Volume 167, number 1 FEBS 1198 February 1984

The complete sequence of human preprocalcitonin

J.M. Le Moullec*, A. Jullienne+, J. Chenais, F. Lasmoles, J.M. Guliana, G. Milhaud and M.S. Moukhtar

LA 163 CNRS et U113 Inserm, Service de Biophysique, Faculté de Médecine Saint-Antoine, 27 rue Chaligny, 75571 Paris Cédex 12 and **Centre de Recherches Roussel Uclaf, Dept. Recherches Biotechnologiques, 111 route de Noisy, 93230 Romainville, France*

Received 8 November 1983; revised version received 6 December 1983

DNA complementary to mRNA extracted from the thyroid glands of patients suffering from medullary carcinoma of the thyroid (MCT), a calcitonin-producing tumour, was inserted in the *Pst* site of pBR 322 by G-C tailing. The recombinant plasmids were used to transform *Escherichia coli* DP 50. Ampicillin-resistant clones were screened using a ^{32}P-labelled cDNA to mRNA extracted from a case of MCT particularly rich in calcitonin (CT) mRNA. Positive clones were subsequently rescreened using a ^{32}P poly(T) probe. Eighty clones were thus purified, and the inserts obtained by digestion with *Pst*I were subjected to positive hybridization selection with subsequent translation in vitro. An insert stimulating synthesis of the protein and containing restriction sites compatible with the previously published complete sequence of calcitonin mRNA from rat was sequenced. This cDNA insert contained the entire coding region of 426 bp, 70 bp at the 5′-end, and 295 bp upstream from the poly(A) tail. The complete amino acid sequence of human preprocalcitonin could thus be deduced.

Preprocalcitonin *mRNA* *Medullary carcinoma of the thyroid* *Cloning* *Complete sequence* *N-terminal peptide*

1. INTRODUCTION

Calcitonin (CT), the hypocalcemic hypophosphatemic hormone [1,2], is the specific marker for medullary carcinoma of the thyroid (MCT) [3] a tumor existing either in sporadic or familial (autosomal dominant) form. The hormone is also produced ectopically by a large number of tumors thought to arise from cells sharing either common histochemical characteristics, the Apud concept [4], or the same embryological origin, the neural crest [5]. Establishing the structure of the calcitonin gene in normal and pathological conditions and producing specific probes are of importance for studying the different forms of MCT, sporadic or hereditary, and in demonstrating the ectopic production of the hormone in other types of cancer. The presence of a prohormone with an approximate M_r of 11800 has been reported [6]. Translation of mRNA extracted from MCT in cell-free systems has resulted in the synthesis of primary translation products of similar M_r, 15000 [7] and 14500 [8], or in products with higher M_r, 21000 [9]. An M_r of 55000 was initially reported in [10] using an intact cell for translation, the frog oocyte. A partial nucleotide sequence of human CT mRNA, extracted from a case of familial MCT and coding for a 21 kDa translation product, has recently been reported [11]. We report here the complete nucleotide sequence of cDNA to human calcitonin mRNA extracted from a case of sporadic MCT. This mRNA directs the translation of 14.5 kDa preprocalcitonin.

+ To whom correspondence should be addressed

Abbreviations: CT, calcitonin; mRNA, messenger RNA; MCT, medullary carcinoma of the thyroid; cDNA, complementary DNA; sscDNA, single-stranded cDNA; dscDNA, double-stranded cDNA; bp, base pairs

Published by Elsevier Science Publishers B.V.

2. MATERIALS AND METHODS

2.1. *Extraction of poly(A)-rich RNA*

Total cytoplasmic RNA was extracted from thyroid glands removed from several patients suffering from medullary carcinoma of the thyroid (MCT) using phenol chloroform extraction. Poly(A)-rich RNA was isolated by chromatography on oligo(dT) cellulose.

2.2. *Synthesis of dscDNA*

Total poly(A) RNA was used as a template for reverse transcriptase for the synthesis of sscDNA. After alkaline hydrolysis the second strand was synthesized using the same enzyme. The hairpin structure was removed by the action of S1 nuclease. After size selection of dscDNA by ultracentrifugation, dC tails were added using terminal deoxynucleotidyl transferase. pBR 322 was linearised by the action of *Pst*I and tailed with dG. dscDNA were annealed to poly(dC)-tailed pBR 322 and the resulting chimeric plasmids used to transform *E. coli* strain DP50 under category II physical containment conditions. Ampicillin sensitive clones were screened using a ^{32}P-labelled cDNA to poly(A) RNA extracted from another case of MCT rich in calcitonin (20% abundance of CT mRNA as estimated by translation and specific immunoprecipitation). Colonies showing a positive response were further screened with a poly(T) ^{32}P-labelled probe. Plasmids obtained from colonies giving a positive response with both probes were screened using positive hybridization selection [12].

