MOLECULAR AND CELLULAR BIOLOGY, Jan. 1989, p. 268–277
0270-7306/89/010268-10$02.00/0

Vol. 9, No. 1

UbiA, the Major Polyubiquitin Locus in *Caenorhabditis elegans*, Has Unusual Structural Features and Is Constitutively Expressed

ROGER W. GRAHAM, DON JONES, AND E. PETER M. CANDIDO*

Department of Biochemistry, Faculty of Medicine, University of British Columbia, Vancouver, British Columbia, V6T 1W5 Canada

Received 25 May 1988/Accepted 6 October 1988

Ubiquitin is a multifunctional 76-amino-acid protein which plays critical roles in many aspects of cellular metabolism. In *Caenorhabditis elegans*, the major source of ubiquitin RNA is the polyubiquitin locus, *UbiA*. *UbiA* is transcribed as a polycistronic mRNA which contains 11 tandem repeats of ubiquitin sequence and possesses a 2-amino-acid carboxy-terminal extension on the final repeat. The *UbiA* locus possesses several unusual features not seen in the ubiquitin genes of other organisms studied to date. Mature *UbiA* mRNA acquires a 22-nucleotide leader sequence via a *trans*-splicing reaction involving a 100-nucleotide splice leader RNA derived from a different chromosome. *UbiA* is also unique among known polyubiquitin genes in containing four *cis*-spliced introns within its coding sequence. Thus, *UbiA* is one of a small class of genes found in higher eucaryotes whose heterogeneous nuclear RNA undergoes both *cis* and *trans* splicing. The putative promoter region of *UbiA* contains a number of potential regulatory elements: (i) a cytosine-rich block, (ii) two sequences resembling the heat shock regulatory element, and (iii) a palindromic sequence with homology to the DNA-binding site of the mammalian steroid hormone receptor. The expression of the *UbiA* gene has been studied under various heat shock conditions and has been monitored during larval moulting and throughout the major stages of development. These studies indicate that the expression of the *UbiA* gene is not inducible by acute or chronic heat shock and does not appear to be under nutritional or developmental regulation in *C. elegans*.

Ubiquitin is a 76-amino-acid protein that has been highly conserved in the evolution of eucaryotes. It has been implicated in vital roles in most fundamental cellular processes (27, 51). In the nucleus, ubiquitin is thought to play important roles in chromatin structure (36), cell cycle events (11, 20), and, potentially, DNA repair (30). In the cytoplasm, ubiquitin is one of the most abundant proteins. It is the mediator of the major ATP-dependent nonlysosomal protein degradation pathway (for a review, see reference 27) and has been found conjugated to actin in insect cytoskeleton (4). Ubiquitin is believed to have a role in cell surface recognition. Antibodies to ubiquitin detect the lymphocyte homing receptor (57), the platelet-derived growth factor receptor (69), growth hormone receptor (35), and several other unidentified external membrane proteins (57). Ubiquitin has also been associated with regulation of the heat shock response (6, 42; U. Bond and M. J. Schlesinger, Abstr. Cold Spring Harbor Meet. Heat Shock 1985, p. 103) and with an autocrine growth factor (44). In most of its roles, monomeric ubiquitin is conjugated to another protein and may act as a recognition signal to guide distinct classes of specialized proteins and enzymes. Recently it has been shown that ubiquitin may itself possess proteolytic activity (22). The X-ray structure of ubiquitin has been determined to a resolution of 0.18 nm (66) and reveals a compact globular protein with a C-terminal tail which extends into the solvent. A hydrophobic core and extensive hydrogen bonding provide ubiquitin with exceptional stability (68), which may be essential for the many cellular environments it occupies.

Much of the recent effort to understand ubiquitin metabolism has concentrated on the study of ubiquitin at the level of gene structure. All organisms examined to date possess a polyubiquitin gene, with the number of tandem repeats varying between organisms, from 3 in *Dictyostelium discoideum* (24) to 15 or more in *Drosophila melanogaster* (2). It is believed that ubiquitin is transcribed as a polycistronic mRNA from these loci, translated into a polyprotein, and then cleaved proteolytically to yield free ubiquitin monomer. Although common in viral systems, this process represents a unique system of gene expression for eucaryotes, whose genes are typically interrupted by introns and are monocistronic (polycistronic mRNAs exist for some secretory proteins and for yeast α factor; however, they all have spacer sequences or a low copy number or both [16, 34]). Since it has been conserved, this unique gene structure probably confers some functional advantage(s) to cells. In *Saccharomyces cerevisiae*, the polyubiquitin locus has been shown to play an essential role in supplying ubiquitin to cells for sporulation and development and in enabling cells to withstand conditions of stress (e.g., high temperature or nutritional deprivation) (21).

We have begun studies on ubiquitin in the nematode *Caenorhabditis elegans*. Using a yeast ubiquitin gene clone (kindly provided by A. Varshavsky, Massachusetts Institute of Technology, Cambridge), we have isolated the *C. elegans* polyubiquitin gene, *UbiA*. We have previously shown that *UbiA* RNA acquires a 22-nucleotide leader sequence by a *trans*-splicing reaction (26). Here we present the complete sequence of *UbiA* and characterize the structure and expression of its transcript. *UbiA* possesses unusual features not seen in ubiquitin genes from other organisms studied to date.

MATERIALS AND METHODS

Screening *C. elegans* genomic libraries. The plasmid pUB2, containing the yeast ubiquitin-coding region, was provided

* Corresponding author.

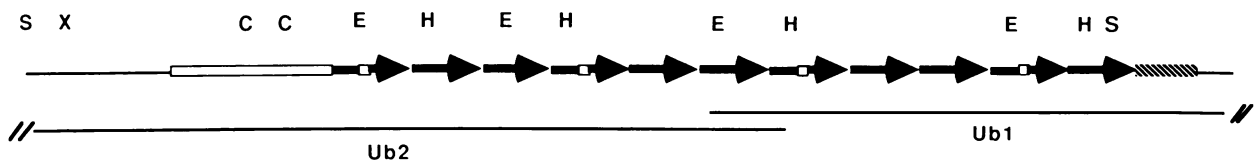


FIG. 1. Structure of the *UbiA* gene of *C. elegans*. Thick arrows indicate tandemly repeated regions encoding monomeric ubiquitin. Open boxes represent introns which are present in the primary transcript of *UbiA*. The shaded box represents the final 2 amino acids of polyubiquitin and the 3′ untranslated region. Lines below the gene represent the two overlapping phage clones, Ub1 and Ub2, which contain *UbiA*. Selected restriction sites are indicated as follows: S, *Sal*I; X, XabI; C, ClaI, H, *Hin*dIII; E, *Eco*RI.

by A. Varshavsky. The ubiquitin-coding region was subcloned into the vector M13mp19 (40), and uniformly radiolabeled probes for library screening were prepared by primer extension of single-stranded recombinant phage DNA as described by Russnak and Candido (53). A λ Charon 4 library of a partial *Eco*RI digest of *C. elegans* genomic DNA was provided by T. Snutch (Simon Fraser University, Burnaby, British Columbia, Canada). Also used was a λgt11 library of a partial *Eco*RI digest provided by R. Barsted (Washington University, St. Louis, Mo.). Both phage libraries were screened by the method of Benton and Davis (5). Positive clones were analyzed by restriction enzyme digestion and Southern blotting (60).

