BRIEF COMMUNICATION

Michael F. Criscitiello · Ralph Benedetto
Aurita Antao · Melanie R. Wilson
V. Gregory Chinchar · Norman W. Miller
L. William Clem · Thomas J. McConnell

β₂-microglobulin of Ictalurid catfishes

Received: 15 December 1997 / Revised: 13 March 1998

Key words  Beta-2 microglobulin · Catfish · MHC · Phylogeny

Major histocompatibility complex (MHC) class I molecules are expressed on the surface of virtually every vertebrate nucleated cell. The MHC-encoded class I heavy chain is non-covalently bound to the β₂-microglobulin (β₂m), an association necessary for proper folding, peptide binding, and surface display of class I antigens (Hansen et al. 1989; Vitiello et al. 1990). Unlike the classical class I heavy chain genes, the β₂m chain gene is not linked to the MHC loci in human and mouse and is not polymorphic (Klein 1986).

Although β₂m sequences have been determined in only a few species of fish and birds, a high degree of sequence conservation has been noted in β₂m among all vertebrate species in which it has been identified. β₂m closely resembles the structure of an immunoglobulin (Ig) constant region domain (Peterson et al. 1972; Smithes et al. 1972) and also exhibits similarity to the MHC class I α3 domain (Williams et al. 1988). Both of the genes encoding these Ig superfamily proteins are thought to have arisen by duplication of a common ancestral gene (Burnet 1970; Gally et al. 1972). Hence, the study of β₂m in lower vertebrates may shed light on the origin of the MHC.

In addition, the fact that the channel catfish is presently the best characterized fish in vitro system with regard to cellular aspects of adaptive immunity makes it a logical choice for detailed studies of MHC and related molecules (Miller et al. 1985, 1986, 1994). For example, previous in vitro studies have shown that catfish have the functional equivalents of T, B, NK, and accessory cells which interact in an alloantigen (presumably MHC)-restricted fashion (Clem et al. 1991; Vallejo et al. 1992). Consequently the work reported here was undertaken to identify and sequence the β₂m chain gene from six North American catfishes: channel catfish, Ictalurus punctatus; Headwater catfish, I. pricei; blue catfish, I. furcatus; I. n.sp. (Ictalurus new species, referred to here as the chihuahua catfish, Humphries and W.W. Miller, personal communication); white catfish, Amia catus; and yellow bullhead Amia catus.

A partial cDNA sequence for β₂m was amplified from the cloned B-cell line, 1B10 (Miller et al. 1994), and used to probe a channel catfish 42TA macrophage cDNA library (Luft et al. 1996). Five phage clones were plaque purified and two were sequenced. Both these clones exhibited significant similarity to known β₂m sequences in BLAST (Altschul et al. 1990) searches of the GenBank database. The longest cDNA clone [1170 base pairs (bp)] contained a 5' untranslated region, the entire β₂m coding sequence, and a 3' untranslated region containing four possible polyadenylation signal sites. Figure 1A shows the nucleotide and inferred amino acid sequences of this cDNA. The channel catfish β₂m mature protein is predicted to be 97 amino acids in length with a 19 aa leader. The characteristic cysteines forming the intradomain disulfide bridge (Williams and Barclay 1988) are found at residues 25 and 80. Based on sequences obtained from polymerase chain reaction (PCR) amplification of genomic DNA from different catfish species (see below) the exon boundaries of channel catfish β₂m are readily identified. At the 3' end of exon 3, i.e., amino acids 90–93, is a potential carbohydrate acceptor site (Asn-Ile-Ser). To date no β₂m chain gene containing an N-linked glycosylation signal sequence(s) has been reported in any species.

