The EMBO Journal vol.4 no.3 pp.755-759, 1985

The human ubiquitin multigene family: some genes contain multiple directly repeated ubiquitin coding sequences

Ove Wiborg, Marianne S.Pedersen, Anette Wind[1], Lars E.Berglund, Kjeld A.Marcker and Jens Vuust

Department of Molecular Biology and Plant Physiology, University of Aarhus, DK-8000 Aarhus C, Denmark

[1]Present address: Statens Seruminstitut, Amager Boulevard 80, DK-2300 Copenhagen S, Denmark

Communicated by K.Marcker

Ubiquitin coding sequences were isolated from a human genomic library and two cDNA libraries. One human ubiquitin gene consists of 2055 nucleotides and codes for a polyprotein consisting of 685 amino acid residues. The polyprotein contains nine direct repeats of the ubiquitin amino acid sequence and the last ubiquitin sequence is extended with an additional valyl residue at the C-terminal end. No spacer sequences separate the ubiquitin repeats and the coding regions are not interrupted by intervening sequences. This particular gene is transcribed since cDNAs corresponding to the genomic sequence have been isolated. At least two more types of ubiquitin genes are encoded in the human genome, one coding for an ubiquitin monomer while another presumably codes for three or four direct repeats of the ubiquitin sequence. Human DNA contains many copies of the ubiquitin sequence. Ubiquitin is therefore encoded in the human genome as a multigene family.

Key words: ubiquitin genes/direct DNA repeats/polyproteins/multigene family

Introduction

Ubiquitin is a small protein consisting of 76 amino acids. It was originally identified in calf thymus, but has subsequently been found in all eukaryotic cells investigated (see reviews by Hershko and Ciechanover, 1982; Hershko, 1983; Ciechanover *et al.*, 1984). Ubiquitin is one of the most conserved proteins known, since the amino acid sequence is identical from insects to human. Ubiquitin is required for ATP-dependent, non-lysosomal intracellular protein degradation, which eliminates the majority of intracellular defective proteins as well as normal proteins with a rapid turnover. The mechanism of this particular pathway of protein degradation involves covalent binding of ubiquitin to the proteins to be degraded, through isopeptide bonds from the C-terminal Gly in ubiquitin to ε-amino groups of lysyl side chains. Presumably, the function of ubiquitin is to label the protein as a substrate for intracellular proteases. The most abundant ubiquitin-protein conjugate, however, is ubiquitin-H2A, in which ubiquitin is bound to Lys 119 in histone H2A (Busch and Goldknopf, 1981). This conjugate is not degraded. Experimental evidence suggests that such ubiquitinated histones are present primarily in actively transcribed chromosomal regions. Ubiquitin may, therefore, play a role in regulation of gene expression (Levinger and Varshavsky, 1982).

Recently, we have screened a cDNA library, constructed from porcine brain RNA, for sequences coding for small brain peptides. A number of overlapping cDNA clones thus isolated represent almost the entire coding sequence for ubiquitin lacking only the part coding for the five C-terminal amino acids. These probes enabled us to isolate and analyze other clones containing ubiquitin-specific sequences from a human genomic library and two human cDNA libraries. We report here that human ubiquitin genes constitute a multigene family and that some of these genes have an unusual DNA sequence organization.

Results

Construction and cloning of porcine ubiquitin cDNA

Plasmid pPUb29 is one of several recombinant plasmids containing ubiquitin coding sequences that were isolated from a porcine brain cDNA library (Wind and Vuust, unpublished). This library was obtained by reverse transcription of poly(A)-containing RNA from porcine brain, primed by the oligodeoxynucleotide mixture d($\frac{A}{G}$-A-A-$\frac{A}{G}$-T-C-C-A-T-C-C-A) (Boel *et al.*,

