In the fruitfly Drosophila, as in all eukaryotes examined so far, some ubiquitin-coding sequences appear fused to unrelated open reading frames. Two of these fusion genes have been previously described (the homologues of UB11-UB12 and UB14 in yeast), and we report here the organization and expression of a third one, the DUb80 gene (the homologue of UB13 in yeast). This gene encodes a ubiquitin monomer fused to an 80-amino-acid extension which is homologous with the ribosomal protein encoded by the UB13 gene. The 5' regulatory region of DUb80 shares common features with another ubiquitin fusion gene, DUb52, and with the ribosomal protein genes of Drosophila, Xenopus and mouse. We also find helix-loop-helix protein-binding sequences (E-boxes). The DUb80 gene is transcribed to a 0.9 kb mRNA which is particularly abundant under conditions of high protein synthesis, such as in ovaries and exponentially growing cells.

INTRODUCTION

Ubiquitin is a highly conserved 76-amino-acid protein present in all eukaryotes [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]. Ubiquitin also appears conjugated with certain nuclear [2-4], cytoplasmic [5] and cell-surface proteins [6-8] without causing their degradation. This raises the possibility that conjugation of ubiquitin to proteins may sometimes exert regulatory functions.

The ubiquitin genes are fusion genes that fall into two structural types. First, a polyubiquitin gene consisting of a variable number of tandem repeats of a 228 bp protein-coding region. Expression of this gene produces a polyprotein as a primary translation product, which is subsequently processed into monomeric ubiquitin units by specific cytoplasmic proteinases [9]. Although the number of repeats in this gene varies among species, its organization is very conserved [9-11]. The second structural type of ubiquitin gene encodes hybrid proteins in which a ubiquitin monomer is fused in frame to unrelated sequences. These sequences encode C-terminal extensions of 52 (Ub52) or 76-80 (Ub80) amino acids [12,14-21]. Both Ub52 and Ub80 amino acid tails are predominantly basic [22], highly conserved among eukaryotes and, in all organisms examined so far, contain two specific regions: a putative nucleic-acid-binding domain and a nuclear-translocation signal [18,23,24]. The differences between these gene classes are mirrored in their specific functions: the polyubiquitin gene has been found to provide free ubiquitin proteins, especially under stress conditions [12,25], whereas the C-terminal extensions have been identified as ribosomal proteins (rp) in yeast [23] and mammals [24]. Although ubiquitin has not been found in the ribosome, it appears that its fusion to the N-terminus of these rps increases the efficiency of their incorporation into the ribosome [23].

In the fruitfly Drosophila, ubiquitin genes constitute a multi-gene family. The polyubiquitin gene has been isolated [26] and characterized in our laboratory [27] and by Lee et al. [20]. This gene is highly polymorphomic among different stocks and strains [27]. It contains either 15 or 18 tandem repeats, depending on the fruity stock analysed [20,26]. In Drosophila, unlike in most other organisms, it is not significantly induced under our heat-shock conditions [26]. The Drosophila ubiquitin-fusion genes have been also isolated and characterized. In previous reports we described the isolation and sequencing of a cDNA corresponding to a DUb52 ubiquitin-fusion gene [28], as well as the nucleotide sequence and expression of a genomic clone from this gene [17]. Additionally, a cDNA corresponding to the DUb80 ubiquitin-fusion gene has also been isolated and characterized by Lee et al. [20].

We report here the nucleotide sequence and expression of a genomic clone from D. melanogaster which contains the sequences encoding the DUb80 tail-fusion protein. The DUb80 promoter region contains similar regulatory elements to those found in the DUb52 ubiquitin-fusion gene, suggesting that both genes could be regulated in the same way as other Drosophila rps.

MATERIALS AND METHODS

Cell culture
S2 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.

Abbreviations used: rp, ribosomal protein; poly(A)*, polyadenylated.
* Present address: Institute of Molecular Biology and Biotechnology, HELLAS, P.O. Box 1527, Heraklion 71110, Crete, Greece.
† To whom correspondence should be addressed.
‡ Present address: Área de Biología Molecular y Virología Vegetal, CIT-INIA, carretera de La Coruña Km 7, 28049 Madrid, Spain.
§ Present address: Departamento de Química Inorgánica, Orgánica y Bioquímica, Facultad de Ciencias Químicas, Universidad de Castilla La Mancha, Ciudad Real, Spain.