Genomic Southern blot analysis. Genomic DNA was prepared from the *C. elegans* Bristol (N2) and Bergerac (BO) strains as described by Jones et al. (31). Southern transfers, hybridization, and washing of filters were performed by the method of Russnak et al. (53). The hybridization signal was found to retain specificity and increase in intensity under conditions of lowered stringency (5× SSPE–30% formamide, 42°C [1× SSPE is 10 mM sodium phosphate {pH 7.4}, 150 mM NaCl, and 1 mM EDTA]). These conditions were therefore used for all Southern, Northern (RNA), and phage library hybridizations described here.

DNA sequencing. Restriction fragments of genomic DNA were subcloned from λ phage into the sequencing vectors M13mp18 and M13mp19 (40) and were sequenced by the dideoxy-chain termination method of Sanger (54) at 50°C. All cloned DNA was sequenced on both strands or multiple times on one strand. DNA and protein sequences were analyzed by using the computer program described by Delaney (15).

Culture of *C. elegans*. *C. elegans* Bristol (N2) was maintained in liquid culture as described previously (63). Synchronous populations were obtained by inoculating cultures with embryos prepared from gravid adult *C. elegans* by alkaline bleaching (19). Cultures were either fed immediately or were allowed to hatch and arrest as L1 larvae prior to feeding to improve the degree of synchrony (14). Life cycle stages were collected and frozen at −70°C in Buffer A (250 mM sucrose, 10 mM $MgCl_2$, 10 mM Tris hydrochloride, pH 8.0, 1 mM EGTA [ethylene glycol-bis(β-aminoethyl ether)-*N*,*N*,*N*′,*N*′-tetraacetic acid]).

RNA analysis. Total cellular RNA was prepared from synchronized populations of all *C. elegans* life cycle stages by the method of Chirgwin et al. (10), which removes contaminating DNA. This allowed direct quantification of the RNA by spectrophotometry and ethidium fluorescence and hence permitted direct comparisons between samples. Northern (RNA) blot analysis was carried out as described by Thomas (64). Dot blot analysis was performed by using a Bio-Rad manifold. The plasmid pCeA7, containing the *C. elegans* actin 1 gene, was provided by M. Krause (Fred Hutchinson Cancer Research Center, Seattle, Wash.).

RESULTS

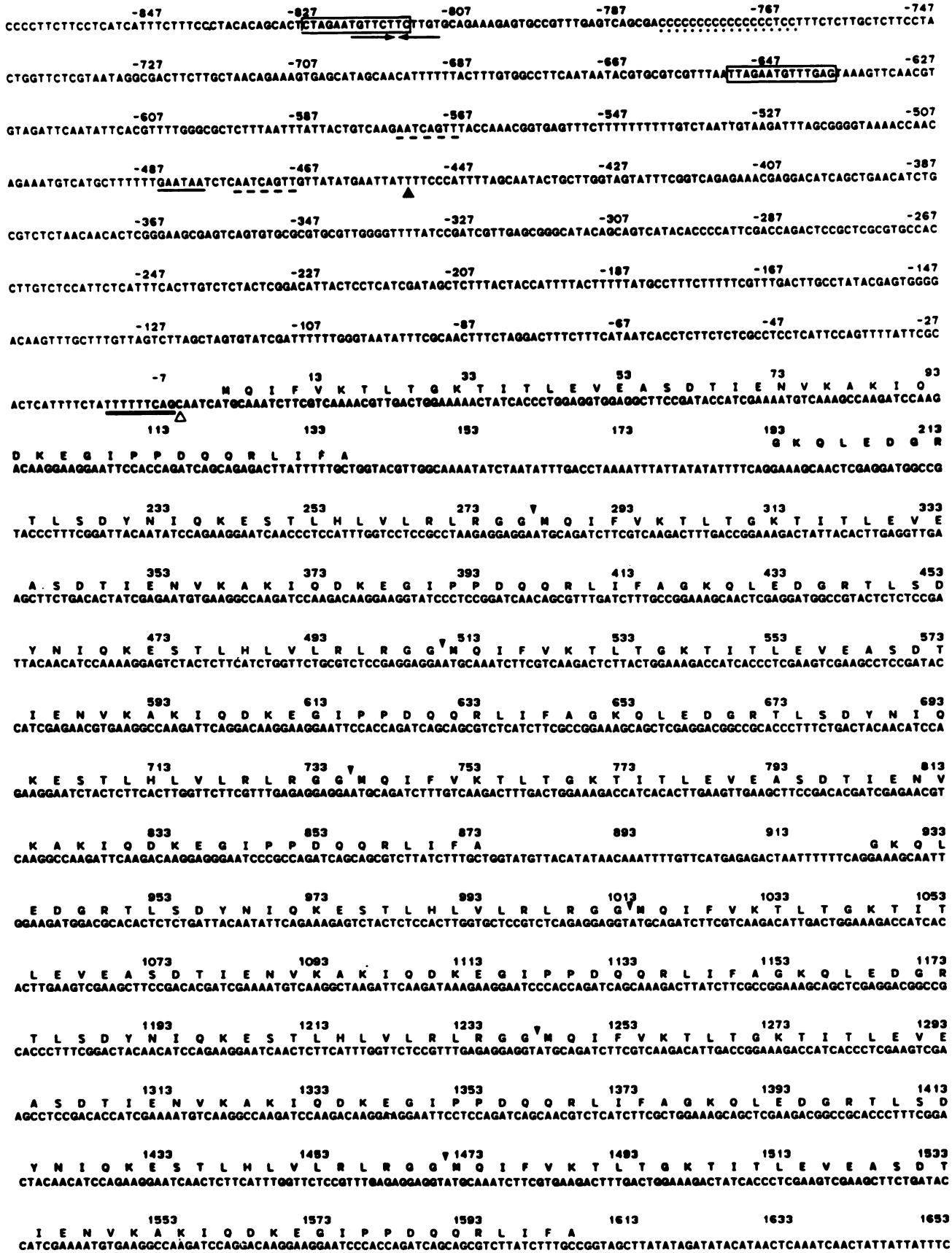
Isolation of genomic clones. Preliminary Southern hybridization analysis confirmed the presence of sequences homologous to yeast ubiquitin in restriction digests of *C. elegans* genomic DNA. A screen of the λ Charon 4 library representing 10 genome equivalents revealed a single class of phage which hybridized to the yeast ubiquitin probe. This phage was termed λUb1. Subclones of λUb1 which hybridized to yeast ubiquitin sequences were subjected to partial DNA sequencing, and it was found that the phage represented an incomplete polyubiquitin gene, lacking the 5′ sequences. We obtained another clone (λUb2) in a subsequent screen of the λgt11 phage library. This clone overlapped λUb1 by 220 base pairs and included the rest of the ubiquitin-coding region as well as the 5′ noncoding sequence. One possible reason for the difficulty we encountered in isolating a complete genomic clone of the polyubiquitin gene may be its tandemly repeated structure, which may be subject to recombinational deletion during the construction of phage libraries.

Localization of the *UbiA* gene. Employing our λUb1 clone, Coulson et al. isolated a series of overlapping cosmids containing the *UbiA* gene as part of their *C. elegans* genome mapping project (12; J. Sulston, personal communication). This cosmid 'contig' extends approximately 100 kilobases (kb) downstream (3′) of the polyubiquitin gene. One of these cosmids, C16A7, was nick translated and used as a probe for in situ hybridization to *C. elegans* chromosomes by D. Albertson of the Medical Research Council Laboratory of Molecular Biology, Cambridge, England). Using this procedure, the *UbiA* gene was mapped to the left half of chromosome III (D. Albertson, personal communication).

