A novel RNA helicase gene tightly linked to the *Triplo-lethal* locus of *Drosophila*

Douglas R.Dorer, Alan C.Christensen* and Daniel H.Johnson[1]
Department of Biochemistry and Molecular Biology, Jefferson Institute of Molecular Medicine, Thomas Jefferson University, 1020 Locust Street, Philadelphia, PA 19107 and [1]Department of Biochemistry and Molecular Biology, University of Miami School of Medicine, 1550 Northwest 10th Avenue, Miami, FL 33136, USA

Received June 5, 1990; Revised and Accepted August 6, 1990 EMBL accession no. X52846

ABSTRACT

The *Triplo-lethal* (*Tpl*) locus of *Drosophila* is the only known locus which is lethal when present in three copies rather than the normal two. After recovering a hybrid-dysgenesis-induced mutation of *Tpl* we used a rapid combination of transposon tagging, chromosome microdissection and PCR to clone the P element that had transposed into the *Tpl* region. That P element is located within the gene for a new and unique member of the RNA helicase family. This new helicase differs from all others known by having glycine-rich repeats at both the amino and carboxyl termini. Curiously, genetic analysis shows that the P element inserted into this gene is not responsible for the *Tpl* mutant phenotype. We present possible explanations for these findings.

INTRODUCTION

One of the presumed selective advantages of diploidy is that the organism is buffered against the consequences of genetic damage by the presence of two copies of each gene. Since most null mutations are recessive, the concentration of most gene products in the cell must generally be in excess of what is required to carry out their function. Therefore, an increase in the copy number of a single gene should rarely be deleterious. This prediction was borne out for a multicellular organism, *Drosophila melanogaster*, when Lindsley, Sandler and coworkers systematically examined the effects of aneuploidy for small regions of the genome (1). Their work demonstrated that there is only one locus in the entire genome which is lethal when present in three copies. This locus was named *Triplo-lethal* (*Tpl*), and is located on the third chromosome at cytological location 83E1,2 (2, our unpublished data). *Tpl* is also haplo-lethal, so any alteration in the dosage of *Tpl* has a dominant lethal phenotype. Individuals carrying either one or three doses of *Tpl* die as late embryos or early first instar larvae with no obvious morphological abnormalities (2–4, our unpublished data).

Genetic analysis of *Tpl* has revealed that it is an unusual or complex locus. Although deficiencies for the locus are easily selected because they block the lethal effect of a duplication of the locus, extensive efforts to obtain point mutations having the same phenotype as deficiencies have been fruitless (3–5). An intermediate class of mutation, referred to as hypomorphic alleles, has been recovered. These mutations block the lethal effect of a duplication of *Tpl* and they are recessive lethals, but they are not dominant lethals as deficiencies are. The recovery of this class of mutation has been sporadic (4). Following P-M hybrid dysgenesis, we recovered fifteen hypomorphic alleles (5). The high rate of recovery of hypomorphic alleles following hybrid dysgenesis, the low rate of recovery with any other mutagen, and the fact that the hypomorphic alleles recovered by Roehrdanz and Lucchesi (4) are hotspots for homologous recombination (6) suggests that transposable elements may be responsible for this class of mutational event at *Tpl*. In addition the apparent failure of all these studies to recover point mutations whose phenotype is the same as a deficiency suggests that *Tpl* may consist of several functionally redundant genes, such that more than one single base change is required to completely eliminate the function of *Tpl* (3–5).

Surprisingly, of the fifteen hypomorphic alleles recovered following hybrid dysgenesis, fourteen have no P element associated with the known cytogenetic location of *Tpl* (5). The fifteenth, Tpl^{J34}, has a P element hybridization signal at band 83E1,2. By a combination of transposon tagging (7) with microdissection and amplification (8) we rapidly cloned the P element present at 83E1,2 along with the flanking genomic DNA. The flanking DNA was then used as a probe to isolate cDNA clones, which were analyzed by DNA sequencing. The cDNA sequence suggests that we have cloned a new member of a growing family of RNA helicase genes (9–23), but one with several unusual features. The protein is homologous to the RNA helicases already found, and appears to be most similar to the human nuclear protein p68 (10). The human and *Drosophila* genes are even interrupted by an intron at the exact same location. However, this *Drosophila* protein, RM62, is probably not the functional equivalent of the human p68 protein since not all

* To whom correspondence should be addressed

putative functional domains are conserved between them, and the *Drosophila* protein has a new glycine-rich motif at both termini that is not present in the human p68 protein.

Sequencing of both genomic and cDNA clones of the helicase shows that the P element in the mutant Tpl^{J34} lies within an intron of the gene. This intron is in an unusual location; it interrupts the initiator AUG codon. Curiously, genetic tests suggest that this P element is not actually responsible for the Tpl^{J34} mutation (5, this work).

Here we report a rapid new method for isolating mobile-element-containing clones from defined chromosomal regions in mutants, and how this approach was used on the Tpl^{J34} mutant. We describe the transcription unit found there, and the new member of the RNA helicase family encoded by the gene. We also discuss the possible relevance of the RNA helicase gene to the mysteries of the *Triplo-lethal* locus of *Drosophila*.

MATERIALS AND METHODS

Drosophila mutants

The origins of the *Tpl* mutations used in this study, and the culture conditions, were as previously described (3, 5). All other genetic markers are described in Lindsley and Grell (24).