The nucleotide sequence data reported will appear in the EMBL, GenBank and DDBJ Nucleotide Sequence databases under the accession number X69119.
Screening recombinant libraries

A library of *D. melanogaster* genomic DNA partially digested with restriction endonuclease Sau3A, cloned into the BamHI site of phage EMBL-3, was obtained from Clontech (Palo Alto, CA, U.S.A.). A cDNA library cloned into the EcoRI site of phage gt10 was obtained from Dr. T. B. Kornberg (University of California, San Francisco). The screening of both libraries was performed by standard procedures. *Escherichia coli* host strains and manipulations involving recombinant phages were as previously described [31].

DNA manipulations and enzyme reactions

Restriction endonucleases were used as described by the manufacturer. Double-stranded DNA was 32P-labelled, by nick translation, by the procedure of Rigby et al. [32]. DNA inserts were isolated and subcloned in pUC18 as described in Sambrook et al. [33], and their nucleotide sequence were determined by the dyeoxy method [34] using T7 DNA polymerase.

Northern-blot analysis

*D. melanogaster* total RNA from different developmental stages and tissues was isolated as described previously [27]. Total RNA from cell cultures was extracted as described by Bunch et al. [35]. Polyadenylated [poly(A)]*+* RNA was isolated with oligo(dT)–cellulose (Boehringer) as described by the manufacturer, and its concentration was measured by its absorbance at 260 nm. RNA samples were denatured in formaldehyde, electrophoresed in agarose/formaldehyde gels [33] and electrotransferred to Nylon membranes. Filters were prehybridized and hybridized as described by Arribas et al. [27]. 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 25-nucleotide primer, complementary to bases 1591–1615 (corresponding to the first 25 nucleotides of the tail sequence; see Figure 2 below), was prepared. The oligonucleotide was labelled at its 5' end with T4 polynucleotide kinase [33]. The primer-extension reaction was performed as described by Ghosh et al. [36].

Reverse transcriptase/PCR analysis

The cDNA-negative band was synthesized from adult-fruitfly total RNA [33] by using the reverse-transcriptase activity from avian myeloblastosis virus and poly(dT) (Boehringer) as primer. The reaction was performed in 100 μl, of which 5 μl were used as a template for PCR reactions. For the amplification reactions the following synthetic primers were prepared (see Figures 2 and 3a below): A, 17-nucleotide primer complementary to bases 1–17; B, 17-nucleotide primer complementary to bases 1405–1421; C, 25-nucleotide primer complementary to bases 1591–1615; D, 17-nucleotide primer complementary to bases 1717–1733; and E, 17-nucleotide primer complementary to bases 2031–2047. The PCR reactions were performed in a total volume of 50 μl with 0.5 unit of *Taq* DNA polymerase (Boehringer) under the conditions suggested by the suppliers. The template was denatured for 5 min at 94 °C, followed by 29 cycles of PCR with incubations of 1 min at 55 °C (annealing), 1 min at 72 °C (polymerization), 1 min at 94 °C (denaturation) and a final extension of 10 min at 72 °C. After reaction the samples were electrophoresed in 0.8 % agarose gels. The fragments of interest were subcloned in the Smal site of pUC18 and sequenced.

RESULTS

Genomic structure of the *Dub80* ubiquitin-fusion gene

We had previously isolated a *D. melanogaster* clone corresponding to the 3' end of the 80-amino-acid ubiquitin-fusion-protein gene (nucleotides 1640–1683 and 1768–2086; Figure 2 (below) and R. Barrio, H. L. Cabrera and C. Arribas, unpublished work). This partial cDNA, called p80Ub+, containing
only part of the coding sequence of the extension protein, was used as a probe to screen a D. melanogaster genomic library. One positive clone was obtained, \( \lambda G80 \), which contained an insert of approx. 14 kb. This phage was digested with XhoI, resulting in two fragments, X1 and X2. Fragment X2 hybridized with both the tail- and the ubiquitin-specific probes, and was subcloned in the SalI site of pUC18, resulting in the plasmid pG80X2 (Figure 1). Fragment X1 was located in the 3' end of X2 and was not further analysed. pG80X2 contained an insert of 4 kb which was mapped with restriction endonucleases and fully sequenced according to the strategy shown in Figure 1. However, pG80X2 lacked the 5' end of the gene, and the genomic library was again screened using the XhoI-BglII 5' fragment of pG80X2. A second positive phage was isolated, \( \lambda G806 \) (Figure 1), whose insert was digested with SauI and subcloned in pUC18. The resulting plasmid, pG806, was also mapped with restriction enzymes and fully sequenced, as shown in Figure 1. A total of 5245 bp were sequenced by both strands.