Analysis of the *UbiA* gene sequence. A schematic illustration of the 4-kb region containing the *C. elegans* polyubiquitin gene *UbiA* is presented in Fig. 1. The overlapping phage clones λUb1 and λUb2 are indicated by lines. Coding regions of *UbiA* were initially assigned by Southern blotting of restriction digests of λUb1 and λUb2 DNA and were confirmed by sequencing. Introns and flanking noncoding regions of *UbiA* have also been indicated.

We sequenced approximately 4,500 base pairs of DNA within and around the *UbiA* gene (Fig. 2). The coding region consists of 11 tandem copies of monomeric ubiquitin sequence with no spacer DNA between repeats. In addition, there is a 2-amino-acid extension on the final repeat unit. Assuming that the *UbiA* mRNA is translated into a polyprotein, its primary translation product would have a mass of 95 kilodaltons; this polyprotein would then be posttranslationally cleaved into monomeric ubiquitin with a molecular weight of 8,600. The 76-amino-acid ubiquitin sequence is identical for each of the 11 repeats; however, there is degeneracy in the wobble position of some codons.

Genomic Southern hybridization analysis was undertaken to confirm the arrangement of the *UbiA* gene in the Bristol



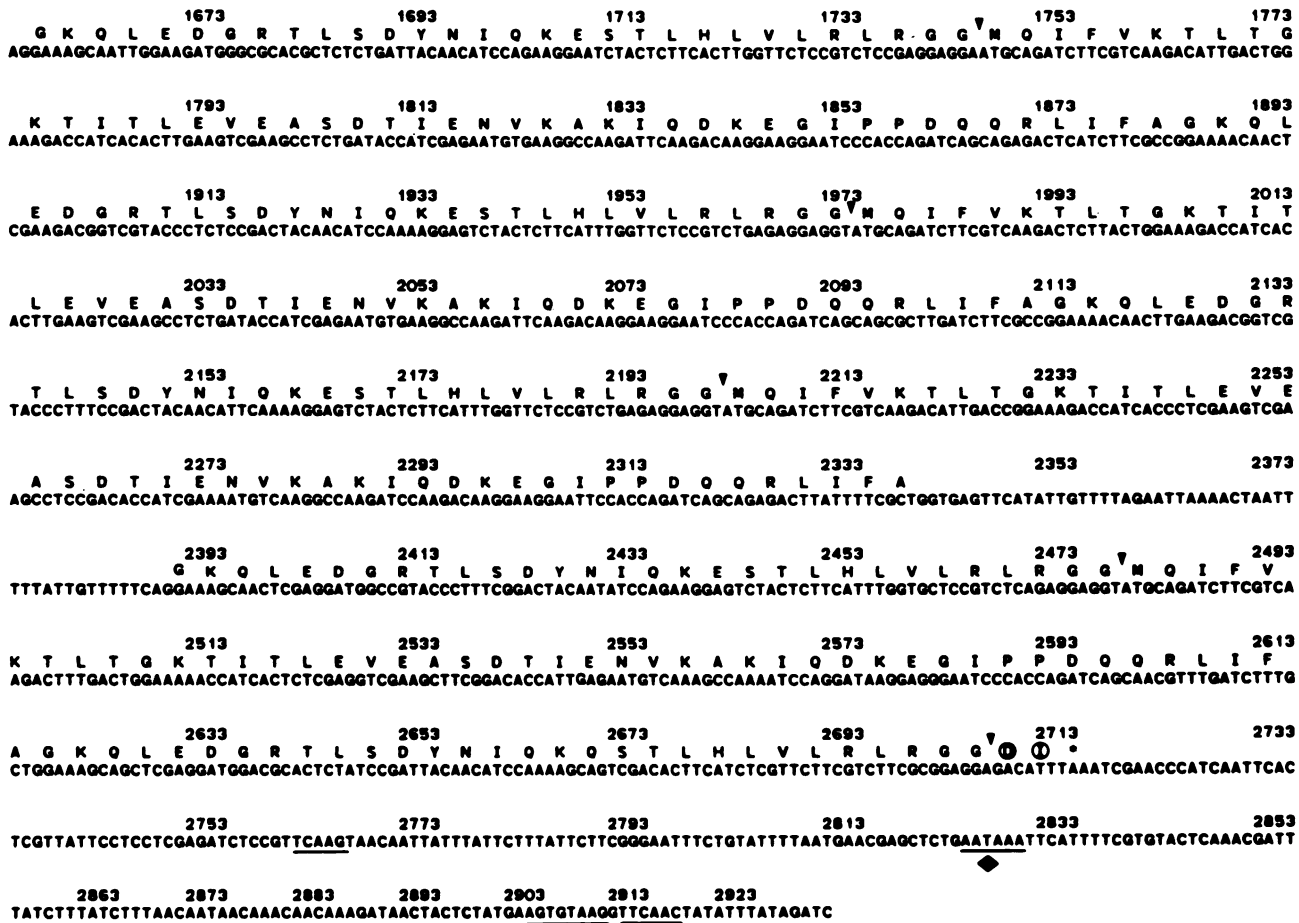


FIG. 2. The sequence of *UbiA*, the polyubiquitin gene of *C. elegans*. The A of the methionine initiator codon of the first ubiquitin repeat unit has been numbered +1; thus, the start site of transcription is at −455 (▲). The translation of *UbiA* into amino acids is given above its nucleotide sequence. Individual repeat units are separated by small arrowheads. The 3′ splice signal sequence used for *trans* splicing is underlined darkly, and the adjacent nucleotide (△), represents the site of ligation of the 5′-spliced leader exon. The TATA box homology is underlined at −487. Various potential regulatory sequences are also indicated; two elements closely matching the HSE consensus are boxed, a sequence resembling the binding site for the mammalian steroid hormone receptor is indicated by two arrows, a block of cytosine nucleotides is underlined with dots, and 8-base-pair direct repeats are indicated by broken underlines. The two nonubiquitin amino acids at the end of polyubiquitin are circled, and the ochre stop codon is indicated by an asterisk. Three sequences potentially involved in efficient mRNA 3′-end formation are underlined, as is the polyadenylation signal (◆).

(N2) strain of *C. elegans* and to determine whether the polyubiquitin structure is conserved in a related *C. elegans* strain, Bergerac (BO). These strains are crossfertile but show a high incidence of DNA polymorphisms due to a higher frequency of insertions of the transposon Tc1 in the Bergerac strain (52). Figure 3 shows a comparison of the two strains of *C. elegans*, BO and N2. Restriction enzymes that cut outside the coding region (*Sal*I) or within the coding region (*Eco*RI) were selected, and digests were probed with an *Eco*RI-*Bgl*II fragment of the coding region which would cross-react with any portion of a ubiquitin repeat sequence. As seen in this figure, the *UbiA* gene was located on a 3.4-kb *Sal*I fragment in the genomic DNA of both strains. Digestion with *Eco*RI yielded the expected hybridizing fragments of 4.0, 1.8 (weak), 0.9, 0.6, and 0.5 kb corresponding to the *UbiA* gene. The 2.3-kb *Eco*RI fragment seen in Fig. 3 originates for a second ubiquitin gene (see below). Similar results were obtained for *Xba*I, *Sst*I, *Pst*I, *Xho*I, *Bgl*II, *Hin*dIII, and *Sau*3AI digest comparisons of DNA from the two strains (data not shown), confirming the organization predicted from λUb1 and λUb2 phage DNA digests.