A clone harboring a plasmid containing an insert of approximately 800 bp and specifically retaining CT mRNA (fig.1) was subjected to restriction analysis in order to compare its digestion pattern with that of sequenced rat CT cDNA. We have designated this plasmid 3271.

2.3. *Determination of nucleotide sequence*

Restriction fragments suitable for sequence analysis were generated by digestion with the restriction endonucleases *Ava*II, *Bgl*II, *Cfo*I, *Dde*I, *Eco*RII, *Hin*fI, *Hpa*II, *Pst*I, *Rsa*I, *Sac*I, *Sph*I. They were either 5′-end labelled by

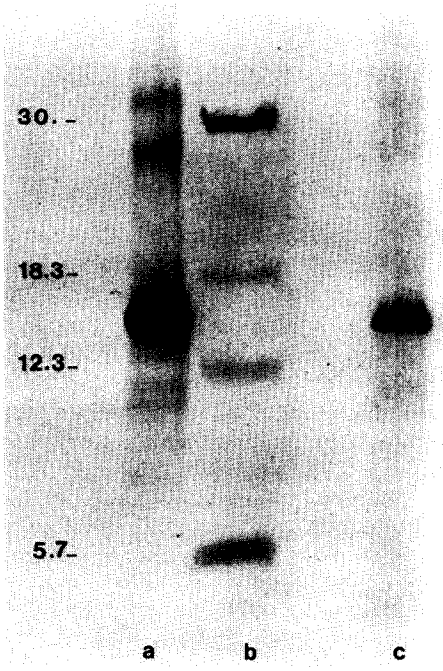

Fig. 1. Hybridization selection of calcitonin mRNA by cloned cDNA. Recombinant plasmid DNA's were cleaved with *Pst*I, and the inserts were purified by polyacrylamide gel electrophoresis and recovered by electroelution. DNA was immobilized on nitrocellulose filters and hybridized to mRNA. mRNA selectively retained was eluted from the filters and translated in a reticulocyte cell-free system. Labelled translation products were separated by SDS-polyacrylamide gel electrophoresis and detected by fluorography performed for 3 days at −80°C. Lane a, total translation products immunoprecipitated with anti-human calcitonin antiserum. Lane b, C-labelled M_r markers, from top to bottom: M_r 30000, carbonic anhydrase; 18367, lactoglobulin A; 12300, cytochrome *c*; 5700, insulin. Lane C, eluted and translated mRNA hybridized to DNA from plasmid 3271.

