Biochimica et Biophysica Acta, 1174 (1993) 285–288
Elsevier Science Publishers B.V.

BBAEXP 90554

Short Sequence-Paper

Sequence of the growth hormone (GH) gene from the silver carp (*Hypophthalmichthys molitrix*) and evolution of GH genes in vertebrates [1]

Yunhan Hong [2] and Manfred Schartl [2]

Gene Center, Max-Planck-Institute for Biochemistry, Martinsried (Germany)

(Received 12 June 1992)

Key words: Recombinant DNA; Growth hormone gene; PCR; (Silver carp); (Fish)

The silver carp (*Hypophthalmichthys molitrix*) growth hormone (GH) gene was isolated and sequenced following amplification from genomic DNA by the polymerase chain reaction. The gene spans a region of approx. 2.5 kb nucleotides (nt) and consists of five exons. The sequence predicts a polypeptide of 210 amino acids (aa) including a putative signal peptide of 22 hydrophobic aa residues. The arrangement of exons and introns is identical to the GH genes of common carp, grass carp, and very similar to mammals and birds, but quite different from that for the GH genes of tilapia and salmonids. The silver carp GH gene shares a high homology at the nt and aa levels with those of grass carp (95.3% nt, 99.5% aa) and of common carp (81% nt, 95.7% aa).

Growth hormone (GH), a single-chain polypeptide of approx. 22 kDa produced and secreted by the pituitary gland, is essential for growth regulation in vertebrates. Because of its potential in animal husbandry, GH have been widely used in transgenic studies in livestocks including fish. The cDNAs for GHs have been cloned and sequenced in several fish species [1,4,7–10,12,13,15,17–23,26]. More recently, the GH genomic sequences have also been described in rainbow trout [2], Atlantic salmon [14], chum salmon [25], common carp (cc) [6], grass carp (gc) [11] and tilapia [3]. GH gene organisation seems to vary more in fish than in mammals, as those for the trout, salmon and tilapia consist of six exons while those for the carps have only five. Here we report the cloning and sequence analysis of the gene coding for the silver carp GH (scGH). The silver carp (*Hypophthalmichthys molitrix*) feeds on phytoplankton and represents one of the most rapid growing fish species. We are interested in comparing its GH sequence with those of other fish and in using its GH gene for the production of fast-growing transgenic fish.

Silver carp was collected from the breeding population reared at the Experimental Station of the Yangtze-River Fisheries Research Institute, Shashi, Hubei, PR China. Genomic DNA was extracted from whole blood cells and subjected to amplification by the polymerase chain reaction (PCR) using two oligonucleotide primers designed according to the published cDNA sequences for the common carp [3,14] and the grass carp [9]. Following *Eco*RI digestion, the PCR amplified product was cloned into pBluescript II + at the *Eco*RI site. The recombinant plasmid DNA was prepared according to standard procedures [15]. The DNA sequence was determined on both strands of the denatured plasmid template by the dideoxy chain-termination method using T7 DNA polymerase. The sequencing strategy is described in Fig. 1.

Based on the sequence obtained (Fig. 2), the scGH gene is 2484 bp long (2432 bp for the region from the ATG translation start codon to the ATTAAA polyadenylation signal) and consists of five exons (exon I, 40 bp; exon II, 140 bp; exon III, 117 bp; exon IV, 162 bp; exon V, 715 bp) and four introns (intron I, 270 bp; intron II, 471 bp; intron III, 423 bp; intron 4, 146). All introns begin with GT and terminate with AG. The

Correspondence to: M. Schartl, Department of Physiological Chemistry I, Biocenter, University of Würzburg, Am Hubland, 97074 Würzburg, Germany.

[1] This paper contains part of the Ph.D. thesis of Y.H.

[2] Present address: Department of Physiological Chemistry I, Biocenter, University of Würzburg, Am Hubland, 97074 Würzburg, Germany.

