Correction. In the article "Immunoglobulin heavy chain variable region gene evolution: Structure and family relationships of two genes and a pseudogene in a teleost fish" by Melanie R. Wilson, Darlene Middleton, and Gregory W. Warr, which appeared in number 5, March 1988, of Proc. Natl. Acad. Sci. USA (85, 1566–1570), the authors request that the following correction be noted. Three immunoglobulin \( V_H \)-related sequences from the goldfish, *Carassius auratus* were presented. These were of two apparently functional genes (gene 3, GenBank accession no. J03616, and gene 5A, GenBank accession no. J03617) and one pseudogene (5B, GenBank accession no. J03618). It has become apparent that two of these sequences (gene 5A and pseudogene 5B) are of higher primate (probably human) origin. This conclusion is based on the observations that (i) the sequences are very close to those of some reported human \( V_H \) sequences (compare 5A with sequence 60P2 of Schroeder et al. (25), (ii) the recombinant \( \lambda \) phage in which genes 5A and 5B are present also contains a typical Alu repeat sequence, and (iii) although we have not been able to find any evidence of contamination that could explain our results, we have also been unable so far to find sequences 5A and 5B (or closely related sequences) in goldfish other than the two individuals reported in this study. Therefore, we have no conclusive evidence for an interspecies gene transfer. The third goldfish sequence that we presented (gene 3), which was cloned from the same library as genes 5A and 5B (9), does not show close similarity to any known human (or other vertebrate) \( V_H \) sequence in GenBank (identity values are 70% or less), and the recombinant \( \lambda \) phage in which gene 3 is present (total insert size, \(~18\) kilobases) does not contain any human repetitive DNA sequences but does hybridize with repetitive goldfish DNA.

Immunoglobulin heavy chain variable region gene evolution: Structure and family relationships of two genes and a pseudogene in a teleost fish

MELANIE R. WILSON, DARLENE MIDDLETON, AND GREGORY W. WARR*

Department of Biochemistry and Molecular Biology, Medical University of South Carolina, Charleston, SC 29425

Communicated by Susumu Ohno, October 2, 1987

ABSTRACT Nucleotide sequences for two immunoglobulin heavy chain variable region (VH) genes and one pseudogene in the goldfish (Carassius auratus) and the family relationships and distribution of these genes in individual fish are presented. Comparison of the nucleotide and inferred amino acid sequences of goldfish and other vertebrate VH genes indicates that goldfish VH genes show the major VH gene regulatory and structural features (5'-putative promoter region, split hydrophobic leader, three framework and two complementarity-determining regions, and 3'-recombination signals for VH to diversity region joining) and that goldfish VH genes are not more closely related to one another than they are to VH genes of evolutionarily distant vertebrates such as the mammals. Goldfish VH genes appear to exist in distinct families, and individual goldfish can carry from none to apparently >15 genes of a given family. These results suggest that whereas the basic structure of VH genes has been conserved in evolution, there may be substantial variation in the nature and population distribution of VH genes in vertebrates. The great diversity of antibody specificities is derived from multiple germ-line copies of the gene segments encoding the antigen-binding site, the variety of possible recombinations of these elements, and somatic events that introduce mutations and junctional imprecision during recombination (1–3). Of the three elements [heavy chain variable region (VH), diversity, and heavy chain joining region] encoding the binding-site-containing domain of the immunoglobulin heavy chain in mammals, the VH genes are present in the greatest number in the germ line and have also been suggested to encode the region that is the primary determinant of binding-site specificity (4). In the course of vertebrate evolution, the immune system appears to have increased in complexity, for example, in terms of the appearance of multiple structurally diverse classes of antibody (5), and in its efficiency, e.g., the appearance of antibodies with higher-affinity binding sites, and in the phenomenon of rapid increase of antibody affinity during an antibody response. A complete understanding of the evolution of the immunoglobulin molecule will be dependent on our knowing the structure, diversity, organization, and expression of the immunoglobulin gene elements in diverse vertebrate groups. Whereas the observation of the unusual organization of VH, diversity, heavy chain constant, and joining region genes in an elasmobranch (6) in multiple repeating clusters of ~10 kilobases is clearly consistent with the restricted antibody response seen in this species (7), it is not known if this pattern of gene organization is restricted to the elasmobranchs or is representative of the ectothermic vertebrates in general. The teleost fish are one of the major vertebrate classes for which large gaps exist in our knowledge. No information on the primary structure of their immunoglobulin molecules or the genes encoding them is available. Here we describe the structure of VH genes in a teleost fish, the goldfish (Carassius auratus), and present evidence that the organization and expression of VH gene families in this species is different from what has been observed in the mammals.