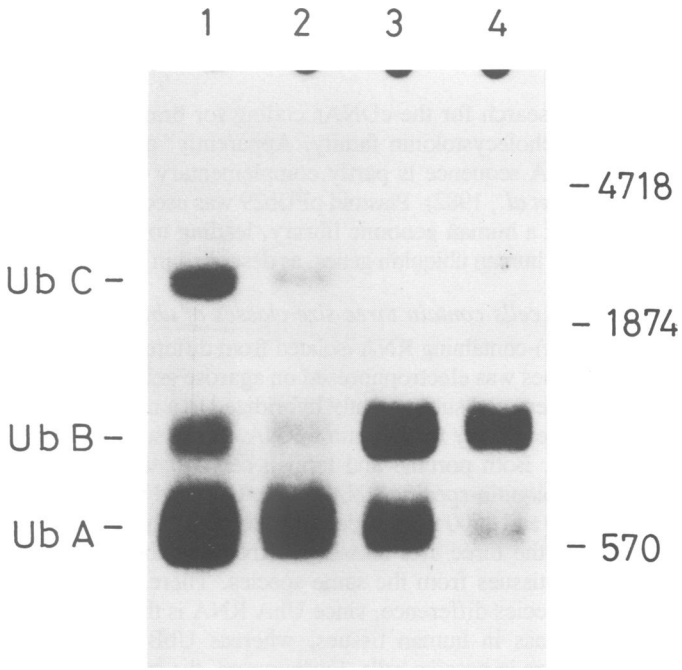

Fig. 1. Northern blotting analysis of poly(A)-containing RNA from human intestinal mucosa **(lane 1)**, human gastrinoma (Boel *et al.*, 1983) **(lane 2)**, porcine intestinal mucosa **(lane 3)** and porcine brain **(lane 4)**. **Lanes 1, 3** and **4** were loaded with ~15 μg of RNA, **lane 2** with ~5 μg of RNA. Hybridization was with the nick-translated *Pvu*II-*Hind*III fragment of λHUb1; hybridization with the 684-bp *Xho*I fragment of pHUb14-38 gave the same result. Autoradiography was for 4 days. Mol. wt. markers are 28S (4718 nucleotides, Chan *et al.*, 1983) and 18S (1874 nucleotides, Chan *et al.*, 1984) rat rRNA, and human gastrin mRNA (570 nucleotides; Wiborg *et al.*, 1984). The amount of porcine UbC RNA is much reduced in comparison with that of human RNA, and is not visible on the photographic reproduction. Longer exposure times revealed that porcine UbC is slightly larger than human UbC.

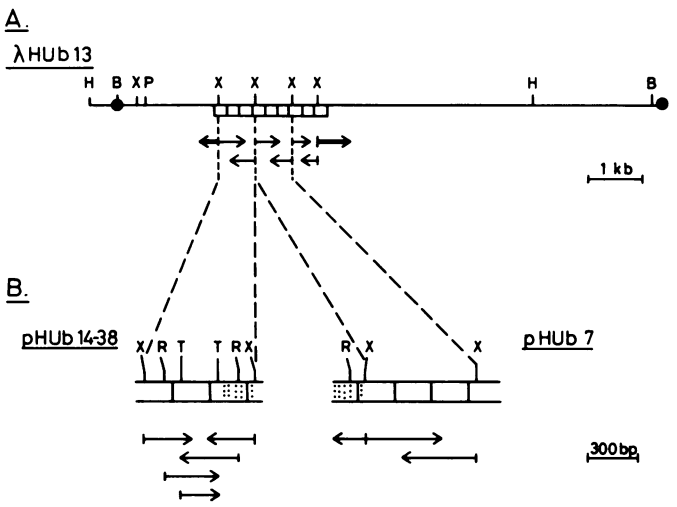

Fig. 2. (A) Physical map of the DNA fragment present in clone λHUb13. The boxed region indicates the coding region of the ubiquitin gene; vertical lines separate the nine direct ubiquitin repeats. The 5′ end of the gene is to the left. Filled circles indicate the ends of the insert. **(B)** Physical maps of the human ubiquitin cDNA inserts in clones pHUb14-38 and pHUb7. Broken lines indicate the corresponding regions in the genomic clone. Dotted areas show the overlapping cDNA sequences, and vertical lines separate the ubiquitin repeats. The strategy for determining the nucleotide sequence is shown by horizontal arrows, indicating the extent and direction of each sequence reading. Thin arrows indicate sequences determined by the dideoxy technique for sequence determination (Sanger *et al.*, 1980); thick arrows indicate sequences determined by the chemical degradation procedure (Maxam and Gilbert, 1977). B, *Bam*HI; H, *Hind*III; P, *Pst*I; R, *Rsa*I. T, *Taq*I; X, *Xho*I.

