Nucleotide and derived amino acid sequences of a cDNA coding for pre-uteroglobin from the lung of the hare (*Lepus capensis*)

Maria S. LÓPEZ DE HARO and Antonio NIETO*

Centro de Biología Molecular (C.S.I.C.–U.A.M.), Universidad Autónoma, Canto Blanco, 28049 Madrid, Spain

An almost full-length cDNA coding for pre-uteroglobin from hare lung was cloned and sequenced. The derived amino acid sequence indicated that hare pre-uteroglobin contained 91 amino acids, including a signal peptide of 21 residues. Comparison of the nucleotide sequence of hare pre-uteroglobin cDNA with that previously reported for the rabbit gene indicated five silent point substitutions and six others leading to amino acid changes in the coding region. The untranslated regions of both pre-uteroglobin mRNAs were very similar. The amino acid changes observed are discussed in relation to the different progesterone-binding abilities of both homologous proteins.

INTRODUCTION

Uteroglobin, a small secretory protein, is synthesized mainly in the uterus and lung of the rabbit. Recent studies have focused considerable attention on this protein as its synthesis is subjected to differential control by steroid hormones in the two mentioned organs (for a review see Beato *et al.*, 1983). These studies have led to the isolation and characterization of the rabbit uteroglobin gene and its 5' flanking sequences containing putative recognition sites for steroid receptors (Bailly *et al.*, 1983; Suske *et al.*, 1983; Cato *et al.*, 1984). The physiological function of uteroglobin remains to be established, although its high expression during the implantation period together with the fact that the protein binds progesterone may implicate its involvement in reproductive functions (Krishnan & Daniel, 1967; Mukherjee *et al.*, 1983).

Uteroglobin appears to be present only in the rabbit and a few other related species (Nieto & Lombardero, 1982). Recently we have isolated and characterized the uteroglobin of one of these species, namely the hare (*Lepus capensis*). Several features appear to distinguish this uteroglobin from that of the rabbit, including its chemical properties, progesterone-binding ability and, possibly, its hormonal regulation (López de Haro & Nieto, 1983).

Therefore it would be of interest to perform a more extensive study of this uteroglobin, as it could throw more light on several aspects of its structure, function and evolution. In this work we present the nucleotide sequence of a hare pre-uteroglobin cDNA, from which the complete amino acid sequence of the protein and other distinctive features have been deduced.

MATERIALS AND METHODS

Materials

Enzymes were obtained from either Bio-Labs or Boehringer–Mannheim. Low-gelling-temperature agarose (type VII) was from Sigma Chemical Co. [α -³²P]dATP and [γ -³²P]ATP were purchased from Amersham Inter-

national. Avian-myeloblastosis virus reverse transcriptase was from Life Sciences.

Preparation and cloning of cDNA

Poly(A)-containing RNA from hare lung was obtained by oligo(dT)–cellulose chromatography as described previously (López de Haro *et al.*, 1984). Double-stranded cDNA was prepared by using reverse transcriptase and *Escherichia coli* DNA polymerase I for the synthesis of the first and second strands respectively (Wickens *et al.*, 1978). A cDNA library was prepared by cloning at the *Pst*I site of the plasmid pUC12 by using the dG/dC tailing procedure (Maniatis *et al.*, 1982).

Screening of the library

Recombinants bearing uteroglobin cDNA sequences were identified by colony hybridization (Grunstein & Hogness, 1975), with rabbit uteroglobin cDNA probes labelled by nick translation. The probes used were pUG10 cloned in plasmid pBR322 (Fernández-Renau *et al.*, 1984) and pUG119 cloned in plasmid pUC18. This latter plasmid contained the sequence coding for the complete mature rabbit protein in addition to non-translated 3' sequences (López de Haro & Nieto, 1985). Plasmid DNA was obtained from positive colonies, digested with restriction enzymes and analysed by agarose-gel electrophoresis. DNA from selected recombinant plasmids was prepared as described by Birnboim & Doly (1979).

Sequence analysis

Plasmid DNA was digested with restriction enzymes and labelled at the 5'-ends with polynucleotide kinase or at the 3'-ends with the Klenow fragment of DNA polymerase I (Maxam & Gilbert, 1980). After secondary digestion with the appropriate enzyme, labelled fragments were isolated by agarose-gel electrophoresis (Langridge *et al.*, 1980) and subjected to chemical degradation by the procedure of Maxam & Gilbert (1980) as modified by Cooke *et al.* (1980).

