Volume 231, number 1, 187–191 FEB 05766 April 1988

Human ubiquitin genes: one member of the UbB gene subfamily is a tetrameric non-processed pseudogene

Jack B. Cowland, Ove Wiborg* and Jens Vuust

Laboratory of Molecular Biology, Statens Seruminstitut, Amager Boulevard 80, DK-2300 Copenhagen S and **Department of Molecular Biology and Plant Physiology, Aarhus University, C.F. Moellers Allé 130, DK-8000 Aarhus C, Denmark*

Received 28 January 1988

The human ubiquitin gene family consists of three subfamilies. One of these, the UbB subfamily, includes a functional gene coding for a polyubiquitin protein that contains three ubiquitin copies tandemly repeated, as well as three pseudogenes of the processed type. We have now isolated a fifth human UbB type gene, different from any of the previously identified ones. This newly isolated gene is a tetrameric pseudogene which has presumably arisen by unequal crossing-over of two ancestral trimeric alleles. Southern blotting data indicate that all members of the human UbB gene subfamily are now accounted for.

Ubiquitin; Pseudogene; Intron; Unequal allelic cross-over

1. INTRODUCTION

Ubiquitin is a highly conserved, 76-amino acid residue protein found in all eucaryotic cells examined [1]. The protein is involved in a number of important cellular functions. In the cytoplasm, ubiquitin participates in the ATP-dependent, non-lysosomal proteolysis of both abnormal and certain short-lived regulatory proteins [1]. In the nucleus, ubiquitin is covalently bound to histone H2A [2]; in this position, it may play a role in determining the gross organization of chromatin [3]. Recent observations also indicate a possible role for ubiquitin as a regulator of the heat shock response [4], as well as a constituent of certain cell-surface receptors [5].

It was demonstrated previously in our laboratories [6] that ubiquitin is encoded in the human genome as a multigene family. Human poly(A)$^+$ RNA contains three distinctly sized ubiquitin gene transcripts (UbA, UbB and UbC) of approximately 650, 1100 and 2500 nucleotides, respectively. We found that the UbC gene transcript is of a remarkable structure as it codes for nine direct repeats of the 76-amino acid ubiquitin unit in a poly-protein format; the UbA gene was found to code for only a single ubiquitin [6]. It was later shown by others [8,9] that at least two UbA transcripts exist, each coding for a polypeptide containing ubiquitin fused to a non-ubiquitin peptide sequence at the C-terminus [8,9].

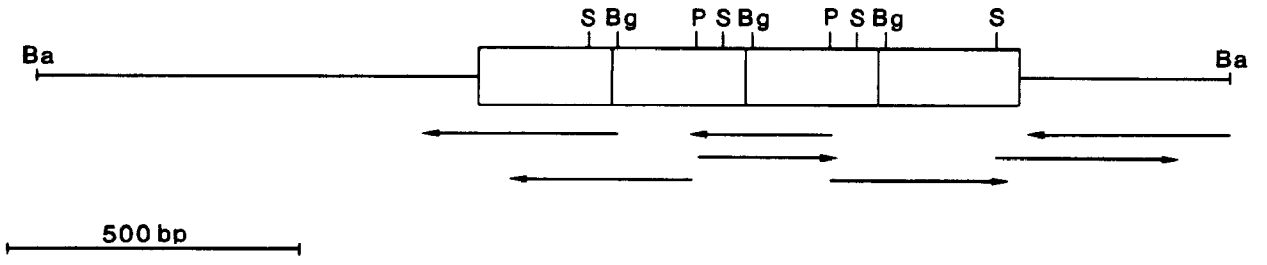
Recently, Baker and Board reported [7] that a functional UbB gene codes for three direct repeats of ubiquitin. These same authors also found three UbB pseudogenes [7–10], all belonging to the processed type of pseudogene.

In the following, we report the isolation and characterization of an UbB pseudogene with an organization quite different from the three previously described pseudogenes of this subfamily [7,10]. The total number of characterized human UbB genes is thus raised to five, which is the number predicted from genomic Southern blot analysis. We also describe the sequence of a cloned UbB cDNA, confirming the proposed structure of the active UbB type gene, and the presence of a 715 nucleotide intron in the 5′ non-coding region, as previously suggested by Baker and Board [7].

