The Journal of Immunology

Two Lineages of Mannose-Binding Lectin-Associated Serine Protease (MASP) in Vertebrates

Yuichi Endo, Minoru Takahashi, Miki Nakao, Hidetoshi Saiga, Hideharu Sekine, Misao Matsushita, Masaru Nonaka and Teizo Fujita

J Immunol 1998; 161:4924-4930; ;
http://www.jimmunol.org/content/161/9/4924

The Journal of Immunology is published twice each month by
The American Association of Immunologists, Inc.,
1451 Rockville Pike, Suite 650, Rockville, MD 20852

Print ISSN: 0022-1767 Online ISSN: 1550-6606.

Two Lineages of Mannose-Binding Lectin-Associated Serine Protease (MASP) in Vertebrates[1, 2]

Yuichi Endo,[3]* Minoru Takahashi,* Miki Nakao,† Hidetoshi Saiga,‡ Hideharu Sekine,* Misao Matsushita,* Masaru Nonaka,§ and Teizo Fujita*

Mannose-binding lectin-associated serine protease (MASP) is a newly identified member of the serine protease superfamily. MASP is involved in host defense against pathogens through a novel system of complement activation, designated the lectin pathway. To elucidate the origin of the lectin pathway and the molecular evolution of MASP, we cloned six MASP cDNAs from five vertebrate species going from mammal to cyclostome. An alignment of the amino acid sequences deduced from the cDNAs revealed the presence of two different lineages of the MASP gene. This classification was supported by the precise correlation with two types of exon organization for the protease domain. One of the two lineages is unique in that a single exon encodes the protease domain, unlike most other serine proteases. All members of this group, termed the AGY type, have an AGY codon at the active site serine. A phylogenetic tree suggests that the AGY type diverged from another lineage, termed the TCN type, before the emergence of primitive vertebrates. Furthermore, the presence of MASP or MASP-like sequences in most vertebrate species suggests that the lectin pathway functions extensively in vertebrates and that its origin is traced back to the invertebrate stage. ***The Journal of Immunology,*** **1998, 161: 4924–4930.**

We previously cloned a novel human serine protease, designated mannose-binding lectin (MBL)[4]-associated serine protease (MASP) (1, 2), which is associated with MBL, a C-type serum lectin involved in a system of complement activation, designated the lectin pathway (2, 3). After MBL recognizes carbohydrate on the surface of a pathogen, the proenzyme form of MASP is converted to the active form, resulting in the proteolytic activation of the complement components C4, C2, and C3 (1, 4, 5) and finally in the killing of the pathogen. Thus, MBL-MASP complexes play a primary role in host defense during the lag period before the onset of adaptive immunity through the generation of specific Ab (6).

Recently, a second form of human MASP (MASP2) was cloned (3), which is ~41% identical to and has the same domain structure as the first form of MASP (MASP1) cloned by us. The precise composition and stoichiometry of MBL/MASP1/MASP2 complexes, as well as the enzymatic properties of the two MASPs, are unknown. The overall structures of the two human MASPs resemble the proteolytic components, human C1r and C1s, of the C1 complex in the classical complement pathway (2, 3, 7, 8). Thus, at least four MASP-related molecules, two MASPs, C1r and C1s, are present in human and are encoded by four different genes. The MASP/C1r/C1s family forms a unique branch of the serine protease superfamily, which includes members with a single exon-encoded protease domain.

It is of considerable interest to determine whether lower animals, especially the lamprey, one of the most primitive vertebrates, have a lectin pathway driven by MASP-like serine protease(s), since the genes encoding several molecules involved in adaptive immunity, such as Ig (9), TCR (10), and MHC class I (11) and II (12), can all be traced back to cartilaginous fish, but not to cyclostomes.

In this paper, we describe the cDNA cloning of six MASPs from vertebrates such as mammals (mouse), amphibians (*Xenopus*), bony fish (carp), cartilaginous fish (shark), and cyclostomes (lamprey) and discuss the molecular evolution of the MASP/C1r/C1s family including two ascidian MASPs reported recently (13). We also discuss the possible presence of a lectin pathway in these vertebrates.

