Lectin Pathway of Bony Fish Complement: Identification of Two Homologs of the Mannose-Binding Lectin Associated with MASP2 in the Common Carp (Cyprinus carpio)

Miki Nakao, Takayuki Kajiya, Yuho Sato, Tomonori Somamoto, Yoko Kato-Unoki, Misao Matsushita, Munehiro Nakata, Teizo Fujita and Tomoki Yano

*J Immunol* 2006; 177:5471-5479; doi: 10.4049/jimmunol.177.8.5471

http://www.jimmunol.org/content/177/8/5471

---

**References**

This article cites 50 articles, 20 of which you can access for free at:

http://www.jimmunol.org/content/177/8/5471.full#ref-list-1

**Subscription**

Information about subscribing to *The Journal of Immunology* is online at:

http://jimmunol.org/subscription

**Permissions**

Submit copyright permission requests at:

http://www.aai.org/About/Publications/JI/copyright.html

**Email Alerts**

Receive free email-alerts when new articles cite this article. Sign up at:

http://jimmunol.org/alerts
Lectin Pathway of Bony Fish Complement: Identification of Two Homologs of the Mannose-Binding Lectin Associated with MASP2 in the Common Carp (Cyprinus carpio)1,2

Miki Nakao,3,* Takayuki Kajiya,* Yuho Sato,* Tomonori Somamoto,* Yoko Kato-Unoki,* Misao Matsushita,† Munehiro Nakata,‡ and Tomoki Yano*†

The lectin pathway of complement is considered to be the most ancient complement pathway as inferred from identification of ancient homologs of mannose-binding lectin (MBL) and MBL-associated serine proteases (MASPs) in some invertebrates. MBL homologs with galactose selectivity and an MASP3-like sequence also occur in bony fish, linking the evolution of the lectin complement pathway from invertebrates to higher vertebrates. However, these cannot be considered authentic complement components until confirmatory functional evidence is obtained. Here, we report the isolation and characterization of two MBL homologs from a cyprinid teleost, the common carp, Cyprinus carpio. One, designated GalBL, corresponds to the MBL-like molecule with the galactose specificity. The other is an authentic MBL with mannose specificity. Both were found to associate with a serine protease that cleaves native human C4 into C4b but not C4i with a hydrolyzed thiostere. Molecular cloning and phylogenetic analysis revealed this C4-activating protease to be carp MASP2, indicating that MASP2 arose before the emergence of bony fish. Database mining of MBL-like genes reveals that MBL and GalBL genes are arranged in tandem in the zebrafish genome and that both lectins are conserved in the distantly related puffer fish. These results imply that bony fish have developed a diverged set of MBL homologs that function in the lectin complement pathway. The Journal of Immunology, 2006, 177: 5471–5479.

The lectin pathway of mammals is an Ab-independent activation route of the complement system, relying on recognition of pathogen-associated molecular patterns by two distinct groups of lectins, mannose-binding lectin (MBL)4 and ficolins. These serum lectins associate with novel serine proteases, termed MBL-associated serine proteases (MASPs), that proteolytically activate the complement components C4, C2, and C3. In humans, the lectin pathway is considered to be crucial in innate immunity, especially in infants and children, providing an immediate defense against microbial infections (reviewed in Refs. 1 and 2).

MBL is an oligomeric lectin. Each of its subunits is composed of four distinct domains: 1) a cysteine-rich N-terminal region responsible for disulfide-linked oligomerization; 2) a collagenous region characterized by Gly-X-Y repeats (X and Y stand for any amino acids); 3) a neck region; and 4) a C-type lectin-like carbohydrate recognition domain (CRD) at the C-terminal that places it within the collectin family. MBL preferentially binds to terminal α-mannose and N-acetyl-β-d-glucosamine (GlcNAc) of oligosaccharides and polysaccharides. Ficolin is also a homo-oligomer of subunits with domain organization similar to MBL, although the C terminus has a fibrinogen-like domain instead of a C-type lectin. This fibrinogen-like domain shows significant amino acid sequence similarity to that of tachylectin 5 from horseshoe crab (3), and is responsible for carbohydrate recognition by ficolins. Like tachylectin 5, the fibrinogen-like domain usually recognizes acetylated sugars, such as GlcNAc and N-acetyl-β-d-galactosamine (4), enabling ficolins to recognize a spectrum of pathogens distinct from that by MBL. Thus, together, MBL and ficolins seem to facilitate pattern recognition of a wide variety of pathogens (5).