1983), in a search for the cDNAs coding for brain peptides of the gastrin-cholecystokinin family. Apparently, a porcine ubiquitin mRNA sequence is partly complementary to the primer (Panabiéres *et al.*, 1982). Plasmid pPUb29 was used for the initial screening of a human genomic library, leading to characterization of some human ubiquitin genes, as described in the following.

Mammalian cells contain three size classes of ubiquitin mRNA

Total poly(A)-containing RNA isolated from different human and porcine tissues was electrophoresed on agarose gels, transferred to GeneScreen, and subsequently hybridized to a nick-translated DNA complementary to ubiquitin mRNA. The results are shown in Figure 1. Both porcine and human cells contain three size classes of ubiquitin-specific RNA (UbA, UbB and UbC) that are ~650, 1100 and 2500 nucleotides long, respectively. The relative amounts of the three size classes of RNA are about the same in different tissues from the same species. There is, however, a striking species difference, since UbA RNA is the most abundant size class in human tissues, whereas UbB RNA is the prevalent form in porcine cells. Furthermore, the quantity of porcine UbC RNA is much reduced in comparison with that of human UbC RNA. It appears that the content of ubiquitin mRNA in porcine cells is somewhat higher in intestinal mucosa than in cerebral cortex (Figure 1).

One human ubiquitin gene contains nine directly repeated ubiquitin coding sequences

To analyze the ubiquitin gene structure, a human genomic library was screened with a ubiquitin-specific cDNA probe. In addition, human ubiquitin cDNA clones were isolated from two human cDNA libraries, constructed from RNA extracted from intestinal mucosa and liver, respectively. In this way, 20 different genomic and 40 cDNA clones containing ubiquitin-specific sequences were isolated. Subsequent restriction endonuclease and Southern blotting analyses revealed that several different genes and gene transcripts were represented in the isolated clones. However, two cDNA clones (pHUb14-38 from human intestinal RNA and pHUb7 from human liver RNA) and the genomic clone λHUb13 yielded fragments of the same size after digestion with restriction endonuclease *Xho*I. Thus, pHUb14-38 and pHUb7 most likely represent transcripts corresponding to the ubiquitin sequence present in clone λHUb13, and consequently these clones were selected for sequencing. Figure 2 shows the overall structure and a partial restriction endonuclease map of these clones, as well as the sequencing strategy employed. Figure 3 shows the nucleotide sequences of the coding region and the non-coding parts of the ubiquitin gene present in λHUb13. The nucleotide sequences of the two cDNA clones are overlapping and are in accordance with the corresponding sequences present in clone λHUb13.

The coding region of the human ubiquitin gene in λHUb13 is 2055 nucleotides long. It codes for a polypeptide consisting of 685 amino acid residues. This polypeptide contains nine direct repeats of the ubiquitin amino acid sequence. No spacer sequences separate the ubiquitin repeats, but for each of the first eight ubiquitin sequences, Gly-76 is followed immediately by Met-1 of the next ubiquitin sequence. Thus, the primary translation product is a poly-ubiquitin. Presumably, the Met-1 of the N-terminal ubiquitin repeat is the initiating Met since all ATG codons present further upstream are followed by stop codons. There is a potential cap site 44 bp upstream from the initiating ATG, and a possible 'TATA' box (AATATGTAA) 30 bp upstream from the putative cap site (Figure 3). There are no intervening sequences in the coding region but the presence of an intervening sequence in the 5′ non-coding region cannot be ruled out since cDNAs extending beyond the initiating ATG codon have thus far not been obtained. The C-terminal ubiquitin repeat in the primary translation product contains an additional valyl residue at the C-terminal end, and a putative polyadenylation signal (Proudfoot and Brownlee, 1976) is present 47 bp downstream from the stop codon.

Because of the highly repetitive nature of this gene, special precautions were necessary to establish the number and order of restriction endonuclease fragments. The difficulties encountered are exemplified by the presence of two 684-bp long *Xho*I fragments, each corresponding to three ubiquitin repeats, and one 456-bp *Xho*I fragment, representing two ubiquitin repeats. The sequences of the cDNA clones pHUb14-38 and pHUb7 established the order of the two 684-bp fragments. To ascertain the number and order of all the *Xho*I fragments in the genomic clone, an analysis using partial restriction endonuclease digestion was carried out as follows: the ~8-kb *Hind*III fragment, containing the entire ubiquitin gene from λHUb13 (Figure 2) was subcloned in the *Hind*III site of pBR322 in such a way that the 5′ end of the ubiquitin-coding region of the insert was adjacent to the *Eco*RI site of pBR322. This recombinant plasmid was digested with *Eco*RI, end-labelled, and subjected to partial digestion with restriction endonuclease *Xho*I. The number and sizes of radioactive fragments thus generated established the number and order of all the *Xho*I fragments of the gene.