Northern analysis. Northern analysis of total *C. elegans* RNA using the *Eco*RI-*Bgl*II coding region probe of *UbiA* revealed two main hybridizing mRNAs (Fig. 4). The more abundant mRNA was 2,500 nucleotides in length, sufficient to encode the 11-repeat *UbiA* precursor polypeptide, while the other, less abundant mRNA (*UbiB*) was approximately 700 nucleotides in length, sufficient to encode either three ubiquitin repeats or a ubiquitin sequence fused to an unrelated DNA sequence. Northern hybridization with a DNA probe derived from the 3′ noncoding region of *UbiA* indicated that the *UbiB* mRNA arose from a different ubiquitin-encoding locus and was not a processed form of *UbiA* (26). Genomic Southern data confirmed that at least two distinct ubiquitin loci were present in *C. elegans* DNA and indicated that the gene encoding the 700-nucleotide mRNA was likely wholly or partially contained on a 2.3-kb *Eco*RI fragment or a 12-kb *Sal*I fragment (see above).

3′ Noncoding region. The polyubiquitin precursor polypeptide has a distinct carboxy-terminal sequence; the final ubiquitin repeat has a 2-amino acid extension prior to the ochre stop codon. The 3′ end of *UbiA* has a typical polyad-

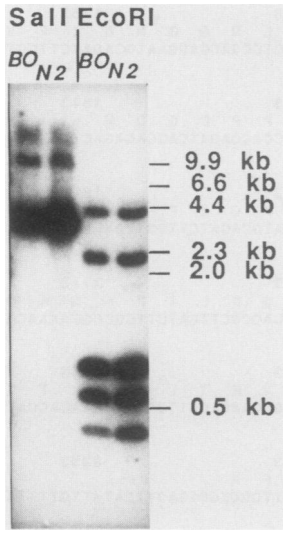


FIG. 3. Analysis of *C. elegans* genomic DNA. Genomic DNA (2 μg) from *C. elegans* Bristol (lanes N2) or Bergerac (lanes BO) were digested with *Sal*I or *Eco*RI, separated by electrophoresis on a 0.7% agarose gel, and transferred to nitrocellulose. These Southern blots were hybridized to a portion of the *UbiA*-coding region (an *Eco*RI-*Bgl*II fragment spanning nucleotides 1344 to 1754) and fluorographed.

enylation signal (AATAAA) 110 base pairs downstream of the stop codon. Also found in this region are many sequences which have previously been shown to be involved in efficient 3′ end formation, including AGTGTAAG (39), TTC AAC, and TCAAG (65) (Fig. 2). This region is also rich in potential secondary structures. The exact site of polyadenylation has not been determined, but is likely to occur at an A residue 20 to 100 nucleotides downstream of the AATAAA sequence.

Intragenic introns. A unique feature of the *C. elegans* polyubiquitin gene is the presence of introns within the coding region. Short introns (50 base pairs) interrupt the Gly-47 codon G/GA in 4 of the 11 *UbiA* repeats; the position of these introns is identical in each repeat. The introns obey the *C. elegans* splice consensus sequences (18).

5′ *trans* intron. The first ubiquitin-encoding repeat is preceded by a sequence which closely resembles a *C. elegans* 3′ splice consensus sequence, TTTCAG (18). We have recently shown that the 3′ splice site in the *UbiA* upstream noncoding region is used for splicing; however, it undergoes a *trans*-splicing reaction, in which a 22-nucleotide leader exon from a separate RNA is joined to this splice acceptor site in the *UbiA* precursor RNA (26). The sequence of this leader is identical to that of the *trans*-spliced leader of the *C. elegans* actin genes, which is derived from the 5′ end of a 100-nucleotide spliced leader RNA (33).

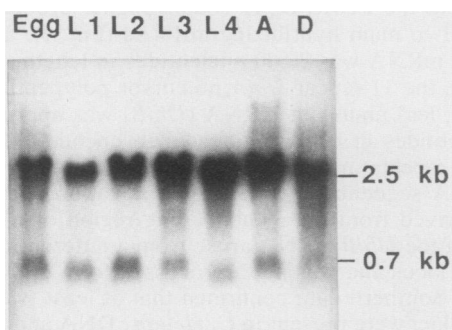


FIG. 4. Northern analysis of ubiquitin mRNA levels during *C. elegans* development. For each life cycle stage, 10 μg of total RNA was separated on a denaturing formaldehyde gel, transferred to nitrocellulose, and hybridized to a radiolabeled fragment of the coding region of *UbiA* (see Fig. 3). The two main ubiquitin-encoding mRNAs are indicated. Lanes: L1-L4, larval stages; A, adult; D, dauerlarva.

5′ End promoter structure. We have previously localized the major site of transcript initiation to a sequence 450 nucleotides upstream of the 3′ splice site used for *trans* splicing (26). The major *UbiA* promoter elements are thus likely to be located immediately upstream of this region. Careful scrutiny of the sequence upstream of the *UbiA* start site reveals no clear TATA motif. The TATA box influences the accuracy of transcript initiation in eucaryotes and is almost always found 20 to 35 nucleotides upstream of the cap site (43). A derivative of the TATA motif, GAATAA, is found 32 nucleotides upstream of the *UbiA* heterogeneous nuclear RNA start site (Fig. 2). This variant of the TATA sequence may account for the inexactness of *UbiA* transcription initiation; minor alternative sites of heterogeneous nuclear RNA initiation can be detected by S1 nuclease analysis (data not shown). Multiple CAAT-box homologies are found within 200 nucleotides of the start site, including one between the GAATAA box and the cap site. The sequence C_{15}TCC at −782 may represent a functionally important regulatory sequence; overall, *C. elegans* DNA is 67% AT rich, so the probability of random occurrence of such a sequence is low.

At positions −653 and −828 are sequences which resemble the heat shock element (HSE) defined by Pelham (C--GAA--TTC--G) (48). This element is found upstream of most eucaryotic heat shock genes and has been shown to confer heat inducibility both in vitro and in vivo (59). The sequence at −653 matches the HSE consensus sequence at six of eight positions, whereas the sequence at −828 matches the HSE consensus at seven of eight positions.

The *UbiA* promoter region also contains an inverted repeat at −816, overlapping the distal HSE. A search of published sequences revealed similarity between this inverted repeat and the binding site(s) for mammalian steroid hormone receptors (55). Whether this inverted repeat functions to confer steroid sensitivity to *UbiA* transcription is not known. This element is a considerable distance from the heterogeneous nuclear RNA start site and hence may be only weakly functional. Another sequence of potential interest in the *UbiA* 5′ end is an 8-nucleotide direct repeat at −477 and −576 (Fig. 2, dashed underline). The significance of this element is unknown.

Expression of ubiquitin. When the cross-reacting *Eco*RI-*Bgl*II sequence from the *UbiA*-coding region was used to probe Northern blots of *C. elegans* RNAs from various life cycle stages, it was found that the level of both ubiquitin mRNAs remained relatively constant throughout development (Fig. 4). In separate titration experiments, we estimated the level of *UbiA* mRNA to be approximately 0.003% of poly(A)$^+$ RNA (data not shown).

