Biochem. J. (1992) 286, 281–288 (Printed in Great Britain)

Structure and expression of the *Drosophila* ubiquitin–52-amino-acid fusion-protein gene

Héctor L. CABRERA,* Rosa BARRIO† and Carmen ARRIBAS‡
Departamento de Biología Molecular, Facultad de Ciencias, Universidad Autónoma de Madrid, 28049 Madrid, Spain

Ubiquitin belongs to a multigene family. In *Drosophila* two members of this family have been previously described. We report here the organization and expression of a third member, the *DUb52* gene, isolated by screening a *Drosophila melanogaster* genomic library. This gene encodes an ubiquitin monomer fused to a 52-amino acid extension protein. There are no introns interrupting the coding sequence. Recently, it has been described that this extension encodes a ribosomal protein in *Saccharomyces*, *Dictyostelium*, and *Arabidopsis*. The present results show that the 5′ regulatory region of *DUb52* shares common features with the ribosomal protein genes of *Drosophila*, *Xenopus* and mouse, including GC- and pyrimidine-rich regions. Moreover, sequences similar to the consensus Ribo-box in *Neurospora crassa* have been identified. Furthermore, a sequence has been found that is similar to the binding site for the TFIIIA distal element factor from *Xenopus laevis*. The *DUb52* gene is transcribed to a 0.9 kb mRNA that is expressed constitutively throughout development and is particularly abundant in ovaries. In addition, the *DUb52* gene has been found to be preferentially transcribed in exponentially growing *Drosophila* cells.

INTRODUCTION

Ubiquitin is a 76-amino acid protein present in all eukaryotic cells examined, its amino acid sequence being the most conserved among all known proteins [1]. In the cytoplasm, ubiquitin binds to proteins that will enter the non-lysosomal ATP-dependent degradation pathway, characteristic of both abnormal and short-lived proteins [1]. The existence of ubiquitin conjugates with certain nuclear [2–4], cytoplasmic [5] and cell-surface proteins [6–8] without causing their degradation raises the possibility that conjugation of ubiquitin to proteins may sometimes exert regulatory functions.

Ubiquitin genes have been cloned and characterized in a variety of species. It has been shown that these genes are organized in the genome in two major structural types. The first one is a polyubiquitin gene that encodes a polyubiquitin precursor protein containing head-to-tail repeats in tandem. This precursor is cleaved into monomeric ubiquitin proteins by specific cytoplasmic proteinases [9]. Although the number of repeats in this gene varies among species, its organization is very conserved [9–11]. In yeast, the polyubiquitin gene has been found to provide free ubiquitin proteins especially during stress conditions [12,13]. The second structural type of ubiquitin genes encode a single ubiquitin moiety fused in frame to unrelated sequences. These sequences encode 3′-*C*-terminal extensions of 52 (Ub52) [12,14–18] or 76–80 (Ub80) [12,18–21] amino acids. Both *Ub52* and *Ub80* encode highly conserved amino acid tails that are predominantly basic [22], which have been identified as ribosomal proteins (rp) in yeast [23] and mammals [24]. The fusion of ubiquitin to the *N*-terminus of these rps apparently increases the efficiency of their incorporation into the ribosome [23].

In *Drosophila*, ubiquitin genes constitute a multigene family. The polyubiquitin gene has been isolated [25] and characterized in our laboratory [26] and by Lee *et al.* [20]. It contains either 15 [24] or 18 [20] tandem repeats, depending on the fly stock analysed. This gene presents a transcriptional pattern that is highly polymorphic in different stocks and strains [26]. Moreover, in most organisms the transcription of the polyubiquitin gene is dramatically increased during heat-shock response [12,13,27], yielding a major heat-shock protein. However, in *Drosophila*, this gene is not significantly induced under our heat-shock conditions [25]. A cDNA corresponding to the *Ub80* ubiquitin fusion gene has also been isolated and characterized by Lee *et al.* [20]. Additionally, a previous report from our laboratory described the isolation and sequencing of a cDNA corresponding to a *DUb52* ubiquitin-fusion gene [28].

We report here the nucleotide sequence and expression of a genomic clone from *Drosophila melanogaster* which contains the sequences encoding the DUb52 tail fusion protein. We also show that the *DUb52* promoter region is very similar to those of the *Drosophila* rp genes. This observation is consistent with the possibility that DUb52 expression could be regulated as other rps [23,24].

MATERIALS AND METHODS

Cell culture

S3 cells from *D. melanogaster* [29] were cultured in M3 media [30] supplemented with heat-inactivated 10% bovine fetal serum. The cultures were always grown in Nunc flasks at 24 °C.

Screening of recombinant libraries

A library consisting of *Sau*3A-partially-digested *D. melanogaster* genomic DNA cloned into the *Bam*HI site of phage EMBL3 was obtained from Clontech (Palo Alto, CA, U.S.A.) and screened according to the manufacturer's recommended protocols. *Escherichia coli* host strains and manipulations involving recombinant phages were as previously described [31].

DNA manipulations and enzyme reactions

Restriction endonucleases were used following the manufacturer's instructions. Double-stranded DNA was ^{32}P-labelled

Abbreviations used: rp, ribosomal protein; TDEF, TFIIIA distal element factor; poly(A)$^+$, polyadenylated.

* Present address: Plant Biology Laboratory, Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, CA 92037, U.S.A.
† To whom correspondence should be addressed.
‡ Present address: Departamento de Química Orgánica, Inorgánica y Bioquímica, Facultad de Ciencias Químicas, Universidad de Castilla la Mancha, Paseo de la Universidad 4, 13071 Ciudad Real, Spain.

The nucleotide sequence data reported will appear in the EMBL Nucleotide Sequence Database under the accession number X59943.

