Immunogenetics (2019) 71:233–249
https://doi.org/10.1007/s00251-018-1090-2

REVIEW

Ancient features of the MHC class II presentation pathway, and a model for the possible origin of MHC molecules

Johannes M. Dijkstra[1] · Takuya Yamaguchi[2]

Received: 16 August 2018 / Accepted: 6 October 2018 / Published online: 30 October 2018

Abstract

Major histocompatibility complex (MHC) molecules are only found in jawed vertebrates and not in more primitive species. MHC class II type structures likely represent the ancestral structure of MHC molecules. Efficient MHC class II transport to endosomal compartments depends on association with a specialized chaperone, the MHC class II invariant chain (aliases Ii or CD74). The present study identifies conserved motifs in the CLIP region of CD74 molecules, used for binding in the MHC class II binding groove, throughout jawed vertebrates. Peculiarly, in CD74a molecules of Ostariophysi, a fish clade including for example Mexican tetra and zebrafish, the CLIP region has duplicated. In mammals, in endosomal compartments, the peptide-free form of classical MHC class II is stabilized by binding to nonclassical MHC class II "DM," a process that participates in "peptide editing" (selection for high affinity peptides). Hitherto, DM-lineage genes had only been reported from the level of amphibians, but the present study reveals the existence of *DMA* and *DMB* genes in lungfish. This is the first study which details how classical and DM lineage molecules have distinguishing glycine-rich motifs in their transmembrane regions. In addition, based on extant MHC class II structures and functions, the present study proposes a model for early MHC evolution, in which, in an ancestral jawed vertebrate, the ancestral MHC molecule derived from a heavy-chain-only antibody type molecule that cycled between the cell surface and endosomal compartments.

Keywords MHC class II · CD74 · Invariant chain · HLA-DM · Transmembrane domain · Evolution

Introduction: overview of MHC class II presentation

Classical major histocompatibility complex (MHC) molecules are noncovalent heterodimers which consist of two similar membrane-distal domains which together form a peptide binding structure, two membrane-proximal immunoglobulin superfamily (IgSF) domains, and one (in class I) or two (in class II) transmembrane plus cytoplasmic tail regions (Bjorkman et al. 1987; Brown et al. 1993; Neefjes et al. 2011). In MHC class II, the IIα and IIβ heterodimer components are similar in organization, whereas MHC class I molecules consist of a large heavy chain and the single IgSF domain molecule β_2-m. By arguments of parsimony, it has been concluded that the more symmetric structure of MHC class II probably is more similar to the ancestral MHC molecule, which is believed to have been a homodimer (e.g., Kaufman et al. 1984a, 1992; Hughes and Nei 1993). The larger similarity of class II to a proposed symmetric homodimer ancestor is also supported by the fact that class II can bind peptides in either direction (Günther et al. 2010), whereas class I-bound peptides are always in the same N-to-C orientation. Both MHC class I and II molecules present peptides at the cell surface for screening by the T cell receptors (TCRs) of T cells (Neefjes et al. 2011).

Classical MHC class I and II functional properties also favor a model in which class II is more reminiscent of the ancestral MHC type. Namely, whereas classical MHC class II binds relaxed peptides in polyproline type II helical conformation, with the ends allowed to extend beyond the class II binding groove, binding of peptides by class I is more demanding, with no well-understood processes ensuring that

This article is part of the Topical Collection on *Biology and Evolution of Antigen Presentation*

✉ Johannes M. Dijkstra
Dijkstra@fujita-hu.ac.jp

[1] Institute for Comprehensive Medical Science, Fujita Health University, Dengaku-gakubo 1-98, Toyoake, Aichi-ken 470-1192, Japan

[2] Laboratory of Fish Immunology, Friedrich-Loeffler-Institute, Boddenblick 5A, 17498 Insel Riems, Greifswald, Germany

both termini of typically bulged peptides are bound in the class I binding groove pockets A and F (Stern and Wiley 1994). Furthermore, whereas classical MHC class II molecules present peptides to helper or regulatory T cells which can help decide on how to proceed with the immune response, classical MHC class I molecules present peptides to cytotoxic T cells which can eliminate the presenting cell (Castellino and Germain 2006; Neefjes et al. 2011; Roche and Furuta 2015). An evolutionary “half-ready” primordial function for an MHC-restricted T cell system can be more easily envisioned for T helper/reg functions than for the more dangerous cytotoxic functions. We assume that only after a refined T cell education system for distinguishing self from non-self has been established, T cells also started to be used for cytotoxic functions.

Among the described MHC molecules, classical MHC class I and II appear to be the oldest, and, with some exceptions (Malmstrøm et al. 2016; Dijkstra and Grimholt 2018), they are stably inherited throughout jawed vertebrates (Dijkstra et al. 2013, 2018; Grimholt et al. 2015). Duplicates of classical MHC genes were used for the establishment of a variety of derived “nonclassical” functions, which have a more limited species distribution (Hughes and Nei 1989; Adams and Luoma 2013; Mellins and Stern 2014; Grimholt et al. 2015; Dijkstra et al. 2013, 2018). The oldest known nonclassical MHC class II lineage is DM, and mammalian DM molecules participate in the peptide loading pathway of classical MHC class II (see below).

The binding of peptides by MHC class II involves a network of conserved hydrogen bonds between the MHC class II molecule and the peptide ligand backbone, plus insertions of several peptide ligand sidechains into the MHC class II peptide binding domain groove pockets (reviewed by Painter and Stern 2012). The core of the bound part of the peptide is 9 aa, and the sidechains that are largely buried within groove pockets are residues P1, P4, P6, and P9 of that core, with P1 and P9 sidechains typically pointing perpendicular into the groove. The pockets are specific for each MHC class II molecule and determine which peptides are preferentially bound. The binding ensures that peptides can be efficiently presented for long times, with binding half-lives of days to weeks, which is very unusual among molecules of this size; even “low affinity” peptides can bind for several hours, which explains the need for regulation of which peptides are bound (Yin et al. 2012; Pos et al. 2013). The important function of MHC class II molecules is the presentation of peptides derived from exogenous antigens. Hence, the place where the molecules are loaded with such peptides are late endosomal compartments, where also antigens internalized by phagocytosis, micropinocytosis, or endocytosis, but also self-antigens from lysosomes, traffic to (Tulp et al. 1994; Blum et al. 2013). However, the MHC class IIα and IIβ chains are assembled in the endoplasmic reticulum (ER), and without bound peptide, they are not very stable (Stern and Wiley 1992; Marks et al. 1995), which is exemplified by the fact that structures of peptide-free classical MHC class II have not been determined yet. In the ER, their promiscuous binding of peptide stretches creates the potential problem of them being able to bind stretches of misfolded or not yet fully folded proteins (Busch et al. 1996; Fortin et al. 2013). Solving these problems, in evolution, a specialized chaperone was created, named MHC class II invariant chain (Ii) or CD74, which appears to be present in all species with MHC class II (Criscitiello et al. 2012, and see below). The CD74 molecules have a non-structured region, the class II-associated invariant chain peptide (CLIP), which binds into the MHC class II binding groove and also have cytoplasmic tail motifs that deliver the CD74/MHC complexes into the endocytic pathway (see the CD74 paragraph below). In the late endosomes, the CD74 molecules are degraded by several proteases including the cathepsins S, L, and F and the homolog signal peptide peptidase-like 2a (SPPL2a) (Bird et al. 2009; Colbert et al. 2009; Hüttl et al. 2016). The last part of CD74 that can be found associated with the MHC class II molecules is a nested set of peptides of 16–24 amino acids in length of the CLIP region, which corresponds to CD74 positions 81–104 (Rudensky et al. 1991; Riberdy et al. 1994). What then, at least in the case of mammals, follows, is a competition between CLIP peptide, antigen peptides, and DM molecules for binding to the MHC class II molecule. The peptides compete directly for the binding space in the groove, whereas the DM molecules and peptides compete through stimulating two different structures of the MHC class II molecules that each are not compatible with efficiently binding both DM and peptide (Fung-Leung et al. 1996; Pos et al. 2012; Schulze and Wucherpfennig 2012). In this way, DM increases the speed with which antigenic peptides can replace CLIP, and enhances the presentation of high-affinity peptides, namely those that can outcompete DM, by MHC class II at the cell surface. This is part of a process called “peptide editing.”

The above-described mechanisms form the major principles of an elegant system, which can readily be understood. However, fine-tuning of the presented peptide arrays is achieved by for example MHC class II polymorphism (Rammensee et al. 1995), by the expression profiles of proteases which can differ between cell types and conditions (Adler et al. 2017), by the levels of co-expression of nonclassical MHC class II (Denzin 2013), and by which CD74 isoform is expressed based on alternative *CD74* transcripts (Fortin et al. 2013). In the present study, we will not discuss those items, except highlighting the fact that throughout jawed vertebrate species, transcripts are found for both CD74 with and without the C-terminal thyroglobulin type-1 (TG) domain, which is thought to affect cathepsin activity (see the CD74 paragraph below).

The important contributions which we believe that the present study makes to the understanding of the evolution of MHC class II loading pathways are (1) an improved analysis of CD74 sequences found throughout jawed vertebrates, detecting CLIP core motifs in all sequences; and two CLIP motifs in CD74a sequences of fish like zebrafish and Mexican tetra; (2) detection of DM lineage sequences in lungfish and better-defining motifs in the transmembrane plus cytoplasmic tail regions that distinguish DM molecules from classical molecules; and (3) a hypothetical model for how in evolution MHC class II-like function might originally have been performed by an antibody from which the ancestral MHC molecules then originated.

Conservation of CD74 sequence motifs from cartilaginous fish to human

CD74 is a non-polymorphic type II transmembrane protein that can form homotrimers. Each of the CD74 molecules within those trimers retains an ability to bind a single MHC class IIαβ complex, and so nonamer complexes can be formed (Jones et al. 1979; Lamb and Cresswell 1992). However, the CD74 trimers can also associate with fewer MHC class II molecules (Koch et al. 2011; Majera et al. 2012; Cloutier et al. 2014; Cresswell and Roche 2014). The functions of CD74, already listed above, are (i) to bind the assembled MHC class II α and β chains in the endoplasmic reticulum and prevent them from aggregating and from binding other peptide fragments in their binding groove, (ii) to deliver the MHC class II molecules to endosomal compartments, and (iii) there to participate in the exchange of CLIP peptide for higher affinity peptides. CD74 knockout mice show reduced levels of MHC class II at cell surfaces, and normal developments/responses of helper T and B cells are impeded (Bikoff et al. 1993; Viville et al. 1993; Shachar and Flavell 1996). CD74 has also been proposed to have chaperone functions for other molecules, for example for CD1 (reviewed by Gelin et al. 2009) and TLR7 (Tohmé and Manoury 2015), and are even believed to have non-chaperone functions at the cell surface (e.g., Naujokas et al. 1993; Gore et al. 2008). However, it is notable that in Atlantic cod and other gadiform fishes which seem to have lost intact *MHC class II A* and *B* genes, also intact *CD74* genes seem to have been lost (Star et al. 2011; Malmstrøm et al. 2016). This suggests that, at least in the ancestor of these fishes, CD74 function was predominantly dedicated to MHC class II presentation. Dedication of CD74 function to MHC class II presentation is also suggested by the co-expression patterns of *CD74* and *MHC class II* genes in species as variable as for example nurse shark (Criscitiello et al. 2012), rainbow trout (Dijkstra et al. 2003), and mouse (Walter et al. 2000). Furthermore, co-precipitation studies have shown/suggested the binding between CD74 and MHC class II in mammals, chicken, and teleost fish (Jones et al. 1979; Guillemot et al. 1986; Bremnes et al. 2000; Ye et al. 2009; Chen et al. 2017). Throughout jawed vertebrates, transcripts for both CD74 with and without the C-terminal thyroglobulin type-1 (TG) domain are found, for example in human (Strubin et al. 1986; O'Sullivan et al. 1987), mouse (Koch et al. 1987), chicken (Bremnes et al. 2000; Fujiki et al. 2003; GenBank accession AJ292038), zebrafish (for CD74a; Yoder et al. 1999), and nurse shark (Criscitiello et al. 2012). In human, but not throughout mammals, additional CD74 forms with ER retention signals are found based on alternative use of start codons (Strubin et al. 1986; O'Sullivan et al. 1987; Lotteau et al. 1987; Fortin et al. 2013), but this article will not further discuss them as they do not seem to represent an ancient CD74 feature.