In order to investigate the role of the HSE-like sequences in the *UbiA* promoter, we prepared RNA from embryos or synchronized populations of larvae subjected to various heat shock regimes. The kinetics of induction of the *C. elegans* heat shock response have been well characterized (31, 53, 58). *C. elegans* normally grows at temperatures between 14 and 25°C. Typically, heating *C. elegans* in liquid at 33°C for

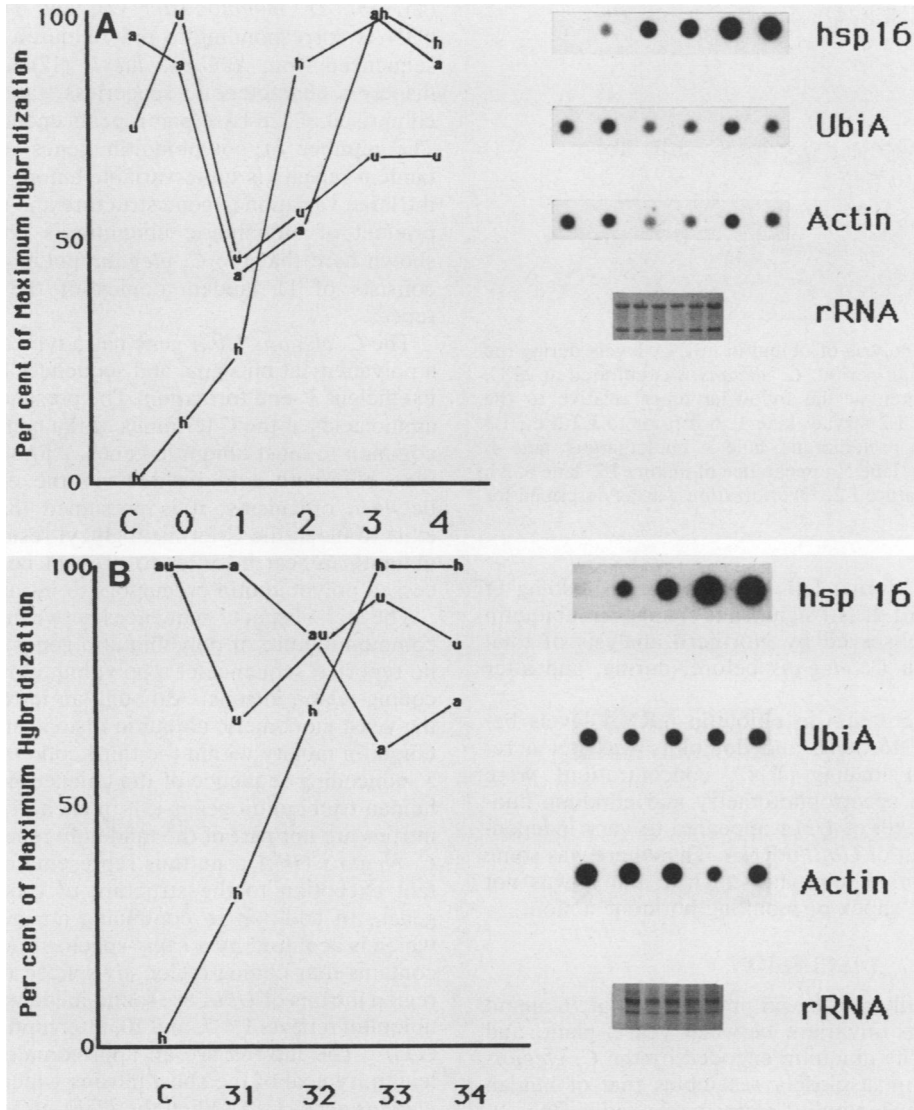


FIG. 5. Dot blot analysis of ubiquitin mRNA levels during heat shock. Total RNA (3 μg) was analyzed at each time point by hybridization to one of the three ^{32}P-labeled DNA probes as indicated. The probe for *UbiA* was a gene-specific *Fnu*DII-*Sst*I fragment from the 3′ end of the *UbiA* locus. Due to the large difference in mRNA abundance, *UbiA* and actin dots were exposed for 3 days, while hsp16 dots were exposed for only 1 day. Hybridization intensity for each dot was determined by densitometry and is expressed graphically as a proportion of the intensity of the sample with maximum hybridization for that probe (a, actin; u, *UbiA*; h, hsp16). Also shown is a photograph of the ethidium fluorescence of the 28S and 18S rRNAs for each RNA sample. (A) *C. elegans* embryos were shifted from 22 to 33°C, and samples were removed at the following times after temperature elevation: 0 min (0), 15 min (1), 30 min (2), 45 min (3), and 60 min (4). C, Control embryos. (B) Embryos were elevated to the indicated temperatures (in degrees Celsius) (C, control, [22°C]), and samples were taken after 45 min.

as little as 15 min induces heat shock mRNAs. We found it necessary to examine multiple samples for a given heat shock experiment; comparisons of single time points for temperature-stressed versus unstressed *C. elegans* was unsatisfactory. A prolonged exposure to 33°C following a slow rise to this temperature resulted in a pronounced induction of hsp16 RNA but did not appear to result in an increase in ubiquitin mRNA levels in *C. elegans* embryos (Fig. 5). The levels of *UbiA* and actin RNA fluctuated somewhat between samples, but there was no linear trend and at most a twofold variation in signal intensity. A similar result was obtained if embryos were rapidly raised to heat shock temperature or when L2 larvae were analyzed. The expression of ubiquitin mRNA in embryos stressed for 45 min at different temperatures is also presented in Fig. 5. Ubiquitin RNA levels fluctuated slightly and nonlinearly as the temperature was increased from 31 to 34°C. This result is similar to the behavior of another noninducible RNA, actin, and in contrast to the behavior of a typical inducible heat shock mRNA such as hsp16, the concentration of which increases linearly over this temperature range. It should be noted that temperatures above 35°C result in significant alterations in cell physiology and viability, and as such do not accurately reflect heat induction (D. Jones, personal communication). It appears from our results that the level of *UbiA* mRNA does not increase significantly during chronic or acute heat stress.

Ubiquitin transcription was also assayed during the larval moulting period from L1 to L2, the transition which represents the greatest change in larval size during *C. elegans* development. Synchronized L1 larvae were sampled prior to

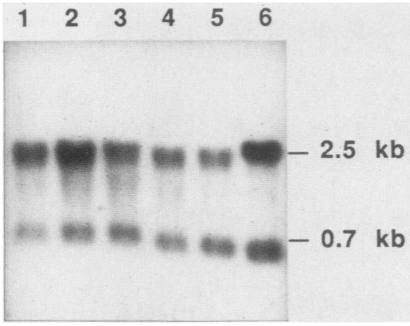


FIG. 6. Northern analysis of ubiquitin mRNA levels during the L1-to-L2 larval moulting period. *C. elegans* was cultured at 22°C, and samples were taken at the following times relative to the appearance of mature L2 larvae: lane 1, 6 h prior to L2 (i.e., L1 stage); lane 2, 45 min prelethargus; lane 3, midlethargus; lane 4, shedding of old cuticle, lane 5, appearance of mature L2; lane 6, 2 h after appearance of mature L2. Hybridization was as described for Fig. 4.

premoult lethargus, during lethargus, through shedding of the cuticle, and into L2. No significant changes in ubiquitin RNA levels were observed by Northern analysis of total RNA prepared from *C. elegans* before, during, and after moulting (Fig. 6).