→

Fig. 2. Nucleotide sequence of human calcitonin precursor mRNA cloned in plasmid 3271. The figure gives the entire nucleotide sequence of human calcitonin precursor and its 3′- and 5′-flanking regions. The amino acid sequence deduced from the nucleotide sequence is indicated above. Below are listed nucleotides and corresponding amino acids of rat preprocalcitonin [15] which differ from the human sequence. Homology between the two sequences was optimized by introducing a limited number of gaps (– – –). Numbers above the human protein sequence refer to amino acid positions: −1 to −84 amino terminal cryptic peptide and leader sequence; 1 to 32 calcitonin; 1 to 25 carboxy terminal cryptic peptide.

```
        5'-CTCTGGCTGG ACGCCGCCGC CGCCGCTGCC ACCGCCTCTG ATCCAAGCCA CCTCCCGCCA
           --ACCAACC  C TGGCT CA TCAG ACC   G A T   A  C      T   T G T A
                                  -80                                -70
                    Met Gly Phe Gln Lys Phe Ser Pro Phe Leu Ala Leu Ser Ile Leu Val
GAGAGGTGTC ATG GGC TTC CAA AAG TTC TCC CCC TTC CTG GCT CTC AGC ATC TTG GTC
 G    CA              T  TG                T          T  G               C
                       Leu                             Val Val            Leu
                                      -60                                 -50
Leu Leu Gln Ala Gly Ser Leu His Ala Ala Pro Phe Arg Ser Ala Leu Glu Ser Ser
CTG TTG CAG GCA GGC AGC CTC CAT GCA GCA CCA TTC AGG TCT GCC CTG GAG AGC AGC
     AC          T  G         G     TT  T  G       A A  T A  A
    Tyr          Cys Gly      Gln    Val     Leu         Thr
                                      -40
Pro Ala Asp Pro Ala Thr Leu Ser Glu Asp Glu Ala Arg Leu Leu Leu Ala Ala Leu
CCA GCA GAC CCG GCC ACG CTC AGT GAG GAC GAA GCG CGC CTC CTG CTG GCT GCA CTG
     GC ATG ---      T           A   A       T       A     ---
    Gly Met ---                      Glu                   ---
-30                                  -20
Val Gln Asp Tyr Val Gln Met Lys Ala Ser Glu Leu Glu Gln Glu --- --- Gln Glu
GTC CAG GAC TAT GTG CAG ATG AAG GCC AGT GAG CTG GAG CAG GAG --- --- CAA GAG
        A       A            A  T   G                       GAG GAA   G
        Asn     Met             Val Arg                     Glu Glu

              -10                                  -1   +1
Arg Glu Gly Ser Ser Leu Asp Ser Pro Arg Ser Lys Arg Cys Gly Asn Leu Ser Thr
AGA GAG GGC TCC AGC CTG GAC AGC CCC AGA TCT AAG CGG TGC GGT AAT CTG AGT ACT
GCT           T     T                                 T   G               C
               10                                      20
Cys Met Leu Gly Thr Tyr Thr Gln Asp Phe Asn Lys Phe His Thr Phe Pro Gln Thr
TGC ATG CTG GGC ACA TAC ACG CAG GAC TTC AAC AAG TTT CAC ACG TTC CCC CAA ACT
                  G       A  A     C                   C
                                   Leu
                     30          +1                              +10
Ala Ile Gly Val Gly Ala Pro Gly Lys Lys Arg Asp Met Ser Ser Asp Leu Glu Arg
GCA ATT GGG GTT GGA GCA CCT GGA AAG AAA AGG GAT ATG TCC AGC GAC TTG GAG AGA
T                           C                       G   AG              C
Ser                                                 Ala Ser             Thr
                                   +20
Asp His Arg Pro His Val Ser Met Pro Gln Asn Ala Asn Stop
GAC CAT CGC CCT CAT GTT AGC ATG CCC CAG AAT GCC AAC TAA  ACTCCTCCCT TTCCTTCCTA
A    C  A    C T   --- --- --- --- --- TT   G       G  -G   CT    C     T C
Asn     His     Tyr --- --- --- --- --- Phe Gly

ATTTCCCTTC TTGCATCCTT CCTATAACTT GATCCATGTG GTTTGGTTCC TCTCTGGTGG CTCTTTGGGC
G     A        T T                          A    -----          CT  T   C

TGGTATTGGT GGCTTTCCTT GTGGCAGAGG ATGTCTCAAA CTTCAGATGG GAGGAAAGAG AGCAGGACTC
 AT   G    TA      A   A    A       GTATCTG GAATCTCCAA TGGGTG  GA GAA   3'

ACAGGTTGGA AGAGAATCAC CTGGGAAAAT ACCAGAAAAT GAGGGCCGCT TTGAGTCCCC CAGAGATGTC

ATCAGAGCTC CTCTGTCCTG CTTCTGAATG TGCTGATCAT TTGAGGAATA AAATTATTTT TCCCC

poly A(28)-3'.
```

polynucleotide kinase with [γ-^{32}P]ATP or 3′-end labelled by terminal transferase with 3′-[^{32}P]cordycepin 5′-triphosphate using a commercial kit (NEN). After cleavage with a second restriction enzyme or strand separation, fragments were separated by gel electrophoresis and submitted to the chemical sequencing reaction in [13].

3. RESULTS AND DISCUSSION

We report here what is to our knowledge the first complete sequence of the primary translation product of human calcitonin mRNA, a polypeptide of 141 amino acids. In addition to the 426 bp of the coding region, plasmid 3271 contains 70 bp in the 5′- and 295 bp in the 3′-untranslated regions (fig.2). Almost total agreement in nucleotide sequence was found between this sequence derived from mRNA extracted from patients with sporadic MCT and the partial sequence, coding for 93 amino acids but lacking the N-terminal region, reported in [11] for mRNA extracted from a case of hereditary MCT. The only differences noted were a substitution of a single nucleotide in the triplet coding for leucine (position −36), where we found CTC instead of GTC and consequently leucine instead of valine, and the absence of one adenine after position 630 in the untranslated region. Whether these discrete differences are due to genetic polymorphism or to spontaneous mutations occurring in the course of the malignant disease is of potential interest.

The M_r of the precursor of human calcitonin which we deduce from the amino acid sequence, 15461, is consistent with the M_r of the primary translation products of human CT mRNA reported by us [8] and in [7], 14500 and 15000, respectively. It is, however, much lower than the relative M_r, 21000, found in [9] after translation of mRNA extracted from a case of familial MCT. Since it was this familial MCT mRNA which was used to establish the partial nucleotide sequence of human calcitonin [11], the apparent contradiction in protein M_r could be due to differences in migration rate of the precursors in polyacrylamide gels. The reported partial sequence [11] is almost identical to ours; thus if differences in the proteins exist, they should be located in the N-terminal region.

Extensive homology in nucleotide sequence in the translated region, and consequently in the amino acid sequence, is found with rat CT precursor [14,15]. However, human preprocalcitonin is longer than that of rat by 5 amino acids which are located 3 amino acids before the stop codon [11]. This C-terminal peptide also contains 8 amino acid substitutions. In both species the N-terminal amino acid extensions have the same length (84 amino acids) but show 21 differences in amino acid sequence.