Library construction and screening

Genomic DNA was prepared from adult flies of the genotype Tpl^{J34} / *Dp(3R)21l73* by the method of Scott *et al* (25). The DNA was partially digested with Sau3A under conditions optimized to produce a large proportion of 15 to 20 kilobase fragments (26), treated with calf intestinal alkaline phosphatase (Boehringer Mannheim), then ligated to Lambda DASH predigested with BamHI and HindIII (Stratagene). The recombinant phage DNA was packaged in vitro with the Gigapack XL kit (Stratagene) and propagated on E. coli LE392 (Stratagene).

Cosmid and lambda libraries were screened by standard protocols (26) using DNA probes labelled with [α-^{32}P]-dCTP by the random hexamer method (27). Plasmid pπ25.7wc DNA (28) was used for the primary screening of the Tpl^{J34} lambda library. The λgt10 cDNA library (29) was screened with EcoRI fragments from λMargo10 subcloned into pVZ-1 (see below). *In situ* hybridization of biotinylated DNA probes to polytene chromosomes was done as described previously (5).

Northern Blotting

Poly(A)$^+$ RNA was purified from 0−12 hour embryos, third instar larvae, adults and SL2 *Drosophila* tissue culture cells using the guanidine isothiocyanate/CsCl method followed by chromatography on oligo-(dT)-cellulose as described (26). RNA was denatured in glyoxal, electrophoresed in 1% agarose gels, transferred to nitrocellulose, and probed with ^{32}P-labelled DNA as described (30).

Microdissection PCR

A collection of MboI restriction fragments from the region encompassing 83C-83F was produced by microdissection of this region from larval salivary gland polytene chromosomes, followed by MboI digestion, linker ligation to provide primer binding sites, and amplification using the polymerase chain reaction as described (8). Radiolabelled probes were made by further amplifying the collection of fragments for 10 cycles of PCR in the presence of [α-^{32}P]-dCTP.

DNA Sequencing

cDNA clones from λgt10 and genomic restriction fragments from cosmid and lambda clones were subcloned in both orientations in the phagemid pVZ-1 for sequencing of single-stranded DNA templates as described (31). A nested series of deletions was generated for each cloned insert with exonuclease III (32) for sequencing using the dideoxy-chain-termination procedure (33).

Isolation of Tpl^{J34} Revertants and Recombinants

Recombinants between ri^+ Tpl^{J34} Ki^+ p^+ and *ri* Tpl^+ *Ki* p^p were isolated as described (5). Spontaneous revertants of Tpl^{J34} were isolated by mass mating Tpl^{J34} / *TM3, Ser* flies to *ri* Tpl^{10} *Ki* p^p / *TM3, Ser*. Any progeny who were Kinked but non-Serrate were mated to *l(3)DTS-2 Sb / TM3, Ser* to establish a line, and then further tested for the presence of a hypomorphic allele as previously described (5, 6).

Database searching

The GenBank release 63.0 and EMBL release 21.0 nucleotide sequence databases and the PIR release 22.0 protein sequence database were searched using the FASTA program and GENEPRO software from Riverside Scientific as described (34, 35).

RESULTS

Cloning the DNA from the Tpl^{J34} mutant

A genomic library was made from a stock of Tpl^{J34} / *Dp(3;3)21l73* in the vector Lambda DASH (Stratagene). After screening the library with a P element probe we identified numerous positive plaques, most of which represented clones containing sequences from the approximately 30 other P elements present in the genome of the stock. The positive plaques were re-screened in duplicate with the P element probe, and the probe made from the 83C−F region by microdissection and amplification (8). From this screening, we recovered one clone, λMargo10, which hybridized to both probes. *In situ* hybridization of λMargo10 to polytene chromosomes from an M strain confirmed that this clone contained unique sequences from 83E1,2. Figure 1 depicts an *in situ* hybridization of λMargo10

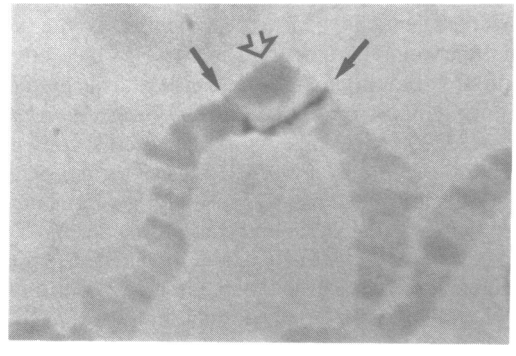

Figure 1. *In situ* hybridization of λMargo10 to a *Tpl* duplication. λMargo10 DNA was biotinylated and hybridized *in situ* to polytene chromosomes from a stock of *Dp(3;3)21l73 / Df(3R)18i77* as described in Materials and Methods. These chromosomes are described in Keppy and Denell (3). The deficiency is subvisible, and the duplication is a tandem reversed duplication of 83E to 84B. This often takes on the appearance of a hairpin or bulge, the center of which is indicated by the open arrow. The dual hybridization signals of λMargo10 in the duplicated region are indicated by the solid arrows.

to polytene chromosomes from a *Dp(3;3)21l73* bearing strain. This duplication, which is a tandem reversed duplication of polytene chromosome bands 83E-84B (3), is the prototypic duplication of *Tpl*, as it is very stable, and has been used with consistent results for a number of years (3–6, 36). The hybridization at the distal end of the duplication shows that the cloned DNA in λMargo10 is contained within the duplication. A restriction map of the DNA insert from λMargo10 is shown in Figure 2, and the location of the P element is indicated. Comparison of the restriction map of λMargo10 with that of wild type DNA suggests that this P element is approximately 2.9kb and may represent a complete P element (37).

