

Structure of Two Developmentally Regulated *Dictyostelium discoideum* Ubiquitin Genes

ROBERTO GIORDA AND HERBERT L. ENNIS*

Roche Institute of Molecular Biology, Roche Research Center, Nutley, New Jersey 07110

Received 17 November 1986/Accepted 10 March 1987

A previously isolated cDNA clone, pLK229, that is specific for mRNA developmentally expressed during *Dictyostelium discoideum* spore germination and multicellular development, was used to screen two genomic libraries. Two genomic sequences homologous to pLK229 were isolated and sequenced. Genomic clone p229 is identical to the cDNA clone pLK229 and codes for a polypeptide of 381 amino acids. This polypeptide is composed of five tandem repeats of the same 76-amino-acid sequence. Clone λ 229 codes for a protein of 229 amino acids, containing three tandem repeats of the identical 76-amino-acid sequence. A computer search for homology to known proteins revealed that the 76-amino-acid repeat was identical to human and bovine ubiquitin except for two amino acid differences.

The simple eucaryote *Dictyostelium discoideum* grows as single cells. Depletion of the food supply sets in motion an orderly succession of developmental events that subsequently results in the formation of a multicellular structure (31). During the transition from single cell to multicellularity, the differential expression of certain genes and their products results in the accumulation of developmentally regulated proteins which may be involved in the program of cytodifferentiation leading to construction of the terminally differentiated fruiting body (1, 4). The fruiting body contains dormant spores which can germinate under the appropriate conditions, each yielding a single amoeba (9). Spore germination in *D. discoideum*, similar to other stages in slime mold development, is accompanied by developmentally regulated changes in both protein and mRNA synthesis (10, 13, 14).

In previously reported work we have identified proteins that are developmentally regulated and specific for spore germination and have cloned a number of mRNAs which are specifically expressed during germination (10, 18). RNA blot analysis revealed that mRNA specific to a cDNA clone named pLK229 was present in a very low concentration throughout the life cycle of *D. discoideum*, but its level increased dramatically in spores and in germinating spores.

In the present study we report on the sequences of two genomic clones homologous to pLK229. The sequences of the deduced proteins are almost identical to that for human ubiquitin, a 76-amino-acid protein whose sequence is conserved in many organisms (26, 28, 35).

MATERIALS AND METHODS

Organism. *D. discoideum* Ax3 was grown axenically in HL-5 medium (32). Spore germination of *D. discoideum* spores was performed as previously described (12).

Isolation of RNA and DNA. Total RNA was isolated from dormant and activated spores and vegetatively growing cells as described before (19). Poly(A)⁺ RNA was purified by oligo(dT)-cellulose chromatography (2).

Plasmid DNA was isolated by the method of Clewell and Helinski (8). Small-scale preparations of plasmid and bacteriophage DNA and large-scale phage DNA isolation were performed as previously described (24). Fragments generated by restriction enzyme digestion of the DNA were isolated by electrophoresis and subsequent electroelution from the gel as previously described (19).

RNA blot and Southern analysis. For RNA blots, poly(A)⁺ RNA was size fractionated on agarose-formaldehyde gels as previously described (19). The gel was then soaked in 25 mM sodium phosphate buffer (pH 6.5) for 40 min, and the RNA was transferred to GeneScreen by using the same buffer.

D. discoideum nuclear DNA was digested to completion with various restriction enzymes, fractionated on a 0.8% agarose-TAE gel (TAE is 0.04 M Tris acetate, 2 mM EDTA), and transferred to a nitrocellulose sheet by the method of Southern (29).

The filters from both of the above procedures were baked for 2 h at 80°C in vacuo, prehybridized for 8 to 14 h at 37°C in 10 ml of 50% formamide-2× SSC (0.3 M NaCl, 0.03 M sodium citrate)-0.12 M sodium phosphate (pH 7.1)-2 mM EDTA-0.2% sodium dodecyl sulfate-1% Sarkosyl (CIBA-GEIGY Corp., Summit, N.J.)-4× Denhardt solution (0.08% Ficoll 400 [Pharmacia Fine Chemicals, Piscataway, N.J.], 0.08% polyvinylpyrrolidone, 0.08% bovine serum albumin), and hybridized at 37°C for 18 to 24 h in 4 ml of the above solution containing 5 × 10⁵ counts per min per ml of the nick-translated probe. The filters were washed with 0.2× SSC containing 0.1% sodium dodecyl sulfate twice for 20 min at room temperature and twice for 20 min at 37°C (low stringency) or 50°C (high stringency). The air-dried filters were exposed at -70°C to X-ray film (Kodak XAR-5; Eastman Kodak Co., Rochester, N.Y.) by using a sandwich of two Dupont Cronex intensifying screens.

Labeling of DNA probe. Plasmid pLK229 was nick translated to a specific activity of about 10⁸ counts per min per μg with [α -³²P]dCTP (specific activity, >400 Ci/mmol) and a nick-translation kit, according to the specifications of the manufacturer.

Construction and screening of genomic libraries. One genomic library was constructed by the insertion of fragments of 4.5 to 5.5 kilobases (kb) from a *Sau*3A partial

* Corresponding author.