Transcription of the λHUb13 ubiquitin gene will yield a RNA ~2500 nucleotides long which is in accordance with the size of the largest ubiquitin RNA found in Northern blotting analyses (UbC, Figure 1).

Some ubiquitin genes code for an ubiquitin monomer

Two other genomic clones, λHUb1 and λHUb6 were analyzed

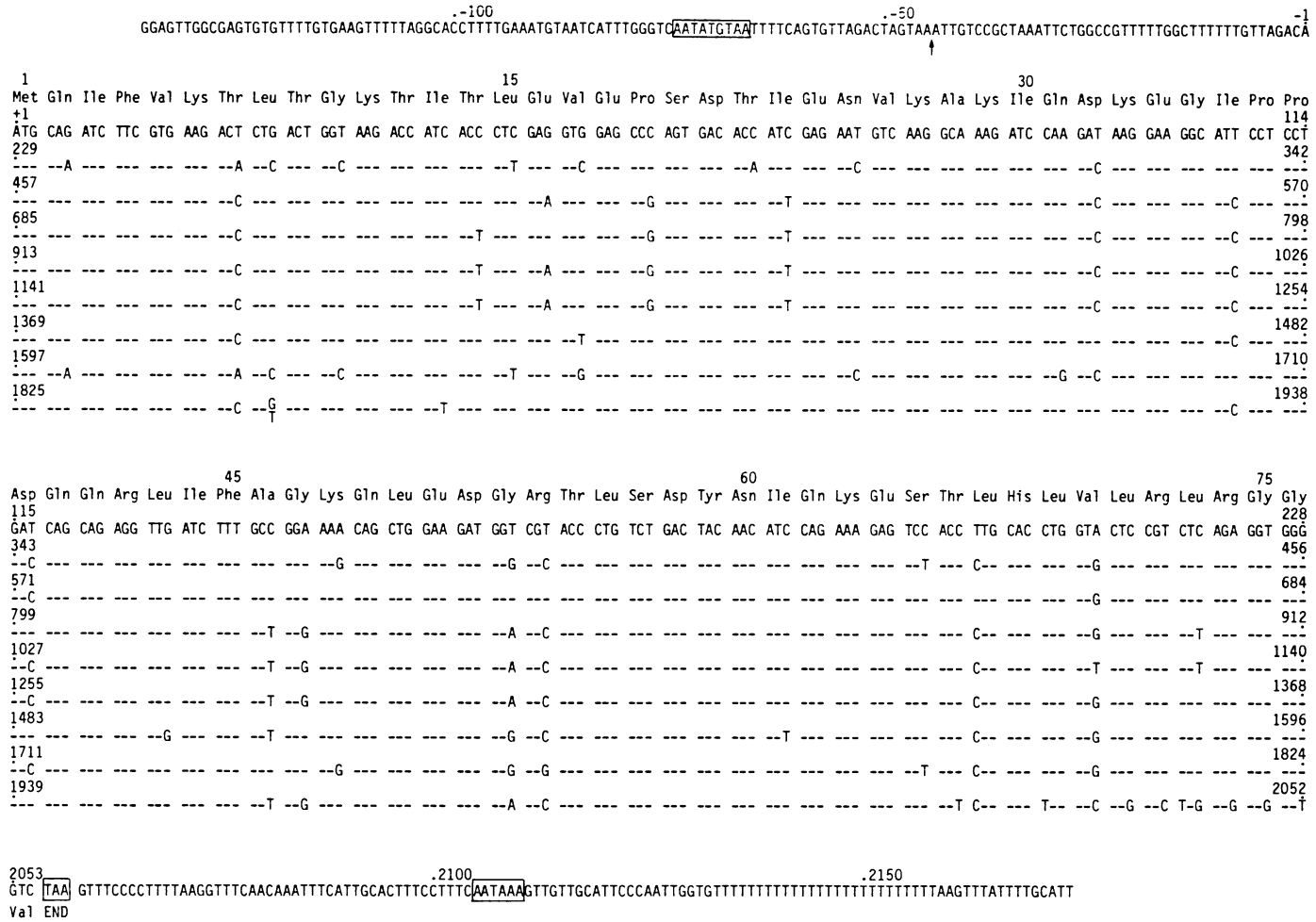

Fig. 3. The nucleotide sequence of the coding region and part of the non-coding 5′ and 3′ regions of the human ubiquitin gene present in the clone λHUb13. The positions are numbered from the A of the presumed initiating methionine codon. Negative numbers are used in the 5′ non-coding region. The nucleotide sequence of the 5′ ubiquitin-coding direct repeat is given in full; in the following repeats, nucleotide identity with the 5′ repeat is indicated by hyphens. The ubiquitin amino acid sequence is shown above the 5′ repeat. The putative TATA box, the presumed polyadenylation signal, and the stop codon following the 3′ additional valine codon are boxed. The presumed cap site is indicated by an arrow.

Fig. 4. Nucleotide sequence comparison of the non-coding 5′ and 3′ regions of the human ubiquitin genes present in clones λHUb13 (also shown in Figure 3.), λHUb1 and λHUb6. Each of the genes in λHUb1 and λHUb6 contains an ubiquitin monomer coding sequence (not shown). In the 5′ regions, positions are numbered as in Figure 3. In the 3′ regions, numbering is from the T of the stop codon. For λHUb6, no further reliable sequence data are presently available. Differences between λHUb1 and λHUb6 are indicated by stars. In the 3′ end of λHUb6, three insertions (dots) and one deletion (C with arrow) were introduced for maximum homology. The putative TATA boxes and cap sites are boxed and indicated by arrows, respectively. The presumed polyadenylation signal of λHUb13 is boxed. The *Hind*III recognition site of λHUb1, used for hybridization probe preparation (see Materials and methods) is indicated and underlined.

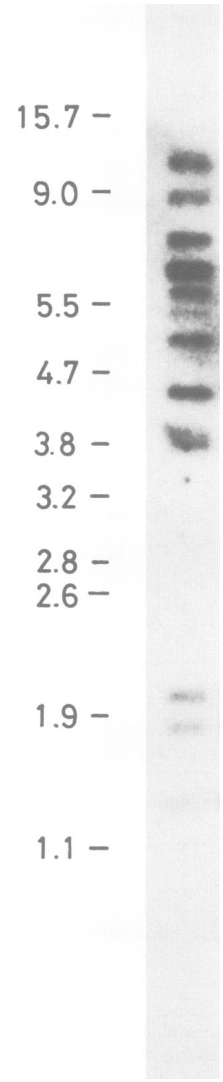

