MOLECULAR AND CELLULAR BIOLOGY, Nov. 1988, p. 4727–4735
0270-7306/88/114727-09$02.00/0

Vol. 8, No. 11

Structure and Expression of Ubiquitin Genes of *Drosophila melanogaster*

HYUNSOOK LEE, JEFFREY A. SIMON,† AND JOHN T. LIS*

Section of Biochemistry, Molecular and Cell Biology, Cornell University, Ithaca, New York 14853

Received 5 May 1988/Accepted 4 August 1988

We isolated and characterized two related ubiquitin genes from *Drosophila melanogaster*, polyubiquitin and *UB3-D*. The polyubiquitin gene contained 18 repeats of the 228-base-pair monomeric ubiquitin-encoding unit arranged in tandem. This gene was localized to a minor heat shock puff site, 63F, and it encoded a constitutively expressed 4.4-kilobase polyubiquitin-encoding mRNA, whose level was induced threefold by heat shock. To investigate the pattern of expression of the polyubiquitin gene in developing animals, a polyubiquitin-*lacZ* fusion gene was introduced into the *Drosophila* genome by germ line transformation. The fusion gene was expressed at high levels in a tissue-general manner at all life stages assayed. The ubiquitin-encoding gene, *UB3-D*, consisted of one ubiquitin-encoding unit directly fused, in frame, to a nonhomologous tail sequence. The amino acid sequence of the tail portion of the protein had 65% positional identity with that of yeast UBI3 protein, including a region that contained a potential nucleic acid-binding motif. The *Drosophila UB3-D* gene hybridized to a 0.9-kilobase mRNA that was constitutively expressed, and in contrast to the polyubiquitin gene, it was not inducible by heat shock.

Ubiquitin is a highly conserved, 76-amino-acid protein that is present in all eucaryotes examined, both as a free polypeptide and covalently joined to various cytoplasmic, nuclear, and cell surface proteins. Proteins known to be conjugated to ubiquitin in vivo include histones H2A and H2B (1, 21, 38), lymphocyte homing receptor (34), platelet-derived growth factor receptor (43), and *Drosophila* actin (3). The various functions of ubiquitin are not completely understood, but the existing data indicate that it has several important roles. First, ubiquitin plays an important role in the ATP-dependent, nonlysosomal protein degradation pathway (10, 16). It is joined covalently via its C-terminal Gly residue to amino groups of short-lived proteins, and this conjugation is required for the selective proteolysis of these proteins (9, 10). Second, ubiquitin is a modifier of chromatin structure through its conjugation to histones H2A and H2B (21, 38). It has been suggested that this modification may facilitate transcription and DNA repair (21). Third, it has been proposed that ubiquitin and its conjugation system serve as a sensor of the heat shock response. In this model, the denatured proteins created by heat shock overload the ubiquitin conjugation system. This results in the net deubiquitination of a sensor protein, which could be either the heat shock transcription factor or a protein that activates this factor (28, 31).

Recently, the ubiquitin genes from various organisms were cloned and characterized (2, 18, 30, 39). These studies revealed that the amino acid sequence of ubiquitin is highly conserved from yeasts (30) to humans (39). A major source of ubiquitin in yeast cells, especially during heat stress and starvation, is provided by a gene that consists of tandem repeats of a 228-base-pair (bp) protein-coding region that is arranged by spacerless head-to-tail joining. Expression of this gene generates a primary translation product consisting of a polyprotein that is subsequently processed into monomeric units (29, 30). The number of unit repeats in polyubiquitin genes varies from species to species: *Xenopus laevis* appears to have as many as 12, whereas yeasts contain 5.

Recently, monomeric ubiquitin genes have also been identified in yeast and human cells (24, 29). Interestingly, some of these genes encode hybrid proteins in which a ubiquitin monomer is fused in frame to a nonhomologous tail protein sequence that is highly conserved between yeasts and humans (24, 29). At least one of the proteins encoded by these hybrid genes is capable of being processed to monomeric ubiquitin, since the pool of free ubiquitin in exponentially growing yeast cells is at normal levels in mutants lacking the polyubiquitin gene (13).

Identification of the *Drosophila* polyubiquitin gene has been reported by Izquierdo and colleagues (2, 18). In this paper, we describe further analyses of the structure and expression of the *Drosophila* polyubiquitin gene. We examined the tissue and temporal patterns of expression of the polyubiquitin gene during the course of *Drosophila* development, using β-galactosidase staining assays of germ line transformants containing a ubiquitin-*lacZ* fusion gene. Also, we describe the isolation and characterization of the *Drosophila* ubiquitin hybrid gene *UB3-D*.

MATERIALS AND METHODS

Plasmids and reagents. Plasmid pPUb has a 7.5-kilobase (kb) *Bam*HI-*Sal*I genomic DNA fragment that contains the intact polyubiquitin gene cloned into *Bam*HI-*Sal*I-digested pUC19 (see Results). Plasmids pUB3-D1 and pUB3-D2 are subclones of a λgt11 cDNA clone of the *UB3-D* gene. pUB3-D1 contains a 0.25-kb fragment that includes the 5′ leader sequence and a portion of the ubiquitin-encoding unit of the *UB3-D* gene (up to the *Eco*RI site within the ubiquitin-encoding unit; see Fig. 3), and pUB3-D2 contains a 0.46-kb fragment from the remainder of the cDNA. Both *UB3-D* fragments are cloned into the *Eco*RI site of pUC19. Plasmid pUB3-NE has a 0.27-kb *Nhe*I-*Eco*RI fragment that contains only the tail portion of the *UB3-D* gene cloned into the *Xba*I-*Eco*RI sites of pUC19.

Restriction enzymes, alkaline phosphatase, T4 DNA ligase, S1 nuclease, and avian myeloblastosis virus reverse

* Corresponding author.
† Present address: Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115.