A good study on CD74 evolution is by Criscitiello et al. (2012), who describe the *CD74* gene in cartilaginous fish. They describe that the *CD74* gene was probably only established in jawed vertebrates together with the MHC system, and our investigations of more recent datasets (not shown) suggest that this may be true. Compared to the Criscitiello study, probably based on newer genomic sequence information (GenBank accession AAVX02033104), the present study was able to deduce a full-length CD74 molecule for primitive elephant shark (a chimera), whereas Criscitiello et al. (2012) could only predict part of the molecule. Furthermore, we newly identified the CD74 sequence of a primitive ray-finned fish, bichir. More importantly, compared to previous studies, including our own study (Dijkstra et al. 2003), the present study made a more compelling alignment which allows a better recognition of some of the conserved motifs (Fig. 1). These motifs will be subsequently discussed below. The most exciting finding probably is that a CLIP motif with likely P1, P4, P6, and P9 residues can be distinguished in all investigated CD74 sequences. For each teleost fish, two CD74 sequences are shown, CD74a and CD74b. The existence of these two types of sequences was first described for zebrafish (Yoder et al. 1999), and later found for other teleost fishes (e.g., Dijkstra et al. 2003; Fujiki et al. 2003). Our analysis of genomic sequence databases (details not shown) suggests that teleost fish *CD74A* and *CD74B* genes were derived from the whole genome duplication (WGD) event early in the teleost lineage. For example, according to the Ensembl database (https://www.ensembl.org/), the *CD74A* and *CD74B* genes reside on chromosomes 14 and 21, respectively, which is a pair of chromosomes for which multiple probable WGD-derived ohnologous gene pairs were reported (Howe et al. 2013). At least in rainbow trout, expression patterns of *CD74A* and *CD74B* genes are similar (Dijkstra et al. 2003). To our knowledge, only for teleost

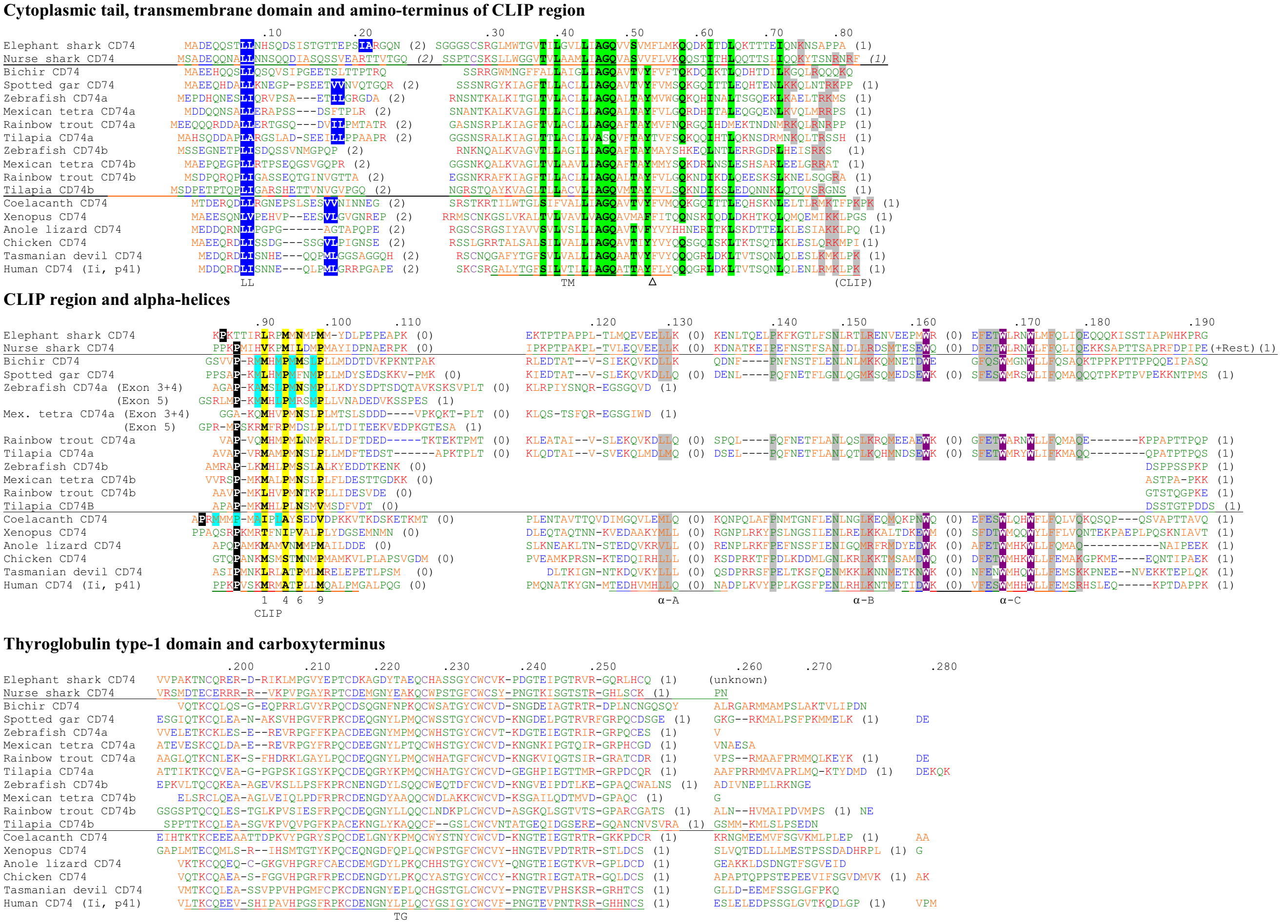


Fig. 1 Alignment of deduced CD74 sequences of representative species. The depicted CD74 sequences were retrieved from the following GenBank accession numbers, or elsewhere: Elephant shark (*Callorhinchus milii*), Criscitiello et al. (2012), and GenBank accession AAVX02033104; Nurse shark (*Ginglymostoma cirratum*), expected based on Criscitiello et al. (2012); Bichir (*Polypterus senegalus*), overlapping SRA reads from set SRX796491; Spotted gar (*Lepisosteus oculatus*), GFIM01016794 and AHAT01005333; Zebrafish (*Danio rerio*) CD74a, NP_571665; Mexican tetra (*Astyanax mexicanus*) CD74a, XP_007232273, GFIE01002800, and NW_019172837; Rainbow trout (*Oncorhynchus mykiss*) CD74a, AAL91668; Tilapia (*Oreochromis niloticus*) CD74a, XP_003455233; Zebrafish CD74b, NP_571447; Mexican tetra CD74b, XP_022531879; Rainbow trout CD74b, AAL91669; Tilapia CD74b, XP_003438357; Coelacanth, XP_014351519; Xenopus (*Xenopus laevis*), AAH59976; Anole lizard (*Anolis carolinensis*), Criscitiello et al. (2012); chicken (*Gallus gallus*), AJ292038; Tasmanian devil (*Sarcophilus harrisii*), XP_012407251; and human (*Homo sapiens*) p41, P04233. The numbers above the alignment follow the human p41 sequence. Between fragments encoded by separate exons, a number for the phase of the corresponding intron is given between brackets; the intron-exon organization was determined by comparison with the appropriate genomic sequences. For nurse shark and bichir CD74, we could not compare with species-specific genomic sequences; in the depicted nurse shark CD74 sequence, the indicated expectations of introns are based on comparison with the whale shark (*Rhincodon typus*) genomic scaffold of GenBank accession LVEK02008796. For the figure, the CD74 sequences were aligned by hand, based on considerations of similarity and phylogeny, intron-exon organization, and expected structural features. Not for all regions the alignments can be deemed reliable, but we do feel certain that the alignment reliably supports the conclusions made in the main text. Motifs indicated by color shading are explained in the main text. Non-shaded residues were colored purple in case of cysteines, and, based on Hopp and Woods (1981), red font was used for basic residues, blue for acidic residues, and green and orange for the other residues with the more hydrophilic ones in green. LL: conserved dileucine(-like) motif; TM (underlined): human CD74 transmembrane domain; upward triangle: cleavage site for SPPL2a; CLIP (underlined), human CLIP region; 1, 4, 6, 9: positions P1, P4, P6, and P9 of human CLIP fragment; α-A, α-B, and α-C: α-helices determined for human CD74 (Jasanoff et al. 1998); TG (underlined): thyroglobulin type 1 domain

CD74A, and not for teleost *CD74B*, transcripts without the TG domain coding sequence have been found.

Below, the different domains of CD74 are discussed from N- to C-terminus, using our Fig. 1 alignment. We believe that this alignment is better than previous alignments because (1) more sequences are known, (2) our alignment was made in agreement with intron-exon organization, (3) the duplication of the exon 3 sequence in fish of the clade Ostariophysi was recognized, and (4), although that was also done in some previous alignments, the molecular structures and functions were considered. The residue numbering used in descriptions below follows the human CD74 sequence as shown in Fig. 1.

The N-terminal cytoplasmic tail with dileucine(-like) motif

By deletion, mutation, and transfer analyses, it was found that the N-terminal cytoplasmic tail of human CD74 has two motifs, involving the LI and ML motifs, shaded blue in Fig. 1; each of these two motifs independently help to target the molecule to endosomal compartments (Pieters et al. 1993; Bremnes et al. 1994). Figure 1 shows that the first of these two dileucine(-like) motifs is conserved throughout the aligned CD74 sequences, with acidic residues positioned in N-terminal vicinity thereof. It can therefore be expected that all these molecules target to endosomal compartments (Pond et al. 1995; Kozik et al. 2010; Santos et al. 2013).

The transmembrane domain

In Fig. 1, the human CD74 transmembrane (TM) domain is underlined based on region estimations of Ashman and Miller for murine CD74 (1999) and also in agreement with predictions by TMpred software (https://embnet.vital-it.ch/software/TMPRED_form.html). Immediately notable is the unusual high level of conservation of this domain at the residue level (noted before by Bremnes et al. 2000; Dixon et al. 2006), whereas typically only the hydrophobic character of TM domains is well-conserved. Among the highly conserved residues, shaded in green, are the non-hydrophobic residues (S/T)38, G47, Q48, and (S/T)51. It was reported that the CD74 TM domains self-associate as trimers and may contribute to CD74 homo-trimerization, and that substitution of the residues Q48 and T51 in human CD74 resulted in a decreased efficiency of that trimerization (Ashman and Miller 1999; Dixon et al. 2006). However, of note is that also CD74 molecules without TM domain were found to form trimers (Majera et al. 2012). The CD74 TM domain is thought to form a left-handed α-helical structure, bringing Q48 and T51 residues into close proximity of each other on the same face of the TM helix (Dixon et al. 2006). Based on a combination of experiments and modeling, Dixon et al. (2006) proposed a model for human CD74 TM trimer interactions in which their interfaces are formed by the residues F37, S38, V41, L44, Q48, and T51, all situated at the same face of the helix. That model may explain the good conservation of (S/T)38, L44, Q48, and T51 (Fig. 1), and, as we interpret their model, also of the conservation of residues L40 and G47. We do not have a possible explanation yet for the conservation of residue A46.

At position 53 of the C-terminal region of the TM domain, although absent in cartilaginous fish CD74, an aromatic tyrosine or phenylalanine is well conserved in CD74 of Osteichthyes ("bony animals"). The position between Y53 and F54 in human CD74 was found to be an important cleavage site for presenilin homolog signal peptide peptidase-like 2a (SPPL2a), although this protease also cleaved at this CD74 site if residues were replaced with alanine (Hüttl et al. 2016).

The region between TM and CLIP

Criscitiello et al. (2012) noted poor conservation of the region between the transmembrane domain and the CLIP region, and this stretch was believed to be flexible/disordered by Jasanoff et al. (1995, 1998). Furthermore, alanine replacement studies for fragments between residues 55 and 78 could not convincingly find an effect on CD74 function (Frauwirth and Shastri 2001). However, despite the low overall sequence similarity in the region, it is quite notable in Fig. 1 that at positions 62, 65, and 72, with very few exceptions, leucines and isoleucines are found (shaded in green). The distances of these hydrophobic residues would agree with a hydrophobic core of a α-helical coiled-coil structure, and their conservation provides partial support for the model by Koch et al. (2011). Namely, Koch et al. (2011) postulated, based on computer modeling and on the positions of leucines in human CD74, that the stretches of residues 7-to-81 of human CD74 molecules form continuous α-helices that in complexes of three CD74 molecules form coiled-coils. However, whereas the conservation of leucines/isoleucines shown in Fig. 1 provides support for a coiled-coil structure of the N-terminal part of the ectodomain, conservation of a suitable pattern in the cytoplasmic tail region is not observed.

Another notable characteristic of the region between TM and CLIP is that, just N-terminally of CLIP, consistently basic residues are found (shaded gray in Fig. 1). For human CD74 K83 and K86, which reside in the CLIP region, an interaction with acidic residues of MHC class II was proposed (Kropshofer et al. 1995), but that probably is not proven yet, and the fact that the basic residues in this region are not conserved at precise locations argues against them having this proposed ability as an important function.