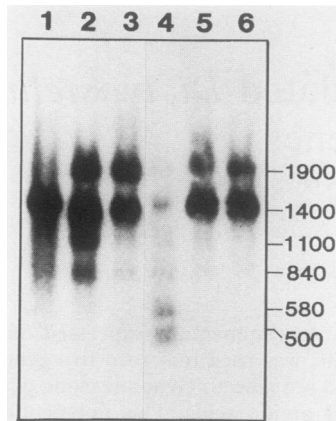


FIG. 1. RNA blot analysis of poly(A)⁺ RNA isolated from dormant spores (lane 1), spores at 1.5 (lane 2) and 3 h (lane 3) of germination, and vegetative cells (lane 4) at 14 h (lane 5) and 20 h (lane 6) of development (see reference 1 for method). The RNAs (1 μ g per lane) were fractionated on a 1.5% agarose–6% formaldehyde gel, transferred to GeneScreen, and hybridized to nick-translated pLK229. The filter was washed at low stringency and exposed for 17 h, as described in Materials and Methods. The approximate sizes of the six bands are indicated.

digestion of *D. discoideum* M2 DNA into the *Bam*HI site of plasmid pAT153 (33). The library was a generous gift of A. Ceccarelli, Torino, Italy. *Escherichia coli* HB101 cells were transformed by the plasmid library, and 10⁵ colonies were screened (24) with nick-translated pLK229 as a probe. The pLK229-specific clone isolated was called p229. A second library was constructed by inserting fragments of 2.2 to 7.4 kb, resulting from a complete *Eco*RI digestion of strain Ax3 DNA, into the *Eco*RI site of λ gt 10 phage. The phage DNA was packaged and used to infect *E. coli* BNN102 cells, and 1.2 \times 10⁵ plaques were screened (24) by using nick-translated pLK229 as a probe. This pLK229-specific clone was called λ 229.

Sequencing. The M13 cloning and sequencing procedures followed were those of the manufacturer of the cloning and sequencing kits, with minor modifications. The cDNA insert of pLK229 was first cloned in the *Pst*I site of M13 mp18 (25) and sequenced from both ends by the dideoxy method (27). Deoxyadenosine 5' [α -³⁵S]thiotriphosphate (650 Ci/mmol) was used because it gave better resolution than [α -³²P]dATP did. The complete sequence was obtained by subcloning and sequencing the *Hpa*II fragments of the insert.

The *Eco*RI fragment of clone p229 (named p229C), which contains the coding region, and the entire insert of clone λ 229 were cloned in both directions into the *Eco*RI site of M13 mp18, and one strand was completely sequenced by the previously described exonuclease III method (16). Where necessary, the second strand was also partially sequenced.

The sequences obtained were stored and processed by use of the Bion program developed by Intelligenetics Inc., Mountain View, Calif. The amino acid composition of the proteins deduced from the obtained DNA sequences was used to search the National Biomedical Research Foundation protein structure data base for homology to known proteins by use of the VAX 11/780 computer (Hoffmann-La Roche Inc., Nutley, N.J.). We thank D. Webb for assistance in using this computer program.

Sources of materials. Restriction endonucleases were obtained from Bethesda Research Laboratories, Inc., Bethesda, Md., and New England BioLabs, Inc., Beverly,

Mass. Bethesda Research Laboratories was the source of nitrocellulose sheets, the nick-translation kit, *E. coli* HB101-competent cells, M13 cloning and sequencing kits, the sequencing gel electrophoresis system, exonuclease III, T4 ligase, and the Klenow fragment of DNA polymerase I. GeneScreen was obtained from New England Nuclear Corp., Boston, Mass.; oligo(dT)-cellulose was from Collaborative Research, Inc., Bedford, Mass.; elutip-d was from; Schleicher & Schuell, Inc., Keene, N.H.; [α -³²P]dCTP and deoxyadenosine-5' [α -³⁵S]thiotriphosphate was from Amer-sham Corp., Arlington Heights, Ill.; and λ gt 10 arms and the Gigapack packaging system were from Stratagene, San Diego, Calif.

RESULTS

Developmental regulation of pLK229 mRNA accumulation.

RNA blot analysis of RNA from vegetative cells, germinating spores, and multicellular development indicates that pLK229-specific mRNA is developmentally regulated. Poly(A)⁺ RNA was isolated from vegetatively growing cells, from dormant and germinating spores (1.5 and 3 h), and from cells after 14 and 20 h of development on filters (1). Poly(A)⁺ RNA (1 μ g) was size fractionated on a formaldehyde-agarose gel as described in the Materials and Methods section. The RNAs were transferred to GeneScreen sheets and hybridized to nick-translated pLK229 plasmid DNA. As was described previously, by using total RNA (18) one species of pLK229-specific mRNA of about 1,400 nucleotides (nt) was found. Its concentration was highest in spores and declined considerably during germination. Little was present in growing cells.

However, under the conditions of hybridization of the present experiment, using poly(A)⁺ RNA instead of total RNA, other mRNA species which hybridized to pLK229 plasmid DNA of about 1,900, 1,400, 1,100, 840, 580, and 500 nt (Fig. 1) were observed. Only the 1,400-nt species was present in spores (lane 1). At 1.5 h (lane 2) of germination four mRNAs were observed of 1,900, 1,400, 1,100, and 840 nt, whereas at 3 h (lane 3) the predominant species were the 1,900- and 1,400-nt mRNAs and only traces of the 1,100- and 840-nt mRNAs were present.

Little of the 1,900- to 840-nt mRNAs was present in growing cells (lane 4), but two new bands at 580 and 500 nt were observed. During multicellular development on filters, these two mRNAs disappeared and the 1,900- and 1,400-nt species predominated (lanes 5 and 6).

The observed difference in intensity of hybridization between the different species of mRNA can be attributed to differences in abundance of the mRNAs or to differences in the number of ubiquitin repeats (see later figures). These two points cannot be resolved from the data presented in Fig. 1.

Sequence of genomic and cDNA clones. Southern blots of *Dictyostelium* DNA digested with various restriction enzymes indicated that the pLK229 genes constitute a multigene family. The blots presented in Fig. 2 indicate at least six genes. It is noteworthy that, as will be seen later, there are no sites for *Eco*RI and *Cl*aI in the coding region of the genomic clones we have isolated.