Channel catfish primers TM301 and TM307, based on the cDNA sequence, were used to amplify the β₂m chain gene from genomic DNA of six species of catfish: I. punctatus, I. n.sp., I. pricei, I. furcatus, A. catus, and...
Fig. 1  A Nucleotide and inferred amino acid sequences of channel catfish \(\beta_{2m}\) cDNA. The sequences of primers TM307 and TM301 are italicized and the possible adenylation signal sites are underlined. The coding sequence of the mature protein is numbered 1–97, with the leader sequence numbered –19 to 1. Exon boundaries are based on comparison of cDNA with the genomic channel catfish \(\beta_{2m}\) sequence. The GenBank accession number for \(I.\) punctatus cDNA is AF016041. GenBank accession numbers for the \(I.\) punctatus genomic \(\beta_{2m}\) sequences are as follows: AF016042 (\(I.\) punctatus genomic DNA), AF016043 (\(I.\) punctatus “new species”, chihuahua catfish genomic DNA), AF016044 (\(I.\) pricei genomic DNA), AF022958 (\(I.\) furcatus genomic DNA), AF022959 (\(A.\) natalis genomic DNA), and AF022960 (\(A.\) catus genomic DNA). To isolate the \(I.\) punctatus cDNA clone, forward primer TM294 (5′-TGYCAYGT-NACNGGNTTYTAYCC-3′ (Y=pC or T; N=pA, C, G or T) with similarity to the conserved regions surrounding the 1st (forward) cysteine of \(\beta_{2m}\) was used with the reverse primer oligo(dT) 17 to PCR amplify a 350 bp fragment from a cDNA pool made from the catfish B-cell line 1B10 (Miller et al. 1994). Thirty-five cycles of amplification (94°C for 1 min, 56°C for 1 min, 72°C for 1 min) were performed using 5 units Taq polymerase (Life Technologies, Gaithersburg, Md.), 0.4 \(\mu\)g of forward and reverse primers, and 1% of the first-strand cDNA pool; reaction conditions were as recommended by the manufacturer. The resultant fragment was cloned into PCRII (Invitrogen, San Diego, Calif.) and sequenced. The initial database searches using NCBI BLASTN (Altschul et al. 1990) found the sequence homologous to known \(\beta_{2m}\) sequences. This 350 bp fragment was subsequently used to screen \(5 \times 10^5\) plaque-forming units of an amplified channel catfish macrophage cell line (42TA) cDNA library (Luft et al. 1996) at high stringency. Filter lifts (Micron Separations, Westboro, Mass.) were hybridized with the 350 bp \(\beta_{2m}\) fragment that was random prime-labeled with \(^{32}\)P-dCTP. Filters were washed to a final stringency of 65°C in 0.1 x standard sodium citrate and 1% sodium dodecyl sulfate. Positive phage clones were plaque purified and subcloned into Bluescript by in vivo excision (Stratagene, La Jolla, Calif.). Primers for amplification of the six species of catfish genomic \(\beta_{2m}\) sequences were designed based on the complete cDNA sequence obtained from one of these positive clones. The forward primer, TM307 (GCGGGATCCAT-GAAGTTTCTGCTTTCCTTCGTC) corresponds to the 5′ end of the leader sequence with the addition of a Bam HI cloning site (underlined). The reverse primer, TM301 (GCGAAGCTTCTG-GAAATGAAGCCCAGGAGGTTTA) corresponds to the 3′ end of the \(\beta_{2m}\) sequence beginning with the stop codon TAA and includes a HindIII site (underlined). Thirty-five cycles of amplification (94°C for 30 s, 58°C for 30 s, 68°C for 4 min) were performed using the Advantage polymerase kit (Clontech, Palo Alto, Calif.), using a high fidelity thermostable polymerase, for \(I.\) punctatus, \(I.\) furcatus and \(A.\) natalis. An annealing temperature of 51°C was employed for \(I.\) pricei and \(I.\) n.sp., and 50°C for \(A.\) catus. All amplified products, except those from \(I.\) pricei and \(I.\) n.sp., which were sequenced directly in PCR, were cloned into pCR-Script (Stratagene) or into pGEM-T (Promega, Madison, Wisc.) and sequenced using an Applied Biosystems 373A automated sequencer and an FS Ready Reaction DyeDeoxy Terminator Cycle Sequencing Kit (Perkin Elmer, Foster City, Calif.). Approximately 100 ng of DNA was used as the template; reaction conditions were as recommended by the manufacturer. B Schematic of channel catfish \(\beta_{2m}\). Exons 1, 2, 3, and intron 1 are labeled. Key amino acid positions are designated on the underside of the figure, as are the intradomain disulfide-bond forming cysteines and the potential N-linked glycosylation site at amino acid position 90. The leader sequence is shaded.
A. natalis. PCR yielded products of 1279, 1547, 1548, 1562, 1438, and 1474 bp sequences, respectively. Figure 1B is a schematic of channel catfish β₂m showing the exon and intron structure. As commonly found for teleost MHC genes, each of the catfish β₂m chain introns are phase one, with the intron interrupting the exon between the first and second nucleotides. The exon/intron splice sites for the six catfish genes show high sequence similarity to one another, particularly in the coding regions. The most similar (I. punctatus and I. n.sp.) differ only in their intron sequences; I. pricei differs from I. punctatus by only two nucleotides (one amino acid) in the coding region; and I. pricei, I. furcatus, A. catus, and A. natalis are 99%, 96%, 87%, and 83% similar to I. punctatus, respectively, in amino acid sequence.

