

## Cloning of LL5, a novel protein encoding cDNA from a rat pituitary library

Linet Levi, Israel Hanukoglu \*, Moshe Raikhinstein, Fortune Kohen  
and Yitzhak Koch

*Department of Hormone Research, Weizmann Institute of Science, Rehovot 76100 (Israel)*

(Received 28 July 1993)

**Key words:** Cloning; cDNA; Single copy gene; Gonadotropin-releasing hormone; Anti-idiotypic antibody; Novel protein;  
(Rat pituitary gland)

While screening a rat pituitary cDNA library for a peptide hormone receptor, we identified a cDNA that represents a novel gene. The 3.8 kb cDNA has an open reading frame of 2.3 kb encoding a protein of 781 amino acids ( $M_r = 87\,507$ ). Southern blot analysis of rat, mouse, bovine and human genomic DNAs revealed that a homologous gene is present in these species probably in a single copy. Northern blot analysis showed that in addition to the pituitary gland, the gene is also expressed in other rat tissues. Scanning of DNA and protein databanks revealed no significant homology to any other sequence. Thus, this gene encodes a heretofore unidentified protein.

The mammalian genome is estimated to contain 50 000 to 100 000 genes. However, currently the sequences of only a few thousand genes are known. The relative merits of genomic versus cDNA sequence analysis is a debated issue in the context of the Human Genome Project [1]. In the absence of cDNA sequence information, the prediction of transcribed regions of genomic sequence is wrought with difficulties [2]. Therefore, the determination of the sequences of novel proteins remains an important task in the analysis of the mammalian genome. In the course of screening a rat pituitary cDNA expression library we isolated a novel cDNA encoding a large protein and characterized its expression patterns. We report here the full predicted sequence of this protein as a novel addition to the protein sequence databases.

A rat pituitary  $\lambda$ gt11 cDNA expression library was screened with an anti-idiotypic antibody raised against an antibody to gonadotropin releasing hormone (GnRH) [3]. One clone that gave unequivocally strong reaction was plaque purified by repeated screenings with the antibody, and the  $\lambda$  DNA was isolated as previously described [4]. A cDNA insert of 3.8 kb was excised by *EcoRI* digestion and subcloned into pBlue-script KS (Stratagene). The entire sequence of the

cDNA (designated LL5) was determined by the dideoxy chain termination method using oligonucleotide primers. Both strands of the cDNA were sequenced by an automated DNA sequencer (Applied Biosystems).

The largest open reading frame of the cDNA was 2343 bp followed by a 3'-non coding region of 1250 bp ending with a poly(A)<sup>+</sup> tract (Fig. 1). This reading frame encoded a predicted protein of 781 amino acids ( $M_r = 87\,507$ ). Analysis of the amino acid sequence did not reveal highly hydrophobic segments that are characteristic of G-protein coupled receptors. Thus, the possibility that the cDNA encoded a membrane receptor for GnRH was eliminated. However, LL5 sequence and the rat GnRH receptor [5] showed identity at two sites of three residues (LEQ and PLT) at the amino-terminal domain of the receptor. This domain probably includes the hormone binding site, and the short sequences of identity may represent a common epitope recognized by the anti-idiotypic antibody.

A search of the GenBank, EMBL and SwissProt databases using the FASTA program did not reveal significant similarity to any known protein or DNA sequence in these databases. However, LL5 sequence showed 82% similarity to a partial cDNA sequence of only 330 bp (GenBank accession No. EST00107) isolated from human brain [1]. LL5 probably represents the full coding length of the rat homolog of this cDNA fragment. The conserved sequence motifs found in the predicted LL5 protein sequence include two *N*-glycosylation sites (at amino acid 413 and 551), a single cAMP phosphorylation site (at 95), a tyrosine phospho-

\* Corresponding author. Fax: +972 8 344116.

The nucleotide sequence data reported in this paper have been submitted to the EMBL/GenBank and DDBJ Nucleotide Sequence Databases under the accession number X74226 as *R. norvegicus* LL5 mRNA.

ylation site (at 746), and many protein kinase C phosphorylation sites.

