Developmental and Comparative Immunology 36 (2012) 521–533

Contents lists available at SciVerse ScienceDirect

Developmental and Comparative Immunology

journal homepage: www.elsevier.com/locate/dci

Shark class II invariant chain reveals ancient conserved relationships with cathepsins and MHC class II

Michael F. Criscitiello [a,*], Yuko Ohta [b], Matthew D. Graham [b,1], Jeannine O. Eubanks [a], Patricia L. Chen [a], Martin F. Flajnik [b]

[a] Department of Veterinary Pathobiology, College of Veterinary Medicine and Biomedical Sciences, Texas A&M University, College Station, TX 77843, USA
[b] Department of Microbiology and Immunology, School of Medicine, University of Maryland at Baltimore, Baltimore, MD 21201, USA

ARTICLE INFO

Article history:
Received 11 September 2011
Revised 16 September 2011
Accepted 16 September 2011
Available online 1 October 2011

Keywords:
MHC
Antigen processing
Invariant chain
Evolution
Shark

ABSTRACT

The invariant chain (Ii) is the critical third chain required for the MHC class II heterodimer to be properly guided through the cell, loaded with peptide, and expressed on the surface of antigen presenting cells. Here, we report the isolation of the nurse shark Ii gene, and the comparative analysis of Ii splice variants, expression, genomic organization, predicted structure, and function throughout vertebrate evolution. Alternative splicing to yield Ii with and without the putative protease-protective, thyroglobulin-like domain is as ancient as the MHC-based adaptive immune system, as our analyses in shark and lizard further show conservation of this mechanism in all vertebrate classes except bony fish. Remarkable coordinate expression of Ii and class II was found in shark tissues. Conserved Ii residues and cathepsin L orthologs suggest their long co-evolution in the antigen presentation pathway, and genomic analyses suggest 450 million years of conserved Ii exon/intron structure. Other than an extended linker preceding the thyroglobulin-like domain in cartilaginous fish, the Ii gene and protein are predicted to have largely similar physiology from shark to man. Duplicated Ii genes found only in teleosts appear to have become sub-functionalized, as one form is predicted to play the same role as that mediated by Ii mRNA alternative splicing in all other vertebrate classes. No Ii homologs or potential ancestors of any of the functional Ii domains were found in the jawless fish or lower chordates.

1. Introduction

The hallmark molecular components of the adaptive immune response have been found in the oldest group of living jawed vertebrates, the cartilaginous fish. Multiple IgH (Flajnik, 2002), IgL (Criscitiello and Flajnik, 2007), and TCR (Criscitiello et al., 2010; Rast et al., 1997) are diversified by RAG (Bernstein et al., 1994) and AID (Conticello et al., 2005) in sharks and rays. Furthermore, these animals are in the oldest extant group of vertebrates having a polymorphic, polygenic major histocompatibility complex (MHC) (Kasahara et al., 1992).

MHC class II glycoproteins present peptides to $CD4^+$ T cells. Newly synthesized class II α and β chains assemble in the endoplasmic reticulum (ER) together with a Type II glycoprotein called the invariant chain (Ii) (Jones et al., 1979). Ii is a chaperone ensuring the correct folding and trafficking of MHC class II proteins (Bikoff et al., 1993). Ii first trimerizes before the sequential addition of three class II α/β dimers (Lamb and Cresswell, 1992). In this nine-chain complex, each Ii blocks the peptide binding groove of one of the three class II heterodimers with a peptide called CLIP (class II-associated invariant chain peptide) (Freisewinkel et al., 1993), which prevents loading of class II with ER-derived proteins and peptides and provides the groove occupancy required for the stability of class II heterodimers (Zhong et al., 1996). In the trans-Golgi the αβIi complex is diverted from the secretory pathway to the endocytic pathway via a conserved motif in the Ii cytoplasmic tail. The complex is transported to an acidified post-Golgi vesicle where first the membrane distal portion of Ii is proteolytically degraded to leave the LIP (leupeptin-induced peptide), which still blocks the peptide binding cleft and retains the Ii transmembrane and cytoplasmic segments that continue to target the complex to the endosomal MHC class II compartment (MIIC) (Blum and Cresswell, 1988). The low pH of MIIC activates proteases to further cleave the membrane-proximal portion of Ii, leaving only CLIP in the peptide binding cleft. Blockade of this progressive cleavage of the Ii results in accumulation of Ii intermediates and reduced class II surface expression (Neefjes and Ploegh, 1992). DM can then bind class

