The EMBO Journal vol.8 no.11 pp.3437 - 3446, 1989

Molecular analysis of *ref(2)P*, a *Drosophila* gene implicated in sigma rhabdovirus multiplication and necessary for male fertility

Sybille Dezelee, Françoise Bras, Didier Contamine, Miguel Lopez-Ferber, Dominique Segretain and Danielle Teninges

Laboratoire de Génétique des Virus, CNRS, 91198 Gif sur Yvette Cedex, France

Communicated by M.Ashburner

The *ref(2)P* gene of *Drosophila melanogaster* is implicated in sigma rhabdovirus multiplication. A permissive allele was cloned and sequenced. The structural gene (3.1 kbp) is divided into three exons. The mRNAs are heterogeneous in size. They differ only in the 5′ end of the first exon. The sequence upstream of the short mRNAs contains classical promoter elements. No TATA and CAAT boxes are appropriately positioned upstream of the initiation sites of the long mRNAs, but several repeats, palindromic sequences and inverted CAAT boxes are present. These observations, together with the tissue-dependent distribution of short and long transcripts, support the hypothesis of the existence of at least two classes of genuine initiation sites. The long size of the untranslated leader RNA region suggests a control of gene expression at the translation level. The same translation product of 599 amino acids (76.3 kd) is predicted for all mRNAs, but the *in vitro* translation product migrates in SDS-PAGE with a higher apparent mol. wt (115-125 kd). The putative *ref(2)P* protein contains internal repeats, PEST regions which may be signals for protein degradation, and interesting structural motifs such as zinc finger and amphiphilic helices. These later motifs could be mitochondrial pre-sequences. The degeneration of mitochondria is observed in the spermatids of sterile male flies homozygous for the loss-of-function alleles. The amino acid sequence of the *ref(2)P* product shows no homology with any known protein from the data banks.

Key words: Drosophila ref(2)P gene/rhabdovirus sigma/male sterility/transcript mapping/translation product

Introduction

The study of genetic resistance to viral infections in eukaryotes may provide an insight not only into the detailed interactions between virus and host products but also into the identification of host gene's functions that have as yet escaped other screening procedures. The *Drosophila* genes that interfere with the multiplication of sigma virus belong to this category. They were identified by the restrictive effect of some of their alleles on sigma virus multiplication and thus were named refractory (*ref*) genes.

Sigma virus is a negative strand RNA virus belonging to the rhabdovirus family (Brun and Plus, 1980; Teninges and Bras-Herreng, 1987). It is generally non-pathogenic to its host in ordinary conditions, but infected flies become irreversibly paralysed upon exposure to high concentrations of carbon dioxide. Such a symptom is also produced by other rhabdoviruses in several dipteran species and has been correlated with viral multiplication in the central nervous system of these insects (Bussereau, 1975; Rosen, 1980). The only mode of propagation known for sigma virus in nature is the transgametal pathway (without chromosomal integration of the viral genome) (Brun and Plus, 1980). Embryogenesis is not disturbed by the presence of viral material in the eggs, implying that viral syntheses have to be kept within limits compatible not only with the host cell survival (as in most persistent infections), but also with the development pattern of the host organism. Such restrictions may be under viral and/or host control.

The different *ref* genes currently known to exist map to five different loci (Gay, 1978). The most extensively studied of these genes is *ref(2)P*, which maps to region 37 EF of the *Drosophila melanogaster* chromosomes (Nakamura *et al.*, 1986). Two classes of *ref(2)P* alleles have been defined in terms of their effect on the sigma virus: restrictive $ref(2)P^{P}$ and permissive $ref(2)P^{O}$ (Nakamura, 1978). They are found in natural populations (Fleuriet, 1976). The meaning of their distribution is still unknown as it seems independent of any selective pressure exerted by sigma virus infection. Several loss-of-function alleles [$ref2(P)^{null}$] were induced either by X-rays, chemical agents or transposon mutagenesis (Contamine *et al.*, 1989). All homozyous $ref(2)P^{null}$ flies are viable, but males are sterile. Females are not affected though the gene is also expressed in their germ cells, as seen by the effects of restrictive alleles on the transovarian transmission of the virus (Gay, 1978).

The study of sigma virus multiplication in the different genetic combinations allowed with the three types of alleles showed: (i) a dosage effect of the restrictive allele; (ii) a co-dominance of the restrictive and permissive alleles and (iii) an equivalence of the permissive and null alleles (Nakamura, 1978; Contamine *et al.*, 1989). These results suggest that the amount of the restrictive product defines the intensity of restriction and that the permissive product might not be indispensable for efficient viral syntheses, while it is able to compete with the restrictive product.

A permissive allele of this gene has been cloned and the transcription unit has been defined. This gene is transcribed into two poly(A)$^+$ mRNAs of ~2300 and 2400 nucleotides. Males contain roughly equal amounts of both, while females contain an excess of the longest species which is the only species found in ovaries (Contamine *et al.*, 1989).

In this report, we analyse the sequence of the gene, the structure of its diverse transcripts and we show that their heterogeneity results from a differential use of promoters. We examine the properties of the predicted protein and the morphological defects in the $ref(2)P^{null}$ sterile males.

Results

Ref(2)P gene sequencing

The *ref(2)P* locus was located by breakpoint mapping in 37E3-F3 and cloned DNAs containing the structural part of the gene were identified by P transposon tagging (Contamine *et al.*, 1989). To define the structure and the size of the *ref(2)P* transcription unit, we subcloned a 4.7 kbp DNA fragment which includes the two original *Eco*RI fragments containing the *ref(2)P* structural gene and flanking regions.

Using the strategy explained in Materials and methods, both strands were sequenced (Figure 1). A large open reading frame (ORF) of 1533 nucleotides started in position +1141; no long (ORF) was found either in the two other frames or in the three others on the opposite strand which was the only strand hybridizing with the *ref(2)P* transcripts.

The intron–exon structure of the ref(2)P gene

S1 nuclease mapping was performed as described schematically in Figure 2. On the autoradiograms of the neutral gels, the *ref(2)P* transcripts were revealed with the A, B and C probes, but not with probes on the left of the A probe or on the right of the C probe (results not shown). The autoradiograms of the alkaline agarose gels (Figure 2, lane 1) showed that the region corresponding to the A probe contained two exons (E1L and E1S) ~450 and 350 nucleotides in length. A large exon E2 (~1500 nucleotides) was revealed with both B and C probes and a short exon E3 (~400 nucleotides) with the C probe.

Digestion of the *ref(2)P* genomic DNA with different restriction enzymes before hybridization with poly(A)$^+$ RNAs allowed to approximately locate the four types of exons on the genomic map. The long E2 exon and the short E3 exon were located using *Bgl*I restriction sites (Figure 2, lane 2). The positions of the E1L and E1S exons were determined by using additional restriction sites (for example, pre-digestion with *Sac*I resulted in a common protected fragment for both exons) and by using the *Eco*RI–*Sac*I probe which revealed only the longest *ref(2)P* transcripts on non-denaturing agarose gels and only the E1L exon on alkaline agarose gels. These results showed that the two *ref(2)P* transcripts were co-linear and had two exons in common (E2 and E3). The E2 exon was connected either to E1L or to E1S, which seemed to differ at their 5′ end.

Intron–exon boundaries and the 3′ end of transcripts were precisely determined by hybridizing the poly(A)$^+$ RNAs to uniformly labelled single-stranded DNA probes and analysing the size of the protected fragments after S1 nuclease digestion on sequencing gels. Relevant experiments are shown in Figure 3. For each extremity mapped, the protected fragments differed only by a few nucleotides. This presumably reflected 'breathing' at the end of the double-stranded domains consisting of A–T or A–U pairs. Messenger RNAs, thus, seemed homogeneous for their 3′ end and splice junctions. These junctions were located on the genomic sequence (Figure 1) and were confirmed by sequencing three independent cDNA clones. Maturation of *ref(2)P* mRNAs results in the splicing of two introns, 631 and 55 nucleotides in length, allowing the connection of three open reading frames: one of 252 nucleotides in the E1 exon, a second which covers the entire E2 exon (1481 nucleotides) and the last one of only 142 nucleotides in the E3 exon. The splice sites fit the intron donor and acceptor consensus sequences (Mount, 1982) and each intron contains, 30–40 nucleotides upstream of its 3′ end, sequences related to the CTAAT sequence of the lariat branchpoints described by Keller and Noon (1984).