* To whom correspondence should be addressed.

RESULTS

The cDNA library from hare lung was screened with rabbit uteroglobin cDNA probes, which were shown to cross-hybridize with hare uteroglobin mRNA sequences (López de Haro *et al.*, 1984). Six out of 1500 recombinants gave positive signals. This number of positive clones could be in the range of those expected, taking into account that in hare lung (López de Haro *et al.*, 1984) the uteroglobin mRNA contents are twice those found in rabbit lung [about 0.1–0.2% of the total mRNA population (Savouret *et al.*, 1980; Heins & Beato, 1981)]. Plasmid DNA from positive clones was digested with *Pst*I restriction endonuclease to release the inserted cDNA,

and the size of the inserts was estimated by agarose-gel electrophoresis. One of the clones, termed pUGh5, containing the longest insert (about 550 base-pairs) was selected for sequence determination.

A restriction map of clone pUGh5 and the sequence strategy is shown in Fig. 1. The sequence was obtained starting at either the *Bam*HI or the *Hind*III sites of the multilinker region of the plasmid pUC12 and reading through a few nucleotides of the plasmid DNA. Alternatively, part of the sequence was obtained starting at an internal *Eco*RI site in the insert. The nucleotide sequence of the pUGh5 clone is shown in Fig. 2. The insert consisted of 434 nucleotide residues in addition to the poly(A) tract and the oligo(dC/dG) tails added for

Fig. 1. Restriction map and sequence strategy of clone pUGh5

Only the restriction sites used in sequencing or in cloning are shown. The sequence was obtained by 5'-end labelling at either the *Bam*HI or the *Hind*III sites of plasmid pUC12 and secondary cleavage with *Hind*III or *Bam*HI respectively. Alternatively, the internal *Eco*RI site was labelled at the 5'- or the 3'-ends and the sequence was obtained after secondary digestion with *Bam*HI or *Hind*III restriction endonuclease. Unbroken and dashed arrows refer to the sense and anti-sense strands respectively. Black boxes indicate the oligo(dC/dG) tails. The hatched area indicates the coding region.

Fig. 2. Nucleotide and derived amino acid sequences of clone pUGh5

The nucleotide changes corresponding to the sequence of the rabbit pre-uteroglobin mRNA are written above the hare sequence for comparison. Also, the corresponding amino acid residues of rabbit pre-uteroglobin are shown underneath and boxed. The arrow indicates the boundary between the signal peptide and the mature uteroglobin. The termination codon, the polyadenylation signal and the putative pre-initiation signal are underlined. Nucleotides deleted in the rabbit sequence are bracketed. Amino acids are numbered starting at Met-21 of the signal peptide and extending up to Met-70 of the mature protein.

cloning. The amino acid sequence of the encoded protein can be deduced from this nucleotide sequence, it being taken into account that hare uteroglobin, like its rabbit counterpart, is synthesized as a precursor containing a signal peptide (López de Haro *et al.*, 1984). Preceded by several nucleotide residues of the 5' untranslated region, the protein starts at the ATG initiation codon and extends, through a 21-amino acid-residue signal peptide, up to the *N*-terminal glycine residue of the mature hare uteroglobin (López de Haro & Nieto, 1983). The mature protein consists of 70 amino acid residues preceding the termination codon.

A comparison of these nucleotide and derived amino acid sequences with those reported for rabbit pre-uteroglobin mRNA (Bailly *et al.*, 1983; Suske *et al.*, 1983) indicates considerable homology between both mRNAs. In the coding region only point substitutions can be observed. These substitutions lead to differences in one amino acid in the signal peptide and five in the mature protein; another five substitutions are silent. The portion of the 5' non-translated region obtained in this clone shared total homology with its rabbit counterpart, and includes the consensus sequence CCACC, which has been implicated in the recognition of the initiation site by the ribosome (Kozak, 1984). The 3' untranslated region shows almost total homology with that of the rabbit mRNA, with only five nucleotide substitutions and a short dinucleotide tract that is deleted in the rabbit (positions 366 and 367). The polyadenylation signal is also located at the same relative position in both homologous mRNAs.

