Structure and expression of the *Drosophila* ubiquitin-80-amino-acid fusion-protein gene

Rosa BARRIO,* Araceli DEL ARCO,† Héctor L. CABRERA‡ and Carmen ARRIBAS§

Departamento de Biología Molecular, Facultad de Ciencias, Universidad Autónoma de Madrid, 28049 Madrid, Spain.

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 *UBI1-UBI2* and *UBI4* in yeast), and we report here the organization and expression of a third one, the *DUb80* gene (the homologue of *UBI3* 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 *UBI3* 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 polymorphic among different stocks and strains [27]. It contains either 15 or 18 tandem repeats, depending on the fruitfly 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, 28040 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.

Figure 1 Physical map of the two genomic clones isolated (pG80X2 and pG806) coding for *DUB80*

The shaded boxes in pG80X2 and pG806 indicate the transcribed region of the ubiquitin-fusion gene. In the detailed map, introns are indicated by white boxes. The black box indicates the ubiquitin-coding region and the striped box the tail-fusion-coding region. 3' and 5' non-translated regions are represented by shaded boxes. The different subclones sequenced are indicated by horizontal bidirectional arrows. Smaller arrows represent synthetic oligonucleotides used for sequencing. A, *AccI*; Bg, *BglII*; E, *EcoRI*; H, *HindIII*; Pst, *PstI*; Pv, *PvuII*; RV, *EcoRV*; S, *SalI*; Sca, *Scal*; X, *XhoI*. The *SacI* and *XhoI* sites come from the EMBL-3 phage subcloning site, and the *SalI* sites from the pUC18 polylinker.

Screening recombinant libraries

A library of *D. melanogaster* genomic DNA partially digested with restriction endonuclease *Sau3A*, cloned into the *Bam*HI 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 32 P-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 was determined by the dideoxy 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, comp-

lementary 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 μ l, of which 5 μ l 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 μ l 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 $^{\circ}$ C, followed by 29 cycles of PCR with incubations of 1 min at 55 $^{\circ}$ C (annealing), 1 min at 72 $^{\circ}$ C (polymerization), 1 min at 94 $^{\circ}$ C (denaturation) and a final extension of 10 min at 72 $^{\circ}$ C. After reaction the samples were electrophoresed in 0.8% agarose gels. The fragments of interest were subcloned in the *Sma*I 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, λ G80, which contained an insert of approx. 14 kb. This phage was digested with *Xho*I, resulting in two fragments, X1 and X2. Fragment X2 hybridized with both the tail- and the ubiquitin-specific probes, and was subcloned in the *Sal*I 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 *Xho*I-*Bgl*II 5' fragment of pG80X2. A second positive phage was isolated, λ G806 (Figure 1), whose insert was digested with *Sac*I 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 a 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 μ E3 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