Subsequently, one of us (D. H. J., manuscript in preparation) has shown that sub-nanogram quantities of DNA from yeast artificial chromosome clones (YACs) of *Drosophila* can also be digested by a restriction endonuclease, linkered, and amplified by the same method as used here for a microdissected chromosome fragment (see Experimental Procedures). This can then be used as the second probe in order to identify a particular mobile element insert from many positive clones.

Isolation of cDNA clones flanking the P element in *Tpl^{J34}*

The two EcoRI fragments which include part of the P element and the flanking regions (see Figure 2) were subcloned and then used as probes of λgt10 cDNA libraries made from embryos and third-instar larvae (29), in an effort to obtain clones from the transcription unit(s) flanking the site of insertion of the P element in *Tpl^{J34}*. A number of clones were recovered which hybridized to M10−2 but not to M10−18. Many of these appeared to be incomplete at the 3· ends, as no poly(A) tract was found when they were subcloned and sequenced. The 5· cDNA clones were then used to probe the cDNA library, and a large number of independent overlapping cDNA clones was obtained. Subcloning and DNA sequencing of several of these revealed that they all appeared to represent a single class of mRNA molecules, with 2 different poly(A) addition sites. A wild type cosmid library in cosPneo (38) was then screened with the cDNA clones, and two overlapping cosmid clones were obtained. *In situ* hybridization of the cosmid clones to polytene chromosomes of *Dp(3;3)21l73 / Df(3R)18i77* (data not shown) allow us to orient the molecular map to the chromosome as indicated in Figure 2.

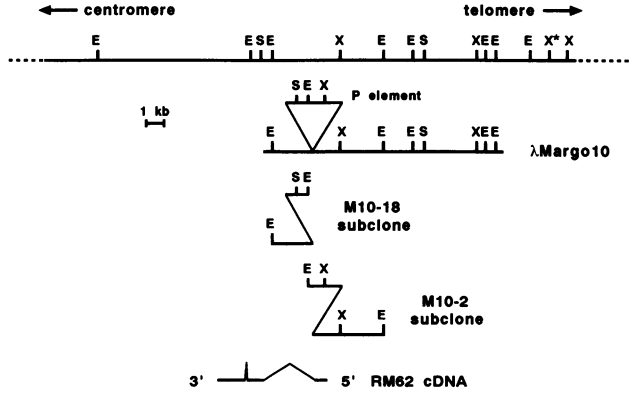

Figure 2. Molecular map of the region surrounding the P element of *Tpl^{J34}*. A wild type restriction map is shown, with abbreviations as follows: E, EcoRI; X, XhoI; S, SalI. The extent of the clone λMargo10 is shown; the position of the P element is indicated in λMargo10. The EcoRI subclones from λMargo10, M10-2 and M10-18, are also depicted below λMargo10. The cDNA encoding RM62 is also indicated and shows the positions and sizes of the two introns described in the text. The polymorphic XhoI site referred to in the text is marked with an asterisk.

