Identification of IgF, a hinge-region-containing Ig class, and IgD in *Xenopus tropicalis*

Yaofeng Zhao*†‡, Qiang Pan-Hammarström†§, Shuyang Yu†, Nancy Wertz¶, Xiaofeng Zhang‖, Ning Li*, John E. Butler¶, and Lennart Hammarström†‡

*State Key Laboratory for AgroBiotechnology, China Agricultural University, Beijing 100094, People's Republic of China; †Division of Clinical Immunology, Karolinska University Hospital, Huddinge, Karolinska Institutet, SE-141 86 Stockholm, Sweden; §Department of Immunology, School of Basic Medical Science, Peking University, Beijing 100083, People's Republic of China; ¶Department of Microbiology, University of Iowa, Iowa City, IA 52242; and ‖Centre for Structural Biochemistry, Karolinska Institutet, SE-141 57 Stockholm, Sweden

Edited by Max D. Cooper, University of Alabama, Birmingham, AL, and approved June 12, 2006 (received for review January 11, 2006)

Only three Ig isotypes, IgM, IgX, and IgY, were previously known in amphibians. Here, we describe a heavy-chain isotype in *Xenopus tropicalis*, IgF (encoded by C_ϕ), with only two constant region domains. IgF is similar to amphibian IgY in sequence, but the gene contains a hinge exon, making it the earliest example, in evolution, of an Ig isotype with a separately encoded genetic hinge. We also characterized a gene for the heavy chain of IgD, located immediately 3′ of C_μ, that shares features with the C_δ gene in fish and mammals. The latter gene contains eight constant-region-encoding exons and, unlike the chimeric splicing of μC_H1 onto the IgD heavy chain in teleost fish, it is expressed as a unique IgD heavy chain. The IgH locus of *X. tropicalis* shows a 5′ V_H-D_H-J_H-C_μ-C_δ-C_χ-C_υ-C_ϕ 3′ organization, suggesting that the mammalian and amphibian Ig heavy-chain loci share a common ancestor.

amphibians | antibody evolution

Immunoglobulins (Igs) are essential components of adaptive immunity and are produced only in gnathostomes such as mammals, birds, reptiles, amphibians, and jawed fish (1, 2). Most mammals express five classes of Igs, IgM, IgD, IgG, IgA, and IgE, each endowed with distinct biological effector functions. Mammalian IgM and IgE heavy chains are composed of four ≈110-aa constant region domains encoded by separate exons, presumably arising from gene duplication during evolution (3). IgD and IgG contain only three domains (rodent IgD contains only two constant region domains) but also encompass a short exon-encoded hinge (genetic hinge) (4–7). IgA is also a three-domain molecule, with a functional hinge encoded by the 5′ end of the heavy-chain constant region domain (C_H) 2 (C_H2) exon (8, 9). The hinge regions contain one or more cysteines that are used to bridge the two heavy chains to form an H_2L_2 antibody structure. Hinge regions are also rich in proline, which confers conformational flexibility that allows waving, rotation of Fab arms, and wagging of the Fc fragment, thus facilitating antigen binding and triggering of effector functions (10, 11). Hinge segments have previously been observed only in mammalian Igs; however, Savan *et al.* (12) recently identified a heavy-chain isotype in fugu fish that contains a hinge region encoded within its C_H2 exon, similar to the hinge of mammalian IgA (8, 9). This putative hinge contains five repeats of VKPT but lacks a cysteine residue to connect the two heavy chains (12).

It is generally believed that mammalian Igs arose from ancestral Igs of lower vertebrates. IgM is found in all vertebrates (13–18); however, the phylogenetic origin of the remaining mammalian Igs is less well established, although Igs referred to as IgA/IgX and IgY have been reported in birds, reptiles, and amphibians (1, 2, 19–22). Cartilaginous fish and lungfish express IgM, IgNAR, and/or IgW/IgX, containing either two or more than four constant region domains (1, 15, 23). Bony fish express three heavy-chain isotypes, IgM, IgD, and IgZ/IgT (24, 25). IgD is found in most mammals but not in birds, amphibians, or reptiles, whereas multidomain encoding C_δ genes have been described in bony fish (1, 22, 26, 27). Thus, a traditional phylogenetic pathway connecting IgD in bony fish and IgD in mammals is missing.

It has long been thought that there are only three Ig classes, IgM, IgA/IgX, and IgY, in lower vertebrates, including birds, reptiles, and amphibians (1, 2, 14, 19–22). The heavy chains that have been characterized to date in these species contain four constant region domains (except for a truncated IgY containing two domains in some species) but no hinge (28). cDNAs encoding the heavy chains of IgM, IgX, and IgY have all been cloned previously in *Xenopus laevis* (14, 19, 20). IgX has been considered to be an analogue of mammalian IgA because a large number of IgX-positive B cells are located in the gut epithelium (29). IgX is structurally distinct from mammalian IgA but is similar to chicken IgA (19, 30). IgY, also consisting of four constant region domains, is found in a variety of birds, amphibians, and reptiles (31) and is regarded as a functional homologue of IgG and the progenitor of both mammalian IgG and IgE (31).

In the last two decades, molecular approaches have facilitated the investigation of the genomes in a variety of species. IgZ was recently discovered in zebrafish (24). The recent assembly of the *Xenopus tropicalis* genome sequence allowed us to perform a search for additional Ig heavy-chain constant region genes in an amphibian.