-2923 GAAATCCAATAATTCGGTTGTTTATTAAGAAACATTTACATATGTTATTACCCACTTGAATTTTGTTTAAATTACTACTAACTGGTTTAGTTTGGCCAT
 -2823 AAGTGTGGGAGAAGTTACTTTCTCATTAAAAATTTATGATCCAATTTTACCTGTCAATCAAAGCTACACTTTAAATCAATTAATTTCTGATTAG
 -2723 AGGGTATAGAAGCTGCTTAGTAAATTATGAACCAATTGAAATTAGAATTGGTAATTAGCACCTGCAATTCATGCTGTCAATCAACAAAAGACCTTACTA
 -2623 TAAAAATGACACAACATTCAGCAGGATAAACAATAGTAAATGAATCTCTTAATCAGTATGTTTATTTTAACTAGCTGCAGGGGAGGGTGAATAATT
 -2523 CGAGTATGCAAGAAATTCGGTTGTAACGAATTTCTGTTAAAGTCTAGTAAAAACCAAGCAAGCAATAGTGTTTTCGGTGTATTATTTAAAGGTAAGCAATGTCG
 -2423 CTAAGTTACAATAAAGGTATTCATAATATATTTATTTGATAGATTTTAAAGAAATCTCTCTGTCATTGATGCGAATTAATCAATTTGTGAGGTAGTTAAC
 -2323 CTTAACCAAGTTATTCTACTAACATCGATATTGACGCGGAAAATTCAGCGCAACTCTCTGTGCTGGTAAATGCCAGATGCAGTGGATCATCTGAACTTC
 -2223 TTTTGAAGCCTCTGAAAATAGGTATTTCAACAAAACCTATCAGTGGTTTCTCTGGGGTGTGATCTOGAGTTCAATCCCTATTTCCTTAAATCTCAACT
 -2123 ATGTTGGTCCCAAGCCAGATAACCTATGTTAATGAAACCTCAGATGGTTATGCGTACTGGGATATCATTGCAAGGACCGCCTTGAAGACGGATTG
 -2023 GAAATAAATTTAATAGCAAGTTGATAAAGCATACATTGAATATAGTTTCGGATATTTTCCATTGTCAAAGCATTGATTTTCGGGACAAATTAATGGAT
 -1923 TAACCTAAGAGGAGCAAGTGTGTAAGTAATGAATAAGTTCTGTAACCTATTAGATTATAGCCATTTCATTTTGGAACTAGATCGACAAGGAAGATA
 -1823 TAATCTCAATGAACAAATGAATCAATCTCTCGGACCAACTAAGAGTGTGGAGTGGCTGGCATTCTAATGATCTACTACGAGCTCTCTGGGACTTCT
 -1723 GAGAGACGTTTCAATTTCAAGGTTAACTTCGAAATCTAGAGGATGGGCTGGAAGTTTGGGCAATACTACCTTTGACTTGGACTCTTGGGATAGTTAT
 -1623 GCGAAATGAGGAGATATAGCCGATCCGGTGGTTTCAATCGAATTAATCGCTTTGCTGAGTTTGATACCATACACCAGAGTTGGGAAATTTGAGACA GC
 -1523 TTTTCTATACCGCTACACTTCGGATTGGATACTCATGGAAGAGTGGTAACCTTCTGTGCCATTGAGGACAGAGTTTGGTTGTTCTGTCTATTAAAC
 -1423 CTGTGTCTTTTAGCATCATTTGGGTGCTATTGAAATATCGACTGAAGATATTGGTAGACGAGTTTATAGTCAAAAGCTGGAACATTTAAATCAGAAA
 -1323 AGTCCGGTGATTGTGCTTAAGACACATG CATTGAAAGCAATTTTACAGACATTTGGAGCCTGTGCAACAGGGATTGGACCCAAATACTGTAGATCGTT
 -1223 CAGTGGCTTTTGGTAATGACACTGTTCTTCTCCCTGGTTATGTACAACATTTATACGCTCTTCGGTGGGGGATTGCTTAGTTCTTCAGATCA
 -1123 GGGTCCCTCTACAGTCGATGAAATAACTGCAAGCCCCCTGAAAATATCATTOGAAGACATTGGTTACTACAAAGCTCTTATTCGGAGTAAGTTGCTGAAA
 -1023 AGATATATATTTATATTACCATTTACTCGATTACTTAAACAGGAGAGTCAAAATCGATCTATTACAAGATTA ATAGAGAAAAGCTGCTC TTGCTCAAGAA
 -923 TTTAAATGAATTGCAATTTTGGCCATATTGAGGATGCTGTCCCATATCTCAAAGCGGGGATTGCTTTTTC ATGCGGAAGTGGTGGAC GCTTATCCG
 -823 GTGATTTTCAAGATATTTTGATGCCAACGAGATCTGTGATCTGGTGAAGTCTCGGCTGTGATGGAG GTCGAAATTTTGAATTTG GATTTTACACAAGAACA
 -723 GCCAGTATACGAGATTTTCAAACAGCGTAAGTATTGAGTCATTGCGATGGCTTAACGCGCAAAAAGAAGCTGTTTTTTTTTAAACCGGCTCTTATCCA
 -623 CATCAGTATATTTCCCAACATTCCTTGGGAAGTAAGATAGTTTATAATTTTATAATAGTTTATAATTTGTAATAAATCTATATTTTCTGCTGTA
 -523 GTATGTGTAATGCCAGGAAAAGGTTTGTAGAGCGCATTTCTCGGCGCGGCAAAATTAAGGCGGCTGCAATCATTATACACCGTTATCTGTA
 -423 AGCTTGGCTCAGGAGTGTACCTGGATTGTTATTTTAAATTTGTAGAGTATTAAACAAATTAATCTAGAACTTAAATTTGTATACATGTTTATTTTAAT
 -323 AGGCGGATTGGAGCATCGCTTCTACTTTTGTGCTTGAAAAAGTATATGCACATTTTGGACCAAGAAATTTTGGGTTTAAATTTTATTTTGT
 -223 TTCTTTATGATGATTTTGTGATTGTTATATTTTATAATTTTGTAAATAAGGAAAAGTGTATTGGAATATGGCTTCTGTTTCATCCAGTTTATGTT
 -123 CCATTCCACGCGTTCCAAAAATAAAGCGGTAGTTCAAAA ATAAACCAAAAGTAACTCCGAATGCAGCCCTGCGCTCAATTTTCAACTTCCTGTACGCG
 -23 TGTGTGTGTGTTCTTTCTTTCT [TGTATTCTGTTGCCGAAGCAAG] gtagtttggggtgttacattttaagcatcggttaagtgaataaataacataca
 76 aatacataactaaagtagtccttacacttgtaaaactcttacgagtgatgacacattgtgattgaacaatgagaatgattgcagtgacatg taaga
 177 ttgacagcaatcaaaattgctaaaagatacccgtaacaaaaaaatattgtgacatttaatttatacattaaggttaagtgaatttcggcggtgattagt
 278 tatctgtggtggtgacacagaggtctccacaggaacaaatttcgcacaccccccaacgctc taataaac taatgatttcagatctggtc
 378 tacagaattaacaattaacaatgccgcagaatgagatatggaacgccatcgcaagctgtatggcggcgattggattacaggaacggaacgcaagaa
 478 ggaagcgctcttcccaaggaccgagcagaaaggctcgcaagttgcgcgccataaggccaaactcttaataaggagcgacgaatgaaagattcaga
 578 ttaagaagaagatccaggcccacgaagaagaaggtaaaaaggaggaggaaggatcgaggatggagccctgccgcatatctgctgcagagaggcat
 678 ccagtcagcgccaaggctcctgctcaatgatcaagcagaagcgcaaggagaaggcaggcaagtgaggacgtgccattcccaagtagcgctcagtc
 778 gatgctgaggtcttcaagtaactaaagaccggaagacaaagcgaaaggcatggaagcgcatggtcacaaaagtcacattcgctggtagaacttcacac
 878 gcaagccaccaaaagttcgagcgttctcattcgacccatgggtctgcatgaagaaggctcagttacgcatccagaactgaaagccacttcaactctg
 978 catcattggcgctcaagaagaaccccgctcgccatgctcacttcttgggtgaattacaaagggtactgtgatcgaagtcaaatctctgagctgggt
 1078 ttggttaacgcaaacgggaaaagttgctggggcaatacgtcaggtcacgaacaatccagaaaacgatgggttatcaatgcagtgctgctgtctaac
 1178 cccgaaaatgctgctacctaatttaagagtttgatttaataaacaacatacataaataacatgaacaatcggttgattctcttagcttattatactatta
 1278 gctgtagcaatgccatataataaacacacatatctataccaa catctttttgtttcttgcgaag [TTTGGTGAATGGAATACAG ATG CAG ATC TTC
 1 M Q I F
 1375 GTT AAG ACC CTC ACG GGC AAG ACC ATC ACT CTT GAG GTC GAA CCG TCG GAT ACT ATC GAA AAT GTC AAA GCC AAG
 5 V K T L T G K T I T L E V E P S D T I E N V K A K

