Nucleotide Sequence Analysis of a cDNA Encoding Human Ubiquitin Reveals That Ubiquitin Is Synthesized as a Precursor*

(Received for publication, December 14, 1984)

P. Kay Luad¶, Billie M. Moats-Staats¶, James G. Simmons¶, Eileen Hoyt¶, A. Joseph D’Ercole¶, Frank Martin¶, and Judson J. Van Wyk¶

From the ¶Departments of Physiology and Pediatrics, The School of Medicine, The University of North Carolina, Chapel Hill, North Carolina 27514 and §Amgen, Thousand Oaks, California 91320-1789

Ubiquitin is a 76-amino acid protein whose sequence is highly conserved throughout evolution from invertebrates to mammals. It is both a cytoplasmic and nuclear protein. In the cytoplasm it is involved in ATP-dependent nonlysosomal proteolysis. In the nucleus, ubiquitin is conjugated to histone 2A and may play a role in regulation of chromatin structure and/or regulation of transcriptional activity. During attempts to identify a cDNA encoding somatomedin-C (insulin-like growth factor I) we screened a fetal human liver cDNA library with a mixture of 17 base oligonucleotides corresponding to a portion of the B chain of somatomedin-C. One oligonucleotide of the mixture hybridized to two cDNAs encoding ubiquitin despite a 2-base pair mismatch. Nucleotide sequence analyses of the 350- and 516-base pair cDNAs revealed that they correspond to the same ubiquitin mRNA. The coding sequence of the 516-base pair cDNA begins at amino acid 5 of the ubiquitin sequence and encodes amino acids 5 through 76 of ubiquitin, an 80-amino acid carboxy-terminal extension, a 3' untranslated region, and a poly(A) tail. The finding that ubiquitin is synthesized as a precursor raises the possibility that the precursor sequence may be important in compartmentalization of ubiquitin or ubiquitin precursors. Analyses of ubiquitin mRNAs in poly(A) RNA extracted from human liver and various rat tissues reveals that there are three distinct mRNAs encoding ubiquitin in humans and four mRNAs in the rat.

Ubiquitin was first isolated from bovine thymus and was reported to stimulate differentiation of B and T lymphocytes (1). It is a 76-amino acid protein (M₀ ≈ 8500) which is highly conserved throughout evolution and is identical in amino acid sequence in organisms as diverse as humans and insects (2, 3). Ubiquitin has been implicated in a variety of cellular functions. In the nucleus it is conjugated to histone 2A to form the nuclear protein A24 (4) which may play a role in regulation of chromatin structure (5) and has been reported to associate preferentially with actively transcribed genes (6). In cytoplasm ubiquitin is part of an ATP-dependent nonlysosomal proteolytic pathway which has been the focus of several recent investigations (7, 8). The precise physiological functions of ubiquitin in nucleus and cytoplasm are, however, still not clearly defined. Information about the structure and organization of ubiquitin genes and mRNAs may enhance our understanding of the multiple biological functions of ubiquitin. We report here the isolation of a human ubiquitin cDNA which encodes 95% of the ubiquitin molecule and a carboxy-terminal precursor sequence. We also demonstrate that multiple mRNAs encode ubiquitin.

**Experimental Procedures**

Oligonucleotide Synthesis and Labeling—The amino acid sequence Phe-Tyr-Phe-Asn-Lys-Pro corresponding to residues 23 to 28 of somatomedin-C/insulin-like growth factor I (9, 10) was chosen for synthesis of oligonucleotide probes complementary to the somatomedin-C/insulin-like growth factor I mRNA. Thirty-two possible 17-base oligomers were provided in four groups of 8 (Fig. 1) by Amgen Corp. Oligomers were end-labeled with 32P using γ-32P[ATP and T4 polynucleotide kinase and purified on 20% polyacrylamide gels containing 8 M urea (11).