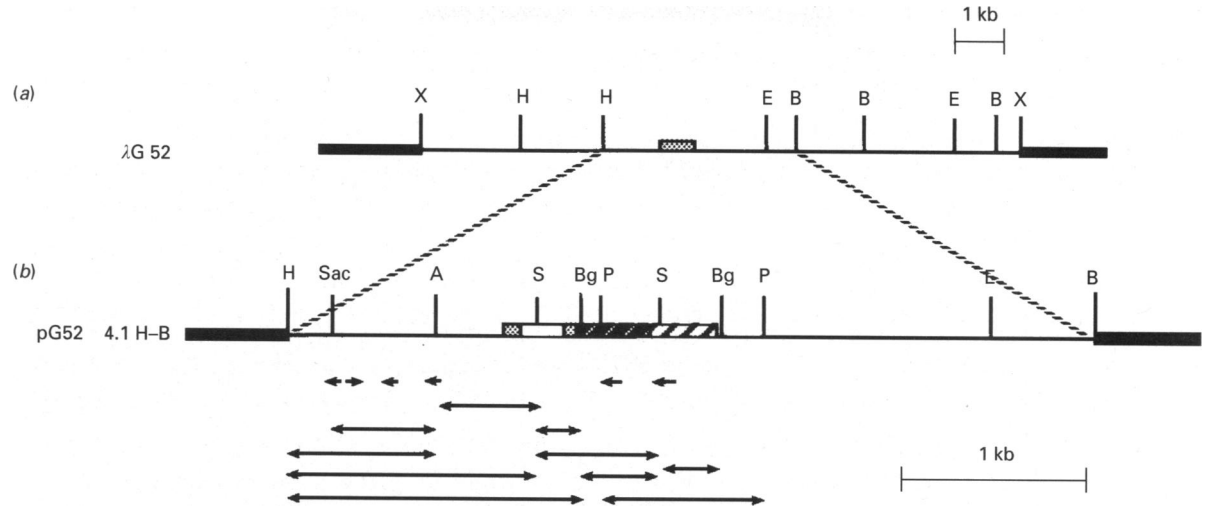

Fig. 1. (*a*) Physical map of the 12.5 kb genomic DNA fragment present in the clone λ G52 and (*b*) physical map of the subclone pG52 4.1, a *Hin*dIII–*Bam*HI positive fragment 4.1 kb long

The boxed region in (*a*) indicates the coding region of the ubiquitin-fusion gene. In (*b*) the black box represents the ubiquitin region and the striped box ▨ the tail-fusion encoding region. The intron is indicated by a white box. Stippled boxes (▩) represent the 5′ non-coding region present in the cDNA. The strategy for sequencing is shown by horizontal arrows, indicating the extension of the subclones and the direction of each sequence reading. Smaller arrows represent synthetic oligonucleotides employed for sequencing. A, *Acc*I; B, *Bam*HI; Bg, *Bgl*II; E, *Eco*RI; H, *Hin*dIII; P, *Pvu*II; S, *Sal*I; Sac, *Sac*I; X, *Xho*I.

by nick translation, by the procedure of Rigby *et al.* [32]. DNA inserts were isolated and subcloned in pUC18 as described in Maniatis *et al.* [31,33] and their nucleotide sequences were determined by the dideoxy method [34], by using T7 DNA polymerase.

Southern-blot hybridization

Genomic DNA was isolated from *D. melanogaster* adults (Canton S) as previously described [26]. The DNA samples were digested with restriction endonucleases, electrophoresed in 0.8% agarose gels, transferred to Nylon filters (Schleicher and Schuell) and hybridized as described by Maniatis *et al.* [31].

Northern-blot hybridization

D. melanogaster total RNA from different developmental stages and tissues was isolated as described by Arribas *et al.* [26]. Total RNA from cell cultures was extracted as described by Bunch *et al.* [35]. Polyadenylated RNA was isolated by using oligo(dT) cellulose (Boehringer), following the manufacturer's instructions. RNA samples were denatured in formaldehyde, electrophoresed in agarose/formaldehyde gels [31] and electro-transferred to Nylon membranes. Filters were prehybridized and hybridized as described by Arribas *et al.* [26]. To quantify the intensity of hybridization bands, RNA autoradiographs were analysed with a laser densitometer (Image Quant 2.0).

Primer-extension analysis

For primer extension, a synthetic 15-nucleotide primer (RB6), complementary to bases 1892–1906 (Fig. 2) (corresponding to the first 15 nucleotides of the tail sequence), was prepared. The oligonucleotide was labelled at its 5′ end with T4 polynucleotide kinase [33]. The primer-extension reaction was achieved as described by Ghosh *et al.* [36].

RESULTS

Genomic structure of the *DUb52* ubiquitin-fusion gene

The isolation of a *D. melanogaster* cDNA clone (pUCUb52), corresponding to a ubiquitin–52-amino acid fusion protein, has been previously reported [28]. For these studies, pUCUb52 was digested with *Rsa*I, and the resulting fragment (nucleotides 263–463, see ref. [28]) was subcloned into the *Sma*I site of pUC18. The new plasmid, p52Ub$^-$, containing only the tail portion of pUCUb52, was used as a probe to screen a *D. melanogaster* genomic library to allow the characterization of the *Drosophila* tail fusion gene. Out of 60000 recombinant phages, we obtained one positive clone, λ G52, that contained an insert of approx. 12.5 kb. This clone hibridized with both the tail and the ubiquitin-specific probes. λ G52 DNA was further digested with *Hin*dIII and *Bam*HI restriction endonucleases. The resulting 4.1 kb positive fragment was subcloned in pUC18. This plasmid, pG52 4.1, was mapped with restriction endonucleases (Fig. 1). The interesting fragments were subcloned and analysed by sequencing.

We sequenced 2306 bp of the clone pG52 4.1 (Fig. 2). The coding region extends from the nucleotides 1663 to 2046. The end of the ubiquitin coding region is located at nucleotide 1890. The non-coding regions are very AT-rich. Thus AT represents the 61% of total nucleotides in the intron region, the 59% in the 3′ region and the 75% from nucleotide 636–1326 in the 5′ region. On the other hand the coding region is enriched in GCs.

