Analysis of α -Globin Gene Family of Carp

I. α -Globin Gene Structure of No. 3, No. 6, and No. 7

Masato Miyata,* Ikuo Hirono,* and Takashi Aoki*

(Received February 8, 1993)

Seven kinds of α -globin genes were isolated from the genomic library of carp *Cyprinus carpio* chromosomal DNA. In this study, the complete nucleotide sequences of three carp α -globin genes (Nos. 3, 6, and 7) were determined. These carp genes had analogue with the arrangement of translated, untranslated, and intervening sequences of the vertebrate α -globin gene, *i.e.* two intervening sequences interrupting the coding region into three parts. The coding region of exon 1, 2, and 3 in Nos. 3, 6, and 7 were 95 base pair (bp), 208 bp, and 129 bp, respectively. The first and second intervening sequences of No. 3 were 159 bp and 96 bp in length, respectively. The intervening sequences of No. 6 were 179 bp and 99 bp, and of No. 7 159 bp and 92 bp in length, respectively. In the 5'- and 3'-untranslated region, sequences homologous with the consensus sequence of the signals for expression were observed. The individuality of the nucleotide sequences suggests that the carp α -globin genes constitute a family structure. The amino acid residues of the predicted carp α -globin concern quaternary structure were extremely conserved with other vertebrate, *e.g.*, human, duck, and frog.

The globin genes have been studied in many vertebrates as a model for genetic regulation and biological evolution. However, there has been little research on fish globin genetics. The first molecular-genetical analysis of fish globin genes was performed by Takeshita *et al.*¹⁾ They cloned carp α -globin gene cDNA coding 432 base pairs (bp) of an open reading frame. However, there were 16 discrepancies between the predicted amino acid sequence of the cDNA and the previously reported α -globin chain of the German mirror carp.²⁾ Subsequently, Nobuta and Aoki cloned seven fragments which hybridized with the carp α -globin cDNA.³⁾ The complete nucleotide sequence of one (No. 4) of the seven clones (No. 1 to No. 7) was determined. The structure of gene No. 4 was similar to a conserved form where the two intervening sequences separate the protein coding region into three exons. Comparing cDNA and gene No. 4, there were 26 nucleotide and 12 deduced amino acid discrepancies. In gene No. 4, the signals required for expression were recognized in the 5' flanking region, 5' non-coding region and the 3' non-coding region.

In this paper, we present the complete nucleotide sequences of the carp genomic DNA fragments Nos. 3, 6, and 7 that hybridized with carp α -

globin cDNA. The restriction site arrangement of these fragments was suitable for determining their nucleotide sequence. The nucleotide and predicted amino acid sequences of the three DNA fragments are compared and discussed.

Materials and Methods

Figure 1 shows the alignment of the restriction enzyme cleavage map of carp genomic DNA fragment clones that hybridized with the carp α -globin cDNA. The regions that were anticipated to include the structural region of fragments Nos. 3, 6, and 7 were subcloned into pUC118 and pUC119,⁴⁾ respectively. The detailed restriction endonuclease cleavage maps of the three fragments were generated to permit additional subcloning to determine nucleotide sequence. The determination of nucleotide sequence was performed by the dideoxy method described by Sanger *et al.*,^{5,6)} and the nucleotide sequences were edited and analyzed using by GENETYX and GENETYX-CD (SDC Software Development Co., Ltd. Tokyo Japan).

* Department of Biological Resources, Faculty of Agriculture, Miyazaki University, Gakuen-Kibanadai Nishi, Miyazaki 889-21, Japan (宮田雅人, 広野育生, 青木 宙: 宮崎大学農学部生物資源利用学科).
Corresponding author: T. Aoki

Fig. 1. Restriction enzyme cleavage map of the seven chromosomal carp α -globin gene clones.

Restriction sites are indicated by vertical lines and letters: (B) *Bam*HI, (E) *Eco*RI, (H) *Hind*III. The thick lines indicate the fragments that hybridized with α -globin cDNA. The structures of each globin gene are represented by filled (protein coding region: Exon) and open (intervening sequences: IVS, and 5'- and 3'-untranslated regions) rectangles. The broken lines indicate unanalyzed regions in this study.