Characteristics of the protein encoded by the cDNA

The one large open reading frame in the cDNA sequence translates into a protein of 62kD with several unusual properties. Of the first 81 amino acids, 38 are glycines, present in runs of up to 7 glycine residues in a row. This region is also highly charged, with a total of 28 charged residues, including 14 arginines and 9 aspartic acid residues. At the carboxyl end, the same pattern occurs; of the final 49 amino acids, 27 are glycine, and occur in repeats of up to 8 glycines. The carboxyl region is also charged, although not as much as the amino end. Between these glycine-rich and highly charged domains at both ends of the protein is a central core which is homologous to a family of RNA helicases. For comparison, the sequences of the *D. melanogaster* RM62 protein, the homologous *D. pseudoobscura* protein (see below), and three other members of the family are presented in Figure 3. The other members shown are the human nuclear antigen p68 (10) which is the most similar member of

```
D. m.                                MA PHDRDFGHSG RGGRGGDRGG DDRRGGGGGG  32
D. p.                                MA PHDRDFGQNS RGGRDGRGGG EDRRGGGGGG
vasa MSDDWDDEPI VDTRGARGGD WSDDEDTAKS FSGEREGDGV GGSGGEGGGY QGGNRDVFGR
-------------------------------------------------------------------------
D. m. -----NRFGG GGGGGDYHGI RNGAVEKRRD DRGGG----- NRFGGGGGF- GDRRGGGGGG  81
D. p. GGGGGSRFGA GGGGGDYHGV RNGRIEKRRD DRGGGGGGGG GRLAGGGGFG GDRRGGGG--
p68                                              FGGSRAG PLSGK----K
vasa IGGGRGGGAG GYRGGNRDGG GFHGGRREGE RDFRGGEGGF RGGQGGSRGG QGGSRGGQGG
-------------------------------------------------------------------------
D. m. ----SQDLPM RPVDFSNLAP FKKNFYQEHP NVANRSPYEV QRYREEQ--- -EITVRGQV- 132
D. p. ----SQDLPM RPVDFSNLTP FKKNFYQEHA TVAARSPYEV QRYRDEH--- -EITVRGQA-
p68  FGNPGEKLVK KKWNLDELPK FEKNFYQEHP DLARRTAQEV ETYRRSK--- -EITVRGHNC
vasa FRGGEGGFRG RLYENEDGDE RRGRLDREER GGERRGRLDR EERGGER(46)IPVKVTGSDV
eIF4A                                                (17)EPEGVIESNW
ID          L   R     L   F KNFYQEH    A R   EV   YR E     EITVRG
-------------------------------------------------------------------------
D. m. PNPIQDFSEV HLPDYVMKEI RRQGYKAPTA IQAQGWPIAM SGSNFVGIAK TGSGKTLGYI 192
D. p. QNPIQDFGEV YLPEYVTKEI RRQGYKEPTP IQAQGWPIAM SGANFVGIAK TGSGKTLGYI
p68  PKPVLNFYEA NFPANVMDVI ARQNFTEPTA IQAQGWPVAL SGLDMVGVAQ TGSGKTLSYL
vasa PQPIQHFTSA DLRDIIIDNV NKSGFKIPTP IQKCSIPVIS SGRDLMACAQ TGSGKTAAFL
eIF4A NEIVDSFDDM NLSESLLRGI YAYGFEKPSA IQQRAILPCI KGYDVIAQAQ SGTGKTATFA
ID     PIQ F E   LP  V   I  RQG K PT  IQAQGWP A  SG   VG A  TGSGKTL Y
-------------------------------------------------------------------------
D. m. LPAIVHINNQ QPLQRGDGPI ALVLAPTREL AQQIQQVATE FGSSSYVRNT CVFGGAPKGG 252
D. p. LPAIVHINNQ QPLQRGDGPI ALVLAPTREL AQQIQQVATE FGSSSYVRNT CVFGGAPKGG
p68  LPAIVHINHQ PFLERGDGPI CLVLAPTREL AQQVQQVAAE YCRACRLKST CIYGGAPKGP
vasa LPILSKLLED PHELELGRPQ VVIVSPTREL AIQIFNEARK FAFESYLKIG IVYGGTSFRH
eIF4A ISILQQI--- --ELDLKATQ ALVLAPTREL AQQIQKVVMA LGDYMGASCH ACIGGTNVRA
ID   LPAIVHIN Q   L RGDGPI  LVLAPTREL AQQIQQVA E     SY   T CV GGAPKG
-------------------------------------------------------------------------
D. m. QMRDLQRGC- EIVIATPGRL IDFLSAGSTN LKRCTYLVLD EADRMLDMGF EPQIRKIVS- 310
D. p. QMRDLQRGC- EIVIATPGRL IDFLSSGGTN LKRCTYLVLD EADRMLDMGF EPQIRKIVS-
p68  QIRDLERGV- EICIATPGRL IDFLECGKTN LRRTTYLVLD EADRMLDMGF EPQIRKIVD-
vasa QNECITRGC- HVVIATPGRL LDFVDRTFIT FEDTRFVVLD EADRMLDMGF SEDMRRIMTH
eIF4A EVQKLQMEAP HIIVGTPGRV FDMLNRRYLS PKYIKMFVLD EADEMLSRGF KDQIYDIFQ-
ID   Q RDLQRGC  EIVIATPGRL IDFL G T N LKR TYLVLD EADRMLDMGF EPQIRKIV
-------------------------------------------------------------------------
D. m. -QIRPDRQTL MWSATWPKEV KQLAEDFLGN YIQINIGSLE LSANHNIRQV VDVCDEFSKE 369
D. p. -QIRPDRQTL MWSATWPKEV KQLAEDFLGN YIQINIGSLE LSANHNIRQV VEVCDEFSKE
p68  -QIRPDRQTL MWSATWPKEV RQLAEDFLKD YIHINIGALE LSANHNILQI VDVCHDVEKD
vasa VTMRPEHQTL MFSATFPEEI QRMAGEFLKL R-FVAIGIVG GRCSDVKQTI YEVNK-YAKR
eIF4A -KLNSNTQVV LLSATMPSDV LEVTKKFMRD PIRILVKKEE LTLEGIRQFY INVEREEWKL
ID    QIRPDRQTL MWSATWPKEV  QLAEDFL    I INIG LE LSANHNI Q  V VC E  K
-------------------------------------------------------------------------
D. m. EKLKTLLSDI YDTSESPGKI IIFVETKRRV DNLVRFIRSF GVRCGAIHGD KSQSERDFVL 429
D. p. EKLKSLLSDI YDTSENPGKI IIFVETKRRV DNLVRFIRSF GVRCGAIHGD KSQSERDFVL
p68  EKLIRLMEEI MSEKEN--KT IVFVETKRRC DELTRKMRRD GWPAMGIHGD KSQQERDWVL
vasa SKLIEILSE- -QADG----T IVFVETKRGA DFLASFLSEK EFPTTSIHGD RLQSQREQAL
eIF4A DTLCDLYE-- -TLTITQ--A VIFINTRRKV DWLTEKMHAR DFTVSAMHGD MDQKERDVIM
ID   EKL  LL  I     E   K  IIFVETKRRV D L RF R   G      IHGD KSQSERD VL
-------------------------------------------------------------------------
D. m. REFRSGKSNI LVATDVAARG LDVDGIKYVI NFDYPQNSED YIHRIGRTGR SNTKGTSFAF 489
D. p. REFRSGKSNI LVATDVAARG LDVDGIKYVI NFDYPQNSED YIHRIGRTGR SNTKGTSFAF
p68  NEFKHGKAPI LIATDVASRG LDVEDVKFVI NYDYPNSSED YIHRIGRTAR STKTGTAYTF
vasa RDFKNGSMKV LIATSVASRG LDIKNIKHVI NYDMPSKIDD YVHRIGRTGC VGNNGRATSF
eIF4A REFRSGSSRV LITTDLLARG IDVQQVSLVI NYDLPTNREN YIHRIGRGGR FGRKGVAINM
ID   REFRSGKS I L ATDVAARG LDV  IK VI N DYP NSED YIHRIGRTGR S  KGT   F
-------------------------------------------------------------------------
D. m. FTKNNAKQAK ALVDVLREAN QEINPALENL ARNSRYDGGG GPSRYGGGGG GG-RFGGGGF 548
D. p. FTKNNAKQAK ALV-VLREAN QEINPALENL ARNSRYDGGG GASRYGGGGG GGGRFGGGGF
p68  FTPNNIKQVS DLISVLREAN QAINPKLLQL VEDRGSGRSR GRGGMKDDRR DRYSAGKRGG
vasa FDPEKDRAIA ADLVKILEGS GQTVPDFLRT CGAGGDGGYS NQNFGGVDVR GR--GNYVGD
eIF4A VTEEDKRTLR DIETFYNTSI EEMPLNVADL I
ID   FT NN KQ   AL   VLREAN QEINP L  L
-------------------------------------------------------------------------
D. m. KKGSLSNGRG FGGGGGGGG- -EG-RHSRFD                                  575
D. p. KKGRLSNGRG FGGGGGGGGG GDGGRHTRFD
p68  FNTFRDRENY DRGYSSLLKR DFGAKTQNGV(80)
vasa ATNVEEEEQW D
```