To examine the presence of homologous genes in the genomic DNA of several other species we carried out Southern blot analyses. For this purpose the cDNA insert was purified by electroelution, labeled by random priming and then reacted with the blots. Southern

blot analyses revealed the presence of a homologous gene in all mammalian species examined (Fig. 2). The sizes of the hybridizing fragments in all species varied between 2 and 20 kb. Considering that the cDNA represents an mRNA of about 4 kb, the Southern blot results suggest that the hybridizing fragments probably represent a single gene with introns.

```

                                CCTGACACTAGGAGCAGGGGCCGAGGACTC 32
GGAGTCTCTCGCCCACTCGGGGAGTCTCTGGCACCTCGCAAGGGCAGCTTCAGTGGCAGGCTAAGCCAGCCTACAGTCTGGGCTCTCTTACTGGGGCTTACCTCGCCAGAGCCCC 152
ATGCTCAAAGGAAGCTTTCCAGTGGGGACTTGCGGTGCCCATCCCAAGGGAGAGGAAAAATAGCATCACGGAGATCAGCGACAATGAGGACGATCTCTGGAGTACCACAGCGGCAGC 272
M L K G S F P V G T C G C P S Q G R G K I A S R R S A T M R T I S W S T T G G S 40
GCCAAGAGCGGCTCCGGGAGCAGGAGATGGAGAAGGTGAACGCCAGCGCTGGAGACCCTCTGAACCTGTGTGCTGAGTACAGCCGAGCTGACGGGGACCTGAGACTGGGGAACGTG 392
A K S G S G S R R W R R L E R Q R L E T I L N L C A E Y S R A D G G P E T G E L 80
CCGAGTATAGGAGAGGCCACGGCAGCACTGGCTTGGCAGGCAGGAGCCCTCAAGAGGCTTGGCAGGAGCCATAGTAGTGTCTGGAAGGAGTGGCGAGGAGTCTGGAGGTGCCTCGCAG 512
P S I G E A T A A L A L A G R R P S R G L A G A I V V S G R S G E E S G G A S Q 120
CGCCTGTGGAGAGCATGGAGCGCTCTGATGAAGAGAATCTCAAAGAGGAGTGCAGCAGCACAGAGACCCAGCAGGAGCAGCAAGATGCTCTAGCAGCAAGCTCAAAGGAGAGGTG 632
R L W E S M E R S D E E N L K E E C S S T E S T Q E H E D A P S T K L Q G E V 160
CTGGCTGTGGAAGAGGAGCGGGCTCAGGTTCTGGGGCCGCTAGAGCAGCTGAAGGTCCGAGTGAAGGAGCTGGAACAGCAGCTGCAGGAGGAGCAGCCGAGAGGCTGAAATGGAGAGAGCA 752
L A V E E E R A Q V L T G R V E Q L K V R V K E L E Q Q L Q E A A R E A E M E R A 200
CTGTGCAGGGGAGAGGGAGCGGAGCGAGCATTGCTGCAGAAGGAACAAAGAGCTATGGACCAGCTGCAGGAGAAGTTGGTGGCCTTGGAGACAGGCATCCAGAAGGAGAGGGACAAG 872
L L Q G E R E A E R A L L Q K E Q R A M D Q L Q E K L V A L E T G I Q K E R D K 240
GAGGCAGACGCTCTAGAGACTGAGACAAAGCTCTTTGAGGACCTGGAGTTCAGCAGCTGGAGCGGAGAGTGGCTGGAGGAGGAGCGGGAGTTGGCCGCCCAAGGGCTGCTTCGCAGC 992
E A D A L E T E T K L F E D L E F Q Q L E R E S R V E E E R E L A G Q G L L R S 280
AAGGCTGAGTGTGTCGCGACGCTGAGCAAGAGAAAGGAGCGCTAGCTGTACTGGACAGTACGCTGGGAGATCCGGGCCCAAGCAGTACAGGAGTACAGCGCTCTGGCCCGGGAAAAG 1112
K A E L L R S V S K R K E R L A V L D S Q A G Q I R A Q A V Q E S E R L A R E K 320
AATGCTGTCTGCAGTACTGCAAAGGAGAAGGAGAAGCTGACTGTGCTTGAGAGGAGATATCATTGCTCACGGGTGGCAGGCCTTTTCCGAAGACCACTCAACCTCAAAGAGGTT 1232
N A V L Q L L Q K E K E K L T V L E R R Y H S L T G G R P F P K T T S T L K E V 360
TACCGCTCAAAGATGAATGGTGACATGGCCAGCCCTGCCAGGACCCGCTAGTGGCCCTCTCCCTCCTCAGGCTCCTCATCTCTTCTTACAGCTCAGCGCTGGCTACCTTGGCC 1352
Y R S K M N G D M A S P L P R T R S G P L P S S S G S S S Q L S V A T L G 400
CGTAGTCTCTCCCAAAGAGTGCCCTGCTCGCCAGAATGGCACCAGTAGTCTCCCTCGAAACCTGGCAGCCAGCTGCAGGACATTGAGACCAAGCGCCAGCTGGCCCTACAGCAAAAG 1472
R S P S P K S A L L A Q N G T S S L P R N L A A T L Q D I E T K R Q L A L Q Q K 440