Results

Identification of the Genomic Sequence Encoding IgM (μ), IgX (χ), and IgY (υ) in *X. tropicalis*. The *X. tropicalis* genome is available in the *X. tropicalis* Genome Sequencing Project database at the Sanger Institute (www.sanger.ac.uk/Projects/X_tropicalis/). By using the published Ig sequences in *X. laevis* as templates, the C_μ, C_χ, and C_υ genes in *X. tropicalis* were all identified in an assembled scaffold (Scaffold_928) where the C_χ is located ≈40 kb downstream of the C_μ. The constant regions were all deduced on the basis of both genomic sequences and EST clones. The organization of the genes is presented in Fig. 1. The amino acid sequences of all three classes displayed a high degree of divergence between *X. tropicalis* and *X. laevis* (Figs. 6–8, which are published as supporting information on the PNAS web site), with sequence similarities of only 74.3%, 82.1%, and 75.0% for IgM, IgX, and IgY, respectively. There is a cysteine located at the carboxyl-terminal end of the IgX of *X. tropicalis* that may be used for binding to the J chain. This cysteine is absent in the IgX sequence of *X. laevis* (Fig. 7) (19).

Identification of an IgD-Encoding Gene (C_δ) in *X. tropicalis*. The long stretch (≈40 kb) of intervening DNA between the C_μ and C_χ genes

Conflict of interest statement: No conflicts declared.

This paper was submitted directly (Track II) to the PNAS office.

Abbreviation: C_H, heavy-chain constant region domain.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. DQ350886 and DQ350887).

‡To whom correspondence may be addressed. E-mail: yaofeng.zhao@ki.se or lennart.hammarstrom@ki.se.

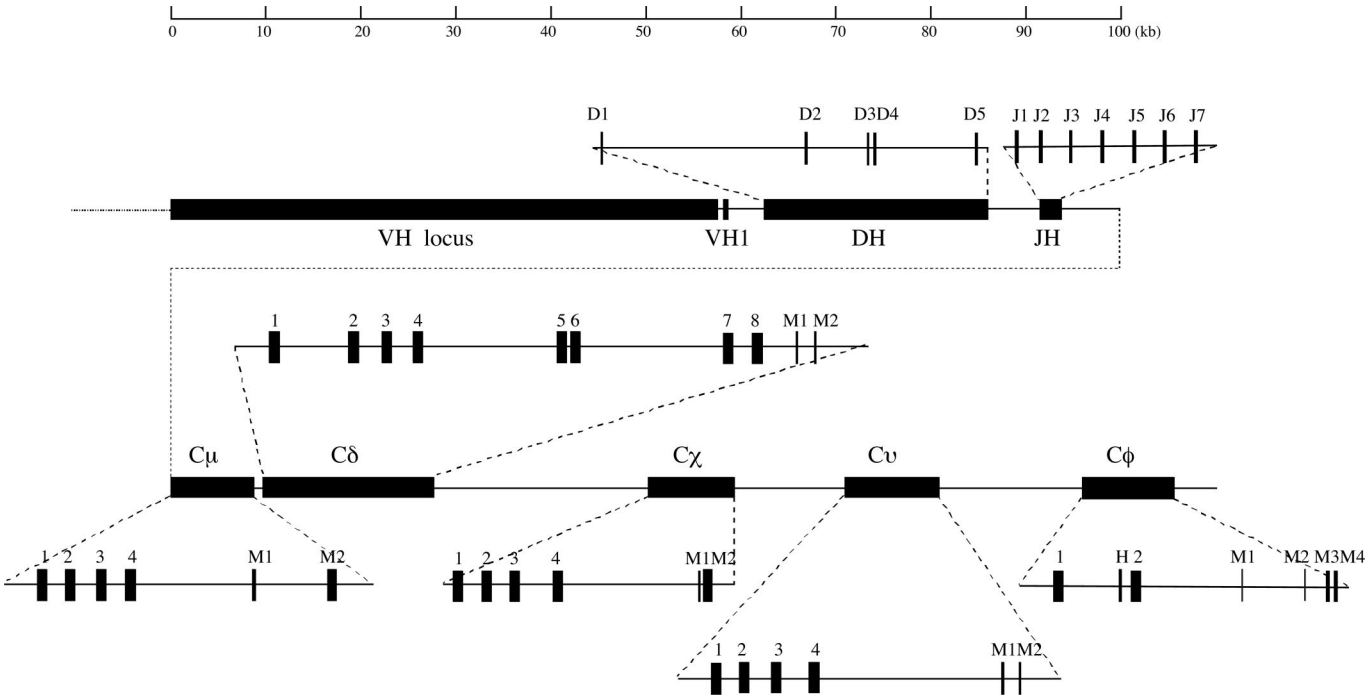

Fig. 1. Assembly of the *X. tropicalis* IgH gene locus. VH, heavy-chain variable genes; DH, heavy-chain diversity gene segments; JH, heavy-chain joining gene segments; C_{μ}, IgM encoding gene; C_{δ}, IgD encoding gene. The filled boxes indicate exons encoding structurally conserved IgC domains: C_{χ}, IgX encoding gene; C_{υ}, IgY encoding gene; C_{ϕ}, IgF encoding gene; M, membrane exon. The domains encoding exons of each constant region gene are indicated with Arabic numbers. The position of the χC_H1 exon is uncertain because it is missing in Scaffold_928 due to a small sequence gap.

encouraged us to look for an IgD-encoding gene. A BLAST search that permitted some local mismatches was performed against the genome database, using the sequence of the transmembrane region of IgM from *X. tropicalis* as a template. This procedure identified a putative transmembrane-encoding region that could not be ascribed to IgM, IgX, or IgY heavy-chain encoding genes in Scaffold_928. This DNA sequence is located between the C_{μ} and C_{χ} genes and encodes a short peptide similar to IgM and IgD transmembrane regions in other species (Fig. 9, which is published as supporting information on the PNAS web site) (32), suggesting the presence of an additional heavy-chain constant region gene. To test this possibility, we used a nested RT-PCR to amplify cDNA transcripts from spleen RNA, using primers that cover the most-expressed J_H genes (J_H3, see below) and the putative exon for the transmembrane region. Sequence analysis of the amplified product revealed a heavy-chain transcript with eight C_H exons, plus exons for M1 and M2 segments (Figs. 1 and 2, and Fig. 10, which is published as supporting information on the PNAS web site).

