with and enhances the nuclear receptor coactivator function, and TGS1, the methyltransferase involved in the formation of the 2,2,7-trimethylguanosine (m_3G) cap of non-coding small RNAs. DTL is expressed throughout all of the developmental stages of *Drosophila*. The *dtl* mRNA has two ORFs (uORF and dORF). The product of dORF is the 60-kDa PIMT/TGS1 homolog protein that is translated from an internal AUG located 538 bp downstream from the 5′ end of the message. This product of *dtl* is responsible for the formation of the m_3G cap of small RNAs of *Drosophila*. Trimethylguanosine synthase activity is essential in *Drosophila*. The deletion in the dORF or point mutation in the putative MTase active site results in a reduced pool of m_3G cap-containing RNAs and lethality in the early pupa stage. The 5′ region of the *dtl* message also has the coding capacity (uORF) for a 178 amino acid protein. For complete rescue of the lethal phenotype of *dtl* mutants, the presence of the entire *dtl* transcription unit is required. Transgenes that carry mutations within the uORF restore the MTase activity but result in only partial rescue of the lethal phenotype. Interestingly, two transgenes bearing a mutation in uORF or dORF in *trans* can result in complete rescue.

An RNA-binding protein, PIMT[1] (peroxisome proliferator-activated receptor-interacting protein with a methyltransferase domain), was recently identified in mammalian cells and described as a component of a transcriptional coactivator complex that enhances the nuclear receptor coactivator peroxisome proliferator-activated receptor-interacting protein function (1, 2). The RNA-binding domain of the 852 amino acid PIMT was mapped to its N-terminal domain, which contains the G*XX*G*XX*I sequence often present in the K homology domain of RNA-binding proteins. Another characteristic of PIMT is the presence of a MTase domain at its C terminus. Indeed, PIMT bound AdoMet; however, a truncated version of the protein that lacked the putative MTase but contained the RNA-binding domain still had the ability to enhance the transcriptional activity of nuclear receptors (1, 2). A homolog of PIMT has been described in *Saccharomyces cerevisiae* as the product of the *YPL157w* gene, an AdoMet-dependent MTase (3). In yeasts, *YPL157w* is not essential. Its deletion affects splicing and results in a cold-sensitive phenotype. By analyzing the effects of mutations in the conserved MTase domain of the encoded protein, Mouaikel *et al.* (3) demonstrated that the gene is essential for the formation of the 2,2,7-trimethylguanosine (m_3G) 5′-cap structure of snRNAs and small nucleolar RNAs and designated the protein trimethylguanosine synthase 1 (TGS1). TGS1 homologs have been identified in a number of organisms, all of them sharing the conserved sequence motifs of MTases (4). The MTase domain, however, comprises only the ~220 amino acid long C-terminal regions of the proteins, and as compared with the 315 amino acid yeast protein, the TGS1 homologs of higher eukaryotes possess an additional extended N-terminal region. Here, we describe the *Drosophila melanogaster* TGS1, which we identified on the basis of its affinity to bind RNA. *D. melanogaster* TGS1 is translated via an internal translation start site from an RNA, which has coding capacity for two polypeptides. We present genetic evidence that *tgs1* is an essential gene in *Drosophila* and that its mutation compromises RNA m_3G cap formation. Identification of the *tgs1* mutation in *Drosophila* opens up the possibility of a detailed analysis of the complex function(s) in which TGS1/PIMT participates in higher eukaryotes.

EXPERIMENTAL PROCEDURES

Recombinant DNA Constructions—All of the reagents used for recombinant DNA work were from Sigma or Promega unless otherwise indicated. Restriction endonucleases and other DNA-modifying enzymes used for recombinant constructions were from Fermentas,

* This work was supported by grants from the Hungarian Science Fund (T046414), the Hungarian Ministry of Health (078/2003), and European Community (LSHG-CT-2004-502950). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¶ Present address: Hematopoesis Department, University of Maryland, Holland Laboratory, 15601 Crabbs Branch Way, Rockville, MD 20855.

** To whom correspondence should be addressed. Tel.: 36-62-544686; Fax: 36-62-544651; E-mail:borosi@bio.u-szeged.hu.

[1] The abbreviations used are: PIMT, peroxisome proliferator-activated receptor-interacting protein with a methyltransferase domain; Mtase, methyltransferase; AdoMet, *S*-adenosyl-L-methionine; m_3G, 2,2,7-trimethylguanosine; TGS1, trimethylguanosine synthase 1; DTL, *Drosophila* Tat-like; RF, reading frame; ORF, open reading frame; dORF, downstream ORF; uORF, upstream ORF; EGFP, enhanced green fluorescent protein; IRES, internal ribosome entry site; TSS, transcription start site; HIV, human immunodeficiency virus; d189, deficiency 189; RACE, rapid amplification of cDNA ends; sn, small nuclear; TAR, transactivation response element.

whereas radioactive preparations were from Izotop Ltd. (Budapest, Hungary). Full-length cDNA clones of *dtl* were isolated from a *D. melanogaster* embryonic cDNA library using as hybridization probe a cDNA fragment recovered in the screen described below. Several independent cDNA clones were sequenced to establish the structure of the *dtl* message as depicted in Fig. 1. The corresponding genomic region was isolated from a λFixII genomic library through consecutive hybridization steps with ^{32}P-labeled cDNA probes.

To generate fragments encompassing the upstream or the downstream open reading frames (ORFs) (uORF and dORF, respectively) of *dtl* by PCR amplification, the following primers were used (numbers are according to the *dtl* cDNA): DTLuF, 5′-ATGGTA*CC*CGCCGCCACCATGTCCCTGGTGCCAG-3′ (ATG *underlined* at +80); DTLdF, 5′-ATGGTA*CC*CGCCGCCACCATGAACACACACCACCTG-3′ (ATG *underlined* at +613); and DTLudR, 5′-CTGCCAGAATTCTTTCC-3′ (the EcoRI site *underlined* is at +739).

The plasmids pET3a-DTL, pET3a-DTLu, and pET3a-DTLd were constructed in order to express DTL proteins in bacteria. For this reason, the *dtl* cDNA and two subfragments containing either the uORF or the dORF (encompassing the regions from nucleotides +80 to +739 and from +613 to +2008) were inserted into pET3a (Novagen). Recombinant proteins were produced in BL21(DE3) cells following isopropyl 1-thio-β-D-galactopyranoside induction according to standard protocols.

For the construction of DTL-enhanced green fluorescent protein (EGFP) fusions, *dtl* genomic DNA fragments were inserted in front of the gene encoding the EGFP in plasmids pEGFP-N3 and pEGFP-N1 (Promega). Plasmids pDTLu-EGFP and pDTLu^s-EGFP both contain the region from +80 (the second ATG of *dtl* mRNA, see Figs. 1 and 4) to +739 inserted in front of EGFP (the numbers indicate the position of the A of ATGs unless otherwise indicated). The reading frames (RFs) of dORF and of EGFP in pDTLu-EGFP are in phase. pDTLu^s-EGFP was constructed from pDTLu-EGFP by filling up the BssHII site at nucleotide 502, which resulted in a +1 shift in the RF, such that in pDTLu^s-EGFP the RF from +80 to the end of EGFP is open. pDTLd-EGFP carries the region from +614 (from the seventh ATG of *dtl* mRNA) to +739 inserted between the cytomegalovirus promoter and EGFP-coding region, such that the RF of dORF and EGFP are in phase. In pDTLu^d-EGFP, a fragment encompassing the region from +80 to +502 is inserted in front of the EGFP, such that the phase of EGFP is the same as in uORF. In pDTL-EGFP, the *dtl*-coding region from nucleotide +80 to +1803 is inserted in-frame in front of the EGFP-coding region. The structures of the plasmid constructs were verified by sequencing. Details on the constructions are available upon request.