Figure 2 For legend see opposite.

```

1450  ATT CAG GAC AAA GAA  GGA ATT CCT CCA GAT CAG CAG CGT CTG ATC TTC GCT GGC AAG CAG CTG GAA GAT GGC CGC
    30  I  Q  D  K  E   G  I  P  P  D  Q  Q  R  L  I  F  A  G  K  Q  L  E  D  G  R
1525  ACC CTC TCT GAC TAT AAC ATT CAG AAG GAG TCC ACT TTG CAC TTG GTG CTC CGC CTG CGT GGT GGT* GCC AAG AAG
    55  T  L  S  D  Y  N  I  Q  K  E  S  T  L  H  L  V  L  R  L  R  G  G  A  K  K
1600  CGC AAG AAG AAG AAC TAC TCC ACT CCC AAG AAA ATC AAG CAC AAG AGG  AAG AAG GTC AAG CTA GCT GTC CTG AAA
    80  R  K  K  K  N  Y  S  T  P  K  K  I  K  H  K  R  K  K  V  K  L  A  V  L  K
1675  TAC TAC AAG] gtaacacacaacaatctacgacacatacgaacacattctgatctgttcttgggaattcttgagctctggcttgctcattggcttaag
    105  Y  Y  K
1772  atttggatatcagagtgcattcaccagtgccctatattgaatgaacatgtatcagacgcagtatatta2 tcactttctatgctctttctttccatag [GTT
    108                                     V
1871  GAC GAG AAC GGC AAG ATC CAC CGT CTC CGT CGC GAG TGC CCC GGT GAG AAC TGC GGC  GCC GGC GTC TTC ATG GCT
    109  D  E  N  G  K  I  H  R  L  R  R  E  C  P  G  E  N  C  G  A  G  V  F  M  A
1946  GCC CAC GAA GAT CGT CAC TAC TGC GGC AAG TGC AAC CTG ACC TTT GTC TTC AGC AAA CCA GAG GAA AAG TAA TTTTG
    134  A  H  E  D  R  H  Y  C  G  K  C  N  L  T  F  V  F  S  K  P  E  E  K  End
2023  CTACATAAGATCATGTACGTTTCCAGAAATCAAATAAAGGTAGTAATTGAATAATAAATCAATCGCTG] AAATTTTCCTTTTATTGTTAGTTAGCAC
2122  TAGCGTGGTTAGTAATGTTTAGCAAAACACTATAGTCTCATTTAGGGTGTATGGTTTATTTGTGAAAAATTTGAATACATTTTGTATATTCGATTTTTA
2222  CTCGACCTTGAGAAATGCAATTTGTTTATTGTTAAACCACTCAAACGTTAACTATGTCGACGTTAAGGCTTTTGGACAAGAAAGTCAGTATATCGGTGGCT
2322  TCCTCAATTCTCTTCGATGTGGCTTGCCAGCTCCATGGCCCGCCTCTGTAGACGCGTAACAGAACCGGATTTTCTGGAATTC