MATERIALS AND METHODS

Cloned Genomic DNA of the Goldfish. The construction of a genomic DNA library in phage λ EMBL4, its screening under P1/EK1 containment, and the initial characterization of seven recombinant phage showing cross-hybridization with the S107 murine VH probe (8) have been described elsewhere (9). DNA fragments of interest were subcloned and sequenced by the dideoxynucleoside triphosphate chain-termination method (10).

DNA Blotting. Genomic DNA (10 µg) or phage DNA (1.0 µg) was digested to completion with restriction enzymes, electrophoresed in one dimension on agarose gels, and blotted onto nitrocellulose sheets by standard techniques (11), as described (9). DNA probes were nick-translated to a specific activity of 2–5 × 10⁶ cpn/µg, and hybridization was carried out exactly as described (9). Filters were washed four times for 30 min under low-stringency conditions (22°C; 0.15 M NaCl/0.015 M sodium citrate/0.1% NaDODSO₄) or high-stringency conditions (65°C; 0.015 M NaCl/0.0015 M sodium citrate/0.1% NaDODSO₄) before drying and exposing to x-ray film at −70°C.

RESULTS AND DISCUSSION

Structure of VH Genes in the Goldfish. The functional regions and sequencing strategy for two VH genes and one pseudogene cross-hybridizing with the murine S107 probe are shown in Fig. 1A. Gene 5A and pseudogene 5B were derived from the same recombinant λ phage, a partial restriction map of which is shown in Fig. 1B. Segments 5A and 5B were separated by ~5.7 kilobases and were in the same transcriptional orientation. The nucleotide sequences (Fig. 2) indicated that the hybridizing regions of 5A and 5B represented apparently functional VH genes. They possessed 5′-flanking sequences typical of VH promoters, including the highly conserved (12) octameric sequence (ATGCAAAT, position −145, Fig. 2) and possible “TATA boxes” (TTAAT, position −117 in gene 5A, and ATGAAAA, position

Abbreviation:VH, immunoglobulin heavy chain variable region gene.

*To whom reprint requests should be addressed: Department of Biochemistry and Molecular Biology, Medical University of South Carolina, 171 Ashley Avenue, Charleston, SC 29425.

†The sequences reported in this paper are being deposited in the EMBL/GenBank data base (Bolt, Beranek, and Newman Laboratories, Cambridge, MA, and Eur. Mol. Biol. Lab., Heidelberg) (accession nos. J03616, goldfish VH gene 5A; J03617, goldfish VH gene 5A; J03618, goldfish pseudogene 5B).
Fig. 1. Partial restriction maps of goldfish V_H-positive DNA showing the sequencing strategy for genes 3 and 5A and pseudogene 5B (A) and the localization of gene 5A and pseudogene 5B on recombinant λ phage 5 (B). (A) , regions encoding mature polypeptide; , regions encoding leader; hatched box, regulatory regions (putative promoter octamer sequence and TATA box and 3'-recombination signals). The arrows indicate sequences determined from cloning sites or by extension priming. Sequence upstream of the 5' Pvu II site on gene 5 was determined on a Pvu II–Pvu II partial digest fragment. Sequence upstream of the 5' Pst I site on pseudogene 5B was determined on an Sst I–Sst I fragment (B). Restriction sites: A, Alu I; B, BamHI; D, Dra I; H, Hae III; S, Sma I; T, Taq I; P, Pvu II. (B) Complete and partial digestion products of end-labeled DNA were used to derive the map. The regions containing the S107 cross-hybridizing DNA, which defined the majority of gene 5A and pseudogene 5B and which were bounded by Pst I sites, are indicated by the solid boxes. Gene orientation (noncoding strand) is indicated by the arrows underneath these boxes. L and R, left and right arms, respectively, of the EMBL4 vector.

