A Novel Functional Class I Lineage in Zebrafish (Danio rerio), Carp (Cyprinus carpio), and Large Barbus (Barbus intermedius) Showing an Unusual Conservation of the Peptide Binding Domains

Corine P. Kruiswijk, Trudi T. Hermsen, Adrie H. Westphal, Huub F. J. Savelkoul and René J. M. Stet

http://www.jimmunol.org/content/169/4/1936

References
This article cites 62 articles, 22 of which you can access for free at: http://www.jimmunol.org/content/169/4/1936.full#ref-list-1

Subscription
Information about subscribing to The Journal of Immunology is online at: http://jimmunol.org/subscription

Permissions
Submit copyright permission requests at: http://www.aai.org/About/Publications/JI/copyright.html

Email Alerts
Receive free email-alerts when new articles cite this article. Sign up at: http://jimmunol.org/alerts
A Novel Functional Class I Lineage in Zebrafish (Danio rerio), Carp (Cyprinus carpio), and Large Barbus (Barbus intermedius) Showing an Unusual Conservation of the Peptide Binding Domains

Corine P. Kruiswijk,* Trudi T. Hermsen,† Adrie H. Westphal,‡ Huub F. J. Savelkoul,* and René J. M. Stet†*

Species from all major jawed vertebrate taxa possess linked polymorphic class I and II genes located in an MHC. The bony fish are exceptional with class I and II genes located on different linkage groups. Zebrafish (Danio rerio), common carp (Cyprinus carpio), and barbus (Barbus intermedius) represent highly divergent cyprinid genera. The genera Danio and Cyprinus diverged 50 million years ago, while Cyprinus and Barbus separated 30 million years ago. In this study, we report the first complete proteincoding class I ZE lineage cDNA sequences with high similarity between the three cyprinid species. Two unique complete proteincoding cDNA sequences were isolated in zebrafish, Dare-ZE*0101 and Dare-ZE*0102, one in common carp, Cyca-ZE*0101, and six in barbus, Bain-ZE*0101, Bain-ZE*0102, Bain-ZE*0201, Bain-ZE*0301, Bain-ZE*0401, and Bain-ZE*0402. Deduced amino acid sequences indicate that these sequences encode bonafide class I proteins. In addition, the presence of conserved potential peptide anchoring residues, exon-intron organization, ubiquitous expression, and polymorphism generated by positive selection on putative peptide binding residues support a classical nature of class I ZE lineage genes. Phylogenetic analyses revealed clustering of the ZE lineage clade with nonclassical cyprinid class I Z lineage clade away from classical cyprinid class I genes, suggesting a common ancestor of these nonclassical genes as observed for mammalian class I genes. Data strongly support the classical nature of these ZE lineage genes that evolved in a trans-species fashion with lineages being maintained for up to 100 million years as estimated by divergence time calculations. The Journal of Immunology, 2002, 169: 1936–1947.

Cartilaginous fish are the most ancient group of vertebrate species possessing MHC genes. In all species studied to date, the MHC is a large chromosomal region, containing many genes encoding proteins of immunological importance (1–3) with the exception of the bony fish that represent about half of all vertebrate species (4). Unlike all other jawed vertebrates, bony fish class I and II genes are located on different linkage groups (5–8).

MHC class I and II genes encode structurally similar proteins that present peptides to T lymphocytes. The class I genes can be subdivided into classical class I and nonclassical class I molecules based on structural and functional differences and expression patterns (9). The MHC classical class I genes are involved in Ag presentation, presenting endogenous derived peptides to CD8 positive T cells. They have been shown to be highly polymorphic and coexpressed on cells in almost all tissues. Class I molecules are composed of a large α-chain, encoded in the MHC, noncovalently associated with a much smaller β2-microglobulin (β2m)2 molecule, encoded outside the MHC. The class I α-chain consists of three extracellular domains with two membrane-distal domains that form the peptide-binding region. Polymorphic residues within this peptide binding region interact with peptides and are under positive Darvinian selection (10). MHC polymorphism evolves in a trans-species fashion (11). In general, MHC class I genes seem to be more divergent and more rapidly evolving than class II genes. HLA class I lineages are recognized in great apes and thus maintained up to 6 million years, while certain HLA class II lineages were recognized in prosimians that diverged from human ~85 million years (12, 13).

The MHC nonclassical class I genes encode molecules with a typical MHC class I structure but do not have the function and tissue distribution of the classical genes. They exhibit low polymorphism, are often expressed in a tissue-specific fashion, and are encoded either in the MHC or outside this complex (9). Functions of nonclassical class I molecules like CD1, HLA-E, HFE, MICA, and MICB are now emerging and the presence and conservation of nonclassical molecules among species underline the importance of their roles (reviewed in Ref. 14). CD1 molecules, an extensively studied group of non-MHC-encoded genes, were shown to present lipid structures, and therefore, play an important role in defense against bacterial infections. The MHC-encoded nonclassical HLA-E molecule modulates NK cell function by presenting peptides derived from classical class I leader sequences, while the MHC-encoded HFE molecule plays a role in iron metabolism and does not bind peptides. Unlike CD1, HLA-E, and HFE, the MICA molecule does not associate with β2m.

To date, nonclassical class I Z lineage genes have only been identified in the teleost (bony fish) species, gibel crucian carp (Carassius auratus langsdorfi) and common carp (Cyprinus carpio) with 18 U.S.C. Section 1734 solely to indicate this fact.

*Cell Biology and Immunology Group, Department of Animal Sciences, and †Laboratory of Biochemistry, Department of Agrotechnology and Food Sciences, Wageningen University, Wageningen, The Netherlands. E-mail address: rene.stet@celb.edc.wau.nl

1 Address correspondence and reprint requests to Dr. Rene J. M. Stet, Cell Biology and Immunology Group, Department of Animal Sciences, Wageningen University, P.O. Box 338, 6700 AH Wageningen, The Netherlands. E-mail address: rene.stet@celb.edc.wau.nl

2 Abbreviations used in this paper: β2m, β2-microglobulin; Δns, number of nonsynonymous substitutions per site; Δsy, number of synonymous substitutions per site.