The intron region. The intron is 262 bp long (lower case in Fig 2). It is located in the 5′ non-coding region, and extends up to 13 bp upstream to the ATG initiation codon. This feature is common to other ubiquitin genes in other species [27,28,37,38]. The intron obeys the *D. melanogaster* consensus sequences [39] (italics in Fig. 2). The 5′ and 3′ splice junctions are located at nucleotides 1388 and 1649, respectively. We identified the branch point sequence at nucleotide 1626, approx. 30 nucleotides upstream from the 3′ splicing site [40] (italics in Fig. 2). The boundaries of the intron and the consensus sequences are coincident with the cDNA sequence of *DUb52* [28].

The 5′ regulatory region. The rp genes in *Drosophila*, *Xenopus* and mouse share a common feature: a pyrimidine-rich region flanking the transcription initiation site, flanked by GC-rich regions (twice underlined in Fig. 2) (reviewed in refs. [41] and [42]). In fact, we can see a CT-rich region at nucleotide 1337, preceded by a GC-rich stretch (nucleotide 1323). In the human

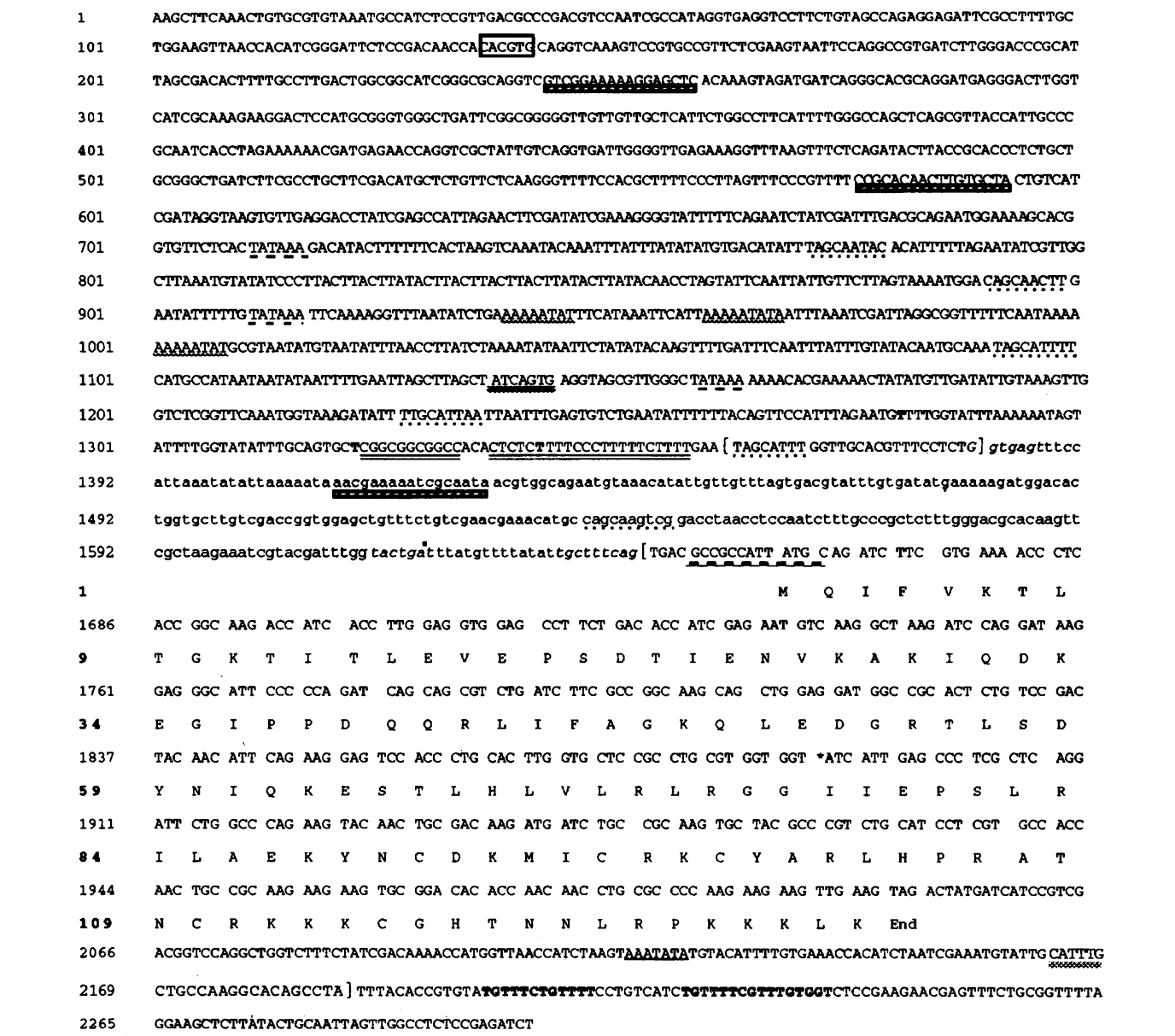

Fig. 2. Nucleotide sequence of the *DUb52* gene

Nucleotides are numbered from a *Hin*dIII site upstream of the gene. Amino acids are indicated by bold numbers. The junction between ubiquitin and tail protein is shown by an asterisk. The fragment overlapping the cDNA sequence is surrounded by square brackets. The 5′ intron is in lower case. The consensus 5′ and 3′ splicing and branch-point sequences are in italics. The *A* implicated in the splicing reaction is marked by a dot. The CT and GC stretches are doubly underlined. The different transcription start points are in **bold** letters. The *Drosophila* transcription start consensus sequence (ATCAGTG) is underlined by a shaded line. The TATA-boxes are marked by broken lines. The TATA-like elements are underlined by wavy lines. The CAAT-boxes are underlined by dots. The elements similar to the Ribo-box sequences of *N. crassa* are underlined by striped boxes. The sequence similar to the binding site of the *Xenopus* TDEF is boxed. The translation start consensus sequence is underlined by thick broken lines. The polyadenylation-like signal (AAATATA) is marked by a single line, and the TG sequence is in bold letters. The other 3′ end formation signal (CATTTG) is marked by a shaded line.