```

Figure 2 Nucleotide sequence of the *DUB80* gene

Nucleotides have been arbitrarily numbered from the first transcription start point. Amino acids are indicated by **bold** numbers. The junction between ubiquitin and the tail protein is shown by an asterisk (*). Both introns are written in lower case. The cDNA sequence published by Lee et al. [20] is surrounded by square brackets ([]). The consensus 5' and 3' splicing and branch-point sequences are in *italics*. The adenines involved in the splicing reaction are marked by a bold overbar ([~]). The CT and GT stretches are doubly underlined (). The different transcription start points are in **bold** letters. The TATA consensus sequence is in a shaded box [■]. The TATA-like elements are underlined by wavy lines (~~~~~). The CAAT-elements are underlined by dotted lines (.....). The translation start consensus sequence and the polyadenylation signal are underlined with single lines (). The Ribo-box homologous sequences are underlined by striped boxes. The E-box sequences are boxed.

higher in ovaries than in other tissues analysed. Likewise, *DUB80* expression was altered depending on the growth conditions. RNA isolated from *Drosophila* cell cultures growing in exponential and stationary phase was hybridized to the tail-specific probe p80Ub⁻ (Figure 4c). Like the *DUB52* transcript [17], the 80-amino-acid fusion-protein transcript was almost undetectable in stationary-phase cells, whereas its concentration during the exponential phase was increased about 15-fold. A similar phenomenon has been observed in expression from other ubiquitin fusion genes [12,15].

DISCUSSION

The isolation and sequencing of a cDNA clone coding for the 80-amino-acid ubiquitin-fusion protein DUB80 has previously been reported by Lee et al. [20]. In the present paper we have described the characterization of a genomic clone from *D. melanogaster* (*DUB80*) encoding this ubiquitin-fusion protein. The amino acid sequence of the ubiquitin-extension proteins is very conserved from yeast to humans, 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 *DUB80* 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 DUB52 [17] and DUB80 [20] fusion proteins are encoded by only one gene [20]. Like *DUB52*, the

DUB80 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 *DUB80* 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 *DUB80* 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 *DUB80* 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 μ g 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 represent 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 μ g/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 μ g) from ovaries (O) and 5 μ g 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 RNA, 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 4c).