Two sequences homologous to pLK229 were isolated by screening about 100,000 colonies from the plasmid library and 120,000 plaques from the λ gt 10 phage library. Clone p229, isolated from the plasmid library, contained an insert of about 5.2 kb (Fig 3). Only the *Eco*RI fragment between the pAT153 *Eco*RI site and the first *Eco*RI site in the *Dictyostelium* insert (1,500 base pairs from the *Dictyostelium* insert

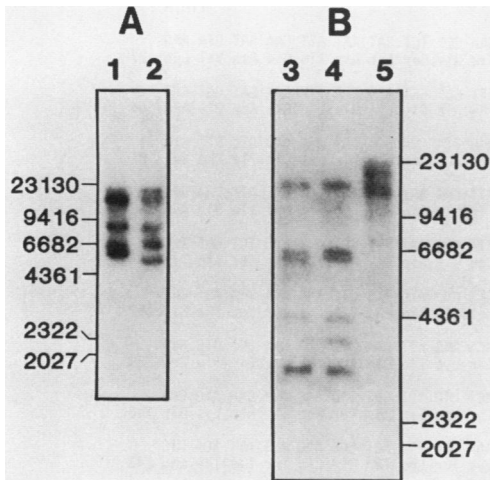


FIG. 2. Southern blot analysis of genomic DNA. *D. discoideum* DNA (5 μ g) was completely digested with *Clal* (lane 1), *Clal* plus *Bam*HI (lane 2), *Eco*RI (lane 3), *Eco*RI plus *Hind*III (lane 4), and *Hind*III (lane 5); it was then size fractionated on a 0.8% agarose-TAE gel, transferred to nitrocellulose, and hybridized to nick-translated pLK229. The filter was washed at high stringency and exposed for 17 h, as described in Materials and Methods. The positions of the relevant molecular weight markers (λ digested with *Hind*III and *phi*X174 digested with *Hae*III) are indicated.

and 375 base pairs from pAT153) hybridized to pLK229 in a Southern blot (data not shown). This fragment, denoted p229C, was sequenced. Clone λ 229, isolated from the λ gt 10 library, contained an *Eco*RI fragment of 2.8 kb corresponding to the smallest *Eco*RI band in Fig. 2.

Figures 4 and 5 give the nucleotide sequences of both genomic clones and the cDNA clone pLK229. The sequence of p229C is identical to that of the cDNA but it lacks 119 bases at the 5' terminus (Fig. 4). The coding region of p229C/pLK229 is 1,143 nt long and codes for a polypeptide of 381 amino acids. This polypeptide is composed of five tandem repeats of the same 76-amino-acid sequence plus an additional asparagine as the carboxyl-terminal amino acid.

Clone λ 229 has an open reading frame of 687 nt and codes for a polypeptide of 229 amino acids. This polypeptide contains three tandem repeats of the same 76-amino-acid sequence in addition to leucine as the terminal amino acid (Fig. 5).

A computer search for homology to known proteins, using the National Biomedical Research Foundation protein struc-

ture data base, revealed that each 76-amino-acid repeat was identical to human and bovine ubiquitin except for a substitution of glycine for proline at position 19 and asparagine for threonine at position 22 (26, 28, 35).

Although the repeats themselves differ in up to 34 of the 228 bases (Fig. 6), it is noteworthy that the amino acid sequence of each repeat is identical except for the substitution of a threonine for the alanine in position 28 of the second, third, and fifth repeat of p229C.

DISCUSSION

As is indicated by its name, ubiquitin, a 76-amino-acid protein, is found in all eucaryotic cells either free or covalently bound to cellular proteins. Its sequence is highly conserved throughout the various genera (29). The protein has three known functions. It is required for ATP-dependent intracellular protein degradation (7, 17). Ubiquitin apparently targets proteins for degradation by covalent linkage. Ubiquitin also binds to certain histones (6, 15) which are believed to be found preferentially within transcribing regions of the nucleosomes. Consequently, a role for ubiquitin in gene expression has been suggested (22). The binding of lymphocytes to and their subsequent migration through lymph node endothelium involves a ubiquitinated lymphocyte-homing receptor (30).

We previously isolated a number of cDNA clones of mRNAs developmentally regulated during *D. discoideum* spore germination (18). These mRNAs were either entirely absent or were present in very low concentrations during vegetative growth and other stages of the developmental sequence, but their levels increased dramatically in spores and during spore germination. In the present study one of these cDNAs, pLK229, was used to select genomic clones which were subsequently sequenced. We report that two pLK229-specific genomic clones code for a protein (deduced from the DNA sequence) identical to human ubiquitin, except for two amino acids. Proline in position 19 and threonine in position 22 of human ubiquitin are glycine and asparagine, respectively, in *D. discoideum* ubiquitin. As has been found for humans (35), *Xenopus* spp. (11), and *Saccharomyces cerevisiae* (26), the ubiquitin genes are generally organized as tandem repeats of the basic 76-amino-acid protein, and they seem to be part of a multigene family. Human genes have been isolated which contain nine, four, three, or only one of the 76-amino-acid repeats (35), and there are six repeats in the *Saccharomyces* gene (26). The two *D. discoideum* genes we studied contain three and five repeats, respectively.