GTGAGTTACCTCTGCCAGCCCCCTCTACCCGAAGACCCAGCAGGCGACCAAGTGATCGAGGAGCAACGGCGACGGCTGGCTGAGCTGAAGCAGAAAGCAGCGGAGAGGCTCAGTGC 1592
V E L P P A E P L S P E D P A G H Q V I E E Q R R R L A E L K Q K A A A E A Q C 480
CAGTGGGATGCTTGCATGGTGACGCGCCTTTCCAGCAGGCGCTGCTGGCTTCCCCACGCTCATGCACCATTCATCTCTGCACCACCTGCCAGCAGGAGGAACTGGAGAGGAGGGC 1712
Q W D A L H G A A A F P A G P S G F P T L M H H S I L H H L P A G R E R G E E G 520
GAGCACGCTTACGACACCCCTGAGCCTGGAGAGTTCCGACAGCATGGAGACCAGCATCTCTACGGGGGGCAACTCCGCTGTTCCCTGACAACATGTCCAGCGCTAGTGGTCTGGACATG 1832
E H A Y D T L S L E S S D S M E T S I S T G G N S A C S P D N M S S A S G L D M 560
GGCAAGATAGAGAAATGGAGAAGATGCTAAAGGAAGCTCACGGGAGAGAGCGGGCTCATGGAGTCCAGGAGCGGGAGATGGAGCTACGGAGGCGAGCCCTGGAGGAGGAACGTCGG 1952
G K I E E M E K A H A E K S R L M E S R E R E M E L R R E L R E E E R R 600
AGGCGGGAGCAAGTGAGCGGAGGCTGCAGAGTGAGAGCGCCCGCAGGCAGCAGCTGGTGGAGAAAGAGTCAAGCTGAGGGAGAAGCAGTTTTCCAGGCTCGACCTCTGACACGTTAT 2072
R R E Q V E R R L Q S E S A R R Q Q L V E K E V K L R E K Q F S Q A R P L T R Y 640
CTTCCCAACCGAAGGAGGACTTTGATTGAAGACCCACATAGAGTCGTAGGCCATGGGGTGGACACCTGCCTACATGTGGTGTCTAGCAGCAAGGTCTGCCGTGGCTACTTAATCAAG 2192
L P N R K E D F D L K T H I E S S G H G V D T C L H V V L S S K V C R G Y L I K 680
ATGGGTGGCAAGATTAATCGTGAAGAAGCGCTGGTTGTCTTTGACCGGCTGAAGCGCACCCCTTCTCTACTATGTGGACAAACACGAGACGAAGCTGAAGGGGGTCTATCTACTCCAG 2312
M G G K I K S W K K R W F V F D R L K R T L S Y Y V D K H E T K L K G V I Y F Q 720
GCCATCGAGGAAGTGTACTATGACCACCTGCGCAGCGCAGCCAAGATCCGAATCCAGCCCTCACCTTCTGTGTGAAGACCATGACCGGCTGTACTACATGGTGGCCCATCTGCAGAG 2432
A I E E V Y Y D H L R S A A K S P N P A L T F C V K T H D R L Y Y M V A P S A E 760
GCCATGCGCATCTGGATGGATGTCATCGTCAGGGAGCTGAGGGCTACACTCAATTCATACTAACCAGGCACCTCTGGGTCCCCCAGGTGAGCCCAAGGACCATGCCCTGCC 2552
A M R I W M D V I V T G A E G Y T Q F M N * 781
ACTTGCTGCTGCTGCTTGTCCCTGGAGACAGAGGCACACTGGGCACGGGCCACAAAGGGTCAGGATGTAGTTAGTACCCAGGGCTTTTGTGCAAGAAAACGAAATTGTTCTATGAGGAG 2672
TTTGGTGTCTGGGAATCAGGCTCCGGTCCCTAAAGAGCCAGCAGAGAGGAAGGAAGGGAGAGGCTTGTGCGCTCCTGGAGTCCAGAAGTGCAGGATTCCTCTGGGTCCAGCAG 2792
GCTGAAGAGGCATAGCTGAGCTGTGACTCTCAGCTCTGTCTGGTACCACCTTAGCCAGCTTACTTCTCTGTGTAAGGACAAATATGGTACCAGAACTGCCAAAGATCCCTTTTC 2912
TCATCTATCCCTCTACAGGGGTCTGGGGGTCTGAGCAGAGCCAGCTGTAGAAGAGGGAGAGGCTCAGGCTGCACCTCCCGCCCCGGTCTGCTCTCTCTCTCCCTCCACCTCCCA 3032
CTTTTCTACTACTTAGCTCAAGGACACAGACCTCAAGTGTCTTCTCGAGGTCCAGCCCGTAATCATGTTACAAGCCCATCATCTTGTCTCCATGGTGGCTGCTTCTCACCATTGGT 3152
ACACACTAATTGCCATGGCCACTCTGTGGCTCTGCCACTGCTTGGCTGTGGGCCACATGAGGGTACATGCCACCATCTCTCCAACCCAGGCTGGGGGCTGTCAATGGTGGGAGG 3272
AGATAGAGTTCTCCCGGCTGCAGTGTGTCTCCAGCCCCAGCCAGCCCTCTCCCTTTTACTGTGCTTGTAGAGCCAGAAGGATGATGTCCAGAGATCCAAGACAGAGAGCAGC 3392
ATCACTGAGGGGAGAGGAAGCTAAGAGGTCTGCCCTCCCTTCTGACAGGACTAGTTCCTAATAAGGAACGAGGGTGGTAGCTCAGAGTGTCTGGAAAGAGCCTGAGTTGGGAGCAGGG 3512
GGGGCAGAGCAGAGGCCCTGAGCTGTGGTAACCTCCCTTCAAGTTTCCCAAGGTGACGGGAAAGAGGACAGAGGATCAAGTCTCCAGCTGGCTGGCTCAGCCTTGGCCTTAAACAT 3632
AAGCCACCTTCCCTGCCCTCCCAAGCACTGGGCCCTTGGTGTGGGCTGGCTGGGTGTTTGCAGTATTTGAAGCATTTACGACAGAACAATAAAGCATTTTGACTATGTAACAAAAA 3752
AAAAAAAAAAAAA 3765

