Immunogenetics 38: 408–420, 1993

Immuno-
genetics

Zebrafish *Mhc* class II α chain-encoding genes: polymorphism, expression, and function

Holger Sültmann[1], Werner E. Mayer[1], Felipe Figueroa[1], Colm O'hUigin[1], Jan Klein[1, 2]

[1] Max-Planck-Institut für Biologie, Abteilung Immungenetik, Corrensstr. 42, 72076 Tübingen, Germany
[2] Department of Microbiology and Immunology, University of Miami School of Medicine, Miami, FL 33101, USA

Received May 3, 1993/Revised version received May 31, 1993

Abstract. Its small size and short generation time renders the zebrafish (*Brachydanio rerio*) an ideal vertebrate for immunological research involving large populations. A prerequisite for this is the identification of the molecules critical for an immune response in this species. In earlier studies, we cloned the zebrafish genes coding for the β chains of the class I and class II major histocompatibility complex (Mhc) molecules. Here, we describe the cloning of the zebrafish α chain-encoding class II gene, which represents the first identification of a class II *A* gene in teleost fishes. The gene, which is less than 3 kilobases (kb) distant from one of the β chain-encoding genes, is approximately 1.2 kb long and consists of four exons interrupted by very short (<200 base pairs) introns. Its organization is similar to that of the mammalian class II *A* genes, but its sequence differs greatly from the sequence of the latter (36% sequence similarity). Among the most conserved parts is the promoter region, which contains X, Y, and TATA boxes with high sequence similarity to the corresponding mammalian boxes. The observed striking conservation of the promoter region suggests that the regulatory system of the class II genes was established more than 400 million years ago and has, principally, remained the same ever since. Like the *DMA*, but unlike all other mammalian class II *A* genes, the zebrafish gene codes for two cysteine residues which might potentially be involved in the formation of a disulfide bond in the α1 domain. The primary transcript of the gene is 1196 nucleotides long and contains 708 nucleotides of coding sequence. The gene is expressed in tissues with a high content of lymphoid/myeloid cells (spleen, pronephros, hepatopancreas, and intestine). The analyzed genomic and cDNA sequences are probably derived from different loci (their overall sequence similarity in the coding region is 73% and their 3′ untranslated regions are highly divergent from each other). The genes are apparently functional. Comparison of genes from different zebrafish populations reveals high exon 2 variability concentrated in positions coding for the putative peptide-binding region. Phylogenetic analysis suggests that the zebrafish class II *A* genes stem from a different ancestor than the mammalian class II *A* genes and the recently cloned shark class II *A* gene.

Correspondence to: J. Klein.

Introduction

Major histocompatibility complex (Mhc) molecules are cell surface-expressed, highly polymorphic, heterodimeric glycoproteins which present peptides to T cells and thus initiate specific immune responses (Klein 1986; Rothbard and Gefter 1991). Mammalian Mhc class I molecules are responsible for the presentation of endogenously processed peptides to cytotoxic (CD8-positive) T cells (Bjorkman and Parham 1990); Mhc class II molecules bind peptides of proteins which entered the cell by the endocytotic pathway and present them to T helper (CD4-positive) cells (Lanzavecchia 1990). A mammalian class II molecule consists of one α and β chain, the two chains being encoded in separate *DXA* and *DXB* genes, respectively, where *D* stands for class II, *X* for unspecified class II family, and *A* or *B* for the α or β subclass. Each chain consists of two extracellular domains, α1, α2, or β1, β2, a connecting peptide, a transmembrane region, and a cytoplasmic tail. According to the model of Brown and co-workers (1988), capturing of peptides is the function of the α1 and β1 domains, and especially of the highly variable amino acid residues in the peptide-binding region (PBR).

Mhc class II *B* genes have been cloned from a number of mammalian species (Klein 1986), and more recently also from a variety of fishes (Hashimoto et al. 1990; Hordvik et al. 1993; Juul-Madsen et al. 1992; D. Klein et al. 1993; Ono et al. 1992; 1993 a, b, d). Class II *A* genes have thus far been identified in mammals (Klein 1986) and one species of cartilaginous fish (Kasahara et al. 1992) only. In an effort to shed light on *Mhc* evolution, we have initiated studies involving several teleost (bony fish) species. One of these species is the zebrafish (*Brachydanio rerio*), which was selected because of its short generation time, large number of offspring, and easy genetic (classical and molecular) manipulation (Laale 1977). In an earlier report (Ono et al. 1992), we described the cloning and properties of zebrafish class II *B* genes; here we have set out to identify zebrafish class II *A* genes.

Materials and methods

Fish. Zebrafish (*Brachydanio rerio*) were obtained from a dealer (Aquarium Pelz, Tübingen, Germany), from the laboratory strain C29 maintained at the Max-Planck-Institut für Entwicklungsbiologie (Tübingen), and from animals caught in North Bengal, India. Carp (*Cyprinus carpio*) and trout (*Oncorhynchus mykiss*) were purchased at a local market.

Preparation of genomic DNA. After the fish were killed, tissues were homogenized in liquid nitrogen and then treated with proteinase K at 45° C in lysis buffer [10 mM Tris/HCl, pH 7.5, 5 mM ethylene diaminetetraacetic acid (EDTA), 0.5% lauryl sarcosinate]. DNA was isolated, using an automated extractor (Applied Biosystems, Weiterstadt, Germany).

Genomic library. The zebrafish genomic library was obtained from Stratagene (La Jolla, CA). Inserts were cloned into the *Xho* I site of the λFIX™ phage vector, used to transfect *Escherichia* coli (*E. coli*) NM 538. The complexity of the original library was 2×10^6 plaque-forming units (PFU); the titer of the amplified library at the time of use was 1.5×10^9 PFU/ml.

Screening of genomic library. Replicate nitrocellulose filters (Sartorius, Göttingen, Germany) containing approximately 6×10^5 clones of the genomic library were screened with a cocktail of ^{32}P-labeled (Feinberg and Vogelstein 1983) zebrafish *Mhc* class II *B* probes *DAB1*01, DAB3*01,* and *Br2.1.* The *DAB1*01* and *DAB3*01* probes were full-length cDNA clones (Ono et al. 1992); the *Br2.1* probe was a polymerase chain reaction (PCR) product encompassing part of exon 3 (Sültmann 1991). The screening was preceded by prehybridization in 50% formamide, 5× SSPE (1× SSPE contains 0.15 M NaCl, 0.01 M NaH_2PO_4, 1 mM EDTA, pH 8.0), 0.1% sodium dodecyl sulfate (SDS), 5× Denhardt's (1× Denhardt's is 0.02% Ficoll, 0.02% bovine serum albumin, 0.02% polyvinylpyrrolidone) at 42° C. The screening itself consisted of hybridization under the conditions given above for 16 h after the addition of the probe. The filters were washed with 0.5× SSPE, 0.1% SDS at room temperature (twice for 20 min each time) and used to expose an X-ray film (XAR5; Kodak, Stuttgart, Germany) overnight with a DuPont Cronex intensifying screen. Single positive clones were isolated after a second screening.

Southern blots. Eleven micrograms of total genomic DNA were digested with *Hin* dIII for 16 h and the resulting fragments were separated by electrophoresis in a 0.8% agarose gel containing 0.5 μg/ml ethidium bromide, with the running buffer consisting of 1× TAE (0.04 M Tris-acetic acid, 2 mM EDTA, pH 8.5). The DNA was transferred to a Hybond N$^+$ nylon membrane (Amersham Buchler, Braunschweig, Germany) by alkaline (0.4 M NaOH) vacuum blotting. Hybridization was carried out in 50% formamide, 5× SSPE, 1% SDS, 5× Denhardt's, 100 μg/ml tRNA (from brewer's yeast) for 16 h and was followed by washing twice with 0.5× SSPE/0.1% SDS at 40° C. The filters were used to expose an X-ray film for 1–5 days.

Restriction mapping of genomic clones. A combination of single, double, and partial digestions with *Not* I, *Eco* RI, *Hin* dIII, *Bam* HI, and *Sma* I endonucleases was used. Agarose gels were blotted as described above and positive fragments were identified by hybridization with *Mhc* probes and probes located at the borders of the polylinker of the λFIX™ vector.

Preparation of RNA. Twenty zebrafish of the C29 strain were kept without feeding for two days, then anesthetized in 0.02% 3-aminobenzoic acid ethylester, killed, and their hepatopancreases and spleens frozen in liquid nitrogen. Total RNA was isolated according to the method described (Zhu and co-workers 1991). Poly(A)$^+$ RNA was isolated using an mRNA purification kit (Pharmacia LKB, Freiburg, Germany). Carp, trout, and cichlid fish RNA was isolated from different organs of a single fish in a similar way.

Construction and screening of cDNA library. A TimeSaver cDNA synthesis kit (Pharmacia LKB) was used to produce cDNA and ligate it, using *Eco* RI/*Not* I adaptors, to the *Eco* RI site of the λgt10 phage vector (Stratagene, Heidelberg, Germany). Packaging was carried out using an in vitro packaging kit (Stratagene) and the recombinant phages were used to transfect *E. coli* NM514. The complexity of the library was 1×10^6 PFU; the library was amplified once to a titer of 1×10^{11} PFU/ml. Replica nitrocellulose filters containing 5×10^5 clones were screened with the probe 369.1.3, a 600 base pair (bp) PCR product obtained by amplification of class II *A* cDNA and encompassing a stretch from the middle of exon 3 to the 3′ end of the 3′ untranslated (UT) region. Hybridization conditions were as described above.

