Biochimica et Biophysica Acta, 1129 (1992) 239-242
© 1992 Elsevier Science Publishers B.V. All rights reserved 0167-4781/92/$05.00

BBAEXP 90302

Short Sequence-Paper

Organization and nucleotide sequence of carp gonadotropin α subunit genes

Chang-Jen Huang [1,2], Fore-Lien Huang [1,3], Yi-Chi Wang [1], Yea-Sha Chang [1,2] and Tung-Bin Lo [1,2]

[1] *Institute of Biological Chemistry, Academia Sinica*, [2] *Graduate Institute of Biochemical Sciences and* [3] *Department of Zoology, National Taiwan University, Taipei (Taiwan)*

(Received 1 November 1991)

Key words: Recombinant DNA; Intron-exon boundary; Molecular evolution; Glycoprotein hormone; (*Cyprinus carpio*)

We have used PCR to amplify and align the sequence of two genes encoding cGTHα. Both genes comprise four exons and three introns. The organization of cGTHα genes is very similar to that of mammalian GTHα genes. However, the cGTHα genes only span a region of 1.2 kb which is much smaller than those mammalian GTHα genes.

Lutropin (LH), follitropin (FSH) and thyrotropin (TSH) are pituitary glycoprotein hormones [1]. The former two are designated as gonadotropin (GTH) because they act on the testes and ovaries to stimulate steroid synthesis, gametogenesis, spermiation and ovulation. Each glycoprotein hormone is dimeric, containing a common α subunit and a hormone-specific β subunit. The hormonal activity is expressed only after noncovalent association of these two subunits.

The amino acid sequences of the α subunits of mammalian GTH from several species [2] and of non-mammalian GTH including several fish species [3-10] are known. Homology analysis reveals that α subunits are highly conserved among all species (> 66% identity). Whether such conservancy of the primary structure of α subunits is also reflected on the gene level remains to be investigated. Up to now, only a single α subunit and α gene had been reported in mammals [11-17]. However, at least two species of fish have more than one α subunit. In the salmon pituitary glands, there are two distinct α subunits with 72% identity in the amino acid sequence and both are active [7-9]. In the carp, there are two extremely similar α cDNAs (about 96% identity and designated as cGTHα1 and cGTHα2) which encode respectively an active (α1) and a biologically inactive (α2) subunit [3,18]. These facts imply that fish may have multiple α genes. In order to understand the evolution of the GTHα genes and the possibility of multiplicity of the cGTHα gene, we investigated the genomic structure of the α gene of carp (*Cyprinus carpio*).

Using SuperCos 1 [19] as cosmid vector, a carp genomic library was constructed which contained $1 \cdot 10^5$ colonies. When both cGTHα1 and cGTHα2 cDNAs were used as probes, 16 positive cosmid clones were obtained (data not shown). The size of these clones was only about 20 kb, much smaller than those of usual cosmid clones (40-50 kb). Through tedious restriction enzyme mapping, Southern blot analysis, subcloning and sequencing, no obvious sequences relating to the intron-exon junction were detected. Therefore, some sequences in the α genes might cause recombination and deletion during cloning. No further work was performed on the above clones.

The polymerase chain reaction (PCR) method was subsequently used to investigate the cGTHα gene. Many sets of primers were designed and synthesized and only those which could produce promising amplification products are shown in Fig. 1. The sizes of amplified products were estimated from agarose gel electrophoresis as shown in Fig. 2. Different sets of primers produced different products: 220 bp for primers

---

The sequence data in this paper have been submitted to the EMBL/Genbank Data Libraries under the accession numbers X56497 and X56498.

Abbreviations: bp, base pair(s); GTH, gonadotropin; GTHα, α subunit of GTH; cGTHα, carp GTHα; kb, kilobase(s); PCR, polymerase chain reaction.

Correspondence: T.B. Lo, Institute of Biological Chemistry, Academia Sinica, P.O. Box 23-106, Taipei, Taiwan, R.O.C.