An alignment of the six catfish β₂m amino acid sequences with those of other vertebrate species is shown in Fig. 2. Twenty-two amino acids that are invariant in all other published vertebrate β₂m sequences are also invariant in the catfish species (Dixon et al. 1993; Ellis and Martin 1993; Ellis et al. 1993, 1995; Ono et al. 1993; Milland et al. 1993; Riegert et al. 1996; Ruiz et al. 1994; Shum et al. 1996). Each of the four Ictalurus species contain the exon 3 glycosylation signal (marked in bold) encoded by the identical nine nucleotide sequence, whereas the Ameirus catfish species, rainbow trout (Oncorhynchus mykiss), and carp (Cyprinus carpio) miss having an N-linked glycosylation signal sequence by one nucleotide (data not shown). When compared with mammalian and bird β₂m, each of the known teleost β₂m (Dixon et al. 1993; Ono et al. 1993; Shum et al. 1996) are two amino acids shorter, i.e., they lack amino acids 85 and 86 in exon 3. This shortening of exon 3 could affect, at least for the Ictalurus species, the location of the glycosylation signal sequence in the folded protein. Using diagrammatic representations of human β₂m as a guide (Bjorkman et al. 1987), one can postulate the structure of β₂m lacking these two amino acids. Since these amino acids occur in the turn between β strands 6 and 7, their absence could shift the location of the potential glycosylation site (amino acids 92–94 by human numbering) to the more exposed turn region.

The appearance of a single band in Southern blot analyses of genomic DNA from five individual fish digested separately with three different restriction enzymes (data not shown) suggest the presence of only one β₂m chain locus in the channel catfish. Although β₂m is usually encoded by a single-copy gene (Klein 1986), recently two examples of multiple loci have been described in teleosts. Dixon and co-workers (1993) identified two loci in gynogenetic carp, probably the result of tetraploidy in this species. Ten different β₂m sequences have been cloned and sequenced from an individual rainbow trout, a number that the tetraploidy of this species cannot account for (Shum et al. 1996). The authors of that study speculate that the β₂m chain gene in this salmonid species has remained in the MHC proper and has consequently been subjected to duplication events that are thought to give rise to the large diversity of MHC alleles. Their Southern blot data also indicate the presence of multiple β₂m chain loci, but the manner in which their sequences fall into homologous groups suggest polymorphism. Perhaps both polymorphism and multiple alleles contribute to the high number of rainbow trout β₂m sequences, as opposed to the simple non-polymorphic locus probably present in channel catfish.
Fig. 3 Phylogenetic tree of \(\beta_2m\) mature peptide-coding nucleotide sequences from different vertebrates. Searches for sequences similar to catfish \(\beta_2m\) were performed and preliminarily aligned using the NCBI BLAST E-mail server (Altschul et al. 1990). The final nucleotide sequence alignments were performed using the PILEUP and PRETTY programs of the Genetics Computer Group (GCG) (Devereux et al. 1984). Pairwise distances were calculated for intron 1 using the \(P\)-distance algorithm \(\beta_2m\) and the neighbor-joining method of Saitou and Nei (1987). Calculations and dendrogram construction were completed using the Molecular Evolutionary Genetic Analysis (MEGA) (Penn State University, University Park, Penn.) programs. Numbers on nodes indicate the frequency with which this node was recovered per 100 bootstrap replications in a total of 500 replications. The MEGA program treated both \(Amerius\) species as a single outgroup, therefore no bootstrap value was generated for this node. References for individual sequences are as in Fig. 2.