The 3′ end of the *ref(2)P* transcripts was found in position +3089 by S1 nuclease mapping, but all the cDNA clones ended in +3104. This discrepancy probably results from overdigestion of hybrids by S1 nuclease due to the high AT content of this region. The transcription termination site (+3104) is 26 bp downstream of an AATAAA box and 24 bp upstream of a GT cluster (Figure 1), as usually found (Birnstiel *et al.*, 1985). Other AATAAA boxes occur in the sequence, but apparently they are not used as termination signals.

The initiation sites of the ref(2)P transcripts

Three different methods were used to determine the 5′ end of the E1 exons: S1 and *Exo*VII nuclease mapping and primer extension. Results are summarized in Figure 4. After nuclease digestions, multiple protected fragments were observed; the smallest corresponded to the expected E1S short exon, the fragment ~120 nucleotides longer to the E1L exon; in addition, other fragments ~170 and 200 nucleotides longer than E1S were present, but in much smaller amounts (these fragments account for the smear observed above the E1L band in Figure 2A). The same distribution was observed after S1 nuclease or exonuclease VII digestion, but with exonuclease VII the fragment sizes were always 10–20 nucleotides longer. These results could be explained both by an excessive exolytic degradation by S1 nuclease and by the incapacity of exonuclease VII to give perfect blunt ends.

The primer extension method also gave numerous extended products (Figure 4) which placed the initiation points between the positions determined by S1 nuclease and exonuclease VII mapping. The position of the longest fragment (ssL) was defined as nucleotide +1. This fragment was followed by a set of three extended products in positions +19, +22, +25 (sL) and the main products in +69, +72 (L) and +195, +198 (S). In the longest cDNA clone which starts in position +25, the sequence of the first exon is in complete agreement with the genomic sequence. Thus, until position +25, the long fragments obtained by primer extension are not due to the connection of upstream sequences to the E1 exons. A set of additional products +49, +52, +55 and other shorter fragments had no counterpart in the nuclease mapping experiments. Examination of the nucleotidic sequence on both sides of these extra-potential

Fig. 1. Nucleotide sequence of the *ref(2)P* gene and its predicted protein. The transcription startpoints determined in primer extension experiments are indicated by arrowheads. Nucleotide position +1 is defined by the startpoint of the longest mRNA. The three exons are underlined. The presumptive TATA and CAAT boxes upstream of the short transcripts are boxed as well as the AATAAA sequence presumably used as the polyadenylation signal, the GT cluster which follows and the lariat branchpoint sequences in the introns. The three GTAAAAATA repeated motifs in the region upstream of the long transcript startpoints are underlined twice. The open reading frame (ORF) of 1878 nucleotides starts in position +291 and ends with a TAA stop codon in position +2855. The deduced amino acid sequence of the presumed *ref(2)P* protein starts with the first MET-codon (numbered 1) in position +372 on the nucleotide sequence, and is 599 amino acids long.

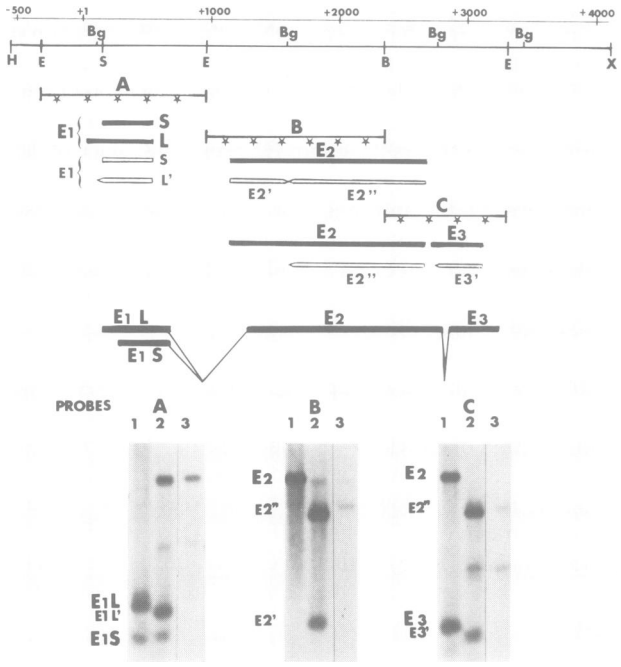

Fig. 2. Structure of *ref(2)P* transcripts. Poly(A)$^+$ RNAs from *Drosophila* females (6 μg) were hybridized to the plasmid DNA containing the *ref(2)P* *Hin*dIII−*Xho*I fragment either intact (**lane 1**) or cut with *Bgl*I restriction enzyme (**lane 2**). **Lane 3** is a DNA−DNA renaturation control: the DNA cut with *Bgl*I was treated as the other samples except that poly(A)$^+$ RNAs were omitted. The hybrids were treated with S1 nuclease and analysed by electrophoresis in non-denaturing or in alkaline agarose gels. After blotting onto nylon membranes, the protected fragments were detected by hybridization with one of the three labelled probes (A, B or C) and autoradiographed. After exposure, the membrane was reused with the second probe and then with the last one. Only the autoradiograms of the alkaline agarose gel are presented (**lower panel**). The experiment is schematically represented (**upper panel**). The DNA to be protected is shown as a thin line; the *Bgl*I (Bg) restriction sites are indicated above and those used to generate the labelled probes below: *Eco*RI (E), *Bam*HI (B) and *Sac*I (S). A, B and C probes are underlined with stars. The DNA fragments corresponding to lanes 1 are represented by filled boxes, those corresponding to lanes 2 by open boxes. The deduced structure of the *ref(2)P* transcripts is presented in the **middle panel**: filled boxes show the limits of the exons; connecting lines show the introns.

startpoints revealed that secondary structures could be formed in the mRNAs and might have impaired reverse transcriptase progression, resulting in artefactual arrests. An example of secondary structure which could explain the presence of the artefactual products +49, +52 and +55 is shown in Figure 4.

The positions of the primer extended products synthesized on poly(A)$^+$ RNAs extracted from ovaries corresponded exclusively to L, sL and ssL fragments, while those corresponding to the short products (S) were undetectable (Figure 4). About half of the poly(A)$^+$ RNAs from females originate from this organ. This may account for the different ratio of long over short products observed in females and males (Figure 4).

The 5′ flanking region of the ref(2)P structural gene

The short *ref(2)P* transcripts were initiated in +195 +198 positions, ~30 and 80 bp downstream of potential TATA and CAAT boxes which could be part of their promoter (Figure 1). Other potential CAAT boxes seemed to be repeated ~140 bp upstream of these initiation sites.

The genomic sequence upstream of the more distal initiation sites was very rich in A + T, but no obvious TATA and CAAT boxes were observed in the putative promoter region, 50 and 80 bp upstream. Sequences which are only consistent in six of seven positions with the eukaryotic promoter consensus sequence TATA(A/T)A(A/T) (Breathnach and Chambon, 1981) were present in position +21 (TATTAAT) and in positions −13, −50 and −100 (TAAAAAT). Whether they can function as promoter sites is unknown. The three last sequences were included in three 9mer repeats, GTAAAAATA. Other sequences were repeated and did not seem to be distributed randomly. The first 9mer repeat was preceded by three motifs: an 8mer palindromic sequence (TATCGATA) in position −144 and two hexamers (TATGTA and TGTGAA) in position −125 and −117 respectively. These three motifs were found again, upstream of the second 9mer repeat, in the same order. Upstream of the third 9mer repeat, only partial homology with the 8mer sequence and with one of the two hexamers was observed, but it was followed by the perfect 8mer palindromic sequence (TATCGATA) in position +3. Potential CAAT boxes in direct or reverse orientation were present in −216, −295 and −459. The former was in complete agreement with the consensus sequence (AGATTGACC).

