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

P. Kay Luad§¶, Billie M. Moats-Staats§¶, James G. Simmonst, 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 ¶Angen, Thousand Oaks, California 91320-1789

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 (Mr = 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 and may play a role in regulation of chromatin structure (4). 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.

* This work was supported by National Institutes of Health Grants AM01022 and HD06299. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¶ Recipient of a Research Career Development Award from the Juvenile Diabetes Foundation. To whom reprint requests should be addressed.

§¶ Recipient of Research Career Development Award HD00435 from the National Institutes of Child Health and Human Development.

The abbreviations used are: bp, base pairs; UBCP, ubiquitin carboxy-terminal precursor sequence.
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 $^{32}$P 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 isolated 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 nucleotide 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 mismatches, 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 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 $^{32}$P-labeled probes generated by digestion of the 516-bp cDNA with the
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.

Ubiquitin is both a nuclear and cytoplasmic protein. In the nucleus it is conjugated to histone 2A by an isopeptide linkage between its carboxy-terminal glycine and the ε-NH₂ group of the lysine at position 119 of histone 2A (4). In the cytoplasm ubiquitin is involved in ATP-dependent, nonlysosomal proteolysis (7). It is possible that different genes and mRNAs encode nuclear and cytoplasmic ubiquitins. Our finding of multiple ubiquitin mRNAs (Fig. 5) is consistent with this possibility. The cDNA characterized here encodes a carboxy-terminal precursor sequence in addition to ubiquitin. Hybridization data indicate that the precursor sequence is specific to one of three mRNAs in human tissues (Fig. 5B) and is conserved across humans and rats (Fig. 5D). These findings indicate a biological role for the ubiquitin carboxyl-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 amino 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...
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
1. Goldstein, G., Scheid, M., Hammerling, U., Boyse, E. A., Schlesinger, D. H., and Niall, H. D. (1975) Proc. Natl. Acad. Sci. U. S. A. 72, 11–15
2. Schlesinger, D. H., and Goldstein, G. (1975) Nature 255, 423–424
3. Gavilanes, J. G., Gonzalez de Buitrago, C., Perez-Castells, R., and Rodriguez, R. (1982) J. Biol. Chem. 257, 10297–10270
4. Goldknopf, I. L., and Busch, H. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 864–868
5. Matsu, S. I., Seon, B. K., and Sandberg, A. A. (1979) Proc. Natl. Acad. Sci. U. S. A. 76, 6386–6390
6. Levinger, L., and Varshavsky, A. (1982) Cell 28, 375–385
7. Hershko, A. (1983) Cell 34, 11–12
8. Ciechanover, A., Finley, D., and Varshavsky, A. (1984) Cell 37, 57–66
9. Rinderknecht, E., and Humbl, R. E. (1978) FEBS Lett. 89, 283–286
10. Klapper, D. G., Svoboda, M. E., and Van Wyk, J. J. (1983) Endocrinology 112, 2215–2217
11. Wallace, R. B., Shaffer, J., Murphy, R. F., Bonner, J., Hirose, T., and Itakura, K. (1979) Nucleic Acids Res. 6, 3543–3557
12. Michelson, A. M., Markham, A. F., and Orkin, S. H. (1983) Proc. Natl. Acad. Sci. U. S. A. 80, 472–476
13. Woods, D. E., Markham, A. F., Ricker, A. T., Goldberger, G. A., and Collen, H. R. (1982) Proc. Natl. Acad. Sci. U. S. A. 79, 5661–5665
14. Clewell, D. B., and Helsinki, D. R. (1970) Biochemistry 9, 4428–4440
15. Thomas, P. S. (1980) Proc. Natl. Acad. Sci. U. S. A. 77, 5201–5206
16. Markham, A., and Gilbert, W. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 560–564
17. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., and Rutter, W. J. (1979) Biochemistry 18, 5294–5299
18. Southern, E. M. (1975) J. Mol. Biol. 98, 503–517