Isolation of D. melanogaster TGS1 cDNA—We isolated a cDNA fragment encoding part of the *Drosophila* homolog of PIMT/Tgs1 in a modified version of a two-plasmid screen developed for the selection of peptide sequences capable of inducing antitermination on lentiviral TAR sequences (5). In this modified version of the screen, the combinatorial peptide library of the *N*-expresser plasmid of the original screening system was replaced by a cDNA library prepared by hexameric random primers from *Drosophila* embryonic poly(A)+ mRNA (the average chain length of the cDNAs was 300–400 bp). This *N*-expresser plasmid library was cotransfected with the *N*-reporter plasmid described by Hamada *et al.* (5) in which the HIV TAR sequence was fused to the N-terminal end of the *lacZ* gene. Colonies with the strongest β-galactosidase activity were selected, and the *Drosophila* cDNA segment they carried was recovered and sequenced. The cDNAs we selected represented an as yet unidentified *Drosophila* gene that we designated *dtl* (*Drosophila* Tat-like).

Fly Stocks, Generation, and Rescue of dtl Mutants—Fly stocks were maintained at 25 °C on standard food. The line *l(3)S096713* (967) was identified in the Szeged *Drosophila* Stock Center, and the position of the P-element insertion was determined by nucleotide sequencing following plasmid rescue. Deficiency 189 (d189) was isolated by identifying improper excision after remobilization of the P-element present in *l(3)S096713*. The P-element was mobilized by the p(Δ2–3) transposase source, and mutations were isolated by scoring for loss of the "mini-white" marker. Imprecise excision lines were identified by genetic crosses based on their phenotype over *Df(3R)P14* and further analyzed by Southern blotting of genomic DNA from hemizygotes and by PCR using primers derived from the 31-bp terminal inverted repeat sequence of the P-element and flanking genomic sequences. Finally, the breakpoints of the deletion were determined by sequencing. To generate *dtl* transgene, a 4.6-kb genomic fragment corresponding to the region between −926 and +3700 (with respect to the transcription initiation site of *dtl*) was inserted into the CaSpeR4 vector (pCaSpeR-DTL) (6), which has the mini-white$^+$ selectable marker. Modified versions of the above plasmid, pCaSpeR-DTLu^m and pCaSpeR-DTLd^m (Fig. 1), were constructed by inserting a synthetic oligonucleotide (CTAGTCTAGACTAG) with stop codons (*underlined*) in all three reading frames at position +502 and by deleting the region between +871 and +1868, respectively. For the construction of *dtl* transgene with a single amino acid change in the MTase active center (pCaSpeR-DTL^{S423R}), a G to C point mutation resulting in a Ser to Arg amino acid change at amino acid position 423 was generated in pCaSpeR-DTL by PCR using the mutagenic primers 5′-TTTCCTGAGG*C*CTCCGTGG-3′ and 5′-CCACGGAG*G*CCTCAAGGAAA-3′ in combination with external primers. The mutation generated a new StuI restriction site (AGGCCT), which was used to verify the construct and the cDNA synthesized on it, following RT-PCR. Transformants were made by P-element-mediated germ-line transformation of w^{1118} host embryos using helper plasmid pΔ2–3 as described by Spradling (7). Independent *mini white*$^+$ transformant lines carrying single insertions were isolated, and the insertions were localized to chromosomes by using balancers. From each construct, at least two independent transformed lines were established and used in this work. To analyze the *dtl* phenotype, *y w*; *d189/TM3*,*Sb Ser y*$^+$ females were crossed to *yw*; *P967/TM3*, *Sb Ser y*$^+$ males. To determine the lethal phase, the numbers of animals reaching the second or third instar or pupal stage were recorded. *d189/P967* animals and their siblings carrying the *TM3* balancer were separated based on the *y* and *y*$^+$ markers. The development of *dtl* mutant and control siblings was followed, and the time needed for 50% of the animals to molt to the next instar was taken as characteristic. At least five independent experiments, each involving 30–60 animals, were performed.

Immunological Techniques—The DTL protein and its derivatives were expressed in bacteria. For antibody production, the fragment of the *dtl* cDNA corresponding to the region from +1243 to + 1612 (encoding amino acids 236–357 of DTLd) was generated by NcoI digestion and inserted in-frame into pET3a to obtain pET3a-DTLnn. The truncated DTL protein expressed in pET3a-DTLnn-containing *Escherichia coli* BL21(DE3) cells was isolated from inclusion bodies, refolded, and injected into rabbits. Antibodies were purified on a protein A-Sepharose affinity column. Protein electrophoresis was performed according to standard protocol. After the transfer of proteins by electroblotting, the membranes were blocked for 1 h in 5% nonfat dry milk in TBST (20 mM Tris-HCl, pH 7.4, 150 mM NaCl, 0.05% Tween 20) and incubated overnight with primary antibody diluted in TBST. Membranes were washed, incubated with horseradish peroxidase-conjugated anti-rabbit secondary antibodies (DACO), washed again extensively, and developed using the ECL (Amersham Biosciences) kit by following the manufacturer's recommendations.

Anti-m$_3$G monoclonal antibody (Ab-1, Oncogene) was used to detect the m$_3$G cap in *Drosophila* tissues by immunostaining and for immunoprecipitation. For immunostaining, larvae in various stages of development were dissected in phosphate-buffered saline and fixed in 10% formaldehyde, 1 mM EGTA solution. Treatment with anti-m$_3$G primary antibody at 4 °C was followed by anti-mouse fluorescein isothiocyanate-conjugated secondary antibodies (DACO) for 2 h at room temperature.

For immunoprecipitation, 10 μg of total RNA samples isolated from wild-type and *dtl* mutant L3 larvae, as indicated in Fig. 6, were incubated with protein A-Sepharose-bound m$_3$G-specific monoclonal antibody, collected, washed, and treated with proteinase K. Supernatants were treated similarly without washes. Proteinase K-treated supernatants and precipitates were extracted with phenol-chloroform and precipitated with ethanol. For Northern hybridization, immunoprecipitated and control RNA samples were separated on 8% denaturing polyacrylamide gel, blotted to nylon membrane by capillary transfer, and hybridized with ^{32}P-labeled probes.

To generate hybridization probes, DNA fragments corresponding to regions of the U2 (CR31850) and U4 (CR32998) snRNA were amplified on a genomic DNA template by PCR using the following primers: U2fw, 5′-CGGATCCATCGCTTCTCGGCCTTAT-3′; U2rv, 5′-CGAATTCAAATCCCGGCGGTACTGC-3′; U4fw, 5′-CGGATCCCTTAGCGCAGTGGCAATAC-3′; and U4rv, 5′-CGAATTCCTCGTAAGGGCTTCCAAAA-3′. The amplified products were cloned into pTZ vector, and probes were synthetized by *in vitro* transcription with T7 RNA polymerase in the presence of [α-^{32}P]UTP.