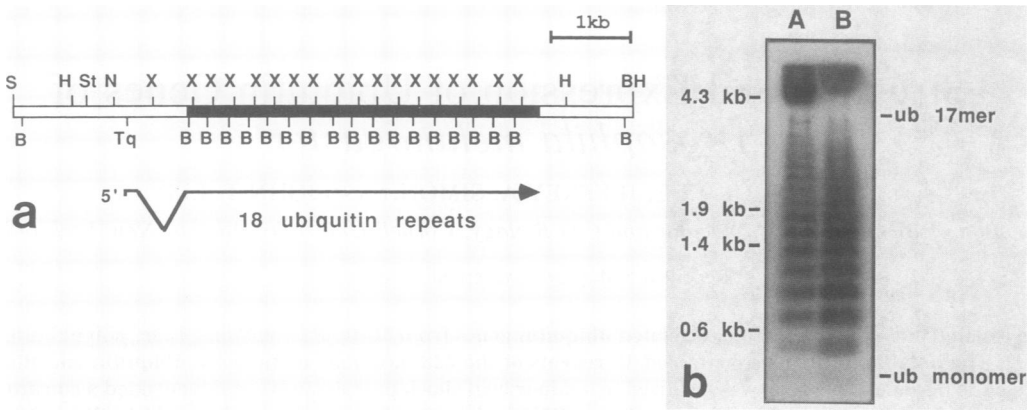

FIG. 1. (a) Restriction map of the cloned *Drosophila* polyubiquitin locus showing the 7.5-kb *Sal*I-*Bam*HI fragment that contains the polyubiquitin gene. Symbols: ■, protein-encoding region of the gene; →, direction and approximate position of the transcript. The kink in the arrow represents the intron within the leader sequence. The following restriction enzyme sites are shown: *Sal*I (S), *Hin*dIII (H), *Sty*I (St), *Nru*I (N), *Xho*I (X), *Bam*HI (BH), *Taq*I (Tq; not all *Taq*I sites are shown), and *Bgl*II (B). (b) Determination of the number of ubiquitin-encoding repeats in the polyubiquitin gene. Plasmid pPUb, containing the polyubiquitin gene shown in panel a, was partially digested with 5 (A) and 10 (B) U of *Xho*I for 15 min at 37°C. The DNA was separated on a 0.8% agarose gel and blotted onto a GeneScreen Plus membrane as described in the text. The filters were hybridized to a nick-translated DNA fragment composed of the 0.23-kb ubiquitin-encoding repeat.

transcriptase were purchased from Boehringer Mannheim Biochemicals. T4 polynucleotide kinase and Klenow polymerase were purchased from New England BioLabs, Inc. Sequencing kits and Sequenase were from U.S. Biochemical. [α-^{32}P]dATP, [α-^{32}P]dCTP, and [γ-^{32}P]ATP (3,000 Ci/mmol) were obtained from Amersham Corp.

Southern hybridization. Genomic DNA was isolated as previously described (37) from *D. melanogaster* adults (the Canton S strain made homozygous as described by Simon et al. [37]). The DNA samples were fractionated electrophoretically in a 0.8% agarose gel and transferred to a GeneScreen Plus filter according to the instructions of the manufacturer. The filter was hybridized with nick-translated DNA probes as described elsewhere (25).

In situ hybridization. Salivary gland squashes were performed as described elsewhere (23). Hybridization and detection were performed as described previously (36), using the 0.23-kb ubiquitin-encoding repeat as a probe.

DNA sequencing. Restriction fragments were cloned into M13 phage and sequenced by the dideoxy-chain termination method, using Sequenase.

RNA analysis. To prepare heat shock RNA, adult flies were heat shocked by immersing a glass vial containing the flies in a water bath at 36.5°C for 1 h. RNA was prepared as previously described (37). Non-heat shock RNA was prepared similarly but without heat shock. RNA samples were electrophoresed and transferred to nitrocellulose filters essentially as described previously (35). Ethidium bromide staining revealed the presence of approximately equivalent amounts of RNA in all lanes.

For S1 mapping, a 0.94-kb *Hin*dIII-*Xho*I fragment (Fig. 1a) from the 5′ upstream region of the polyubiquitin gene was cloned into pUC19, and this subclone was used to generate a 0.63-kb *Taq*I fragment. This fragment was end labeled and then cleaved with *Sty*I (Fig. 1a) to generate a 5′-end-labeled 0.47-kb fragment that was complementary to the 5′ end of the ubiquitin RNA. This labeled fragment was isolated by gel electrophoresis and used in hybridizations to detect complementary RNA. A sample of total RNA (150 μg) or of poly(A)$^+$ RNA (20 μg) from heat-shocked or non-heat-shocked flies was ethanol precipitated and suspended with 20 μl of hybridization buffer (80% formamide, 0.4 M NaCl, 1 mM EDTA, 40 mM MOPS [morpholinepropanesulfonic acid], pH 7.0). After the labeled DNA probe was added, the samples were covered with 50 μl of mineral oil, heated at 85°C for 15 min, and quickly transferred to 51°C for annealing. After 15 h, 180 μl of S1 buffer (30 mM sodium acetate, 0.1 M NaCl, 2 mM $ZnCl_2$, 5% glycerol, pH 4.5 to 5.0) and S1 nuclease [800 U for total RNA and 320 U for poly(A)$^+$ RNA samples] was added to each tube. After incubation at 15°C for 90 min, the samples were ethanol precipitated and analyzed on a 7% denaturing polyacrylamide gel. *Hpa*II-digested pBR322 DNA was used as a size marker.

For primer extension, a 23-nucleotide primer, complementary to bases 915 through 937 (Fig. 2) (located 4 bp upstream of the intron donor site), was prepared. The 23-mer was labeled at its 5′ end with T4 polynucleotide kinase as described elsewhere (41). The end-labeled primer was annealed to 20 μg of total RNA by being heated at 70°C for 5 min and cooled to room temperature. Reverse transcription was carried out as described previously (26). A 7% denaturing polyacrylamide gel was used to analyze the resulting extension products. *Hpa*II-cut pBR322 DNA was used as a size marker.

Ubiquitin-*lacZ* fusion. The plasmid pPUb (Fig. 1a) was digested with *Bgl*II to liberate a 2-kb fragment containing the upstream region. The 2-kb fragment was then treated with Klenow enzyme to fill in sticky ends and cloned into *Sma*I-digested pMC1871 (8), which contains the β-galactosidase protein-coding region. A 5.1-kb *Sal*I fragment containing the ubiquitin-*lacZ* gene was cloned into SalI-digested Carnegie 20 vector for germ line transformation. Transformation was performed as described elsewhere (37), using *Drosophila melanogaster* $Adh^{fn6}cn$; ry^{502} (ACR). Third-instar larvae, early- and mid-stage pupae, and adults were dissected, stained for β-galactosidase activity, and examined as described previously (15). Embryos of ages 4 to 8 and 12 to 16 h were permeabilized and stained whole as described elsewhere (X. Sun, Ph.D. thesis, Cornell University, Ithaca, N.Y., 1987).