Predicted amino acid sequence of the ref(2)P gene product

Ref(2)P transcript maturation resulted in the connection of three open reading frames in the three exons. The first ATG in the (+)strand DNA is expected to be the initiator codon as in 95% of the mRNAs tabulated by Kozak (1984). This ATG, in position +372, was in frame with this ORF and preceded by a sequence, GAAA−ATG, close to the consensus sequence germane to translation initiation in *Drosophila* (C/A)AA(A/C)−ATG (Cavener, 1987). The following ATG triplets did not fulfil these two conditions, except the fourth ATG in exon 2 (position +1249). Translation from the first ATG would produce a protein of 599 amino acid residues with a calculated mol. wt of 76.3 kd.

In vitro translation of *ref(2)P* was studied with the longest cDNA clone isolated from a cDNA library constructed in the expression vector PNB40. This cDNA was sequenced; it contained all the coding region, started in position +25 and ended in +3104. Four divergences with the genomic sequence were found (Figure 5), three of them were located in the coding region and induced amino acid changes: Ser188 was replaced by Arg, Asp 542 by Glu and Thr 575 by Ser. This cDNA was transcribed *in vitro* by the SP6 RNA polymerase and synthesized RNAs translated *in vitro* with a rabbit reticulocyte lysate. Results are shown in Figure 5. Apart from the endogenous products of the lysate preparation seen in the 'control' and '*Hin*dIII' lanes, a major polypeptide was obtained with an apparent mol. wt of 115−125 kd instead of the 76 kd as deduced from the sequence. A ladder of faster migrating products was also observed. When the RNAs were synthesized from a template cleaved with *Bam*HI, the major product migrated like a 78 kd polypeptide instead of the 54 kd expected. The minor products were not displaced, which suggests that they resulted from premature termination.

Anomalous electrophoretic migrations have been described for proteins rich in proline and containing long stretches of

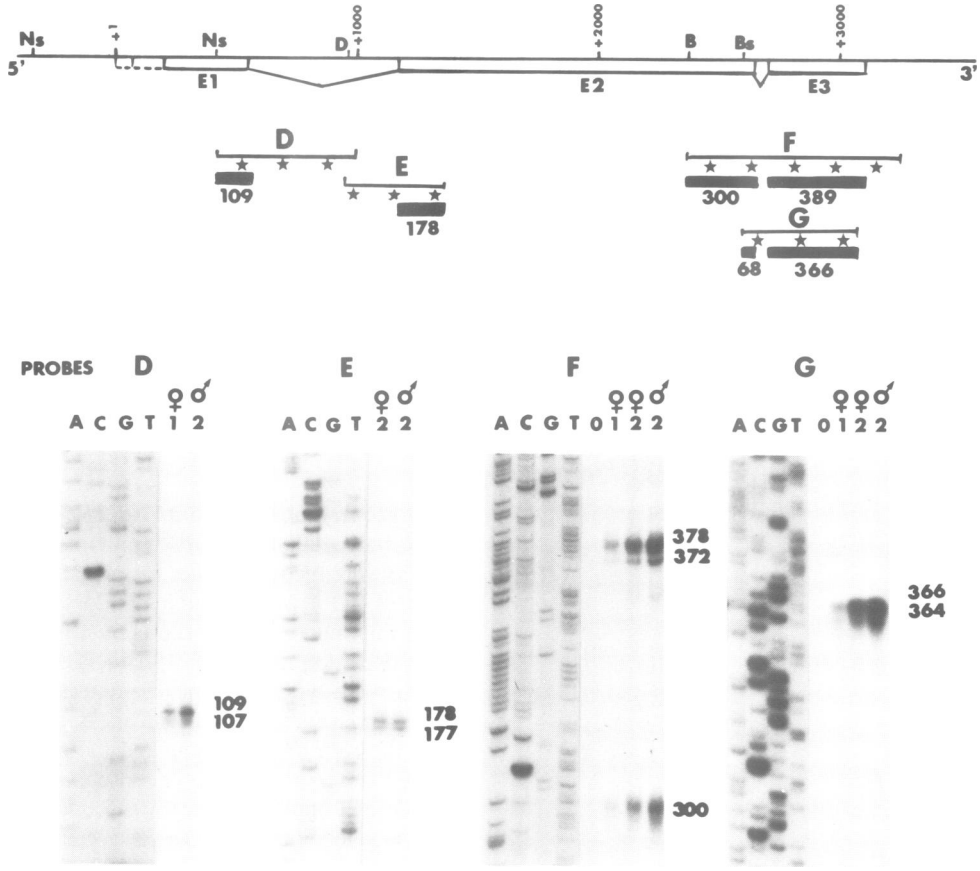

Fig. 3. Mapping of the splice junctions of *ref(2)P* transcripts. **Upper panel**: genomic DNA is represented as a continuous line; open boxes show the limits of the three exons (E1, E2 and E3); the cleavage sites for restriction endonucleases *Bam*HI (B), *Bst*EII (Bs), *Dra*I (D) and *Nsi*I (Ns) used to generate the DNA probes are indicated. **Middle panel**: the single-stranded uniformly labelled DNA probes (D, E, F and G: thin lines underlined with stars) and the fragments protected from nuclease digestion (thick lines) are shown to scale. The sizes in nucleotides of these protected fragments are indicated. **Lower panel**: autoradiograms of S1 nuclease protection experiments. Probes ($2-5 \times 10^4$ c.p.m.) were hybridized with 1 μg (1) or 2 μg (2) of poly(A)$^+$ RNAs prepared from adult *Drosophila* males or females or with tRNA as control (0); experiments were performed as described in Materials and methods. The protected fragments were analysed next to the dideoxynucleotide sequencing reactions used as size markers, on sequencing gels containing either 8% (D and E) or 5% (F and G) acrylamide. The protected fragment corresponding to the intact E3 exon is smaller (372–378 nucleotides) than the size deduced from cDNA sequencing (389 nucleotides). This discrepancy is due to over-digestion of the hybrids as explained in the text.

basic and acidic residues (e.g. c-*myc*, Persson *et al.*, 1984; Hann and Eisenman, 1984; *EIA*, Gingeras *et al.*, 1982; polyoma large T antigen, Soeda *et al.*, 1980; v-*fos*, Verma, 1984; 70K U1 snRNP, Query *et al.*, 1989). *Ref(2)P* protein also contains long stretches of charged amino acids; when compared with the average in the NBRF Library (Dayhoff *et al.*, 1978), it appears rich in proline, 8.5% (mean value 5.25% in the Library), and up to 20% from amino acid 240 to 317 and 15% from 360 to 433. In addition, three regions (P1, P2 and P3 in Figure 6) are enriched in proline (P), but also in glutamic acid (E), serine (S) and threonine (T). They exhibited PEST−SCORES >5, calculated according to Rogers *et al.* (1986). Such PEST regions were found in proteins with short half-life. In spite of a relatively high level of transcripts, the *ref(2)P* protein is difficult to detect in *Drosophila* (F.Wyers, personal communication); thus this protein could have a short half-life. This property, associated with an anomalous electrophoretic mobility, is also met in c-*myc*, *EIA* and v-*fos* proteins.

A detailed analysis of the amino acid sequence points out other interesting features. Analysis of the hydrophobicity profile did not reveal regions which could be transmembrane domains. In fact, the sequence is mainly hydrophilic. The distribution of charged amino acids and of cysteine clearly divides the putative protein into two domains: the N-terminal half where basic and acidic regions alternate, and the C-terminal half where three acidic regions are separated by two stretches devoid of any charge residues (Figure 6).

