# Processing of fish Ig heavy chain transcripts: Diverse splicing patterns and unusual nonsense mediated decay

Sylvie Quiniou, Melanie Wilson, Pierre Boudinot

Developmental and Comparative Immunology 35 (2011) 949–958

Contents lists available at ScienceDirect

Developmental and Comparative Immunology

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

Review

# Processing of fish Ig heavy chain transcripts: Diverse splicing patterns and unusual nonsense mediated decay

Sylvie M.A. Quiniou[a], Melanie Wilson[b], Pierre Boudinot[c,*]

[a] USDA-ARS, Catfish Genetics Research Unit, Stoneville, MS 38776, USA
[b] University of Mississippi Medical Center, Dept of Microbiology, Jackson, MS 39216, USA
[c] INRA, Unité de Virologie et Immunologie Moléculaires, 78352 Jouy-en-Josas Cedex, France

ARTICLE INFO

*Article history:*
Available online 17 December 2010

*Keywords:*
Fish
Immunoglobulin
Splicing
NMD
IgM
IgD
IgT

ABSTRACT

While the diversification of the antigen-binding sites is realized by genomic VDJ rearrangements during B cell differentiation, different forms of immunoglobulin (Ig) heavy (H) chains can be produced through multiple splicing pathways. In most vertebrates, the secreted (S) and membrane (Mb) forms of IgM chain are created by alternative splicing through usage of a cryptic splice site in Cμ4 allowing the junction to the TM exon. The processing pattern for Igμ is different in teleosts, which generally use the Cμ3 donor site instead. In ancient fish lineages, multiple unusual splicing patterns were found for Ig H chain, involving donor sites that do not always follow the classical consensus. The production of IgD versus IgM H chains seems to be generally realized by alternative splicing in all vertebrates, but typical teleost IgD H chains are chimeric and contains a Cμ1 domain. Together, these observations raise questions on how different fish regulate RNA splicing and if their splicing machinery is especially complex. A preliminary scan of the zebrafish and stickleback genomes provides evidence that gene orthologs to the mammalian main splice factors are highly conserved as single copy genes, while the snRNPs U repertoire may be different and may explain other particular features of RNA processing in fish.



**Contents**



## 1. Introduction

The protein coding sequences of the majority of eukaryotic genes are interrupted by non-coding sequences termed introns. Since its discovery in 1977, the process of RNA splicing, which removes introns from newly synthesized RNAs, appears as the major regulatory and diversifying mechanism in gene expression.

* Corresponding author. Tel.: +33 1 34652585; fax: +33 1 34652591.
*E-mail address:* pierre.boudinot@jouy.inra.fr (P. Boudinot).

0145-305X/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.dci.2010.12.007

The splicing process is operated by the spliceosome, a large and dynamic RNP-based machinery, and it is this RNA/protein complex that insures not only accuracy of splice site recognition, but through the choice of the site conveys a flexibility that leads to a diversity of alternatively spliced transcripts. Many multi-exon genes are spliced into multiple isoforms and as expected the resulting transcriptome is dependent upon the cell type and expression context. Thus, alternative splicing controls cellular processes by adjusting protein isoform repertoire in response to varying cellular conditions. This tissue- and context-dependent alternative splicing is determined by trans-acting proteic factors, cis-acting motifs present in the transcripts, and also by other parameters such as RNA secondary structure. However, the control of the alternative "splicing code" is still not entirely deciphered (Barash et al., 2010). Currently, even though the first global attempts to establish a comprehensive set of predictive rules for alternative splicing are very promising, they do not account for 100% of observed transcribed splice variants. To be complete, the splicing code should also predict the alternative splicing induced by environment effects on cell biology. For example, this is particularly relevant for immune response genes that change their splicing pattern depending upon the cell activation state.

The genes of immunity, especially those involved in pathogen sensing or cell–cell interactions, are frequently highly diversified to cope with the huge variability of the ligands that they must recognize. In vertebrates, diversification mechanisms are not dependent on alternative splicing, but rather typified by those that generate immunoglobulins (Ig), i.e. RAG based VDJ rearrangement, gene conversion and somatic hypermutation, processes first discovered in mammals (reviewed in Max (2009)). In contrast, the best example of diversification generated by alternative splicing is perhaps the crustacean and insect Down syndrome cell adhesion molecule (Dscam) gene, which uses combinatorial splicing of alternative exons to generate tens of thousands of protein isoforms (Schmucker et al., 2000; Wojtowicz et al., 2004; Brites et al., 2008) and have variable functions as axon-guidance proteins and as antigen receptors possibly involved in immunity (Dong et al., 2006). Alternative splicing in Dscam can modulate the extracellular receptor moiety, the transmembrane segment as well as the cytoplasmic tail by differential usage of exons containing activating or inhibitory tyrosine based immuno-modulatory motifs (Brites et al., 2008). However, it is also important to emphasize that alternative splicing plays a key role in gnathostome Ig expression in determining the fate of membrane (Mb) bound Ig versus secreted (S) Ig and in some cases the expressed Ig heavy (H) chain isotype (Igμ versus Igδ) of the molecule (reviewed in Max (2009)).

Among gnathostomes, the splicing rules of Ig H transcripts in "fishes"[1] show interesting peculiarities, which may shed light on the mechanisms and variations of the vertebrate splicing machinery: (1) although fish possess a limited number of isotypes (IgM, IgD and IgT/Z) they show a large diversity of splicing patterns leading to diverse IgH isoforms. (2) It has been observed that Ig pre-mRNA containing premature STOP codons introduced by out-of-frame VDJ recombination are not as efficiently eliminated in teleost fish as they are in mammals. The process responsible for this elimination – the nonsense mediated decay (NMD) – is in fact tightly connected to the splicing process and protects the cell from the detrimental effects of truncated – often dominant negative – proteins (Wilkinson, 2005).

This review will focus on the particularities of the Ig transcript processing by splicing and NMD pathways in fish and their evolutionary implications.

[1] The term "Fish" designates here the Actinopterygii, including ancient lineages such as Chondrostei (Bichirs, Sturgeons), Holostei (gar, bowfin) as well as the large group of teleosts.

## 2. Multiple Igμ heavy chain transcripts are produced by alternative splicing in fish

### 2.1. Membrane bound Ig and alternative splicing of the IgH chain transcripts in mammals

In mice and humans, S- and Mb-IgM H chains are produced from the same transcript through alternative splicing. This process is regulated during B cell development; mRNA encoding the Mb-IgM H chain predominates at the first stages of B cell differentiation while S-Igμ transcripts are most abundant in plasma cells. This "switch" is the end result of a regulation process involving splice site choice as well as cleavage and polyadenylation site choice. An S-Igμ transcript is produced when the mRNA is cleaved and polyadenylated between the constant (C) region domain Cμ4 and the transmembrane (TM) exons, while the use of a cryptic splice site located within Cμ4 with the 3′ acceptor site of the TM1 exon leads to Mb-IgM expression. Furthermore, regulation of Igμ expression does not require Igμ gene-specific splice sequences, since the balance is maintained when the Igμ splice sites are replaced by exogenous splice sequences (Peterson, 1992; Seipelt et al., 1998). Also, the balance of S-Igμ and Mb-Igμ expression is controlled by both cis-acting sequences and regulatory components of the splicing machinery acting in trans (Edwalds-Gilbert et al., 1997; Zhao et al., 1999; Takagaki et al., 1996).

### 2.2. Diverse splicing pathways of membrane bound Igμ in teleost fish

While *Xenopus* and cartilaginous fish also possess a cryptic splice site encoded within their Cμ4 and process S-Igμ and Mb-Igμ in a matter similar to that of mammals (Ross et al., 1998), another processing pattern for Igμ was discovered in teleosts in 1990. Mb-Igμ transcripts were shown to have the TM exons spliced directly to the donor site located at the 3′ end of the Cμ3 exon instead of having the TM1 exon spliced to a cryptic donor splice within Cμ4. This pathway was first described in the channel catfish, *Ictalurus punctatus*, where it was also shown that the Cμ4 internal cryptic splicing site is missing. Thus, catfish Mb-Igμ H chain lacks the Cμ4 domain, while the S-Igμ H chain still contains it. This pathway was later described in many other teleost species, including rainbow trout, *Oncorhynchus mykiss*, Atlantic cod, *Gadus morhua*, and common carp, *Cyprinus carpio* L., to name a few, and this "atypical Mb-Igμ splicing" appears to be the general rule followed by teleost fish (van Ginkel et al., 1994; Lee et al., 1993; Bengten et al., 1991) since the Cμ4 exon lacks an effective cryptic splice site. The only exceptions reported to date are (1) a VDJ–Cμ1–TM1–TM2 Mb-Igμ transcript recently identified in zebrafish, which is probably produced by alternative splicing (Hu et al., 2011) and (2) the particular pattern reported in Antarctic fishes (see below). Importantly, as was demonstrated in the catfish model, the VDJ–Cμ1–Cμ2–Cμ3–TM1–TM2 Mb-Igμ form does not inhibit B cell signal transduction since peripheral blood leukocytes (PBL) exhibit rapid intracellular phosphorylation events, calcium flux and proliferation in response to anti-IgM stimulation (van Ginkel et al., 1994). Also, a typical conserved antigen receptor transmembrane (CART) motif (Campbell et al., 1994), which is critical for the interactions of membrane bound Igs with B cell signaling accessory molecules, can be observed in the TM region of all teleost Mb-Igμ cDNAs examined to date and since the intracytoplasmic tail of teleost Mb-Igμ H chains resembles mammalian Mb-Igμ and consists of only three amino acids, it was predicted that like all the mammalian Mb-Ig isotypes, teleost Mb-Igμ cannot alone transduce an activation signal. In other words, fish Mb-Igμ must associate with signaling transducing molecules comparable to CD79a and CD79b. More recently, orthologs of CD79a and CD79b containing conserved immune-receptor tyrosine-based

activation motifs (ITAM) have been identified in pufferfish, *Fugu rubripes*, rainbow trout, stickleback, *Gasterosteus aculeatus*, catfish and zebrafish (Guselnikov et al., 2003; Sahoo et al., 2008; Lundqvist et al., 2009; Hu et al., 2010) and in studies using catfish clonal B cells it was demonstrated that epitope-tagged CD79a and CD79b are non-covalently associated with Mb-Igμ. Therefore, teleost fish do express functional B cell receptors (BCR).