MASPs are homologous to the subcomponents of the complement components C1r, C1s, and C1s, because they share the same domain organization: two C1r/C1s/Uegf/bone morphogenetic protein 1 domains that are interrupted by an epidermal growth factor-like domain, followed by two short consensus repeat modules and a serine protease domain. Three different forms of MASPs have been identified in mammals, namely MASP1, MASP2, and MASP3 (2). MASP1 and MASP3 share the N-terminal five domains but differ in the C-terminal serine protease domain as a result of alternative splicing (6). The serine protease domain of MASP1 is a common type encoded by multiple exons and contains the histidine loop structure and the active serine residue encoded by a TCN codon. Conversely, the MASP3 serine protease domain is encoded by a single intronless exon, in which the histidine loop is missing, and an AGY codon is used for the active serine (2). MASP2 has a MASP3-like serine protease domain but is encoded by a distinct gene, MASP2, that also produces a truncated molecule,
termed small MBL-associated protein or MASP1, by alternative poly-
adenylation and splicing (7–9). Several lines of functional evidence have shown that MASP2 is responsible for activation of C4 and C2 to form C3-convertase, a C3-cleaving enzyme complex, whereas MASP1 directly activates C3 (2). However, the natural substrate of MASP3 is still unknown and the exact molecular composition of MBL/ficolins and MASPs have yet to be determined.

Recently, complement components, C3 and factor B, have been identified in the horseshoe crab (an arthropod) (10), and a C3-like gene has been found in coral (a cnidarian) (11). These findings indicate that some complement components arose well before the divergence of the protostomes and deuterostomes, and a complement pathway, albeit in a primordial form, may have a much longer evolutionary history than previously thought. Although neither an MBL-like lectin nor a MASP-like protease has been found in these primitive animals, the lectin pathway is now widely regarded as the ancestral mechanism of complement activation, and is the form seen in several extant deuterostome invertebrates (12). Structural and functional analyses of the complement components C3 (13), an MBL-like novel glucose-binding lectin (14), MASP1-like serine proteases (15), and a mammalian CR3-like receptor (16, 17) in the Japanese solitary ascidian suggest that these components constitute a prototype complement system and that the lectin pathway plays an essential role in C3-mediates opsonization for phagocytes in primitive animals. The presence of the lectin pathway has also been indicated in cephalochordates by the identification in Amphioxus of the lectin pathway components, MASP1, MASP3, and C3 (18). These close ancestral relatives of vertebrates are believed to lack MASP2-like components, because an MASP2-like gene is absent from the genome of another member of urochordates, the sea squirt, Ciona intestinalis (19, 20). This is consistent with the lack, in the urochordates, of C4 and C2, natural substrates of MASP2 in mammals. In the lamprey, an agnathan that occupies a more ancient taxonomic position than cartilaginous or bony fish, both an MBL-like and a novel C1q-like molecule have been reported to function as lectins that activate C3 through an MASP3-like serine protease, MASP-A (21, 22). This points to the lectin pathway playing a crucial role in complement activation in animals lacking adaptive immunity (21).

Although cartilaginous and bony fish are thought to possess a functional lectin pathway (2, 23), this belief is supported only by some fragmentary structural and functional data at the molecular level: principally, the presence of a putative MBL-like collectin in carp, zebrafish, and goldfish (24) and a putative MASP3-like molecule in banded hound shark and carp (25). However, the primary structure of the teleost MBL-like lectin predicts that it has selectivity for galactose, unlike mammalian MBL, which recognizes mannose and GlcNAc (24), so this lectin might be better designated as galactose-binding lectin (GalBL), as it is throughout this paper. With regard to MASP-like molecules, the putative carp C1r/C1s/MASTP2-like sequences may be more closely allied to mammalian C1r (26, 27) than MASP. To date, there is no evidence at the protein level to indicate the presence of functional lectin pathway in either cartilaginous or bony fish. By contrast, amphibians have acquired a full lectin pathway similar to that of mammals, because ficolin, MASP1/3, and MASP2 genes have all been identified in Xenopus (25, 28). In birds, although MBL, MASP2, and MASP3 have been found, MASP1 has not and may be absent (29–31). Thus, cartilaginous and bony fish represent the missing link in the evolution of the lectin pathway from the minimum prototype present in invertebrates to the more sophisticated one present in mammals.

The present study was aimed at obtaining molecular and functional data to ascertain the presence of a true lectin pathway in fish, using the common carp (Cyprinus carpio) as the model species.

**Materials and Methods**

### Carps

Carp IgM and anti-carp IgM

Carp IgM was purified from serum as described for rainbow trout IgM (33). Briefly, carp serum was fractionated by precipitation with 5–15% polyethylene glycol 4000, and the resulting precipitate was dissolved in 100 mM Tris-HCl (pH 8.2) before gel filtration through a Superdex 200 pg column (1.6 × 60 cm), pre-equilibrated with the same buffer. The IgM-rich fractions in the void volume were pooled, diluted 10-fold with water, and then subjected to anion-exchange chromatography on a linear NaCl gradient ranging from 0 to 300 mM in 10 mM Tris-HCl buffer (pH 8.0) in a Q-Sepharose FF column (1.6 × 5 cm). The carp IgM, eluting at ~100 mM NaCl, was homogenous as judged by SDS-PAGE under reducing conditions, giving a 74-kDa H chain band and a 25-kDa L chain band (data not shown). The purified IgM was then emulsified with CFA and injected s.c. into a rabbit at weekly intervals over 3 wk. A month after the last injection, the rabbit was bled, and IgG in the antisera was purified by affinity chromatography on a HiTrap protein A column (Amersham Biosciences) as described in the manufacturer’s instructions. The anti-carp IgM rabbit IgG (IgM-A) was then fixed to a hydroxylysine-citrate-activated HiTrap column (1 ml) (Amersham Biosciences), following the manufacturer’s instructions.