Fig. 5. Southern blotting analysis of human chromosomal DNA, digested with restriction endonuclease *Hind*III. Hybridization was with the nick-translated *Pvu*II-*Hind*III fragment of λHUb1. Autoradiography was for 7 days. An identical result was obtained with a DNA preparation from a different individual. Size markers were fragments from wild-type phage λ DNA digested with *Eco*RI and *Bam*HI; their molecular sizes are given in kbp.

in detail. They both contain only one ubiquitin sequence. Interestingly, their 5′ and 3′ non-coding regions, although mutually homologous, are completely different from the corresponding regions of the large gene in λHUb13 (Figure 4). The potential cap site and TATA box of λHUb1 are more in agreement with the consensus sequences for these sites (Breathnach and Chambon, 1981) than are the equivalent sites in λHUb13. Presumably, the 650 nucleotide long RNA (UbA, Figure 1) represents transcription products of genes similar to the ones present in λHUb1 and λHUb6. The gene in λHUb1 is most likely a pseudogene, since it contains a deletion near the 3′ end of the coding region. A gene corresponding to the 1100 nucleotide RNA (UbB, Figure 1) has not been isolated with certainty so far.

The human ubiquitin genes constitute a multigene family

Human chromosomal DNA from two individuals were digested with *Hind*III, electrophoresed in an agarose gel, blotted onto nitrocellulose and hybridized with a ubiquitin-specific probe (Figure 5). In both preparations, at least 12 hybridizing fragments of varying intensity are visible. Since no *Hind*III sites have been found in any of the ubiquitin sequences analyzed so far, these data suggest that there are at least 12 different DNA fragments containing ubiquitin genes, each composed of either one or multiple ubiquitin coding sequences. This result is also in agreement with the large number of different genomic and cDNA clones isolated. The exact number of ubiquitin coding sequences present in the human genome has not been determined. However, it is clear that there are many such copies and the human ubiquitin genes therefore constitute a multigene family. At present it is not known whether these genes are clustered or whether they are dispersed throughout the genome.