The CLIP region

Previously, we and others wrestled with the alignments of several diverged parts of the CD74 sequences, including the CLIP region, of widely divergent species (Yoder et al. 1999; Dijkstra et al. 2003; Fujiki et al. 2003; Silva et al. 2007; Criscitiello et al. 2012). However, if exon-intron organization and molecular structural features are considered, and if CD74 comparisons include sequences from many animal clades, we feel that the core part of the CLIP region can rather reliably be aligned (Fig. 1). Making this alignment, it also became immediately obvious that the exon 5 sequence of CD74a in the fish superorder Ostariophysi, which includes species like for example zebrafish (*Danio rerio*, belonging to the order Cypriniformes) and Mexican tetra (*Astyanax mexicanus*, belonging to the order Characiformes), is a duplication of the exon 3 sequence which encodes the core part of CLIP. In previous papers, that had not been realized, and they only reported that the zebrafish *CD74A* exon 5 sequence did not encode the expected CD74-characteristic α-helical trimerization region. Our alignment also shows that, rather consistently, in the region directly N-terminal of a proposed P1-to-P9 core, a basic residue and a proline are found. Already in 1994, the Rammensee research group (Falk et al. 1994) described that at the penultimate N-terminal position of peptides bound to human MHC class II molecules, a proline is often found, which they speculated to protect the peptide from digestion by exopeptidases. This may also explain the proline found at the penultimate N-terminal position of the human CLIP sequence, position 82 in Fig. 1. In regard to the better-conserved proline at position 87 (shaded black in Fig. 1), Kropshofer et al. (1995) found that this proline could have a function in reducing affinity of CLIP for the MHC class II molecule, since replacement for an alanine substantially increased the binding half-life. We speculate that in CD74 of some species, which do not have a proline at position 82 or similar position, the P87 residue is used both for blocking exopeptidase activity and for modifying the binding affinity for MHC class II. Ghosh et al. (1995) determined the structure of the CLIP 87–101 fragment bound to HLA-DR3. That structural report nicely shows how the charged sidechain of the lysine at P-1 (K90 in Fig. 1) points away from the complex. However, possibly because the 81–86 region was not included in the investigated structure, the P87 residue seems not to interfere with any other residues, and from this structural report, the reason for the P87 conservation probably cannot be understood.

According to theory, the CLIP peptide should be able to bind into the peptide binding groove of all classical MHC class II molecules in a species. However, that binding should not be too strong, especially not under the low pH of endosomal compartments, as otherwise the fragment cannot be replaced with peptides derived from pathogens. Sidechains of the residues at the human CD74 CLIP positions P1, P4, P6, and P9 (1, 4, 6, and 9, and indicated in yellow shading in Fig. 1) are small to medium-sized hydrophobic residues, which probably fit in the respective pockets without optimizing the possible binding energy. In the Fig. 1 alignment, the four residues that we consider to be situated at the evolutionary equivalent positions (not necessarily the functionally equivalent positions) of the human CLIP P1, P4, P6, and P9 residues, are highlighted in yellow in all the CD74 molecules if deemed sufficiently reasonable for having the expected characteristics. We assume that the characteristics of asparagine (small and hydrophilic, but not charged) are sufficiently unpronounced for allowing to be an inserted residue in the P6 pocket without disturbing CLIP function. However, three of the aligned residues at the proposed ancestral CLIP P6 position are difficult to accept as agreeing with CLIP function, namely F96 of spotted gar CD74, and R and D at the equivalent of position 96 in the zebrafish and Mexican tetra CD74a exon 5 encoded CLIP regions. With blue shading, for some of the molecules aligned in Fig. 1, we have indicated potential alternative functional sets of CLIP P1, P4, P6, and P9 residues, which also, or even better, would agree with CLIP consensus features, including having a charged or hydrophilic residue at the P1 position. Even among human MHC class II molecules, the use of alternative (overlapping) CLIP P1–P9 regions for binding into the class II groove has been observed (Nguyen et al. 2017), thus a small shift in CLIP P1–P9 region usage between species, or, like in humans, the use of multiple overlapping CLIP P1–P9 regions within single species, may explain these motifs. It remains to be determined whether both the exons 3 and 5 encoded CLIP sequences of teleost fish CD74a can bind into MHC class II binding grooves, whether they are promiscuous or selective, and whether teleost CD74a molecules can bind one or two MHC class II molecules at the time.

The α-helical region involved in CD74 homo-trimerization

With the exceptions of teleost fish CD74b, and CD74a in Ostariophysi, the CD74 molecules appear to have three well-conserved α-helices with a set of characteristic tryptophans (the α-helices of human CD74 are underlined in Fig. 1). The symmetric structure formed by the α-helical regions of three interacting CD74 molecules was revealed by nuclear magnetic resonance (NMR) analysis by Jasanoff et al. (1998), who determined the structures formed by fragments of human CD74 from positions 118 to 192. In the structure determined by Jasanoff, the three protomers all have the same conformation, with a short α-helix A going upwards, then a long loop region, followed by a short α-helix B going upwards making a compact elbow shape with a closely connected long α-helix C pointing diagonally downwards. Of each protomer, the C-terminal region of α-helix C inserts like a pin into a socket formed by all domains of one other protomer and by the N-terminal region of α-helix A of a third protomer. We suggest the interested readers to access the PDB database accession 1IIE, because especially in regard to the importance of the conserved residues the structural organization of this region is difficult to explain in words or in few pictures. Good analyses of the sequence and predicted structural conservation of this region in CD74 from fish to mammals have been reported by Silva et al. (2007) and Criscitiello et al. (2012). In Fig. 1, conserved sequence characteristics in this region are highlighted with gray shading, except for the conserved tryptophans which, because of their pronounced character, are in indicated with purple shading. For several of the shaded residues, Yasanoff et al. (1998) reported involvement in intramolecular contacts between α-helices B and C, namely for L150, L153, M157, W162, and F165. In addition, Yasanoff et al. (1998) observed prolines in the AB strand, of which P140 is conserved between species, and they assumed that the prolines favored the extended structure of this loop region. For the shaded residues L129, L130, L150, L153, K154, M157, W162, F165, W168, W172, and F175, Yasanoff et al. observed participation in interprotomer contacts. The W168 residues in α-C of each of the three protomers interact tightly with each other at the center of the homo-trimeric complex. Participation in interprotomer contacts can also be observed for N149, E166, and R179, whereas the conservation of residue (S/T)167 can not readily be understood.

The C-terminal end of the CD74 exon 6 encoded sequence beyond the α-helical region is believed to be flexible/disordered (Yasanoff et al. 1998), and in nurse shark has an extension which is not shown in Fig. 1 (see Criscitiello et al. 2012). This region, which is poorly conserved among species, has no known function. We have no explanation for the absence of the α-helical region in CD74a of Ostariophysi and in CD74b throughout the investigated teleost fish.

The thyroglobulin type-1 domain

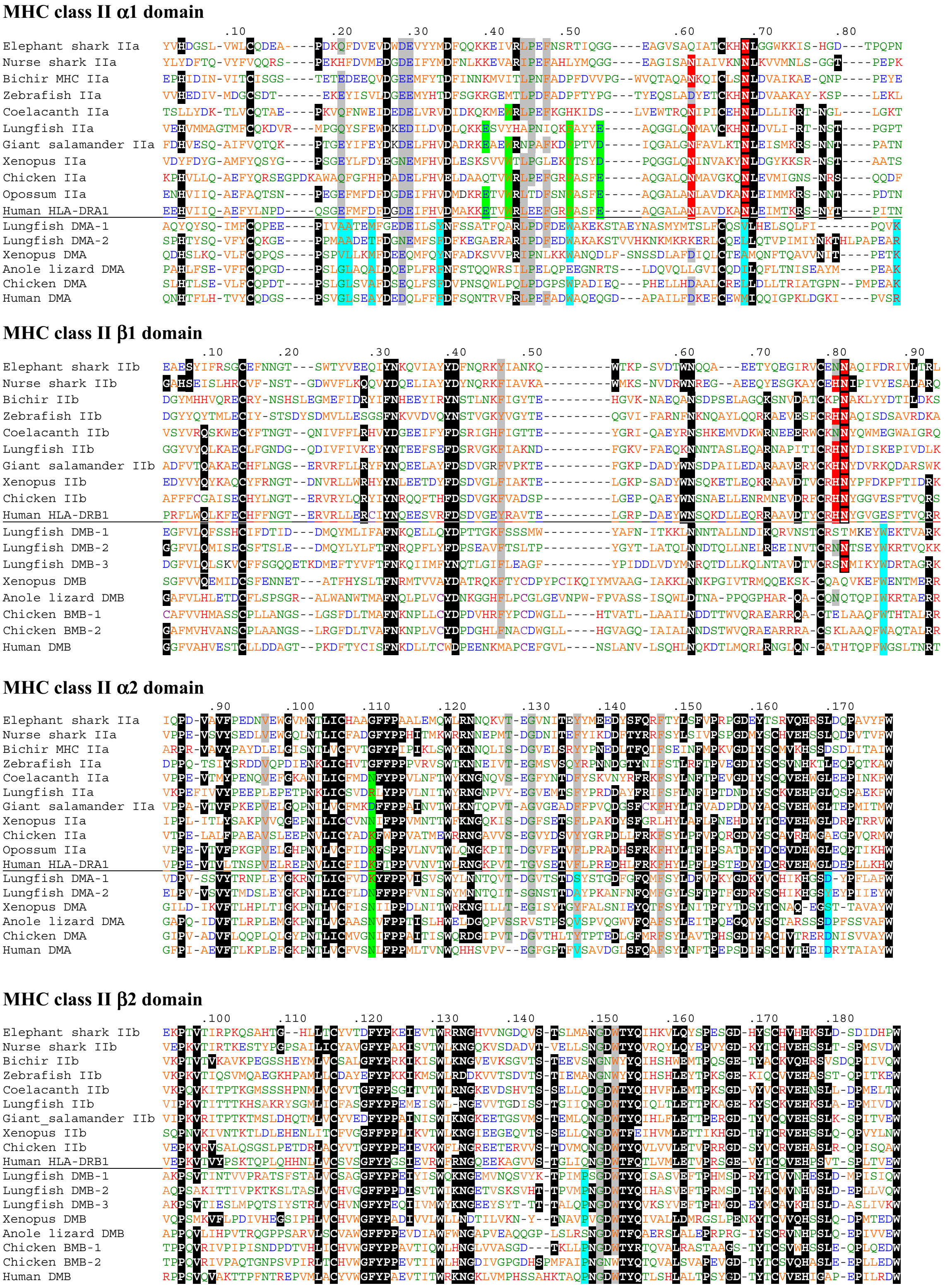
The large C-terminal domain of full-size CD74 is a thyroglobulin type-1 (TG) domain, which is a type of domain with highly conserved sequence which can be found in a number of molecules (Mihelic and Turk 2007). We found it difficult to distinguish sequence features in the CD74 TG domain which are unique for CD74 molecules and will not discuss the TG sequences in detail; for a more detailed sequence comparison of the CD74 TG domains between divergent vertebrates, we refer to Silva et al. (2007) and Criscitiello et al. (2012). The important function of the TG domain in the CD74 context appears to be that it can bind cathepsin L and inhibit its activity, whereas it is hardly efficient in blocking cathepsin S activity (Bevec et al. 1996; Fineschi et al. 1996; Guncar et al. 1999; Mihelic et al. 2008; Li et al. 2015). Furthermore, a chaperone function of CD74 TG domain for cathepsin L has been proposed (Lennon-Duménil et al. 2001). Throughout investigated species, there seem to be CD74 molecules with and without the TG domain (see above), and at least in mammals, both forms can support MHC class II function (Fineschi et al. 1995, 1996). However, the peptides which are presented by MHC class II are different depending on CD74 having a TG domain, which probably is caused by the modified proteolytic environment affecting both the speed of CD74 digestion (and the therewith associated MHC transport), and the available peptide pool (Fineschi et al. 1995, 1996; Bikoff et al. 1998; Fortin et al. 2013). Despite that some differences are known, the real importance of having both CD74 with and without TG domain should probably be considered as not well understood yet.

The carboxy-terminal region

The C-terminal region of CD74 beyond the TG domain differs largely between species (Fig. 1). This fragment is thought to be flexible/disordered in human CD74 (mentioned in Jasanoff et al. 1998), is virtually absent in some species (Fig. 1), and probably does not have a single conserved function.

Identification of DMA and DMB lineage sequences in lungfish

DM molecules are nonclassical MHC class II molecules which at least in mammals, and presumably in other tetrapod species, can not bind peptide antigens themselves (Mosyak et al. 1998; Chazara et al. 2011; Pos et al. 2012; Schulze and Wucherpfennig 2012; Dijkstra et al. 2013). In humans, HLA-DM facilitates peptide exchange on MHC class II proteins and is required for efficient peptide loading in vivo (Morris et al. 1994; Denzin and Cresswell 1995; Sloan et al. 1995; Kropshofer et al. 1996). How HLA-DM can do this was elucidated by the determination of the structure of HLA-DM bound to HLA-DR1 with an empty P1 pocket (Pos et al. 2012), and its comparison with structures of HLA-DM-free HLA-DR1 loaded with full-length peptide. Several good reviews discussed the pathways/mechanisms of how HLA-DM presumably is involved in peptide editing (Schulze and Wucherpfennig 2012; Blum et al. 2013; Pos and Wucherpfennig 2013; Mellins and Stern 2014). HLA-DM can only bind classical MHC class II efficiently when the binding groove of the latter is (partially) free of peptide. The major interactions between HLA-DM and HLA-DR1 involve the "P1-side" of the α1 domain of HLA-DR1, the side of the HLA-DM α1 domain where the C-terminus of the α1 helix region is situated, and the HLA-DMA α2 domain (Pos et al. 2012). In agreement with the evolutionary analysis of MHC class II molecules already done by Pos et al. (2012), we concluded that residues necessary for potentially allowing the important hydrogen bonding observed between HLA-DR1 α1 domain residue W43, and HLA DM α2 domain residue N125 (position 111 in Fig. 2), were present from the level of coelacanth and not possible in more primitive fishes (Dijkstra et al. 2013). The interaction between HLA-DR1 αW43 and HLA-DM αN125 pulls away the αW43 residue from its P1 pocket-supporting position, which disrupts the normal HLA-DR1 P1 pocket and induces the insertion of the HLA-DR1 residues αF51 and αE55 into the peptide binding groove. In this way, DM molecules compete with peptides for the conformation/occupancy of the classical MHC class II binding groove. So far, DM lineage genes were only reported for tetrapod species, including also chicken (Kaufman et al. 1999; Chazara et al. 2011) and frog (Ohta et al. 2006).