The N-terminal half of the sequence contains most of the tyrosine residues and all the cysteine (nine residues). Six of these cysteine, localized from residue 127 to 154, are organized in three doublets followed by a histidine triplet (Figure 6). This arrangement evocates 'zinc-finger' structures (Berg, 1986; Evans and Hollenberg, 1988). In the putative *ref(2)P* protein, one or two fingers could be formed. In Figure 6, we show examples of sequence alignment with fingers found in proteins implicated in nucleic acid interactions and with the consensus sequence derived from the multiple fingers of the transcription factor of *Xenopus laevis*, TFIIIA (Miller *et al.*, 1985).

In the C-terminal half of the sequence, repeated units occur between residues 386 and 413. The motif (SANQSXXP) is exactly reiterated three times (Figure 6). Several other imperfect repeats where the changes are generally conservative, can be seen farther on the sequence. The repetition of this motif, even degenerated, was not found in the PSEQIP Library (Claverie and Bricault, 1986).

The morphological anomalies in mitochondria observed

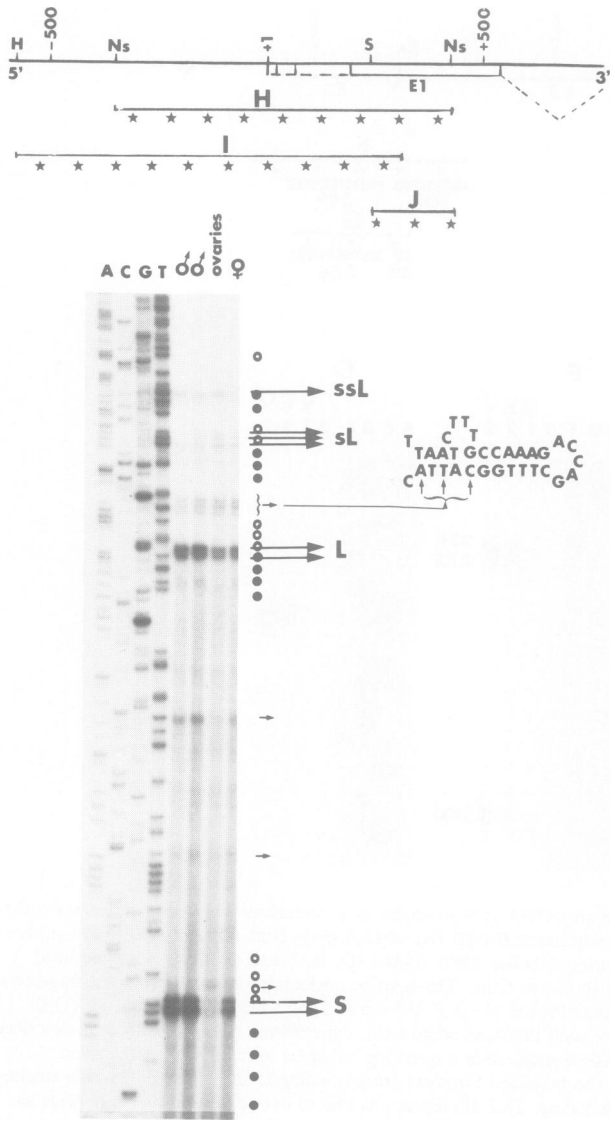

Fig. 4. Determination of transcription initiation sites by primer extension and nuclease mapping. **Upper panel**: genomic DNA is shown as a continuous line; the open box represents the short E1 exon and the open dashed box indicates the limits of the long E1 exons; the cleavage sites for restriction endonucleases *Hin*dIII (H), *Nsi*I (Ns) and *Sac*I (S) used to generate the single-stranded uniformly labelled probes and primer (thin lines underlined with stars) are indicated. **Lower panel**: autoradiogram of a primer extension experiment. The J primer (219 nucleotides) was hybridized with 4 μg of poly(A)$^+$ RNAs from *Drosophila* males or females or from ovaries and extended with reverse transcriptase as described in Materials and methods. The dideoxynucleotide sequencing reactions used as size markers were performed on the single-stranded DNA clone used to prepare the J primer, so the position of the transcription initiation sites can be read directly on the sequence. The extended products are designated by long arrows. The initiation site positions determined by S1 nuclease mapping (H probe) and exonuclease VII mapping (I probe) are respectively represented by filled and open circles. Reverse transcriptase pause sites are indicated by small arrows. The predicted secondary structure at the level of the sites found between position +49 and +55 is presented.

during spermiogenesis of sterile flies carrying null alleles of the *ref(2)P* gene (described below) led us to investigate the presence of mitochondrial pre-sequences. These pre-sequences are expected to form amphiphilic helices with high hydrophobic moments, free of acidic residues and enriched in arginine (Eisenberg *et al.*, 1984; Roise *et al.*, 1986; von Heijne, 1986). In the *ref(2)P* protein sequence, two stretches of 18 amino acids (7−24 and 19−35) could be folded into amphiphilic helices. The amino acid distribution along the helical wheel of Schiffer and Edmundson (1967) is shown in Figure 6. In the first hypothetical helix, positive charges do not consist of arginine, but of lysine. In both helices, proline and glycine which are amino acids known to destabilize helical structure, are present, but it should be noted that these amino acids are not excluded from many sequences which have been shown to be functional mitochondrial pre-sequences (Baker and Schatz, 1987).

The predicted *ref(2)P* protein sequence was aligned with the 6000 protein collection of the non-redundant PSEQIP Library (Claverie and Bricault, 1986). No significant homology was found, thus we cannot take advantage of data banks for the identification of *ref(2)P* function.

Male sterility phenotype of ref(2)P^{null} mutants

Three independent *ref(2)P^{null}* mutants (ODP1, ODP2 and ODP3) were recovered after mutagenesis with a bialkylating agent (1-2-3-4 diepoxybutane) (Contamine *et al.*, 1989). Since this mutagen has been shown to induce deletions in DNA (Reardon *et al.*, 1987), sequence divergences with the *ref(2)P* wild-type allele may be detected by S1 nuclease in hybrid molecules formed with messenger RNAs extracted from the *ref(2)P^{null}* mutants and the *ref(2)P* cloned DNA. Using the experimental procedure explained in the legend of Figure 2, one divergence was found for each mutant and approximately located on the gene sequence: around +1750, +395 and +2265 for ODP1, ODP2 and OPD3 respectively. All three mutations affected the coding region.

Homozygous *ref(2)P^{null}* mutants were viable, but males were sterile. These males exhibited an apparently normal reproductive system. After squashing the testes in saline solution, numerous elongated sperm bundles were observed but, unlike in wild-type flies, spermatozoa were not motile and the tail region frequently appeared packed into a single rope-like mass.

Electron microscope observations of cross-sections of the testes showed that spermatogonia, spermatocytes and the young spermatids (until stage 5 of spermiogenesis as defined by Stanley *et al.*, 1972) were identical in the wild-type and in the three *ref(2)P^{null}* mutants. Important differences were observed after stage 5.

In wild-type cysts at stage 6 (Figure 7a), spermatids were homogenous in size and were connected together by thin cytoplasmic bridges. Almost all spermatids contained only one axoneme and two mitochondria. In the largest mitochondria (the major mitochondrial derivative), an electron dense structure, the paracrystal, was in close contact with the axoneme. The minor mitochondrial derivative, the smallest one, never contained such structure.