Correspondence address: J. Cowland, Department of Molecular Biology, Statens Seruminstitut, Amager Boulevard 80, DK-2300 Copenhagen S, Denmark

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/88/$3.50 © 1988 Federation of European Biochemical Societies

```
                                                                    -91 GATGCCAACTTTGAAG  -76

-75  CATATTAGATGTAAAAGCAGAAATACAA[C/T]CATCCTGAGATGACACGCTTATGTTTTACTTTTAATCTAGGTCAAA  -1

     Met Arg Ile Phe Val Lys Thr Leu Thr Gly Lys Ile Ile Thr Leu Glu Val Glu Pro
   1 ATG CGA ATC TTC GTG AAA ACC CTT ACC GGC AAG ATC ATC ACC CTT GAA GTG GAG CCC   57
         Gln                                     Thr
 229 --- -AG --- --- --- --G --- --G --- --- --- -C- --- --- --G --- --- --- ---  285
         Gln                                     Thr
 457 --- -AG --- --- --- --G --- --G --- --- --- -C- --- --- --G --- --- --- ---  513
         Lys                                     Thr
 685 --- AAG --- --- --- --G --- --G --T --A --- -C- --- --- --G --G --- --- ---  741

     Ser Ala Thr Ile Glu Asn Val Lys Ala Lys Ile Gln Asp Lys Glu Gly Asn Pro Cys
  58 AGT GCC ACT ATC GAA AAT GTC AAA GCC AAA ATC CAA GAT AAA GAA GGC AAT CCC TGT  114
         Asp                                             Lys         Ile     Pro
 286 --- -A- --C --- --- --- --G --G --- --G --- --- --- --- --- --- -TC --- CCC  342
         Asp                                             Lys                 Pro
 514 --- -A- --C --- --- --- --G --G --- --G --- --- --- --- --- --- -*- --- CCC  570
         Asp                                             Glu         Ile     Pro
 742 --- -A- --C --- --- --- --G --- --- --G --- --G --- G-- --- --- -TC --- CCC  798

     Asp Gln Gln Arg Leu Ile Phe Ala Gly Lys Gln Arg Glu Asp Gly Arg Ser Leu Ser
 115 GAC CAG CAG AGG CTC ATC TTT GCA GGC AAG CAG CGG GAA GAT GGC CGC AGT CTT TCT  171
     Gly Arg                     Ser             Gln Leu                 Ser
 343 -G- -G- --- --- --- --- -C- --- --- --- --- -T- --- --- --- --- --- --- ---  399
     Asp Gln                     Phe             Gln Leu                 Thr
 571 --T --- --- --- --- --- --- --- --- --- --- -T- --- --- --- C/G -T -C- --- ---  627
     Asp Gln                     Phe             Lys Leu                 Thr
 799 --T --- --- --- --- --- --- --- --- --- --- A-- TT- --- --- --- --- -C- --- ---  855

     Asp end Asn Ile Gln Lys Glu Ser Thr Leu His Leu Val Leu Arg Arg Arg Gly Gly
 172 GAC TAG AAC ATC CAG AAA GAG TCG ACC CTG CAT CTG GTT CTG CGT CGG AGA GGT GGT  228
         Tyr Asn                                                 Leu
 400 --- --C --- --- --- --- --- --- --- --- --- --- --C --- --- -T- --- --- ---  456
         Tyr Asn                                                 Leu
 628 --- --C --- --T -A/T- --G -A/T- --- --- --- --C --- --- --- --- -T- --- --- ---  684
         Tyr Ser                                                 Leu
 856 --- --C -G- --- --- --- --- --- --- --- --- --C --- --C --- --C -T- --G --- ---  912
     Cys END
 913 TGT TAA TTCTTCAGTCTTGCATTAGCAGTGCCCTCTAGTGGCATTACTCTGCACTATAGCCATTTGCCCCAAC  985

 986 TTAGGTTTAGAAATTACAAGTTTCAGTAATAGCTGAACCTGTTCAAAATGTTAATAAAGGTTTCGTTGAATGGTA 1060

1061 GCATACTTGGTGTTTTGTCATGATATTCTCTAGTGATGTGTGGGTACACTTAAAACTGGTGAAAATGTCTAGGGA 1135

1136 GGGATTTAATTTTGAGATTGGTAATGTGCCCAAAGTTAAGTCACTTGACTCTGGTATACACTTGGGTGGGCTGTG 1210

1211 GGGTAAGAGCCTTCTTT[G/A]CTGTAAGTCATTATTA 1244
```