The constant region heavy-chain gene we describe is located immediately downstream of the C_{μ} gene (1.3 kb downstream of the μM2 exon). The 1.3-kb DNA seems too short to accommodate DNA elements such as an I region promoter, an I exon, and a switch region, all of which are required for class switch DNA recombination. Nor was any repetitive sequence, suggestive of a switch region, observed by dot-plot analysis of the 1.3-kb sequence (see below). Thus, the identified gene is most likely expressed through cotranscription with the C_{μ} gene, supporting the notion that it may be the homologue of fish and mammalian C_{δ}. This hypothesis was further supported by a phylogenetic analysis in which the gene clustered with fish and mammalian C_{δ} genes (Fig. 3). Moreover, the deduced transmembrane region of this heavy-chain class, which is encoded by two separate exons, shows a high similarity to IgD in cows and sheep (Fig. 9). Taken together, these data strongly suggest that the identified gene is a C_{δ} gene in *X. tropicalis*. Surprisingly, a BLAST analysis showed that the *X. tropicalis* IgD had the highest overall similarity to the lungfish IgW, indicating that IgD and IgW may have a common origin (33).

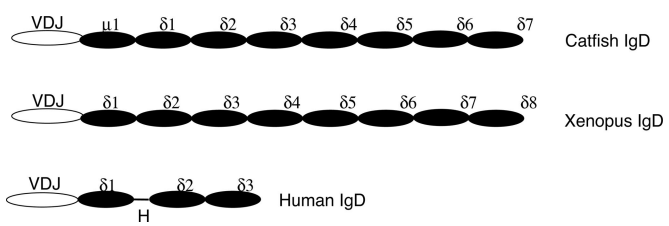

Fig. 2. Structure of IgD heavy chains in different vertebrates. Catfish IgD (GenBank accession no. U67437); human IgD (GenBank accession no. AAB21246); H, hinge region.

Unlike fish IgD, in which the μC_H1 is spliced onto the IgD sequences to form a chimeric heavy chain (27), the cloned *X. tropicalis* IgD sequences showed that the J_H-encoding gene segments were spliced directly onto the δC_H1 exon, generating a unique IgD heavy chain (Fig. 2). Correspondingly, there are two cysteines in the *Xenopus* δC_H1 (Fig. 10) that could be used to link heavy chains and light chains; these cysteines are absent in catfish δC_H1 (27).

Structural analysis of the *X. tropicalis* C_{δ} gene showed that it spanned ≈18 kb of genomic DNA and consisted of eight C_H exons, where exons 5 and 6 are homologous to exons 7 and 8 (82% and 76% homology at the DNA level, respectively), suggesting intragene exon duplication similar to that observed in fish IgD (34). We furthermore performed domain-to-domain sequence comparisons to determine the sequence homology of the *X. tropicalis* IgD with its equivalents in catfish and humans. This analysis revealed homology of the *Xenopus* δC_H1 with the δC_H1 domains of catfish and humans and homology of the *Xenopus* δC_H7 and the catfish δC_H6 with the human δC_H3 (27) (Table 1, which is published as supporting information on the PNAS web site).

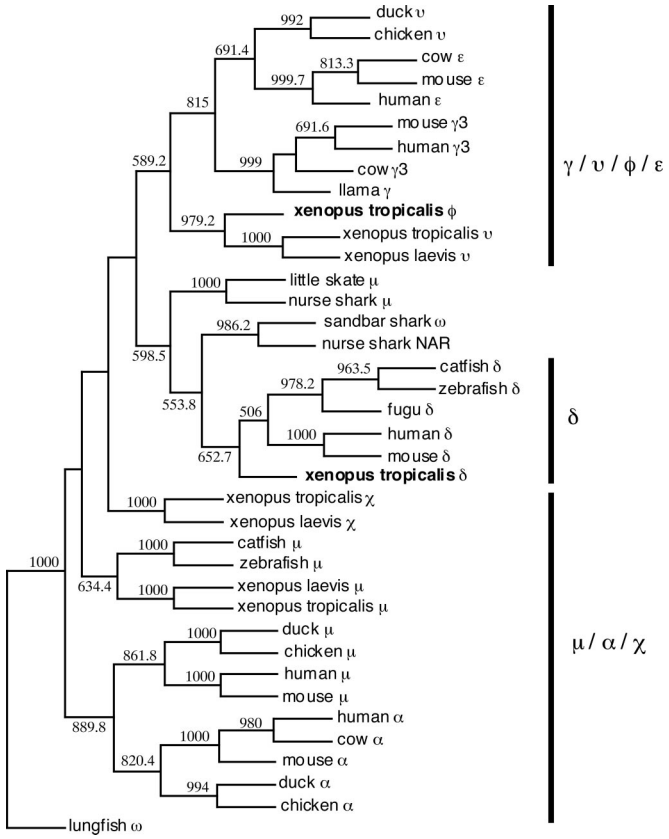

Fig. 3. An unrooted phylogenetic tree of Igs in vertebrates. The tree was constructed by using protein sequences of the first and last C_H domains (fish δC_H6 and *Xenopus* δC_H7 were used as the last domain) of all heavy-chain classes. Except for the Ig sequences obtained in this study, all other sequences were taken from the GenBank database, with the following accession numbers: C_δ gene: catfish (AF363450), fugu (AB159481), human (BC021276), mouse (J00449), and zebrafish (BX510335); C_μ gene: catfish (M27230), chicken (X01613), duck (AJ314750), human (X14940), mouse (V00818), nurse shark (M92851), little skate (M29679), *X. laevis* (BC084123), and zebrafish (AY643751); C_α gene: chicken (S40610), cow (AF109617), duck (U27222), human (P01877), and mouse (BC010324); C_ε gene: cow (BTU63640), human (AK130825), and mouse (X01857); C_γ gene: cow (S82407), human (BX640623), llama (AF305955), and mouse (AY498569); C_υ gene: chicken (X07174), duck (AJ314754), and *X. laevis* (X15114); and C_χ gene: *X. laevis* (BC072981), nurse shark NAR (U18701), sandbar shark IgW (U40560), and lungfish IgW (AF437727).