For the *ref(2)P^{null}* mutants ODP1, ODP2 and ODP3, we frequently observed in a given cytoplasm more than one axoneme exhibiting an apparently normal structure; these axonemes—up to seven in some cases—were found in cysts which appeared to correspond to stage 6. Indications of degeneration occurred with the appearance of large myelin figures around the spermatids. The most striking difference was observed in the mitochondria, which varied in size and appeared degenerate (Figure 7b). The paracrystal was still present in one of the two mitochondria associated with an axoneme, allowing the major and minor derivatives—both

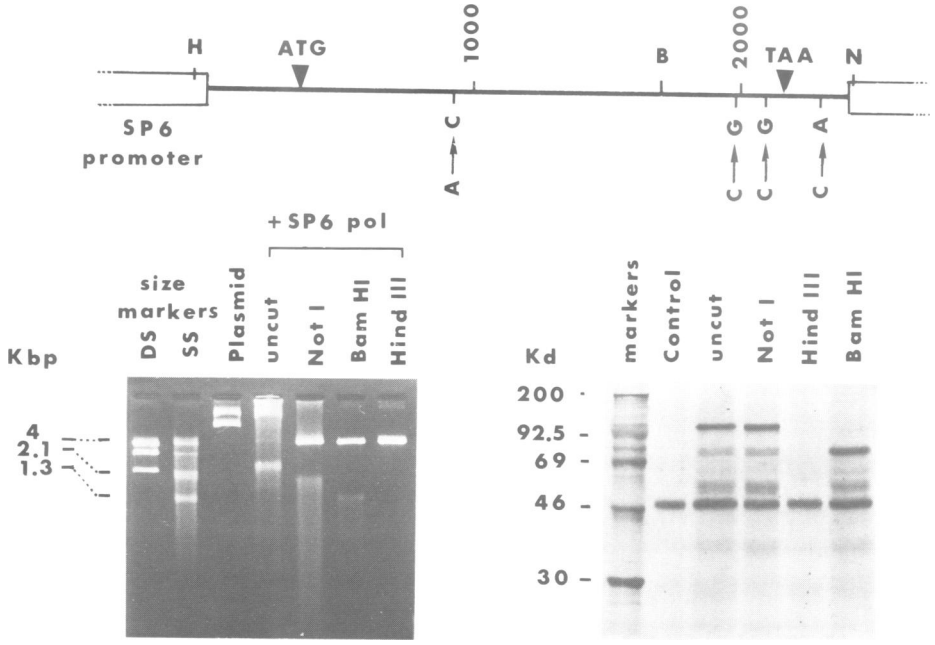

Fig. 5. *In vitro* transcription and translation of *ref(2)P* cDNA. **Upper panel**: the *ref(2)P* cDNA is represented as a continuous line. The restriction sites used are indicated above: *Hin*dIII (H), *Bam*HI (B) and *Not*I (N). The first ATG of the long ORF and the TAA stop codons are also shown. The four divergences found between the cDNA and the genomic sequence (in +1564, +2628, +2781 and +2986 in Figure 1) are indicated below. **Left panel**: *in vitro* transcription products were analysed on a 1.5% agarose gel. PNB40 plasmid DNA containing the *ref(2)P* cDNA insert was used as a template either intact or previously linearized with *Not*I (downstream of the cDNA 3′ end) or with *Bam*HI (inside the insert) or with *Hin*dIII (upstream of the cDNA 5′ end) restriction enzyme. Three DNA restriction fragments either double-stranded (DS) or previously denatured with 0.2 N NaOH (SS) were used as size markers. **Right panel**: the *in vitro* translation products of the transcribed RNAs were synthesized in a rabbit reticulocyte lysate and analysed by SDS–PAGE (10% acrylamide and 0.8% bis-acrylamide). Control for the endogenous products of the lysate was performed with a transcription mix containing no DNA. Size markers are radioactive Rainbow^{TM} proteins from Amersham.

of which were affected—to be distinguished. At later stages, the spermatids contained numerous axonemes and were devoid of any mitochondria. This degenerative process was limited to spermatids: mitochondria in the interstitial cells of the cysts and in the cells of the testis sheath appeared unaffected.

Discussion

The *D.melanogaster ref(2)P* gene structure has now been elucidated. The 3.1 kb transcription unit encodes mRNAs which are heterogeneous in length. This heterogeneity is not due to alternative splicing or to transcription termination at multiple sites. The only difference is in the transcription initiation sites. Whether the different 5′ ends correspond to the actual transcription initiation sites, rather than to points of artefactual nuclease digestion or artefactual release of the reverse transcriptase (due to strong secondary structures), cannot be completely ascertained. Nevertheless, the following observations argue in favour of the existence of at least two classes of initiation sites. (i) The ratio between short and long transcripts is not constant. Only the long ones are present in ovaries, indicating a tissue specificity. The ratio and total amount of transcripts vary during *Drosophila* development, suggesting a temporal regulation (Contamine, unpublished data). (ii) An appropriately-positioned TATA box and several CAAT boxes are present upstream of the short transcripts, as in classical promoters (Breathnach and Chambon, 1981). This is not observed upstream of the long transcripts. Several degenerated TATA boxes are present, but their distances to the initiation sites are longer than the usual 30 bp. Eukaryotic gene promoters which are characterized by an important heterogeneity at the level of transcription initiation also lack TATA boxes (Reynolds *et al.*, 1984; Ishii *et al.*, 1985; Huckaby *et al.*, 1987). They are exceptionally G + C rich and often contain GC boxes, GGGCGG, which may be important regulatory elements (Dynan and Tjian, 1985; McKnight and Tjian, 1986). On the contrary, the region upstream of the long *ref(2)P* transcripts is very rich in A + T and contains no GC boxes. Nevertheless the several repeats, palindromic sequences and inverted CAAT boxes could be promoter elements for long transcripts.

Interestingly, the heterogeneity of the mRNAs does not affect the large open reading frame: the same protein is predicted for all transcripts. This has also been observed for the four mRNAs of the *Drosophila* homeotic *Antennapedia* gene, which are under the control of two remote promoters (Schneuwly *et al.*, 1986). Such a situation is encountered for the maternal and zygotic transcripts of *Caudal* (Mlodzik and Gehring, 1987) and *Hunchback* (Tautz *et al.*, 1987). These genes exhibit a maternal effect, like *ref(2)P* for sigma virus multiplication. Another feature common to these four genes is the existence of long non-coding sequences in their mRNAs (the leader region of the *ref(2)P* mRNAs is 174–372 nucleotides in length and the trailer is 250 nucleotides). These sequences are thought to play a role in the regulation of gene expression at the translation level (Pelletier and Sonenberg, 1985).

The *ref(2)P* gene is expressed in the female germ cell line. However, *ref(2)P^{null}* homozygous flies are perfectly viable and the females are fertile. The only tissue where the absence of *ref(2)P* expression has severe effects is the male germ

