The immunoglobulin M heavy chain constant region gene of the channel catfish, *Ictalurus punctatus*: an unusual mRNA splice pattern produces the membrane form of the molecule

Melanie R. Wilson, Ann Marcuz, Frederick van Ginkel¹, Norman W. Miller¹, L. William Clem¹, Darlene Middleton² and Gregory W. Warr^{2,*}

Basel Institute for Immunology, Grenzacherstrasse 487, CH-4005 Basel, Switzerland, ¹Department of Microbiology, University of Mississippi Medical Center, Jackson, MS 39216 and ²Department of Biochemistry and Molecular Biology, Medical University of South Carolina, 171 Ashley Avenue, Charleston, SC 29425, USA

Received May 2, 1990; Revised and Accepted July 26, 1990

EMBL accession no. X52617

ABSTRACT

The immunoglobulin (IgM) heavy chain constant region gene of the channel catfish, *Ictalurus punctatus*, has been cloned and characterized. The gene contains four constant region domain-encoding exons (CH1 to CH4) expressed in the secreted form of the immunoglobulin, and two exons encoding the transmembrane (TM) domain utilized in the lymphocyte membrane receptor form of the immunoglobulin. The sequence of a cDNA clone encoding the 3' region of the message for the membrane receptor form of the μ chain indicates that the TM1 exon is spliced directly to the CH3 exon, and not into a site within the CH4 exon, as occurs in the mammals, a shark and an amphibian. This unusual pattern of splicing, which produces a membrane heavy chain that is characteristically smaller than the secreted heavy chain, may be common to all teleost fish.

INTRODUCTION

Investigations of the structure of the immunoglobulins (Igs) have revealed that in all vertebrate classes a molecule identifiable as IgM is present. In some of the vertebrates, e.g., mammals, *Xenopus* (an amphibian) and *Heterodontus*, (an elasmobranch) the structure of the μ -chain gene is established: in other vertebrates, e.g., chickens, our knowledge is partial or obtained by inference from cDNA sequences (1–7). When the structure of the μ gene is known, it has always been found to consist of four exons (CH1 to CH4) that encode the secreted form of the heavy chain (μ s) found in the serum IgM. In addition to these exons, an additional two exons are also always found, coding for the transmembrane (TM) domain of the lymphocyte membrane receptor form of the IgM heavy chain, μ m, which includes an approximately 26 residue stretch of hydrophobic or uncharged amino acid residues that is believed to span the lipid

bilayer of the lymphocyte plasma membrane (2,3,5,8,9). Alternate mRNA processing events determine whether or not the secretory or the membrane form of the gene is expressed (2,8). The secretory form of the message normally terminates at its own poly(A) site, but if the membrane form is to be expressed, the 3' region of the CH4 is spliced out along with the CH4-TM1 intron, resulting in the formation of a heavy chain with a hydrophobic 'tail' to its CH4. This μ m heavy chain is directed to the plasma membrane by the appropriate stop-transfer signals present in this different C-terminus (9). The relative expression of μ m and μ s message appears to be controlled by competition between splicing of the TM1 exon into the CH4 exon, and utilization of the μ s poly(A) site (8).

In the teleost fish, which express both a lymphocyte-membrane and a secreted, serum IgM (10–14) the sequences of the complete μ s cDNAs from 2 species are known: *Elops saurus*, the ladyfish and *Ictalurus punctatus*, the channel catfish, (12–14). It is also known that the organization of the V, D, J and C regions encoding the μ s is essentially similar to that of the mammals (12,14), rather than to that of the elasmobranchs (3). We present here a detailed analysis of a constant region gene of a teleost fish IgH locus, and show that although it possesses the same exon structure as other known vertebrate μ constant region genes, its pattern of exon-splicing in the expression of μ s and μ m forms is different from that employed by the mammals, shark and the amphibian, *Xenopus* (2,3,5,8 and Du Pasquier and Schwager, personal communication).

METHODS

cDNA library construction.