The sequence data reported in this paper have been submitted to the EMBL/Genbank Data Libraries under the accession number M94348.

introns are significantly longer than those of mammalian GH genes. As with GH genes of other fish except for tilapia, the scGH 3′ untranslated region are substantially larger than that of mammalian and chicken genes (500 bp in fish vs. 100 bp in mammals and chicken). The scGH gene is rich in AT bases (61.8%), especially in the introns (68.4%) and in the 3′ untranslated region (63%), with the protein coding sequence having no such a bias (52.1% GC).

Comparison of the scGH gene structure with those from other animals (Fig. 3) reveals that the scGH gene, in contrast to the GH genes of rainbow trout, Atlantic salmon and tilapia, has only five exons. Similar to the common carp and grass carp (gs), exon V is not splitted as in tilapia and the two salmonid species. All mammalian GH genes characterized so far have the same number of exons and introns as the carps. Also the chicken GH gene has recently been reported to have five exons. Thus, the introduction of intron V into fish GH gene should have taken place after the divergence of fish and tetrapodes. Furthermore, among the three carp species, silver carp and grass carp are essentially identical in the GH transcription units and in the sizes of exon and introns, suggesting a more recent common origin for the two species than for silver carp and common carp.

No data are available concerning the aa sequence of the authentic scGH polypeptide. However, since the first 22 aa residues from the N-terminus are highly hydrophobic and also highly homologous to the signal peptide sequence of other fish, especially that of the common carp [3], this region is therefore assumed to represent the signal peptide of the pre-GH, which will be cleaved during maturation and secretion of the hormone. The first aa residue in mature scGH should then be a serine as is the case for the common carp GH.

Comparison of the scGH gene sequence with those from other fish indicates that there is a high degree of homology, both at the nt and aa level, among the three carp species. The scGH and gcGH genes share an extremely high overall homology (95.3%) throughout the entire gene sequence (98.4% identity for the protein coding sequence, 96.8% for the 3′ untranslated region, 94.3% for intron 1, 91.5% for intron 2, 88.4% for intron 3 and 95.1% for intron 4, respectively), while the comparison between scGH and ccGH genes reveals a lower overall homology (81%). The homology regions are unevenly distributed, with the protein coding sequence having a high identity (92.4%), followed by the 3′ untranslated region (83% identity), while the intron sequences having varied much more extensively (74.4% identity for intron 1, 54.4% for intron 2, 44.4% for intron 3 and 58.7% for intron 4). Although all the three species belong to the same family Cyprinidae, they are classified into three separate subfamilies. The GH gene structure and aa sequence (see below) indicate a much closer relationship of silver carp with grass carp than with common carp.

The preGHs of the three carp species consist of 210 aa residues. The scGH differs by only one aa from gcGH, resulting in 99.5% homology, and by 7 and 16 aa from ccGH1 and ccGH2, leading to 95.7% and 92.4% homology (common carp have at least two GH genes, the one whose aa sequence is more homologous to the scGH is arbitrarily referred to ccGH1, whose genomic sequence is not yet available; while the other one whose gene sequence has been determined is termed ccGH$_2$ which is more divergent to the scGH than the ccGH1). The silver carp preGH is identical to