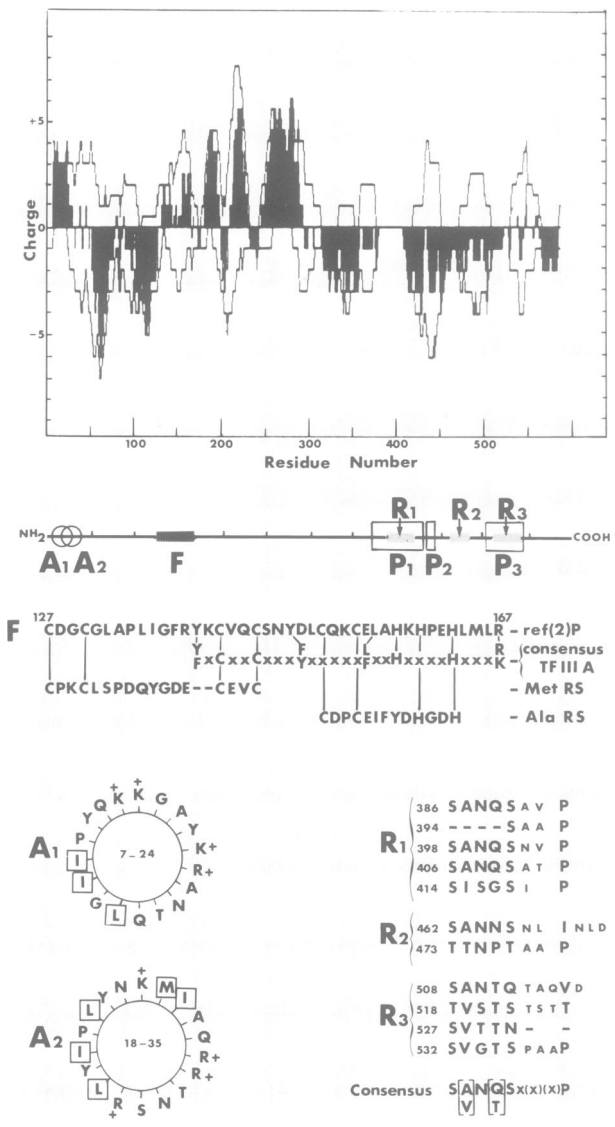

Fig. 6. Features of the putative *ref(2)P* protein. **Upper panel**: the distribution of charged residues is presented, using an 18 amino acid size window (charges were calculated on the basis of +1 for lys and arg, +0.5 for his and −1 for asp and glu); the black profile corresponds to the net charges. **Middle panel**: relative positions of the key elements A, F, R and P. **Lower panel**: their detailed characteristics are summarized. The two circles, A1 and A2, correspond to the two amphiphilic helices predicted according to Eisenberg (1984): A1, from residue 7 to 24 [hydrophobic moment, $\mu H = 6.66$; hydrophobicity of its non-polar face (seven neighbouring residues), $H_{max} = 3.83$]; A2, from 18 to 35 ($\mu H = 7.87$; $H_{max} = 3.35$). The amino acid projections along the helical wheel of Schiffer and Edmundson (1967) are presented. The dark box F corresponds to the potential finger region (from residue 127 to 167). Its amino acid sequence is compared to the consensus sequence of TFIIIA fingers and to the finger sequences of two *E.coli* enzymes [methionyl- and alanyl-tRNA synthetase (Berg, 1986)]. The shaded boxes R1, R2 and R3 correspond to the conserved repeated motifs. Their sequences are aligned and a consensus is derived for positions at which at least nine of the 11 repeats contain identical or conservatively substituted residues. X are generally apolar residues, frequently A. Open boxes, P1, P2 and P3 indicate PEST regions displaying a score >5: 6.2 for P1, from residue 372 to 431; 23.4 for P2, from 433 to 443; 9.2 for P3, from 504 to 542 (PEST−FIND scores were calculated as described by Rogers *et al.*, 1986).

cells: *ref(2)P*^{null} homozygous males are sterile. Since the sterility does not occur in females, the *ref(2)P* gene does not belong to the class of genes implicated during meiosis, such as *Drosophila* ms(2)3R (Romrell *et al.*, 1972). Moreover, failure of cell division during meiosis in this mutant results in four spermatids in a common cytoplasm (Romrell *et al.*, 1972), while in *ref(2)P*^{null} mutants a random number of axonemes per cytoplasm is observed.

In other mutants, mitochondrial malformations have also been described, but they appear earlier during spermiogenesis and are associated with axonemal disruption (Anderson, 1967; Wilkinson *et al.*, 1975; Habliston *et al.*, 1977). In *ref(2)P*^{null} mutants, the organization of the axonemes seems normal, even when the mitochondria have completely degenerated. The fact that mitochondrial degeneration is limited to spermatids and does not affect other testis tissues indicates that the *ref(2)P* gene product interacts directly or indirectly with products appearing specifically during spermiogenesis. The presence of specific motifs in the amino acid sequence suggests two types of possible interactions: (i) the potential targetting pre-sequences of the N-terminal end of the protein could interact with the mitochondrial membrane and (ii) the potential Zn finger could interact with nucleic acids.

The *ref(2)P* gene product is necessary in the male germ line, while in somatic tissues and in the female germ line its absence has no apparent effect. However, interactions between the *ref(2)P* protein and the sigma virus clearly occur in these tissues where, moreover, the molecular data have confirmed the expression of the *ref(2)P* gene. It is difficult to assume that, in somatic and female germ line cells, the *ref(2)P* expression is only related to the sigma virus. Another still unknown function of the gene may exist.

Materials and methods

Bacterial strains and vectors

Transformations with pEMBL recombinant DNAs were carried out in *Escherichia coli* JM103 or DIH101 (a HB101 derivative: rec A, r$^-$, m$^-$, F'kanamycin, from David Ish-Horowicz, IRCF, London). Plasmids pNB40 carrying the cDNA clones were grown in the *E.coli* DH5α strain of Bethesda Res. Lab. Inc. Plasmids pEMBL 18 and 19 (Dente *et al.*, 1985) were used for subcloning and sequencing. The M13 derivative M13K07 (Chang and Cohen, 1978; Dotto and Zinder, 1984) was the helper phage used to produce single-stranded DNA from pEMBL plasmids.

DNA sequencing

The genomic clone which spans two restriction sites, *Hin*dIII and *Xho*I, originates from a genomic DNA library of *D.melanogaster* Oregon R in λEMBL4 (Contamine *et al.*, 1989). It was subcloned in the pEMBL 18 and 19 vectors. Serial unidirectional deletions were created in the inserts as described by Yanisch-Perron *et al.* (1985). After transformation of *E.coli* cells (DIH101), clones were characterized by partial restriction mapping. Clones harbouring plasmids whose inserts vary in length by ~200 bp were grown overnight, infected with the helper phage M13K07 (~5 × 10^9 p.f.u. per 20 μl of stationary cell culture), diluted with 100 volumes of fresh LB medium containing ampicillin, kanamycin and thiamin and grown again overnight. Single-stranded DNA was prepared from the cell culture supernatant as described by Messing (1983).

DNA sequencing was accomplished by the dideoxy chain termination procedure (Sanger *et al.*, 1977), using Amersham's M13 sequencing kit. For some fragments, compressions due to strong secondary structures in the labelled product were overcome by using gels containing 7 M urea plus 40% formamide.

Isolation of ref(2)P cDNA clones

A cDNA library of Oregon R *D.melanogaster* imaginal discs in the SP6-T7 expression vector pNB40 (Brown and Kafatos, 1988) was screened for *ref(2)P* cDNA according to Maniatis *et al.* (1982). From ~60 000 colonies tested, three clones containing at least a part of the 5'-untranslated region of the mRNA were selected. For sequencing the *ref(2)P* cDNAs were inserted in pEMBL cloning site and deletions were created taking advantage of the restriction map deduced from the genomic sequence.