### Purification of carp MBL

Carp MBL was purified closely following the method for lamprey MBL and C1q (21). Briefly, pooled carp serum (300 ml) was made 7% in polyethylene glycol 4000, and the precipitate was separated by centrifugation at 10,000 × g for 20 min at 4°C. It was dissolved in 50 mM Tris-HCl buffer (pH 7.8) containing 200 mM NaCl and 10 mM CaCl2, and applied to a column (1.5 × 3 cm) of GlcNAc-agarose (Sigma-Aldrich) equilibrated with the same buffer. After washing the column with the same buffer, adsorbed proteins were successively eluted with buffers containing 5, 50, and 300 mM methyl-α-D-mannopyranoside (MetMan).

The elute with 50 mM MetMan-containing buffer was applied to the anti-carp IgM-HiTrap column equilibrated with TBS containing 10 mM CaCl2. After washing the column with this buffer, the adsorbed IgM was eluted with 0.2 M glycine-HCl (pH 2.5). The purified carp MBL, present in the flow-through fraction, was dialyzed against 50 mM Tris-HCl buffer (pH 7.4) containing 100 mM NaCl, 2 mM CaCl2, and 0.02% NaN3.

### Purification of a GalBL

To purify a GalBL, a column (1.5 × 2 cm) of acid-treated agarose was prepared according to a published method (34) except that Bio-Gel A 1.5m (Bio-Rad) was used instead of Sepharose 6B and equilibrated with the same starting buffer as used for GlcNAc-agarose. The flow-through fraction from the GlcNAc column was applied to the acid-treated agarose column and eluted with buffer containing 20 mM p-galactose, before passing through the anti-carp IgM column as described above.

### C4 activation assays

The ability of carp MBL-MASP complex to cleave C4 into C4b was examined as follows: 2 μg of human C4, purified as described before (35), was incubated with varying amount of the purified carp lectins in 50 mM Tris-HCl buffer (pH 7.4) containing 100 mM NaCl and 2 mM CaCl2 at 25°C for 1 h. C4 activation was monitored by appearance of α′-chain after SDS-PAGE under reducing conditions. For the negative control substrate, C4i, an inactivated form of C4 in which the thioester is hydrolyzed, was prepared by incubating C4 with 5 mM methyamine in 20 mM Tris-HCl buffer (pH 9.0) at 37°C for 2 h (36).

### General protein analyses

SDS-PAGE was performed under either reducing or nonreducing conditions as previously described (37), and the protein bands were visualized with Coomassie Blue R-250. Two-dimensional SDS-PAGE (first dimension under nonreducing conditions and second dimension under reducing conditions) was performed as described in Ref. 38. N-terminal amino acid
sequences of the separated polypeptides were determined after electrophototography onto a polyvinylidene difluoride membrane (Bio-Rad) using a gas-phase protein sequencer (PPSQ-21; Shimadzu).

Assay of carp MBL binding to oligosaccharides and neoglycoproteins

Carp MBL was labeled with 125I and its ability to bind neoglycolipids and neoglycoproteins was analyzed as described elsewhere (39). The neoglycolipids were prepared by reductive amination from dipalmitoylphosphatidylcholine and were prepared by reductive amination from dipalmitoylphosphatidylcholine. The test neoglycoproteins included BSA conjugated with mannose, glucose, GlcNAc, galactose, maltose, cellobiose, or lactose. Human MBL purified as described (40) was used as control.

Complementary DNA cloning of carp MBL and MASP2

Poly(A)⁺ RNA was isolated from carp hepatopancreas using a Quick Prep mRNA purification kit (Amersham Biosciences) and reverse-transcribed with Moloney murine leukemia virus reverse transcriptase and oligo(dT) primer (Invitrogen Life Technologies), according to the manufacturer's instructions. First-strand cDNA served as a template for PCR amplification using two pairs of degenerated oligonucleotide primers: P1 (CARGGCNCNCCNGNAGA encoding QGPPGK) and P2 (CANGGNACRTCRTATCACTCATCCTGAT ATATCCAC) for MBL, and P3 (ATNCCTGGCARGTNATGAT coding for IPWQVM) and P4 (CCCCCANACGATNCNCCncomplementarily encoding GGVIS) for MASP2. PCR amplification was conducted using AmpliTaq Gold DNA polymerase (PerkinElmer) on Astec Thermocycler PC800 under the following conditions: initial denaturation at 95°C for 10 min, 40 cycles of (95°C for 0.5 min, 45°C for 0.5 min, and 72°C for 1 min), and a final extension at 72°C for 5 min. The amplified products were gel-purified and subcloned into pGEM-T vector (Promega). Nucleotide sequencing was performed on a model 377 sequencer (Applied Biosystems).