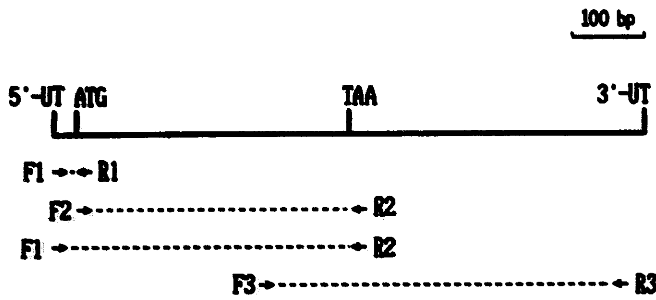

Fig. 1. The locations of primers and the strategy of using PCR to align the sequences of cGTHα gene. The forward (F) and reversed (R) primers were synthesized by a DNA synthesizer, Model 381A from Applied Biosystems, Foster City, CA, U.S.A. Sequences of the primers are indicated in Fig. 3. 5' UT, 5' untranslated region; 3' UT, 3' untranslated region; ATG, start codon; TAA, stop codon.

of F1 and R1 (lane 1); 520 bp for primers F2 and R2 (lane 2); 740 bp for primers F1 and R2, which was equal to the products from primers F1/R1 and F2/R2 (lane 3); 460 and 560 bp for primers F3 and R3 (lane 4).

All amplified DNA fragments were purified from agarose gel and were ligated into the *Sma*I site of plasmid vector pUC 18 [20]. For clones containing the

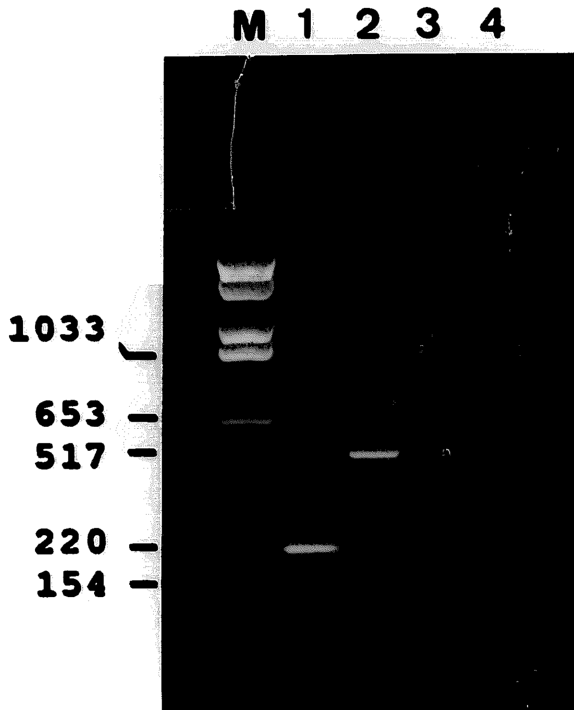

Fig. 2. PCR-amplified products from different set of primers. The products were separated by electrophoresis on a 1.7% agarose gel. 2 μg genomic DNA (prepared according to Ref. 22) and 10 μg primer were used. The samples were subjected to 50 cycles of amplification using a programmable heat block, model No. IHB 101 (Hybaid Ltd., Middlesex, U.K.) under the conditions of denaturation at 94°C for 1.5 min, annealing at 50°C for 1 min and polymerization at 72°C for 3 min. M, molecular weight standard as shown in bp. Lane 1–4, amplified products from primers F1/R1, F2/R2, F1/R2 and F3/R3, respectively.