Figure 3. RNA helicase sequences. A comparison of the protein sequences of the RM62 gene from *D. melanogaster* (D. m.) and the homologous gene from *D. pseudoobscura* (D. p.) with the human nuclear antigen p68 (10), the *D. melanogaster vasa* gene (16), and the mouse eIF4AI gene (39). Numbers in parentheses indicate amino acid residues not shown: 46 amino acids from within *vasa*, 17 amino acids from the N-terminus of eIF4AI, and 80 amino acids from the C-terminus of p68. The row labelled ID lists amino acids present in both the *D. melanogaster* and *D. pseudoobscura* RM62 proteins and also present in at least one of the other proteins, up to the carboxy terminus of eIF4AI. Bold face indicates that an amino acid is present in all 5 proteins.

the family, the *D. melanogaster vasa* protein (12, 16), which is the only other member of the family described in *Drosophila*, and the mouse translation initiation factor eIF4AI (39). Many striking similarities can be seen, particularly the motifs described previously for the RNA helicase family (11, 13, 15, 19). Among these are the sequences PT(A/P)IQ at 160−164, A−TGSGKTL at 181−189, PTRELAQQIQQ at 218−228, TPGR−−D at 267−273, VLDEADRML (the DEAD box− (19)) at 289−297, SAT-P at 322−326, and YIHRIGR at 470−476. The *D. melanogaster* RM62 protein is 66% identical to the human p68 protein between residues 95 and 520. Inspection of the sequences reveals other identities and conservative amino acid replacements between the *Drosophila* proteins and the other members of the family, especially p68.

A cDNA library generated from *D. pseudoobscura* (40) was also screened using the *D. melanogaster* cDNA clone as a probe. A homologous clone was obtained, subcloned and sequenced. The open reading frame found in the *D. pseudoobscura* cDNA clone is very similar to the one found in *D. melanogaster*. Over the entire length of the protein, the sequence is 91% identical between *D. melanogaster* and *D. pseudoobscura*. However, most of the changes are at the termini of the molecule; the central two-thirds of the protein is 98% identical between the two species. The evolutionary conservation of this sequence between the two species (divergence time ~45MYr) is similar to the extent of conservation seen between other genes from these species (41). The glycine-rich motifs at the two ends of the protein are conserved in terms of their general sequence, but the lengths of the glycine runs vary between the two species. This suggests that the overall structure conferred by the glycine repeats and the charged residues is important, but the lengths of the glycine repeats can vary within limits without affecting the function of the protein.

Gene structure

Comparison of partial DNA sequence from the wild type cosmid clones with the cDNA sequence revealed two introns. The first intron interrupts the initiator ATG codon, as indicated in Figure 4. Restriction analysis of the cDNA and the genomic DNA shows that this intron is approximately 2.9kb long. Further DNA sequencing revealed a second intron (tentatively labelled Intron 2) of 107 nucleotides. This intron interrupts the codon for Lysine 371. This corresponds exactly to the location of the only sequenced intron-exon junction of the human p68 gene (42). The sequences of the splice junction regions of Intron 2 and of the corresponding p68 intron are also shown in Figure 4. Iggo et al (1989) do not report the sequence of the 3′ splice junction of p68. The first 6 nucleotides of the intron (including the consensus GT) are also identical in RM62 and p68. Conservation of an intron location for 600 million years (the approximate time since the divergence of vertebrates from invertebrates) suggests a common ancestral origin of both RM62 and p68. The splice junctions and polypyrimidine tracts of both introns are in agreement with the consensus sequences for introns (43). As the genomic sequence is not complete, there may be other, as yet undetected, introns in the gene.

```
Intron 1:    GAAGGAGTAATgtgagtacg...(~2.9kB)...tttcccagGGCACCACAC
                    Me                                    tAlaProHis

Intron 2:    AAGGAGGAGAAgtaagtgga... (77bp) ...caatacagATTGAAGACC
             LysGluGluLy                                sLeuLysThr

p68 intron:  AAGGATGAAAAgtaagtttt...(~370bp)...nnnnnnnnNNNNNNNNNN
             LysAspGluLy                                sLeuIleArg
```