Copyright © 2002 by The American Association of Immunologists, Inc. 0022-1767/02/$02.00
carpio) (15–17). Despite extensive searches for Z lineage sequences in genomic DNA and cDNA libraries of the cyprinid species zebrafish (Danio rerio), no evidence was obtained for the existence of such genes (6, 18, 19).

Previously, Southern blot hybridization performed on restriction-enzyme digested high m.w. DNA of carp of different geographical origins, using a probe to class I Z exon 4, detected 9–12 hybridizing fragments at extremely low stringency (20). These data suggested the existence of additional class I Z sequences in carp. A recent attempt to identify novel class I sequences in common carp revealed the partial coding sequence of the extracellular domains of new class I Z lineage sequences, Cyca-Zr2 and Cyca-Zr3 (21). In phylogenetic analyses, these sequences clustered with other cyprinid class I Z lineage sequences, but formed a separate clade; and therefore, are renamed to Cyca-ZE*0101 and Cyca-ZE*0201. That study also revealed two other unique class I Z lineage sequences, Cyca-Zr1 and Cyca-Zr4. However, these sequences formed a clade with Cyca-ZB (16) and therefore, are renamed to Cyca-ZB*0201 and Cyca-ZB*0301.

In this study, we identified the complete coding sequence of Cyca-ZE*0101. Furthermore, we analyzed the presence of these class I ZE molecules in zebrafish (2n = 50; Ref. 22) and barbus (Barbus intermedius; 2n = 150; Ref. 23), representing highly divergent cyprinid genera that separated from common carp (2n = 100; Ref. 24) 50 and 30 million years ago, respectively. Analyses of the complete protein-coding cyprinid sequences indicated a more classical nature of the ZE lineage genes. Therefore, we investigated their expression, intron-exon organization, and the characteristics of polymorphic residues for peptide binding and evidence for positive selection.

Materials and Methods

Fish

A λcaip cDNA library (Cyca-λcaip) prepared from PMA-stimulated phagocytes from C. carpio L. was available (25) to characterize the full-length cDNA sequences. Genomic DNA was extracted from a blood sample of a C. carpio L. R3 × R8 F1, hybrid individual (26) and total RNA was extracted from thymus, head kidney, spleen, kidney, and intestine sample to study gene expression.

A liver sample from a Lake Tana B. intermedius individual was collected to extract genomic DNA and total RNA for identification of class I Z sequences. Samples of muscle, kidney, liver, and thymus from a Lake Tana Barb of acutirostris individual were collected for extraction of total RNA to study gene expression.

A muscle tissue sample of one D. rerio wild-type individual (hatched according to standard procedures; Ref. 27) was collected for total RNA extraction to identify expressed class I Z sequences. In addition, muscle samples of four D. rerio F1 hybrid individuals (Dinamma, Brakel, The Netherlands) were collected to extract genomic DNA to study positive selection.

Genomic DNA and total RNA extraction

Genomic DNA was isolated using a Wizard Genomic DNA Purification kit (Promega, Madison, WI) according to the protocol provided. Total RNA extraction from tissue samples was performed according to the protocol described by Dixon et al. (28). DNA and RNA concentrations were determined using the GeneQuant system (Amersham Pharmacia Biotech, Roosendaal, The Netherlands).

PCR and Expand Long Template PCR conditions

Standard PCR conditions were 1× reaction buffer, 1.5 mM MgCl₂, 1 U of Taq polymerase (Goldstar; Eurogentec, Seraing, Belgium), 0.2 mM dNTPs, 0.2 μM of each primer, and 5 μl phage suspension or 100 ng genomic DNA. The cycling profile was 1 cycle at 94°C for 5 min followed by 30 cycles consisting of denaturing at 94°C for 30 s, annealing at 55°C or 55°C for 30 s, polymerization at 72°C for 1 min, and a final cycle of 10 min at 72°C. Expand Long Template PCR was performed according to the protocol described for amplification of cDNA (Boehringer Mannheim, Ingelheim, Germany). The standard and Expand Long Template PCR was performed using a GeneAmp PCR system 9700 (PE Applied Biosystems, Foster City, CA).

Amplification of expressed class I ZE genes in C. carpio

To amplify the missing 3′ end of class I Z sequences, anchored PCR was performed on a Cyca-λcaip library. A λ-specific anti-sense primer was used (77) in combination with a class I ZE lineage-specific sense primer A (Table I) matching the end of exon 3 of known cyprinid class I ZE sequences.

Amplification of expressed class I ZE genes in B. intermedius and D. rerio

Expressed class I ZE lineage sequences in barbus and zebrafish were isolated using the GeneRacer kit for full-length, RNA ligase-mediated rapid amplification 5′ and 3′ ends (Invitrogen, Carlsbad, CA). Full-length barbus

Table I. Primers used for amplification of genomic and cDNA and gene expression studies