Screening of a cDNA Library Prepared from Fetal Human Liver poly(A) RNA—Labeled oligomers were used to screen a cDNA library provided by Dr. Stuart Orkin and prepared from poly(A) RNA isolated from fetal human liver (12). Approximately 120,000 tetracycline-resistant transformants of Escherichia coli MC1061 were hybridized with a mixture of the 32 labeled oligomers. Hybridization conditions were essentially as described by Woods et al. (13) with modifications in hybridization and washing temperatures based on the length and base composition of oligomers. Melting temperatures (Tm) of oligomers were calculated by the summation of 4°C for every guanine or cytosine residue and 2°C for every adenine or thymine residue. This resulted in a calculated Tm range of 36 to 48°C for potential oligomer-cDNA hybrids. Hybridization and washing of filters were carried out at 36°C. Colonies showing a positive hybridization signal were selected from master plates. Recombinant plasmids were purified by the cleared lysate technique (14) followed by two centrifugations in cesium chloride/ethidium bromide gradients. Positive hybridization signals were confirmed by dot blot assays. Duplicate dot blots of plasmid DNA were prepared as described by Thomas (15), and were each hybridized with one of the four mixtures of 8 oligomers (Fig. 1). Hybridizations were carried out as described by Woods et al. (13), except that the hybridization temperature was 35°C and successive washings were performed at 30, 40, and 42°C. Four plasmids gave unambiguous hybridization signals (Fig. 2).

Nucleotide Sequence Analyses of cDNA Inserts—The four plasmid DNAs showing positive hybridization signals were digested with a variety of restriction endonucleases. Two plasmids contained cDNA inserts which had similar restriction maps. The two cDNA inserts of 350 and 516 base pairs (bp) were sequenced by the method of Maxam and Gilbert (16) following the sequencing strategy depicted in Fig. 3. Analyses of mRNAs Encoding Ubiquitin by Filter Hybridization (Northern blot analyses)—Aliquots of poly(A) RNA extracted from human liver and a variety of rat tissues (17) were denatured in glyoxal...
possible mRNA sequences were synthesized in four sets of 8 oligo-
nucleotides. Thirty-two 17-base oligonucleotides complementary to all
B chain of somatomedin-C was chosen for synthesis of oligonucleo-
tides. The predicted mRNA sequence encoding the amino acids
contained 5 degenerate codons to give 32 possible mRNA coding
nucleotides as shown.

Hybridizations were for 24 h at 36°C. Blots were exposed to
silver X-ray film (Kodak XAR-5) for 16 h at -70°C with intensifying screens.
Set 1 and set 3 oligomers showed positive hybridization signals. (20
X 3 0.4% NaCl, 0.3 M Na citrate, pH 7.0. 20 X Denhardt's = 0.4%
each bovine serum albumin, Ficoll type '70, and polyvinylpyrrolidone).

and dimethyl sulfoxide (15) and size fractionated by electrophoresis
on 1% agarose gels. The RNAs were transferred from the gels to Gene
Screen (New England Nuclear) by capillary transfer (18). Blots were
hybridized with restriction fragments of the 516-bp cDNA labeled with
32P by nick translation (19). Hybridizations were for 16 h at
42°C as previously described (20).

RESULTS AND DISCUSSION

Four of 120,000 bacterial clones hybridized with the mixture of
32 17-base oligomers (Fig. 1). Recombinant plasmids iso-
lated from the clones were analyzed by dot blot assays with
each of four different mixtures of 8 17-base oligomers. Three
out of the four plasmids hybridized with set 1 oligomers and
one plasmid hybridized with set 3 oligomers (Figs. 1 and 2).
Nucleotide sequence analyses were performed on two cDNA
inserts (350 and 516 bp, respectively) which hybridized to set
1 oligomers and showed similar restriction maps. The nucleo-
tide sequences of each revealed that the cDNAs corresponded
to the same mRNA with no ambiguities in the nucleotide
sequence. Comparisons of the cDNA nucleotide sequences
with those of the synthetic oligomers revealed two base pair
mismatches (Fig. 4). That positive hybridization signals were
obtained between cDNAs and oligomers, despite these mismat-
ches, is presumably because hybridization conditions were
designed to allow specific hybridization of all oligomers in a
mixture of 8 oligomers with different G/C and A/T content.
This finding demonstrates one of the problems in the use of
mixtures of oligomers for identification of cDNAs. Identification
of false positives in this manner could be avoided by
use of oligomers corresponding to at least two different regions
of a protein sequence.