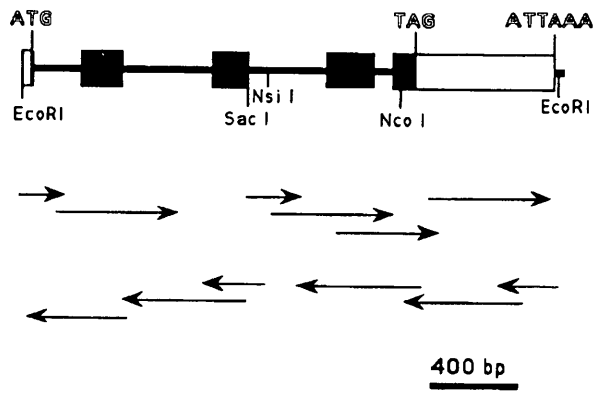

Fig. 1. The general structure, partial restriction map and sequencing strategy of the silver carp GH gene. Genomic DNA was prepared from blood cells of silver carp (*Hypophthalmichthys molitrix*). Two synthetic oligonucleotides (GH51 and GH31) were used as PCR primers. GH51 (5′-ggaattctagaTGTGTCTACCCTGAGCGAAATG-G-3′) is derived from the 5′ region of the two cDNAs [4,15] and one genomic sequence [6] for the common carp GHs, GH31 (5′-ttcgaattcatTGGATGCAATTTAAAACTTTAAT-3′) is complementary to the 3′ region of all three sequences. The primers contain an additional *Eco*RI site (small letters) at their 5′ ends for further cloning. Polymerase chain reactions (PCR) were performed in a 50 μl total volume containing 1×PCR buffer (67 mM Tris-Cl, pH 8.8, 6.7 mM $MgCl_2$, 16.6 mM $(NH_4)_2SO_4$, 10 mM 2-mercaptoethanol), 50 pM of each primer, 100 μM of each of dATP, dTTP, dGTP and dCTP (Pharmacia), 100–300 ng of the silver carp genomic DNA, and 2 units of Taq DNA polymerase (Perkin-Elmer). The samples were denatured for 5 min at 92°C, followed by 35 reaction cycles (1 min denaturation at 92°C, 1 min annealing at 50°C, and 5 min extension at 72°C) with the final cycle having 15 min extension. The PCR-amplified fragment was digested with *Eco*RI, separated on a 1% agarose gel, recovered from the gel using the Geneclean II kit (Bio 101 Inc.), and cloned into plasmid pBluescript II KS+ (Stratagene) which had been cut with *Eco*RI. Plasmid DNA was sequenced in both directions by the dideoxynucleotide method using T7 DNA polymerase. For sequencing several overlapping subclones were generated using the sites illustrated. Several synthetic oligonucleotides were used for the primer walking. Exons are shown by boxes, with solid boxes being the translated regions, open boxes untranslated regions. Introns are shown by the lines between the boxes. The translation start codon (ATG), the stop codon (TAG) and the polyadenylation signal (ATTAAA) are indicated. Arrows indicate the extents and directions of sequencing by the dideoxy method.

preGH of human (36.7%), chicken (43. 1%), bullfrog (41.1%), flounder (51.6%), tilapia (55%) tuna (56.7%), rainbow trout (65.4%), Atlantic salmon (66%), common carp (95.7%) and grass carp (99.5%), respectively.