Tissue Culture and Transfections—To assess DTL expression in eukaryotic cells, plasmids carrying DTL-EGFP fusions were transfected into HeLa cells. 3–5 μg of purified DNA was transfected from each plasmid into 10^6 cells by the calcium phosphate coprecipitation method, as described previously (8). 40 h after transfection, cells were harvested, observed under a microscope, and divided into two fractions. One fraction was sorted by fluorescence-activated cell sorter to estimate the transfection efficiency and green fluorescent protein expression. The other fraction was lysed, the protein concentration was determined by

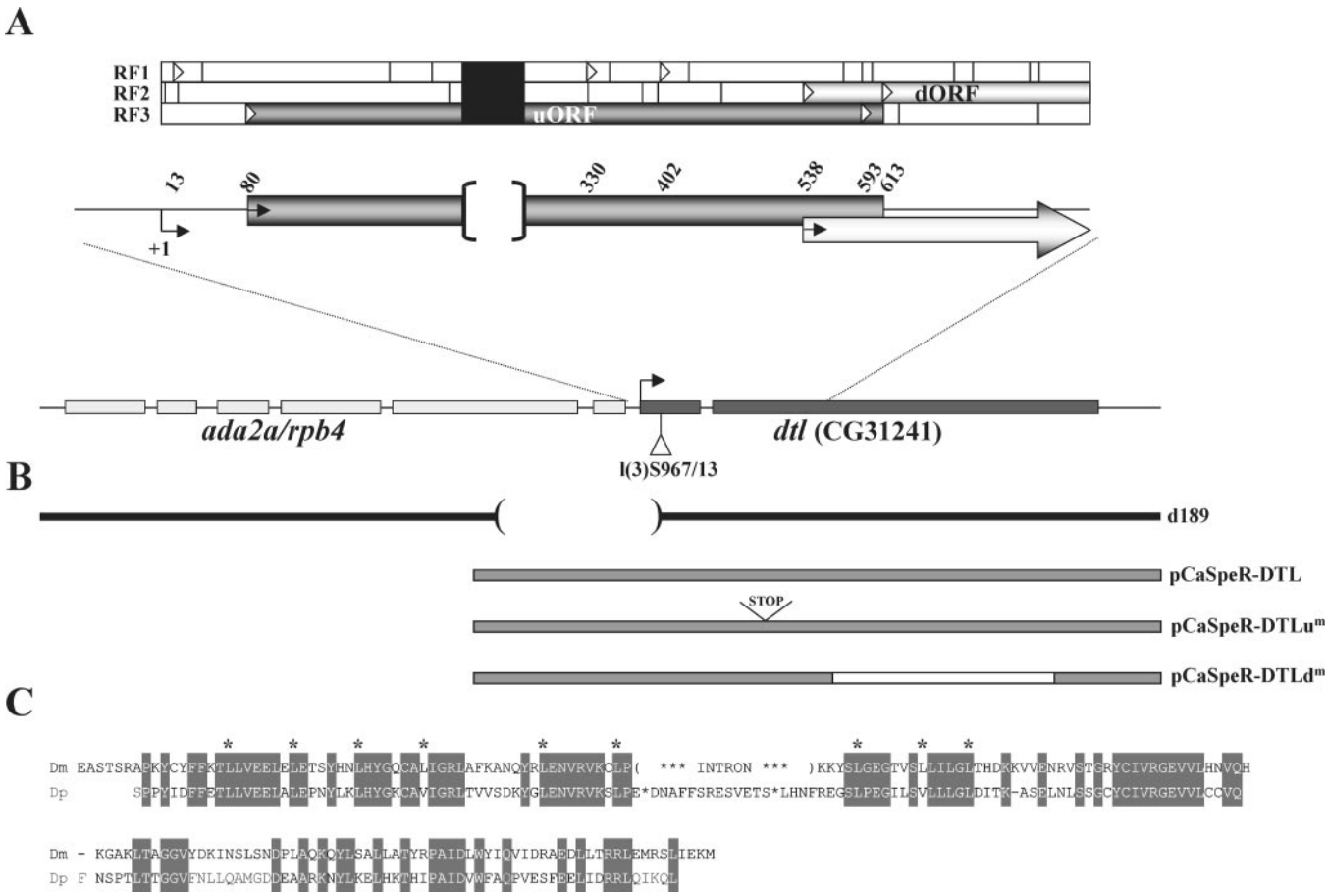

FIG. 1. **Structures of the *dtl* gene (*A*) and *dtl* transgenes (*B*) and the alignment of the translation product of uORF and the related putative protein of *D. pseudoobscura*.** *A*, the genomic structure of *dtl* and the adjacent *ada2a/rpb4* genes is shown with the upstream region enlarged. The *heavy bars* in the *middle* represent the two ORFs; *numbers* indicate the positions of A in the AUG codons. Above, the AUGs (*triangles*) and stop codons (*vertical lines*) in relation to the three possible reading frames (RF1, RF2, and RF3) are shown. uORF (*gray*) and dORF (only the first part is shown in *lighter gray*) correspond to RF3 and RF2, respectively. *Brackets* and the *black box* represent intron. The transcription start site of *dtl* and the position of P-element insertion in line *l(3)S96713* are indicated by +1 and a *triangle*, respectively. The *numbers* refer to the *dtl* cDNA. *B*, the extension of deletion in *d189* and the genomic fragments inserted into CaSpeR4 to generate *dtl* rescue clones. *C*, alignment of DTLu and the related putative protein of *D. pseudoobscura*. Identical amino acids are in a *gray background*. *Asterisks* mark conserved Leu residues occurring in periodic repeat in the N-terminal regions.

the Bradford method, and the proteins were analyzed on denaturing polyacrylamide gel. The proteins were blotted onto nitrocellulose membrane, and Western blots were developed with a green fluorescent protein-specific antibody (a generous gift of Dr. Ferenc Nagy). Blots were developed using the ECL kit (Amersham Biosciences) as recommended and exposed to x-ray films.

RESULTS

DTL Is the Drosophila Homolog of PIMT/Tgs1—In a screen developed for the identification of RNA-binding proteins (5), we recovered several independent clones of a *Drosophila* cDNA fragment encoding peptides able to bind to the HIV TAR region. Subsequent nucleotide sequence analysis revealed that the identified cDNAs represented fragments of an as yet unidentified *Drosophila* gene that we designated *dtl*. To learn more about *dtl*, we isolated further cDNA and genomic clones, sequenced them, and carried out *in silico* searches to identify *dtl*-related sequences in databases. These analyses revealed that *dtl* is a unique gene in the *D. melanogaster* genome located in the 90F region of the cytological map. According to the FlyBase annotation (CG31241), the *dtl* message has an unusually long 5′-untranslated region and gives rise to a 491 amino acid protein product (Fig. 1). Using primer extension and the RACE technique to clone cDNA fragments containing the 5′ end of the *dtl* message, we mapped the transcription start site of *dtl* (9). Northern blots demonstrated that *dtl* gives rise to a single mRNA of 2600 nucleotides, which can be detected in every stage of *Drosophila* development (Fig. 2).