* Corresponding author. Address: Texas A&M University, Mailstop 4467, College Station, TX 77843, USA. Tel.: +1 979 845 4207, mobile: +1 305 299 2522; fax: +1 979 862 1088.
E-mail addresses: mcriscitiello@cvm.tamu.edu (M.F. Criscitiello), yota@som.umaryland.edu (Y. Ohta), jeubanks@cvm.tamu.edu (J.O. Eubanks), pchen@cvm.tamu.edu (P.L. Chen), mflajnik@som.umaryland.edu (M.F. Flajnik).
[1] Deceased.

0145-305X/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.dci.2011.09.008

II and release LIP or CLIP, facilitating the exchange for endosomal peptides before transport to the cell surface (Schafer et al., 1996). Maturation of phagosomes containing non-self cargo (versus apoptotic self-cargo) may be enhanced by toll-like receptor-mediated vesicular signaling, marking those phagosomes with pathogenic contents for fusion into the MIIC compartment (Blander and Medzhitov, 2006).

The enzymes that cleave Ii are related to papain and known as cathepsins, which are the same proteases that degrade lysosomal contents for antigen loading (Riese and Chapman, 2000). The activation of cathepsins requires an acidic environment, and they can be divided into four categories depending on the critical component of their active sites: cysteine, aspartate, serine, or metal ions (Turk et al., 2001). Cysteine cathepsins are primarily involved in antigen processing, specifically cathepsins L and S are dedicated to this function. Cathepsin activity is regulated by several protein inhibitors, including cystatins, thyropins and even one domain of Ii itself. The thyroglobulin-like (Tg) domain of longer Ii isoforms is a strong inhibitor of cathepsin L but not cathepsin S (Bevec et al., 1996).

The mammalian Ii gene has an exon organization that largely corresponds to its structural protein domains. The first exon encodes the amino-terminal cytoplasmic tail including the endosomal targeting motifs, the second exon encodes the transmembrane domain, and the third exon encodes a linker between the membrane and the trimerization domain that includes CLIP. The fourth, fifth and sixth exons contribute to the three alpha helices and connecting strands of the trimerization domain. The seventh exon, alternatively spliced out in short isoforms, encodes the Tg domain that presumably inhibits cathepsin proteolytic action. The eighth exon nearly encodes the entire carboxy-terminal end, which is often rich in charged residues but has unclear function. The ninth protein-coding exon encodes the final amino acid or two and contains the stop codon.

Unlike MHC genes, Ii genes do not display high allelic polymorphism, but four variants of the protein are found in human as p33, p35, p41 and p43 (O'Sullivan et al., 1987). Use of an alternative start codon accounts for the small molecular weight differences between the (predominant) p33 and p41, and p35 and p43 forms. The p35 and p43 forms contain an ER-retention motif lacking in the shorter forms from the alternative initiation site; this signal is concealed upon $\alpha\beta$ binding and allows transport of the nonamer to the Golgi (Schutze et al., 1994). The 10 kDa distinction between the p33/p35 and p41/p43 forms results from the alternatively spliced Tg domain exon, mentioned above. The Tg domain is a structural motif found in several functionally unrelated proteins (e.g., testican, equistatin, thyroglobulin) and sometimes functions as an inhibitor of cysteine proteases, often with higher target specificity than the better studied cystatins (Mihelic and Turk, 2007).