Northern blots. Twenty micrograms of carp RNA, isolated from heart, spleen, hepatopancreas, intestine, pronephros, and ovary were dried under vacuum and dissolved in RNA-loading buffer [48% formamide, 6% formaldehyde, 5% bromophenol blue, 6% glycerol, 1× MOPS (1× MOPS contains 0.2 M 3-N-morpholinopropane sulfonic acid, 50 mM sodium acetate, pH 6, 10 mM EDTA)]. The RNA samples were then electrophoresed in 1% agarose denaturing gel containing 1× MOPS, 2% formaldehyde at 40 V for 16 h in 1× MOPS running buffer. The gel was stained with 0.5 μg/ml ethidium bromide and blotted on a Hybond N$^+$ membrane by alkaline transfer (0.05 M NaOH) under vacuum. Hybridization was carried out as described above.

Polymerase chain reaction. Amplification, using the cloned total cDNA as a template, was carried out with two primers. One was a degenerate sense-primer [TU 369: 5′-TT(C/T)G(C/T)TGAT-(G/A)(G/A)ATTCT(A/T)(T/C)CCTCC-3′] designed on the basis of data from Cho and co-workers (1991) and Kasahara and co-workers (1992) and expected to anneal in the 5′ half of class II *A* gene exon 3. The other was an antisense primer (#1244: 5′-ACAAGCTTG-TATTTCTTCCAGGGTAA-3′) annealing in the vicinity of the *Eco* RI cloning site in the left arm of λgt10. Approximately 10^8 phages

were added to the reaction mixture containing 1 × PCR buffer (Perkin Elmer, Überlingen, Germany), 1 μM each of sense and antisense primer, 0.1 mM each of the four deoxynucleoside triphosphates (Pharmacia), and 2.5 units of *Taq* polymerase (Beckman, Fullerton, CA). Each of the total of 40 cycles consisted of 1 min denaturation at 94° C, 1 min annealing at 50° C, followed by 2 min primer extension at 72° C. The final extension was for 10 min at 72° C. PCR with the genomic library was carried out under identical conditions. The products were phenol/chloroform extracted, phosphorylated with T4 polynucleotidekinase (New England Biolabs, Schwalbach/Ts., Germany) for 30 min, and blunt ends were generated by a Klenow fragment of DNA polymerase I (BRL, Gaithersburg, MD). The products were separated on a low-melting point agarose gel (BRL). Selected bands were cut out from the gel and centrifuged for 10 min at 15 000 rpm, and the DNA from the supernatant was ligated into M13 vectors. Amplification of exon 2 from total genomic DNA of zebrafish was obtained with TU461 (sense, 5′-GCTCAAGCTGAGCACAGG-3′) and TU450 (antisense, 5′-TTGCTCTTCTGGAGAGTT-3′) or TU579 (sense, 5′-TGAGCACAGGGATTTTGA-3′) and TU580 (antisense, 5′-CTTCTGGAGAGTTGTAT-3′) primers designed on the basis of the cDNA sequences obtained. Two-hundred-and-fifty nanograms of genomic DNA were subjected to 35 PCR cycles under the conditions described above, except for the annealing temperature (47° C).

For amplification of exon 2 sequences from Southern blots, 22 μg of total genomic DNA were digested with *Hin* dIII, the digest was divided into two aliquots, and each aliquot was subjected to electrophoresis in 0.8% agarose gel. After rinsing in 0.5 M NaOH/0.15 M NaCl and 0.5 M Tris/HCl, pH 8.0, 0.15 M NaCl (30 min in each solution), the gel was dried in a vacuum drier and cut into two parts. One part was hybridized with the zebrafish class II *A* exon 2 probe MW1273 (a PCR product generated by using primers TU461 and TU450) and positive bands were cut out from the second gel. The slices were suspended in 50 μl of TE buffer (10 mM Tris HCl, pH 8.0, 1 mM EDTA, pH 8.0) and stored at 4° C. Two to four microliters of the supernatant were then used as template source in PCR amplification.

DNA sequencing and analysis. DNA was purified by electrophoresis in low-melting point agarose gels, bands were cut out from the gel, centrifuged, and the supernatant was used for ligation into M13 vectors. A replicative form of M13 was isolated as described (Sambrook et al. 1989), treated with RNase A, and sequenced by the dideoxy chain-termination method (Sanger et al. 1977) using the Sequenase kit (Renner, Dannstadt, Germany).

Construction of dendrograms. Sequences were aligned, genetic distances between them were calculated by Kimura's two-parameter method (Kimura 1980), and they were arranged into matrices. A computer algorithm [neighbor-joining method of Saitou and Nei (1987)] was then applied to organize the sequences in such a way that those with the shortest distances separating them became neighbors. The result was depicted graphically in the form of a phylogenetic tree (dendrogram). Sequences were analyzed using the GCG (Genetics Computer Group) software package, version 7. GenEMBL was used as the database for the comparison with known sequences.

Results

Isolation of a class II A probe from the zebrafish. Fluid phage suspension containing the zebrafish cDNA library was used as substrate for PCR with primers TU369 and #1244. Electrophoresis of the PCR-amplified product yielded two major diffuse bands of approximately 400 bp and 600 bp. Fragments from both bands were subcloned in the *Sma* I site of the M13 phage vector and sequenced. Sequences of the smaller fragments showed no significant similarity to any of the known *Mhc* genes. In the larger fragments, however, the sequence of the major clone (369.1.3) showed significant similarity with exon 3 of class II *A* genes of different mammalian species.

Isolation of zebrafish class II A cDNA clones. Screening of the zebrafish cDNA library with the 600 bp long insert of the clone 369.1.3 yielded 50 positive signals on both replica filters. Thirty of the 50 clones were selected, and, after rescreening, ten of these were chosen for subcloning in M13. The ten clones were of five different sizes: 0.9, 1.0, 1.3, 1.4, and 1.9 kb. Three of the clones 1.3.4, 1.4.3, and 2.1.4, were sequenced in their entirety and another five partially, but of these, three turned out to have a sequence not related to the class II *A* gene. The sequences of the 1.3.4 and 1.4.3 clones are identical in the overlapping region except for two nucleotide substitutions (G↔A and A↔T); the sequence of the 2.1.4 clone differs from that of 1.4.3 by 57/711 substitutions in the coding region (Fig. 1). The sequence of the 2.1.4 clone is 966 nucleotides long. From a comparison with sequences of mammalian class II *A* genes, we deduce that it consists of 41 bp of 5′UT region, 711 bp of coding region, and 214 bp of 3′UT region. The coding region consists of 51 bp specifying 17 amino acid residues of the leader (signal) peptide and 660 bp specifying 219 amino acid residues of the mature protein. Although there is a methionine-specifying codon upstream from the one we identify as the start site of translation, it is most probably not used, because it is followed immediately downstream by an in frame stop codon. The codon we identified as the translation initiation site is flanked by the sequence TGA*A*CTATG*G*, which accords well with the consensus initiation sequence (Kozak 1984). The size of the primary transcript of the zebrafish class II genes has not been determined. In a related species, the carp, the size has been estimated as 1800 nucleotides at least (see below). If the zebrafish transcripts were also of this size, it would indicate that a substantial part of the 5′UT region is missing in the 2.1.4 cDNA clone.