We thank Zhizhou Hu (the Yangtze-River Fisheries Research Institute, Shashi, Hubei, PR China) for kindly providing the silver carp genomic DNA. This work was supported by grants to M.S. provided by G.I.F.

```
                                M   A   R                                                          3
ggaattctagatcTGTCTACCCTGAGCGAA ATG GCT AGA G gtatggatgtgatgctttgtctcatttgtggatttaagacttatttttaactcc  95
ttatttcatttagatgttttttttttgtgctttcctcttatttgtaaaactaatttcatagttgtaaaagaagtttattggagcatggaaaacaaaaca  195
aatgtatttgatctccaaaatctgttttgattaatccaaaatcttgatttcattggttattttatcaggaacgatgttgcagaggaaacaaacctctttc  295

                 A   L   V   L   L   S   V   V   L   V   S   L   L   V   N   Q   G   T   A   S   E    24
ttttccttctcctag  CA TTA GTG CTG TTG TCG GTG GTG CTG GTT AGT TTG TTG GTG AAC CAG GGG ACG GCC TCA GAG  372

 N   Q   R   L   F   N   N   A   V   I   R   V   Q   H   L   H   Q   L   A   A   K   M   I   N   D    49
AAC CAG CGG CTT TTC AAC AAC GCA GTC ATC CGT GTT CAA CAC CTG CAC CAG CTG GCT GCA AAA ATG ATT AAC GAC  447

 F                                                                                                  50
TTT gtaagattcccattaaaatctattcaaaactagcagattctaacctgtcatttagaacaaaacgaacgatccttggcctattcaatgggaaatatg  546
aggctttaggacttaaattaaaattctgtcatcatttattcaccctcagtttgttcccaaaccctgtatgaatttctttcttctgctgaacacaaaagaa  646
gatattttgaagaatatgggtagccaaacagttgatggtccccagtgacttccatagtatggaaaaaaaaaaatactatggaagtcaccaacattcttc  746
aaaatatagtcttttgtgatcagcagaagaaactcatacaggtttggaacaacttcagggtgagtaaatgatgactgaatttttcctttagtaagtgttc  846

                                                                              E   D   N   L   L   P    56
tttgcatttccatcagatgtatttattgattatgttatttgagtgagtaaatactattttcctttgttattaag GAG GAC AAC CTG TTG CCT  939

 E   E   R   R   Q   L   S   K   I   F   P   L   S   F   C   N   S   D   S   I   E   A   P   T   G    81
GAG GAA CGC AGA CAG CTG AGT AAA ATC TTT CCT CTG TCT TTC TGC AAC TCT GAC TCC ATT GAG GCG CCC ACT GGA 1014

 K   D   E   T   Q   K   S   S                                                                      89
AAA GAT GAA ACG CAG AAG AGC TCT gtaagtacacaatgtcattaaaaatgaaaccccaacctcttatgttttggtcaaaaaaatgtagaact 1106
atgaaagaaataatagtacatgcatgcaactaaatttttttttttgttttcacaattgctcaaaaatattacactgtacactgtaaaaaacatttttttct 1206
tttgtcaaatcaacttaattaatgtggttcagatagcataatattgtgagtttctgctgattaaaccaatcaccttcattcattgtattaactcaaattt 1306
taaggcaaccatgtaacttatttttttttaaaaccaacaattttttttacagtgcagcatttttcttgcagtttggaaaacagaaccaaccacataaggca 1406

                                                          M   L   K   L   L   R   I   S   F   R   L   100
tcaaaaatggctaggttttctgaaggccttccggtattatttttttctttcttcag ATG TTG AAG CTC CTT CGC ATC TCT TTC CGC CTC  1494

 I   E   S   W   E   F   P   S   Q   T   L   S   G   A   V   S   N   S   L   T   V   G   N   P   N   125
ATT GAG TCC TGG GAG TTC CCC AGC CAG ACC CTG AGC GGA GCC GTC TCA AAC AGC TTG ACC GTC GGG AAC CCC AAC  1569

 Q   I   T   E   K   L   A   D   L   K   V   G   I   S   V   L   I   K                              143
CAG ATC ACT GAG AAG CTG GCC GAC TTG AAA GTG GGC ATC AGC GTG CTC ATC AAG gtgagagagaccagatttattctaggct 1651
actctgtttttttatacagtgtgtctaagtggtcataccaaggagacgtgatgacatctttacatcttcagggcttatttggatagctatgacaaattg 1751

                    G   C   L   D   G   Q   P   N   M   D   D   N   D   S   L   P   L   P   F   E   163
ccatttgtatctgcacag  GGA TGT CTG GAT GGT CAA CCA AAC ATG GAT GAT AAC GAC TCC CTG CCG CTG CCT TTT GAG  1829

 D   F   Y   L   T   M   G   E   S   S   L   R   G   S   F   R   L   L   A   C   F   K   K   D   M   188
GAT TTC TAC TTG ACC ATG GGG GAG AGC AGC CTC AGA GGG AGC TTC CGT CTT CTG GCT TGC TTC AAG AAG GAC ATG  1904

 H   K   V   E   T   Y   L   R   V   A   N   C   R   R   S   L   D   S   N   C   T   L   *           210
CAC AAG GTG GAA ACT TAC CTA AGG GTT GCA AAT TGC AGG AGA TCC CTG GAT TCA AAC TGC ACC CTG TAG AGGGCGCC 1981

AATGTATTGCTAGCCAAAGCCTGTGACACACTTTGCTGCAAATCTAAAACCAGTTTAAGTCCTCAAAATCTCCTAATATAATTATTATCTGGTTTTATAT 2081

ATGCAGGAAATGTCAACCAGGCATGGCTAGGTCTGTTCTCTAGCTCCCTCCCATATCTAAACCCTACCTAACACTATTGTATTTATTCTTCTCATTGGGG 2181

AGTGCTCGTAAATTAAAGACATTAAGATCTGATTTAACACTTTCACAGTGGTGCTAAGCAATTTATGGCGATATATTTTAAAATGTGCCCAAATCGCTTT 2281

GACTCTAGTATTTTATGGCTCCAAAAATGGCTAAAGATGCCTTTTGTCGAAACTGTCATTTGGATGGATGGGTTCACTCCCAACCAAGTGTATTAATGTA 2381

AACATTTGTCTGTCTGATAGGTTATGTCCATATTATTAGCTCATGCTGTTCTCTTGAAGCTGTGTGTCTTTATCCATTAAATTTTATTGCATCCAATgaa 2481

ttc                                                                                                2484
```