UbA52, the cap site is found in a palindromic pyrimidine tract, also very close to the intron start point [17]. On the other hand, our data of transcription start studies (see below) by primer-extension analysis indicate that three different transcription initiation points seem to be present (Fig. 3). The first one is at thymidine 1342, in the CT stretch. The other two points are located 20 and 62 nucleotides upstream. These data are in agreement with those reported for other ubiquitin-fusion proteins [17] and *Drosophila* rp genes [43]. However, we can also see a *D. melanogaster* transcription start consensus sequence, ATCAGTG [44], at the nucleotide 1137 (indicated by a shaded line in Fig. 2). The meaning of this sequence is so far unknown.

With regard to regulatory elements in the 5′ region, it has been described in *Drosophila* that rp genes lack the canonical TATA-box sequence 20–30 nucleotides upstream from the transcription start site, but they contain TATA-like elements at unusual positions. We have found TATA-like elements, AAATATA [45], at positions 938, 960 and 1001 (wavy lines in Fig. 2). There are also three canonical TATAAA sequences at nucleotides 711, 911 and 1158 (broken lines in Fig. 2).

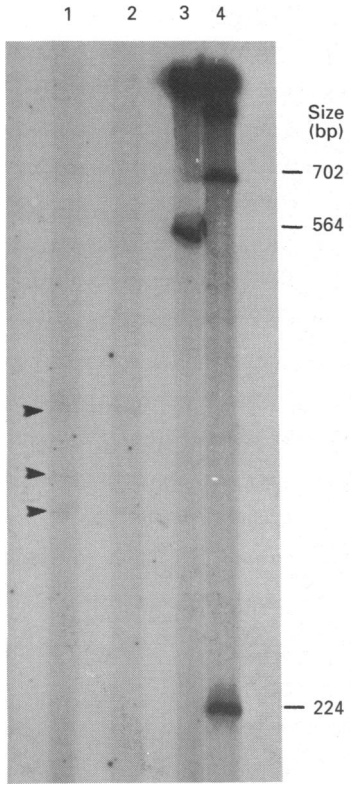

Fig. 3. Primer-extension analysis

A 15 bp-pair long synthetic oligonucleotide, complementary to nucleotides 1891–1905 (5′-CGAGGGCTCAATGAT-3′) was 5′ end labelled and annealed to 10 μg (lane 1) and 5 μg (lane 2) of poly(A$^+$) RNA from *D. melanogaster* (Canton S) adult flies. Primer-extended products are indicated by arrowheads. Size markers are ^{32}P-labelled phage λ DNA digested with *Hin*dIII (lane 3) and *Bst*EII (lane 4). Their molecular sizes are given.

Table 1. Nucleotide sequences in *DUb52* gene similar to the consensus Ribo-box in *N. crassa*

Position	Sequence	Mismatches with the consensus
	c	
243	GT GGAAAAa-GG-AGCTC	3
577	CcGcAcAACTtG-tGCTa	6
1417	aacGAAAAaTCGCA-ATa	6
Ribo-box	C G C C G A	
consensus	TAGAAAA T GG G TC	
	G T G G A C	

Multiple CAAT-box [46] similarities have been found at nucleotides 771, 891, 1092, 1227, 1362 and 1544 (marked by dots in Fig. 2). These sequences are quite similar to the proposed *Drosophila* CAAT consensus sequence A(A/T)GCA(A/T)-N(A/T)N, which differs slightly from the mammalian consensus sequence [47].

At nucleotides 243, 577 and 1417 (indicated by striped lines in Fig. 2), we detected sequences identical with the consensus Ribo-box described in *N. crassa* [48]. The sequences are given in Table 1.

In the *crp-1* and *crp-2* genes of *N. crassa*, three Ribo-box sequences were found [48]. One of them overlaps the transcription start point. In our case, one of them was located inside the intron, and the others were positioned far upstream from the cap sites.

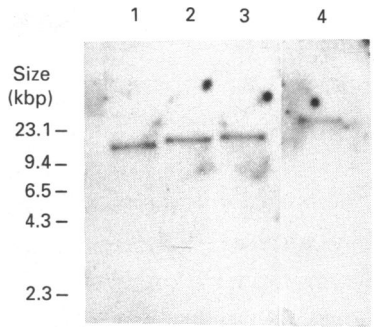

Fig. 4. Genomic Southern-blot analysis of the ubiquitin-fusion gene *DUb52*

Genomic DNA from *D. melanogaster* (Canton S) adult flies (8 μg) was digested with *Eco*RI (lane 1), *Bam*HI (lane 2), *Hin*dIII (lane 3) and *Xho*I (lane 4), and hybridized with the 0.2 kb tail-specific clone p52Ub$^-$. Size markers were fragments from wild-type phage λ DNA digested with *Hin*dIII.

At 1200 bp upstream from the CT stretch, we found a sequence (boxed in Fig. 2) similar to the binding site of TFIIIA distal element factor (TDEF) of *Xenopus laevis* [49]. This cis-element (CACGTG) stimulates transcription from the TFIIIA promoter in oocytes but not in somatic cells. We do not know if this sequence has any physiological relevance in the *DUb52* gene.

The 3′ non-coding region. At the 3′ end of the sequence, we found a polyadenylation-like sequence AATATA (nucleotide 2118, indicated by a single underline in Fig. 2) [47], but we could not find other sequences involved in the efficient 3′ end formation [50,51]. However, there are related sequences at positions 2164 [51], 2201 and 2222 [50]. At approx. 100 bp downstream from the AATATA, we observed a GT-rich element (bold letters in Fig. 2). In yeast, it has been shown that the factor CF1 binds this element, which is necessary for the stabilization of the poly-adenylation machinery [52]. In *DUb52*, the exact site of polyadenylation has not been determined, but it is likely that it might be far downstream from this sequence, bearing in mind the size of the *DUb52* mRNA (see below).

The coding region. The sequence corresponding to the coding region coincides with that previously described in cDNA pUCUb52 [28]. The ATG translation initiation codon is included in a *Drosophila* consensus sequence GCCGCC(A/G)CC**ATG**G [53]. The purine 5′ upstream of the ATG codon is conserved, as well as the cytosines at positions -4 and -5, although, at positions -1 and -2, they are replaced by thymidines (underlined by dots in Fig. 2).