The predicted length of the leader-encoding sequence (17 codons) is among the shortest of the class II *A* genes sequenced thus far but it exceeds the length of the shortest leader sequence in eukaryotic genes (13 codons; see von Heijne 1983). The putative leader-sequence cleavage site of the 2.1.4-specified protein deviates from the consensus in that it does not have a residue with a small, uncharged side-chain at position –1, a bulky aromatic residue at position –2, and Gly-Pro at positions –4 and –5 (von Heijne 1983).

```
11.2     PROMOTER REGION
         AAGCTTTTGTCAGTAGAAAGAAATTAAAATGCCAAGAAACAGATGTGTCATACTGAGTGGTAAATTCTGATTGGCTGAG
         GTGATTCTGAGGCAGTTATATGGCTCACAAGAGATAGTT

         5´UT REGION
2.1.4                TATTTGATATGGTGTTTACTTCAGTCCCCTCTCATTGAACT
1.4.3    AGCCGCTGTT-G-------------A-G--------TA-------------
1.3.4    AGCCGCTGTT-G-------------A-A--------TA-------------
11.2     AGAGCTAGTAA-AC-T-AT-T--AG-GTA-GTTC-T-TTGGCTC-CAG-AAG

         CODING REGION
         -17                          -10
2.1.4    ATG GAT CTG TTT GGT TTT CTG CTG ATA TTC ACT GTT ATT GTG AGT AAT AGG
1.4.3    --- --- --C --- --- --- T-- --- -C- --- --- --- --- T-- --- --- GT-
1.3.4    --- --- --C --- --- --- T-- --- -C- --- --- --- --- T-- --- --- GT-
11.2     --- --G G-- -A- -TG --C A-C --A -CT C-T -G- --G T-- --T TC- TCA GA- GTT AAT G

         INTRON 1
11.2     GTAATCGTATTTTTTTAATTACTTCATCTCTAATTTTACTTTTTGCTGCATTTGTCTTGCAGTTGTGTTCATCTCATAG
11.2     ACGATGTCTTTTGACACTTGACATGGATGATATTTTATCTGTTTCAG

         EXON 2
           1                                        10                                  20
2.1.4    GCT CAA GCT GAG CAC AGG GAT TTT GAC TTT ACT GGG TGT TCT GAT ACA GAG AAA GAT GAT
1.4.3    --- --- --- --- --- --- --- G-- --- --- TT- --A --- --- --C --- --A --- --A T--
1.3.4    --- --- --- --- --- --- --- G-- --- --- TT- --A --- --- --C --- --A --- --A T--
11.2               T- -TA --T GA- --- A-- -TT A-G GAC --A --C --A --- --- --- --- --G T--

                                         30                                  40
2.1.4    TTG ACT GGA TTT GAT GGA GAG GAG GAG TAC CAT ACA GAC TTC ATT AGA AAA GAA GGA GTA
1.4.3    --- CAA --- --- --- --- --- --- CT- --- --- T-- --- --- --- --- --- -T- --- ---
1.3.4    --- CAA --- --- --- --- --- --- CT- --- --- T-- --- --- --- --- --- -T- --- ---
11.2     A-A T-- -TT C-- --- --- --- --A AT- --- --- --- --- --- -G- G-- --- CGG --- -AG

                                         50                                  60
2.1.4    ATG ACA CTG CCT GAC TTT GCA GAT CCT TAC AGC TAT CCT GGA GGT TAT GAG GCT GGT GTT
1.4.3    G-- --- GC- --- --- --- --- --- --- ATG --- --- --- --- TT- --- --- AA- A-- ---
1.3.4    G-- --- GC- --- --- --- --- --- --- ATG --- --- --- --- TT- --- --- AA- A-- ---
11.2     --- --- T-- --- --- --- --- --- --- -TT -C- --- --- --T AC- --- --- CAG A-- C--

                                         70                                  80
2.1.4    GCT AAT ATG GAG ATC TGC AAA CAA AAT TTA GCC ACA GAC ATT AAG GCA TAC AAC TCT CCA
1.4.3    --- C-A --- --- G-- --- --- --- G-- --- --- --- --- --C --- --- --- --- --- ---
1.3.4    --- C-A --- --- G-- --- --- --- G-- --- --- --- --- --C --- --- --- --- --- ---
11.2     --- G-- TAT --A -CA --- --- --T --C --- -AT GTT -C- GCC --- --C --- --A --A --T

                  84
2.1.4    GAA GAG CAA CTA
1.4.3    --- --- --- ---
1.3.4    --- --- --- ---
11.2     TT- --- A-- TTG G

         INTRON 2
11.2     GTGAGACTTTATACTTAGTTTTATGGTACTGCTTCTTTACAAAGTGTAAGTATATTAAATCGAGTCATATAAATCCAAA
11.2     TGTTTAAGCAAAGGCAGTTATTTTTTGATCTACCTTGCTTCTTTTTTTAG

         EXON 3                90                                  100
2.1.4    GAC CCA CCT GTG ACA TCC ATC TAT TCA GAA GAT GAA GTG GTT CTG GAT GAA AAG AAC ACA
1.4.3    --- --T --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- -G- --- ---
1.3.4    --- --T --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- -G- --- ---
11.2      -T --- --- CA- --T --A --- --- --- AGG --- --T --- CAG -CA --- ATT G-A --T -AG

                               110                                 120
2.1.4    CTC ATC TGT CAT GTG ACT GGA TTC TTC CCT CCA CCT GTC AAT GTC TCC TGG ACG AAA AAC
1.4.3    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
1.3.4    --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
11.2     --- --- --- --- --- --- --T --- --- --- --- --- --- --- -GA --- --- --- --A --G ---

                               130                                 140
2.1.4    AAT GAT ATC GTG ATG GAG GAC ATT AGT TTT AGC CAG TAC CGT GAC AAC AGT GAC GGC ACC
1.4.3    --- --- --T --- -CA --- --A --- --- --- --- --A --- --C AGA --- --- --T --- ---
1.3.4    --- --- --T --- -CA --- --A --- --- --- --- --A --- --C AGA --- --- --T --- ---
11.2     --C --G --T --- -CA --- -GG --G --- G-A --- --- --T --A CCA --T -AC --T --- --A

                               150                                 160
2.1.4    TTC AAC ATG TTC TCC GCT CTG AAG TTC ACA CCT GCT GAA GGA GAC ATT TAC AGC TGC ACA
1.4.3    --- --- --- --- --T --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
1.3.4    --- --- --- --- --T --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
11.2     -AT --- --T --- --T A-- --A -GA --- --T --- -TG --- --- --- --- --- --- --- -GT
```

Fig. 1. (For legend see p. 412)

However the putative leader does have the expected hydrophobic core.

The 219 amino acid residues of the predicted mature protein can be divided (again, based on analogy to mammalian class II α chains) into α1 (84 residues), α2 (95 residues), connecting-peptide (11 residues), transmembrane (23 residues), and cytoplasmic (6 residues) domains (Fig. 2). The length of the zebrafish α1 domain is the same as that of the human DR α1 domain (Lee et al. 1982) but is three residues shorter than the length of the human DQ α1 domain (Auffray et al. 1984). In comparison to mammalian class II α

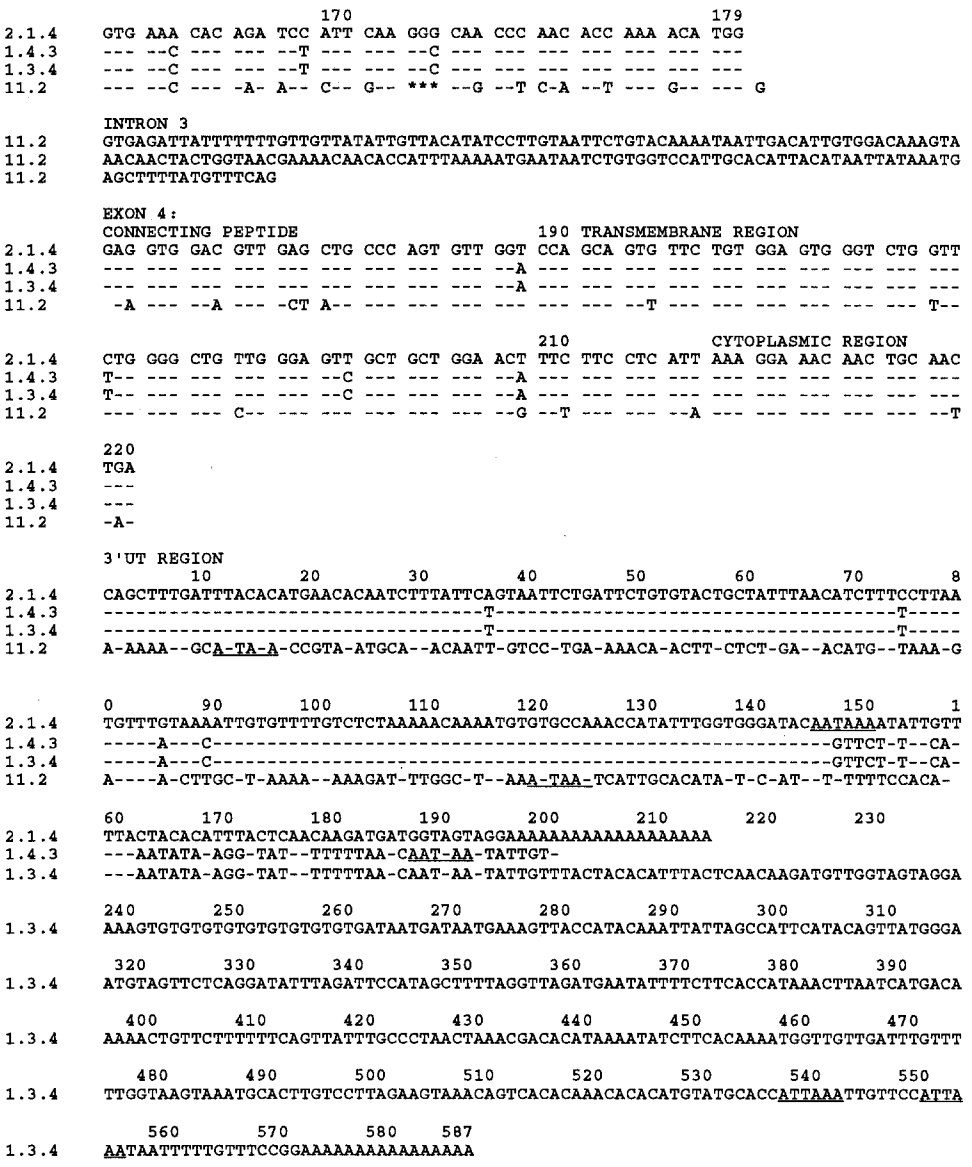

Fig. 1. Nucleotide sequence of three cDNA clones (2.1.4, 1.3.4, and 1.4.3) and one genomic clone (11.2) of zebrafish *Mhc* class II *A* genes. 2.1.4 is used as a reference; identities with the reference sequence are indicated by *dashes* (–). Exon boundaries in 11.2 were determined from the position of putative splicing signals. Promoter elements in the 5′UT region and potential polyadenylation signals in the 3′UT region are *underlined*. The sequences were submitted to the GenBank nucleotide sequence database. Accession codes are L19445 (2.1.4), L19446 (1.4.3), L19450 (1.3.4), and L19451 (11.2).

sequences (excluding DMA), the zebrafish sequences suffered a two-residue deletion at positions 21 and 22 (in the numbering system used in Figure 2) and two one-residue insertions at positions 56 and 58. The nurse shark class II α chain shows only one residue insertion at position 58 (Kasahara et al. 1992); both insertions are shared by zebrafish α and mammalian DMα chains (Cho et al. 1991; Kelly et al. 1991).