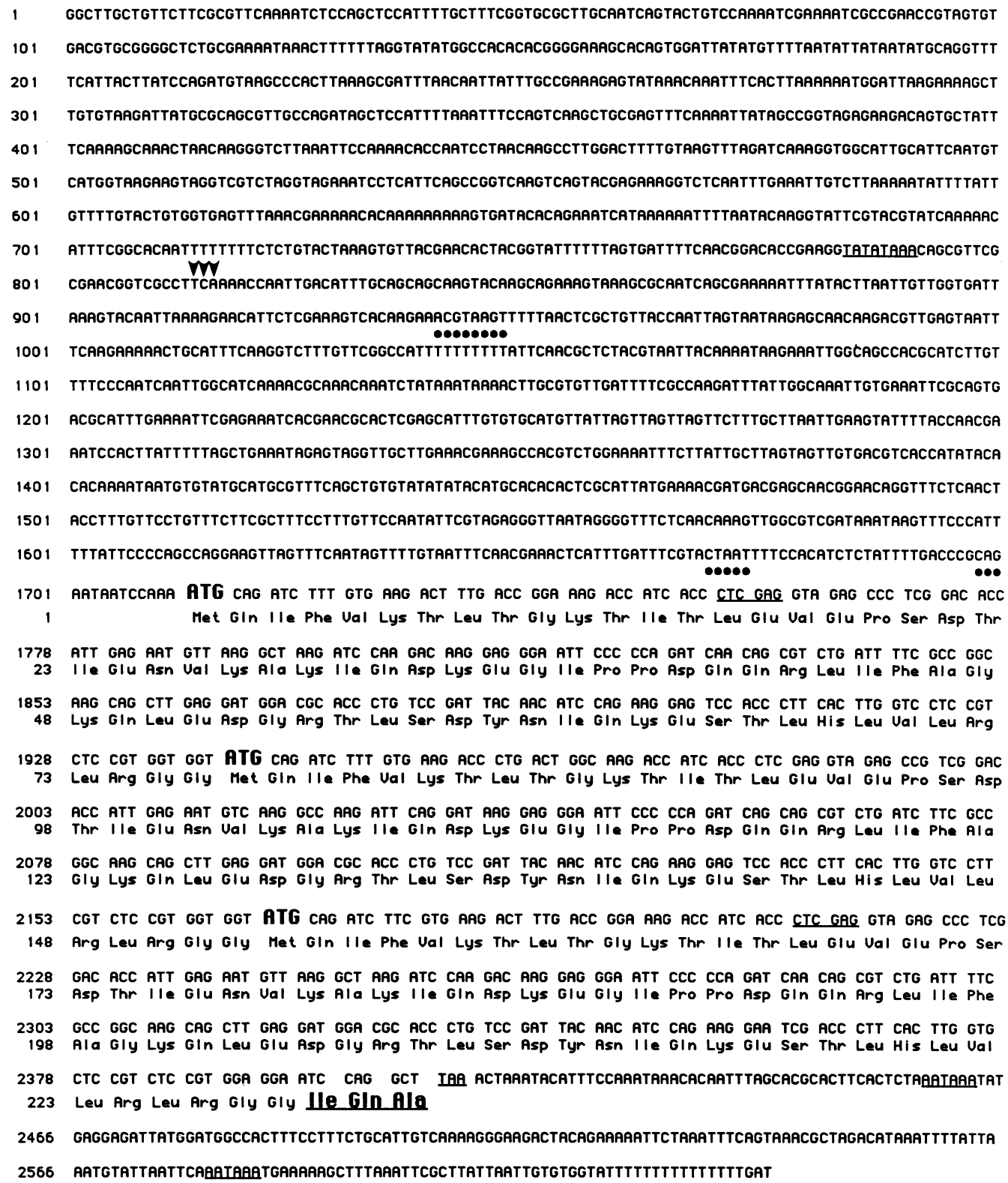

FIG. 2. Nucleotide and deduced amino acid sequences of the polyubiquitin locus. The DNA sequence, which includes the 5′-flanking region through codon 16 of the first ubiquitin-encoding repeat, is fused at the *Xho*I site (underlined) to the sequence determined from a subcloned dimer of internal ubiquitin-encoding repeats. (We sequenced several dimers and monomers of repeats that encode identical amino acid sequences, but we did not determine the order of the 18 repeats and present only a representative dimer.) This sequence is joined at the 3′ end to a sequence starting at amino acid 17 of the terminal repeat (underlined *Xho*I site) and a 3′-flanking region. The first amino acid codon of each repeat is indicated in boldface. Intron donor and acceptor sites and the TACTAAC box (19) are marked by dots. The transcription start sites defined by primer extension and S1 nuclease mapping analysis are indicated by arrows. The putative TATA box is underlined in the 5′-flanking region. Possible polyadenylation signals are underlined in the 3′-flanking region. The three extra amino acid residues following amino acid 76, Gly, in the terminal repeat are boldfaced and underlined.

RESULTS

Genomic structure of the polyubiquitin gene. We isolated the cDNA clone adm63F.1 in a previous search for heat shock gene cDNAs (22). The corresponding RNA was found to be constitutively expressed in cell culture and was induced 4.5-fold by heat shock; hence, this gene, which maps to the cytological locus 63F, has been termed a minor heat shock gene (22). Here, we report that the DNA sequence of adm63F.1 establishes it as a polyubiquitin gene. To isolate an intact genomic clone, the adm63F.1 DNA was used as a probe for Southern hybridization with electrophoretically separated *Bam*HI-*Sal*I-digested genomic DNA. A genomic DNA fragment of approximately 7.5 kb showed hybridization to the ubiquitin cDNA, and DNA of this size range was purified and cloned into pUC19 to generate a small genomic library. With the cDNA fragment used as a probe, the library was screened to obtain a plasmid, pPUb, that contained the intact polyubiquitin gene on a 7.5-kb fragment (Fig. 1a).

A partial *Xho*I digestion of pPUb showed a ladder of bands with a unit size of 228 bp (Fig. 1b). The number of such bands indicated that the *Drosophila* polyubiquitin gene contained 18 tandem ubiquitin-encoding repeats. This determination of the repeat number agrees with an independent measurement derived from quantitative Southern hybridization (23) performed with variable amounts of a fragment containing only one ubiquitin-encoding repeat unit as a standard (data not shown).

The sequences of several repeats of the *Drosophila* ubiquitin-encoding region have been independently determined by Arribas et al. (2). We have cloned monomers and dimers of the ubiquitin-encoding repeat of plasmid pPUb and sequenced three randomly chosen monomers and four dimers of the repeat. The predicted amino acid sequences of the ubiquitin-encoding repeats sequenced were identical (Fig. 2), although the nucleotide sequences varied, especially at the third position of the codon. The predicted amino acid sequence was also identical to that for five of the nine ubiquitin-encoding repeats previously reported (2), but we did not see the frameshift mutation and the amino acid substitution that had been reported for several sequenced repeats (2). In addition, this amino acid sequence was identical to that of human (39) and chicken (5) ubiquitin. Interestingly, the C-terminal ubiquitin-encoding repeat of *D. melanogaster* (which has not been examined previously) encoded 79 residues rather than the 76 characteristic of mature ubiquitin. The extra C-terminal residues were Ile-Gln-Ala following residue 76, Gly. This contrasts with yeast and chicken polyubiquitin, in which a single extra amino acid has been found (5, 30). By blocking the carboxy terminus of the preceding Gly residue, the extra C-terminal residues may serve to prevent participation of unprocessed polyubiquitin in reactions of ubiquitin-protein conjugation (30).

The sequence reported in Fig. 2 also includes sequences, not previously reported, of regions upstream and downstream of the gene and of the RNA leader. During analysis of the sequencing data, we found potential intron donor and acceptor sites within the 5′ noncoding sequence. The existence of a 758-bp intror within the 5′ untranslated leader region was confirmed by sequence analysis of five cDNA clones (Fig. 1a). The intron donor site was located 126 bp downstream of the transcription start site (see below), and the acceptor site was 11 bp upstream of the first codon, AUG. Interestingly, the position of the intron closely resembled that of an intron in the chicken ubiquitin gene, *UBI* (6).