DTL Is Translated via an Internal Initiation Mechanism—The structure of the *dtl* message, as deduced from a comparison of cDNA clones and genomic sequences, is unusual in that the RNA product synthesized from it has two ORFs (Fig. 1*A*). The promoter-proximal one, uORF, starts at the second ATG, 80 bp downstream from the transcription start site (TSS), and has a coding capacity for a 178 amino acid polypeptide. The putative protein-encoded, DTLu (Fig. 1*C*), has no other homologs in the databases with the exception of putative proteins of *Drosophila* species. Its most prominent feature is the presence of Leu residues in a periodical repeat arrangement resembling that seen in leucine zippers. Indeed, the 20-kDa protein product of this ORF forms dimers when expressed in bacteria (Fig. 3, *lanes 1* and *2*). The distal ORF of *dtl*, dORF, has a coding capacity for a 491 amino acid protein. This ORF starts at the fifth ATG, 538 nucleotides downstream from the transcription initiation site (Fig. 1*A*). DTLd, the putative polypeptide encoded by dORF, contains an Arg-rich region and amino acid blocks, characteristic of MTases. The two ORFs of the *dtl* message have a 76-bp overlap. At nucleotides from +614 to +616 (with respect to the TSS), a UGA stop codon closing the uORF overlaps with an AUG, the second in-frame in dORF (AUGA) (Fig. 1*A*). Consequently, depending on the translation

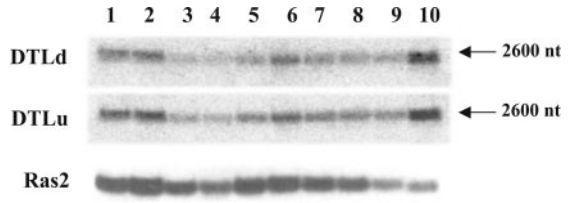

FIG. 2. **Detection of *dtl* transcript by Northern blots.** Total cellular RNA isolated from adult females (*lane 1*), adult males (*lane 2*), late pupa (*lane 3*), middle pupa (*lane 4*), early pupa (*lane 5*), late third instar larva (*lane 6*), early third instar larva (*lane 7*), second instar larva (*lane 8*), first instar larva (*lane 9*), and embryos (*lane 10*) were separated on denaturing gel, blotted onto nylon membrane and hybridized with ^{32}P-labeled *dtl* cDNA. On the blot shown in the *upper panel*, the hybridization probe was a cDNA fragment corresponding to the region of *dtl* from +51 to +739. In the *middle panel*, hybridization with a cDNA fragment from the downstream ORF of *dtl* (from +739 to +2000) is shown. *Bottom panel*, hybridization control with *ras2* cDNA. *nt*, nucleotide.

initiation site(s) used, the message can give rise to protein product(s) in several ways (Fig. 1*A*). Firstly, a −1 slip of the ribosome at the overlap could result in a 644 amino acid product encompassing both ORFs. Secondly, translation reinitiation could result in two separate proteins of 178 and 466–491 amino acids. Thirdly, internal translation initiation could result in a protein of 491 or 466 amino acids. To test experimentally which translation initiation is used *in vivo*, we expressed the *dtl* cDNA in bacteria and generated DTL-EGFP fusions to study their expression in eukaryotic cells. Both *dtl* ORFs directed protein expression when inserted into bacterial expression vectors (Fig. 3). The full-length *dtl* cDNA directed the synthesis of proteins of 20 and 60 kDa (Fig. 3, *lanes 5* and *8*). The molecular weights of these proteins suggested that they correspond to products of the uORF and dORF. Indeed, antibodies specific for DTLu (*lanes 3–5*) and DTLd (*lanes 6–8*) recognized these proteins specifically. It was noteworthy that no product resulting from the cotranslation of the two ORFs was detected (Fig. 3, *lanes 5* and *8*).

To determine whether translation initiated at an ATG in the uORF can proceed into dORF in eukaryotic cells, a fragment corresponding to the +80 to +739 region of *dtl* was inserted in front of the EGFP-coding region to generate plasmid pDTLu-EGFP (Fig. 4*A*). The reading frames of dORF and of EGFP in pDTLu-EGFP are in phase. To test whether initiation can take place at the overlapping start/stop codons at 613, another construct, pDTLd-EGFP, was produced containing the region from +614 to +739 in front of the EGFP-coding region (Fig. 4*A*). In this plasmid, the RF of dORF is in phase with the EGFP. Western blots of extracts of HeLa cells transfected with either pDTLd-EGFP or pDTLu-EGFP and developed by EGFP-specific antibodies showed that pDTLu-EGFP directed the production of a fusion protein ~4–6 kDa larger than EGFP and EGFP, whereas pDTLd-EGFP directed only EGFP production (Fig. 4*B*, *lanes 1* and *2*). This result suggested that, as in bacteria, translation initiation in eukaryotic cells also started at the ATG located at the start of dORF (at +538) and that no fusion product of uORF-dORF-EGFP was produced. To test whether translation occurring on the uORF had any effect on the translation of dORF, we constructed additional EGFP fusions. In pDTLu^{d}-EGFP, the EGFP-coding region was fused in phase to the region coding for the N-terminal part of uORF at position 502. This plasmid has the ATG opening the uORF of *dtl* but lacks the ATG at the start of the dORF. In pDTLu^{s}-EGFP, the reading frame within the uORF was altered by the insertion of four nucleotides at position +502. Consequently, in this last construct, the ATG at +80 opens a reading frame extending to the end of the EGFP-coding region. HeLa cells transfected with pDTLu^{d}-EGFP produced no fusion protein but EGFP. In contrast, in extracts of pDTLu^{s}-EGFP-transfected cells, we detected proteins identical to those seen in pDTLu-EGFP, *e.g.* EGFP and a 4–6-kDa larger fusion protein. These results led us to conclude that translation of the *dtl* message starts at an internal translation initiation site located at 538 nucleotides from the 5′ end of the message. Accordingly, the major product of *dtl* is a 491 amino acid protein. Western blots of *Drosophila* cell extracts developed by antibodies raised against bacterially expressed DTL support this result by showing the production of a 60-kDa protein (Fig. 5*A*). However, it must be noted here that our antibodies also detected other immunoreactive bands (both larger and smaller) in animals in different developmental stages (data not shown).

To detect the cellular localization of the DTL protein translated in eukaryotic cells, we constructed a further fusion pDTL-EGFP in which the EGFP-coding region is fused in-frame to the C-terminal region of DTLd after amino acid 410 (Fig. 4*A*). In HeLa cells transfected with pDTL-EGFP, we detected intense fluorescence in the nucleus (Fig. 5*B*). Furthermore, within the nucleus, we often observed several small spots exhibiting very intense fluorescence.

Dtl Has an Essential Function in Drosophila—For the functional analysis of *dtl*, we initiated genetic studies to identify and characterize *dtl* mutation. A lethal P-element insertional mutant *D. melanogaster* line, *(3)S096713 P967*, identified in an independent screen, carries a P-element insertion 35 bp downstream from the transcription initiation site of *dtl* (Fig. 1*B*). *P967* homozygotes and *P967* mutation over deficiency *Df(3R)P14*, which uncovers the 90FG region, are lethal. The P-element insertion in *P967* is only 107 nucleotides upstream from the transcriptional start of the adjacent *ada2a/rpb4* gene, and consequently, the function of *ada2a/rpb4* could also be affected by the insertion (9). Because P-element insertions are often hypomorphic, in order to facilitate further genetic studies, we generated an additional *dtl* allele by remobilizing the P-element in line *l(3)S096713*. In line *d189* isolated by this technique, improper excision of the P-element resulted in a deletion extending from +35 to −1235 with respect to the transcription start site of *dtl* (Fig. 1). In *P967/d189* second instar larvae, we could not detect wild-type *dtl*-specific mRNAs by RT-PCR analysis. On Western blots of protein extracts of *P967/d189* L1 larvae, a reduced amount of the 60-kDa protein recognized by anti-DTLd polyclonal antibody was detected (data not shown). We believe that the DTLd protein present in *P967/d189* animals at L1 represents a fraction of the maternally provided DTLd still present in this stage of development.