Another feature shared by zebrafish α and DMα chains is the presence of two cysteine residues 52 (zebrafish α) or 54 (DMα) residues apart. In the model of the α chain three-dimensional structure proposed by Brown and co-workers (1988), the two cysteines are close enough to each other to form an intradomain disulfide bond, in the zebrafish α chain, Cys 13 (zebrafish numbering) would be located near the first loop of the β-pleated sheet, and Cys 66 in the α-helix. These are also approximately the positions at which cysteines have been found in all class II β chains thus far analyzed (reviewed in Klein 1986). The nurse shark class II α chain lacks these two cysteine residues (Kasahara et al. 1992). The overall amino acid

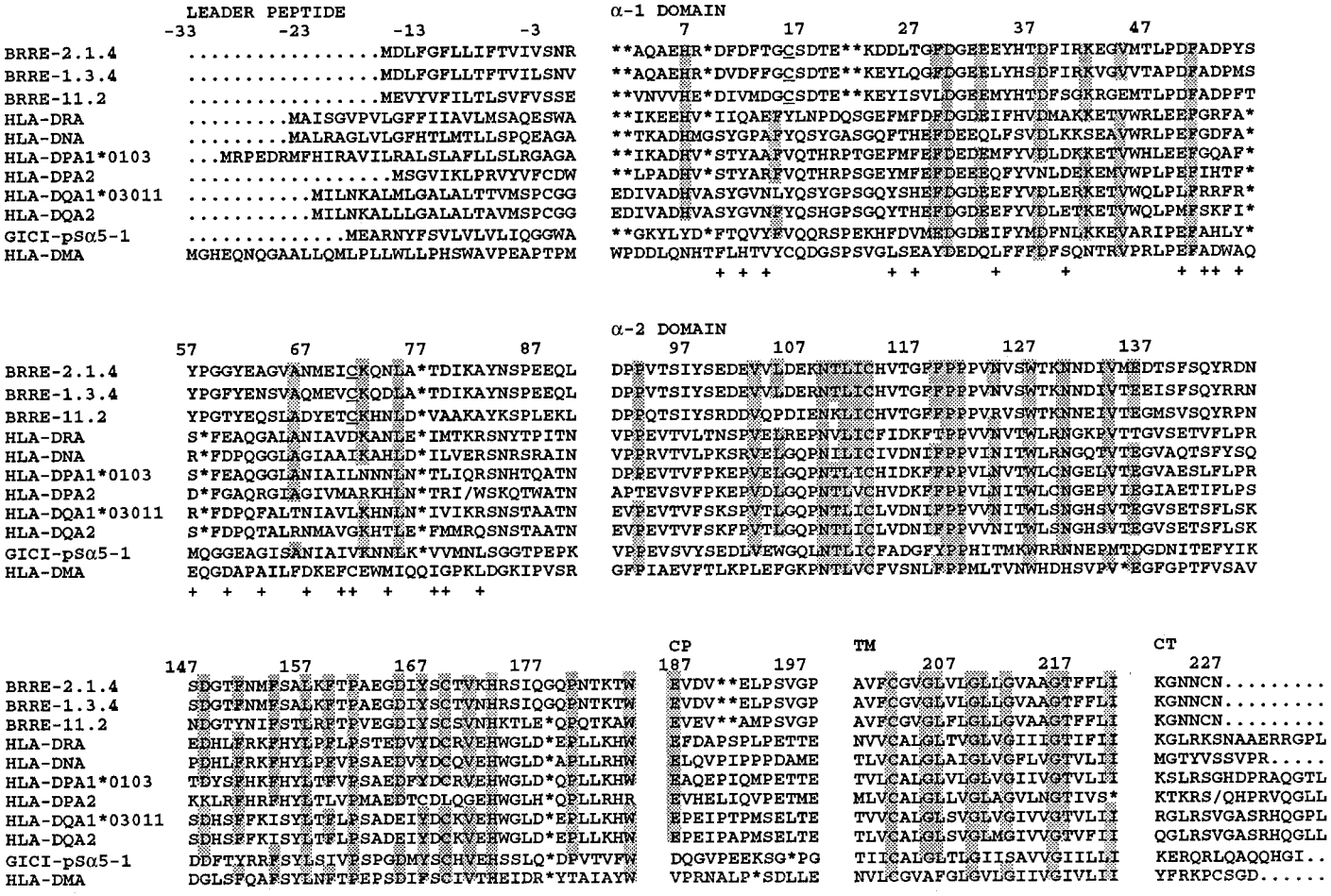

Fig. 2. Amino acid sequence of selected *Mhc* class II α chains. Presumed insertions/deletions are indicated by *asterisks*, shared residues between zebrafish and most of the mammalian and shark sequences are *highlighted*. The + sign is shown below residues of the putative PBR. Cysteine residues forming a putative disulfide bridge are *underlined*. The *horizontal bar* indicates putative glycosylation sites. Sources of the sequences are: Gici-pSα5-1, Kasahara and co-workers (1992); HLA-DRA, Schamboek and co-workers (1983); HLA-DNA, Jonsson and Rask (1989); HLA-DPA1*0103 and HLA-DPA2, Lawrance and co-workers (1985); HLA-DQA1*03011 and HLA-DQA2, Jonsson and co-workers (1989); HLA-DMA, Kelly and co-workers (1991). CP, connecting peptide; TM, transmembrane region; CT, cytoplasmic domain. Amino acid residues are given in the international one-letter code. The numbering is based on the HLA-DQA sequence.

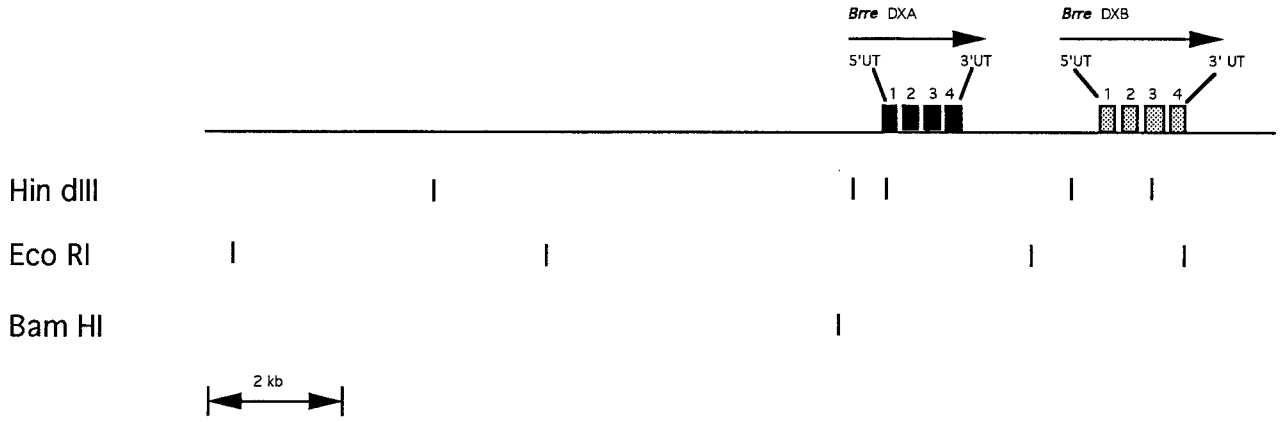

Fig. 3. Restriction map of the genomic clone 11.2. The length of the clone (in kb pairs) is indicated by the *horizontal line*, orientation of the genes by *arrows*, and exons by *filled* (*DXA*) or *shaded* (*DXB*) *rectangles*. Restriction endonucleases used in the mapping are indicated on the *left-hand* side. *Vertical lines* indicate positions of restriction sites. UT, untranslated region; 1,2,3,4 exons.