DISCUSSION

The results reported here show the obtainment of an almost complete cDNA for hare lung pre-uteroglobin. In fact, if the 5' untranslated regions of hare and rabbit pre-uteroglobins were of similar size, about 30 nucleotide residues of this region would have been lost during the cloning procedure. Comparison of the nucleotide sequences of the coding regions of rabbit and hare pre-uteroglobins indicate five silent or synonymous substitutions among 273 nucleotide residues. This proportion suggests a time of divergence between the two species of about 4000000 years, taking into account the estimated rate of synonymous mutations in the coding regions of other genes (Miyata *et al.*, 1980). This estimated time of divergence would support evidence obtained from palaeontological records (López Martínez, 1977). On the other hand, the proportion of nucleotide substitutions leading to an amino acid change compared with that of silent ones appears to be higher than that expected from other data (Miyata *et al.*, 1980). This fact might suggest a rapid evolution of the protein.

The 3' untranslated regions of both pre-uteroglobin mRNAs are fairly similar, a finding in agreement with the postulation that these regions are highly conserved in homologous proteins and perhaps serve as regulatory elements in gene expression (Yaffe *et al.*, 1985).

Determinations of the apparent molecular mass of hare uteroglobin by SDS/polyacrylamide-gel electrophoresis suggested that it could be slightly smaller than rabbit uteroglobin (López de Haro & Nieto, 1983). The amino acid sequence shown here indicates that both uteroglobins contain the same number of residues. Since the *N*-terminal amino acid of the mature hare uteroglobin

is glycine (López de Haro & Nieto, 1983), it could be that hare pre-uteroglobin is intracellularly cleaved between Pro-4 and Gly-5 instead of between Ala-1 and Gly-1, thus resulting in a mature uteroglobin shorter than its rabbit counterpart. However, this seems unlikely, as the Cys-3 would then be lost and, therefore, hare uteroglobin would contain only one cysteine residue per monomeric molecule, whereas the amino acid analysis shows that it contains two (López de Haro & Nieto, 1983).

It should also be mentioned that the similarity in size of both uteroglobins was suggested by careful determinations of their respective Stokes radii (López de Haro & Nieto, 1983). Therefore, it appears most likely that hare uteroglobin presents an aberrant behaviour on SDS/polyacrylamide-gel electrophoresis.

One striking and well-studied property of uteroglobin is its ability to bind progesterone (Beato *et al.*, 1983). X-ray-crystallographic analyses have shown that the dimeric form of the protein contains a hydrophobic cavity inside the molecule with a size fitting that of progesterone (Mornon *et al.*, 1980). The residues delimiting this cavity are also known (Mornon *et al.*, 1980; Morize, 1984). Hare uteroglobin also binds progesterone, although to a much lesser extent (López de Haro & Nieto, 1983). It is possible that the amino acid substitutions in hare uteroglobin could change the secondary and tertiary structures of the protein, leading to variations in the shape or size of the internal cavity. Also, changes in the residues delimiting the cavity could lead to a diminished affinity for progesterone. Examination of the hare uteroglobin sequence indicated that none of the residues of the internal cavity was changed. Application of the predictions for secondary structure (Chou & Fasman, 1974) suggests that most of the amino acid changes in hare uteroglobin should not result in important variations of the secondary structure. However, the substitution of Arg-5 by glycine could be of importance. This amino acid change appears to occur very infrequently, possibly as a consequence of the role that certain glycine residues play in the packing of helical regions (Schulz & Schirmer, 1979).

Thus this substitution in hare uteroglobin might result in a greater flexibility of the chain at this point, close to the Phe-6 and Ile-2 residues, which form part of the internal cavity. This greater flexibility might alter the size or shape of the cavity, leading to the decreased capacity for progesterone binding of hare uteroglobin. In this respect it should be mentioned that a single substitution of glycine by other amino acids produces alterations in the secondary structure of the p21 protein (Pincus & Brandt-Rauf, 1985), which, indeed, changes its physiological properties and causes malignant transformation of cells.

Uteroglobin is synthesized in the lung and the uterus. In this study we used hare lung as source of poly(A)-containing RNA, as it has a high content of uteroglobin mRNA (López de Haro *et al.*, 1984). Nevertheless, it appears likely that hare uteroglobins from uterus and lung were the same protein, on taking into account that, in rabbit, the gene has a single copy and the uterine and lung uteroglobins present the same amino acid sequence (López de Haro & Nieto, 1985), thus ruling out the possibility of differential tissue expression.