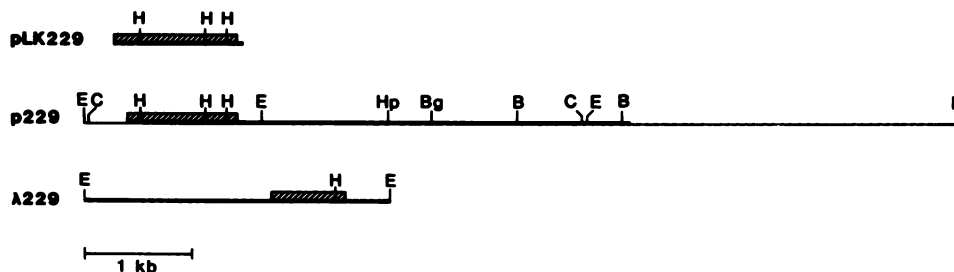


FIG. 3. Restriction map of the cDNA clone pLK229 and of the genomic clones p229 and λ 229. The entire plasmid p229 is shown, but only the inserts of pLK229 and λ 229 are given. The p229 insert is defined by the thick line, and the plasmid is defined by the thin line. pLK229 and p229 are aligned to show the identities of their restriction sites. p229C is that *Eco*RI fragment between the pAT153 *Eco*RI site (furthest left in the figure) and the first *Eco*RI site to the right of it in the *Dictyostelium* insert. The coding region is represented by the hatched box. The orientation in the sense strand is 5' left to 3' right. Restriction sites: B, *Bam*HI; Bg, *Bgl*III; C, *Clal*; E, *Eco*RI; H, *Hpa*I; Hp, *Hpa*I.

-5 TTATA

1 ATG CAA ATT TTT GTT AAA ACA CTT ACT GGA AAG ACC ATT ACT CTT GAA GTA GAA GGT TCT GAT AAT ATT GAA AAT GTA AAG
Met Gln Ile Phe Val Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu Glu Val Glu Gly Ser Asp Asn Ile Glu Asn Val Lys 27

82 GCT AAA ATT CAA GAC AAA GAA GGT ATC CCA CCA GAT CAA CAA CGT CTT ATT TTT GCT GGT AAA CAA TTG GAA GAT GGT CGT
 Ala Lys Ile Gln Asp Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu Ile Phe Ala Gly Lys Gln Leu Glu Asp Gly Arg 54

163 ACT CTC TCT GAC TAT AAC ATT CAA AAG GAA TCC ACT CTC CAT TTA GTA CTT AGA TTA AGA GGT GGT ATG CAA ATC TTT GTA
 Thr Leu Ser Asp Tyr Asn Ile Gln Lys Glu Ser Thr Leu His Leu Val Leu Arg Leu Arg Gly Gly Met Gln Ile Phe Val 81

244 AAG ACC CTC ACC GGT AAG ACT ATT ACA CTC GAA GTT GAA GGT AGT GAC AAC ATT GAG AAT GTA AAA ACA AAG ATT CAA GAT
 Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu Glu Val Glu Gly Ser Asp Asn Ile Glu Asn Val Lys Thr Lys Ile Gln Asp 108

325 AAA GAA GGT ATT CCA CCA GAT CAA CAA CGT CTC ATT TTC GCT GGT AAA CAA TTA GAA GAT GGT CGT ACT CTC TCT GAT TAC
 Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu Ile Phe Ala Gly Lys Gln Leu Glu Asp Gly Arg Thr Leu Ser Asp Tyr 135

406 AAT ATC CAA AAA GAA TCA ACT CTC CAC TTG GTC CTC AGA TTA AGA GGT GGT ATG CAA ATC TTT GTC AAA ACA TTA ACT GGT
 Asn Ile Gln Lys Glu Ser Thr Leu His Leu Val Leu Arg Leu Arg Gly Gly Met Gln Ile Phe Val Lys Thr Leu Thr Gly 162

487 AAG ACC ATC ACT CTT GAA GTT GAG GGC TCA GAT AAT ATT GAA AAT GTT AAA ACA AAG ATT CAA GAC AAA GAA GGT ATC CCA
 Lys Thr Ile Thr Leu Glu Val Glu Gly Ser Asp Asn Ile Glu Asn Val Lys Thr Lys Ile Gln Asp Lys Glu Val Glu Thr Leu 189

568 CCA GAC CAA CAA AGA TTA ATT TTC GCT GGT AAA CAA TTG GAA GAT GGT CGT ACT CTC TCT GAT TAC AAC ATT CAA AAG GAA
 Pro Asp Gln Gln Arg Leu Ile Phe Ala Gly Lys Gln Leu Glu Asp Gly Arg Thr Leu Ser Asp Tyr Asn Ile Gln Lys Glu 216

649 TCC ACT CTC CAT TTA GTC CTC AGA TTA AGA GGT GGT ATG CAA ATT TTT GTC AAA ACC CTC ACT GGT AAG ACT ATT ACA CTC
 Ser Thr Ile Thr Leu His Leu Val Leu Arg Leu Arg Gly Gly Met Gln Ile Phe Val Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu 243

730 GAA GTT GAA GGT TCT GAT AAC ATT GAG AAT GTT AAA GCC AAA ATC CAA GAC AAA GAA GGT ATT CCA CCA GAT CAA CAA CGT
 Glu Val Glu Gly Ser Asp Asn Ile Glu Asn Val Lys Ala Lys Ile Gln Asp Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg 270

811 CTC ATT TTC GCC GGT AAA CAA TTG GAG GAT GGT CGT ACT CTC TCT GAT TAC AAC ATT CAA AAG GAA TCC ACT CTC CAT TTA
 Leu Ile Phe Ala Gly Lys Gln Leu Glu Asp Gly Arg Thr Leu Ser Asp Tyr Asn Ile Gln Lys Glu Ser Thr Leu His Leu 297

892 GTT CTC AGA TTA AGA GGT GGT ATG CAA ATT TTT GTA AAA ACT CTC ACT GGT AAA ACC ATT ACA CTC GAA GTT GAA GGT AGT
 Val Leu Arg Leu Arg Gly Gly Met Gln Ile Phe Val Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu Glu Val Glu Thr Leu 324