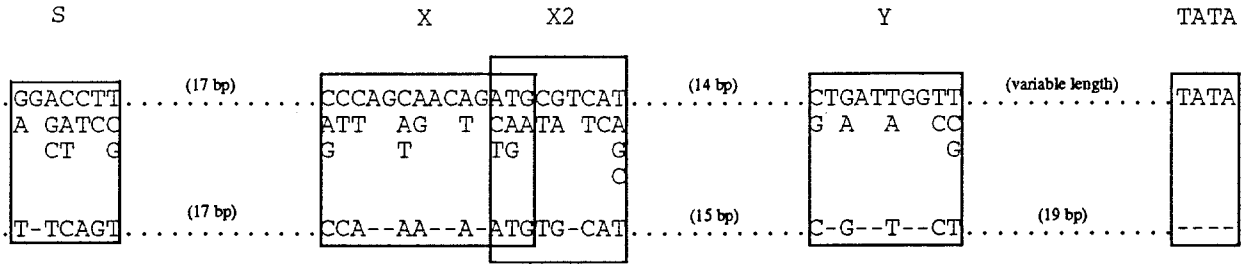

Fig. 4. Sequences of regulatory elements in the promoter region of the zebrafish class II *A* gene (lower sequence) compared to the sequence of mammalian class II *A* genes (Benoist and Mathis 1990; Cogswell et al. 1991). Individual elements are *boxed*, distances between them are indicated in bp. Identities with the top sequence are indicated by *dashes* (–).

sequence similarity between the zebrafish and mammalian (DM excluded) α chains is 29%. Several residues are shared by all or most of the known α chains (Fig. 2).

The α2 domain of the zebrafish sequence is one residue longer (position 178) than the mammalian DN, DP, DQ, and DR α2 domains. The zebrafish sequence shares a potential glycosylation signal at positions 124–126 with the mammalian α chains, with the exception of DMα (Fig. 2). The shark sequence also lacks the glycosylation signal at this position. With the exception of DPA2, the mammalian DNα, DPα, DQα, and DRα chains have a second glycosylation signal in the α1 domain. The zebrafish α chains lack this signal, as do all known class II β chains. The sequence similarity between the zebrafish and mammalian α2 domains (excluding DMα) is 36%. At a number of positions, α2 residues (including the cysteines at positions 113 and 169) are fully conserved among all the known α chains (Fig. 2).

The connecting-peptide segment of the zebrafish α chain is two residues shorter than the corresponding segments of the DNα, DPα, DQα, and DRα chains, and one residue shorter than those of the DMα and the shark α chains. The sequence similarity among the various chains is very low in this region.

The transmembrane domain of the zebrafish α chain is of the same length as that of the mammalian α chains. The motif GXXXGXXGXXXG (X being a hydrophobic residue) is shared by the zebrafish and mammalian α chains. The G residues of this motif have been postulated by Cosson and Bonifacino (1992) to form the face of an α-helix that interacts with the corresponding phase of the β chain in the αβ heterodimer. However, at position 198, where mammalian α chains have a negatively charged residue (Glu), the zebrafish and the shark α chains have uncharged residues (Pro and Gly, respectively). The Glu198 has been postulated by Cosson and Bonifacino (1992) to interact with the positively charged Lys at the corresponding position of the β chain.

The cytoplasmic domain of the zebrafish α chain is only six amino acid residues long and thus one of the shortest cytoplasmic tails of the class II proteins on record (Klein 1986). It shares Lys and Gly at positions 222 and 223 with some of the mammalian α chains, but otherwise the sequences are very different from one another.

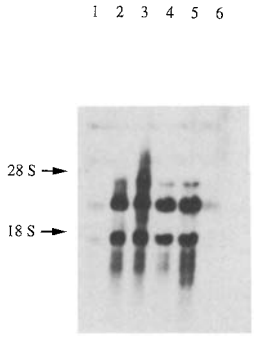

Fig. 5. Northern blot of RNA from different carp tissues hybridized with the cDNA probe 2.1.4. The sizes of the ribosomal RNA are indicated by *arrows*. Lane numbering: 1, heart; 2, intestine; 3, spleen; 4, hepatopancreas; 5, pronephros; 6, ovary.

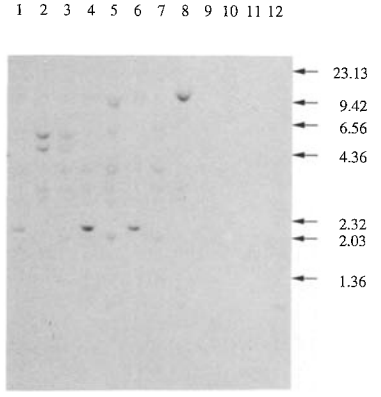

Fig. 6. Southern blot of total genomic DNA from nine zebrafish individuals. (Lanes 1,2,3,8,9, wild-type; 4, albino; 5, albino × frankei; 6, frankei; 7, homozygous laboratory strain; 10, trout (*Oncorhynchus mykiss*); 11, cichlid (*Melanochromis auratus*); 12, carp (*Cyprinus carpio*). The DNA was digested with *Hin* dIII and hybridized with the cDNA probe 2.1.4. Sizes of the marker in kb pairs are indicated by *arrows*.

The 3′UT region of the zebrafish clone 2.1.4 has a polyadenylation signal AATAAA 145 bp downstream from the termination signal, followed by a poly(A) tail 45 bp downstream of the polyadenylation signal. The 1.4.3 sequence, which is 12 bp longer in the 5′UT region but otherwise very similar to the 2.1.4 sequence, loses its similarity just before the polyadenylation signal of the latter; it contains a potential polyadenylation signal 40 bp downstream of this point.

Isolation and characterization of a genomic fragment bearing the zebrafish class II A gene. Screening of the zebrafish phage library with the class II *B* gene probe cocktail yielded several *DXB* gene-bearing clones (H. Sültmann, W. Mayer, and J. Klein, in preparation). One of these clones, 11.2, also hybridized with the class II *A* cDNA probe. Further characterization of the 11.2 clone revealed that it contained one class II *B* and one complete class II *A* gene, the two genes being arranged in a head-to-tail orientation (Fig. 3). The class II *A* gene occupied a segment of the clone approximately 1.2 kb in length. This segment was sequenced and the disposition of exons and introns was determined by the presence of putative splicing signals (Breathnach and Chambon 1981) and by homology with zebrafish class II *A* cDNA sequences. The zebrafish *A* gene consists of four exons interrupted by three relatively short introns (126 bp, 129 bp, and 174 bp in length). The distance between the *A* and *B* genes is less than 3 kb.

Comparison with the consensus sequences (Kozak 1984) led to the identification of two putative transcription initiation sites, one of which is more conserved than the other. Assuming that the more conserved one is the true initiation signal, the primary transcript could be expected to be 1196 nucleotides long and to contain 708 nucleotides of coding sequence. The putative exon 1 consists of 52 bp of 5′UT region, 51 bp of sequence coding for the 17 amino acid residues of the putative leader peptide, and 7 bp coding for two amino acid residues of the α1 domain. The 246 bp of exon 2 code for the remainder of the α1 domain, the 282 bp of exon 3 code for the α2 domain, and the 132 or 240 bp of exon 4 code for the connecting peptide, the transmembrane region, the cytoplasmic domain, and the 3′UT region (there are two polyadenylation signals in this particular sequence, one 10 bp and the other 118 bp downstream from the stop codon and it is is not clear which of the two is used). Sequencing upstream from the transcription-initiation signals revealed the presence of at least four regulatory elements: the X-box (14 bp), the X2-box (9 bp, partially overlapping with the X-box), the Y box (10 bp), and the TATA box (4 bp; see Fig. 4). The distance between the X- and Y-boxes is 21 bp and that between the Y- and the TATA-boxes is 19 bp; the TATA box is separated from the transcription initiation signal by 19 bp of sequence. The zebrafish class II *A* gene is thus rather compact, a characteristic which it shares with other fish *Mhc* class II genes already characterized (Ono et al. 1992, 1993a, b).

The overall sequence similarity between the coding sequence of the genomic clone and that of the cDNA clone is 73%. The similarities of the individual exons are as follows: exon 1, 53%; exon 2, 69%; exon 3, 74%; and exon 4, 89%. The 3′UT regions of the two sequences differ strikingly from each other, suggesting that the cDNA and genomic sequences might be derived from different class II *A* loci.