A channel catfish lymphocyte cDNA library was constructed in λ gt11, using polyA positive mRNA obtained by the guanidinium thiocyanate procedure and oligo dT chromatography from 8 day

* To whom correspondence should be addressed

1 TAAATGTTGGTGTGTCTCGGCTATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTATCTATCTAAGTATCTATCATACAGCT
CH1
101 GTGCAAAGCGCCCCGAAATCCCTGTTTCCCCTGTGGCAGTGCGGCTCGGCCTCGGACGGTTTAGTCACTCTTGGCTCGCTCACGCGCGATTGGCCCTCGG
V Q S A P K S L F P V W Q C G S A S D G L V T L G C V T R D L A S A
201 CCGACGGACTGAGCTTCATATGAAGGATGCGAGCGGGAGCGCGCTGACTGACGTCGTGCAATACCCGGCGGTGCAAGCGACCGGAGGTTACACCTCGGT
D G L S F I W K D A S G S A L T D V V Q Y P A V Q A T G G Y T S V
301 GAGCCATGTGCGCGTCAAGGCTTCTGACTGGAACGGGAACAAGAATTCACGTGCGAAGTCAAAAATGGCCCTAGGATCTAAAGACGCGTCTTGC AAAAG
S H V R V K A S D W N G N K F T C E V K N G L G S K D A S L Q K
401 CCAGGTAGGCGAGAAACAACCTGAAAAACCCCTTTTGTGTGTCATTCCAATATGCAATCAGTGCCACATTATTAGGTAACCTACACCTGCCCATACA
P
501 AACACTTTACTAAACTTTCCTTGATACTGGGGTGATACAGTCATGGGTAGTTTTTGTACCTTTTAACTCTGTATGTTTTACCTGGAAACATGAAACCAGT
601 GCACCTTTTAAAAATAATTTAAGGCACACAATTTGGACGTGAAGGCCAGTGATGTACATTTACAGCTTTATTCTAAAAGCTAATTTAAGGTACTTCACGT
701 GCACCTCCCTCGGTAAATATACATATTTAAAGATATCAAAGACTTGATTGTATGTTACCTAGATTCTGACACCGGTGTGATGAATGGTCAATG
801 CACACACACACATACACACACACACACACACACACACACATCATATTTATACCTTAGCAAGCAAAACAAAAGTAATATATATGCTAAAAAC
901 AAATAAACATATGTTTATTATTGTTTACTCTGCTTAAACATTATGATTCATTACAGATATAGTACAGAAGTGTTTATTCTTTATAACACCGTTTTAGC
1001 TGTAACCTATAAAATTCCTGTAAAACACATTATTTAAATAGAAAAGTTGAAATCAGCAGCAACCATTGAATTGTTAAAAAAAAAAAAAAAAAGGTTATTTT
1101 CTATACTGGCACCACATCAGAGTTTCTGGTTATGCAATCAGAGAGATTA AAAAATT CAGATACAACACTTTGT TCCCCACTAGCCTCAAAGATGATATAGAT
1201 GGTTACAAATTTATAATAATTTAGTGTTAATATTTCTCTTAGCTACATACACAATCTAATTACAGTCTGAGTGTGAGCATTATTCTTTAAGAAAAATGATGC
1301 AAAGGATGTGAAATTTGACTGTAAACAGAATAATTGTCTACAAAAGTTTCAATTCCTTCAAGTCCTTTTTTAAAGAAGATGTTGCATGTGTAATCAGACTG
1401 AGACAGTCTGCTTTGGCCATTTAAAGTGCTGACTAAAAGTTTATTATTCTTAAATGTGTATCCATTTTTTTCTGACTTTTCAGTCCAGCTAGCAATTTT
1501 TTATGTCAGAGTCAAACATTTTAAATTTCAAAGCTTACGTATACGTGTGGTCTTTCTCTGCCAATGCTAAACAAAGCCTGTTCCCTTTTTACCTGTG
1601 CCTTATGACTTCTTGCATGTTTGTGTCACAGTATGTCTATCAAACATATGCAAAGTTTGCTGTGGCATCAACAGTTCTAGGTGATAAATTAGATTTTA