The development of *P967/d189* animals is nearly normal until they reach stage L3. The earliest difference we observed between *dtl* mutants (*P967/d189*) and their control siblings (*TM3* heterozygotes) was that the mutants needed a slightly longer time to complete the L2 instar (23.1 ± 1.2 h *versus* 26.5 ± 0.9). Nonetheless, 84.6% *P967/d189* animals completed the larval stages of development and formed puparium. However, none of them emerged as an adult, all perishing in stages P3-P4. To provide definite proof that the observed phenotype of *P967/d189* animals resulted from the lack of *dtl* function, we constructed and analyzed *Drosophila* lines carrying *dtl* transgenes. A 4.6-kb genomic fragment corresponding to the *dtl*-coding region and adjacent chromosomal DNA was inserted into pCaSpeR4 to generate pCaSpeR4-DTL (Fig. 1), which was injected into embryos. Following appropriate genetic crosses, two independent transgene carrier lines (*P967/d189* P[*DTL*]) were established. The two lines gave identical results in the following experiments. The P[*DTL*] transgene completely restored the wild-type phenotype of *P967/d189* animals. Thus, we

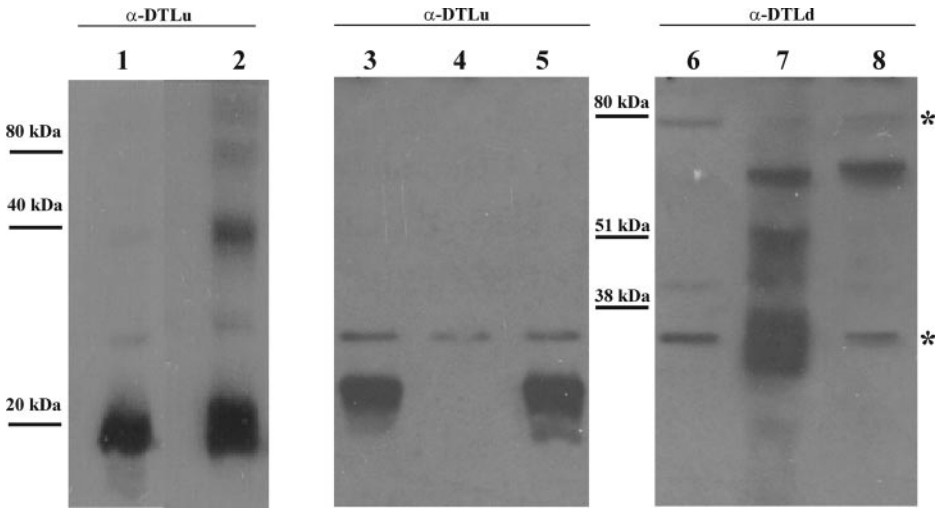

FIG. 3. **Expression of DTL proteins in *E. coli*.** Western blots of protein extracts prepared from *E. coli* transformed with pET3a-DTLu (*lanes 1–3* and *6*), pET3a-DTLd (*lanes 4* and *7*), and pET3a-DTL (*lanes 5* and *8*). The same samples were loaded onto *lanes 1* and *2* following treatment according to the standard protocol (*lane 1*) or without denaturing (*lane 2*) to detect protein dimer formation. Blots with *lanes 1–2* and *3–5* were developed with antibodies specific for DTLu (α-*DTLu*). *Lanes 6–8* were developed with DTLd-specific polyclonal antibody (α-*DTLd*). Estimated molecular weights and bands corresponding to DTLu, DTLd, and the dimer of DTLu are labeled. *Asterisks* indicate the positions of nonspecifically interacting bacterial proteins. The protein expressed from pET3a-DTLd is shorter and more extensively degraded than the protein translated from the full-length cDNA (compare *lanes 7* and *8*). Note that blots with *lanes 1–2* and *3–8* represent different gels of 12 and 10% acrylamide, respectively.

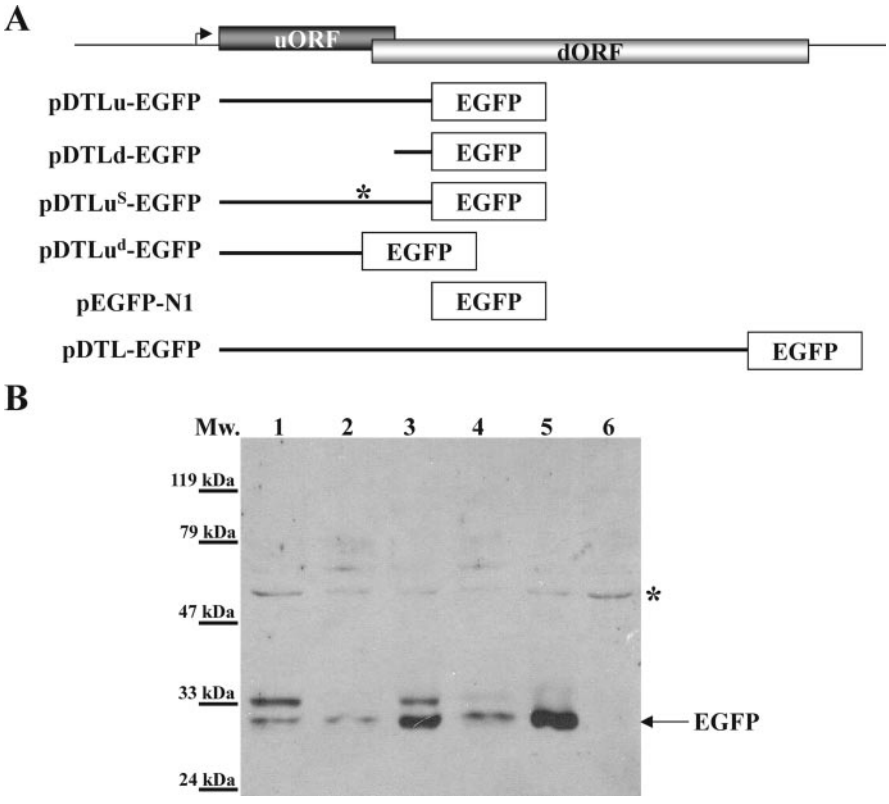

FIG. 4. **Expression of *dtl* uORF-EGFP and dORF-EGFP fusion proteins in HeLa cells.** *A*, schematic structures of the two ORFs of *dtl* and the *dtl*-EGFP fusions used. Detailed descriptions of the constructs are given under "Experimental Procedures." The *asterisk* indicates the position of a +1 frameshift generated in pDTLu-EGFP to obtain pDTLu^{s}-EGFP. *B*, Western blot of protein extracts of HeLa cells transfected with plasmid pDTLu-EGFP (*lane 1*), pDTLd-EGFP (*lane 2*), pDTLu^{s}-EGFP (*lane 3*), pDTLu^{d}-EGFP (*lane 4*), pEGFP-N3 (*lane 5*), or an extract of mock-transfected cells (*lane 6*). The blot was developed with anti-GFP monoclonal antibody. Molecular masses (*Mw.*) are indicated on the *left*, and the position of EGFP is indicated on the *right*. The *asterisk* indicates a nonspecifically interacting protein present in HeLa cells.

established that the early pupal lethality observed in *P697/d189* animals is a result of *dtl* mutation.