Identification of IgF, a Hinge-Containing Ig Heavy-Chain Isotype in *X. tropicalis*. The large EST database of *X. tropicalis* available in the National Center for Biotechnology Information (NCBI) GenBank database (1,037,711 EST clones as of April 30, 2006) enabled analysis of EST clones harboring rearranged V(D)J sequences. Most of these clones contained IgM, IgX, or IgY heavy chains. However, two EST clones (GenBank accession nos. DR836854 and CF378719) contained a rearranged V(D)J segment and a C_H-like sequence distinct from that of the IgM, IgD, IgX, and IgY heavy chains. A further search in the GenBank database revealed a completely sequenced, but not annotated, cDNA clone (GenBank accession no. BC087793) derived from the U.S. Department of Energy Joint Genome Institute *X. tropicalis* EST project. This cDNA clone would code for an Ig heavy-chain sequence containing a 137-aa variable region (including a signal peptide) and a short constant region (of 230 aa) (Fig. 11, which is published as supporting information on the PNAS web site). On the basis of an NCBI conserved domain search, the latter region is divided into two Ig constant region domains and a short interconnecting polypeptide.

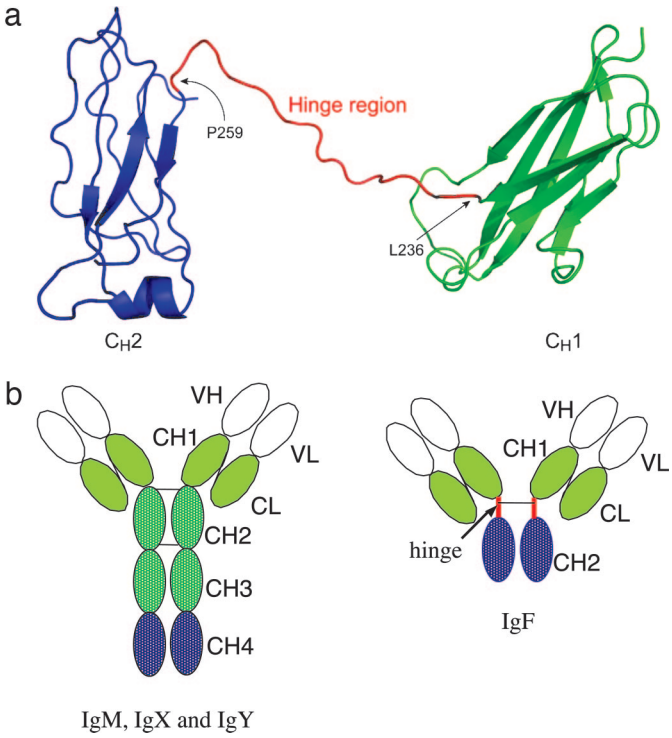

Fig. 4. Structure of Igs in *X. tropicalis*. (*a*) A ribbon representation of the predicted structural model of the *X. tropicalis* IgF heavy chain. The C_H1 and C_H2 domains are colored green and blue, respectively. The putative hinge region between the two domains is colored red. Note that the hinge between C_H1 and C_H2 contains a gap (Ser-248 to Gly-252), which is due to the absence of corresponding residues in the template structure. The figure was prepared with PyMOL software. (*b*) Domain structure of IgF as compared with IgM, IgX, and IgY. There is only one cysteine in the C terminus of the C_H2 domain of IgM for potential inter-heavy-chain disulfide bonding. CH, heavy-chain constant region domain; CL, light-chain constant region domain; VH, heavy-chain variable region; VL, light-chain variable region.

Marked differences between the deduced amino acid sequences and those of *X. tropicalis* IgM, IgD, IgX, and IgY suggest that the cDNA clone represents a secreted form of a heavy-chain isotype in *X. tropicalis*. We termed this isotype "IgF" (encoded by the C_ϕ gene).

The genomic sequence encoding the heavy-chain constant region of IgF was obtained by searching the *X. tropicalis* Genome Sequencing Project database. The C_ϕ gene was found to be present in another assembled scaffold (Scaffold_972). Alignment of the IgF cDNA sequence (Fig. 11) with the genomic sequence showed that the C_ϕ gene consists of three exons, two of which (C_H1 and C_H2) encode constant region domains. The short polypeptide between these domains is encoded by a separate exon, suggesting the presence of a putative hinge region (Figs. 1 and 11). We subsequently performed a protein structure prediction based on comparative modeling (i.e., protein fold recognition) by using 1D and 3D sequence profiles and employing the web-based software 3D-PSSM (35). A loop linking the C_H1 and C_H2 domains was identified that involved residues 236–259 (Fig. 4), in agreement with the amino acid sequence alignment (Fig. 11). This loop may potentially serve as a hinge between the C_H1 and C_H2 domains (Fig. 4).

The putative hinge region of the *X. tropicalis* IgF contains a conserved cysteine that may covalently link the two IgF heavy chains. The presence of three prolines is also reminiscent of the amino acid sequences in the hinge regions of mammalian Igs.

The transmembrane and cytoplasmic regions of the membrane-bound form of IgF were cloned by using 3′ RACE PCR (Fig. 12, which is published as supporting information on the PNAS web

IMMUNOLOGY

site). As seen in Fig. 1, the C terminal of the IgF membrane-bound form is encoded by four exons, with the last two exons (M3 and M4) encoding the transmembrane and cytoplasmic domains, respectively. Two tandem repeats of a nine-amino-acid unit [DL(G/R)AWITGP] are encoded by two other short exons (M1 and M2) located between the C_H2 and transmembrane domains. PCR amplification of two different transcripts indicated that IgF may be expressed in both secreted and membrane-bound forms.