The following gene specific primers were designed to known carp-specific and MASP-like cDNA sequences and synthesized for RACE: P5 (ATACTGACCCGGGGTTGTTGC) and P6 (AGATTTAGTCTGCTGGCTCATCCTGAT ATATCCAC) for MBL, and P7 (GGCCATCGGCATTACCGTG) and P8 (GGCCATCTCTCAGCTGACGTT) for MASP2. A SMART RACE cDNA Amplification kit (Clontech) was used for both 3’-RACE and 3’-RACE, following manufacturer’s protocols. A high fidelity DNA polymerase (KOD plus; Toyobo) was used for the RACE-PCR, and the amplified products might be misassembled. Authenticity of the MASP2 sequence assembled from the RACE products was confirmed by PCR using primers corresponding to the 5’-end (P10, ATGACTATGACTGTCACATGCAAAGC) and the 3’-end (P11, ATATGCTATGACTGTCACATGCAAAGC).

Southern hybridization

Carp genomic DNA (10 µg) isolated from the erythrocytes was digested to completion with BamHI, HindIII, or EcoRI, electrophoresed on a 1% agarose gel, transferred to a Hybond N+ membrane (Amersham Biosciences), and hybridized with digoxigenin-labeled cDNA probe, essentially as described (41). The probes were prepared with PCR-Dig Probe Synthesis kit (Roche Diagnostics).

Phylogenetic analyses and database mining

ClustalX 1.82 software (42) was used to generate multiple sequence alignments and to construct neighbor-joining phylogenetic trees (43) with the aid of Nplot software. All gaps inserted in the alignment were ignored for the tree construction.

Online database searches were conducted using the following web sites: NCBI blast server (www.ncbi.nlm.nih.gov/BLAST/), Fugu genome blast server (http://fugu-sg.org/), and GenomeNet blast server (http://blast. genome.jp/). In all the searches, BLOSUM62 was used as a scoring matrix.

Results

Isolation and identification of carp serum lectins

Carp serum was fractionated by polyethylene glycol precipitation and GlcNAc-agarose chromatography in a similar manner to that used for lamprey MBL (22). As shown in Fig. 1, lane 1, the fractions eluting from the affinity column with 50 mM MetMan show two major bands with molecular masses of 30 and 74 kDa, respectively, indicating contamination with IgM. Accordingly, the eluate was passed through the anti-IgM column to remove this contamination. The final MBL preparation was confirmed to be free of IgM by SDS-PAGE. This produced a major band of 30 kDa and a faint band of 55 kDa under reducing conditions (Fig. 1, lane 2). The N-terminal amino acid sequence of the 30-kDa polypeptide shows a significant similarity to that of MBL-like lectin from carp reported previously (24) because it contains four repeats of the collagen-like sequence signature (Gly-X-Y). Because digestion of carp MBL with collagensenase (type III from Clostridium histolyticum; Sigma-Aldrich), performed as described (21), generates a 22-kDa polypeptide with an N-terminal sequence of GVAGLKG-DKG (data not shown), carp MBL probably contains a collagenous region similar to that of other collagens.

To examine whether the MBL homolog (24) thought to be specific for galactose is actually expressed as a mature protein, we searched for the corresponding protein in carp serum by affinity chromatography using acid-treated agarose, which exposes terminal galactose residues. As shown in Fig. 1, lane 3, the protein eluted from the acid-treated agarose column gave a 32-kDa band after removal of IgM. N-terminal sequence of the 32-kDa polypeptide contains three repeats of the Gly-X-Y motif and shows striking similarity to that deduced from the cDNA sequence of the reported carp MBL-like lectin (24).
Binding specificity of carp MBL

Binding specificity of carp MBL was compared with that of human MBL using neoglycolipids bearing oligosaccharides and BSA-based neoglycoproteins as ligands. Both lectins show similar oligosaccharide-binding spectra (Fig. 2A) and also recognize the same set of neoglycoproteins conjugated with mannose, glucose, GlcNAc, maltose, or cellobiose (Fig. 2B). These results are suggestive of carp MBL purified in this study being a direct functional counterpart of mammalian MBL.

C4 activation by carp MBL and GalBL

Because the MBL and GalBL purified above were expected to associate with MASP-like proteases, we next examined whether or not the two lectin preparations initiate MASP-mediated lectin pathway activation. The purified lectin preparation was incubated with human C4 or C4i and cleavage of their α-chain into α'-chain was monitored by SDS-PAGE. As shown in Fig. 3, both lectin preparations yielded α'-chain (85 kDa) from human C4 but not from human C4i. This indicates that carp MBL and GalBL are associated with a MASP2/1s-like serine proteases, which is active against only native C4 with an intact thioester.

cDNA cloning of carp MBL

RT-PCR amplification from carp hepatopancreas RNA with degenerate primers (P1 and P2) yields a single band of the expected size for MBL (~550 bp). The DNA was subcloned into pGEM-T vector. Among a random selection of 10 clones, one clone, designated CM1, showed high similarity to mammalian MBL over the middle collagenous region and CRD (data not shown). Gene-specific primers, P5 and P6, based on the CM1 sequence were used for 3'-RACE and 5'-RACE, respectively, in combination with UPM supplied in the SMART RACE kit. Three and five distinct cDNA sequences were isolated by 3'-RACE and 5'-RACE, respectively. Next, full-length cDNA encoding carp MBL was amplified from the 3'-RACE-ready first-strand cDNA using a sense strand primer, P9, which corresponds to 5'-untranslated sequence shared by all of the 3'-RACE products, and UPM. After cloning the RACE product (~1 kbp long) into pGEM-T vector, 12 clones were sequenced over the entire region, resulting in three distinct sequences, designated MBL-A, MBL-B, and MBL-C.