973 GAC AAC ATT GAG AAT GTA AAA ACA AAA ATT CAA GAC AAA GAA GGT ATC CCA CCA GAT CAA CAA CGT CTC ATT TTC GCC GGT
 Asp Asn Ile Glu Asn Val Lys Thr Lys Ile Gln Asp Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu Ile Phe Ala Gly 351

1054 AAA CAA TTA GAA GAT GGT CGT ACT CTT TCT GAT TAC AAC ATT CAA AAG GAA TCA ACA CTC CAT TTA GTT CTC AGA TTA AGA
 Lys Gln Leu Glu Asp Gly Arg Thr Leu Ser Asp Tyr Asn Ile Gln Lys Glu Ser Thr Leu His Leu Val Leu Arg Leu Arg 378

1135 GGT GGA AAT TAA AATTACCCACACAAAATTTATCTTATGTGACCTATTCACAAAAAATATATATTTTCAATAACAAAATAAAAATCTTTTT
 Gly Gly Asn *** 381

1238 TTCAAATAGTAAAAAATCAATTTGTTTATTAATTTAAATTTATTATTATTTGTTTTTTTTTTTTTTTTTCCCAAGTAAATCCTTTAGTAAATTAATTTT

1345 TAATTTNNTTTTTTTTTTTTTTTTTCTCAAGGTTAATTAATTTCAAAACATATGAATTTTCGTTAATATCTGCAAATTTTCGAATCAGTTTATTTGAAAAA

1452 AAGTTAAAAAATATTATTATTATTTTTTTTTTGAATTAATATTTTTTTTTTTTTTTTTTTTATAAAAAAATCACAACACCATCATGTTACCAT

1559 AAAAAAATAAAAAATAAAAAAGAGATAAAAAATCAATAGAATTC

FIG. 4. DNA sequence of the *Dictyostelium* ubiquitin p229C/pLK229 gene. The entire sequence of the sense strand of the pLK229 insert (except the synthetic GC linkers), from positions -5 to 1195, and the entire sequence of the p229C insert, from the *Sau3A* site at position 114 to the *EcoRI* site at position 1619, are shown. Position 1 is defined as the first base of the initiation codon. A possible polyadenylation signal is underlined. The protein designation is given beneath each codon in the coding region, and the first methionine in each repeat is boxed.

In addition to the 1,143-nt-long coding region (made up of five repeats), the p229C/pLK229 gene contains at least 52 nt of 3' untranslated sequence (because this is present in the pLK229 cDNA clone). This region might be longer because a possible polyadenylation site starting at position 1223 was found in the sequence of the genomic clone p229C (Fig. 4). This would make the 3' noncoding region at least 85 nt long. Unfortunately, there is no information on the 5' untranslated region of the gene. In addition to an open reading frame of 687 nt (three repeats of the basic 228-nt unit), clone λ 229 (Fig. 5), contains 1,741 nt of 5' and 444 nt of 3' noncoding sequences. It has a typical *Dictyostelium* TATA box (TATAAATA) at position -123, the usual T stretch at -99 to -84, and CAAA and CATT boxes at positions -83 and -78, respectively, in the region 5' to the first ATG (3, 20, 21). A curious feature is a pair of perfect 10-nt-long repeats (TAATTGTATA) at positions -62 and -38. The clone also has three possible polyadenylation sites in the 3'-untranslated region immediately following the stop codon at position 692, one of which overlaps the other two. There are other possible polyadenylation sites 3' to these. There is no significant open reading frame 5' to the first coding ATG.

The last repeat unit in each gene contains one additional amino acid at the carboxyl terminus, asparagine for p229C and leucine for λ 229. This terminal amino acid varies among the ubiquitins found in other organisms and seems to be

cleaved off, because it is not found in the isolated protein (11, 26, 28, 35).

It is obvious from both RNA blot and Southern analysis that there are more *D. discoideum* ubiquitin genes than we have sequenced. There are at least six genomic fragments in Southern gels of the *Dictyostelium* genome (Fig. 2) and there are six mRNA species disclosed by RNA blot analysis (Fig. 1). If we add 100 (noncoding) nt to the 5' end and 50 (noncoding) nt to the 3' end of the basic repeat unit of 228 nt, we can conclude that the two smaller mRNA species seen in Fig. 1 (about 500 to 580 nt) contain one (or perhaps two) repeat. Likewise, the 840-nt mRNA is estimated to contain three repeats, and the 1,100-, 1,400-, and 1,900-nt species are estimated to contain four, five, and seven or eight repeats, respectively. If our calculation is correct, the 840-nt species is the mRNA for clone λ 229, and the 1,400-nt species is the mRNA for clone p229C and the cDNA clone pLK229. The fact that all the cDNA clones we isolated from a 1.5-h library are identical to pLK229 and p229 supports the conclusion that the predominant mRNA species at this time (1,400 nt) corresponds to that for the p229 gene (data not presented). This says that only one discrete-sized mRNA is transcribed from each gene and therefore the different size mRNAs seen are transcribed from genes containing different numbers of repeat units. This has yet to be proved.

