Major histocompatibility complex genes in the common carp, *Cyprinus carpio* L.

Promotor:	dr. W.B. van Muiswinkel
	hoogleraar in de zoölogische celbiologie
Co-promotoren:	dr. R.J.M. Stet
	universitair docent in de zoölogische celbiologie/immunogenetica
	dr. E. Egberts
	universitair hoofddocent in de zoölogische celbiologie/immunologie

Major histocompatibility complex genes in the common carp, *Cyprinus carpio* L.

Saskia H.M. van Erp

Proefschrift
ter verkrijging van de graad van doctor
in de landbouw- en milieuwetenschappen,
op gezag van de rector magnificus,
dr. C.M. Karssen,
in het openbaar te verdedigen
op woensdag 29 mei 1996
des namiddags te half twee in de aula
van de Landbouwuniversiteit te Wageningen

CIP-GEGEVENS KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Erp, Saskia H.M. van

Major histocompatibility complex genes in the common carp, *Cyprinus carpio* L. / Saskia H.M. van Erp. - [S.l. : s.n.]. - Ill.
Thesis Landbouwuniversiteit Wageningen. - With ref. - With summary in Dutch.
ISBN: 90-5485-509-6
Subject headings: carp / major histocompatibility complex / comparative immunology.

Abstract

Van Erp, S.H.M. 1996. Major histocompatibility complex genes in the common carp, *Cyprinus carpio* L. Doctoral thesis, Dept. Experimental Animal Morphology and Cell Biology, Wageningen Agricultural University, P.O. Box 338, 6700 AH Wageningen, The Netherlands.

This thesis describes a study of the major histocompatibility complex (*Mhc*) genes of the common carp (*Cyprinus carpio* L.). The molecules encoded by *Mhc* genes play an essential role in the specific immune response, by presenting antigens to T lymphocytes. Knowledge of the *Mhc* of carp, therefore, contributes to our understanding of the immune response mechanisms in this species. In addition, it may give important insights in the phylogenetic development of these genes. The common carp was found to contain several distinct lineages of *Mhc* class I genes, denoted as *Cyca-U*, *Cyca-Z*, *Cyca*-TC16 and *Cyca*-C4. The *Cyca-U* sequences probably represent classical *Mhc* class I genes, of which most likely only a single locus is expressed in each individual. *Cyca-Z* genes are present in multiple polymorphic copies in the genome, but it is not clear whether these genes are expressed. The sequence of *Cyca*-TC16 is most similar to the class I genes of the coelacanth, a fish which is thought to be a representative of the evolutionary lineage leading to the tetrapods. It is, however, not clear whether *Cyca*-TC16 is expressed. In addition, the sequence encoding carp β_2-microglobulin was isolated. Although two *B2m* genes were detected in each individual, apparently only one of these is expressed. In contrast, at least four class II *B* genes may be expressed in a single animal. These genes are linked in two pairs, which, however, segregate independently. In addition, two expressed class II *A* sequences were identified, most likely derived from two separate loci. Both the class II *A* and *B* genes are likely to encode *bona fide* class II chains, the components of the cell-surface class II α-β heterodimer. Although carp thus possess a complete set of class I and class II genes, it is not yet clear whether these genes reside in a single genetic *Mhc* region.

Stellingen

1. De aanwezigheid van zowel *Mhc* klasse I als klasse II genen in de karper wijst erop dat al bij vissen de strategische tweedeling aanwezig is tussen immuunresponsen tegen intra- en extracellulaire antigenen. Waarschijnlijk zijn er dan ook CD4- en CD8-achtige T-cel co-receptoren aanwezig die hierin een sturende rol spelen. *dit proefschrift*

2. Naamgeving aan *Mhc* genen is gelijk de kip en het ei: een nomenclatuur die de relaties tussen genen helder en juist weerspiegelt is van groot belang voor het succes van het onderzoek, maar zij kan slechts juist zijn wanneer ditzelfde onderzoek de relaties volledig heeft opgehelderd. *dit proefschrift*

3. In termen van zijn *Mhc* is het juister de karper evolutionair gezien een vroegere vertebraat te noemen dan een lagere. *dit proefschrift*

4. Twijfel is een eerbetoon aan de waarheid
   - Ernest Renan -

5. In evolutionair opzicht is het ongunstig als carrièrevrouwen kinderloos blijven.
   - W.A.M. van Erp -

6. Dat karpers een hoge aaibaarheidsfactor hebben kan experimenteel worden aangetoond in het hoofdgebouw van de Landbouwuniversiteit Wageningen.

7. AIO's zijn als jonge hondjes: Hoe vaak ze ook kwispelend al experimenterend tegen een boom aan lopen, na de ergste duizeligheid kwispelen ze alweer.

8. Het is een schrijnende tegenstelling dat voortschrijdende verlichting door de wetenschap juist geloof in het occulte lijkt te stimuleren.

9. Bezuinigingen die tot gevolg hebben dat wetenschappers eigenhandig boekenplanken voor hun kamer moeten kopen en ophangen, leveren waarschijnlijk alleen de Gamma geld op.

10. Hoogstwaarschijnlijk is de volksgezondheid meer gebaat bij een radicale vernietiging van sigaretten dan van Brits rundvee. Uit het uitblijven van eerstgenoemde maatregel kan dan ook worden opgemaakt dat een gezonde economie al snel belangrijker is dan gezond volk.

Stellingen behorend bij het proefschrift
"Major histocompatibility complex genes in the common carp (Cyprinus carpio L.)"
van Saskia H.M. van Erp, Wageningen, 29 mei 1996.

Voorwoord

Mijn proeve van bekwaamheid. Die houdt U vast. U snapt hoe bekwaam ik me voel met dit boekje. Maar dit werkje betekent veel meer dan die proeve alleen. Het staat voor 5 jaar "nog even naar het lab gaan", voor 5 jaar adrenaline, vreugde en veel ontzetting. Zo gaat dat in de wetenschap. En gelukkig was daar altijd weer het "En? Is het gelukt?" van een van de mensen om me heen. Zonder hun aanwezigheid zou de Major Hysterical Crisis heel wat hoger tij hebben gevierd. Om te beginnen natuurlijk René, een rots in de branding, een baken op zee. Met bijbehorend aanvaringsgevaar natuurlijk. Maar samen hebben we het gered, René. Dankjewel. En natuurlijk ook Egbert, Brian, Pedro, Trudi en Marie-José. Het was goed werken en goed toeven in jullie gezelschap. Wim, dank je wel voor je blinde vertrouwen in het succes van mijn werk. De AIO's van EDC (inclusief doctor Geert) had ik voor geen goud willen missen. Veel eten, veel wijn, veel idealen. De wereld verklaard, de ziel gesterkt. Mooie avonden. Ook dank aan de studenten die slechts achteraf genoemd in dit boekje voorkomen. Erik en Joris, jullie waren gezellige maatjes op het lab. Ik heb minstens zoveel van jullie geleerd als jullie (hopelijk) van mij. Robert, Jeroen en Bart, hoewel slechts vanaf de zijlijn heb ik van jullie werk dankbaar enkele belangrijke graantjes meegepikt. Jan (Osse), jouw belangstellende bezoekjes aan het eind van een werkdag heb ik erg gewaardeerd. En verder natuurlijk de rest van EDC, dank jullie wel voor jullie interesse, koffiepraat, valentijnsrozen, gezelschap, enz. enz.

Minstens de helft van mijn leven als AIO had ik het genoegen te vertoeven tussen de Veefokkers op het MHC lab. Richard, je bent van goud. Je was onmisbaar voor me, in zowel praktische als persoonlijke zin. Jan, Martien, Tony (ik mis je confronterende grappen nog steeds, "Ik dacht dat ik een kameel hoorde...."), Carolien, Beja, Rosilde, Tineke, Esther, en alle anderen die in de jaren kwamen en gingen, ik heb genoten in jullie midden.

Peter, jouw opgewektheid is legendarisch. Dankjewel voor alle zonnige hulp aan de copiëermachine. Wim Valen, dank je wel voor alle plaatjes die je in de jaren voor me hebt geschoten. Mijn paranimfen, Gertjan en Richard, dank jullie voor alle steun en goede zorgen.

A few words of special thanks to Felipe Figueroa and Jan Klein, for their hospitality, support and volleyball-training during my stay in Tübingen.

Papa en mama, Floris en Jacintha, dank jullie voor de liefdevolle basis en het onvoorwaardelijk vertrouwen dat er altijd voor me is. Paul, je weet ondertussen evenveel van PCR en allelische lijnen als ik, ik hoop dat je nog lang mijn onderwijzerskwaliteiten zult waarderen. Dank je voor alle geduld en liefde.

En dan zijn daar nog de hoofdrolspelers. Stil op de achtergrond, lijdend voorwerp en onderwerp tegelijk. De karpers. Ze hebben geen idee wat voor geheimen ze hebben prijsgegeven. Zelfs niet die ene op de voorkant, met zijn neus er bovenop. Dank je wel visjes!

Contents

Chapter 1

General introduction

All warm- and cold-blooded vertebrates possess a more or less integrated immune system by which the organism defends itself against unwanted invaders like parasites, bacteria, and viruses, but also against dangers from within, like tumor cells. Harmful attacks by the immune system on self antigens have to be avoided, however, and a mechanism should therefore exist by which the immune system can discriminate between self and non-self. This discrimination is performed by T lymphocytes which are selected early in their development for tolerance to self, but reactivity to non-self. Antigen recognition by T cells is performed through their T-cell receptors which, however, can only recognize antigens when these are presented as peptides by *Mhc* molecules, the transmembrane glycoproteins encoded by the genes of the major histocompatibility complex (*Mhc*).

In mammals, there are two major classes of *Mhc* molecules, class I and class II, which are distinguished by their protein structure, function and tissue distribution. *Mhc* class I molecules are heterodimers of a heavy class I α chain (M_r 45,000), complexed with a light chain, β_2-microglobulin (M_r 12,000). The α chain consists of three extracellular domains, a transmembrane region and a cytoplasmic tail. The membrane proximal α_3 domain has an immunoglobulin (Ig)-like fold, whereas the membrane distal α_1 and α_2 domains together shape a peptide-binding groove, made up of a floor of anti-parallel ß strands lined by two α helices, which form the rims of the groove. In this groove a peptide can be anchored, which is restricted in length to 8-12 amino acids because the ends of the groove are occluded (Bjorkman *et al.* 1987a,b; Saper *et al.* 1991; Fig.1).

Mhc class II molecules are also transmembrane heterodimers of an α chain and a ß chain (both approximately M_r 30,000), each consisting of two extracellular domains, a transmembrane region and a cytoplasmic tail. The membrane proximal α_2 and β_2 domains have a tertiary structure resembling an Ig-fold, whereas the membrane distal α_1 and β_1 domains combine to form a peptide binding groove, very similar to that of the *Mhc* class I molecule. However, the class II groove allows peptides to extend beyond the ends of the cleft, so that it can accomodate peptides varying from 12 to 25 amino acids in length (Brown *et al.* 1993).

The tertiary structure of the two classes of *Mhc* proteins, as elucidated by X-ray crystallography, is highly similar. The two classes differ in their function and, as a consequence, in their tissue distribution. *Mhc* class I molecules mainly present intracellularly derived peptides, including self peptides and viral peptides, to CD8-positive T cells. Activation of these cells elicits a cytotoxic response. The role of class I molecules, therefore, is to mirror the internal affairs of the cell they are located on, and consequently, all nucleated cells in mammals carry class I molecules. *Mhc* class II molecules, on the other hand, mainly present extracellularly derived peptides, for example from internalized and degraded bacteria, to CD4-positive T-helper cells. When activated, these cells provide help and guidance to B cells which results in a

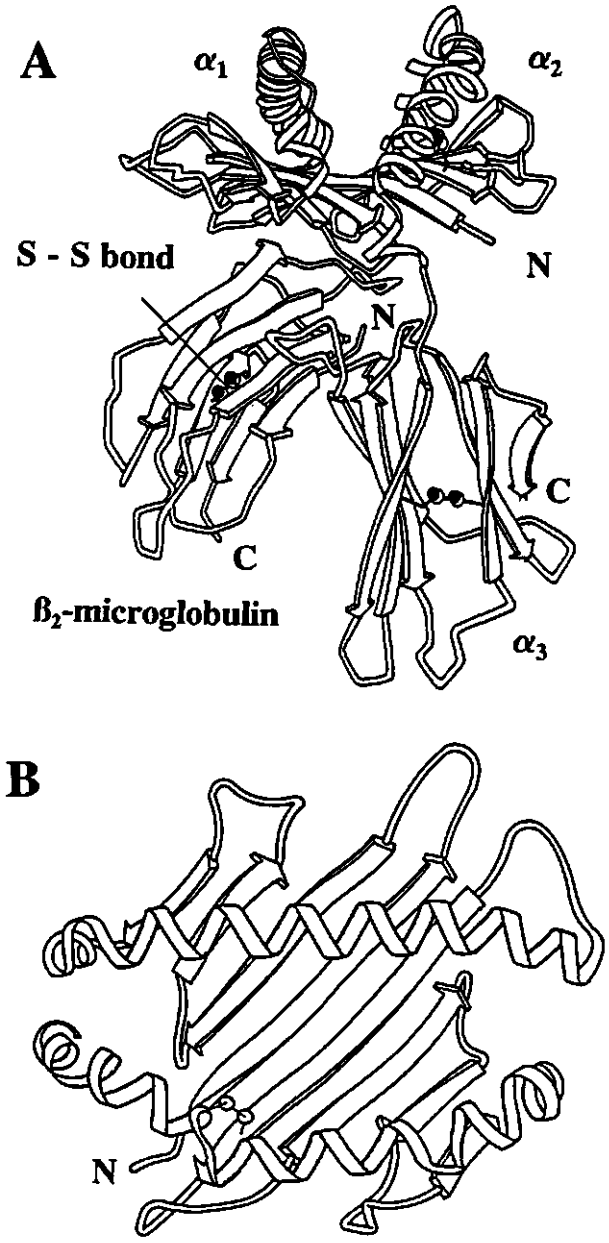


Figure 1. Schematic representation of the three dimensional structure of an *Mhc* class I molecule, **A.** as seen from the side, and **B.** from the top. (Modified after Austyn and Wood 1993)

humoral response. *Mhc* class II molecules are mainly present on antigen presenting cells of the immune system, such as B lymphocytes, dendritic cells and macrophages (Klein 1986).

The human major histocompatibility complex

The major histocompatibility complexes of many mammalian species have been studied, and although certain differences exist between them, they are sufficiently similar to limit ourselves to describe only the one that attracted the most attention, the human *Mhc*. Other mammalian *Mhc*'s are reviewed by Klein (1986) and Trowsdale (1995).

Mhc molecules derive their name from the fact that the genes encoding them are clustered together in a single genetic region, which was first implicated to determine strong allograft rejection. Hence, the name major histocompatibility complex was assigned to this cluster of genes. The human *Mhc*, or *HLA* (Human Leukocyte Antigen) complex, spans approximately 4 Mb of chromosome 6, and it is generally divided into three regions, in order of physical location from centromeric to telomeric: class II, class III and class I (Fig. 2).

The class II region of humans contains several polymorphic *Mhc* class II loci (*DP*, *DQ*, *DR*), each composed of one or more *A* (α chain encoding) and *B* (ß chain encoding) genes. The *A* and *B* genes are not necessarily present in a one to one ratio, so that the product of a single *A* gene may combine with any of several *B* gene products (*e.g.* the *DR* locus, Fig. 2). Some of the genes are pseudogenes, and for the products of some loci, *e.g. DNA* and *DOB*, although related in sequence to *DR*, *DQ* and *DP*, it is not yet clear whether they encode genuine antigen presenting molecules. Interspersed between the class II genes are a number of genes encoding molecules that are involved in processing antigens. The latter include the *DM* genes that encode molecules involved in loading peptides onto class II molecules, the TAP transporter loci, whose products are involved in transporting endogenous peptide into the endoplasmatic reticulum, and the LMP2 and LMP7 genes, encoding proteasome components, which are involved in protein cleavage (reviewed by Trowsdale 1995).

The class III region does not contain *Mhc* genes proper. It encompasses a variety of tightly clustered genes, some of which encode products involved in the immune system, *e.g.* complement components and TNF, but many of which encode proteins, such as collagen, with functions that so far seem to bear no relationship to the immune response proper.

The class I region contains several loci, including the polymorphic *HLA-A*, *-B* and *-C*, which encode the classical antigen presenting *Mhc* class I α chains. In addition, numerous so-called non-classical *Mhc* genes are present that are distantly related in sequence to the polymorphic *A*, *B* and *C* loci, but which are expressed in an aberrant manner, and which are invariably oligomorphic (Klein and O'hUigin, 1994). Many of these non-classical sequences are in fact pseudogenes. Those that are expressed do in some cases present antigenic peptides.

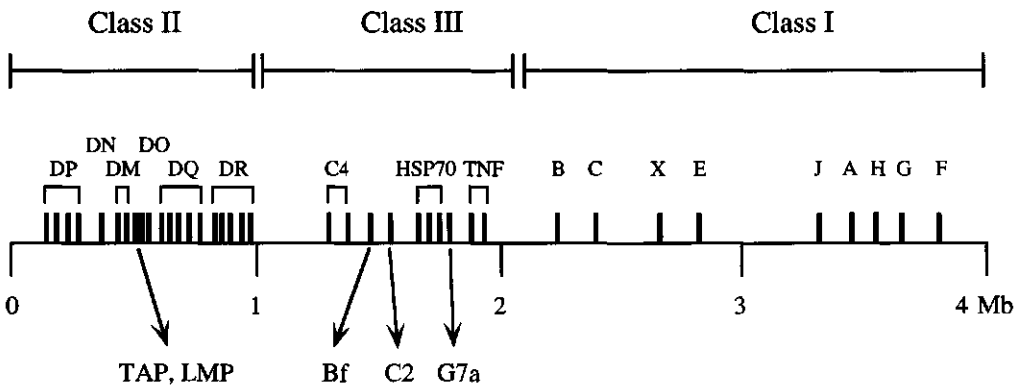


Figure 2. Genetic organization of the *HLA* complex. (Modified after Salter-Cid and Flajnik 1995)

These peptides, however, generally are of a different nature than those of classical class I molecules, resulting in different activation patterns (reviewed by Shawar *et al.* 1994).

Comparison of the genetic organization of the *HLA* region with that of the mouse *Mhc*, the *H-2* complex, showes that the class II and III regions of these two species are similarly organized, and to such an extent that orthologous relationships can be found. In contrast, the class I region differs greatly in the number and organization of genes, and there are no orthologous relationships between human and mouse class I genes. This suggests that the class I region is less stable than the class II and III regions, and more prone to recombinations, duplications and deletions (Klein *et al.* 1993a). As a result, the class I genes of different species have different origins, and seldom orthologous relationships are present.

Mhc-related genes outside of the *Mhc*

A number of genes with sequence similarity to *Mhc* class I genes are found outside the *Mhc* proper, even on different chromosomes. A group of five *Mhc* class I related genes, encoding the non-polymorphic CD1 antigens (CD1a, b, c, d, e), are located on chromosome 1. CD1 molecules are expressed in a restricted group of tissues, and it has been speculated that the CD1 molecules are involved in antigen presentation. Recently, CD1b has been found to present lipid antigens derived from the cell wall of *Mycobacterium tuberculosis* and *M. leprae* to T cells (Beckman *et al.* 1994; Sieling *et al.* 1995). In addition, CD1d appears to be able to present peptides to T cells (Castaño *et al.* 1995).

Another class I related gene outside the *Mhc* is that encoding an Fc receptor expressed in human placenta, which is related to the Fc receptor expressed in the gut of new-born rats (FcRn). Both molecules have been implied to take part in the process of the delivery of maternal IgG to the immunologically immature human fetus and newborn rat, respectively. Interestingly, the three-dimensional structure of the rat FcRn, as determined by X-ray crystallography, shows overall similarity to the structure of *Mhc* class I molecules, with two exceptions: (i) the groove of the FcRn molecule is closed due to a different structural position of the α helices, and can not accomodate peptides (Burmeister *et al.* 1994a), and (ii) the FcRn binds the Fc portion of IgG in a manner distinct from the interactions of *Mhc* molecules with peptides or CD8 (Burmeister *et al.* 1994b).

Other class I related genes are found for which no function has as yet been found. These include the gene encoding Zn-binding α_2-glycoprotein (Araki *et al.* 1988), and a recently isolated gene, MR1, which, most interestingly, was mapped to chromosome 1: the same chromosome to which also CD1 and other members of the immunoglobulin superfamily have been mapped (Hashimoto *et al.* 1995).

It seems clear that some class I-like molecules have been enrolled to perform various antigen presenting tasks, in addition to other functions. It is most likely, that many other class I related genes are present in the genome, but finding these genes depends for a large part on chance, or results of the human genome project. Judging from the presence of these genes in different species (for example both CD1 and FcRn homologues have been identified in humans and rodents), it is apparent that diversification of class I-like genes has been happening since early mammalian evolution, and insight into the structural and functional relationships of these genes will benefit from studies on other vertebrates.

Polymorphism

A remarkable feature of *Mhc* molecules is their polymorphism. In a population many alleles of a single *Mhc* gene are present, with large sequence variability between the alleles (Klein and Figueroa 1986). Most of the diversity between alleles is found in those amino acids whose side chains point into the cleft to interact with the peptide (Bjorkman and Parham 1990). At the DNA level, Hughes and Nei (1988, 1989) showed that the ratio of replacement over silent nucleotide substitutions (d_n/d_s) is significantly higher at positions involved in peptide binding than at other positions in the protein. This apparently is the result of positive selection for variability. The selective force driving the evolution of polymorphism presumably is the advantage from being able, as a population, to present and respond to various peptides from many pathogens, leading to increased fitness. However, other selective forces have also been proposed, including reproductive mechanisms like mating preferences, their function being

either to enhance immunocompetence (by increasing *Mhc* heterozygosity), or to avoid inbreeding (Potts *et al.* 1994).

Much of the *Mhc* polymorphism of a species is derived from the ancestral population, *i.e.*, polymorphism is transspecifically inherited, a phenomenon described in the trans-species hypothesis (Klein 1987). In general, after speciation, only minor modifications accumulate in already existing allelic lineages (Lawlor *et al.* 1988; Mayer *et al.* 1988). The enormous diversity observed thus can be attributed mainly to the age of allelic lineages, rather than to a high mutation rate or a high intensity of natural selection. In fact, *Mhc* genes are found to mutate at rather low rates (Klein *et al.* 1993b), and data from Satta and co-workers (1994) indicate that also the intensity of natural selection is low.

The *Mhc* of non-mammalian vertebrates

Next to mammalian species, *Mhc* genes and/or gene products have been identified in birds, reptiles, amphibia, and recently fish. However, only in the chicken and the clawed toad, *Xenopus*, evidence has been obtained for the presence of an *Mhc*, in the sense of a genetic region encoding class I and class II molecules, directing the specificity of graft rejection and other phenomena dependent on T-cell recognition.

The locus controlling these traits in chicken, named the *B* complex, has been mapped to a microchromosome, and it has been found to contain two to four class I (*B-F*) and two class II (*B-L*) genes. The only, and non-polymorphic, class II *A* gene identified in chicken, has been mapped at some distance from these genes (Kaufman *et al.* 1995). Besides the *B*-complex genes, the segregation of which correlates with serological typing, a second group of *Mhc* genes has been found not to be linked to the *B* complex. This so-called *Rfp-Y* system contains two class I and two class II *B* genes, the latter of which appear to be less polymorphic, and less strongly transcribed than the class II *B* genes in the *B* complex (Miller *et al.* 1994; Zoorob *et al.* 1993).

Exclusive for the chicken is the presence of another highly polymorphic multigene family, named *B-G*, which is unrelated to *Mhc* genes. The *B-G* genes are strongly linked to the *B* complex, and one member was even found on the same cosmid cluster as the *B* complex. The proteins encoded by these genes, so-called *B-G* antigens, are cell surface disulfide-linked heterodimers, which presumably play a role in the immune system, but whose exact function is still unknown (Goto *et al.* 1988; Kaufman and Salomonsen 1992).

The *B-F* and *B-L* genes contain very short introns, on average 100 bp, and intergenic distances are also very short. Both phenomena could be due to the location of the *B* complex on a microchromosome. In addition, no evidence has so far been obtained for the presence in the *B* complex of homologues of the intervening genes of the *HLA* class III region, except perhaps

the chicken homologue of the human *HLA*-linked G9a (BAT8) gene (Spike and Lamont 1995). In combination with the limited number of *B-F* and *B-L* genes, this makes the chicken *Mhc* compact and simple. This has prompted the hypothesis by Kaufman and co-workers (1995) that the *B* complex may represent a 'minimal essential *Mhc*', containing only the minimal number of genes that is absolutely essential for the functioning of a complex integrated immune system.

Descending the evolutionary ladder towards the ectothermic classes, the reptilia are the first encountered. In comparison with the knowledge of the *Mhc* in chicken, however, very little is known of the *Mhc* in reptiles. *Mhc* sequences have been isolated from two species, the Northern water snake (*Nerodia sipedon*) and the Amieva lizard (*Amieva amieva*) (Grossberger and Parham 1992). In addition, molecules that are structurally very similar to known class I or class II molecules have been immunoprecipitated from snake, caiman and turtle. However, since there is still controversy about the strength of T-cell related responses in reptiles, it is as yet unclear what role the *Mhc* molecules play in reptiles, and neither is it clear whether the genes encoding these molecules reside in an *Mhc* proper (Kaufman *et al.* 1990a,b).

The best-studied ectothermic vertebrate in terms of its *Mhc* is the anuran amphibian *Xenopus laevis*. The *Xenopus Mhc* has been defined at the biochemical and functional levels (reviewed by Kaufman *et al.* 1990a), as well as at the molecular-genetic level. Three class II *B* genes, and a single polymorphic class I gene are expressed in *Xenopus*, and the segregation of these genes correlates perfectly with serological typing (Shum *et al.* 1993; Kobari *et al.* 1995; Sato *et al.* 1993). Similar to the situation in mammals, several class III genes have been found to be linked to the frog *Mhc* genes: two to three *HSP70* genes, two copies of a complement factor B (Bf) encoding gene, and probably also the gene encoding complement component C4 (Salter-Cid *et al.* 1994; Kato *et al.* 1994, 1995; Nakamura *et al.* 1986).
Some of the major interest in the frog *Mhc* stems from the natural occurrence of allopolyploid *Xenopus* species. *Xenopus laevis* itself is a tetraploid, and although it still expresses many duplicated loci, it seems to have functionally diploidized its *Mhc*, apparently by deletion, or perhaps by silencing one of the diploid sets (Shum *et al.* 1993). Similar functional diploidization is observed in other naturally occurring polyploid *Xenopus* species, like *X. vestitus* (8n), but not in laboratory-made polyploids, which express *Mhc* genes of each parental species co-dominantly (Du Pasquier *et al.* 1977). This indicates that the diploidization process takes place over evolutionary time, probably as the result of a selection pressure working against an increase in the number of expressed *Mhc* genes.

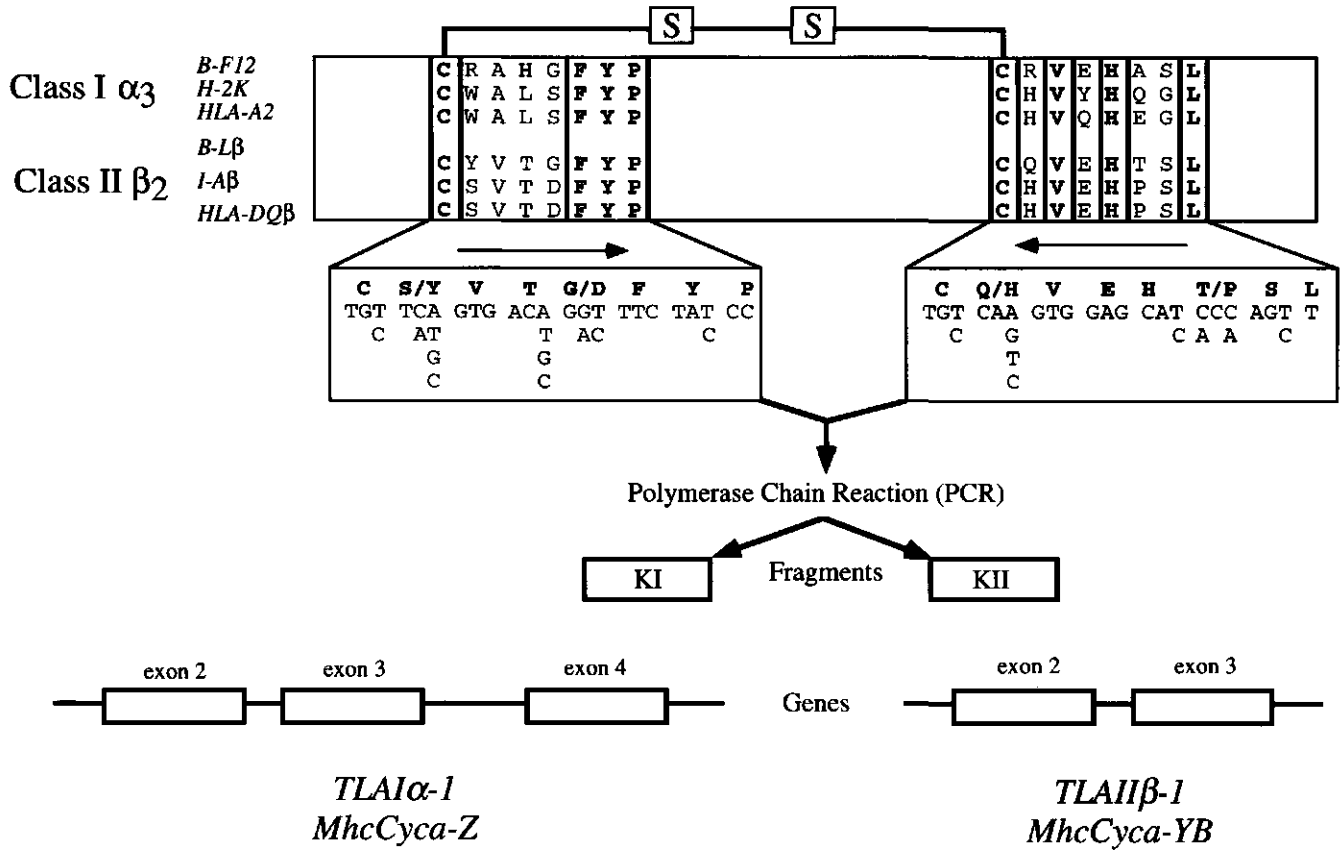


Figure 3. Schematic representation of the strategy applied by Hashimoto and co-workes (1990) in the isolation of carp *Mhc* class I (*TLAIα-1*) and class II (*TLAIIβ-1*) genes.

Another interesting observation made in the studies on *Xenopus Mhc* is the presence of a large family of non-classical class I genes, mapped as a single linkage group to a different chromosome than the classical genes. This non-classical gene family (named XNC) is composed of at least nine subfamilies, and all of these are expressed at the RNA level. Similar to the characteristics of mammalian non-classical genes, however, the XNC genes are expressed at much lower levels than the classical molecules, and show a different tissue distribution. In addition, polymorphism of these genes is minimal, and the amino acid sequences of these genes lack some of the conserved residues involved in peptide binding found in classical class I molecules (Flajnik *et al.* 1993; Salter-Cid and Flajnik 1995).

The *Mhc* of fish

Fish represent the majority of all living vertebrate species; of approximately 43,000 vertebrate species over 30,000 are fish, which have been diversifying for more than 400 million years. Extant fish species can roughly be divided into three major groups: (i) the jawless fish, or Agnatha, represented by lampreys and hag-fishes; (ii) the cartilaginous fish, or Chondrichthyes, including the Elasmobranchii (sharks and rays) and the Holocephali (*e.g.*, rabbitfishes); and (iii) the bony fish, or Osteichthyes, consisting of the lobe-finned fishes (the coelacanth, and lungfishes), and the ray-finned fishes (with primitive representatives like the sturgeons (Chondrostei) and gars (Holostei), but also including all modern bony fishes or Teleostei). Teleostean fish make up the largest group, with 20,000 representative species in an immense variety, ranging from perch, plaice and pufferfish to cod, carp and catfish, and many more.

In an immunological sense, the teleostean fish occupy an interesting position, as they are the most primitive group of species displaying acute graft rejection. For a long time however, this graft rejection and other *Mhc*-related traits, like mixed leukocyte reactivity and cell-mediated lymphocytotoxicity, provided the only evidence for the existence of an *Mhc* in teleosts. Strategies that were successful in the identification of *Mhc* molecules in other species, like the use of cross-reactive DNA probes, monoclonal antibodies or xenoantisera, all failed to identify fish *Mhc* molecules (Stet and Egberts, 1991; Kaufman *et al.* 1990a). With the advent of the polymerase chain reaction (PCR), however, times changed. Using this technique, the first fish *Mhc* genes were identified in the common carp (*Cyprinus carpio* L.) by Hashimoto and co-workers (1990). The strategy exploited the presence of two evolutionary conserved amino acid stretches around the cysteine-residues forming the disulfide bonds in the α_3 domain of class I and the $ß_2$ domain of class II molecules (Fig. 3). Highly degenerate oligonucleotides were designed, complementary in sequence to the conserved stretches, and these were used as primers in PCR amplifications on genomic DNA from carp. Fragments of the expected

size were sequenced, and among them two fragments, KI and KII, were identified that showed similarity to known *Mhc* class I and class II sequences, respectively. Screening a genomic library with these fragments then yielded two genomic clones, λTLAI-1, which contained part of a putative class I gene (*TLAIα-1*), and λTLAII-1, which contained part of a putative class II gene (*TLAIIβ-1*). Similarity with other *Mhc* sequences was low, 30-33% identity for the exon encoding the class I α_3 domain, and 30-40% for the exon encoding the class II β_2 domain. In light of the general failure to isolate these molecules using cross-reactive reagents, however, this did not come as a surprise.

Nevertheless, the carp genes appeared to be enigmatic, as was experienced later. First, the genes reported were not complete. Only the exons encoding the class I α_1, α_2 and α_3 domains and class II β_1 and β_2 domains were identified, and although major stretches of the genomic clones were sequenced, no exons encoding the leader, transmembrane or cytoplasmic domains were identified. Surprisingly, the authors did not report on the expression of these genes, which would have been an obvious part of this study, and which would also have facilitated the identification of any remaining exons, if present.

Soon after these genes were reported, we started attempts to obtain the corresponding cDNA clones from cDNA libraries prepared from lymphoid tissues (spleen and head-kidney) of the carp. Initially, the KI and KII fragments were used to screen the libraries, however, no true positive clones were found. This could have been due to the small size of the KI and KII probes (approximately 190 bp), and therefore we subsequently tried to obtain cDNA sequences through anchored PCR on the libraries. This, however, also failed, and finally we had to conclude that the *TLAIα-1* and *TLAIIβ-1* genes were probably not expressed in these lymphoid tissues, although this was a reasonable assumption. Nevertheless, three years later Okamura and co-workers (1993) reported the finding of cDNA fragments spanning all three extracellular domains of the *TLAIα-1* genes (which at that time were already renamed to *Cyca-Z*), from carp kidney. And in addition, we later obtained a truncated cDNA fragment encoding part of the *TLAIα-1* α_3 domain from a carp thymus cDNA library by anchored PCR. However, no complete *TLAIα-1* or *TLAIIβ-1* cDNA sequences, encoding all of the domains that constitute a genuine *Mhc* molecule, have as yet been found.

Aim and outline of this thesis

The carp and other cyprinid fish species constitute the majority of fish grown in aquaculture systems, resulting in a contribution of 6.7 million tons to the total of 9.5 million tons of fish cultured worldwide in 1992 (FAO 1995). However, the high stocking densities increase the risk of infections and subsequent disease outbreaks, reducing the yields of production. To gain better control over disease it is essential to acquire an understanding of

the immune response of cyprinid fish. *Mhc* molecules play a leading role in the immune response, due to their central position in the process of antigen presentation and T-cell activation. A thorough knowledge of the structure and functioning of the carp *Mhc* can therefore be of great value to the understanding of disease control by vaccination, and selection for disease resistance. In addition, the study of the *Mhc* in carp may give important insights into the evolution of these genes, which helps in understanding their basic functions, and the mechanisms which have led to the evolution of a complex, integrated immune system as it is observed in higher vertebrates.

The study presented in this thesis aims at the identification of the *Mhc* genes of the common carp (*Cyprinus carpio* L.). At the starting point of this study, only the *TLAIα-1* (*Cyca-Z*) and *TLAIIβ-1* (*Cyca-YB*) genes described by Hashimoto and co-workers were known, and although expression of these genes could not be detected, we decided to study their presence in laboratory strains of carp, using Restriction Fragment Length Polymorphism (RFLP) analysis (**Chapter 2**). The existence of class I genes in carp suggested that a β_2-microglobulin molecule could be present as well, and, to verify previous biochemical data that this indeed was the case (Shalev *et al.* 1984), we decided to isolate the corresponding *B2m* cDNA sequences (**Chapter 3**).

Meanwhile, evidence was accumulating that the *Cyca-Z* and *Cyca-YB* sequences probably do not represent *bona fide Mhc* genes, and based on the hypothesis that perhaps other class I sequences could be found in carp, we used a heterologous zebrafish class I probe to isolate a novel class I gene, *Cyca-UA*, from a gynogenetic carp clone A410. To ascertain that this gene was functional, we studied its expression at the protein level, and its polymorphism (**Chapter 4**).

During a study into the expression of class I genes, a serendipitous finding of a class I-like fragment, *Cyca*-TC16, which was completely different in sequence from *Cyca-Z* and *Cyca-UA*, made it clear that multiple distinct class I lineages were present in the carp (**Chapter 5**). The only member of the carp *Mhc* genes not yet identified, was the class II *A* gene, and we set out to isolate this gene from the gynogenetic carp clone A410. In addition, to type this carp clone for its *Mhc* genes, we set out to analyse the expressed class II *B* sequences, and performed preliminary analyses of genetic linkage of these genes (**Chapter 6**). Finally, in the last chapter, the pieces of data obtained in each chapter are combined and discussed, and future directions are given (**Chapter 7**).

References

Araki, T., Gejyo, F., Takagaki, K., Haupt, H., Schwick, H.G., Bürgi, W., Marti, T., Schaller, J., Rickli, E., Brossmer, R., Atkinson, P.H., Putnam, F.W. and Schmid, K. 1988. Complete amino acid sequence of human plasma Zn-α_2-glycoprotein and its homology to histocompatibility antigens. *Proc. Natl. Acad. Sci. USA 85:* 679-683

Austyn, J.M. and Wood, K.J. 1993. Principles of cellular and molecular immunology. Oxford University Press, New York

Beckman, E.M., Porcelli, S.A., Morita, C.T., Behar, S.M., Furlong, S.T. and Brenner, M.B. 1994. Recognition of a lipid antigen by CD1-restricted $\alpha\beta^+$ T cells. *Nature 372:* 691-694

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

Bjorkman, P.J., Saper, M.A., Samraoui, B., Bennett, W.S., Strominger, J.L. and Wiley, D.C. 1987a. Structure of the human class I histocompatibility antigen, HLA-A2. *Nature 329:* 506-512

Bjorkman, P.J., Saper, M.A., Samraoui, B., Bennett, W.S., Strominger, J.L. and Wiley, D.C. 1987b. The foreign antigen binding site and T cell recognition regions of class I histocompatibility antigens. *Nature 329:* 512-518

Brown, J.H., Jardetzky, T.S., Gorga, J.C., Stern, L.J., Urban, R.G., Strominger, J.L. and Wiley, D.C. 1993. Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. *Nature 364:* 33-39

Burmeister, W.P., Gastinel, L.N., Simister, N.E., Blum, M.L. and Bjorkman, P.J. 1994a. Crystal structure at 2.2 Å resolution of the MHC-related neonatal Fc receptor. *Nature 372:* 336-343

Burmeister, W.P., Huber, A.H. and Bjorkman, P.J. 1994. Crystal structure of the complex of rat neonatal Fc receptor with Fc. *Nature 372:* 379-383

Castaño, A.R., Tangri, S., Miller, J.E.W., Holcombe, H.R., Jackson, M.R., Huse, W.D., Kronenberg, M. and Peterson, P.A. 1995. Peptide binding and presentation by mouse CD1. *Science 269:* 223-226

Du Pasquier, L., Miggiano, V.C., Kobel, H.R. and Fischberg, M. 1977. The genetic control of histocompatibility reactions in natural and laboratory-made polyploid individuals of the clawed toad *Xenopus*. *Immunogenetics 5:* 129-141

FAO. 1995. Review of the state of world fishery resources: aquaculture. *FAO Fish. Circ.* No. 886., Rome

Flajnik, M.F., Kasahara, M., Shum, B.P., Salter-Cid, L., Taylor, E. and Du Pasquier, L. 1993. A novel type of class I gene organization in vertebrates: a large family of non-MHC-linked class I genes is expressed at the RNA level in the amphibian *Xenopus*. *EMBO J. 12:* 4385-4396

Goto, R., Miyada, C.G., Young, S., Wallace, R.B., Abplanalp, H., Bloom S.E., Briles, W.E. and Miller, M.M. 1998. Isolation of a cDNA clone from the B-G subregion of the chicken major histocompatibility (*B*) complex. *Immunogenetics 27:* 102-109

Grossberger, D. and Parham, P. 1992. Reptilian class I major histocompatibility complex genes reveal conserved elements in class I structure. *Immunogenetics 36:* 166-174

Hashimoto, K., Hirai, M. and Kurosawa, Y. 1995. A gene outside the human MHC related to classical HLA class I genes. *Science 269:* 693-695

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

Hughes, A.L. and Nei, M. 1988. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature 335:* 167-170

Hughes, A.L. and Nei, M. 1989. Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *Proc. Natl. Acad. Sci. USA 86:* 958-962

Kato, Y., Salter-Cid, L., Flajnik, M.F., Kasahara, M., Namikawa, C., Sasaki, M., and Nonaka, M. 1994. Isolation of the *Xenopus* complement factor B complementary DNA and linkage of

the gene to the frog MHC. *J. Immunol. 153:* 4546-4554

Kato, Y., Salter-Cid, L., Flajnik, M.F., Kasahara, M., Namikawa, C., Sasaki, M., and Nonaka, M. 1995. Duplication of the MHC-linked *Xenopus* complement factor B gene. *Immunogenetics 42:* 196-203

Kaufman, J. and Salomonsen, J. 1992. B-G: we know what it is, but what does it do? *Immunology Today 13:* 1-3

Kaufman, J., Skjoedt, K. and Salomonsen, J. 1990a. The MHC molecules of nonmammalian vertebrates. *Immunol. Rev. 113:* 83-117

Kaufman, J., Skjoedt, K., Salomonsen, J., Simonsen, M., Du Pasquier, L., Parisot, R. and Riegert, P. 1990b. MHC-like molecules in some nonmammalian vertebrates can be detected by some cross-reactive xenoantisera. *J. Immunol. 144:* 2258-2272

Kaufman, J., Völk, H. and Wallny, H.J. 1995. A "minimal essential Mhc" and an "unrecognized Mhc": two extremes in selection for polymorphism. *Immunol. Rev. 143:* 63-88

Klein, J. 1986. Natural history of the major histocompatibility complex. John Wiley and Sons, New York

Klein, J. 1987. Origin of major histompatibility complex polymorphism: the trans-species hypothesis. *Hum. Immunol. 19:* 155-162

Klein, J. and Figueroa, F. 1986. Evolution of the major histocompatibility complex. *CRC Crit. Rev. Immunol. 6:* 295-386

Klein, J. and O'hUigin, C. 1994. The conundrum of nonclassical major histocompatibility complex genes. *Proc. Natl. Acad. Sci. USA 91:* 6251-6252

Klein, J., Ono, H., Klein, D. and O'hUigin, C. 1993a. The accordion model of *Mhc* evolution. In: *Progress in Immunology* (ed. by J. Gergely and G. Petranyi), p.137. Springer, Heidelberg

Klein, J., Satta, Y. and O'hUigin, C. 1993b. The molecular descent of the major histocompatibility complex. *Annu. Rev. Immunol. 11:* 269-295

Kobari, F., Sato, K., Shum, B.P., Tochinai, S., Katagiri, M., Ishibashi, T., Du Pasquier, L, Flajnik, M.F. and Kasahara, M. 1995. Exon-intron organization of *Xenopus* MHC class II ß chain genes. *Immunogenetics 42:* 376-385

Lawlor, D.A., Ward, F.E., Ennis, P.D., Jackson, A.P and Parham, P. 1988. HLA-A,B polymorphisms predate the divergence of humans and chimpanzees. *Nature 335:* 268-271

Mayer, W.E., Jonker, D., Klein, D., Ivanyi, P., Van Seventer, G. and Klein, J. 1988. Nucleotide sequence of chimpanzee MHC class I alleles: evidence for trans-species mode of evolution. *EMBO J. 7:* 2765-2774

Miller, M.M., Goto, R., Bernot, A., Zoorob, R., Auffray, C., Bumstead, N. and Briles, W.E. 1994. Two *Mhc* class I and two *Mhc* class II genes map to the chicken *Rfp-Y* system outside the *B* complex. *Proc. Natl. Acad. Sci. USA 91:* 4397-4401

Nakamura, T., Sekizawa, A., Fujii, T. and Katagiri, C. 1986. Co-segregation of the polymorphic C4 with the MHC in the frog, *Xenopus laevis*. *Immunogenetics 23:* 181-186

Potts, W.K., Manning, C.J. and Wakeland, E.K. 1994. The role of infectious disease, inbreeding and mating preferences in maintaining MHC genetic diversity: an experimental test. *Phil. Trans. R. Soc. Lond.* B. *346:* 369-378

Salter-Cid, L. and Flajnik, M.F. 1995. Evolution and developmental regulation of the major histocompatibility complex. *CRC Crit. Rev. Immunol. 15:* 31-75

Salter-Cid, L., Kasahara, M. and Flajnik, M.F. 1994. *Hsp 70* genes are linked to the *Xenopus* major histocompatibility complex. *Immunogenetics 39:* 1-7

Saper, M.A., Bjorkman, P.J. and Wiley, D.C. 1991. Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 Å resolution. *J. Mol. Biol. 219:* 277-319

Sato, K., Flajnik, M.F., Du Pasquier, L., Katagiri, M. and Kasahara, M. 1993. Evolution of the MHC: isolation of class II ß-chain cDNA clones from the amphibian *Xenopus laevis*. *J. Immunol. 150:* 2831-2843

Satta, Y., O'hUigin, C., Takahata, N. and Klein, J. 1994. Intensity of natural selection at the major histocompatibility complex loci. *Proc. Natl. Acad. Sci. USA 91:* 7184-7188

Shalev, A., Caspi, R.R., Lögdberg, L., Björck, L., Segal, S. and Avtalion, R.R. 1984. Specific activity of carp antisera against $ß_2$-microglobulin ($ß_2$m) and evidence for a $ß_2$m homologue in carp (*Cyprinus carpio*). *Dev. Comp. Immunol. 8:* 639-648

Shawar, S.M., Vyas, J.M., Rodgers, J.R., and Rich, R.R. 1994. Antigen presentation by major histocompatibility complex class I-B molecules. *Annu. Rev. Immunol. 12:* 839-880

Shum, B.P, Avila, D., Du Pasquier, L., Kasahara, M. and Flajnik, M.F. 1993. Isolation of a classical MHC class I cDNA from an amphibian. Evidence for only one class I locus in the *Xenopus* MHC. *J. Immunol. 151:* 5376-5386

Sieling, P.A., Chatterjee, D., Porcelli, S.A., Prigozy, T.I., Mazzaccaro, R.J., Soriano, T., Bloom, B.R., Brenner, M.B., Kronenberg, M., Brennan, P.J. and Modlin, R.L. 1995. CD1-restricted T cell recognition of microbial lipoglycan antigens. *Science 269:* 227-230

Spike, C.A. and Lamont, S.J. 1995. Genetic analysis of three loci homologous to human G9a: evidence for linkage of a class III gene with the chicken MHC. *Anim. Gen. 26:* 185-187

Stet, R.J.M. and Egberts, E. 1991. The histocompatibility system in teleostean fishes: from multiple histocompatibility loci to a major histocompatibility complex. *Fish Shelfish Immunol. 1:* 1-16

Trowsdale, J. 1995. "Both man & bird & beast": comparative organization of MHC genes. *Immunogenetics 41:* 1-17

Zoorob, R., Bernot, A., Renoir, D.M., Choukri, F. and Auffray, C. 1993. Chicken major histocompatibility complex class II B genes: analysis of interallelic and interlocus sequence variance. *Eur. J. Immunol. 23:* 1139-1145

Chapter 2

Polymorphism and estimation of the number of *MhcCyca* class I and class II genes in laboratory strains of the common carp (*Cyprinus carpio* L.)

René J.M. Stet[1], Saskia H.M. van Erp[1],
Trudi Hermsen[1], Holger A. Sültmann[2] and Egbert Egberts[1]

[1]Department of Experimental Animal Morphology and Cell Biology, P.O. Box 338, 6700 AH Wageningen, Wageningen Agricultural University, The Netherlands
[2]Max-Planck Institut für Biologie, Abteilung Immungenetik, D7400 Tübingen, Germany

Reprinted from: **Developmental and Comparative Immunology 17, 141-156 (1993)**

Summary

Restriction fragment length polymorphisms (RFLPs) have been identified in the *Mhc* of the carp (*MhcCyca*) using class I (*Cyca-Z*) and class II *B* (*Cyca-YB*) specific probes. The K1-5 and K2-1 probes were obtained as polymerase chain reaction products after amplification of genomic DNA from a European carp using primers deduced from genomic sequences, and were shown to be 90% and 80% similar to *Cyca-Z* exon 3 and *Cyca-YB* exon 2 sequences, respectively. Six carp strains of different geographical origins and genomic status were studied. In homozygous gynogenetic carp strains the class I probe K1-5 hybridized to 9-12 fragments, whereas the class II *B* probe K2-1 hybridized to 3-5 fragments. Thus, the *Cyca* consists of multiple class I and class II *B* genes. The level of polymorphism of the *Cyca* genes of the strains studied was calculated as the percentage of polymorphic fragments among the total number of fragments observed, and was shown to be 70% for class I and 40-66% for class II *B* genes. In addition, a possible correlation was investigated between a serologically defined locus *K*, which was demonstrated previously to incorporate class I-like characteristics, and molecular genotyping using the class I probe. Two gynogenetic families, which were serologically typed K1 and K2 homozygous, also differed in their RFLPs using a class I probe. This would suggest that the *K* locus is part of the *Cyca* complex.

Key words: *MhcCyca*, *Cyprinus carpio* L., restriction fragment length polymorphism, serology, gynogenesis.

The nucleotide sequences reported in this paper have been submitted to the EMBL/GenBank nucleotide sequence databases and have been assigned the accession numbers S62610 and S62611.

Introduction

The major histocompatibility complex (*Mhc*) genes and gene products have been studied extensively in a limited number of mammalian species (reviewed in Klein 1986). Of the approximately 39,000 remaining non-mammalian vertebrate species, so far only one bird (*Gallus gallus var. domesticus*) and one anuran amphibian (*Xenopus laevis*) have been shown to possess a similar complex based upon the identification and characterization of its genes or gene products (Kaufman *et al.* 1990a). The initial identification of the *Mhc* of the latter species resulted from cellular, serological and biochemical studies, but the corresponding genomic and cDNA sequences have only recently been reported (Bourlet *et al.* 1988; Guillemot *et al.* 1988; Flajnik *et al.* 1991). With respect to the chicken *Mhc*, or *B* complex, the class I (*B-F*) and especially the class II (*B-L*) sequences show a reasonable degree of homology to the corresponding mammalian sequences. Therefore, the *B-L* genes could be identified using heterologous probes, whereas *B-F* genes were isolated by screening of cosmid clones with a *B-L* gene. Such strategies, however, have remained without success in lower vertebrates such as amphibia and Osteichthyes

(bony fishes), possibly due to too much divergence between the classes of Aves and mammalia, and these comparatively old phylogenetic classes. Indeed, a *Xenopus* class I sequence, isolated by screening a cDNA library using monoclonal antibodies and polyclonal antibodies raised against immunoaffinity-purified class I molecules, shows only a low degree of homology with mammalian *Mhc* sequences (Flajnik *et al.* 1991).