FIGURE 2. Binding specificity of carp MBL. A, Binding of carp and human MBL to neoglycolipids bearing oligosaccharides released from various glycoproteins and developed by TLC. OVA and RNase B indicate lanes where neoglycolipids prepared from desialylated oligosaccharides derived from IgG, fibrinogen, ferritin, and α1-acid glycoprotein, respectively, were developed. B, Binding of carp and human MBL to neoglycolipids bearing oligosaccharides: BSA conjugated with mannose (Man), glucose (Glc), N-acetyl-d-glucosamine (GlcNAc), galactose (Gal), maltose (Mal), cellobiose (Cel), and lactose (Lac).

FIGURE 3. Activation of C4 by MBL and GalBL purified from carp serum. Purified carp MBL and GalBL (2 µg each) were incubated with 2 µg of human C4 or C4i at room temperature for 30 min, and then analyzed by SDS-PAGE to monitor the appearance of α'-chain of C4.

FIGURE 4. Multiple alignment of the deduced amino acid sequences of MBL homologs. Dots show residues identical with carp MBL1, and dashes denote gaps introduced for maximum matching. Assignments and bound-
possess conserved amino acid residues responsible for coordination of Ca\(^{2+}\) and ligand binding in the CRD (44). Notably, carp MBL1/MBL2 and human MBL both possess an EPN motif.

**Identification of zebrafish MBL-like genes by database mining**

The entire amino acid sequence of carp MBL1 was used as a query for the TBLASTN search of GenBank nucleotide sequence database, yielding a best hit of zebrafish genomic contig, AL954692. Four distinct genes that show close similarities with carp MBL were identified in its 101,677-bp sequence and are designated gene 1, gene 2, gene 3, and gene 4, respectively. Their putative exon-intron organizations are predicted primarily on the basis of the so-called GT/AG rule, as shown in Table I and Fig. 5A. All of the genes are arrayed in the same transcriptional direction.

As shown in Fig. 5B, the predicted sequences of gene 1 and gene 2 are very similar to each other but different from gene 3. The gene 4-encoding sequence diverges more from the other three. Carp MBL1 shows a substantially higher similarity (81%) to gene 4 than to genes 1–3 (46–49%), suggesting that gene 4 encodes zebrafish MBL. This interpretation is supported by the presence of an EPN motif in the putative CRD sequence of gene 4 that predicts an MBL-like sugar specificity (Fig. 5C). By contrast, the deduced sequence of zebrafish GalBL is almost identical with that deduced from gene 3, indicating that gene 3 encodes GalBL (24). Interestingly, gene 1 and gene 2 show moderate amino acid identity (63–67%) to zebrafish GalBL and contain the QPD motif present in GalBL. These results indicate that zebrafish has at least a single copy of MBL gene and three copies of GalBL, probably as a result of tandem gene duplications. The GalBL sequence from carp is more similar to gene 1 and gene 2 than to gene 3.

**Phylogenetic tree of MBL homologs**

A phylogenetic tree of MBL homologs was constructed using the neighbor-joining method with ascidian glucose-binding lectin included as an outgroup. As shown in Fig. 6, bony fish MBL-like lectins form a clade that probably diverged from the clade that led to MBLs of higher vertebrates. In the bony fish clade, MBLs form a cluster separate from that of GalBLs indicating that GalBL is a creation of the bony fish lineage. Because of their extremely high identity, zebrafish GalBL (24), gene 4 sequence predicted here, and the sequence with accession number Q5043 probably represent allotypic variants of the same gene. In the cluster of GalBLs,