<table>
<thead>
<tr>
<th>Primer</th>
<th>Nucleotide sequence</th>
<th>Accession</th>
<th>Gene Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>5′-TGTGGTGGACCTGGCATCACAACAAAT-3′</td>
<td>M24113</td>
<td>F</td>
</tr>
<tr>
<td>B</td>
<td>5′-ACTTCTGCTTGGTAGAAACCAGT-3′</td>
<td></td>
<td>G</td>
</tr>
<tr>
<td>C</td>
<td>5′-ATGGAAACACGGCTCGACAGT-3′</td>
<td></td>
<td>H</td>
</tr>
<tr>
<td>D</td>
<td>5′-CATTCTCCTGCTTGGTGTTGTG-3′</td>
<td></td>
<td>I</td>
</tr>
<tr>
<td>E</td>
<td>5′-AGACATGATGACCATTAGGA-3′</td>
<td></td>
<td>J</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Primer</th>
<th>Nucleotide sequence</th>
<th>Accession</th>
<th>Gene Expression</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>5′-TGTGGTGGACCTGGCATCACAACAAAT-3′</td>
<td>M24113</td>
<td>F</td>
</tr>
<tr>
<td>B</td>
<td>5′-ACTTCTGCTTGGTAGAAACCAGT-3′</td>
<td></td>
<td>G</td>
</tr>
<tr>
<td>C</td>
<td>5′-ATGGAAACACGGCTCGACAGT-3′</td>
<td></td>
<td>H</td>
</tr>
<tr>
<td>D</td>
<td>5′-CATTCTCCTGCTTGGTGTTGTG-3′</td>
<td></td>
<td>I</td>
</tr>
<tr>
<td>E</td>
<td>5′-AGACATGATGACCATTAGGA-3′</td>
<td></td>
<td>J</td>
</tr>
</tbody>
</table>
liver cDNA and zebrafish muscle cDNA was synthesized according to the protocol described. The 5′ ends were amplified by PCR using an antisense primer B (Table I) based on a conserved part of exon 4 of several common carp and gilthead bream cDNA class I Z nucleotide sequences in combination with the GeneRacer 5′ primer. The 3′ ends of barbus and zebrafish class I ZE sequences were amplified by an expand long template PCR using the GeneRacer 3′ primer in combination with a sense primer C, D, E, F (barbus), or G (zebrafish) matching exon 1 of the amplified 5′ end sequences (Table I).

**Amplification of genomic cyprinid class I ZE genes**

To amplify common carp genomic class I ZE sequences, expand long template PCR was performed using sense primer H matching exon 1 of Cyca-ZE*0101 in combination with a specific antisense primer I matching the 3′ untranslated region of this sequence (Table I).

Class I ZE sequences from zebrafish genomic DNA were amplified using a sense primer J matching aa 7–13 of the α1 domain and antisense primer matching aa 163–171 of the α2 domain (Table I).

**Cloning and DNA sequencing**

PCR products were ligated and cloned using the pGEM T-easy kit (Promega) following the manufacturer’s description. Plasmid DNA was isolated from cells using the QIAprep Spin miniprep kit (Qiagen, Valencia, CA) according to the protocol provided. Subsequently, plasmid DNA was sequenced using the ABI Prism BigDye Terminator Cycle Sequencing Ready Reaction kit and analyzed using an ABI 377 sequencer (PE Applied Biosystems).

**Accession numbers and nomenclature**

The new sequences reported in this study were deposited in the EMBL database under the following accession numbers: Bain-ZE*0101–0501; AJ420274-AJ420284; Cyca-ZE*0101: AJ420951, AJ420952, AJ420957, AJ420958; Dare-ZE*0101–1401: AJ420953-AJ420956 and AJ420975-AJ420984.

The nomenclature used to assign new sequences or rename existing sequences adheres to the recommendations described in the HLA factbook (29). The analysis and combination of the identified sequence in a separate clade together with known common carp Cyca-ZA, -ZB, -ZC, and goldfish Cuaa-ZD, the locus designation ZE was given. An asterisk and four digits follow the locus name. The first two digits describe the lineage and the third and fourth digits that follow describe alleles.

**Expression of Z lineage genes in B, intermedius and C. carpio**

To prepare cDNA of several tissues of carp and barbus, 20 µg total RNA was reverse transcribed using the Universal RiboClone cDNA Synthesis system (Promega) according to the protocol described. Subsequently, 2.5 µl cDNA was used to study gene expression by PCR using gene-specific sense and antisense primers (Table I). Two positive controls were included, classical MHC class I gene Cyca-UA1*01 gene expression and β-actin expression. The results were analyzed by agarose gel electrophoresis.

**Nucleotide sequence, amino acid sequence, and phylogenetic analyses**

Genomic and cDNA cyprinid class I ZE nucleotide sequences were represented by at least two identical clones. Sequence data obtained using the ABI sequencer were analyzed with Sequencer 4.1 software (Gene Codes, Ann Arbor, MI). Multiple alignments were done using the program clustal W version 1.8 (30). Signal peptide prediction analyses were performed using PSORT II (http://www.psort.nibb.ac.jp; Ref. 31).

Phylogenetic analyses and synonymous and nonsynonymous distance estimations were performed using Mega 2.1 software (32). Phylogenetic relationships using p-distances for amino acid sequences were constructed by the neighbor-joining algorithm (33). Synonymous and nonsynonymous distances were estimated by the modified Nei-Gojobori method (34) with p-distances or Jukes-Cantor correction.

**Protein modeling of the Dare-ZE*0101 amino acid sequence**

The mouse MHC class I H-2Ld model was predicted to be a suitable modeling template by the SWISS-MODEL Blast tool (the ExPaSY proteomic server of the Swiss Institute of Bioinformatics; http://www.expasy.ch). The sequence of Dare-ZE*0101 was aligned to the sequence mouse MHC H-2Ld, with ClustalX (35) using the PAM 350 matrix. Model building of Dare-ZE*0101 was performed with MODELLER (36) using the CVFF forcefield (37). The mouse H-2Ld structure (PDB entry: 1LPD) was used as a template. The model was verified after several rounds of energy minimization. The stereochemical quality of the homology model was verified by PROCHECK (38), and the protein folding was assessed with PROSAII (39), which evaluates the compatibility of each residue to its environment independently.

**Results**

**Highly similar class I Z lineage genes in three teleost genera**

This is the first report of complete protein-coding Z lineage sequences identified in three different cyprinid species including zebrafish. The complete protein-coding sequences were all generated by PCR using cDNA as template, ensuring that all sequences were transcribed.

Anchored PCR on a cDNA library of common carp revealed the missing membrane proximal, transmembrane, and cytoplasmic regions of a common carp class I Z gene, Cyca-ZE*0101 (21). The complete deduced amino acid sequence of Cyca-ZE*0101 encoded a putative cleavable signal peptide of 24 N-terminal amino acids, three extra cellular domains similar in length to other class I molecules, a transmembrane and cytoplasmic region (Fig. 1).