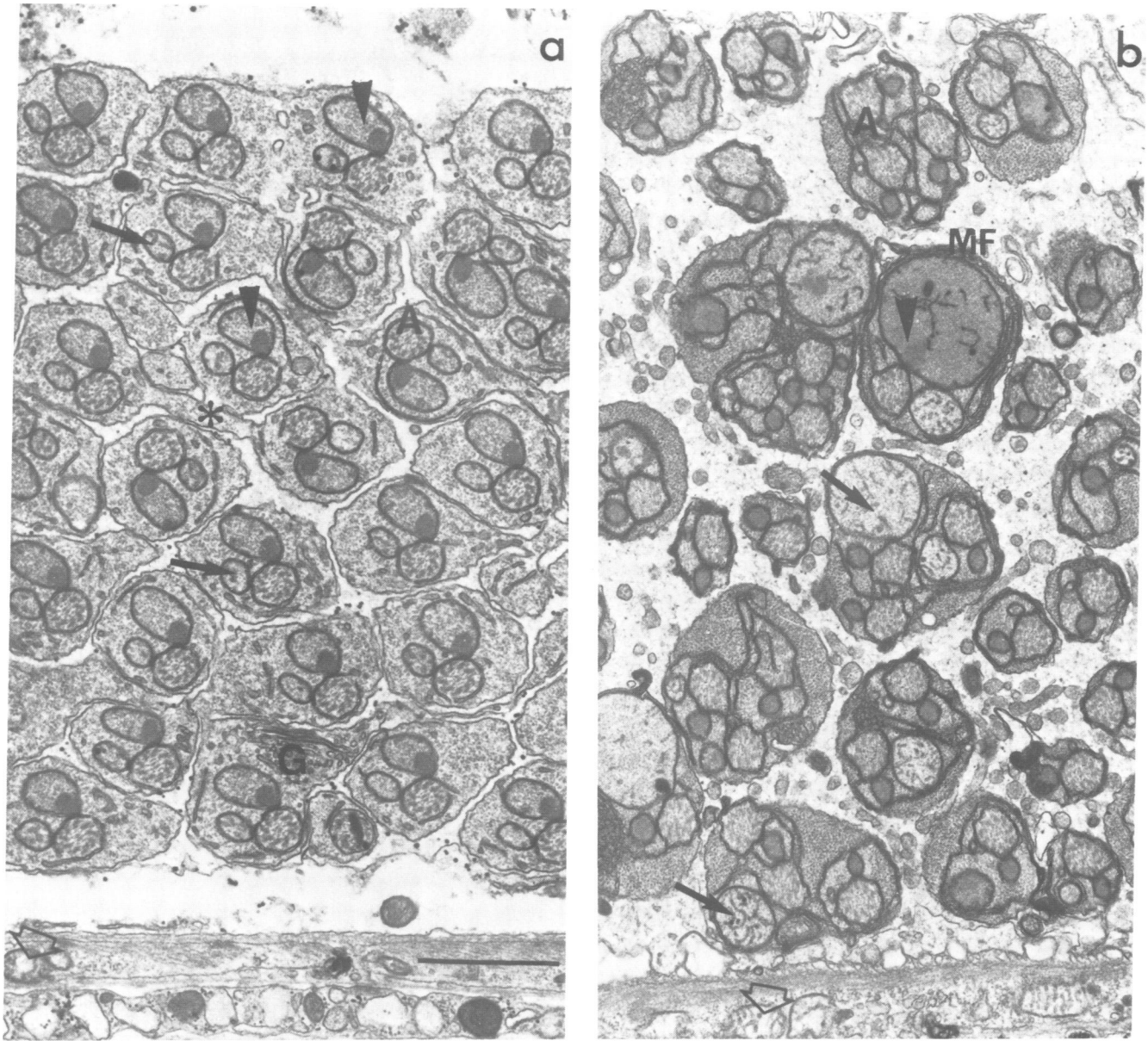

Fig. 7. Panel a: cross-section of a cyst of wild-type fly (stage 6): spermatids contain an axoneme (A) associated with a major (arrowhead) and a minor (arrow) mitochondrial derivative; they are connected together by thin cytoplasmic bridges (*); a Golgi apparatus (G) and a mitochondria of the testis sheath cell (open arrow) are indicated. **Panel b**: cross-section of a cyst of the ODP1 mutant, at the same stage of spermiogenesis: spermatids are randomly fused, showing one or more axonemes (A); major mitochondrial derivative (arrowhead) recognizable by its paracrystal and minor derivative (arrow), both exhibit signs of degeneration; large myelin figures (MF) are present; mitochondria of the testis sheath cell (open arrow) are similar to those observed in (a). Magnification for (a) and (b) is ×21600.

RNA analysis

Poly(A)$^+$ RNAs were purified from adult males or females or hand-dissected ovaries of *D.melanogaster* (strain Oregon M) as described by Contamine *et al.* (1989).

To determine the *ref(2)P* gene structure, S1 nuclease protection experiments were performed as described by Maniatis *et al.* (1982), with only slight modifications. Poly(A)$^+$ RNAs (6 μg) of *Drosophila* females were hybridized with 0.4 μg of unlabelled plasmid DNA containing the *Hin*dIII−*Xho*I *ref(2)P* genomic fragment. This DNA was previously linearized in a unique restriction site of pEMBL or the insert was fragmented using various restriction enzymes. Hybridization was carried out for 3 h at 52°C. S1 nuclease digestion was performed for 30 min at 37°C (at this temperature, partial digestion of RNAs occurred at the level of introns, thus the analysis of the protected fragments on neutral gels allowed the order of the exons to be deduced). Neutral or alkaline gels were blotted onto Hybond membranes (Amersham) and probed as described by Contamine *et al.* (1989). The A, B and C probes were restriction fragments of the *ref(2)P* genomic clone, purifed on agarose gels and ^{32}P-labelled by nick translation (~2 × 10^8 c.p.m./μg).

To determine the precise exon junctions and the transcription initiation sites, nuclease mapping and primer extension analyses were performed as follows. Poly(A)$^+$ RNAs (0.75−4 μg) from *Drosophila* males, females or ovaries were co-precipitated with uniformly labelled single-stranded probe or primer. Pellets were dissolved in 25 μl of hybridization buffer (80% formamide, 0.4 M NaCl, 1 mM EDTA and 40 mM PIPES, pH 6.4). After heat denaturation, hybridization was performed at 47°C for 17 h. For nuclease mapping, hybridization mixes were diluted with 250 μl of the appropriate buffer (30 mM Na acetate pH 4.6, 50 mM NaCl, 1 mM ZnSO4, 5% glycerol for S1 nuclease; 50 mM potassium phosphate pH 7, 8 mM EDTA, 1 mM DTT for exonuclease VII). Incubations were carried out for 30 min at 30°C with 580 U of S1 nuclease or for 30 min at 37°C with 1.1 U of exonuclease VII. Incubation was followed by extraction with phenol−chloroform and ethanol precipitation. Nucleic acids were dissolved in 100 μl of 10 mM Tris−HCl pH 8, 1 mM EDTA, dialysed 2 h against water on Millipore VS filters (0.025 mμ), concentrated by ethanol precipitation, dissolved in formamide dye mix (40% formamide, 0.04% xylene cyanol FF, 0.04% bromophenol blue and 8 mM EDTA) and analysed on sequencing gels.

For primer extension experiments, hybridization mixes were precipitated with ethanol. Pellets were dissolved in 25 μl of reverse transcriptase buffer (5 mM Tris–HCl pH 8.3, 0.1 mM DTT, 10 mM $MgCl_2$, 50 mM KCl and 0.1 mM of each deoxynucleotide) and incubation was carried out for 40 min at 42°C with 20 U of reverse transcriptase. The reactions were stopped by addition of EDTA (10 mM final concentration). After ethanol precipitation, nucleic acids were analysed on sequencing gels.

The probes and the primer shown in Figures 3 and 4 were synthesized by the Klenow enzyme in the presence of ^{32}P-labelled deoxynucleotides, using (+) strand DNA prepared from deleted plasmids as a template. After cleavage with an appropriate restriction enzyme and denaturation in 0.2 N NaOH, the relevant single-stranded radioactive fragment was purified by electrophoresis through a 1.5 or 2% low melting point agarose gel and extracted as described by Maniatis *et al.* (1982).

In vitro transcription and translation

PNB40 plasmids containing the *ref(2)P* cDNA insert were prepared as described by Brown and Kafatos (1988). Supercoiled or linearized templates were transcribed with the SP6 polymerase using standard conditions (Melton *et al.*, 1984). Transcripts were translated in the presence of [^{35}S]methionine in a message dependent rabbit reticulocyte lysate under the conditions described by the supplier (Amersham). SDS–PAGE analysis of translation products followed the method of Laemmli (1970). After electrophoresis, polypeptides were electrotransferred to nitrocellulose and labelled components were visualized by autoradiography on Hyperfilm-βMax from Amersham.

Morphological studies

Testes were dissected in a Ringer solution. They were squashed under a coverslip and observed through a light microscope. For electron microscopic observations, the testes were treated as previously described (Segretain and Roussel, 1988). Thin sections were counterstained for 2 min with lead citrate and examined with a Philips 300 electron microscope.