Loss of dtl Results in a Loss of m_3G Cap-containing RNA Pool—The strongest homology that DTLd protein displays with proteins characterized so far relates to TGS1 of yeast and mammals in its C-terminal region (4). To ascertain whether the loss of enzyme activity needed for m_3G cap formation of small non-coding RNAs could be detected in *dtl* mutants, we compared the immunostaining of dissected larval tissues of wild-type and *dtl* mutant (*P967/d189*) animals with m_3G cap-specific antibody. The immunostaining of tissues of larvae in different stages of development with m_3G-specific monoclonal

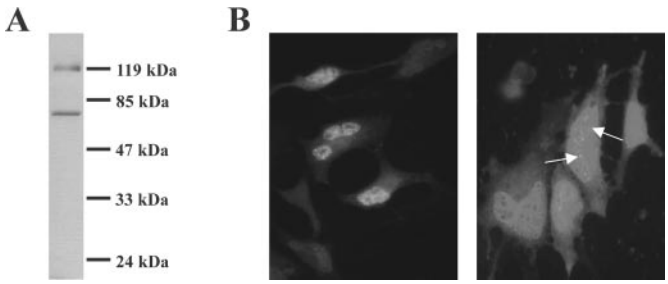

FIG. 5. **The major product of DTL is a 60-kDa protein localized in the nucleus.** *A*, Western blots of S2 cell extracts developed by DTLd-specific polyclonal antibody. *B*, fluorescence microscopy images of HeLa cells transfected with pDTL-EGFP. The accumulation of the fusion protein in the nucleus is well visible. *Arrows* point to intensively fluorescent aggregates often observed in cells expressing the DTL-EGFP fusion protein but not in cells expressing only the short N-terminal part of DTLd encoded in pDTLu-EGFP or in cells expressing only EGFP.

antibody indicated that, in *dtl* animals, the pool of m_3G cap-containing RNAs gradually diminished. As assessed by the staining of dissected larval tissue with m_3G-specific antibody, by the end of the third larval stage, shortly before the mutant animals perished, virtually no m_3G cap-containing RNAs were detected (Fig. 6*A*). By comparing the RNAs of wild-type and mutant animals immunoprecipitated with m_3G-specific antibody, we also observed the loss of m_3G cap-containing snRNAs from *dtl* (*d189/967*) mutants. Northern blot hybridization to immunoprecipitated samples of total RNA isolated from third instar larvae revealed a decrease in the level of m_3G cap-containing U2 and U4 snRNA (Fig. 6*B*). The immunostaining of *189/967* P[*DTL*] larvae demonstrated that a transgene carrying the entire *dtl* region P[*DTL*], which rescued the phenotype, also restored the m_3G cap-containing RNA pool (data not shown). We next considered whether the loss of m_3G cap and lethality were linked and resulted from the loss of *dtl* uORF or dORF or both. To answer these questions, we further tested transgenes for their ability to rescue the *dtl* phenotype (pupal lethality) and restore the m_3G capping of snRNAs. Transgenes P[$DTLu^m$] and P[$DTLd^m$] are modified versions of P[*DTL*], generated by the insertion of stop codons in the uORF (at nucleotide 502) and the deletion of a region of the dORF (between nucleotides 871 and 1868), respectively. Animals carrying transgene P[$DTLd^m$] in the *d189/967* background display a phenotype identical to that seen in *d189/967* (pupal lethality) and the loss of m_3G cap-containing RNAs as determined by either immunostaining or immunoprecipitation (Fig. 6, *A* and *B*). In contrast, in animals carrying transgene P[$DTLu^m$] in an identical genetic background, we detected m_3G cap-containing RNAs at a similar level as in wild-type animals by both immunostaining and immunoprecipitation (Fig. 6, *A* and *B*). However, although the P[$DTLu^m$] transgene rescued P3-P4 lethality, this rescue was not complete and animals carrying the transgene died as pharat adults or immediately after hatching. Interestingly, the transgenes P[$DTLd^m$] and P[$DTLu^m$] together in *trans* position resulted in a complete rescue, leading to the development of fertile adult animals without obvious defects. These data suggest that the product of *dtl* downstream ORF (DTLd) is TGS1, the fission yeast homolog of which has been demonstrated recently to catalyze methyl transfer from AdoMet to M^7GTP (10), and that the function determined by this ORF is essential for *Drosophila* development. Nevertheless, the loss of a function other than that of TGS1 should also be considered because the product of DTL dORF is structurally more similar to its mammalian than to its yeast homolog and the mammalian counterpart has been shown to have a transcriptional coactivator function (1). Accordingly, we set out to test whether the lethality is a result of the loss of TGS1 activity or any other function of DTLd. For this reason, we generated a *dtl* transgene with a single amino acid change at the MTase active site, P[DTL^{S423R}]. An identical mutation of the yeast homolog has been reported to abolish enzyme activity (4). The mutation also generated an StuI cleavage site, which we used to verify that the mutant TGS1 mRNA was expressed in the established transgenic lines (data not shown). Insertion of the P[DTL^{S423R}] transgene did not change the lethal phase of *d189/967* animals. Similarly, no m_3G cap-containing RNAs can be detected in *d189/967* P[DTL^{S423R}] animals by immunostaining (Fig. 6*A*). From these data, we concluded that the TGS1 encoded by the downstream ORF of the *dtl* gene of *Drosophila* is essential for normal fly development.

DISCUSSION

In a screen developed to isolate cDNAs encoding RNA-binding proteins capable of binding to the HIV TAR RNA, we recovered cDNA fragments of an as yet unknown *Drosophila* gene *dtl*. Our analysis of further cDNAs and the corresponding genomic regions revealed that *dtl* has the coding potential for two proteins. The unusual structure of the gene was surprising; hence, we tested the types of transcript(s) and translation products arising from it. Northern blots developed by probes corresponding to either the first or the second ORF of the gene demonstrated an identical RNA species of 2600 nucleotides, indicating that the gene is transcribed as a polycistronic unit. Overlapping translational stop and start codons at the end of the first ORF are another characteristic feature of the *dtl* message. We sequenced several independent cDNA clones and found this region invariably. Furthermore, ESTs present in the data base also revealed the stop and start codons in identical positions. The translated amino acid sequence of the first ORF of *dtl* has hardly any homologs in the existing databases. Protein BLAST searches of the known genomic sequence of *Drosophila pseudoobscura* identified two putative translation products similar to the two products of *dtl* ORFs both in amino acid sequence and in relation to each other. Although there is considerable divergence between the putative products of *dtl* uORF and the related protein of *D. pseudoobscura*, the periodic occurrence of Leu residues in the N-terminal regions of the proteins is preserved and alterations that result in nonsense codons are present only in the intronic region (Fig. 1*C*). Nonetheless, at present, we do not have definite proof that the putative protein product of the first ORF of *dtl* is produced at any time during *Drosophila* development.

The second ORF of *dtl* has a coding capacity for a protein of 491 amino acids. Each cDNA fragment that we isolated in the screen was from this part of the gene and contained the region corresponding to amino acids 200–350. As we isolated the *dtl* cDNAs based on the RNA binding ability of the encoded peptides, we assume that this region carries the RNA binding motif of DTLd. This part of the protein includes an Arg-rich amino acid sequence KKKRRQRQI similar to the RNA-binding region of HIV TAT (RKKRRQRRR) (11, 12).