Using the GeneRacer kit for full-length RNA ligase-mediated rapid amplification, 5′ and 3′ end sequences highly similar to Cyca-ZE*0101 were identified in a barbus and a zebrafish individual. We determined six complete protein-coding sequences in barbus, Bain-ZE*0101, Bain-ZE*0201, Bain-ZE*0301, Bain-ZE*0401, and Bain-ZE*0501, and two complete protein-coding sequences in zebrafish, Dare-ZE*0101 and Dare-ZE*0201 (Fig. 1). The extracellular domains of the complete protein-coding barbus and zebrafish sequences exhibited 75–87% amino acid sequence identity to Cyca-ZE*0101. The α1 domains of the zebrafish and barbus sequences showed 84–94% identity to the Cyca-ZE*0101 α1 domain. The α2 and α3 domains of the barbus and zebrafish ZE sequences were 77–86% and 65–78% identical with the Cyca-ZE*0101 α2 and α3 domains.

Signal peptide prediction analyses identified putative cleavable signal peptide of 26, 19, and 19 N-terminal amino acids in length for Bain-ZE*0201, Dare-ZE*0101, and Dare-ZE*0102, respectively. Bain-ZE*0101, Bain-ZE*0201, and Bain-ZE*0301 sequences only showed putative cleavage sites between the glutamine (Q → −4) and threonine (T → −3) residues. Bain-ZE*0401 and Bain-ZE*0501 sequences possess a leader peptide of 100 aa. However, a putative cleavages site is only predicted between the alanine (A → −29) and glutamic acid (E → −28) residues, suggesting an extension of the α1 domain with 28 aa (Fig. 1).

Analyses of the presence of transmembrane and cytoplasmic regions in the complete protein-coding of Bain-ZE*0301, Bain-ZE*0401, and Bain-ZE*0501 cDNA sequences revealed an in-frame termination codon a few codons downstream of the α3 domain, resulting in absence of these regions.

**Cyprinid class I ZE sequences encode bona fide class I molecule**

Deduced amino acid sequence of all ZE lineage genes were aligned with common carp, zebrafish, shark, and human classical class I sequences (Fig. 1). Most features known to be conserved in classical and nonclassical class I molecules (40, 41) are present in ZE sequences. All ZE sequences possessed the conserved cysteine residues (C109, C176, C217, C276) in the α2 and α3 domains to form disulfide bonds within these domains, conserved residues (H3, D31, H101, D130) in α1 and α2 to form two salt bridges within these domains, and the FYP (222–224) motif in α3 domain.

Three acidic residues in an exposed loop in the α3 domain (aa 237–243) form a major CD8 binding site in mammals. The ZE-lineage sequences all possess at least three acidic residues within this region. Four residues (T10, Q104, D130, Q257) known to be involved in βm binding of human class I molecules are conserved.
FIGURE 1. Comparison of the deduced class I ZE amino acid sequences identified in common carp (Cyca), barbus (Bain), and zebrafish (Dare) with MHC classical class I sequences of common carp, zebrafish, horned shark (Gici), horned shark (Hefr), and human (accession nos. regiven in the legend of Fig. 5). Numbering is based on the Cyca-ZE*0101, including gaps. Dashes indicate identity to the Cyca-ZE*0101 sequence and asterisks indicate gaps.

- Highly conserved residues, cysteine residues, residues involved in salt bridges, and the FYP motif.
- Nine putative evolutionary conserved residues important for anchoring peptide termini. The putative CD8 binding loop is shown in bold in the outlined box, putative N-linked glycosylation sites are underlined, and conserved putative βm contact residues are marked in bold and in italic. Amino acid motifs within the zebrafish ZE sequences, characterizing a group of sequences used to calculate positive selection (Table III), are marked by outlined boxes.
- Premature termination codons.
sequences are comparable to the plot for Cyca-UA*01 in the cyprinid lineage. Hydrophobicity plots of cyprinid class I sequences compared with human classical class I sequences reveal the presence of the conserved nonmammalian motif (Table II; Fig. 1: Y7, Y62, R92, F134, T154, K157, W158, Y171, F183/FIGURE 2. A Wu-Kabat variability analysis (47) for the deduced amino acid sequences of α1 and α2 from all Dare-ZE sequences. Residues with >20% variability are marked by numbers. The numbers along the x-axis denote the amino acid position starting with the first amino acid of the α1 domain of the Dare-ZE molecule. Variability of the first 10 aa is not available.

Cyprinid class I ZE sequences possess an evolutionary conserved peptide-binding motif

The presence of nine evolutionarily conserved putative peptide-anchoring residues in the amino acid sequence of classical I genes has shown to be a useful criterion in discriminating classical class I genes and nonclassical class I genes in many vertebrates (42–46). In classical mammalian class I molecules, this motif of nine aa is YYYYTKWYY, while in nonmammalian vertebrates it slightly changed to YYRFTKWYY (Table II). Alignment of cyprinid ZE sequences with a human classical class I sequence revealed the presence of the conserved nonmammalian motif (Table II; Fig. 1: Y7, Y62, R92, F134, T154, K157, W158, Y171, F183/
Y186) with a minor difference. The tyrosine residue (Y186) was located three amino acids downstream in all cyprinid ZE sequences when aligned with human classical class I (Fig. 1). All other carp Z lineage sequences differ at least at four positions and up to eight when compared with the conserved motif in mammalian class I sequences.

Polymorphism of class I ZE sequences in D. rerio

Analyses of five zebrafish individuals revealed 14 unique Dare-ZE sequences (Fig. 1). Four zebrafish ZE sequences, Dare-ZE*0101, Dare-ZE*0201, Dare-ZE*0301, and Dare-ZE*0401 were identified in mRNA of a zebrafish individual. Ten zebrafish ZE lineage sequences (Dare-ZE*0501 to Dare-ZE*1401; Fig. 1) were generated by PCR on genomic DNA of four zebrafish individuals using primers designed to the start of α1 and the end of α2 of Dare-ZE sequences identified in mRNA. Agarose gel electrophoresis of the genomic PCR products revealed three fragments, ~650, ~800, and ~1000 bp in length in each individual. Subsequent cloning of these PCR products and sequence analyses revealed 10 unique sequences only representing the 650 bp fragment (intron data not shown).