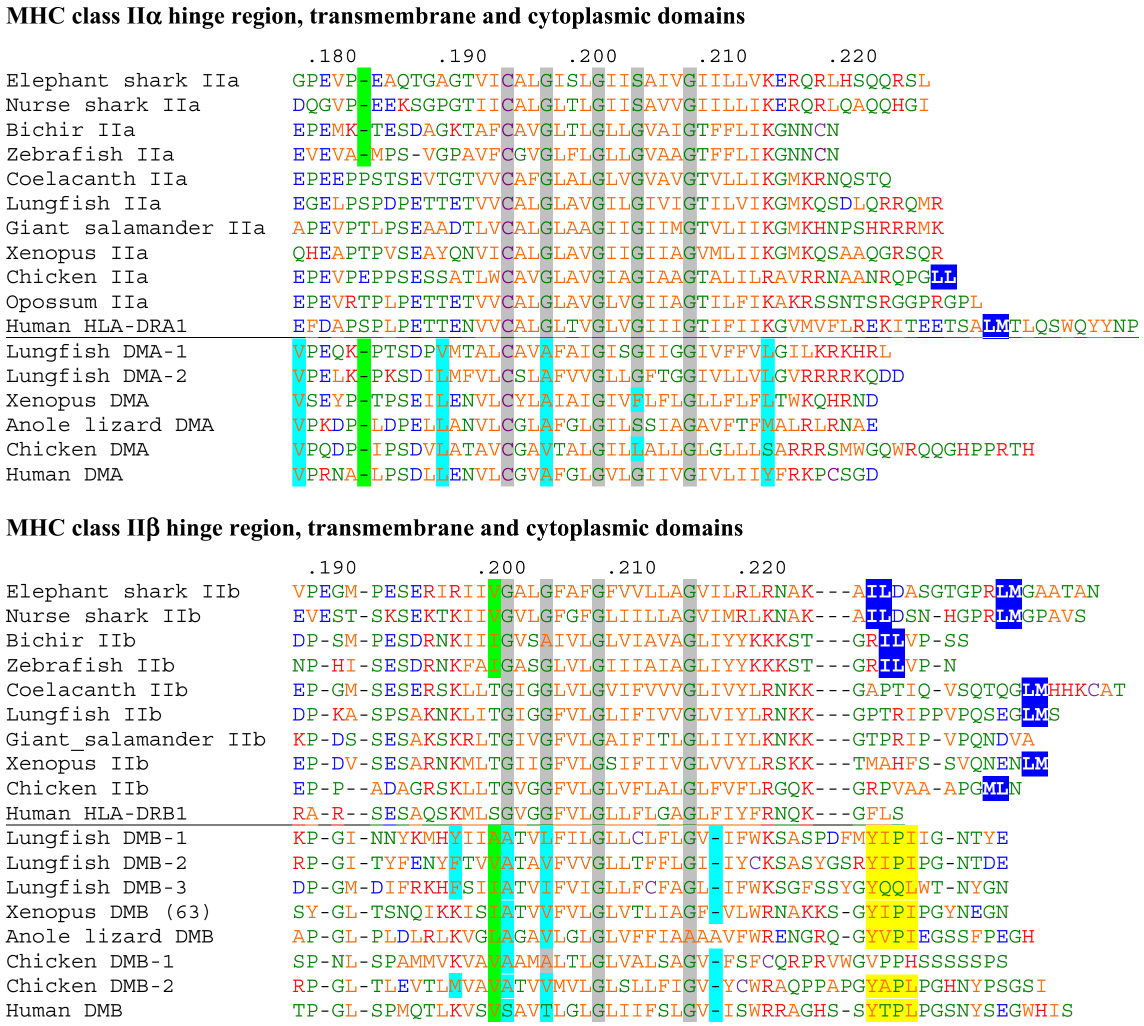
In the present study, we present several DMA and DMB lineage sequences from the South American lungfish (*Lepidosiren paradoxa*), which were retrieved as TSA cDNA sequences from the Transcriptome Shotgun Assembly Sequence Database (https://www.ncbi.nlm.nih.gov/genbank/tsa/). They are shown in the Fig. 2 alignment as Lungfish DMA-1, DMA-2, DMB-1, DMB-2, and DMB-3, together with other representative DM sequences, and with a number of classical MHC class II sequences including one alpha and one beta sequence of South American lungfish. For

◀ **Fig. 2** Alignment of deduced DM and classical MHC class II sequences of representative species. The α and β chain sequences are aligned per domain, separate from each other. The numbers above the alignment indicate the positions in the HLA-DRA1 or HLA-DRB1 sequence, respectively. The non-human classical MHC class II sequences are indicated as IIa or IIb. The GenBank accession numbers of the depicted MHC class IIα sequences are, from top to bottom in the alignment figure: Elephant shark (*Callorhinchus milii*), AFM88468; nurse shark (*Ginglymostoma cirratum*), AAA49311; Bichir (*Polypterus senegalus*), overlapping SRA reads from set SRX796491; zebrafish (*Danio rerio*) AAA72019; coelacanth (*Latimeria chalumnae*), XP_006014228; lungfish (*Lepidosiren paradoxa*) GEHZ01055957; giant salamander (*Andrias davidianus*), AGY55996; xenopus (*Xenopus laevis*), AF454374; chicken (*Gallus gallus*), AAR14673; opossum (*Monodelphis domesticus*) XP_007483702; human (*Homo sapiens*) HLA-DRA1, EAX03629; lungfish DMA-1, GEHZ01056980; lungfish DMA-2, GEHZ01029155; Xenopus DMA, AAH61681; Anole lizard (*Anolis carolinensis*) DMA, XP_008107441; chicken DMA, CAA18966; Human HLA-DMA, CAA54170. The GenBank accession numbers for the depicted MHC class IIβ sequences are, from top to bottom in the alignment figure: Elephant shark, JW875734; Nurse shark, L20274; Bichir, overlapping SRA reads from set SRX796491; Zebrafish, AAA50043; Coelacanth, XP_006010591; Lungfish, GEHZ01012825; Xenopus, BAA08759; Giant salamander, AGY56015; chicken, AAA48948; human HLA-DRB1, CAM34749; Lungfish DMB-1, GEHZ01000527; Lungfish DMB-2, GEHZ01067647; Lungfish DMB-3, GEHZ01038869; Xenopus DMB, ABB8533657; Anole lizard DMB, XP_008122576; chicken BMB1, BAG69311; chicken BMB2, CAA18967; and human HLA-DMB, AAB60387. Sequences were aligned by hand, based on considerations of similarity and phylogeny, and expected structural features. Not for all regions the alignments can be deemed reliable, but we do feel certain that the alignment reliably supports the conclusions made in the main text. Motifs indicated by color shading are explained in the main text; for the alignment of the transmembrane and cytoplasmic domains, it probably can be discussed whether more residues should be colored gray or black. Non-shaded residues were colored as explained in the Fig. 1 legend

several of the aligned classical MHC class II sequences additional information on polymorphism or function is not available, and it might be better to describe them as "classical type," but for convenience, we simply use "classical" throughout this manuscript. The leader sequences and ectodomain N-terminal extensions from MHC consensus (which in mammalian DMα are long, see the alignment figure in Chazara et al. 2011), are not included in the Fig. 2 alignment. The DMA lineage identity of lungfish DMA-1 and DMA-2 was readily confirmed by reciprocal top-matches upon database similarity searches and phylogenetic tree analyses (not shown) and by presence or absence of characteristic motifs throughout the sequence (Fig. 2). Upon phylogenetic tree analyses, the lungfish DMB sequences clustered only weakly (without good bootstrap values) with the other DMB sequences (not shown), but convincing evidence for DMB lineage identity of lungfish DMB-1, DMB-2, and DMB-3, comes from motifs in their transmembrane (TM) and cytoplasmic tail (CY) domains (Fig. 2). In addition, in the β1 and β2 domain sequences, there are also a few residues which support DMB lineage identity of the lungfish sequences (Fig. 2). The fact that the important function of the DM α chain is to interact with classical MHC class II, may explain the higher number of conserved DM characteristic residues in the α1 and α2 domain sequences than found in the β1 and β2 domains. Meanwhile, the important function of the DM β chain is to deliver the DM molecules to the endosomal compartments by using a *YXXΦ* motif (*X* can be any residue, Φ denotes a hydrophobic residue) (Lindstedt et al. 1995; Kozik et al. 2010), which is probably why the DMB lineage signature is most pronounced in that region.

In Fig. 2, motifs are highlighted. The black shading is used for residues that we assume to be characteristic for, and that were possibly inherited from, evolutionary stages in which the MHC class IIA or IIB lineages had not separated yet from their class I counterparts (see also Dijkstra et al. 2018). The gray shading is used for residues that we deem characteristic for the MHC class IIA or IIB lineages. The black shaded residues, and the gray shaded residues with the exception of the TM domain motifs, will not be discussed in the present paper. Red shading is used for the αN62, αN69, βH81, and βN82 residues, which are well-conserved among classical MHC class II sequences, and whose sidechains participate in an important hydrogen bond network with the backbone of the peptide ligand (Painter and Stern 2012); of these four residues, the βN82 residue appears to be most important for peptide binding (Zhou et al. 2009). For the level of conservation of these four residues among classical MHC class II sequences throughout divergent species, see Table S3 in Dijkstra et al. (2013). In Fig. 2, light blue shading is used to highlight residues characteristic for the DMA or DMB lineages, green shading is used to highlight other notable patterns, and the yellow and blue shading in the cytoplasmic tail sequences highlight potential endocytic motifs. Below, we discuss a few of those motifs, while in this study, some light blue motifs are only highlighted as phylogenetic markers and not further discussed. Residue numbering in the text below agrees with the numbering used in the Fig. 2 alignment, which is based on the positions in mature HLA-DR1 chains. In some cases, the numbers are followed by numbers between square brackets, and those are the HLA-DMα or HLA-DMβ residue numbers as used by Pos et al. (2012) so that readers can conveniently compare between studies.

As discussed before (Dijkstra et al. 2013), the αW43 residue and residues with hydrogen bonding potential at position α111 [α125] can be found in classical MHC class IIα sequences from the level of Sarcopterygii (lobe-finned fish plus tetrapods). Already in MHC class IIα of coelacanth, a fish belonging to Actinistia which separated from the other Sarcopterygii around 409 million years ago (MYA) (Broughton et al. 2013), these residues are found (Fig. 2 and Dijkstra et al. 2013). We speculate that the αY43 residue found in the depicted lungfish classical MHC class II



MHC class IIα hinge region, transmembrane and cytoplasmic domains

```
                       .180      .190      .200      .210      .220
Elephant shark IIa    GPEVP-EAQTGAGTVICALGISLGIISAIVGIILLVKERQRLHSQQRSL
Nurse shark IIa       DQGVP-EEKSGPGTIICALGLTLGIISAVVGIILLIKERQRLQAQQHGI
Bichir IIa            EPEMK-TESDAGKTAFCAVGLTLGLLGVAIGTFFLIKGNNCN
Zebrafish IIa         EVEVA-MPS-VGPAVFCGVGLFLGLLGVAAGTFFLIKGNNCN
Coelacanth IIa        EPEEPPSTSEVTGTVVCAFGLALGLVGVAVGTVLLIKGMKRNQSTQ
Lungfish IIa          EGELPSPDPETTETVVCALGLAVGILGIVIGTILVIKGMKQSDLQRRQMR
Giant salamander IIa  APEVPTLPSEAADTLVCALGLAAGIIGIIMGTVLIIKGMKHNPSHRRRMK
Xenopus IIa           QHEAPTPVSEAYQNVICALGLAVGIIGIIAGVMLIIKGMKQSAAQGRSQR
Chicken IIa           EPEVPEPPSESSATLWCAVGLAVGIAGIAAGTALILRAVRRNAANRQPGLL
Opossum IIa           EPEVRTPLPETTETVVCALGLAVGLVGIIAGTILFIKAKRSSNTSRGGPRGPL
Human HLA-DRA1        EFDAPSPLPETTENVVCALGLTVGLVGIIIGTIFIIKGVMVFLREKITEETSALMTLQSWQYYNP
Lungfish DMA-1        VPEQK-PTSDPVMTALCAVAFAIGISGIIGGIVFFVLGILKRKHRL
Lungfish DMA-2        VPELK-PKSDILMFVLCSLAFVVGLLGFTGGIVLLVLGVRRRRKQDD
Xenopus DMA           VSEYP-TPSEILENVLCYLAIAIGIVFLFLGLLFLFLTWKQHRND
Anole lizard DMA      VPKDP-LDPELLANVLCGLAFGLGILSSIAGAVFTFMALRLRNAE
Chicken DMA           VPQDP-IPSDVLATAVCGAVTALGILLALLGLGLLLSARRRSMWGQWRQQGHPPRTH
Human DMA             VPRNA-LPSDLLENVLCGVAFGLGVLGIIVGIVLIIYFRKPCSGD
```