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 María 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

- 1 Rechsteiner, M. (1988) Ubiquitin, Plenum Press, New York
- 2 Levinger, L. and Varshavsky, A. (1982) Cell **28**, 375–385
- 3 Thorne, A. W., Suatieri, P., Briand, G. and Crane-Robinson, C. (1987) EMBO J. **6**, 1005–1010
- 4 Varshavsky, A., Levinger, L., Sundin, L., Barsoum, J., Özkaynak, E., Swerdlow, P. and 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. and Fyrberg, E. A. (1987) Cell **51**, 221–228
- 6 Siegelman, M., Bond, M. W., Gallant, W. M., St. John, T., Smith, H. T., Fried, V. A. and 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. and 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. and Wood, W. I. (1987) Nature (London) **330**, 537–543
- 9 Özkaynak, E., Finley, D. and Varshavsky, A. (1984) Nature (London) **312**, 663–666
- 10 Wiborg, O., Pedersen, M. S., Wind, A., Berglund, L. E., Marcker, K. A. and Vuust, J. (1985) EMBO J. **4**, 755–759
- 11 Giorda, R. and Ennis, H. L. (1985) Mol. Cell. Biol. **6**, 2097–2103
- 12 Özkaynak, E., Finley, D., Solomon, M. J. and Varshavsky, A. (1987) EMBO J. **6**, 1429–1439
- 13 Salvensen, G., Lloyd, C. and Farley, D. (1987) Nucleic Acids Res. **15**, 5485
- 14 Ohmachi, T., Giorda, R., Show, D. R. and Ennis, H. L. (1989) Biochemistry **28**, 5226–5231
- 15 Swindle, J., Ajioka, J., Eisen, H., Sanwal, B., Jacquemot, C., Browder, Z. and Buck, G. (1988) EMBO J. **7**, 1121–1127
- 16 Baker, R. T. and Board, P. G. (1991) Nucleic Acids Res. **19**, 1035–1040
- 17 Cabrera, H. L., Barrio, R. and Arribas, C. (1992) Biochemical J. **286**, 281–288
- 18 Callis, J., Raasch, J. A. and Vierstra, R. D. (1990) J. Biol. Chem. **265**, 12486–12493
- 19 Lund, P. K., Moast-Staats, B. M., Simmons, J. G., Hoy, E., D'Ercole, A. J., Martin, F. and Van Wyk, J. J. (1985) J. Biol. Chem. **260**, 7609–7613
- 20 Lee, H., Simon, J. A. and Lis, J. T. (1988) Mol. Cell. Biol. **8**, 4727–4735
- 21 Bishoff, S. T. and Schwartz, L. M. (1990) Nucleic Acids Res. **18**, 6039–6043
- 22 Redman, K. L. and Rechsteiner, M. (1988) J. Biol. Chem. **263**, 4926–4931
- 23 Finley, D., Bartel, B. and Varshavsky, A. (1989) Nature (London) **338**, 394–401
- 24 Redman, K. L. and Rechsteiner, M. (1989) Nature (London) **338**, 438–440
- 25 Finley, D., Özkaynak, E. and Varshavsky, A. (1987) Cell **48**, 1035–1046
- 26 Izquierdo, M., Arribas, C., Galceran, J., Burke, J. and Cabrera, V. M. (1984) Biochim. Biophys. Acta **783**, 114–121
- 27 Arribas, C., Sampedro, J. and Izquierdo, M. (1986) Biochim. Biophys. Acta **868**, 119–127
- 28 Cabrera, H. L., Arribas, C. and Izquierdo, M. (1990) Nucleic Acids Res. **18**, 3994
- 29 Schneider, I. (1972) J. Embryol. Exp. Morphol. **27**, 353–365
- 30 Shields, G. and Sang, J. H. (1977) Drosophila Inf. Serv. **52**, 161
- 31 Maniatis, T., Fritsch, E. F. and 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. and Berg, P. (1977) J. Mol. Biol. **113**, 237–251
- 33 Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