The relationship between the structural organization of the two ORFs of *dtl* resembles the features of complex transcription units of retroviruses and retrotransposons where a specific translational switch often ensures the synthesis of *gag-pol* polyproteins from one polycistronic RNA. In retroviruses, among them HIV, the utilization of two partially overlapping ORFs requires a −1 ribosome slip (13, 14). We addressed the question of whether there is any evidence of a similar mechanism at the overlapping start/stop codons of the *dtl* message. Because signals directing ribosome slip at overlapping codons within viral RNAs are effectively recognized by the bacterial

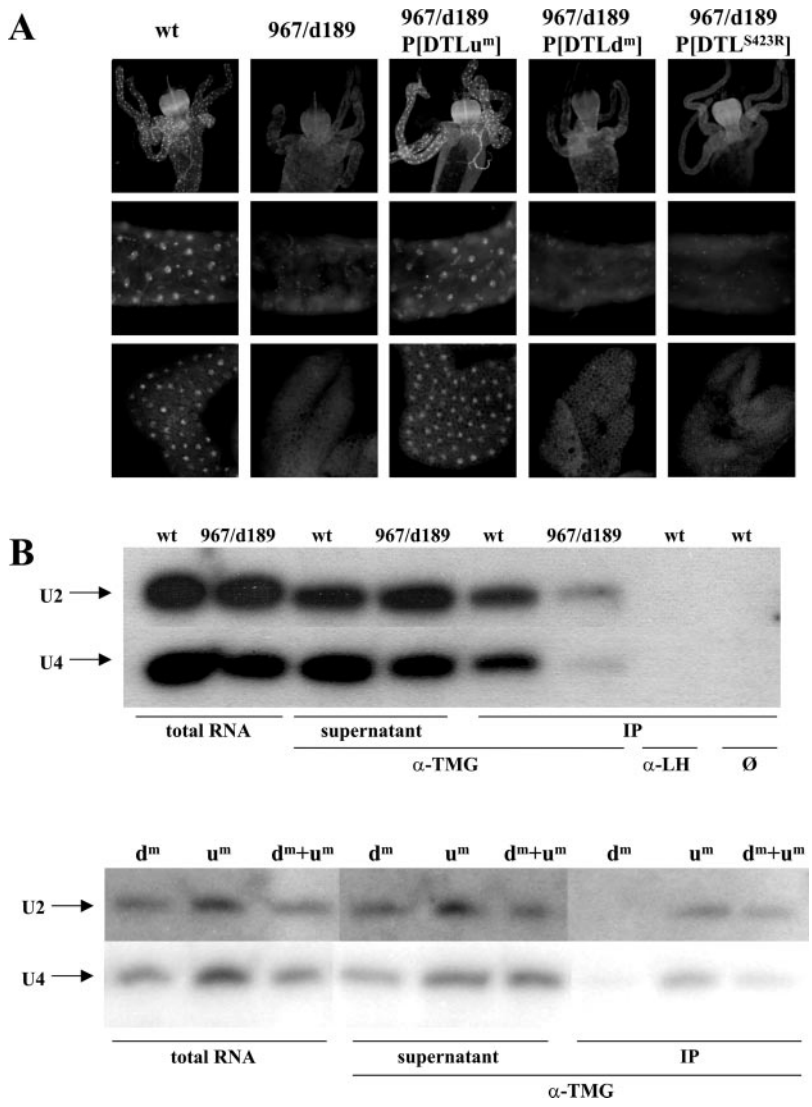

FIG. 6. **The lack of *dtl* function results in a decrease of m_3G cap-containing RNA pool.** *A*, detection of m_3G cap-containing RNAs by m_3G-specific monoclonal antibody (Ab-1) in dissected third instar larva. The genotypes of dissected animals are indicated as wild-type (*wt*) and *l(3)S096713/d189* (*967/d189*). *Top row*, first part of midgut; *middle*, section of midgut at higher magnification; *bottom row*, fat body. *B*, Northern blots of RNA samples precipitated with m_3G-specific monoclonal antibody. Hybridization of U2- and U4-specific probes to total RNA prepared from third instar larvae, supernatants of immunoprecipitations, and precipitated material (*IP*) are shown. *α-TMG* and *α-LH* indicate samples precipitated with m_3G-specific and leghemoglobin-specific antibody as control, respectively. No antibody was added to sample 0. The labels of genotypes are as in *A*. d^m, transgene P[*DTLd*^{m}]; u^m, P[*DTLu*^{m}].

translational machinery (15), we first studied the translation of the *dtl* message in bacteria. As shown in Fig. 3, Western blot analysis of *dtl* proteins expressed in *E. coli*, indicated that, whereas both ORFs of the *dtl* cDNA were translated in bacteria, no cotranslation of the two ORFs occurred and translation initiation of the second ORF started at an internal AUG. In summary, these experiments allowed the conclusion that, under the conditions tested, a ribosome slip resulting in fusion protein formation did not occur in *E. coli*. In eukaryotic cells, under the conditions tested, the first ORF was either not translated or was translated with such low efficiency that its product was not detectable. Plasmid constructs containing the entire uORF and the first part of dORF fused to the EGFP-coding region directed the synthesis of a protein only a few kDa larger than EGFP itself (Fig. 4). This result suggests that translation of the messages synthesized from the fusion genes started only at internal AUGs. Frameshift mutations within the uORF did not affect the synthesis of the fusion protein, indicating that translation of the upstream part of the message is not a prerequisite for initiation at an internal AUG.

Initiation of translation at a downstream AUG recently has been described in a number of eukaryotic messages and appears to be far more frequent than previously believed (16–18). Although our knowledge of the mechanisms is far from complete, it is well accepted that upstream AUGs, upstream ORFs, and internal ribosome entry sites (IRESs) often play critical roles in the expression of genes encoding key regulatory proteins (19, 20). In *Drosophila*, the activity of the IRESs of some genes is developmentally controlled (21). The internal initiation at IRESs is believed to be augmented by RNA-binding proteins. However, the roles of protein factors suspected of being involved are largely unknown. This together with other uncertainties concerning the mechanism prompted Kozak (22) to suggest the redefinition of IRESs as "internal regulatory expression sequences."

The mechanism by which translation of the dORF of *dtl* at an internal AUG of the message is initiated is unclear. A high degree of secondary structure in the 5′ region of *dtl* is not predicted, suggesting that the presence of a stable secondary structure does not inhibit ribosome scanning. Surprisingly, in HeLa cells transfected with plasmids carrying DTL-EGFP fusions, we detected the production of a protein, which we believe is the authentic EGFP. One reason for EGFP expression in the reporter constructs could be that, in the presence of a *dtl* upstream region, ribosomes recognize AUGs within the message at high efficiency. Whether this is a result of a structural feature of the 5′ region of the *dtl* message or a translated peptide from the 5′ region remains to be clarified.

Homology searches for relatives of the 491 amino acid major product of *dtl* in the databases revealed that the closest homolog of the *Drosophila* protein is Tgs1, the 315 amino acid RNA-hypermethylase first identified in *S. cerevisiae* (3). Yeast

Tgs1 and DTL exhibit 41% identity and 58% similarity in their C-terminal 200 amino acid regions. This is the evolutionarily conserved MTase catalytic domain present in Tgs1-like proteins of other organisms as well (4). The large N-terminal domain present in DTL is absent from yTgs1 but is present in Tgs1 orthologs from animals and plants. The related mammalian protein is hTgs1, also identified as PIMT, an 852 amino acid nuclear receptor coactivator-interacting protein (1). Besides the homology between PIMT/hTgs1 and DTL in their C-terminal regions, they also display limited similarity in their central regions. Although DTL clearly has RNA binding affinity and its RNA binding motif is similar to the Arg-rich RNA binding motifs present in many RNA-binding proteins (among others, HIV Tat), we believe that this similarity is coincidental and that no functional homology exists between DTL and HIV TAT.