Assembly of the Ig Heavy-Chain Gene Locus in *X. tropicalis*. In mammals and birds, the IgH genes are organized as a large cluster containing V_H-D_H-J_H-C_H genes and spanning hundreds to thousands of kilobases. To assemble the *X. tropicalis* IgH gene locus, we used sequence data generated by the *X. tropicalis* Genome Sequencing Project. The V_H, D_H, J_H, and four constant region genes, including C_μ, C_δ, C_χ, and C_υ, are all contained within the 298-kb Scaffold_928, whereas the C_ϕ gene is located on a second scaffold (Scaffold_972). The presence of ≈5 kb of overlapping sequences at their ends suggested that these two scaffolds might be linked. To confirm this notion, we performed a long-distance PCR amplification (Fig. 13, which is published as supporting information on the PNAS web site). End-sequencing of the amplified 15-kb PCR product clearly showed that Scaffold_972 is positioned downstream of Scaffold_928. The C_ϕ gene is located ≈15 kb downstream of the C_υ gene; the deduced organization of the *X. tropicalis* Ig heavy-chain gene locus is shown in Fig. 1. The entire IgH locus was thoroughly annotated except for the V_H gene locus. As shown in Fig. 1, the D_H locus spans ≈24 kb of DNA and is located 3 kb downstream of the most 3′ V_H gene. It contains only five D_H gene segments, which is fewer than in most other species (36), but there are sequence gaps between D_H1 and D_H2 and between D_H4 and D_H5, and some D_H genes may thus be missing. This observation may explain why none of the five identified D_H genes is found in the V(D)J junction of the IgF presented in Fig. 3. Each D_H segment is flanked on both sides by classical nonamer and heptamer recombination signal sequences, separated by a conserved 12-bp spacer (Fig. 14, which is published as supporting information on the PNAS web site). Seven J_H gene segments were identified ≈5 kb downstream of the D_H locus (Figs. 1 and 14). According to a BLAST search against the *X. tropicalis* EST database, the J_H3 gene segment is the most frequently used segment in the expressed V(D)J sequences (19/28), followed by J_H1 (4/28) and J_H2 (3/28). We did not find sequences characteristic of the conserved 5′ intronic enhancer that is located between the J_H and C_μ genes in mammalian Ig heavy-chain constant region gene loci (37).

To identify putative switch region sequences for the constant region genes in *X. tropicalis*, we performed sequence comparisons with dot-plot analysis. An ≈2.7-kb region containing short repetitive sequences (63% A+T content) could be identified ≈1.4 kb upstream of the μC_H1 exon, suggesting the presence of a switch μ region (S_μ) (Fig. 15*a*, which is published as supporting information on the PNAS web site). The S_μ region is abundant in AGCT motifs but is shorter than the previously reported S_μ (≈5 kb) in *X. laevis* (38). Such a long repetitive block could not be found in the DNA regions upstream of C_υ or C_ϕ, although these regions all show a high AT content (>60%) and contain some repetitive sequences (Fig. 15*d* and *e*). An ≈750-bp DNA region containing a repetitive sequence was, however, observed upstream of the C_χ gene (Fig. 15*c*). The 1.3-kb intron sequence between the C_μ and C_δ genes is devoid of any possible candidate sequence for a switch region according to the dot-plot analysis (Fig. 15*b*).

Expression of IgF and IgD in *X. tropicalis*. With RT-PCR, we could show that both IgF and IgD are expressed mainly in the spleen (Fig. 5), whereas IgM is expressed in nearly all tissues investigated. In accordance with previous observations (29, 39), IgX was detected in the spleen, intestine, and stomach. Expression of IgF and IgD in the spleen is weaker than expression of IgM, IgX, and IgY, which probably explains the rareness of IgF and IgD EST clones in the NCBI EST database. Hsu *et al.* (39) have previously shown that a short Ig with an estimated size similar to IgF (slightly longer than two domain light chains) could be precipitated by rabbit anti-*Xenopus* Igs in *X. laevis*, suggesting that IgF is a functional heavy-chain isotype in *Xenopus*. However, their results did not show any secreted Ig corresponding to IgD in size (nine domains), indicating that, as in mammals, IgD is expressed at a very low level in serum.

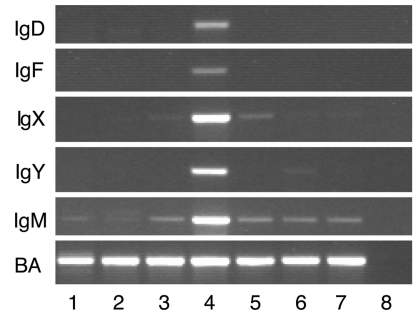

Fig. 5. Expression of *X. tropicalis* IgD, IgF, IgX, IgY, and IgM in different organs as detected by RT-PCR. BA, β actin; 1, kidney; 2, thymus; 3, intestine; 4, spleen; 5, stomach; 6, liver; 7, caecum; 8, negative control.

Discussion

In the present study, we have identified two Ig isotypes, IgD and IgF, in *X. tropicalis.* To our knowledge, these Ig classes have not previously been described in amphibians. We have also shown that the *X. tropicalis* IgH locus is organized in V_H-D_H-J_H-C_μ-C_δ-C_χ-C_υ-C_ϕ order, thus conforming to the typical translocon configuration observed in mammals and fish (24, 37).