Figure 4. Sequence of intron-exon junctions in RM62 and human p68. The DNA sequences derived from genomic and cDNA clones are given in the rows labelled Intron 1 and Intron 2. Only the coding strand is shown. Exon sequences are listed in upper case, introns in lower case. The amino acid sequence is listed below the DNA sequence. The numbers in parentheses indicate the total length of the intron. The third line gives the sequence of an intron found in the human p68 gene (42). The 'N's indicate that the DNA sequence for that region has not been reported, although the protein sequence shown below it has been deduced from the unpublished cDNA sequence and reported (10).

DNA sequencing of the M10−2 and M10−18 subclones (see Figure 2) using oligonucleotide primers derived from the P element sequence (37), and primers derived from the cDNA sequence allowed us to locate the site of insertion of the P element in Tpl^{J34}. The P element is located in Intron 1, 77 nucleotides downstream from the 5′ splice site of the intron. The P element is oriented in such a way that the transposase gene is transcribed in the same direction as RM62. The addition of the 2.9kb P element to Intron 1 makes this intron approximately 5.8kb in the mutant gene.

Analysis of RNA transcripts

RNA was extracted from flies at various stages of development and transferred to a Northern blot. Figure 5 shows the blot hybridized with the complete RM62 cDNA clone. The lanes all contain poly(A)+ selected mRNA except the one labeled total adult, which contains total RNA from adults. Transcripts are seen at all stages of development, including tissue culture cells. The major band is just smaller than 2.9kb although there are other longer forms visible, which may represent the use of the

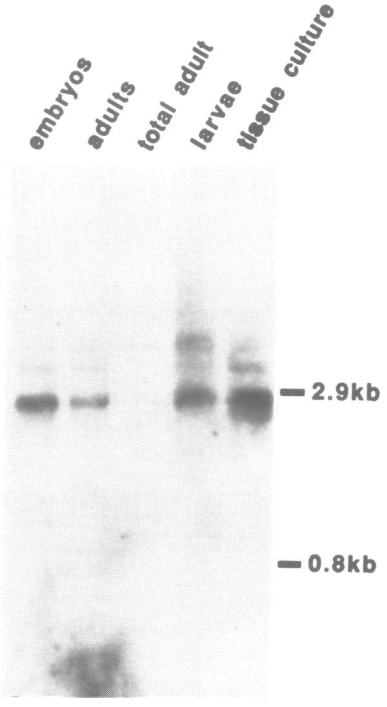

Figure 5. Analysis of transcripts. RNA was extracted from 0−12 hour embryos, third instar larvae, adults, and Schneider Line 2 tissue culture cells and Northern blotted. All lanes contain 10μg poly(A)+ RNA except the lane labelled total adult, which contains 10μg of total RNA. The blot was hybridized to a ^{32}P-labelled RM62 cDNA clone. The position of DNA markers loaded on the same gel is also shown.

alternative polyadenylation sites seen in some of the cDNA clones. The pattern of transcript sizes is qualitatively the same in all stages of development and in cultured *Drosophila* cells.

Mapping *Tpl* mutations to the region

Subsequent to the cloning work described above, we obtained data suggesting that the P element present in Tpl^{J34} may not be responsible for the mutant phenotype. Since some *Tpl* hypomorphic alleles have proven to be hotspots for homologous recombination (6) we looked for recombinants between the ri^+ Tpl^{J34} Ki^+ p^+ chromosome present in the mutant stock and a marked tester chromosome, *ri* Tpl^+ *Ki* p^p. An ri^+ Tpl^{J34} *Ki* p^p recombinant was obtained, as previously described (5). This chromosome, which is indistinguishable in its *Tpl* phenotype from the Tpl^{J34} parent chromosome, was shown by *in situ* hybridization to be missing the P element (5). In addition, we have now obtained the reciprocal recombinant, *ri* Tpl^+ Ki^+ p^+, which is phenotypically wild type at *Tpl*, and which still has the P element insertion. We also isolated a spontaneous revertant, Tpl^{J34R}, which still has the P element intact. The *in situ* hybridizations have also been confirmed by genomic Southern blotting (data not shown). We now have three independently derived lines in which there is no correlation between the presence of the P element at 83E1,2 and the hypomorphic phenotype of Tpl^{J34}. This confirms and extends our previous finding that the mutant alleles of *Tpl* that were recovered following hybrid dysgenesis are not due to P element insertions (5). The recombination data show that the actual lesion that causes the Tpl^{J34} mutation is centromere-proximal to the P element. We have also analyzed the recombinant chromosomes obtained in our earlier study (6), and find that the lesions causing the hypomorphic alleles Tpl^{10} and Tpl^{17} are centromere-proximal to a polymorphic XhoI site which is located 14kb distal to the RNA helicase transcription unit (see Figure 2). We are in the process of further analyzing our collection of deletions, duplications, and hypomorphic alleles by Southern blotting in order to place them on the molecular map.