Southern blots of restriction endonuclease-cleaved *Drosophila* DNA were hybridized with labeled DNA containing all of the polyubiquitin-encoding repeats of pPUb. One major and one or two minor bands of hybridization were observed in each lane (Fig. 3). The intensity of the minor bands was highly variable when nonhomozygous lines were analyzed. We suggest that the minor signals resulted from length polymorphisms among alleles of the polyubiquitin gene. The variability among the different fly stocks observed by us and by Izquierdo et al. (2, 18) suggests that the length of the polyubiquitin gene is flexible, expanding or contracting by unequal crossing over between tandemly arranged repeats. However, we cannot exclude the possibility that the weak hybridization signals were due to additional minor polyubiquitin loci that contained only one or a few ubiquitin repeats or repeats having low homology.

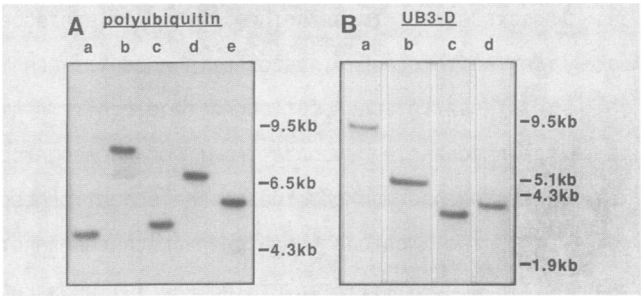

FIG. 3. Southern blot analysis of genomic DNA from Canton S adult flies. (A) Genomic DNA (10 μg) digested with *Sma*I-*Xmn*I (lane a), *Bam*HI (lane b), *Hin*dIII (lane c), *Sal*I (lane d), and *Sma*I-*Nru*I (lane e). Hybridization was carried out as described in the text, using the 0.23-kb ubiquitin-encoding repeat DNA as a probe. (B) Genomic DNA (8 μg) digested with *Bam*HI (lane a), *Hin*dIII (lane b), *Sal*I (lane c), and *Nru*I (lane d). Hybridization was carried out by using a 0.27-kb tail-specific fragment isolated from pUB3-NE as a probe.

The *Drosophila* ubiquitin gene family includes a hybrid gene encoding a ubiquitin monomer fused to an 80-amino-acid tail sequence. The screening of a λgt11 adult fly library with a probe from the ubiquitin-encoding region revealed both strongly and weakly hybridizing plaques. Cloning and sequencing of DNA from these latter plaques revealed a novel ubiquitin-related gene, designated *UB3-D* (Fig. 4). *UB3-D* encoded a protein of 156 amino acids in which a ubiquitin monomer (76 amino acids) was fused to a heterologous tail polypeptide of 80 amino acids. The amino acid sequence of the ubiquitin moiety of the fusion protein was identical to that of polyubiquitin, although the nucleotide sequence was more divergent (83% positional identity). The amino acid sequence of the tail moiety of the protein encoded by *UB3-D* was also homologous to that of a similar ubiquitin hybrid gene product described in other organisms. It had 65% identity to the sequence of yeast UBI3 tail protein (29) and 82% identity to the tail moiety of an analogous human protein (24) (Fig. 4). The tail protein was very basic, containing 30% Arg and Lys, and had the characteristically positioned four Cys residues that could potentially form a zinc finger, DNA-binding motif (20, 29). The fact that the amino acid sequence of the *UB3-D*-encoded protein is conserved among different organisms suggests that it possesses a conserved function.

The number of *UB3-D* genes per genome was determined by hybridization of Southern blots of electrophoretically fractionated genomic DNA from *D. melanogaster* Canton S, using a 0.27-kb tail-specific fragment isolated from plasmid pUB3-NE as a probe (Fig. 3). The existence of one band in

```
                                                                                GAATTCCCTTTTTTGAATTTATTATTCAATTACTACCTTTATTTGATTTCTGGA   54

          AACGTACATGATCTTATGTAGCAAAATTACTTTTCCTTTGATTCTGTTGCCGAAGCAAGTTTGGTGACTGGAATACAAG  133

          ATG CAG ATC TTC GTT AAG ACC CTC ACG GGC AAG ACC ATC ACT CTT GAG GTC GAA CCG TCG  193
D.        Met Gln Ile Phe Val Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu Glu Val Glu Pro Ser   20
H.         -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -
Y.         -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -  Ser  -
                                                                          EcoRI
          GAT ACT ATC GAA AAT GTC AAA GCC AAG ATT CAG GAC AAA GAA GGA ATT CCT CCA GAT CAG  253
D.        Asp Thr Ile Glu Asn Val Lys Ala Lys Ile Gln Asp Lys Glu Gly Ile Pro Pro Asp Gln   40
H.         -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -
Y.         -   -   -  Asp  -   -   -  Ser  -   -   -   -   -   -   -   -   -   -   -   -

          CAG CGT CTG ATC TTC GCT GGC AAG CAG CTG GAA GAT GGC CGC ACC CTC TCT GAC TAT AAC  313
D.        Gln Arg Leu Ile Phe Ala Gly Lys Gln Leu Glu Asp Gly Arg Thr Leu Ser Asp Tyr Asn   60
H.         -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -
Y.         -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -

          ATT CAG AAG GAG TCC ACT TTG CAC TTG GTG CTC CGC CTG CGT GGT GGT GCC  AAG AAG CGC  373
D.        Ile Gln Lys Glu Ser Thr Leu His Leu Val Leu Arg Leu Arg Gly Gly Ala Lys Lys Arg   80
H.         -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   -
Y.         -   -   -   -   -   -   -   -   -   -   -   -   -   -   -   - Gly  -   -   -

          AAG AAG AAG AAC TAC TCC ACT CCC AAG AAA ATC AAG CAC AAG AGG AAG AAG GTC AAG CTA  433
D.        Lys Lys Lys Asn Tyr Ser Thr Pro Lys Lys Ile Lys His Lys Arg Lys Lys Val Lys Leu  100
H.         -   -   - Ser  -  Thr  -   -   -   - Asn  -   -   -   -   -   -   -   -   -
Y.         -   -   - Val  -  Thr  -   -   -   -   -   -   -   - His  -   -   -   -   -

          GCT GTC CTG AAA TAC TAC AAG GTT GAC GAG AAC GGC AAG ATC CAC CGT CTC CGT CGC GAG  493
D.        Ala Val Leu Lys Tyr Tyr Lys Val Asp Glu Asn Gly Lys Ile His Arg Leu Arg Arg Glu  120
H.         -   -   -   -   -   -   -   -   -   -   -   -   -   - Ser  -   -   -   -   -
Y.         -   -   - Ser  -   -   -   -   - Ala Glu  -   - Val Thr Lys  -   -   -   -

          TGC CCC GGT GAG AAC TGC GGC GCC GGC GTC TTC ATG GCT GCC CAC GAA GAT CGT CAC TAC  553
D.        Cys Pro Gly Glu Asn Cys Gly Ala Gly Val Phe Met Ala Ala His Glu Asp Arg His Tyr  140
H.         -   - Ser Asp Glu  -   -   -   -   -   -   -   - Ser  - Phe  -   -   -   -
Y.         - Ser Asn Pro Thr  -   -   -   -   -   - Leu  - Asn  - Lys  -   - Leu  -

          TGC GGC AAG TGC AAC CTG ACC TTT GTC TTC AGC AAA CCA GAG GAA AAG TAA TTTTGCTACAT  615
D.        Cys Gly Lys Cys Asn Leu Thr Phe Val Phe Ser Lys Pro Glu Glu Lys                  156
H.         -   -   -   - Cys  -   - Tyr Cys  - Asn  -   -   - Asp  -
Y.         -   -   -   - His Ser Val Tyr Lys Val Asn Ala                                152

          AAGATCATGTACGTTTCCAGAAATCAAATAAAGGTACTAATTGAATAATAAATTCAATCGCTGAAAAAAAAAAAAAAAA  694

          AAAAA                                                                          699
```