Tissue distribution. To find out in which tissues the fish class II *A* genes are transcribed, we carried out northern blot analysis. The small size of the zebrafish (the adult has a length of no more than 5 cm), which would make it difficult to isolate RNA from some of the organs, compelled us to use carp RNA instead. Carp belongs to the same Cyprinidae family as the zebrafish, and the zebrafish class II *A* probes crosshybridize readily with carp DNA. In the northern blot, the zebrafish class II A probe (cDNA probe 2.1.4) hybridized with carp

```
            10        20        30        40        50        60        70        80        90
2.1.4   AQAEHRDFDFTGCSDTEKDDLTGFDGEEEYHTDFIRKEGVMTLPDFADPYSYPGGYEAGVANMEICKQNLATDIKAYNSPEEQL
1.4.3   -------V--F--------EY-Q------L--S-----V--V-A------M----F--NS--Q--V---D--------------
MW1273        -V--F--------EY-Q------L--S-----V--V-A------M----F--NS--Q--V---D--------
158.1         ---------------S------V--------------------A-K----------I---------S--**
158.2         ---------------S------V--------------------V-K----------I---------Y----
159.1         ----L---E---EY-Q------LF-S-----V--V-A------V-F--F--NS--Q--V------------
167.1         ----Y---E---EY-E------L--S-----V--V-A------K----F--NS--Q--V------------
170.3         --Q-V----------N-------------------------------------V-------Y-E--
171.1         --Q-M---E----VFI-------------------------------SFDIV-GQ--A-------Y----
171.2         --Q-M---E---EF-L---------A---------------------SFDIV-GQ--A-------Y----
S4            -V--F--------EF-Q------L--S-----V--V-A------M----F--NS--Q--V---D--------
S5.3             -F--------EF-Q------P--S-----V--V-A------M----F--NS--Q--V---D--------
G1            --S------------S------V--------------------A-R-----------A------------
G17           ---------------S------V--------------------A-R-----------A------------
```

Fig. 7. Amino acid sequences translated from exon 2 PCR product nucleotide sequences. The products were obtained with TU579/TU580 (S5.3) and TU450/TU461 (all others) primers. Identity with the top sequence (2.1.4) is indicated by *dashes* (–); *asterisks* in 158.1 are deletions probably due to a PCR artifact. Individuals 158, 159, 167, 170, and 171 were caught at different locations in North Bengal, India; individual MW1273 originated from a laboratory strain. S4 and S5.3 are amplification products from a 3 kb *Hin* dIII fragment of a genomic Southern blot; G1 and G17 were amplified from a zebrafish genomic library (Stratagene). Several of the sequences were found repeatedly in different clones from one PCR. Numbering of residues corresponds to that used in *Figure 2*.

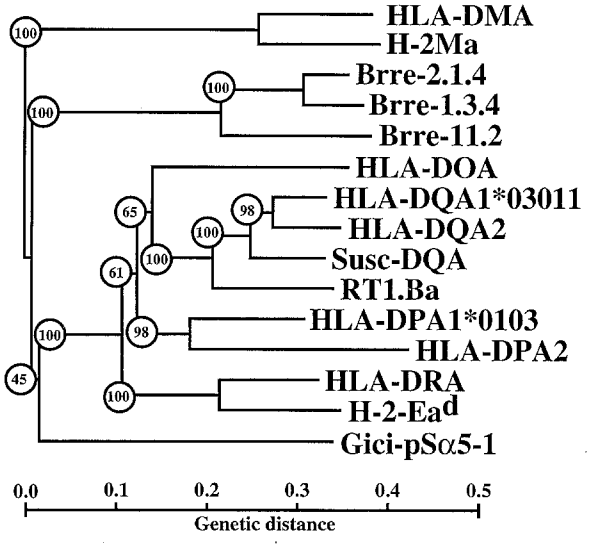

Fig. 8. Dendrogram based on genetic distances between protein sequences from the nurse shark (Gici), zebrafish (Brre), mouse (H-2), pig (Susc), rat (RT1), and human (HLA). Genetic distances were determined from percent sequence identity between proteins. Numbers on each node indicate frequency with which this node has been recovered per 100 bootstrap replications in a total of 500 replications. References for the individual sequences are: H-2Ma, Cho and co-workers (1991); Susc-DQA, Hirsch and co-workers (1990); Rt1.Ba, Barran and McMaster (1987); H2-EA^d, Hyldig-Nielsen and co-workers (1983); for the remaining sequences, references are as in Figure 2.

RNA isolated from spleen, hepatopancreas, pronephros, and intestine; it failed to hybridize with RNA isolated from heart and ovaries (Fig. 5). This distribution of tissue expression is the same as that observed for the class II *B* genes (Ono et al. 1992) and it correlates with the relative abundance of lymphoid/myeloid cells in the different organs. Like those of mammals (Klein 1986), fish class II genes are presumably primarily transcribed in lymphoid cells. The northern blots showed three hybridizing bands, two strong ones approximately 1800 and 2500 nucleotides in length, and a weaker one approximately 4500 nucleotides in length.

Variability of class II A genes. To determine how polymorphic the zebrafish class II *A* genes might be, we isolated genomic DNA from individual fishes of different origin and used it for Southern blot and PCR analysis. In the former method, we isolated DNA from nine individuals, digested it with the *Hin* dIII restriction endonuclease, blotted the digests onto filters, and hybridized the blots with the cDNA probe 2.1.4. Included in the blots were also digests from carp, trout, and cichlid fish. In the zebrafish, the number of hybridizing fragments ranged from one to eight and all individuals gave different hybridizing patterns (Fig. 6). Crosshybridization (two bands) with the carp DNA was also obtained but no hybridization was observed with DNA of the trout (family Salmonidae) or of the cichlid fish (family Percidae). The variation in intensity and number of hybridizing bands among the zebrafish suggested the existence of several class II *A* loci and the occurrence of haplotype polymorphism in this species. Crosshybridization of the zebrafish probe with DNA from carp indicated relative sequence conservation between the fish class II *A* genes of these species.

In the PCR analysis, we amplified DNA from six zebrafish, subcloned the amplification products, and sequenced the inserts of the subclones. The primers used for the amplification annealed to the 5′ end (primers TU461 and TU579) and 3′ end (primers TU450 and TU580) of exon 2, so that the product of specific amplification was approximately 250 bp in length. Altogether we obtained 15 sequences, eight of which proved to be different from one another (Fig. 7). The sequence similarity of the individual subclones ranged from 84%–98% (excluding identical sequences) and this translated into a range of 72%–97% amino acid sequence similarity. When combined, the sequences of the PCR products and of the cDNA

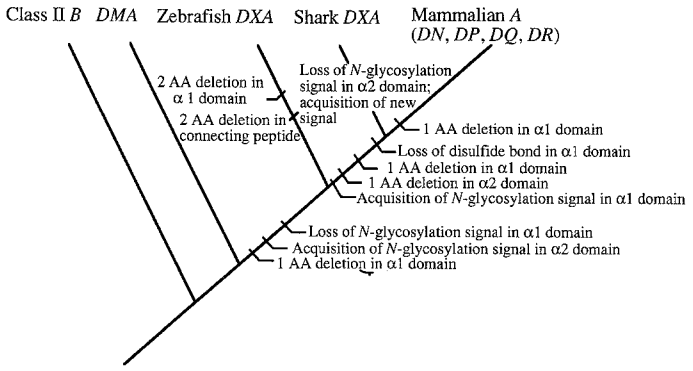

Fig. 9. Cladogram depicting the individual events in the evolution of the *Mhc* class II *A* genes as inferred from amino acid sequences (AA). For explanation, see text.

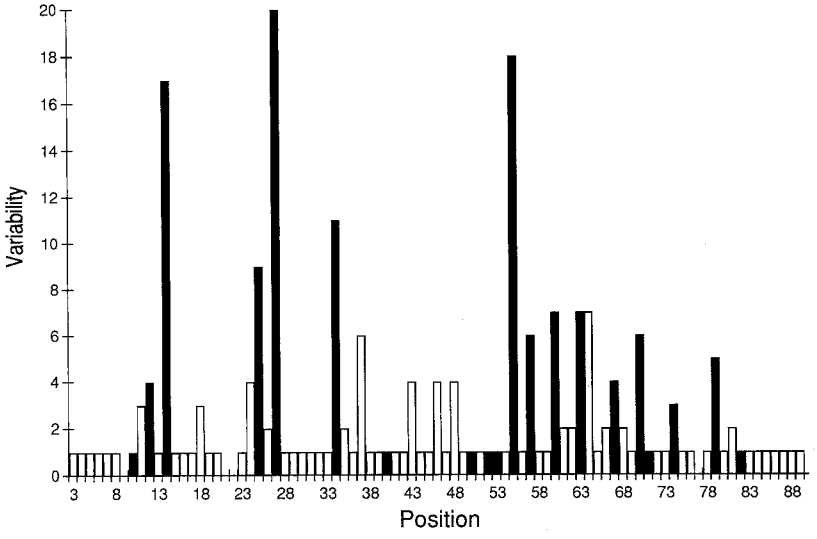

Fig. 10. Wu-Kabat variability plot of Brre class II α1 domain amino acid sequences from Figure 7. Variability score (ordinate) is given by the number of different amino acid residues/frequency of the most common residue for each site (Wu and Kabat 1970). Horizontal axis shows position of each site; the numbering corresponds to that used in *Figure 2*. *Solid bars* indicate residues corresponding to the PBR of the mammalian class II α chains (Brown et al. 1988).

clones can be divided into six groups with intragroup similarity of 93%–97% and intergroup similarity of 67%–91%. The groups roughly correlated with the place of origin of the DNA donors, suggesting that at least part of the observed variability may represent allelic polymorphism.

Discussion

There are two approaches to analyzing evolutionary relationships among molecules such as the class II α chains. The first involves comparing the sequences and either determining similarities among them or reconstructing the history of substitutions as far as possible and then expressing the results of the analysis in the form of a dendrogram, using one or several of the algorithms available for this purpose (Nei 1987). In Figure 8 we used the neighbor-joining algorithm of Saitou and Nei (1987) to obtain a dendrogram depicting possible phylogenetic relationships among the class II α chains analyzed thus far. The phylogenetic tree reveals that DMα (thus far known from mammals only; Cho et al. 1991; Kelly et al. 1991) emerges first, followed by zebrafish class II α, nurse shark class II α, and mammalian DN, DP, DQ, and DR α chains. This topology is confirmed using class II β chain as an outgroup (not shown).