Materials and Methods

Materials

Xenopus laevis was purchased from Nippon Bio-Supp. Center (Nagoya, Japan). Carp (*Cyprinus carpio*), shark (*Triakia scyllium*), and lamprey (*Lampetra japonica*) were obtained from local dealers in Fukuoka, Mie, and Fukushima, Japan, respectively. Solitary ascidian, *Halocynthia roretzi*, was harvested in Mutsu Bay, Japan.

Restriction enzymes were purchased from Toyobo (Osaka, Japan) and Boehringer Mannheim (Mannheim, Germany). [α-^{32}P]dCTP and the Megaprime DNA labeling system were from Amersham Japan (Tokyo). The ligation kit was from Takara Shuzo (Kyoto, Japan). The pGEM-T vector and λZAP II were from Promega (Madison, WI) and Stratagene (La Jolla, CA), respectively.

Nucleotide sequence analysis

DNA sequence was determined by the dideoxy chain termination method (14) using a DNA sequencer (Model 4000; LI-COR, Lincoln, NE). The labeling reaction was conducted using the SequiTherm Long-Read cycle

*Department of Biochemistry, Fukushima Medical College, 1-Hikarigaoka, Fukushima, Japan; †Department of Fisheries Science, Kyusyu University, Hakozaki, Fukuoka, Japan; ‡Department of Biological Sciences, Graduate School of Science, Tokyo Metropolitan University, Hachioji, Japan; and §Department of Biochemistry, Nagoya City University Medical School, Mizuho-ku, Nagoya, Japan

Received for publication March 23, 1998. Accepted for publication June 22, 1998.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[1] This work was supported in part by grants from the Ministry of Education, Science, and Culture of Japan.

[2] The nucleotide sequence data reported in this paper will appear in the DDBJ, EMBL, and GenBank nucleotide sequence databases with the following accession numbers: AB009459 for muMASP2 cDNA, D83276 for xeMASP1 cDNA, AB009072 for xeMASP2 cDNA, AB009073 for caMASP cDNA, AB009074 for shMASP cDNA, AB009075 for laMASP cDNA, AB10813–22 for huMASP1 genomic DNAs, and AB009076 for huC1s genomic DNA.

[3] Address correspondence and reprint requests to Dr. Yuichi Endo, Department of Biochemistry, Fukushima Medical College, 1-Hikarigaoka, Fukushima 960-1295, Japan.

[4] Abbreviations used in this paper: MBL, mannose-binding lectin; MASP, MBL-associated serine protease; mu/hu/ca/sh/la/xe/asMASP, mouse/human/carp/shark/lamprey/*Xenopus*/ascidian MASP; RACE, rapid amplification of cDNA end; SCR, short consensus repeat.

0022-1767/98/$02.00

sequencing kit (Epicentre Technologies, Madison, WI). Sequencing primers were synthesized by Nisinbo (Tokyo).

Cloning of MASP cDNAs

RNA was isolated from liver using the acid guanidine thiocyanate method (15), and the poly(A)$^+$ fraction was purified by passage through an oligo(dT)-cellulose column (Clontech, Palo Alto, CA). The liver cDNAs and the cDNA libraries in λZAP II were prepared as described (16, 17).

A partial cDNA sequence of the mouse counterpart (muMASP2) of human MASP2 (huMASP2) was first obtained from the mouse expression sequence tags (EST) gene bank library (GenBank) by a homology search with the huMASP2 cDNA sequence (3). To complete the 5′ coding sequence, a 5′ RACE (rapid amplification of cDNA end) was conducted using a kit (Marathon; Clontech) with mouse liver cDNA as a template.