1701 CAGACTAGTAACTATAATTTAAGGGTTTCTCAGCCTCAGTTCAGAGAAATTAACAGTGCAAAATCCTCTGTCTTTCCATGGTTCTGAAACTGGAGACT
1801 TCTTCCATAAAATGTTAAATAATGTCTCTTATAGAAAGCTTTACAATGTCAATGATTGTAACATTTTTTGATAAATAACACATTTTTTTAAAGTGTTT
1901 ATTATTAGGCTTATAGAAATGTCAACTACACCAGTACCTGCATAATTGTTTACTATAGAAACAATAACATATCGAGTGCATTAATACAAAATATAGATAATT
2001 TATTAACATTAATTGAGAATTTGACAGCATTCTGATATAAAATTTAGATATTATGTGGTGATAAAAAATAGCATTTTTGTGACCTTATATACTGTGGCAACAC
CH2
2101 AACCTATTTCTAAACTGCAATACCTGCTGTGCCCATTTGCTGCTGATATGGATCTGTGTGTTTGTGTTTTGCAGTTGAGAGAGAGCTCCATGCATCTCTGCTT
V E R E L H A S L L
2201 CTAACAACTCCAACCCAAACAGAAATAGACAATGGAACAGCTACCTTCGTCTGCTTAGCTACACCATTTTCACCTAAATCACACACATTTAAGTGGACTC
L T T P T Q T E I D N G T A T F V C L A T P F S P K S H T F K W T L
2301 TTGAAAAGACGACATCAGTAATAAGGTCAAAGAGAACATAGTAAGCCAGAATAAAGGTAACTTCTAGTCCATAAGTGTTTTGGAAGTACAGCGCCAGTGA
E K T D I S N K V K E N I V S Q N K G N F T A I S V L E L S A S E
2401 ATGCAACAGCTCAACTTCTCCAGTTAAGTGCGAATTCAGCAGAAGAACATAATGTGTTCAAAGAAGCGAGTTATGGTATGTATTATATTATTCTATT
W T S S T S P V K C E F Q Q K N H N V F K E A S Y
2501 CATTA AAAAAAAAATTTTTTAGCAACTTTCAATTTGACAGACCGTCTGAACAGCTATATTACTA ACTTCTTCAAAGAAGCAGTCTATCTCAATTGATTAA
2601 AACTTCTCTATCTGGGTTTTAGCTTGATAGTATACCTACACGATAATATGAGGATCTGTTTTTGATGAAAACTACAACACACTTGCTAAATGCTGTAA
2701 ATGTAAATTTGCTTTAAAGTAATGATTTGCATTCTAGTTTGAATAAAAATGTCAATTATGTATCAGTTATATGTTTATTATTAATGGTAATGTGCACTAA
2801 AACTCTAGATTTTATGGCACATTTTCAGATTATACATTCCTGTGTTTTTGTGTTTTGTTTATTGTTTTTCAAATACTCCTTGCTTCTTTAAAAAA
CH3
2901 TGTATTCTACAACCTCTTTACAGCACCAGGTGACACAAAACAGCCACAGGTGAAAATAACTGGACCTTCCACCGAGGACATTCTGATCAAAAGAGCCGG
A P G D T K Q P Q V K I T G P S T E D I L I K R A G
3001 CCAGCTCGAGTGCAGGGCCGAGGAGACACGGGTTTCAAGAGCATTAATGGCTTATTGGAAATAGAGAGATCTCTTCTATCAAATCTATCTTTCGAAA
Q L E C R A E G D T G F K S I K W L I G N R E I S S L S S K
3101 ACGACGGTTTCTACTCCAAACCCACATCGTGTTCGAAGAGTGGATCAATGGCACCGAATTCATCTGTGAGGTGGAACATGAAGCATTCACTCAACAGTATG
T T V S L Q T H I G F E E W I N G T E F I C E V E H E A F T Q Q Y E
3201 AAAAAGTAACCTTCAAAGAGAAAATGGTAGGAAATGCTTTTACTACTTATACATATTCTTGCTCTTAATTTGTATGAAAAGCTATGAGGATTTTTAAG
K V T F K R E N
3301 AAGCAGAAAATCACATTTTTAAAGCTATCTAATTATATTGAAATTATATGCTAATTATTTTTGTTTCGTTTACATATGACCAGTGTTCCCAACATACGG
CH4