The second approach is to reconstruct the branching sequence of phylogeny through an analysis of derived characters, that is characters shared by two or more taxa under consideration but absent in earlier ancestors (Wiley et al. 1991). The result is a cladogram as depicted in Figure 9. The hierarchies of the class II α chains in the dendrogram (Fig. 8) and the cladogram (Fig. 9) are the same. The characters taken into consideration in the construction of the cladogram are the presence and disposition of cysteine residues that could form a disulfide bond in the α1 domain, the loss of amino acid residues in the α1, α2, and connecting-peptide domains, and the presence or absence and the disposition of *N*-glycosylation sites in α1 and α2 domains. The consistency of the cladogram with the dendrogram is, of course, not surprising, because two of the three

characters on which the former is based are determined by the sequence used to construct the latter (deletions were not taken into account in the construction of the dendrogram). Since, however, the characters used in the cladistic analysis may be functionally more important than many of the substitutions entered in the dendrogram analysis, the consistency does add weight to the implied phylogeny of the class II α chains.

The phylogeny of the α chains does not contradict the implied evolutionary history of studied organisms, with one important exception: the Chondrichthyes (cartilaginous fishes), represented by the nurse shark (Kasahara et al. 1992), appeared on the evolutionary scene before the Osteichthyes (bony fishes), here represented by the zebrafish (Carroll 1988). The discrepancy must be an indication that the genes for the zebrafish and shark class II α chains are not orthologous. We postulate, therefore, that the bony fishes either still have a gene orthologous to the known shark class II *A* gene or that they had such a gene which was subsequently lost. Similarly, the cartilaginous fishes must have had (and perhaps still have) another, more "primitive" class II *A* gene which encodes an α chain more similar to the known zebrafish α chain than the one known thus far. The postulated additional class II *A* genes in the bony fishes and the cartilaginous fishes may differ to such an extent from the genes already identified in representatives of these taxa that they are not detected by crosshybridization with available probes. All the known mammalian class II *A* genes, with the exception of *DMA*, appear to be derived from a common ancestor which emerged after the separation of fish and tetrapods. The *DMA* gene appears to be very ancient, as has been postulated by Cho and co-workers (1991) and Kelly and co-workers (1991).

The zebrafish *DXA* genes described in this communication are probably functional. We base this conclusion on three observations. First, we found no obvious defect in the entire sequence of the genes, which includes the regulatory region. The genes are apparently transcribed in selected tissues and cells (probably lymphoid and myeloid, as indicated by the northern blot experiments) and the short introns are apparently spliced out correctly from the primary transcripts (as indicated by the presence of canonical splicing signals). Second, the *DXA* loci are polymorphic and the bulk of the variation is concentrated in exon 2. The Wu-Kabat plot of variability in the sequenced genes (Fig. 10) reveals that variability is highest in exon 2 at positions predicted by Brown and co-workers (1988) to code for amino acid residues of the PBR. Although PBR variability is not an absolute requirement for functionality of α chain-encoding class II genes (the mammalian *DRA* genes are not polymorphic but functonal; see Carson and Trowsdale 1986), when it does occur, it is indicative of functionality. Third, the ratio of nonsynonymous (at the PBR-encoding part) to synonymous substitutions of the zebrafish *DXA* genes is greater than one and can be as high as seven for some gene pairs (data not shown). This observation suggests that balancing selection is effecting preferential retention of nonsynonymous substitutions in the PBR (Hughes and Nei 1988). Presence of balancing selection is, in turn, indicative of functionality.

In mammals, the families of class II loci (*DO, DP, DQ, DR*) apparently emerged after the separation of marsupial and placental mammals more than 120 million years (my) ago, but most probably at the time of the great radiation of placental mammals 65 my ago (Schneider et al. 1991). Only the *DM* family might be more ancient (Cho et al. 1991; Kelly et al. 1991). The fact that the different gene families possess similar regulatory elements in their promoter region [reviewed by Benoist and Mathis (1990) and Cogswell et al. (1991)] indicates that the elements were established more than 65 my ago. Our observation that very similar regulatory elements are also found in a representative of the teleost fishes suggests that the basic organization of the class II gene promoter region was established before the fish-tetrapod split over 400 my ago. The similarities in the class II regulatory region between teleost fishes and mammals include the presence of the same elements (boxes), the same order of the boxes, the same spacing between the boxes, and high sequence similarity of homologous boxes in the two taxa. The conservation of the spacing supports site-directed mutagenesis experiments which demonstrate reduction of promoter activity in mutants with altered stereospecific alignment of the boxes (Vilen et al. 1991). The correct spacing of the regulatory elements is believed to be necessary to orient the boxes two turns apart on the same face of the DNA double helix for correct attachment of the DNA-binding proteins. Our data indicate that the promoter is among the most conserved regions of the class II genes.

One striking difference between the organization of the zebrafish and mammalian class II *A* genes is in the length of the introns. Short introns have now been documented for the zebrafish class II *A* (this communication), the zebrafish (Ono et al. 1992; H. Sültmann, W. Mayer and J. Klein, unpublished data), carp (Ono et al. 1993d; Hashimoto et al. 1990), and cichlid (Ono et al. 1993b) class II *B*, and for the zebrafish class I *B* (Ono et al. 1993c) and carp class I *A* (Hashimoto et al. 1990) genes. Although introns several kb in length are common in the *Mhc* genes in mammals (reviewed by Klein 1986), in all the fish *Mhc* genes studied so far, nearly all the introns are less than 1 kb long and most are less than 100 bp long. This difference between fish and mammals may not be limited to the *Mhc* genes (e.g., Muller-Schmid et al. 1992; Hong and Schartl 1992; Huang et al. 1992; Ber and Daniel 1992; Scott et al. 1988; Sterrer et al. 1990; Wilson et al. 1991; Liu et al. 1990; Chang et al. 1991; Chiou et al. 1990;

Adam et al. 1991) but could reflect a principal difference in the organization of the genomes in the two vertebrate classes (Bernardi and Bernardi 1991). A practical consequence of this observation is that molecular characterization of fish genes is relatively easy compared to that of mammalian genes.

Acknowledgments. We thank Dr. Stefan Schulte-Merker, National Institute of Medical Research, London, U.K., for supplying fish strains, Dr. Herbert Tichy for providing some of the fish, Ms. Stefanie Zughorst for preparation of zebrafish organs, Ms. Anica Milosev for the computer graphics, and Ms. Lynne Yakes for preparation of the manuscript. The work was supported, in part, by grant AI 23667 from the National Institutes of Health, Bethesda, MD.

References

Adam, D., Maeueler, W., and Schartl, M. Transcriptional activation of the melanoma inducing Xmrk oncogene in *Xiphophorus*. *Oncogene 6:* 73–80, 1991

Auffray, C., Lillie, J. W., Arnot, D., Grossberger, D., Kappes, D., and Strominger, J. L. Isotypic and allotypic variation of human class II histocompatibility antigen α-chain genes. *Nature 308:* 327–333, 1984

Barran, P. A. and McMaster, W. R. DNA sequence analysis of the rat RT1.B alpha gene. *Immunogenetics 26:* 56–62, 1987

Benoist, C. and Mathis, D. Regulation of major histocompatibility complex class II genes: X, Y, and other letters of the alphabet. *Annu Rev Immunol 8:* 681–715, 1990

Ber, R. and Daniel, V. Structure and sequence of the growth hormone-encoding gene from *Tilapia nilotica*. *Gene 113:* 245–250, 1992

Bernardi, G. and Bernardi, G. Compositional properties of nuclear genes from cold blooded vertebrates. *J Mol Evol 33:* 57–67, 1991

Bjorkman, P. J. and Parham, P. Structure, function, and diversity of class I major histocompatibility complex molecules. *Annu Rev Biochem 59:* 253–288, 1990

Breathnach, R. and Chambon, P. Organization and expression of eucaryotic split genes coding for proteins. *Annu Rev Biochem 50:* 349–383, 1981

Brown, J. H., Jerdetzky, T., Saper, M. A., Samraoui, B., Bjorkman, P. J., and Wiley, D. C. A hypothetical model of the foreign antigen binding site of class II histocompatibility molecules. *Nature 332:* 845–850, 1988

Carroll, R. L. *Vertebrate Paleontology and Evolution*, W. H. Freeman, New York, 1988

Carson, S. and Trowsdale, J. Molecular organization of the class II genes of the human and mouse major histocompatibility complexes. *Oxford Surv Euk Genes 3:* 63–94, 1986

Chang, T., Lin, C. L., Chang, W. C., and Chen, P. H. Gamma crystallin genes in carp: cloning and characterisation. *Biochim Biophys Acta 1090:* 261–264, 1991

Chiou, C. S., Chen, H. T., and Chang, W. C. The complete nucleotide sequence of the growth-hormone gene from the common carp (*Cyprinus carpio*). *Biochim Biophys Acta 1087:* 91–94, 1990