Another group of teleost fish, the Antarctic Notothenioids, exhibits yet a different Mb-Igμ mRNA splicing pattern. In 2003, it was found that rockcod, *Notothenia coriiceps*, Mb-Igμ transcripts lack both Cμ3 and Cμ4 domains and the TM exons were spliced directly to the 3′ end of the Cμ2 domain (Ota et al., 2003). More recently, a comprehensive sequencing study of 13 species of Notothenoids (Coscia et al., 2010) showed that, like the rockcod, the majority of these species also expressed Mb-Igμ transcripts lacking both Cμ3 and Cμ4 exons. In addition, genomic analyses also revealed the presence of two exons consisting of 39-nucleotides (RA and RB), between the Cμ3 and TM1 exons (Coscia et al., 2010). These exons are flanked by type-1 splicing signals (predicted to be spliced by U1, U5, U2 snRNPs) and in Mb-Igμ transcripts these short exons are spliced to Cμ2 on their 5′ side and to TM1 on the 3′ side. The inclusion of RA and RB in Mb-Igμ creates an elongated extracellular membrane-proximal domain. A third putative exon termed RC has also been found upstream of RB, but it is not included in any of the mature mRNA transcripts sequenced to date. This unusual Ig structure and splicing pattern is due to a large genomic insertion in the Igμ gene locus that occurred during the radiation of the teleost and the major cooling period of the Antarctic. The finding of this splicing pattern in most of the Antarctic Notothenioids supports the idea of adaptive selection of IgM during Notothenioid evolution proposed by Ota et al. (2003).

### 2.3. Diverse patterns of splicing of Igμ chain transcripts in ancient fish lineages and atypical splicing sites: Do fish have a special splicing code?

When species belonging to ancient fish lineages were examined, a diversity of splicing patterns of the Mb-Igμ transcripts was observed. Two holosteans, the bowfin, *Amia calva*, and the longnose gar, *Lepisosteus osseus*, use the teleostean pathway to produce Mb-Igμ (Cμ3 → TM1) via a conserved cryptic site located in the middle of the Cμ3 exon to produce a shorter Mb-Igμ transcript as well as the mammalian pathway (Cμ4 → TM1) (Wilson et al., 1995a, 1995b). The splicing signals of the cryptic or canonical sites in the Mb-Igμ sequence of these two species are generally compatible with the type 1 consensus splice sites which suggests that there is conservation of splicing machinery. However, it is currently unknown how the ratio between the different spliced forms of Igμ is regulated by the splicing machinery. In another lineage of ancient fish, the chondrosteans, it was also found that Mb-Igμ transcripts were spliced using several pathways. In Siberian sturgeon Mb-Igμ transcripts, the TM1 exon can be spliced to three possible donor sites: (1) a cryptic site at the end of Cμ4, (2) a cryptic site at the end of Cμ3, or (3) the donor splice at the 3′-end of Cμ1. The resulting IgM H chains therefore would contain four, two and a half, or only one Cμ domain(s), respectively (Lundqvist et al., 2009). The shortest Mb-Igμ splice variant, Cμ1 → TM1, was found to be restricted to Mb-Igμ transcripts expressing VH2 segments, while the other two Mb-Igμ variants were found to express members of all of the sturgeon VH families (Lundqvist et al., 2009). In general, the splicing of all vertebrate Igμ exons follows the classical GT/AG rule. However, in the Siberian sturgeon the cryptic Cμ3 → TM1 pathway uses a unique combination of donor, branch, and acceptor sites. The Cμ3 donor site is reminiscent of a type 2 donor site, but type 2 branching and acceptor sites were not found at the expected positions upstream or at the 5′ end of the TM1 exon, respectively. Since type 2 splice sites are usually conserved even in fish, the sturgeon Cμ3–TM1 intron may belong to a new third set of introns, that has evolved in the polyploid context of the Siberian sturgeon genome (Vasil'yev et al., 1981). Fig. 1 shows the different vertebrate Mb-Igμ splicing patterns as represented in the context of the phylogenetic tree of tetrapods and "fish" groups (Janvier et al., 1996). Not only is the TM1 exon splice to the Cμ4 exon internal cryptic donor site the generally accepted Mb-Igμ splicing rule in all tetrapods, it also appears to be an ancestral characteristic. The only exception occurs in the fishes (i.e. Actinopterygii). Notably, the finding of additional Mb-Igμ splicing pathways in the actinopterygians suggests that the Mb-Igμ splicing pathways were subjected to various modifying processes several times during evolution. This diversification of Mb-Igμ splicing pathways in "ancient fish," followed by a general standardization in teleosts may be related to the genome duplication episode that occurred after the separation of fish and tetrapod lineages (Jaillon et al., 2004). Briefly, the genome duplication could have modified the repertoire of splicing factors or snRNAs involved (see below). Interestingly, transfection studies in catfish and mouse plasmacytoma cell lines demonstrated that mouse B cells process normal catfish Mb-Igμ transcripts, i.e. VDJ–Cμ1–Cμ2–Cμ3–TM1–TM2 mRNA, while catfish B cells could not process mouse Mb-Igμ transcripts (Ross et al., 1998). These findings indicate that even though the core machinery is indeed partly conserved between mammals and fish, a specialization of both machinery and genomic motifs must have occurred during evolution.

## 3. Diversity of IgD: duplicated loci, multiple domains and (rare) alternative splicing

### 3.1. IgD forms vary in different vertebrates

IgD was first identified in humans and rodents and early on it was shown that IgM and IgD are co-produced through alternative splicing of a long pre-mRNA containing the VDJ region, the Cμ exons and the Cδ exons (Goding and Layton, 1976; Martin et al., 1976; Blattner and Tucker, 1984; Maki et al., 1981; Rowe and Fahey, 1965; Van Boxel et al., 1972; Abney and Parkhouse, 1974). Accordingly, IgD was found expressed mainly on naive mature B cells, which are $IgM^+/IgD^+$. Even so, it has been demonstrated that humans have a population of $IgD^+$-only B cells ($IgM^-/IgD^+$) predominantly associated with the upper respiratory tract mucosa that undergo class switch to express IgD by using a cryptic switch region found between the Cμ and Cδ genes (White et al., 1990; Owens et al., 1991; Kluin et al., 1995). Also, since IgD had not been found in birds or in other mammals, it was long considered to be a recent evolved Ig isotype, only expressed by primates and rodents. However, in the 1990s, IgD-like sequences were identified in teleosts (Wilson et al., 1997; Hordvik et al., 1999), but because these IgD molecules were predicted to lack a true hinge and were larger in size than mouse and human IgD, there was some debate if teleost IgD was a true ortholog of mammalian IgD. Notably, several lines of evidence have now established that IgD constitutes the second primordial isotype expressed in most gnathostomes (Ohta and Flajnik, 2006). First in 2002 and 2003, the discovery of IgD in cow, sheep, and pig, clearly demonstrated that IgD in mammals is not restricted to rodents and primates (Zhao et al., 2002, 2003). Second, long forms of IgD, which also lack a hinge have been described in *Xenopus* (Zhao et al., 2006) and in reptiles (Gambon-Deza and Espinel, 2008; Wei et al., 2009), which suggests that the IgD isotype is probably general to tetrapods, albeit it is apparently missing in birds and some mammalian species (e.g. rabbits (Lanning et al., 2003)). Third, more recently, a homolog of IgD with 10 Cδ domains and no hinge was identified in platypus, *Ornithorhynchus anatinus*, which confirmed