The 516-bp cDNA insert contains a long open-reading
frame of 457 bases followed by the stop codon, TAA, a short
3' untranslated region of 28 nucleotides, and a tract of 28
adenine residues (Fig. 4). The sequence AATAAA found 12
nucleotides upstream from the poly(A) tract is characteristic
of polyadenylation signals found in other eukaryotic mRNAs
(21). The amino acid sequence of the protein derived from
decoding the longest open reading frame of the 516-bp cDNA
insert was compared with a mammalian protein sequence data
base by Dr. Russell Doolittle (University of California, San
Diego). The first 72 amino acids were found to correspond
identically to amino acids 5 to 76 of ubiquitin (1-3). The
carboxy-terminal glycine of ubiquitin is followed by
an
sequence. Comparisons of the cDNA nucleotide sequences
with those of the synthetic oligomers revealed two base pair
mismatches (Fig. 4). That positive hybridization signals were
obtained between cDNAs and oligomers, despite these mismat-
ches, is presumably because hybridization conditions were
designed to allow specific hybridization of all oligomers in a
mixture of 8 oligomers with different G/C and A/T content.
This finding demonstrates one of the problems in the use of
mixtures of oligomers for identification of cDNAs. Identification
of false positives in this manner could be avoided by
use of oligomers corresponding to at least two different regions
of a protein sequence.

The 516-bp cDNA insert contains a long open-reading
frame of 457 bases followed by the stop codon, TAA, a short
3' untranslated region of 28 nucleotides, and a tract of 28
adenine residues (Fig. 4). The sequence AATAAA found 12
nucleotides upstream from the poly(A) tract is characteristic
of polyadenylation signals found in other eukaryotic mRNAs
(21). The amino acid sequence of the protein derived from
decoding the longest open reading frame of the 516-bp cDNA
insert was compared with a mammalian protein sequence data
base by Dr. Russell Doolittle (University of California, San
Diego). The first 72 amino acids were found to correspond
identically to amino acids 5 to 76 of ubiquitin (1-3). The
carboxy-terminal glycine of ubiquitin is followed by an
extension of 80 amino acids which demonstrates that ubiquitin is
synthesized as a large molecular weight precursor (Fig. 4).

The 516-bp cDNA represents only a partial length copy of
1 ubiquitin mRNA as the coding sequence for the first 4
amino acids of ubiquitin (Met-Gln-Ile-Phe) is absent. We
therefore proceeded to analyze the size and complexity of
mRNAs encoding ubiquitin in poly(A) RNA extracted from
human liver and a human mammary carcinoma cell line (HS-
0578T). RNAs were hybridized with two different 32P-labeled
probes generated by digestion of the 516-bp cDNA with the
**Human Ubiquitin cDNA and Precursor**

**Fig. 4.** Nucleotide sequence of the ubiquitin cDNA (5) and corresponding amino acid sequence of a ubiquitin precursor. The cDNA sequence shown was obtained by sequence analyses of two cDNAs of 350 and 516 bp. The amino acid sequence of the ubiquitin precursor is shown above the nucleotide sequence. Codons for the first 4 amino acids of ubiquitin (underlined) are absent from the cDNA. The carboxy-terminal amino acid (glycine, position 76) is followed by an extension of 80 amino acids, the UBCP. This sequence contains a high proportion of basic amino acids (boxed). The underlined DNA sequence AATAAA at the 3’ end of the cDNA is characteristic of sites involved in addition of the poly(A) tract to eukaryotic mRNAs (21). The region of the ubiquitin cDNA which base-paired with set 1 somatomedin-C oligomers is shown.