What is very intriguing is the observation that various

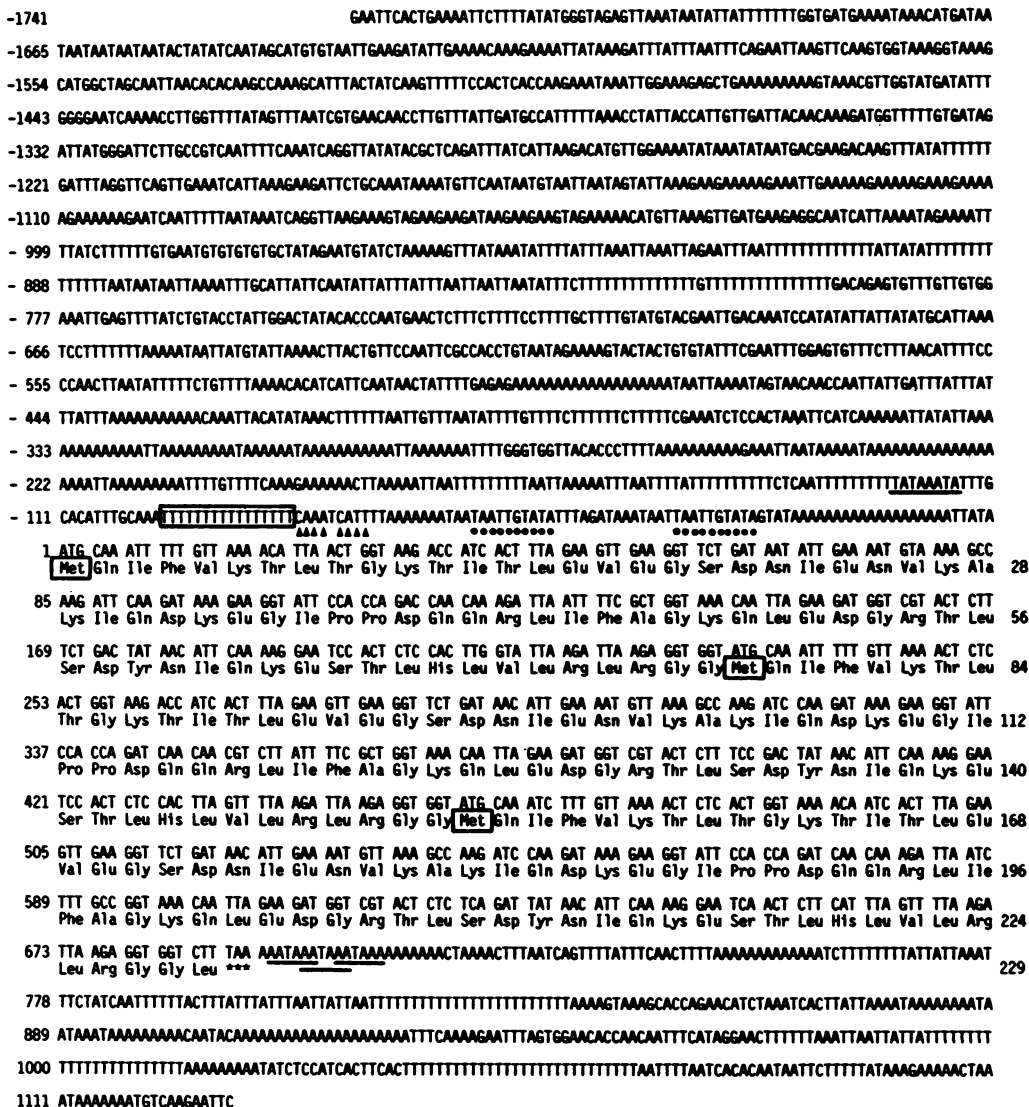


FIG. 5. DNA sequence of the *Dictyostelium* ubiquitin λ 229 gene. The entire sequence of the sense strand of the λ 229 insert is presented. Position 1 is defined as the first base of the initiation codon. The TATA box (TATAAATA) and polyadenylation signals (AATAAA) are underlined; the oligo(dT) stretch is boxed; the direct repeats in the 5' untranslated region of the gene are underlined with dots; and the CAAA and CATT boxes are underscored with triangles. The protein designation is given beneath each codon in the coding region, and the first methionine in each repeat is boxed.

ubiquitin mRNA species are developmentally expressed during *Dictyostelium* spore development (Fig. 1). Little mRNA of sizes 1,900 to 840 nt was found in growing cells. The 580- and 500-nt species were the predominant species present during this stage. Each mRNA species is differentially expressed during different stages of development. A developmental regulation of ubiquitin during early *Xenopus* development was also observed (11). In this organism ubiquitin mRNA is found in the highest concentration in unfertilized eggs and early embryos, and the level decreases subsequently.

The fact that ubiquitin mRNA is developmentally regulated (and by inference the synthesis of the protein) argues for a function for ubiquitin in development. Since ubiquitin is known to be necessary for the decay of a large number of proteins, perhaps this indicates that more protein degradation occurs during spore germination than during other

developmental stages or that ubiquitin is necessary for the degradation of specific proteins that are unique to spore germination.

Ubiquitin is known to form a covalent bond with histone H2A (22). It was suggested by these investigators that the ubiquitin-histone conjugates might modify chromosomal structure by inducing a disorganization of nucleosomes. Since actively transcribed genes have been correlated with loss in nucleosomal organization, ubiquitin might be important in the regulation of gene expression in *Dictyostelium* development.