Results

Structure of Fragments Encoding Carp α -Globin Gene

Figure 2 shows the alignment of the nucleotide sequences of fragments Nos. 3, 6, and 7 analyzed in this study, and also those of previously reported carp α -globin cDNA¹⁷ and carp α -globin gene No. 4.⁸⁾

The total nucleotide sequence of fragment No. 3 was 2178 bp in length. Fragment No. 3 includes 673 bp of non-coding region upstream to the 5' end of the ATG initiation codon and 818 bp of non-coding region downstream to the 3' end of the TAA termination codon. The fragment No. 3 also encodes open reading frame (Exon) in three discontinuous segments interrupted by

two intervening sequences. The positions of the first and second intervening sequences were specified at predicted amino acid positions 31 and 100/101 by the consensus nucleotide sequences of the intron-exon junction.

Fragment No. 6 contained a 1245 bp length nucleotide sequence that was anticipated to encode the α -globin. The nucleotide sequence No. 6 was enclosed by non-coding regions of 295 and 240 bp upstream to the ATG initiation codon and downstream to the TAA termination codon, respectively. The structure of sequence No. 6 was also constructed from three coding regions and two intervening sequences. The position of the intron-exon junction were exactly same as in fragment No. 3.

The complete nucleotide sequence of fragment No. 7 was 1133 bp length recognized as an α -globin encoding domain. The nucleotide sequence No. 7 was enclosed by 293 and 157 bp of non coding tracts, that upstream and downstream of the coding region, respectively. The structure of fragment No. 7 was also constructed from three exons and two intervening sequences with the same junction positions.

Upstream Region

In fragment No. 3, a TATA box (TATAAA) and -100 region (GGGTG) were located 81 and 128 bp upstream from the ATG initiation codon, respectively. The CCAAT box (CCGAT) was found 9 bp downstream from the -100 region. A simple sequence region (TAAA or TAAAA)_n $n=8$ was observed 383 to 345 bp upstream from the ATG initiation codon.

In fragments Nos. 6 and 7, the three promoter elements of -100 region, CCAAT box (CCGAT, TCGA) and TATA box were located 129 bp, 81 bp, and 120 bp upstream from the ATG initiation codon, respectively.

Downstream Region

The dinucleotide GC for poly(A) addition was located 68 nucleotides downstream from the TAA termination codon in fragment No. 3. Also in this fragment, two hexanucleotides AATAAA considered to be the poly(A) signal were found from 21 bp upstream to the predicted poly(A) addition site GC.

In fragment No. 6, the poly(A) addition site GC was located only 60 nucleotides downstream from the TAA termination codon, because a deletion of 8 bp was recognized between the TAA

(Fig. 2)

termination codon and two poly(A) signals.

In fragment No. 7, the poly(A) addition site GC and the two poly(A) signals (AATAAA) were located at the same position as in fragment No. 3.

The various lengths of alternating T (T)_n ($n \leq 15$) tracts were recognized downstream from the poly(A) additional sites in fragments Nos. 3, 6, and 7.

Intervening Sequence

Fragment No. 3 contains large and small intervening sequences in the protein coding region. The first (IVS1) was of 159 bp and occurred in amino acid codon 31. The second (IVS2) was of 96 bp, and occurred between the positions of codons 100 and 101. Tendencies for intervening sequences to begin with the dinucleotide GT and end with the dinucleotide AG,⁷⁾ were conserved in both intervening sequences of fragment No. 3. These dinucleotides were also enclosed in the sequences corresponding to consensus Δ AG|GT Δ AGT (donor site) and YYYYYYN Δ AG|G Δ (acceptor site) (|the site

of exon-intron junction, Y: pyrimidine, N: any nucleotide).⁷⁾ Furthermore, the sequences agree with the conserved sequence that is closely involved with forming the lasso structure during splicing; YNYTRAY⁷⁾ (R: purine) were found next to the acceptor site of the IVS1 and IVS2 of fragment No. 3.

IVS1 and IVS2 of fragment No. 6 were 179 bp and 99 bp in length, respectively. In comparison with IVS1 of fragments Nos. 3 and 7, there were twenty nucleotides in the insertion in IVS1 of fragment No. 6. Consensus sequences for the splicing were also conserved in both intervening sequences of fragment No. 6.

In fragment No. 7, the lengths of IVS1 and IVS2 were 159 bp and 92 bp, respectively. Both intervening sequences of fragment No. 7 also included conserved sequences for splicing.