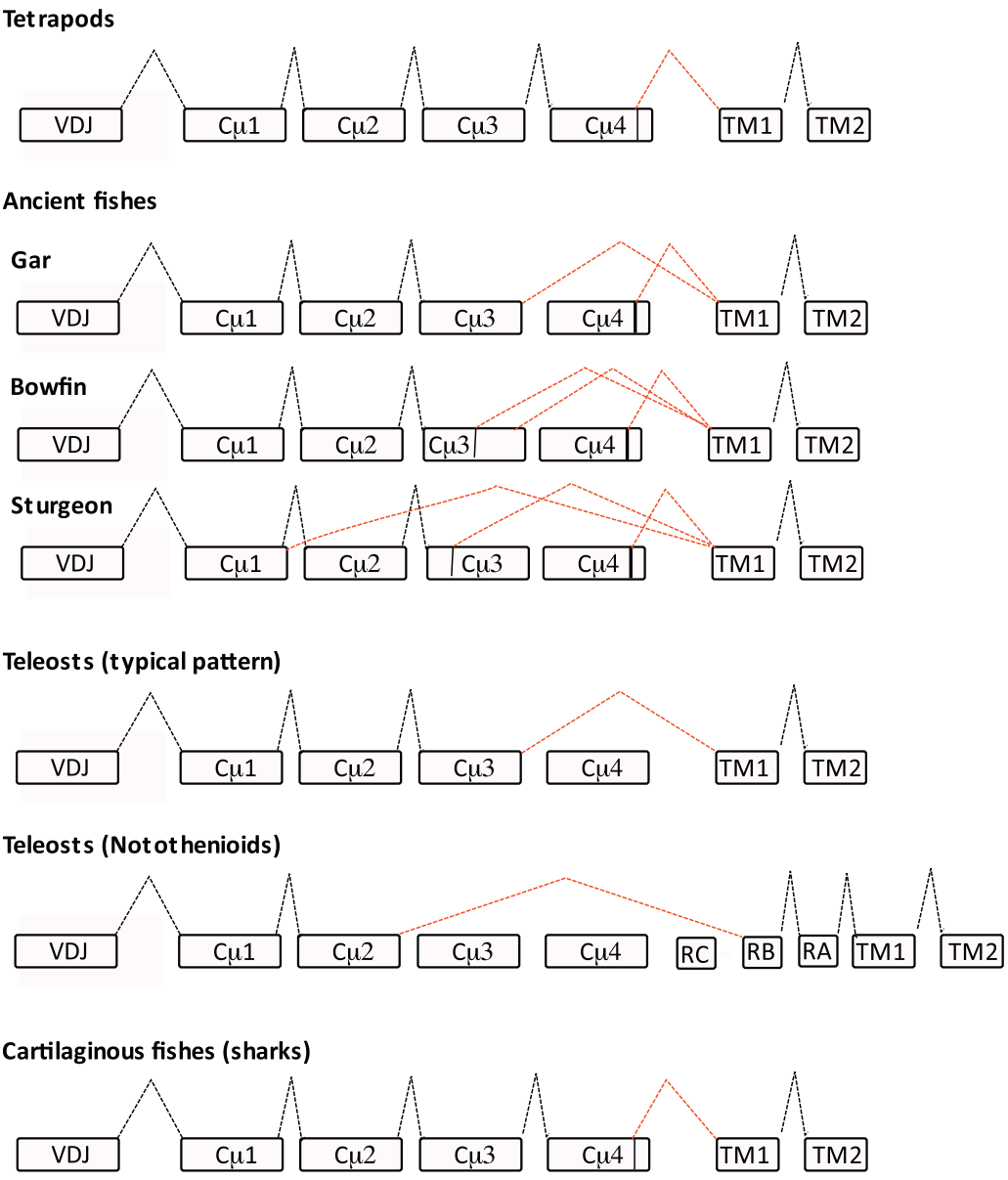


**Fig. 1.** Schematic representation of RNA splicing pathways for the transcripts of the Igμ genes in different fishes and in tetrapods. The transcripts for the heavy chain of the secreted forms of IgM contains the four Cμ domains in all vertebrates studied so far, and a polyA site located 3′ of the Cμ4 exon is used. The transcripts for the heavy chain of the membrane bound forms of IgM are spliced following diverse patterns represented in the figure.

that IgD was present in the common ancestor of mammals (Zhao et al., 2009). Finally, the demonstration that the genes encoding the IgW H chain of sharks and skates were orthologs of IgD H chains provided definitive evidence of the ancestral nature of the IgD isotype (Ohta and Flajnik, 2006).

Interestingly, while IgM exhibits a relatively stable structure in all gnathostomes, IgD shows a much higher plasticity (Ohta and Flajnik, 2006). Briefly, IgD H chains show important structural variations between the different lineages of vertebrates, including varying numbers of Cδ domains, the presence or absence of a hinge region, and the inclusion of the Cμ1 or a Cμ1-like domain as the first C domain (see below). Furthermore, in some species, the IgD H chains are also diversified by usage of multiple alternative splice sites in B cells, following different pathways. This is specifically demonstrated in the nurse shark, *Ginglymostoma cirratum*, and horn shark, *Heterodontus francisci*, where the alternative splicing of TM exons to the Cδ4 or Cδ2 exons leads to two distinct Mb-IgD (or Igw) transcripts. Additionally, other splicing pathways produce two forms of S-IgD transcripts. Here it is important to note that while it is well documented that the chondrychtian Ig isotype repertoire is enriched by alternative splicing of IgD-like, NAR, and even IgM pre-mRNAs (Rumfelt et al., 2004), much less is known concerning to what extent splicing diversification contributes to the Ig isotype repertoire in teleost fish.

### *3.2. Diversification of IgD heavy chains by alternative splicing and combinatory usage of Cδ exons is rare in teleosts*

The first teleost IgD sequence identified was from the channel catfish. It was originally identified as a cDNA consisting of a rear-

ranged VDJ spliced to the Cμ1 domain, followed by seven CHδ domains, some of which exhibited similarity to mammalian Igδ domains, a TM region, and a short charged cytoplasmic tail (Wilson et al., 1997). Reasons for classifying this cDNA as an Igδ transcript included the co-expression of Igδ and Igμ in some, but not all, catfish B cells, and the finding of the Igδ gene approximately 1.5 kb downstream of the Igμ gene in the catfish IgH locus (Martin et al., 1976; Hordvik et al., 1999). Moreover, the fact that catfish IgD was chimeric and that all membrane Igδ transcripts consistently included the first Cμ1 domain between the rearranged VDJ and the Cδ1 domain implied that catfish IgD was produced by an unusual pathway of RNA processing, a process similar to the alternative pathway of RNA processing of mammalian mature IgD transcripts. However, catfish IgD differed from eutherian IgD by the inclusion of the Cμ1 domain, the large number (seven) of Cδ domains, and by the absence of a hinge. The inclusion of seven or more Igδ domains and the absence of a hinge are also conserved features in amphibian, reptile, shark and platypus IgD. While the elongated stalk-like structures of these IgD forms are predicted to be un-flexible, it seems likely that these molecules could bend. In comparison, mammalian (i.e. eutherians) IgD molecules are shorter with very flexible hinge regions. Another unique feature concerning catfish IgD is that Mb-IgD and S-IgD forms are encoded by different Igδ genes separated by approximately 735 kb (Bengten et al., 2006). These two genes are almost identical, except for their terminal exons. The Mb-IgD is encoded by an Igδ gene termed IGHD1, which ends in a typical transmembrane exon and the S-IgD is encoded by an Igδ gene termed IGHD3, which ends in a secreted exon. Both genes are associated with functional enhancers (Bengten et al., 2002). A similar IGH organization also occurs in the pufferfish where the exons encoding secreted Igδ are found to be linked to a separate array of D and JH genes found 5′ of the canonical array of D and JH genes linked to the Igμ and membrane Igδ genes (Aparicio et al., 2002). Thus, catfish are not unique in their Igδ gene arrangement and in catfish and pufferfish membrane IgD and secreted IgD are produced through the specialization of duplicated loci and not by alternative splicing. Importantly, anti-IgD monoclonal antibody (mAb) specific for the secreted IgD tail exon readily identifies IgD in catfish serum and a second anti-IgD mAb specific for the Cδ2 domain detects membrane IgD on catfish B cells (Edholm et al., 2010).

IgD transcripts and/or genes have been identified in variety of teleosts, including Atlantic salmon, *Salmo salar*, and Atlantic halibut, *Hippoglossus hippoglossus* (Hordvik, 2002), Atlantic cod, *G. morhua* (Stenvik and Jorgensen, 2000), pufferfish (Aparicio et al., 2002), Japanese flounder, *Paralichthys olivaceus* (Srisapoome et al., 2004) and grass carp, *Ctenopharyngodon idella* (Xiao et al., 2010). To date all teleost functional Mb-IgD transcripts contain the Cμ1 exon making them chimeric. In Atlantic salmon, approximately 10% of the transcripts involving Cδ do not contain Cμ1, however these transcripts seem to be sterile or partially spliced transcripts without VH and/or JH sequence (Yasuike et al., 2010). The inclusion of the Cμ1 exon was originally hypothesized by Wilson et al. (1997) to be necessary since it permits the IgD H chain to covalently associate with catfish IgL chains and more recently through the use of anti-catfish IgD mAbs, catfish Mb-IgD was shown to associate with IgL chains (Edholm et al., 2010). This inclusion of a Cμ1 or Cμ1-like domain in IgD chains is not unique to the teleosts since the Cδ1 domains of cow, sheep and pig exhibit >93% amino acid identity to the respective Cμ1 domain in each of these species (Zhao et al., 2002, 2003). Moreover, in the pig, IgD transcripts containing the Cμ1 domain instead of the Cδ1 domain can also be found. In this regard, it is interesting that affinity binding studies of different mammalian Ig isotypes suggests that CH1 domains may be involved in defining antigen–antibody affinity (Pritsch et al., 1996; Adachi et al., 2003).