A Wu-Kabat variability plot (47) based on 162 aa of the α1 and α2 from 14 unique Dare-ZE sequences showed the presence of putative polymorphic residues. A total of 21 aa residues within the α1 and α2 domains showed 20% or higher variability, while 9 of these 21 residues (aa: 11, 41, 90, 95, 106, 110, 129, 142, 154) showed 30% or higher variability (Fig. 2).

Protein modeling of Dare-ZE*0101

Protein model construction of the Dare-ZE*0101 amino acid sequence using the mouse H-2Ld crystal structure as template resulted in several putative protein models. Fig. 3A shows the mouse template (blue) in complex with a peptide mainly containing alanine residues (APAAAAAAM) and the βm superimposed on the constructed models of the Dare-ZE*0101 sequence. Only four major putative deviations from the mouse crystal were observed in four different loops (Fig. 3A, arrows). Two putative loop structures were located in the α3 domain and two in the α2 domain. The flexible loop structure in the α helix of the α2 domain, due to an amino acid insertion in the Dare-ZE*0101 sequence compared with the mouse sequence, may have major implication for peptide binding. The position of this loop, either pointing toward the peptide binding pocket, away from the pocket, or a position between binding. The position of this loop, either pointing toward the peptide binding pocket, away from the pocket, or a position between binding. The position of this loop, either pointing toward the peptide binding pocket, away from the pocket, or a position between binding. The position of this loop, either pointing toward the peptide binding pocket, away from the pocket, or a position between binding. The position of this loop, either pointing toward the peptide binding pocket, away from the pocket, or a position between binding. The position of this loop, either pointing toward the peptide binding pocket, away from the pocket, or a position between binding. The position of this loop, either pointing toward the peptide binding pocket, away from the pocket, or a position between binding. The position of this loop, either pointing toward the peptide binding pocket, away from the pocket, or a position between binding.
be involved in binding of peptide of mammalian class I molecules. The tyrosine residue (Fig. 3C, orange) located three aa downstream of the phenylalanine residue is located near the end of one of the α2 domain α helices. The position of the polymorphic residues exhibiting between 20 and 30% variability (Fig. 3D, light blue) and >30% variability (Fig. 3D, orange) in Wur-Kabat plots are shown to be located in the β-strands, α-helices, and loops.

**Exon-intron organization of cyprinid class I ZE sequences**

Sequence-specific primers were designed to the leader peptide and to the 3’ untranslated region of Bain-ZE*0401, Bain-ZE*0501, and Cyca-ZE*0101 to generate genomic PCR products. Two unique genomic barbus sequences similar to the cDNA nucleotide sequence of Bain-ZE*0401 and Bain-ZE*0501 and one genomic sequence similar to the cDNA nucleotide sequence of Cyca-ZE*0101 were identified (Fig. 4). Both genomic barbus sequences consisted of seven exons and six introns. The leader peptide is encoded by exon 1a, 1b, and 1c, the α1, α2, and α3 domains are encoded by exon 2, 3, and 4, respectively. The connecting peptide, the transmembrane, and the cytoplasmic tail were not present in these genomic sequences. Exon 5 encoded three (Bain-ZE*0401) or nine (Bain-ZE*0501) in-frame amino acids followed by a termination codon. The remainder of exon 5 contained the 3’ untranslated region. All introns start with GT and end with AG, and were all phase 1 introns, a codon split between the first and the second base (48).

The genomic common carp sequence consists of nine exons and eight introns (Fig. 4). The leader peptide and the three extra cellular domains, α1, α2, and α3 are encoded by separate exons (exons 1–4). The connecting peptide, the transmembrane region and the cytoplasmic tail are encoded by exon 5 and the cytoplasmic tail by exon 6, 7, and 8. Exon 8 also encoded the start of the 3’ untranslated region, while exon 9 contained for the remainder of the 3’ untranslated region. All introns start with GT and end with AG and were all phase 1 introns.

**Phylogenetic analyses**

Phylogenies were constructed separately for the α1, α2, and α3 domains of the cyprinid ZE sequences, with representatives of class I genes from other vertebrate taxa. The branching order and major groupings are similar to those documented in previous studies (17, 21, 44, 49). In trees constructed of the α1, α2, or α3 domain sequences, all the cyprinid ZE sequences cluster together in single clades supported by high bootstrap values (Fig. 5) and the cyprinid ZE clades clustered on a single branch with the clades comprising all other cyprinid Z lineage sequences, ZA, ZB, ZC, and ZD. However, the latter topology is only supported by a high bootstrap value in the tree of the α3 domain, while medium and low bootstrap values are observed in trees of α2 and α1 domains, respectively.

Stet al. (21) extensively studied nonclassical fish class I ZA, ZB, ZC, and ZD sequences in comparison with classical ones. In that study, comparison of classical fish class I sequences, including Cyca-UA, showed a very high degree of diversity in the α1 domains, a moderate diversity in the α2 domains, and conservation of the α3 domains. This diversity is reflected in the phylogeny, with longer branch lengths indicating higher diversity. Branch lengths in the cyprinid ZE clade of the α1 domain tree are remarkably short in relation to branch length of the α2 and α3 domain trees. This observation is in stark contrast to the paradigm of highly divergent class I α1 and α2 domains.

Both the α1 and α2 trees show a large zebrafish ZE subcluster supported by high bootstrap values. The subcluster in the tree of α1 comprised all 10 genomic zebrafish ZE sequences and one cDNA sequence. These sequences also clustered together in the α2 tree with the exception of Dare-ZE*0901. The Dare-ZE cDNA sequences, Dare-ZE*0101 and Dare-ZE*0201, formed a second, and all barbus and carp ZE sequences a third subcluster in the α1 tree supported by high bootstrap values. However, this topology was dissolved in the tree of the α2 domain. Three clear subclusters in the cyprinid ZE clade, supported by high bootstrap values, were formed in the tree of the α3 domain. Barbus ZE sequences formed two clusters comprising Bain-ZE*0201, Bain-ZE*0401, and Bain-ZE*0501, or Bain-ZE*0101, Bain-ZE*0102, and Bain-ZE*0401 with the latter cluster including Cyca-ZE*0101. The two cDNA zebrafish ZE sequences formed a third subcluster.