Fig. 2. The nucleotide and deduced amino acid sequence of the silver carp GH gene. Exons are shown by capitals and introns by lowercase letters. The ATTAAA polyadenylation signal is shown in bold. Encoded amino acid residues are represented by one-letter codes above the codons. The asterik marks the stop codon. The sequences used for the PCR primers are underlined. The bases shown by lower-case letters at both ends were introduced by the PCR primers. Note that the sequence immediately downstream of the ATTAAA is different from the PCR primer (GH31; see Fig. 1). The reason for this strange phenomenon is unknown. The sequence has been deposited in GenBank (accession No. M94348).

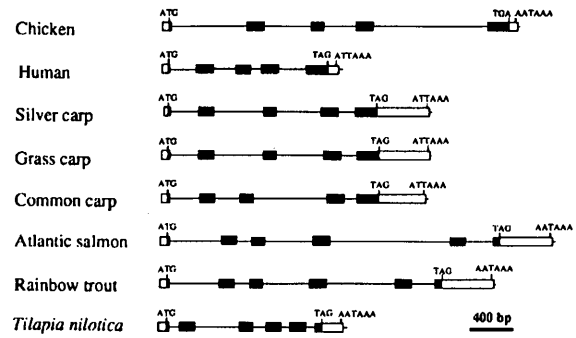

Fig. 3. Comparison of the GH gene structures of silver carp, chicken [24] human [7], grass carp [11], common carp [6], rainbow trout [2], Atlantic salmon [14] and tilapia [3]. Boxes = exons; open boxes = untranslated regions; solid boxes = translated regions; lines between exons = introns. Translation start codons, stop codons and polyadenylation signals are indicated.

(No. I-18-127.3/87) and the Bundesministerium für Forschung und Technologie through 'Schwerpunkt: Grundiagen und Anwendungen der Gentechnologie' (No. 26). Y.H. is a fellow of the Deutsche Akademische Austauschdienst.