Fig. 1. Nucleotide and inferred amino acid sequence of the catfish IgM heavy chain gene. The exons encoding the four secretory domains (CH1, 2, 3, and 4) and the two C-terminal transmembrane domains (TM1 and 2) are indicated. The inferred amino acid sequence (single letter code) is indicated under the first base of each codon, and stop codons are indicated by an asterisk. The splice sites, inferred by comparison with the cDNA sequence of ref 14, are indicated by underlining. The first amino acid of exons CH2, CH3 and CH4 and TM1 is indicated for the codons formed at the splice junctions. The first amino acid encoded by exon CH1 is unknown since the first codon of CH1 is determined by the J region utilized during VDJ rearrangement. The polyadenylation signal sequences for the secretory and transmembrane mRNAs are indicated by underlining. The two clones isolated from the cDNA library by antibody screening corresponded to the coding regions between bases 2432 and 3154 (clone 104, encoding 15 residues of CH2 and 77 residues of CH3) and bases 3470 and 4022 (clone 105, encoding 127 residues of CH4 and 172 bases of 3' untranslated sequence).

Fig. 1. Nucleotide and inferred amino acid sequence of the catfish IgM heavy chain gene. The exons encoding the four secretory domains (CH1, 2, 3, and 4) and the two C-terminal transmembrane domains (TM1 and 2) are indicated. The inferred amino acid sequence (single letter code) is indicated under the first base of each codon, and stop codons are indicated by an asterisk. The splice sites, inferred by comparison with the cDNA sequence of ref 14, are indicated by underlining. The first amino acid of exons CH2, CH3 and CH4 and TM1 is indicated for the codons formed at the splice junctions. The first amino acid encoded by exon CH1 is unknown since the first codon of CH1 is determined by the J region utilized during VDJ rearrangement. The polyadenylation signal sequences for the secretory and transmembrane mRNAs are indicated by underlining. The two clones isolated from the cDNA library by antibody screening corresponded to the coding regions between bases 2432 and 3154 (clone 104, encoding 15 residues of CH2 and 77 residues of CH3) and bases 3470 and 4022 (clone 105, encoding 127 residues of CH4 and 172 bases of 3' untranslated sequence).

DNA sequencing

The dideoxynucleoside triphosphate chain-termination method (21) was used to derive the nucleotide sequence (from both strands) of cDNA and genomic clones of interest. DNA was subcloned into commercially available vectors (M13Rf, pGEM 4, Promega, Madison, WI, and pBluescript, Stratagene, La Jolla, Ca.). Both double-stranded (lambda phage, plasmid) and single-stranded (M13 phage) sequencing were performed using a) priming from vector sites flanking the inserted DNA, or b) sites internal to the sequence using synthesized oligonucleotides. ^{35}S - α -thio dATP and modified T7 DNA polymerase (Sequenase, United States Biochemical Corp., Cleveland, OH) were used for the sequencing reactions.

Analysis of leucocyte surface-labelled proteins

Isolated peripheral blood leucocytes from the channel catfish were surface radio-iodinated with ^{125}I as iodide using the lactoperoxidase method and solubilized in 0.5% NP40 containing 2.0M urea at pH 8.0 as described elsewhere (11). Labelled Ig was then isolated from the lysate using monoclonal antibodies (directed to determinants on the heavy chain of the serum IgM), and analysed by sodium dodecylsulfate polyacrylamide gel electrophoresis (11,22).

RESULTS

Cloning of the Catfish IgH Constant Region Gene

The antigen-stimulated channel catfish lymphocyte cDNA library was screened with a rabbit antibody to catfish IgM and a number of positive clones were isolated. Two of these were sequenced and compared to the published sequence of the secreted catfish heavy chain (μs) cDNA (13,14). One of the clones (104) corresponded to 15 codons of CH2 plus 77 codons of CH3, and another clone (105) corresponded to the terminal 127 codons of CH4 and 172 bases of untranslated 3' sequence (the sequences of these clones are indicated in the legend to Fig 1). These pieces of μs cDNA were used to screen the channel catfish genomic library, and from the positive phage one clone was identified (12c) that proved to contain the entire coding region for the constant region of the catfish IgM heavy chain: the four exons encoding domains CH1 through CH4 of μs and the two exons encoding the transmembrane (TM) domain that is found in the μm molecule. The sequence of this region is shown in Fig 1, and the physical map of the genomic insert in recombinant phage 12c is shown in Fig 2. The four exons encoding the μs , their splice sites and the polyadenylation signal, were identified by comparison with the published complete heavy chain cDNA sequence (13,14). One characteristic structural feature of μ constant region genes, the CH4 internal splice site for CH4 to TM1 joining (8), i.e. AGGT or its consensus, found about 20 codons from the 3' end of CH4, appears to be absent (Fig. 1: cf refs 8,13). The transmembrane exons were identified by the following means. The genomic sequence 3' of the μs polyadenylation site was searched for open reading frames of approximately 40 amino acids showing the typical TM1 pattern of a cluster of acidic residues, followed by approximately 26 hydrophobic or uncharged residues followed by a basic residue. Such a region was identified between bases 5463 and 5876 (Fig 1). This region falls within a small SalI to EcoRI fragment of 156 bp. This fragment was isolated and used as a probe to screen the original cDNA library. One strongly-hybridizing clone (18.1) was identified, isolated and sequenced. The sequence (Fig 3) is