The cDNA of the *Xenopus* homologue (xeMASP1) of human MASP1 (huMASP1) was cloned by screening a *Xenopus* liver cDNA library with a ^{32}P-labeled cDNA fragment of huMASP1 (2.8-kbp of *Eco*RI fragment; Ref. 2) as a probe. A positive clone, X43, was subcloned in pBluescript II SK by in vivo excision according to the manufacturer's instructions (Stratagene), then further subcloned into pBluescript II KS at internal restriction sites and sequenced. 5′ RACE and 3′ RACE were performed to complete the cDNA sequence.

The cDNA of the *Xenopus* homologue (xeMASP2) of huMASP2 was amplified by RT-PCR with degenerated primers derived from the conserved amino acid sequences in huMASP1, mouse MASP1 (muMASP1; mouse P-100 in Ref. 8) and xeMASP1: IKLYFMH (YF1, 5′-ATIAAA/GCTITAT/CTTT/CATGCA-3′, where I represents inosine); YLCEYDY (YF2, 5′-TAT/CCTITGT/CGAA/GTAT/CGAT/CTA-3′), CGEKSP (YR1, 5′-GGIG/CA/T/CT/CTTT/CTCICCA/GCA-3′); and CPYDYL/IK (YR2, 5′-TTIAG/TA/GTAA/GTCA/GTAIGGA/GCA-3′). The PCR products with the expected sizes were cloned into pGEM-T vector and sequenced. A *Xenopus* liver cDNA library was then screened using the insert DNA as a probe, and positive clones (X112 and X161) were sequenced.

A nested RT-PCR with the same degenerated primers was performed to isolate carp MASP (caMASP) cDNA. Carp liver cDNA library was then screened using the ^{32}P-labeled PCR product as a probe. A positive clone, 2A6, was subcloned and sequenced. A 5′ RACE was conducted to complete the 5′ coding sequence.

Shark MASP (shMASP) cDNA was isolated by a similar RT-PCR. 5′ RACE and 3′ RACE were then conducted with the primers corresponding to the nucleotide sequence of the PCR product and shark liver cDNA as a template. All of the sequences of the PCR and 5′/3′ RACE products were determined by sequencing three or more clones to rule out PCR error.

A similar RT-PCR was employed to amplify a portion of lamprey MASP (laMASP) cDNA. A lamprey liver cDNA library was then screened with the ^{32}P-labeled PCR product as a probe. Three positive clones, L402, L410, and L412, were subcloned and sequenced.

Cloning of genomic DNA of the huMASP1 and human C1s genes

To determine the exon-intron structure of the huMASP1 gene, the genomic DNA was cloned by screening EMBL-3 (Clontech) and λFIX II phage libraries (Stratagene) with a ^{32}P-labeled full length of huMASP1 cDNA (2) as a probe. The overlapping clones (F419, F302, F210, F301, F21, and A1) were subjected to restriction mapping, Southern blot hybridization with ^{32}P-labeled fragment of huMASP1 cDNA, and subcloning at the internal restriction sites and then sequenced.

The human C1s (huC1s) gene was analyzed by PCR using the sequences of huC1s cDNA as primers and the genomic DNA from human peripheral mononuclear cells as a template. To design primer sequences, the putative positions of intron insertion in the huC1s gene were estimated by referring to those in the huMASP1 gene. The PCR products obtained were cloned in pGEM-T vector and sequenced.