Finally, ubiquitin was reported to be a heat shock protein in chicken embryo fibroblasts (5). We heat shocked growing *D. discoideum* cells (23), isolated the RNA, and searched for pLK229-specific mRNA by RNA blot analysis. We could find no increase in pLK229 mRNA concentration after heat shock (data not presented). In addition, we could not find a

Amino acid sequence	Met	Gln	Ile	Phe	Val	Lys	Thr	Leu	Thr	Gly	Lys	Thr	Ile	Thr	Leu	Glu	Val	Glu	Gly
λ 229 Repeat 1	ATG	CAA	ATT	TTT	GTT	AAA	ACA	TTA	ACT	GGT	AAG	ACC	ATC	ACT	TTA	GAA	GTT	GAA	GGT
2	---	---	---	---	---	---	---	--T	C-C	---	---	---	---	---	---	---	---	---	---
3	---	---	--C	---	---	---	---	--T	C-C	---	---	--A	--A	---	---	---	---	---	---
p229 C Repeat 1	---	---	---	---	---	---	---	C-T	---	--A	---	---	--T	---	C-T	---	--A	---	---
2	---	---	--C	---	--A	--G	--C	C-C	--C	---	---	--T	--T	--A	C-C	---	---	---	---
3	---	---	--C	---	--C	---	---	---	---	---	---	---	---	---	C-T	---	---	--G	--C
4	---	---	---	---	--C	---	--C	C-C	---	---	---	--T	--T	--A	C-C	---	---	---	---
5	---	---	---	---	--A	---	--T	C-C	---	---	--A	---	--T	--A	C-C	---	---	---	---
Ser Asp Asn Ile Glu Asn Val Lys Ala Lys Ile Gln Asp Lys Glu Gly Ile Pro Pro	TCT	GAT	AAT	ATT	GAA	AAT	GTA	AAA	GCC	AAG	ATT	CAA	GAT	AAA	GAA	GGT	ATT	CCA	CCA
2	---	---	--C	---	---	---	--T	---	---	---	---	--C	---	---	---	---	---	---	---
3	---	---	--C	---	---	---	--T	---	---	---	---	--C	---	---	---	---	---	---	---
1	---	---	---	---	---	---	---	--G	--T	--A	---	---	--C	---	---	---	--C	---	---
2	AG-	--C	--C	---	--G	---	---	---	Thr	A-A	---	---	---	---	---	---	---	---	---
3	--A	---	---	---	---	---	--T	---	Thr	A-A	---	---	--C	---	---	---	--C	---	---
4	---	---	--C	---	--G	---	--T	---	---	--A	--C	---	--C	---	---	---	---	---	---
5	AG-	--C	--C	---	--G	---	---	---	Thr	A-A	--A	---	--C	---	---	---	--C	---	---
Asp Gln Gln Arg Leu Ile Phe Ala Gly Lys Gln Leu Glu Asp Gly Arg Thr Leu Ser	GAC	CAA	CAA	AGA	TTA	ATT	TTC	GCT	GGT	AAA	CAA	TTA	GAA	GAT	GGT	CGT	ACT	CTT	TCT
2	--T	---	---	C-T	C-T	---	---	---	---	---	---	---	---	---	---	---	---	---	--C
3	--T	---	---	---	---	--C	--T	--C	---	---	---	---	---	---	---	---	---	---	--A
1	--T	---	---	C-T	C-T	---	--T	---	---	---	---	--G	---	---	---	---	---	---	--C
2	--T	---	---	C-T	C-C	---	---	---	---	---	---	---	---	---	---	---	---	---	--C
3	---	---	---	---	---	---	---	---	---	---	---	---	--G	---	---	---	---	---	--C
4	--T	---	---	C-T	C-C	---	---	--C	---	---	---	--G	--G	---	---	---	---	---	--C
5	--T	---	---	C-T	C-C	---	---	--C	---	---	---	---	---	---	---	---	---	---	---
Asp Tyr Asn Ile Gln Lys Glu Ser Thr Leu His Leu Val Leu Arg Leu Arg Gly Gly	GAC	TAT	AAC	ATT	CAA	AAG	GAA	TCC	ACT	CTC	CAC	TTG	GTA	TTA	AGA	TTA	AGA	GGT	GGT
2	---	---	---	---	---	---	---	---	---	---	---	---	---	--A	--T	---	---	---	---
3	--T	---	---	---	---	---	---	--A	---	--T	--T	--A	--T	---	---	---	---	---	---
1	---	---	---	---	---	---	---	---	---	--T	--A	---	C-T	---	---	---	---	---	---
2	--T	--C	--T	--C	---	--A	---	--A	---	---	---	---	--C	C-C	---	---	---	---	---
3	--T	--C	---	---	---	---	---	---	---	--T	--A	--C	C-C	---	---	---	---	---	---
4	--T	--C	---	---	---	---	---	---	---	--T	--A	--T	C-C	---	---	---	---	---	---
5	--T	--C	---	---	---	---	--A	--A	---	--T	--A	--T	C-C	---	---	---	---	---	--A

FIG. 6. Comparison of nucleotide sequences of the ubiquitin repeats of *D. discoideum* clones λ 229 and p229C. The number of the repeats refers to the order in which they appear in the genes. Only differences are given. A dashed line means that the base at that position is identical to that of the first repeat in clone λ 229. The first repeat in clone λ 229 was arbitrarily chosen as the reference sequence.

heat shock promoter in the 5' noncoding region of genomic clone λ 229, a gene which seems to be transcribed.

During revision of this report, we became aware of a recent publication (34) describing an unusual cDNA clone which contained a peptide-coding sequence linked to a portion of the 3' end of the coding region of a single ubiquitin repeat. This repeat has a different nucleotide sequence than the eight repeats we have sequenced. Westphal et al. also

showed that ubiquitin mRNA was developmentally regulated during multicellular development.