**Class I Z mRNA expression in tissues of cyprinid fish**

Expression of class I Z mRNA was studied in several tissues of a carp and barbus individual by RT-PCR with sequence-specific primers (Fig. 6). Both β-actin mRNA and Cyca-UA1*01 mRNA are expressed in all tissues studied.

Expression of mRNA of Cyca-ZB*0201 and Cyca-ZE*0101 was detected in thymus, head kidney, spleen, kidney, and intestine.

---

**FIGURE 4.** The genomic organization of carp Cyca-ZE*0101 and barbus Bain-ZE*0401 and Bain-ZE*0501. Exons encoding for the leader peptide, extracellular domains, α1, α2, and α3, the transmembrane (Tm) and cytoplasmic region (Ct), and the 3’ untranslated region (3’ ut) are indicated respectively by ■, ■, ■, ■, ■, ■, ■, ■, ■, ■, ■. Exon lengths are placed above the different protein domains, while intron lengths are placed below. *, Intron lengths estimated by PCR.
Bain-ZE*0401 mRNA and Bain-ZE*0101 and Bain-ZE*0102 mRNA were expressed in anterior and posterior kidney, liver, thymus, and muscle (data not shown).

Amplification specificity of the class I Z lineage genes was verified by sequencing and analyzing the amplified PCR products. Analyzing the amplified RT-PCR products revealed the presence of both Bain-ZE*0101 and Bain-ZE*0102 sequences in all tissues investigated. The amplified region of the nucleotide sequences of Bain-ZE*0101 and Bain-ZE*0102 sequences differ by two synonymous nucleotide substitutions.

Positive selection acting on zebrafish class I ZE sequences

Positive selection plays an important role in generating polymorphism in the peptide binding region (α1 and α2 domains) of classical class I molecules. Particularly, residues involved in peptide binding are under positive Darwinian selection. To search for evidence of positive selection (number of nonsynonymous substitutions per site (d_\text{N}) : number of synonymous substitutions per site (d_\text{S})) ratio > 1; Ref. 10), the ratio between the rates of d_\text{S} and d_\text{N} was calculated using two different distance methods (Table III).
Although the evolutionary conserved peptide anchor residues of HLA class I sequences are superimposable on cyprinid ZE sequences, HLA polymorphic peptide binding residues are not superimposable on zebrafish ZE variability. Therefore, putative residues involved in peptide binding were identified based on the variability they possessed in the Wu-Kabat analysis (Fig. 2). Positions possessing variability >20% were assigned as putative peptide binding residues and all others as nonpeptide binding residues. The cyprinid ZE sequences represent an ancient lineage; therefore, noise between different loci accumulated over a long period of time may obscure the signal for positive selection. However, when the genomic organization of loci in multigene families is unknown, their identity can only be inferred using multiple criteria. Figureoa et al. (18) used five criteria to estimate a minimum number of loci in swordtail fishes. Applying three of the five criteria identified a group of 10 sequences, Dare-ZE*0301, Dare-ZE*0501–0801, and Dare-ZE*1001–1401 belonging to a putative similar locus. These 10 sequences formed a distinct clade in phylogenetic analyses of α1 and α2 domains (Fig. 6); they shared characteristic motifs in exon sequences (Fig. 1) and possessed comparable intron 2 sequences of 92–99 bp in length (data not shown). Therefore, $d_o/d_s$ ratios were calculated for 14 zebrafish sequences that may represent multiple loci, and for 10 sequences that most likely represents only one locus.

The $d_o/d_s$ ratio based on 14 zebrafish ZE sequences (Table III, 14 taxa) revealed purifying or neutral selection for nonpeptide binding and positive selection for putative peptide binding residues for both distance methods. Although ratios for positive selection were slightly above 1 (1.4051, 1.3569) at putative PBR positions, they were supported by $p$ values <0.05. Using only the 10 sequences (Table III, 10 taxa) substantially increased the $d_o/d_s$ ratios of the putative peptide binding and nonpeptide binding residues (Table III). Similar evidence was found for assigning residues exhibiting a variability of 30% or higher as putative peptide binding residues and all others as nonpeptide binding.

### Table III. Calculation of $d_o/d_s$ ratios

<table>
<thead>
<tr>
<th>Variability</th>
<th>14 taxa</th>
<th>10 taxa</th>
<th>14 taxa</th>
<th>10 taxa</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_o/d_s$</td>
<td>0.3686</td>
<td>1.1141</td>
<td>0.4589</td>
<td>0.8810</td>
</tr>
<tr>
<td>$p$ value*</td>
<td>1.0000</td>
<td>0.1689</td>
<td>1.0080</td>
<td>1.0000</td>
</tr>
<tr>
<td>SE</td>
<td>0.0871</td>
<td>0.8105</td>
<td>0.1467</td>
<td>0.8510</td>
</tr>
<tr>
<td>$d_o/d_s$</td>
<td>0.03050</td>
<td>1.1192</td>
<td>0.3794</td>
<td>0.1711</td>
</tr>
<tr>
<td>$p$ value*</td>
<td>0.0993</td>
<td>1.0000</td>
<td>0.1689</td>
<td>1.0000</td>
</tr>
<tr>
<td>SE</td>
<td>0.0866</td>
<td>0.8309</td>
<td>0.4187</td>
<td>0.9008</td>
</tr>
</tbody>
</table>

*Test for positive selection using Mega software: $H_0$: $d_o = d_s$; $H_1$: $d_o > d_s$; reject $H_0$ when $p$ value < 0.05.