Fig. 2. Partial restriction endonuclease map of recombinant genomic channel catfish lambda phage 12c. The boxes indicate the exons encoding the first four heavy chain constant region domains (CH1–4) and the two exons encoding the alternate transmembrane anchoring C-terminus of the heavy chain (TM1 and 2). Restriction sites are indicated by S, Sal I; E, Eco RI; H, Hind III. L and R indicate the left and right arms, respectively, of the EMBL 3 vector phage. Polyadenylation signal sequences are indicated by filled O.

clearly that of the 3' end of the message for the μm and allowed identification, within the genomic sequence, of the full TM1 exon, the small TM2 exon which encodes the valine and lysine residues typical of the C-terminus of μm chains (15) and the polyadenylation signal for the μm message.

Unusual Splicing Creates the μm Message

An unusual feature of the cDNA sequence of 18.1 is that it clearly indicated that the TM1 exon was joined directly to the splice site 3' of the CH3 exon rather than into an internal splice site near the 3' end of the CH4 exon (Fig 3). However, the μm message of the channel catfish appeared to be quite rare (18.1 was the only TM-encoding clone that we were able to detect) and it was therefore decided to determine if the expressed μm polypeptide showed properties consistent with its production from a message spliced as suggested by the sequence of 18.1.

If the μm of the channel catfish was in fact encoded only by exons CH1-CH2-CH3-TM1-TM2, the polypeptide product would be expected to be 83 amino acids shorter than the μs , a calculated difference in M_r of approximately 9000. In order to determine if this difference could be observed in the μm polypeptide, proteins of the peripheral blood leucocytes of the catfish were surface radio-iodinated with ^{125}I as iodide using lactoperoxidase, lysed, and the surface Ig isolated by using four individual monoclonal antibodies directed to catfish serum IgM. The isolated polypeptides were reduced and analyzed by sodium dodecylsulfate polyacrylamide gel electrophoresis (Fig 4A). The four lanes in which material isolated using the monoclonal antibodies was electrophoresed, all showed prominent bands whose migration was consistent with their identification as Ig heavy and light chains. The putative μm heavy chains migrated significantly faster than the heavy chain of the control, serum IgM from the channel catfish. This is also shown in Fig 4B, where the heavy chain region of a gel in which the μs and μm were run in adjacent lanes, is detailed. This apparent difference in M_r between the putative μm and the μs was calculated to be 10,000 (± 1000) from the results of three separate experiments. The monoclonal antibodies to channel catfish IgM precipitated, in addition to bands identified as μm and light chains, material of unknown nature of approximate M_r 180,000 (Fig 4A). This material has been observed previously in association with catfish lymphocyte surface IgM: it was identified in both reduced and unreduced sodium dodecylsulfate polyacrylamide gels, and may be a polypeptide that associates non-covalently with membrane IgM in the lymphocyte membrane (11). The IgM of channel catfish contains three species of light chain, separable by sodium dodecylsulfate polyacrylamide gel electrophoresis (22). These

Fig. 3. The nucleotide and inferred amino acid sequence of catfish cDNA clone 18.1. The boundaries between the CH3, TM1 and TM2 exons are deduced from the complete genomic sequence (Fig 1) and are shown with their splice sites at the bottom of this figure. Numbering refers to the genomic sequence (see Fig. 1).

Fig. 4 A) Analysis of surface radio-labelled polypeptides isolated from channel catfish leucocytes using monoclonal antibodies to the serum IgM. IgM in lysates of surface radioiodinated channel catfish leucocytes was isolated using monoclonal antibodies to the serum IgM (11,22). Lane 1, antibody 3E11; Lane 2, antibody 1H6; Lane 3, antibody 3D11; Lane 4, antibody 9E1; Lane 5, control mouse ascites; Std, 14 C-labelled standards, (myosin H chain, 200,000; phosphorylase b, 97,000; bovine serum albumin, 68,000; ovalbumin, 43,000; alpha chymotrypsin, 25,700); IgM, 125 I-labelled channel catfish serum IgM precipitated with antibody 9E1. Samples were analyzed under reducing conditions on gels of 10% acrylamide concentration. The band at the top of the Std lane is insoluble material that did not penetrate the resolving gel. B) Comparison of μ_s and μ_m . Radiolabelled serum IgM and surface IgM of channel catfish lymphocytes were precipitated using antibody 9E1 and analyzed, as described in Fig 4A, in adjacent lanes of a 10% acrylamide gel. The heavy-chain region of the autoradiograph is shown. The migration of Coomassie blue stained bovine serum albumin (BSA) is indicated.