119 in gene 5A). Genes 3 and 5A, encoded (Fig. 3) in open reading frame, complete V_H genes comprising three framework regions, two complementarity-determining regions, and a hydrophobic 19-amino acid NH_2-terminal signal peptide, the coding region for which was split by a single intron. The DNA on the 3' side of the coding region contained sequences typical of recombination signals [conserved heptamer-23-base-pair (bp) spacer-conserved nonamer] for V_H to diversity region joining. The S107 cross-hybridizing region 5B was an apparent pseudogene (Fig. 2). It contained a stop codon in the first framework-encoding region and lacked the putative promoter region (conserved octamer and TATA boxes), although in all other aspects it resembled a typical V_H gene. Thus, goldfish V_H genes resemble, in their major structural features, the apparently universal pattern of V_H gene structure seen not only in mammals but also in reptiles and elasmobranchs (13, 14).

Translation of the nucleotide sequences for genes 3 and 5A into inferred amino acids (Fig. 3) showed, in an alignment with S107, a clustering of sequence identities in the framework regions (especially near the cysteines) and the much greater divergence of sequence in the complementarity-
Genetics: Wilson et al.  


Fig. 2. Nucleotide sequences of goldfish V_h genes 3 and 5A and pseudogene 5B are aligned with that of mouse V_h gene S107 (8). Within the coding regions, gaps were introduced only within complementarity-determining region 2 (CDR2). Gaps outside the coding region were introduced to allow alignment of the conserved octameric sequence ATGCRAAT (position -145), the start codon, and the conserved variable-diversity recombination signals on the 5' side of the coding region. In addition, some gaps were introduced to allow alignment of obviously similar sequences of region in the intron splitting the leader region. The numbering refers to gene 3. The stop codon of gene 5B in framework region 1 (FRI) is boxed.

determining regions. Alignment of the sequences could be accomplished without the introduction of gaps in any part of the coding region except for the second complementarity-determining region. Variations in the length of the second complementarity-determining region are characteristic of distinct V_h gene families (15) and suggest that genes 3 and 5A could be regarded as representatives of distinct families. They are 73% identical in their nucleotide sequences in the coding regions and 67% identical in their inferred amino acid sequence which is below the level (80% identity) suggested by Brodeur and Riblet (16) as characteristic of members of V_h families in the mammals. The close similarity in structure of gene 5A and pseudogene 5B (96% nucleotide identity in the coding region) and their close association in the DNA
DNA during no cal recombinant phage analysis to related determine the number and processes of April/May function, and always splits the is not notable the data in present closest other hand at hand essential amino acid leader S107, in the case of gene 3, there was a 76% identity with KOL (17), and in the case of gene 5B there was an 86% identity with BUT (18). However, there was an even closer similarity observed in a comparison of leader regions, involving gene 5A and human gene H11 (19) in which there was 96.5% identity at the nucleotide level and 89.5% identity at the inferred amino acid level. While the strong conservation of an 18- or 19-amino acid leader peptide in heavy chains is notable since apparently approximately one in five of randomly generated peptides can function as signal peptides for protein secretion (20), it is perhaps more surprising that there is conservation, in the V<sub>H</sub> genes, of the intron that always splits the leader-encoding region after the first base of the codon for amino acid 4. This intron has no known function, and its presence suggests that, in the course of V<sub>H</sub> gene evolution, mechanisms other than those acting to maintain on the one hand functionally essential structures and on the other hand a sufficient diversity of binding sites to meet antigenic challenge may have operated. One possible mechanism is molecular drive (21) that could have acted by processes of nonreciprocal exchange to maintain V<sub>H</sub> gene structures independently of their selective value.