The similarity of the nucleotide sequence of IVS1 between fragments Nos. 3 and 7 was notable. There were only two nucleotide discrepancies in each IVS1 between fragments Nos. 3 and 7.

Fig. 3. Homologous alignment of carp α -globin sequences with the corresponding chains from human, duck *Carinia moschata* and frog *Xenopus laevis*.

All amino acid sequences except that of German mirror carp²³ are predicted from nucleotide sequences. Boxes indicate the helical regions, as established for mammalian hemoglobins. Letters below the sequences represent the residues that have structural or functional importance in the human $\alpha 1$ -globin chain. The positions of functional residues are indicated by letters: (H) haem binding,²⁴ (1) $\alpha 1\beta 1$ and (2) $\alpha 1\beta 2$ interfaces,²⁴ and (B) Bohr effect residues.²⁵

Protein-Encoding Region

The predicted coding regions of 'fragments Nos. 3, 6, and 7 have no gaps or insertions that cause a frameshift. The coding regions of fragments Nos. 3, 6, and 7 encode 142 amino acid residues. Comparing the nucleotide sequences of carp α -globin cDNA and of fragments Nos. 3, 6, and 7, there were 12 (No. 3), 11 (No. 6), and 6 (No. 7) nucleotide discrepancies.

Discussion

In this study, we have determined the three kinds of nucleotide sequences of carp genomic DNA fragments that hybridize with carp α -globin cDNA.¹³ We have shown that the DNA fragments Nos. 3, 6, and 7 are carp α -globin gene, for

the following reasons. 1) The arrangement of translated, untranslated, and intervening sequences were completely consistent with those of vertebrate α -globin gene. 2) In comparison with the previously reported α -globin chain of the German mirror carp,²³ the deduced amino acid sequences of fragments Nos. 3, 6, and 7 have equivalent length, and there were over 84.5% homologies. A further characteristic property was that the nucleotide and deduced amino acid sequences of the carp α -globin genes Nos. 3, 4,²³ 6, and 7 did not correspond with those of cDNA as determined by Takeshita *et al.*¹³ Therefore, these facts suggest that the carp α -globin genes constitute a family structure.

In general, the sequences important for transcription belonged to three components; TATA

box (TATAAA),^{8,9)} CCAAT box (CCAAT),^{8,10)} and -100 region (GGGYG, CRCCC).^{8,10)} In the genomic carp α -globin genes, Nos. 3, 4,³⁾ 6, and 7, the TATA box (TATAAA), -100 region (GGGTG) and CCAAT box (CCGAT, TGCAT) were recognized. However, the locations of the carp CCAAT boxes were similar to those of other vertebrate globin genes, although the sequences have discrepancies with consensus sequence CCAAT. Consequently, the carp α -globin genes are thought to have different components that replace the role of the CCAAT box, or inconsistencies of the CCAAT box regulate their expression level. The simple sequence (TAAA or TAAAA)_n encoded upstream of gene No. 3 was complementary to the simple sequence of human β - and ζ -globin genes.¹¹⁾ Therefore, this correspondence reveals the potential importance of carp (TAAA or TAAAA)_n tract.

In the downstream region, the two copies of the poly(A) signal AATAAA were seen in genes Nos. 3, 4,³⁾ 6, and 7. However, only one poly(A) signal AATAAA was observed in other carp genes, e.g. growth hormone¹²⁾ and prolactin gene,¹³⁾ or other vertebrate α -globin genes including human,¹⁴⁾ duck *Carinia moschata*,¹⁵⁾ and frog *Xenopus laevis*.¹⁶⁾ Therefore, we suggest the two poly(A) signals are peculiar properties of the carp α -globin gene.