PCR to amplify the gene encoding the protease domain of MASP

Genomic DNA was prepared as described (18). To determine whether the protease domain of each MASP is encoded by a single exon or by split exons, PCR was performed using the genomic DNA as a template. A primer set was designed to amplify the entire protease domain: 5′-TTTGTGGACTATCAGCCCGCACAAC-3′ and 5′-CCTTGAGTCAATGGGTAAGGCTGGA-3′ for huMASP2; 5′-TTTCCAGGTCTGCTCTAGCAAGA-3′ and 5′-AGACACTGGGCTGTCACTTCTAC-3′ for xeMASP1; 5′-AGCAGTAGCTCGCGCACGTATAG-3′ and 5′-AGTCAGTAATATGTATTTCCATATCA-3′ for xeMASP2; 5′-TCTGTGGAAGGCCTGCTCGACC-3′ and 5′-ATGTTCAATCAGCAGGATCCTTGC-3′ for shMASP; 5′-GCCCAAGCACGTGATAATGGAGTTG-3′ and 5′-ACTGACAAACGCGCGTCGTGATCG-3′ for laMASP; 5′-GAAGAAGCTCTACAACGATCCTGC-3′ and 5′-ACAAAGCGGGTCATTCAATCCAGG-3′ for ascidian MASPa (asMASPa; Ref. 13); and 5′-AGCCGAAGACAATCCACTACTGGC-3′ and 5′-ATGATCCATTCAACGACTGTTTGC-3′ for asMASPb. The PCR product was visualized on an agarose gel following electrophoresis by staining with ethidium bromide.

Construction of phylogenetic tree

The 14 members of the MASP/C1r/C1s family including huMASP1 (2, 7), huMASP2 (3), muMASP1 (8), muMASP2, xeMASP1, xeMASP2, caMASP, shMASP, laMASP, asMASPa and -b (13), huC1r (19, 20), huC1s (21, 22), and golden hamster C1s (ghC1s; CASP in Ref. 23) were aligned at the primary structure level using Clustal w software (EMBL Data Library, Heidelberg, Germany). A pairwise distance matrix was obtained by calculating the proportion of the different amino acids. The matrix was then used to construct trees by the neighbor-joining method (24). Bootstrap analysis was used to assess the reliability of branching patterns. For each tree, 1000 bootstrap replications were performed.

Results and Discussion

Primary structures of MASPs deduced from the cDNAs

The cDNAs we obtained of muMASP2, xeMASP1 and -2, caMASP, shMASP, and laMASP encoded 685, 698, 688, 745, 720, and 722 amino acids, respectively. An alignment was made of the amino acid sequences of these proteins in addition to those of huMASP1 and -2 and muMASP1 (Fig. 1). All of the deduced amino acid sequences consisted of putative signal peptides followed by the secreted proenzyme forms of MASP. The latter consisted of six domains, which are also conserved in the previously cloned members of the MASP/C1r/C1s family. Three amino acid residues (His^{489}, Asp^{551}, and Ser^{645}, numbering in xeMASP1) essential for the active center of serine protease (25) and the proteolytic cleavage site between Arg^{446} and Ile^{447}, essential for generating the active form of MASP comprising two polypeptides (heavy and light chains), were also conserved. All members have an aspartic acid residue at −6 relative to the active site serine, a substrate-related residue (25), suggesting their trypsin-type substrate specificity.

A comparison of the protease domains revealed two different types in vertebrates. The xeMASP1 has two additional cysteines (Cys^{474} and Cys^{490}) forming the "histidine loop" disulfide bridge (26) in their protease domains in the same manner as huMASP1 and muMASP1, whereas the other members lacked these residues, as do huC1r/C1s. Two types of codons, TCN and AGY, for the active site serine were found: TCT in xeMASP1, AGC in muMASP2, and laMASP and AGT in xeMASP2, caMASP, and shMASP.

The calculated percentages of identity of each MASP to other members at the amino acid sequence level are shown in Figure 2. The high percentages of identity of huMASP1 vs muMASP1 and huMASP2 vs muMASP2 strongly suggest that muMASP1 and -2 are mouse counterparts of huMASP1 and -2, respectively. The significantly higher percentages of identity of xeMASP1 vs huMASP1 and xeMASP2 vs huMASP2 also indicate that xeMASP1 and -2 are *Xenopus* counterparts of huMASP1 and -2, respectively. Both caMASP and shMASP have slightly higher percentages of identity to the MASP1 group rather than to the MASP2 group, although the structural features of their protease domains, such as the absence of the histidine loop disulfide bridge and an AGY codon at the active site serine, are commonly shared with the MASP2 group. This suggests that the evolutionary history of MASP is not simple and that both caMASP and shMASP are direct counterparts of neither MASP1 nor MASP2 (see below). The laMASP showed 36.6 to 41.3% identity with other vertebrate MASPs and 37.3 to 37.5% with huC1r/C1s, respectively. This