The observed fluctuation in ubiquitin mRNA levels between samples on Northern and dot blots was not a reflection of unequal loading. RNA concentrations were determined by both spectrophotometry and ethidium fluorescence, and the level of *UbiA* appeared to vary independently relative to that of *UbiB* mRNA. Thus there was some variation in the level of ubiquitin mRNA, but it was not correlated with heat shock or moulting hormone action.

DISCUSSION

Ubiquitin is a highly conserved protein. Out of 76 amino acid residues, 72 are invariant between yeast, plant, and animal ubiquitins. The ubiquitin encoded by the *C. elegans* polyubiquitin gene most closely resembles that of human ubiquitin (67), differing in only 1 amino acid residue (Fig. 7). There is no intragenic variation in amino acid sequence encoded within the nematode polyubiquitin gene, although there is some divergence at the level of codon usage.

Polyubiquitin genes have been isolated from a number of organisms, including humans (3, 67), chickens (7), yeasts (45, 46), *D. melanogaster* (2), and *D. discoideum* (24); cDNAs corresponding to polyubiquitin mRNAs have been sequenced from *Xenopus laevis* (17) and barley (23). All display a characteristic spacerless, tandem-array structure comprised of 228-base-pair repeats encoding 76 amino acids. The number of polyubiquitin genes and the number of tandem repeats is quite variable among organisms. Despite the large variation in gene structure and expression, the final product of monomeric ubiquitin is conserved. We have shown here that the *C. elegans* polyubiquitin gene (*UbiA*) consists of 11 tandem copies of the ubiquitin-encoding repeat.

The *C. elegans UbiA* gene has a typical 3′ end, containing a polyadenylation signal and sequences believed to function in efficient 3′-end formation. The presence of an additional 2 amino acids at the C terminus of the nematode *UbiA* gene is common to most ubiquitin genes. Although the number and type of amino acid residues in the extension may vary between organisms, it is presumed that these sequences exist to block the C-terminal glycyl residues of the adjacent ubiquitin repeat in order to prevent conjugation of unprocessed polyubiquitin precursors to intracellular proteins.

The lack of spacer sequences between tandem repeats is a common feature of polyubiquitin genes. To date, however, no reported sequence of a polyubiquitin gene has contained coding-region introns. Although an intron occurs in two of the yeast monomeric ubiquitin fusion genes (interrupting the ubiquitin moiety within the third codon) (45) and within the 5′ noncoding sequence of the chicken polyubiquitin (7) and human triubiquitin genes (3), in each of these examples the introns are not part of the tandemly repeated sequence. The *C. elegans UbiA* gene thus represents a unique and important exception to the structure of other known ubiquitin genes. In addition to containing an unusual leader intron which is acquired by a *trans*-splicing mechanism, *UbiA* also contains four coding-region *cis*-spliced introns. The coding-region introns of *UbiA* are found in identical locations within ubiquitin repeats 1,4,7, and 10, interrupting the Gly-47 codon G/GA. The introns are all approximately 50 base pairs in length, typical of the short introns which predominate in *C. elegans* genes (18). When the *UbiA cis* intron sequences are aligned, they show much greater similarity to each other than to other short introns in *C. elegans*. The conserved position of introns in the tandem repeats strongly suggests that the nematode polyubiquitin gene evolved by a series of duplication and crossing-over events (56, 61). The repeating,

Human	**P**	**E**	**A**	**S**
Barley	**S**	**D**	**A**	**A**
Yeast	**S**	**D**	**S**	**S**
C. elegans	**A**	**E**	**A**	**S**

MQIFVKTLTGKTITLEVE SDTI NVK KIQDKEGIPPDQQRLIFAGKQLEDGRTL DYNIQKESTLHLVLRLRGG (asterisk over the G of KEG**I**P, Gly-47)

FIG. 7. Interspecies comparison of ubiquitin amino acid sequences. Invariant residues are shown on the central line. Positions of interspecies variation are indicated as gaps; the corresponding amino acid substitutions in human, barley, yeast, and *C. elegans* ubiquitins are indicated. The asterisk indicates the glycine codon interrupted by an intron in 4 of the 11 ubiquitin repeat units of the *C. elegans* polyubiquitin gene.

+ − − structure implies that the duplicating cassette may have been a + − dimer or a + − − trimer.

No consensus TATA motif is evident in the proximity of the transcriptional start site of *UbiA*; however, a GAATAA sequence present at −487 is the likely analog of this element. The sequence of C_{15}TCC located at −782 may have some functional significance to *UbiA* gene regulation. A similar sequence exists upstream of the chicken β-globin gene (38) and is thought to be a site of SP1 transcription factor binding (32). Intriguingly, poly(A) or poly(T) tracts upstream of yeast genes have been implicated in the control of constitutive expression (62).

All organisms respond to chemical or thermal stress by synthesizing heat shock proteins which are thought to transiently protect labile cellular structures and processes (8, 13, 25, 37). Heat shock protein induction is linked to the appearance of abnormal proteins within the cell during stress. For instance, heat shock protein synthesis can be induced at normal growth temperatures in *Drosophila* mutants expressing large amounts of abnormal actin (28) and in *Xenopus* oocytes by the injection of denatured proteins (1). A mammalian temperature-sensitive mutant cell line (ts85) defective in ubiquitin conjugation synthesizes heat shock proteins at the nonpermissive temperature (11, 20), and ubiquitin itself is heat inducible in chicken (6, 7) and yeast (21) cells. There is a shift of ubiquitin toward more conjugated forms, accompanied by an increase in the rate of proteolysis, when rat hepatoma cells are subjected to heat shock (47). Both chicken (7) and yeast (21) polyubiquitin genes possess sequences in their 5′ regulatory region which resemble the HSE consensus sequence defined by Pelham (48), and their mRNA levels increase three- to fivefold during heat or chemical stress. It has also been demonstrated genetically that the yeast *Ubi4* polyubiquitin gene is essential for thermal tolerance (21).

There are two HSE sequences upstream of the nematode *UbiA* gene transcriptional start site, at −653 and −826 (−198 and −371 relative to the transcription start). A kinetic analysis was undertaken to look for changes in ubiquitin mRNA synthesis upon temperature elevation, but no significant effects were seen. These experiments lead us to conclude that *UbiA* transcription is not under thermal control and thus that the HSE sequences present are nonfunctional. Ubiquitin may be regulated at the level of the free ubiquitin protein pool in a cell. As such, chicken cells (with four ubiquitin repeats per mRNA) and yeast cells (with five such repeats per mRNA) might require more mRNA to replenish depleted levels of free ubiquitin during heat shock than do *C. elegans* cells which contain 11 ubiquitin repeats per mRNA. *C. elegans* cells might be able to increase the level of free ubiquitin monomer sufficiently by translational or posttranslational means. Alternatively, one can envision thermal regulation of the synthesis or activity of the enzymes involved in ubiquitin conjugation. The *Drosophila* (2) and *Dictyostelium* (24) polyubiquitin genes lack both HSE sequences and thermal induction of transcription.