- 34 Sanger, F., Nicklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463–5467
- 35 Bunch, T. A., Grinblat, Y. and Goldstein, L. S. B. (1988) *Nucleic Acids Res.* **16**, 1043–1059
- 36 Ghosh, P., Reddy, V., Piatak, M., Lebowitz, P. and Weissman, S. (1980) *Methods Enzymol.* **65**, 580–595
- 37 Mount, S. M., Burks, C., Hertz, G., Stormo, G. D., White, O. and Fields, C. (1992) *Nucleic Acids Res.* **20**, 4255–4262
- 38 Mager, W. H. (1988) *Biochim. Biophys. Acta* **949**, 1–15
- 39 Larson, D. E., Zahradka, P. and Sells, B. H. (1991) *Biochem. Cell Biol.* **69**, 5–22
- 40 Tyler, B. M. and Harrison, K. (1990) *Nucleic Acids Res.* **18**, 5759–5765
- 41 Blacwell, T. K. and Weintraub, H. (1990) *Science* **250**, 1104–1110
- 42 Lassar, A. B., Davis, R. L., Wright, W. E., Kadesch, T., Murre, C., Voronova, A., Baltimore, D. and Weintraub, H. (1991) *Cell* **66**, 305–315
- 43 Millar, S. E., Lader, E., Liang, L. F. and Dean, J. (1991) *Mol. Cell. Biol.* **11**, 6197–6204
- 44 Anthony-Cahill, S. J., Benfield, P. A., Fairman, R., Wasserman, Z. R., Brenner, S. L., Stafford, W. F., Altenbach, C., Hubbell, W. L. and DeGrado, W. F. (1992) *Science* **255**, 979–983
- 45 Murre, C., McCaw, P. S., Vaessin, H., Caudy, M., Jan, L. Y., Cabrera, C. V., Buskin, J. N., Hauschka, S. D., Lassar, A. B., Weintraub, H. and Baltimore, D. (1989) *Cell* **58**, 537–544
- 46 Gregor, P. D., Sawadogo, M. and Roeder, R. G. (1990) *Genes Dev.* **4**, 1730–1740
- 47 Müller-Taubenberger, A., Graack, H. R., Grohmann, L., Schleicher, M. and Gerisch, G. (1989) *J. Biol. Chem.* **264**, 5319–5322
- 48 Gausing, K. and Jensen, C. B. (1990) *Gene* **94**, 165–171
- 49 Graham, R. W., Jones, D. and Candido, E. P. M. (1989) *Mol. Cell. Biol.* **9**, 268–277
- 50 Snyder, M., Hunkapiller, M., Yuen, D., Silvert, D., Fristrom, J. and Davidson, N. (1982) *Cell* **29**, 1027–1040
- 51 Tamura, T. and Mikoshiba, K. (1991) *FEBS Lett.* **282**, 87–90
- 52 Jolliff, K., Li, Y. and Jonhson, L. F. (1991) *Nucleic Acids Res.* **19**, 2267–2274
- 53 Levy, S., Avni, D., Hariharan, N., Perry, R. P. and Meyuhas, O. (1991) *Proc. Natl. Acad. Sci. U.S.A.* **88**, 3319–3323
- 54 Qian, S., Zhang, J. Y., Kay, M. A. and Jacobs-Lorena, M. (1987) *Nucleic Acids Res.* **15**, 987–1003
- 55 O'Connell, P. and Rosbash, M. (1984) *Nucleic Acids Res.* **12**, 5495–5513
- 56 Brown, S. J., Rhoads, D. D., Stewart, M. J., Van Slyke, B., Chen, I. T., Johnson, T. K., Denell, R. E. and Roufa, D. J. (1988) *Mol. Cell. Biol.* **8**, 4314–4321
- 57 Jones, N. (1990) *Cell* **61**, 9–11
- 58 Torres, R., Schreiber-Agus, N., Morgenbesser, S. D. and DePinho, R. A. (1992) *Curr. Opin. Cell. Biol.* **4**, 468–474
- 59 Sun, X. H. and Baltimore, D. (1991) *Cell* **64**, 459–470
- 60 Hall, R. K. and Taylor, W. L. (1989) *Mol. Cell. Biol.* **9**, 5003–5011
- 61 Prendergast, G. C. and Ziff, E. B. (1992) *Trends Genet.* **8**, 91–96
- 62 Stewart, M. J. and Denell, R. (1993) *Mol. Cell. Biol.* **13**, 2524–2535
- 63 Planta, R. J. and Raué, H. A. (1988) *Trends Genet.* **4**, 64–68