Southern-blot analysis. In order to determine whether the *DUb52* gene is present as a single-copy gene, Southern blots of digested *D. melanogaster* Canton S DNA were hybridized with ^{32}P-labelled DNA p52Ub$^-$ (Fig. 4). Single hybridization bands, ranging between 9 and 23 kb, were obtained when DNA samples were digested with *Eco*RI, *Bam*HI, *Hin*dIII and *Xho*I enzymes (see Fig. 1). These data suggest that *DUb52* gene exists as a single copy in the *Drosophila* genome.

Pattern of ubiquitin-fusion gene expression

***DUb52* encodes a 0.9 kb RNA.** Polyadenylated [poly(A)$^+$]RNA from different developmental stages of *D. melanogaster* was isolated and analysed by hybridization to a labelled tail-specific DNA p52Ub$^-$. Fig. 5 shows one single band of 0.9 kb in embryonic, larval, pupal and adult RNA. This RNA is

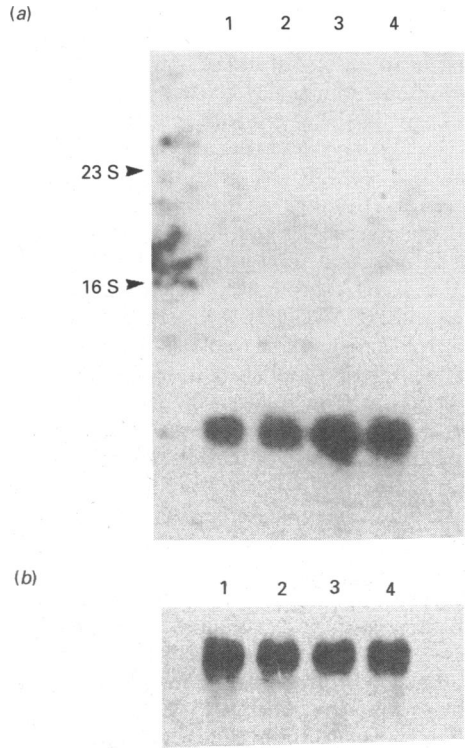

Fig. 5. Northern-blot analysis of different developmental stages from *D. melanogaster* (Canton S)

(*a*) Poly(A$^+$) RNA (2 μg/lane) was isolated from embryos (lane 1), third-instar larvae (lane 2), pupae (lane 3) and adult flies (lane 4), and hybridized with tail-specific plasmid p52Ub$^-$. The arrowheads show the position of 23S and 16S molecular-size marker rRNAs from *E. coli*. (*b*) The same filter was hybridized to an α-tubulin-specific probe, to check for RNA loading.

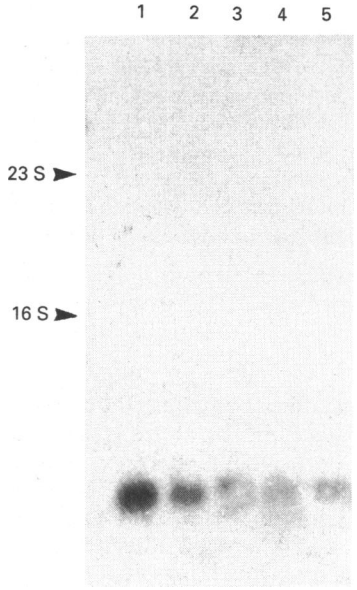

Fig. 6. Northern-blot analysis of different tissues from *D. melanogaster*

Total RNA (2.5 μg) from ovaries (lane 1) and 5 μg of total RNA from Malpigian tubules and intestines (lane 2), imaginal discs (lane 3), fat bodies (lane 4) and salivary glands (lane 5) were hybridized to the tail-specific probe p52Ub$^-$.

constitutively expressed throughout development as can be observed when comparing the intensity of the bands from different developmental stages (Fig. 5).

As Lee *et al.* [20] had previously observed, we also detected specific hybridization of the ubiquitin–80 amino acid tail fusion cDNA to a 0.9 kb mRNA (results not shown). Thus it appears that both genes transcribe a 0.9 kb messenger (see the Discussion section).

Differential expression of *DUb52* gene. To determine the distribution of ubiquitin-fusion gene expression, total RNA from Malpigian tubules and intestine-associated tissues, imaginal discs, fat bodies and salivary glands were isolated from third-instar larvae. In addition, total RNA from adult ovaries was analysed. Samples were fractionated, blotted and hybridized to the p52Ub$^-$ probe. Fig. 6 shows a single band of 0.9 kb which is abundantly expressed in all the samples. However, the level of expression varied depending on the tissue examined. From densitometer analysis it can be deducted that the steady-state concentrations of the transcript for the ubiquitin–52-amino acid fusion gene in ovaries is at least four times higher than in other tissues analysed.

In order to determine whether *DUb52* gene expression is altered depending on the growth conditions, RNA isolated from *Drosophila* cell cultures growing in exponential and stationary phases were analysed. Filters were then hybridized to the tail-specific probe p52Ub$^-$ (Fig. 7*a*). The 52-amino acid fusion-protein transcript is almost undetectable in stationary-phase cells, whereas its concentration during the exponential phase is 15-fold increased. A similar phenomenon has been observed in

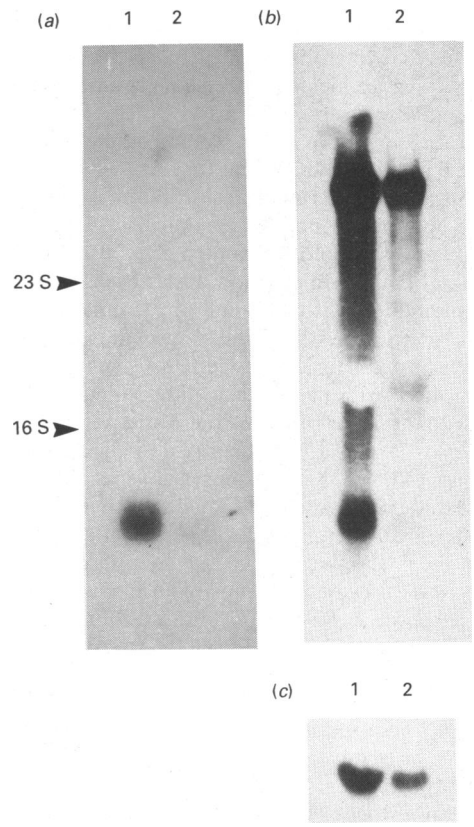

Fig. 7. Northern-blot analysis of RNA from exponentially (lane 1) and stationary (lane 2)-phase growing *Drosophila* cultured cells