```

Fig. 1. Sequence of the LL5 cDNA insert and the predicted protein encoded by the largest open reading frame. The plasmid pLL5 sequence also includes *NorI* and *EcoRI* linker sequences at both 5' and 3' ends of the sequence shown in the figure. Polyadenylation signal is underlined.

Since the cDNA was isolated from a rat pituitary cDNA library we wanted to determine whether the gene is also expressed in other tissues. In Northern blot analysis an mRNA of 5.9 kb was observed in all tissues examined (Fig. 3). Only in the pituitary an additional transcript of 3.2 kb was also present which may suggest a differential regulation. The function of the protein encoded by LL5 is yet unknown. However, its expression in a variety of tissues suggests a general cellular function. Further studies are required to express this protein and to characterize its functional properties.

We are grateful to Dr. Y. Amir-Zaltsman for the

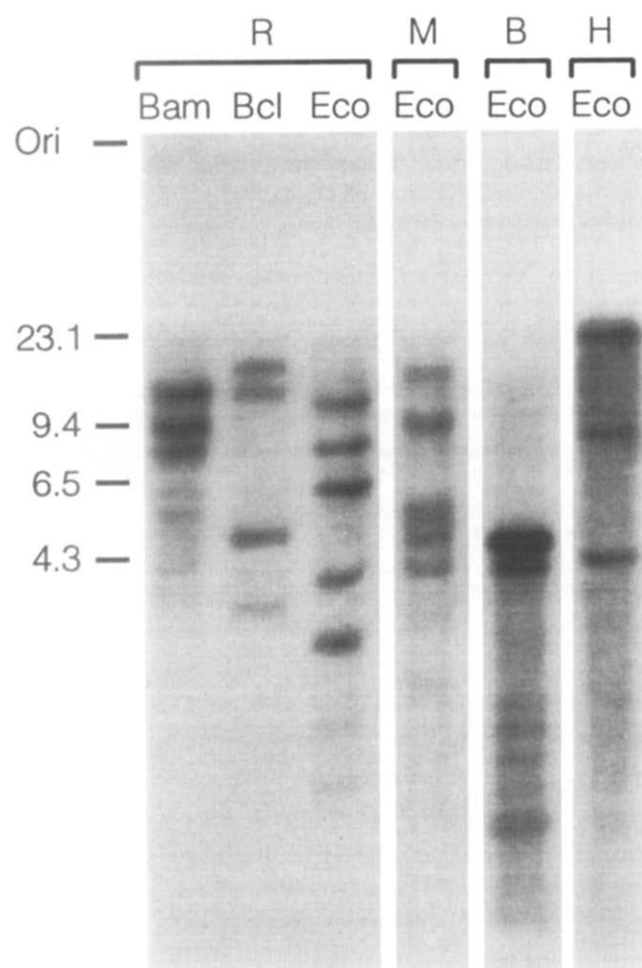


Fig. 2. Hybridization of LL5 cDNA to rat (R), mouse (M), bovine (B) and human (H) genomic DNAs. Each lane of the gel contained 8  $\mu$ g of genomic DNA digested with *Bam*HI, *Bcl*I, or *Eco*RI, enzymes that did not have a recognition site in the LL5 cDNA. The DNAs were blotted onto GeneScreen Plus membrane and hybridized with the radiolabelled cDNA as described [6]. Washes: 2 $\times$ SSC, 20°C, 10 min; 2 $\times$ SSC, 1% SDS, 42°C, 30 min; 0.2 $\times$ SSC, 1% SDS, 42°C, 30 min.

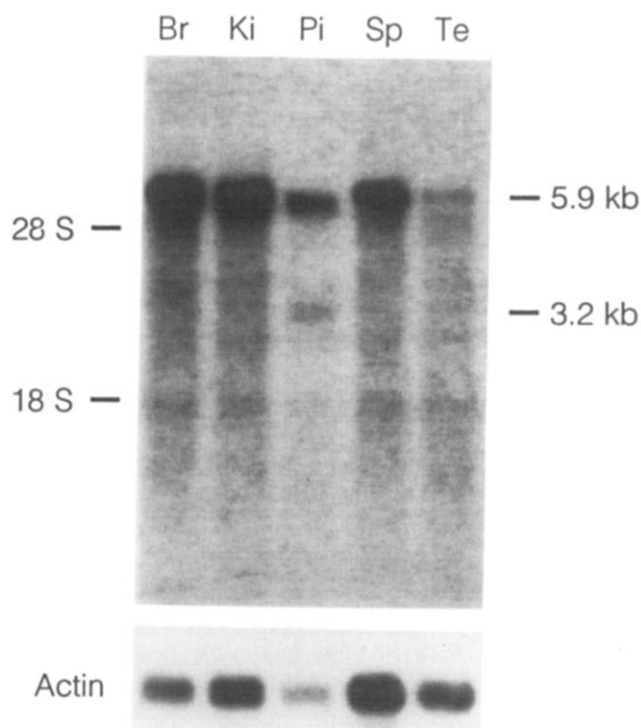