In each intervening sequence of genes Nos. 3, 6, and 7, the three conserved sequences that are required to accomplish the splicing were found. These conserved sequences are thought to play a role in splicing their intervening sequences. We have demonstrated the notably conserved intervening sequences of the carp genomic α -globin genes Nos. 3, 6, and 7. However, in comparing intervening sequences between the constituent genes of the vertebrate globin gene family, e.g., human,^{14,17)} duck,¹⁵⁾ and frog,¹⁶⁾ their lengths are disparate and they show almost no homologies. Thus, we suggest that the structural homogeneity of the three carp α -globin genes Nos. 3, 6, and 7 is a distinctive feature of the carp α -globin gene family. This must be due to the fact that the carp genome is of tetraploid origin.¹⁸⁾

The previously reported amino acid sequence of the German mirror carp α -globin chain,²⁾ and deduced amino acid sequences from carp chromosomal genes Nos. 3, 4,³⁾ 6, and 7, and carp α -globin cDNA¹⁾ were aligned (Fig. 3). The predicted amino acid sequences from human α 1,¹⁴⁾ duck adult α A,¹⁵⁾ and frog adult α 1¹⁶⁾ globin genes

were also aligned. In comparison with these distinct α -globin polypeptides, the carp α -globins have one insertion of amino acid residue at position 47(Arg or Asp). The insertion of amino acid residue position 47(Arg or Asp) was also found in a tentative α -globin sequences constructed from amino acid composition data in other fish including trout *Salmo irideus*,¹⁹⁾ mountain sucker *Catostomus clarkii*,²⁰⁾ shark *Heterodontus portusjacksoni*,²¹⁾ and coelacanth *Latimeria chalumnae*,²²⁾ but was not seen in the human,¹⁴⁾ duck,¹⁵⁾ or frog.¹⁶⁾ This position corresponds to the interhelical region between the C and E helix. We suggest the insertion of position 48 allows only inconsequential changes in function and the three-dimensional structure of hemoglobin. In the human α 1-globin polypeptide, the functional roles of individual amino acid residues have been closely investigated.²²⁻²⁵⁾ In the five predicted carp α -globin amino acid sequences of cDNA, Nos. 3, 4,³⁾ 6, and 7, 15 of haem interfaces,²⁴⁾ 15 of 16 α 1 β 1 interfaces,²⁴⁾ 13 of 15 α 1 β 2 interfaces,²⁴⁾ and 4 of 6 residues involved in the Bohr effect²⁶⁾ were preserved. In comparison with the human α 1 and predicted carp α -globin polypeptide, of the amino acids that were expected to have a relationship with a Pertz mechanism,²³⁾ 12 of 15 haem interfaces, 8 of 16 α 1 β 1 interfaces, and 11 of 15 α 1 β 2 interfaces were similar, but all 6 Bohr effect residues were inconsistent (see Fig. 3). This suggests that there were inconsequential structural and functional differences, but that there was some characteristic variation among human α 1 and five carp predicted α -globins.

Currently, we are determining the nucleotide sequences of fragment Nos. 1, 2, and 5 that are constituent of carp α -globin gene family, to ascertain the peculiarities of each component. In addition, we are undertaking further studies to determine the linkage map of the carp α -globin gene cluster (if they exist) to elucidate the relationship of the genetic structure and gene expression.

Acknowledgements

This research was supported in part by a grant for Pioneering Research Projects in Biotechnology from the Ministry of Agriculture, Forestry and Fisheries, Japan.