LITERATURE CITED

- Alton, T. H., and H. F. Lodish. 1977. Developmental changes in messenger RNAs and protein synthesis in *Dictyostelium discoideum*. *Dev. Biol.* 60:180-206.
- Aviv, H., and P. Leder. 1972. Purification of biologically active

- globin messenger RNA by chromatography on oligothymidylic acid-cellulose. *Proc. Natl. Acad. Sci. USA* **69**:1408-1412.
3. **Barklis, E., B. Pontius, K. Barfield, and H. F. Lodish.** 1985. Structure of the promoter of the *Dictyostelium discoideum* prespore EB4 gene. *Mol. Cell. Biol.* **5**:1465-1472.
 4. **Blumberg, D. D., and H. F. Lodish.** 1980. Changes in the messenger RNA population during differentiation of *Dictyostelium discoideum*. *Dev. Biol.* **78**:285-300.
 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. **Busch, H., and I. L. Goldknopf.** 1981. Ubiquitin-protein conjugates. *Mol. Cell. Biochem.* **40**:173-187.
 7. **Ciechanover, A., D. Finley, and A. Varshavsky.** 1984. The ubiquitin-mediated proteolytic pathway and mechanisms of energy-dependent intracellular protein degradation. *J. Cell. Biochem.* **24**:27-53.
 8. **Clewell, D. B., and D. R. Helinski.** 1969. Supercoiled circular DNA-protein complex in *Escherichia coli*: purification and induced conversion to an open circular DNA form. *Proc. Natl. Acad. Sci. USA* **62**:1159-1166.
 9. **Cotter, D., and K. B. Raper.** 1966. Spore germination in *Dictyostelium discoideum*. *Proc. Natl. Acad. Sci. USA* **56**:880-887.
 10. **Dowbenko, D. J., and H. L. Ennis.** 1980. Regulation of protein synthesis during spore germination in *Dictyostelium discoideum*. *Proc. Natl. Acad. Sci. USA* **77**:1791-1795.
 11. **Dworkin-Rasti, 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.
 12. **Ennis, H. L., and M. Sussman.** 1975. Mutants of *Dictyostelium discoideum* defective in spore germination. *J. Bacteriol.* **124**:62-64.
 13. **Giri, J. G., and H. L. Ennis.** 1978. Developmental changes in RNA and protein synthesis during germination of *Dictyostelium discoideum* spores. *Dev. Biol.* **67**:189-201.
 14. **Giri, J. G., and H. L. Ennis.** 1979. Protein and RNA synthesis during spore germination in the cellular slime mold *Dictyostelium discoideum*. *Biochem. Biophys. Res. Commun.* **77**:282-289.
 15. **Goldknopf, I. L., and H. Busch.** 1977. Isopeptide linkage between nonhistone and histone 2A polypeptides of chromosomal conjugate-protein A24. *Proc. Natl. Acad. Sci. USA* **74**:864-868.
 16. **Henikoff, S.** 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* **28**:351-359.
 17. **Hershko, A., and A. Ciechanover.** 1982. Mechanisms of intracellular protein breakdown. *Annu. Rev. Biochem.* **51**:335-364.
 18. **Kelly, L. J., R. Kelly, and H. L. Ennis.** 1983. Characterization of cDNA clones specific for sequences developmentally regulated during *Dictyostelium discoideum* spore germination. *Mol. Cell. Biol.* **3**:1943-1948.
 19. **Kelly, R., L. J. Kelly, and H. L. Ennis.** 1985. *Dictyostelium discoideum* mRNAs developmentally regulated during spore germination have short half-lives. *Mol. Cell. Biol.* **5**:133-139.
 20. **Kimmel, A., and R. Firtel.** 1982. The organization and expression of the *Dictyostelium* genome, p. 233-324. In W. F. Loomis (ed), *The development of Dictyostelium discoideum*. Academic Press, Inc., New York.
 21. **Kimmel, A., and R. Firtel.** 1983. Sequence organization in *Dictyostelium*: unique structure at the 5'-ends of protein coding genes. *Nucleic Acids Res.* **11**:541-552.
 22. **Levinger, L., and A. Varshavsky.** 1982. Selective arrangement of ubiquitinated and D1 protein-containing nucleosomes within the *Drosophila* genome. *Cell* **28**:375-385.
 23. **Loomis, W. F., and S. Wheeler.** 1980. Heat shock response of *Dictyostelium*. *Dev. Biol.* **79**:399-408.
 24. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
 25. **Norrander, J., T. Kempe, and J. Messing.** 1983. Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. *Gene* **26**:101-106.
 26. **Ozkaynak, E., D. Finley, and A. Varshavsky.** 1984. The yeast ubiquitin gene: head-to-tail repeats encoding a polyubiquitin precursor protein. *Nature (London)* **312**:663-666.
 27. **Sanger, F., N. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463-5467.
 28. **Schlesinger, D. H., G. Goldstein, and H. D. Niall.** 1975. The complete amino acid sequence of ubiquitin, an adenylate cyclase stimulating polypeptide probably universal in living cells. *Biochemistry* **14**:2214-2218.
 29. **Southern, E.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**:503-517.
 30. **St. John, T., W. M. Gallatin, M. Siegelman, H. T. Smith, V. A. Fried, and I. L. Weissman.** 1986. Expression cloning of a lymphocyte homing receptor cDNA: ubiquitin is the reactive species. *Science* **231**:845-850.
 31. **Sussman, M., and R. Sussman.** 1969. Patterns of RNA synthesis and of enzyme accumulation and disappearance during cellular slime mould cytodifferentiation, p. 403-435. In *Symposia of the Society for General Microbiology, number XIX, Microbial growth*. Society for General Microbiology, Cambridge University Press, Cambridge.
 32. **Sussman, R., and M. Sussman.** 1967. Cultivation of *Dictyostelium discoideum* in axenic medium. *Biochem. Biophys. Res. Commun.* **29**:53-55.
 33. **Twigg, A. J., and D. Sherratt.** 1980. *trans*-complementable copy-number mutants of plasmid ColEI. *Nature (London)* **283**:216-218.
 34. **Westphal, M., A. Müller-Taubenberger, A. Noegel, and G. Gerisch.** 1986. Transcript regulation and carboxyterminal extension of ubiquitin in *Dictyostelium discoideum*. *FEBS Lett.* **209**:92-96.
 35. **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.