Total RNA from 10^6 cells was hybridized to the tail-specific probe p52Ub$^-$ (*a*), ubiquitin probe pUC63F [24,25] (*b*) and α-tubulin-specific probe (*c*).

the expression of the 52-amino acid fusion tail from *S. cerevisiae* [12], and *Trypanosoma cruzi* [16]. However, the hybridization to the polyubiquitin gene pUC63F [25,26] reveals that the polyubiquitin expression is not apparently altered by growth conditions (Fig. 7*b*). Thereby, and despite the results observed with both *Saccharomyces* and *Trypanosoma* [12,16], where the transcription of the polyubiquitin gene is stimulated during stationary-phase cultures, in *Drosophila* this expression does not appear to be significantly induced under these conditions.

DISCUSSION

The isolation and sequencing of a cDNA clone coding for the 52-amino acid ubiquitin-fusion protein DUb52 has previously been reported [28]. In the present work, we describe the characterization of a genomic clone from *D. melanogaster* encoding this ubiquitin-fusion protein (*DUb52*). The amino acid sequence of the extension proteins is very conserved from yeast to humans. This finding could indicate a common function of these proteins among eukaryotes. Recently, it has been shown that these are ribosomal proteins in yeast [23], mammals [24] and other organisms [18,54].

Structure of the *DUb52* gene

In some organisms, ubiquitin-fusion protein genes belong to a multigene family. Moreover, the sequence that is identical with that in *DUb52* in yeast [12], *Arabidopsis* [18], barley [55] and humans [17] is present in more than one copy per haploid genome. However, our Southern-blot analysis results suggest that in the genome of *D. melanogaster* the 52-amino acid ubiquitin-fusion protein is encoded by only one gene (Fig. 4).

The *DUb52* genomic sequence shares some similarities with those of ubiquitin-fusion proteins in other organisms and also with other *Drosophila* rp genes. Some common features among higher eukaryote rp genes (except for human) are the presence of a pyrimidine-rich transcriptional start point, surrounded by GC-rich boxes; and the absence of a canonical TATA-box [41]. These features are also seen in other non-rp genes transcribed by RNA polymerase II [56,57]. In addition, it has been described that the oligopyrimidine tract at the 5′ end of mammalian rp mRNA is important for translational control mechanisms [58]. Likewise, the human ubiquitin-fusion protein UbA_{52} gene [17] presents the cap site located within a 13-nucleotide palindromic pyrimidine tract, close to the 5′ end of the intron. Moreover, some *Drosophila* rp genes also start in a CT-rich stretch [43], although other rp genes do not [47,59]. We also show here that *DUb52* presents a CT-rich region preceded by a GC-rich stretch in its 5′ region. Our primer-extension experiments locate three putative cap sites (Fig. 3). One of them is in the CT-rich stretch, although the other two points are several nucleotides upstream from this sequence. On the other hand, a *Drosophila* transcription start consensus sequence [44] can be observed 186 bp upstream from the CT tract. The significance of this sequence is, however, unknown. Similarly, we found multiple TATA and TATA-like boxes [45], even though these sequences are too distant to be significant. There were also multiple CAAT-like boxes [47], as in other *Drosophila* rp genes.

A different element occurring in the *DUb52* gene is a region similar to the so-called Ribo-box sequence of *N. crassa* [48]. The co-ordinated expression of ribosomal components involves all three RNA polymerases. The Ribo-box sequence is present in the 5S promoter transcribed by the RNA polymerase III, in the 40S promoter transcribed by the RNA polymerase I, and in the *crp-1* and *crp-2* rp genes, transcribed by RNA polymerase II. It has been suggested that the Ribo-box sequence may co-ordinate transcription of rp genes with those of 5S and 40S rRNA genes [48]. It should be interesting to test the function of these sequences, if any, in the *Drosophila* genome.

Hall & Tylor [49] have described a transcription factor, TDEF, which stimulated the TFIIIA gene transcription threefold in *Xenopus* oocytes. Although TDEF can bind to DNA in both oocytes and somatic cells, in somatic cells the TFIIIA gene does not exhibit an increased transcription. TDEF binds to a distal region in the TFIIIA promoter, which includes a 6 bp palindromic sequence (CACGTG). A yeast cell protein [60] and the adenovirus type-2 major late transcription factor [61] possess DNA-binding properties similar to those of TDEF. Interestingly, we can also find this element, even though it is located far away from the transcription start sites (Fig. 2). The presence of this sequence and the possibility of the existence of an ovary-specific transcription factor in *Drosophila* are in agreement with our expression results. Thus, as we can see in Fig. 6, the *DUb52* mRNA concentration is higher in ovaries, at least four times with respect to other tissues analysed.

The amino acid sequences of the ubiquitin tails have been extremely conserved during evolution, suggesting an important role for these proteins in eukaryotes. In yeast [23] and mammals [24], these extension proteins are located in the cytoplasm, where they are contained in the ribosomes. This location could explain the presence of a nucleic acid-binding domain in the primary structure of both Ub52 and Ub80. The function of this domain could be binding or interaction with mRNA or rRNA molecules within the ribosome.

It has been suggested that these ribosomal tails, as well as other ribosomal proteins, are extremely short-lived if they are not properly assembled in the ribosome [62]. Hence, the transient binding of ubiquitin to ribosomal-extension proteins could increase the rate at which the ribosomal tails are transported and/or assembled into the ribosome. This phenomenon could decrease the proportion of tails synthesized *de novo* that should be degraded during the transition to the ribosome [23].