The IgD genes have been analyzed in detail in several teleost species and the typical arrangement consists of seven unique Igδ exons, with a repeated Cδ2–Cδ3–Cδ4 block, followed by either a single TM exon or two TM exons (Bengten et al., 2006, 2002; Hordvik, 2002). In Atlantic salmon two IgD loci were identified, with a common structure of Cδ1–(Cδ2–Cδ3–Cδ4)$_n$–Cδ5–Cδ6–Cδ7–TM1–TM2, and the Cδ2–Cδ3–Cδ4 block repeated three times (IgH-A) or four times (IgH-B) (Yasuike et al., 2010). While transcripts containing both TM1 and TM2 appeared to be the Mb-IgD form mainly expressed by Atlantic salmon, a splicing variant was identified where the TM2 exon was spliced directly to the Cδ6 exon using the regular splicing sites (Hordvik, 2002). This alternative splicing is reminiscent of the pathway which produces a second form of IgE in human (Batista et al., 1996). In the grass carp the Igδ genes also contain a repeated Cδ2–Cδ3–Cδ4 block, but like in the catfish, there in only a single TM exon and the IgD H chain consists of (δ1–(δ2–δ3–δ4)$_2$–δ5–δ6–δ7–TM) (Xiao et al., 2010). Also in the grass carp, the intron between Cδ4.1 and Cδ2.2 has been lost, and the intron between Cδ3.1 and Cδ4.1 is much shorter than the one between Cδ3.2 and Cδ4.2. In contrast, in the Atlantic cod the IgD locus seems to have been subjected to dramatic rearrangement events leading to the loss of the Cδ3, Cδ4, Cδ5, and Cδ6 exons and a tandem duplication of the Cδ1–Cδ2 region (Stenvik and Jorgensen, 2000). Also, downstream of the two TM exons a truncated Cδ7 exon is found and this remnant may be the result of another recombination event. The intron between the duplicated regions contains a short exon (δy), which exhibits some similarity to the junctional sequence found between the hydrophilic spacer and the hydrophobic region of the TM sequence. Full-length Atlantic cod IgD transcripts therefore consist of VDJ–Cμ1–Cδ1.1–Cδ2.1–δy–Cδ1.2–Cδ2.2–Cδ7–TM (Stenvik and Jorgensen, 2000). This IgD locus (and Mb-IgD splice form) is quite different from the typical IgD gene arrangement found in most teleosts studied to date, but such extensive rearrangement of the IgD locus may have occurred in other fish lineages. For example, in the Japanese flounder, no exon duplication was found in the IgD H chain locus (Srisapoome et al., 2004). In addition, a second "atypical" IgD locus is found in the pufferfish, in which the Cδ1–Cδ2–Cδ3–Cδ4–Cδ5–Cδ6 exons are duplicated instead of the Cδ2–Cδ3–Cδ4 exons. However, as previously shown for catfish, salmon, halibut and cod, pufferfish Mb-IgD transcripts also contain the Cμ1 exon and no alternative Mb-IgD transcripts with varying numbers of Igδ exons have been found (Saha et al., 2004). Here it should also be stated that S-IgD transcripts have not been found.

Thus, the variability/diversity of teleost IgD H chains sequenced to date appears to be mainly the result of gene duplications within or of the Igδ locus, rather than by specific RNA splicing pathways. Furthermore, the duplicated loci found in certain species can specialize in the production of the S-δ or Mb-δ, respectively, as in catfish. Overall, the number of Cδ exons also seems to be relatively stable in the different teleost species. Together, these observations suggest that the selective pressures leading to the different forms of IgW, NAR and IgM in sharks (and or skates) have not affected teleost IgD in the same way. Indeed, with so many Cδ exons included in teleost Mb-IgD and S-IgD transcripts combined with the occurrence of having very similar, or duplicated Cδ exons in transcripts – it would seem that the potential for IgD isotypic variation by alternative splicing would be very high. However, whether the current IgD data truly reflects a strict regulation of splicing of pre-mRNA transcribed from multiple duplicated loci can only be determined with more detailed analyses of different teleost B cell transcriptomes. In the future, it would also not only be of interest to examine the diversity of IgD transcripts in holosteans and chondrosteans, but to revisit or determine if teleosts other than the giant grouper,

*Epinephelus itaira*, express monomeric or (Fab)2-like IgM forms (Clem and McLean, 1975).

The third isotype identified in teleost is termed IgT/IgZ and recent studies demonstrate that it is important in mucosal responses against parasites (Zhang et al., 2010). In zebrafish and salmonids, IgT/Z transcripts appear to follow the Ig splicing rule since they consist of VDJ–Cτ/ζ1–Cτ/ζ2–Cτ/ζ3–Cτ/ζ4 – a TM or secreted tail (Hu et al., 2010; Danilova et al., 2005; Hansen et al., 2005; Tadiso et al., 2011). However, in other teleost species IgT/Z gene structure and organization can vary. For example, in fugu, the IgT/Z CH region consist of Cτ/ζ1 and Cτ/ζ4 domains separated by a hinge, followed by a TM region (Savan et al., 2005a) and in the common carp two Mb-IgT/Z subclasses have been described (Savan et al., 2005b). Common carp IgZ1 transcript resembles zebrafish and rainbow trout IgT/Z transcripts, while IgZ2 transcripts consists of VDJ–Cμ1–Cτ/ζ4 – followed by a TM or secreted tail (Ryo et al., 2010). Since sequence analysis suggests that these two subclasses are encoded by different loci, it appears that the chimeric IgZ2 subclass is encoded by a rearranged duplicated loci rather than a product of alternative splicing. Therefore, it seems likely that the true picture for this Ig isotype, as for IgD, will likely become more complicated as more Ig sequences from different teleost become available and more duplicated loci are discovered (Yasuike et al., 2010).

## 4. NMD and abundance of sterile transcripts in fish

### *4.1. The NMD pathway and the elimination of Ig and TCR transcripts with premature stop codons in mammals*

Premature stop codons in ORF resulting in the production of truncated polypeptides may interfere with the protein functions, e.g. by forming dominant negative variants. To cope with this problem, the mutant mRNAs are quickly degraded through a pathway known as nonsense mediated mRNA decay (NMD) (Wilkinson, 2005; Chang et al., 2007a). The NMD involves a number of factors and is also regulated by cis-acting sequences. The first signal for NMD is a stop codon, but the process needs a second signal to recognize the stop codon as premature and to trigger RNA degradation. After splicing, a complex of proteins – the exon junction complex (EJC) – deposited at the exon–exon junction site, serves as the second signal required for NMD when it is found downstream of a stop codon. In mammals, the true STOP codon is generally located in the last exon and the translation complex therefore does not meet any EJC downstream. Premature stop codons located between the last exon/exon junction and the true STOP codon are therefore immune to NMD. In addition, due to its size the ribosome reaches the EJC when it reads the codon located 55 nt upstream of the exon/exon junction, therefore the complex cannot detect stop codons located 55 nt upstream the exon/exon junction as premature.

T-cell receptor (TCR) and Ig transcripts commonly harbor premature nonsense codons due to out of frame VDJ rearrangements and constitute prime targets of efficient NMD mechanisms. Indeed, in mammals these premature nonsense codon-bearing TCR (and Ig) transcripts are efficiently targeted to protect lymphocytes from the potentially adverse effects of the truncated receptors that would otherwise be produced: while theoretically 2/3 of VDJ junctions are out of frame, less than 5% are observed among cDNA sequences (Lozano et al., 1993). The 55 nt rule explained above is generally followed by all transcripts except those for TCRs (and most probably immunoglobulins). In TCR transcripts, nonsense codons have a "polar" effect on the strength of the NMD response that is determined by the distance to the last exon–exon junction rather than by a precise boundary position: the most 5′ premature stop codons trigger a more robust response than the most 3′ (Wang et al., 2002). These observations indicated that the NMD pathway has probably several branches with variable efficiencies depending on the transcript class.

In contrast to the splicing machinery, the NMD pathway is not based on RNPs and various proteins have been demonstrated to be essential for NMD. For example, the UPF1, UPF2, and UPF3 proteins form the core NMD machinery, while SMG-1, SMG-5, SMG-6, and SMG-7 proteins mediate the phosphorylation and dephosphorylation cycle of UPF1 (reviewed in Chang et al. (2007a)). The EJC components have also been identified, and the EJC core consists of four proteins: Y14, MAGOH, eIF4AIII, and MLN51 (also known as BTZ or CASC3; reviewed in Chang et al. (2007a); or for a complete list, see Table 1). Other levels of regulation are also critical for the control of mRNA containing premature termination codons (PTCs). For example, a post-transcriptional switch tightly regulates UPF3A levels and the balance between UPF3A and UPF3B regulates the efficiency of the NMD process (Chan et al., 2009). Furthermore, a second pathway – distinct from the cytoplasmic NMD phenomenon – is also involved in the control of mouse TCRβ mRNAs containing PTCs (Bhalla et al., 2009). This pathway results in nonsense TCRβ mRNA being retained in the nucleus, because of a dramatic partitioning shift in the nuclear to cytoplasmic nonsense TCRβ transcripts ratio, this response is termed the nonsense codon-induced partitioning shift (NIPS). Together, NIPS and NMD drastically reduce the levels of transcripts with premature stop codons in the cytoplasm. In human cells, NIPS was not exclusively observed just in TCRβ transcripts, but was found in various mRNAs, including β-globin and triose-phosphate isomerase (Chang et al., 2007b). As with NMD, the NIPS response involves factors UPF1 and eIF4AIII factors. Finally, a third control of nonsense mutations is the up-regulation of alternatively spliced transcripts that eliminates the PTC containing exon. In fact, nonsense mediated alternative splicing (NAS) is efficiently triggered when nonsense mutations disrupting the reading frame affects the exonic splicing enhancers (ESEs) (Chang et al., 2007b).