### Table I. Zebrafish MBL-like genes identified in a genomic contig AL954692

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Base Number</th>
<th>Length (bp)</th>
<th>Encoding Domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene 1</td>
<td></td>
<td></td>
<td>M-Col</td>
</tr>
<tr>
<td>Exon 1</td>
<td>22,836–23,043</td>
<td>208</td>
<td></td>
</tr>
<tr>
<td>Exon 2</td>
<td>23,137–23,214</td>
<td>78</td>
<td>Col</td>
</tr>
<tr>
<td>Exon 3</td>
<td>23,981–24,712</td>
<td>746 + 256b</td>
<td>N + CRD + 3'-UT (to 2nd polyA)</td>
</tr>
<tr>
<td>Gene 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exon 1</td>
<td>25,575–25,782</td>
<td>208</td>
<td>M-Col</td>
</tr>
<tr>
<td>Exon 2</td>
<td>25,875–25,952</td>
<td>78</td>
<td>Col</td>
</tr>
<tr>
<td>Exon 3</td>
<td>26,281–26,686</td>
<td>473 + 933b</td>
<td>N + CRD + 3'-UT (to 1st polyA)</td>
</tr>
<tr>
<td>Gene 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exon 1</td>
<td>31,053–31,260</td>
<td>208</td>
<td>M-Col</td>
</tr>
<tr>
<td>Exon 2</td>
<td>31,338–31,415</td>
<td>78</td>
<td>Col</td>
</tr>
<tr>
<td>Exon 3</td>
<td>31,549–32,185</td>
<td>467 + 171b</td>
<td>N + CRD + 3'-UT (to 2nd polyA)</td>
</tr>
<tr>
<td>Gene 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exon 1</td>
<td>36,570–36,756</td>
<td>187</td>
<td>M-Col</td>
</tr>
<tr>
<td>Exon 2</td>
<td>36,871–36,957</td>
<td>87</td>
<td>Col</td>
</tr>
<tr>
<td>Exon 3</td>
<td>38,894–38,977</td>
<td>84</td>
<td>N</td>
</tr>
<tr>
<td>Exon 4</td>
<td>39,074–39,655</td>
<td>377 + 205b</td>
<td>N + CRD + 3'-UT (to 2nd polyA)</td>
</tr>
</tbody>
</table>

a Abbreviations for domain names: M, the initiation methionine; Col, collagenous region; N, neck region; CRD, carbohydrate recognition domain; UT, untranslated region.

b Length of the amino acid coding region.

---

**FIGURE 5.** Identification of zebrafish MBL and GalBL genes. A. Arrangement of genes encoding zebrafish MBL and GalBL. A part of a zebrafish genomic contig (GenBank accession number AL954692) is shown as a horizontal line with nucleotide numbers from the beginning of the contig. Boxes stand for exons predicted on the basis of sequence alignment and the GT/AG rule for intron-exon boundaries. B. Amino acid sequence identity between MBL-like lectins of zebrafish and carp. C. Alignment of the amino acid sequences of a part of the MBL-like lectin domain encoded by the zebrafish genes 1–4 in A. Primary determinant residues for sugar specificity are shown in boldface. Consensus residues for the C-type lectin domain and residue numbers in human MBL-C (Swiss-Prot accession no. P11226) are shown above and below in the alignment, respectively.
GalBL and genes 1 and 2 probably diverged before the separation of zebrafish, goldfish, and carp from their common ancestor. Thus, GalBL probably arose during the early stages of bony fish evolution and after the MBL/GalBL duplication.

Identification of carp MASP2 associated with MBL and GalBL

To identify the protease responsible for C4 activation by purified carp MBL, carp MBL was analyzed by two-dimensional SDS-PAGE followed by N-terminal sequencing. As shown in Fig. 7A, Comassie blue staining reveals two minor polypeptide spots with molecular masses of 55 kDa (spot 2) and 29 kDa (spot 3) alongside a major 30-kDa spot of MBL (spot 1). N-terminal sequences of spot 2 and spot 3, respectively, show significant similarities to those of H chain and L chain of human MASP2 (Fig. 7B).

A cDNA fragment encoding a part of the L chain was amplified using a pair of degenerated primers (P3 and P4). P3 encodes an amino acid sequence stretch (IPWQVMI) in the N-terminal sequence of the L chain, whereas P4 complementarily encodes a well-conserved sequence (GGIVSWG) near the active serine residue of the L chain, which is the serine protease domain in human MASP2. RT-PCR amplification from carp hepatopancreas poly(A)−/H11001 RNA yielded a single band with expected size (630 bp). The amplified product was cloned and sequenced, resulting in the isolation of a cDNA fragment similar to human MASP2 (data not shown). A full-length cDNA sequence was then obtained by 5′-RACE and 3′-RACE performed using SMART RACE cDNA amplification kit and gene-specific primers, P7 and P8, designed from the cDNA fragment sequence. After the sequences of the RACE products were assembled on the basis of a 100-bp-long overlap, authenticity of the assembly was verified by amplification of the entire coding region with the primer pair, P10 and P11.
Table II.  Amino acid sequence identity of carp MASP2 with other members of the MASP/C1r/C1s family

<table>
<thead>
<tr>
<th>Species</th>
<th>MASP1</th>
<th>MASP2</th>
<th>MASP3</th>
<th>C1r</th>
<th>C1s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human</td>
<td>38.1</td>
<td>46.5</td>
<td>40.6</td>
<td>37.7</td>
<td>35.1</td>
</tr>
<tr>
<td>Mouse</td>
<td>39.5</td>
<td>45.5</td>
<td>41.3</td>
<td>34.8</td>
<td>35.8</td>
</tr>
<tr>
<td>Xenopus</td>
<td>39.8</td>
<td>43.2</td>
<td>42.8/42.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carp</td>
<td>37.4</td>
<td>34.8/33.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trout</td>
<td>35.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shark</td>
<td>39.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lamprey</td>
<td>37.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ascidian</td>
<td>29.0/29.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amphioxus</td>
<td>36.5</td>
<td>34.5</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Calculated on the basis of pair-wise alignments.
* Identity with the a/b isotypes.
* Identity with the A/B isotypes.