V<sub>H</sub> Families Related to Genes 3 and 5A in the Goldfish. To determine the number and distribution of V<sub>H</sub> genes closely related to genes 3 and 5A, Southern blot hybridization analysis under high-stringency conditions was undertaken. Of 35 independently derived, unique S107 cross-hybridizing recombinant phage clones recovered from the library, probes for genes 3, 5A, and 5B hybridized only with the phage in which they were cloned (data not shown). Hybridization of clone 5A back onto restriction digest fragments of the original DNA from which the library was constructed (Fig. 4) showed only two hybridizing components of identical size to those present in the digest of the phage, suggesting that no rearrangement of the DNA in this region had occurred during construction or propagation of the library and that phage clone 5 contained all the members of the V<sub>H</sub> gene family present in the original genomic library.

When the hybridization of probes for genes 3 and 5A on DNA from other individual goldfish was tested, it was observed that 14 out of 15 fish lacked sequences cross-hybridizing under high stringency conditions with these probes. The DNA from the 15th fish, however, showed multiple restriction fragments hybridizing with probes for gene 3 and gene 5A (Fig. 5). The patterns of hybridizing fragments for the two probes appeared to overlap significantly but not completely (Fig. 5). The Southern blot hybridization patterns for this DNA were essentially identical under low- and high-stringency conditions (data not shown) and always rather blurred (in six repeated experiments), indicating the possibility of multiple hybridizing fragments with mobilities close enough to form broad overlapping bands. There appeared, by conservative estimate, to be a minimum of 15 bands hybridizing with the gene 5A probe and 9 bands hybridizing with the gene 3 probe.

Thus it appears that the goldfish V<sub>H</sub> genes differ significantly from those of the mammals in terms of their family relationships, both within and among individuals. In the mouse, for example, there appears to exist a much larger number of V<sub>H</sub> genes than we were able to detect with heterologous or homologous probes in the goldfish. The number of mouse V<sub>H</sub> genes, originally estimated in the hundreds (16), now appears likely to be >1000 (22). These V<sub>H</sub> genes have been divided into nine families on the basis of

---

Fig. 3. Inferred amino acid sequences of goldfish V<sub>H</sub> genes 3 and 5A and pseudogene 5B. The alignment of the sequence with that for the mouse V<sub>H</sub> S107 (8) is identical to that shown in Fig. 2. The one-letter amino acid code is used.

---

Fig. 4. Southern blot hybridization analysis of genomic DNA with goldfish V<sub>H</sub> gene 5A. DNA (10 μg) from which the library was constructed and DNA (1 μg) from λ phage clone 5 was digested to completion with Kpn I and Sst I, electrophoresed on 1% agarose gels, and transferred to nitrocellulose for hybridization with the 32P-labeled nick-translated 584-bp Par I-Par I DNA fragment containing the coding sequence for gene 5A. Lanes: A, λ phage 5 DNA (autodigraph exposed for 2 hr); B, genomic DNA (autodigraph exposed for 24 hr). Size markers are in kilobases.

---

sequence similarities (16, 23, 24), and these families can also be defined, operationally, by Southern blot hybridization analysis under high-stringency conditions, which detects degrees of nucleotide sequence similarity corresponding to that used to define a $V_H$ family (16). Our studies clearly suggest that distinct families of $V_H$ genes exist in the goldfish. However, in view of the relatively close similarity between gene 3 and gene 5A, it may be premature to conclude that they are not members of the same family. In the mouse, all individuals appear to possess genes of all $V_H$ families (24), and, although the size of the $V_H$ gene families varies substantially, variations in the number of genes of any one family possessed by different mice is usually small, leading to the conclusion that in the mammals $V_H$ gene families have evolved by the gradual gain and loss of small numbers of $V_H$ genes (24). The situation in the goldfish appears quite different. Genes 3 and 5A seem to belong to a $V_H$ gene family (or families) with numbers that vary greatly within the population of goldfish, even being apparently absent in significant numbers of individuals.

Thus, while we can conclude that the goldfish possesses apparently typical vertebrate $V_H$ genes, our initial observations on their organization into $V_H$ families would lead us to suggest that the mode of evolution of these families, on a population basis, seems likely to have been quite different, in a teleost fish, from what has been seen for the mammals.

This work was supported by grants from the National Science Foundation (PCM 8408484 and 8709877).