In summary, the decay of TCR and Ig transcripts containing premature stop codons is operated by several pathways (NMD, NIPS and NAS), which cooperate and use some common factors. It is also influenced by cis-acting sequences present in the transcript itself.

### *4.2. High frequency of sterile transcripts of Ig/TCR in fish is due to out of frame junctions. Is the fish NMD system altered or differently regulated?*

Surprisingly, the amount of Ig and TCR sterile transcripts in fish seems to be significantly higher than the one observed in mammals. While the proportion of transcripts with premature nonsense codons due to an out-of-frame junction is around 2–3% in human (Jores and Meo, 1993) as well as in the mouse, it is often as high as 20–30% in fish (Boudinot et al., 2001; Charlemagne et al., 1998; PB unpublished data). For example, 20–30% of rainbow trout TCRβ transcripts using Vβ4 and Jβ1 segments were found out of frame in spleen of naïve adult animals (Boudinot et al., 2001, 2004). This proportion seems to be much lower among strongly selected T cell subsets post-viral infection (Boudinot et al., 2001), suggesting that the regulation mechanisms involved might be induced by the lymphocyte activation state. While the high proportion of TCR sterile transcripts has been observed for all tested fish Vβ rearrangements, the proportion of sterile Mb-Igμ transcripts seems to be variable depending on the VH segment used, suggesting that the efficiency of NMD (or other mechanisms involved in PTC recognition/control) may be different for Ig versus TCR transcripts.

Similarly, studies in the axolotl, *Ambystoma mexicanum*, also showed that NMD efficiency for Ig and TCR transcripts is much lower than what is observed for mammalian Ig and TCR transcripts. Approximately 30% of axolotl TCRβ junctions were found to be

**Table 1**
Zebrafish and stickleback orthologues of the main splice/NMD factors described in humans and mice (from Ensembl release 58, zebrafish Zv8 and stickleback BroadS1 genome assembly).

| *Homo sapiens* | | | | *Danio rerio* | | *Gasterosteus aculeatus* | |
|---|---|---|---|---|---|---|---|
| Gene | Function[a] | ID | Location | ID | Location | ID | Location |
| ACINUS-ACIN1 | EJC | NP_055792.1 | 14q11.2 | ENSDARG00000054290 | Chr2:36,399,241-36,373,765 | ENSGACG00000014063 | GroupIII-2,804,742-2,816,630 |
| | | | | ENSDARG00000029313 | Chr2: 36,389,154-36,399,494 | | |
| | | | | ENSDARG00000026842 | Chr7:24,123,858-24,148,043 | ENSGACG00000019196 | GroupVII-2,931,416-2,945,152 |
| MAGOH | NMD, EJC | NP_002361.1 | 1p34-p33 | ENSDARG00000038635 | Chr8:21,515,227-21,515,915 | ENSGACG00000003869 | GroupVIII-2,120,287-2,119,007 |
| | | | | Unassigned copy | | | |
| PININ-PNN | EJC | BAF85268.1 | 14q21.1 | ENSDARG00000015851 | Chr17:11,249,791-11,258,822 | ENSGACG00000012434 | GroupXV:13,042,671-13,046,138 |
| | | | | ENSDARG00000043485 | Chr20:12,621,069-12,630,870 | | |
| SFRS18[b] | | NP_056306.1 | 6q16.3 | ENSDARG00000069855 | Chr16:31,477,853-31,485,868 | ENSGACG00000004989 | GroupXX: 3,265,741-3,273,201 |
| SFRS16[b] | | NP_008987.2 | 19q13.3 | ENSDARG00000061742 | Chr18:37,074,308-37,104,039 | ENSGACG00000009826 | GroupI: 9,749,862-9,760,409 |
| C1orf6[b] | | NP_064713.3 | 1p36.13-p35.1 | ENSDARG0000003044 | Chr13:44,843,051-44,849,896 | ENSGACG00000007326 | GroupXV: 4,313,005-4,314,429 |
| REF-ALY-THOC4 | EJC | NP_005773.3 | 17q25.3 | ENSDARG00000077732 | Chr3:52,819,615-52,825,868 | ENSGACG00000013328 | GroupXI:12,858,726-12,856,280 |
| SAP18 | EJC | NP_005861.2 | 13q12.11 | ENSDARG00000057854 | Chr9:20,622,009-20,620,804 | ENSGACG00000014077 | Group20,840,545-20,841,159 |
| SRm160-SRRM1 | EJC | CAH73089.1 | 1p36.11 | ENSDARG00000001244 | Chr17:23288522-23275871 | ENSGACG00000007266 | GroupXV: 4,162,657-4,170,605 |
| UAP56-BAT1 | EJC | AAP36788.1 | 6p21.3 | ENSDARG00000036069 | Chr19:26490024-26478266 | NF | |
| DDX39[c] | | NM_005804 | 19p13.12 | ENSDARG00000006225 | Chr1:42750400-42745770 | ENSGACG00000019335 | GroupIX:16,546,572-16,549,442 |
| | | | | ENSDARG00000015111 | Chr21:12750314-12768981 | ENSGACG00000006096 | GroupXI:2,607,467-2,605,426 |
| DDX20[c] | | NP_009135.3 | 1p21.1-p13.2 | ENSDARG00000061204 | Chr 8:29010991-29027531 | NF | |
| UPF2 | NMD,EJC | NP_056357.1 | 10p14-p13 | ENSDARG00000074658 | Chr: not assigned | ENSGACG00000001080 | Group: not assigned |
| Y14-RBM8A | NMD,EJC | EAW71418 | 1q22 | ENSDARG00000016516 | Chr16:43177553-43170582 | ENSGACG00000013157 | GroupXX:15,702,803-15,704,881 |
| SMG-5 | NMD | BAA83041.2 | 1q21.2 | ENSDARG00000041481 | Chr16:43,402,835-43,373,296 | ENSGACG00000013142 | GroupXX:15,623,279-15,633,639 |
| SMG-6 | NMD | NP_060045.4 | 17p13.3 | ENSDARG00000075957 | Chr10:33,893,609-33,868,333 | ENSGACG00000020521 | GroupVII:19,484,133-19,476,057 |
| | | | | ENSDARG00000075921 | Chr10-34282012-34298894 | | |
| SMG-7 | NMD | NP_775179.1 | 1q25 | ENSDARG00000060767 | Chr2:34,095,278-34,118,081 | ENSGACG00000001634 | Group: not assigned |
| eIF4AIII | NMD, EJC | 2HYIC | 17q25.3 | ENSDARG00000014472 | Chr11:43,879,541-43,896,018 | ENSGACG00000020004 | GroupIX: 20,154,669-20,158,059 |
| MLN51-CASC3 | NMD, EJC | NP_031385.2 | 17q11-q21.3 | ENSDARG00000029911 | Chr3:20,445,734-20,437,098 | ENSGACG00000015032 | GroupXI:16,704,669-16,705,192 |
| PYM-WIBG | NMD, EJC | NP_115721.1 | 12q13.2 | ENSDARG00000046024 | Chr8:704,851-704,532 | ENSGACG00000010703 | GroupXII:13,619,159-13,619,281 |
| RNPS1 | NMD, EJC | NM_080594.1 | 16p13.3 | ENSDARG00000015853 | Chr3:7,963,064-7,963,240 | ENSGACG00000005641 | GroupV:7,047,733-7,047,918 |
| | | | | | | ENSGACG00000011261 | GroupXI:9,641,414-9,644,697 |
| SMG-1 | NMD | NP_055907.3 | 16p12.3 | ENSDARG00000054570 | Chr3-28,540,592-28,508,976 | ENSGACG00000004970 | GroupXI:572,306-613,886 |
| | | | | ENSDARG00000079212 | Chr3:28,505,367-28,508,818 | | |
| MTOR | | NP_004949.1 | 1p36.2 | ENSDARG00000053196 | Chr8: 46,508,711-46,752,918 | ENSGACG00000004869 | GroupXII:5,335,336-5,401,047 |

Table 1 (*Continued*)

| *Homo sapiens* | | | | *Danio rerio* | | *Gasterosteus aculeatus* | |
|---|---|---|---|---|---|---|---|
| Gene | Function[a] | ID | Location | ID | Location | ID | Location |
| TAP-NXF1 | EJC | NP_006353.2 | 11q12-q13 | ENSDARG00000055076 | Chr21-28,979,765-28,970,533 | ENSGACG00000016934 | GroupIV-5,242,302-5,248,974 |
| | | | | ENSDARG00000073829 | Chr14-9,305,238-9,234,555 | | |
| | | | | ENSDARG00000078064 | Chr14:9,234,232-9,268,192 | | |
| UPF1 | NMD | BAA19664.2 | 19p13.2-13.11 | ENSDARG00000016302 | Chr2-55,618,866-55,678,956 | ENSGACG00000015504 | GroupIII-8,302,996-8,294,138 |
| | | | | | | ENSGACG00000008078 | GroupVIII-9,327,826-9,318,387 |
| AQR | | NP_055506.1 | 15q14 | ENSDARG00000016775 | Chr17:47,340,862-47,358,381 | ENSGACG00000010437 | GroupXV:9,048,147-9,086,994 |
| ZNFX1 | | NP_066363.1 | 20q13.13 | ENSDARG00000074028 | Chr 6: 59,827,808-59,885,975 | ENSGACG00000006038 | GroupXVII:4531155-4544248 |
| UPF3a | NMD, EJC | NP_075387.1 | 13q34 | ENSDARG00000069297 | Chr9-35229316-35219721 | ENSGACG00000001972 | GroupXVI-3202482-3204459 |
| UPF3b | NMD, EJC | AAI21018.1 | Xq25-q26 | ENSDARG00000000489 | Chr14-34697643-34687954 | ENSGACG00000017435 | GroupIV-6674465-6676457 |

NF, not found.