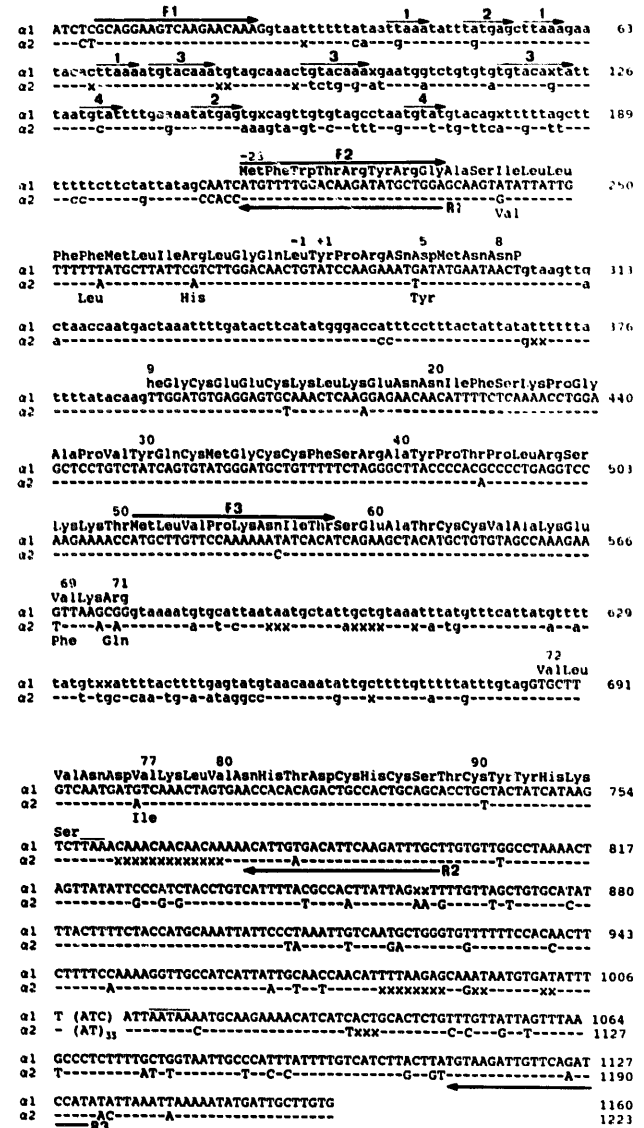

Fig. 3. The nucleotide sequences of cGTHα1 and cGTHα2 genes. Dashes in cGTHα2 gene indicate nucleotides identical to those of cGTHα1 gene. Crosses denote deletions used for the maximal match alignment of sequence. Exons are shown by capitals and introns by lower case letters. The unusual AT-repeat sequences in cGTHα2 gene are represented as $(AT)_{33}$ with the repeat number denoted by the subscript. Several direct repeats are shown and numbered 1 to 5. Locations and sequences of PCR primers are also shown. F1, F2 and F3 are forward primers; R1, R2 and R3 are reversed primers. TAA, stop codon; AATAA, polyadenylation signal.

same amplified DNA fragment, at least eight clones were subjected to DNA sequence analysis using Sanger's dideoxy chain termination method [21]. All the PCR amplified products of each set of primers contain two types of DNA sequences; one matches with the cGTHα1 cDNA while the other matches with the cGTHα2 cDNA which had been determined previously [3]. After analyzing the sequence of PCR products from different sets of primer, the sequence of cGTHα1 and cGTHα2 genes were aligned and are shown in Fig. 3.

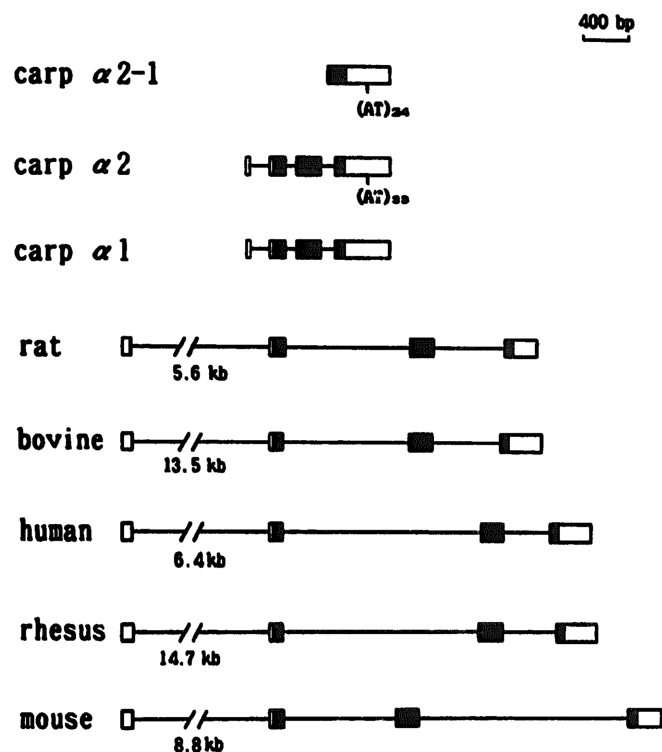

Fig. 4. Comparison of α gene structure of human [11], bovine [12], rat [13], mouse [14], rhesus [15] and carp. Boxes and lines represent exons and introns. Coding regions are shown as solid and noncoding regions as open boxes.

Both cGTHα1 and cGTHα2 genes comprise four exons and three introns. Intron I is located in the 5′ untranslated region and several tandem repeats are present in this intron as indicated in Fig. 3. It is 177 bp in length for both genes. The sequence homology of intron I between cGTHα1 and cGTHα2 genes is 74%. The other two introns, II and III, are located in the coding region and are 82 bp and 108 bp in the cGTHα1 gene, compared to 80 bp and 101 bp in the cGTHα2 gene. The sequence homology between cGTHα1 and cGTHα2 genes was 94% for intron II and 65% for intron III. An unusual AT-repeat is present in the 3′ untranslated region of the cGTHα2 gene but not in the cGTHα1 gene. Several kinds of cGTHα2 cDNA with varying length of AT-repeats as represented by $(AT)_{24-33}$ were found. A third cGTHα gene, α2-1, 460 bp in length, was detected by PCR using primers F3 and R3 (Fig. 4). Nucleotide sequence analysis of this fragment indicates that it is identical to part of the cGTHα2 gene except that intron III is lacking. Whether it is a pseudogene or not is unknown. The above results indicate that multiple α genes are present in carp as opposed to the single α gene as found in mammals [11–17].