2. MATERIALS AND METHODS

2.1. *Human genomic and cDNA libraries*

A genomic library prepared from adult human leucocyte DNA as described [6] was a gift from Dr J.P. Hjorth, Aarhus University. A human liver cDNA library was kindly donated by Dr Claudio Schneider, EMBL, Heidelberg. Screening of these libraries was carried out by plaque hybridizations essentially as described [11]. Initial screenings were carried out using the 684-bp *Xho*I fragment of pHUb14-38 or the 210-bp *Pvu*II-*Hin*dIII fragment of λHUb1 [6], both of which contain ubiquitin-coding sequences. In a subsequent screening, a 747-bp *Bam*HI-*Bgl*II fragment from λHUb-5 (see below), containing the intron of the UbB gene, was used to isolate the UbB pseudogene-containing clone λHUb-3A. Phage DNA prepared from positive plaques was digested with relevant restriction endonucleases, and the resulting fragments analyzed by Southern blotting.

2.2. *Nucleotide sequence determination*

Appropriate restriction fragments were end-labelled with [α-^{32}P]dATP and *Escherichia coli* DNA polymerase I 'Klenow fragment', and the nucleotide sequence was determined by the chemical cleavage procedure of Maxam and Gilbert [12].

2.3. *Genomic Southern blot*

5 μg of human leucocyte DNA prepared as described [6] was digested to completion with *Hin*dIII and separated on a 0.7% agarose gel. Following transfer to nitrocellulose filter (Schleicher and Schuell, BA 85), the DNA was hybridized with a 373 bp *Sal*I-*Bam*HI fragment from the 3′ non-coding region of the UbB gene in clone λHUb-3A essentially as described [13].

3. RESULTS AND DISCUSSION

Screening of 2×10^6 recombinants from an adult human leucocyte genomic library resulted in the isolation of 22 λ-phage clones containing ubiquitin specific sequences. Mapping of restriction enzyme recognition sites indicated that two of these clones (λHUb-3A and λHUb-5) contained sequences coding for polymeric ubiquitin, different from the previously characterized nonameric ubiquitin gene [6]. Results of analyses of the nucleotide sequence, as well as Northern blots (data not shown) revealed that both genes belong to the UbB gene subfamily. Clone λHUb-5 was found to contain a major portion of the functional UbB gene, described by Baker and Board [7], fused to the right arm of the vector.

The nucleotide sequence of the ubiquitin gene present in λHUb-3A is shown in fig.1. This gene differs from the previously described UbB genes [7] by possessing four ubiquitin copies. However, the λHUb-3A gene is surely a member of the UbB gene family since approx. 90% homology is found between the 5′ and 3′ non-coding regions of this gene and those of the human functional trimeric ubiquitin gene (fig.2). By contrast, an alignment with the corresponding regions of the monomeric [6,8,9] and nonameric [6] human ubiquitin genes shows no homology. Also, the 3′ non-coding sequence of the λHUb-3A gene binds only to the UbB transcripts in a Northern blotting analysis (not shown).

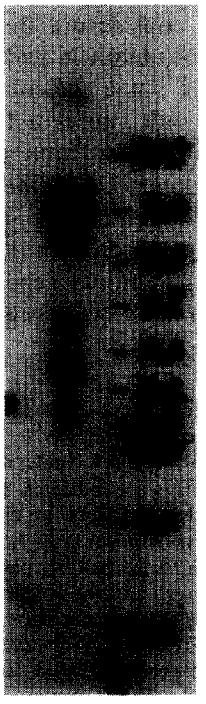
Unlike the trimeric gene, the tetrameric gene is incapable of encoding a complete poly-ubiquitin protein due to a stop-codon in the first ubiquitin-coding unit terminating translation after amino acid 58 (see fig.1). Additionally, the tetrameric gene contains a large number of amino acid substitutions that probably will render any protein product invalid, considering the high degree of conservation of the ubiquitin amino acid sequence. Finally, a single base deletion in the third ubiquitin-coding unit causes a frameshift of the correct ubiquitin reading frame.