have been reported to associate differentially with the heavy chains recognized by the monoclonal antibodies used in these experiments (22) a result that the data of Fig 4A would appear to confirm. The serum IgM standard (radiolabelled and precipitated by a monoclonal antibody) shows two contaminants, of Mr approximately 31,000 and 43,000, whose nature is unknown (Fig 4A). In several of this series of experiments, the Ig isolated from lysates of surface-labelled leucocytes showed, in addition to the putative μ_m , a second band in the heavy chain region. This second band was identical in mobility to the μ_s . In all experiments with surface-labelled leucocyte lysates, the putative μ_m was seen: the heavier μ_s was seen only in some experiments (about one third, data not shown), but always in combination with the putative μ_m chain and never alone. It seems likely that the slower heavy chain was authentic μ_s : either serum

IgM absorbed to leucocyte plasma membranes by Fc receptors (reviewed in 23), or internal IgM, destined for secretion, that was labelled by the lactoperoxidase reaction in damaged or permeable cells.

DISCUSSION

The channel catfish IgM constant region gene conforms to the pattern of organization seen in all other vertebrate μ chain genes examined so far: there are four exons encoding the four protein domains of the secreted heavy chain, μ_s , as predicted by Ghaffari and Lobb (13,14), and in addition there are two exons that encode the transmembrane region of the membrane receptor form of the heavy chain (μ_m).

The lengths of the introns separating these six exons seem to

be quite variable in the vertebrates, and not to conform to any pattern obviously related to the phylogenetic position of the species. For example, in the mouse, the introns separating CH1, 2, 3 and 4 are all less than 500bp (1,2). In the shark, *Heterodontus*, there is a larger intron (about 1.4kb) between CH2 and CH3, (3) whereas in the channel catfish there is a long intron (approximately 1750 bp) between CH1 and CH2. *Xenopus* also shows a relatively long CH1 to CH2 intron, of approximately 1680 bp (5, and du Pasquier and Schwager, personal communication). In all species there seems to be a substantial distance between CH4 and the first of the TM exons (e.g., 2kb in the mouse, 3kb in the horned shark, 1.9kb in the channel catfish), and in *Xenopus* and the channel catfish (but not mouse and shark) there is a large intron separating TM1 from the small (2 codons) TM2 (1.1 kb in *Xenopus*, (ref 5) and 2.3 kb in the catfish). The physical map of the μ -encoding region of the channel catfish IgH constant region locus (Fig 2) is consistent with the southern blot restriction fragment mapping of the CH1 to CH4 exons carried out by Ghaffari and Lobb (14) and this, taken with the sequence identity between the cDNA (14) and our genomic DNA, confirms that we have cloned the gene for the major expressed IgM heavy chain in this species. Ghaffari and Lobb (14) established that while the channel catfish possesses only a single haploid copy of this gene it has multiple copies of the V_H genes, and Amemiya and Litman (12) have shown in a more formal manner that another teleost, the ladyfish (*Elops saurus*), has an IgH locus arranged analogously to that of the mammals. On the basis of these two reports, this is a conclusion that can probably now be generalized to all teleost fish.

The degree of similarity between the inferred amino acid sequences of the CH1, 2, 3 and 4 domains of the channel catfish and those of other vertebrates has been examined previously (14) and identity values from around 21%, to as high as 36% for CH4s were reported. In this regard it is interesting that a comparison of the channel catfish inferred amino acid sequences with those recently published (12) for the constant domains of the ladyfish μ s reveal that while the identity values for CH1, 2 and 3 are 33%, 33% and 25% respectively, for the CH4s it is 42%. Thus, while there appears to be a higher conservation of structure in vertebrate μ s CH4 domains than in CH1–3, the reasons for this are not clear, although they may relate to the role of CH4 in polymerization of the IgM molecule. The presence of three cysteines in both the CH2 and CH3 domains of the ladyfish μ s (12) suggests that these domains may be involved in inter-heavy chain disulfide bonding, a function that they probably do not have in the channel catfish, where the two cysteines in each of these two domains would both probably be involved in an intra-domain disulfide bond. The identity values for the transmembrane domains of diverse vertebrate μ chains are comparable to those for CH4s or even higher: for TM1 the channel catfish shows between 40% and 42% identity (amino acids) to the TM1 of mouse, horned shark and *Xenopus* (2,3,5 and Du Pasquier and Schwager, personal communication), and for the comparison of TM2s the figure is 100%. The TM2 sequence of Val-Lys is almost invariant in vertebrate μ chains (15), and even the nucleotide sequence is highly conserved: the sequence of the channel catfish TM2 codons (Val, Lys, Stop) is identical to that of the mouse (but not of other species). The high identity values rapidly disappear in the TM1 to TM2 intron and the 3' untranslated region of the TM2, and the reason for the conservation of the TM exon (particularly TM2) nucleotide and amino acid sequences is unknown.