DISCUSSION

While aneuploidy for whole chromosome arms is severely deleterious, in general this is due to the additive effects of aneuploidy for all of the genes within the duplicated or deleted region. The only known case of a single locus being lethal when present in non-diploid copy number is *Tpl* (1). It is interesting to note that trisomics for the×chromosome (metafemales) and for chromosome arm 2L often survive much longer than flies carrying three copies of *Tpl* (44−46). One possible explanation for why dosage alterations of *Tpl* are so severe is that *Tpl* is involved in such a fundamental cellular process that disruption of that process has pleiotropic effects on expression of the entire genome. Some aspect of RNA metabolism, such as splicing, turnover or localization, might be such a process, so the discovery of an RNA helicase gene in the region of *Tpl* is therefore intriguing.

The genetics of *Tpl* suggest that it is a complex locus, possibly involving functionally redundant genes. RM62 by itself can not explain the complex genetics of *Tpl*. Functional redundancy of several RNA helicases including RM62 could explain the difficulty in generating point mutations at *Tpl*, since eliminating only one member of the group would have a minimal effect compared to deleting all of them at once. Duplications of *Tpl*, being rather large in molecular terms, would presumably duplicate the entire group of RNA helicases simultaneously. In a different context, Tartof and coworkers point out that extreme gene dosage sensitivity is easily explained by mass action if the gene products are involved in an assembly driven process (47). Interestingly, one of the members of the RNA helicase family, the E. coli *srmB* gene, appears to be involved in ribosome assembly, and shows a dosage-dependent interaction with mutations in a ribosomal protein gene (21, 48).

The cellular role of RM62 is not obvious from the protein sequence. For some members of the RNA helicase family, genetic or biochemical analysis has allowed a function to be postulated (12, 16, 23, 39). The most similar member of the family to RM62 is the human nuclear antigen p68, and although p68 has been shown to be an RNA-dependent ATPase, and an ATP-dependent RNA helicase (14, 49), the role of p68 in the cell is unclear. It is also important to note that even though p68 may be the most similar to RM62, there are significant differences between the sequences of p68 and RM62 that make it impossible to draw direct correlations between them. For example, a carboxy-terminal epitope shared between p68 and SV40 Large T antigen is believed to be part of an important functional domain (10), but it is not conserved in RM62. The striking sequence identities between p68 and RM62, and the identical intron position described above, leave little doubt that these proteins have a common origin. However the equally striking differences between them at the amino and carboxyl termini open the question of whether there is a different *Drosophila* gene that is the functional equivalent of p68, and a homologous human protein that is the equivalent of RM62.

The glycine-rich motifs at both ends of RM62 are not found in p68 nor any of the other RNA helicases, nor did we find any similar sequences in the GenBank or EMBL databases. While a number of proteins have superficially similar glycine-rich domains (50−52), none of them have the large number of both acidic and basic residues within the glycine-rich domain as RM62 does. In general, glycine-rich domains are posited to confer flexibility on the polypeptide chain, perhaps forming 'hinges' between domains (53). In the cytokeratins, glycine-rich regions may be important in chain-chain associations for assembly into intermediate filaments (54). The highly charged nature of the glycine-rich domains of RM62, and the fact that they occur at both ends of the molecule lead us to suggest that they are extremely flexible 'tails' that extend away from the RNA helicase core of the protein into the solvent. If these tails interact with other proteins, such as cytoskeletal components, or allow RM62 to polymerize, the flexibility conferred by the glycine runs may allow the protein to be oriented independently or to swivel, which might be important in allowing long double-helical RNA molecules to be unwound or translocated by it. Further work on the RM62 protein and *Tpl* will help to clarify the roles of RNA helicases in cell biology, and may provide a model system for the study of lethal gene interactions and gene dosage effects.

ACKNOWLEDGEMENTS

We are grateful to Marilyn Cadden and Helene Klapper for skilled technical assistance, to Jonathan Schiffman for isolating the Tpl^{J34} revertants, to Tom Kornberg for the cDNA libraries, to Mark Champe and Charles Laird for the cosmid library, and to Steve Henikoff for pointing out the vertical DEAD box at position 146. This work was supported in part by grants from the National

Institutes of Health to A.C.C. (GM38483) and D.H.J. (HG00081), and a predoctoral fellowship to D.R.D. from the Percival E. and Ethel Brown Foerderer Foundation.