Intron regions

\( DUb80 \) contains two introns (lower case in Figure 2). Like \( DUb52 \) [17], it presents an intron in the 5' non-coding region (intron 1). This is a feature common to ubiquitin genes in other species, as well as to other Drosophila genes. \( DUb80 \) also presents an 184 bp intron (intron 2) in the tail-coding region. The 5' and 3' splicing sites, as well as the branch point sites, matched the D. melanogaster consensus sequences [37]. The existence of both introns has been proved by PCR (Figure 3a). The different size of the PCR products using genomic DNA (lanes 3' and 4') or cDNA (lanes 3 and 4) as templates reflected the presence of the introns 2 and 1 respectively (lanes 1 and 1' are positive controls, while lane 2 shows the negative control for the reactions). The presence of the small intron is also corroborated by the sequence of the cDNA isolated by Lee et al. [20], and by the partial cDNA clone previously isolated in our laboratory. Moreover, the product of the reaction shown in lane 4 was subcloned in pUC18 and sequenced to demonstrate the correct boundaries of the intron 1. Our genomic sequencing data were corroborated by the analysis of the different PCR clones, no difference in the sequence being found. However, our data differ from those of Lee et al. [20] in the 5' end of the sequence. This sequence presented an 'artefactual' inverted repetition of the 3' end of the cDNA placed in its 5' end (from nt 594 to nt 672 of the 3' end, and from nt 13 to nt 91 of the 5' end). Nevertheless, the following 20 nucleotides (nucleotides 92-113) from the first exon of \( DUb80 \) were well defined in the published sequence [20]. No other differences were found between the sequence found by Lee et al. [20] and that found by ourselves.

5' regulatory region

Primer-extension analysis (Figure 3b) indicated two transcription start areas. The first one contains two different points which are included in a CT-rich stretch (doubly underlined in Figure 2). The first point was arbitrarily called 'nucleotide 1'. \( DUb52 \), as well as other rp genes from Drosophila, Xenopus and mouse (reviewed by Mager [38] and Larson et al. [39]), also presents a CT-rich region which includes the transcription start points. This region is preceded by a GT-rich one. The second area is a region rich in TATA-like elements and includes four different transcription start points (nucleotides -118, -154, -168 and -219; shown in bold in Figure 2). Nevertheless, Northern-blot analysis revealed one single transcript which did not change along developmental stages or among different tissues. It is possible that those different transcription start points are used under different physiological conditions. We have not examined this point.

At nucleotides -1542, -952, -849 and -757 (indicated by striped lines in Figure 2), we detected sequences identical with the consensus 'Ribo-box' described in the crp-1 and crp-2 genes of Neurospora crassa [40]. In these genes three Ribo-box sequences were found. One of them overlaps the transcription start point [40]. In our case, the Ribo-box sequences were positioned far upstream from the cap sites. We also found helix-loop-helix protein-binding sequences. Helix-loop-helix proteins are transcription regulatory factors which bind the consensus sequence CANNTG, called the 'E-box'. \( DUb80 \) presented the sequence CACCTG (nt -2665; boxed in Figure 2) which is also present in the promoter region of human and mouse zona-pellucida Zp-2 genes [43]. Moreover, the sequence CACCTG is identical with the MyoD-binding site on the muscle creatine kinase gene [44] and to the E12 and E47 binding site on human immunoglobulin heavy chain [45]. A different E-box (CACATG) was found in the 5' region (nt -1300; boxed in Figure 2) and in the intron 1 (nucleotides 166, 338 and 352; boxed in Figure 2). This sequence element has been defined previously as part of the binding sites for the human transcription factors USF and TFE3, which bind specifically to the adenovirus major late promoter and the pE3 motif within the immunoglobulin heavy-chain enhancer [46].

3' non-coding region

The 3' end of the transcript was defined by the end of the cDNA isolated by Lee et al. [20]. Moreover, we have sequenced different cDNA clones coding for \( DUb80 \). The 3' end varied among them and from the Lee et al. [20] clone, but the differences consist of a few nucleotides. Major differences in length were not found.

Coding region

The sequence corresponding to the coding region is identical with the one described by Lee et al. [20] and also with the partial cDNA clones isolated in our laboratory. The sequence surrounding the translation start codon matched the Drosophila consensus sequence.

Moreover, the ubiquitin coding regions from \( DUb80 \) and \( DUb52 \) are identical at the amino acid level [17].