The UbB pseudogenes described [7] appeared to belong to the processed type of pseudogene, as inferred from a comparison with the functional UbB gene. Thus, the functional gene was assumed to contain an intervening sequence from nucleotide

Fig.1. Sequence and physical map of the human tetrameric pseudogene. Top: nucleotide sequence of the coding region and parts of the 5′ and 3′ non-coding regions of the gene present in clone λHUb-3A. The positions are numbered from the A of the start codon; negative numbers are used in the 5′ non-coding region. The nucleotide sequence of the first ubiquitin coding repeat is given in full, while only differing nucleotides are indicated for the following repeats. Identical nucleotides are indicated by hyphens. A gap is introduced at position 563 (*) to preserve the reading frame. The amino acid sequence, deduced from the nucleotide sequence, is written in full for the first repeat. When two or more different residues are found at the same position in the repeats, they are all shown. Amino acid residues, differing from the known sequence of human ubiquitin, are underlined. Bottom: physical map of a 2.2 kb *Bam*HI fragment carrying the tetrameric pseudogene. The boxed region indicates the coding region of the ubiquitin gene; vertical lines separate the four direct ubiquitin repeats. The 5′ end of the gene is to the left. The strategy of sequencing is shown by horizontal arrows. The number of repeats was confirmed by a partial *Sal*I digest after inserting the fragment in the *Bam*HI site of pBR322, digesting with *Hin*dIII and *Pst*I, end-labeling with ^{32}P using Klenow polymerase, and retrieval of the largest fragment Ba, *Bam*HI; Bg, *Bgl*II; P, *Pvu*II; S, *Sal*I.

```
5` END
cDNA                          -72 TGGGTGAGCTTGTTTGTGTCCCTGTGGGTGGACGTGGTTGGTGATTGGCAGGATCCTGGTATCCGCTAACAG -7
UBI3 -815 ATTTAGGGGCGGTTGGCTTTGTTGGGTGAGCTTGTTTGTGTCCCTGTGGGTGGACGTGGTTGGTGATTGGCAGGATCCTGGTATCCGCTAACAGgtac -715

cDNA                                                                                                  -6 GTCAAA ATG 3
UBI3 -95 atagttgctaattttgaagaatattaggtgtaaaagcaagaaatacaatgatcctgaggtgacacgcttatgttttacttttaaactagGTCAAA ATG 3
UBI4   -91 GATGCCAACTTTGAAGCATATTAGATGTAAAAGCA GAAATACAA[T/C]CATCCTGAGATGACACGCTTATGTTTTACTTTTAATCTAGGTCAAA ATG 3

3` END
UBI3  688  TAA TTCTTCAGTCATGGCATTCGCAGTGCCCAGTGATGGCATTACTCTGCACTATAGCCATTTGCCCCAACTTAAGTTTAGAAATTACAAGTTTCAGT  785
UBI4  916  TAA TTCTTCAGTCTT GCATTAGCAGTGCCCTCTAGTGGCATTACTCTGCACTATAGCCATTTGCCCCAACTTAGGTTTAGAAATTACAAGTTTCAGT 1012

UBI3  786  AATAGCTGAACCTGTTCAAAATGTTAATAAAGGTTTCGTTGCATGGTAGCATACTTGGTGTTTTGTCATGAAATTCTCTAGTGATGTGTGGGTACGCTT  884
UBI4 1013  AATAGCTGAACCTGTTCAAAATGTTAATAAAGGTTTCGTTGAATGGTAGCATACTTGGTGTTTTGTCATGATATTCTCTAGTGATGTGTGGGTACACTT 1111

UBI3  885  AAAACTGGTGAAAATGTTTAGGGA    TTTAATTTTGAGATTGGTAATGTGCTCAAAGTTAAGTCACTTGACTTTGGTATACACTTGGGTGGGCTGAG  979
UBI4 1112  AAAACTGGTGAAAATGTCTAGGGAGGGATTTAATTTTGAGATTGGTAATGTGCCCAAAGTTAAGTCACTTGACTCTGGTATACACTTGGGTGGGCTGTG 1210

UBI3  980  GGGCAAGAGCCTTCTTTGCTGTTTAAGTCATTACAA 1015
UBI4 1211  GGGTAAGAGCCTTCTTT[G/A]CTGT  AAGTCATTATTA 1244
```

Fig.2. Nucleotide sequence comparison between the 5′ and 3′ regions of the trimeric human ubiquitin gene in clone λHUb-5 (UBI3), the tetrameric pseudogene in clone λHUb-3A (UBI4), and the 5′ non-coding region of the cDNA clone cHUb3-1100 (cDNA). Differing bases are indicated with an asterisk. The nucleotides of the intron of the trimeric gene are written in lower case, and the proposed cap-site is shown with an arrow. The 3′ non-coding sequence, and the sequence from −718 to −815, of the trimeric gene is from [7].