### Discussion

In the past, extensive search for nonclassical class I sequences and additional classical class I sequences in zebrafish failed to provide evidence for existence of such loci in this species (6,18,19). However, this study clearly shows the presence of a novel bonafide class I ZE lineage in zebrafish that is also present in barbus and common carp. The class I ZE lineage sequences identified in zebrafish and in two other cyprinid species exhibit four important features in favor to consider them as classical class I molecules. First, they possess the conserved amino acid residues involved in peptide binding of mammalian classical class I molecules. Second, they show ubiquitous expression in barbus and common carp. Third, class I ZE sequence variability is observed among sequences from several zebrafish individuals. Fourth, the variability seen among zebrafish class I ZE sequences is generated by positive selection acting upon putative peptide binding residues. However, unlike mammalian classical class I genes, the cyprinid class I ZE α1 domains show extraordinarily high amino acid conservation between and within the three divergent species studied, while less...
conservation is observed in the α2 and α3 domains. Phylogenetically, the class I ZE lineage seems to be more related to the common carp (Cyca-ZA, ZB, ZC) and goldfish (Caae-ZD) class I Z lineage sequences, which were assigned as nonclassical class I genes (21). The presence of this additional class I ZE lineage in zebrafish and in two other cyprinids, either classical or nonclassical in function, may have implications for the observation that in bony fish class I and II genes identified to date are located in different linkage groups (5–8). The novel zebrafish ZE lineage may be linked to one of the class II regions identified in three different linkage groups (7). This linkage would implicate a complex of class I and II genes like the MHC observed in all other jawed vertebrate species studied.

To date, unlike the mammalian counterparts, the function of class I molecules in fish has not been formally demonstrated. Their classical or nonclassical nature is inferred from amino acid sequence analyses, expression patterns and the extent of polymorphism, and compared with their mammalian counterparts. The presence of conserved peptide anchoring residues (18, 40) in the cyprinid class I ZE lineage favors assigning them as classical, although one of the tyrosine residues is replaced by a phenylalanine residue. The substitution of a tyrosine residue by a phenylalanine residue is seen in most nonmammalian classical class I sequences at position Y123. This indicates that such a replacement may not have major implications for the ability to bind peptide termini. Another possibility might be that the tyrosine residue located three aa residues downstream (Fig. 1, Y186) functions as the conserved peptide anchoring residue in cyprinid class I ZE molecules instead of the phenylalanine residue (Fig. 1, F183). However, protein modeling suggests that this extends the peptide binding groove. This may implicate binding of somewhat larger peptides or molecules similar to CD1 that binds lipid Ags in a substantially larger binding groove (50). The binding groove of cyprinid class I ZE molecules possesses hydrophilic residues in the α1 domain as indicated by hydrophobicity plots (data not shown). Protein modeling indicates that the hydrophilic residues in this domain comprise two β-strands and the α helix that follows (Fig. 3), suggesting that one side of the peptide binding groove is extremely hydrophilic.

Ubiquitous expression is another feature in favor of the classical nature of cyprinid class I ZE sequences. However, ubiquitous expression is also seen for another nonclassical class I ZE sequence (Cyca-ZB*0201) that thus does not possess the conserved peptide binding motif. In the past, characteristics of nonclassical class I genes led to the hypothesis that these genes are nonfunctional relics of ancient classical class I genes, whose ultimate extinction is inevitable (51, 52). However, many nonclassical class I genes have been reported in mammals and several studies revealed the functionality of these molecules (reviewed in Ref. 14). A possible explanation for ubiquitous expression of this nonclassical class I ZE gene in common carp might be that fish possess a variety of nonclassical functional class I molecules similar to the situation seen in mammals. This is supported by the fact that nonclassical class I sequences are not limited to common carp and goldfish (reviewed in Ref. 21) as previously suggested (18). To date, nonclassical fish class I genes have also been identified in coelacanth (53), shark (3, 44, 54), salmonids (49), and catfish (55).

With the variety of class I-like genes that now have been identified in different mammalian species, the nonclassical label has become ambiguous in mammals. Thus, it is suggested that classical class I genes, presenting peptides to cytotoxic T lymphocytes, are only those highly expressed MHC encoded loci that are subject to balancing selection which favors polymorphism at the positions that function as peptide binding residues (56, 57). The evidence that zebrafish class I ZE loci are subject to balancing selection generating variability at specific positions in the α1 and α2 domain can be considered as the most important feature supporting the classical nature of these genes. In contrast to the evolutionary conserved peptide anchor residues of HLA class I sequences that are superimposable on cytadin ZE sequences, HLA polymorphic peptide binding residues (58) are not superimposable on zebrafish class I ZE variability. However, binding of peptides or other small molecules of a different chemical nature might have favored variability at positions other than those identified in HLA class I sequences.

Protein modeling indicates that only 10 of 21 variable residues are located in a β-strand or an α helix at a position that might be involved in ligand binding. The remaining 11 residues are located in loops where they might play a role in receptor binding such as NK receptor and TCR. Two polymorphic residues in the loop between the two α helices of the α2 domain of the class I ZE molecules may have an undefined implication for peptide binding. This loop possesses high flexibility created through insertion of two aa residues compared with the mouse model. However, only cocrystallization with what is bound into the groove and functional studies will provide data on the actual structure of the molecule and the biological role of this novel class I ZE lineage.

The genomic organization of the carp class I Cyca-ZE*0101 gene is similar to mammalian MHC classical class I genes (48) with the exception of an additional intron which is located in the 3' untranslated region. Such an intron is observed for carp classical class I genes. However, in the case of Cyca-UA1*01, the cytoplasmic region of this gene is encoded by two exons, exon 6 and 7 (17). Although mammalian classical class I genes possess a characteristic intron-exon organization, it is not a criterion for distinguishing them from nonclassical class I genes. They may have similar or different organizations from that of classical class I genes.