Discussion

Human and presumably other mammalian genomes contain a ubiquitin multigene family with a highly unusual DNA sequence organization. Some of the genes in this family code for ubiquitin polyproteins, containing nine direct repeats of ubiquitin. Others may code for three or four ubiquitin repeats, while some genes certainly code for only one ubiquitin sequence. There are several precedents of polyprotein precursors containing a number of identical or very similar polypeptide sequences from which several active products are liberated by proteolytic processing. These examples include the corticotropin-β-lipotropin precursor (Nakanishi *et al.*, 1979), the preproenkephalins (Noda *et al.*, 1982; Guebler *et al.*, 1982; Comb *et al.*, 1982), the yeast α-factor precursor (Kurjan and Herskowitz, 1982), the glucagon precursor (Lopez *et al.*, 1983) and the caerulein precursors (Hoffmann *et al.*, 1983). In all these cases, the coding sequences are separated by spacer sequences of varying lengths. In contrast, the human ubiquitin precursor deduced from the structure of the genomic DNA and cDNA shown in Figures 2 and 3 is unique, since it contains nine exact, direct repeats of the ubiquitin polypeptide sequence. These ubiquitin repeats are not separated by any spacer sequences but are immediately adjacent to one another. Another unusual feature of the ubiquitin genes is the lack of intervening sequences in the coding regions, similar to the histone (Kedes, 1979) and interferon genes (Nagata *et al.*, 1980).

Each ubiquitin-coding repeat of the poly-ubiquitin gene in λHUb13 has an unique nucleotide sequence. Strikingly, none of the several base changes results in amino acid substitutions. Two amino acid substitutions are present in one of the ubiquitin monomer genes (λHUb1), but this gene is most likely a pseudogene. It is possible that some of the many ubiquitin genes present in the human genome contain amino acid substitutions. This would imply that ubiquitin is in reality a mixture of very homologous proteins. However, the complete conservation of amino acid sequence observed in the functional genes sequenced so far is in accordance with the fact that ubiquitin is completely conserved through evolution. Thus, a very strong selection pressure must operate in order to maintain the conservation of the amino acid sequence of ubiquitin.

Presumably, ubiquitin polyproteins are processed to ubiquitin monomers by a proteolytic enzyme recognizing the Gly-Met bond linking the ubiquitin repeating units, as well as the Gly-Val bond of the C-terminal repeat. It is not known whether ubiquitin polymers are required for some purposes other than merely to act as precursors for ubiquitin monomer.

Ubiquitin isolated from calf thymus consists of only 74 amino acid residues, lacking the two COOH-terminal Gly residues. It has been a matter of speculation whether this ubiquitin derivative is an *in vitro* proteolytic artefact, or whether it occurs *in vivo*

(Ciechanover *et al.*, 1984). The possibility that the two Gly residues are added post-translationally is ruled out by the finding that they are encoded in the genomic sequences. The results of the Northern blotting analysis of RNAs from different tissues suggest that cerebral cortex contains less ubiquitin mRNA than does intestinal mucosa. This may be due to a lower rate of cell division in the brain tissue leading to a relatively low requirement for ubiquitin for binding to histone H2A.

Materials and methods

Enzymes and reagents

Restriction endonucleases were from New England Biolabs, Amersham, or Boehringer. Reverse transcriptase was purchased from Life Sciences, St. Petersburg, FL. DNA polymerase I 'Klenow fragment' was obtained from P-L Biochemicals. [α-^{32}P]dATP (~3000 Ci/mmol) was from Amersham.

Preparation of ubiquitin-specific hybridization probes

Plasmid pPUb29, containing a porcine ubiquitin cDNA insert (see Results), was used for screening a human genomic library, as described below. Following initial characterization of the human genomic and cDNA clones, two fragments from these were used as probes for colony and plaque hybridization as well as for Southern and Northern blotting analyses. One was a *Pvu*II-*Hind*III fragment from λHUb1, subcloned in pBR322; this 210-bp fragment contains 147-bp of the ubiquitin coding sequence and 63-bp of the 5′ non-coding region of the presumed ubiquitin pseudogene. The 684-bp *Xho*I fragment of pHUb14-38 was also used. The fragments were excised, purified and self-ligated as described (Marcker *et al.*, 1984) prior to labelling by nick-translation (Maniatis *et al.*, 1975).