Since serological reagents for *bona fide Mhc* gene products are not available yet in fish, another strategy for the isolation of *Mhc* genes was adopted, based on the polymerase chain reaction (PCR). From comparisons of the amino acid (aa) sequences of class I α_3 and class II β_2 domains from man, mouse and chicken, highly degenerate oligonucleotide primers were constructed, based on the sequences surrounding the conserved cysteine residues involved in the intramolecular disulfide bond of these domains (Hashimoto *et al.* 1990). After amplification of genomic carp DNA, two different fragments KI and KII were cloned. Using the KI and KII fragments as probes, two different clones could be identified from a genomic library, one containing a partial class I gene sequence (*TLAIα-1*), and the other a partial class II *B* gene sequence (*TLAIIß-1*). The KI and KII fragments correspond to the class I exon 3 and class II *B* exon 2 of the putative carp *Mhc*.

In order to reduce confusion in future nomenclature of the *Mhc* of different teleostean fishes, the gene designation, as used by Hashimoto and co-workers (1990), has been altered in compliance with the proposal by Klein and co-workers (1990). Thus, the carp (*Cyprinus carpio* L.) *Mhc* will be symbolized by *MhcCyca*, while the class I gene *TLAIα-1* designation is changed to *Cyca-Z* and the *TLAIIß-1* class II *B* gene to *Cyca-YB*. This new nomenclature was cleared through the register with the consent of Hashimoto and co-workers.

The availability of the KI and KII probes recognizing the most conserved exons of the two classes of *Cyca* genes has prompted the study of a number of characteristics of the *Cyca*, such as multiplicity and polymorphism of its genes. Both traits are important hallmarks of the *Mhc* in higher vertebrates, and have revealed important implications for the erection of hypotheses concerning the evolution of the *Mhc* (Figueroa and Klein 1986; Klein and Figueroa 1986; Klein 1987). In this study different class I and class II *B* probes have been used to estimate the number of class I and class II *B* genes in the *Cyca*. Also, the polymorphism of these genes has been investigated by restriction fragment length polymorphism (RFLP) analyses of a limited number of carp lines with different geographical origins. Although most of these carp lines are probably derived from Denube carp, and thus may have a common ancestral origin dating back to the 16th century (Berka 1985), it should be stressed that these cultured carp strains are hybrids of different carp races, European feral, wild, and even Chinese carp. In addition, these hybrids have also undergone deliberate selection for growth and natural selection for disease resistance in adaptation to local aquacultural circumstances in the past

100 years. This has probably resulted in largely unknown bottle necks in all carp populations. In this study, therefore, no attempts have been made to study the polymorphism of the *Cyca* in terms of gene frequencies, but instead the polymorphism of this complex in a number of laboratory strains was investigated.

In a previous study a major histocompatibility locus *K* was identified by alloantisera produced within a gynogenetic carp family (Kaastrup *et al.* 1989). This locus was shown to incorporate class I-like characteristics, based on skin transplantation data and distribution of its gene products as seen in flowcytometric analyses. Using a *Cyca* class I-specific probe, it is also investigated in this study whether the *K* locus is part of the *Cyca*, in order to establish a correlation between molecular genotyping and expressed polymorphism of the *Cyca*.

Materials and methods

Animals.

The laboratory strains of the common carp (*Cyprinus carpio* L.) used in this study were originally obtained from different geographical stocks. The A strain originated from Israel and is known as the DOR70 (Wohlfarth *et al.* 1980), the W strain was obtained from the Netherlands, the D strain from Germany, the R8 from Hungary, and the R3 from Poland. The latter two have been bred by means of brother-sister matings for five and six generations, respectively. In our laboratory, second-generation meiotic gynogynetic families were obtained, essentially as described by Nagy and co-workers (1978) and Komen and co-workers (1988). For the A strain, two second generation families A4.19 and A4.3 were produced from first-generation meiotic females A4-19 and A4-3. These carp had been previously serologically typed K1- and K2-homozygous using alloantisera defining allelic specificities of a putative major histocompatibility locus *K* (Kaastrup *et al.* 1989). For the W strain, two gynogenetic families W11.49 and W11.52 were generated from randomly selected females from a first-generation gynogenetic W11 family, without prior knowledge of serologically identified major histocompatibility *K* locus alleles within this family.

In addition to these strains, two F1 hybrid families, R3 x R8 and E20 x E6 were used. The latter F1 hybrid family was a cross between two individuals of a homozygous clone (Komen *et al.* 1991). This clone was originally produced by mitotic gynogenesis using a D x W F1 hybrid female carp. Random outbred carp (WK) with unknown family history, were obtained from the OVB, Lelystad, The Netherlands. These carp, however, showed different phenotypic characteristics, with respect to scaling and shape.

DNA isolation.

Blood samples were obtained from the caudal vein of three to six individuals from each carp family or group, under proper anesthesia. Erythrocytes were isolated by centrifugation at 400 x *g* for 10 min and washed twice in cTBS (20 mM Tris, 150 mM NaCl, 0.8 mM $MgCl_2$, 0.2 mM $CaCl_2$, pH 7.6, and adjusted to 270 mOsmol). Packed cells (50 μl) were resuspended in 500 μl cell lysis buffer (10 mM Tris/HCl pH 8.0, 1 mM $MgCl_2$, 150 mM NaCl, and 1% CHAPS) and incubated for 1 hour on ice. Nuclei were collected by centrifugation at 13,000 x *g* for 10 min. Recovered nuclei were subsequently lysed in 50 mM Tris/HCl pH 9.5, 100 mM EDTA, 1% SDS and incubated overnight (o.n.) at 50°C in the presence of proteinase-K (1 mg/ml). DNA was extracted by consecutive extractions using phenol, phenol/chloroform/iso-amyl alcohol (IAA), chloroform/IAA, and chloroform, and finally precipitated in 0.1 vol 3M sodium acetate and 2 vol 100% ethanol at -20°C. The DNA was washed in 70% ethanol, dried, and dissolved in TE (10 mM Tris/HCl pH 7.6, 1 mM EDTA).

DNA samples of the R3 x R8 F1 hybrid family were obtained from liver samples. Small samples were homogenized in TEN (100 mM Tris/HCl pH 8.0, 10 mM EDTA, 250 mM NaCl, 1% SDS) buffer, and incubated o.n. at 50°C in the presence of proteinase-K (1 mg/ml). Further isolation of high molecular weight DNA was performed as described above with the exception of an additional precipitation after the extractions using 0.6 vol 2-propanol.

DNA digestion.

DNA samples (20 μg) were digested in a large volume (400 μl) with *Pst*I, *Taq*I, or *Eco*RI (generally 5 units/μg) according to the manufacturer's (Boehringer, Mannheim, Germany) specifications in the presence of RNAse. DNA was precipitated and dissolved in TE and used directly for electrophoresis.

Electrophoresis and Southern blotting.

Digested DNA samples (10 μg/lane) were separated in 0.8 or 1.0% horizontal agarose gels in a TBE (90 mM Tris, 90 mM boric acid, 0.25 mM EDTA, pH 8.5) buffer system. Electrophoresis was carried out o.n. at 15°C in non-recirculating buffer tanks.

Following depurination in 0.25 N HCl, denaturation in 0.5 M NaOH/1.5 M NaCl, neutralization in 1 M Tris/HCl pH 7.4, 1.5 M NaCl, and equilibration in 10 x SSC (SSC; 150 mM NaCl, 150 mM Na-citrate pH 7.0), the DNA was transferred to a nylon filter (Hybond N+, Amersham, UK) by vacuum blotting in 10 x SSC at 40 mbar for 1 hour using a Vacugene XL system (Pharmacia, Uppsala, Sweden). Subsequently, nylon filters were placed on filter paper soaked in 0.4 M NaOH for 1 min, and briefly washed in 5 x SSC. Filters were air dried and either used immediately or stored at 4°C.

Probes and hybridizations.

The probes KI and KII hybridizing with *Cyca-Z* exon 3 and *Cyca-YB* exon 2, respectively, (Hashimoto *et al.* 1990) were a kind gift of Dr. Hashimoto, Toyoake, Japan. In addition, fragments comparable to these probes were obtained by PCR amplification of genomic DNA from a European carp using primers that were constructed for the amplification of class I and class II *B Cyca* genes. The primers for *Cyca* class I were 5'-d[TGTCTGGTCACTGGTTTCTACCC]-3' and 5'-d[AGGCTGCTGTGAATCACATGACA]-3' and for *Cyca* class II *B* 5'-d[TGCAGTGCCTATGACTTCTACCC]-3' and 5'-d[GAGCTGGCGTGCTCCACCACACA]-3'. PCR amplification was performed on 500 ng genomic DNA, in a mix containing 200 μM of each dNTP, 1 μM of each primer, 2.5 U Taq polymerase (Perkin Elmer, Emeryville CA, USA) and reaction buffer in a final volume of 100 μl. The mixture was subjected to a thermal cycle profile (1 min 94°C, 2 min 55°C, 2 min 72°C) for 50 cycles with an additional extension step at 72°C for 10 min. The PCR products of about 190 base pairs (bp) were cloned into the *Sma*I site of pBluescript SK- (Stratagene, La Jolla, CA) and sequenced.

The probes were excised from pBluescript using *Eco*RI and *Bam*HI restriction digestion. The probe was separated from the plasmid on agarose gel and recovered by excising the DNA from the gel, followed by centrifugation and precipitation (Heery *et al.* 1990). Radiolabeling of the probes (200 ng) was performed by multiprime labeling, essentially according to Feinberg and Vogelstein (1983). Probes were recovered by precipitation in 1 vol 5 M ammonium acetate and 6 vol ethanol at -20°C. After a brief wash in 70% ethanol, the radiolabeled probe was dried, and dissolved in TE containing 0.01% SDS.

The nylon filters were prehybridized in 5 x SSC, 5 mM EDTA, 0.1% SDS, 5 x Denhardt's, and 30% or 35% formamide for the KI (K1-5) and KII (K2-1) probes, respectively. The prehybridization mix was filter sterilized through a 0.2 μm filter and 100 μg of denatured *Escherichia coli* DNA was added. Prehybridization was carried out for 4-5 hours at 42°C, and hybridization o.n. at 42°C by adding 1 x 10^7 cpm. For both K1-5(KI) and K2-1(KII) hybridizations the filters were washed at low stringency, *i.e.*, 4 x SSC, 0.1% SDS at 42°C for 10 min, and 4 x SSC, 0.1% SDS at 45°C for 10 min. Autoradiography was performed at -70°C using Kodak X-OMAT AR imaging films with intensifying screens. Fragment size was scored using an imaging system (Cybertech CS-1, Berlin, Germany).

In a number of experiments, filters were rehybridized with another probe after removing the first probe. Filters were washed in boiling 0.5% SDS, and were allowed to cool down to room temperature. Subsequently, filters were washed twice briefly in 5 x SSC, air dried and reused for hybridization.

A.

```
K1-5      TGT CTG GTA ACT GGT TTC TAC CCC AGA GAT ATT GAG ATG AAC ATC AGA
λTLAI-1   --- --- --C --- --- --- --- --- --- --- --- --- --- --- --- ---

K1-5      CTG AAC AGA ATT AAC ATT GAG AGC CAG ATA TCT TCT GGA ATC AGA CCA
λTLAI-1   --- --- --- --- --- --- --- --- --- --- --- --A --- --- --- ---

K1-5      AAT GAT AAT GAA AGC TTT CAG CTC AGA TCC AGT GTG AAG ATC GAC AGA
λTLAI-1   --- --- G-- --- --- --- --- A-G --- --- --- --- --- --- --- ---

K1-5      AAC CAC AGA GGA TCT TAT GAC TGT CAT GTG ATT CAC AGC AGC CA
λTLAI-1   --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --
```

B.

```
K1-5      CLVTGFYPRD IEMNIRLNRI NIESQISSGI RPNDNESFQL RSSVKIDRNNH
λTLAI-1   ---------- ---------- ---------- ----D----M -----------

K1-5      RGSYDCHVIH SS
λTLAI-1   ---------- --
```

C.

```
K2-1      TGC AGT GCC TAT GAC TTC TAC CC* AAA CCC ATT AAA CTG ATG TGG ATG
λTLAII-1  --- --- --- --- --- --- --- --C --- --- --- --- --- -CA --- ---

K2-1      AGA GAT GAA ATG AAA AGT GGC AGC TGA TGT GAT GTT CAT TGA GGA GAT
λTLAII-1  --- --- --T -*A -G- G-- -A- -A- --- --- --C --C --C --- --- -C-

K2-1      GGC TAA TGG AGA CTG GTA TTA TCA AAT CCA CTC CCA CCT GGA ATA TTT
λTLAII-1  --- -G- --- --- --- --- C-- C-- --- --- --- --- --- --- --- C--

K2-1      TCC CAA ACC TGG AGA GAA GAT CTC CTG TGT GGT GGA GCA CGC CAG CTC
λTLAII-1  C-- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
```

D.

```
K2-1      CSAYDFYPKP IKLMWMRDEA KVAADVMFIE EMANGDWYYQ IHSHLEYFPK
λTLAII-1  ---------- -----T--DK E-TT--TST- -L-D------ ----------

K2-1      PGEKISCVVE HAS
λTLAII-1  ---------- ---
```

Figure 1. Alignment of K1-5 (**A**) and K2-1 (**C**) with λTLAI-1 and λTLAII-1. For inferred aa sequences of K1-5 (**B**) and K2-1 (**D**) the reading frames of the λTLAI-1 and λTLAII-1 were used.

Results

In order to obtain a realistic estimate of the number of putative class I and class II *B* genes present in carp, probes should be used which, with a high probability, will hybridize to the conserved parts of most of the class I genes present. These conditions can be fulfilled by the PCR products KI and KII described by Hashimoto and co-workers (1990), which were obtained after amplification of carp genomic DNA using degenerate primers based on the conserved aa sequences surrounding the cysteine residues in the third and second domain of class I and class II *B*, respectively.

In order to validate the general use of the KI and KII PCR fragments in assessing the number of *Cyca* genes and their polymorphism, it was investigated whether similar fragments could be obtained from a European carp. To this end a PCR was performed with oligonucleotide primers, which sequences were based on the conserved elements of the genomic sequences reported for the Japanese carp. After amplification two fragments were cloned from the 190 bp band. Sequence analysis of one of the cloned fragments (K1-5), excluding the primer sequences, demonstrated extensive similarity to carp class I genomic clone λTLAI-1 (Hashimoto *et al.* 1990) of 97% at the nucleotide level (Fig. 1A), whereas another cloned fragment, designated K2-1, was shown to be 87% similar to carp class II genomic clone λTLAII-1 (Fig. 1C). In the K1-5 sequence six substitutions were observed of which three were non-synonymous, resulting in two aa substitutions, one conservative and one non-conservative (Fig. 1B). However, the number of non-synonymous nucleotide substitutions in the K2-1 sequence was much higher, giving rise to nine non-conservative and two conservative aa substitutions (Fig. 1B,D).

K1-5 and K2-1 were used as probes for the *Cyca* genes in all experiments described in this study. In addition, in a limited number of experiments hybridizations were repeated with KI and KII. Hybridizations were performed on restriction-enzyme-digested high molecular weight DNA of carp of different geographical origins and genomic status. A total number of five different geographical origins, *i.e.*, Dutch (W and WK), German (E), Hungarian (R3), Polish (R8), and Israeli (A), were represented in the essentially homozygous mitotic gynogenetic clone (E20 x E6), partly homozygous second generation meiotic gynogenetic families (A4 and W11), inbred lines (R3 and R8), and outbred population (WK) used in this study.

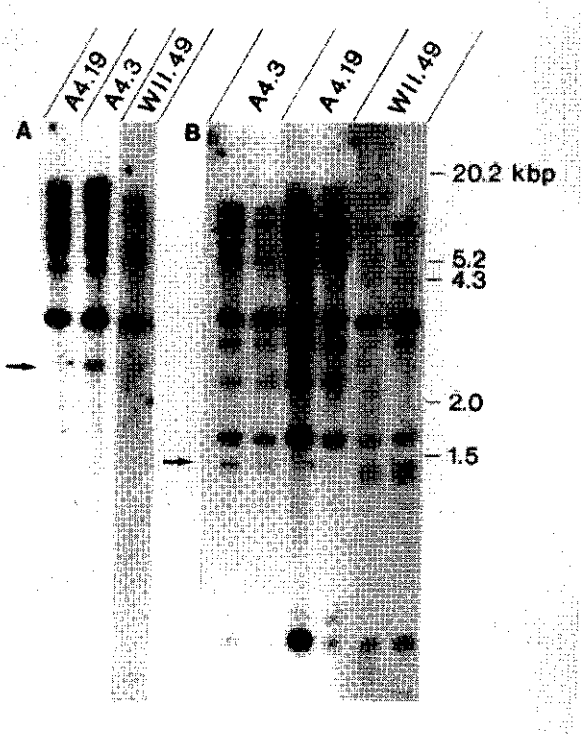


Figure 2. RFLP patterns of *Eco*RI- (**A**) and *Pst*I-digested (**B**) DNA of different carp strains, hybridized with carp class I probe K1-5. Arrows indicate presence of hybridizing fragment correlating with serological typing for the *K* locus of the A4.3 and A4.19 families.

Table 1. Hybridization patterns of *Eco*RI-digested DNA, hybridized with *Cyca-Z* exon 3 (K1-5) probe.

Carp line	Fragment Size (bp)									#
	17500	11700	8150	7100	6100	5750	5000	3150	2450	
A4.19	1*	1	1	1	1	0*	1	1	0	7
A4.3	1	1	1	1	1	0	1	1	1	8
W11.49	1	1	0	1	0	1	1	1	0	6

* 1 indicates presence, 0 absence of fragment
\# number of fragments

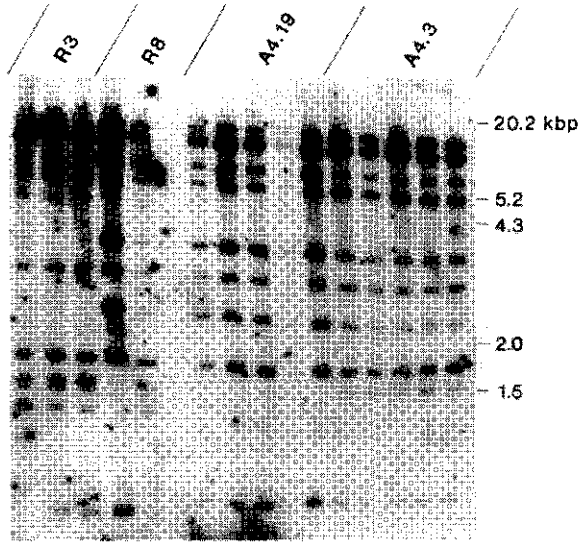


Figure 3. RFLP patterns of *Pst*I-digested DNA of different carp strains, hybridized with carp class I probe K1-5.

Class I genes.

Southern blot hybridization patterns of *Eco*RI-digested DNA obtained with the ^{32}P-labeled K1-5 probe, homologous to a *Cyca-Z* exon 3, showed that in the gynogenetic carp families A4.19, A4.3, and W11.49 six to eight hybridizing fragments were detected (Fig. 2A, Table 1), five of which were present in all individuals tested. Although only three gynogenetic families were analyzed in this experiment, they all appeared to be characterized by a unique hybridization pattern.

The analysis was subsequently extended with the inclusion of other carp strains using digestion with *Pst*I. In these experiments 9-12 hybridizing fragments were observed in the carp families studied (Figs. 2B and 3, Table 2), with four fragments seen in each individual. No differences were detected between individuals within each family. In contrast, most families were characterized by a unique hybridization pattern, except that no differences were seen between the W11.49 and W11.52, and between the R8 and A4.19 families. Particular families were characterized by the presence or absence of certain *Pst*I fragments. The A4.19 was lacking

Table 2. Hybridization patterns of *Pst*I-digested DNA, hybridized with *Cyca-Z* exon 3 (K1-5) probe.

Carp line	Fragment Size (kb)														#
	14.6	10.6	7.0	6.2	5.7	4.7	3.45	3.0	2.6	2.25	2.1	2.05	1.95	0.85	
A4.19	1*	1	1	0*	1	0	1	1	1	1	0	0	0	1	9
A4.3	1	1	1	0	1	0	1	1	1	1	1	0	0	1	10
W11.49	0	1	0	0	1	1	1	1	0	1	1	1	0	1	9
W11.52	0	1	0	0	1	1	1	1	0	1	1	1	0	1	9
E20xE6	0	1	1	0	1	1	1	1	0	1	1	1	0	0	9
R3	0	1	1	1	1	1	0	1	0	1	1	1	1	1	11
R8	1	1	1	0	1	0	1	1	1	1	0	0	0	1	9

*# Symbols same as Table 1.

the 2100 bp, the E20 x E6 the 850 bp, and the R3 the 3450 bp fragment, respectively. In addition, the R3 was the only family showing a 6200 bp *Pst*I fragment, compared to the other families studied (Table 2). In a separate experiment where the segregation of *Pst*I fragments was studied in an F1 hybrid family R3 x R8, hybridization of K1-5 to *Pst*I-digested DNA of R3 and R8 revealed that three out of the eight polymorphic fragments (14,600; 3,450 and 2,600 bp)(see Table 2) were seen to segregate in the R3 x R8 progeny. With respect to the number of fragments, essentially the same observations were made in hybridizations of *Taq*I-digested DNA from the gynogenetic families A4.3, W11.49, W11.52, and E20 x E6. Thus, from the total number of eight to nine fragments per family, five were seen in all individuals (Figs. 4A and 5A). In contrast to the analyses with *Pst*I, however, where the R3 and R8 inbred lines showed differences between but not within the particular families, now individuals were also different within the families. The number of fragments (12) seen in the WK individual was higher compared to that observed in the gynogenetic families (eight or nine). In all the above-reported hybridization experiments no differences were observed between the use of the K1-5, or the KI probe.

The level of polymorphism expressed as the percentage of polymorphic fragments among the total number of fragments observed was calculated to be 70%, when using *Pst*I or *Taq*I digestions. However, in hybridization experiments with K1-5 performed on *Eco*RI-digested DNA of individuals from only three families, no more than 45% of the number of fragments detected could be classified as polymorphic.

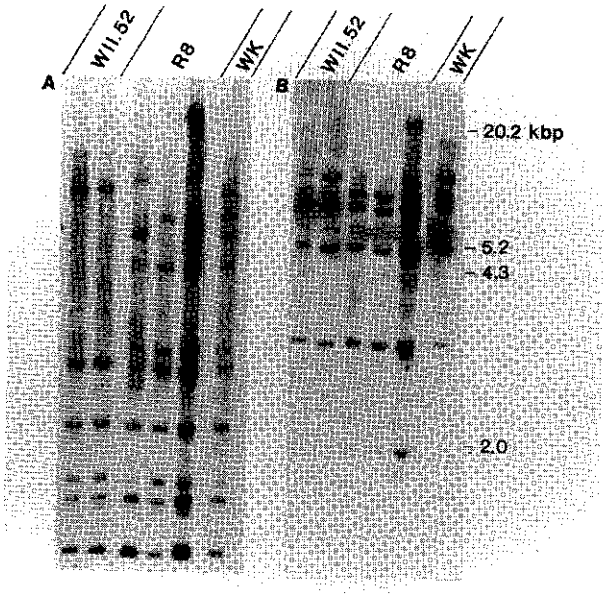


Figure 4. RFLP patterns of *Taq*I-digested DNA of different carp strains, hybridized with class I probe K1-5 (**A**) and class II *B* probe K2-1 (**B**).

Class II *B* genes.

The hybridization patterns of restriction-enzyme-digested DNA with probe K2-1 were assessed in a number of carp families under low stringency conditions. A difficulty in interpreting the hybridization patterns was the frequent occurrence of weakly hybridizing fragments. The hybridization patterns of *Pst*I-digested DNA showed three to five strongly hybridizing fragments, with an additional number of weakly hybridizing fragments. If the latter were included in the analysis, a total number of six to nine fragments could be observed in the four carp families studied (Fig. 6, Table 3). The same was found for *Taq*I-digested DNA hybridized with K2-1, revealing three to five strongly hybridizing fragments, and accumulating in 7-10 if the other fragments were also scored (Figs. 4B and 5B). Each family demonstrated a characteristic hybridization pattern, except those of *Taq*I-digested DNA of the E20 x E6 and R3 carp, which appeared to be similar. No differences in hybridization patterns of *Pst*I-digested DNA were observed between individuals of the W11.52, W11.49, A4.3, A4.19, and E20 x E6 families. However, WK individuals from an outbred population showed differences in both K1-5 and K2-1 hybridizations, whereas individuals from R3 and R8 families only differed in their K1-5 probe hybridization patterns.

The level of polymorphism of the K2-1 hybridizing fragments among the families studied,

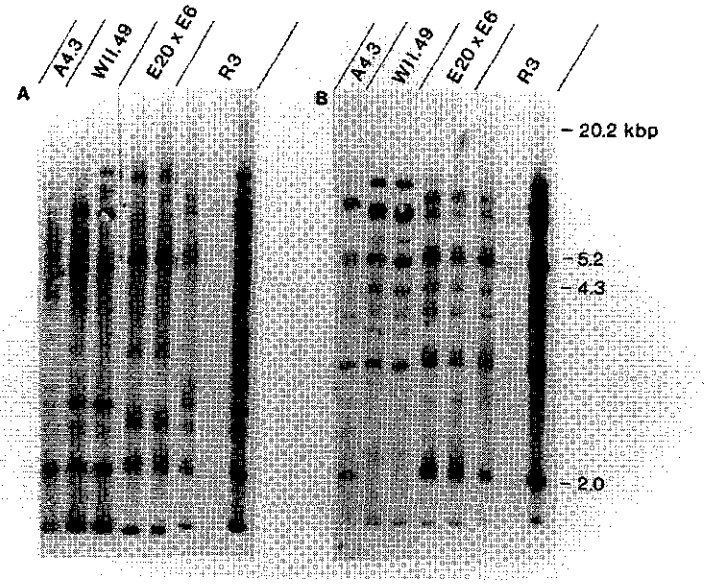


Figure 5. RFLP patterns of *Taq*I-digested DNA of different carp strains, hybridized with class I probe K1-5 (**A**) and class II *B* probe K2-1 (**B**).

Table 3. Hybridization patterns of *Pst*I-digested DNA, hybridized with *Cyca-YB* exon 2 (K2-1) probe.

Carp line	Fragment Size (kb)													#
	12.1	8.6	7.75	7.45	+ 6.6	6.2	5.2	+ 4.75	+ 3.95	+ 3.6	+ 3.4	+ 2.6	+ 2.55	
A4.3	1*	1	0*	1	0	1	1	0	0	1	0	1	1	5(8)
W11.49	0	1	1	0	0	0	1	1	1	0	1	0	0	3(6)
E20xE6	1	1	0	0	1	1	1	1	1	0	0	0	0	4(7)
R3	1	1	0	0	0	1	1	1	1	1	0	1	1	4(9)

Total number of strong fragments, and weak fragments in parentheses.
\+ Weakly hybridizing fragment
* 1 indicates presence and 0 indicates absence of fragment.

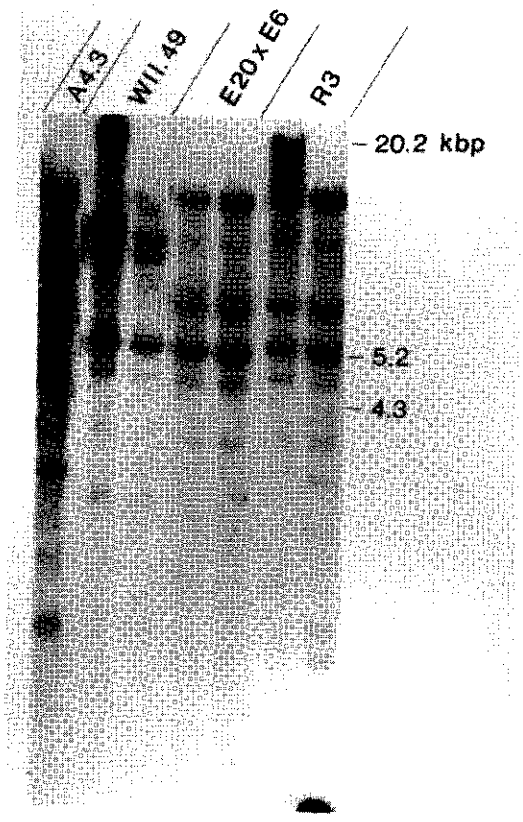


Figure 6. RFLP patterns of *Pst*I-digested DNA of different carp strains, hybridized with class II *B* probe K2-1.

when expressed as the percentage of polymorphic fragments, was calculated to be 40% and 66% for *Taq*I and *Pst*I, respectively.

Serological correlation.

In a previous study (Kaastrup *et al.* 1989) alloantisera were raised between carp from a first gynogenetic generation of the female A4. Two alloantisera were described that were shown to correspond to two co-dominantly expressed allelic products of a single major histocompatibility locus *K*, incorporating class I-like characteristics. These alloantisera were used to type carp from this gynogenetic A4 progeny serologically. Two families, A4.19 and

A4.3, were produced by gynogenesis from female carp typed K1 and K2 homozygous, respectively. High molecular weight DNA from individuals from the A4.19 and A4.3 families was digested with *Eco*RI and *Pst*I and hybridized with the *Cyca-Z* exon 3-specific K1-5 probe to investigate whether a serological typing for this putative major histocompatibility locus *K* would correlate with the molecular genotyping of *Cyca-Z* genes. The A4.19 and A4.3 differed in both the *Eco*RI and *Pst*I hybridization patterns (Fig. 2). In the *Eco*RI and *Pst*I hybridization patterns the A4.3 showed an additional fragment of 2450 bp and 2100 bp, respectively.

Discussion

The *Mhc* of bony fishes (Osteichthyes) has remained elusive for more than two decades following the first description of allograft rejection in this class of vertebrates (Hildemann 1970; Kallman 1970). Although the results from genetic analyses of tissue transplantation, mixed leukocyte reactivity, and serologically detectable alloreactivity, in combination with the presence of an integrated immune system were seen to be indicative of the presence of an *Mhc* at this phylogenetic level, these observations have not led to a proper identification and characterization of *Mhc* genes and gene products in teleostean fishes (reviewed in Stet and Egberts 1991). Also, antibodies recognizing *bona fide Mhc* molecules from other vertebrate species were unable to unequivocally precipitate teleostean *Mhc* molecules resulting from putative cross-reactivity (Kaufman *et al.* 1990b, c). It was not until the advent of PCR technology that a beginning could be made with the identification of the *Mhc* in a teleostean fish (Hashimoto *et al.* 1990). Two PCR fragments, designated KI and KII, were shown to correspond to carp class I (*Cyca-Z* exon 3) and a class II *B* (*Cyca-YB* exon 2) specific sequences, respectively. Invariably, these exons have been shown to be the most conserved among distinct loci of different mammalian species (Klein and Figueroa 1986; Figueroa and Klein 1986). To assess the conserved nature of these exons among different haplotypes of the *Cyca*, specific primers were constructed based on the genomic sequences of the *Cyca-Z* and *Cyca-YB* genes. Two fragments designated K1-5 and K2-1 were obtained after amplification of genomic DNA of a European carp. Sequence analyses of these fragments showed that the corresponding exons are indeed well conserved as indicated by the similarity to the *Cyca-Z* and *Cyca-YB* sequences obtained from a Japanese carp. The inferred aa sequence of the K1-5 probe shows a similar level of substitutions as observed in comparisons of mammalian class I sequences (Sood *et al.* 1985). In contrast, the numbers of non-synonymous substitutions in the K2-1 sequence resulting in clustered and mainly non-conservative substitutions seems remarkable. Whether this reflects interallelic or interlocus variability remains to be investigated.

The availability of probes identifying the most conserved exons of the class I and class II *B*

Cyca genes makes it feasible to study a number of characteristics of the *Cyca*, such as the number of genes and the level of polymorphism. Thus, the number of genes may be estimated from the number of hybridizing fragments in Southern hybridizations using 6 bp-restriction-endonuclease digestion of genomic DNA (Cami *et al.* 1981). Since a most accurate estimate can be obtained from homozygous carp, where the complexity of the hybridization pattern should be reduced, carp families have been included which were reproduced by gynogenesis, either by inhibition of the second meiosis or by inhibition of the first mitosis (Komen *et al.* 1988; Komen *et al.* 1991). Both techniques will lead to fast inbreeding and inferred homozygosity of the *Cyca* genes.

The gynogenetic carp families were shown to possess 6-10 fragments hybridizing with class I probes. The number of hybridizing fragments in the partly inbred strains R3 and R8 is comparable to the number observed in the gynogenetic families. However, some residual heterozygosity of the *Cyca* genes is expected to be present in these strains, because in the R3 x R8 F1 hybrids segregation was observed of some class I-hybridizing *Pst*I fragments. Moreover, the hybridization patterns of *Taq*I-digested DNA of R3 and R8 obtained with the K1-5 probe, the sequence of which incorperates a *Taq*I restriction site, also revealed differences between individuals of each of these inbred strains. This phenomenon was not observed in the gynogenetic strains (A4, W11, and E20 x E6).

The number of class I genes seems to fall outside the range of 17-36 class I genes given for mammalian species such as man and mice (Srivastava *et al.* 1985; Steinmetz *et al.* 1982). However, the number of carp class I genes is comparable to that of other mammals such as rabbit (Rebière *et al.* 1987) and swine (Satz *et al.* 1985). Of the non-mammalian vertebrates, only the number of class I genes of the chicken (6 *B-F* genes in the *B12* haplotype) is known (Guillemot *et al.* 1988). In *Xenopus* only recently a class I cDNA clone has been described, and therefore no data are as yet available on the number of class I genes (Flajnik *et al.* 1991), although biochemical data suggest that only one classical class I gene is expressed (Flajnik *et al.* 1984).

The number of class II *B* genes, as estimated from the number of hybridizing *Pst*I fragments with the class II *B* exon 2-specific probes, is about half of that of the class I genes, namely three to five, if only the strongly hybridizing fragments are taken into account. Apart from these fragments, an additional three to five weakly hybridizing fragments were observed. The significance of these faint bands is hard to evaluate due to the low stringency conditions under which the hybridization had to be performed. No apparent explanation can be given for the anomalous hybrid stability of these homologous probes. However, these low stringency hybridizations were not only a peculiarity of the K1-5 and K2-1 probes, but also of the KI and KII probes. Therefore, it is difficult to assess whether these weakly hybridizing fragments

are indeed class II genes, or represent irrelevant signals due to cross-hybridizations with other genes as a result of sequence similarity (Santos *et al.* 1983). Thus, the number of class II *B* genes as estimated from the strongly hybridizing fragments is comparable to the estimates for mice (Steinmetz *et al.* 1982), humans (Bodmer and Bodmer, 1984), and chicken (Guillemot *et al.* 1988). Also, in *Xenopus* five different class II ß chains have been identified by two-dimensional gelelectrophoresis of immunoprecipitated membrane glycoproteins (Kaufman *et al.* 1985).

The analyses of the *Mhc* in mammals have revealed that duplication is a main event in the evolution of this complex, leading to its contemporary genetic organization (Klein and Figueroa 1986; Figueroa and Klein 1986). It appears that duplication of *Mhc* genes is also seen in fish, be it, to a lesser degree, at least for class I genes. Whether this reflects a more primordial state of the *Cyca*, or is the result of a more recent contraction remains unclear. Hashimoto and co-workers (1990) have raised the suggestion, based on aa sequence comparisons deduced from *Cyca-YB* exon 1, that this exon encodes a domain resembling a V-set domain, unlike that of higher vertebrates. This might suggest that the *Cyca* is preserved in a more primordial state compared to the *Mhc* of higher vertebrates.

Although the number of carp strains used in this study has been limited, the level of polymorphism of the *Cyca* genes is remarkable. Almost all strains exhibit different hybridization patterns with both class I and class II *B* probes. In Europe, pond culture of carp dates back to the 16th century (Berka 1985). Especially in Germany and Bohemia, several carp races were developed for use in intensive pond culture, whereas in China no specific carp races were developed. Four main races could be distinquished namely, Aischgrunder, Lausitzer, Bohemian and Galician, but between the two world wars most carp races became extinct, while the remaining were extensively mixed. Nowadays, most German carp were reconstructed from Galician x Lausitzer. Because pre- and post-war exports were mainly carp with a Galician stem, in many carp populations in other countries this race can be found.

The W strain originates from the only carp farm in the Netherlands. This farm started in 1899, breeding with a mixture of German races. In 1956 mirror carp were selected for breeding under stringent mass selection, which resulted in the establishment of the W strain. The Hungarian R8 and the Polish R3 have undergone similar selections. Also, Israeli carp farming started by the introduction of a number of European races. Between 1965 and 1970 a large selection experiment starting with five different lines from different carp farms resulted in the A (DOR70) strain. Although the carp strains A, W, E, R3 and R8 all had a different geographical origin, it is by no means clear from this study what the level of polymorphism of the *Cyca* is in the local populations in terms of gene frequencies. However, the observed polymorphism might reflect variability between populations consisting of different races present

in these geographical regions. This is substantiated by the observation that all phenotypically different carp (WK group) showed different hybridization patterns (data not shown).

The level of polymorphism can be expressed as the proportion of polymorphic fragments among the total number of fragments detected (Nei and Li 1979), thus enabling an interspecies comparison. RFLP analyses of class I genes in a number of mammalian species has revealed that their polymorphism ranges from 30-80% (Orr 1983; Palmer *et al.* 1983; Nižetić *et al.* 1985). The level of *Cyca* class I polymorphism of 70% for both *Pst*I and *Taq*I, as detected in this study with a limited number of different carp strains, falls well within the range given for mammalian species. Moreover, extending the analysis to more strains can only lead to an increase of the level of restriction fragment length polymorphism. The polymorphism of 40%-60% as calculated for class II *B* genes is slightly lower compared to that of class I. Still, this observation indicates that the *Cyca* class II *B* genes are also highly polymorphic. No information is as yet available on the genetic organization of the class I and class II regions of the *Cyca* in terms of the presence of different loci. Future studies revealing this organization will be needed to enable us to study the polymorphism of the class II region *A* and *B* genes using locus-specific probes (Bakura *et al.* 1985; Andersson and Rask 1988).

In a number of species it has been possible to establish a correlation between molecular genotyping using RFLP analyses, and serological, cellular or biochemical typing of the *Mhc* (Hála *et al.* 1988; Joosten *et al.* 1990). The latter three define expressed polymorphism, which will allow functional studies to be performed, and the results to be interrogated properly. It is therefore important to establish whether a similar correlation exists between different *Cyca*-typing methods in carp.

In a previous study on the presence of an *Mhc* in carp, we have described two co-dominantly expressed allelic specificities K1 and K2 of a single histocompatibility locus *K*, which were identified with alloantisera produced within a first-generation gynogenetic offspring from the female carp A4 (Kaastrup *et al.* 1989). Although this locus *K* was shown to incorporate class I-like characteristics, it was not clear whether it is part of a major histocompatibility complex.

In order to substantiate the observation that the *K* locus may represent a *Cyca* class I locus, the hybridization patterns obtained with the KI and K1-5 probes of *Eco*RI- and *Pst*I-digested DNA from two families, which had been serologically typed K1/1 and K2/2, respectively, were analyzed. Hybridizations with the class I probes revealed in both digestions an additional fragment in the K2 homozygous individuals (A4.3), as compared to the K1 homozygous family (A4.19). As the A4.3 and A4.19 families were produced by gynogenesis and therefore theoretically should only have maternal genes, the difference observed in the hybridization patterns may indeed reflect a difference in *K* haplotypes, as this had been the

basis for the selection of the female carp A4-3 and A4-19. Such a difference was absent in hybridization patterns of the two other second generation gynogenetic families W11.49 and W11.52, which were produced from two serologically unselected first-generation gynogenetic female carp (W11-49 and W11-52) sharing the same mother W11. Thus, a correlation is observed between molecular genogyping for *Cyca* class I genes, and expression of serologically defined alleles of a single histocompatibility locus *K*. No data are available on the co-segregation of these traits, since the A4 female carp is not available anymore. However, a gynogenetic progeny from a K1/K2-heterozygous first-generation gynogenetic A4 carp is currently being produced. Unfortunately, such data are as yet not available for class II typing due to the low reproducibility of MLR analyses (Stet and Egberts 1991).

Future functional studies of polymorphic *Cyca* gene products would be greatly facilitated by the availability of antibodies recognizing monomorphic determinants, similar to W6/32 used in mammalian studies (Neefjes *et al.* 1986). Concurrently, elucidation of the genetic organization of the *Cyca*, providing locus-specific probes, might refine such functional studies.

Acknowledgements

We are grateful to Adrie Groeneveld and Geert Wiegertjes for the production of the gynogenetic carp families, and to Annette Regelink and Jessica ter Steege for assistence in the experiments. Drs. Keiichiro Hashimoto and Yoshikazu Kurosawa are acknowledged for their kind gift of the KI and KII probes, and Dr. Felipe Figueroa for critical reading of the manuscript.

References

Andersson, L. and Rask, L. 1988. Characterization of the MHC class II region in cattle. The number of *DQ* genes varies between haplotypes. *Immunogenetics 27:* 110-120

Bakura, M., Vincek, V., Figueroa, F. and Klein, J. 1985. How polymorphic are class II loci of the mouse *H-2* complex? *Immunogenetics 21:* 569-579

Berka, R. 1985. A brief insight into the history of Bohemian carp pond management. In: *Aquaculture of Cyprinids* (eds: Billard, R. and Marcel, J.), INRA, Paris, pp. 35-40

Bodmer, J.G. and Bodmer, W.F. 1984. Histocompatibility testing 1984. *Immunol. Today 5:* 251-254

Bourlet, Y., Béhar, G., Guillemot, F., Fréchin, N., Billault, A., Chaussé, A.M., Zoorob, R. and Auffray, C. 1988. Isolation of chicken major histocompatibility complex class II (*B-L*) ß chain sequences: comparison with mammalian ß chains and expression in lymphoid organs. *EMBO J. 7:* 1031-1039

Cami, B., Brégégère, F., Abastado, J.P. and Kourilsky, P. 1981. Multiple sequences related to classical histocompatibility antigens in the mouse genome. *Nature 291:* 673-675

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

Figueroa, F. and Klein, J. 1986. The evolution of class II MHC genes. *Immunol. Today 7:* 78-81

Flajnik, M.F., Canel, C., Kramer, J. and Kasahara, M. 1991. Evolution of the major histocompatibility complex: molecular cloning of major histocompatibility complex class I from the amphibian *Xenopus*. *Proc. Natl. Acad. Sci. USA 88:* 537-541

Flajnik, M.F., Kaufman, J.F., Riegert, P. and Du Pasquier, L. 1984. Identification of class I major histocompatibility complex encoded molecules in the amphibian *Xenopus*. *Immunogenetics 20:* 433-442

Guillemot, F., Billault, A., Pourquié, O., Béhar, G., Chaussé, A.M., Zoorob, R., Kreibich, G. and Auffray, C. 1988. A molecular map of the chicken major histocompatibility complex: the class II ß genes are closely linked to the class I genes and the nucleolar organizer. *EMBO J. 7:* 2775-2785

Hála, K., Chaussé, A.M., Bourlet, Y., Lassila, O., Hasler, V. and Auffray, C. 1988. Attempts to detect recombination between *B-F* and *B-L* genes within the chicken *B* complex by serological typing, in vitro MLR, and RFLP analyses. *Immunogenetics 28:* 433-438

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

Heery, D.M., Gannon, F. and Powell, R. 1990. A simple method for subcloning DNA fragments from gel slices. *Trends Genet. 6:* 173

Hildemann, W.H. 1970. Transplantation immunity in fishes: Agnatha, Chondrichthyes and Osteichthyes. *Transplant. Proc. 2:* 253-259

Joosten, I., Hensen, E.J., Sanders, M.F. and Andersson, L. 1990. Bovine MHC class II restriction fragment length polymorphism linked to expressed polymorphism. *Immunogenetics 31:* 123-126

Kaastrup, P., Stet, R.J.M., Tigchelaar, A.J., Egberts, E. and Van Muiswinkel, W.B. 1989. A major histocompatibility locus in fish: serological identification and segregation of transplantation antigens in the common carp (*Cyprinus carpio* L.). *Immunogenetics 30:* 284-290

Kallman, K.D. 1970. Genetics of transplantation in Teleostei. *Transplant. Proc. 2:* 263-271

Kaufman, J.F., Flajnik, M.F. and Du Pasquier, L. 1985. *Xenopus* MHC class molecules. II. Polymorphism as determined by two-dimensional gel electrophoresis. *J. Immunol. 134:* 3258-3264

Kaufman, J.F., Skjoedt, K. and Salomonsen, J. 1990a. The MHC molecules of nonmammalian vertebrates. *Immunol. Rev. 113:* 83-117

Kaufman, J.F., Skjoedt, K., Salomonsen, J., Simonsen, M., Du Pasquier, L. and Parisot, R.

1990b. MHC-like molecules in some nonmammalian vertebrates can be detected by some cross-reactive xenoantisera. *J. Immunol. 144:* 2258-2272

Kaufman, J.F., Ferrone, S., Flajnik, M.F., Kilb, M., Völk, H. and Parisot, R. 1990c. MHC-like molecules in some nonmammalian vertebrates can be detected by some cross-reactive monoclonal antibodies. *J. Immunol. 144:* 2273-2280

Klein, J. 1986. Natural history of the Major Histocompatibility Complex. John Wiley and Sons, New York

Klein, J. 1987. Origin of the major histocompatibility complex polymorphism: the trans-species hypothesis. *Human Immunol. 19:* 152-162

Klein, J., Bontrop, R.E., Dawkins, R.L., Erlich, H.A., Gyllensten, U.B., Heise, E.R., Jones, P.P., Parham, P., Wakeland, E.K. and Watkins, D.I. 1990. Nomenclature for the major histocompatibility complexes of different species: a proposal. *Immunogenetics 31:* 217-219

Klein, J. and Figueroa, F. 1986. The evolution of class I MHC genes. *Immunol. Today 7:* 41-44

Komen, J., Bongers, A.B.J., Richter, C.J.J., Van Muiswinkel, W.B. and Huisman, E.A. 1991. Gynogenesis in common carp (*Cyprinus carpio* L.). II. The production of homozygous gynogenetic clones and F1 hybrids. *Aquaculture 92:* 127-142

Komen, J., Duynhouwer, J., Richter, C.J.J. and Huisman, E.A. 1988. Gynogenesis in common carp (*Cyprinus carpio* L.). I. Effects of genetic manipulation of sexual products and incubation condition of eggs. *Aquaculture 69:* 227-239

Nagy, A., Raijki, K., Horváth, L. and Csányi, V. 1978. Investigations on carp *Cyprinus carpio* L. gynogenesis. *J. Fish. Biol. 13:* 215-224

Neefjes, J.J., Hensen, E.J., De Kroon, T.I.P. and Ploegh, H.L. 1986. A biochemical characterization of feline MHC products: unusually high expression of class II antigens on peripheral blood lymphocytes. *Immunogenetics 23:* 341-347

Nei, M. and Li, W.H. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. USA 76:* 5269-5273

Nižetić, D., Figueroa, F., Nevo, E. and Klein, J. 1985. Major histocompatibility complex of the mole-rat. II. Restriction fragment length polymorphism. *Immunogenetics 22:* 55-67

Orr, H.T. 1983. Use of the Southern blotting to analyse the size and restriction fragment polymorphism of HLA class I DNA in the human population. *Transplant. Proc. 15:* 1900-1906

Palmer, M., Wettstein, P.J. and Duncan, W.R. 1983. Evidence for extensive polymorphism in class I genes of the major histocompatibility complex (RT-1). *Proc. Natl. Acad. Sci. USA 72:* 1184-1188

Rebière, M.C., Marche, P. and Kindt, T.J. 1987. A rabbit class I major histocompatibility complex gene with a T cell specific expression pattern. *J. Immunol. 139:* 2066-2074

Santos, E., Reddy, E.P., Pulciani, E.P., Feldmann, R.J. and Barbacid, M. 1983. Spontaneous activation of a human proto-oncogene. *Proc. Natl. Acad. Sci. USA 80:* 4679-4683

Satz, M.L., Wang, L.C., Singer, D.S. and Rudikoff, S. 1985. Structure and expression of two porcine genomic clones encoding class I MHC antigens. *J. Immunol. 135:* 2167-2175

Sood, A.K., Pan, J., Biro, P.A., Pereira, D., Srivastava, R., Reddy, V.B., Duceman, B.W., Weissman, S.M. 1985. Structure and polymorphism of class I MHC antigen mRNA. *Immunogenetics 22:* 101-121

Srivastava, B., Duceman, B.W., Biro, P.A., Sood, A.K. and Weissman, S.M. 1985. Molecular organization of the class I genes of the human major histocompatibility complex. *Immunol. Rev. 84:* 93-121

Steinmetz, M., Winoto, A., Minard, K. and Hood, L. 1982. Clusters of genes encoding mouse transplantation antigens. *Cell 28:* 489-498

Stet, R.J.M. and Egberts, E. 1991. The histocompatibility system in teleostean fishes: from multiple histocompatibility loci to a major histocompatibility complex. *Fish Shellfish. Immunol. 1:* 1-16

Wohlfarth, G., Lahman, M., Hulata, G. and Moav, R. 1980. The story of DOR-70: a selected strain of Israeli common carp. *Bamidgeh 32:* 3-5

Chapter 3

Characterization of β_2-microglobulin transcripts from two teleost species

Brian Dixon[1], René J.M. Stet[2], Saskia H.M. van Erp[2] and Bill Pohajdak[1]

[1]Marine Gene Probe Laboratory, Department of Biology, Dalhousie University, Halifax, Nova Scotia, Canada, B3H 4J1
[2]Department of Experimental Animal Morphology and Cell Biology, P.O. Box 338, 6700 AH Wageningen, Wageningen Agricultural University, The Netherlands

Reprinted from **Immunogenetics 38, 27-34 (1993)**
with kind permission of ©Springer-Verlag

Summary

Using degenerate primers based on published $ß_2$-microglobulin sequences we were able to obtain an expected 111-base-pairs (bp) polymerase chain reaction (PCR) fragment from tilapia genomic DNA. The sequence of this fragment showed a high degree of similarity to mouse $ß_2$-microglobulin at the protein level. We used these primers in an "anchored PCR" to obtain a 213 bp PCR fragment from a carp cDNA library. This was then used to clone a full-length $ß_2$-microglobulin cDNA from carp. The carp sequence showed the highest similarity to rabbit $ß_2$-microglobulin. Both sequences showed strong similarities to all previously published vertebrate $ß_2$-microglobulin sequences. The predicted protein secondary structure of both the carp and tilapia clones was almost identical to the corresponding regions of previously known vertebrate $ß_2$-microglobulin protein sequences. When either the carp or tilapia probes were used against corresponding Northern blots they hybridized to a message of approximately 800-1000 bases long, which corresponds to the previously published lengths of $ß_2$-microglobulin mRNAs. Southern blotting indicated that $ß_2$-microglobulin was encoded by a single copy gene in both cases. Phylogenetic analysis indicated that the sequences were related to the $ß_2$-microglobulins of higher vertebrates but grouped together in an ancestral position.

The nucleotide sequences reported in this paper have been submitted to the EMBL/GenBank nucleotide sequence databases and have been assigned the accession numbers L05536 (carp) and L05537 (tilapia).

Introduction

The polymerase chain reaction (PCR) technique is one of the most recent techniques developed to isolate and clone DNA rapidly (Saiki *et al.* 1988). The major advantages of PCR over other techniques is the ability to amplify DNA from very minute DNA sources (Saiki *et al.* 1988). Since the PCR primers only have to hybridize to very short stretches of DNA, this technique has proved to be extremely valuable in cloning genes from different species across large evolutionary time differences (Sakanari *et al.* 1989). In these types of experiments, degenerate primers are often synthesized based on conserved regions of the homologous proteins from various species. Since the expected product size is often known, the PCR product may be purified from non-specific products by a separating gel.

The major histocompatibility complex (*Mhc*) is a group of genes which encode extremely polymorphic proteins essential to the function of the immune system of all vertebrates. There are two major types or classes of *Mhc* proteins. Class I *Mhc* molecules are located on the surface of most cells and are involved in presenting foreign antigens to cytotoxic T cells. Class I *Mhc* molecules consist of two polypeptide chains: an M_r 45,000 alpha chain which associates with an M_r 12,000 chain, referred to as $ß_2$-microglobulin (reviewed in Klein 1986). Class II *Mhc* molecules are restricted to antigen-presenting cells, and they are also composed of two chains, α and ß, both of which are approximately M_r 30,000 in size (Klein 1986).