MHC class IIβ hinge region, transmembrane and cytoplasmic domains

```
                       .190         .200      .210      .220
Elephant shark IIb    VPEGM-PESERIRIIVGALGFAFGFVVLLAGVILRLRNAK---AILDASGTGPRLMGAATAN
Nurse shark IIb       EVEST-SKSEKTKIIVGVLGFGFGLIILLAGVIMRLKNAK---AILDSN-HGPRLMGPAVS
Bichir IIb            DP-SM-PESDRNKIIIGVSAIVLGLVIAVAGLIYYKKKST---GRILVP-SS
Zebrafish IIb         NP-HI-SESDRNKFAIGASGLVLGIIIAIAGLIYYKKKST---GRILVP-N
Coelacanth IIb        EP-GM-SESERSKLLTGIGGLVLGVIFVVVGLIVYLRNKK---GAPTIQ-VSQTQGLMHHKCAT
Lungfish IIb          DP-KA-SPSAKNKLITGIGGFVLGLIFIVVGLVIYLRNKK---GPTRIPPVPQSEGLMS
Giant_salamander IIb  KP-DS-SESAKSKRLTGIVGFVLGAIFITLGLIIYLKNKK---GTPRIP-VPQNDVA
Xenopus IIb           EP-DV-SESARNKMLTGIIGFVLGSIFIIVGLVVYLRSKK---TMAHFS-SVQNENLM
Chicken IIb          EP-P--ADAGRSKLLTGVGGFVLGLVFLALGLFVFLRGQK---GRPVAA-APGMLN
Human HLA-DRB1        RA-R--SESAQSKMLSGVGGFVLGLLFLGAGLFIYFRNQK---GFLS
Lungfish DMB-1        KP-GI-NNYKMHYIIAATVLFILGLLCLFLGV-IFWKSASPDFMYIPIIG-NTYE
Lungfish DMB-2        RP-GI-TYFENYFTVVATAVFVVGLLTFFLGI-IYCKSASYGSRYIPIPG-NTDE
Lungfish DMB-3        DP-GM-DIFRKHFSIIATVIFVIGLLFCFAGL-IFWKSGFSSYGYQQLWT-NYGN
Xenopus DMB (63)      SY-GL-TSNQIKKISIATVVFVLGLVTLIAGF-VLWRNAKKS-GYIPIPGYNEGN
Anole lizard DMB      AP-GL-PLDLRLKVGLAGAVLGLGLVFFIAAAVFWRENGRQ-GYVPIEGSSFPEGH
Chicken DMB-1         SP-NL-SPAMMVKVAVAAMALTLGLVALSAGV-FSFCQRPRVWGVPPHSSSSSPS
Chicken DMB-2         RP-GL-TLEVTLMVAVATVVMVLGLSLLFIGV-YCWRAQPPAPGYAPLPGHNYPSGSI
Human DMB             TP-GL-SPMQTLKVSVSAVTLGLGLIIFSLGV-ISWRRAGHS-SYTPLPGSNYSEGWHIS
```

Fig. 2 continued.

sequence can function in DM-mediated peptide editing in a manner similar to that as known for αW43 residues.

The classical MHC class II residues α(F/Y)51 and α(D/E)55, which according to Pos et al. (2012) form an important part of the DM-mediated peptide editing system by inserting into the groove of peptide-free DM-bound classical molecules, are only found as pair from the level of Rhipidistia (Dipnomorpha plus tetrapods). The residue pair can be found in lungfish (belonging to Dipnomorpha) as well as in tetrapods, two groups of species which separated around 375 MYA (Broughton et al. 2013). However, in some classical coelacanth MHC class IIα sequences, a reminiscent αL51 + αE55 motif can be found (e.g., GenBank accession XP_006014287, not shown here).

Judging from HLA-DM structures (PDB accessions 1HDM and 4FQX; not shown here), the DM-characteristic residues αW51 and βW87 seem to stabilize that part of the combined DM α1 + β1 domains where in peptide-free classical MHC class II molecules the unstable P1 pocket region is located.

Both lungfish DMA-1 and DMA-2 lack both αN62 and αN69, arguing against their involvement in classical type peptide binding function. As found in DMα sequences of other species, in lungfish DMA-1 and DMA-2 the αN69 residue is replaced by a hydrophobic residue.

The lungfish DMB-1 sequence lacks both βH81 and βN82, thus cannot bind peptide ligand in the way done by classical DMβ molecules. However, the lungfish DMB-2 and DMB-3 sequences have the important βN82 residue (Zhou et al. 2009), whereas DMB-2 residue βN81 instead of βH81 is also found in some classical sequences (Dijkstra et al. 2013). Therefore, also depending on with which α chain the different lungfish DMβ molecules form a complex, some may and some may not bind peptide ligands. As common among nonclassical MHC molecules (Dijkstra et al. 2018), at some point in evolution, DM lineage molecules presumably derived from classical molecules and at that moment of origin, they presumably retained classical features. So, it may very well be that in primitive lungfish, the DM lineage molecules are not only used for the DM function known in tetrapods, but that they are used for several functions including ones that are more reminiscent of classical functions. Future research will have to clarify that.

Pos et al. (2012) concluded that HLA-DM residue αR84 [αR98] is among the few key amino acids important for DR-DM interaction. However, in the DR-DM complex, the residue makes a salt bridge with HLA-DR αE40, and this glutamic acid is not perfectly conserved among classical IIα sequences (Fig. 2 and Pos et al. 2012). Nevertheless, it is notable that at position 84, the DMα sequences, including those of lungfish, have an arginine or lysine, whereas such residue is unusual in classical IIα sequences, and that lungfish classical IIα does have αE40 (Fig. 2).

DM molecules are transported to endosomal compartments due to their *YXXΦ* motif in the DMβ cytoplasmic tail (Lindstedt et al. 1995), which is highlighted with yellow shading in Fig. 2. The motif is conserved among the aligned DMβ sequences, except in chicken DMB-1 (Fig. 2). The function of the two chicken DMβ molecules is not known. In Fig. 2, dileucine(-similar) motifs in the cytoplasmic tail are highlighted with dark blue shading, but it is questionable whether their context is sufficiently favorable for being efficient endocytic targeting motifs (Pond et al. 1995; Brunet et al. 2000; Kozik et al. 2010), especially since classical MHC class II complexes are known to depend on CD74 for efficient transport (see above).

At positions α184 and β201, the DM sequences show interesting similarities with classical sequences of cartilaginous and ray-finned fishes (Fig. 2). Because the DM α111 [α125] residues agree with residues found at that position in classical class IIα of Sarcoptherygii, we speculate that the DMA and DMB lineages separated from the sarcoptherygian classical IIA and IIB lineages around the time of origin of Sarcoptherygii. Therefore, DM lineage genes in coelacanth would not be unexpected, but we have not found them so far.

The βK198 residue of HLA-DR is believed to be the C-terminal residue of the ectodomain, and there are data supporting its importance for binding between α and β chains (Cosson and Bonifacino 1992), as well as for recognition by membrane-associated RING-CH 8 (MARCH8) (Jahnke et al. 2013). Conservation of a basic residue at the β198 position appears to be less stringent among DMβ sequences than among classical IIβ sequences (Fig. 2), which, in regard of the possible MARCH8 function, would be consistent with different demands on intracellular transport.

Typically, in classical IIα and IIβ sequences at the matching positions α215 and β222, a lysine, arginine or glutamine is found, which in many cases is the start of the cytoplasmic tail (Fig. 2; Harton et al. 2016). For unknown reason, in DMα sequences, a hydrophobic residue tends to be found at this position, while in most DMβ sequences there is a one amino acid deletion between the TM core domain and the conserved basic residue at β222.

There is evidence for the classical molecules that the αC195 residue, which is highly conserved in both classical IIα and in DMα sequences, can be palmitoylated, and so helps to target the molecules to lipid rafts (Kaufman et al. 1984b; Komaniwa et al. 2009; Harton et al. 2016). It should also be noted that the cysteine is part of a set of small residues which are expected to have their sidechains at the same side of the TM α-helix (Travers et al. 1984; Cosson and Bonifacino 1992; gray shaded residues in Fig. 2).

Gray shading in the hydrophobic transmembrane regions of classical MHC class II in Fig. 2 is used to show conserved sets of small residues being either glycine, alanine, serine, or cysteine (G/A/S/C), approximately positioned one or more complete turns apart at the same side of the respective TM domain α-helix. In the Fig. 2 alignment, although for the ectodomains we did not try to properly align the IIα and IIβ sequences with each other, the depicted TM domain sequences of both α and β sequences can be considered as a single alignment. The alignment shows that the gray-shaded residues are at similar locations in the α and β sequences, but that at the position matching α205 in the β sequences, the G/A/S/C residues are lacking (Fig. 2). The importance of the motifs for helix packing and binding between the α and β chains, and the fact that the β chains have one fewer G/A/S/C position, are well established; however, there is still discussion possible on which of the motif residues are important and how they interact (Travers et al. 1984; Cosson and Bonifacino 1992; Moore et al. 2008; King and Dixon 2010; Dixon et al. 2014; Drake and Drake 2016; Harton et al. 2016). The only thing we want to add to the discussion is the observation that the motifs in the DMα and DMβ TM and bordering regions are different from those in the classical molecules, something which has so far only be realized partially (Harton et al. 2016). At position α198, in the DMα sequences an alanine or valine is found, whereas in the classical IIα sequences, a glycine is common (Fig. 2). Furthermore, in comparison to classical IIα, among DMα sequences, there seems to be a reduced evolutionary pressure to maintain a G/A/S/C residue at position α205. In classical IIβ sequences, at position β202, a glycine is common, but in the analyzed DMβ sequences, a serine or alanine is found (Fig. 2). Finally, whereas at position β205 of classical IIβ sequences, G/A/S/C residues are found, most of the analyzed DMβ sequences have larger residues like leucine, isoleucine, valine, and threonine. For classical MHC class II molecules, it has been proposed that the important TM units for interaction are the glycines of GxxxG motifs, and that the glycines of a single β chain GxxxG motif, namely βG205-to-βG209, can alternatively interact with the glycines of α chain TM motif αG198-to-αG202 or motif αG205-to-αG209, and that the resulting shift between TM regions causes functionally relevant conformational changes in the ectodomains (Dixon et al. 2014; Drake and Drake 2016). If DM molecules would not need similar options, which may partially explain the change in TM domain motifs compared to classical molecules, but that is speculation only.

For a discussion of how MHC class II systems in primitive vertebrates might work in the absence of DM, we refer to our previous study (Dijkstra et al. 2013). We speculate that the establishment of DM function may have involved an optimization of a stabilizing, and therefore potentially peptide-editing, effect that even (some) classical MHC class II molecules might have on each other (for examples of binding affinity among classical MHC class II molecules, see Schafer et al. 1995).

A hypothesis for the origin of MHC molecules

As explained in the introduction, we believe, like several other researchers, that the first MHC molecule was a homodimer MHC class II-like structure. It is more difficult to deduce from what type of non-MHC molecule the MHC molecules originated. The membrane-proximal IgSF domains are of the C1 category (Halaby et al. 1999), which are only found in a few molecule families. Well-known molecules with IgSF C1 category domains are antibodies and TCRs. In contrast to their IgSF domains, the MHC membrane-distal domains ($\alpha1$ and $\beta1$ in the case of MHC class II) have unique structures (Bjorkman et al. 1987; Brown et al. 1993) for which ancestral candidates are quite speculative. Nevertheless, based on instinct but also on analysis of the sequences, we believe, like several other researchers (explained in Ohnishi 1984; Hood et al. 1985; Hashimoto et al. 1990; Hashimoto 2003), that most likely, the molecule ancestral to the MHC molecules was a transmembrane molecule with a membrane-distal domain of the IgSF variable (V) category and at least one membrane-proximal domain of the IgSF C1 category. In the present study, we will not discuss this matter at the sequence level, but in Fig. 3, we present (as far as we know) a new model that aims to explain, based on functional arguments, how MHC molecules may have derived from antibodies.

In Fig. 3, from steps 1 to 7, we propose how, in the evolution of jawed vertebrates, first antibodies, then TCRs, and then MHC molecules may have arisen. We assume that antibodies arose first (Fig. 3, step 1), because they basically can function independently of TCR and MHC molecules, and some antibodies can function in homomeric form. We assume that the established pool of antibody genes contained both single exons per IgSF *V* domain, as well as genes using somatic recombination of V, D, and J (-like) fragments. Such situation has been described for the horned shark *Heterodontus francisci*, where both types of genes possibly are used together in "cluster type" organization (Kokubu et al. 1988; Litman et al. 1993). The genomic situation in which a single continuous germline sequence encodes a complete antibody IgSF *V* domain has been coined "VDJ germline joined" (Litman et al. 1993). In the evolutionary stage proposed in Fig. 3, step 1, all antibodies function as part of B cell receptors (BCRs) and are secreted upon proper stimulation by antigen and co-factors.