In the full-length sequence of MASP2-like carp cDNA (2307 bp), an ORF of 2055 bp specifies 685 aa. As shown in Fig. 8, the alignment of the deduced protein sequences of carp MASP2 with other members of MASP/C1r/C1s family reveals conservation of their domain structure. Importantly, carp MASP2 shows a closer similarity to that of human MASP2 than to human MASP1, MASP3, C1r, or C1s at the amino acid level (Table II). It should be noted that the serine protease domain of carp MASP2 lacks the two cysteine residues that form the histidine loop, and an AGY codon encodes the active center serine residue as does human MASP1, MASP2, and C1s.

The assignment of carp MASP2 is also supported by the neighbor-joining phylogenetic tree (Fig. 9), in which carp MASP2 forms a cluster with MASP2 molecules from higher vertebrates, such as Xenopus, mouse, and humans, with a high bootstrap value (97%).

**FIGURE 9.** Phylogenetic tree of the MASP/C1r/C1s family drawn by neighbor-joining method. Only bootstrap percentages <100% are inserted. Abbreviations for animal species are as in Fig. 6, except for Onmy (rainbow trout), Xela (Xenopus laevis), Laja (lamprey), Trsc (banded houndshark), and Brbe (amphioxus).

**FIGURE 10.** Genomic Southern blotting analysis of carp MBL, GalBL, and MASP2. Carp genomic DNA was digested with BamHI (B), HindIII (H), or EcoRI (E), and electrophoresed on a 1% agarose gel, followed by blotting to a Hybond N+ membrane. The digests on the membrane were hybridized with DIG-labeled cDNA probes corresponding to the C-type lectin domains of carp MBL (nucleotide number 489–820 in AB110826) and GalBL (nucleotide number 469–837 in AF227737) and to the serine protease domain of carp MASP2 (nucleotide number 1393–2082 in AB234294). After stringent washing, the hybridized probes were visualized with alkaline phosphatase-conjugated anti-DIG and Lumiphos Plus chemiluminescent substrate. Sizes and positions of α/HindIII marker DNA are shown on the left. The black and white arrowheads show MBL-specific bands (5.2 kbp in BamHI-digest and 2.7 kbp in HindIII digest) and GalBL-specific band (3.8 kbp in HindIII digest), respectively, in comparison between MBL and GalBL probes.

**Genomic Southern hybridization of carp MBL, GalBL, and MASP2**

Genomic Southern hybridization was performed to estimate the number of MBL, GalBL, and MASP2 genes in carp, which has a pseudotetraploid origin. As shown in Fig. 10, MBL and GalBL probes hybridized with one to two common bands in each digest, in addition to minor bands specific for each probes (marked by the arrowheads in Fig. 10). These results suggest that MBL and GalBL genes, which may be present as a single or duplicated copies, are tandemly located in close proximity, as in the zebrafish genome (Fig. 5A). In contrast, MASP2-probe hybridized with two bands in each digest, indicating that carp have duplicated MASP2 genes.

**Discussion**

Among the lectin pathway components identified in mammals, an MBL homolog or GalBL (24) and a MASP3-like serine protease (25) have been characterized at the cDNA level in bony fish. Unlike mammalian MBL, the bony fish MBL homolog is synthesized mainly in spleen and predicted to have a selectivity for galactose. (Accordingly, the lectin is designated GalBL in this paper.) Because the natural substrates of MASP3 have not been determined for any animal species, it remains unclear whether or not bony fish have a functional lectin pathway linked with C3 or C4 activation and, if so, which components construct this pathway. In this report, we show that a MBL-MASP2 complex and a GalBL-MASP2 complex purified from carp serum activate C4, suggesting that at least two distinct collectins, which differ in the binding specificity, comprise the lectin complement pathway of carp. The identification of MASP2 is supported by both structural and functional lines of evidence. Certainly, sequence similarity and phylogeny are sufficiently convincing to assign the carp protease as MASP2. In addition, specific cleavage of C4, but not of C4i, into fragment C4b by MBL- or GalBL-MASP2 complexes is a reliable indicator of MASP2 activity and rules out the possibility that fragmentation could be catalyzed by nonspecific proteases contaminating the purified MBL and GalBL. Because no protease inhibitor was included in the MBL purification procedures, the MASP2 associated with the purified MBL and GalBL...
were in an activated from, as shown by the two-chain structure seen in Fig. 7.

Thus, the present study clearly shows that bony fish have an authentic MBL in the sense that it shows the same carbohydrate specificity with human MBL. Recently, a similar authentic MBL molecule associated with MASP-A has been identified from lamprey and shown to activate lamprey C3 (22). Therefore, it is likely that MBL was created before the emergence of lamprey and has been conserved as a recognition molecule of the lectin pathway throughout vertebrate evolution. Moreover, MBL-A and MBL-C must have diverged after the common ancestor of tetrapods had separated from bony fish, so the MBL-A/MBL-C-duplication was probably an independent event in the mammalian lineage.