Expression of the *DUb52* gene

Analysis of *DUb52* gene expression reveals the presence of only one 0.9 kb mRNA constitutively expressed in all developmental stages (Fig. 5). Previous reports [20] had shown that the *Ub80* fusion gene in *Drosophila* also encodes an mRNA with a similar molecular size. Our findings confirm this observation (results not shown). In humans, both genes homologous to *DUb52* and *DUb80* [17] also have the same transcript size.

However, the *DUb52* gene expression varies considerably among different larval and adult tissues. Thus, adult ovaries show a high accumulation of this fusion-protein mRNA as compared with other tissues examined (Fig. 6). This result is in accordance with the existence of a high rate of ribosome synthesis during *Drosophila* oogenesis. A *Drosophila* egg contains about 5×10^{10} ribosomes [63]. Therefore it is not surprising to find a higher level of ribosomal proteins in ovaries. A similar result has been observed in *Manduca sexta* [21].

It has been proposed in yeast that the fusion genes represent the major source of ubiquitin when ribosome synthesis is active, i.e. during the exponential growth phase. However, the yeast polyubiquitin gene is activated later in the cell cycle, during nitrogen starvation and under cellular stress conditions [12,54]. In *Drosophila* cell culture, this differential expression occurs in the *DUb52* gene, but does not occur in the polyubiquitin gene. In fact, cells growing in the exponential phase accumulate the *DUb52* transcript very actively, whereas it is repressed when cells enter the stationary phase (Fig. 7). In contrast, the polyubiquitin

gene does not seem to be differentially regulated during the two growth conditions. Additionally, it had previously been observed that the *Drosophila* polyubiquitin gene is not significantly induced to transcribe under heat-shock conditions [25]. The data reported here confirm this observation. However, Lee *et al.* [20] found a threefold increase in polyubiquitin concentrations in heat-shocked cells. Clearly, in *Drosophila* the heat-shock induction of polyubiquitin transcript concentrations does not reach the level observed in other species [12,27], where ubiquitin is considered a major heat-shock protein. In *Caenorhabditis elegans*, Graham *et al.* [38] did not find significant changes in polyubiquitin expression on temperature elevation. Both organisms contain a large number of ubiquitin repeats in the polyubiquitin mRNA (11 in *Caenorhabditis* and 15–18 in *Drosophila*), which could provide enough ubiquitin moieties to replenish the depleted concentrations of free ubiquitin during heat-shock conditions. However, other organisms have to increase the rate of transcription in order to provide enough ubiquitin to reach the appropriate concentrations in the cell.

In conclusion, in *Drosophila* the genes encoding ubiquitin belong to a multigene family. In this work we have contributed to its characterization by isolating and analysing a genomic clone that encodes a ubiquitin-fusion protein and a 52-amino acid ribosomal protein. Until now, the ubiquitin genes of the ubiquitin multigene family isolated in *Drosophila* were the polyubiquitin gene, characterized in our laboratory [25,26,64], and the 80-amino acid fusion gene [20]. With the *DUb52* gene, the repertoire of sequences encoding ubiquitin in *Drosophila* seems to be complete.

We are very grateful to Juan García, Araceli del Arco and Javier Llanos for their expert technical assistance, and to Lucía Herranz and Peter J. Boynton for a critical reading of the manuscript. This work was performed at Marta Izquierdo laboratory and was supported by a grant from the D.G.I.C.Y.T. (Spain). R.B. is a fellow from the Gobierno Vasco.