In the proposed stage of Fig. 3, step 2, some "B cell" types stopped releasing their "BCRs," and are specialized in cytokine releasing function upon antigen recognition. We therefore call the cells "T cells" instead, using quotation marks to express that at this stage, there is no full-blown T cell system yet. The genes used for these membrane-fixed BCRs were duplications of the antibody type genes with genomically separated coding fragments for variable domains *V*, *D*, and *J* regions, and evolved to become TCR genes. Compared with the original Fig. 3, step 1 situation, the adaptive immune system in Fig. 3, step 2 improved by allowing a "second opinion" analysis concerning the antigen bound by the B cell. Second opinion analysis is exactly where helper plus regulatory T cells are all about when interacting with B cells. They may either stimulate or downregulate potential B cell responses. The close relatedness of antibodies and TCRs is well recognized (e.g., Litman et al. 2010; Flajnik and Kasahara 2010), although discussion is possible on whether BCRs or TCRs were first because of the existence of TCR molecules of the TCR$\gamma\delta$ category which seem to function largely independent of MHC molecules.

In the evolution stage depicted in Fig. 3, step 3, a further improvement of the system was introduced. Namely, rather than keeping the antigen bound on the surface, the BCR now first recycles through the endosomal compartments. There, proteases expose interior parts of the bound antigen, causing that, when the complex returns to the cell surface, instead of intact antigen, now partially digested antigen is presented. This increases the number of epitopes of an antigen, which the B and T cells together can examine. Basically, that is the same principle as known for extant functions and mechanisms of B and T helper/reg cells (Roche and Furuta 2015; Adler et al. 2017), except that at this proposed evolutionary stage, the antigens were not released from the BCR, not cut into very small pieces, and not reloaded onto MHC molecules.

The proposed evolutionary stage depicted in Fig. 3, step 4 is similar to the Fig. 3, step 3 stage, except that the antigen is cut into smaller pieces. How rapid and with which intermediate stages, the system most likely changed from the Fig. 3, step 2 to the Fig. 3, step 4 situation, is up for debate, and has to consider whether complexes are favorable/possible in regard to energy terms. In the proposed Fig. 3, step 4 situation, the TCRs of the T cell recognize the combination of antigen plus presenting BCR, which is quite reminiscent of the function of extant TCR$\alpha\beta$ T cells which recognize antigen peptides together with the presenting MHC molecules.

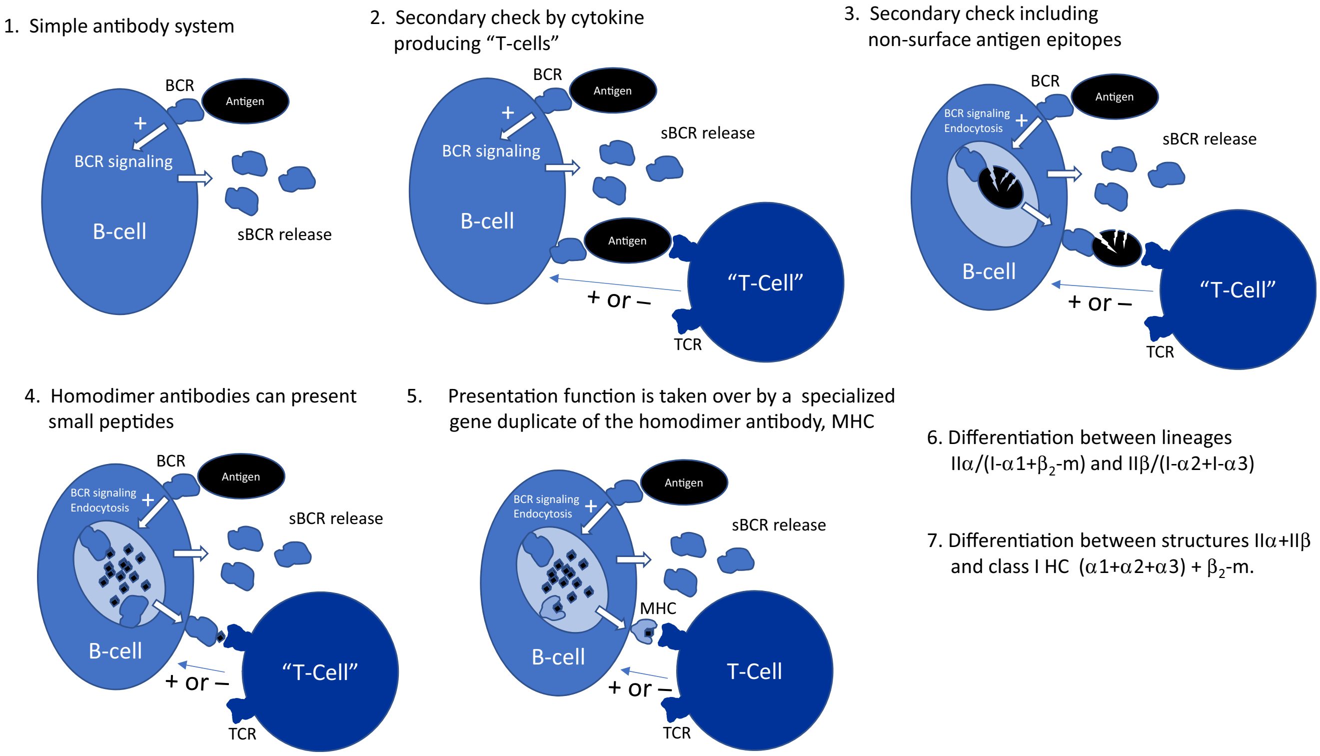


Fig. 3 This model schematically shows in stages 1 to 7, how we envision that MHC class II evolved from an antibody type molecule. Detailed explanations are given in the main text. Antigen, protein antigen; BCR, B cell receptor; sBCR, secreted B cell receptor (antibody); T cell, cell with T cell like function; TCR, T cell receptor; HC, heavy chain. The drawings in step 3 versus steps 4 and 5 aim to show that in step 3, in endosomal compartments, the BCR-bound antigen is only partially digested before the BCR-antigen complex recycles to the surface, whereas in steps 4 and 5, the antigen is digested into small peptides

In Fig. 3, step 5, finally, a system emerges which is essentially identical to extant B-T cell interactions. After endocytosis, the BCR releases its antigen, which is cut into small peptides, which are now loaded onto a specialized molecule encoded by a gene duplication of the BCR gene; this specialized molecule eventually evolves to an MHC molecule. In order for such gene duplication to possibly generate the MHC ancestral gene, it is necessary that the BCR gene which was duplicated, encoded a "heavy-chain-only" type antibody, as is known in some species (Flajnik et al. 2011) and that the V domain coding sequence was "VDJ germline joined" (see above). Obviously, those two preconditions are within the range of known antibody situations.

To Fig. 3, in order to be complete, we also added the subsequent steps 6 and 7. In the Fig. 3, step 5 stage, the ancestral BCR homodimer became an MHC homodimer with membrane-distal domains which lost many of their recognizable IgSF *V* characteristics. In the Fig. 3, step 6 stage, the MHC homodimer changed into a heterodimer, with the IIα/(I-α1 + β2-m) and IIβ/(I-α2 + I-α3) lineages each acquiring their own characteristic residues and IgSF domain orientations (Bjorkman et al. 1987; Brown et al. 1993; Stern and Wiley 1994). At the stage of Fig. 3, step 7, after an exon shuffling event giving rise to MHC class I organized genes, the MHC class I and II molecules started to diverge from each other. Possibly, MHC-restricted cell mediated cytotoxicity emerged in evolution after the acquisition of MHC class I molecules, but that is speculation only.

Other theories, which we deem less attractive, have also been proposed for the origin of MHC molecules (reviewed in Hashimoto 2003; Kaufman 2018). The critical absence in all the models, including the one we present here, is a convincing explanation for how the unique structures of the MHC peptide binding domains were generated. A functional model of evolution, as presented here, should be of great help for efficiently searching evolutionary MHC precursor forms that still may be present in some species. Future findings of molecules with structural features, or functions, intermediate to those of antibodies and MHC molecules, would support our model.

Conclusion

We present a new model for how, in evolution, MHC molecules may have originated. MHC class II molecules function by presenting digested parts of exogenous antigens, taken up by professional antigen presenting cells, to helper and regulatory T cells for a second opinion. We assume that TCRs originated from BCRs, and that a primitive communication system "discussing" endocytosed BCR-bound antigens, between BCR expressing B cells and TCR expressing T cells, formed the ancestral situation from which the MHC presenting system emerged. We assume that the MHC class II molecules derived from a "VDJ germline joined" heavy-chain-only BCR. The advantages of our model are the functional simplicity, some support from sequence comparisons (Hashimoto 2003), and consistency with the facts that the first MHC molecules presumably were homodimer MHC class II like molecules, and that extant B cells efficiently express both BCR and MHC class II. We hope that future research can provide more evidence for our model.

Relatively early in MHC evolution, known because the molecule is found in all species with MHC class II molecules, the MHC class II molecules acquired a specialized chaperone, CD74 (MHC class II invariant chain). The CD74 alignment presented in this study shows that CLIP motifs for binding into the MHC class II peptide binding groove can be found in CD74 from cartilaginous fish to mammals. Previous studies had already noted that the deduced zebrafish CD74 molecules lack the CD74-characteristic α-helical region for homotrimerization. However, the present study is the first to recognize that, while they have lost exons for those α-helical regions, the *CD74A* genes in Ostariophysi (a superorder of fish including, for example, zebrafish and Mexican tetra) acquired a duplication of the CLIP core coding exon, and the resulting CD74a molecules might bind two MHC class II molecules. We hope that the CD74 alignment figure may also draw attention to highly conserved residues, such as the leucines or isoleucines at positions 62, 65, and 72 (Fig. 1), for which functions are not known.

Much later in evolution, namely within Sarcoptherygii (lobe-finned fish plus tetrapods), a newly established nonclassical MHC class II molecule, DM, acquired a function for assisting classical MHC class II to bind high-affinity peptides. In the present study, we found DMA and DMB lineage genes in lungfish. Analysis of sequence motifs in the lungfish DMα and DMβ sequences, and also in the lungfish classical MHC class II sequences, indicate that DM function as known in mammals may also exist in lungfish.

As with the CD74 alignment, our MHC class II alignment figure may draw attention to highly conserved residue patterns which otherwise tend to be ignored. Especially fascinating are the differences in the motifs in the transmembrane plus cytoplasmic regions between classical MHC class II and DM lineage molecules. That does not only concern the readily observable tyrosine-based endocytic motif in DMβ cytoplasmic sequences, but also a number of subtle differences in conservation of, for example, glycine-rich motifs in the transmembrane domains (Fig. 2). Looking at the distinct conservations of these motifs, it can only be concluded that we do not yet properly understand the function of the classical MHC class II and DM transmembrane domains. We hope that this article will give the readers more insights into the deep evolution of the MHC class II system.

References

Adams EJ, Luoma AM (2013) The adaptable major histocompatibility complex (MHC) fold: structure and function of nonclassical and MHC class I-like molecules. Annu Rev Immunol 31:529–561

Adler LN, Jiang W, Bhamidipati K, Millican M, Macaubas C, Hung SC, Mellins ED (2017) The other function: class II-restricted antigen presentation by B cells. Front Immunol 8:319

Ashman JB, Miller J (1999) A role for the transmembrane domain in the trimerization of the MHC class II-associated invariant chain. J Immunol 163:2704–2712

Bevec T, Stoka V, Pungercic G, Dolenc I, Turk V (1996) Major histocompatibility complex class II-associated p41 invariant chain fragment is a strong inhibitor of lysosomal cathepsin L. J Exp Med 183:1331–1338

Bikoff EK, Huang LY, Episkopou V, van Meerwijk J, Germain RN, Robertson EJ (1993) Defective major histocompatibility complex class II assembly, transport, peptide acquisition, and CD4+ T cell selection in mice lacking invariant chain expression. J Exp Med 177(6):1699–1712

Bikoff EK, Kenty G, Van Kaer L (1998) Distinct peptide loading pathways for MHC class II molecules associated with alternative II chain isoforms. J Immunol 160:3101–3110

Bird PI, Trapani JA, Villadangos JA (2009) Endolysosomal proteases and their inhibitors in immunity. Nat Rev Immunol 9:871–882

Bjorkman PJ, Saper MA, Samraoui B, Bennett WS, Strominger JL, Wiley DC (1987) Structure of the human class I histocompatibility antigen, HLA-A2. Nature 329:506–512

Blum JS, Wearsch PA, Cresswell P (2013) Pathways of antigen processing. Annu Rev Immunol 31:443–473

Bremnes B, Madsen T, Gedde-Dahl M, Bakke O (1994) An LI and ML motif in the cytoplasmic tail of the MHC-associated invariant chain mediate rapid internalization. J Cell Sci 107:2021–2032