Acknowledgements

We thank Clément Mettling for kindly providing *E.coli* DIH101 strain, pEMBL plasmids and M13K07 helper phage; Nicolas Brown for Oregon R cDNA libraries; Jean-François Gibrat and Isabelle Sauvaget for help with the computer analyses of sequences; Françoise Wyers for pointing out the PEST regions in *ref(2)P* protein sequence and the anomalous electrophoretic migration of proline-rich polypeptides. We are grateful to Anne-Marie Petitjean for supplying *Drosophila* poly(A)$^+$ RNAs and Claudine Vigier for excellent technical assistance. M.L.F. was the recipient of a postdoctoral fellowship from Ministerio Educacion y cienca and Ministère de la Recherche et de l'Enseignement Supérieur. This work was supported by the Centre National de la Recherche Scientifique through the UPR no. A.2431 and by a grant no. 861009 from the Institut National de la Santé et de la Recherche Médicale.

References

Anderson,W.A. (1967) *Mol. Gen. Genet.*, **99**, 257–273.
Baker,A. and Schatz,G. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 3117–3121.
Berg,J.M. (1986) *Science*, **232**, 485–487.
Birnstiel,M.L., Busslinger,M. and Strub,K. (1985) *Cell*, **41**, 349–359.
Breathnach,R. and Chambon,P. (1981) *Annu. Rev. Biochem.*, **50**, 349–383.
Brown,N.H. and Kafatos,F.C. (1988) *J. Mol. Biol.*, **203**, 425–437.
Brun,G. and Plus,N. (1980) In Ashburner,M. and Wright,T.R.F. (eds), *The Genetics and Biology of Drosophila*. Academic Press, New York, Vol. 2d, pp. 625–702.
Bussereau,F. (1975) *Ann. Microbiol.*, **126B**, 389–403.
Cavener,D.R. (1987) *Nucleic Acids Res.*, **15**, 1353–1361.
Chang,A.C.Y. and Cohen,S.M. (1978) *J. Bacteriol.*, **134**, 1141–1156.
Claverie,J.M. and Bricault,L. (1986) *Proteins: Structure, Function and Genetics*, **1**, 60–65.
Contamine,D., Petitjean,A.M. and Ashburner,M. (1989) *Genetics*, in press.
Dayhoff,M.O., Hunt,L.T., and Hurst-Calderone,S. (1978) In Dayhoff,M.O. (ed.), *Composition of Proteins. Atlas of Protein Sequence and Structure*. Washington: National Biomedical Research Foundation, Vol. 5, suppl. 3, pp. 363–373.
Dente,L., Ciliberto,G. and Cortese,R. (1985) *Nucleic Acids Res.*, **13**, 3941–3952.
Dotto,G.P. and Zinder,N.D. (1984) *Nature*, **311**, 279–280.
Dynan,W.S. and Tjian,R. (1985) *Nature*, **316**, 774–778.
Eisenberg,D., Schwarz,E., Komaromy,M. and Wall,R. (1984) *J. Mol. Biol.*, **179**, 125–142.
Evans,R.M. and Hollenberg,S.M. (1988) *Cell*, **52**, 1–3.
Fleuriet,A. (1976) *Evolution*, **30**, 735–739.
Gay,P. (1978) *Mol. Gen. Genet.*, **159**, 269–283.
Gingeras,T.D., Sciaky,D., Gelinas,R., Bing-Dong,J., Yen,C., Kelly,M., Bullock,P., Parsons,B., O'Neil,R. and Roberts,R. (1982) *J. Biol. Chem.*, **257**, 13475–13491.
Habliston,D.L., Stanley,H.P. and Bowman,J.T. (1977) *J. Ultrastruct. Res.*, **60**, 221–234.
Hann,S.R. and Eisenman,R.N. (1984) *Mol. Cell. Biol.*, **4**, 2486–2497.
von Heijne,G. (1986) *EMBO J.*, **5**, 1335–1342.
Huckaby,C.S., Conneely,O.M., Beattie,W.G., Dobson,A.D.W., Tsai,M.J. and O'Malley,B.W. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 8380–8384.
Ishii,S., Xu,Y.H., Stratton,R.H., Roe,B.A., Merlino,G.T. and Pastan,I. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 4920–4924.
Keller,E.B. and Noon,W.A. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 7417–7420.
Kozak,M. (1984) *Nucleic Acids Res.*, **12**, 857–872.
Laemmli,U.K. (1970) *Nature*, **227**, 680–685.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
McKnight,S. and Tjian,R. (1986) *Cell*, **46**, 795–805.
Melton,D.A., Krieg,P.A., Rebagliati,M.R., Maniatis,T., Zinn,K. and Green,M.R. (1984) *Nucleic Acids Res.*, **12**, 7035–7056.
Messing,J. (1983) *Methods Enzymol.*, **101**, 20–78.
Miller,J., McLachlan,A.D. and Klug,A. (1985) *EMBO J.*, **4**, 1609–1614.
Mlodzik,M. and Gehring,W.J. (1987) *Cell*, **48**, 465–478.
Mount,S.M. (1982) *Nucleic Acids Res.*, **10**, 459–472.
Nakamura,N. (1978) *Mol. Gen. Genet.*, **159**, 285–292.
Nakamura,N., Gay,P. and Contamine,D. (1986) *Biol. Cell*, **56**, 227–238.
Pelletier,J. and Sonenberg,N. (1985) *Cell*, **40**, 515–526.
Persson,H., Hennighausen,L., Taub,R., Degrado,W. and Leder,P. (1984) *Science*, **225**, 687–693.
Query,C.C., Bentley,R.C. and Keene,J.D. (1989) *Cell*, **57**, 89–101.
Reardon,J.T., Liljestrand-Golden,C.A., Dusenbury,R.L. and Smith,P.D. (1987) *Genetics*, **115**, 323–331.
Reynolds,G.A., Basu,S.K., Osborne,T.F., Chin,D.J., Gil,G., Brown,M.S., Goldstein,J.L. and Luskey,K.L. (1984) *Cell*, **38**, 275–285.
Rogers,S., Wells,R. and Rechsteiner,M. (1986) *Science*, **234**, 364–369.
Roise,D., Horvath,S.J., Tomich,J.M., Richards,J.H. and Schatz,G. (1986) *EMBO J.*, **5**, 1327–1334.
Romrell,L.J., Stanley,H.P. and Bowman,J.T. (1972) *J. Ultrastruct. Res.*, **38**, 563–577.
Rosen,L. (1980) *Science*, **207**, 989–991.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Schiffer,M. and Edmundson,A.B. (1967) *Biophys. J.*, **7**, 121–135.
Schneuwly,S., Kuroiwa,A., Baumgartner,P. and Gehring,W.J. (1986) *EMBO J.*, **5**, 733–739.
Segretain,D. and Roussel,C. (1988) *Gamete Res.*, **21**, 451–463.
Soeda,E., Arrand,J., Smolar,N., Walsh,J. and Griffin,B. (1980) *Nature*, **283**, 445–453.
Stanley,H.P., Bowman,J.T., Romrell,L.J., Reed,S.C. and Wilkinson,R.F. (1972) *J. Ultrastruct. Res.*, **41**, 443–466.
Tautz,D., Lehmann.,R., Schnurch,H., Schuh,R., Seifert,E., Kienlin,A., Jones,K. and Jackle,H. (1987) *Nature*, **327**, 383–389.
Teninges,D. and Bras-Herreng,F. (1987) *J. Gen. Virol.*, **68**, 2625–2638.
Verma,J. (1984) *Nature*, **308**, 317.
Wilkinson,R.F., Stanley,H.P. and Bowman,J.T. (1975) *J. Ultrastruct. Res.*, **53**, 354–365.
Yanisch-Perron,C., Vieira,J. and Messing,J. (1985) *Gene*, **33**, 103–119.

Received on March 28, 1989; revised on July 6, 1989