−7 to nucleotide −721, relative to the initiation codon ATG. We have proven the existence and location of this intron by analysing the sequence of a cDNA clone (cHUb3-1100), isolated from a human cDNA library (fig.2).

The tetrameric pseudogene differs from the three previously described pseudogenes of the UbB gene family [7,10] by containing intron-coding sequences. Therefore, this particular pseudogene does not belong to the processed category of pseudogenes. The fact that it carries an additional ubiquitin coding unit, compared to the active trimeric UbB gene, indicates that this pseudogene has arisen from an unequal crossing-over event between two ancestral trimeric alleles.

Fig.3. Southern blotting analysis of human DNA digested with *Hin*dIII. Hybridization was with the nick-translated *Sal*I-*Bam*HI 3′ non-coding fragment of λHUb-3A. Size markers are fragments from λ phage DNA digested with *Eco*RI and *Bam*HI; their sizes are given in kb. The faint band of >25 kb corresponds to hybridization to residual undigested genomic DNA.

The finding of this fourth pseudogene raises the total number of characterized members of the UbB gene subfamily to five, which is exactly the number of bands found on a Southern blot of *Hin*dIII digested genomic human DNA, hybridized with a 373 bp *Sal*I-*Bam*HI 3′ non-coding fragment of the UbB gene of λHUb-3A (fig.3). The intensities of the bands match a UbB gene population of three processed pseudogenes, containing only UbB-type 3′ noncoding sequences as far as the poly(A)-addition site, i.e. 142 bp, and two full length genes: the 'active' trimer gene and the tetrameric pseudogene both containing the entire sequence corresponding to the 373 bp probe. We conclude that all members of the human ubiquitin B gene subfamily have now been accounted for.

Acknowledgements: We thank Marianne Nielsen for skillful technical assistance and Karen Langner for typing the manuscript. This work was supported by grants from The Danish Cancer Society, the Danish Medical Research Council, Direktør Ib Henriksens Fond, Fonden af 17-12-1981, and Novo Industri A/S.

REFERENCES

[1] Hershko, A. and Ciechanover, A. (1987) Prog. Nucleic Acids Res. 33, 19–56.
[2] Goldknopf, I.L. and Busch, H. (1977) Proc. Natl. Acad. Sci. USA 74, 5492–5495.
[3] Matsui, S., Seon, B.K. and Sandberg, A.A. (1979) Proc. Natl. Acad. Sci. USA 76, 6386–6390.
[4] Ananthan, J., Goldberg, A.L. and Voellmy, R. (1986) Science 232, 522–524.
[5] Siegelman, M., Bond, M.W., Gallatin, W.M., St John, T., Smith, H.T., Fried, V.A. and Weissman, I.L. (1986) Science 231, 823–829.
[6] Wiborg, O., Pedersen, M.S., Wind, A., Berglund, L.E., Marcker, K.A. and Vuust, J. (1985) EMBO J. 4, 755–759.
[7] Baker, R.T. and Board, P.G. (1987) Nucleic Acids Res. 15, 443–463.
[8] Lund, P.K., Moats-Staats, B.M., Simmons, J.G., Hoyt, E., D'Ercole, J., Martin, F. and Van Myk, J.J. (1985) J. Biol. Chem. 260, 7609–7613.
[9] Salvesen, G., Lloyd, S. and Farley, D. (1987) Nucleic Acids Res. 15, 5485.
[10] Baker, R. and Board, P.G. (1987) Nucleic Acids Res. 15, 4352.
[11] Benton, W.D. and Davis, R.W. (1977) Science 196, 180–182.
[12] Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560–564.
[13] Southern, E.M. (1975) J. Mol. Biol. 98, 503–517.