Bremnes B, Rode M, Gedde-Dahl M, Nordeng TW, Jacobsen J, Ness SA, Bakke O (2000) The MHC class II-associated chicken invariant chain shares functional properties with its mammalian homologs. Exp Cell Res 259:360–369

Broughton RE, Betancur-R R, Li C, Arratia G, Ortí G (2013) Multi-locus phylogenetic analysis reveals the pattern and tempo of bony fish evolution. PLOS Curr Tree Life Ed 1. https://doi.org/10.1371/currents.tol.2ca8041495ffafd0c92756e75247483e

Brown JH, Jardetzky TS, Gorga JC, Stern LJ, Urban RG, Strominger JL, Wiley DC (1993) Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. Nature 364:33–39

Brunet A, Samaan A, Deshaies F, Kindt TJ, Thibodeau J (2000) Functional characterization of a lysosomal sorting motif in the cytoplasmic tail of HLA-DObeta. J Biol Chem 275(47):37062–37071

Busch R, Cloutier I, Sékaly RP, Hämmerling GJ (1996) Invariant chain protects class II histocompatibility antigens from binding intact polypeptides in the endoplasmic reticulum. EMBO J 15:418–428

Castellino F, Germain RN (2006) Cooperation between CD4+ and CD8+ T cells: when, where, and how. Annu Rev Immunol 24:519–540

Chazara O, Tixier-Boichard M, Morin V, Zoorob R, Bed'hom B (2011) Organisation and diversity of the class II DM region of the chicken MHC. Mol Immunol 48:1263–1271

Chen FF, Lin HB, Li JC, Wang Y, Li J, Zhang DG, Yu WY (2017) Grass carp (Ctenopharyngodon idellus) invariant chain of the MHC class II chaperone protein associates with the class I molecule. Fish Shellfish Immunol 63:1–8

Cloutier M, Gauthier C, Fortin JS, Thibodeau J (2014) The invariant chain p35 isoform promotes formation of nonameric complexes with MHC II molecules. Immunol Cell Biol 92:553–556

Colbert JD, Matthews SP, Miller G, Watts C (2009) Diverse regulatory roles for lysosomal proteases in the immune response. Eur J Immunol 39:2955–2965

Cosson P, Bonifacino JS (1992) Role of transmembrane domain interactions in the assembly of class II MHC molecules. Science 258:659–662

Cresswell P, Roche PA (2014 Jul) Invariant chain-MHC class II complexes: always odd and never invariant. Immunol Cell Biol 92(6): 471–472

Criscitiello MF, Ohta Y, Graham MD, Eubanks JO, Chen PL, Flajnik MF (2012) Shark class II invariant chain reveals ancient conserved relationships with cathepsins and MHC class II. Dev Comp Immunol 36:521–533

Denzin LK (2013) Inhibition of HLA-DM mediated MHC class II peptide loading by HLA-DO promotes self tolerance. Front Immunol 4: 465

Denzin LK, Cresswell P (1995) HLA-DM induces CLIP dissociation from MHC class II alpha beta dimers and facilitates peptide loading. Cell 82:155–165

Dijkstra JM, Grimholt U (2018) Major histocompatibility complex (MHC) fragment numbers alone—in Atlantic cod and in general—do not represent functional variability. F1000Res 7:963

Dijkstra JM, Kiryu I, Köllner B, Yoshiura Y, Ototake M (2003) MHC class II invariant chain homologues in rainbow trout (Oncorhynchus mykiss). Fish Shellfish Immunol 15:91–105

Dijkstra JM, Grimholt U, Leong J, Koop BF, Hashimoto K (2013) Comprehensive analysis of MHC class II genes in teleost fish genomes reveals dispensability of the peptide-loading DM system in a large part of vertebrates. BMC Evol Biol 13:260

Dijkstra JM, Yamaguchi T, Grimholt U (2018) Conservation of sequence motifs suggests that the nonclassical MHC class I lineages CD1/PROCR and UT were established before the emergence of tetrapod species. Immunogenetics 70:459–476

Dixon AM, Stanley BJ, Matthews EE, Dawson JP, Engelman DM (2006) Invariant chain transmembrane domain trimerization: a step in MHC class II assembly. Biochemistry 45:5228–5234

Dixon AM, Drake L, Hughes KT, Sargent E, Hunt D, Harton JA, Drake JR (2014) Differential transmembrane domain GXXXG motif pairing impacts major histocompatibility complex (MHC) class II structure. J Biol Chem 289:11695–11703

Drake LA, Drake JR (2016) A triad of molecular regions contribute to the formation of two distinct MHC class II conformers. Mol Immunol 74:59–70

Falk K, Rötzschke O, Stevanović S, Jung G, Rammensee HG (1994) Pool sequencing of natural HLA-DR, DQ, and DP ligands reveals detailed peptide motifs, constraints of processing, and general rules. Immunogenetics 39:230–242

Fineschi B, Arneson LS, Naujokas MF, Miller J (1995) Proteolysis of major histocompatibility complex class II-associated invariant chain is regulated by the alternatively spliced gene product, p41. Proc Natl Acad Sci U S A 92:10257–10261

Fineschi B, Sakaguchi K, Appella E, Miller J (1996) The proteolytic environment involved in MHC class II-restricted antigen presentation can be modulated by the p41 form of invariant chain. J Immunol 157:3211–3215

Flajnik MF, Kasahara M (2010) Origin and evolution of the adaptive immune system: genetic events and selective pressures. Nat Rev Genet 11:47–59

Flajnik MF, Deschacht N, Muyldermans S (2011) A case of convergence: why did a simple alternative to canonical antibodies arise in sharks and camels? PLoS Biol 9:e1001120

Fortin JS, Cloutier M, Thibodeau J (2013) Exposing the specific roles of the invariant chain isoforms in shaping the MHC class II Peptidome. Front Immunol 4:443

Frauwirth K, Shastri N (2001) Mutation of the invariant chain transmembrane region inhibits II degradation, prolongs association with MHC class II, and selectively disrupts antigen presentation. Cell Immunol 209:97–108

Fujiki K, Smith CM, Liu L, Sundick RS, Dixon B (2003) Alternate forms of MHC class II-associated invariant chain are not produced by alternative splicing in rainbow trout (Oncorhynchus mykiss) but are encoded by separate genes. Dev Comp Immunol 27:377–391

Fung-Leung WP, Surh CD, Liljedahl M, Pang J, Leturcq D, Peterson PA, Webb SR, Karlsson L (1996) Antigen presentation and T cell development in H2-M-deficient mice. Science 271:1278–1281

Gelin C, Sloma I, Charron D, Mooney N (2009) Regulation of MHC II and CD1 antigen presentation: from ubiquity to security. J Leukoc Biol 85:215–224

Ghosh P, Amaya M, Mellins E, Wiley DC (1995) The structure of an intermediate in class II MHC maturation: CLIP bound to HLA-DR3. Nature 378:457–462

Gore Y, Starlets D, Maharshak N, Becker-Herman S, Kaneyuki U, Leng L, Bucala R, Shachar I (2008) Macrophage migration inhibitory factor induces B cell survival by activation of a CD74-CD44 receptor complex. J Biol Chem 283:2784–2792

Grimholt U, Tsukamoto K, Azuma T, Leong J, Koop BF, Dijkstra JM (2015) A comprehensive analysis of teleost MHC class I sequences. BMC Evol Biol 15:32

Guillemot F, Turmel P, Charron D, Le Douarin N, Auffray C (1986) Structure, biosynthesis, and polymorphism of chicken MHC class II (B-L) antigens and associated molecules. J Immunol 137:1251–1257

Guncar G, Pungercic G, Klemencic I, Turk V, Turk D (1999) Crystal structure of MHC class II-associated p41 II fragment bound to cathepsin L reveals the structural basis for differentiation between cathepsins L and S. EMBO J 18:793–803

Günther S, Schlundt A, Sticht J, Roske Y, Heinemann U, Wiesmüller KH, Jung G, Falk K, Rötzschke O, Freund C (2010) Bidirectional binding of invariant chain peptides to an MHC class II molecule. Proc Natl Acad Sci U S A 107:22219–22224

Halaby DM, Poupon A, Mornon J (1999) The immunoglobulin fold family: sequence analysis and 3D structure comparisons. Protein Eng 12:563–571

Harton J, Jin L, Hahn A, Drake J (2016) Immunological functions of the membrane proximal region of MHC class II molecules. F1000Res. Faculty Rev-368

Hashimoto K (2003) The transformation hypothesis for the origin of the major histocompatibility complex (MHC) family molecules and the elucidation of the fundamental ancient duplication unit (FADU) in the genome. Recent Res Dev Immun 1:55–80

Hashimoto K, Nakanishi T, Kurosawa Y (1990) Isolation of carp genes encoding major histocompatibility complex antigens. Proc Natl Acad Sci U S A 87:6863–6867

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

Hopp TP, Woods KR (1981) Prediction of protein antigenic determinants from amino acid sequences. Proc Natl Acad Sci U S A 78:3824–3828

Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, Collins JE, Humphray S, McLaren K, Matthews L, McLaren S, Sealy I, Caccamo M, Churcher C, Scott C, Barrett JC, Koch R, Rauch GJ, White S, Chow W, Kilian B, Quintais LT, Guerra-Assunção JA, Zhou Y, Gu Y, Yen J, Vogel JH, Eyre T, Redmond S, Banerjee R, Chi J, Fu B, Langley E, Maguire SF, Laird GK, Lloyd D, Kenyon E, Donaldson S, Sehra H, Almeida-King J, Loveland J, Trevanion S, Jones M, Quail M, Willey D, Hunt A, Burton J, Sims S, McLay K, Plumb B, Davis J, Clee C, Oliver K, Clark R, Riddle C, Eliott D, Threadgold G, Harden G, Ware D, Mortimer B, Kerry G, Heath P, Phillimore B, Tracey A, Corby N, Dunn M, Johnson C, Wood J, Clark S, Pelan S, Griffiths G, Smith M, Glithero R, Howden P, Barker N, Stevens C, Harley J, Holt K, Panagiotidis G, Lovell J, Beasley H, Henderson C, Gordon D, Auger K, Wright D, Collins J, Raisen C, Dyer L, Leung K, Robertson L, Ambridge K, Leongamornlert D, McGuire S, Gilderthorp R, Griffiths C, Manthravadi D, Nichol S, Barker G, Whitehead S, Kay M, Brown J, Murnane C, Gray E, Humphries M, Sycamore N, Barker D, Saunders D, Wallis J, Babbage A, Hammond S, Mashreghi-Mohammadi M, Barr L, Martin S, Wray P, Ellington A, Matthews N, Ellwood M, Woodmansey R, Clark G, Cooper J, Tromans A, Grafham D, Skuce C, Pandian R, Andrews R, Harrison E, Kimberley A, Garnett J, Fosker N, Hall R, Garner P, Kelly D, Bird C, Palmer S, Gehring I, Berger A, Dooley CM, Ersan-Ürün Z, Eser C, Geiger H, Geisler M, Karotki L, Kirn A, Konantz J, Konantz M, Oberländer M, Rudolph-Geiger S, Teucke M, Osoegawa K, Zhu B, Rapp A, Widaa S, Langford C, Yang F, Carter NP, Harrow J, Ning Z, Herrero J, Searle SMJ, Enright A, Geisler R, Plasterk RHA, Lee C, Westerfield M, de Jong PJ, Zon LI, Postlethwait JH, Nüsslein-Volhard C, Hubbard TJP, Crollius HR, Rogers J, Stemple DL (2013) The zebrafish reference genome sequence and its relationship to the human genome. Nature 496:498–503

Hughes AL, Nei M (1989) Evolution of the major histocompatibility complex: independent origin of nonclassical class I genes in different groups of mammals. Mol Biol Evol 6:559–579

Hughes AL, Nei M (1993) Evolutionary relationships of the classes of major histocompatibility complex genes. Immunogenetics 37:337–346

Hüttl S, Helfrich F, Mentrup T, Held S, Fukumori A, Steiner H, Saftig P, Fluhrer R, Schröder B (2016) Substrate determinants of signal peptide peptidase-like 2a (SPPL2a)-mediated intramembrane proteolysis of the invariant chain CD74. Biochem J 473:1405–1422

Jahnke M, Trowsdale J, Kelly AP (2013) Ubiquitination of HLA-DO by MARCH family E3 ligases. Eur J Immunol 43:1153–1161

Jasanoff A, Park SJ, Wiley DC (1995) Direct observation of disordered regions in the major histocompatibility complex class II-associated invariant chain. Proc Natl Acad Sci U S A 92:9900–9904

Jasanoff A, Wagner G, Wiley DC (1998) Structure of a trimeric domain of the MHC class II-associated chaperonin and targeting protein II. EMBO J 17:6812–6808

Jones PP, Murphy DB, Hewgill D, McDevitt HO (1979) Detection of a common polypeptide chain in I–A and I–E sub-region immunoprecipitates. Mol Immunol 16:51–60

Kaufman J (2018) Unfinished business: evolution of the MHC and the adaptive immune system of jawed vertebrates. Annu Rev Immunol 36:383–409