Recently Hashimoto and co-workers (1990) used the degenerate-primer strategy to clone both class I and II *Mhc* from carp. They synthesized primers based on similar sequences of the α_3 and $ß_2$ domain from chicken, mouse and human *Mhc* classes I and II. The isolated PCR products were sequenced and used as probes to screen a carp genomic library. The coding region from these clones had some sequence similarity to class I and II *Mhc* molecules from other species but, more importantly, contained several conserved amino acids (aa) in key positions and shared structural similarity with class I and II *Mhc* molecules from other species. The fish class II *Mhc* also showed characteristics similar to the immunoglobulin (*Ig*) superfamily V set, thus providing more indications that *Ig*s, T-cell receptors, and *Mhc* evolved from a common protein. The cloning of carp *Mhc* using PCR has facilitated the rapid cloning of *Mhc* molecules from rainbow trout (Glamann *et al.* 1991; Juul-Madsen *et al.* 1992), Atlantic salmon (Fosse *et al.* 1991) and sharks (Kasahara *et al.* 1992; Hashimoto *et al.* 1992).

The isolation of $ß_2$-microglobulin protein from distantly related vertebrates and invertebrates has been the goal of several groups (Shalev *et al.* 1981, 1983; Warr *et al.* 1984; Roch *et al.* 1983). The importance that $ß_2$-microglobulin plays together with class I molecules in presenting endogenous peptides (Nuchtern *et al.* 1989; Townsend *et al.* 1989,1990) and preventing the presentation of exogenous peptides (Rock *et al.* 1990, 1991) should theoretically be common to all organisms with an immune system capable of recognizing non-self antigens. Understanding the evolution of $ß_2$-microglobulin from various taxa may provide insights into how the *Mhc* function arose. In most of these earlier studies $ß_2$-microglobulin was detected serologically using xenogenic antibodies in goldfish (Warr *et al.* 1984), earthworms (Roch *et al.* 1983), *Drosophila* (Shalev *et al.* 1983), and cod, plus several invertebrates (Shalev *et al.* 1981). These experiments are not totally conclusive, as they only indicate the presence of a protein structurally similar to $ß_2$-microglobulin. In addition, the co-precipitation of putative class I molecules using their putative $ß_2$-microglobulin was never observed. Further proof, such as sequencing these proteins or their corresponding genes, is required to conclude that they are indeed $ß_2$-microglobulin. We attempted to obtain cDNA clones of $ß_2$-microglobulin from two fish species, tilapia (*Oreochromis niloticus* L.) and carp (*Cyprinus carpio* L.) in order to prove conclusively that this protein exists and is functional in lower vertebrates.

Materials and methods

Fish stocks

The tilapia used in this experiment were cultured in the Marine Gene Probe Laboratory (MGPL). The carp were A4.10, a second-generation gynogenetic carp line produced at the Department of Experimental Animal Morphology and Cell Biology (Wageningen, The Netherlands; Kaastrup *et al.* 1989).

Tilapia PCR reactions and sequencing of products

Tilapia genomic DNA was isolated from fresh red blood cells as described by Moreno and co-workers (1989). Oligodeoxyribonucleotide primers for PCR were synthesized by the MGPL on an Applied Biosystems PCR-Mate Model 391 DNA Synthesizer, and purified on Nensorb Prep columns (Du Pont de Nemours, Den Bosch, The Netherlands). The primer sequences are B2M-1: 5'-d[CA(A,G)GT(A,T,C,G)TA(T,C)(T,A)(C,G)(A,T,C,G)(C,A)G(A,T,C,G)CA]-3'; and B2MB: 5'-d[CTC(C,G,T)CCGTTCTTCAGCAG]-3', with sequence degeneracies indicated in parentheses. PCR was performed in a DNA Thermocycler (Perkin Elmer, Norwalk,CT) with 45 cycles consisting of denaturation at 94°C (first cycle 5 min; 30 s thereafter), annealing at 55°C (30 s), and polymerization at 72°C (1 min; 5 min, final cycle). All reaction volumes were 100 µl overlaid with 50 µl light mineral oil, and contained the following final concentrations: 10 mM tris-HCl pH 8.3 at 25 °C; 50 mM KCl; 1.5 mM $MgCl_2$; 200 µM each of dATP, dCTP, dGTP, dTTP; 1 µM each of the primers; 50 ng genomic DNA; and 2.5 units *Taq* DNA polymerase (Perkin Elmer). PCR products were fractionated on a vertical 7% polyacrylamide gel run in 1 x Tris-borate (TBE). The DNA was visualized by staining the gel in 0.5 µg/ml ethidium bromide. DNA products were isolated by excising the bands, crushing the gel, and extracting the DNA overnight in sterile dH_2O. The DNA was subcloned into the *Hin*cII site of M13mp18; single-stranded phage DNA was sequenced by the dideoxy chain termination method (Sanger *et al.* 1977), using the modified T7 DNA polymerase version 2.0 Sequenase kit (US Biochemicals, Cleveland, OH).

Sequence analysis

The DNA sequences (and derived protein sequences) were analyzed against the GenBank, the National Biomedical Research Foundation, and the Protein Identification Resource and Swiss Protein databases, using the computer search program Blast (Altschul *et al.* 1990) and local similarity alignment program FastA (Pearson and Lipman 1988). Matched protein sequences were globally aligned using Align (Myers and Miller 1988). The sequences were analyzed using several programs from the GCG package (Genetics Computer Group, Madison,

WI). The phylogram was constructed from the predicted fish aa sequences plus previously published vertebrate sequences, using the program Clustal V (Higgins *et al.* 1992) which uses the neighbor-joining method of Saitou and Nei (1987).

Southern blotting

For the Southern analysis (Sambrook *et al.* 1989) of tilapia, 15 μg of the indicated DNAs were digested to completion with *Eco*RI or *Hin*dIII, and electrophoresed at 1.2 V/cm for 13 hours (h) in a 0.8% agarose gel run in 1 x TBE pH 8.3. The gel was stained with 0.5 μg/ml ethidium bromide and photographed. The gel was transferred to nylon (Amersham, Oakville, Canada; Sambrook *et al.* 1989) and crosslinked with ultraviolet light for 5 min. The radiolabeled DNA probe was prepared by PCR from the tilapia 111 bp PCR product in M13mp18, using the M13 forward and reverse sequencing primers (Perfect Match Cloning Systems: Stratagene, La Jolla, CA), and substituting 33.3 μmoles of [alpha-^{32}P]dCTP (approximately 3000 Ci/mmol; Amersham) for dCTP in the deoxyribonucleotide mixture (Bellamy *et al.* 1990). PCR conditions were 45 cycles at 94°C (first cycle 3 min; 15 s thereafter), 50°C (15 s), and 72°C (30 s). The blot was probed (approximately 2 x 10^6 cpm per 10 ml) at 42°C for 18 h in hybridization solution (5 x saline sodium phosphate-EDTA (SSPE), 0.1% sodium dodecyl sulfate (SDS), 1% w/v bovine serum albumin, 150 μg/ml tRNA, 10% sodium dextran sulfate, 50% formamide, and 1 x Denhardt's solution). The blot was washed four times with 0.2 x SSPE pH 7.4, 0.1% w/v SDS at 55°C for 30 min, and exposed to XAR5 film (Kodak, Rochester, NY) for 18 h at -80°C with an intensifying screen. The blot was then further washed with 0.1 x SSPE pH 7.4, 0.1% w/v SDS at 55°C, and exposed as above (see figures 3 A, B). High molecular weight carp DNA was obtained from liver. Small samples were homogenized in TEN (100 mM Tris/HCl pH 8.0, 10 mM EDTA, 250 mM NaCl, 1% SDS) buffer, and incubated overnight at 50°C in the presence of proteinase-K (1 mg/ml). DNA was isolated by subsequent phenol extractions and ethanol precipitations. Southern analysis of carp DNA was performed similarly, except that the filter was probed with WAG2B2, a 213 bp carp PCR fragment, in hybridization solution containing 50% formamide, 5 x SSC, 0.1% SDS, 5 mM EDTA, 5 x Denhardt's, and 100 μg denatured *Escherichia coli (E. Coli)* DNA, and the filter was washed for 15 min in 4 x SSC, 0.1% SDS at 50°C.

Tilapia RNA preparation

10 ml of blood were obtained from the caudal vein of tilapia. The blood was centrifuged over 3 ml Histopaque 1119 ficoll (Sigma, St. Louis, MO) for 30 min at 300 x g. Lymphocytes were collected and stimulated with 10 μg/ml poly I:C in RPMI 1640 plus glutamine, 20 mM Hepes, pH 7.5 and 2 g/l sodium bicarbonate (Flow Laboratories, McClean, WV) containing

10% heat-inactivated fetal calf serum (Flow Laboratories) for 30 min at 30°C. Total cellular RNA was obtained from the white blood cells by the method of Chirgwin and co-workers (1979). The integrity and quantity of the RNA was analyzed by electrophoretic fractionation on a denaturing 0.8% agarose gel at 90 V for 2 h. Carp RNA isolation is outlined below in the cDNA library construction section.

Northern Analysis

Approximately 15 μg of total RNA were denatured and loaded on a denaturing 0.8% agarose gel and electrophoresed at 90 V for 2 h. RNA was transferred to a nylon membrane (Amersham). The blot was probed with ^{32}P-dCTP-labeled random-primed (Feinberg and Vogelstein 1983) tilapia 111 bp PCR product. The probe was hybridized in hybridization solution (see Southern blotting of tilapia) overnight and the blot was washed at 55°C in 0.1 x SSPE, 0.1% SDS for 30 min. Blots were autoradiographed with XAR5 film (Kodak) overnight at -80°C with an intensifying screen. Northern analysis of carp RNA was performed similarly, except that 3-6 μg of poly(A)$^+$-RNA (see RNA isolation in next section) were run per lane and the filter was probed with a 213 bp carp PCR fragment (WAG2B2, see Results and discussion) in hybridization solution containing 50% formamide, 5 x SSC, 0.1% SDS, 5 mM EDTA, 5 x Denhardt's, and 100 μg denatured *E. coli* DNA, and the filter was washed for 15 min in 4 x SSC, 0.1% SDS at 50°C.

Preparation of carp mRNA and cDNA library

Total RNA was isolated from pronephros (head kidney) and spleen by homogenizing and selective precipitations in the presence of high molar concentrations of lithium chloride and urea. After proteinase-K treatment and phenol/chloroform extractions, total RNA was precipitated in ethanol, washed, and dissolved in water. cDNA, containing *Eco*RI adaptors, was prepared from poly(A)$^+$-RNA using an mRNA purification kit (Promega, Madison, WI) and a cDNA synthesis kit (Pharmacia, Uppsala, Sweden) according to the manufacturer's specifications. The obtained cDNA was ligated into *Eco*RI-digested λgt11 vector, and packaged. The resulting primary library contained 8 x 10^6 recombinant PFU. The library was amplified and λgt11 phage particles were isolated using a plate lysate method, followed by a standard lambda DNA extraction.

PCR screening of carp library.

Aliquots (300 ng) of λgt11-cDNA-library DNA were used in an anchored PCR by combining λgt11 sequence primers (λ forward: 5'-d[GGTGGCGACGACTCCTGGAGCCCG]-3' or λ reverse: 5'-d[TTGACACCAGACCAACTGGTAATG]-3' (Promega)) with tilapia

$ß_2$-microglobulin degenerate oligonucleotides (B2M-1 or B2M-B). The reaction was performed in *Taq* buffer, using 1 unit of *Taq* polymerase (Promega), supplemented with 1.5 mM $MgCl_2$ and 200 μM of each dNTP in a final volume of 100 μl. The mixtures were subjected to a thermal cycle profile (1 min 94°C, 2 min 55°C, 1 min 72°C) for 30 cycles, with an additional extension step at 72°C for 10 min.

Cloning and sequencing.

The appropriate PCR products were excised from agarose gels and recovered by centrifugation and precipitation (Heery *et al.* 1990). Possible non-template extensions present in the PCR fragments were filled in using T4 polymerase (Pharmacia) in the presence of dNTPs. Fragments thus obtained were cloned in the *Eco*RV site of pBluescript SK (Stratagene). The nucleotide sequence was determined with automatic sequencing (Applied Biosytems, Maarssen, The Netherlands) using DyePrimers (M13RP1). Sequence data were analyzed using the GCG package.

Screening of the carp cDNA library.

Hybond-N+ filters (Amersham) containing a total of 8 x 10^5 PFU from a carp λgt11 cDNA library were hybridized with the 213 bp WAG2B2 probe labeled to a specific activity of 2 x 10^8 cpm/μg using the random priming method (Feinberg and Vogelstein 1983). Filters were prehybridized in a solution containing 40% formamide, 5 x SSC, 0.1% SDS, 5 mM EDTA, 5 x Denhardt's, and 100 μg denatured *E. coli* DNA at 42°C for 6 h. Hybridization was performed at 42°C for 16 h in the presence of a labeled probe (1 x 10^7 cpm per filter). Filters were washed for 30 min in 4 x SSC, 0.1% SDS at 45°C, followed by a high-stringency wash in 1 x SSC, 0.1% SDS at 65°C for 15 min. Filters were exposed to XAR5 film (Kodak) and positive plaques were identified. These plaques were again hybridized to the WAG2B2 probe, and single positive plaques were isolated. cDNA inserts from positive plaques were obtained by PCR using λ forward and reverse sequence primers (Promega). Appropriate DNA fragments were isolated and cloned into pTZ18R and pTZ19R. ssDNA was obtained and sequenced using M13 reverse sequence primer, Sequenase version 2.0 (US Biochemicals) and 35S-dATP. Positive clones were verified using an internal PCR primer (OL85: 5'-d[CTGCCATGTCAGTGGCTTCCA]-3', based on the sequence of the 213 bp WAGB2B PCR fragment) and the λ forward and reverse sequence primers.

Results and discussion

Cloning and Sequencing.

We synthesized degenerate primers to regions of the ß$_2$-microglobulin genes based on conserved protein regions in several mammalian species (Klein 1986; Gussow *et al.* 1987; Gold *et al.* 1987). The primer B2M-1 was designed to correspond to aa positions 8-14 of ß$_2$-microglobulin which are conserved between all mammalian species (Klein 1986). B2M-B was designed to match one of the conserved regions described by Gold and co-workers (1987), which are conserved among all members of the *Ig* superfamily including human ß$_2$-microglobulin aa positions 39-44 (Gussow *et al.* 1987). In both cases we reduced degeneracy at the 3' end of the primer by designing them so that there were aa's with fewer degenerate codons there. These primers were used in PCR reactions with tilapia DNA and resulted in an expected 111 bp fragment. This fragment was sequenced (Fig. 1) and these primers were used in conjunction with the λ forward and reverse primers to obtain a 213 bp PCR fragment (WAG2B2) from a carp cDNA library (78.4% identity, six out of eight substitutions conservative in a 37 aa residue overlap with the tilapia sequence: 79% identity and five out of five substitutions conservative in the 24 aa region between the primers). The 213 bp fragment was obtained with the combination of B2M-B and λ forward primers. Reactions using B2M-1 were unsuccessful as there is a region of poor similarity between the tilapia and carp sequences at the 3' end of the primer. This would result in poor primer annealing and thus no amplification.

The 213 bp carp PCR fragment was used to screen a carp cDNA library. The first screening produced four positive plaques. DNA from these four clones was amplified in a PCR using λ forward and reverse sequence primers. The clones were tested by performing a PCR reaction with the primer OL-85, a primer derived from the carp 213 bp PCR product sequence, and λ forward or reverse sequence primer. Three clones produced fragments of the predicted size; WAGB2M-1, WAGB2M-2, and WAGB2M-3. Sequencing of these clones revealed that the size of the cDNA insert of WAGB2M-1 was 1095 bp, and WAGB2M-2 and WAGB2M-3 both contained inserts of 879 bp. WAGB2M-1 and WAGB2M-2 were sequenced completely, while WAGB2M-3 was partially sequenced to confirm that it was similar to WAGB2M-2. The sequence of WAGB2M-1 is presented in figure 1. The sequences of WAGB2M-1 and WAGB2M-2 were identical with the exceptions that in WAGB2M-2 position 165 is a T and position 406 is a G. The T at position 165 alters the predicted aa to serine, which is the predicted aa from the tilapia sequence (Fig. 1) and the mammalian sequences at this position (Fig. 2). WAGB2M-2 and WAGB2M-3 end at position 879, but unlike WAGB2M-1, positions 877 to 879 are all A's. The presence of multiple polyadenylation sites

```
   1    AAACTCTGGACTAAATCACCAAAA

  25    ATG AGA GCA ATC ATC ACT TTT GCC CTG TTC TGT GTG CTG TAC GTC
 -20     m   r   a   i   i   t   f   a   l   f   c   v   l   y   v

                                                      Q   V   Y   W
                                                    CAG GTG TAT TGG
  70    ACT GTA CAA GGG AAA ACA TCC AGT CCC AAG GTT CAG GTG TAC AGC
  -5     t   v   q   g   K   T   S   S   P   K   V   Q   V   Y   S

         R   H   P   G   E   Y   G   K   E   D   V   L   I   C   H
        AGG CAT CCA GGA GAG TAT GGG AAG GAA GAT GTC TTG ATC TGC CAC
 115    CAT TTT CCC GGA GAG TAC GGA AAG GAG AAC ACC CTG ATC TGC CAT
  12     H   F   P   G   E   Y   G   K   E   N   T   L   I   C   H

         V   S   N   F   H   P   P   D   I   T   I   T   L   L   K
        GTG AGT AAT TTC CAC CCC CCT GAC ATC ACC ATC ACA CTG CTG AAG
 160    GTC AGG GGC TTC CAC CCT CCT GAT ATC ACC ATT GAA CTG CTG AAG
  27     V   A   G   F   H   P   P   D   I   T   I   E   L   L   K

         N   G   E
        AAC GGC GAG
 205    GAT GGC GAG ATT CTC CCA AAT ACC CAG CAG ACG GAC CTG GCC TTC
  42     D   G   E   I   L   P   N   T   Q   Q   T   D   L   A   F

 250    GAA AAG GGC TGG CAG TTT CAC CTC ACC AAG AGC GTC ACC TTC AAA
  57     E   K   G   W   Q   F   H   L   T   K   S   V   T   F   K

 295    CCA GAG AGA GGA CAG AAC TAC GCT TGC AGT GTT CGA CAC ATG AAC
  72     P   E   R   G   Q   N   Y   A   C   S   V   R   H   M   N

 340    AAT AAA AAC ATC TAT TCT TGG GAG CCC AAC ATG TAA
  87     N   K   N   I   Y   S   W   E   P   N   M

 376    AACTAAAACAGTCAAAGGCGGAGGAAAGGCGGATAATTTCACATGTTTTGGACATTTTC
 435    CATCATCTGCAGAAACAATAATATCCTGCTGTTCGATATTACAAACCTGATGTTGCTCA
 494    AGAAGCACTATTCATACAGATGACTTTTGTAACTTCTGCAATTCATCGATTTTAGCTTT
 553    TTCTTTAGTTACCTTTTTTTTTTCTTTTTCATAAGTGTAAGGATAATTTTACCAAAATA
 612    GCTGCTACGTGTTTTTGAATAAGTTTTGATCAGTTTCCGATTCTTACCCTGTTTCATTC
 671    CTCTCTGTGTAGTATGCAATACTGGGCCATATTTCTGCTTACTTTTGGCCTTTATTTTG
 730    TAATTAAATACCAATTTCACAATCAATTTTCTTTTTAAAACAAATAAGACTACTGCAAA
 789    CTGGGTATTTAATGATATCAATAATTCAATTTGACACTTTTCAGAGAAATGTCATTCTA
 848    AGTGGGGAAAATAAAGTTTAGAAAACCTAATGGAACAGTTGTGTTTATTTGCTTCAAAT
 907    TTTCTGTACTAATGCCTGTACTTTTAAGTTAAATCAACGTGATAGTGAGTTTCAGTTAT
 966    TTGTCTAACCTAAAACATAAACGTCTGCCTAGCAACTGAGCAACTAAGGTGGTGATTGG
1025    ATTAAAAAAAACAAGTAGAGACAAGTAGACTTTCGTTTTTCTGTGAGGCTATTAAAAAG
1084    GTAAAGTCCTCG
```

Figure 1. Complete sequence of WAGB2M-1, the longest carp β_2-microglobulin cDNA, aligned with the tilapia 111 bp PCR product. Predicted protein sequence of tilapia sequence displayed above DNA sequence; predicted protein sequence of carp DNA shown below DNA sequence. Underlined regions of tilapia sequences indicate location of degenerate PCR primers used to obtain fragment. Putative carp 19 aa residue hydrophobic leader sequence shown in lower case letters. Nucleotide number and aa number shown on left-hand side beside corresponding sequence. Polyadenylation sequence underlined (positions 857-862), as is adenine at position 877 after which poly(A)-tail is added.

in the 3' untranslated region (UT) of mouse ß$_2$-microglobulin produces several species of mRNAs which differ in size (Parnes *et al.* 1983). However, the carp cDNA sequence, like human ß$_2$-microglobulin (Gussow *et al.* 1987), contains only one polyadenylation signal sequence (positions 856 to 861, underlined in figure 1) which is 15 bases upstream from the end of clones WAGB2M-2 and WAGB2M-3. We deduced that the the poly A tail is added after position 876 (underlined) and that WAGB2M-2 and WAGB2M-3 are polyadenylated, while WAGB2M-1 represents a transcript which has not yet been polyadenylated. This clone was probably isolated along with the poly(A)$^+$-RNA due to the long adenine stretches in its 3' end.

All the carp sequences contained a 348 bp open reading frame encoding a predicted 116 aa precursor protein and a 97 aa mature protein which would have a relative mass of 13,300. This is consistent with previously published ß$_2$-microglobulin sequences (Gates *et al.* 1979; Wolfe and Cebra 1980; Parnes and Seldman 1982; Gussow *et al.* 1987). The tilapia and carp sequences are shown with their predicted aa sequences in figure 1. The predicted aa sequences of both DNA sequences show features characteristic of ß$_2$-microglobulins of vertebrate species. ß$_2$-microglobulin is a member of the *Ig* superfamily. Proteins from this family of genes always contain a pair of cysteine residues which form a disulfide bridge producing a 65 aa residue loop (Williams 1987). There are two cysteine residues 65 residues apart in the predicted carp sequence (residues 25 and 80). Residue 25 is conserved in the corresponding position in the tilapia sequence (Figs. 1, 2), indicating that these sequences encode proteins which are members of the *Ig* superfamily. Further evidence that the carp sequence encoded a member of the *Ig* superfamily of genes was obtained using the program Profilescan (Gribskov *et al.* 1988) from the GCG program package. This program indicated that aa residues around the cysteine residue at position 80 (YACSVRH) matched those found in the consensus sequence found in all members of the *Ig* superfamily ([F,Y]xCx[V,A]xH).

Similarity searches.

Both sequences show a high degree of similarity to the corresponding sequences of previously published ß$_2$-microglobulin sequences. The tilapia sequence was most similar to mouse ß$_2$-microglobulin both with (62% identity, 5 out of 14 substitutions conservative in a 37 aa residue overlap) and without the primer sequences included (54% identity, 4 out of 10 substitutions conservative in a 22 aa overlap; Fig. 2). The primers may be included in this comparison as they are degenerate and thus do not absolutely determine the sequences at the end of the DNA fragment.

The carp sequence showed the highest similarity to rabbit ß$_2$-microglobulin (50% identity, 17 out of 43 substitutions conservative in an 86 aa residue overlap; Fig. 2). The cysteine residues at positions 25 and 80 of the carp sequence align with cysteine residues in

```
          -20       -10         1        10        20        30        40
           |         |          |        |         |         |         |
TILAPIA                                  QVYWRHPGEYGKEDVLICHVSNFHPPDITITLL
CARP      MRAIITFALFCVLYVT-VQGKTSSPKVQVYSHFPGEYGKENTLICHVAGFHPPDITIELL
RABBIT                             VQRAPNVQVYSRHPAENGKPNFLNCYVTSGHPPQIDIELM
MOUSE     MARSVTLVFLVLVSLTGLYAIQKTPQIQVYSRHPPENGKPNILNCYVTQFHPPHIEIQLM
HUMAN     MSRSVALAVLALLSLSGLEAIQRTPKIQVYSRHPAENGKSNFLNCYVSGFHPSDIEVDLL
Consensus *                        *  ******* * ** * * * *  **** * * *

                  50        60        70        80        90
                   |         |         |         |         |
TILAPIA   KNGE
CARP      KDGEILPNTQQTDLAFEKGWQFHLTKSVTFKPERGQNYACSVRH--MNNKNIYSWEPNM
RABBIT    KNGVKIENVEQSDLSFNKDWSFYLLVHTEFTPNNKNEYSCRVKHVTLKEPMTVKWDRDY
MOUSE     KNGKKIPKVEMSDMSFSKDWSFYILAHTEFTPTETDTYACRVKHASMAEPKTVYWDRDM
HUMAN     KNGERIEKVEHSDLSFSKDWSFYLLYYTEFTPTEKDEYACRVNHVTLSQPKIVKWDRDM
Consensus ***          *  * * * *       * *     * * * *          *
```

Figure 2. Alignment of fish $ß_2$-microglobulin with known vertebrate $ß_2$-microglobulin protein sequences. Rabbit and mouse $ß_2$-microglobulin sequences were obtained from Klein (1986). Human $ß_2$-microglobulin sequence from Gussow and co-workers (1987). Consensus sequence indicates aa residues conserved in at least four out of five sequences (*).

$ß_2$-microglobulins of higher vertebrate species (Fig. 2). In addition, the conserved proline residues at positions 14, 32 and 72 which are important for the secondary structure of $ß_2$-microglobulin are all highly conserved in the fish sequences (Fig. 2). A Chou and Fasman (1978) analysis of the secondary structure using the GCG package program Peptide Structure (Jameson and Wolf 1988) of the predicted carp aa sequence indicated that ß-sheet regions extend from aa residues 7-13, 21-27, 35-40, 60-70 and 78-83. These align with ß-sheet regions in mammalian $ß_2$-microglobulin sequences which are thought to be important for secondary structure (Wolfe and Cebra 1980). Previous reports of the detection of a $ß_2$-microglobulin-like protein in goldfish using xenogeneic antibodies (Warr *et al.* 1984) suggested the potential presence of $ß_2$-microglobulin in teleosts, but the high degree of similarity between our clones and vertebrate sequences confirms that fish indeed produce $ß_2$-microglobulin.

The rabbit sequence was obtained by protein sequencing and thus lacks the hydrophobic leader sequence (Gates *et al.* 1979). The human and mouse hydrophobic leaders are both cleaved after aa residue -1 (Fig. 2; Parnes and Seldman 1982; Gussow *et al.* 1987). Based on this alignment the carp sequence should have a potential 19 aa hydrophobic signal sequence. The program Signal Sequence, which uses the signal cleavage site prediction method of Von Heijne

(1986), predicted that the carp sequence would be cleaved between the 19th and 20th aa (S value = 1.0000). The alignment of the predicted fish $ß_2$-microglobulin sequence with three mammalian $ß_2$-microglobulins is shown in figure 2. Aa which are conserved in at least four out of the five sequences are denoted with an asterisk (*). Most mammalian $ß_2$-microglobulins have a mature protein length of 99 aa residues with a 20 aa residue leader sequence. The carp sequence is 116 aa's long. The alignment obtained with the program Clustal V (Higgins *et al.* 1992) shown in figure 2 indicates that the difference in length is due to the presence of two additional aa residues near the carboxy terminus (positions 85 and 86) and one additional aa residue in the hydrophobic leader sequence (position -4) in mammalian $ß_2$-microglobulin sequences. Either fish $ß_2$-microglobulin has deleted aa's which were found in the $ß_2$-microglobulin precursor of the common ancestor of all vertebrates, or these residues represent the insertion of extra sequences into the $ß_2$-microglobulin of the ancestral terrestrial vertebrates following their divergence from fish.

Northern Analysis.

While this is the first report of a $ß_2$-microglobulin from a fish, *Mhc* class I and II genes from carp have already been reported (Hashimoto *et al.* 1990). The report of the carp class I and II *Mhc* genes, however, details only genomic clones and provides no evidence that the reported DNA sequences are expressed. Obtaining a cDNA clone of $ß_2$-microglobulin from fish is strong evidence for expression, but to further validate these results and characterize our clones we performed northern analysis (Fig. 3). The tilapia PCR fragment hybridized to an mRNA between 800 and 1000 bases long (Fig. 3A) and the WAG2B2 PCR fragment detected a message of similar size (Fig. 3B). This is consistent with the previously reported sizes of $ß_2$-microglobulin mRNAs (Suggs *et al.* 1981; Parnes *et al.* 1983; Gussow *et al.* 1987). In addition, the WAG2B2 fragment detected a smaller band which may represent a breakdown product of $ß_2$-microglobulin mRNA (Fig. 3B).

The detection of an mRNA consistent in size with other $ß_2$-microglobulin mRNAs supports the conclusion that our sequences do indeed encode $ß_2$-microglobulin. The fact that the RNA used in these two blots was obtained from different tissues is consistent with these mRNAs encoding $ß_2$-microglobulin, since its close association with class I *Mhc* requires it to be expressed in all tissues (Klein 1986). The function of $ß_2$-microglobulin is to ensure that endogenous peptides are held in the cleft of the *Mhc* class I molecule, then to dissociate following the loss of the endogenous peptide to prevent the class I molecule from obtaining and presenting exogenous peptides (Rock *et al.* 1990, 1991). The association of class I *Mhc* and $ß_2$-microglobulin indicates that the previously reported carp *Mhc* genes are probably expressed. This is in agreement with the reports of the isolation of cDNA clones of class I *Mhc* genes

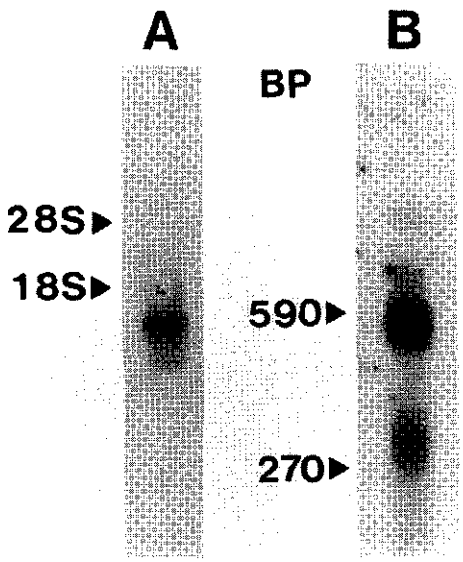


Figure 3. Northern blot of carp and tilapia RNA probed with corresponding β_2-microglobulin clones. **A.** Tilapia total RNA probed with 111 bp PCR product. Arrows indicate the position of human 28s and 18s rRNA bands used as size standards. **B.** Carp total RNA probed with 213 bp WAG2B2 PCR product. Arrows indicate position of DNA size standards.

from carp (Van Erp *et al.* 1996), Atlantic salmon (Fosse *et al.* 1991), and class II *Mhc* genes from several other fish species (Glamann *et al.* 1991; Juul-Madsen *et al.* 1992; Kasahara *et al.* 1992). The presence of expressed *Mhc* genes in fish indicates that the immune system of fishes functions in a manner consistent with that of higher vertebrates.

Southern analysis.

Unlike the other *Mhc* proteins, β_2-microglobulin is usually encoded by a single-copy gene in higher vertebrates (Parnes *et al.* 1983; Klein 1986). Southern analysis using the tilapia and carp PCR fragments indicated that the tilapia sequence is encoded by a single-copy gene (Fig. 4A), but that two copies of the gene are present in carp (Fig. 4C). The two bands in figure 4C, lane 4, are not consistent with the internal *Pst*I restriction site observed within the carp cDNA sequence, as this is outside the 213 bp probe region, and the *Hin*dIII restriction

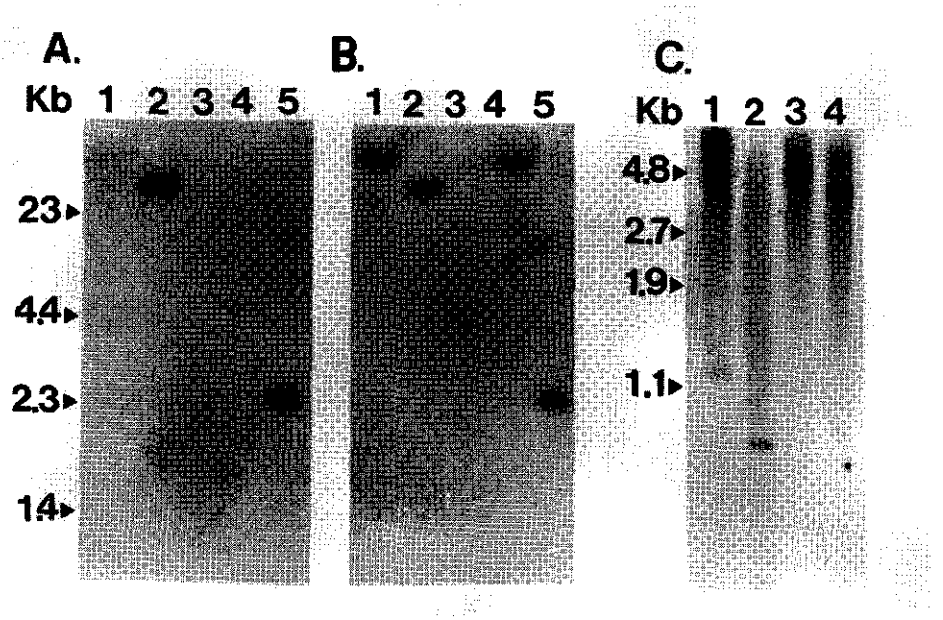


Figure 4. Southern blots probed with fish ß$_2$-microglobulin clones. **A.** Southern blot of rainbow trout, brook trout and tilapia DNA probed with tilapia 111 bp PCR product and washed under stringent conditions (see methods). Arrows indicate size of DNA size standards. Lane 1; rainbow trout DNA digested with *Hin*dIII. Lane 2; tilapia DNA digested with *Hin*dIII. Lane 3; brook trout DNA digested with *Eco*RI. Lane 4; rainbow trout digested DNA with *Eco*RI. Lane 5; tilapia DNA digested with *Eco*RI. **B.** The same Southern blot from A washed under less stringent conditions (see methods). **C.** Southern blot of carp DNA probed with 213 bp WAG2B2 PCR product. Arrows indicate size of DNA size standards. Lane 1; carp DNA digested with *Eco*RI. Lane 2; carp DNA digested with *Hae*III. Lane 3; carp DNA digested with *Hin*dIII. Lane 4; carp DNA digested with *Pst*I.

enzyme digest produces two bands in both tilapia (Fig. 4A, lane 2) and carp (Fig. 4C, lane 2). The multiple bands produced in the carp DNA cannot be alleles as the fish used in the experiment were inbred stock produced gynogenetically (Kaastrup *et al.* 1989), and thus must represent the presence of two copies of the gene. The presence of two copies of this gene in carp is not unusual, since they are regarded as tetraploid species (Ohno *et al.* 1967). This is also consistent with the observation that the two *Pst*I sites do not segregate when tested on Southern blots containing full sibs and outgroup representatives (data not shown). Since tilapia are not tetraploid and our fish were outbred, we suggest that the multiple bands produced by *Hin*dIII digestion probably represent allelic variation between the maternal and paternal copies

of this gene. These polymorphisms may be useful in genetic studies of fish and brood-stock analysis in tilapia aquaculture.

The Southern blot presented in figure 4A was washed under stringent conditions, but a less stringent wash produced the results presented in figure 4B. The tilapia PCR product cross hybridizes with rainbow trout (*Onchorynchus mykiss*) and with brook trout (*Salvelinus fontanalis*) under these less stringent conditions. This indicates the presence of β_2-microglobulin genes in these species, with a high degree of sequence similarity to tilapia β_2-microglobulin. This is not surprising as the β_2-microglobulin sequence is highly conserved over evolutionary time periods (Klein 1986). This observation is consistent with the reports detailing the cloning of *Mhc* class I genes from Atlantic salmon (Fosse *et al.* 1991).

Phylogenetic analysis.

An alignment of the 37 aa residue region defined by the tilapia PCR fragment with carp and seven other species was used to produce a phylogram (Fig. 5A). A second phylogram was produced using the full predicted protein sequence of the carp cDNA clone and β_2-microglobulin sequences from seven other species (Fig. 5B). In figure 5A the tilapia and carp sequences group together and the other species appear to branch in an order consistent with accepted evolutionary pathways. The turkey and chicken sequences are identical over this 37 aa residue stretch and thus group as one branch (Skjodt *et al.* 1986). The phylogram produced with the full sequence (Fig. 5B) was different only in the branching order of some mammalian species. Fish β_2-microglobulin appears to be related to other β_2-microglobulins, but groups apart from them in a phylogenetic analysis.

The data presented here provide evidence that fish contain genes for β_2-microglobulin and actively express them, indicating that the fish immune system works in a fashion similar to that of higher vertebrates. These data also provide a further understanding of how the immune system evolved. So far, the fish immune system appears to be quite similar in its complexity to those of higher vertebrates. Fish do possess an IgM-like immunoglobulin and *Mhc* genes, but an equivalent to the T-cell receptor has not been identified to date. Also, there is no definitive marker for fish T cells available. The identification of fish T-cell markers, or a T-cell receptor would immensely advance our understanding of fish immune systems. In order to determine the origins of this complex system, important vertebrate immunoproteins should be sought in lower chordates and invertebrates.

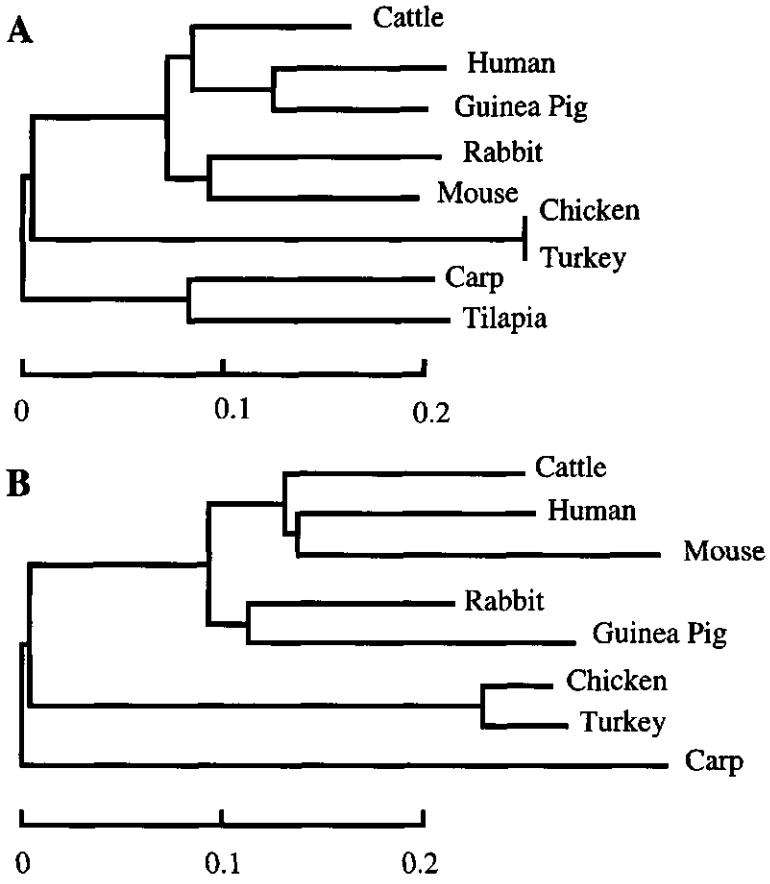


Figure 5. Phylogenetic trees for vertebrate ß$_2$-microglobulin sequences. **A.** Based on the 37 aa region encoded by the tilapia 111 bp sequence; **B.** based on full-length sequences. Scales at the bottom represent genetic distance. Rabbit, mouse and human sequences were obtained from references noted in the legend to figure 2. Chicken and turkey sequences are from Skjodt and co-workers (1988) and Welinder and co-workers (1991). Bovine sequence is from Groves and Greenberg (1982). Guinea pig sequence is from Wolfe and Cebra (1980).

Acknowledgements

We thank Gregory Stuart and Trudi Hermsen for technical assistance. This work was supported by grants to B. P. and the Marine Gene Probe Laboratory from the Ministry of Trade and Technology (N.S. government), the Network of Centres of Excellence program and the Natural Sciences and Engineering Research Council.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. 1990. Basic local alignment search tool. *J. Mol. Biol. 215:* 403-410

Bellamy, R., Inglehearn, C., Lester, D., Hardcastle, A. and Bhattacharya, S. 1990. Better fingerprinting with PCR. *Trends Genet. 6:* 32

Chirgwin, J.M., Przybyla, A.E., MacDonald, R.J. and Rutter, W.J. 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry 18:* 5294-5299

Chou, P.Y. and Fasman, G.D. 1978. Prediction of the secondary structure of proteins from their amino acid sequence. *Adv. Enzymol. 47:* 45-147

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

Fosse, V. M., Hordvik, I., Bergersen, O., Endresen, C. and Lie, Ø. 1991. Cloning and characterization of the *Mhc* class I and class II genes from Atlantic salmon (*Salmo salar*). *Dev. Comp. Immunol. 15:* S38

Gates III, F.T., Coligan, J.E. and Kindt, T.J. 1979. Complete amino acid sequence of rabbit $ß_2$-microglobulin. *Biochemistry 18:* 2267-2272

Glamann, J.V., Juul-Madsen, H. and Simonsen, M. 1991. Isolation of a cDNA clone encoding *Mhc* class II ß chain in rainbow trout. *Dev. Comp. Immunol. 15:* S37

Gold, D.P., Clevers, H., Alarcon, B., Dunlap, S., Novotny, J., Williams, A.F. and Terhorst, C. 1987. Evolutionary relationship between the T3 chains of the T-cell receptor complex and the immunoglobulin supergene family. *Proc. Natl. Acad. Sci. USA 84:* 7649-7653

Gribskov, M., Homyak, M., Edenfield, J. and Eisenberg, D. 1988. Profile scanning for three-dimensional structural patterns in protein sequences. *Comput. Appl. Biosci. 4:* 61-66

Groves, M.L. and Greenberg, R. 1982. Complete amino acid sequence of bovine $ß_2$-microglobulin. *J. Biol. Chem. 257:* 2619-2626

Gussow, D., Rein, R., Ginjaar, I., Hochstenbach, F., Seeman, G., Kottman, A. and Ploegh, H.L. 1987. The human $ß_2$-microglobulin gene; primary structure and definition of transcriptional unit. *J. Immunol. 139:* 3132-3138

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

Hashimoto, K., Nakanishi, T. and Kurosawa, Y. 1992. Identification of a shark sequence resembling the major histocompatibility complex class I α_3 domain. *Proc. Natl. Acad. Sci. USA 89:* 2209-2212

Heery, D.M., Gannon, F. and Powell, R. 1990. A simple method for subcloning DNA fragments from gel slices. *Trends Genet. 6:* 173

Higgins, D.G., Bleasby, A.J. and Fuchs, R. 1992. CLUSTAL V: improved software for multiple sequence alignment. *Comput. Appl. Biosci. 8:* 189-191

Jameson, B. and Wolf, H. 1988. The antigenic index; A novel algorithm for predicting antigenic determinants. *Comput. Appl. Biosci. 4:* 181-186

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

Kaastrup, P., Stet, R.J.M., Tigchelaar, A.J., Egberts, E. and Van Muiswinkel, W. B. 1989. A major histocompatibility locus in fish: serological identification and segregation of transplantation antigens in the common carp (*Cyprinus carpio* L.). *Immunogenetics 30:* 284-290

Kashahara, M., Vazquez, M., Sato, K., McKinney, C. and Flajnik, M.F. 1992. Evolution of the major histocompatibility complex: isolation of class II A cDNA clones from the cartilaginous fish. *Proc. Natl. Acad. Sci. USA 89:* 6688-6692

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

Moreno, R.F., Booth, F.R., Hung I.H., and Tilzer, L.L. 1989. Efficient DNA isolation within a single gel barrier tube. *Nucl. Acids Res. 17:* 8393

Myers, E. and Miller, W. 1988. Optimal alignments in linear space. *Comput. Appl. Biosci. 4:* 11-17

Nuchtern, J.G., Bonifacino, J.S., Biddison, W.E. and Klausner, R.D. 1989. Brefelden A implicates egress from endoplasmic reticulum in class I restricted antigen presentation. *Nature 339:* 223-226

Ohno, S., Muramoto, J., Christian, L. and Atkin, N.B. 1967. Diploid-tetraploid relationship among old-world members of the fish family Cyprinidae. *Chromosoma 23:* 1-9

Parnes, J.R., Robinson, R.R. and Seidman, J.G. 1983. Multiple mRNA species with distinct 3' termini are transcribed from the ß$_2$-microglobulin gene. *Nature 302:* 449-452

Parnes, J.R. and Seldman, J.G. 1982. Structure of wild-type and mutant mouse ß$_2$-microglobulin genes. *Cell 29:* 661-669

Pearson, W.R. and Lipman, D.J. 1988. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA 85:* 2444-2448

Roch, P., Cooper, E.L. and Eskinazi, D.P. 1983. Serological evidences for a membrane structure related to human ß$_2$-microglobulin expressed by certain earthworm leukocytes. *Eur. J. Immunol. 13:* 1037-1042

Rock, K.L., Gamble, S., Rothstein, L.,Gramm C. and Benacerraf, B. 1991. Dissociation of ß$_2$-microglobulin leads to a substantial pool of inactive class I *Mhc* heavy chains on the cell surface. *Cell 65:* 611-620

Rock, K.L., Rothstein, L.E., Gamble, S.R. and Benacerraf, B. 1990. Reassociation with ß$_2$-microglobulin is necessary for K^b class I major histocompatibility complex binding of exogenous peptides. *Proc. Natl. Acad. Sci. USA 87:* 7517-7521

Saiki, R. K., Gelfand, D.H., Stoffel, S., Scharf, S., Higuchi, R., Horn, G., Mullis, K. B. and Erich, H.A. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science 239:* 487-491

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

Sakanari, J.A., Staunton, C.E., Eakin, A.E., Craik, C.S. and McKerrow, P. 1989. Serine proteases from nematode and protozoan parasites: isolation of sequence homologs using generic molecular probes. *Proc. Natl. Acad. Sci. USA 86:* 4863-4867

Sambrook, J., Fristch, E.F. and Maniatis, T. 1989. Molecular Cloning: A Laboratory Manual. 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor

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

Shalev, A., Greenberg, A.H., Logdberg, L. and Bjorck, L. 1981. ß$_2$-microglobulin-like molecules in low vertebrates and invertebrates. *J. Immunol. 127:* 1186-1191

Shalev, A., Pla, M., Ginsburger-Vogel, T., Echalier, G., Logdberg, L., Bjorck, L., Colombani, J. and Segal, S. 1983. Evidence for ß$_2$-microglobulin-like and H-2-like determinants in Drosophila. *J. Immunol. 130:* 297-302

Skjodt, K., Welinder, K.G., Crone, M., Verland, S., Salomonsen, J. and Simonsen, M. 1986. Isolation and characterization of chicken and turkey beta2-microglobulin. *Mol. Immunol. 23:* 1301-1309

Suggs, S.V., Wallace, R.B., Hirose, T., Kawashima, E.H. and Itakura, K. 1981. Use of synthetic oligonucleotides as hybridization probes: isolation of cloned cDNA sequences for human ß$_2$-microglobulin. *Proc. Natl. Acad. Sci. USA 78:* 6613-6617

Townsend, A., Elliott, T., Cerundolo, V., Foster, L., Barber, B. and Tse, A. 1990. Assembly of *Mhc* class I molecules analysed in vitro. *Cell 62:* 285-295

Townsend, A., Ohlen, C., Bastin, J., Ljunggren, H.-G., Foster, L. and Karre, K. 1989. Association of class I major histocompatibility heavy and light chains induced by viral peptides. *Nature 340:* 443-448

Van Erp, S.H.M, Dixon, B., Figueroa, F., Egberts, E. and Stet, R.J.M. 1996. Identification and characterization of a new major histocompatibility complex class I gene in carp (*Cyprinus carpio* L.). *Immunogenetics* in press

Von Heijne, G. 1986. A new method for predicting signal sequence cleavage sites. *Nucl. Acids Res. 14:* 4683-4690

Warr, G.W., Rothstein, I.L. and DeLuca, D. 1984. β_2-microglobulin-like polypeptide of the goldfish, *Carassius auratus*. *Mol. Immunol. 21:* 105-108

Welinder, K.G., Jespersen, H.M., Walther-Rasmussen, J. and Skjodt, K. 1991. Amino acid structures of chicken and turkey beta-2-microglobulin. *Mol. Immunol. 28:* 177-182

Williams, A. 1987. A year in the life of the immunoglobulin superfamily. *Immunol. Today 8:* 298-303

Wolfe, P.B. and Cebra, J.J. 1980. The primary structure of guinea pig β_2-microglobulin. *Mol. Immunol. 17:* 1493-1505

Chapter 4

Identification and characterization of a new major histocompatibility complex class I gene in carp (*Cyprinus carpio* L.)

Saskia H.M. van Erp[1], Brian Dixon[1],
Felipe Figueroa[2], Egbert Egberts[1] and René J.M. Stet[1]

[1]Department of Experimental Animal Morphology and Cell Biology, P.O. Box 338, 6700 AH Wageningen, Wageningen Agricultural University, The Netherlands
[2]Max-Planck Institut für Biologie, Abteilung Immungenetik, D7400 Tübingen, Germany

Accepted for publication in: **Immunogenetics (1996)**

Summary

In this study we report the finding of three representatives of a new group of major histocompatibility complex class I sequences in carp: *Cyca*-12 (*Cyca-UA1*01*), a full-length cDNA; *Cyca*-SP1 (*Cyca-UAW1*), a PCR fragment from cDNA; and *Cyca*-G11 (*Cyca-UA1*02*), a partial genomic clone. Comparison of the amino acid sequences of *Cyca*-12, *Cyca*-SP1 and *Cyca*-G11 with classical and non-classical class I sequences from other species shows considerable conservation in regions that have been shown to be involved in maintaining the structure and function of class I molecules. The genomic organization of *Cyca*-12 has been elucidated by analysis of a partial genomic clone *Cyca*-G11, in combination with PCR amplifications on genomic DNA of a homozygous individual. Although the genomic organization is similar to that found in class I genes from other species, the 3' untranslated region contains an intron which is unprecedented in class I genes, and intron 2 is exceptionally large (± 14 kilobases). Southern blot analysis indicates the presence of multiple related sequences. In phylogenetic analyses, the *Cyca-UA* sequences cluster with class I genes from zebrafish and Atlantic salmon, indicating that the ancestral gene arose before the salmonid/cyprinid split, approximately 120-150 million years ago. The previously reported class I *Cyca-Z* genes from carp, and *Caau-Z* genes from goldfish, cluster as a completely separate lineage. A polyclonal antiserum (anti-*Cyca*12) was raised against a recombinant fusion protein containing most of the extracellular domains of *Cyca*-12. The antibodies showed substantial reactivity with the recombinant protein and an M_r 45,000 protein in membrane lysates of spleen and muscle, as well as to determinants present on leukocytes in fluorescence-activated cell sorter analyses. Erythrocytes and thrombocytes were found to be negative.

The nucleotide sequences reported in this paper have been submitted to the EMBL/GenBank nucleotide sequence databases and have been assigned the accession numbers X9105-X91028.

Introduction

Products encoded by the genes of the major histocompatibility complex (*Mhc*) play a major role in antigen presentation. For a long time the presence of *Mhc* molecules in fish was inferred from indirect evidence, such as the presence of an integrated immune system, mixed leukocyte reactivity (Stet and Egberts 1991), and the detection of a co-dominantly expressed histocompatibility alloantigen that correlated with skin transplant rejection in carp (Kaastrup *et al.* 1989). However, it was not until 1990 that sequences of *Mhc* genes from fish were reported (Hashimoto *et al.* 1990). The *Cyca-Z* (class I) and *Cyca-YB* (class II *B*) partial genomic sequences from common carp that were described in this first publication have over the years been followed by class I and class II genes in a growing number of fish species (Dixon *et al.* 1995). New *Mhc* genes have also been identified in carp, including class II *B* genes (*Cyca-DAB*) (Ono *et al.* 1993a), β_2-microglobulin encoding sequences (Dixon *et al.* 1993), and recently

class II *A* genes *(Cyca-DXA)* (Van Erp *et al.* 1996a). Thus, representatives of all the sequences encoding both heterodimeric *Mhc* molecules have been identified in the carp.