A dendrogram was constructed employing the neighbor-joining method on distances calculated using the \(p\)-distance algorithm (MEGA, Penn State University, University Park, Penn.) for the intron 1 sequence of the six catfish species (Fig. 3). Intron 1 is about 900 bp long in zebrafish (incomplete sequence) compared with 1078, 1079, 1091, 974, and 1010 bp in \(I.\ punctatus\), \(I.\ n.sp\), \(I.\ pricei\), \(I.\ furcatus\), \(A.\ catus\), and \(A.\ natalis\), respectively. Insertions/deletions in the sequence comparisons were not weighted in the calculations used for generation of the dendrogram. The topology of this dendrogram, unlike a dendrogram based on nucleotide coding regions, is identical to one generated using 384 nucleotides of the 5\(^\prime\) end of the mitochondrial cytochrome \(B\) gene (data not shown). This dendrogram (Fig. 3) is consistent with phylogenetic relationships predicted using morphological characteristics (Lundberg 1992). This analysis places \(I.\ n.sp\). with \(I.\ pricei\), and places \(I.\ punctatus\) outside of and ancestral to these two. The phylogeny of these closely related \(Ictalus\) species has never been resolved. The phylogeny offered in Fig. 3 is consistent with the fishes’ present geographic range: \(I.\ n.sp\). and \(I.\ pricei\) are restricted in geographic range to Mexican and Southwestern arid regions, while the native range of \(I.\ punctatus\) includes much of North America (Peterson 1991). \(I.\ furcatus\) places as ancestral to the other \(Ictalus\) species as expected (Lundberg 1992). The white catfish (\(A.\ catus\)) is placed with \(A.\ natalis\), supporting the movement of the white catfish from the \(Ictalus\) to the \(Amerius\) genus (Hubbs and Lagler 1958; Lundberg 1982). Also, \(A.\ natalis\) and \(A.\ catus\) \(\beta_2m\) do not contain the glycosylation signal found in the four \(Ictalus\) catfish sequences.

In conclusion, we identified a unique glycosylation signal sequence of \(\beta_2m\) in \(Ictalus\ punctatus\) and found the site in the only three other members of the genus that were sampled. Two species of closely related \(Amerius\) catfishes lack the encoded consensus glycosylation site. No previously reported \(\beta_2m\) chain genes contain sequences signaling for glycosylation. Experiments are underway to determine whether or not the catfish \(\beta_2m\) protein product is glycosylated. A glycosylated \(\beta_2m\) could possibly associate with calnexin and/or calreticulin in the endoplasmic reticulum, as proposed for class I molecules during folding and assembly (Parrham 1996), in \(Ictalus\) catfishes. Also, a phylogenetic tree generated from the catfish \(\beta_2m\) sequences corroborates and extends recent morphological analyses of the phylogeny of these fishes, subsequent work should further elucidate the phylogeny of the family Ictaluridae.

Acknowledgments We thank B. Jenson (Dexter National Fish Hatchery, Dexter, New Mexico), J. Layton (Layton and Smith Fishery, Windsor, North Carolina), and A.M. Snyder (Museum of Southwestern Biology, University of New Mexico, Albuquerque, New Mexico) for generously providing fish samples. We thank R. Miller (Museum of Zoology, University of Michigan) for discussions about the chihuahua catfish. This work was supported by grants from ECU (95-RS-37), NIH (R37-AI-19530), and USDA (95–37204–2225).