Ii knockout mice show impeded class II transport and surface expression. Class II found on the surface of Ii-deficient cells has an unstable conformation due to the lack of endogenously processed peptide, but the dimers can bind peptide added to the medium. Accordingly, cells from these mice do not present whole exogenous antigen well and the animals have greatly reduced numbers of thymic and peripheral $CD4^+$ T cells (Bikoff et al., 1993; Viville et al., 1993). These transgenic mice studies suggest that Ii prevents class II from binding floppy, incompletely folded proteins in the ER (rather than preventing the binding of peptides transported into the ER by TAP) and stabilize the heterodimer. Ii knockout mice with restoration of either p31 or p41 (containing the Tg domain) have shown that both forms participate in class II folding and assembly, can reconstitute the $CD4^+$ T cell population, and rescue immune responses to protein antigen (Shachar et al., 1995; Takaesu et al., 1995). The complete functions of each isoform are not known; however p41 has been shown to be necessary for airway hyperresponsiveness and IgE responses in the lung (Ye et al., 2003).

Although crucial to class II antigen presentation, Ii and cathepsins are encoded outside of the MHC (Long et al., 1983). However, cathepsins S and L are found in MHC paralogous regions (Flajnik and Kasahara, 2010), one of many linkages that contribute to hypotheses of an ancestral "pre-adaptive immune complex" encoding antigen receptors, NK receptors and antigen processing and presentation components (Ohta et al., 2011). Ii and homologous genes have been identified in several divergent vertebrate model species, although such reports are few in comparison to class II α/β chains. Amongst poikilothermic vertebrates, annotated Ii sequences have been submitted to public databases from reptiles and amphibians and studies have been conducted on Ii from bony fish species. The cloning of the first Ii from lower vertebrates was done in zebrafish, and this work confirmed that, like in mammals, fish Ii-like transcripts exist in multiple forms (Yoder et al., 1999). Work in rainbow trout also found two Ii products (Dijkstra et al., 2003) that are encoded by two paralogous genes (Fujiki et al., 2003) as opposed to alternative splicing. Structure of sea bass Ii was modeled more recently with analysis of potential interactions with class II and cathepsins (Silva et al., 2007). Here, we report the first description of Ii from the cartilaginous fish, the oldest vertebrate group with MHC-based adaptive immunity. We set out to determine whether the gene and its expression were phylogenetically conserved, and attempted to find genes related to precursors of Ii and cathepsins in jawless vertebrates and lower deuterostomes.

2. Methods

2.1. Cloning of nurse shark Ii chain and cathepsins

A *Ginglymostoma cirratum* (nurse shark) spleen/pancreas cDNA library was constructed in the pDONR222 vector using the Gateway cloning system (Invitrogen). From this an ~8000 clone expressed sequence tag (EST) database was created after removing known housekeeping, MHC, immunoglobulin and TCR clones by subtractive colony hybridization using 137 mm Magna membranes (Osmonics) for probing and high stringency washing techniques described previously (Criscitiello et al., 2004). DNA was prepared with 96-Turbo plasmid miniprep kits (Qiagen) or TempliPhi rolling circle DNA amplification (GE Healthcare) and single dye-terminator based sequencing runs were performed at the University of Maryland Biopolymer Core Facility using the universal M13 reverse primer. ESTs were used as queries against the non-redundant protein sequence database with blastx (NCBI). Gene-specific primers (Supplemental Table 1) were designed to complete sequencing of clones with high identity to Ii and cathepsins. Additional cDNA libraries from shark lymphoid tissues were assayed by 5′ and 3′ rapid amplification of cDNA ends (RACE) PCR with gene-specific Ii primers to identify all expressed splice variants. These were cloned and sequenced as above or with Zyppy plasmid DNA miniprep kit (Zymo Research), extended with BigDye XTerminator (Applied Biosystems), purified and sequenced by the Texas A&M DNA Technologies Core Laboratory.

2.2. Blotting

Total RNA was prepared for northern blotting as described (Bartl et al., 1997), and 10 μg was loaded in each lane. The nurse shark nucleotide diphosphate kinase (NDPK) probe used as a loading control was amplified with primers NDPKF and NDPKR (Kasahara et al., 1992) (Supplemental Table 1). A probe for nurse shark Ii was amplified from primers NSIiF1 and NSIiR1 which generate a probe from cDNA encoding the endosomal targeting sequence of the cytoplasmic tail to CLIP. Northern blotting and