Preparation of RNA

RNA was prepared by a modification of the procedure of Chirgwin *et al.* (1979), as previously described (Boel *et al.*, 1983). Poly(A)-containing RNA was obtained by chromatography on oligo(dT)-cellulose (Aviv and Leder, 1972).

Northern blotting analysis

RNA was separated by electrophoresis in 1.5% agarose gels containing 2.2 M formaldehyde (Maniatis *et al.*, 1982). Subsequent transfer to GeneScreen (New England Nuclear) as well as hybridization and washing procedures followed the protocol of Marcker *et al.* (1984).

Isolation of human genomic ubiquitin DNA

A human genomic library was kindly donated by Dr J.P.Hjorth. It was prepared from a limited *Sau*3A digest of adult human leukocyte DNA that was cloned in the *Bam*HI sites of the bacteriophage λL47.1 (Loenen and Brammar, 1980). Screening of 10^6 phages was carried out essentially according to Benton and Davis (1977), using either pPUb29 or a fragment from λHUb1 as a probe (see above). Phage DNA from positive plaques was characterized by Southern blotting analysis, following digestion with appropriate restriction endonucleases.

Construction and cloning of double-stranded human ubiquitin cDNA

Poly(A)-containing RNA was prepared from human intestinal mucosa from colon sigmoideum, surgically removed from a patient suffering from intestinal obstruction due to chronic diverticulitis. Oligo(dT)-primed cDNA was synthesized and cloned in the *Pst*I site of pBR322 as described previously (Boel *et al.*, 1983), using *Escherichia coli* K803 (Wood, 1966) as a host. Bacterial colonies containing recombinant plasmids were screened with a ubiquitin-specific DNA probe after transfer to Colony-Plaque Screen (New England Nuclear). Subsequent treatment of the filters as well as hybridization and washing procedures were according to the manufacturers' recommendations. A human liver cDNA library (Woods *et al.*, 1982) was screened in the same way in collaboration with Dr R.Cortese, EMBL, Heidelberg.

DNA sequence analysis

The dideoxy chain termination method of Sanger *et al.* (1980) was used as described (Wiborg *et al.*, 1982). In a few cases, restriction fragments were 3′ end-labelled with [α-^{32}P]dATP and DNA polymerase I, and the DNA sequence was determined by the chemical cleavage procedure of Maxam and Gilbert (1977).

DNA preparation for genomic Southern blotting

Freshly drawn human blood, 50 ml, was made 2 mM in EDTA, and the cells were lysed by mixing with 200 ml of a 10 mM Tris-HCl buffer (pH 8) containing 5 mM $MgCl_2$, 320 mM sucrose, 1% Triton X-100 and 2 mM EDTA. Nuclei were pelleted by centrifugation (3000 *g*, 15 min), and lysis of nuclei was achieved by suspension in 9.5 ml of a buffer containing 24 mM EDTA and 75 mM NaCl. Proteinase K (final concentration 2 μg/ml) and 1% SDS was added, and the volume adjusted to 10 ml. The mixture was incubated overnight at 4°C, then extracted once with phenol and once with chloroform. One half volume of 7.5 M ammonium acetate was added, and DNA was precipitated by addition of one and a half volumes of isopropanol. Aliquots of 5 μg were digested with the appropriate restriction endonuclease and the resulting fragments analyzed by Southern blotting analysis (Southern, 1975).

Acknowledgements

We thank Marianne Nielsen and Ole Nymann for skilful technical assistance and Birthe Hother Nielsen for typing the manuscript. We are also grateful to Dr R. Cortese for collaboration in the screening of one of the cDNA libraries, and to Dr N.P.Fiil, Dr E.Boel, Dr J.P.Hjorth and Dr S.E.Petersen for their help and advice during part of this work. This work was supported by grants from NOVO Industri A/S, from P.Carl Petersens Fond, from Arvid Nilsson's Fond and from the Danish Medical and Natural Science Research Councils.