Fig. 5. Autoradiogram of Northern blot hybridizations of 32P-labeled DdeI fragments of the ubiquitin cDNA with mRNAs in human and rat tissues. Aliquots of mRNAs were fractionated by electrophoresis on a 1% agarose gel and transferred to Gene Screen. Blots were hybridized to 32P-labeled (nick-translated) DdeI fragments of the ubiquitin cDNA followed by washing and autoradiography. In A and C, blots were hybridized with a 218-bp fragment encoding amino acids 5 through 76 of ubiquitin. In B and D, blots were hybridized with a 300-bp fragment encoding the 80 amino acid carboxyl-terminal precursor sequence, 3' untranslated region, and poly(A) tail. Sizes of mRNA were estimated by comparison of migration of mRNAs with ethidium bromide-stained DNA molecular weight markers. A, mRNA isolated from human tissues: lane 1, liver; lane 2, liver; and lane 3, mammary carcinoma cell line. (The largest migrating band in lane 2 co-migrates with 28 S ribosomal RNA and probably represents hybridization of the ubiquitin cDNA with 28 S rRNA sequences present in this mRNA preparation.) B, mRNAs isolated from human tissues: lane 1, liver; lane 2, mammary carcinoma cell line. C, mRNAs isolated from rat tissues: lane 1, testis; lane 2, pancreas; lane 3, testis; lane 4, brain; and lane 5, thymus. D, mRNAs isolated from rat tissues: lane 1, testis; lane 2, brain, and lane 3, thymus.

Fig. 6. Comparison of the amino acid sequences of preprosomatomedin-C and the UBCP. The amino acid sequence of UBCP was compared with all regions of the somatomedin-C precursor (29). Gaps were introduced to maximize homology. Shown are regions of homology between UBCP and a short region of the signal sequence of the somatomedin-C precursor, and UBCP and the B chain of somatomedin-C. Identical amino acids are boxed.

in nucleus and cytoplasm may, however, require distinct signals to regulate the distribution of ubiquitin between the two compartments.

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
19. Rigby, P. W. J., Diekmann, M., Rhodes, C., and Berg, P. (1977) *J. Mol. Biol.* **113**, 237–251
20. Lund, P. K., Goodman, R. H., Montminy, M. R., Dee, P. C., and Habener, J. F. (1983) *J. Biol. Chem.* **258**, 3280–3284
21. Proudfoot, N. J., and Brownlee, G. G. (1975) *Nature* **253**, 211–214
22. Walker, J. M. (1982) in *The HMG Chromosomal Proteins: Primary Structures* (Johns, E. W., ed) pp. 69–87, Academic Press, London
23. Lanford, R. E., and Butel, J. S. (1984) *Cell* **37**, 801–813
24. Kalderon, D., Richardson, W. D., Markham, A. F., and Smith, A. E. (1984) *Nature* **31**, 33–38
25. Gurdon, J. B. (1970) *Proc. R. Soc. London. B Biol. Sci.* **176**, 303–314
26. Dingwall, C., Sharnick, S. V., and Laskey, R. A. (1982) *Cell* **30**, 449–458
27. Dabauvalle, M. C., and Franke, W. W. (1982) *Proc. Natl. Acad. Sci. U. S. A.* **79**, 5302–5306
28. De Robertis, E. M. (1983) *Cell* **32**, 1021–1025
29. Jansen, M., Van Schaik, F. M. A., Ricker, A. T., Bullock, B., Woods, D. E., Gabbay, K. H., Nussbaum, A. L. Sussench, J. S., and Van den Brande, J. L. (1983) *Nature* **306**, 609–611
30. Doolittle, R. F. (1981) *Science* **214**, 149–159
31. Ozkaynak, E., Finley, D., and Varshavsky, A. (1984) *Nature* **312**, 663–666
32. Dworkin-Rastl, E., Shrutkowski, A., and Dworkin, M. B. (1984) *Cell* **39**, 321–330