FIG. 4. Nucleotide and deduced amino acid sequences of the *Drosophila UB3-D* cDNA clone (D) and comparison with the amino acid sequences of the homologous human (H) (24) and yeast *UBI3* (Y) (29) genes. Bold characters indicate the first amino acid of the tail moiety of the protein. The arrowhead indicates the site of proteolytic cleavage that would be required to generate mature ubiquitin from the primary translation product. A stretch rich in basic amino acid residues that resembles a nuclear localization signal (11) is underlined in the tail sequence. The amino acid residues of the three clones are numbered from 1 to 156 (H and D) or 152 (Y), beginning with the N termini.

each lane suggests that there is only one copy of the *UB3-D* gene in the genome.

Analysis of RNA encoded by the polyubiquitin and *UB3-D* genes. The sizes and relative amounts of RNAs containing ubiquitin-encoding units were estimated by hybridization of the polyubiquitin-encoding repeat probe to Northern (RNA) blots of total RNA from adult flies. An RNA species of 4.4 kb produced the most prominent signal; however, several smaller and less abundant RNAs were also observed (Fig. 5). The major 4.4-kb mRNA was of a size consistent with that predicted for the full-length polyubiquitin-encoding RNA. The minor band of 0.9 kb in length was apparently the *UB3-D* mRNA, since it hybridized strongly to a probe made from the *UB3-D* gene. The other minor band may represent another monoubiquitin hybrid gene homologous to that found in yeast cells (29) or a minor polymorphic variant of the polyubiquitin gene in the population of flies from which the RNA was isolated (see above).

A previous study established that the original polyubiquitin cDNA clone hybridized to a constitutively expressed RNA species in *Drosophila* cell cultures that was inducible to a 4.5-fold-higher rate of synthesis by heat shock (22).

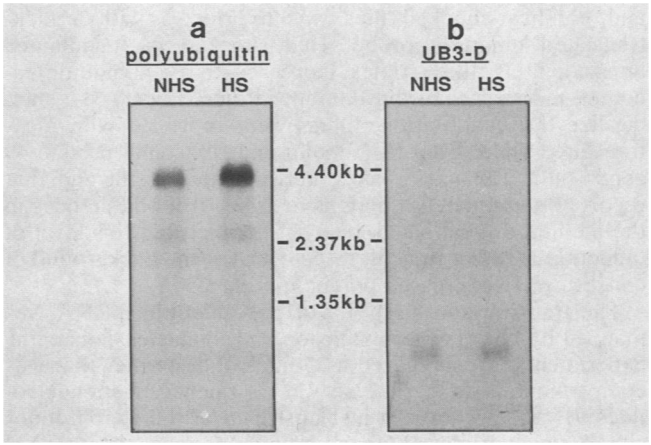

FIG. 5. Northern blot analysis of heat shock and control RNA. Total RNA (30 μg) isolated from non-heat-shocked or heat-shocked flies was fractionated and hybridized with the 0.23-kb ubiquitin-encoding repeat DNA (a) or pUB3-D2 cDNA (b) fragment as probes.

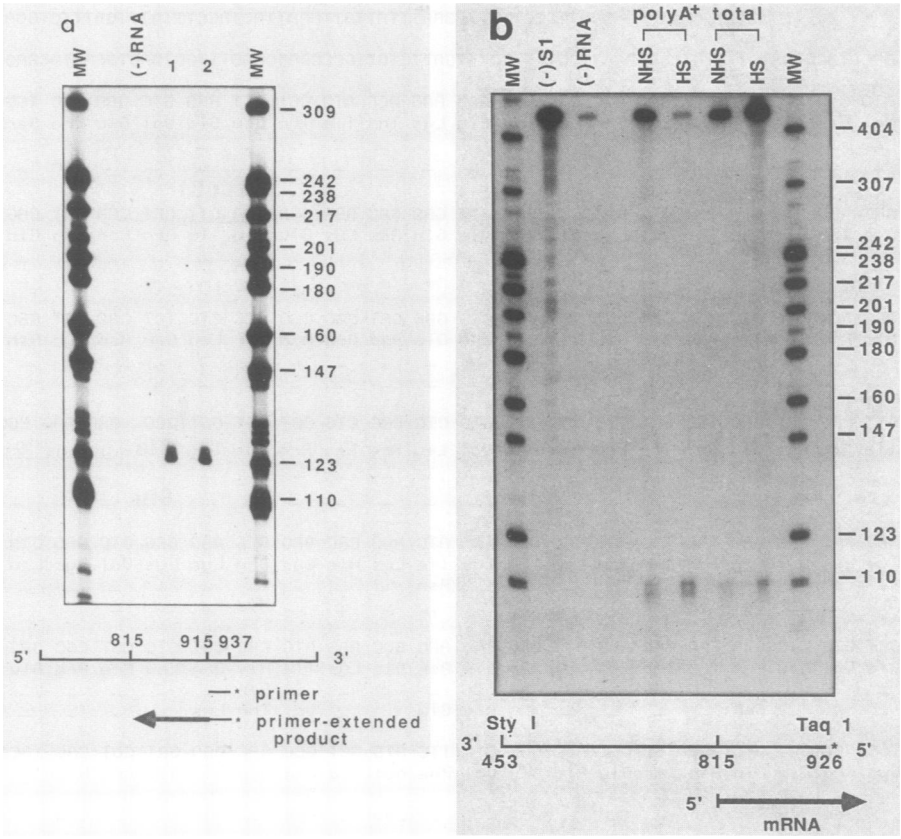

FIG. 6. (a) Primer extension mapping of the transcription start. A 23-nucleotide synthetic primer (5'-CTTGTGACTTTCGAGAATGTTCT-3') complementary to nucleotides 915 through 937 was end labeled and annealed to total RNA (20 μg) from two preparations (lanes 1 and 2). A control primer was mock annealed in the absence of RNA [lane (−)RNA]. Reverse transcription was carried out as described in the text. The same experiment done separately with heat shock RNA yielded a single primer-extended product with the same size as that for non-heat shock RNA (not shown). *Hpa*II-cut pBR322 was used as a size marker, and the fragment sizes (in nucleotides) are indicated in the MW lanes. (b) S1 nuclease mapping of the transcription start. A 0.47-kb *Sty*I-*Taq*I restriction fragment, end labeled with kinase at the *Taq*I site, was hybridized to non-heat shock (NHS) or heat shock (HS) total (150 μg) or poly(A)$^+$ (20 μg) RNA. As controls, hybridizations without RNA plus S1 treatment [lane (−)RNA] and with RNA minus S1 treatment [lane (−)S1] were performed. *Hpa*II-cut pBR322 DNA was used as a size marker; fragments sizes are shown as in panel a.