However, data on the carp class I *Cyca-Z* genes are limited, as no sequences other than cDNA fragments encoding the extracellular α_1, α_2 and α_3 domains have been reported to date, and expression has thus far only been detected in kidney (Okamura *et al.* 1993). Moreover, many of the characteristics present in classical, and to a lesser extent in non-classical class I amino acid sequences from other species are not shared by the *Cyca-Z* sequences, and this has raised doubts as to their function. In contrast, a full-length cDNA sequence from Atlantic salmon (*Sasa*-p30) does possess many of the conserved class I features that are absent in *Cyca-Z* (Grimholt *et al.* 1993). Recently, a full-length class I transcript was isolated from zebrafish (*Brre*-U1), which was more closely related to the sequence from salmon than to the *Cyca-Z* sequences (Takeuchi *et al.* 1995). The higher degree of similarity is in contrast with the close phylogenetic relationship between zebrafish and carp. This led to the hypothesis that a gene equivalent to *Sasa*-p30 and *Brre*-U1 could be present in the carp as well.

Despite the wealth of accumulating sequence data on fish *Mhc* genes, studies of expression and function of the encoded products are so far absent. It is generally assumed that *Mhc* molecules in fish perform the same function as their mammalian classical counterparts, but this still needs to be proven experimentally. For such studies it is essential that antibodies directed against fish *Mhc* molecules become available.

The objective of this study, therefore, was to isolate new class I sequences from carp, and to raise polyclonal antibodies against the encoded protein.

Materials and methods

Fish

The fish used in these experiments were laboratory strains of common carp (*Cyprinus carpio* L.). R3, R8, W11-49 and WK were described by Stet and co-workers (1993). A4.10me2 is a second-generation meiotic gynogenetic family, generated from a first-generation meiotic female, A4-10me1 (Kaastrup *et al.* 1989). Homozygous gynogenetic carp were produced by mitotic gynogenetic reproduction from pooled eggs from three A4.10me2 females, according to previously described methods (Komen *et al.* 1990), and subsequent meiotic reproduction of one individual of the mitotic offspring led to the generation of clone A410.

Carp cDNA libraries

A λgt11 cDNA library was prepared from pronephros and spleen of 6 individuals of

2nd generation gynogenetic carp A4.10me2 as previously described (Dixon *et al.* 1993). A thymus cDNA library in λgt10 was constructed from thymi from ten 5-month old individuals of the clone A410. Poly(A)$^+$-RNA was isolated, and cDNA was synthesized using a kit from Pharmacia (Uppsala, Sweden). The initial complexity of the library was 2.6 x 10^6 plaque-forming units (PFU). The library was amplified once in *E. Coli*-strain NM514.

Screening of thymus cDNA library

A total of 1 x 10^6 PFU were transferred to 20 nitrocellulose filters (Schleicher & Schuell, Dassel, Germany), and screened with an *Mhc* class I cDNA clone (U1) from zebrafish (*Brachydanio rerio*) (Takeuchi *et al.* 1995). The probe was labeled by the random primer method (Feinberg and Vogelstein 1983) to a specific activity of 4 x 10^7 cpm/μg. The filters were prehybridized and hybridized as previously described (Stet *et al.* 1993), in a solution containing 40% formamide and 4 x 10^5 cpm of probe per filter. The filters were washed to a final stringency of 0.5 x SSC; 0.1% SDS at 60°C, and after exposure for 16 hours at room temperature, positive plaques were identified. Single positive plaques were isolated after a second screening, using the same procedure.

Screening of carp genomic library

Nitrocellulose filters (n=20) (Schleicher & Schuell) containing approximately 5 x 10^5 PFU/filter of a commercially obtained genomic carp library in λfixII (Stratagene, La Jolla, CA) were screened with the carp cDNA probe. Screening protocols were essentially the same as used for the screening of the cDNA library, with the exception of using a hybridization solution containing 45% formamide. Single positive clones were isolated after a second screening.

DNA sequencing

Selected fragments were subcloned in pTZ18R and pTZ19R (Stratagene), or in pUC18 using the Sureclone ligation kit (Pharmacia). The subclones were sequenced using the Cycle Sequencing Kit (Pharmacia), or the Sequenase DNA sequencing kit (USB, Cleveland, OH). Sequence data were analyzed with several programs from the GCG-package (Genetics Computer Group, Madison WI, USA), ClustalV (Higgins *et al.* 1992) and MEGA (Kumar *et al.* 1993).

Polymerase chain reaction (PCR)

Amplification by PCR was performed as described (Dixon *et al.* 1993), using a thermal cycle profile (1 min 94°C, 1 min 55°C, 2.5 min 72°C) for 30 cycles, with an additional extension step of 10 min at 72°C. Amplification of large fragments from genomic DNA was performed using the Expand Long Template PCR System (Boehringer, Mannheim, Germany),

according to the manufacturer's specifications. λgt10- and λgt11-specific primers (Promega, Madison, WI) were used to amplify cDNA inserts from the libraries.

Southern blot analysis

Genomic DNA was extracted from tissues, and Southern blots, containing 5 μg of *Pst*I-digested DNA per lane, were prepared as described (Stet *et al.* 1993). The filter was hybridized as performed in the screening procedures (*vide supra*), in 40%-45% formamide (depending on the probe used) using 1 x 10^7 cpm of radiolabeled probe, prepared by the random primer method to a specific activity of 1.5 x 10^8 cpm/μg. The filter was washed to a final stringency of 0.5 x SSC at 45°C for 15 minutes, and used to expose X-omat S films (Kodak, Rochester, NY) for 3-5 days at -80°C with intensifying screens.

Expression of recombinant *Cyca*-12, production of a polyclonal antiserum

A 770 basepair (bp) fragment spanning most of exons 2, 3 and 4 was amplified by PCR and cloned in frame into expression vector pRSET (Invitrogen, Leek, The Netherlands). The polyhistidine containing fusion protein was expressed in *E.coli* BL21 (DE3) plysS (Novagen, Madison, WI) in a culture induced with 0.8 mM IPTG, and subsequently isolated under denaturing conditions, according to the Xpress System protein purification procedure, using Ni^{2+}-affinity chromatography (Invitrogen). Pre-immune serum was collected from a female rabbit (New Zealand white), which was subsequently primed by s.c. injection of 200 μg of the recombinant protein, diluted 1:1 in complete Freund's adjuvant.

A new batch of recombinant protein was isolated using the same method, but with a subsequent additional purification using the Prep cell preparative gel electrophoresis system (Biorad, Hercules, CA). The rabbit was boosted 4 weeks after the first immunization by s.c. injection of 90 μg of the highly purified protein, diluted 1:1 in incomplete Freund's adjuvant. Serum (denoted anti-*Cyca*12) was collected 2 weeks after the second immunization. To eliminate aspecific reactivities, the serum was absorbed with carp erythrocytes. To this end, the serum was mixed 1:1 with packed erythrocytes for 2 hours, and the absorbed serum was collected after centrifugation.

Preparation of membrane lysates

Single-cell suspensions were resuspended in cRPMI (RPMI adjusted to 270 mOsmol) with 1mM protease inhibitors to approximately 1x10^8 cells/ml. Cells were homogenized on ice, nuclei and other debris were removed, and the membrane fraction was collected by centrifugation at 100,000xg for 1 hour at 4°C. The membrane pellet was washed in cTBS (18 mM Tris/HCl pH 7.6, 0.7 mM $MgCl_2$, 0.18 mM $CaCl_2$), and solubilized in cTBS with

1% CHAPS. After centrifugation, the supernatant was stored at -20°C.

Western blotting

Western blotting of protein samples and immunostaining was performed as described (Rombout *et al.* 1993), using the rabbit sera (1:200) and Goat-anti-Rabbit (GAR) second antibody (Ab) conjugated to alkaline phosphatase (AP) (1:3000; Biorad).

Isolation of carp cells and flow cytometry

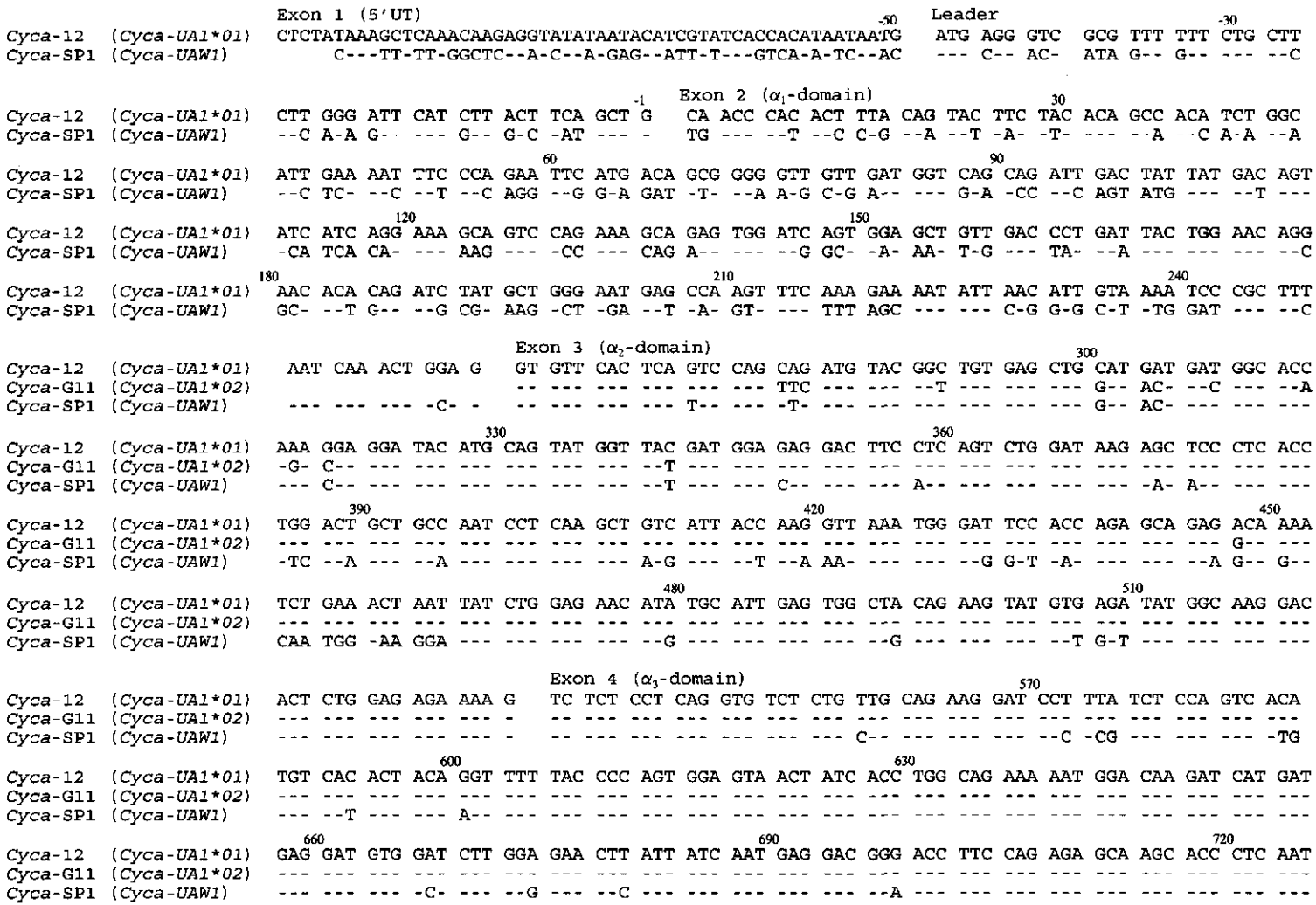
Peripheral blood leukocytes (PBL) were isolated according to Komen and co-workers (1990), and lymphocyte-enriched cell suspensions of various organs were isolated as previously described (Koumans-Van Diepen *et al.* 1994), using, however, a discontinuous Percoll gradient of densities 1.06 and 1.07 g/cm^3. 10^6 cells were single and double labeled using the polyclonal antiserum anti-*Cyca*12 (1:200) in combination with either monoclonal antibody (mAb) WCI12, recognizing carp surface immunoglobulin (sIg) (Secombes *et al.* 1983), or mAb WCL6, which specifically reacts with carp thrombocytes (Rombout *et al.* 1994). Labeling procedures were essentially as previously described (Rombout *et al.* 1993), using Goat-anti-Mouse (GAM) conjugated to fluorescein isothiocyanate (FITC) (1:200) and GAR conjugated to phycoerythrin (PE, 1:200) (DAKO, Glostrup, Denmark), as second Ab. Flowcytometry was carried out as described by Koumans-Van Diepen and co-workers (1995).

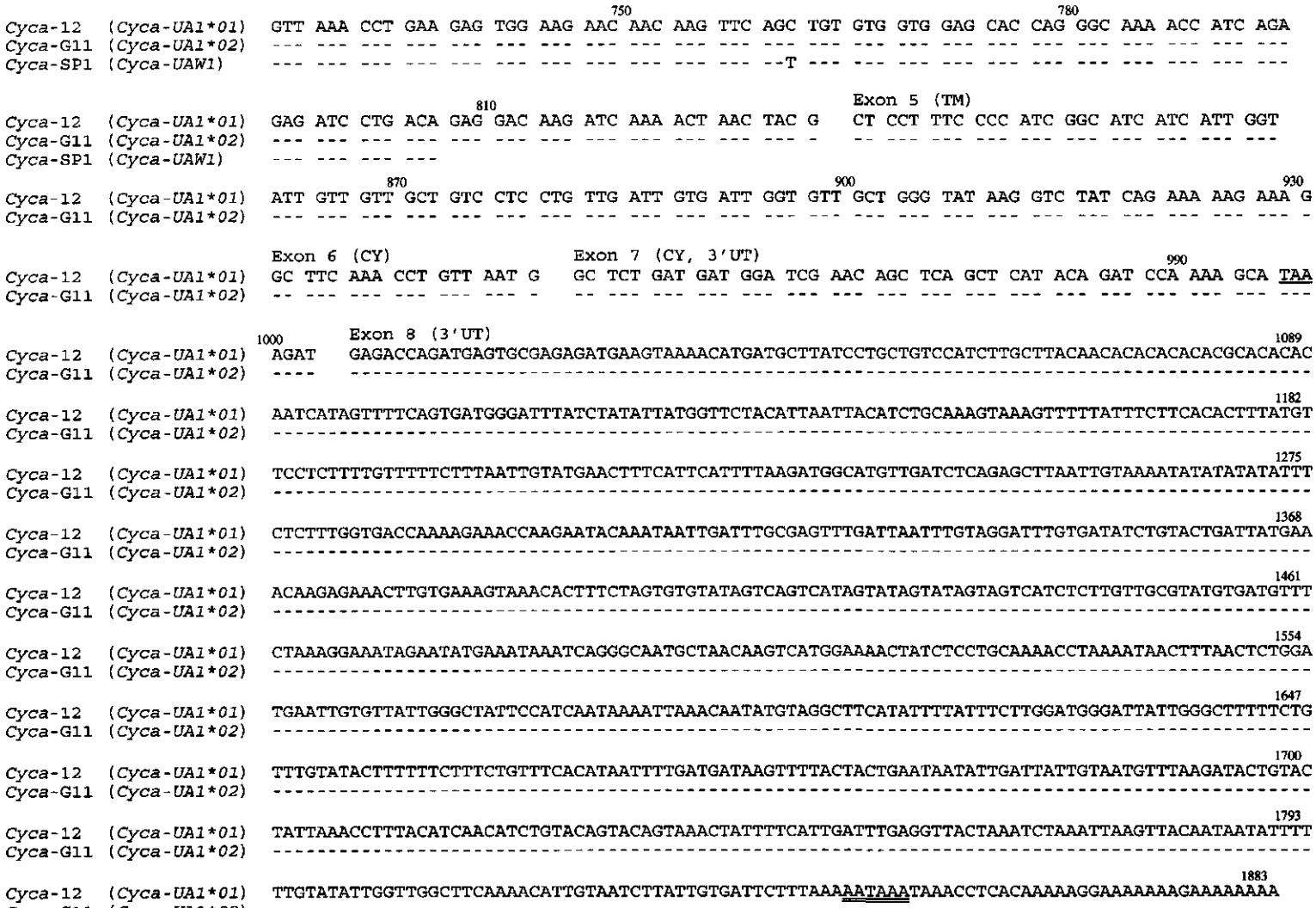
Results

Isolation of class I cDNA sequences from carp

In order to isolate class I sequences from carp we used a full-length class I cDNA probe U1 from zebrafish (*Brachydanio rerio*; Takeuchi *et al.* 1995). A thymus cDNA-library prepared from ten individuals of a homozygous clonal carp line A410 was screened with U1, which yielded a large number of positive plaques (0.01% of the PFU). Positive plaques (n=20) were re-screened, upon which 15 remained positive. Restriction mapping of the inserts of these positive clones indicated the presence of identical sequences. This was also confirmed by partial sequencing of nine of these positive clones, and consequently, only one of the cDNA clones, *Cyca*-12, was completely sequenced (Fig. 1).

The sequence encompasses 2026 basepairs (bp), consisting of 1050 bp of coding sequence, flanked by 51 bp and 925 bp of 5' and 3' untranslated sequence (5'UT, 3'UT), respectively. The start site of translation, however, is ambiguous because of the presence of two adjacent


Exon 1 (5'UT)
Leader
Exon 2 (α_1-domain)
Exon 3 (α_2-domain)
Exon 4 (α_3-domain)
Cyca-12 (Cyca-UA1*01) CTCTATAAAGCTCAAACAAGAGGTATATAATACATCGTATCACCACATAATAATG ATG AGG GTC GCG TTT TTT CTG CTT
Cyca-SP1 (Cyca-UAW1) C---TT-TT-GGCTC--A-C--A-GAG--ATT-T---GTCA-A-TC--AC --- C-- AC- ATA G-- G-- --- --C
Cyca-12 (Cyca-UA1*01) CTT GGG ATT CAT CTT ACT TCA GCT G CA ACC CAC ACT TTA CAG TAC TTC TAC ACA GCC ACA TCT GGC
Cyca-SP1 (Cyca-UAW1) --C A-A G-- --- G-- G-C -AT --- - TG --- --T --C C-G --A --T -A- -T- --- --A --C A-A --A
Cyca-12 (Cyca-UA1*01) ATT GAA AAT TTC CCA GAA TTC ATG ACA GCG GGG GTT GTT GAT GGT CAG CAG ATT GAC TAT TAT GAC AGT
Cyca-SP1 (Cyca-UAW1) --C TC- --C --T --C AGG --G G-A GAT -T- --A A-G C-G A-- --- G-A -CC --C AGT ATG --- --T ---
Cyca-12 (Cyca-UA1*01) ATC ATC AGG AAA GCA GTC CAG AAA GCA GAG TGG ATC AGT GGA GCT GTT GAC CCT GAT TAC TGG AAC AGG
Cyca-SP1 (Cyca-UAW1) -CA TCA CA- --- AAG --- -CC --- CAG A-- --- --G GC- -A- AA- T-G --- TA- --A --- --- --- --C
Cyca-12 (Cyca-UA1*01) AAC ACA CAG ATC TAT GCT GGG AAT GAG CCA AGT TTC AAA GAA AAT ATT AAC ATT GTA AAA TCC CGC TTT
Cyca-SP1 (Cyca-UAW1) GC- --T G-- --G CG- AAG -CT -GA --T -A- GT- --- TTT AGC --- --- C-G G-G C-T -TG GAT --- --C
Cyca-12 (Cyca-UA1*01) AAT CAA ACT GGA G GT GTT CAC TCA GTC CAG CAG ATG TAC GGC TGT GAG CTG CAT GAT GAT GGC ACC
Cyca-G11 (Cyca-UA1*02) -- --- --- --- --- --- --- TTC --- --- --T --- --- --- G-- AC- --C --- --A
Cyca-SP1 (Cyca-UAW1) --- --- --- -C- - -- --- --- --- T-- --- -T- --- --- --- --- --- --- G-- AC- --- --- ---
Cyca-12 (Cyca-UA1*01) AAA GGA GGA TAC ATG CAG TAT GGT TAC GAT GGA GAG GAC TTC CTC AGT CTG GAT AAG AGC TCC CTC ACC
Cyca-G11 (Cyca-UA1*02) -G- C-- --- --- --- --- --- --- --- --T --- --- --- --- --- --- --- --- --- --- --- --- ---
Cyca-SP1 (Cyca-UAW1) --- C-- --- --- --- --- --- --- --- --T --- --- C-- --- --- A-- --- --- --- --- -A- A-- --- ---
Cyca-12 (Cyca-UA1*01) TGG ACT GCT GCC AAT CCT CAA GCT GTC ATT ACC AAG GTT AAA TGG GAT TCC ACC AGA GCA GAG ACA AAA
Cyca-G11 (Cyca-UA1*02) --- G-- ---
Cyca-SP1 (Cyca-UAW1) -TC --A --- --A --- --- --- --- A-G --- --T --A AA- --- --- --G G-T -A- --- --- --A G-- G--
Cyca-12 (Cyca-UA1*01) TCT GAA ACT AAT TAT CTG GAG AAC ATA TGC ATT GAG TGG CTA CAG AAG TAT GTG AGA TAT GGC AAG GAC
Cyca-G11 (Cyca-UA1*02) ---
Cyca-SP1 (Cyca-UAW1) CAA TGG -AA GGA --- --- --- --- --G --- --- --- --- --G --- --- --- --T G-T --- --- --- ---
Cyca-12 (Cyca-UA1*01) ACT CTG GAG AGA AAA G TC TCT CCT CAG GTG TCT CTG TTG CAG AAG GAT CCT TTA TCT CCA GTC ACA
Cyca-G11 (Cyca-UA1*02) --- --- --- --- --- - -- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Cyca-SP1 (Cyca-UAW1) --- --- --- --- --- - -- --- --- --- --- --- --- C-- --- --- --- --C -CG --- --- --- -TG
Cyca-12 (Cyca-UA1*01) TGT CAC ACT ACA GGT TTT TAC CCC AGT GGA GTA ACT ATC ACC TGG CAG AAA AAT GGA CAA GAT CAT GAT
Cyca-G11 (Cyca-UA1*02) ---
Cyca-SP1 (Cyca-UAW1) --- --T --- --- A-- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---
Cyca-12 (Cyca-UA1*01) GAG GAT GTG GAT CTT GGA GAA CTT ATT ATC AAT GAG GAC GGG ACC TTC CAG AGA GCA AGC ACC CTC AAT
Cyca-G11 (Cyca-UA1*02) ---
Cyca-SP1 (Cyca-UAW1) --- --- --- -C- --- --G --- --C --- --- --- --- --- --A --- --- --- --- --- --- --- --- ---


Cyca-12 (Cyca-UA1*01) GTT AAA CCT GAA GAG TGG AAG AAC AAC AAG TTC AGC TGT GTG GTG GAG CAC CAG GGC AAA ACC ATC AGA
Cyca-G11 (Cyca-UA1*02)
Cyca-SP1 (Cyca-UAW1)
Exon 5 (TM)
Cyca-12 (Cyca-UA1*01) GAG ATC CTG ACA GAG GAC AAG ATC AAA ACT AAC TAC G CT CCT TTC CCC ATC GGC ATC ATC ATT GGT
Cyca-G11 (Cyca-UA1*02)
Cyca-SP1 (Cyca-UAW1)
Cyca-12 (Cyca-UA1*01) ATT GTT GTT GCT GTC CTC CTG TTG ATT GTG ATT GGT GTT GCT GGG TAT AAG GTC TAT CAG AAA AAG AAA G
Cyca-G11 (Cyca-UA1*02)
Exon 6 (CY)
Exon 7 (CY, 3'UT)
Cyca-12 (Cyca-UA1*01) GC TTC AAA CCT GTT AAT G GC TCT GAT GAT GGA TCG AAC AGC TCA GCT CAT ACA GAT CCA AAA GCA TAA
Cyca-G11 (Cyca-UA1*02)
Exon 8 (3'UT)
Cyca-12 (Cyca-UA1*01) AGAT GAGACCAGATGAGTGCGAGAGATGAAGTAAAACATGATGCTTATCCTGCTGTCCATCTTGCTTACAACACACACACACGCACACAC
Cyca-G11 (Cyca-UA1*02)
Cyca-12 (Cyca-UA1*01) AATCATAGTTTTCAGTGATGGGATTTATCTATATTATGGTTCTACATTAATTACATCTGCAAAGTAAAGTTTTTATTTCTTCACACTTTATGT
Cyca-G11 (Cyca-UA1*02)
Cyca-12 (Cyca-UA1*01) TCCTCTTTTGTTTTTCTTTAATTGTATGAACTTTCATTCATTTTAAGATGGCATGTTGATCTCAGAGCTTAATTGTAAAATATATATATTTT
Cyca-G11 (Cyca-UA1*02)
Cyca-12 (Cyca-UA1*01) CTCTTTGGTGACCAAAAGAAACCAAGAATACAAATAATTGATTTGCGAGTTTGATTAATTTGTAGGATTTGTGATATCTGTACTGATTATGAA
Cyca-G11 (Cyca-UA1*02)
Cyca-12 (Cyca-UA1*01) ACAAGAGAAACTTGTGAAAGTAAACACTTTCTAGTGTGTATAGTCAGTCATAGTATAGTATAGTAGTCATCTCTTGTTGCGTATGTGATGTTT
Cyca-G11 (Cyca-UA1*02)
Cyca-12 (Cyca-UA1*01) CTAAAGGAAATAGAATATGAAATAAATCAGGGCAATGCTAACAAGTCATGGAAAACTATCTCCTGCAAAACCTAAAATAACTTTAACTCTGGA
Cyca-G11 (Cyca-UA1*02)
Cyca-12 (Cyca-UA1*01) TGAATTGTGTTATTGGGCTATTCCATCAATAAAATTAAACAATATGTAGGCTTCATATTTTATTTCTTGGATGGGATTATTGGGCTTTTTCTG
Cyca-G11 (Cyca-UA1*02)
Cyca-12 (Cyca-UA1*01) TTTGTATACTTTTTTCTTTCTGTTTCACATAATTTTGATGATAAGTTTTACTACTGAATAATATTGATTATTGTAATGTTTAAGATACTGTAC
Cyca-G11 (Cyca-UA1*02)
Cyca-12 (Cyca-UA1*01) TATTAAACCTTTACATCAACATCTGTACAGTACAGTAAACTATTTTCATTGATTTGAGGTTACTAAATCTAAATTAAGTTACAATAATATTTT
Cyca-G11 (Cyca-UA1*02)
Cyca-12 (Cyca-UA1*01) TTGTATATTGGTTGGCTTCAAAACATTGTAATCTTATTGTGATTCTTTAAAAATAAATAAACCTCACAAAAAGGAAAAAAAGAAAAAAAA
Cyca-G11 (Cyca-UA1*02)

ATG codons. Based on the signal cleavage site prediction method of Von Heijne (1986), the coding region can be divided into 51 bp encoding 17 amino acid residues of the leader peptide (under the assumption that the first ATG is used as the start codon), and 999 bp specifying 332 amino acids of the mature protein. The derived amino acid sequence of the mature protein can be divided into protein domains inferred from alignments with other class I sequences (Fig. 2). These alignments may demonstrate conservation in *Cyca*-12 of structural and functional features of class I molecules.

Binding of antigenic peptides by class I molecules has been found to be dependent, among other things, on the presence of eight extremely conserved residues in pockets A and F of the peptide-binding groove, which bind the amino- and carboxyl-termini of the peptide (Kaufman *et al.* 1994). In *Cyca*-12 these residues (Y7, Y57, Y156, Y168, R82, T139, K142, W143) are conserved, except that tyrosine at position 82 is replaced by arginine. Three known contact pairs between β_2m and class I α_1/α_2 domains (HLA positions H_β31-Q96, W_β60-D122, W_β60-Q96), and one contact pair between β_2m and α_3 (S_β11-Q242) are found to be conserved in *Cyca-B2m* (Dixon *et al.* 1993) and *Cyca*-12 (*Cyca* pos. H_β31-Q93, W_β60-D118, W_β60-Q93 and S_β11-Q236), by comparison with the class I and β_2m sequences from humans and mice (Fremont *et al.* 1992). In antigen presentation to $CD8^+$ T cells, CD8 acts as a co-receptor by binding to the α_3 domain of class I. An exposed loop with a high content of acidic residues in the α_3 domain of *HLA-B27* (HLA residues 223-229) was found to be a major CD8-binding site (Salter *et al.* 1990; Kaufman *et al.* 1994). The acidic nature of this functionally important region is well conserved in classical class I molecules, and the homologous stretch in *Cyca*-12 (positions 217-223, Fig. 2) also contains many acidic residues (D217, D219, E220, D221, D223).

In addition many other residues are found that are conserved, with only a few exceptions, in the sequences of mammals, amphibia, reptiles and fish. In the α_1 and α_2 domain these include C98 and C161, presumably forming a disulfide bridge, pairs of residues involved in the formation of salt-bridges (D28/H3, H90/D115) (Saper *et al.* 1991), and the potential N-linked glycosylation site N84. The α_3 domain contains many residues characteristic of the Ig-fold (Williams and Barclay 1988). The cytoplasmic region of *Cyca*-12 contains a highly conserved serine at position 325, the homologous site of which is phosphorylated *in vivo* in HLA molecules

Figure 1 (previous pages). *Mhc* class I nucleotide sequences of *Cyprinus carpio* L. *Cyca*-12 and *Cyca*-SP1 are derived from cDNA, *Cyca*-G11 from a genomic clone. Numbers above the sequence indicate nucleotide positions within the mature protein. The stop codon and poly-adenylation signal are singly and doubly underlined, respectively. Dashes denote identity to *Cyca*-12.

(Guild and Strominger 1984). This serine is invariant in the class I sequences compared, with the exception of the sequence from Atlantic salmon. Overall, the sequence of *Cyca*-12 contains many conserved characteristics that are found to be functionally or structurally important for class I molecules.

Genomic organization

A commercially obtained genomic carp library was screened with a probe encompassing basepairs 60 - 1923 of the *Cyca*-12 cDNA sequence. First screening yielded 45 positive plaques, 25 of which were re-screened, upon which 17 were shown to be weakly positive, whereas three remained strongly positive. The inserts of the three strongly hybridizing clones were mapped by restriction analysis, which indicated overlapping inserts. However, from hybridization experiments it was concluded that none of the three genomic clones contained the 5' region of the gene, and only one, designated *Cyca*-G11, contained both the α_2 domain and the 3' end of the gene, whereas the other two did not contain the former. Therefore, only *Cyca*-G11 was chosen for partial sequence analysis (Fig. 3-I).

The insert of *Cyca*-G11 had a length of approximately 15 kilobasepair (kb), and restriction mapping and hybridization experiments showed that it contained a *Cyca*-12-like gene, covering 2.7 kb of intron 2 to the end of the gene. The sequence of *Cyca*-G11 proved to be identical to *Cyca*-12 in all identified exons, except for 14 nucleotide substitutions in exon 3.

Alignment with the cDNA clone *Cyca*-12 revealed that the α_2 and α_3 domains and the connecting peptide/transmembrane region were encoded in separate exons, provisionally designated exons 3, 4 and 5. Exon 6 encoded the first six amino acids of the cytoplasmic region, while exon 7 encoded the rest of this region, the stop codon and 4 bp of the 3' UT. Finally, exon 8 encoded the remainder of the 3'UT. The introns (3-7) were shown to start with GT and end with AG, and they were all phase 1 (*i.e.*, the codon is split between the first and second base).

None of the three obtained genomic clones, however, contained the 5' end of the gene, encompassing the exons encoding the 5'UT, leader peptide and α_1 domain, and further screening failed to produce clones containing the 5' end of the gene. Therefore we decided to unravel the genomic organization of *Cyca*-12 by using PCR on genomic DNA from carp clone A410. Because in this carp clone only a single cDNA sequence was found in the thymus library of ten individuals, this clone can be considered to be homozygous for this gene. Hence, amplification of allelic variants was expected to be unlikely. To make sure, however, that the amplified products could be properly identified as fragments of the gene encoding *Cyca*-12, pairs of specific primers were designed, based on the sequence of *Cyca*-12, each spanning an intron with a substantial part of the two flanking exons (Fig. 3-II). In addition, of each


LEADER
Cyca-12 MRVLAFFLLGIHLTSA
Cyca-G11 ....
Cyca-SP1 --TIVVL--RV-VAY-
Brre-UBA*01 MQSLIGLLLVVCLQYA-G
Sasa-p30 MKGFIL-VLGIGLL-TA--
Cyca-ZA1 ....
Lach-UB*01 ....
Xela-UAA1f MDLRL-PILLT-W-SAVYS
Xela-XNC1.1 MAFLFLL-S-ETSAVYC
Amam-LC1 LATSWXDLGXAET-RGS
Gaga-B-F12 LDLRPAAVERPEFRDLRVQRCEAMGPCGALGLGLLLAAVCGAAAP
HLA-A2 MAVMAPRTLVL-LSGA-ALTQ-V-
ALPHA-1
(1) (10) (20) (30) (40) (50) (60) (70) (80) (90)
1 10 20 30 40 50 60 70 80
Cyca-12 ATHTLQYFYTATSGIE*NFPEFMTAGVVDGQQIDYYDS**IIRKAVQKAEWISGAV***DPDYWNRNTQIYAGNEPSF*KENINIVKSRFNQTG*
Cyca-G11 ....
Cyca-SP1 V------YF---T--S*---RLVDV-MLN-EP-SM---**TSQ-K-P-QK-MAENL***-YE---SA-EMRKARDQV-*FS--QVLMD-----A*
Brre-UBA*01 ---SWKAY--G-T-LT*E----VALNLI-D-LMG-F--**KTNRFKSQFQ-MEDNL***GKE-DEQQ-N-LL-YPEV-*-N--KV-ME-----Q*
Sasa-p30 V--A-K-----S-EVP*-----VVV-----V-MVH---**NSQR--P-QD-VNK-A***--Q--E---G-FK-SQQT-*-A--D-A-Q----S-*
Cyca-ZA1 *K-Y-R-MF-VMLKAD*TL-V-SAV-ES-HI--SH-SI******EE**QV-MRENL***TE-D-DQAPEGPP-TRDWY*LDL-K-LSN CTESS*
Lach-UB*01 .G-IWKEYI-F-KETD*GHLKI-AVDM--DV-TA---D**M-SRH-L-QP-MEELGRVE--NS-ASKATEFLLGNKE-N-GV-GRL-KET-HSS*
Xela-UAA1f GS-S-R-Y---V-DRAFGL---S-V-Y--DT--ER-S-**DTGRDEPATQ-MKQKE***G-E--E-E--KSK-I-AT-*-H-VKVAMD----ST*
Xela-XNC1.1 RS---HFHI-LA-APIPGLSQYAIIVYM--V-YGR-N-**D--R-Q**FYSA-LNP***LSV-LDMQ-KFAQTF-VWQ*RHRL-FLMGV--T-NG
Amam-LC1 PS-S------GV-EPGQGL-Q-IVV-Y----LFVQ---**NT-EMLPRVS--KDNE***-SK--EGQ--NLQ-A--V-*RG---TAMN-Y----*
Gaga-B-F12 EL---R-IQ--MTDPGPGQ-W-V-V-Y---ELFVH-N-**TA-RY-PRT---AAKA***-QQ--DGQ---GQ---QID*R--LG-LQR-Y----*
HLA-A2 GS-SMR--F-SV-RPGRGE-R-IAV-Y--DT-FVRF--DAASQRMEPR-P--EQE****G-E--D-E-RKVKAHSQTH*RVDLGTLRGYY--SEA
ALPHA-2
(100) (110) (120) (130) (140) (150) (160) (170) (180)
90 100 110 120 130 140 150 160 170
Cyca-12 GV**HSVQQMYGCELH***DDGTK*GGYMQYGYDGEDFLSLDKSSLTWTAANPQAVITKVKWDSTR*AETKSETNYLENICIEWLQKYVRYGKDTLERK
Cyca-G11 --**----F------D***T---R*R----------*--A--Q----------
Cyca-SP1 --**--F-L------D***T----*R---------Q--I----NT--F-------M---N--EAN-*--AEQWKG----M---------G---------
Brre-UBA*01 --**-TF-F-----MD***---N-*QVHW-I-------I----KT-------S--MT--------G*--ANYWKG----E----V----G---------
Sasa-p30 --**-VN-W-----WD***-EAGVTE-FE-W-------IAF-LKTKS-I-PTHRS----L----DT*-QNEHRK--YTQ------K---D---S--M-T
Cyca-ZA1 EL**YVL-RLI----EKLP-GTVNLTPVDE--FG---CVAFNSDT-QGSDKS-NVKE-EI-R-****HQV-LQA**FLKN-LD-ISTFNN******TK-
Lach-UB*01 --**FIY---H----N***--E-T*TVLWKTSFQ----S-F-GDRNLY-P-AGK-Q-RSKVCPVDE*-WDSYFSEDWKQE---G-K--LL---E-----
Xela-UAA1f -T**-M--W------G***---SI*R--E-HV---RE-FA--TEEWVYVPSVRE-QL-TQ--N-PEVNAPERNK---Q-----G-KR-LS--QAE---R
Xela-XNC1.1 NGHI-VY-RKSA----***----I*---QDIAF--KE--VF--ERAVYITSTQE--MVSHL-NKR**YDSTDSKMF--ID--KHMKM-LP-ISND--K-
Amam-LC1 -L**-TW-W------R***G--S-*---S-F----R--VA---ET------DSE-QV--S---ALG*-MNQGRKF---KF--------L---NK--L--
Gaga-B-F12 -S**-T--W----DIL***EG-PI*R--Y-MA---R--TAF--GTM-F---V-E--P--R--EEES**-PERWK----ET-V---RR--E---AE-G-R
HLA-A2 -S**-T--R----DVG***S-WRFLR--H--A---K-YIA-KEDLRS----DMA-QT--H--EAAHV--*QLRA*---GT-V---RR-LEN--E--Q-T

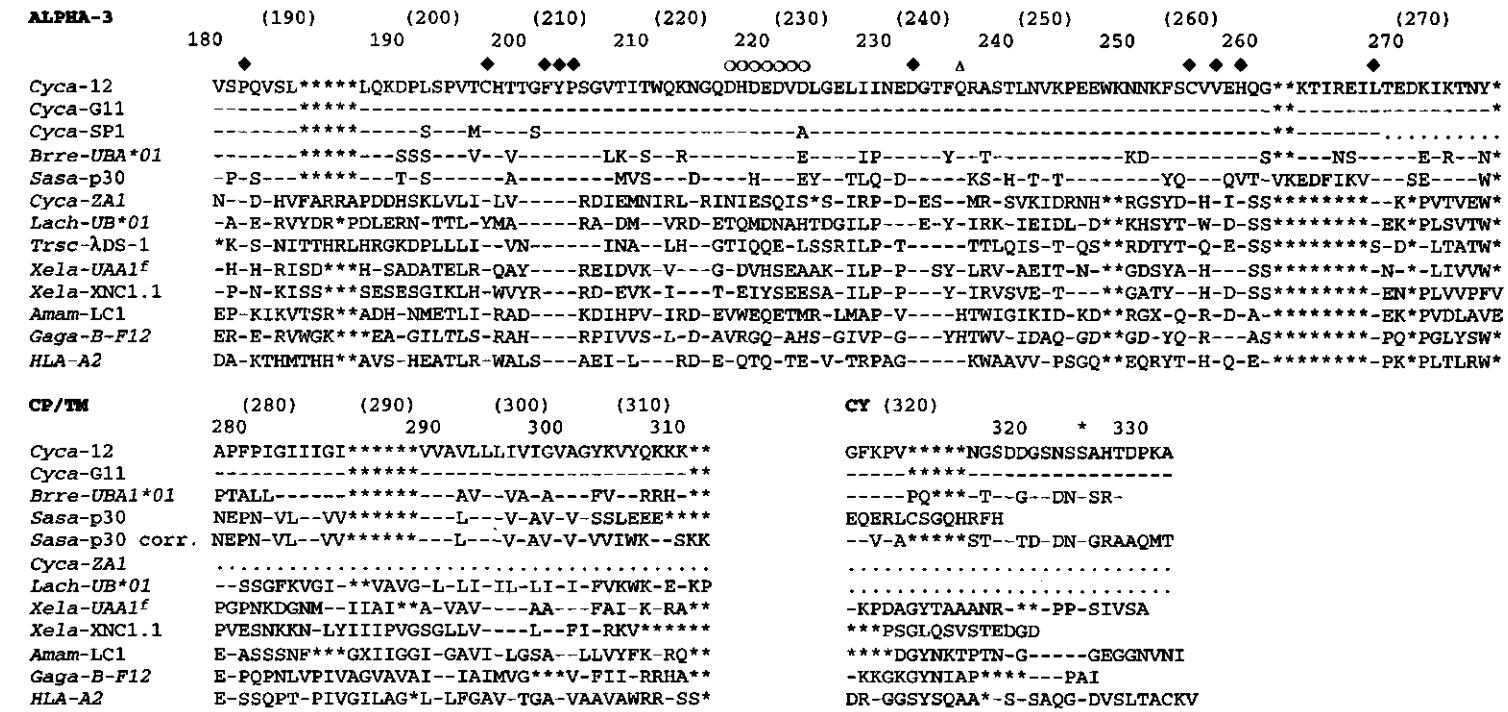


Figure 2. Comparison of carp *Cyca*-12/SP1/G11 amino acid sequences with *Mhc* class I sequences from other vertebrates. *Brre-UBA*01* (U1), zebrafish (Takeuchi *et al.* 1995); *Sasa*-p30, Atlantic salmon (Grimholt *et al.* 1993); *Sasa*-p30 corr., corrected sequence, frame shifted by insertion of two nucleotides (U. Grimholt personal communication); *Cyca-ZA1*, common carp (Okamura *et al.* 1993); *Lach-UB*01*, coelacanth (Betz *et al.* 1994); *Trsc*-λDS-1, nurse shark (Hashimoto *et al.* 1992); *Xela-UAA1*^{f}, *X. laevis*, classical gene (Shum *et al.* 1993); *Xela*-XNC1.1, *X. laevis*, non-classical (Flajnik *et al.* 1991); *Amam*-LC1, lizard (Grossberger and Parham 1992); *Gaga-B-F12*, domestic fowl (Guillemot *et al.* 1988); *HLA-A2*, human (Koller and Orr 1985). Numbering is based on the *Cyca*-12 sequence, for comparison *HLA-A2* numbering is shown in parentheses. Dashes indicate identity to *Cyca*-12, dots denote absence of sequence information. Flowers indicate residues binding peptide termini, the spade indicates a potential N-linked glycosylation site, triangles indicate conserved $ß_2$m-contact residues, circles designate position of CD8-binding loop in *HLA-A2*, diamonds indicate other highly conserved residues, and the asterisk indicates a potential kinase site.

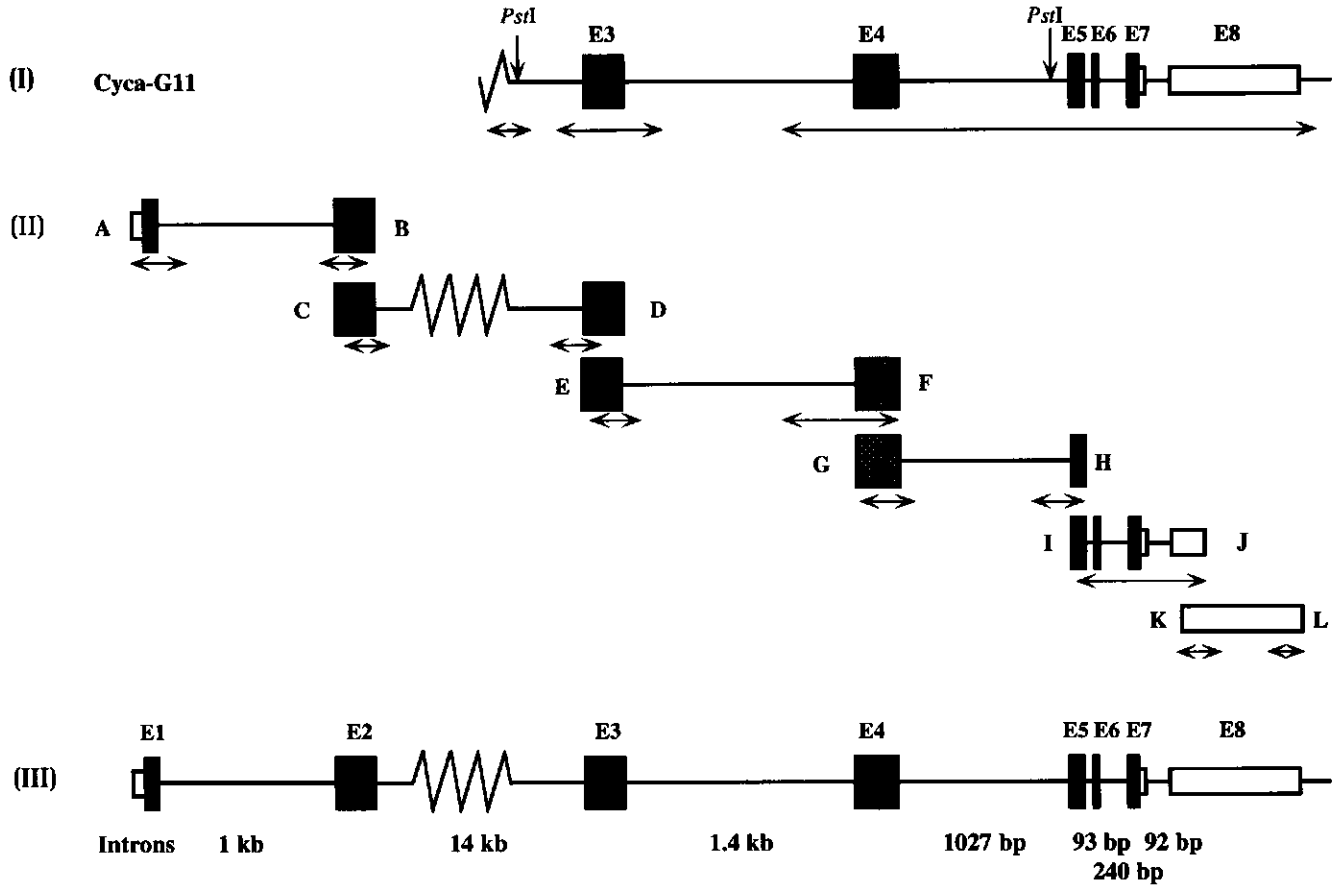


Figure 3. Genomic organization of *Cyca*-G11 and *Cyca*-12: (I) genomic clone *Cyca*-G11; (II) PCR amplified *Cyca*-12 gene fragments, A-B, C-D, E-F, G-H, I-J, K-L; (III) genomic organization of *Cyca*-12, deduced from (II). Boxes indicate exons 1 to 8 (denoted E1-E8) (solid boxes: coding regions; open boxes: 5' and 3' untranslated regions). Horizontal lines indicate introns; intron sizes are denoted in (III). Horizontal arrows indicate stretches for which sequence was determined.

amplified fragment multiple recombinant clones were sequenced, to identify any co-amplified sequence variants of *Cyca*-12.

Each of the amplifications designated A-B, C-D, E-F, G-H, I-J and K-L (see fig. 3-II), yielded a single band. Fragment C-D was amplified using a long-template PCR system, and because of its size, this fragment was not cloned in one piece. Each of the bands was shown to contain only a single sequence by sequencing multiple recombinant clones. In all cases the exon sequences were identical to the *Cyca*-12 cDNA sequence. In addition, the partial sequences of introns 3 to 7 (fragments E-F to I-J) were found to be identical to those found in *Cyca*-G11, and the intron sizes, inferred from electrophoretic separation, corresponded perfectly to those identified in *Cyca*-G11. The size of PCR product K-L confirmed the absence of further introns in the 3'UT region, as found earlier in *Cyca*-G11. As previously found for introns 3 to 7 in genomic clone *Cyca*-G11, introns 1 and 2 (contained by fragments A-B and C-D) were now found to be phase 1 as well, starting with GT and ending with AG.

Gene multiplicity and polymorphism

To determine the number of related genes present in the genome, we hybridized *Pst*I-digested genomic DNA from several carp strains with a cDNA probe containing exons 3 and 4. In all cases multiple bands were observed (Fig. 4A). In the DNA from clone A410 (Fig. 4A, lane 4) the major band of approximately 4.5 kb corresponds to the *Pst*I-fragment found in the genomic clone *Cyca*-G11, spanning part of intron 2, exon 3, intron 3, exon 4 and part of intron 4 (Fig. 3-I). The six less strongly hybridizing bands probably derive from related loci, as strain A410 is presumably homozygous for *Cyca*-12. Comparison with DNA obtained from carp strains with expected lower levels of inbreeding (Fig. 4A, lanes 1-3) showed that the major 4.5 kb band is invariably present in all of these strains. Of the minor bands detected only two (approximately 1.1 kb and 4.7 kb) are shared by all strains, whereas the remaining bands are polymorphic.

To analyze polymorphism of the exon 2 sequences among the strains used, we hybridized the same southern blot to a probe encompassing only exon 1 and the first 129 nucleotides of exon 2. Only in the individuals from strains R3xR8 and A410 (Fig. 4B, lanes 2 and 4) a hybridizing *Pst*I fragment of approximately 15 kb was observed. In contrast, no hybridization could be detected in the other strains, not even when lower stringency conditions were applied.

To obtain more insight in the diversity of this gene, we analyzed the presence of *Cyca*-12-related sequences in a λgt11 cDNA library from spleen/pronephros prepared from six individuals of a 2nd generation meiotic gynogenetic family (A4.10me2). To this end, a fragment spanning exons 1 to 4 was amplified using anchored PCR with antisense primer sve9404 located at positions 808-824 in exon 4 in combination with λgt11 primers. A single band with the expected

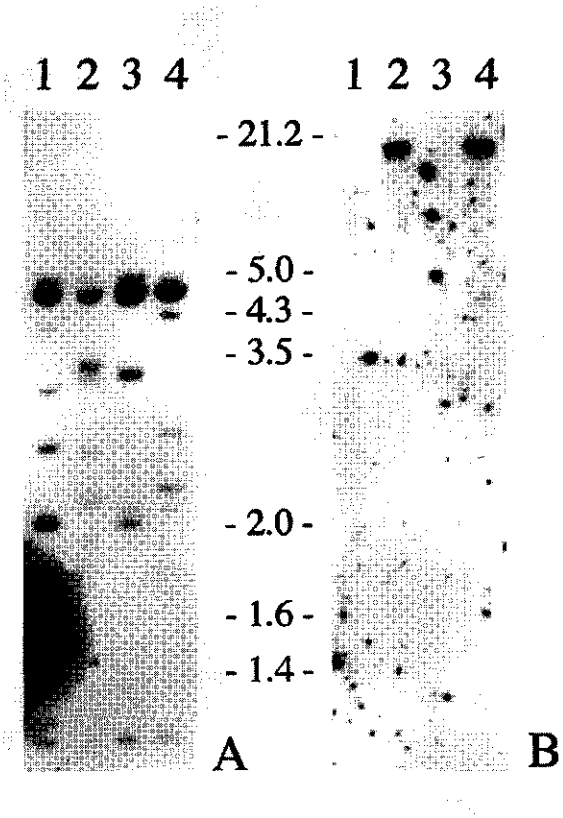


Figure 4. Southern blot analysis of *Pst*I-digested genomic carp DNA, **(A)** hybridized with a cDNA probe spanning exon 3 and 4 of cDNA clone *Cyca*-12, and **(B)** hybridized with a probe spanning exon 1 and 129 bp of exon 2 from cDNA clone *Cyca*-12. Lane 1: WK; lane 2: F1 individual R3 x R8; lane 3: W11.49; lane 4: clone A410. Size markers are indicated between panels A and B.

size of 900 bp was cloned and sequenced. Analysis of 13 recombinant clones resulted in only a single sequence, *Cyca*-SP1 (Figs. 1 and 2). Sequence comparison with the sequence of *Cyca*-12 showed a high similarity in exon 4 (95% nucleotide (nt), and 96% amino acids (aa) identity), with markedly decreased degrees of similarity in exons 3, 2 and 1 of 84% nt (75% aa), 56% nt (43% aa) and 63% nt (38% aa) identity, respectively.