However, perhaps the most striking observation of this study is the alternate splicing that occurs in the catfish heavy chain constant region gene transcript, to produce mRNA for the secretory and transmembrane μ -chain forms, and which is quite different from that described for the mammals and the sharks (2,3,8). In the typical vertebrate constant region gene, the μ m message is generated when the TM1 and 2 exons are spliced into the CH4 exon. The results presented here suggest that in the catfish μ m, the TM exons encode 47 residues which are substituted for the entire CH4 domain. The evidence for this is that 1) the typical consensus 5' splice site in the vertebrate CH4 (8) is absent in the channel catfish, 2) the cDNA sequence of a clone (18.1) spanning the 3' end of the μ m fits exactly a splice pattern of CH3 to TM1. The studies of Ghaffari and Lobb (14) have clearly indicated that there is present in the channel catfish haploid genome only a single copy of the CH3 and CH4 exons. We have confirmed this by Southern blot-hybridization experiments (data not shown) and can therefore conclude with some confidence that the sequence of cDNA clone 18.1 arises from an alternate mRNA processing event, rather than expression of a different gene. 3) Surface labelling studies indicate that channel catfish leucocytes express on their plasma membrane a putative μ m heavy chain with a mobility on SDS polyacrylamide gel electrophoresis consistent with the predicted size of the polypeptide encoded under this suggested μ m message splice pattern. Although the possibility exists that the apparent size difference between the mm and ms polypeptides is due to differential glycosylation, the observed Mr difference of around 10000 would require the mm polypeptide to be almost entirely lacking in carbohydrate (which accounts for 12–17% of vertebrate m chain mass). Since two thirds of the potential N-glycosylation sites reside in the CH2 and CH3 domains (14), which would both be common elements of the mm and ms polypeptides, differential glycosylation does not appear to be a likely explanation of the observed size differences between μ m and μ s.

This available evidence would seem to lead to the conclusion that, in the channel catfish, μ m is indeed produced by an mRNA splicing pattern that has not previously been observed in mammals, the horned shark, or the amphibian, *Xenopus* (2,3,5,8). The likely effect of this splice pattern on the covalent quaternary structure of the catfish Ig molecule can be noted. The catfish CH4 appears to encode all the cysteines involved not only in linking together the two μ s chains of each $(\mu L)_2$ subunit, but also the linking of these $(\mu L)_2$ subunits to each other to form the complex variety of covalent IgM structures, up to $(\mu L)_8$, that are seen in catfish plasma (13). It appears that the cysteine encoded near the beginning of the TM1 exon (before the membrane-spanning region) could compensate for the loss of the CH4 cysteines, by providing the possibility of covalent linkage between the two heavy chains of the $(\mu L)_2$ structure that is typical of the membrane form of IgM. In support of this model is the fact that channel catfish B-cell surface IgM has been shown to have a mobility in sodium dodecylsulfate polyacrylamide gel electrophoresis consistent with a covalent $(\mu L)_2$ structure (11). The alternate splicing of the TM exons into CH4 in the mammals (2) serves a similar function: the cysteines forming the covalent inter- $(\mu L)_2$ subunit disulfides are eliminated, while those responsible for the covalent heavy chain to heavy chain linkage in the $(\mu L)_2$ subunit remain.

The pattern of mRNA splicing by which the channel catfish generates the message for its μ m heavy chain may be general

to the teleost fish. The evidence for this is that 1) in the goldfish, *Carassius auratus*, the μ m form of the IgM heavy chain was clearly shown by double-label techniques to migrate faster (an apparent Mr difference of 10,000) in sodium dodecylsulfate polyacrylamide gel electrophoresis than did the μ s form of the heavy chain (24). In the trout (*Onchorhynchus mykiss*), analysis of lymphocyte surface-labelled IgM by similar methods also showed significant μ -chain associated radioactivity migrating faster than the μ s standard (25). 2) In the ladyfish, *Elops saurus*, the CH4 exon lacks an identifiable internal donor splice site that could join to the TM1 acceptor splice site (12), as does the CH4 exon in the channel catfish (13,14 and Fig 1).