The importance of ubiquitin in development has been elegantly demonstrated by a genetic analysis of the yeast *Ubi4* polyubiquitin gene (21). Deletion studies have revealed a requirement for a functional polyubiquitin locus to ensure proper spore development and viability. Ubiquitin-dependent proteolysis has been postulated to play important roles in erythrocyte maturation (50) and eye lens development (29). Developmental induction of one or more ubiquitin mRNA(s) has been documented in several organisms (2, 17, 23, 24, 41).

The levels of *UbiA* and *UbiB* mRNA are relatively constant in all nematode life cycle stages, at both normal and heat shock temperatures. We investigated whether a transient rise in *UbiA* mRNA occurred during the transition from one larval stage to another in *C. elegans*. It was reasoned that the moult should represent a period of increased turnover of intracellular structures and regulatory proteins, which in turn might lead to an increased demand for ubiquitin. We saw only slight fluctuations in ubiquitin mRNA levels, with no correlation to the onset of moulting.

It has been demonstrated that yeast *Ubi4* mRNA levels respond to the nutritional status of the cell, increasing four-to fivefold in stationary-phase (starved) cells relative to exponentially growing vegetative cells (21). Under conditions of starvation or nutritional deprivation, *C. elegans* L2 larvae may enter a resistant life cycle stage termed the dauerlarva (9). Dauerlarvae are resistant to starvation, desiccation, and harmful chemicals such as detergents. We investigated *UbiA* expression in dauerlarvae, on the hypothesis that levels of *UbiA* mRNA might be higher given the stressful environment and the likelihood that abnormal proteins might accumulate in the absence of nutritional input. There was no apparent change in *UbiA* mRNA level, even upon heat shock of dauerlarvae (not shown). Thus dauerlarvae must have an adequate pool of free ubiquitin to fulfill their metabolic requirements.

The level of ubiquitin mRNA in *C. elegans* thus remains relatively constant through development, and during heat stress. This suggests (i) that *UbiA* mRNA is relatively stable and present at sufficient concentration to accommodate temporary changes in the requirement for free ubiquitin, and (ii) that the concentration of free ubiquitin may be regulated at the level of translation or protein processing. Alternatively, the ubiquitin system could be regulated via activity of the conjugating enzymes which attach ubiquitin to target proteins (49).

ACKNOWLEDGMENTS

This work was supported by grants from the Medical Research Council of Canada and the British Columbia Health Care Research Foundation. R.W.G. is a recipient of a Natural Science and Engineering Research Council 1967 Centennial Scholarship.

We are grateful to A. Varshavsky for providing the yeast ubiquitin clone pUB2, to M. Krause for a *C. elegans* actin gene, and to T. Snutch and R. Barsted for *C. elegans* genomic libraries.