With regard to phylogeny of MASP family, carp have been reported to possess MASP3-like and C1r-like serine proteases among the members of MASP/C1r/C1r family (25, 26). However, in the present study, neither of these was detected by N-terminal sequencing performed on the purified MBL and GalBL preparations under the conditions used, although MASP2 was clearly identified. MASP-related protein (MRP) (45), a truncated form of MASP3, was also undetectable. It remains to be discovered whether the complexes contain MASP1, MASP3, and MRP, even at a very low level. The absence of MASP1 has recently been reported for chicken; MASP1-specific serine protease exons are missing in the chicken genome (31). Possibly this could also be the case for carp, although it has yet to be proven by genomic analysis.

One of the unique features of the bony fish complement system is the diversity of its components. In various fish species, multiple copies of C3 and factor B genes have been isolated and suggested to produce functionally differentiated isoforms (23, 46). Carp complement, in particular, has multiple copies of gene coding for C3 (47), factor B/C2 (48), C1r/s (26), MASP3/MPR (45), C4 (49), C5 (50), and factor I (51). It is therefore intriguing that two homologous, but distinct, coding sequences of MBL have been cloned from carp. As two clones (nos. 3 and 10) differing only in UTs can be grouped into MBL2, which has a coding sequence distinct from that of MBL1, it is likely that carp MBL1 and MBL2 are encoded by distinct genes, whereas clones nos. 3 and 10 may represent allotypes of MBL2 gene. The phylogenetic tree (Fig. 6) indicates that MBL1 and MBL2 diverged after carp and zebrafish had separated, and therefore the MBL1/MBL2 duplication is likely to have occurred during tetraploidization of the ancestral carp, or could be an even more recent event. Because the deduced amino acid sequences of MBL1 and MBL2 share 97% similarity and both have consensus residues responsible for ligand binding, it is possible that the two isoform proteins do not functionally diverged. However, MBL1 and MBL2 could differ in their expression with MBL2 as the dominant form, because collagenase digestion of purified MBL gives a single N-terminal sequence matching the collagenous region of MBL2 rather than MBL1 (Fig. 4).

The gene cluster of the MBL-homolog found in zebrafish (Fig. 5) demonstrates the diversified isoforms of MBL with distinct carbohydrate specificity. Sequence similarity (Fig. 5B) and phylogenetic tree analysis (Fig. 6) of zebrafish MBL-like genes 1–4 imply the mechanism of the MBL-like gene diversification as follows: Bony fish duplicated the MBL gene in a certain stage and mutated the extra copy to acquire galactose-binding capability, thereby establishing GaBL and making it the common ancestor of genes 1–3. Then the GaBL gene duplicated into gene 3, the direct ancestor of genes 1 and 2. The gene 1/gene 2 duplication event probably occurred independently from carp in the zebrafish lineage. Although it has yet to be established whether the GaBL homologs also duplicated in carp, the branching pattern of the GaBL members in the phylogenetic tree (Fig. 6) predicts that carp also possess a zebrafish gene 3-like GaBL in addition to the gene 1/2-like GaBL. Because bony fish MBL and GalBL sequences have been identified only from cyprinid species (a family of carp relatives), confirmation of the presence of the two MBL homologs in other bony fish taxa awaits data from other teleost groups, e.g., salmonids (trout), tetradoontes (puffer fish), etc. Notwithstanding, BLAST searches of Tetraodon nigroviridis (the green puffer) and Takifugu rubripes (the tiger puffer) genome databases using carp GaBL as the query yield putative transcripts encoding both a collagenous region and a QPD-type CRD (GSTENP00034526001 from Tetraodontes and SINFURUP0000158981 from Takifugu). In contrast, similar searches using carp MBL as the query give a putative transcript that spans a neck region and an FNP-type CRD (SINFURUP0000158981) from Takifugu. These in silico findings imply that both MBL and GaBL are conserved in the puffer fish species. The apparent lack of a collagenous region-encoding sequence of MBL in the Takifugu draft genome (the fourth assembly) could be due to an incomplete assembly of the contigs and scaffolds, which might ensue from the presence of repetitive sequences, such as those encoding collagenous regions (Gly-X-Y repeats).

In conclusion, carp complement has a functional lectin pathway, which is armed with MBL and GaBL, both of which associate with MASP2 to catalyze C4 activation. The biological significance of the MBL/GaBL diversity in fish is of great relevance to gaining an insight into evolution of the lectin pathway-mediated innate immunity in vertebrates.

Acknowledgments
We thank Hayami Maeda for technical assistance and Dr. Valerie J. Smith (University of St. Andrews, St. Andrews, Scotland) for critically reading the manuscript. Dr. Kenichiro Honjo helped us in developing x-ray films in Southern hybridization. The BoxShade server (www.ch.embnet.org/software/BOX_form.html) was utilized to format the multiple alignments.

Disclosures
The authors have no financial conflict of interest.

References