Remarkable is the presence of barbus sequences lacking a transmembrane and cytoplasmic region at the mRNA level. The exon-intron organization clearly showed the absence of these regions at the genomic level. Thus, the soluble class I Bain-ZE*0401 and ZE*0501 molecules are not due to alternative splicing. These sequences also exhibit three introns within the much longer hydrophilic leader peptide compared with other class I sequences. Although these soluble molecules are expressed, it remains to be seen whether they are functional.

The birth-and-death model of evolution assumes that over the long-term evolution of MHC molecules, new genes are created by repeated gene duplications. Some of the duplicated genes are maintained in the genome, while others are deleted or become nonfunctional through deleterious mutations (59–61). Klein et al. (62) described a similar mechanism designated the accordion model. The three aberrant Bain-ZE sequences, encoding soluble class I molecules may be the remnants of previous gene duplications that underwent one or two deletion events after duplication resulting in sequences coding for soluble molecules with normal or aberrant MHC leader sequences. The aberrant Bain-ZE sequences then, are duplicated genes that lost the exon coding for connecting peptide, transmembrane, and cytoplasmic regions. This results in genes encoding soluble molecules like Bain-ZE*0201 that possess a putative signal peptide of 26 aa similar to the other class I ZE sequences. A second deletion event resulted in aberrant leader sequences like Bain-ZE*0501. This is supported by the observation that the Bain-ZE*0201 and Bain-ZE*0501 possess a normal or an aberrant leader sequence, respectively, but the remainder of sequences shows, including the 3' untranslated region and intron 6, 98% nucleotide similarity (data not shown). Exon 1a of the aberrant leader peptide Bain-ZE*0501 shows 75% nucleotide similarity to exon 4 of Bain-ZE*0101, suggesting a deletion event in a
region between two closely linked class I ZE genes that arose from gene duplication. The hexaploid status of barbus (23) may explain the presence of these aberrant class I ZE sequences in this species. Deletions, insertions, and amino acid replacements in a single gene will not result in immediate lethality due to the presence of multiple gene copies in this species. Redundancy in the genome of polyploid species allows duplicated genes to persist as functional genes over a long period of time, although they may accumulate non-detrimental mutations or alternatively become pseudogenes (63).

The phylogeny of cyprinid fish class I genes is reviewed by Stet et al. (21). At that time we only identified two partial class I ZE lineage sequences (Cyca-Zr2 and Cyca-Zr3) in common carp. These sequences were considered to be nonclassical based on clustering of these sequences with common carp and goldfish nonclassical class I Z genes. With more data available now, the classical nature of these sequences becomes clear. The relatively close relationship of the ZE lineage with common carp and goldfish nonclassical class I Z genes (ZA, ZB, ZC, and ZD) suggests a common ancestor. However, nonclassical Z lineage genes have not been identified in zebrafish, suggesting that the nonclassical Z lineage arose after divergence of the genera Danio and Cyprinus. This would suggest that these Z lineage genes should be present in barbus. However, this species has not been studied to that extent. An alternative for the absence of these Z lineage genes in zebrafish may be that they are simply not yet identified. An explanation for the lack of evidence of these genes might be the approach that assumes a conservation of the α3 domain as is observed for mammalian class I sequences. In this study, we demonstrated that the class I lineage ZE sequence shows considerable divergence in the α3 domains; and therefore, the above approach is flawed.

The cyprinid ZE sequences evolved in a trans-species fashion, suggesting an orthologous relationship. The number of ZE loci within a species and whether some ZE loci arose from gene duplications within species after divergence of the genera Danio, Cyprinus, and Barbus is unclear from these data. However, two zebrafish subclusters within the ZE clade suggest at least two loci in zebrafish. The four different ZE sequences identified in single zebrafish individuals and the presence of two is consistent with the fact that this species is diploid (22).

Several studies estimated that the genera, Danio, Cyprinus, and Barbus diverged 50 and 30 million years ago, respectively (28, 64, 65). The synonymous mutation rate per primate MHC loci seems to be similar to other loci with an average rate of 2.3 *10^-9 synonymous substitutions per synonymous site per year for HLA loci A, B, and C (66). Divergence time calculations using this mutation rate and the rate of synonymous substitution per synonymous site between cyprinid ZE lineage sequences estimated ~100 million years for zebrafish and barbus or common carp, and ~40 million years for common carp and barbus. Divergence time calculation using only exon 2, 3, or 4 sequences all resulted in comparable divergence time estimates. Analyses of intron sequences, which may provide more reliable divergence time estimates, also resulted in comparable values when a substitution rate of 2.3 *10^-9 substitutions per site was used. These time estimates suggest the presence of an ancestral class I ZE gene before separation of the three cyprinid genera, Danio, Cyprinus, and Barbus.

What distinguishes the class I ZE genes from the counterparts in primates is presence of this lineage in three teleost genera being maintained for up to 100 million years. In contrast, HLA class I lineages can only be recognized in great apes, which diverged from humanoids ~6 million years ago (12). Ancient classical class I lineage maintained up to 20 million-years-old are also described in other bony fish (67). Bony fish class I and II genes, unlike other jawed vertebrate species, are located on different linkage groups (5–8). The lack of linkage between the class I and II genes must have influenced the evolution of these genes. Impervious in this respect is the fact whether the class I ZE sequences are linked to other class I genes or class II genes. This will be clarified within the near future by the zebrafish genome project that will reveal the linkage group of these class I ZE lineage genes.

The maintenance of the ZE lineage for up to 100 million years and the unusual conservation of the peptide binding domains not only within species but also across species highlight the importance of their function. Although these domains show an unusually high conservation at the amino acid level, each domain exhibited a high degree of nucleotide diversity as shown by divergence time estimates based on the level of synonymous substitutions. The conservation of the α1 and α2 domains may relate to recognition of highly conserved molecular patterns derived from pathogens common to the three cyprinid species. Recognition of these conserved molecular structures might be the driving force to conserve the α1 and α2 domains in cyprinids.

Acknowledgments

We thank Martin de Graaf and Nand Sibbing for providing tissue samples of a Lake Tana B. intermedius individual and Johan Koelman for his assistance in identifying the novel sequences in common carp.

References