More recently, a study of polyubiquitin RNA levels in flies suggested that the gene was not induced by heat shock (2). To reexamine this question, RNA from both heat-shocked and non-heat-shocked flies was hybridized with a nick-translated ubiquitin probe. Heat-shocked flies contained approximately three times more 4.4-kb polyubiquitin-encoding message than did uninduced flies (Fig. 5). To standardize the quantitation, filters were reprobed with nick-translated DNA from β-1-tubulin and ribosomal protein 49 genes (40). The discrepancy between this result and that reported previously (2) may have arisen from differences in the method of heat shock or from the fact that this level of induction is rather difficult to detect over the background of constitutively expressed polyubiquitin.

The transcription start site of polyubiquitin mRNA was mapped by both primer extension and S1 nuclease mapping experiments. We prepared a 23-nucleotide primer that was complementary to the known RNA sequences from nucleotides 915 to 937, located just upstream of the intron donor site. A single primer-extended product of approximately 123 bases was observed with non-heat-shock RNA (Fig. 6a) and heat shock RNA (data not shown).

S1 nuclease mapping was also performed. An end-labeled *Sty*I-*Taq*I 0.47-kb fragment was annealed to poly(A)$^+$ RNA or total RNA as described in Materials and Methods. S1 nuclease digestion resulted in an approximately 112-bp protected fragment (Fig. 6b). Since the *Taq*I site used for end labeling is 11 bp upstream of the 3' end of the 23-nucleotide primer, the S1-protected fragment is 11 bp shorter than the primer-extended fragment. Thus, the estimated transcription start site is at nucleotide 815 ± 3 (Fig. 2) for RNAs made both in heat shock-induced and uninduced flies. Sequence analysis of four cDNA clones confirmed this assignment of the start site. In addition, we noted the presence of a TATA box sequence, TATATAAA, 31 bp upstream of this transcription start site and within a GC-rich environment (Fig. 2).

The sequence of one cDNA clone contained an additional 20 bp upstream of the major transcription start site. This may be explained by a minor start site(s) upstream of the major site. This explanation is also consistent with the minor bands observed when the film was exposed longer than that shown in Fig. 6b.

To examine *UB3-D* mRNA levels, the filter was hybridized with a nick-translated pUB3-D2 cDNA fragment. It was found that the *UB3-D* gene was expressed at about a fivefold-higher level than was the β-1-tubulin gene, and the expression of the *UB3-D* gene was not detectably increased upon heat shock (Fig. 5).

FIG. 7. β-Galactosidase activity in polyubiquitin-*lacZ* transformants. Larvae of transformant (left) and parent strain ACR (right) were dissected and stained in X-Gal buffer for 2 h (15).

Expression of polyubiquitin gene is tissue general. The polyubiquitin RNA is present in *Drosophila* cell culture and in whole animals at embryonic, larval, pupal, and adult stages (22; Fig. 5 and data not shown). To facilitate more precise analysis of the tissue-specific and temporal pattern of expression of the polyubiquitin gene, the gene was marked by fusion to the protein-coding region of β-galactosidase. This hybrid gene contained the 2-kb, 5′ *Bgl*II fragment of the polyubiquitin gene (Fig. 1a), which had 1.1 kb of sequences upstream of the transcription start. This hybrid gene was introduced into the *Drosophila* genome by germ line transformation (33). Five independent transformants and the parental ACR flies were dissected in X-Gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) buffer as described in Materials and Methods. The blue reaction product was first observed within 30 min in the transformants, while the endogenous β-galactosidase activity in ACR flies was detected in only a few tissues and only after overnight development as described previously (15). Some variability in staining intensity was observed, but most if not all tissues stained dark blue within 2 h in all of the transformants. No difference in the level of expression was observed between female and male flies. We also compared the expression of the fusion gene in heat-shocked and non-heat-shocked transformants. In heat-shocked flies, blue staining appeared more rapidly but the final pattern was identical. To examine the developmental specificity of the expression, transformant third-instar larvae and early- and mid-stage pupae were dissected as described above. Groups of 4- to 8-h and 12- to 16-h embryos were permeabilized and stained as described in Materials and Methods. We observed a tissue-general pattern of staining at all life stages assayed (Fig. 7), supporting the view that the polyubiquitin gene is expressed constitutively and in a developmentally nonspecific manner.

DISCUSSION

Ubiquitin is encoded by an unusual family of genes. This family has been most extensively characterized in yeast cells (29); however, the basic composition of this family is similar in humans (39) and, as described above, in *D. melanogaster*. A prominent source of the ubiquitin polypeptide is a polyubiquitin gene that contains tandem arrays of ubiquitin-encoding units; in yeast cells, the polyubiquitin gene provides by far the major source of ubiquitin in heat-stressed and stationary-phase cells. Additional ubiquitin is provided by hybrid genes that encode a ubiquitin monomer fused to an unrelated amino acid tail sequence. Both the polyprotein and the hybrid proteins are processed to yield the monomeric units (29, 30) that contribute to the pool of free ubiquitin (13).

The protein sequence of ubiquitin is extremely conserved. The predicted amino acid sequence of the unit repeat of the product of the *Drosophila* polyubiquitin gene is identical to that of chicken and human ubiquitin. The organization of the polyubiquitin genes is also conserved. Ubiquitin-encoding repeats in the polyubiquitin gene are organized in a head-to-tail tandem array with no spacer or termination codon between the repeats. Each ubiquitin unit is directly joined to an adjacent unit by a Gly-Met junction, which in yeast cells has been shown to be cleaved by a specific processing enzyme to generate mature monomeric ubiquitin (29, 30). Although the general structure of the polyubiquitin gene has been conserved, the number of repeats, the number of genes per genome, and the identity of the final amino acid residue in the terminal repeat of the polyprotein are different among species (5, 39). There are 18 repeats in the polyubiquitin gene of our *D. melanogaster* Canton S line; this is the longest ubiquitin gene identified (there are 4 repeats in chicken cells [6], 5 in yeast cells [30], 9 in human cells [39], and approximately 12 in *X. laevis* [12]).

The *Drosophila UB3-D* gene is a ubiquitin hybrid gene encoding a protein with an 80-amino-acid tail sequence. The amino acid sequence of the tail moiety of the *UB3-D* gene product has over 65% homology with that of yeast UBI3 protein and 82% homology with that of the human homolog. The tail amino acid sequences are thus conserved to a high degree over great evolutionary distances and therefore are likely to have similar functions. The tail protein is rich in basic amino acids and has a highly basic stretch of seven residues at the beginning of the tail (Fig. 4) which resembles a sequence motif required for protein localization to the nucleus (11). It has also been suggested that the tail sequence contains a generalized consensus sequence originally described in TFIIIA (20) that forms distinct Zn^{+}-binding domains which interact with nucleic acids (29). These data suggest that either the intact hybrid protein or the 80-residue tail may function as a nucleic acid-binding protein (29).