Detection of cell surface expression of *Cyca*-12

To study cell surface expression of the molecules encoded by the *Cyca*-12 gene, a polyclonal antiserum was raised against the extracellular domains of *Cyca*-12. To this end, the region of the *Cyca*-12 cDNA encoding these domains was cloned in frame into expression vector pRSET. The poly-histidine-containing recombinant fusion protein was isolated and used to immunize a rabbit. The collected immune serum (designated anti-*Cyca*12) was tested in western blot experiments, and was found to possess strong reactivity with the recombinant protein, as compared with the pre-immune serum (Fig. 5A). In addition, the immune serum detected a molecule of approximately M_r 45,000 on western blots of membrane lysates of spleen and muscle, which was not detected by the pre-immune serum (Fig. 5B). An additional protein of approximately M_r 84,000 was aspecifically detected by the immune and pre-immune serum in muscle, but only by the immune serum in spleen. Nevertheless we believe that also in spleen this protein is aspecifically detected, and that the reason for it not being detected by the pre-immune serum in spleen is the generally lower reactivity of the serum towards the membrane lysate of spleen.

The polyclonal serum anti-*Cyca*12 was then used to label erythrocytes, PBL, and lymphocyte fractions from spleen, pronephros and thymus, all derived from an A410 individual, followed by fluorescence-activated cell sorter (FACS)-analysis. Negative controls were reactions using either pre-immune serum, or the second antibody only. Labeling of erythrocytes with either immune- or pre-immune serum showed no difference in reactivity between the two sera, and only weak binding was observed. This justified absorption of the immune serum with red blood cells to remove the background staining. After absorption no reactivity of the serum with erythrocytes could be detected. In contrast, when the absorbed serum was used to label PBL and density-separated lymphocyte fractions from spleen, pronephros and thymus, substantial reactivity was still observed, identifying two populations, positive and negative. The density-separated fractions from spleen and pronephros showed essentially the same pattern of reaction as did PBL, although in both organs a higher fraction of the cells was anti-*Cyca*12 positive, as compared with PBL (Fig. 6A).

To gain further insight into the identity of the anti-*Cyca*12-positive cell population, double labelings were performed using the monoclonal antibody WCI12, recognizing carp surface immunoglobulin (sIg), and WCL6, reactive with thrombocytes. FACS analysis of PBL revealed that 69% of the cells were anti-*Cyca*12 brightly positive, and 31% were negative (all data given as percentages of the total number of cells). All WCI12$^+$ (sIg$^+$) cells were bright,

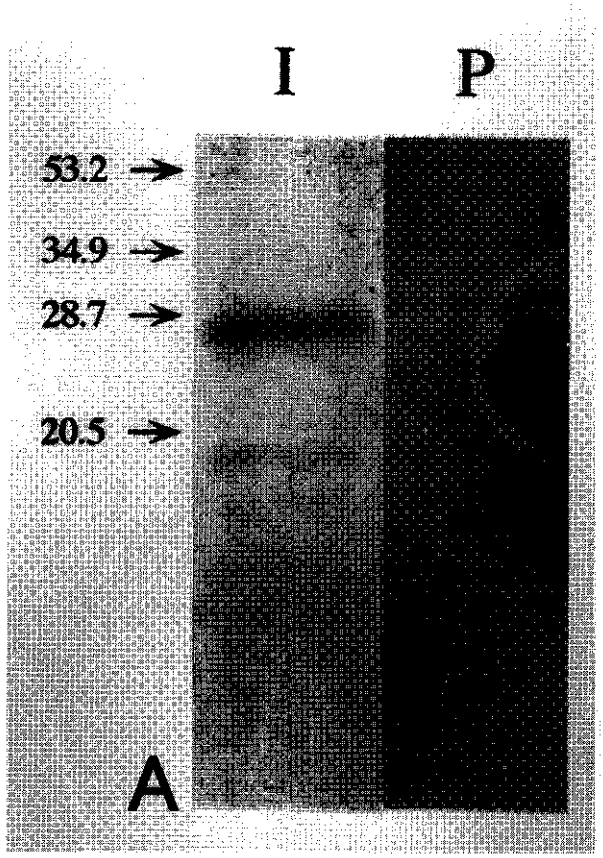
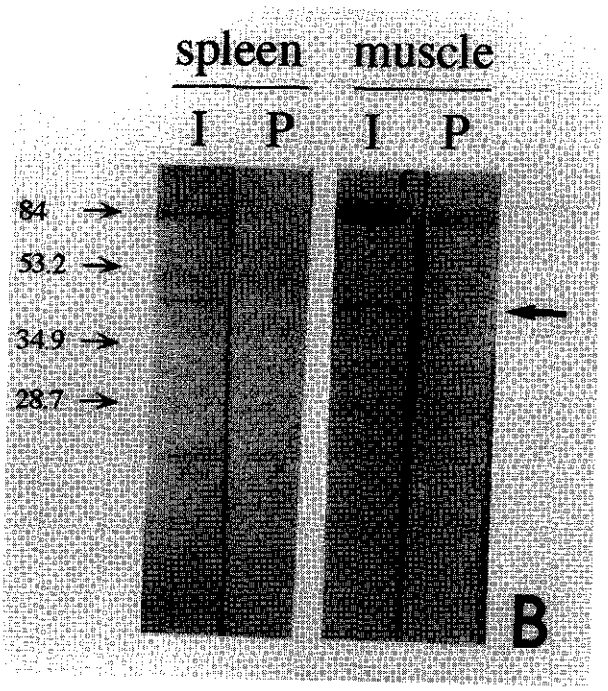


Figure 5. Western blot analysis of (**A**) SDS-PAGE-separated purified recombinant *Cyca*-12-$\alpha_1/\alpha_2/\alpha_3$ protein, and (**B, right page**) SDS-PAGE-separated membrane lysates of spleen and muscle. Lanes I stained with absorbed immune polyclonal rabbit serum anti-*Cyca*12, and lanes P stained with rabbit pre-immune serum. Arrows indicate molecular weight markers (kD).

while the sIg^- cells were present in two populations, anti-*Cyca*12 bright (21%) and negative (29%) (Fig 6B). In contrast, $WCL6^+$ cells in PBL were found almost entirely within the anti-*Cyca*12-negative population (Fig. 6C).

In spleen, 88% of the cells were recognized by anti-*Cyca*12, and all of the sIg^+ cells were found within this anti-*Cyca*12-bright population. The sIg^- cells consisted of anti-*Cyca*12-positive cells (64%) and anti-*Cyca*12-negative cells (11%). All $WCL6^+$ cells (thrombocytes) were present within the anti-*Cyca*12-negative population (data not shown). In pronephros, also



88% of the cells were anti-*Cyca*12 positive. Again all WCI12$^+$ cells were found within this anti-*Cyca*12-positive population. The remainder, WCI12$^-$ cells, consisted of two populations, anti-*Cyca*12 positive (77%) and negative (11%). In this organ, however, no WCL6$^+$ cells were detected (*data not shown*).

Double staining of thymocytes using anti-*Cyca*12 and WCI12 showed that the large sIg$^-$ thymocyte population was recognized by anti-*Cyca*12 as a single positive population, whereas only a limited number of cells were sIg$^+$. Of the latter, almost all cells were also brightly positive for the anti-*Cyca*12 serum (Fig. 6D).

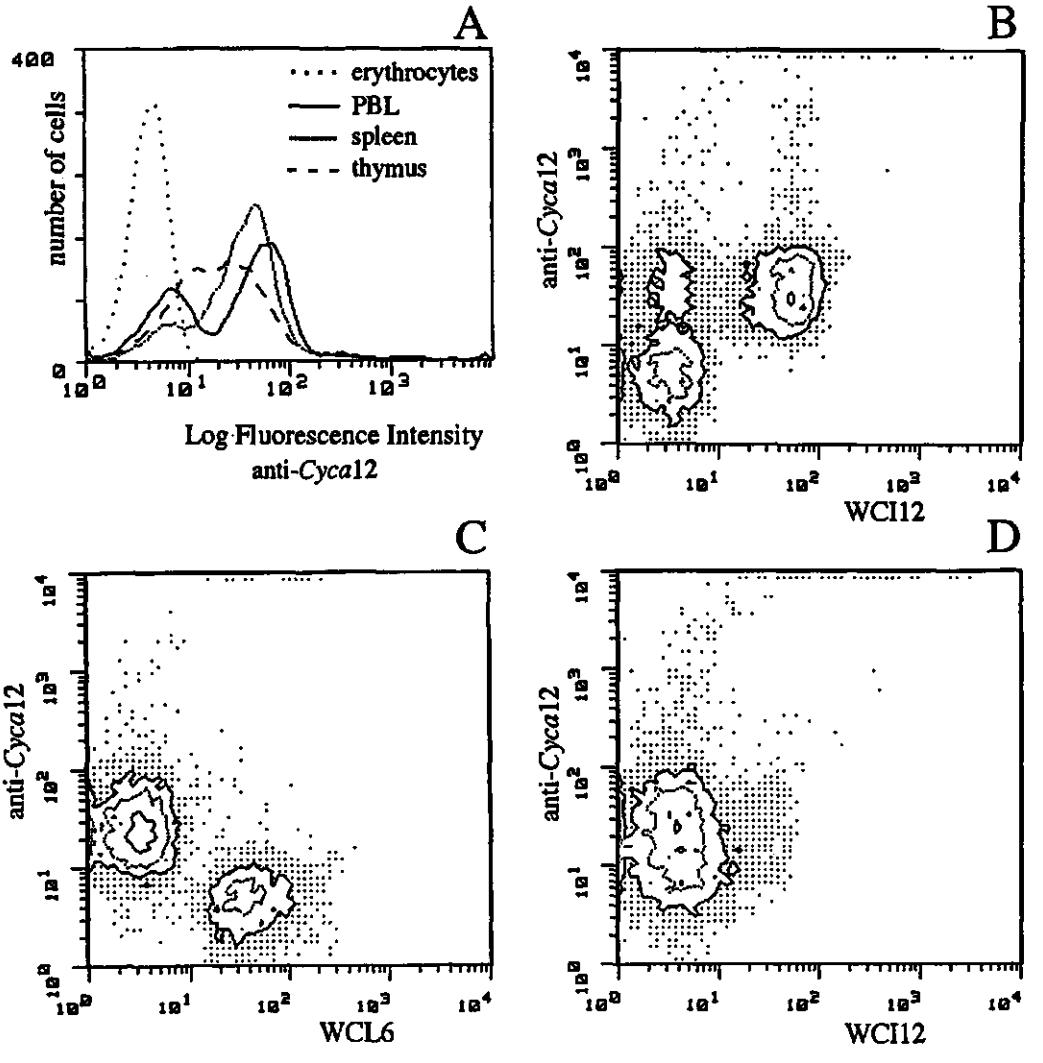


Figure 6. Flowcytometric analyses of anti-*Cyca*12-stained cells from various tissues. (A) Histogram of anti-*Cyca*12-stained erythrocytes, PBL, lymphocyte-enriched spleen cells, and lymphocyte-enriched thymocytes. (B) Contour graph of PBL, double stained with anti-*Cyca*12 and mAb WCI12. (C) Contour graph of PBL, double stained with anti-*Cyca*12 and mAb WCL6. (D) Contour graph of lymphocyte-enriched thymocytes, double stained with anti-*Cyca*12 and WCI12.

Discussion

Before the present study, two distinct types of class I-like sequences were reported to be present in the carp, the *Cyca-Z* genes (Okamura *et al.* 1993), and C4, a cDNA-derived PCR fragment (Grossberger and Parham 1992). Both types of sequences exhibit a low degree of conservation of functional features of class I molecules. Full-length cDNA sequences of neither *Cyca-Z* nor C4 have been reported, and data on expression are still lacking. In the present study a third type of class I sequences was identified in carp. *Cyca*-12, a full-length cDNA was obtained by screening a thymus cDNA library from carp with a heterologous class I cDNA probe, *Brre-UBA1*, from zebrafish. *Cyca*-SP1, a PCR fragment comprising exons 1 through 4, was amplified from a spleen/pronephros cDNA library. Finally, *Cyca*-G11, a genomic clone containing part of intron 2 to the end of a *Cyca*-12-like gene, was obtained by screening a commercial library with *Cyca*-12. The amino acid sequences of *Cyca*-12, *Cyca*-SP1 and *Cyca*-G11 show considerable conservation in regions that are involved in maintaining the function and structure of class I molecules (Fig. 2). The conserved peptide-binding residues (PBR) in the α_1 and α_2 domains are all present but one. The only substitution, arginine at position 82, where tyrosine is found in mammalian sequences, is, however, consistently present in classical class I sequences of cold-blooded vertebrates and chicken. It is noteworthy that, in contrast to the new sequences presented in this study, in the *Cyca-Z* sequences only Y7 in pocket A, and T143 and K146 in pocket F (*HLA-A2* numbering) are conserved. The region homologous to the CD8-binding loop in the α_3 domain has a conserved acidic nature, which suggests the presence of a carp homologue of CD8 with positively charged surface regions binding to the negatively charged loop of the class I α chain. Nevertheless, CD8 in fish has, as yet, not been reported. The four potential $ß_2$m-class I α contact pairs that are identical in the human, mouse and carp heterodimers are among the mere seven contact pairs that are identical between humans and mice (Fremont *et al.* 1992).

In conclusion, the molecules encoded by *Cyca*-12/SP1/G11 may well be capable of binding peptides, CD8 and $ß_2$m, and in addition they show many other characteristics described for class I molecules, as discussed by Kaufman and co-workers (1994). However, an additional remark should be made concerning the conservation of kinase site S325 in the cytoplasmic region. This serine is conserved in all sequences compared, except in the sequences from chicken and salmon (Fig. 2). The divergence of the salmon sequence (*Sasa*-p30) in the cytoplasmic region is surprising, when compared with the high degree of similarity to the carp and zebrafish sequences in the other regions of the transcript. However, a frameshift in the salmon sequence around amino acid position 302 (carp numbering, Fig. 2) results in a cytoplasmic amino acid sequence which now also contains the conserved S325, and which

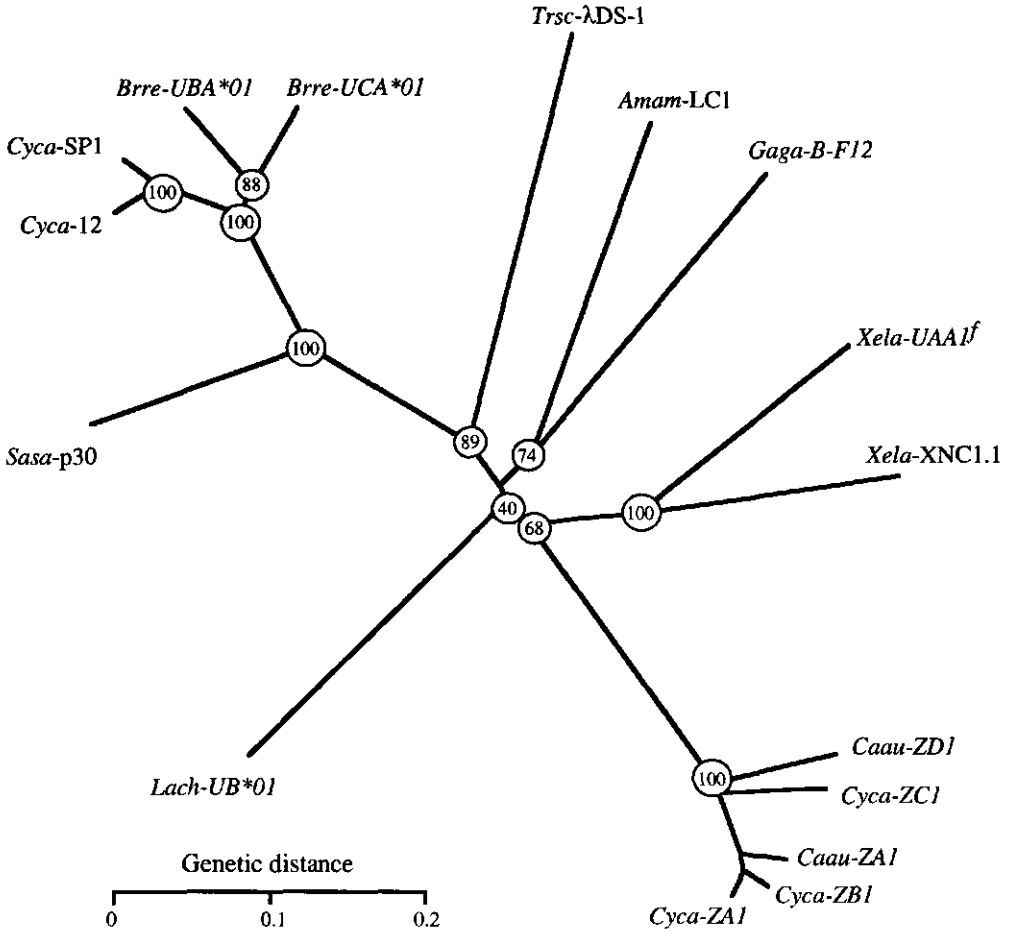


Figure 7. Phylogenetic tree constructed by the neighbor-joining method (Saitou and Nei 1987), based on uncorrected p-distances (proportion of differences) of exon 4 nucleotide sequences of *Mhc* class I genes from various vertebrates. Sources of sequences not described in the legend to Figure 2 are as follows: *Brre-UCA*01*, zebrafish (Takeuchi *et al.* 1995); *Caau-ZA1*, *Caau-ZD1*, Ginbuna crucian carp, and *Cyca-ZB1*, and *Cyca-ZC1*, common carp (Okamura *et al.* 1993). Numbers on nodes represent percentages of 1000 bootstraps supporting each partitioning.

A striking observation is that the 14 substitutions (of which 4.33 are synonymous) in exon 3 are not accompanied by any substitutions in 2335 bp of coding and non-coding sequence downstream of exon 3. A similar situation was observed in the comparison of *Cyca-DAB*01* and *Cyca-DAB*02* (Ono *et al.* 1993a), class II *B* cDNA clones from nurse shark (Bartl and Weissman 1994), and new class I alleles in Xenopus (M. Flajnik personal communication). This could be explained by a type of recombination, bringing together parts of sequences from different origins, which seems more likely than effective purifying selection at all sites in coding and non-coding regions.

To study the phylogenetic relationship of *Cyca*-12/SP1/G11 to class I sequences from other fish and lower vertebrates, a neighbor-joining tree was constructed from an alignment of exon 4 sequences (Fig. 7). *Cyca*-12/SP1/G11 cluster together with the class I sequences from zebrafish *(Brre-UB/UC)*, and Atlantic salmon (*Sasa*-p30). Within this cluster, the salmon sequence branches separately from the carp and zebrafish sequences, which in turn are grouped in separate clusters. However, this species-specific clustering is not observed in a phylogenetic tree constructed from an alignment of exons 3 of the same sequences, where *Cyca*-12 and *Cyca*-G11 cluster with *Brre-UA*, while *Cyca*-SP1 is located on a separate branch, not grouping with any of the other zebrafish or salmon class I sequences (data not shown). Therefore, we propose the designation *Cyca-UA1*01* for *Cyca*-12, and *Cyca-UA1*02* for *Cyca*-G11, while *Cyca*-SP1 is denoted *Cyca-UAW1*, since it can not yet be assigned to a particular locus. Phylogenetic analysis using only exons 2 of these sequences illustrates the divergence between these sequences of *Cyca*-12 and *Cyca*-SP1; in this analysis exon 2 of *Cyca*-12 is found to be more closely related to that of *Sasa*-p30, than to exon 2 of *Cyca*-SP1 (data not shown). Although this grouping is supported by only 44% of the bootstrap replications, it may indicate that exons 2 of *Cyca*-12 and *Sasa*-p30 share a more recent common ancestor than the rest of these genes. This may support the suggestion that a recombination has brought exons 2 and 3 of *Cyca*-12 together, as the closer relationship of *Cyca*-12 with *Sasa*-p30 is not found when comparing exons 3 (data not shown).

The clustering of class I genes from carp with those of zebrafish and Atlantic salmon confirms the finding by Takeuchi and co-workers (1995), that the ancestral gene arose before the salmonid/cyprinid split, probably as long as 120-150 million years (my) ago. Also, the tree illustrates clearly that the *Cyca-Z* and *Caau-Z* genes are members of a completely separate lineage, at a large genetic distance. Although these genes do not contain obvious defects, the lower conservation of functionally important features, and the difficulties in establishing expression of full-length cDNA clones may indicate that they do not exert a classical class I function. The fact that they are still expressed in the common and crucian carp may, however, indicate that they have another function. The expansion of class I like sequences in carp appears

to extend even further, since additional class I-like partial sequences, including C4, distinct from *Cyca-Z* and *UA*, have been isolated from carp cDNA (Grossberger and Parham 1992; Van Erp *et al.* 1996b).

To clarify the complex picture emerging from these distinct class I gene lineages in carp, studies on expression and function are becoming of great importance. Judging from the sequence data and the genomic organization alone, the *Cyca-UA* genes may well represent a group of classical class I sequences. However, routinely used criteria to discern non-classical from classical class I sequences are polymorphism and expression pattern (Shawar *et al.* 1994). Expression of *Cyca-UA* genes at the transcript level was detected in thymus and spleen/head kidney by the isolation of cDNA clones derived from these organs. Also, PCR amplifications on cDNA of adherent cells from spleen and head kidney confirmed the presence of transcripts in both organs (data not shown). However, functional studies can be performed using antibodies, and for this reason we proceeded to raise a polyclonal antiserum (anti-*Cyca*12) against a recombinant fusion protein containing most of the extracellular domains of *Cyca*-12.

The resulting antibodies showed substantial reactivity with the recombinant protein, and an M_r 45,000 molecule in membrane lysates of spleen and muscle, as well as to cell surface determinants present on cells in FACS analyses. Lymphocytes, both sIg^+ and sIg^-, were anti-*Cyca*12 positive. Erythrocytes and thrombocytes, although nucleated, were anti-*Cyca*12 negative. Class I expression on erythrocytes has been found to vary with species, whereas platelets, the mammalian counterparts of thrombocytes, were reported to express class I molecules (Klein 1986). The possibility remains, however, that carp erythrocytes and thrombocytes express other class I molecules, different from the *Cyca-UA* group. *Cyca*-12-like sequences thus appear to be expressed in a variety of tissues, consistent with the expression pattern of classical class I molecules.

Acknowledgements

We thank Joris Veltman for assistance in the isolation and sequence analysis of *Cyca*-12. Jeroen Roelofs and Pedro Rodrigues are acknowledged for their part in the preparation of the polyclonal antiserum, and the FACS analyses. Brian Dixon is supported by a postdoctoral fellowship from the Natural Sciences and Engineering Research Council of Canada.

References

Bartl, S. and Weissman, I.L. 1994. Isolation and characterization of major histocompatibility complex class IIB genes from the nurse shark. *Proc. Natl. Acad. Sci. USA 91:* 262-266

Betz, U.A.K., Mayer, W.E. and Klein, J. 1994. Major histocompatibility complex class I genes of the coelacanth *Latimeria chalumnae*. *Proc. Natl. Acad. Sci. USA 91:* 11065-11069

Dixon, B., Stet, R.J.M., Van Erp, S.H.M. and Pohajdak, B. 1993. Characterization of $ß_2$-microglobulin transcripts from two teleost species. *Immunogenetics 38:* 27-34

Dixon, B., Van Erp, S.H.M., Rodrigues, P.N.S., Egberts, E. and Stet, R.J.M. 1995. Fish major histocompatibility complex genes: an expansion. *Dev. Comp. Immunol. 19:* 109-133

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

Figueroa, F., Ono, H., Tichy, H., O'hUigin, C. and Klein, J. 1995. Evidence for insertion of a new intron into an *Mhc* gene of perch-like fish. *Proc. R. Soc. Lond. B 259:* 325-330

Flajnik, M.F., Canel, C., Kramer, J. and Kasahara, M. 1991. Evolution of the major histocompatibility complex: molecular cloning of major histocompatibility complex class I from the amphibian *Xenopus*. *Proc. Natl. Acad. Sci. USA 88:* 537-541

Fremont, D.H., Matsumura, M., Stura, E.A., Peterson, P.A. and Wilson, I.A. 1992. Crystal structures of two viral peptides in complex with murine MHC class I H-$2K^b$. *Science 157:* 919-927

Grimholt, U., Hordvik, I., Fosse, V.M., Olsaker, I., Endresen, C. and Lie, Ø. 1993. Molecular cloning of major histocompatibility complex class I cDNAs from Atlantic salmon (*Salmo salar*). *Immunogenetics 37:* 469-473

Grossberger, D. and Parham, P. 1992. Reptilian class I major histocompatibility complex genes reveal conserved elements in class I structure. *Immunogenetics 36:* 166-174

Guild, B.C. and Strominger, J.L. 1984. Human and murine class I MHC antigens share conserved serine 335, the site of HLA phosphorylation *in vivo*. *J. Biol. Chem. 259:* 9235-9240

Guillemot, F., Billault, A., Pourquie, O., Behar, G., Chausse, A.M., Zoorob, R., Kreiblich, G. and Auffray, C. 1988. A molecular map of the chicken major histocompatibility complex: the class II ß genes are closely linked to the class I genes and the nucleolar organizer. *EMBO J. 7:* 2775-2785

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

Hashimoto, K., Nakanishi, T. and Kurosawa, Y. 1992. Identification of a shark sequence resembling the major histocompatibility complex class I α_3 domain. *Proc. Natl. Acad. Sci. USA 89:* 2209-2212

Higgins, D.G., Bleasby, A.J. and Fuchs, R. 1992. CLUSTAL V: improved software for multiple sequence alignment. *Comput. Appl. Biosci. 8:* 189-191

Kaastrup, P., Stet, R.J.M., Tigchelaar, A.J., Egberts, E. and Van Muiswinkel, W.B. 1989. A major histocompatibility locus in fish: serological identification and segregation of transplantation antigens in the common carp (*Cyprinus carpio* L.). *Immunogenetics 30:* 284-290

Kaufman, J., Salomonsen, J. and Flajnik, M. 1994. Evolutionary conservation of MHC class I and class II molecules-different yet the same. *Seminars in Immunol. 6:* 411-424

Klein, J. 1986. Natural History of the Major Histocompatibility Complex. Wiley and Sons, New York

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

Koller, B.H. and Orr, H.T. 1985. Cloning and complete sequence of an HLA-A2 gene: analysis of two HLA-A alleles at the nucleotide level. *J. Immunol. 134:* 2727-2733

Komen, J., Van den Dobbelsteen, P.J.M., Slierendrecht, W.J. and Van Muiswinkel, W.B. 1990. Skin grafting in gynogenetic common carp (*Cyprinus carpio* L.). The development of

histocompatible clones. *Transplantation 49:* 788-793

Koumans-Van Diepen, J.C.E., Egberts, E., Peixoto, B.R., Taverne, N. and Rombout, J.H.W.M. 1995. B cell and immunoglobulin heterogeneity in carp (*Cyprinus carpio* L.), an immuno(cyto)chemical study. *Dev. Comp. Immunol. 19:* 97-108

Koumans-Van Diepen, J.C.E., Taverne-Thiele, J.J., Van Rens, B.T.T.M. and Rombout, J.H.W.M. 1994. Immunocytochemical and flow cytometric analysis of B cells and plasma cells in carp (*Cyprinus carpio* L.); an ontogenetic study. *Fish Shellfish Immunol. 4:* 19-28

Kumar, S., Tamura, K. and Nei, M. 1993. MEGA: Molecular Evolutionary Genetics Analysis, version 1.01. The Pennsylvania State University, University Park

Okamura, K., Nakanishi, T., Kurosawa, Y. and Hashimoto, K. 1993. Expansion of genes that encode MHC class I molecules in cyprinid fishes. *J. Immunol. 151:* 188-200

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

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

Ono, H., Figueroa, F., O'hUigin, C. and Klein, J. 1993b. Cloning of the $ß_2$-microglobulin gene in the zebrafish. *Immunogenetics 38:* 1-10

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

Ono, H., O'hUigin, C., Tichy, H. and Klein, J. 1993d. Major histocompatibility complex variation in two species of cichlid fishes from Lake Malawi. *Mol. Biol. Evol. 10:* 1060-1072

Rombout, J.H.W.M., Koumans-Van Diepen, J.C.E., Taverne, N. and Taverne-Thiele, J.J. 1994. Analysis of blood leucocytes and thrombopoiesis with monoclonal antibodies specific for carp thrombocytes. *Dev. Comp. Immunol. 18 Suppl. 1:* S61

Rombout, J.H.W.M., Taverne, N., Van de Kamp, M. and Taverne-Thiele, J.J. 1993. Differences in mucus and serum immunoglobulin of carp (*Cyprinus carpio* L.). *Dev. Comp. Immunol. 17:* 309-317

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

Salter, R.D., Benjamin, R.J., Wesley, P.K., Buxton, S.E., Garrett, T.P.J., Clayberger, C., Krensky, A.M., Norment, A.M., Littman, D.R. and Parham, P. 1990. A binding site for the T-cell co-receptor CD8 on the α_3 domain of HLA-A2. *Nature 345:* 41-46

Saper, M.A., Bjorkman, P.J. and Wiley, D.C. 1991. Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 Å resolution. *J. Mol. Biol. 219:* 277-319

Secombes, C.J., Van Groningen, J.J.M. and Egberts, E. 1983. Separation of lymphocyte subpopulations in carp, *Cyprinus carpio* L., by monoclonal antibodies: immunohistochemical studies. *Immunology 48:* 165-175

Shawar, S.M., Vyas, J.M., Rogers, J.R. and Rich, R.R. 1994. Antigen presentation by major histocompatibility complex class I-b molecules. *Annu. Rev. Immunol. 12:* 839-880

Shum, B., Avila, D., Du Pasquier, L., Kasahara, M. and Flajnik, M. 1993. Isolation of a classical MHC class I cDNA from an amphibian. evidence for only one class I locus in the *Xenopus* MHC. *J. Immunol. 151:* 5376-5386

Stet, R.J.M. and Egberts, E. 1991. The histocompatibility system in teleostean fishes: from multiple histocompatiblity loci to a major histocompatibility complex. *Fish Shellfish Immunol. 1:* 1-16

Stet, R.J.M., Van Erp, S.H.M., Hermsen, T., Sültmann, H.A. and Egberts, E. 1993. Polymorphism and estimation of the number of *MhcCyca* class I genes in laboratory strains of the common carp (*Cyprinus carpio* L.). *Dev. Comp. Immunol. 17:* 141-156

Sültmann, H., Mayer, W.E., Figueroa, F., O'hUigin, C. and Klein, J. 1993. Zebrafish Mhc class II α chain-encoding genes: polymorphism, expression and function. *Immunogenetics 38:* 408-420

Sültmann, H., Mayer, W.E., Figueroa, F., O'hUigin, C. and Klein, J. 1994. Organization of *Mhc* class II *B* genes in the zebrafish (*Brachydanio rerio*). *Genomics 23:* 1-14

Takeuchi, H., Figueroa, F., O'hUigin, C. and Klein, J. 1995. Cloning and characterization of class I *Mhc* genes of the zebrafish, *Brachydanio rerio*. *Immunogenetics 42:* 77-84

Trowsdale, J. 1995. "Both man & bird & beast": comparative organization of MHC genes. *Immunogenetics 41:* 1-17

Van Erp, S.H.M., Egberts, E. and Stet, R.J.M. 1996a. Characterization of major histocompatibility complex class II *A* and *B* genes in a gynogenetic carp clone. *Immunogenetics* in press

Van Erp, S.H.M., Egberts, E. and Stet, R.J.M. 1996b. Evidence for multiple distinct histocompatibility complex class I lineages in teleostean fish. *Eur. J. Immunogenetics* in press

Von Heijne, G. 1986. A new method for predicting signal sequence cleavage sites. *Nucleic Acids Res. 14:* 4683-4690

Williams, A.F. and Barclay, A.N. 1988. The immunoglobulin superfamily - domains for cell surface recognition. *Annu. Rev. Immunol. 6:* 381-405

Chapter 5

Evidence for multiple distinct major histocompatibility complex class I lineages in teleostean fish

Saskia H.M. van Erp, Egbert Egberts and René J.M. Stet

Department of Experimental Animal Morphology and Cell Biology, P.O. Box 338, 6700 AH Wageningen, Wageningen Agricultural University, The Netherlands

Accepted for publication in: **European Journal of Immunogenetics (1996)**

Summary

In the context of studies on expression of *MhcCyca-Z* sequences of the common carp, PCR amplifications of exon 4 were performed on cDNA obtained from pooled thymi of 20 carp F1 individuals. Five recombinant clones (*Cyca*-TC3, -TC13, -TC15, -TC17 and -TC18) were found to be 96% similar to exon 4 region of *Cyca-ZA1*. Each of the five sequences was unique, and differed in a few positions in both the nucleotide and derived amino acid sequence from any of the *Cyca-Z* sequences known to date. These data suggest that multiple Z genes per locus are present in the carp, which are transcribed in the thymus. In the course of analyzing the amplified *Cyca-Z* sequences, serendipity yielded a clone, *Cyca*-TC16, containing a class I-like sequence substantially different from any other carp class I sequence. The predicted amino acid sequence of *Cyca*-TC16 was most similar to the class I genes (*Lach-U*) from the coelacanth (42-46% amino acid identity). *Cyca*-TC16 contains three conserved $ß_2$-microglobulin contact residues, and the secondary structure was predicted by computer algorithms to be similar to that of the α_3 domain of *HLA-A2*. Phylogenetic analysis shows that carp class I sequences reside in four distinct clusters: (i) *Cyca-Z*, *Cyca*-TC3, -TC13, -TC15, -TC17, and -TC18 together with *Caau-Z* from ginbuna crucian carp; (ii) *Cyca-U* with *Brre-U* (zebrafish) and *Sasa*-p30 (Atlantic salmon); (iii) *Cyca*-TC16 with *Lach-U* (coelacanth); and (iv) *Cyca*-C4.

The nucleotide sequences reported in this paper have been submitted to the EMBL/GenBank nucleotide sequence databases and have been assigned the accession numbers X95182-87.

Introduction

The major histocompatibility complex is a volatile part of the genome, and this characteristic is one of the features of this complex that emerged from extensive studies on the *Mhc* of mammals, especially of humans and mice (Klein *et al.* 1993). The genes within the *Mhc* apparently duplicate, are deleted, mutate, recombine, and as a consequence lose or acquire sometimes extremely diversified functions. The result is a large number of genes with apparently related origins, but which encode molecules with a variety of functions, including antigen presentation in the classical or in the more restricted non-classical sense (Shawar *et al.* 1994). Sometimes these molecules have completely unrelated functions, such as the rat neonatal Fc receptor (Burmeister *et al.* 1994). Now that *Mhc* genes have been identified in a rapidly growing number of cold-blooded vertebrates (Betz *et al.* 1994; Dixon *et al.* 1995; Takeuchi *et al.* 1995; Grossberger and Parham 1992; Van Erp *et al.* 1996a; Flajnik *et al.* 1991; Shum *et al.* 1993) it is to be expected that in these species a similar abundance of related genes will be found. Indeed, in the common carp, *Cyprinus carpio* L., three distinct *MhcCyca* class I genes and gene-fragments were previously reported, namely (1) the *Cyca-Z* partial sequences from genomic DNA and cDNA (Hashimoto *et al.* 1990; Okamura *et al.* 1993);

(2) *Cyca*-C4, a PCR fragment from cDNA spanning part of exon 3 and 4 (Grossberger and Parham 1992); and (3) the recently identified *Cyca-UA* cDNA and genomic sequences (Van Erp *et al.* 1996a). The nomenclature of the carp *Mhc* genes is in accordance with the proposal made by Klein and others (1990), such that the name *Cyca* is derived from the first two letters from the genus and species names for the common carp, *Cyprinus carpio*.

Cyca-Z, *Cyca*-C4 and *Cyca-U* sequences are substantially different from each other (15 - 20% identity at the amino acid level). Whereas *Cyca*-C4 so far is a unique sequence of unclear status with only limited similarity to class I sequences, *Cyca-Z* and *Cyca-U* are represented by multiple sequence variants in carp and other teleosts, thereby constituting *Z* and *U* sequence lineages. Of the carp class I sequences, *Cyca-U* genes are the most likely candidates for encoding class I molecules with a classical function, based on the presence of characteristic classical key residues, and expression patterns (Van Erp *et al.* 1996a). *Cyca-Z* on the other hand lacks several features of classical class I molecules, and expression has so far only been detected in the kidney.

In the context of studies on expression of *Cyca-Z* sequences in other organs, the present study was aimed at identification of these sequences in thymus by performing PCR on cDNA.

Materials and methods

Fish

The fish used in this study were laboratory strains of common carp (*Cyprinus carpio* L.). R3xR8 are the offspring of a hybrid cross between a female of Hungarian origin (R8 strain), and a male of Polish origin (R3 strain). Fish were held at 25°C in recirculation systems, fed on pelleted food (Provimi, Rotterdam, The Netherlands).

Preparation of thymus cDNA

Total cellular RNA from thymi of 20 individuals of R3xR8 carp was isolated using the lithium precipitation method of Palmiter (1974), incorporating a DNAse I treatment. RNA was converted into cDNA with a poly-dT primer, using a cDNA synthesis kit according to the manufacturer's specification (Boehringer, Mannheim, Germany),

Polymerase Chain Reaction (PCR)

PCR was carried out in 100 μl of a solution containing 100 ng of DNA template in 1 x *Taq* buffer, 0.2 mM of each dNTP, 0.2 μM of each primer, and 1 unit of *Taq* polymerase (Perkin-Elmer, Norwalk, CT). Amplification was performed using 30 cycles of a thermal cycle profile

(1 min 94°C, 1 min 55°C, 2 min 72°C), with a final extension step of 10 min at 72°C. Oligonucleotides used in the amplifications were degenerate primers I and II:
I: 5'-TG(C/T)CT(A/G/C/T)GT(C/G)AC(A/G/C/T)GGTTTCTACCC-3'; and
II: 5'-AG(G/A)CT(T/G)(G/C)(T/G)(G/A)TG(C/A)ATCACATGACA-3'
in part based on the *Cyca-ZA1* sequence, modified from Hashimoto and co-workers (1990).

Cloning and sequence analysis

PCR fragments were blunt-end ligated into pTZ19R, and plasmids were used to transform *E.coli* strain JM101. Single-stranded phagemid DNA was obtained and sequenced by the dideoxy chain termination method (Sanger *et al.* 1977), using the T7 DNA polymerase sequenase kit (US Biochemicals, Cleveland, OH). Sequence data were analyzed using several programs from the GCG package (Genetics Computer Group, Madison, WI). Multiple sequence alignment and phylogenetic trees were constructed using the program ClustalV (Higgins *et al.* 1992), and MEGA (Kumar *et al.* 1993).

Southern transfer and hybridization

PCR amplified fragments were separated by agarose gel electrophoresis, and subsequently transferred to Hybond-N+ nylon membranes, according to the manufacturer's protocols (Amersham, Amersham, UK). Probes were labeled to a specific activity of 1.5×10^8 cpm/μg, and hybridization was carried out for 16 hours at 42°C, in a solution of 40% formamide, 5 x SSC, 5 mM EDTA, 0.1% SDS, 5 x Denhardt's and 100 μg denatured *E.coli*-DNA, using 2×10^6 cpm of the labeled probe. Membranes were washed to a final stringency of 1 x SSC, 0.05% SDS at 60°C for 15 min., and used to expose X-omat S films (Kodak, Rochester, NY), for 24 hours at -80°C with intensifying screens.

A.

```
Cyca-ZA1  TGTCTGGTCACTGGTTTCTACCCCAGAGATATTGAGATGAACATCAGACTGAACAGAATTAAC
Cyca-ZA2  ------------------------------------------------------------
Cyca-ZB1  ----------------------------G-----------------T--------------
Cyca-ZC1  ------------------------------------------------T--G------C-GCA
K1-5      --------A-----------------------------------------------------
Cyca-TC3      A--G--G-----------------G-----------------T--------------
Cyca-TC13 -----C--G--G-----------------G-----------------T--------------
Cyca-TC15 ----------A-----------------G-G---------------T--------------
Cyca-TC17 -----A--G--C-----------------------------------------------
Cyca-TC18 --C--------C-----------------G-----------------T--------------

Cyca-ZA1  ATTGAGAGCCAGATATCATCTGGAATCAGACCAAATGATGATGAAAGCTTTCAGATGAGATCC
Cyca-ZA2  ----------A-C-------------------------A-----------------------
Cyca-ZB1  ---C---A-------------------------------------------------------
Cyca-ZC1  C---G--A---A----TT----A---------C----C-----G-TC-------C-C------
K1-5      ----------------T---------------------A-------------C-C------
Cyca-TC3  ---C---A-----------------------------------------C-------------
Cyca-TC13 ---C---A-----------------------------------------------T-------
Cyca-TC15 ---C---A-------------------------------------------------------
Cyca-TC17 ------G----------C---------------------A-------------C-C------
Cyca-TC18 -C-C---A-------------------------------------------------------

Cyca-ZA1  AGTGTGAAGATCGACAGAAACCACAGAGGATCTTATGACTGTCATGTGATTCACAGCAGCCT
Cyca-ZA2  ------------------------------------------------------------
Cyca-ZB1  ------------------------------------------------------------
Cyca-ZC1  ----------------------T-A--------T-------TT---C-A------------
K1-5      ------------------------------------------------------------A
Cyca-TC3  ------------------------------------------------TC-------
Cyca-TC13 ------------------------------------------------CC------
Cyca-TC15 ------------------------------------------------C-------
Cyca-TC17 ------------------------------------------------C-------
Cyca-TC18 ------------------------------------------G---CC------
```

B.

```
                10        20        30        40        50        60
Cyca-ZA1  CLVTGFYPRDIEMNIRLNRINIESQISSGIRPNDDESFQMRSSVKIDRNHRGSYDCHVIHSS
Cyca-ZA2  ------------------------T--------N-------------------------
Cyca-ZB1  ----------V-----------QN-------------------------------------
Cyca-ZC1  --A--------------D-TALGN--F-E----A-G---L----------K--F--F-N---
K1-5      ........-------------------------N----L--------------.......
Cyca-TC3  .........--V-----------QN-------------L-----------------.....
Cyca-TC13 .........--V-----------QN-------------H-----------------.....
Cyca-TC15 .........-GV-----------QN-------------------------------.....
Cyca-TC17 ........---------------G----------N----L--------------.......
Cyca-TC18 .........--V----------TQN-------------------------------.....
```

Figure 1. (**A**) Nucleotide and (**B**) amino acid alignments of *Cyca-Z* partial exon 4 sequences. *Cyca-ZA1*, *-ZA2*, *-ZB1* and *-ZC1* from (Okamura *et al.* 1993), K1-5 (Stet *et al.* 1993), *Cyca*-TC3 to -TC18 this study. Dashes denote identity to *Cyca-ZA1*, primer-encoded sequences are underlined. Dots denote absence of sequence information.

```
          10          20          30          40          50
TGC CTA GTG ACA GGT TTC TAC CCC AAG GAC ATC TCT GTT AAA TGG GAG CTG
cys leu val thr gly phe tyr pro lys asp ile ser val lys trp glu leu

          61                                  91
GAT GGG AAG CCA ACC TCT CTA CAT GTG ACA ACA GAC ATC TTA CCT AAT CAT
asp gly lys pro thr ser leu his val thr thr asp ile leu pro asn his

                    121                                       151
GAT TCG ACT TAC CAG GTG CAT AAA ACC ATC TTC ATC TCT GGT TCT AAA CAC
asp ser thr tyr gln val his lys thr ile phe ile ser gly ser lys his

                                  181
AAC TAC TCA TGT CAT GTG ATG CAC CGC AGC CT
asn tyr ser cys his val met his arg ser
```

Figure 2. Nucleotide and derived amino acid sequence of *Cyca*-TC16. Primers used in the isolation are underlined.

Results

To study the expression of *MhcCyca-Z*-like sequences especially in the carp thymus, degenerate PCR primers were designed, complementary to the conserved regions flanking the cysteine codons in exon 4 of *Cyca-ZA1*. cDNA was obtained from pooled thymi of 20 R3xR8 carp individuals. Pooling of thymi was required to obtain sufficient RNA for cDNA synthesis, because of the small size of this organ in carp. PCR amplifications performed with the degenerate primers on this cDNA yielded a single band with a size of approximately 200 bp. Five of the resultant recombinant clones (denoted *Cyca*-TC3, -13, - 15, -17, and -18) were found to contain sequences that were 96% similar to the region between the cysteine codons in exon 4 of *Cyca-ZA1*. Each of these five sequences was unique, and differed in only a few positions in both the nucleotide and derived amino acid sequence from any of the *Cyca-Z* sequences known to date (Fig. 1).

In addition a recombinant clone was isolated, designated *Cyca*-TC16, containing an insert of 189 bp (Fig. 2), of which the predicted amino acid sequence showed the highest similarity to the recently described class I genes (*Lach-UA* and *-UB*) from the coelacanth, *Latimeria chalumnae*. Identities ranged from 42% to 46% depending on the coelacanth gene used in the comparison. Alignment of the predicted amino acid sequence of *Cyca*-TC16 with class I sequences from other fish and cold-blooded vertebrates shows that overall similarity is low (Fig. 3). The lowest similarities were found in the comparisons with *Cyca-Z* and *Cyca-UA* sequences from carp (both 20% identity). Nevertheless, *Cyca*-TC16 possesses several residues and motifs that are conserved between most of the *Mhc* class I sequences from cold-blooded vertebrates in the comparison. Three potential β_2-microglobulin (β_2m) contact residues (P32, D35 and Q39) are conserved between *Cyca*-TC16, *HLA-A2* and *H-2K*^{b} (Saper *et al.* 1991;

```
                    10        20        30        40        50        60
                              ooooooo   Δ  Δ  Δ
Cyca-TC16     CLVTGFYPKDISVKWELDGKPTSLHVTTD*ILPNHDSTYQVHKTIFI*SGS**KHNYSCHVMHRS
Cyca-UA1*01   -HT-----SGVTIT-QKN-QDHDED-DLGELII-E-G-F-RAS-LNVKPEEWKNNKF--V-E-QG
Brre-UBA*01   -H------SGLKIS-QRN-QDHDED-ELGELM--E-G---RTS-LNVKPEEWK-DKF--V-E-QG
Sasa-p30      -HA-----SGVM-S-QK--QDHHED-EYGET-Q-D-G-F-KSSHLTVTPEEWKNNK-Q-V-QVTG
Cyca-ZA1      --------R--EMN*IRLNRINIESQISSG-R--D-ESF-MRSSVK-DRNH**RGS-D---I-S-
Caau-ZA1      --------R--EMN*IRLNRINIQNQISSGVR--D-E-F-MRSSVK-DRNH**RGS-D---I-S-
Lach-UA*01    -M----H-RA-D-T-IR--ETRMDNAH--G----E-G---IK---E-G-DD**-RS-A-E-D-G-
Lach-UB*01    YMA-----RAVDMT-VR--ETQMDNAH--G----E-E---IR---E-DLED**--S-T-W-D-S-
Trsc-λDS-1    -H-N----SG-NAT-LHN-GTIQQE-LSSR----T-G-F-TTLQ-SVTPQ-**RDT-T-Q-E-S-
Xela-UAA1f    -QAY----RE-D---VKN-GDDVHSEAAKE----P-GS--LRV-AE-TPNE**GDS-A---E-S-
Xela-XNC1.1   -W-YR---R-VE---IKN-TDEIYSEESAE----P-G---IRVSVEVTPEE**GAT-----D-S-
Amam-LC1      -RAD-------HPV-IR--EVWEQETMRGLMA--V-G-FHTWIG-K-DPKD**RGRFQ-R-D-AG
```

Figure 3. Comparison of carp *Cyca*-TC16 derived amino acid sequence with *Mhc* class I partial exon 4 sequences from cold blooded vertebrates. Sources of sequences not mentioned in legend to Figure 1: *Cyca-UA1*01*, common carp (Van Erp *et al.* 1996a), *Brre-UBA*01*, zebrafish (Takeuchi *et al.* 1995); *Sasa*-p30, Atlantic salmon (Grimholt *et al.* 1993); *Caau-ZA1*, ginbuna crucian carp (Okamura *et al.* 1993); *Lach-UA*01, Lach-UB*01*, coelacanth (Betz *et al.* 1994); *Trsc*-λDS-1, nurse shark (Hashimoto *et al.* 1992); *Xela-UAA1*^{f}, clawed frog, classical gene (Shum *et al.* 1993); *Xela*-XNC1.1, clawed frog, non-classical (Flajnik *et al.* 1991); *Amam*-LC1, lizard (Grossberger and Parham 1992); Residues encoded by primers are underlined; Dashes indicate identity to *Cyca*-TC16; Triangles denote putative $ß_2$m-contacting residues; Circles denote position of region homologous to CD8-binding loop in *HLA-A2*;

Fremont *et al.* 1992). Also, the secondary structure of the *Cyca*-TC16 encoded protein, as predicted by the program Peptidestructure, using the Chou-Fasman-Prevelige and Garnier-Osguthorpe-Robson algorithms, contains ß strands in overlapping positions with those shaping the ß pleated sheets in the Ig-fold of the α_3 domain of *HLA-A2* (data not shown). In contrast, CD8-binding residues (Salter *et al.* 1990) are not conserved in *Cyca*-TC16.