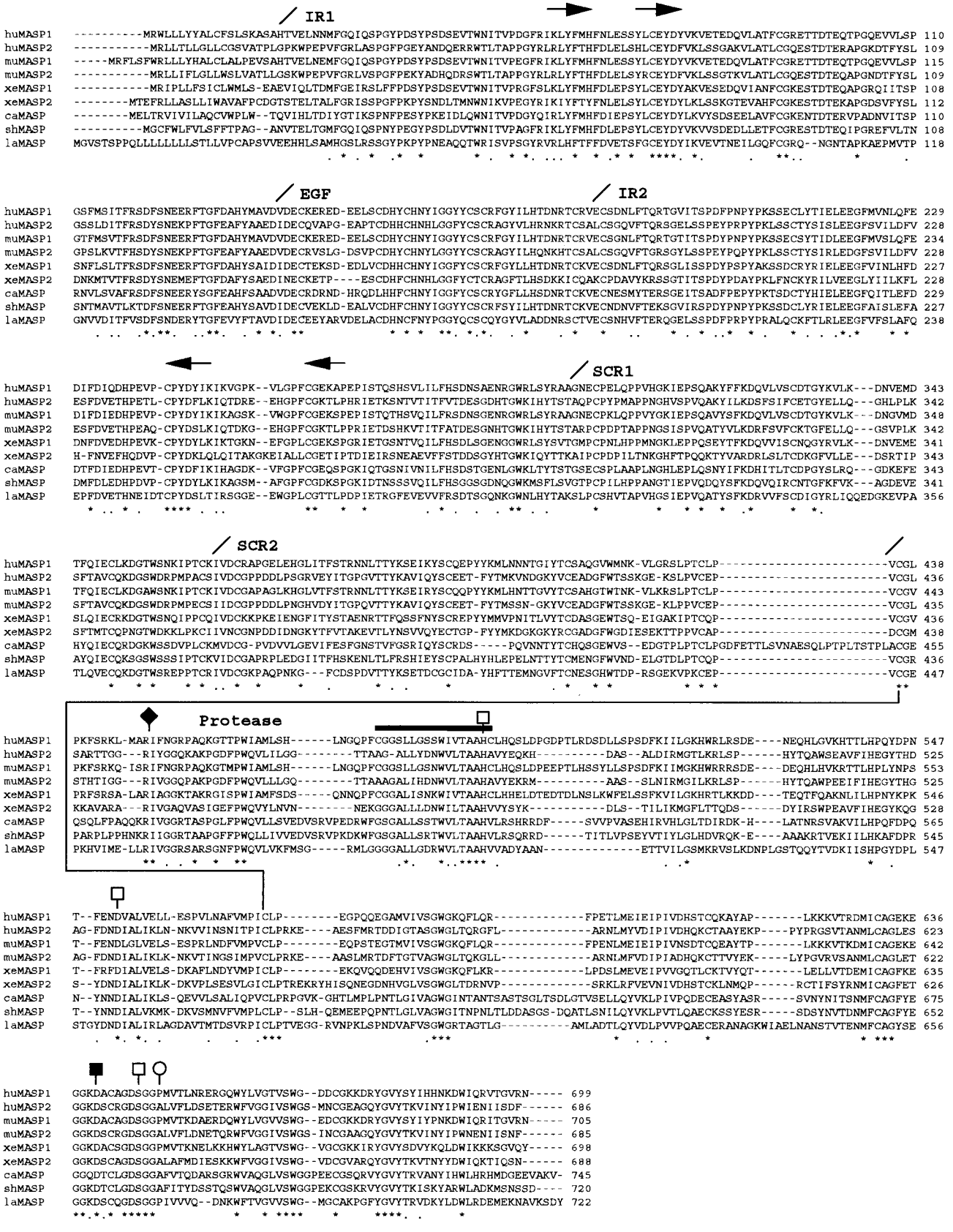

FIGURE 1. Alignment of the amino acid sequences of six vertebrate MASPs with those of previously cloned mammalian MASPs (huMASP1 and -2 and muMASP1). Gaps introduced to increase identity are shown by dashes. The completely conserved residues and the residues of similar nature among these nine sequences are indicated by asterisks and dots, respectively, below the sequences. The N termini of each domain are marked by slashes above the sequences: the first internal repeat (IR1), epidermal growth factor (EGF)-like, the second internal repeat (IR2), two SCRs (SCR1 and -2), and protease domains. Three residues essential for the catalytic site, the residue at the substrate specificity crevice, and the residue at site 3 relative to the active site serine (see text) are marked by open squares, closed square, and open circle, respectively. The proteolytic cleavage site between arginine and isoleucine, which generates heavy and light chains, is marked by a closed diamond. The thick, horizontal line represents the histidine loop disulfide bridge. PCR primer sites used for cDNA cloning are indicated by arrows above the sequences.

	huMASP2	muMASP1	muMASP2	xeMASP1	xeMASP2	caMASP	shMASP	laMASP	asMASPa	asMASPb	huC1r	huC1s
huMASP1	40.7	86.5	38.0	61.5	38.0	52.2	50.2	41.1	28.6	31.2	36.1	35.7
huMASP2		42.1	81.1	40.5	50.2	45.2	41.0	40.4	32.5	31.6	38.4	38.3
muMASP1			39.1	60.4	38.3	50.7	52.7	40.8	29.6	31.0	36.5	36.2
muMASP2				38.2	49.1	42.0	43.5	41.0	33.6	34.4	39.5	39.8
xeMASP1					37.9	55.2	52.8	41.3	31.3	30.0	35.6	36.1
xeMASP2						40.5	39.5	36.6	28.8	29.8	37.5	37.5
caMASP							39.1	40.7	27.0	29.2	35.5	38.3
shMASP								40.5	28.1	29.2	34.1	36.6
laMASP									29.4	29.8	37.3	37.5