Schematic presentations of human, bovine, rat, mouse, rhesus and carp GTHα gene structure are shown in Fig. 4. The basic organization of all α genes are the same, consisting of four exons and three introns. The intron I is located in the 5′ untranslated region and the introns, II and III, are in the coding region. The splicing sites of α genes among all species thus determined are very similar: intron I, at a few bases from the start codon, ATG; intron II, at codon +9 or +10 for bovine, carp, mouse and rhesus, at codon +6 or +7 for human and rat; intron III, at the site between codon +70 and +71 or between codon +71 and +72. The size of exons of cGTHα genes is almost the same as those mammalian α genes. On the other hand, the size of introns of cGTHα genes is much smaller than those of mammalian α genes: intron I, 177 bp vs. 5.6 to 14.7 kb; intron II, 80 to 82 bp vs. 1.1 to 1.7 kb; intron III, 101 to 108 bp vs. 0.4 to 1.7 kb. The above data imply that the length and complexity of the introns are increased during molecular evolution of α genes.

We thank Dr. G.D. Chang for valuable discussions. This work was supported by grants from the National Science Council, Taiwan, Republic of China.

References

1 Pierce, J.G. and Parsons, T.F. (1981) Annu. Rev. Biochem. 50, 465–495.
2 Sairam, M.R. (1983) in Hormonal proteins and peptides: Gonadotropic Hormones (Li, C.H., ed.), pp.1–79, Academic Press, New York.
3 Chang, Y.S., Huang, C.J., Huang, F.L. and Lo, T.B. (1988) Int. J. Pept. Prot. Res. 32, 556–564.
4 Liu, C.S., Huang, F.L., Chang, Y.S. and Lo, T.B. (1989) Eur. J. Biochem. 186, 105–114.
5 Sekine, S., Saito, A., Itoh, H., Kawauchi, H. and Itoh, S. (1989) Proc. Natl. Acad. Sci. USA 86, 8645–8649.
6 Chang, Y.S., Huang, C.J., Huang, F.L., Liu, C.S. and Lo, T.B. (1990) Gen. Comp. Endorcrinol. 78, 23–33.
7 Kawauchi, H., Suzuki, K., Itoh, H., Swanson, P., Naito, N., Nagahama, Y., Nozaki, M., Nakai, Y. and Itoh, S. (1989) Fish Physiol. Biochem. 7, 29–38.
8 Suzuki, K., Kawauchi, H. and Nagahama, Y. (1988) Gen. Comp. Endorcrinol. 71, 292–301.
9 Swanson, P., Suzuki, K., Kawauchi, H. and Dickhoff, W.W. (1991) Biol. Repr. 44, 29–38.
10 Querat, B., Jutisz, M., Fontaine, Y.A. and Counis, R. (1990) Mol. Cell. Endocrinol. 71, 253–257.
11 Fiddes, J.C. and Goodman, H.M., (1981) J. Mol. Appl. Gen. 1, 3–18.
12 Boothby, M., Ruddon, R.W., Anderson, C., McWilliams, D., and Boime, I. (1981) J. Biol. Chem. 256, 5121–5127.
13 Goodwin, R.G., Moncman, C.L., Rottman, F.M. and Nilson, J.H. (1983) Nucleic Acids Res. 11, 1910–1920.
14 Burnside, J., Buckland, P.R. and Chin, W.W. (1988) Gene 70, 67–74.
15 Gordon, D.F., Wood, W.M. and Ridgway, E.C. (1988) DNA 7, 679–690.
16 Golos, T.G., Durning, M. and Fisher, J.M. (1991) DNA Cell Biol. 10, 367–380.
17 Steger, D.J., Altschmied, J., Buscher, M. and Mellon, P.L. (1991) Mol. Endocrinol. 5, 243–255.

18 Huang, C.J., Huang, F.L., Chang, G.D., Chang, Y.S., Lo, C.F., Fraser, M.J. and Lo, T.B. (1991) Proc. Natl. Acad. Sci. USA 88, 7486–7490.
19 Evans, G.A., Lewis, K. and Rothenberg, B.E. (1989) Gene 79, 9–20.
20 Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene 33, 103–119.
21 Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
22 Signer, E., Kuenzle, C.C., Thomann, P.E. and Hubscher, U. (1988) Nucleic Acids Res. 16, 7738.