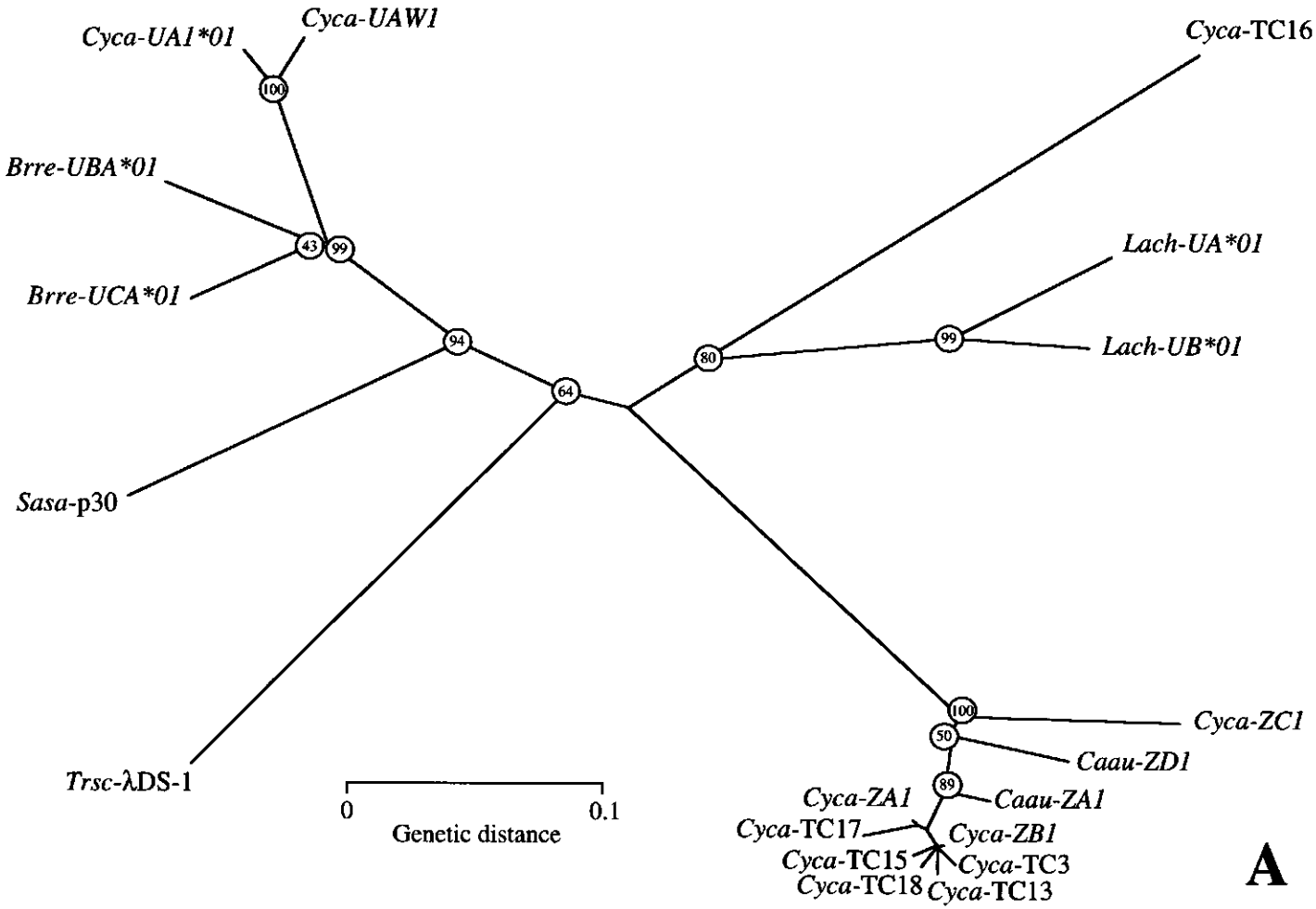
To study the evolutionary relationships to other class I exon 4 sequences from fish, a neighbor-joining tree was constructed (Fig. 4A). *Cyca*-TC16 clusters with the class I sequences from coelacanth (*Lach-UA/-UB*). In a separate cluster, *Cyca*-TC3, -TC13, -TC15 and -TC18 branch with the carp class I *Cyca-Z* sequences and the *Caau-Z* genes from ginbuna crucian carp, whereas the carp class I *Cyca-UA* sequences group together with the *Brre-U* genes from zebrafish and *Sasa*-p30 from Atlantic salmon. The sequence from nurse shark (*Trsc*-λDS-1) branches away from the three Osteichthyan class I clusters. In order to be able to include the fourth carp sequence, *Cyca*-C4, in the phylogenetic analyses, a neighbor-joining tree was constructed using an alignment of class I nucleotide sequences spanning the region between the C-terminal cysteine codon of exon 3 and the N-terminal cysteine codon of exon 4 (Fig. 4B). Although *Cyca*-TC16 and the sequence from nurse shark could not be included in this tree,

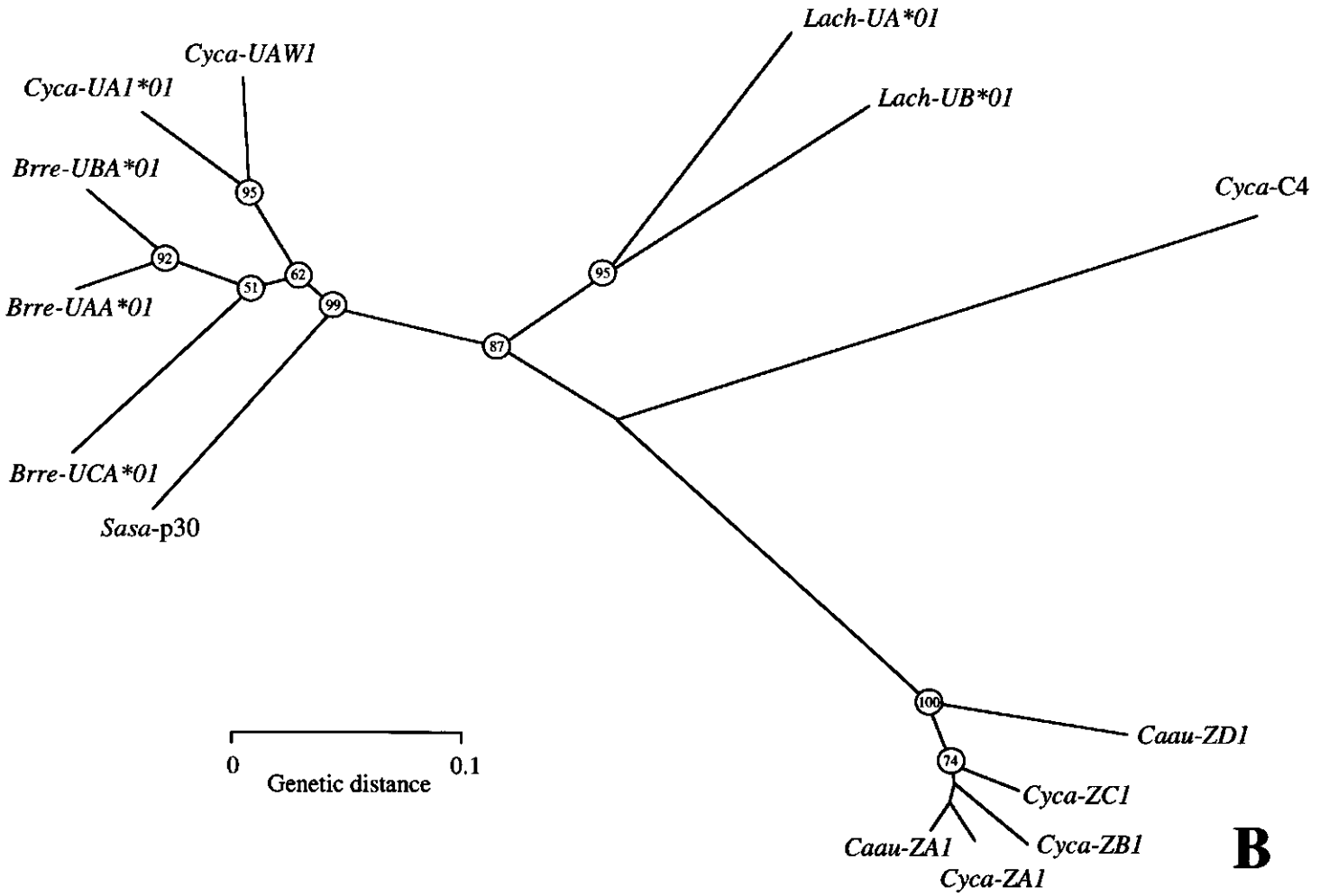
because they do not span this region of the class I sequence, the topology of this second tree basically follows that of the other. As observed in the exon 4 tree, *Cyca-Z* and *Cyca-UA* sequences are found in separate clusters, and now also *Cyca*-C4 is found on a distinct branch, at a large genetic distance from the other lineages.

In order to obtain the remainder of the transcript of *Cyca*-TC16, three cDNA libraries and a commercially obtained genomic library (Stratagene, La Jolla, CA) were screened. The cDNA libraries were prepared from (1) spleen/head kidney (Ono *et al.* 1993), (2) thymus (Van Erp *et al.* 1996a), and (3) activated macrophages. However, neither screening using *Cyca*-TC16 as a probe, nor screening of the cDNA libraries using anchored PCR with *Cyca*-TC16-specific primers, resulted in the isolation of *Cyca*-TC16-encoding clones.

Nevertheless, the presence of *Cyca*-TC16 in 14 individuals of the R3xR8 family was confirmed by PCR amplification using both a set of *Cyca*-TC16-specific primers, and in six of these individuals, using degenerate primers I and II, designed to amplify *Cyca-Z* sequences (Fig. 5). Although amplification efficiency was low, hybridization of the PCR products to *Cyca*-TC16 revealed that *Cyca*-TC16-like sequences were successfully amplified from each of the individuals tested. The PCR product obtained from one of the individuals using *Cyca*-TC16-specific primers was cloned and sequenced, which proved it to be identical to *Cyca*-TC16. In addition, also the PCR products amplified with primers I and II hybridized to *Cyca*-TC16, showing that similar sequences could reliably be amplified with these primers.

Figure 4 (next pages). Phylogenetic tree created by the neighbor-joining method (Saitou and Nei 1987) using uncorrected p-distances, of *Mhc* class I nucleotide sequences from various vertebrates, using
A. sequences spanning the region between the two cysteine codons in exon 4, and
B. sequences spanning the region between the C-terminal cysteine codon of exon 3 and the N-terminal cysteine codon of exon 4. Gaps were eliminated from the comparisons. Sources of sequences not mentioned in legend to Figures 1 or 3 are as follows: *Cyca-UAW1* (Van Erp *et al.* 1996a); *Cyca*-C4 (Grossberger and Parham 1992); *Brre-UAA*01*, *Brre-UCA*01* zebrafish (Takeuchi *et al.* 1995); *Caau-ZD1*, ginbuna crucian carp (Okamura *et al.* 1993). Numbers on nodes indicate the percentage of 1000 bootstraps supporting each partitioning.


Cyca-UA1*01
Cyca-UAW1
Cyca-TC16
Brre-UBA*01
Lach-UA*01
Brre-UCA*01
Lach-UB*01
Sasa-p30
Trsc-λDS-1
Cyca-ZC1
Caau-ZD1
Cyca-ZA1
Caau-ZA1
Cyca-TC17
Cyca-ZB1
Cyca-TC15
Cyca-TC3
Cyca-TC18
Cyca-TC13
100
43
99
94
64
80
99
100
50
89
0
0.1
Genetic distance
A


Lach-UA*01
Cyca-UAW1
Lach-UB*01
Cyca-UA1*01
Brre-UBA*01
Cyca-C4
95
92
51
62
99
95
87
Brre-UAA*01
Brre-UCA*01
Sasa-p30
100
Caau-ZD1
74
Cyca-ZC1
Caau-ZA1
Cyca-ZB1
Cyca-ZA1
0
0.1
Genetic distance
B

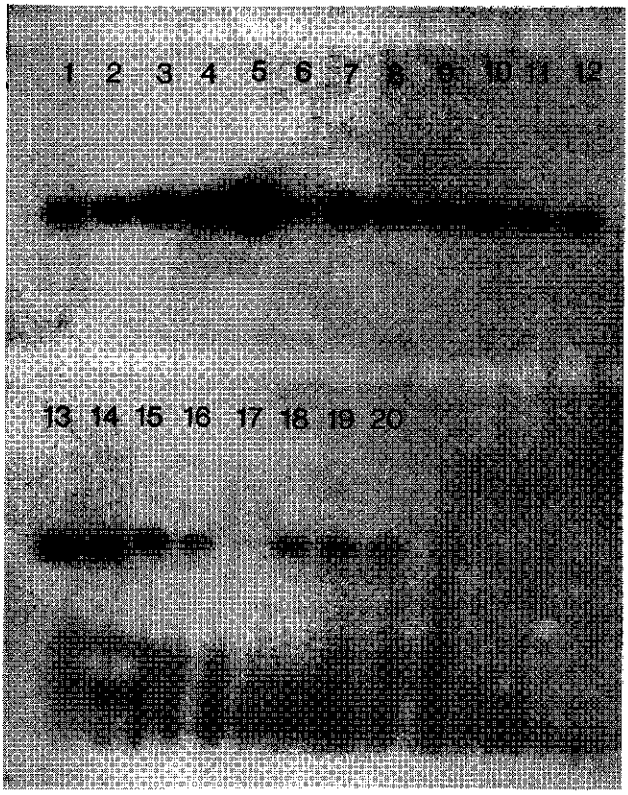


Figure 5. PCR fragments amplified from genomic DNA of R3xR8 F1 individuals, detected by hybridization with *Cyca*-TC16.
Lanes 1-14: using *Cyca*-TC16 specific primers on DNA of individuals no. 6-12, 14-20;
Lanes 15-20: using degenerate primers I and II on DNA of individuals no. 6, 8, 10, 12, 14, 16.

Discussion

In the context of a study on the expression of *Cyca-Z* sequences in various organs, we performed PCR amplifications on cDNA derived from thymus using degenerate primers, complementary to the regions surrounding the cysteine codons in exon 4 of *Cyca-Z* sequences (Hashimoto *et al.* 1990). Among the sequences amplified, five unique *Cyca-Z* sequences were identified. Although only one clone for each of the sequences was analyzed, it however seems unlikely that PCR artefacts could create five different variants. *Cyca*-TC3, -TC13, -TC15 and -TC18 all share residue V11, and the motif Q23N24, found in the previously reported *Cyca-ZB1* locus, whereas *Cyca*-TC17 shares I11 and E23 with the *Cyca-ZA* sequences and K1-5. The finding of five different sequences in five recombinant clones sequenced, suggests the presence of a large number of *Cyca-Z* sequences in the R3xR8 F1 individuals from which the thymus cDNA was prepared. This would be consistent with the suggestion by Okamura and co-workers (1993) that at least three, if not more, *Cyca-Z* genes per locus are present in the carp genome. Calculation of the genetic distances between the *Cyca-Z* sequences gives an indication of the divergence times between the various loci. We applied a synonymous substitution rate of 2.9×10^{-9} substitutions per synonymous site per year, which is based on an approximate divergence time of 50 my (Stroband *et al.* 1995) between the two most closely related carp and zebrafish class II *B* sequences, *Cyca-DAB4*01* (Van Erp *et al.* 1996b) and Brre-DAB3*01 (Ono *et al.* 1992). Using this rate, the carp *Cyca-ZA1* and *Cyca-ZB1* are estimated to have diverged during approximately 6 my, whereas the *Cyca-ZC1* locus diverged from the *ZA1/ZB1* ancestral locus approximately 30 my ago.

The successful amplification of *Cyca-Z* variants in the present study indicates that, besides the previously reported expression in kidney (Okamura *et al.* 1993), these sequences are also transcribed in the thymus. This was further confirmed by the finding of partial transcripts in the thymus cDNA library using anchored PCR (Van Erp and co-workers, unpublished data).

In the course of analyzing the amplified *Cyca-Z* sequences, serendipity yielded a clone, *Cyca*-TC16, containing a class I-like sequence substantially different from any other carp class I sequence reported to date. *Cyca*-TC16 was identified as class I-like based on its amino acid sequence similarity to the class I sequences of the coelacanth. In retrospect, the level of serendipity of this finding can be rationalized by the high similarity between the degenerate primers used in the amplification of the coelacanth sequences (Betz *et al.* 1994) and *Cyca*-TC16.

The conservation of three potential β_2-microglobulin contacting residues (P32, D35, Q39) suggests that the product of *Cyca*-TC16 may bind β_2m. The congruity of the positions of the ß strands in *Cyca*-TC16, as predicted by computer algorithms, with those found in the crystal structure of *HLA-A2* supports the conclusion that the *Cyca*-TC16 encoded protein could be

folded in an Ig-like manner. The region in *Cyca*-TC16 corresponding to the CD8-binding loop in mammalian class I molecules is, however, not conserved, as was also found for the *Lach-U* sequences from the coelacanth. In contrast, in the *Cyca-UA* sequences this region does seem to be functionally conserved by a high content of acidic residues (Van Erp *et al.* 1996a). The lack of conservation of this region in *Cyca*-TC16 may therefore indicate that this molecule is not involved in binding CD8. One should bear in mind, however, that CD8 of fish still remains unidentified.

All attempts to obtain the remainder of the *Cyca*-TC16 sequence failed. Failure to isolate the sequence from cDNA libraries could be due to the possiblity that *Cyca*-TC16 is not expressed in the tissues used to construct the libraries, or at extremely low levels. Although *Cyca*-TC16 was initially amplified from thymus cDNA, both primers were positioned within exon 4 around the cysteine codons, and as a consequence amplification from traces of genomic DNA can not be excluded. Failure to isolate positive clones from a genomic library may have been caused by haplotypic differences between the R3xR8 carp and the carp individual used in preparation of the genomic library, as allelic differences are rare in α_3 domains. In addition, technical difficulties in using *Cyca*-TC16 as a probe may have generally hampered screenings. This explanation seems to be supported by difficulties experienced in performing Southern analyses on R3xR8 genomic DNA using *Cyca*-TC16 as a probe (data not shown). A similar case of aberrant hybridization properties of homologous probes was observed in a study using *Cyca-Z*- and *Cyca-YB*-derived exon 4 probes in RFLP analyses of common carp, in which only extremely low stringency could be applied (Stet *et al.* 1993).

Phylogenetic analyses of *Mhc* class I exon 4 sequences of fish, including *Cyca*-TC16 and members of two of the other class I lineages identified in carp, *Cyca-Z*, and *Cyca-UA*, illustrate the large genetic distance between each of the lineages (Fig. 4A). *Cyca*-C4, the fourth carp class I sequence, appears to be a member of yet another lineage, distinct from the *Cyca-Z* and *-U* lineages, and most likely also distinct from the *Cyca*-TC16 lineage, considering the clustering of *Cyca*-TC16 with the *Lach-U* sequences, which in turn branch separately from *Cyca*-C4 (Fig. 4B). Thus, four distinct lineages of *Mhc* class I sequences appear to be present in the carp.

The apparent relationship between *Cyca*-TC16 and the *Lach* genes clustering together in the tree may have been caused by convergent evolution, or alternatively it indicates that *Cyca*-TC16 is encoded by an old gene, the origin of which dates back to the last common ancestor of carp and coelacanth, approximately 400 million years (my) ago (Norman and Greenwood 1975). The considerable amino acid identity (46%) between *Cyca*-TC16 and *Lach-UA*01* would then have to be the result of purifying selection aimed at maintaining the

protein sequence. In contrast, under the assumption that *Cyca*-TC16 could be an old pseudogene, 400 my of neutral evolution would have eliminated any similarity between these sequences. The *Lach-U* sequences are present in high numbers in the genome of the coelacanth (at least 34 sequences in one individual), and these sequences possess the conserved residues involved in binding antigenic peptides and $ß_2$-microglobulin. Nevertheless, absence of the CD8-binding loop, and Cys203 in some *Lach-UB* sequences and *Cyca*-C4, were in the former case taken as indications for a non-classical function (Betz *et al.* 1994). Whether functional similarities between the proteins encoded by the *Lach-U* sequences and *Cyca*-TC16 were also retained through 400 my of evolution, awaits the isolation of the remainder of *Cyca*-TC16.

Genes of the *Cyca-U* lineage have so far been found in three cyprinids (carp (Van Erp *et al.* 1996a), zebrafish (Takeuchi *et al.* 1995) and large African barbel (Dixon *et al.* unpublished data), and two salmonids (Atlantic and pink salmon (Grimholt *et al.* 1993; Katagiri *et al.* unpublished data)), indicating that the *U* genes were present before the cyprinid-salmonid split, probably as long as 120-150 my ago. Genes of the *Z* lineage have so far been found only in the common carp and the closely related ginbuna crucian carp. However, under the assumption that lineages arise through repeated duplications from an ancestral gene, the sequence dissimilarity of *Cyca-Z* with the *U* and *Cyca*-TC16 sequences seems to point at an old age of the *Z* lineage as well. In this scenario the age of each of these lineages implies that representatives of *U* and perhaps *Z* were possibly present in the early representatives of all cyprinids and salmonids. However, genes of the *Z* lineage may have been lost from all other but the common and ginbuna crucian carp. The status of *Cyca*-TC16 in comparison to the *Lach-U* sequences from the coelacanth implies that orthologous genes could potentially be present in all actinopterygians. In many cases the descendants of the ancestral gene will undoubtedly have been lost or diverged beyond recognition, but finding of sequences of this lineage in other fish species will further elucidate the nature and function of these old sequences.

Acknowledgements

We would like to thank Dr. Jan Klein for fruitful discussions, and for his hospitality during my (SvE) work in Tübingen. The critical advice and assistance by Dr. Felipe Figueroa is gratefully appreciated.

References

Betz, U.A.K., Mayer, W.E. and Klein, J. 1994. Major histocompatibility complex class I genes of the coelacanth *Latimeria chalumnae*. *Proc. Natl. Acad. Sci. USA 91:* 11065-11069

Burmeister, W.P., Gastinel, L.N., Simister, N.E., Blum, M.L. and Bjorkman, P.J. 1994. Crystal structure at 2.2 Å resolution of the MHC-related neonatal Fc receptor. *Nature 372:* 336-343

Dixon, B., Van Erp, S.H.M., Rodrigues, P.N., Egberts, E. and Stet, R.J.M. 1995. Fish major histocompatibility complex genes: an expansion. *Dev. Comp. Immunol. 19:* 109-133

Flajnik, M.F., Canel, C., Kramer, J. and Kasahara, M. 1991. Evolution of the major histocompatibility complex: molecular cloning of major histocompatibility complex class I from the amphibian *Xenopus*. *Proc. Natl. Acad. Sci. USA 88:* 537-541

Fremont, D.H., Matsumura, M., Stura, E.A., Peterson, P.A. and Wilson, I.A. 1992. Crystal structures of two viral peptides in complex with murine MHC class I H-2K^b. *Science 257:* 919-927

Grimholt, U., Hordvik, I., Fosse, V.M., Olsaker, I., Endresen, C. and Lie, Ø. 1993. Molecular cloning of major histocompatibility complex class I cDNAs from Atlantic salmon (*Salmo salar*). *Immunogenetics 37:* 469-473

Grossberger, D. and Parham, P. 1992. Reptilian class I major histocompatibility complex genes reveal conserved elements in class I structure. *Immunogenetics 36:* 166-174

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

Hashimoto, K., Nakanishi, T. and Kurosawa, Y. 1992. Identification of a shark sequence resembling the major histocompatibility complex class I α_3 domain. *Proc. Natl. Acad. Sci. USA 89:* 2209-2212

Higgins, D.G., Bleasby, A.J. and Fuchs, R. 1992. CLUSTAL V: improved software for multiple sequence alignment. *Comput. Appl. Biosci. 8:* 189-191

Katagiri, T., Hirono, I., Sakaki, M., Aoki, T., Hashimoto, K. and Kurosawa, Y. Isolation of major histocompatibility complex antigen class I cDNA from pink salmon *Oncorhynchus gorbuscha*. Unpublished. EMBL/genbank accession numbers D58379-D58395.

Klein, J., Bontrop, R.E., Dawkins, R.L., Erlich, H.A., Gyllensten, U.B., Heise, E.R., Jones, P.P., Parham, P., Wakeland, E.K. and Watkins, D.I. 1990. Nomenclature for the major histocompatibility complexes of different species: a proposal. *Immunogenetics 31:* 217-219

Klein, J., Ono, H., Klein, D. and O'hUigin, C. 1993. The accordion model of *Mhc* evolution. In: *Progress in Immunology* (ed. by J. Gergely and G. Petranyi), p.137. Springer, Heidelberg.

Kumar, S., Tamura, K. and Nei, M. 1993. *MEGA: Molecular Evolutionary Genetics Analysis, version 1.01*. The Pennsylvania State University, University Park.

Norman, J.R. and Greenwood, P.H. 1975. *A History of Fishes*. 3rd edn, p.356. Ernest Benn, London.

Okamura, K., Nakanishi, T., Kurosawa, Y. and Hashimoto, K. 1993. Expansion of genes that encode MHC class I molecules in cyprinid fishes. *J. Immunol. 151:* 188-200

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

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

Palmiter, R.D. 1974. Magnesium precipitation of ribonucleoprotein complexes: expedient techniques for isolation of undegraded polysomes and messenger ribonucleic acid. *Biochemistry 13:* 3606-3615

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

Salter, R.D., Benjamin, R.J., Wesley, P.K., Buxton, S.E., Garrett, T.P.J., Clayberger, C., Krensky, A.M., Norment, A.M., Littman, D.R. and Parham, P. 1990. A binding site for the T-cell co-receptor CD8 on the α_3 domain of HLA-A2. *Nature 345:* 41-46

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

Saper, M.A., Bjorkman, P.J. and Wiley, D.C. 1991. Refined structure of the human histocompatibility antigen HLA-A2 at 2.6 Å resolution. *J. Mol. Biol. 219:* 277-319

Shawar, S.M., Vyas, J.M., Rogers, J.R. and Rich, R.R. 1994. Antigen presentation by major histocompatibility complex class I-b molecules. *Annu. Rev. Immunol. 12:* 839-880

Shum, B., Avila, D., Du Pasquier, L., Kasahara, M. and Flajnik, M. 1993. Isolation of a classical MHC class I cDNA from an amphibian. Evidence for only one class I locus in the *Xenopus* MHC. *J. Immunol. 151:* 5376-5386

Stet, R.J.M., Van Erp, S.H.M., Hermsen, T., Sültmann, H.A. and Egberts, E. 1993. Polymorphism and estimation of the number of *MhcCyca* class I genes in laboratory strains of the common carp (*Cyprinus carpio* L.). *Dev. Comp. Immunol. 17:* 141-156

Stroband, H.W.J., Stevens, C., Te Kronnie, G., Samallo, J., Schipper, H., Kramer, B. and Timmermans, L.P.M. 1995. Expression of *carp-cdx1*, a *caudal* homolog, in embryos of the carp, *Cyprinus carpio*. *Roux's Arch. Dev. Biol. 204:* 369-377

Takeuchi, H., Figueroa, F., O'hUigin, C. and Klein, J. 1995. Cloning and characterization of class I *Mhc* genes of the zebrafish, *Brachydanio rerio*. *Immunogenetics 42:* 77-84

Van Erp, S.H.M., Dixon, B., Figueroa, F., Egberts, E. and Stet, R.J.M. 1996a. Identification and characterization of a new *Mhc* class I gene in carp (*Cyprinus carpio* L.). *Immunogenetics* in press.

Van Erp, S.H.M., Egberts, E. and Stet, R.J.M. 1996b. Characterization of major histocompatibility complex class II *A* and *B* genes in a gynogenetic carp clone. *Immunogenetics* in press

Chapter 6

Characterization of major histocompatibility complex class II *A* and *B* genes in a gynogenetic carp clone

Saskia H.M. van Erp, Egbert Egberts and René J.M. Stet

Department of Experimental Animal Morphology and Cell Biology, P.O. Box 338, 6700 AH Wageningen, Wageningen Agricultural University, The Netherlands

Accepted for publication in **Immunogenetics (1996)**

Summary

A prerequisite for performing functional studies on major histocompatibility complex molecules of fish is the availability of genetically well-defined homozygous strains. Previously we have applied gynogenetic reproduction to generate isogenic carp, denoted clone A410. This clone has recently been demonstrated to express a single class I gene, *Cyca-UA1*01*, and in the present study two class II *B* and two class II *A* transcripts were obtained. The two class II *B* transcripts, *Cyca-D*(CB3)*B* and *Cyca-D*(CB4)*B*, as well as the class II *A* transcripts, *Cyca-D*(10A)*A* and *Cyca-D*(15A)*A* appear to be *bona fide* class II transcripts based on the presence of conserved protein characteristics of the inferred class II molecules. With the isolation of class II *A* sequences representatives of all major classes of *Mhc* genes have been identified in the carp. To assess the relationship between the different class II genes, segregation studies, comparison of cDNA and intron 1 sequence data, and phylogenetic analyses were performed. These studies show that the class II *B* transcripts, *Cyca-D*(CB3)*B* and *Cyca-D*(CB4)B, are derived from related, closely linked loci. In addition, these studies indicate that the previously described *Cyca-DAB*01* and *Cyca-DAB*02* are also closely linked, but that this linked pair segregates independently from the *Cyca-D*(CB3)*B* and *Cyca-D*(CB4)*B* loci. The class II *A* transcripts most likely are derived from separate loci and do not represent alleles, as they were found not to segregate in the individuals of the clone which was generated by meiogenetic gynogenesis.

The nucleotide sequences reported in this paper have been submitted to the EMBL/GenBank nucleotide sequence databases and have been assigned the accession numbers X95431-X95435, Z47730-Z47733, and Z47757.

Introduction

Major histocompatibility complex (*Mhc*) sequences have been isolated from numerous species of fish, mostly from modern bony fishes (Teleostei) (Dixon *et al.* 1995; Takeuchi *et al.*1995; Van Erp *et al.* 1996a), but also from cartilaginous fish (Elasmobranchii) (Dixon *et al.* 1995), and a representative of the lobe-finned fishes (sarcopterygii), the coelacanth (Betz *et al.* 1994). In the common carp (*Cyprinus carpio* L.) *Mhc* sequences found to date include two lineages of class II *B* genes (*Cyca-YB* (Hashimoto *et al.* 1990) and *Cyca-DAB* (Ono *et al.* 1993)), sequences encoding $ß_2$-microglobulin (*Cyca-B2m*) (Dixon *et al.* 1993), and sequences from multiple lineages of class I genes (*Cyca-Z* (Hashimoto *et al.* 1990; Okamura *et al.* 1993), *Cyca-UA* (Van Erp *et al.* 1996a), *Cyca*-TC16 (Van Erp *et al.* 1996b) and *Cyca*-C4 (Grossberger and Parham 1992)), some of which are probably non-classical or perhaps even pseudogenes. Transcription has so far only been firmly established for the class II *B Cyca-DAB* genes and class I *Cyca-UA* genes, which therefore potentially are the most likely candidates to encode functional *Mhc* molecules.

As far as can be judged at this stage, not only do *Mhc* sequences of both class I and II

appear to be present in all jawed fishes, but comparison of the deduced protein sequences with those of *Mhc* molecules of mouse and man also has suggested major similarities in the functioning of *Mhc* molecules of fish and mammals (Stet and Egberts 1991; Dixon *et al.* 1995). Nevertheless, functional studies on fish *Mhc* molecules have so far not been reported. For such studies the availability of genetically well-defined homozygous strains is a crucial prerequisite. Production of such strains by conventional inbreeding would require 20 generations of brother-sister matings to reach acceptable levels of homozygosity. In contrast, inbreeding by gynogenetic or androgenetic reproduction reduces the number of generations to a minimum of two. In our laboratory, gynogenesis in the common carp has been studied and used for many years (Kaastrup *et al.* 1989; Komen *et al.* 1991; Wiegertjes *et al.* 1994), and we have applied this technique to generate a strain of homozygous identical fish, denoted clone A410. This strain was generated by four successive gynogenetic reproductions, starting with a single female A4 from the Israelian DOR70-strain (Van Erp *et al.* 1996a). We have now undertaken to type this carp strain for the expression of *Mhc* genes, in particular of *Cyca-UA* and *Cyca-DAB*, which most likely encode functional classical *Mhc* molecules. In a previous study we have already described the *Cyca-UA* sequence present in this strain (Van Erp *et al.* 1996a), and the aim of the present study is to characterize the class II *A* and *B* sequences expressed in this strain. In addition, class II *A* sequences have not yet been identified in carp, and their isolation and characterization would thus complete the set of *Mhc* genes in carp.

Materials and methods

Fish

The fish used in these experiments were all laboratory strains of the common carp (*Cyprinus carpio* L.). Clone A410 is an isogenic carp line obtained through gynogenesis (Van Erp *et al.* 1996a). A4.10me2 is a second-generation meiotic gynogenetic carp family, generated from a first-generation meiotic female, A4-10me1 (Kaastrup *et al.* 1989). R3xR8 are the F1-hybrid offspring of a cross between a male of Polish origin (R3 strain) and a female of Hungarian origin (R8 strain)(Wiegertjes *et al.* 1994).

Screening of the A410 thymus cDNA library

A cDNA library in λgt10 was constructed from pooled thymi from ten 5-months old individuals of the clone A410, as previously described (Van Erp *et al.* 1996a). A total of 1×10^6 PFU from the A410 thymus cDNA library were screened with the *Cyca-D*(RdB1)*A* fragment essentially as described by Van Erp and co-workers (1996a), now using hybridization solutions containing 50% formamide. Positive plaques were identified after exposure for 24 hours at

-80°C using intensifying screens, and single positive plaques were isolated after a second screening using the same procedure.

Polymerase chain reaction

Genomic DNA was extracted from tissues as described (Stet *et al.* 1993). Amplification by PCR was performed in 100 μl of a solution containing 200 ng of DNA template in 1 x Goldstar reaction buffer, 1.5 mM $MgCl_2$, 0.2 mM of each dNTP, 0.2 μM of each primer and 1 unit of Goldstar thermostable DNA polymerase (Eurogentec, Seraing, Belgium). Amplification was carried out using a thermal cycle profile (1 min 94°C, 1 min 55°C, 2 min 72°C) for 30 cycles, with an additional extension step of 10 min at 72°C.

Primers used in the amplifications were:

OL93-139: 5'-CTGATGCTGTCTGCTTTCACTGGAGCA-3'
OL93-140: 5'-CTGTTTTATCACGGATCGCTGACTG-3'
OL94-23 : 5'-GATTTGAGCATTATGTTTGCA-3'
exIDAB4 : 5'-ATGCTGTCTGCATTTACTGGAACAG-3'
DAB04s : 5'-CTCTGCTGCAGTTCTGCC-3'
DAB06s : 5'-TGTCCACTGAAGTTTTCAGA-3'
DXA-fw : 5'-GCTCAAGCTGAGCACAGGG-3'
DXA-rev : 5'-CTCTTCTGGAGAGTTGTATGC-3'

In addition, λgt10 and λgt11 specific primers (Promega, Madison WI, USA) were used in anchored PCR in combination with specific primers, and in amplifications of cDNA inserts from positive cDNA clones.

Sequence analysis

Selected fragments were cloned in pTZ18R and pTZ19R (Stratagene, La Jolla CA, USA), in pUC18 using the Sureclone ligation kit (Pharmacia, Uppsala, Sweden), or in pGEM (Promega) according to the manufacturers' protocols. Recombinant clones were sequenced using the Mn^{2+} Sequenase DNA sequencing kit (USB, Amersham, UK). Sequence data were analyzed with several programs from the GCG-package (Genetics Computer Group, Madison WI, USA), ClustalV (Higgins *et al.* 1992) and MEGA (Kumar *et al.* 1993).

***Mhc* haplotyping**

CycaD(CB3)*B* and *Cyca-D*(CB4)*B* were identified by PCR using forward primer exIDAB4 (complementary to the end of exon 1 of *CycaD*(CB3)*B* and *Cyca-D*(CB4)*B*) in combination with reverse primer OL94-23 (complementary to the end of exon 2 of both *CycaD*(CB3)*B* and *Cyca-D*(CB4)*B*). Fragments obtained were sequenced. Subsequently, typing was performed

by PCR using sequence-specific reverse primers (primer DAB04s to *CycaD*(CB3)*B*, and primer DAB06s to *Cyca-D*(CB4)*B*), in combination with forward primer exIDAB4. Identification was further confirmed by size analysis of the products.

*Cyca-DAB*01*- and *Cyca-DAB*02*-like sequences were identified using PCR amplification with primers specific to the end of exon 1 (OL93-139) and the end of exon 2 (OL93-140), followed by sequence analysis. Subsequently, fish were typed by PCR-RFLP using restriction endonuclease *Rsa*I.

Results

Identification of class II *B* transcripts

In order to isolate *MhcCyca* class II *B* transcripts from clone A410, anchored PCR was performed on a thymus cDNA library derived from this strain. Amplification using anti-sense primer OL94-23, complementary to codons 80-86 of *Cyca-DAB*02* (Ono *et al.* 1993), in conjunction with λgt10-specific primers, yielded a fragment of approximately 300 bp. Sequence analysis of this fragment showed that it contained a single *Cyca-DAB*-like sequence (denoted *Cyca-D*(CB3)*B*), which was different from both the previously reported *Cyca-DAB*01* and *Cyca-DAB*02* (Ono *et al.* 1993). To obtain the 3' end of *Cyca-D*(CB3)*B*, anchored PCR was subsequently performed using a sense primer, denoted EXIDAB4, complementary to codons -5 to 3 of *Cyca-D*(CB3)*B*. The resulting fragment, approximately 1400 bp in size, contained two different sequences (Fig. 1). One was identical to *Cyca-D*(CB3)*B* in the overlapping region, whereas the other (denoted *Cyca-D*(CB4)*B*) was different from *Cyca-D*(CB3)*B*, and also different from *Cyca-DAB*01* and *Cyca-DAB*02*. Finally, the 5' end of *Cyca-D*(CB4)*B* was obtained by anchored PCR using a specific anti-sense primer, denoted DAB06s, complementary to codons 72-77 of *Cyca-D*(CB4)*B*. The sequences of *Cyca-D*(CB3)*B* and *Cyca-D*(CB4)*B*, obtained in this manner, are 1457 and 1467 bp in size, respectively, of which 720 and 747 bp constitute the coding regions, flanked by 3' untranslated regions (Fig. 1). Neither of the two clones appear to contain the complete leader-encoding sequence and the startcodon. *Cyca-D*(CB3)*B* and *Cyca-D*(CB4)*B* are 92% identical in their coding regions, whereas in the 3'UT regions they are 99% identical. Similarity of *Cyca-D*(CB3)*B* and *Cyca-D*(CB4)*B* to *Cyca-DAB*01* and *Cyca-DAB*02* is considerable when comparing the coding sequences (79% identity), however, the 3'UT regions of *Cyca-D*(CB3)*B*/*Cyca-D*(CB4)*B* are not easily alignable to those of *Cyca-DAB*01*/*Cyca-DAB*02*, which are identical.

Comparison of the derived amino acid sequence of *Cyca-D*(CB3)*B* and *Cyca-D*(CB4)*B* with other fish class II *B* sequences, shows that *Cyca-D*(CB3)*B* and *Cyca-D*(CB4)*B* share many of the protein features found in fish class II ß chains (Fig. 2): (i) a putative N-linked

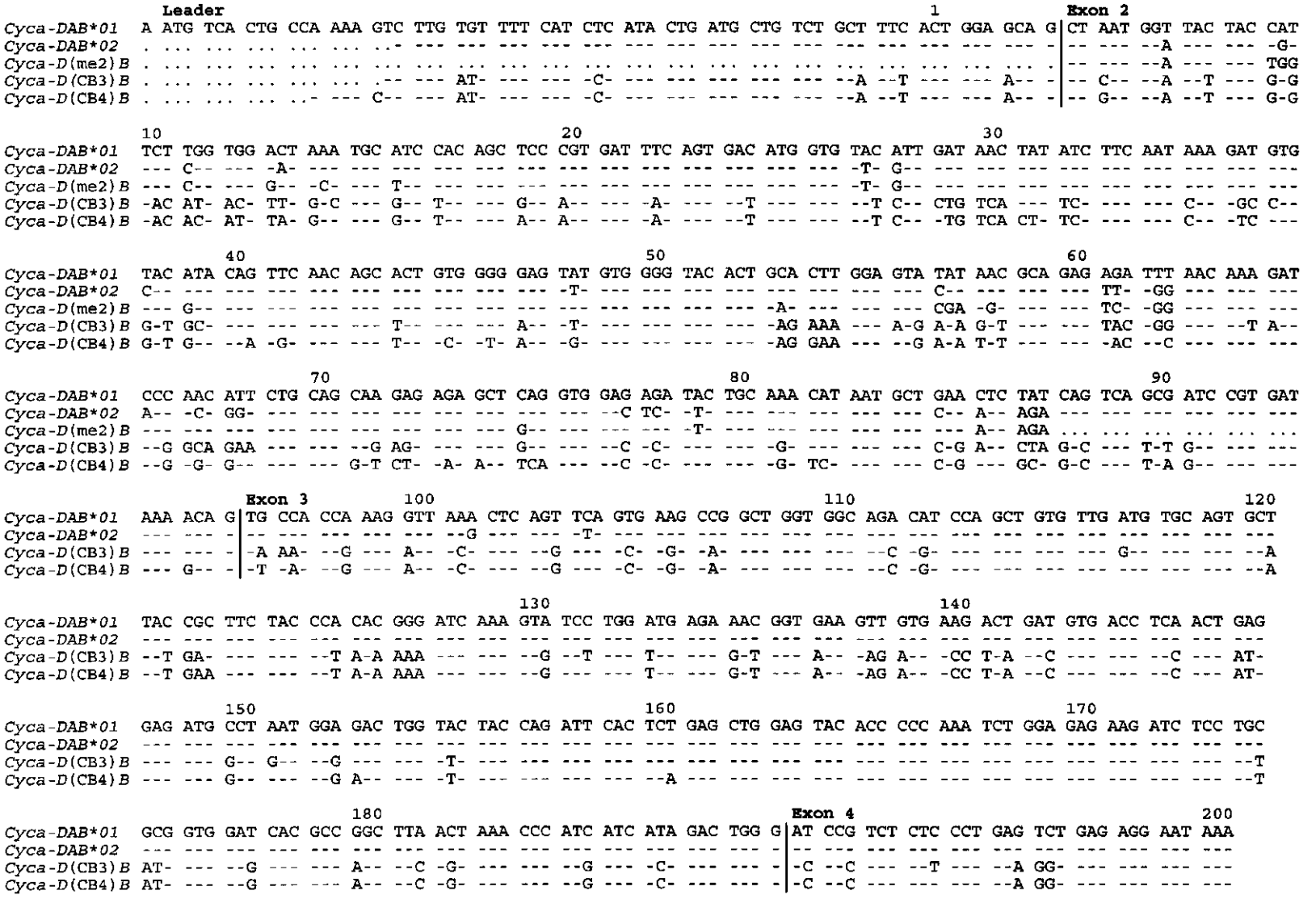
glycosylation site is present at amino acid (aa) positions 42-44, as found in all teleosts except striped bass and cichlid, which contain a glycosylation site at aa positions 18-20; (ii) cysteine residues are present in both the β_1 and β_2 domain (C15, C80, C118, C174), as commonly found in class II ß chains, but *Cyca-D*(CB4)*B* possesses two additional cysteine residues (aa positions 41 and 48) in the β_1 domain, which were also observed in *Brre-DAB2*01* from zebrafish; (iii) conserved residues found to be involved in peptide binding in mammalian class II molecules (W61, H81, N82, *HLA-DR1* numbering) are well conserved in *Cyca-D*(CB3)*B* and *Cyca-DAB*02* (W62, H82, N83, carp numbering), however in *Cyca-D*(CB4)*B* only one of these residues (N83) is present, and in *Cyca-DAB*01* only two (H82 and N83); (iv) a highly conserved stretch is found at aa positions 142-146, which corresponds to part of the region found in mammals to be involved in CD4 binding; and (v) in the transmembrane region, regularly spaced glycines are present, hypothesized to be involved in pairing with the transmembrane region of the class II α chain (Fig. 2).

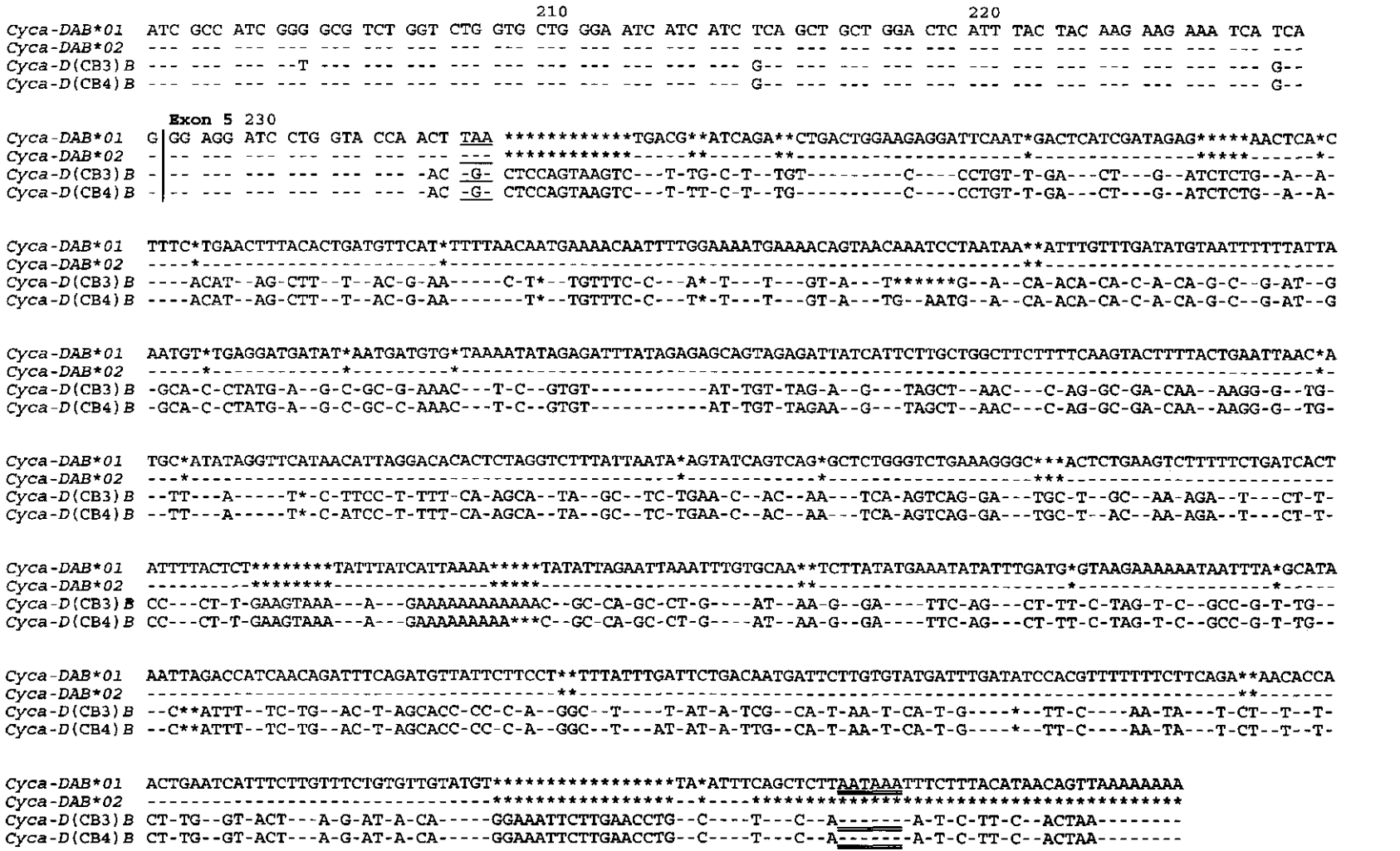
In analyzing the relationships between *Cyca-D*(CB3)*B*, *Cyca-D*(CB4)*B*, *Cyca-DAB*01* and *Cyca-DAB*02* in terms of being alleles or loci, comparison of intron sequences can be highly informative. Therefore, genomic fragments spanning intron 1 and exon 2 were amplified from each of the four class II *B* sequences. Amplification of *Cyca-D*(CB3)*B* and *Cyca-D*(CB4)*B* was performed on genomic DNA from the parental female of clone A410, whereas *Cyca-DAB*01* and *Cyca-DAB*02* genomic fragments were amplified from DNA from an individual from the A4.10me2 family.

The introns 1 of *Cyca-D*(CB3)*B* and *Cyca-D*(CB4)*B* were found to be 669 bp and 561 bp in size, respectively (Fig. 3A). A remarkable feature shared by these introns is the presence of a non-consensus 5'-splice site GCAAGT instead of the consensus GTAAGT. In spite of

Figure 1 (next two pages). Nucleotide alignment of *Cyca-DAB* sequences. Dashes denote identity to *Cyca-DAB*01*, asterisks denote gaps included for optimal alignment. Dots indicate absence of sequence information. Putative exon boundaries are based on personal observation, and on the genomic organisation of zebrafish *Brre-DAB* (Sültmann *et al.* 1994). Stop codons and poly-adenylation signals are singly and doubly underlined, respectively. Numbering denotes codon numbers within putative mature protein. Sources of sequences as in figure 6.

Figure 2 (page 116). Alignment of teleost class II *B*-derived amino acid sequences. Dashes denote identity to *Cyca-DAB*01*, asterisks denote gaps included for optimal alignment. Dots indicate absence of sequence information. Triangles indicate conserved cysteine residues, squares indicate an N-linked glycosylation site, spades indicate positions of conserved peptide-binding residues in mammalian class II *B* sequences, circles denote the region involved in CD4 binding in mammals, and diamonds denote conserved residues of the transmembrane region. Numbering is as in figure 1. Sources of sequences as in figure 6.


Leader
Exon 2
Exon 3
Exon 4
Cyca-DAB*01
Cyca-DAB*02
Cyca-D(me2)B
Cyca-D(CB3)B
Cyca-D(CB4)B


210
220
Exon 5 230
Cyca-DAB*01
Cyca-DAB*02
Cyca-D(CB3)B
Cyca-D(CB4)B

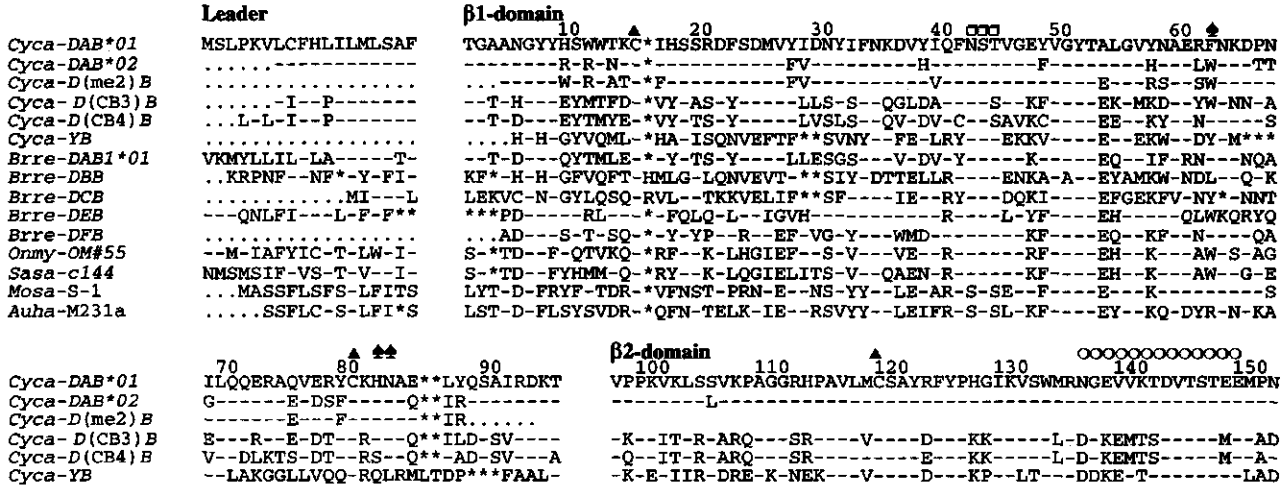


Figure 2. (legend page 113)

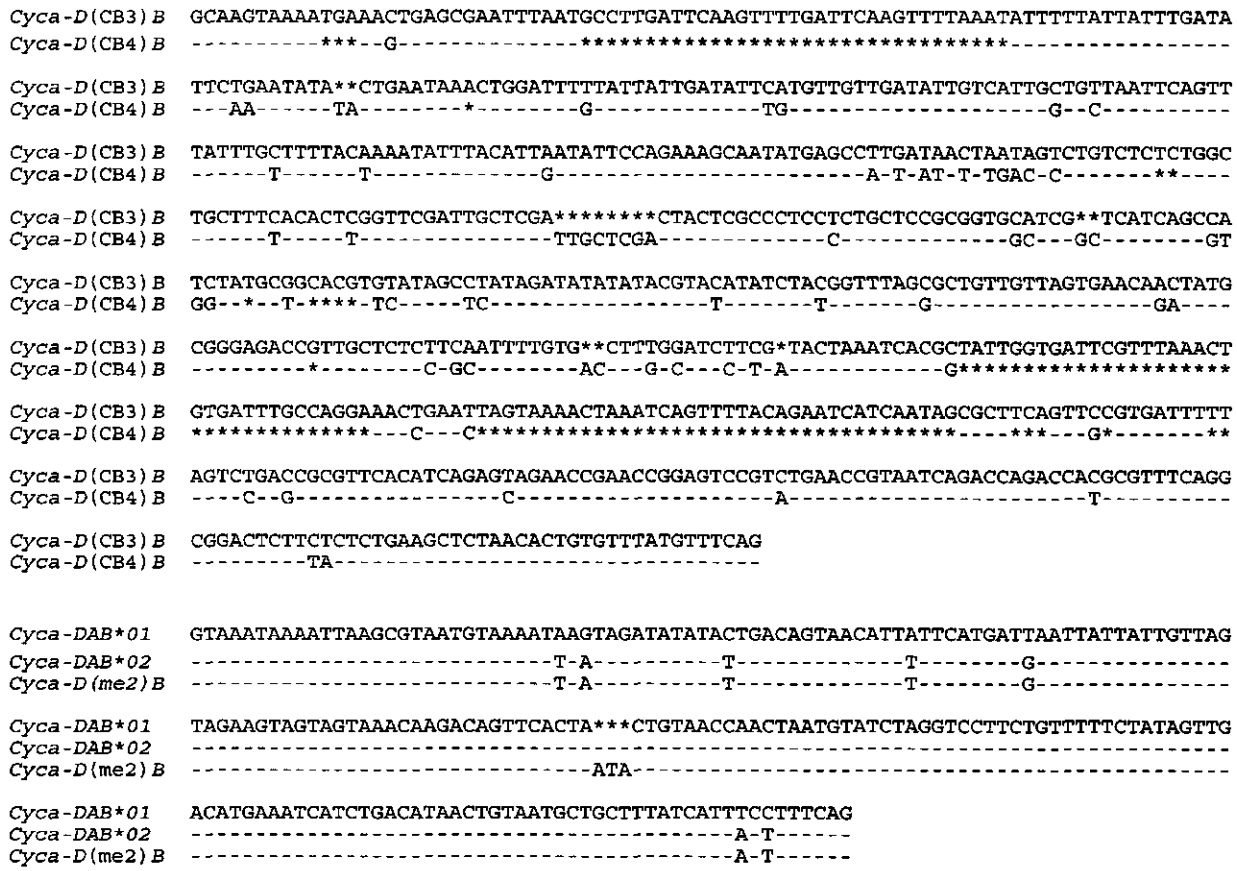


Figure 3. Alignment of intron 1 sequences of
A. *Cyca-D*(CB3)*B* and *Cyca-D*(CB4)*B*. Dashes denote identity to *Cyca-D*(CB3)*B*.
B. *Cyca-DAB*01*, *Cyca-DAB*02* and *Cyca-D*(me2)*B*. Dashes denote identity to *Cyca-DAB*01*.
Asterisks indicate gaps introduced for optimal alignment.

the difference in size, the intron 1 sequences of *Cyca-D*(CB3)*B* and *Cyca-D*(CB4)*B* are well alignable (89% identity). In contrast, the introns 1 of *Cyca-DAB*01* and *Cyca-DAB*02* were both found to be 204 bp in size, and they were 96.6% identical. Besides *Cyca-DAB*01* and *Cyca-DAB*02* a third genomic sequence was amplified from the A4.10me2 individual. This sequence, denoted *Cyca-D*(me2)*B*, showed high similarity to *Cyca-DAB*01* and *Cyca-DAB*02*, both in coding and intron 1 sequence, which was, however 207 bp in size (Fig. 2 and 3B). In contrast, none of these introns could be aligned with the intron 1 sequences of *Cyca-DAB*(CB3)*B* and *Cyca-DAB*(CB4)*B* (Fig. 3).