REFERENCES

1. Rechsteiner, M. (1988) Ubiquitin. Plenum Press, New York
2. Levinger, L. & Varshavsky, A. (1982) Cell **28**, 375–385
3. Thorne, A. W., Sautiere, P., Briand, G. & Crane-Robinson, C. (1987) EMBO J. **6**, 1005–1010
4. Varshavsky, A., Levinger, L., Sundin, L., Barsoum, J., Özkaynak, E., Swerdlow, P. & Finley, D. (1983) Cold Spring Harbor Symp. Quant. Biol. **47**, 511–528
5. Ball, E., Karlik, C. C., Beall, C. J., Saville, D. L., Sparrow, J. C., Bullard, B. & Fyrberg, E. A. (1987) Cell **51**, 221–228
6. Siegelman, M., Bond, M. W., Gallantin, W. M., St. John, T., Smith, H. T., Fried, V. A. & Weissman, I. L (1986) Science **231**, 823–829
7. Yarden, Y., Escobedo, J. A., Kuang, W.-J., Yang-Feng, T. L., Daniel, T. O., Tremble, P. M., Chen, E. Y., Ando, M. E., Harkins, R. N., Francke, U., Friend, V. A., Ullrich, A. & Williams, L. T. (1986) Nature (London) **323**, 226–232
8. Leung, D. W., Spencer, S. A., Cachianes, G., Hammonds, R. G., Collins, C., Henzel, W., Barnard, R., Waters, M. J. & Wood, W. I. (1987) Nature (London) **330**, 537–543
9. Özkaynak, E., Finley, D. & Varshavsky, A. (1984) Nature (London) **312**, 663–666
10. Wiborg, O., Pedersen, M. S., Wind, A., Berglund, L. E., Marcker, K. A. & Vuust, J. (1985) EMBO J. **4**, 755–759
11. Giorda, R. & Ennis, H. L. (1985) Mol. Cell. Biol. **6**, 2097–2103
12. Özkaynak, E., Finley, D., Solomon, M. J. & Varshavsky, A. (1987) EMBO J. **6**, 1429–1439
13. Finley, D., Özkaynak, E. & Varshavsky, A. (1987) Cell **48**, 1035–1046
14. Salvensen, G., Lloyd, C. & Farley, D. (1987) Nucleic Acids Res. **15**, 5485
15. Ohmachi, T., Giorda, R., Show, D. R. & Ennis, H. L. (1989) Biochemistry **28**, 5226–5231
16. Swindle, J., Ajioka, J., Eisen, H., Sanwal, B., Jacquemot, C., Browder, Z. & Buck, G. (1988) EMBO J. **7**, 1121–1127
17. Baker, R. T. & Board, P. G. (1991) Nucleic Acids Res. **19**, 1035–1040
18. Callis, J., Raasch, J. A. & Vierstra, R. D. (1990) J. Biol. Chem. **265**, 12486–12493
19. Lund, P. K., Moats-Staats, B. M., Simmons, J. G., Hoy, E., D'Ercole, A. J., Martin, F. & Van Wyk, J. J. (1985) J. Biol. Chem. **260**, 7609–7613
20. Lee, H., Simon, J. A. & Lis, J. T. (1988) Mol. Cell. Biol. **8**, 4727–4735
21. Bishoff, S. T. & Schwartz, L. M. (1990) Nucleic Acids Res. **18**, 6039–6043
22. Redman, K. L. & Rechsteiner, M. (1988) J. Biol. Chem. **263**, 4926–4931
23. Finley, D., Bartel, B. & Varshavsky, A. (1989) Nature (London) **338**, 394–401
24. Redman, K. L. & Rechsteiner, M. (1989) Nature (London) **338**, 438–440
25. Izquierdo, M., Arribas, C., Galceran, J., Burke, J. & Cabrera, V. M. (1984) Biochim. Biophys. Acta **783**, 114–121
26. Arribas, C., Sampedro, J. & Izquierdo, M. (1986) Biochim. Biophys. Acta **868**, 119–127
27. Bond, U. & Schlesinger, M. J. (1986) Mol. Cell. Biol. **6**, 4602–4610
28. Cabrera, H. L., Arribas, C. & Izquierdo, M. (1990) Nucleic Acids Res. **18**, 3994
29. Schneider, I. (1972) J. Embryol. Exp. Morphol. **27**, 353–365
30. Shields, G. & Sang, J. H. (1977) Dros. Inf. Serv. **52**, 161
31. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
32. Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. (1977) J. Mol. Biol. **113**, 237–251
33. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
34. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463–5467
35. Bunch, T. A., Grinblat, Y. & Goldstein, L. S. B. (1988) Nucleic Acids Res. **16**, 1043–1059
36. Ghosh, P., Reddy, V., Piatak, M., Lebowitz, P. & Weissman, S. (1980) Methods Enzymol. **65**, 580–595
37. Baker, R. T. & Board, P. G. (1987) Nucleic Acids Res. **15**, 443–463
38. Graham, R. W., Jones, D. & Candido, E. P. M. (1989) Mol. Cell. Biol. **9**, 268–277
39. Snyder, M., Hunkapiller, M., Yuen, D., Silvert, D., Fristrom, J. & Davidson, N. (1982) Cell **29**, 1027–1040
40. Smith, C. W. J., Patton, J. G. & Nadal-Guinard, B. (1989) Annu. Rev. Genet. **23**, 527–577
41. Mager, W. H. (1988) Biochim. Biophys. Acta **949**, 1–15
42. Larson, D. E., Zahradka, P. & Sells, B. H. (1991) Biochem. Cell Biol. **69**, 5–22
43. Qian, S., Zhang, J.-Y., Kay, M. A. & Jacobs-Lorena, M. (1987) Nucleic Acids Res. **15**, 987–1003
44. Hultmark, D., Klemenz, R. & Gehring, W. J. (1986) Cell **44**, 429–438
45. Voelker, R. A., Gibson, W., Graves, J. P., Sterling, J. F. & Eisenberg, M. T. (1991) Mol. Cell. Biol. **11**, 894–905
46. Benoist, C., O'Hare, K., Breathnach, R. & Chambon, P. (1980) Nucleic Acids Res. **8**, 127–142
47. O'Connell, P. & Rosbash, M. (1984) Nucleic Acids Res. **12**, 5495–5513
48. Tyler, B. M. & Harrison, K. (1990) Nucleic Acids Res. **18**, 5759–5765
49. Hall, R. K. & Tylor, W. L. (1989) Mol. Cell. Biol. **9**, 5003–5011
50. McLauchlan, J., Gaffney, D., Witton, J. L. & Clements, J. B. (1985) Nucleic Acids Res. **13**, 1347–1368
51. Urano, Y., Watanabe, K., Sakai, M. & Tamaoki, T. (1986) J. Biol. Chem. **261**, 3244–3251
52. Gilmartin, G. M. & Nevins, J. R. (1991) Mol. Cell. Biol. **11**, 2432–2438
53. Feng, Y., Gunter, L. E., Organ, E. L. & Cavener, D. R. (1991) Mol. Cell. Biol. **11**, 2149–2153
54. Müller-Taubenberger, A., Graack, H. R., Grohmann, L., Schleicher, M. & Gerisch, G. (1989) J. Biol. Chem. **264**, 5319–5322
55. Gausing, K. & Jensen, C. B. (1990) Gene **94**, 165–171
56. Tamura, T. & Mikoshiba, K. (1991) FEBS Lett. **282**, 87–90
57. Jolliff, K., Li, Y. & Johnson, L. F. (1991) Nucleic Acids Res. **19**, 2267–2274

58. Levy, S., Avni, D., Hariharan, N., Perry, R. P. & Meyuhas, O. (1991) Proc. Natl. Acad. Sci. U.S.A. **88**, 3319–3323
59. Brown, S. J., Rhoads, D. D., Stewart, M. J., Van Slyke, B., Chen, I.-T., Johnson, T. K., Denell, R. E. & Roufa, D. J. (1988) Mol. Cell. Biol. **8**, 4314–4321
60. Chodosh, L. A., Buratowski, S. & Sharp, P. A. (1989) Mol. Cell. Biol. **9**, 820–822
61. Carthew, R. W., Chodosh, L. A. & Sharp, P. A. (1985) Cell **43**, 439–448
62. Planta, R. J. & Raué, H. A. (1988) Trends Genet. **4**, 64–68
63. Houng-Evans, B. R., Jacobs-Lorena, M., Cummings, M. R., Britten, R. J. & Davidson, E. M. (1980) Genetics **95**, 81–94
64. Izquierdo, M., Arribas, C. & Alonso, C. (1981) Chromosoma **83**, 353–366

Received 7 August 1991/3 January 1992; accepted 4 February 1992