Kaufman JF, Auffray C, Korman AJ, Shackelford DA, Strominger J (1984a) The class II molecules of the human and murine major histocompatibility complex. Cell 36:1–13

Kaufman JF, Krangel MS, Strominger JL (1984b) Cysteines in the transmembrane region of major histocompatibility complex antigens are fatty acylated via thioester bonds. J Biol Chem 259:7230–7238

Kaufman J, Andersen R, Avila D, Engberg J, Lambris J, Salomonsen J, Welinder K, Skjødt K (1992) Different features of the MHC class I heterodimer have evolved at different rates. Chicken B-F and beta 2-microglobulin sequences reveal invariant surface residues. J Immunol 148:1532–1546

Kaufman J, Milne S, Göbel TW, Walker BA, Jacob JP, Auffray C, Zoorob R, Beck S (1999) The chicken B locus is a minimal essential major histocompatibility complex. Nature 401:923–925

King G, Dixon AM (2010) Evidence for role of transmembrane helix-helix interactions in the assembly of the class II major histocompatibility complex. Mol BioSyst 6:1650–1661

Koch N, Lauer W, Habicht J, Dobberstein B (1987) Primary structure of the gene for the murine Ia antigen-associated invariant chains (II). An alternatively spliced exon encodes a cysteine-rich domain highly homologous to a repetitive sequence of thyroglobulin. EMBO J 6:1677–1683

Koch N, Zacharias M, König A, Temme S, Neumann J, Springer S (2011) Stoichiometry of HLA class II-invariant chain oligomers. PLoS One 6:e17257

Kokubu F, Litman R, Shamblott MJ, Hinds K, Litman GW (1988) Diverse organization of immunoglobulin VH gene loci in a primitive vertebrate. EMBO J 7:3413–3422

Komaniwa S, Hayashi H, Kawamoto H, Sato SB, Ikawa T, Katsura Y, Udaka K (2009) Lipid-mediated presentation of MHC class II molecules guides thymocytes to the CD4 lineage. Eur J Immunol 39:96–112

Kozik P, Francis RW, Seaman MN, Robinson MS (2010) A screen for endocytic motifs. Traffic 11:843–855

Kropshofer H, Vogt AB, Hämmerling GJ (1995) Structural features of the invariant chain fragment CLIP controlling rapid release from HLA-DR molecules and inhibition of peptide binding. Proc Natl Acad Sci U S A 92:8313–8317

Kropshofer H, Vogt AB, Moldenhauer G, Hammer J, Blum JS, Hämmerling GJ (1996) Editing of the HLA-DR-peptide repertoire by HLA-DM. EMBO J 15:6144–6154

Lamb CA, Cresswell P (1992) Assembly and transport properties of invariant chain trimers and HLA-DR-invariant chain complexes. J Immunol 148:3478–3482

Lennon-Duménil AM, Roberts RA, Valentijn K, Driessen C, Overkleeft HS, Erickson A, Peters PJ, Bikoff E, Ploegh HL, Wolf Bryant P (2001) The p41 isoform of invariant chain is a chaperone for cathepsin L. EMBO J 20:4055–4064

Li Q, Ao J, Mu Y, Yang Z, Li T, Zhang X, Chen X (2015) Cathepsin S, but not cathepsin L, participates in the MHC class II-associated invariant chain processing in large yellow croaker (Larimichthys crocea). Fish Shellfish Immunol 47:743–750

Lindstedt R, Liljedahl M, Péléraux A, Peterson PA, Karlsson L (1995) The MHC class II molecule H2-M is targeted to an endosomal compartment by a tyrosine-based targeting motif. Immunity 3:561–572

Litman GW, Rast JP, Shamblott MJ, Haire RN, Hulst M, Roess W, Litman RT, Hinds-Frey KR, Zilch A, Amemiya CT (1993) Phylogenetic diversification of immunoglobulin genes and the antibody repertoire. Mol Biol Evol 10:60–72

Litman GW, Rast JP, Fugmann SD (2010) The origins of vertebrate adaptive immunity. Nat Rev Immunol 10:543–553

Lotteau V, Teyton L, Burroughs D, Charron D (1987) A novel HLA class II molecule (DR alpha-DQ beta) created by mismatched isotype pairing. Nature 329:339–341

Majera D, Kristan KČ, Neefjes J, Turk D, Mihelič M (2012) Expression, purification and assembly of soluble multimeric MHC class II-invariant chain complexes. FEBS Lett 586:1318–1324

Malmstrøm M, Matschiner M, Tørresen OK, Star B, Snipen LG, Hansen TF, Baalsrud HT, Nederbragt AJ, Hanel R, Salzburger W, Stenseth NC, Jakobsen KS, Jentoft S (2016) Evolution of the immune system influences speciation rates in teleost fishes. Nat Genet 48:1204–1210

Marks MS, Germain RN, Bonifacino JS (1995) Transient aggregation of major histocompatibility complex class II chains during assembly in normal spleen cells. J Biol Chem 270:10475–10481

Mellins ED, Stern LJ (2014) HLA-DM and HLA-DO, key regulators of MHC-II processing and presentation. Curr Opin Immunol 26:115–122

Mihelic M, Turk D (2007) Two decades of thyroglobulin type-1 domain research. Biol Chem 388:1123–1130

Mihelic M, Dobersek A, Guncar G, Turk D (2008) Inhibitory fragment from the p41 form of invariant chain can regulate activity of cysteine cathepsins in antigen presentation. J Biol Chem 283:14453–14460

Moore DT, Berger BW, DeGrado WF (2008) Protein-protein interactions in the membrane: sequence, structural, and biological motifs. Structure 16:991–1001

Morris P, Shaman J, Attaya M, Amaya M, Goodman S, Bergman C, Monaco JJ, Mellins E (1994) An essential role for HLA-DM in antigen presentation by class II major histocompatibility molecules. Nature 368:551–554

Mosyak L, Zaller DM, Wiley DC (1998) The structure of HLA-DM, the peptide exchange catalyst that loads antigen onto class II MHC molecules during antigen presentation. Immunity 9:377–383

Naujokas MF, Morin M, Anderson MS, Peterson M, Miller J (1993) The chondroitin sulfate form of invariant chain can enhance stimulation of T cell responses through interaction with CD44. Cell 74:257–268

Neefjes J, Jongsma ML, Paul P, Bakke O (2011) Towards a systems understanding of MHC class I and MHC class II antigen presentation. Nat Rev Immunol 11:823–836

Nguyen TB, Jayaraman P, Bergseng E, Madhusudhan MS, Kim CY, Sollid LM (2017) Unraveling the structural basis for the unusually rich association of human leukocyte antigen DQ2.5 with class-II-associated invariant chain peptides. J Biol Chem 292:9218–9228

Ohnishi K (1984) Domain structures and molecular evolution of class I and class II major histocompatibility gene complex (MHC) products deduced from amino acid and nucleotide sequence homologies. Orig Life 14:707–715

Ohta Y, Goetz W, Hossain MZ, Nonaka M, Flajnik MF (2006) Ancestral organization of the MHC revealed in the amphibian Xenopus. J Immunol 176:3674–3685

O'Sullivan DM, Noonan D, Quaranta V (1987) Four Ia invariant chain forms derive from a single gene by alternate splicing and alternate initiation of transcription/translation. J Exp Med 166:444–460

Painter CA, Stern LJ (2012) Conformational variation in structures of classical and non-classical MHCII proteins and functional implications. Immunol Rev 250:144–157

Pieters J, Bakke O, Dobberstein B (1993) The MHC class II-associated invariant chain contains two endosomal targeting signals within its cytoplasmic tail. J Cell Sci 106:831–846

Pond L, Kuhn LA, Teyton L, Schutze MP, Tainer JA, Jackson MR, Peterson PA (1995) A role for acidic residues in di-leucine motif-based targeting to the endocytic pathway. J Biol Chem 270:19989–19997

Pos W, Sethi DK, Call MJ, Schulze MS, Anders AK, Pyrdol J, Wucherpfennig KW (2012) Crystal structure of the HLA-DM-HLA-DR1 complex defines mechanisms for rapid peptide selection. Cell 151:1557–1568

Pos W, Sethi DK, Wucherpfennig KW (2013) Mechanisms of peptide repertoire selection by HLA-DM. Trends Immunol 34:495–501

Rammensee HG, Friede T, Stevanoviíc S (1995) MHC ligands and peptide motifs: first listing. Immunogenetics 41:178–228

Riberdy JM, Avva RR, Geuze HJ, Cresswell P (1994) Transport and intracellular distribution of MHC class II molecules and associated invariant chain in normal and antigen-processing mutant cell lines. J Cell Biol 125:1225–1237

Roche PA, Furuta K (2015) The ins and outs of MHC class II-mediated antigen processing and presentation. Nat Rev Immunol 15:203–216

Rudensky AY, Preston-Hurlburt P, Hong SC, Barlow A, Janeway CA Jr (1991) Sequence analysis of peptides bound to MHC class II molecules. Nature 353:622–627

Santos MS, Park CK, Foss SM, Li H, Voglmaier SM (2013) Sorting of the vesicular GABA transporter to functional vesicle pools by an atypical dileucine-like motif. J Neurosci 33:10634–10646

Schafer PH, Pierce SK, Jardetzky TS (1995 Dec) The structure of MHC class II: a role for dimer of dimers. Semin Immunol 7(6):389–398

Schulze MS, Wucherpfennig KW (2012) The mechanism of HLA-DM induced peptide exchange in the MHC class II antigen presentation pathway. Curr Opin Immunol 24:105–111

Shachar I, Flavell RA (1996) Requirement for invariant chain in B cell maturation and function. Science 274:106–108

Silva DS, Reis MI, Nascimento DS, do Vale A, Pereira PJ, dos Santos NM (2007) Sea bass (Dicentrarchus labrax) invariant chain and class II major histocompatibility complex: sequencing and structural analysis using 3D homology modelling. Mol Immunol 44:3758–3776

Sloan VS, Cameron P, Porter G, Gammon M, Amaya M, Mellins E, Zaller DM (1995) Mediation by HLA-DM of dissociation of peptides from HLA-DR. Nature 375:802–806

Star B, Nederbragt AJ, Jentoft S, Grimholt U, Malmstrøm M, Gregers TF, Rounge TB, Paulsen J, Solbakken MH, Sharma A, Wetten OF, Lanzén A, Winer R, Knight J, Vogel JH, Aken B, Andersen Ø, Lagesen K, Tooming-Klunderud A, Edvardsen RB, Tina KG, Espelund M, Nepal C, Previti C, Karlsen BO, Moum T, Skage M, Berg PR, Gjøen T, Kuhl H, Thorsen J, Malde K, Reinhardt R, du L, Johansen SD, Searle S, Lien S, Nilsen F, Jonassen I, Omholt SW, Stenseth NC, Jakobsen KS (2011) The genome sequence of Atlantic cod reveals a unique immune system. Nature 477:207–210

Stern LJ, Wiley DC (1992) The human class II MHC protein HLA-DR1 assembles as empty alpha beta heterodimers in the absence of antigenic peptide. Cell 68(3):465–477

Stern LJ, Wiley DC (1994) Antigenic peptide binding by class I and class II histocompatibility proteins. Structure 2:245–251

Strubin M, Berte C, Mach B (1986) Alternative splicing and alternative initiation of translation explain the four forms of the Ia antigen-associated invariant chain. EMBO J 5:3483–3488

Tohmé M, Manoury B (2015) Invariant chain is a new chaperone for TLR7 in B cells. Mol Immunol 68:102–105

Travers P, Blundell TL, Sternberg MJ, Bodmer WF (1984) Structural and evolutionary analysis of HLA-D-region products. Nature 310(5974):235–238

Tulp A, Verwoerd D, Dobberstein B, Ploegh HL, Pieters J (1994) Isolation and characterization of the intracellular MHC class II compartment. Nature 369:120–126

Viville S, Neefjes J, Lotteau V, Dierich A, Lemeur M, Ploegh H, Benoist C, Mathis D (1993) Mice lacking the MHC class II-associated invariant chain. Cell 72:635–648

Walter W, Scheuer C, Lingnau K, Reichert TE, Schmitt E, Loos M, Maeurer MJ (2000) H2-M, a facilitator of MHC class II peptide loading, and its negative modulator H2-O are differentially expressed in response to proinflammatory cytokines. Immunogenetics 51:794–804

Ye H, Xu FZ, Yu WY (2009) The intracellular localization and oligomerization of chicken invariant chain with major histocompatibility complex class II subunits. Poult Sci 88:1594–1600

Yin L, Calvo-Calle JM, Dominguez-Amorocho O, Stern LJ (2012) HLA-DM constrains epitope selection in the human CD4 T cell response to vaccinia virus by favoring the presentation of peptides with longer HLA-DM-mediated half-lives. J Immunol 189:3983–3994

Yoder JA, Haire RN, Litman GW (1999) Cloning of two zebrafish cDNAs that share domains with the MHC class II-associated invariant chain. Immunogenetics 50:84–88

Zhou Z, Callaway KA, Weber DA, Jensen PE (2009) Cutting edge: HLA-DM functions through a mechanism that does not require specific conserved hydrogen bonds in class II MHC-peptide complexes. J Immunol 183:4187–4191