By remobilization of a P-element (*P967*) integrated in the vicinity of *dtl* TSS, we generated a *dtl* allele *d189*, which lacks the entire *dtl* regulatory region. *P967*/*d189* animals are early pupa lethal, demonstrating that *dtl* has an essential function. Mutant animals need a slightly longer time than their siblings for completion of the L2 stage. The severe phenotype, early pupal lethality, seen in the absence of *dtl* is in contrast with the mild phenotype observed in the absence of Tgs1 in *S. cerevisiae* (3). Immunohistochemical staining with an antibody specific for m_3G cap indicated that, in *dtl* animals as well as in TGS1 mutant yeast cells, the pool of RNAs with m_3G cap is reduced compared with the wild-type controls. The loss of TGS1 activity results in pupal lethality of *dtl* animals (*d189*/*967*). The fact that the loss of MTase activity is the underlying cause of lethality is clearly indicated by the failure of transgenes mutated in the MTase active site to rescue the phenotype. The difference in severity of the mutations of *dtl* and its yeast counterpart yTgs1 might also reflect functional differences between the m_3G cap-containing RNAs in the two organisms. One intriguing question that arises is whether the loss of TGS1 in *dtl* mutants leads to lethality because of a general failure of vital functions such as splicing and rRNA maturation in which snRNA and small nucleolar RNAs are involved or because of depletion of a specific type of m_3G cap-containing RNA. Our immunoprecipitation-hybridization studies demonstrated that RNA samples obtained from *dtl* animals shortly before they died contained practically no m_3G-capped U2 and U4 snRNAs. In accord with this finding, immunostaining of larval tissues also indicated a nearly complete loss of the pool of m_3G cap-containing RNAs. The last residual m_3G cap-specific antibody-reactive material that we could observe was in imaginal histoblasts (Fig. 6). We believe that *967*/*d189* animals are *dtl* null and that the DTLd protein present in mutants in L1 represents a fraction of the maternally provided DTLd still present at this stage of development. However, we cannot exclude the possibility that the *967* allele is hypomorphic. Thus, the roles of maternal and zygotically expressed TGS1 in the survival of *dtl* animals cannot be determined from our data. Even so, it is surprising that, as can be judged from the immunostaining results, a considerable amount of the m_3G cap-containing RNA pool is present in *967*/*d189* mutants even at a late larval stage. Our data do not permit a conclusion regarding whether the loss of TGS1 activity in a higher eukaryote such as *Drosophila* is lethal due to a general failure of RNA metabolism or because of its effect on one or a few specific type(s) of small RNA(s). Nonetheless, the essential role of the enzyme is underlined by the observation that a single amino acid change in the MTase domain results in lethality. As the product of *dtl*, dORF is more similar in its structure to the mammalian PIMT protein than it is to the yeast TGS1, which has only the MTase domain. It is also conceivable that the enzyme in higher eukaryotes has a function requiring MTase activity other than modification of the small RNA cap structure. To our surprise, *dtl* function can be provided by a combination of two transgenes carrying intact uORF and dORF, neither of which alone is capable of complete rescue. The most probable explanation for this observation could be that the two ORFs of *dtl* complement each other in *trans*. Whether the complementation results from the interaction of two proteins translated from the two ORFs of *dtl* or from the interaction of RNA and protein molecules remains to be clarified. At present, we cannot exclude the possibility that the upstream region of *dtl* itself has a function other than influencing the production of the major DTLd protein. Genetic and molecular studies are underway to disclose the relationship that exists between the two products of the *dtl* gene.

Acknowledgments—We thank Katalin Ökrös and Bettina Nagy for their expert technical help. We are grateful to Dr. Ferenc Nagy for kindly providing the green fluorescent protein-specific antibody.

REFERENCES

1. Zhu, Y., Qi, C., Cao, W. Q., Yeldandi, A. V., Rao, M. S, and Reddy, J. K. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98,** 10380–10385
2. Misra, P., Qi, C., Yu, S., Shah, S. H., Cao, W. Q., Rao, M. S., Thimmapaya, B., Zhu, Y., and Reddy, J. K. (2002) J. *Biol. Chem.* **277,** 20011–20019
3. Mouaikel, J., Verheggen, C., Bertrand, E., Tazi, J., and Bordonne, R. (2002) *Mol. Cell* **9,** 891–901
4. Mouaikel, J., Bujnicki, J. M., Tazi, J., and Bordonne, R. (2003) *Nucleic Acids Res.* **31,** 4899–4909
5. Hamada, K., Martin, S. S., and Frankel, A. D. (1996) *Nature* **380,** 175–179
6. Thummel, C. S., Boulet, A. M., and Lipshitz, H. D. (1988) *Gene* (*Amst.*) **74,** 445–456
7. Spradling, A.C. (1986) in *Drosophila, a Practical Approach*, (Roberts, D. B., eds) pp. 175–198, IRL Press, Oxford
8. Kingston, R. E. (1987) in *Current Protocols in Molecular Biology* (Asubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A., and Struhl, K., eds.) pp. 9.1.1–9.6.3, John Wiley & Sons, Inc., New York
9. Pápai, G., Komonyi, O., Tóth, Z. S., Pankotai, T., Muratoglu, S., Udvardy, A., and Boros, I. (2005) *Gene* (*Amst.*), in press
10. Hausmann, S., and Shuman, S. (2005) *J. Biol. Chem.* **280,** 4021–4024
11. Gatignol, A., and Jeang, K. T. (2000) *Adv. Pharmacol.* **48,** 209–227
12. Rana, T. M., and Jeang, K. T. (1999) *Arch. Biochem. Biophys.* **365,** 175–185
13. Jacks, T., Madhani, H. D., Masiarz, F. R., and Varmus, H. E. (1988) *Cell* **55,** 447–458
14. Jacks, T., Power, M. D., Masiarz, F. R., Luciw, P. A., Barr, P. J., and Varmus, H. E. (1988) *Nature* **331,** 280–283
15. Yelverton, E., Lindsley, D., Yamauchi, P., and Gallant, J. A. (1994) *Mol. Microbiol.* **11,** 303–313
16. Morris, D. R., and Geballe, A. P. (2000) *Mol. Cell. Biol.* **20,** 8635–8642
17. Vagner, S., Galy, B., and Pyronnet, S. (2001) *EMBO Rep.* **10,** 893–898
18. Bonnal, S., Boutonnet, C., Prado-Lourenco, L., and Vagner, S. (2003) *Nucleic Acids Res.* **31,** 427–438
19. Morris, D. R. (1995) *Prog. Nucleic Acids Res. Mol. Biol.* **51,** 339–363
20. Holcik, M., Sonenberg, N., and Korneluk, R. G. (2000) *Trends Genet.* **16,** 469–473
21. Ye, X., Fong, P., Iizuka, N., Choate, D., and Cavener, D. R. (1997) *Mol. Cell. Biol.* **17,** 1714–1721
22. Kozak, M. (2003) *Gene* (*Amst.*) **318,** 1–23

Genes: Structure and Regulation:

DTL, the *Drosophila* Homolog of PIMT/Tgs1 Nuclear Receptor Coactivator-interacting Protein/RNA Methyltransferase, Has an Essential Role in Development

Orbán Komonyi, Gábor Pápai, Izzet Enunlu, Selen Muratoglu, Tibor Pankotai, Darija Kopitova, Péter Maróy, Andor Udvardy and Imre Boros

J. Biol. Chem. 2005, 280:12397-12404.
doi: 10.1074/jbc.M409251200 originally published online January 31, 2005