LITERATURE CITED

1. **Ananthan, J., A. L. Goldberg, and R. Voellmy.** 1986. Abnormal proteins serve as eukaryotic stress signals and trigger the activation of heat shock genes. Science **232**:522–524.
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. **Baker, R. T., and P. G. Board.** 1987. The human ubiquitin gene family: structure of a gene and pseudogenes from the Ub B subfamily. Nucleic Acids Res. **15**:443–463.
4. **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.
5. **Benton, W. D., and R. W. Davis.** 1977. Screening λgt recombinant clones by hybridization to single plaques *in situ*. Science **196**:180–182.
6. **Bond, U., and M. Schlesinger.** 1985. Ubiquitin is a heat shock protein in chicken embryo fibroblasts. Mol. Cell. Biol. **5**:949–956.
7. **Bond, U., and M. 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.
8. **Burdon, R. H.** 1986. Heat shock and the heat shock proteins. Biochem. J. **240**:313–314.
9. **Cassada, R. C., and R. L. Russell.** 1975. The Dauerlarva, a postembryonic developmental variant of the nematode *Caenorhabditis elegans*. Dev. Biol. **46**:326–342.
10. **Chirgwin, J. M., A. E. Przbyla, R. J. McDonald, and W. F. Rutter.** 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry **18**:5294–5299.
11. **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.
12. **Coulson, A., J. Sulston, S. Brenner, and J. Karn.** 1986. Towards a physical map of the genome of the nematode *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. USA **83**:7821–7825.
13. **Craig, E. A.** 1985. The heat shock response. Crit. Rev. Biochem. **18**:239–280.
14. **Cruzen, M., and T. Johnson.** 1987. Preparation and maintenance of age-synchronous mass cultures. Worm Breeders Gazette **10**: 141.
15. **Delaney, A. D.** 1982. A DNA sequence handling program. Nucleic Acids Res. **10**:61–67.
16. **Douglass, J., O. Civelli, and E. Herbert.** 1984. Polyprotein gene expression: generation of diversity of neuroendocrine peptides. Annu. Rev. Biochem. **53**:665–715.
17. **Dworkin-Rastl, E., A. 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.
18. **Emmons, S. W.** 1988. The genome, p. 47–79. *In* W. B. Wood (ed.), The nematode *Caenorhabditis elegans*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
19. **Emmons, S. W., M. R. Klass, and D. Hirsh.** 1979. Analysis of the constancy of DNA sequences during development and evolution of the nematode *Caenorhabditis elegans*. Proc. Natl. Acad. Sci. USA **76**:1333–1337.
20. **Finley, D., A. Ciechanover, and A. Varshavsky.** 1984. Thermolability of ubiquitin-activating enzyme from the mammalian cell cycle mutant ts85. Cell **37**:43–55.
21. **Finley, D., E. Ozkaynak, and A. Varshavsky.** 1987. The yeast polyubiquitin gene is essential for resistance to high temperatures, starvation, and other stresses. Cell **48**:1035–1046.
22. **Fried, V. A., H. T. Smith, E. Hildebrandt, and K. Weiner.** 1987. Ubiquitin has intrinsic proteolytic activity: implications for cellular regulation. Proc. Natl. Acad. Sci. USA **84**:3685–3689.
23. **Gausing, K., and R. Barkardottir.** 1986. Structure and expression of ubiquitin genes in higher plants. Eur. J. Biochem. **158**: 57–62.
24. **Giorda, R., and H. L. Ennis.** 1987. Structure of two developmentally regulated *Dictyostelium discoideum* ubiquitin genes. Mol. Cell. Biol. **6**:2097–2103.
25. **Goff, S. A., and A. L. Goldberg.** 1985. Production of abnormal proteins in *E. coli* stimulates transcription of *lon* and other heat shock genes. Cell **41**:587–595.
26. **Graham, R. W., K. Van Doren, S. Bektesh, and E. P. M. Candido.** 1988. Maturation of the major ubiquitin gene transcript in *Caenorhabditis elegans* involves the acquisition of a trans-spliced leader. J. Biol. Chem. **263**:10415–10419.
27. **Hershko, A., and A. Ciechanover.** 1986. The ubiquitin pathway for the degradation of intracellular proteins. Prog. Nucleic Acid Res. Mol. Biol. **33**:19–56.
28. **Hiromi, Y., H. Okamoto, W. J. Gehring, and Y. Hotta.** 1986. Germline transformation with *Drosophila* mutant actin genes induces constitutive expression of heat shock genes. Cell **44**: 293–301.
29. **Jahngen, J. H., A. L. Haas, A. Ciechanover, J. Blondin, D. Eisenhauer, and A. Taylor.** 1986. The eye lens has an active ubiquitin-protein conjugation system. J. Biol. Chem. **261**:13760–13767.
30. **Jentsch, S., J. P. McGrath, and A. Varshavsky.** 1987. The yeast DNA repair gene RAD6 encodes a ubiquitin-conjugating enzyme. Nature (London) **329**:131–134.
31. **Jones, D., R. H. Russnak, R. J. Kay, and E. P. M. Candido.** 1986. Structure, expression, and evolution of a heat shock gene locus in *Caenorhabditis elegans* that is flanked by repetitive elements. J. Biol. Chem. **261**:12006–12015.
32. **Jones, K. A., K. R. Yamamoto, and R. Tjian.** 1985. Two distinct transcription factors bind to the HSV thymidine kinase promoter *in vitro*. Cell **42**:559–572.
33. **Krause, M., and D. Hirsh.** 1987. A trans-spliced leader sequence on actin mRNA in *C. elegans*. Cell **49**:753–761.
34. **Kurjan, J., and I. Herskowitz.** 1982. Structure of a yeast pheromone gene (MFα): a putative α-factor precursor contains four tandem copies of mature α-factor. Cell **30**:933–943.
35. **Leung, D. W., S. A. Spencer, G. Cachianes, R. G. Hammonds, C. Collins, W. J. Henzel, R. Barnard, M. J. Waters, and W. I. Wood.** 1987. Growth hormone receptor and serum binding protein: purification, cloning and expression. Nature (London) **330**:537–543.
36. **Levinger, L., and A. Varshavsky.** 1982. Selective arrangement of ubiquitinated and D1 protein-containing nucleosomes within the *Drosophila* genome. Cell **28**:375–385.
37. **Lindquist, S.** 1986. The heat shock response. Annu. Rev. Biochem. **55**:1151–1191.
38. **McGhee, J. D., W. I. Wood, M. Dolan, J. D. Engel, and G. Felsenfeld.** 1981. A 200 base pair region at the 5′ end of the chicken adult β-Globin gene is accessible to nuclease digestion. Cell **27**:45–55.
39. **McLauchlin, J., D. Gaffney, J. L. Witton, and J. B. Clements.** 1985. The consensus sequence YGTGTTYY located downstream from the AATAAA signal is required for efficient formation of mRNA 3′ termini. Nucleic Acids Res. **13**:1347–1368.
40. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–78.
41. **Mezquita, J., R. Oliva, and C. Mezquita.** 1987. New ubiquitin mRNA expressed during chicken spermiogenesis. Nucleic Acids Res. **15**:9604.
42. **Munro, S., and H. Pelham.** 1985. What turns on heat shock genes. Nature (London) **317**:477–478.
43. **Nevins, J. R.** 1983. The pathway of eukaryotic mRNA formation. Annu. Rev. Biochem. **52**:441–466.
44. **Okabe, T., M. Fujisawa, A. Mihara, S. Sato, N. Fujiyoshi, and F. Takuku.** 1986. Amino-terminal amino acid sequence of a novel autocrine growth factor: homology with ubiquitin. J. Cell Biol. **103**:442A.
45. **Ozkaynak, E., D. Finley, M. Solomon, and A. Varshavsky.** 1987. The yeast ubiquitin genes: a family of natural gene fusions. EMBO J. **6**:1429–1439.
46. **Ozkaynak, E., D. Finley, A. Varshavsky.** 1984. The yeast ubiquitin gene: head-to-tail repeats encoding a polyubiquitin precursor protein. Nature (London) **312**:663–666.
47. **Parag, H. A., B. Raboy, and R. G. Kulka.** 1987. Effect of heat shock on protein degradation in mammalian cells: involvement of the ubiquitin system. EMBO J. **6**:55–61.
48. **Pelham, H. R. B.** 1982. A regulatory upstream promoter element in the *Drosophila* HSP70 heat-shock gene. Cell **30**:517–528.
49. **Pickart, C. M., and I. A. Rose.** 1985. Functional heterogeneity of ubiquitin carrier proteins. J. Biol. Chem. **260**:1573–1581.
50. **Rapoport, S. M., W. Dubiel, and M. Muller.** 1981. The mechanism of maturation-dependent breakdown of mitochondria in reticulocytes. Acta Biol. Med. Germ. **40**:1277–1283.
51. **Rechsteiner, M.** 1987. Ubiquitin-mediated pathways for intracellular proteolysis. J. Biol. Chem. **3**:1–30.
52. **Rosenzweig, B., L. W. Liao, and D. Hirsh.** 1983. Sequence of the *C. elegans* transposable element Tc1. Nucleic Acids Res. **11**: 4201–4209.
53. **Russnak, R. H., and E. P. M. Candido.** 1985. Locus encoding a family of small heat shock genes in *Caenorhabditis elegans*: two genes duplicated to form a 3.8-kilobase inverted repeat. Mol. Cell. Biol. **5**:1268–1278.
54. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
55. **Scheidereit, C., S. Geisse, H. M. Westphal, and M. Beato.** 1983. The glucocorticoid receptor binds to defined nucleotide se-

quences near the promoter of mouse mammary tumour virus. Nature (London) **304**:749–752.
56. **Sharp, P. M., and W. Li.** 1987. Ubiquitin genes as a paradigm of concerted evolution of tandem repeats. J. Mol. Evol. **25**:58–64.
57. **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.
58. **Snutch, T. P., and D. L. Baillie.** 1983. Alterations in the pattern of gene expression following heat shock in the nematode *Caenorhabditis elegans*. Can. J. Biochem. Cell Biol. **61**:480–487.
59. **Sorger, P. K., M. J. Lewis, and H. R. B. Pelham.** 1987. Heat shock factor is regulated differently in yeast and HeLa cells. Nature (London) **329**:81–84.
60. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–518.
61. **Stone, E. M., K. N. Rothblum, and R. J. Schwartz.** 1985. Intron dependent evolution of chicken glyceraldehyde phosphate dehydrogenase gene. Nature (London) **313**:498–500.
62. **Struhl, K.** 1985. Naturally occurring poly(dA-dT) sequences are upstream promoter elements for constitutive transcription in yeast. Proc. Natl. Acad. Sci. USA **82**:8419–8423.
63. **Sulston, J. E., and S. Brenner.** 1974. The DNA of *Caenorhabditis elegans*. Genetics **77**:95–104.
64. **Thomas, P. S.** 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. USA **77**:5201–5205.
65. **Urano, Y., K. Wantanbe, M. Sakai, and T. Tamaoki.** 1986. The human albumin gene. Characterization of the 5′ and 3′ flanking regions and the polymorphic gene transcripts. J. Biol. Chem. **261**:3244–3251.
66. **Vijay-Kumar, S., C. E. Bugg, and W. J. Cook.** 1987. Structure of ubiquitin refined at 1.8 Å resolution. J. Mol. Biol. **194**:531–544.
67. **Wiborg, O., M. S. Pedersen, 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.
68. **Wilkinson, K. D., M. J. Cox, L. B. O'Connor, and R. Shapira.** 1986. Structure and activities of a variant ubiquitin sequence from Bakers' yeast. Biochemistry **25**:4999–5004.
69. **Yarden, Y., J. A. Escobedo, W.-J. Kuang, T. L. Yang-Feng, T. O. Daniel, P. M. Tremble, E. Y. Chen, M. E. Ando, R. N. Harkine, 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.