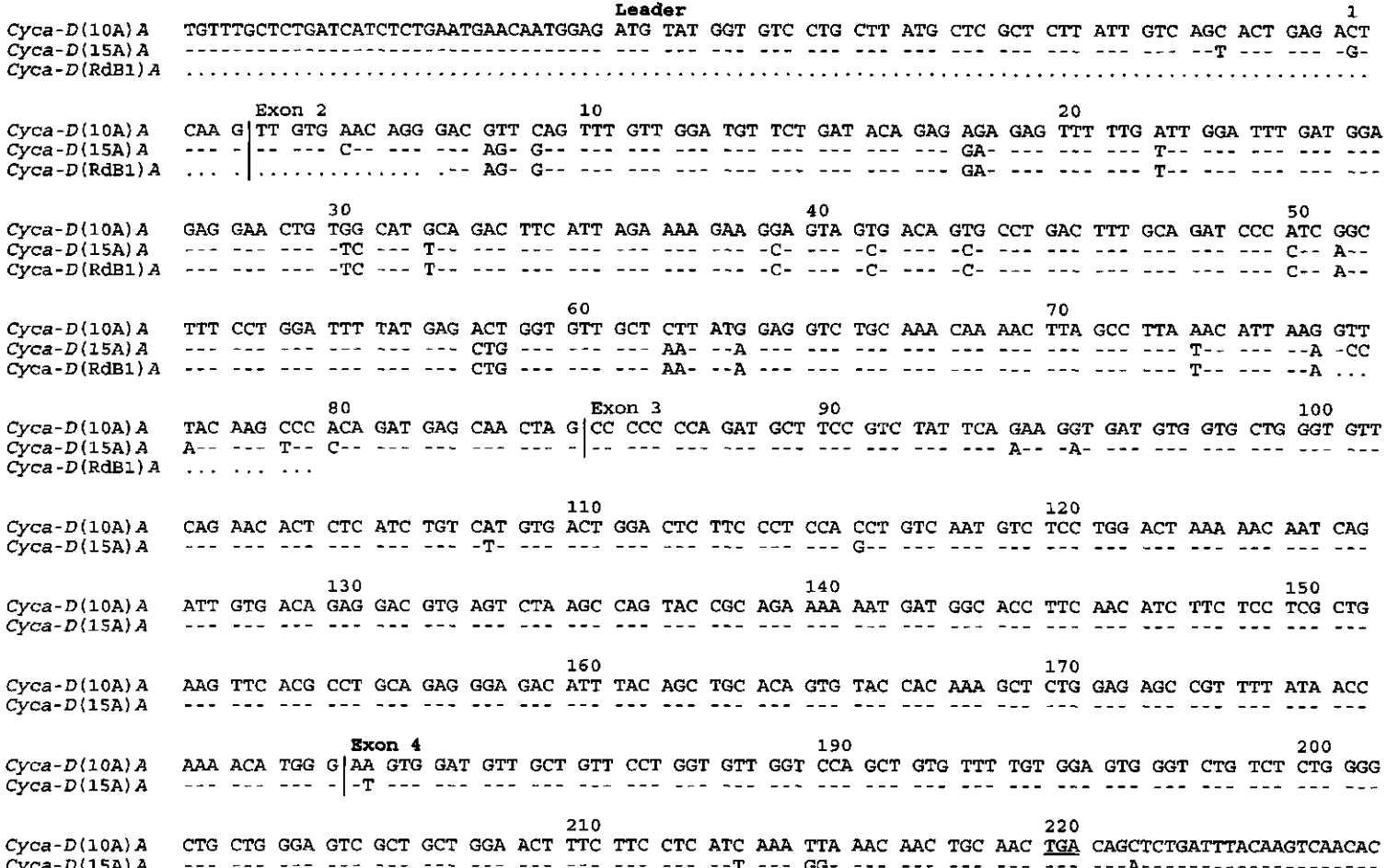
Preliminary segregation studies were subsequently performed using the information obtained from the genomic sequences. To this end, two gynogenetic carp families, clone A410 and family A4.10me2, and their respective female parent were typed. In addition, segregation analyses were performed in a hybrid carp family R3xR8 (described by Wiegertjes and co-workers, 1995). In clone A410, all individuals tested (n=8), as well as their parental female carp were typed to contain *Cyca-D*(CB3)*B* and *Cyca-D*(CB4)*B*. Within the A4.10me2 family, 12 individuals as well as their parent were typed. The parental female was shown to contain *Cyca-DAB*01*, *Cyca-DAB*02* and *Cyca-D*(me2)*B*. Of the 12 meiotic gynogenetic daughters typed, three individuals were found to contain *Cyca-DAB*01* and *Cyca-DAB*02*, four individuals contained only *Cyca-D*(me2)*B*, whereas five individuals contained *Cyca-DAB*01*, *Cyca-DAB*02* and *Cyca-D*(me2)*B*. Thus, *Cyca-DAB*01* and *Cyca-DAB*02* were consistently found to co-segregate, separately from *Cyca-D*(me2)*B*. Within the R3xR8 family 70 individuals as well as the R3 and R8 parental individuals were typed. Individual R3 was found to contain *Cyca-DAB*01* and *Cyca-DAB*02*, whereas individual R8 contained *Cyca-DAB*01*, *Cyca-DAB*02*, *Cyca-D*(CB3)*B* and *Cyca-D*(CB4)*B*. In the offspring three genotypes were observed: 34 out of 70 individuals contained *Cyca-DAB*01* and *Cyca-DAB*02*, 11 out of 70 individuals contained *Cyca-D*(CB3)*B* and *Cyca-D*(CB4)*B*, and 25 out of 70 individuals contained *Cyca-DAB*01*, *Cyca-DAB*02*, *Cyca-D*(CB3)*B* and *Cyca-D*(CB4)*B*. Thus, a consistent co-segregation was observed of *Cyca-DAB*01* with *Cyca-DAB*02*, and of *Cyca-D*(CB3)*B* with *Cyca-D*(CB4)*B*.

Isolation and identification of class II *A* transcripts.

For the isolation of class II *A* transcripts, anchored PCR was performed on the A410 thymus cDNA library, using anti-sense primer *DXA*-rev, complementary to codons 76-82 in exon 2 of zebrafish class II *A* cDNA clone 2.1.4 (Sültmann *et al.* 1993). The PCR product contained several bands, of which a fragment of 430 bp was isolated. This fragment was subsequently used as template in a nested PCR, again using anti-sense primer *DXA*-rev, but

now in conjunction with sense primer *DXA*-fw, complementary to codons 1-6 of zebrafish class II *A* clone 2.1.4. The product of this nested PCR consisted of a single fragment, with a size of approximately the expected 246 bp. This fragment was cloned and sequenced, and one out of the three recombinant clones analyzed (denoted *Cyca-D*(RdB1)*A* contained a sequence that was 84% identical to exon 2 of zebrafish class II *A* (Fig. 4). In order to obtain full-length carp class II *A* transcripts, *Cyca-D*(RdB1)*A* was then used as a probe to screen the A410 thymus cDNA library. In the first screening a large number of positive plaques were identified (approximately 0.1% of the PFU). Sixteen of these were chosen for second screening, upon which 14 remained positive. However, from only three of these λ-clones a hybridizing insert could be amplified using λgt10-specific primers. Two of these positive clones, denoted *Cyca-D*(10A)*A* and *Cyca-D*(15A)*A*, were completely sequenced (Fig. 4), whereas the third clone was only partially sequenced and was found to be identical to *Cyca-D*(10A)*A*. The transcripts were 1004 bp and 1360 bp in size, and they proved to be different from each other (94% nucleotide (nt) identity). The sequence of *Cyca-D*(15A)*A* was identical with the *Cyca-D*(RdB1)*A* fragment in the entire overlapping region. Both sequences could be identified as class II *A* transcripts based on high similarity to class II *A* sequences from zebrafish (76% nt identity) and striped bass (63% nt identity).

The predicted amino acid sequences of *Cyca-D*(10A)*A* and *Cyca-D*(15A)*A* could be divided into protein domains based on alignments with class II *A* sequences from the above mentioned species including nurse shark (Fig. 5). Similarity among the teleostean class II *A* sequences is high, with substantial stretches of identical sequence. The alignment further illustrates that many protein features of teleostean class II *A* protein sequences are also found in *Cyca-D*(10A)*A* and *Cyca-D*(15A)*A*; in the α_1 domains of *Cyca-D*(10A)*A* and *Cyca-D*(15A)*A* two cysteine residues are present that have also been found in the other teleostean sequences, but not in the class II *A* sequence from nurse shark. A potential glycosylation site is found in the α_2 domain at residues 118-120 only in carp and zebrafish, whereas in striped bass and nurse shark glycosylation sites are found at aa positions 127-129 and 75-77, respectively. Residues that bind the conserved termini of antigenic peptides are less well conserved in teleostean class II *A* sequences. Out of the five such residues found in mammalian class II α chains, only one residue, Nα69, is conserved in both *Cyca-D*(10A)*A* and *Cyca-D*(15A)*A*, whereas another, Nα62, is present only in *Cyca-D*(15A)*A*. *Cyca-D*(10A)*A* and *Cyca-D*(15A)*A* contain the typical regularly spaced glycine residues in the transmembrane region, that play a role in the interaction with the class II ß chain. Also, the residue directly preceding this transmembrane section is proline (P190) instead of the glutamic acid commonly found in mammalian sequences.


Leader
Exon 2
Exon 3
Exon 4
Cyca-D(10A)A
Cyca-D(15A)A
Cyca-D(RdB1)A

```
Cyca-D(10A)A   ATTCTTGTTTTATAATTCAGATTATTTGTACTTCTAATTAACATTCTCCACCAGTTAAAATTCCTTAGATTAAACAATTTGGTACTAAAAACAATAACTT
Cyca-D(15A)A   ----------------------CC----------------------------------------TG--A-------CT-------------C---------

Cyca-D(10A)A   TGTGTACCATTGTACATGACAAACCATGTTCAAATTTGGCTGTCGTTCTAGTCG*AATTATTGCTCTGACACAACATGATTTATTCATTTGTTCAAATAT
Cyca-D(15A)A   ----------------------------------------------CC-------A-----------------------------------------G-

Cyca-D(10A)A   TCATAAGTGATTAAAATTTAATTTTTCTTTTGAAATTTAC.....................................................
Cyca-D(15A)A   ------A------------C--------------------AGCAAAGACACCATGCAATGGATTTTATGAAGTTCACGAAAACATTCAATACTTATCTAA

Cyca-D(15A)A   AATGAAGATAATAACTCATCATCATCGACAGTTTATGTTTCATCGTAAAGTAAGAGTGTATGTTTCAAGATAATCAAGTCAGTGAAATCAAAAATAATAT

Cyca-D(15A)A   TAATTTCTAGATATTTGTTCACAGACACAAGGCTCTGACACATTTCAAACACTGGACAAAAAATATTGTTTTACTGCACATTTATTTGTTCAAAACAATG

Cyca-D(15A)A   TCATCTATATATATATATATATATATATAGGCCTATAATTATTTTTTTTCTGCCATTTTTGCCTGGATTGGACAAAACAGTTGAAGAAAGAGGGGAAGA
```

Figure 4. Nucleotide alignment of carp class II *A* sequences. Dashes denote identity to *Cyca-D*(10A)*A*. Asterisk indicates gap included for optimal alignment. Dots indicate absence of sequence information. Putative exon boundaries are based on the genomic organization of zebrafish class II *A* (Sültmann *et al.* 1993). Numbering denotes codon numbers within putative mature protein. Stopcodons are underlined.

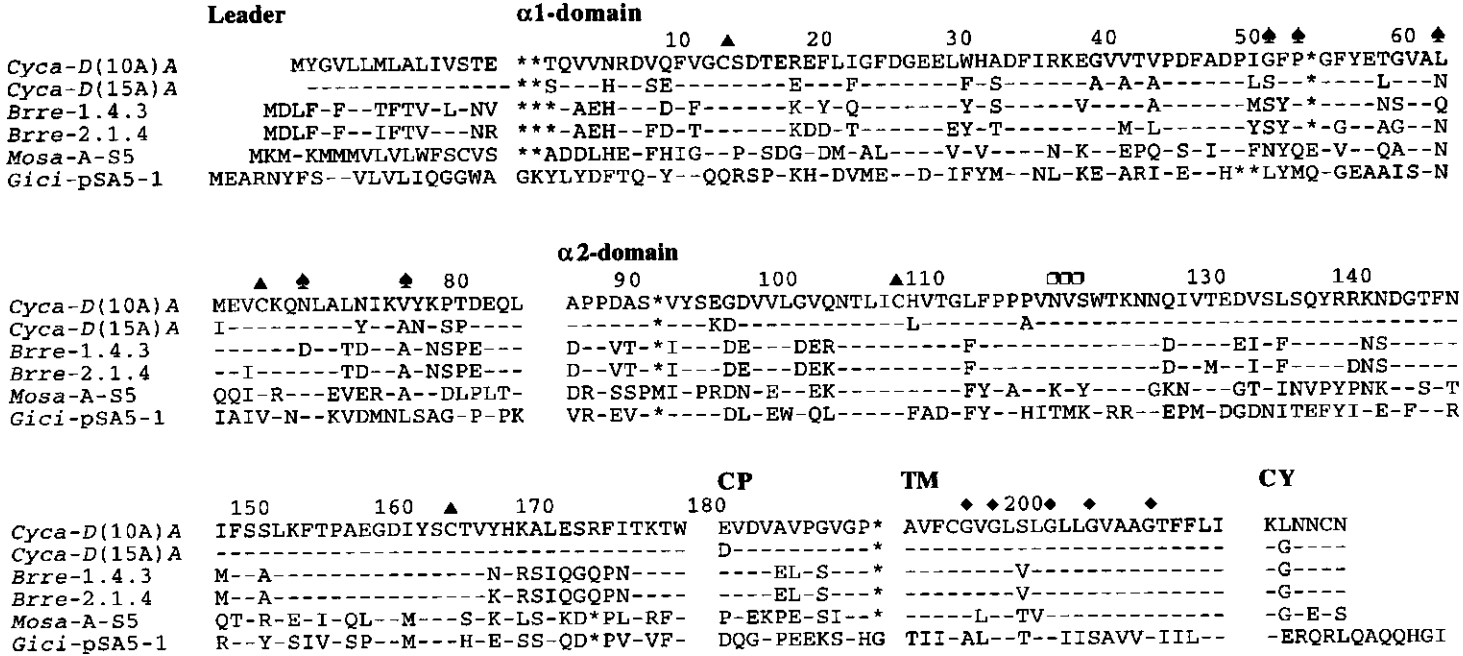


Figure 5. Alignment of fish class II *A* derived amino acid sequences. Dashes denote identity to *Cyca-D*(10A)*A*, asterisks denote gaps included for optimal alignment. Triangles indicate conserved cysteine residues, squares denote an N-linked glycosylation site, and diamonds indicate conserved residues of the transmembrane region. Numbering is as in figure 4. Sources of sequences as in figure 6.

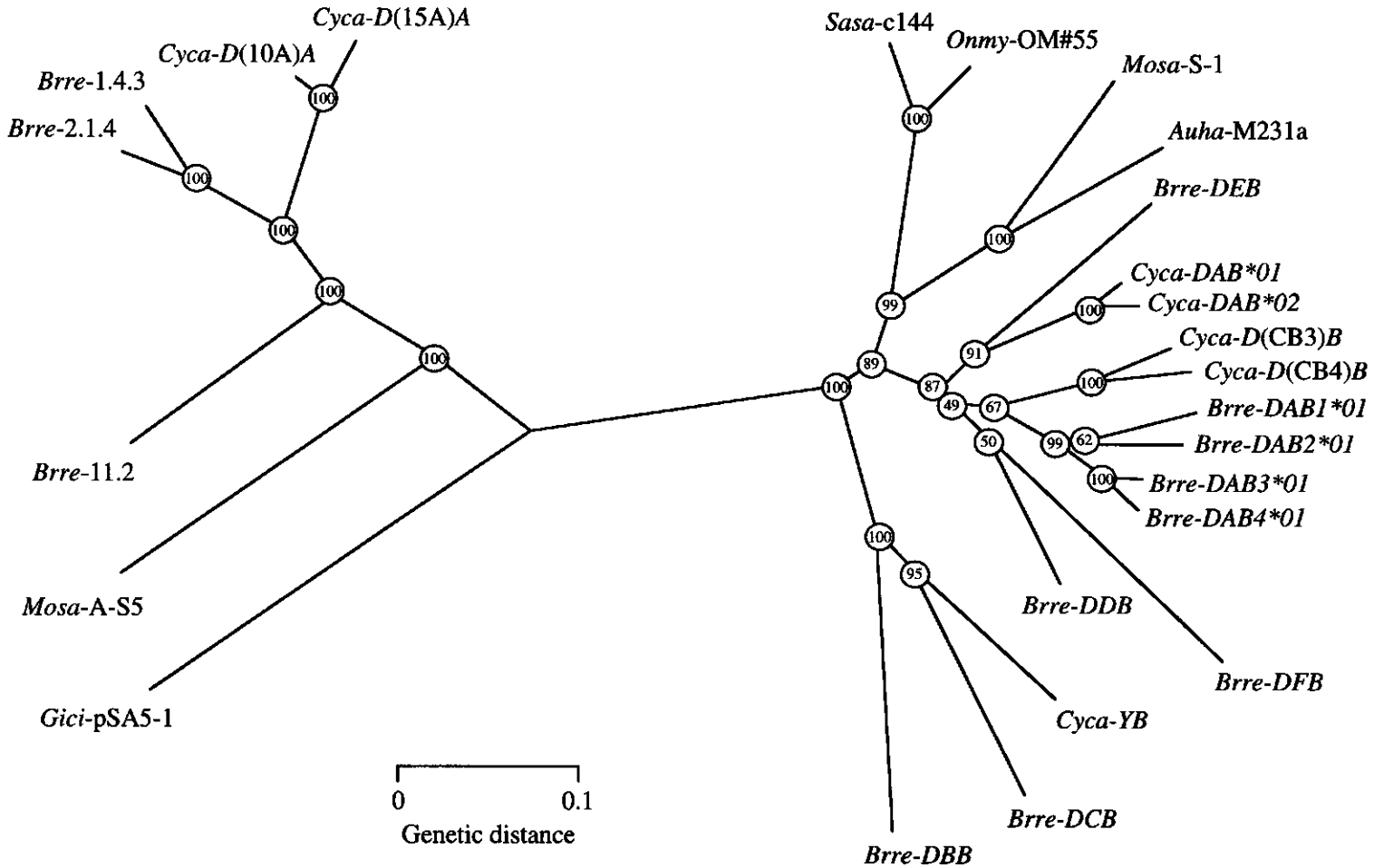


Figure 6. (legend next page)

Phylogenetic analysis

To further analyze the relationship of *Cyca-D*(CB3)*B*, *Cyca-D*(CB4)*B*, *Cyca-D*(10A)*A* and *Cyca-D*(15A)*A* with other class II sequences from teleostean fishes, a neighbor-joining tree was constructed using complete coding sequences of class II genes (Fig. 6). As expected, class II *A* and class II *B* sequences form separate groups. The class II *B* sequences in turn split into three main clusters; the *Cyca-YB* and *Brre-DBB* and *-DCB* form one cluster, as do the class II *B* sequences from Atlantic salmon, rainbow trout, striped bass and the African great lake cichlid *Aulonocara hansbaenshi*. The third cluster consists of the remainder of the carp and zebrafish class II *B* sequences. *Cyca-D*(CB3)*B* and *Cyca-D(CB4)B* group together, in a cluster that also contains all *Brre-DAB* sequences. *Cyca-DAB*01* and *Cyca-DAB*02* cluster with *Brre-DEB*, as previously observed. The *MhcCyca* class II *A* sequences, *Cyca-D*(10A)*A* and *Cyca-D*(15A)*A* group together, in a cluster with zebrafish class II *A* transcripts *Brre*-1.4.3 and *Brre*-2.1.4. The third zebrafish class II *A* sequence, *Brre*-11.2, branches at a considerable distance from this cluster, and no carp orthologue has as yet been found.

Figure 6 (previous page). Phylogenetic tree constructed by the neighbor-joining method (Saitou and Nei 1987) using percentages of nucleotide differences, based on coding sequences of fish class II *A* and class II *B* genes. Numbers on nodes represent percentages of 1000 bootstraps supporting each partitioning. Sources of sequences: *Cyca-DAB*01*, *Cyca-DAB*02*, common carp (Ono *et al.* 1993); *Cyca-YB*, common carp (Hashimoto *et al.* 1990); *Brre-DAB* sequences, zebrafish (Ono *et al.* 1992); *Brre-DBB*, *-DCB*, *-DDB*, *-DEB*, *-DFB*, zebrafish (Sültmann *et al.* 1994); *Sasa*-c144, Atlantic salmon (Hordvik *et al.* 1993), *Onmy-DAB*01*, rainbow trout (Glamann 1995), *Mosa*-S-1, striped bass (Walker and McConnell 1994); *Auha*-M231a, cichlid *Aulonocara hansbaenshi* (Klein *et al.* 1993); *Brre*-1.3.4, -2.1.4, -11.2, zebrafish (Sültmann *et al.* 1993); *Mosa*-A-S5, striped bass (Hardee *et al.* 1995); *Gici*-pSA5-1, nurse shark (Kasahara *et al.* 1992).

Discussion

Experimental studies on function and structure of the Mhc in carp require the availability of well-defined clonal carp lines. The gynogenetic carp clone A410 produced in our laboratory has previously been characterized to express a single class I *A* gene, *Cyca-UA1*01*, and in the present study two class II *B* and two class II *A* transcripts were obtained from this carp line. The two class II *B* transcripts, *Cyca-D*(CB3)*B* and *Cyca-D*(CB4)*B*, appear to be *bona fide* class II ß chain-encoding transcripts based on the presence of conserved protein characteristics of class II ß molecules (Dixon *et al.* 1995; Stet *et al.* in press; Kaufman *et al.* 1994). In order to obtain more insight into the relationship between *Cyca-D*(CB3)*B* and *Cyca-D*(CB4)*B*, as well as their relationship to the previously isolated carp class II *B* transcripts *Cyca-DAB*01* and *Cyca-DAB*02* (Ono *et al.* 1993), the intron 1 sequences of these transcripts were studied. The intron 1 sequences of *Cyca-D*(CB3)*B* and *Cyca-D*(CB4)*B* are different in size, 669 and 561 bp respectively, but besides this the sequences are well alignable. In addition, these introns share the presence of a non-consensus 5' splice-site. Splice-sites which have GC at the 5' intron border have been observed in only 0.2% of 7500 splice-sites examined (Senepathy *et al.* 1990). In only one other *Mhc* sequence, namely in intron 2 of *HLA-DPA2 (SXα)*, has a similar non-consensus 5' splice-site been found, however, *HLA-DPA2* probably is a pseudogene (Boss *et al.* 1985). In contrast, *Cyca-D*(CB3)*B* and *Cyca-D*(CB4)*B* have initially been isolated as correctly spliced cDNA transcripts, and these showed no obvious defects that would suggest them to be pseudogenes. It was, however, hypothesized that the presence of a non-consensus splice-site could cause slower upregulation of expression of the encoded protein (Haviland *et al.* 1991). Notably, transcribed class II *B* sequences which were closely related to *Cyca-D*(CB3)*B* and *Cyca-D*(CB4)*B* and which also contained a similar non-consensus splice-site have been isolated from two closely related cyprinids *Barbus intermedius intermedius* and *Barbus bocagei* (Dixon *et al.* unpublished data). In contrast, the intron 1 sequences of the previously reported carp class II *B* transcripts *Cyca-DAB*01* and *Cyca-DAB*02* contain consensus 5' splice-sites, and the size of these introns is much smaller than those of *Cyca-D*(CB3)*B* and *Cyca-D*(CB4)*B*.

Taken together the intron sequence data seem to suggest that *Cyca-D*(CB3)*B* and *Cyca-D*(CB4)*B* are derived from different loci, as their introns 1 are different in size, and that *Cyca-DAB*01* and *Cyca-DAB*02* may be derived from a single locus, based on only minor differences in intron and untranslated regions.

However, to be more conclusive about this identification, we performed initial segregation studies. Within the clone, A410, all individuals as well as their mother carried both *Cyca-D*(CB3)*B* and *Cyca-D*(CB4)*B*. As the mother of the clone was produced by mitotic

gynogenesis, genetically she is the mere duplication of a haploid egg, and therefore homozygous. The finding that she contains both *Cyca-D*(CB3)*B* and *Cyca-D*(CB4)*B* indicates that these genes are either linked on the same chromosome, or, alternatively, that they are located on two different, non-homologous, homozygous chromosomes. Her offspring, clone A410 was produced by meiotic gynogenesis, and therefore either of both genetic configurations proposed above could account for the presence of both *Cyca-D*(CB3)*B* and *Cyca-D*(CB4)*B* in each of the offspring.

The Jukes-Cantor corrected genetic distance (d_s) calculated from synonymous substitutions between *Cyca-D*(CB3)*B* and *Cyca-D*(CB4)*B* is 0.0665, which implies an approximate divergence time of 12 my, using a 50 my divergence time (Stroband *et al.* 1995) between the two closest carp and zebrafish class II *B* sequences, *Cyca-D*(CB4)*B* and *Brre-DAB3*01*, as a reference (resulting in a rate of $2{,}9 \times 10^{-9}$ substitutions/synonymous site per year).

Within the A4.10me2 family, *Cyca-DAB*01* and *Cyca-DAB*02* were found to co-segregate. This is surprising, as these sequences were previously identified as alleles of a single locus, based on a high sequence similarity between the two transcripts, a finding which in the present study is extended to their intron 1 sequences. However, because of the observed co-segregation in the A4.10me2 family, this identification has to be considered erroneous, and more likely, these transcripts are derived from two genes which arose very recently by a duplication event. The genetic distance (d_s) between the coding regions of *Cyca-DAB*01* and *Cyca-DAB*02* is 0.0119, which implies a divergence time of approximately 1 my. In family A4.10me2 a third sequence, *Cyca-D*(me2)*B*, was found to segregate separately from *Cyca-DAB*01* and *Cyca-DAB*02*. Transcripts of *Cyca-D*(me2)*B* gene have however as yet not been observed, although a cDNA library was screened which was prepared from six individuals from this family.

Further elucidation of the relationships of *Cyca-D*(CB3)*B*, *Cyca-D*(CB4)*B*, *Cyca-DAB*01* and *Cyca-DAB*02* was possible by the fortunate presence of these sequences in a hybrid family R3xR8 as described by Wiegertjes and co-workers (1995). Although this extensive sharing of sequences between fish from Poland (R3), Hungary (R8) and Israel (A4, DOR70) is remarkable, it can be explained by a recent common origin of these carp in eastern Europe (Wohlfarth *et al.* 1980). The segregation patterns in the R3xR8 family also showed the consistent co-segregation of *Cyca-DAB*01* with *Cyca-DAB*02*, and of *Cyca-D*(CB3)*B* with *Cyca-D*(CB4)*B*. This is consistent with the hypothesis that the former pair of sequences are derived from two linked loci, and the same holds true for *Cyca-D*(CB3)*B* with *Cyca-D*(CB4)*B*. Interestingly, it was observed that *Cyca-D*(CB3)*B* and *Cyca-D*(CB4)*B* were not linked to *Cyca-DAB*01*/*Cyca-DAB*02*, the two sets of linked genes segregated independently. The frequencies of the observed genotypes in the offspring indicate that the R3 parental individual carried *Cyca-DAB*01* and

*Cyca-DAB*02* on a single chromosome of a homologous pair, whereas possible class II *B* sequences on the other chromosome were not detected. The R8 parental individual probably also carried *Cyca-DAB*01* and *Cyca-DAB*02* on a single chromosome, whereas the homologous chromosome contained *Cyca-D*(CB3)*B* and *Cyca-D*(CB4)*B*. The confound differences between the coding sequences and introns from, on the one hand, *Cyca-D*(CB3)*B* and *Cyca-D*(CB4)*B*, and on the other hand, *Cyca-DAB*01* and *Cyca-DAB*02*, seem to indicate that these sequences are derived from two non-orthologous sets of loci, rather than being alleles of two linked loci. This is also supported by the position of the two sets in the phylogenetic tree: whereas *Cyca-D*(CB3)*B* and *Cyca-D*(CB4)*B* cluster with the *Brre-DAB* sequences, *Cyca-DAB*01* and *Cyca-DAB*02* cluster with *Brre-DEB*. Based on these findings we propose to denote *Cyca-D*(CB3)*B* and *Cyca-D*(CB4)*B* as *Cyca-DAB3*01* and *Cyca-DAB4*01*, respectively. Further, we propose to change the allelic designations of *Cyca-DAB*01* and *Cyca-DAB*02* to *Cyca-DAB1*01* and *Cyca-DAB2*01*. Although these sequences cluster with *Brre-DEB* in the phylogenetic tree, we have chosen not to change the designation *DAB* to *DEB*, as proposed by Sültmann and co-workers (1994), because of the possibility that a more closely related, but as yet unidentified, locus may exist in another fish species, which would then unduly necessitate us to rename the sequences again.

The apparent segregation of the two pairs of loci (*Cyca-DAB1*01/Cyca-DAB2*01* vs. *Cyca-DAB3*01/Cyca-DAB4*01*) seems to point at haplotype polymorphisms. Nevertheless, we can not exclude that we have simply failed to amplify and detect the apparently lacking alleles when we performed PCR using primers complementary to conserved regions of the transcripts. The suggestion of haplotypic differences is in a sense reminiscent of the situation found in zebrafish where a genomic library that contained *Brre-DEB* sequences, apparently did not contain *Brre-DAB* genes, whereas zebrafish individuals that expressed *Brre-DAB* sequences could not be shown to express *Brre-DEB*. Haplotype polymorphisms may well add to the possible explanations presented by the authors, *i.e.*, non-expression of *Brre-DEB* and incomplete genomic libraries.

An alternative explanation for segregation of loci would be the presence of more than one complex of *Mhc* genes. Although the carp is a tetraploid species as suggested by the number of chromosomes, all of these are visible as bivalents in the metaphase of cell-divisions, except the 4 micro-chromosomes which are observed as quadrivalents (Ohno *et al.* 1967). *Mhc* genes of carp could therefore in theory be present as a tetrasomically inherited *Mhc* (present on the micro-chromosomes), two *Mhc* complexes which segregate independently, as the result of functional diploidization, or perhaps as a normal disomically inherited *Mhc*.

Nevertheless, the segregation studies presented here do not conclusively support any

of these *Mhc* configurations. The independent segregation of *Cyca-DAB*01/Cyca-DAB*02* and *Cyca-D*(me2)*B* observed in the gynogenetic family A4.10me2 appears to refute the possibility of a normal disomically inherited *Mhc*, as, in that case, it is expected that the gynogenetic offspring would be homozygous. In contrast, the segregation pattern in the R3xR8 hybrid family is fully consistent with a disomically inherited *Mhc*. Further segregation studies are clearly needed, and one should perhaps even take into account that inheritance of *Cyca* genes in the carp may not follow basic Mendelean rules.

With the isolation of class II *A* transcripts from strain A410 in the present study, representatives of all 4 classes of *Mhc* genes have been identified in the carp (Ono *et al.* 1993; Van Erp *et al.* 1996a; Dixon *et al.* 1993). *Cyca-D*(10A)*A* and *Cyca-D*(15A)*A* appear to be *bona fide* class II *A* transcripts. Also, the protein characteristics commonly found in class II *A* proteins are all present, and the encoded proteins may well combine with the *Cyca-DAB*-encoded ß chains to form a class II α-ß heterodimer. *Cyca-D*(10A)*A* and *Cyca-D*(15A)*A* did not segregate in the individuals of the carp clone, and because this strain was generated by meiotic gynogenetic reproduction of a maternal fish which also contained the two class II *A* genes, these are most likely to be derived from separate loci. Hence we propose to designate *Cyca-D*(10A)*A* as *Cyca-DXA1*01* and *Cyca-D*(15A)*A* as *Cyca-DXA2*01*. The Jukes-Cantor corrected genetic distance between the coding regions of these two sequences is 0.0248, which implies an approximate divergence time of 4my. Indicative in this respect is that *Brre*-11.2 was found on a genomic clone which also contained *Brre-DEB*. To find the carp equivalent of *Brre*-11.2 we may therefore have to look in a strain that contains the carp equivalent of *Brre-DEB*, namely those carrying the *Cyca-DAB1*01* and *Cyca-DAB2*01* genes.

Summarizing, carp clone A410 has so far been typed to express two class II *B* genes (*Cyca-DAB3*01* and *Cyca-DAB4*01*) and two class II *A* genes (*Cyca-DXA1*01* and *Cyca-DXA2*01*), in addition to a single class I gene (*Cyca-UA1*01*) (Van Erp *et al.* 1996a). It remains to be established whether these *Cyca* genes actually are present as a complex of genes. Preliminary linkage studies seem to indicate that *Cyca-DAB3*01* and *Cyca-DAB4*01* are indeed linked to *Cyca-UA1*01*, however, future studies will be needed for further elucidation of the configuration and, thereupon, the evolution of *Mhc* genes in carp.

Acknowledgements

We thank Robert de Boer and Bart Keijser for their assistance in the isolation and sequence analysis of the *Cyca-DXA* clones.

References

Betz, U.A.K., Mayer, W.E. and Klein, J. 1994. Major histocompatibility complex class I genes of the coelacanth *Latimeria chalumnae*. *Proc. Natl. Acad. Sci. USA 91:* 11065-11069

Boss, J.M., Mengler, R., Okada, K., Auffray, C. and Strominger, J.L. 1985. Sequence analysis of the human major histocompatibility gene SX-alpha. *Mol. Cell. Biol. 5:* 2677-2683

Dixon, B., Stet, R.J.M., Van Erp, S.H.M. and Pohajdak, B. 1993. Characterization of β_2-microglobulin transcripts from two teleost species. *Immunogenetics 38:* 27-34

Dixon, B., Van Erp, S.H.M., Rodrigues, P.N., Egberts, E. and Stet, R.J.M. Fish major histocompatibility complex genes: an expansion. *Dev. Comp. Immunol. 19:* 109-133, 1995

Grossberger, D. and Parham, P. 1992. Reptilian class I major histocompatibility complex genes reveal conserved elements in class I structure. *Immunogenetics 36:* 166-174

Glamann, J. 1995. Complete coding sequence of rainbow trout Mhc IIß chain. *Scand. J. Immunol. 41:* 365-372

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

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

Hardee, J.J., Godwin, U., Benedetto, R. and McConnell, T.J. 1995. Major histocompatibility complex class II *A* gene polymorphism in the striped bass. *Immunogenetics 41:* 229-238

Higgins, D.G., Bleasby, A.J. and Fuchs, R. 1992. CLUSTAL V: improved software for multiple sequence alignment. *Comput. Appl. Biosci. 8:* 189-191

Haviland, D.L., Haviland, J.C., Fleisher, D.T. and Wetsel, R.A. 1991. Structure of the murine fifth complement component (C5) gene. A large, highly interrupted gene with a variant donor splice site and organizational homology with the third and fourth complement component genes. *J. Biol. Chem. 18:* 11818-11825

Kaufman, J., Salomonsen, J. and Flajnik, M. 1994. Evolutionary conservation of MHC class I and class II molecules-different yet the same. *Seminars in Immunol. 6:* 411-424

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

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

Kaastrup, P., Stet, R.J.M., Tigchelaar, A.J., Egberts, E. and Van Muiswinkel, W.B. 1989. A major histocompatibility locus in fish: serological identification and segregation of transplantation antigens in the common carp (*Cyprinus carpio* L.). *Immunogenetics 30:* 284-290

Kumar, S., Tamura, K. and Nei, M. 1993. *MEGA: Molecular Evolutionary Genetics Analysis, version 1.01.* The Pennsylvania State University, University Park

Komen, J., Bongers, A.B.J., Richter, C.J.J., Van Muiswinkel, W.B. and Huisman, E.A. 1991. Gynogenesis in common carp (*Cyprinus carpio* L.) II. The production of homozygous gynogenetic clones and F1 hybrids. *Aquaculture 92:* 127-142

Okamura, K., Nakanishi, T., Kurosawa, Y. and Hashimoto, K. 1993. Expansion of genes that encode MHC class I molecules in cyprinid fishes. *J. Immunol. 151:* 188-200

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

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

Ohno, S., Muramoto, J. and Christian, L. 1967. Diploid-tetraploid relationship among old-world members of the fish family *Cyprinidae. Chromosoma 23:* 1-9

Sültmann, H., Mayer, W.E., Figueroa, F., O'hUigin, C. and Klein, J. 1993. Zebrafish *Mhc* class II α chain-encoding genes: polymorphism, expression, and function. *Immunogenetics 38:* 408-420

Sültmann, H., Mayer, W.E., Figueroa, F., O'hUigin, C. and Klein, J. 1994. Organization of *Mhc* class II *B* genes in the zebrafish (*Brachydanio rerio*). *Genomics 23:* 1-14

Stet, R.J.M., Van Erp, S.H.M., Hermsen, T., Sültmann, H.A. and Egberts, E. 1993. Polymorphism and estimation of the number of *MhcCyca* class I genes in laboratory strains of the common carp (*Cyprinus carpio* L.). *Dev. Comp. Immunol. 17:* 141-156

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

Stet, R.J.M., Dixon, B., Van Erp, S.H.M., Van Lierop, M.C., Rodrigues, P.N.S. and Egberts, E. Inference of structure and function of fish Major Histocompatibility Complex (MHC) molecules from expressed genes. *Fish Shellfish Immunol.* in press

Stet, R.J.M. and Egberts, E. 1991. The histocompatibility system in teleostean fishes: from multiple histocompatiblity loci to a major histocompatibility complex. *Fish Shellfish Immunol. 1:* 1-16

Senapathy, P., Shapiro, M.B. and Harris, N.L. 1990. Splice junctions, branch point sites, and exons: sequence statistics, identification, and applications to genome project. *Meth. Enzymol. 183:* 252-278

Stroband, H.W.J., Stevens, C., Te Kronnie, G., Samallo, J., Schipper, H., Kramer, B. and Timmermans, L.P.M. 1995. Expression of *carp-cdx1*, a *caudal* homolog, in embryos of the carp, *Cyprinus carpio. Roux's Arch. Dev. Biol. 204:* 369-377

Takeuchi, H., Figueroa, F., O'hUigin, C. and Klein, J. 1995. Cloning and characterization of class I *Mhc* genes of the zebrafish, *Brachydanio rerio. Immunogenetics 42:* 77-84

Van Erp, S.H.M., Dixon, B., Figueroa, F., Egberts, E. and Stet, R.J.M. 1996a. Identification and characterization of a new major histocompatibility complex class I gene in carp (*Cyprinus carpio* L.). *Immunogenetics* in press

Van Erp, S.H.M., Egberts, E. and Stet, R.J.M. 1996b. Evidence for multiple distinct major histocompatibility complex class I lineages in teleostean fish. *Eur. J. Immunogenetics* in press

Walker, R.A. and McConnell, T.J. 1994. Variability in an MHC*Mosa* class II ß chain encoding gene in striped bass *(Morone saxatilis). Dev. Comp. Immunol. 18:* 325-342

Wiegertjes, G.F., Stet, R.J.M., Bongers, A.B.J., Voorthuis, P., Zandieh Doulabi, B., Groeneveld, A. and Van Muiswinkel, W.B. 1995. Investigation into the immune responsiveness of F1 hybrids of homozygous carp (*Cyprinus carpio* L.) selected for high or low antibody production: indication for immune gene control. In: G.F. Wiegertjes, *Immunogenetics of disease resistance in fish.* PhD-thesis, Wageningen Agricultural University, Wageningen

Wiegertjes, G.F., Stet, R.J.M. and Van Muiswinkel, W.B. 1994. Divergent selection for antibody production in common carp (*Cyprinus carpio* L.) using gynogenesis. *Anim. Genet. 25:* 251-257

Wohlfarth, G., Lahman, M., Hulata, G. and Moav, R. 1980. The story of Dor-70: a selected strain of the Israeli common carp. *Bamidgeh. 32:* 3-5

Chapter 7

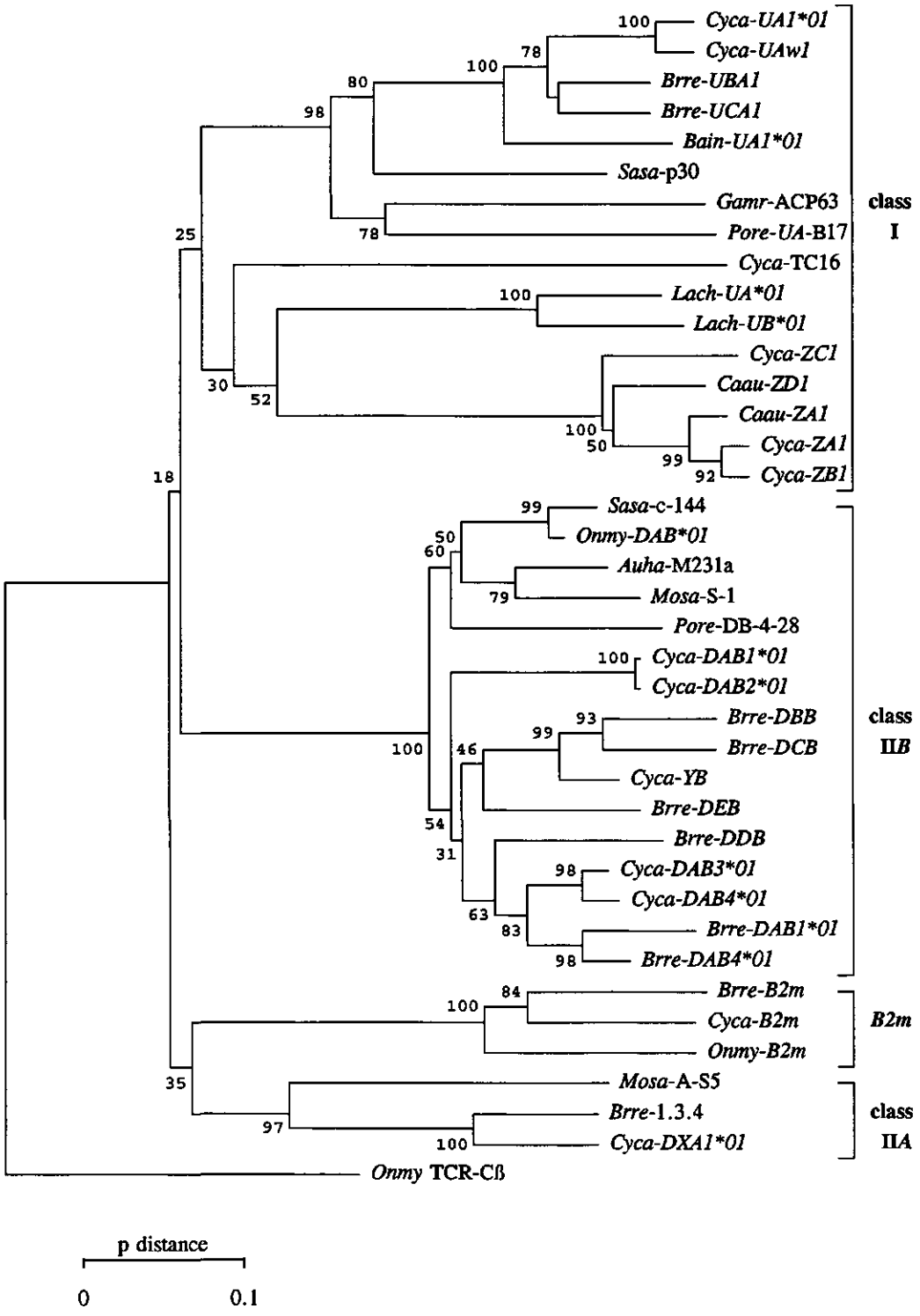
General discussion.

The five years of work described in this thesis coincide with the five years of reports on *Mhc* genes from fish. In 1990 the first fish *Mhc* sequences were found in carp, and since then the number of fish species in which *Mhc* sequences have been identified has risen steeply to 37 at the end of 1995. In these five years many aspects of the fish *Mhc* genes have surfaced, and the progress has made some unexpected turns along the way. In this general discussion, we will try to combine the pieces found in each separate chapter, assess them in retrospect, and add future directions.

Class I genes in carp

The area of class I genes in carp has turned out to be much more complex than initially expected. At this moment, four distinct types of class I α-chain encoding genes have been found in the carp, in chronological order of appearance in literature, *Cyca-Z*, *Cyca*-C4, *Cyca-U* and *Cyca*-TC16. In spite of the very large genetic distance between the different lineages (p-distances between exons 4 are in the order of 0.5 - 0.6, *Cyca*-C4 excluded, becauses this sequence does not comprise exon 4), all of these class I sequences cluster together in a phylogenetic tree containing also class II and *B2m* genes (Fig. 1). This may indicate that all carp class I lineages are derived from a single class I ancestral gene. However, the level of genetic distance between these lineages, and the sharing of some of these lineages by fish species from different superorders, suggest that the duplications leading to the four types of genes have happened early in the evolution of teleostean fish.

Of the four types, the *U* lineage appears to be the common class I lineage of teleost fish, as to date *U* sequences have been found in representatives of four euteleost superorders: (i) the Protacanthopterygii, represented by the salmonids Atlantic salmon and pink salmon; (ii) the Ostariophysi, represented by the cyprinids common carp, the zebrafish, and large African barbel; (iii) the Paracanthopterygii, represented by the cod; and (iv) the Acanthopterygii, as represented by the guppy (Figs. 1 and 2). This omnipresence in euteleosts combined with the observed sequence characteristics which indicate that the encoded proteins may bind antigenic peptides, CD8 and $ß_2$-microglobulin, and the strong, easily detectable expression in a variety of tissues, strongly argues that these genes encode classical class I molecules. This conclusion is further strengthened by the outcome of flowcytometric experiments performed with two polyclonal antisera, one raised to recombinant *Cyca-UA1*01*, and the other to recombinant *Cyca-B2m*. These two antisera produced very similar staining patterns on lymphocytes from various lymphoid organs, suggesting that *Cyca-UA1*01* and *Cyca-B2m* are present on the same cells. Also, the detection of two distinct subpopulations of class I-positive lymphocytes, namely B cells (sIg positive) and putative T cells (sIg negative), adds to the conclusion that *Cyca-U* sequences encode classical class I molecules.


Cyca-UA1*01
Cyca-UAw1
Brre-UBA1
Brre-UCA1
Bain-UA1*01
Sasa-p30
Gamr-ACP63
Pore-UA-B17
Cyca-TC16
Lach-UA*01
Lach-UB*01
Cyca-ZC1
Caau-ZD1
Caau-ZA1
Cyca-ZA1
Cyca-ZB1
Sasa-c-144
Onmy-DAB*01
Auha-M231a
Mosa-S-1
Pore-DB-4-28
Cyca-DAB1*01
Cyca-DAB2*01
Brre-DBB
Brre-DCB
Cyca-YB
Brre-DEB
Brre-DDB
Cyca-DAB3*01
Cyca-DAB4*01
Brre-DAB1*01
Brre-DAB4*01
Brre-B2m
Cyca-B2m
Onmy-B2m
Mosa-A-S5
Brre-1.3.4
Cyca-DXA1*01
Onmy TCR-Cß
class I
class IIB
B2m
class IIA
p distance
0
0.1

So what could representatives of the other three lineages encode? The presence of these lineages in species other than the carp appears to be far more restricted than that of the *U* genes. So far, *Z* genes have only been found in the carp and the closely related Ginbuna crucian carp (*Carassius auratus langsdorfii*) (Hashimoto *et al.* 1990; Okamura *et al.* 1993), the lineage represented by *Cyca-TC16* may also contain the *Lach-U* class I genes from the crossopterygian coelacanth (Betz *et al.* 1994), and *Cyca*-C4 has been observed exclusively in a San Francisco China-town market fish, held to be a common carp (Grossberger and Parham 1992). The divergence between the *U* sequences and the other three lineages is reminiscent of the divergence between mammalian classical *Mhc* molecules and distantly related molecules like the FcRn, Zn-α_2-glycoprotein and CD1. However, none of these non-classical molecules shows significant similarity to *Cyca-Z*, *Cyca*-TC16 or *Cyca*-C4.

In the absence of experimental studies on the function of these carp molecules, one has to rely on circumstantial evidence, such as data on expression, or the presence of certain residues in the amino acid sequence. Expression of *Cyca*-Z, *Cyca*-TC16, and *Cyca*-C4 is not firmly established, neither by northern hybridizations, nor after extensive attempts in our laboratory to obtain *Cyca-Z* and *Cyca*-TC16 full-length cDNA clones from libraries prepared from carp lymphoid organs. Nevertheless *Cyca-Z* fragments have been successfully amplified from carp kidney cDNA by Okamura and co-workers (1993), *Cyca*-C4 was originally amplified from cDNA (Grossberger and Parham 1992), and so was *Cyca*-TC16 in our laboratory. Also, the absence of conserved peptide-binding residues and a CD8-binding site may indicate that *Cyca-Z* and *Cyca*-TC16 molecules do not function as classical antigen-presenting molecules. On the other hand, *Cyca-Z* genes are present in multiple copies in the genome, which appear to be linked (see also section: "Do fish have an *Mhc*?"), and which exhibit RFLP. In addition,

Figure 1. Phylogenetic tree created by the neighbor-joining method (Saitou and Nei 1987), based on uncorrected p-distances between *Mhc* sequences from bony fish, using exon 4 for class I, exon 3 for class II and exon 2 for *B2m*. The sequences of the TCR Cß region of rainbow trout (*Onmy* TCR Cß; Partula *et al.* 1995), was used as an outgroup. Numbers on nodes indicate the percentages of 1000 bootstraps supporting each partitioning. Sources of sequences not described in this thesis: *Brre-DAB* sequences, zebrafish (Ono *et al.* 1992); *Brre-DBB*, *-DCB*, *-DDB*, *-DEB* (Sültmann *et al.* 1994); *Cyca-YB* (*TLAIIß-1*) (Hashimoto *et al.* 1990); *Cyca-DAB1*, *Cyca-DAB2*, common carp (Ono *et al.* 1993b); *Pore-DB*-4-28, *Pore-UA*-A30, guppy (Sato *et al.* 1995); *Sasa*-c144, Atlantic salmon (Hordvik *et al.* 1993) *Onmy-DAB*01*, rainbow trout (Glamann 1995); *Auha*-M231a, cichlid (Klein *et al.* 1993); *Mosa*-S-1, striped bass (Walker and McConnell 1994); *Brre*-B2m (Ono *et al.* 1993a); *Onmy-B2m* (P. Parham, pers. comm.); *Mosa*-A-S5 (Hardee *et al.* 1995); *Brre*-1.3.4 (Sültmann *et al.* 1993); *Cyca-Z*, common carp, and *Caau-Z*, Ginbuna crucian carp (Okamura *et al.* 1993); *Lach-U*, coelacanth (Betz *et al.* 1994); *Gamr*, cod (A.C. Persson, pers. comm.); *Sasa*-p30, (Grimholt *et al.* 1993); *Bain-UA1*01*, large African barbel (Accession no. X94107); *Brre-U*, (Takeuchi *et al.* 1995).