[a] As defined by Chang et al. (2007a).

[b] Paralog of PININ as such potentially involved in EJC.

[c] Paralog of UAP56-BAT1 as such potentially involved in EJC.

defective at all stages of development (Golub et al., 1997). Also, approximately 25% of the Igμ junctions were out of frame in three and a half month old animals, while most rearrangements were in frame in 10 and 24 month-old animals, showing that the control of the NMD can change during the development and is different in B and T lymphocytes (Kerfourn et al., 1996). These findings suggest that NMD efficiency is much lower for Ig and TCR transcripts in ectothermic vertebrates than it is in human and mouse, with the effect being more pronounced for TCR than Ig transcripts.

Together, the above observations may be explained by the particularities of the splicing/NMD machinery in fish and amphibians. Since teleosts have been subjected to one or more rounds of global genome duplication(s) during their evolution, the repertoire of splicing factors may be more diverse than that found in mammals. Moreover, a recent comprehensive study shows that although U12 introns are generally well conserved in vertebrates, there is already a significant divergence between fish and eutherians (Lin et al., 2010). Thus, to gain an insight into the splicing/NMD machinery in fish, we searched the zebrafish and stickleback databases (Ensembl assemblies, Stickleback v1.0 and zebrafish Zv8, release 58) for orthologs of the main splice/NMD factors that have been identified in humans and mice (Table 1). At first glance, the resulting preliminary list suggests that the genes involved in splicing and/or NMD are generally highly conserved and are found as single copies in the genome. However, there are two or three copies of a few genes, e.g. ACIN1, PNN, DDX39, SMG-6, SMG-1, NXF1 and UPF1 but there is no clear amplification/diversification of these genes. Here it should be noted that most of the duplications are found in the zebrafish and are not retrieved in the stickleback. Therefore, because the ancestral fish genome has been subjected to global genome duplication(s), it appears that the duplicated copies have been generally eliminated to maintain a canonical machinery.

Since the spliceosome is a RNP-based machinery, snRNAs are probably also involved in the NMD, either directly or indirectly. The snRNAs involved in the splicing machinery are encoded at several sites in the human genome. For example, the snRNA U2 locus represents arrays of repeat with one gene in each unit and present in variable numbers. It is therefore rather difficult to get a precise idea of the number of genes encoding the snRNAs from the fish genome sequences available. However, while a number of very close counterparts can be found for snRNAs U1, U2, U6, it seems that close homologs of the snRNAs U4, U5, U12, are missing in both zebrafish and stickleback.[2] This observation needs to be confirmed, but may provide a starting point to dissect the differences between the mammalian and teleost splicing/NMD machineries.

## 5. Conclusion

The structure of the IgM H chain is very conserved among most vertebrates and consists of four Cμ domains in both Mb-IgM and S-IgM forms. One exception is found in most teleosts where the splicing pattern produces Mb-IgM with three Cμ domains. The frequency of unusual or alternative splicing of Igμ transcripts does not seem to follow the variations of alternative splicing observed in the global transcriptome of diverse fish species: the average frequency of alternative splicing is rather low in the large zebrafish genome which contains many duplicated regions and genes and very high in the compact pufferfish genome (Lu et al., 2010), but the same IgμM versus IgμS splicing pathway has been found in both species. Only a few species make an exception to this rule: (1) in different lineages of ancient fishes, Mb-IgM chains may contain different numbers of Cμ domains due to diverse splicing pathways. Such diversity of Mb-IgM forms could constitute an additional level of regulation of their effector functions, i.e. each H chain could play a definite role in antigen capture and presentation, in complement activation and in B cell maturation; (2) in Notothenioids, two new exons located between Cμ4 and TM1 have been

[2] In their detailed study showing the coregulation of different spliceosome components in zebrafish, Trede et al. show Northern blot experiments using probes specific to U4 and U6. The corresponding zebrafish sequences are not available, but the probes detected signals with all expected characteristics (Trede, N.S., Medenbach, J., Damianov, A., Hung, L.H., Weber, G.J., Paw, B.H., et al., 2007. Network of coregulated spliceosome components revealed by zebrafish mutant in recycling factor p110. Proc. Natl. Acad. Sci. U.S.A. 104, 6608–6613). The lack of homologous sequences in the current assembly may therefore be due to the incompleteness of the genomic data.

acquired by the locus and are integrated into the Mb-Igμ transcript. This diversity of splicing of Mb-IgM transcripts may represent a response to particular selection pressures towards isotypic diversification. However, the gain of exons observed in Notothenioids is not necessarily due to of an adaptive phenomenon to the cold water of the Antarctic ocean, but may also be a consequence of genomic events leading to modifications of the splicing machinery. Such an explanation is probably involved – at least in part – in the pattern observed in the ancient fish lineages, holosteans and chondrosteans. Thus, the global duplication which occurred in the beginning of fish evolution as well as the later one(s) – for example in sturgeons – may explain the modifications of the splicing mechanisms. How did this diversification of splicing pathways occur in ancient fishes and lead to the CH3 → TM1 pattern present in all teleosts? A frequent explanation for switch from constitutive to alternative splicing seems to be the relaxation of the 5′ end of an exon, followed by fixation of an exonic splicing regulatory sequence (Lev-Maor et al., 2007), a motif that is conserved from fish to mammals. It would be interesting to test this type of hypothesis using Igμ transcripts of ancient fishes, but the genomic sequences of the Ig μ/δ loci required for such an analysis are not yet available for holosteans, chondrosteans and other ancient fish lineages. In addition, a different evolution of the repertoires of splicing factors and Short RNAs participating in the spliceosome may have also contributed to the diversification of the splicing patterns.

While diverse Mb-Igμ transcripts are observed in certain fish species due to unusual splicing pathways, it appears that it is not in general the case for the Igδ transcripts. However, the structure of the IgD H chain is highly variable between the different groups of vertebrates (in the number of Cδ domains and presence or absence of hinge) due to various patterns of partial duplication of the IgD locus during evolution. In this context suggesting a selection pressure towards diversification of IgD, it was somewhat surprising not to observe an additional diversity of Igδ transcripts due to alternative splicing in fishes. Basically, the available data indicate that only one membrane bound and one secreted IgD H chain need to be produced in each fish species, of which the structure is adapted specifically to each teleost group. This also appears to hold true in other vertebrates. The only exception to this has been observed in sharks, skates and lungfish where the constant part of IgD-like IgW H chain is diversified through alternative splicing (Ohta and Flajnik, 2006).

While the general characteristics of splicing regulation in a genome may be determined by its global and structural properties – such as length and degree of duplication – the Ig splicing appears to have evolved under a strong selective constraint linked to the isotype functional specialization. This evolution led to very different patterns for each isotype: the IgM have a well conserved structure among vertebrates, while the IgD was subjected to lineage specific diversification. The different selection pressures exerted on IgD and IgM remains to be precisely understood, but the diverse splicing patterns appear as an additional pathway which may give rise to further isotypic diversity from the genetic potential defined by the evolution of the locus.

Additionally, the processing of the Ig transcripts meets another level of complexity due to the V–D–J recombination which generates a large amount of RNA containing premature STOP codons. The mechanisms leading to the degradation of these “sterile” transcripts are tightly connected to the splicing machinery. Therefore, the regulation of the isotypic diversity by splicing probably exerts significant constraints on the evolution of the NMD and related pathways. The particular features of the fish NMD and splicing for Ig transcripts thus provide an extraordinary model to dissect both the plasticity and the connection of these processes in their genomic and evolutionary context.

## Acknowledgements

We thank Louis du Pasquier for discussions and critical reading of the manuscript. This work was supported by Institut National de la Recherche Agronomique.