Pattern of \( DUb80 \) fusion-gene expression

Poly(A)\(^{+}\) RNA from different developmental stages of D. melanogaster was isolated and analysed by hybridization to the labelled tail-specific DNA, p80Ub\(^{+}\). Figure 4(a) shows one single band of 0.9 kb in embryonic, larval, pupal and adult RNA that is constitutively expressed throughout development. We had previously observed specific hybridization of the ubiquitin–52-amino-acid tail fusion cDNA to a 0.9 kb mRNA [17]; thus both genes are transcribed to a 0.9 kb messenger. To determine the distribution of ubiquitin-fusion gene expression, total RNA from Malpighian 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 with the tail-specific p80Ub\(^{+}\) probe. Figure 4(b) shows a single band of 0.9 kb which is expressed in all the samples. However, the level of expression varied depending on the tissue examined. From densitometer analysis it can be deduced that the steady-state concentrations of the transcript for the ubiquitin–80-amino-acid fusion gene is, as with the \( DUb52 \) gene [17], at least four times
Figure 2  For legend see opposite.
The E-box was p80Ub-probe higher in characterization among proteins exponential and encoding (DUbS0) from yeast sequence of the RNA proteins ribosomal 80-amino-acid organisms, proteins are more organisms, nevertheless, proteins are in yeast, ubiquitin-fusion proteins are about in yeast, clone [12,16,18,48], indicating a common function of these proteins among eukaryotes. It has been shown that they are ribosomal proteins in yeast [23], mammals [24] and other organisms [18,47].

Structure of the DUbS0 gene
In some organisms, ubiquitin-fusion protein genes are present in more than one copy per haploid genome [12,16,18,48]. Nevertheless, in Drosophila, both DUbS2 [17] and DUbS0 [20] fusion proteins are encoded by only one gene [20]. Like DUbS2, the DUbS0 genomic sequence shares some similarities with those of ubiquitin-fusion proteins in other organisms and also with other Drosophila rp genes. One of them is the presence of intron sequences in the 5' regulatory region. Thus DUbS0 contains one intron in its 5' region (intron 1) and another one in the tail-coding region (intron 2). We have demonstrated the existence of both introns by a PCR technique (Figure 3a).

Other features common to rp genes from higher eukaryotes (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 [49,50]. These features are also seen in other non-rp genes transcribed by the RNA polymerase II [51,52]. In addition, it has been shown that the oligopyrimidine tract at the 5' end of mammalian rp mRNA is important for translational control mechanisms [53]. Moreover, some Drosophila rp genes also start in a CT-rich stretch [54], although other rp genes do not [55,56]. We show here that DUbS0 presents in its 5' regulatory sequence a CT-rich region, but preceded by a GT-rich stretch rather than the previously described GC-rich one. Our primer-extension experiments located two transcription start areas. The first one contains two putative cap sites which are included in the CT-rich region (Figure 2). The second one is a region rich in TATA-like elements and contains four different transcription start points which are several nucleotides upstream from the pyrimidine tract (Figure 2). Since the DUbS0 gene encodes a unique mRNA (see below), we are unable with our data to comment on the possible physiological significance of the different start sites.

The TATA consensus sequence we have found at −84 (Figure 2) is likely to be too distant to be functional. There are also
Figure 3  Introns and transcription-start-point localization

(a) PCR analysis of the genomic structure of DuB80 cDNA synthesized from adult-fruitfly total RNA (lanes 1, 2, 3 and 4) or adult-fruitfly genomic DNA (lanes 1', 3' and 4') were used as templates for the PCR reactions. Oligonucleotides B and C were used in the PCR reactions of lanes 1 and 1'; B and E in lanes 3 and 3'; A and C in lanes 4 and 4' (see the Materials and methods section for the sequence of the synthetic oligonucleotides used). The symbols for the genomic scheme of DuB80 are as in Figure 1. The PCR products are represented below the scheme. The size bands are indicated in bp. (b) Primer-extension analysis of the transcription start points of DuB80. Lane 1 shows the primer-extended products using 10 μg of adult-fruitfly total RNA. Estimated sizes are indicated in bp by arrows according to the sequencing reaction showed in panel 2. The genomic scheme of DuB80 is indicated by the same symbols as in Figure 1. The small arrow indicates the tail-specific synthetic oligonucleotide used (see the Materials and methods section) and the big one represents the different products of the extension reaction.

Figure 4  Expression analysis of the DuB80 gene

(a) Northern-blot analysis of different developmental stages from D. melanogaster. Poly(A)+ RNA (2 μg/lane) was isolated from embryos (E), third-instar larvae (L), pupae (P) and adult fruitflies (A) and hybridized with the tail-specific plasmid p80Ub-. To the left is shown the position of the 23 S and 16 S rRNA molecular-size markers from Escherichia coli. The same filter was hybridized with an α-tubulin-specific probe to check for RNA loading (bottom of the Figure). (b) Northern-blot analysis of different tissues from D. melanogaster. Total RNA (2.5 μg) from ovaries (O) and 5 μg of total RNA from Malpighian tubules and intestines (I), imaginal discs (ID), fat bodies (FB) and salivary glands (SG) were hybridized to the tail-specific probe p80Ub-. (c) Northern-blot analysis of RNA from cultured Drosophila cells. Total RNA from 10⁶ exponentially growing (Ex) and stationary-phase (St) cultured cells was hybridized to the tail-specific probe p80Ub- (upper part of panel) and to an α-tubulin-specific probe (lower part of panel).
multiple CAAT-like boxes [55], as in other Drosophila rp genes.