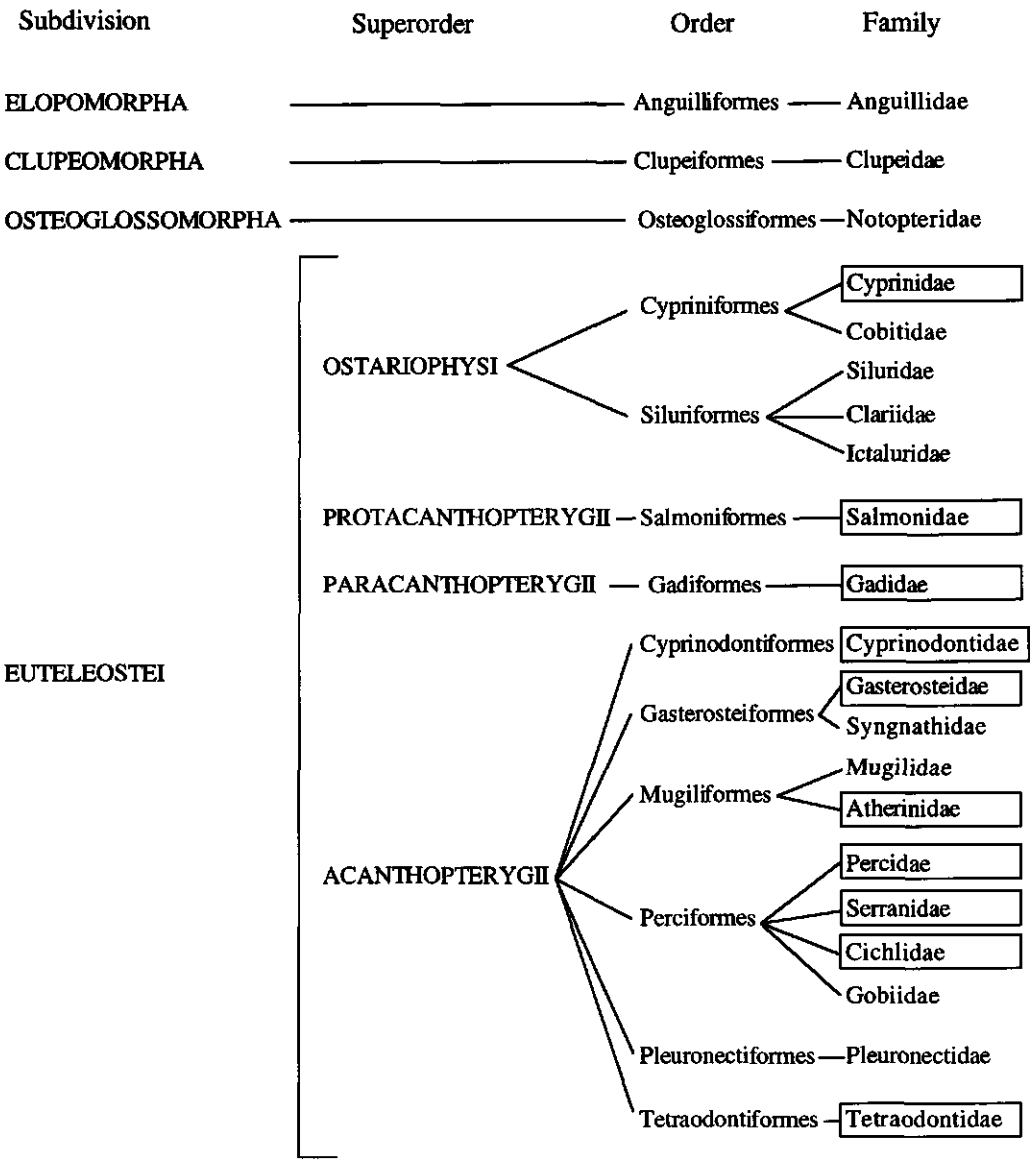


Figure 2. Classification of teleostean fish, modified after Nelson (1994). Boxed family names indicate that *Mhc* sequences were identified in species representing that family.

RFLP patterns, using a *Cyca-Z* probe in two gynogenetic families A4.3 and A4.19, were found to correlate with histo-incompatibility. For this, either the *Cyca-Z* genes themselves, or perhaps a closely linked histocompatibility gene could account. As for the function of *Cyca*-TC16, under the assumption that this sequence shares a common origin with the *Lach-U* sequences from the coelacanth, one can only speculate that during a long time of its existence *Cyca*-TC16 must have had a conserved function within the *Ig* superfamily, for certainly a non-functional gene would have diverged beyond recognition during 400 million years (my) of evolution.

Class II genes in carp

In contrast to the class I story, teleostean class II genes form a much more homogeneous group. At this moment, class II *B* genes are by far the best represented among reported fish *Mhc* sequences, with representatives from species from three euteleostean superorders: the Protacanthopterygii, represented by the salmonids Atlantic salmon, 6 species of Pacific salmon (accession no. U34716-20, U34692-713), and rainbow trout; the Ostariophysii, represented by the cyprinids common carp, Ginbuna crucian carp, zebrafish, large African barbel, and Portuguese barbel (accession no. X93896-7); and finally the Acanthopterygii as represented by the striped bass, guppy, pufferfish (accession no. X87413), and multiple species of African cichlids, but also in perch, silver side, and three-spined stickle back (Figueroa *et al.* 1995). Yet, all of the class II *B* sequences are clearly related, giving rise to a homogeneous cluster in phylogenetic trees and, in contrast to the class I situation, the genetic distances between the sequences are much smaller (the largest genetic p-distance between two carp class II *B* exon 3 sequences is 0.2) (Fig. 1).

In carp, evidence is so far obtained for the existence of at least 5 class II *B* loci: *Cyca-YB* by Hashimoto and co-workers (1990), *Cyca-DAB1*, *Cyca-DAB2* (Ono *et al.* 1993b) and *Cyca-DAB3* and *Cyca-DAB4* from our laboratory. This set of genes could account for the three to five hybridizing fragments observed in the Southern blot hybridizations using a *Cyca-YB* exon 4 probe, K2-1, because the level of sequence similarity of the *Cyca-YB* probe to the *Cyca-DAB* loci (80% nt. identity) is enough to ensure hybridization under the low stringency conditions applied.

Phylogenetic analysis indicates that the *Cyca-DAB1* and *Cyca-DAB2* loci are the result of a recent duplication (approximately 1 my ago, calculated from the Jukes-Cantor corrected proportion of synonymous differences (d_s) in exons 3, and a rate of 2.9×10^{-9} substitutions/synonymous site per year). Also *Cyca-DAB3* and *Cyca-DAB4* loci are the result of a relatively recent duplication of an ancestral gene, about 10-12 my ago, and this gene was already present in the ancestor of carp and zebrafish, judging from the clustering of *Cyca-DAB3* and *Cyca-DAB4* with *Brre-DAB*. The genetic distance between the *Cyca-DAB1*/*Cyca-DAB2*

group of genes and the *Cyca-DAB3/Cyca-DAB4/Brre-DAB* group of genes indicates that the duplication leading to the ancestral genes of these two groups happened 80-120 my ago. The *Cyca-YB* gene is still incomplete, as to date only exons 2 and 3 have been identified. In the phylogenetic tree, this gene clusters with the putatively non-functional *Brre-DBB* and *Brre-DCB* genes from zebrafish.

Reports of class II *A* genes are somewhat lagging behind. Such genes have to date been identified in only three species, namely the cyprinids common carp and zebrafish, and a representative of the perciformes, the striped bass. In the carp two class II *A* loci appear to be present, showing a high level of sequence identity (95%, exons 3 d_s=0.02), which, using the same synonymous substitution rate as above, implies a divergence time of approximately 3.5 my. The presence of multiple class II *A* loci is expected because of the presence of multiple class II *B* loci, and also in the zebrafish evidence was obtained for the presence of multiple loci (Sültmann *et al.* 1993). Further conclusions regarding the evolution and relationship of the teleost class II *A* genes, however, awaits the identification of these sequences in more species.

The class II α and ß chains encoded by the *Cyca-DAB* and *Cyca-DXA* genes, respectively, may well combine to form a functional class II α-ß heterodimer. Yet, experimental proof for this has so far not been obtained, and this assumption is only based on indirect indications, concerning expression and characteristics of the protein sequence (Stet *et al.* in press). In carp, expression of class II *B* genes was detected by the isolation of full-length cDNA clones from lymphoid tissues and, for *Cyca-DAB1* and *Cyca-DAB2* by northern hybridizations and semi-quantitative PCR on cDNA in various lymphoid organs and lymphoid cells (Rodrigues *et al.* 1995). A possible exception is the *Cyca-YB* sequence, which most likely is not expressed. To date, expression of class II *A* genes in carp is only indicated by the isolation of full-length cDNA clones.

Features of the protein sequences of the carp class II *B* and class II *A* genes suggest that the encoded protein chains may function as a classical class II heterodimer. Significant indications that products encoded by *Cyca-DAB* encode functional antigen presenting molecules can be found in analyses of the amino acid variability in the derived protein sequences of the $ß_1$ domains. Highly variable residues in *Cyca-DAB* protein sequences correspond with those positions of human *Mhc* proteins that interact with peptides (data not shown). The same observation was made in a much more extensive study on the variability of class II *B* genes in the species flock of large African barbels (*Barbus intermedius*) of Lake Tana (Dixon *et al.* submitted). The *Bain-DAB* sequences identified in these barbels are all very much related to either the *Cyca-DAB1/2* or the *Cyca-DAB3/4* genes, and in these *Bain-DAB* sequences

variability was found to be the highest at sites corresponding to mammalian peptide-binding residues. In addition, sometimes very high d_n/d_s ratios were observed between sequences. This strongly suggests that the peptide-binding regions of these genes experience positive selection, and that the molecules encoded by these genes are involved in binding various peptides. Suggestive in this respect are also the results obtained by Wiegertjes and co-workers in our laboratory (1995), who found a correlation between the *Cyca-DAB* haplotype of F1 hybrid carp, and the level of their humoral immune response to the hapten-carrier DNP-KLH.

Origin of *Mhc* genes.

The obvious analogies in the structure and functioning of *Mhc* class I and class II molecules has raised interest in the evolutionary interrelationships between the different classes of *Mhc* genes. Two main hypotheses regarding the origin of class I and class II genes have been proposed. They differ mainly in their explanation regarding the origin of the membrane distal domains of *Mhc* molecules, and as a consequence, in their conclusion with regard to which class of *Mhc* molecules arose first. The first hypothesis speculates that class I was produced by an exon-shuffling event that combined an immunoglobulin constant domain and the peptide-binding region of HSP70 (Flajnik *et al.* 1991). The class II genes subsequently arose by recombining the class I α_1 domain with a separate immunoglobulin C-domain.
The second hypothesis suggests that class II arose first, each chain consisting of an immunoglobulin C-domain combined with an immunoglobulin V-domain. The Ig V-domains of the class II α and class II ß molecules then refolded together to shape the antigen binding groove. Class I and $ß_2$m subsequently arose by a chromosomal inversion event combined with a splice acceptor site mutation, which eliminated the transmembrane and cytoplasmic domains of the resultant *B2m* gene (Kaufman *et al.* 1984).

The clustering of the fish *Mhc* sequences in the phylogenetic tree presented in figure 1 indeed suggests a relationship between the class II $ß_2$ domains and class I α_3 domains, and also between class II α_2 and $ß_2$m. This phylogeny has been suggested before by Hood and co-workers (1985) and it was also observed in phylogenetic analyses by Hughes and Nei (1993), who mainly included sequences from mammals, fowl, and *Xenopus*. One should bear in mind, however, that the bootstrap values in these trees are generally low. In any case, the observed clustering renders equal support for either hypothesis.

Yet, the finding that in teleostean fish all class II $ß_1$ and, in contrast to the mammalian molecules, also class II α_1 domains contain the two cysteine residues spaced approximately 53 amino acids apart, does exclusively support the second hypothesis. It is much more likely that these residues are the legacy of an Ig-fold in which they once took part, than that these residues evolved independently in both HSP70 domains. If so, then the cysteine residues in

class II α_2 domain were lost at least twice in evolution, namely in the lineage leading to mammals, and in the shark (Kasahara *et al.* 1992).

Also Hughes and Nei (1993) have presented arguments that support the second hypothesis. They showed that: (i) class II β_1 domains are statistically more similar to immunoglobulin V domains than to HSP70 domains; (ii) class II α_1 domains are significantly more similar to class II β_1 domains, than to HSP70 α_1; and (iii) the genetic distance between class II α_1 and class II β_1 domains is much smaller than would be expected if these originated from HSP70. In addition, Dixon and co-workers (1995) have demonstrated that the intron/exon structure of HSP70 genes cannot offer support to the theory that an HSP70 peptide-binding region was donated to the first *Mhc* genes.

Taken all this together, we think the evidence is overwhelmingly in favor of the hypothesis that the ancestral *Mhc* molecule had a class II-like structure.

Do fish have an *Mhc*?

This question has eluded the minds of several researchers before 1990, and it seemed to be answered by the finding of *Mhc* sequences from both class I and class II, in a multitude of teleost species. Nevertheless, it is still appropriate to pose the same question, for it is still unclear whether fish really have histocompatibility genes that are organized in a genetic complex. The data on this subject are so far extremely scarce.

Only for class II genes some firm linkage data are available: *Cyca-DAB1*01* is linked to *Cyca-DAB2*01*, and *Cyca-DAB3*01* is linked to *Cyca-DAB4*01*, but these two pairs of linked genes segregate. In addition, preliminary data suggest that *Cyca-UA1*01* and *Cyca-DXA1* are linked to *Cyca-DAB3* and *Cyca-DAB4*, which would already complete a set of linked class I, class II*B* and class II*A* genes. This, however, requires further confirmation through segregation studies.

Some evidence also exists for linkage of the multiple *Cyca-Z* genes. Hashimoto and co-workers (1990) obtained evidence for the presence of two *Cyca-Z* genes on the genomic clone λTLAI-1. In addition, linkage of at least some of the *Cyca-Z* genes was indicated by the results of RFLP studies, in which the number of hybridizing *Z* fragments were found to vary depending on the restriction endonuclease used to digest the DNA (*cf.* Figures 2 and 3, chapter 2).

For a large part, however, we are still probing in the dark. It is for example not clear whether the expressed *Cyca-U* locus is linked to the related, but apparently non-expressed, sequences that are detected on Southern blots probed with *Cyca-UA1*01*. Nor is it clear whether these classical *Cyca-U* genes are linked to the non-classical class I-like sequences *Cyca-Z*, *Cyca-TC16* and *Cyca-C4*, or whether the latter are perhaps located on separate chromosomes, like some of the mammalian non-classical genes, including FcRn, Zn-α_2-glycoprotein, and

CD1. To answer these questions, future studies including RFLP analyses, other genomic typing studies (PCR combined with RFLP or sequence analysis) or perhaps genomic mapping using cosmid or YAC libraries, will be required.

Conjectural is also the suggestion that in teleost fish the *B2m* gene is perhaps located in the *Mhc*. It is generally assumed that the *B2m* gene moved out of the *Mhc* to a different chromosome at some time point in evolution during the emergence of mammals, and the possibility remains that this happened after the evolutionary branch leading to teleostean fish split off from the branch leading to the tetrapods. But, so far, no evidence has been obtained that indicate a shared chromosomal localization of *B2m* and *Mhc* genes in fish. Recent findings in rainbow trout, however, indicate the presence of multiple *B2m* genes in the genome of this species. These genes all have an identical coding sequence, but the flanking sequences are different between genes and, more importantly, between trout individuals. This could therefore provide an important tool in studying linkage of *B2m* genes to the other *Mhc* genes of trout (Shum *et al.* in press).

A problem that announced itself during the linkage studies performed so far in carp, is the unresolved ploidy status of this species. Based on the number of chromosomes (100 or 104), carp is considered to be tetraploid, presumably as the result of auto-tetraploidization (genome doubling) or, more likely, allo-tetraploidization (species hybridization), which probably occurred less than 16 my ago (Larhammar and Risinger 1994). Allo-tetraploidization appears to be the more likely scenario because all except four of the carp chromosomes form bivalents in the metaphase of cell divisions, whereas only the four micro-chromosomes have been observed to form a quadrivalent (Ohno *et al.* 1967). Also, on estimation 52% of the genes are present as two active, but diverged copies (*i.e.*, as two sets of two homologues), whereas the remainder of the genes is present as a single copy, *i.e.*, in a 'normal' diploid state (Larhammar and Risinger 1994). Furthermore, a recent study on microsatellite markers suggests that most of the loci detected are diploid, *i.e.*, having at most two alleles, while only a minority of the loci as seen by the markers showed more than two alleles (Crooijmans *et al.* in preparation). The *Cyca-B2m* gene appears to be present in two copies in the carp genome: in RFLP studies two hybridizing bands were consistently observed, which did not show variation between several outbred individuals or F1 hybrids. The carp, therefore, seems to have retained two copies of the *B2m* gene after the tetraploidization event, although the finding of only a single cDNA sequence suggests that it may have silenced one of these copies, which would result in a functional diploidization.

Taken together, it seems that the carp is in an intermediate stage of diploidization. Consequently, if there is an *Mhc* region in carp, it could be present in different configurations: (i) in four allelic copies (tetrasomically inherited); (ii) in two times two allelic copies (*i.e.*,

two *Mhc*'s, each on a pair of disomically inherited chromosomes); or perhaps (iii) as a single disomically inherited *Mhc* (a 'normal' diploid *Mhc*), while the other copy of the *Mhc* was lost by silencing or deletion, as observed in polyploid *Xenopus* species (Shum *et al.* 1993). As yet we did not detect more than two alleles at any *Cyca* locus. On the contrary, for most loci we could not identify more than a single, yet segregating, gene, which suggests the presence of null-alleles or haplotype polymorphisms. This, in combination with the level of inbreeding encountered in the carp, highly complicates the elucidation of the chromosomal configuration of the *Mhc*.

So where to go from here?

The *Mhc* genes of carp have turned out to constitute an extended genetic family, with classical and non-classical class I genes, *B2m* genes, class II *A* and *B* genes (Table 1). The initial expectations that fish *Mhc* genes, if existing at all, would perhaps be primitive and as a consequence dissimilar from the mammalian ones, has clearly been refuted. Moreover, the fish immune system is proving to be strikingly similar to that of mammals, with *Mhc* genes, and, as recently shown, rearranging T-cell receptor genes as well as B-cell receptor genes. The presence of both class I and class II genes suggests that a strategic differentiation in the responses to intracellularly and extracellularly derived antigens is already present in fish. In line with this, it seems plausible to expect that, also in fish, these differential responses are guided by co-receptors like CD4 and CD8, which, however, still await discovery. Taking all this into account, it may be time to abolish the term lower-vertebrate when referring to fish and their immune system. Nevertheless, the knowledge about the functioning of the immune system of fish, and especially the mechanisms of antigen presentation, is based for a large part on indirect evidence. With the set of carp *Mhc* genes now available, a valuable basis is provided for future research on the immune response of carp, and the first experiments using antibodies directed against carp *Mhc* molecules are already being performed.

At the same time, undoubtedly new fish *Mhc* sequences will keep flowing in. To ensure some future clarity in the relationships between the multitude of genes from the multitude of fish species, the need to pay more attention to nomenclature is increasing. We believe that the nomenclature first adopted by Shum and co-workers (1993) and subsequently by others, is probably efficient. In this nomenclature each class I locus is designated by *U* (for *Una*), followed by a letter denominating the family of loci, and a number indicating the exact locus, with an asterisk and number to specify the allele at that locus. Nonetheless, confusion is already rising, and it seems important that the second letter designating the family of loci is chosen with care. This is illustrated by the similar names given to the coelacanth class I sequences (*Lach-UA*), and the zebrafish and carp class I sequences (*Brre-UA* and *Cyca-UA*), although

Table 1. *Mhc* sequences of the common carp

Class I		*B2m*		Class II *A*		Class II *B*	
Cyca-KI	1	*Cyca-B2m*	3	*Cyca-DXA1*01*	3	*Cyca-DAB1*01*	1,3
Cyca-K1-5	1			*Cyca-DXA2*01*	3	*Cyca-DAB2*01*	1,3
Cyca-TC3	4					*Cyca-D*(me2)*B*	1
Cyca-TC13	4					*Cyca-D*(clc)*B*	1
Cyca-TC15	4					*Cyca-DAB3*01*	1,3
Cyca-TC17	4					*Cyca-DAB4*01*	1,3
Cyca-TC18	4					*Cyca-YB*	5
Cyca-ZA1	5					*Cyca*-KII	1
Cyca-ZA2	4					*Cyca*-K2-1	1
Cyca-ZB1	4						
Cyca-ZC1	4						
Cyca-TC16	4						
Cyca-C4	4						
*Cyca-UA1*01*	2,3						
*Cyca-UA1*02*	5						
Cyca-UAw1	4						

1.: Genomic, PCR fragment
2.: Genomic, full-length sequence
3.: cDNA, full-length clone
4.: cDNA, PCR fragment
5.: Genomic, partial clone

these sequences apparently belong to completely different lineages. In line with these thoughts it would perhaps be prudent to rename the *Cyca-Z* sequences to *Cyca-UZ*, although a consequence of this would be that *Cyca-ZA* would have to be renamed to *Cyca-UZ1*, *Cyca-ZB* to *Cyca-UZ2* and so on. Similar ponderings can be spent on the class II nomenclature, using a *D* (Duo) as the first letter to designate class II gene names. In line with this, *Cyca-YB* would have to be renamed to *Cyca-DYB*, which would make its kind immediately clear by name. Hopefully, joint efforts will soon create commonly accepted rules.

Another important issue to be addressed in the future is the question whether or not these genes are in fact localized in a complex. Linkage studies, haplotyping, and mapping of genes will undoubtedly yield valuable information. The complex genetics of carp will, however, complicate straightforward interpretation of the results, and for this reason it may prove necessary to resort to a technique like fluorescence activated *in situ* hybridization, which, most attractively, is abbreviated as FISH.

References

Betz, U.A.K., Mayer, W.E. and Klein, J. 1994. Major histocompatibility complex class I genes of the coelacanth *Latimeria chalumnae*. *Proc. Natl. Acad. Sci. USA 91:* 11065-11069

Crooijmans, R.P.M.A., Bierbooms, V., Komen, J., Van der Poel, J.J., Strijk, J.A. and Groenen, M.A.M. Microsatellite markers in common carp (*Cyprinus carpio* L.). in preparation

Dixon, B., Van Erp, S.H.M., Rodrigues, P.N.S., Egberts, E. and Stet, R.J.M. 1995. Fish major histocompatibility complex genes: an expansion. *Dev. Comp. Immunol. 19:* 109-133

Dixon, B., Nagelkerke, L.A.J., Egberts, E. and Stet, R.J.M. Evolution of cyprinid major histocompatibility complex class II beta genes as reflected in the large barbels (*Barbus intermedius* complex) of Lake Tana. submitted

Figueroa, F., Ono, H., Tichy, H., O'hUigin, C. and Klein, J. 1995. Evidence for insertion of a new intron into an *Mhc* gene of perch-like fish. *Proc. R. Soc. Lond. B 259:* 325-330

Flajnik, M.F., Canel, C., Kramer, J. and Kasahara, M. 1991. Which came first, MHC class I or class II? *Immunogenetics 33:* 295-300

Glamann, J. 1995. Complete coding sequence of rainbow trout Mhc IIß chain. *Scand. J. Immunol. 41:* 365-372

Grimholt, U., Hordvik, I., Fosse, V.M., Olsaker, I., Endresen, C. and Lie, Ø. 1993. Molecular cloning of major histocompatibility complex class I cDNAs from Atlantic salmon (*Salmo salar*). *Immunogenetics 37:* 469-473

Grossberger, D. and Parham, P. 1992. Reptilian class I major histocompatibility complex genes reveal conserved elements in class I structure. *Immunogenetics 36:* 166-174

Hardee, J.J., Godwin, U., Benedetto, R. and McConnell, T.J. 1995. Major histocompatibility complex class II *A* gene polymorphism in the striped bass. *Immunogenetics 41:* 229-238

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

Hood, L., Kronenberg, M. and Hunkapiller, T. 1985. T cell antigen receptors and the immunoglobulin supergene family. *Cell 40:* 225-229

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

Hughes, A.L. and Nei, M. 1993. Evolutionary relationships of the classes of major histocompatibility complex genes. *Immunogenetics 37:* 337-346

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

Kaufman, J.F., Auffray, C., Korman, A.J., Shackelford, D.A. and Strominger, J. 1984. The class II molecules of the human and murine major histocompatibility complex. *Cell 36:* 1-13

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

Larhammar, D. and Risinger, C. 1994. Molecular genetic aspects of tetraploidy in the common carp *Cyprinus carpio*. *Mol. Phyl. Evol. 3:* 59-68

Nelson, J.S. 1994. Fishes of the world. Wiley & Sons, New York

Ohno, S., Muramoto, J., Christian, L. and Atkin, N.B. 1967. Diploid-tetraploid relationship among old-world members of the fish family *Cyprinidae*. *Chromosoma 23:* 1-9

Okamura, K., Nakanishi, T., Kurosawa, Y. and Hashimoto, K. 1993. Expansion of genes that encode MHC class I molecules in cyprinid fishes. *J. Immunol. 151:* 188-200

Ono, H., Figueroa, F., O'hUigin, C. and Klein, J. 1993a. Cloning of the $ß_2$-microglobulin gene in the zebrafish. *Immunogenetics 38:* 1-10

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

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

Partula, S., De Guerra, A., Fellah, S. and Charlemagne, J. 1995. Structure and diversity of the T cell antigen receptor ß-chain in a teleost fish. *J. Immunol. 155:* 699-706

Rodrigues, P.N.S., Hermsen, T.T., Rombout, J.H.W.M., Egberts, E. and Stet, R.J.M. 1995. Detection of *Mhc* class II transcripts in lymphoid tissues of the common carp (*Cyprinus carpio* L.). *Dev. Comp. Immunol.* 19: 483-496

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

Sato, A., Figueroa, F., O'hUigin, C., Reznick, D.N. and Klein, J. 1995. Identification of major histocompatibility complex genes in the guppy, *Poecilia reticulata. Immunogenetics 43:* 38-49

Shum, B., Avila, D., Du Pasquier, L., Kasahara, M. and Flajnik, M. 1993. Isolation of a classical MHC class I cDNA from an amphibian. Evidence for only one class I locus in the *Xenopus* MHC. *J. Immunol. 151:* 5376-5386

Shum, B.P., Azumi, K., Zhang, S., Kehrer, S.R., Raison, R.L., Detrich, H.W. and Parham, P. Unexpected $ß_2$-microglobulin sequence diversity in individual rainbow trout. *Proc. Natl. Acad. Sci. USA* in press

Stet, R.J.M., Dixon, B., Van Erp, S.H.M., Van Lierop, M.C., Rodrigues, P.N.S. and Egberts, E. Inference of structure and function of fish Major Histocompatibility Complex (MHC) molecules from expressed genes. *Fish Shellfish Immunol.* in press

Stroband, H.W.J., Stevens, C., Te Kronnie, G., Samallo, J., Schipper, H., Kramer, B. and Timmermans, L.P.M. 1995. Expression of *carp-cdx1*, a *caudal* homolog, in embryos of the carp, *Cyprinus carpio. Roux's Arch. Dev. Biol. 204:* 369-377

Sültmann, H., Mayer, W.E., Figueroa, F., O'hUigin, C. and Klein, J. 1993. Zebrafish *Mhc* class II α chain-encoding genes: polymorphism, expression, and function. *Immunogenetics 38:* 408-420

Sültmann, H., Mayer, W.E., Figueroa, F., O'hUigin, C. and Klein, J. 1994. Organization of *Mhc* class II *B* genes in the zebrafish (*Brachydanio rerio*). *Genomics 23:* 1-14

Takeuchi, H., Figueroa, F., O'hUigin, C. and Klein, J. 1995. Cloning and characterization of class I *Mhc* genes of the zebrafish, *Brachydanio rerio. Immunogenetics 42:* 77-84

Walker, R.A. and McConnell, T.J. 1994. Variability in an MHC*Mosa* class II ß chain encoding gene in striped bass *(Morone saxatilis). Dev. Comp. Immunol. 18:* 325-342

Wiegertjes, G.F., Stet, R.J.M., Bongers, A.B.J., Voorthuis, P., Zandieh Doulabi, B., Groeneveld, A. and Van Muiswinkel, W.B. 1995. Investigation into the immune responsiveness of F1 hybrids of homozygous carp (*Cyprinus carpio* L.) selected for high or low antibody production: indication for immune gene control. In: G.F. Wiegertjes, *Immunogenetics of disease resistance in fish.* PhD-thesis, Wageningen Agricultural University, Wageningen

In **chapter 1** an overview is given of the structure and function of *Mhc* molecules and *Mhc* genes in several vertebrate species. Molecules encoded by genes of the major histocompatibility complex play an essential role in the specific immune response by presenting antigen fragments to T lymphocytes. Two classes of polymorphic *Mhc* molecules can be distinguished, the class I molecules that present endogenous peptides to CD8-positive cytotoxic T cells, and the class II molecules that present exogenous antigens to CD4-positive T-helper cells. The *Mhc* molecules appear to have very similar functions in several species studied. In mammals, the genes encoding the *Mhc* molecules are clustered together in a single genetic region, the major histocompatibility complex. This region consists of classical and non-classical class I genes, class II genes, and a number of unrelated genes, some of which encode proteins that are involved in the functioning of the immune system.

In non-mammalian vertebrates, the *Mhc* has been most extensively studied in the chicken and the clawed toad, *Xenopus laevis*. In chicken, *Mhc* genes are organized differently from the way they are in mammals. Chicken *Mhc* genes reside in two major clusters, first, in the *B* complex, containing polymorphic class I and class II genes which segregate in accordance with serological typing, and second, the *Rfp-Y* complex. The latter also contains class I and class II genes, but these appear to be less polymorphic and show a lower level of transcription. In the clawed toad, *Xenopus*, classical *Mhc* genes are found in a complex together with unrelated genes, which are homologous to those that are also present in the *Mhc* in mammals. In naturally occurring polyploid species of *Xenopus* a functional diploidization of the *Mhc* is observed, probably due to deletion of all *Mhc* but a single diploid set, most likely as the result of selection pressure against an increase in the number of *Mhc* genes. In addition, *Xenopus* contains a large family of linked non-classical class I genes, which reside on a different chromosome than the classical genes.

In fish, for a long time, the presence of *Mhc* genes has been deduced from indirect evidence, such as acute graft rejection and mixed leukocyte reactivity. The first formal proof for the presence of *Mhc* genes was presented in 1990 by Hashimoto and co-workers who cloned partial class I and class II genes from carp. These *TLAIα-1* (*Cyca-Z*) and *TLAIIβ-1* (*Cyca-YB*) genes, however, lacked the exons encoding the leader, transmembrane and cytoplasmic domains, and expression of these genes was not reported. In a later study, Okamura and co-workers reported kidney cDNA-derived PCR fragments spanning the exons encoding the extracellular domains of *Cyca-Z*. Attempts in our laboratory to isolate full-length cDNA clones of *Cyca-Z* and *Cyca-YB* from cDNA libraries of carp lymphoid organs, have also failed so far.

Chapter 2 describes a study on the number and restriction fragment length polymorphism (RFLP) of *Cyca-Z* and *Cyca-YB* genes in laboratory strains of the common carp, using probes

encoding the α_3 domain of *Cyca-Z* or the β_2 domain of *Cyca-YB*. Six carp strains of different geographical origins were studied. In homozygous gynogenetic carp strains the *Cyca-Z* probe hybridized to 9 - 12 fragments, whereas the *Cyca-YB* probe hybridized to 3-5 fragments. Thus, multiple *Cyca-Z* and *Cyca-YB* genes are present in the genome of carp. The levels of RFLP of the *Cyca-Z* and *Cyca-YB* genes in the strains studied were calculated as the percentage of polymorphic fragments among the total number of fragments, and are shown to be 70% for *Cyca-Z* and 40-66% for *Cyca-YB* genes. In addition, RFLP patterns, using the *Cyca-Z* probe in two gynogenetic families A4.3 and A4.19, were found to correlate with histo-incompatibility. For this, either the *Cyca-Z* genes themselves, or perhaps a closely linked histocompatibility gene could account.

In all species in which *Mhc* molecules were thus far identified, class I molecules are found to consist of a class I α heavy chain, complexed with a light chain, the β_2-microglobulin. The association of the α chain with β_2-microglobulin has been shown to be a prerequisite for the presentation of endogenous peptides. It was, therefore, to be expected that also the carp would possess β_2-microglobulin. The presence of this molecule was previously inferred from immunoprecipitations, but to obtain more definitive proof, we isolated and characterized a genomic fragment from tilapia and the full-length cDNA sequence encoding the β_2-microglobulin of carp (**chapter 3**). Both sequences show strong similarities to all previously published vertebrate β_2-microglobulin sequences. The predicted protein secondary structure of the carp amino acid sequence is almost identical to the corresponding regions of previously known vertebrate β_2-microglobulin protein sequences. In northern hybridizations a message of 800 - 1000 bases was detected. Southern blotting revealed two hybridizing fragments in the carp, while a single hybridizing fragment was detected in DNA of tilapia. Phylogenetic analyses indicate that the fish sequences are related to the β_2-microglobulins of higher vertebrates but group together in an ancestral position.

Since the initial identification of the first class I gene from carp, *Cyca-Z*, doubts have been raised about the functionality of these genes. Even now, 5 years later, still no sequences other than kidney cDNA fragments encoding the extracellular α_1, α_2 and α_3 domains, have been reported. Also, *Cyca-Z* sequences do not share many of the characteristics present in classical and, to a lesser extent, in non-classical class I amino acid sequences from other species. Following the first report on the *Cyca-Z* sequences, however, full-length class I cDNA clones were then obtained from Atlantic salmon and later zebrafish. These class I genes proved to be highly dissimilar from the *Cyca-Z* genes, and in addition, they contain much more of the conserved class I features which are absent in *Cyca-Z*. This prompted us to study whether such a class I gene, more similar to the Atlantic salmon and zebrafish gene, would be present in carp too (**chapter 4**). We succeeded in isolating three representatives of a new group of

carp class I genes, designated *Cyca-UA1*01* (*Cyca*-12), *Cyca-UAW1* (*Cyca*-SP1), and *Cyca-UA1*02* (*Cyca*-G11). These sequences show considerable conservation in known structurally and functionally important regions of class I molecules. The genomic organization of *Cyca-UA1*01* was elucidated, and it was found to be similar to class I gene structures in mammals. Nevertheless, the 3' untranslated region was found to contain an intron, which is unprecedented in class I genes, and intron 2 was found to be exceptionally large (approximately 14 kilobasepairs). Southern blot analyses indicate the presence of multiple related sequences, of which presumably only a single gene is expressed. Phylogenetic analyses indicate that the *U* class I-lineage of genes evolved before the salmonid/cyprinid split, approximately 120-150 my ago, and that this lineage clusters away from the *Cyca-Z* genes. A polyclonal antiserum was raised against a recombinant fusion protein containing the extracellular domains of *Cyca-UA1*01*. The antibodies detected a protein of M_r 45,000 in membrane lysates of spleen and muscle. In FACS analyses, antibody-specific determinants could be identified on leukocytes, but not erythrocytes and thrombocytes. Taken together, the characteristics of the *Cyca-U* sequences and the molecules they encode, strongly indicate that these genes encode classical *Mhc* class I molecules.

In addition to the three class I lineages identified in carp (*Cyca-Z*, *Cyca-U*, and *Cyca*-C4), another completely distinct class I sequence, *Cyca*-TC16, was serendipitously obtained by PCR on thymus cDNA (**chapter 5**). This sequence appeared to be most similar to exon 4 of the class I genes (*Lach-U*) from the coelacanth (42-46% amino acid identity), and the predicted secondary structure was similar to the α_3 domain of *HLA-A2*. In addition, three β_2-microglobulin contact residues were found to be conserved in *Cyca*-TC16. Phylogenetic analyses of fish class I sequences reveals the presence of four distinct clusters: (i) *Z* genes from carp and Ginbuna crucian carp, (ii) *U* genes from carp, zebrafish and Atlantic salmon, (iii) *Cyca*-TC16 with *Lach-U*, and (iv) *Cyca*-C4.

To complete the set of *Mhc* genes from carp, we set out to isolate cDNA sequences encoding the class II α chain, from a gynogenetic carp clone A410, which had already been typed to be *Cyca-UA1*01* homozygous. In addition, we isolated the class II *B* sequences expressed in this strain to complete its typing (**chapter 6**). Two class II *A* cDNA sequences (*Cyca-DXA1* and *Cyca-DXA2*), and two class II *B* cDNA sequences (*Cyca-DAB3*01* and *Cyca-DAB4*01*) were isolated from this strain. Preliminary segregation studies have been performed to analyse the interrelations of the four carp class II *B* cDNA sequences thus far isolated. The four genes were found to be linked in pairs: *Cyca-DAB1*01* linked to *Cyca-DAB2*01* (refuting the previous assumption that these two sequences were alleles of a single locus), and *Cyca-DAB3*01* to *Cyca-DAB4*01*. These two linked pairs, however,

segregate independently, although they most likely do not represent alleles of each other. This indicates the presence of haplotype polymorphisms. The two class II *A* sequences also most likely represent two different loci, as these were found not to segregate among members of the clone, produced by meiogynogenesis.

In **chapter 7**, the results obtained in the previous chapters are combined and discussed. The carp contains several old, distinct lineages of class I genes, but this species most likely expresses only a single classical class I gene, the *Cyca-U* gene. The function of the other lineages remains yet to be established. In contrast to the large genetic distances between the class I lineages, the class II genes of fish are a much more homogeneous group, and orthologous loci appear to be present in carp and zebrafish. The presence of two cysteine residues in the α_1 domain of class II α chains of fish supports the hypothesis that, in the evolution of *Mhc* genes, the class II genes arose first, giving rise to class I genes later. In future studies, it will be interesting to study the ploidy status of the carp *Mhc*. The carp is considered to be a tetraploid fish, of which a large number of loci have been functionally diploidized. The expression of most likely only a single *Cyca-U* locus, therefore, raises the question whether the carp has perhaps silenced or deleted one of its diploid sets, or that it has retained a tetraploid *Mhc*. The set of class I, *B2m*, class II *A* and class II *B* genes now identified in carp, provides a solid basis for future studies on the function of the encoded molecules in the immune response of carp.

Samenvatting

In **hoofdstuk 1** wordt een overzicht gegeven van de structuur en functie van het major histocompatibility complex (*Mhc*) van verschillende diersoorten, en van de moleculen die door de genen van het *Mhc* gecodeerd worden. *Mhc* moleculen spelen een essentiële rol in de specifieke immuunrespons, doordat ze antigenen in fragmentvorm presenteren aan T lymfocyten. Er bestaan twee klassen polymorfe *Mhc* moleculen, nl. klasse I moleculen die endogene peptiden aan CD8-positieve cytotoxische T cellen presenteren, en klasse II moleculen die exogene peptiden aan CD4-positieve T-helper cellen presenteren. Uit onderzoek in verschillende diersoorten is niet of nauwelijks verschil gebleken in de functie van klasse I en klasse II *Mhc* moleculen. De genen die voor *Mhc* moleculen coderen liggen bij zoogdieren bij elkaar op het chromosoom in een gencomplex, het major histocompatibility complex. Het zoogdier *Mhc* bevat klassieke en niet-klassieke klasse I genen, klasse II genen, en een aantal daaraan niet verwante genen. Een aantal genen van deze laatste groep coderen wel voor eiwitten die een rol spelen in het immuun systeem.

Van de overige vertebraten is het *Mhc* het best bestudeerd bij de kip en de klauwpad, *Xenopus laevis*. Anders dan bij zoogdieren, zijn de *Mhc* genen van de kip gelegen in twee clusters. De eerste, het *B* complex, bevat polymorfe klasse I en klasse II genen, waarvan de segregatie overeenkomt met serologische typeringen. De tweede cluster, het *Rfp-Y* complex, bevat ook klasse I en klasse II genen, maar deze genen zijn minder polymorf, en ze komen ook minder sterk tot expressie. Bij de klauwpad, *Xenopus*, komen naast de *Mhc* genen een aantal niet verwante genen in hetzelfde gencomplex voor. Deze laatstgenoemde genen zijn homoloog aan vergelijkbare genen zoals die ook bij zoogdieren in het *Mhc* worden aangetroffen. Bij natuurlijk voorkomende polyploide *Xenopus* soorten blijkt het *Mhc* functioneel diploied te zijn geworden, waarschijnlijk als gevolg van deletie van alle overige copieën van het *Mhc*. Deze deletie heeft in de loop van de evolutie plaatsgevonden, vermoedelijk als gevolg van negatieve selectiedruk op een toename van het aantal *Mhc* genen dat tot expressie komt. Naast de klassieke *Mhc* genen, bezit *Xenopus* verder een grote familie van niet-klassieke genen. Deze liggen ook in een cluster, maar op een ander chromosoom dan waarop de klassieke genen voorkomen.

Het bestaan van een *Mhc* bij vissen kon gedurende lange tijd slechts worden afgeleid uit indirekte aanwijzingen, zoals het verschijnsel van acute afstoting van allo-transplantaten, en het optreden van celproliferatie in gemengde leukocytenkweken. Pas in 1990 werd door de groep van Hashimoto het eerste harde bewijs geleverd voor de aanwezigheid van *Mhc* genen in vis. Zij isoleerden een partiëel klasse I gen en een partiëel klasse II gen uit de karper. Deze genen, *TLAIα-1* (*Cyca-Z*) en *TLAIIß-1* (*Cyca-YB*), bevatten slechts delen van hetgeen bij de *Mhc* genen van hogere gewervelden werd aangetroffen, en ook werd er geen expressie van aangetoond. In het vervolg hierop beschreef de groep van Okamura enkele jaren later PCR

fragmenten van cDNA uit de nier, maar ook hier was sprake van een partiële sequentie, nl. bestaande uit exonen 2, 3 en 4 van *Cyca-Z*. Pogingen die werden ondernomen in onze groep om complete cDNA sequenties van *Cyca-Z* of *Cyca-YB* uit banken van lymphoide organen te isoleren, waren eveneens niet succesvol.

In **hoofdstuk 2** wordt een studie beschreven naar het aantal *Cyca-Z* and *Cyca-YB* genen bij de karper, en naar het restrictiefragment lengte polymorfisme (RFLP) dat deze genen vertonen in een aantal in ons laboratorium voor immunologisch onderzoek ontwikkelde karperlijnen. Hiertoe zijn hybridisaties uitgevoerd met probes die coderen voor het α_3 domein van *Cyca-Z* of het $ß_2$ domein van *Cyca-YB*. Zes karperlijnen van verschillende geografische herkomst werden bestudeerd. In homozygote gynogenetische karperlijnen hybridiseerde de *Cyca-Z* probe met 9-12 DNA fragmenten, terwijl de *Cyca-YB* probe 3-5 fragmenten detecteerde. In het genoom van de karper zijn dus meerdere *Cyca-Z* en *Cyca-YB* genen aanwezig. De graad van RFLP van deze genen is uitgedrukt als het percentage polymorfe fragmenten ten opzichte van het totaal aantal fragmenten. Voor *Cyca-Z* bleek 70% van de fragmenten polymorf, terwijl dit voor *Cyca-YB* 40-66% was. De RFLP patronen van *Cyca-Z* in twee gynogenetische karper families, A4.3 en A4.19, correleerden met histo-incompatibiliteit. Deze histo-incompatibiliteit kan dus een direkt gevolg zijn van de variabiliteit van een *Cyca-Z* gen, of van een ander gen dat vlakbij dit *Cyca-Z* gen gelegen is.

Bij alle diersoorten waar *Mhc* moleculen zijn geïdentificeerd, is gevonden dat klasse I α ketens op de membraan van de cel geassocieerd zijn met een lichte keten, het $ß_2$-microglobuline. Deze associatie is essentiëel gebleken voor de presentatie van endogene peptiden door klasse I moleculen. Bij de karper was het bestaan van een $ß_2$-microglobuline molecuul aannemelijk gemaakt op grond van de resultaten van immunoprecipitatie experimenten, maar om dit bestaan met meer zekerheid vast te stellen hebben we zowel een genomisch DNA fragment uit tilapia als een complete cDNA sequentie uit karper geïsoleerd, die beide coderen voor het $ß_2$-microglobuline (**hoofdstuk 3**). Op eiwitniveau vertonen beide sequenties grote gelijkenis met de $ß_2$-microglobuline sequenties van andere gewervelde dieren. In northern hybridisaties detecteerde de $ß_2$-microglobuline probe een transcript van circa 800-1000 bp. Bij Southern hybridisaties op karper DNA werden twee hybridiserende fragmenten gevonden, terwijl slechts één enkel fragment hybridiseerde in het DNA van tilapia. Fylogenetische analyse op eiwitniveau maakte duidelijk dat de $ß_2$-microglobuline sequenties van karper en tilapia weliswaar verwant zijn aan $ß_2$-microglobuline sequenties van hogere gewervelde dieren, maar dat ze wel een aparte tak vormen.

Na de eerste identificatie van een karper klasse I gen (*Cyca-Z*), rezen er twijfels over de functionaliteit van dit gen. Zelfs nu, 5 jaar later, zijn er nog geen andere cDNA sequenties van dit type beschreven dan die geïsoleerd uit nier, en coderend voor extracellulaire α_1, α_2

en α_3 domeinen. Daarbij komt dat veel van de karakteristieke eigenschappen van klassieke en in mindere mate de niet-klassieke klasse I genen van andere diersoorten, afwezig zijn in *Cyca-Z*. Na de eerste publicatie van de *Cyca-Z* genen, werden echter spoedig complete klasse I cDNA sequenties geïsoleerd, eerst uit zalm en later uit de zebravis. Deze sequenties verschillen sterk van die van de *Cyca-Z* genen, en bovendien bezitten ze ten opzichte van *Cyca-Z* veel meer bij alle klasse I genen geconserveerde eigenschappen. Dit bracht ons er toe aan na te gaan of de karper behalve *Cyca-Z* misschien nog een klasse I gen bezat, waarvan de sequentie meer op die van zalm en zebravis leek (**hoofdstuk 4**). We hebben drie vertegenwoordigers van een nieuwe groep karper klasse I genen gevonden, aangeduid met *Cyca-UA1*01* (*Cyca*-12), *Cyca-UAW1* (*Cyca*-SP1) en *Cyca-UA1*02* (*Cyca*-G11). De aminozuur residuen die belangrijk zijn voor de structuur en functie van klasse I moleculen, zijn in deze sequenties sterk geconserveerd. Vervolgens is de genomische organisatie van *Cyca-UA1*01* opgehelderd, en het blijkt dat deze sterk vergelijkbaar is met die van klasse I genen bij zoogdieren. Uitzonderlijk is echter de aanwezigheid van een intron in het onvertaalde deel aan de 3' kant, en bovendien blijkt intron 2 bijzonder groot (ongeveer 14 kilobaseparen). Uit Southern blot analyses kon opgemaakt worden dat er meerdere verwante *Cyca-U* sequenties in het genoom van de karper voorkomen, waarvan er waarschijnlijk maar één tot expressie komt. Uit fylogenetische analyses wordt de conclusie getrokken dat de groep van klasse I *U*-genen waarschijnlijk al is ontstaan voor de scheiding tussen cypriniden en de salmoniden tijdens de evolutie van de Teleostei, ongeveer 120-150 miljoen jaar geleden. Bovendien blijkt dat de *U*-genen fylogenetisch beschouwd ver af staan van de *Z*-genen. Een polyclonaal antiserum werd opgewekt tegen een recombinant fusie eiwit, dat de extracellulaire domeinen van *Cyca-UA1*01* omvat. De antilichamen detecteerden een eiwit van M_r 45,000 in membraanlysaten van milt en spier, en bovendien reageerden ze in FACS analyses met een antigene determinant aanwezig op leukocyten, maar niet op erythrocyten en thrombocyten. Dit alles samen leidt tot de conclusie dat *Cyca-U* genen coderen voor klassieke *Mhc* klasse I moleculen.

Naast de drie al geïdentificeerde types van klasse I genen in de karper (*Cyca-Z*, *Cyca-U*, en de afzonderlijk beschreven *Cyca*-C4), is ten gevolge van serendipiteit een vierde compleet verschillende klasse I sequentie gevonden, aangeduid met *Cyca*-TC16 (**hoofdstuk 5**). Deze sequentie vertoont de grootste gelijkenis met exon 4 van de klasse I genen (*Lach-U*) van de coelacanth (42-46% identiek in aminozuursequenties), en de voorspelde secundaire eiwitstructuur van *Cyca*-TC16 lijkt op die van het α_3 domein van *HLA-A2*. Ook bevat de sequentie van *Cyca*-TC16 drie geconserveerde aminozuur residuen die bij zoogdieren belangrijk zijn voor het contact van de klasse I α keten met $ß_2$-microglobuline. Uit een fylogenetische analyse van klasse I sequenties van vissen blijken er vier zeer verschillende clusters te bestaan: (i) de *Z* genen van de karper en de Ginbuna kroeskarper, (ii) *U* genen van de karper, zebravis en zalm,

(iii) *Cyca*-TC16 samen met *Lach-U*, en (iv) *Cyca*-C4.

Om de set van karper *Mhc* genen compleet te maken, zijn vervolgens uit een gynogenetische karperkloon cDNA sequenties coderend voor de klasse II α keten geïsoleerd. Deze kloon was al eerder getypeerd voor klasse I, en bleek daarbij homozygoot voor *Cyca-UA1*01*. Om de typering van deze kloon compleet te maken, werden naast klasse II *A* ook de klasse II *B* cDNA sequenties geïdentificeerd (**hoofdstuk 6**). De karperkloon bleek twee klasse II *A* (*Cyca-DXA1*01* en *Cyca-DXA2*01*) en twee klasse II *B* cDNA sequenties (*Cyca-DAB3*01* en *Cyca-DAB4*01*) tot expressie te brengen. Om de relatie tussen de vier tot nu toe geïdentificeerde klasse II *B* genen van de karper vast te stellen, werd een beperkte segregatiestudie uitgevoerd. De vier genen bleken in twee gekoppelde paren voor te komen: *Cyca-DAB1*01* aan *Cyca-DAB2*01* (hetgeen niet in overeenstemming is met de aanname dat deze sequenties allelen van één locus zouden zijn), en *Cyca-DAB3*01* aan *Cyca-DAB4*01*. De twee paren segregeerden echter onafhankelijk van elkaar, hetgeen geinterpreteerd werd als haplotype polymorfismen. Ook de twee klasse II *A* sequenties zijn waarschijnlijk afkomstig van twee verschillende loci, omdat geen segregatie optrad in een d.m.v. meiogynogenese gegenereerde karperkloon nakomelingenschap.

In **hoofdstuk 7** worden de resultaten uit de vorige hoofdstukken gecombineerd en bediscussiëerd. In de karper bestaan meerdere, zeer verschillende en oude types klasse I sequenties, maar hoogstwaarschijnlijk komt maar één enkel klassiek klasse I gen tot expressie, het *Cyca-U* gen. Het is nog onduidelijk wat voor functie de andere types hebben. In tegenstelling tot de grote genetische afstanden die de klasse I types van elkaar scheiden, vormen de klasse II genen van vissen een meer homogene groep, en sommige genen van karper en zebravis lijken zelfs ortholoog te zijn. In het α_1 domein van klasse II α ketens van vissen zijn twee cysteine residuen aanwezig. Dit ondersteunt de hypothese dat klasse II genen in de evolutie van het *Mhc* het eerst zijn ontstaan, en dat klasse I genen vervolgens uit de klasse II genen zijn voortgekomen. Het is van belang om in vervolgstudies meer duidelijkheid te verkrijgen omtrent de ploidie-status van de karper. De karper is een tetraploide vis, waarin een groot aantal loci naar een diploiede status zijn geëvolueerd. De expressie van slechts één enkel *Cyca-U* locus roept vragen op over het aantal aanwezige copieën van het *Mhc* in de karper. Het zou kunnen zijn dat deze soort een diploiede set van het *Mhc* heeft uitgeschakeld of zelfs heeft verwijderd uit zijn genetisch materiaal.

De complete set van *Mhc* klasse I, *B2m*, klasse II *A* en klasse II *B* genen die nu geïdentificeerd zijn bij de karper vormen een goede basis voor verder onderzoek naar de functie van *Mhc* moleculen in de immuunrespons van de karper.

Curriculum vitae

Saskia Henriëtte Martina van Erp werd geboren op 19 april 1967 in Nijmegen. In 1985 deed zij eindexamen aan het Stedelijk Gymnasium te Nijmegen, en begon zij met haar studie Moleculaire Wetenschappen aan de toenmalige Landbouwhogeschool te Wageningen. In juni 1991 studeerde zij af, met als afstudeervakken Biochemie en Celbiologie & Immunologie, en een biochemische stage aan het Bakh-instituut voor biochemie van de Sovjet Academie van Wetenschappen in Moskou. Per 15 mei 1991 werd zij aangesteld als assistent-in-opleiding bij de vakgroep Experimentele Diermorfologie en Celbiologie van de Landbouwuniversiteit te Wageningen. De resultaten van het promotieonderzoek dat aldaar werd uitgevoerd vindt U beschreven in dit proefschrift.