REFERENCES

1. Lindsley, D.L., Sandler, L., Baker, B.S., Carpenter, A.T.C., Denell, R.E., Hall, J.C., Jacobs, P.A., Miklos, G.L.G., Davis, B.K., Gethmann, R.C., Hardy, R.W., Hessler, A., Miller, S.M., Nozawa, H., Parry, D.M. and Gould-Somero, M. (1972) *Genetics*, **71**, 157–184.
2. Denell, R.E. (1976) *Genetics*, **84**, 193–210.
3. Keppy, D.O. and Denell, R.E. (1979) *Genetics*, **91**, 421–441.
4. Roehrdanz, R.L. and Lucchesi, J.C. (1980) *Genetics*, **95**, 355–366.
5. Dorer, D.R. and Christensen, A.C. (1990) *Genetics*, **125**, 795–801.
6. Dorer, D.R. and Christensen, A.C. (1989) *Genetics*, **122**, 397–401.
7. Bingham, P.M., Levis, R. and Rubin, G.M. (1981) *Cell*, **25**, 693–704.
8. Johnson, D.H. (1990) *Genomics*, **6**, 243–251.
9. Chang, T.-H., Arenas, J. and Abelson, J. (1990) *Proc. Natl. Acad. Sci. USA*, **87**, 1571–1575.
10. Ford, M.J., Anton, I.A. and Lane, D.P. (1988) *Nature*, **332**, 736–738.
11. Gorbalenya, A.E., Koonin, E.V., Donchenko, A.P. and Blinov, V.M. (1988) *Nature*, **333**, 22.
12. Hay, B., Jan, L.Y. and Jan, N.Y. (1988) *Cell*, **55**, 577–587.
13. Hodgman, T.C. (1988) *Nature*, **333**, 22–23.
14. Iggo, R.D. and Lane, D.P. (1989) *EMBO J.*, **8**, 1827–1831.
15. Lane, D.P. (1988) *Nature*, **334**, 478.
16. Lasko, P.F. and Ashburner, M. (1988) *Nature*, **335**, 611–617.
17. Leroy, P., Alzari, P., Sassoon, D., Wolgemuth, D. and Fellous, M. (1989) *Cell*, **57**, 549–559.
18. Linder, P. and Slonimski, P.P. (1988) *Nucleic Acids Res.*, **16**, 10359.
19. Linder, P., Lasko, P.F., Ashburner, M., Leroy, P., Nielsen, P.J., Nishi, K., Schnier, J. and Slonimski, P.P. (1989) *Nature*, **337**, 121–122.
20. Linder, P. and Slonimski, P.P. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 2286–2290.
21. Nishi, K., Morel-Deville, F., Hershey, J.W.B., Leighton, T. and Schnier, J. (1988) *Nature*, **336**, 496–498.
22. Sachs, A.B. and Davis, R.W. (1990) *Science*, **247**, 1077–1079.
23. Seraphin, B., Simon, M., Boulet, A. and Faye, G. (1989) *Nature*, **337**, 84–87.
24. Lindsley, D.L. and Grell, E.H. (1968) In: Genetic Variations of *Drosophila melanogaster*. Carnegie Institution of Washington, Washington, D.C.
25. Scott, M.P., Weiner, A.J., Hazelrigg, T.I., Polisky, B.A., Pirrotta, V., Scalenghe, F. and Kaufman, T.C. (1983) *Cell*, **35**, 763–776.
26. Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) In: Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
27. Feinberg, A.P. and Vogelstein, B. (1984) *Anal. Biochem.*, **137**, 266–267.
28. Karess, R.E. and Rubin, G.M. (1984) *Cell*, **38**, 135–146.
29. Poole, S.J., Kauvar, L.M., Drees, B. and Kornberg, T. (1985) *Cell*, **40**, 37–43.
30. Thomas, P.S. (1983) *Methods Enzymol.*, **100**, 255–266.
31. Henikoff, S. and Eghtedarzadeh, M.K. (1987) *Genetics*, **117**, 711–725.
32. Henikoff, S. (1984) *Gene*, **28**, 351–359.
33. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
34. Pearson, W.R. and Lipman, D.J. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 2444–2448.
35. Henikoff, S. and Wallace, J.C. (1988) *Nucleic Acids Res.*, **16**, 6191–6204.
36. Roehrdanz, R.L. and Lucchesi, J.C. (1981) *Dev. Genet.*, **2**, 147–158.
37. O'Hare, K. and Rubin, G.M. (1983) *Cell*, **34**, 25–35.
38. Steller, H. and Pirrotta, V. (1985) *EMBO J.*, **4**, 167–171.
39. Nielsen, P.J. and Trachsel, H. (1988) *EMBO J.*, **7**, 2097–2105.
40. Johnson, D.H. and Henikoff, S. (1989) *Mol. Cell. Biol.*, **9**, 2220–2223.
41. Blackman, R.K. and Meselson, M. (1986) *J. Mol. Biol.*, **188**, 499–515.
42. Iggo, R., Gough, A., Xu, W., Lane, D.P. and Spurr, N.K. (1989) *Proc. Natl. Acad. Sci. USA*, **86**, 6211–6214.
43. Mount, S.M. (1982) *Nucleic Acids Res.*, **10**, 459–472.
44. Birchler, J.A., Hiebert, J.C. and Krietzman, M. (1989) *Genetics*, **122**, 869–879.
45. Devlin, R.H., Holm, D.G. and Grigliatti, T.A. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 1200–1204.
46. Neeley, J.C. (1971) *J. Hered.*, **62**, 334–338.
47. Locke, J., Kotarski, M.A. and Tartof, K.D. (1988) *Genetics*, **120**, 181–198.
48. Nishi, K. and Schnier, J. (1988) *Molec. Gen. Genet.*, **212**, 177–181.
49. Hirling, H., Scheffner, M., Restle, T. and Stahl, H. (1989) *Nature*, **339**, 562–564.
50. Haynes, S.R., Rebbert, M.L., Mozer, B.A., Forquignon, F. and Dawid, I.B. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 1819–23.
51. Haynes, S.R., Raychaudhuri, G. and Beyer, A.L. (1990) *Mol. Cell. Biol.*, **10**, 316–323.
52. Schimmang, T., Tollervey, D., Kern, H., Frank, R. and Hurt, E.C. (1989) *EMBO J.*, **8**, 4015–4024.
53. Beachy, P.A., Helfand, S.L. and Hogness, D.S. (1985) *Nature*, **313**, 545–551.
54. Steinert, P.M., Rice, R.H., Roop, D.R., Trus, B.L. and Steven, A.C. (1983) *Nature*, **302**, 794–800.