Gene expression in a particular tissue is determined by the action of specific transcription factors. The helix-loop-helix proteins have been described as tissue-specific activators of certain genes [reviewed by Jones [57] and Torres et al. [58]]. Family members of these proteins occur in diverse eukaryotes, ranging from mammals to yeast, and are important in regulating metabolism, cell differentiation and development [59]. Many of these helix-loop-helix factors bind the consensus sequence CANNTG (also referred as an ‘E-box’, for enhancer box), which is present in the 5’ regulatory region from tissue-specific genes [44–46,60,61].

We have described previously the presence of an E-box (CACGTG) in the 5’ regulatory region of the DUb52 gene [17], which has been proved to be functional using a reporter gene in S2 Drosophila cells (R. Barrio, unpublished work). Likewise, we have found two different E-boxes in the 5’ non-coding region of DUb80, located far away from the transcription start sites (Figure 2). The presence of these sequences in both DUb52 and DUb80, and the possible existence of an ovary-specific transcription factor in Drosophila, are in agreement with our expression results. Like the DUb52 gene, DUb80 is much more abundantly expressed in ovaries than in other tissues (Figure 4b).

Although, these sequence elements are present in other Drosophila rp genes such as S14A [56], rp49 [55], S6 [62] and rpA1 [54], it is not known whether all these proteins are over-expressed in ovaries or whether those sequences are functional in those genes.

Expression of the DUb80 gene

Previous reports [20] had shown that the Ub80 fusion gene in Drosophila encodes an mRNA with a molecular size of 0.9 kb. Our findings confirm this observation (Figure 4). Moreover, the size of this mRNA, as deduced by the genomic sequence, would vary from 587 bp (considering nt 1 in Figure 2 as the transcription start site) to 755 bp if we consider the farthest start point (position –168 in Figure 2). These sizes are smaller than that observed by Northern analysis [20] (Figure 4a). A similar event occurs in the mRNA coding the DUb52 protein [17]. It is possible that the 3’-polyadenine stretch in both mRNAs contributes to increase their estimated sizes.

This transcript is constitutively expressed in all developmental stages (Figure 4a). However, DUb80 gene expression varies considerably among different larval and adult tissues. Thus adult ovaries show a higher accumulation than other tissues examined (Figure 4b). This result is in accordance with the existence of a high rate of ribosome synthesis during Drosophila oogenesis. A similar result has been observed for the expression of the Ub52 gene from both Drosophila and Manduca sexta [21].

It has been proposed in yeast that the fusion genes represent the major source of ubiquitin when the ribosome synthesis is active, i.e., during the exponential phase of growth [12]. In Drosophila cell culture this differential expression occurs both in Ub52 [17] and Ub80 (Figure 4c) fusion genes. In fact, cells growing in the exponential phase accumulate the DUb80 transcript very actively, whereas it is repressed when cells enter the stationary phase (Figure 4e).

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 ribosomes. This location could explain the presence of a nucleic-acid-binding domain in the primary structure of both Ub52 and Ub80 proteins. This domain could bind to, or interact with, mRNA or rRNA molecules within the ribosome. It has been suggested that the ribosomal tails, as well as other rps, are extremely short-lived if they are not properly assembled in the ribosome [63]. 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 would be degraded during the transition to the ribosome [23].

Drosophila ubiquitin genes belong to a multigene family. In the present study we have isolated and analysed the genomic clone that encodes the last gene of this family to be characterized.

We are very grateful to Juan García, Manuel Soto and Javier Llanos for their expert technical assistance, to Javier Sampedro for a critical reading of the manuscript before its submission, and to Jose Maria Requena for his helpful comments on our work. This work was performed at the Marta Izquierdo laboratory and was supported by a grant from the D.G.I.C.Y.T. (Spain). R.B. was a fellow from the Gobierno Vasco and the Rich Foundation, and A.A. was the recipient of an FPI fellowship.

REFERENCES

30 Shields, G. and Sang, J. H. (1977) Drosophila Int. Serv. 52, 161
Received 18 January 1994/15 March 1994; accepted 21 March 1994