## References

Abney, E.R., Parkhouse, R.M., 1974. Candidate for immunoglobulin D present on murine B lymphocytes. Nature 252, 600–602.
Adachi, M., Kurihara, Y., Nojima, H., Takeda-Shitaka, M., Kamiya, K., Umeyama, H., 2003. Interaction between the antigen and antibody is controlled by the constant domains: normal mode dynamics of the HEL-HyHEL-10 complex. Protein Sci. 12, 2125–2131.
Aparicio, S., Chapman, J., Stupka, E., Putnam, N., Chia, J.M., Dehal, P., et al., 2002. Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. Science 297, 1301–1310.
Barash, Y., Calarco, J.A., Gao, W., Pan, Q., Wang, X., Shai, O., et al., 2010. Deciphering the splicing code. Nature 465, 53–59.
Batista, F.D., Efremov, D.G., Burrone, O.R., 1996. Characterization of a second secreted IgE isoform and identification of an asymmetric pathway of IgE assembly. Proc. Natl. Acad. Sci. U.S.A. 93, 3399–3404.
Bengten, E., Leanderson, T., Pilstrom, L., 1991. Immunoglobulin heavy chain cDNA from the teleost Atlantic cod (*Gadus morhua* L.): nucleotide sequences of secretory and membrane form show an unusual splicing pattern. Eur. J. Immunol. 21, 3027–3033.
Bengten, E., Quiniou, S.M., Stuge, T.B., Katagiri, T., Miller, N.W., Clem, L.W., et al., 2002. The IgH locus of the channel catfish, *Ictalurus punctatus*, contains multiple constant region gene sequences: different genes encode heavy chains of membrane and secreted IgD. J. Immunol. 169, 2488–2497.
Bengten, E., Quiniou, S., Hikima, J., Waldbieser, G., Warr, G.W., Miller, N.W., et al., 2006. Structure of the catfish IGH locus: analysis of the region including the single functional IGHM gene. Immunogenetics 58, 831–844.
Bhalla, A.D., Gudikote, J.P., Wang, J., Chan, W.K., Chang, Y.F., Olivas, O.R., et al., 2009. Nonsense codons trigger an RNA partitioning shift. J. Biol. Chem. 284, 4062–4072.
Blattner, F.R., Tucker, P.W., 1984. The molecular biology of immunoglobulin D. Nature 307, 417–422.
Boudinot, P., Boubekeur, S., Benmansour, A., 2001. Rhabdovirus infection induces public and private T cell responses in teleost fish. J. Immunol. 167, 6202–6209.
Boudinot, P., Bernard, D., Boubekeur, S., Thoulouze, M.I., Bremont, M., Benmansour, A., 2004. The glycoprotein of a fish rhabdovirus profiles the virus-specific T-cell repertoire in rainbow trout. J. Gen. Virol. 85, 3099–3108.
Brites, D., McTaggart, S., Morris, K., Anderson, J., Thomas, K., Colson, I., et al., 2008. The Dscam homologue of the crustacean Daphnia is diversified by alternative splicing like in insects. Mol. Biol. Evol. 25, 1429–1439.
Campbell, K.S., Backstrom, B.T., Tiefenthaler, G., Palmer, E., 1994. CART: a conserved antigen receptor transmembrane motif. Semin. Immunol. 6, 393–410.
Chan, W.K., Bhalla, A.D., Le Hir, H., Nguyen, L.S., Huang, L., Gecz, J., et al., 2009. A UPF3-mediated regulatory switch that maintains RNA surveillance. Nat. Struct. Mol. Biol. 16, 747–753.
Chang, Y.F., Imam, J.S., Wilkinson, M.F., 2007a. The nonsense-mediated decay RNA surveillance pathway. Annu. Rev. Biochem. 76, 51–74.
Chang, Y.F., Chan, W.K., Imam, J.S., Wilkinson, M.F., 2007b. Alternatively spliced T-cell receptor transcripts are up-regulated in response to disruption of either splicing elements or reading frame. J. Biol. Chem. 282, 29738–29747.
Charlemagne, J., Fellah, J.S., De Guerra, A., Kerfourn, F., Partula, S., 1998. T-cell receptors in ectothermic vertebrates. Immunol. Rev. 166, 87–102.
Clem, L.W., McLean, W.E., 1975. Phylogeny of immunoglobulin structure and function. VII. Monomeric and tetrameric immunoglobulins of the margate, a marine teleost fish. Immunology 29, 791–799.
Coscia, M.R., Varriale, S., De Santi, C., Giacomelli, S., Oreste, U., 2010. Evolution of the Antarctic teleost immunoglobulin heavy chain gene. Mol. Phylogenet. Evol. 55, 226–233.
Danilova, N., Bussmann, J., Jekosch, K., Steiner, L.A., 2005. The immunoglobulin heavy-chain locus in zebrafish: identification and expression of a previously unknown isotype, immunoglobulin Z. Nat. Immunol. 6, 295–302.
Dong, Y., Taylor, H.E., Dimopoulos, G., 2006. AgDscam, a hypervariable immunoglobulin domain-containing receptor of the Anopheles gambiae innate immune system. PLoS Biol. 4, e229.
Edholm, E.S., Bengten, E., Stafford, J.L., Sahoo, M., Taylor, E.B., Miller, N.W., et al., 2010. Identification of two IgD+ B cell populations in channel catfish, *Ictalurus punctatus*. J. Immunol. 185, 4082–4094.
Edwalds-Gilbert, G., Veraldi, K.L., Milcarek, C., 1997. Alternative poly(A) site selection in complex transcription units: means to an end? Nucleic Acids Res. 25, 2547–2561.
Gambon-Deza, F., Espinel, C.S., 2008. IgD in the reptile leopard gecko. Mol. Immunol. 45, 3470–3476.
Goding, J.W., Layton, J.E., 1976. Antigen-induced co-capping of IgM and IgD-like receptors on murine B cells. J. Exp. Med. 144, 852–857.
Golub, R., Fellah, J.S., Charlemagne, J., 1997. Structure and diversity of the heavy chain VDJ junctions in the developing Mexican axolotl. Immunogenetics 46, 402–409.