References

- 1) S. Takeshita, T. Aoki, Y. Fukumaki, and Y. Takagi: Cloning and sequence analysis of a cDNA for the α -globin mRNA of carp, *Cyprinus carpio*. *Biochem. Biophys. Acta*, **783**, 265–271 (1984).
- 2) K. Hilse and G. Braunitzer: Die aminosäuresequenz der α -Ketten der beiden hauptkomponenten des Karpfenhaemoglobins. *Hoppe-Seyler's Z. Physiol. Chem.*, **349**, 433–450 (1968).
- 3) K. Nobuta and T. Aoki: Cloning and complete nucleotide sequence of carp α -globin gene. *Nippon Suisan Gakkaishi*, **58**, 49–54 (1992).
- 4) J. Vieira and J. Messing: Production of single-strand plasmid DNA, in "Methods in Enzymology" (ed. by R. Wu and L. Grossman), Vol. 153, Academic Press, London, 1987, pp. 3–11.
- 5) F. Sanger, S. Nicklen, and A. R. Coulson: DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467 (1977).
- 6) M. Hattori and Y. Sakaki: Dideoxy sequencing method using denatured plasmid templates. *Anal. Biochem.*, **152**, 232–238 (1986).
- 7) R. Breathnach and P. Chambon: Organization and expression of eucaryotic split genes coding for proteins. *Annu. Rev. Biochem.*, **50**, 349–383 (1981).
- 8) G. C. Grosfeld, C. K. Shewmaker, P. Jat, and R. A. Flavell: Localization of DNA sequence necessary for transcription of the rabbit β -globin gene in vitro. *Cell*, **25**, 215–226 (1981).
- 9) P. Dierks, A. van Ooyen, M. D. Cochran, C. Dobkin, J. Reiser, and C. Weissman: Three regions upstream from the cap site are required for efficient and accurate transcription of the rabbit α -globin gene in mouse 3T6 cells. *Cell*, **32**, 695–706 (1983).
- 10) R. M. Myers, K. Tilly, and T. Maniatis: Fine structure genetic analysis of β -globin promoter. *Science*, **232**, 613–618 (1986).
- 11) M. Poncz, E. Schwartz, M. Ballantine, and S. Surrey: Nucleotide sequence analysis of the ζ -globin gene region in humans. *J. Biol. Chem.*, **258**, 11599–11609 (1983).
- 12) C. S. Chiou, H. T. Chen, and W. C. Chang: The complete nucleotide sequence of the growth-hormone gene from the common carp (*Cyprinus carpio*). *Biochem. Biophys. Acta*, **1087**, 91–94 (1990).
- 13) H. T. Chen, C. S. Chiou, and W. C. Chang: Cloning and characterization of the carp prolactin gene. *Biochem. Biophys. Acta*, **1088**, 315–318 (1991).
- 14) A. M. Michelson and S. H. Orkin: The 3' untranslated regions of the duplicated human α -globin genes are unexpectedly divergent. *Cell*, **22**, 371–377 (1980).
- 15) J. Niessing, C. Erbil, and V. Neubauer: The isolation and partial characterization of linked α -A- and α -D-globin genes from a duck DNA recombinant library. *Gene*, **18**, 187–191 (1982).
- 16) J. Stalder, U. Wirthmueller, J. Beck, A. Gruber, W. Meyerhof, W. Knoechel, and R. Weber: Primary structure and evolutionary relationship between the adult α -globin genes and their 5'-flanking regions of *Xenopus laevis* and *Xenopus tropicalis*. *J. Mol. Evol.*, **28**, 64–71 (1988).
- 17) J. Lauer, C. K. J. Shen, and T. Maniatis: The chromosomal arrangement of human α -like globin genes: Sequence homology and α -globin gene deletions. *Cell*, **20**, 119–130 (1980).
- 18) W. Engel, J. Schmidtke, and U. Wolf: Diploid-tetraploid relationships in teleostean fishes, in "Isozymes" (ed. by C. L. Markert), Vol. 4, Academic press, New York, 1975, pp. 449–462.
- 19) F. Bossa, D. Barra, R. Petruzzelli, F. Martini, and M. Brunori: Primary structure of hemoglobin from trout (*Salmo irideus*); Amino acid sequence of α chain of Hb trout I. *Biochem. Biophys. Acta*, **536**, 298–305 (1978).
- 20) D. A. Powers and A. B. Edmundson: Multiple hemoglobins of catostomid fish; II. The amino acid sequence of the major α chain from *Cetostomus clarkii* hemoglobins. *J. Biol. Chem.*, **247**, 6694–6707 (1972).
- 21) A. R. Nash, W. K. Fisher, and E. O. P. Thompson: Haemoglobins of the shark, *Heterodontus portusjacksoni*; II. Amino acid sequence of the α -chain. *Aust. J. Biol. Sci.*, **29**, 73–97 (1976).
- 22) T. Gorr, T. Kleinschmidt, and H. Fricke: Close tetrapod relationships of the coelacanth *Latimeria* indicated by haemoglobin sequences. *Nature*, **351**, 394–397 (1991).
- 23) M. F. Perutz: Stereochemistry of cooperative effects in haemoglobin. *Nature*, **228**, 726–739 (1970).
- 24) J. Baldwin and C. Chothia: Haemoglobin: The structural changes related to ligand binding and its allosteric mechanism. *J. Mol. Biol.*, **129**, 175–220 (1979).
- 25) G. Fermi: Three-dimensional fourier synthesis of human deoxyhaemoglobin at 2.5 Å resolution: Refinement of the atomic model. *J. Mol. Biol.*, **97**, 237–256 (1975).