Unexpectedly, the IgF heavy chain we identified contains a hinge region and only two constant region domains. Antibodies containing two constant region domains have previously been found in cartilaginous fish, lungfish, and ducks (13, 23, 28, 40); however, these antibodies are generated through the use of different transcription termination sites or through alternative RNA splicing of the full-length transcript (13, 28). Identification of a two-domain Ig heavy-chain isotype has also recently been reported in fugu (12). This isotype corresponds to zebrafish IgZ in terms of position relative to IgM but is structurally different because it contains a hinge-like sequence within the C_H2 domain-encoding exon (12). The discovery of IgF in *X. tropicalis*, together with the recent findings in fugu described above (12), suggest that the Ig hinge region had already evolved in lower vertebrates.

The flexibility of the Fab arms of antibodies is important for their functional properties. Thus, most mammalian Igs contain a sequence for a hinge, encoded either by a dedicated exon (genetic hinge) or by part of a regular exon (functional hinge). Because they have previously been found only in mammals, hinge regions are believed to have emerged independently in C_δ, C_γ, and C_α after the divergence of mammals from other tetrapods (3, 33, 41). It has previously been proposed that the hinge may either be evolutionarily condensed from an ancestral constant region domain encoding exon (3, 31, 42) or that it evolved by duplication, leading to incorporation of an acceptor RNA splice site (rich in the pyrimidines that are required to encode prolines) into the 5′ portion of the C_H exon (5, 8, 9). If the original splice site of the C_H exon is still used for RNA splicing, the newly incorporated splice site would thus encode a proline-rich hinge segment attached to the N terminus of the C_H domain (5, 8, 9). According to the latter model, further mutations creating a donor splice site may lead to detachment of the hinge exon and formation of a domain relic (the C_H exon was inactivated into an intronic sequence) (5, 8, 9). Sequence analysis of the IgF hinge seems to support the latter hypothesis. First, the 3′ end of the IgF hinge exon (CCTCCATAATGCCAG) is very

similar to a 3′ intronic splice site. Second, the intron (347 bp) between the hinge and C_H2 exons shows homology with the C_H2 exon (Fig. 16, which is published as supporting information on the PNAS web site) and may thus be a domain relic as a consequence of evolution to a detached hinge exon (8). Interestingly, there appears to be another 3′ intronic splice site in the C_H2 exon, immediately upstream of the homologous sequence of the domain relic (Fig. 16), which provides additional support for the notion that a shifting of splice sites has been involved in the formation of the hinge exon.

When separate IgF C_H domains were used for BLAST searches, the IgF C_H2 showed the highest homology to the C_H3 domain of llama and camel IgG in non-*Xenopus* Igs, whereas the IgF C_H1 was similar to the C_H1 of dog and panda IgG, strongly suggesting that IgF is related to mammalian IgG. This finding also explains why IgF clusters with IgY and mammalian IgG in a phylogenetic analysis (Fig. 3). In addition to the sequence similarity, hinge regions of both IgG and IgF are encoded by a separate exon, whereas the mammalian IgA hinge is usually encoded within the C_H2 exon.

The high degree of sequence homology between the υC_H1 and ϕC_H1 exons (Fig. 17, which is published as supporting information on the PNAS web site) indicates that the C_ϕ gene may originally have been duplicated from the C_υ gene. This notion is also supported by a sequence analysis that suggests that both the C_υ and C_ϕ genes are evolutionarily close to the C_γ and C_ε genes of mammals (Fig. 3). IgY is regarded as a precursor of both IgE and IgG (31). Structural similarity (both display a four-domain structure), sequence homology, and shared biological properties of the IgY and IgE heavy chains support the view that IgY or an IgY-like Ig was the immediate predecessor of IgE (31, 43). The IgY-to-IgG transition must, however, have been accompanied by structural changes that led to the creation of a hinge region in modern IgG molecules. Thus, IgF may share a common hinge-containing ancestor with the mammalian IgG. However, hinge formation in IgF in *Xenopus* and hinge formation in IgG in mammals appear to be independent events, inasmuch as the domain relic is located immediately downstream of the IgF hinge exon but upstream of the hinge exon in the mouse $C_{\gamma 2b}$ encoding gene (5, 8). Accepting that the IgG hinge developed after the emergence of mammalian species, the IgF hinge may also have been formed after the divergence of amphibians. Thus, the hinge regions of IgF and IgG may be a consequence of convergent evolution.

Because the C_H2 of IgF shows a high homology with llama and camel IgG, we further compared the composition of the hinges in these Igs and the recently reported hinge in fugu (12, 44). The comparison showed that fugu and camel hinges are both characterized by distinct repeats (VKPT in fugu and PKPQP in camel) that are slightly similar to the C terminus of the IgF hinge (NTKP) (12, 44). The low sequence similarity of hinges in different species is not surprising because the hinge regions appear to have evolved rapidly (45). The fugu hinge lacks the cysteine that is used to bridge the two heavy chains (12, 44). It is thus likely that additional steps (either mutation or generation of another cysteine-containing segment) were involved in the formation of some hinges that are based on a preexisting segment (or sometimes duplicated segments).

IgD has previously been found only in mammals and fish but not in birds and amphibians (6, 27). The discovery of IgD in *X. tropicalis* partially fills this evolutionary gap. The genomic organization of the C_δ gene appears to be similar to its equivalent in fish, inasmuch as the IgD-encoding genes are all encoded by more than four C_H exons (24, 27). However, splicing of the μC_H1 exon onto the IgD sequences, a mechanism that is used to express IgD in fish (27, 46), is not observed in *X. tropicalis*. Rather, the expression of IgD is similar to that in mammals, where rearranged V(D)J sequences are joined directly to the C_δ sequence. The fact that the transmembrane portion of the *X. tropicalis* IgD displays a higher degree of homology with the IgD of mammals than of fish (Fig. 9) suggests that *X. tropicalis* IgD is an evolutionary intermediate between the IgD of fish and mammals and that the C_δ gene has undergone a condensing process in mammals.