References

1 Agellon, L.B. and Chen, T.T. (1986) DNA 5, 463–471.
2 Agellon, L.B., Davies, S.L., Chen, T.T. and Powers, D.A. (1988) Proc. Natl. Acad. Sci. USA 85, 5136–5140.
3 Ber, R. and Daniel, V. (1992) Gene 113, 245–250.
4 Chao, S.-C., Pan, F.-M. and Chang, W.-C. (1989) Biochim. Biophys. Acta 1007, 233–236.
5 Chen, T.T., Agellon, L.B., Lin, C.M., Tsai, H.J., Zhang, P., Gonzalez-Villasenor, L.I. and Powers, D.A. (1989). Fish Physiol. Biochem. 7, 381–385.
6 Chiou, C.-S., Chen, H.-T. and Chang, W.-C. (1990) Biochim. Biophys. Acta 1087, 91–94.
7 DeNoto, F.M., Moore, D.D. and Goodman, H.M. (1981) Nucleic. Acids Res. 9, 3719–3730.
8 Funkenstein, B., Chen, T.T., Powers, D.A. and Benzion, C. (1991) Gene 103, 243–247.
9 Gonzalez-Villasenor, L.I., Zhang, P., Chen, T.T. and Powers, D.A. (1988) Gene 65, 239–246.
10 Ho, W.K.K., Tsang, W.H. and Dias, N.P. (1989) Biochem. Biophys. Res. Commun. 161, 1239–1243.
11 Ho, W.K.K., Wong, M.W. and Chan, A.P.Y. (1991) Biochim. Biophys. Acta 1090, 245–248.
12 Hew, C.L., Trinh, K.Y., Du, S.J. and Song, S. (1989) Fish. Physiol. Biochem. 7, 375–380.
13 Johanson, B. and Valla, S. (1987) In EIFAC/FAO on Selection, Hybridization, and Genetic Engineering in Aquaculture, Vol. II, (Tiews, K., ed.), Bordeaux, 27–30 May, 1986, Heenemann, Berlin.
14 Johansen, B., Christian O. and Valla, S. (1989) Gene 77, 317–324.
15 Koren, Y., Sarid, S., Ber, R. and Daniel V. (1989) Gene 77, 309–315.
16 Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory manual, Cold Spring Habor Laboratory, Cold Spring Habor.
17 Momota, H.,Kosugi, R., Ohgai, H., Hara, A. and Ishioka, H. (1988) Nucleic. Acid Res. 16, 10368.
18 Momota, H.,Kosugi, R., Hiramatsu, H., Ohgai, H., Hara, A. and Ishioka, H. (1988) Nucleic. Acids Res. 16, 3107.
19 Nicoll, C.S., Steiny, S.S., King, D.S., Nishioka, R.S., Mayer, G.L., Eberthardt, N.L., Baxter, J.D., Yamanaka, M.K., Miller, J.A., Steilhamer, J.J., Schilling, J.W. and Johnson, L.K. (1987) Gen. Comp. Endocrinol. 68, 387–399.
20 Rentier-Delrue, E., Swennen, D., Philippart, J. C, L'hoir, C., Lion, M., Benrubi, O. and Martial, J.A. (1989) DNA 8, 271–278.
21 Saito, A., Sekine, S., Komatsu, Y., Sato, M., Hirano, T. and Itoh, S. (1988) Gene 73, 545–551.
22 Sato, N., Watanabe, K., Murata, K., Sakaguchi, M., Kariya, Y., Kimura, S., Nonaka, M. and Kimura, A. (1988) Biochim. Biophys. Acta 949, 35–42.
23 Sekine, S., Mizukami, T., Nishi, T., Kuana, Y., Saito. A., Sato, M., Itoh, S. and Kawauhi, H. (1985) Proc. Natl. Acad.Sci. USA 82, 4306–4310.
24 Tanaka, M., Hosokawa, Y. and Nakashima, K. (1992) Gene 112, 235–239.
25 Wang, Y., Li, H., Liu, D. and Shen, X. (1991) Abstracts of the 4th International Symposium on Genetics in Aquaculture, April 29–May 3, 1991, Wuhan, PR China.
26 Watahiki, M., Tanaka, M., Masuda, N., Yamakawa, M., Yoneda, Y. and Nakashima, K. (1988) Gen. Comp. Endcrinol. 70, 401–406.