References

Aviv,H. and Leder,P. (1972) *Proc. Natl. Acad. Sci. USA*, **69**, 1408-1412.
Benton,W.D. and Davis,R.W. (1977) *Science (Wash.)*, **196**, 180-182.
Boel,E., Vuust,J., Norris,F., Norris,K., Wind,A., Rehfeld,J.E. and Marcker, K.A. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 2866-2869.
Breathnach,R. and Chambon,P. (1981) *Annu. Rev. Biochem.*, **50**, 349-383.
Busch,H. and Goldknopf,I.L. (1981) *Mol. Cell Biochem.*, **40**, 173-187.
Chan,Y.-L., Olvera,J. and Wool,I.G. (1983) *Nucleic Acids Res.*, **11**, 7819-7831.
Chan,Y.-L., Gutell,R., Noller,H.F. and Wool,I.G. (1984) *J. Biol. Chem.*, **259**, 224-230.
Chirgwin,J.M., Przybyla,A.E., McDonald,R.J. and Rutter,W.J. (1979) *Biochemistry (Wash.)*, **18**, 5294-5299.
Ciechanover,A., Finley,D. and Varshavsky,A. (1984) *J. Cell Biochem.*, **24**, 27-53.
Comb,M., Seeburg,P.H., Adelman,J., Eiden,L. and Herbert,E. (1982) *Nature*, **295**, 663-666.
Guebler,U., Seeburg,P., Hoffman,B.J., Gage,L.P. and Udenfriend,S. (1982) *Nature*, **295**, 206-208.
Hershko,A. (1983) *Cell*, **34**, 11-12.
Hershko,A. and Ciechanover,A. (1982) *Annu. Rev. Biochem.*, **51**, 335-364.
Hoffmann,W., Bach,T.C., Seliger,H. and Kreil,G. (1983) *EMBO J.*, **2**, 111-114.
Kedes,L.H. (1979) *Annu. Rev. Biochem.*, **48**, 837-870.
Kurjan,J. and Herskowitz,I. (1982) *Cell*, **30**, 933-943.
Levinger,L. and Varshavsky,A. (1982) *Cell*, **28**, 375-385.
Loenen,W.A.M. and Brammar,W.J. (1980) *Gene*, **10**, 249-259.
Lopez,L.C., Frazier,M.L., Su,C.-J., Kumar,A. and Saunders,G.F. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 5485-5489.
Maniatis,T., Jeffrey,A. and Kleid,D. (1975) *Proc. Natl. Acad. Sci. USA*, **72**, 1184-1188.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning, A Laboratory Manual*, published by Cold Spring Harbor Laboratory Press, NY.
Marcker,A., Lund,M., Jensen,E.Ø. and Marcker,K.A. (1984) *EMBO J.*, **3**, 1691-1695.
Maxam,A. and Gilbert,W. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 560-564.
Nagata,S., Mantei,N. and Weissmann,C. (1980) *Nature*, **287**, 401-408.
Nakanishi,S., Inoue,A., Kita,T., Nakamura,M., Chang,A.C.Y., Cohen,S.N. and Numa,S. (1979) *Nature*, **278**, 423-427.
Noda,M., Furutani,Y., Takahashi,H., Toyosato,M., Hirose,T., Inayami,S., Nakanishi,S. and Numa,S. (1982) *Nature*, **295**, 202-206.
Panabières,F., Fort,P., Piechaczyk,M., Blanchard,J.M. and Jeanteur,P. (1982) *Gene*, **19**, 321-326.
Proudfoot,N.J. and Brownlee,G.G. (1976) *Nature*, **263**, 211-214.
Sanger,F., Coulson,A.R., Barrell,B.G., Smith,A.J.H. and Roe,B.A. (1980) *J. Mol. Biol.*, **143**, 161-178.
Southern,E.M. (1975) *J. Mol. Biol.*, **98**, 503-517.
Wiborg,O., Hyldig-Nielsen,J.J., Jensen,E.Ø., Paludan,K. and Marcker,K.A. (1982) *Nucleic Acids Res.*, **10**, 3483-3494.
Wiborg,O., Berglund,L., Boel,E., Norris,F., Norris,K., Rehfeld,J.F., Marcker, K.A. and Vuust,J. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 1067-1069.
Wood,W.B. (1966) *J. Mol. Biol.*, **16**, 118-133.
Woods,D.E., Markham,A.F., Ricker,A.T., Goldberger,G. and Colten,H.R. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 5661-5665.

Received on 19 December 1984