The availability of the genome sequence of a species made the present study possible and has confirmed and extended information already known from *X. laevis*. The sequence also has provided evidence for IgD in a tetrapod considered more primitive than mammals and has also provided the evolutionarily earliest evidence of an Ig isotype with a separately encoded hinge exon. Presence of five Ig heavy-chain isotypes in *X. tropicalis* suggests that its IgH locus shares a common ancestor with mammals.

Materials and Methods

RNA and DNA Isolations. Frogs (*X. tropicalis*) were purchased from NASCO (Fort Atkinson, WI). RNA isolations were conducted by using either the RNeasy mini kit (Qiagen, Hilden, Germany) or Trizol (Molecular Research Center, Cincinnati, OH), in accordance with standard manufacturer's instructions. The dissected tissues were homogenized by using iron beads and used directly in RNA isolations. DNA isolations were performed with either DNAzol or phenol extraction. First-strand cDNA synthesis was conducted with either random or NotI-d$(T)_{18}$ primers.

RT-PCR Detection of Transcriptions of the *X. tropicalis* Ig Heavy-Chain Genes in Different Organs. The synthesized cDNA samples with RNA isolated from different organs were used in RT-PCR to detect expression of IgF, IgY, IgX, IgM, and IgD; the *Xenopus* β actin gene was used as a control. (The primers are listed in Table 2, which is published as supporting information on the PNAS web site.)

Rapid Amplification of the IgF cDNA 3′ End (3′ RACE). Approximatley 400 ng of spleen total RNA was used to synthesize first-strand cDNA with a First-Strand cDNA synthesis kit (Amersham Pharmacia Biotech, Uppsala, Sweden). The primers used in the 3′ RACE are listed in Table 3, which is published as supporting information on the PNAS web site. The resulting PCR products were purified by using the QIAquick gel extraction kit (Qiagen, Valencia, CA) and subsequently cloned into a T vector and sequenced (MWG Biotech, Ebersberg, Germany).

Amplification and Cloning of the IgD cDNA. A nested RT-PCR was used to amplify the IgD heavy-chain cDNA, using the primers Tropicalis JHS2 (5′ GGG GAC CAG GGA CCA CGG TCA C 3′), Tropicalis JHS3 (5′ ACC ATG GTC ACC GTC ACT TCA G 3′), Tropicalis IgDFullas1 (5′ GTG CAG GTA AAG TAG AAT AGT T 3′), and Tropicalis IgDFullas2 (5′ ATG GTC AGT TTC CTT CTT GGT A 3′). The resulting PCR products were cloned into a T vector and sequenced.

Long-Distance PCR Amplification of the DNA Fragment Between the C_υ and C_ϕ Genes. To determine the position of Scaffold_972 relative to Scaffold_928 and the distance between the C_υ and C_ϕ genes, we performed a long-distance PCR amplification by using one primer, IgYTMs (5′ GAC CAC GGC TAT CAC ATT TAT CTC 3′), derived from the IgY transmembrane region, and IgFC_H1as (5′ GAA ATC CAG AAG CAA AGC ATC CAA 3′), derived from the IgF C_H1 exon by using the Expand Long Template PCR system (Roche Diagnostics). The amplified 15-kb DNA fragment was recovered and directly sequenced from both ends by using the original PCR primers to confirm the sequence identity.

Annotation of the *X. tropicalis* IgH Gene Locus. Whereas the NCBI EST database is used in BLAST searches for expressed sequences (www.ncbi.nlm.nih.gov/BLAST/), the genome database we used is built by the Sanger Institute (www.sanger.ac.uk/DataSearch/). D_H gene segments were identified by using an online software program (FUZZNUC; (http://bioweb.pasteur.fr/seqanal/interfaces/fuzznuc.html) and searching a consensus sequence motif (CACT-

GTG-N_{12}-ACAAAAACC) allowing five mismatches. Another sequence motif (GGTTTTTGT-$N_{21–23}$-CACTGTG) was used to identify J_H gene segments.

DNA and Protein Sequence Computations. DNA and protein sequence editing, alignments, and comparisons were performed with the MegAlign software (DNASTAR). A phylogenetic tree was constructed by using Protpars from the PHYLIP software package. A consensus tree was taken from 1,000 bootstrapped phylogenetic trees. Multiple sequence alignments for the tree construction were performed with ClustalW. The 3D structure prediction was performed with the 3D-PSSM software (35). The resulting structure, based on the template structure of mouse IgG1 (Protein Data Bank code 1IGY, 29% sequence identity), has a PSSM *E*-value of 0.0828, indicating a prediction certainty >90% (35). The first 20 residues on the N terminus and the last 13 residues on the C terminus of the predicted structure were omitted because of the low sequence identity.

We thank Dr. Daniel L. Weeks for providing samples from *X. tropicalis*, Prof. Rudolf Ladenstein for structure modeling, and Dr. Sicheng Wen for technical help in preparing the figures. This work was supported by the Swedish Research Council and the National Natural Science Fund of China.