The polyubiquitin gene has a design that is adapted for the rapid production of large amounts of ubiquitin. However, the existence of multiple genes in the ubiquitin family leaves open the possibility that the polyubiquitin gene is active mainly at specific developmental stages or in specific cell types of multicellular organisms. In yeast cells, the hybrid genes are a major source of ubiquitin in vegetatively growing cells, whereas the polyubiquitin gene is the major source of ubiquitin in stationary-phase cells. Several ubiquitin mRNA species are expressed in both *Dictyostelium discoideum* (14) and *X. laevis* (12), and the relative abundance of these RNAs varies during development. In chickens, a ubiquitin gene that has 94% sequence identity to the coding region of the major polyubiquitin gene (*Ub1*) is expressed in germinal cell

lines during spermiogenesis but has not been detected in other tissues (27).

Examination of the developmental and spatial expression of the polyubiquitin gene in *D. melanogaster* cells in situ by either antibody staining of ubiquitin or hybridization with a ubiquitin cDNA probe is complicated by the multiplicity of genes. To overcome this complication, the polyubiquitin gene was marked by fusion to the *Escherichia coli lacZ* gene and introduced into the germ line by transformation. A large amount of the ubiquitin upstream sequence (1.1 kb) was included on this construct to increase the probability that the hybrid gene would show the normal pattern of polyubiquitin gene expression. This strategy had been successfully applied to generate patterns of expression of the hybrid gene which correspond with those of the endogenous gene (15, 17, 35). Expression of the ubiquitin-*lacZ* gene was examined by using a histochemical stain for β-galactosidase activity and found to be constitutively produced throughout all stages (embryonic, larval, pupal, and adult) of development examined. Since the β-galactosidase activity was relatively stable, we cannot rule out the possibility that quantitative fluctuations in the rate of ubiquitin-*lacZ* expression were masked by β-galactosidase that accumulated during the course of development. However, measurements of the 4.4-kb ubiquitin mRNA by Northern analysis in whole (untransformed) animals confirmed that this gene is expressed constitutively throughout development and is a major ubiquitin-encoding RNA (H. Lee, unpublished data). This tissue-general and developmentally nonspecific expression of polyubiquitin is consistent with its role as one of the major sources of the multifunctional ubiquitin polypeptide.

The levels of free ubiquitin and ubiquitinated H2A rapidly decrease in response to heat shock and other stress treatments (4, 7, 31). Ubiquitin becomes conjugated to a variety of other proteins and in so doing is thought to participate either in the degradation of protein damaged by the stress or in the restoration of these proteins to their native conformations (13). Concomitant with this reduction in free ubiquitin is the enhanced rate of synthesis of polyubiquitin RNA. The level of polyubiquitin RNA has been shown to increase severalfold over normal levels in yeast (13) and chicken (5) cells in response to heat shock. The *Drosophila* polyubiquitin gene is also induced by heat shock, as has been shown in the analysis of pulse-labeled RNA in cell cultures (22) and as shown here by the analysis of RNA levels in flies by Northern blots and S1 assays. Although the polyubiquitin gene encodes a 4.4-kb mRNA which shows threefold induction by heat shock, the 0.9-kb mRNA of *UB3-D* does not show any detectable heat induction.

The observation that the polyubiquitin gene is heat inducible provoked us to search for heat shock consensus elements in its 5′ upstream region. We sequenced 0.8 kb upstream from the transcription start site but failed to find good matches to the heat shock consensus element, either to the original 14-bp sequence (32) or to the modified version consisting of multimers of a 10-bp repeating unit (42). The absence of good heat shock elements in the proximal upstream region is not inconsistent with the fact that the ubiquitin RNA made during heat shock is at a level 5- to 20-fold lower than that of the major heat shock genes (22). This lower level of heat shock expression could be explained by the existence of heat shock consensus elements farther upstream of the sequenced 0.8-kb flanking region or by matches to 1.5 copies of the 10-bp repeating unit (42) present in the leader at position 916 or in the intron at position 1209 (Fig. 2). Alternatively, poor matches to the consensus sequences in the proximal region may weakly bind heat shock transcription factor, resulting in modest stimulation of expression. Finally, the polyubiquitin gene may be induced by a mechanism that is distinct from that of other heat shock genes. Interestingly, both the chicken *UBI* (6) and the yeast polyubiquitin (29) genes contain heat shock consensus elements in their upstream regions, although it is not known whether these elements are functional.

ACKNOWLEDGMENTS

We are grateful to Janis Werner for embryo microinjection, to R. Glaser and V. Vogt for helpful comments on the manuscript, and R. Glaser for help with microscopy.

This work was supported by Public Health Service grant GM25232 to J.T.L. from the National Institutes of Health.

LITERATURE CITED

1. **Anderson, M. W., N. R. Ballal, I. L. Goldknopf, and H. Busch.** 1981. Protein A24 lyase activity in nucleoli of thioacetamide-treated rat liver releases histone 2A and ubiquitin from conjugated protein A24. Biochemistry **20**:1100–1104.
2. **Arribas, C., J. Sampedro, and M. Izquierdo.** 1986. The ubiquitin genes in *D. melanogaster*: transcription and polymorphism. Biochim. Biophys. Acta **868**:119–127.
3. **Ball, E., C. C. Karlik, C. J. Beall, D. L. Saville, J. C. Sparrow, B. Bullard, and E. A. Fyrberg.** 1987. Arthrin, a myofibrillar protein of insect flight muscle, is an actin-ubiquitin conjugate. Cell **51**:221–228.
4. **Bond, U., N. Agell, A. L. Haas, K. Redman, and M. J. Schlesinger.** 1988. Ubiquitin in stressed chicken embryo fibroblast. J. Biol. Chem. **263**:2384–2388.
5. **Bond, U., and M. J. Schlesinger.** 1985. Ubiquitin is a heat shock protein in chicken embryo fibroblasts. Mol. Cell. Biol. **5**:949–956.
6. **Bond, U., and M. J. Schlesinger.** 1986. The chicken ubiquitin gene contains a heat shock promoter and expresses an unstable mRNA in heat-shocked cells. Mol. Cell. Biol. **6**:4602–4610.
7. **Carlson, N., S. Rogers, and M. Rechsteiner.** 1987. Microinjection of ubiquitin: changes in protein degradation in HeLa cells subjected to heat-shock. J. Cell Biol. **104**:547–555.
8. **Casadaban, M. J., A. Martines-Arias, S. A. Shapira, and J. Chou.** 1983. β-Galactosidase fusions for analyzing gene expression in E. coli and yeast. Methods Enzymol. **100**:293–308.
9. **Ciechanover, A., D. Finley, and A. Varshavsky.** 1984. Ubiquitin dependence of selective protein degradation demonstrated in the mammalian cell cycle mutant ts85. Cell **37**:57–66.
10. **Ciechanover, A., D. Finley, and A. Varshavsky.** 1984. The ubiquitin mediated proteolytic pathway and the mechanisms of the energy-dependent intracellular protein degradation. J. Cell. Biochem. **24**:27–53.
11. **Dingwall, C., and R. A. Laskey.** 1986. Protein import into cell nucleus. Annu. Rev. Cell Biol. **2**:366–390.
12. **Dworkin-Rasti, E., E. Shrutkowski, and M. B. Dworkin.** 1984. Multiple ubiquitin mRNAs during Xenopus laevis development contain tandem repeats of the 76 amino acid coding sequence. Cell **39**:321–325.
13. **Finley, D., E. Ozkaynak, and A. Varshavsky.** 1987. The yeast polyubiquitin gene is essential for resistance to high temperature, starvation, and other stresses. Cell **48**:1035–1046.
14. **Giorda, R., and H. L. Ennis.** 1987. Structure of two developmentally regulated *Dictyostelium discoideum* ubiquitin genes. Mol. Cell. Biol. **7**:2097–2103.
15. **Glaser, R. L., M. F. Wolfner, and J. T. Lis.** 1986. Spatial and temporal pattern of hsp26 expression during normal development. EMBO J. **5**:747–754.
16. **Hershko, A., E. Leshinsky, D. Ganoth, and H. Heller.** 1984. ATP-dependent degradation of ubiquitin-protein conjugates. Proc. Natl. Acad. Sci. USA **81**:1619–1623.
17. **Hiromi, Y., A. Kuroiwa, and W. J. Gehring.** 1985. Control elements of the Drosophila segmentation gene fushi tarazu. Cell **43**:603–613.