Fig. 3. Hybridization of LL5 cDNA to poly(A)<sup>+</sup> RNA from rat brain, kidney, pituitary, spleen, and testis. Total RNA was prepared from the indicated tissues using guanidinium thiocyanate [7]. Poly(A)<sup>+</sup> RNA was isolated using Dynabeads Oligo(dT)25. Each lane of the gel contained 4  $\mu$ g of poly(A)<sup>+</sup> RNA. The RNAs were blotted onto GeneScreen membrane and hybridized with the radiolabelled cDNA as described [6]. After stripping the first probe, the same blot was also reacted with an actin cDNA probe as a control. Note that the relative intensity of staining with the LL5 cDNA is similar to that of the actin probe.

preparation of the anti-idiotypic antibodies. This research was supported in part by grants from The Center for Neurosciences and The Leo and Julia Forchheimer Center for Molecular Genetics at the Weizmann Institute of Science.

- 1 Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merril, C.R., Wu, A., Olde, B., Moreno, R.F., Kerlavage, A.R., McCombie, W.R. and Venter, J.C. (1991) *Science* 252, 1651–1656.
- 2 Fickett, J.W. (1982) *Nucleic Acids Res.* 10, 5303–5318.
- 3 Amir-Zaltsman, Y., Ausher, Y., Gayer, B., Lichter, S., Serour, F., Birkenfeld, S. and Kohen, F. (1990) *J. Endocrinol.* 124, 261–268.
- 4 Raikhinstein, M. and Hanukoglu, I. (1992) *Methods Mol. Cell. Biol.* 3, 108–111.
- 5 Kaiser, U.B., Zhao, D., Cardona, G.R. and Chin, W.W. (1992) *Biochem. Biophys. Res. Commun.* 189, 1645–1652.
- 6 Hanukoglu, I. and Gutfinger, T. (1989) *Eur. J. Biochem.* 180, 479–484.
- 7 Chomczynski, P. and Sacchi, N. (1987) *Anal. Biochem.* 162, 156–159.