1. Flajnik, M. F. (2002) *Nat. Rev. Immunol.* **2,** 688–698.
2. Stavnezer, J. & Amemiya, C. T. (2004) *Semin. Immunol.* **16,** 257–275.
3. Lin, L. C. & Putnam, F. W. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 504–508.
4. Tucker, P. W., Liu, C. P., Mushinski, J. F. & Blattner, F. R. (1980) *Science* **209,** 1353–1360.
5. Tucker, P. W., Marcu, K. B., Newell, N., Richards, J. & Blattner, F. R. (1979) *Science* **206,** 1303–1306.
6. Zhao, Y., Kacskovics, I., Pan, Q., Liberles, D. A., Geli, J., Davis, S. K., Rabbani, H. & Hammarström, L. (2002) *J. Immunol.* **169,** 4408–4416.
7. Zhao, Y., Pan-Hammarström, Q., Kacskovics, I. & Hammarström, L. (2003) *J. Immunol.* **171,** 1312–1318.
8. Tucker, P. W., Slightom, J. L. & Blattner, F. R. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 7684–7688.
9. Flanagan, J. G., Lefranc, M. P. & Rabbitts, T. H. (1984) *Cell* **36,** 681–688.
10. Brekke, O. H., Michaelsen, T. E. & Sandlie, I. (1995) *Immunol. Today* **16,** 85–90.
11. Roux, K. H., Strelets, L. & Michaelsen, T. E. (1997) *J. Immunol.* **159,** 3372–3382.
12. Savan, R., Aman, A., Sato, K., Yamaguchi, R. & Sakai, M. (2005) *Eur. J. Immunol.* **35,** 3320–3331.
13. Ota, T., Rast, J. P., Litman, G. W. & Amemiya, C. T. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 2501–2506.
14. Schwager, J., Mikoryak, C. A. & Steiner, L. A. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 2245–2249.
15. Kokubu, F., Hinds, K., Litman, R., Shamblott, M. J. & Litman, G. W. (1988) *EMBO J.* **7,** 1979–1988.
16. Dahan, A., Reynaud, C. A. & Weill, J. C. (1983) *Nucleic Acids Res.* **11,** 5381–5389.
17. Wilson, M. R., Marcuz, A., van Ginkel, F., Miller, N. W., Clem, L. W., Middleton, D. & Warr, G. W. (1990) *Nucleic Acids Res.* **18,** 5227–5233.
18. Turchin, A. & Hsu, E. (1996) *J. Immunol.* **156,** 3797–3805.
19. Haire, R. N., Shamblott, M. J., Amemiya, C. T. & Litman, G. W. (1989) *Nucleic Acids Res.* **17,** 1776.
20. Amemiya, C. T., Haire, R. N. & Litman, G. W. (1989) *Nucleic Acids Res.* **17,** 5388.
21. Zhao, Y., Rabbani, H., Shimizu, A. & Hammarström, L. (2000) *Immunology* **101,** 348–353.
22. Lundqvist, M. L., Middleton, D. L., Hazard, S. & Warr, G. W. (2001) *J. Biol. Chem.* **276,** 46729–46736.
23. Harding, F. A., Amemiya, C. T., Litman, R. T., Cohen, N. & Litman, G. W. (1990) *Nucleic Acids Res.* **18,** 6369–6376.
24. Danilova, N., Bussmann, J., Jekosch, K. & Steiner, L. A. (2005) *Nat. Immunol.* **6,** 295–302.
25. Hansen, J. D., Landis, E. D. & Phillips, R. B. (2005) *Proc. Natl. Acad. Sci. USA* **102,** 6919–6924.
26. Butler, J. E. (2006) *Dev. Comp. Immunol.* **30,** 1–17.
27. Wilson, M., Bengten, E., Miller, N. W., Clem, L. W., Du Pasquier, L. & Warr, G. W. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 4593–4597.
28. Magor, K. E., Higgins, D. A., Middleton, D. L. & Warr, G. W. (1994) *J. Immunol.* **153,** 5549–5555.
29. Mussmann, R., Du Pasquier, L. & Hsu, E. (1996) *Eur. J. Immunol.* **26,** 2823–2830.
30. Mansikka, A. (1992) *J. Immunol.* **149,** 855–861.
31. Warr, G. W., Magor, K. E. & Higgins, D. A. (1995) *Immunol. Today* **16,** 392–398.
32. Zhao, Y. & Hammarström, L. (2003) *Immunology* **108,** 288–295.
33. Hsu, E., Pulham, N., Rumfelt, L. L. & Flajnik, M. F. (2006) *Immunol. Rev.* **210,** 8–26.
34. Hordvik, I., Thevarajan, J., Samdal, I., Bastani, N. & Krossoy, B. (1999) *Scand. J. Immunol.* **50,** 202–210.
35. Kelley, L. A., MacCallum, R. M. & Sternberg, M. J. (2000) *J. Mol. Biol.* **299,** 499–520.
36. Litman, G. W., Anderson, M. K. & Rast, J. P. (1999) *Annu. Rev. Immunol.* **17,** 109–147.
37. Zhao, Y., Kacskovics, I., Rabbani, H. & Hammarström, L. (2003) *J. Biol. Chem.* **278,** 35024–35032.
38. Mussmann, R., Courtet, M., Schwager, J. & Du Pasquier, L. (1997) *Eur. J. Immunol.* **27,** 2610–2619.
39. Hsu, E., Flajnik, M. F. & Du Pasquier, L. (1985) *J. Immunol.* **135,** 1998–2004.
40. Magor, K. E., Warr, G. W., Middleton, D., Wilson, M. R. & Higgins, D. A. (1992) *J. Immunol.* **149,** 2627–2633.
41. Hsu, E. (1994) *Semin. Immunol.* **6,** 383–391.
42. Sakano, H., Rogers, J. H., Huppi, K., Brack, C., Traunecker, A., Maki, R., Wall, R. & Tonegawa, S. (1979) *Nature* **277,** 627–633.
43. Mussmann, R., Wilson, M., Marcuz, A., Courtet, M. & Du Pasquier, L. (1996) *Eur. J. Immunol.* **26,** 409–414.
44. Nguyen, V. K., Hamers, R., Wyns, L. & Muyldermans, S. (1999) *Mol. Immunol.* **36,** 515–524.
45. Sumiyama, K., Saitou, N. & Ueda, S. (2002) *Mol. Biol. Evol.* **19,** 1093–1099.
46. Bengten, E., Clem, L. W., Miller, N. W., Warr, G. W. & Wilson, M. (2006) *Dev. Comp. Immunol.* **30,** 77–92.