The reasons for the deletion of the CH4 exon from the μ m message in the catfish (and probably other teleost fish as well) are unknown. The functional significance of the loss of the CH4 domain from the μ m chains of the catfish B-lymphocyte membrane receptor is also unknown, but obviously does not result in severe compromise of its effectiveness as a receptor for antigen.

ACKNOWLEDGEMENTS

This research was supported in part by grants from the NSF (PCM 8709877) and NIH (1-R01-AI-19530). We thank Drs. Du Pasquier and Schwager for helpful discussion and access to their unpublished data, and Drs. C.-A. Reynaud and R. Perry for helpful comments on the manuscript.

The Basel Institute for Immunology was founded and is supported by Hoffmann-La Roche and Co. Ltd., Basel, Switzerland.

REFERENCES

- McGuire, K.L., Duncan, W.R. and Tucker, P.W. (1985) *Nucleic Acids Res.* **13**, 5611–5628.
- Early, P., Rogers, J., Davis, M., Calame, K., Bond, M., Wall, R. and Hood, L. (1980) *Cell* **20**, 313–319.
- Kokubu, F., Hinds, K., Litman, R., Shambloot, M.J. and Litman, G.W. (1988). *Embo. J.* **7**, 1979–1988.
- Schwager, J., Mikoryak, C.A., and Steiner, L. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 2245–2249.
- Du Pasquier, L., Grossberger, D. and Schwager, J. (1988) Basel Institute for Immunology Annual Report, 40–41.
- Dahan, A., Reynaud, C.-A. and Weill, J.-C. (1983) *Nucleic Acids Res.* **11**, 5381–5389.
- Kawakami, T., Takahashi, N. and Honjo, T. (1980) *Nucleic Acids Res.* **8**, 3933–3945.
- Peterson, M.L. and Perry, R.P. (1989) *Mol. Cell. Biol.* **9**, 726–738.
- Yost, C.S., Hedgpeth, J. and Lingappa, V.R. (1983) *Cell*, **34**, 759–766.
- Warr, G.W., DeLuca, D. and Marchalonis, J.J. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 2476–2480.
- Miller, N.W., Bly, J.E., van Ginkel, F., Ellsaesser, C.F. and Clem, L.W. (1987) *Dev. Comp. Immunol.* **11**, 739–747.
- Amemiya, C.T. and Litman, G.W. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 811–815.
- Ghaffari, S.H. and Lobb, C.J. (1989) *J. Immunol.* **142**, 1356–1365.
- Ghaffari, S.H. and Lobb, C.J. (1989) *J. Immunol.* **143**, 2730–2739.
- Kabat, E.A., Wu, T.T., Reid-Miller, M., Perry, H.M. and Gottesman, K.S. (1987) Sequences of Proteins of Immunological Interest, 4th edition, USDHHS, Public Health Service, National Institutes of Health.
- Miller, N.W. and Clem, L.W. (1984) *J. Immunol. Methods* **72**, 367–379.
- Blin, N. and Stafford, D.W. (1976) *Nucleic Acids Res.* **3**, 2303–2308.
- Weis, J.H. (1987) In Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Smith, J.A., Seidman, J.G. and Struhl, K. (Eds), Current Protocols in Molecular Biology, Wiley Interscience, New York, pp 532–535.
- Feinberg, A.P. and Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13.
- Wilson, M.R., Middleton, D., Alford, C., Sullivan, J.T., Litman, G.W., and Warr, G.W. (1986) *Vet. Immunol. Immunopath.* **12**, 21–28.
- Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
- Lobb, C.J. and Olsen, M.O.J. (1988) *J. Immunol.* **141**, 1236–1245.
- Moretta, L., Moretta, A., Canonica, G.W., Bacigalupo, A., Mingari, M.C. and Cerottini, J.-C. (1981) *Immunol. Rev.* **56**, 141–162.
- Warr, G.W. and Marchalonis, J.J. (1977) *Dev. Comp. Immunol.* **1**, 15–22.
- Warr, G.W., DeLuca, D. and Griffin, B.R. (1979) *J. Immunol.* **123**, 910–917.