Leader sequences of human and murine preprocalcitonin seem to be of similar length since the potential amino acid cleavage site, alanine-valine, in murine preprocalcitonin [14] corresponds to alanine-alanine in the human molecule. The amino acid substitutions observed concern amino acids having similar side functions, with the exceptions of leucine for tyrosine and glycine for cysteine. In the N-terminal cryptic peptide also, substitutions involve amino acids with similar side functions. However, the N-terminal region of human precalcitonin is poorer in methionine (−2 residues) and richer in proline (+1 residue). The total number of acidic amino acids is equal in the two molecules. Therefore, it would seem that the cryptic N-terminal peptide has been more highly conserved than the C-terminal cryptic peptide during evolution.

As for the untranslated regions, a higher frequency of nucleotide substitutions is found in both 3′- and 5′-flanking regions as compared with the coding region. This difference could be due simply to species differences especially in the case of the 3′-region. In the 5′-region, however, important differences are found not only between human and murine mRNA, but also intraspecies [14,15]. As no other sequence of the 5′-untranslated region of the human calcitonin mRNA is known, the presence or absence of polymorphism and its eventual consequences are still to be determined. Screening of a genomic bank with the complete cDNA probe and determining the structure of the calcitonin gene (in progress) will determine the sequence of the precursor of calcitonin in normal tissues as well as eventual modifications of the 5′- and 3′-non-coding regions.

Furthermore, elucidation of the complete sequence of preprocalcitonin will allow development of a region-specific radioimmunoassay for detec-

ting the N-terminal region of the molecule, as has already been achieved for the C-terminal region [16]. Such an assay will help in elucidating the steps in biosynthesis of the hormone and the role of the N-terminal region in this process.

ACKNOWLEDGEMENTS

We thank Miss M.O. Delaunay, Miss N. Segond, and Miss A.M. Pascaud for expert technical assistance; Dr C. Calmettes for the histological examination of the tissues; and Dr C. Desplan for pertinent criticism and advice. This work was supported by INSERM grant 813032. Biohazards associated with this work were previously examined by the French National Committee.

REFERENCES

[1] Copp, D.H., Cameron, E.C. Cheney, B.A., Davidson, A.G.F. and Henze, K.G. (1962) Endocrinology 70, 638–649.
[2] Hirsch, P.F., Gauthier, G.F. and Munson, P.L. (1963) Endocrinology 73, 244–252.
[3] Milhaud, G., Tubiana, M., Parmentier, C. and Coutris, G. (1968) C.R. Acad. Sci. Paris 266, 608–610.
[4] Pearse, A.G.E. (1968) Proc. Roy. Soc. Lond. B 170, 71–80.
[5] Le Douarin, N. and Lelievre, C. (1970) C.R. Acad. Sci. Paris, 270, 2857–2860.
[6] Julliene, A., Segond, N., Calmettes, C., Moukhtar, M.S. and Milhaud, G. (1980) Biochem. Biophys. Res. Commun. 95, 932–937.
[7] Goodman, R.H., Jacobs, J.W. and Habener, J.F. (1979) Biochem. Biophys. Res. Commun. 91, 932–938.
[8] Desplan, C., Benicourt, C., Jullienne, A., Segond, N., Calmettes, C., Moukhtar, M.S. and Milhaud, G. (1980) FEBS Lett. 117, 89–92.
[9] Allison, J., Hall, L., MacIntyre, I. and Craig, R.K. (1981) Biochem. J. 199, 725–731.
[10] Lips, C.J.M., Van der Sluys Veer, J., Van der Donk, J.R., Van Dam, R.H., and Hackeng, W.H.L. (1978) Lancet I, 16–18.
[11] Craig, R.K., Hall, L., Edbrooke, M.R., Allison, J. and MacIntyre I. (1982) Nature 295, 345–347.
[12] Ricciardi, R.P., Miller, J.S. and Roberts B.E. (1979) Proc. Natl. Acad. Sci. USA 76, 4927–4931.
[13] Maxam, A. and Gilbert, W. (1977) Proc. Natl. Acad. Sci. USA 74, 560–564.
[14] Jacobs, J.W., Goodman, R.H., Chin, W.W., Dee, P.C., Habener, J.F., Bell, N.H. and Potts J.T. (1981) Science 213, 457–459.
[15] Amara, S.G., Jonas, V., Rosenfeld, M.G., Ong, E.S. and Evans, R.M. (1982) Nature 298, 240–244.
[16] MacIntyre, I., Hillyard, C.J., Murphy, P.K., Reynolds, J.J., Gaines-Das, R.E. and Craig, R.K. (1982) Nature 300, 460–462.