Cho, S., Attaya, M., and Monaco, J. New class II-like genes in the murine MHC. *Nature 353:* 573–576, 1991

Cogswell, J. P., Zeleznik-Le, N., and Ting, J. P.-Y. Transcriptional regulation of the HLA-DRA gene. *Crit Rev Immunol 11:* 87–112, 1991

Cosson, P. and Bonifacino, J. S. Role of transmembrane domain interactions in the assembly of class II MHC molecules. *Science 258:* 659–662, 1992

Feinberg, A. P. and Vogelstein, B. A. technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal Biochem 132:* 6–13, 1983

Hashimoto, K., Nakanishi, T., and Kurosawa, Y. Isolation of carp genes encoding major histocompatibility complex antigens. *Proc Natl Acad Sci USA 87:* 6863–6867, 1990

Hirsch, F., Sachs, D. H., Gustafsson, K., Pratt, K., Germana, S., and LeGuern, C. Class II genes of miniature swine. III. Characterization of an expressed pig class II gene homologous to *HLA-DQA*. *Immunogenetics 31:* 52–56, 1990

Hong, Y. and Schartl, M. Structure of the rainbow trout metallothionein A gene. *Gene 120:* 277–279, 1992

Huang, C. J., Huang, F. L., Wang, Y. C., Chang, Y. S., and Lo, T. B. Organization and nucleotide sequence of carp gonadotropin alpha subunit genes. *Biochim Biophys Acta 1129:* 239–242, 1992

Hordvik, I., Grimholt, U., Fosse, V. M., Lie, O., and Endresen, C. Cloning and sequence analysis of cDNAs encoding the MHC class II β chain in Atlantic salmon, *Salmo salar*. *Immunogenetics 37:* 437–441, 1993

Hughes, A. L. and Nei, M. Pattern of nucleotide substitutions at major histocompatibility complex class I loci reveals overdominant selection. *Nature 355:* 167–170, 1988

Hyldig-Nielsen, J. J., Schenning, L., Hämmerling, U., Widmark, E., Heldin, E., Lind, P., Servenius, B., Lund, T., Flavell, R., Lee, J. S., Trowsdale, J., Schreier, P. H., Zablitzky, F., Larhammar, D., Peterson, P. A., and Rask, L. The complete nucleotide sequence of the I-E-alpha-d immune response gene. *Nucleic Acids Res 11:* 5055–5071, 1983

Jonsson, A. K. and Rask, L. Human class II *DNA* and *DOB* genes display low sequence variability. *Immunogenetics 29:* 411–413, 1989

Jonsson, A. K., Andersson, L., and Rask, L. Complete sequences of DQA and DQB1 cDNA clones corresponding to the DQw4 specificity. *Immunogenetics 30:* 232–234, 1989

Juul-Madsen, H. R., Glamann, J., Madsen, H. O., and Simonsen, M. MHC class II beta-chain expression in the rainbow trout. *Scand J Immunol 35:* 687–694, 1992

Kasahara, M., Vazquez, M., Sato, K., McKinney, E. C., and Flajnik, M. F. Evolution of the major histocompatibility complex: isolation of a class II *A* gene from the cartilaginous fish. *Proc Natl Acad Sci USA 89:* 6688–6692, 1992

Kelly, A. P., Monaco, J. J., Cho, S., and Trowsdale, J. A new human HLA class II-related locus, *DM*. *Nature 353:* 571–573, 1991

Kimura, M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol 16:* 111–120, 1980

Klein, D., Ono, H., O'hUigin, C., Vincek, V., Goldschmidt, T., and Klein, J. Extensive MHC variability in cichlid fishes of Lake Malawi. *Nature 364:* 330–334, 1993

Klein, J. *Natural History of the Major Histocompatibility Complex*. John Wiley, New York, 1986

Kozak, M. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids Res 12:* 857–872, 1984

Laale, H. W. The biology and use of zebrafish, *Brachydanio rerio*, in fisheries research. A literature review. *J Fish Biol 10:* 121–173, 1977

Lanzavecchia, A. Receptor-mediated antigen uptake and its effect on antigen presentation to class II-restricted T lymphocytes. *Annu Rev Immunol 8:* 773–793, 1990

Lawrance, S. K., Das, H. K., Pan, J., and Weissman, S. M. The genomic organisation and nucleotide sequence of the HLA-SB (DP) alpha gene. *Nucleic Acids Res 13:* 7515–7528, 1985

Lee, J. S., Trowsdale, J., Travers, P. J., Carey, J., Grosveld, F., Jenkins, J., and Bodmer, W. F. Sequence of an *HLA-DR* α-chain cDNA clone and intron-exon organization of the corresponding gene. *Nature 299:* 750–752, 1982

Liu, Z., Zhu, Z., Roberg, K., Faras, A., Guise, K., Kapuscinski, A. R., and Hackett, P. B. Isolation and characterization of beta-actin gene of carp (*Cyprinus carpio*). *DNA Seq 1:* 125–136, 1990

Muller-Schmid, A., Rinder, H., Lottspeich, F., Gertzen, E. M., and Hoffmann, W. Ependymis from the cerebrospinal fluid of salmonid fish: gene structure and molecular characterization. *Gene 118:* 189–196, 1992

Nei, M. *Molecular Evolutionary Genetics*, Columbia University Press, New York, 1987

Ono, H., Klein, D., Vincek, V., Figueroa, F., O'hUigin, C., Tichy, H., and Klein, J. *Mhc* class II genes of zebrafish. *Proc Natl Acad Sci USA 89:* 11886–11890, 1992

Ono, H., O'hUigin, C., Tichy, H., and Klein, J. Major histocompatibility complex variation in two species of cichlid fishes from Lake Malawi. *Mol Biol Evol*, in press, 1993 a

Ono, H., O'hUigin, C., Vincek, V., and Klein, J. Exon-intron organization of fish major histocompatibility complex class II *B* genes. *Immunogenetics 38:* 223–234, 1993 b

Ono, H., Figueroa, F., O'hUigin, C., and Klein, J. Cloning of the β_2-microglobulin gene in the zebrafish. *Immunogenetics 38:* 1–10, 1993 c

Ono, H., O'hUigin, C., Vincek, V., Stet, R. J. M., Figueroa, F., and Klein, J. New β chain-encoding *Mhc* class II genes in the carp. *Immunogenetics 38:* 146–149, 1993 d

Rothbard, J. B. and Gefter, M. L. Interactions between immunogenetic peptides and Mhc proteins. *Annu Rev Immunol 9:* 527–565, 1991

Saitou, N. and Nei, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol 4:* 406–425, 1987

Sambrook, J., Fritsch, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989

Sanger, F., Nicklen, S., and Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA 74:* 5463–5467, 1977

Schamboeck, A., Korman, A. J., Kamb, A., and Strominger, J. L. Organization of the transcriptional unit of a human class II histocompatibility antigen: HLA-DR heavy chain. *Nucleic Acids Res 11:* 8663–8675, 1983

Schneider, S., Vincek, V., Tichy, H., Figueroa, F., and Klein, J. MHC class II genes of a marsupial, the red-necked wallaby (*Macropus rufogriseus*): identification of new gene families. *Mol Biol Evol 8:* 753–766, 1991

Scott, G. K., Hayes, P. H., Fletcher, G. L., and Davies, P. L. Wolffish antifreeze protein genes are primarily organized as tandem repeats that each contain two genes in inverted orientation. *Mol Cell Biol 8:* 3670–3675, 1988

Sterrer, S., Koenigstorfer, A., and Hoffmann, W., Biosynthesis and expression of ependymin homologous sequences in zebrafish brain. *Neuroscience 37:* 277–284, 1990

Sültmann, H. *Der Haupthistokompatibilitätskomplex in Fischen.* Diplomarbeit, Eberhard-Karls-Universität Tübingen, Tübingen, 1991

Vilen, B. J., Cogswell, J. P., and Ting, J. P.-Y. Requirement for stereospecific alignment of the X and Y elements in MHC class II DRA function. *Mol Cell Biol 11:* 2406–2415, 1991

Von Heijne, G. Patterns of amino acids near signal-sequence cleavage sites. *Eur J Biochem 133:* 17–21, 1983

Wiley, E. O., Siegel-Causey, D., Brooks, D. R., and Funk, V. A. *The Compleat Cladist. A Primer of Phylogenetic Procedures*, University of Kansas Museum of Natural History, Lawrance, Kansas, 1991

Wilson, M. R., Middleton, D., and Warr, G. W. Immunoglobulin V(H) genes of the goldfish, *Carassius auratus*: a reexamination. *Mol Immunol 28:* 449–457, 1991

Wu, T. T. and Kabat, E. A. An analysis of the sequences of the variable regions of Bence Jones proteins and myeloma light chains and their implications for antibody complementarity. *J Exp Med 132:* 211–250, 1970

Zhu, Z., Vincek, V., Figueroa, F., Schönbach, C., and Klein, J. *Mhc-DRB* genes of the pigtail macaque (*Macaca nemestrina*): implications for the evolution of human *DRB* genes. *Mol Biol Evol 8:* 563–578, 1991