Guselnikov, S.V., Najakshin, A.M., Taranin, A.V., 2003. *Fugu rubripes* possesses genes for the entire set of the ITAM-bearing transmembrane signal subunits. Immunogenetics 55, 472–479.
Hansen, J.D., Landis, E.D., Phillips, R.B., 2005. Discovery of a unique Ig heavy-chain isotype (IgT) in rainbow trout: implications for a distinctive B cell developmental pathway in teleost fish. Proc. Natl. Acad. Sci. U.S.A. 102, 6919–6924.
Hordvik, I., 2002. Identification of a novel immunoglobulin delta transcript and comparative analysis of the genes encoding IgD in Atlantic salmon and Atlantic halibut. Mol. Immunol. 39, 85–91.
Hordvik, I., Thevarajan, J., Samdal, I., Bastani, N., Krossoy, B., 1999. Molecular cloning and phylogenetic analysis of the Atlantic salmon immunoglobulin D gene. Scand. J. Immunol. 50, 202–210.
Hu, Y.L., Xiang, L.X., Shao, J.Z., 2010. Identification and characterization of a novel immunoglobulin Z isotype in zebrafish: implications for a distinct B cell receptor in lower vertebrates. Mol. Immunol. 47, 738–746.
Hu, Y.L., Zhu, L.Y., Xiang, L.X., Shao, J.Z., 2011. Discovery of an unusual alternative splicing pathway of the immunoglobulin heavy chain in a teleost fish, *Danio rerio*. Dev. Comp. Immunol. 35, 253–257.
Jaillon, O., Aury, J.M., Brunet, F., Petit, J.L., Stange-Thomann, N., Mauceli, E., et al., 2004. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. Nature 431, 946–957.
Janvier, P., Lecointre, G., Le Guyader, H., 1996. Early Vertebrates. Oxford University Press, London.
Jores, R., Meo, T., 1993. Few V gene segments dominate the T cell receptor beta-chain repertoire of the human thymus. J. Immunol. 151, 6110–6122.
Kerfourn, F., Charlemagne, J., Fellah, J.S., 1996. The structure, rearrangement, and ontogenic expression of DB and JB gene segments of the Mexican axolotl T-cell antigen receptor beta chain (TCRB). Immunogenetics 44, 275–285.
Kluin, P.M., Kayano, H., Zani, V.J., Kluin-Nelemans, H.C., Tucker, P.W., Satterwhite, E., et al., 1995. IgD class switching: identification of a novel recombination site in neoplastic and normal B cells. Eur. J. Immunol. 25, 3504–3508.
Lanning, D.K., Zhai, S.K., Knight, K.L., 2003. Analysis of the 3′ Cmu region of the rabbit Ig heavy chain locus. Gene 309, 135–144.
Lee, M.A., Bengten, E., Daggfeldt, A., Rytting, A.S., Pilstrom, L., 1993. Characterisation of rainbow trout cDNAs encoding a secreted and membrane-bound Ig heavy chain and the genomic intron upstream of the first constant exon. Mol. Immunol. 30, 641–648.
Lev-Maor, G., Goren, A., Sela, N., Kim, E., Keren, H., Doron-Faigenboim, A., et al., 2007. The "alternative" choice of constitutive exons throughout evolution. PLoS Genet. 3, e203.
Lin, C.F., Mount, S.M., Jarmolowski, A., Makalowski, W., 2010. Evolutionary dynamics of U12-type spliceosomal introns. BMC Evol. Biol. 10, 47.
Lozano, F., Rada, C., Jarvis, J.M., Milstein, C., 1993. Affinity maturation leads to differential expression of multiple copies of a kappa light-chain transgene. Nature 363, 271–273.
Lu, J., Peatman, E., Wang, W., Yang, Q., Abernathy, J., Wang, S., et al., 2010. Alternative splicing in teleost fish genomes: same-species and cross-species analysis and comparisons. Mol. Genet. Genomics 283, 531–539.
Lundqvist, M., Stromberg, S., Bouchenot, C., Pilstrom, L., Boudinot, P., 2009. Diverse splicing pathways of the membrane IgHM pre-mRNA in a Chondrostean, the Siberian sturgeon. Dev. Comp. Immunol. 33, 507–515.
Maki, R., Roeder, W., Traunecker, A., Sidman, C., Wabl, M., Raschke, W., et al., 1981. The role of DNA rearrangement and alternative RNA processing in the expression of immunoglobulin delta genes. Cell 24, 353–365.
Martin, L.N., Leslie, G.A., Hindes, R., 1976. Lymphocyte surface IgD and IgM in non-human primates. Int. Arch. Allergy Appl. Immunol. 51, 320–329.
Max, E., 2009. Immunoglobulins: molecular genetics. In: Paul, W. (Ed.), Fundamental Immunology. Lippincott, New York.
Ohta, Y., Flajnik, M., 2006. IgD, like IgM, is a primordial immunoglobulin class perpetuated in most jawed vertebrates. Proc. Natl. Acad. Sci. U.S.A. 103, 10723–10728.
Ota, T., Nguyen, T.A., Huang, E., Detrich 3rd., H.W., Amemiya, C.T., 2003. Positive Darwinian selection operating on the immunoglobulin heavy chain of Antarctic fishes. J. Exp. Zool. B: Mol. Dev. Evol. 295, 45–58.
Owens Jr., J.D., Finkelman, F.D., Mountz, J.D., Mushinski, J.F., 1991. Nonhomologous recombination at sites within the mouse JH-C delta locus accompanies C mu deletion and switch to immunoglobulin D secretion. Mol. Cell. Biol. 11, 5660–5670.
Peterson, M.L., 1992. Balanced efficiencies of splicing and cleavage-polyadenylation are required for mu-s and mu-m mRNA regulation. Gene Expression 2, 319–327.
Pritsch, O., Hudry-Clergeon, G., Buckle, M., Petillot, Y., Bouvet, J.P., Gagnon, J., et al., 1996. Can immunoglobulin C(H)1 constant region domain modulate antigen binding affinity of antibodies? J. Clin. Invest. 98, 2235–2243.
Ross, D.A., Wilson, M.R., Miller, N.W., Clem, L.W., Warr, G.W., 1998. Evolutionary variation of immunoglobulin mu heavy chain RNA processing pathways: origins, effects, and implications. Immunol. Rev. 166, 143–151.
Rowe, D.S., Fahey, J.L., 1965. A New Class of Human Immunoglobulins. Ii. Normal Serum Igd. J. Exp. Med. 121, 185–199.
Rumfelt, L.L., Diaz, M., Lohr, R.L., Mochon, E., Flajnik, M.F., 2004. Unprecedented multiplicity of Ig transmembrane and secretory mRNA forms in the cartilaginous fish. J. Immunol. 173, 1129–1139.
Ryo, S., Wijdeven, R.H., Tyagi, A., Hermsen, T., Kono, T., Karunasagar, I., et al., 2010. Common carp have two subclasses of bonyfish specific antibody IgZ showing differential expression in response to infection. Dev. Comp. Immunol. 34, 1183–1190.
Saha, N.R., Suetake, H., Kikuchi, K., Suzuki, Y., 2004. Fugu immunoglobulin D: a highly unusual gene with unprecedented duplications in its constant region. Immunogenetics 56, 438–447.
Sahoo, M., Edholm, E.S., Stafford, J.L., Bengten, E., Miller, N.W., Wilson, M., 2008. B cell receptor accessory molecules in the channel catfish *Ictalurus punctatus*. Dev. Comp. Immunol. 32, 1385–1397.
Savan, R., Aman, A., Sato, K., Yamaguchi, R., Sakai, M., 2005a. Discovery of a new class of immunoglobulin heavy chain from fugu. Eur. J. Immunol. 35, 3320–3331.
Savan, R., Aman, A., Nakao, M., Watanuki, H., Sakai, M., 2005b. Discovery of a novel immunoglobulin heavy chain gene chimera from common carp (*Cyprinus carpio* L.). Immunogenetics 57, 458–463.
Schmucker, D., Clemens, J.C., Shu, H., Worby, C.A., Xiao, J., Muda, M., et al., 2000. Drosophila Dscam is an axon guidance receptor exhibiting extraordinary molecular diversity. Cell 101, 671–684.
Seipelt, R.L., Spear, B.T., Snow, E.C., Peterson, M.L., 1998. A nonimmunoglobulin transgene and the endogenous immunoglobulin mu gene are coordinately regulated by alternative RNA processing during B-cell maturation. Mol. Cell. Biol. 18, 1042–1048.
Srisapoome, P., Ohira, T., Hirono, I., Aoki, T., 2004. Genes of the constant regions of functional immunoglobulin heavy chain of Japanese flounder, *Paralichthys olivaceus*. Immunogenetics 56, 292–300.
Stenvik, J., Jorgensen, T.O., 2000. Immunoglobulin D (IgD) of Atlantic cod has a unique structure. Immunogenetics 51, 452–461.
Tadiso, T.M., Lie, K.K., Hordvik, I., 2011. Molecular cloning of IgT from Atlantic salmon, and analysis of the relative expression of tau, mu and delta in different tissues. Vet. Immunol. Immunopathol. 139, 17–26.
Takagaki, Y., Seipelt, R.L., Peterson, M.L., Manley, J.L., 1996. The polyadenylation factor CstF-64 regulates alternative processing of IgM heavy chain pre-mRNA during B cell differentiation. Cell 87, 941–952.
Van Boxel, J.A., Paul, W.E., Terry, W.D., Green, I., 1972. Communications. IgD-bearing human lymphocytes. J. Immunol. 109, 648–651.
van Ginkel, F.W., Miller, N.W., Cuchens, M.A., Clem, L.W., 1994. Activation of channel catfish B cells by membrane immunoglobulin cross-linking. Dev. Comp. Immunol. 18, 97–107.
Vasil'yev, V., Sokolov, L., Serebryakova, Y., 1981. Karyotypes of the Siberian sturgeon, *Acipenser baeri*, of the Lena river and some aspects of karyotype evolution in Acipenseriformes. J. Ichthyol. 20, 37–45.
Wang, J., Gudikote, J.P., Olivas, O.R., Wilkinson, M.F., 2002. Boundary-independent polar nonsense-mediated decay. EMBO Rep. 3, 274–279.
Wei, Z., Wu, Q., Ren, L., Hu, X., Guo, Y., Warr, G.W., et al., 2009. Expression of IgM, IgD, and IgY in a reptile, *Anolis carolinensis*. J. Immunol. 183, 3858–3864.
White, M.B., Word, C.J., Humphries, C.G., Blattner, F.R., Tucker, P.W., 1990. Immunoglobulin D switching can occur through homologous recombination in human B cells. Mol. Cell. Biol. 10, 3690–3699.
Wilkinson, M.F., 2005. A new function for nonsense-mediated mRNA-decay factors. Trends Genet. 21, 143–148.
Wilson, M.R., van Ravenstein, E., Miller, N.W., Clem, L.W., Middleton, D.L., Warr, G.W., 1995a. cDNA sequences and organization of IgM heavy chain genes in two holostean fish. Dev. Comp. Immunol. 19, 153–164.
Wilson, M.R., Ross, D.A., Miller, N.W., Clem, L.W., Middleton, D.L., Warr, G.W., 1995b. Alternate pre-mRNA processing pathways in the production of membrane IgM heavy chains in holostean fish. Dev. Comp. Immunol. 19, 165–177.
Wilson, M., Bengten, E., Miller, N.W., Clem, L.W., Du Pasquier, L., Warr, G.W., 1997. A novel chimeric Ig heavy chain from a teleost fish shares similarities to IgD. Proc. Natl. Acad. Sci. U.S.A. 94, 4593–4597.
Wojtowicz, W.M., Flanagan, J.J., Millard, S.S., Zipursky, S.L., Clemens, J.C., 2004. Alternative splicing of Drosophila Dscam generates axon guidance receptors that exhibit isoform-specific homophilic binding. Cell 118, 619–633.
Xiao, F.S., Wang, Y.P., Yan, W., Chang, M.X., Yao, W.J., Xu, Q.Q., Wang, X.X., Gao, Q., Nie, P., 2010. Ig heavy chain genes and their locus in grass carp *Ctenopharyngodon idella*. Fish Shellfish Immunol. 29, 594–599.
Yasuike, M., de Boer, J., von Schalburg, K.R., Cooper, G.A., McKinnel, L., Messmer, A., et al., 2010. Evolution of duplicated IgH loci in Atlantic salmon, *Salmo salar*. BMC Genomics 11, 486.
Zhang, Y.A., Salinas, I., Li, J., Parra, D., Bjork, S., Xu, Z., et al., 2010. IgT, a primitive immunoglobulin class specialized in mucosal immunity. Nat. Immunol. 11, 827–835.
Zhao, J., Hyman, L., Moore, C., 1999. Formation of mRNA 3′ ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. Microbiol. Mol. Biol. Rev. 63, 405–445.
Zhao, Y., Kacskovics, I., Pan, Q., Liberles, D.A., Geli, J., Davis, S.K., et al., 2002. Artiodactyl IgD: the missing link. J. Immunol. 169, 4408–4416.
Zhao, Y., Pan-Hammarstrom, Q., Kacskovics, I., Hammarstrom, L., 2003. The porcine Ig delta gene: unique chimeric splicing of the first constant region domain in its heavy chain transcripts. J. Immunol. 171, 1312–1318.
Zhao, Y., Pan-Hammarstrom, Q., Yu, S., Wertz, N., Zhang, X., Li, N., et al., 2006. Identification of IgF, a hinge-region-containing Ig class, and IgD in Xenopus tropicalis. Proc. Natl. Acad. Sci. U.S.A. 103, 12087–12092.
Zhao, Y., Cui, H., Whittington, C.M., Wei, Z., Zhang, X., Zhang, Z., et al., 2009. *Ornithorhynchus anatinus* (platypus) links the evolution of immunoglobulin genes in eutherian mammals and nonmammalian tetrapods. J. Immunol. 183, 3285–3293.