18. **Izquierdo, M., C. Arribas, J. Galceran, J. Burke, and V. M. Cabrera.** 1984. Characterization of a Drosophila repeat mapping at the early ecdysone puff 63F and present in many eukaryotic genomes. Biochim. Biophys. Acta **783**:114–121.
19. **Keller, E. B., and W. A. Noon.** 1985. Intron splicing: a conserved internal signal in introns of Drosophila pre-mRNAs. Nucleic Acids Res. **13**:4971–4981.
20. **Klug, A., and D. Rhodes.** 1987. 'Zinc fingers': a novel protein motif for nucleic acid recognition. Trends Biol. Sci. **12**:464–469.
21. **Levinger, L., and A. Varshavsky.** 1982. Selective arrangement of ubiquitinated and D1 protein-containing nucleosomes within the Drosophila genome. Cell **28**:375–385.
22. **Lis, J. T., W. Neckameyer, R. Dubensky, and N. Costlow.** 1981. Cloning and characterization of nine heat-shock-induced mRNAs of Drosophila melanogaster. Gene **15**:67–80.
23. **Lis, J. T., L. Prestidge, and D. S. Hogness.** 1978. A novel arrangement of tandemly repeated genes at a major heat shock site in D. melanogaster. Cell **14**:901–919.
24. **Lund, P. K., B. M. Moats-Staats, J. G. Simmons, E. Hovt, A. J. D'Ercole, F. Martin, and J. J. Van Wyk.** 1985. Nucleotide sequence analysis of a cDNA encoding human ubiquitin reveals that ubiquitin is synthesized as a precursor. J. Biol. Chem. **260**: 7607–7613.
25. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
26. **Mansson, P. E., D. Hsu, and D. Stalker.** 1985. Characterization of fruit specific cDNAs from tomato. Mol. Gen. Genet. **200**:356–361.
27. **Mezquita, J., R. Oliva, and C. Mezquita.** 1987. New ubiquitin mRNA expressed during chicken spermiogenesis. Nucleic Acids Res. **15**:9604.
28. **Munro, S., and H. Pelham.** 1985. What turns on heat shock genes? Nature (London) **317**:447–448.
29. **Ozkaynak, E., D. Finley, M. J. Solomon, and A. Varshavsky.** 1987. The yeast ubiquitin genes; a family of natural gene fusions. EMBO J. **6**:1429–1439.
30. **Ozkaynak, E., D. Finley, A. Varshavsky.** 1984. The yeast ubiquitin genes: head-to-tail repeats encoding a polyubiquitin precursor protein. Nature (London) **312**:663–666.
31. **Parag, H. A., B. Raboy, and R. G. Kulka.** 1986. Effect of heat shock on protein degradation in mammalian cells: involvement of the ubiquitin system. EMBO J. **6**:55–61.
32. **Pelham, H. R. B.** 1982. A regulatory upstream promoter element in the Drosophila hsp70 heat shock gene. Cell **30**:517–528.
33. **Rubin, G. M., and A. C. Spradling.** 1982. Genetic transformation of Drosophila with transposable element vectors. Science **218**:348–353.
34. **Siegelman, M., M. W. Bond, W. M. Gallatin, T. St. John, H. T. Smith, V. A. Fried, and I. L. Weissman.** 1986. Cell surface molecule associated with lymphocyte homing is a ubiquitinated branched-chain glycoprotein. Science **231**:823–829.
35. **Simon, J. A., and J. T. Lis.** 1987. A germline transformation analysis reveals flexibility in the organization of heat shock consensus elements. Nucleic Acids Res. **15**:2971–2987.
36. **Simon, J. A., C. A. Sutton, and J. T. Lis.** 1985. Localization and expression of transformed DNA sequences within heat shock puffs of Drosophila melanogaster. Chromosoma **93**:26–30.
37. **Simon, J. A., C. A. Sutton, R. B. Lobell, R. L. Glaser, and J. T. Lis.** 1985. Determinants of heat shock-induced chromosome puffing. Cell **40**:805–817.
38. **Thorne, A. W., P. Sautiere, G. Briand, and C. Crane-Robinson.** 1987. The structure of ubiquitinated histone H2B. EMBO J. **6**: 1005–1010.
39. **Wiborg, O., M. S. Pederson, A. Wind, L. E. Berglund, K. A. Marcker, and J. Vuust.** 1985. The human ubiquitin multigene family: some genes contain multiple directly repeated ubiquitin coding sequences. EMBO J. **4**:755–759.
40. **Wong, Y.-C., P. O'Connell, M. Rosbash, and S. C. R. Elgin.** 1981. DNase I hypersensitive sites of the chromatin for Drosophila melanogaster ribosomal protein 49 gene. Nucleic Acids Res. **9**:6749–6762.
41. **Woods, D.** 1984. Oligonucleotide screening of cDNA libraries. Focus **6(3)**:1–2.
42. **Xiao, H., and J. T. Lis.** 1988. Germline transformation used to define key features of the heat shock response element. Science **239**:1139–1142.
43. **Yarden, T., J. A. Escobedo, W. J. Kuang, T. L. Yang-Feng, T. O. Daniel, P. M. Tremble, E. Y. Chen, M. E. Ando, R. N. Harkins, U. Francke, V. A. Fried, A. Ullrich, and L. T. Williams.** 1986. Structure of the receptor for platelet derived growth factor helps define a family of closely related growth factor receptors. Nature (London) **323**:226–232.