Restriction endonuclease DdeI (Fig. 3). A 218-bp cDNA fragment encoding amino acids 5 to 76 of ubiquitin, hybridized to three mRNAs of estimated sizes 600, 900, and 2400 bases in human liver and mammary carcinoma RNAs (Fig. 5A). A 300-bp fragment encoding the carboxy-terminal precursor sequence, 3’ untranslated region, and poly(A) tail, hybridized strongly to the 600-base mRNA, and showed no significant hybridization with the 900 and 2400-base mRNAs (Fig. 5B). These data suggest that the 516-bp cDNA characterized here corresponds to the 600 base mRNA. In addition, these data demonstrate that human liver and mammary carcinoma synthesize three distinct mRNAs encoding ubiquitin, and indicate that only one of these, the 600-base mRNA, encodes the carboxy-terminal precursor sequence. By similar hybridization analyses, a variety of rat tissues were shown to contain multiple ubiquitin mRNAs (Fig. 5C). In all tissues analyzed, mRNAs of estimated sizes 670, 1500, 1600, and 3000 bases hybridized to the 216-bp probe encoding ubiquitin (Fig. 5C). The 300-bp probe encoding the human carboxy-terminal precursor sequence hybridized strongly to the 670-base rat mRNA indicating conservation of this sequence across humans and rats (Fig. 5D). These findings indicate a biological role for the ubiquitin carboxy-terminal precursor sequence (UBCP). The precursor sequence is highly basic containing 33.3% basic amino acids and a high ratio of lysine to arginine (Fig. 4). The highly basic nature is characteristic of nuclear proteins such as histones and the high mobility group nuclear proteins (22) which may indicate a nuclear function for UBCP. In this regard it is of interest to note a stretch of six consecutive basic amino acids near the carboxy terminus of UBCP (Fig. 4). Recent evidence has indicated that a similar stretch of five basic amino acids in the T antigen of SV40 is involved in transport of T antigen to the nucleus or retention within the nucleus (23, 24). Although speculative, one possibility is that UBCP plays a role in transport of the ubiquitin precursor to the nucleus. The concept of the involvement of precursor sequences in nuclear compartmentalization is a novel one. Other nuclear proteins such as histones (25) and the frog oocyte proteins nucleoplasmmin (25) and N1,2, and 4 (27) have been shown to contain within their mature sequences the necessary information for migration to the nucleus and are not apparently synthesized as precursor forms (27, 28). The dual functions of ubiquitin

---

**Table:** Pro-ubiquitin cDNA and ubiquitin cDNA sequences. The sequences are aligned with the corresponding amino acid sequences, and the underlined bases indicate the presence of a TATAA signal at the 3’ end of the cDNA.
8aP-labeled DdeI fragments to Gene Screen. Blots were hybridized to 32P-labeled (nick-translated) D, blots were hybridized with a 300-bp fragment encoding the 80 fractionated by electrophoresis on a 1% agarose gel and transferred autoradiography. In A and carcinoma cell line.

The finding that ubiquitin is synthesized as a precursor also raises questions about the mechanism of post-translational processing of the precursor. At this point it is not possible to predict the pathway of cleavage of the precursor to form ubiquitin. Processing of ubiquitin from the precursor would, however, result in the formation of UBCP or a fragment thereof, and the possibility exists that this molecule represents a biologically active molecule in its own right. Since the cDNA encoding ubiquitin was identified by hybridization to an oligomer encoding a portion of the B chain of somatomedin-C (Figs. 1 and 4), we compared the sequences of the B chain of somatomedin-C and UBCP. Only limited amino acid sequence homology was found (as shown in Fig. 6) and this was not statistically significant by the criteria proposed by Doolittle (30). That any homology exists between somatomedin-C, a growth factor (10), and UBCP, nonetheless suggests a starting point for investigation of a biological role of UBCP.

After submission of this manuscript, Ozkaynak et al. (31) published the sequence of a yeast ubiquitin gene and Dworkin-Rastl et al. (32) published the sequence of a xenopus ubiquitin cDNA. The DNA sequences predict synthesis, in both species, of ubiquitin precursors which contain several repeats of the ubiquitin amino acid sequence (polyubiquitin precursors) (31, 32). The yeast and xenopus ubiquitin precursors differ from the human precursor reported here (Fig. 4). However, the existence of polyubiquitin precursors in these species raises the possibility that one or more of the human or rat ubiquitin mRNAs, which hybridize to a cDNA fragment encoding human ubiquitin (Fig. 5), encode a polyubiquitin precursor. Characterization of the multiple human and rat ubiquitin mRNAs by nucleotide sequence analysis of cDNAs will establish whether this is the case. This information will also provide a basis to investigate whether the multiple human and rat ubiquitin mRNAs are the products of different genes and whether ubiquitins in precursors of different configurations may be destined for functions in different cellular locations.

Acknowledgments—We are indebted to Dr. Russell Doolittle for homology searches, Dr. Stuart Orkin for providing the fetal liver cDNA library, Carol Hubbard for technical assistance, and Kym Shilanta for editorial assistance.

**REFERENCES**