This work was supported by grants from Comisión Asesora para el Desarrollo de la Investigación Científica y Técnica and Fondo de Investigaciones Sanitarias (F.I.S.). M.S.L.H. is a fellow of F.I.S.

REFERENCES

- Bailly, A., Atger, M., Atger, P., Cerbon, M., Alizon, M., Vu Hai, M. T., Logeat, F. & Milgrom, E. (1983) *J. Biol. Chem.* **258**, 10384–10389
- Beato, M., Arnemann, J., Menne, C., Müller, H., Suske, G. & Wenz, M. (1983) in *Regulation of Gene Expression by Hormones* (McKerns, K. W., ed.), pp. 151–175, Plenum Press, New York
- Birnboim, H. C. & Doly, J. (1979) *Nucleic Acids Res.* **7**, 1513–1523
- Cato, A. C. B., Geisse, S., Westphal, H. M. & Beato, M. (1984) *EMBO J.* **13**, 2771–2778
- Chou, P. Y. & Fasman, G. D. (1974) *Biochemistry* **13**, 211–245
- Cooke, N. E., Coit, D., Weiner, R. I., Baxter, J. D. & Martial, J. A. (1980) *J. Biol. Chem.* **255**, 6502–6510
- Fernández-Renau, D., Lombardero, M. & Nieto, A. (1984) *Eur. J. Biochem.* **144**, 523–527
- Grunstein, M. & Hogness, D. S. (1975) *Proc. Natl. Acad. Sci. U.S.A.* **72**, 3961–3965
- Heins, B. & Beato, M. (1981) *Mol. Cell. Endocrinol.* **21**, 139–150
- Kozak, M. (1984) *Nucleic Acids Res.* **12**, 857–872
- Krishnan, R. S. & Daniel, J. C., Jr. (1967) *Science* **158**, 490–492
- Langridge, J., Langridge, P. & Berquist, P. L. (1980) *Anal. Biochem.* **103**, 264–271
- López de Haro, M. S. & Nieto, A. (1983) *Arch. Biochem. Biophys.* **226**, 539–547
- López de Haro, M. S. & Nieto, A. (1985) *FEBS Lett.* **193**, 247–249
- López de Haro, M. S., Fernández-Renau, D. & Nieto, A. (1984) *Comp. Biochem. Physiol. B* **78**, 885–890
- López Martínez, N. (1977) *Trabajos N/Q* (C.S.I.C., Madrid) **3**, 7–45
- Maniatis, T., Fritsch, E. & Sambrook, J. (1982) *Molecular Cloning*, pp. 239–242, Cold Spring Harbor Laboratory, Cold Spring Harbor
- Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560
- Miyata, T., Yasunaga, T. & Nishida, T. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 7328–7332
- Morize, I. (1984) Ph.D. Thesis, Université Paris VII
- Mornon, J. P., Fridlansky, F., Bally, R. & Milgrom, E. (1980) *J. Mol. Biol.* **137**, 415–429
- Mukherjee, D. C., Agrawal, A. K., Manjunath, R. & Mukherjee, A. B. (1983) *Science* **219**, 989–991
- Nieto, A. & Lombardero, M. (1982) *Comp. Biochem. Physiol. B* **71**, 511–514
- Pincus, M. R. & Brandt-Rauf, P. W. (1985) *Proc. Natl. Acad. Sci. U.S.A.* **82**, 3596–3600
- Savouret, J. F., Loosfelt, H., Atger, M. & Milgrom, E. (1980) *J. Biol. Chem.* **255**, 4131–4136
- Schulz, G. E. & Schirmer, R. H. (1979) in *Principles of Protein Structure* (Cantor, C. R., ed.), pp. 166–205, Springer-Verlag, New York
- Suske, G., Wenz, M., Cato, A. C. B. & Beato, M. (1983) *Nucleic Acids Res.* **11**, 2257–2271
- Wickens, M. P., Buell, G. N. & Schimke, R. T. (1978) *J. Biol. Chem.* **253**, 2483–2495
- Yaffe, D., Nudel, U., Mayer, Y. & Neuman, S. (1985) *Nucleic Acids Res.* **13**, 3723–3737

Received 23 December 1985/13 February 1986; accepted 28 February 1986