asMASPa										43.9	28.0	28.4
asMASPb											29.4	26.4
huC1r												37.4

FIGURE 2. A calculated amino acid identity among the MASP/C1r/C1s family. Percent identity was calculated based on the alignment of amino acid sequences.

result, together with an apparent lack of the classical complement pathway in the lamprey (27), suggests that the divergence of cyclostomes predated the gene duplication between MASP and C1r/C1s.

Transcripts of MASP in liver

The presence of transcripts in liver corresponding to the above cDNAs was confirmed by Northern blot hybridization for liver mRNA using our cDNAs as probes: 3.6-, 2.7-, and 1.2-kb-long transcripts of muMASP2; 5.0 and 2.6 kb of xeMASP1; 2.9 and 1.2 kb of xeMASP2; 4.7, 3.0, and 1.7 kb of caMASP; 4.0, 2.9, 2.3, and 1.8 kb of shMASP; and 5.0, 3.1, and 1.0 kb of laMASP (data not shown).

Exon-intron structures of the huMASP1 and huC1s genes

The entire huMASP1 gene spans more than 50 kbp and contains at least 16 exons (Fig. 3). The first internal repeat, epidermal growth factor (EGF)-like, the second internal repeat, and the first and the second short consensus repeat (SCR) domains are encoded by 2, 1, 2, 2, and 2 exons, respectively. As reported previously (28), the protease domain of huMASP1 is encoded by six exons. All of the splice acceptor and donor sequences were consistent with the AG/GT rule.

We found that the huC1s gene spans ~13 kbp and contains 12 exons (Fig. 3). Compared with the huMASP1 gene, the huC1s gene is very compact owing to the short introns. It was also found that the positions of introns were completely conserved between the huMASP1 and the huC1s genes throughout the regions encoding the heavy chain (five N-terminal domains). As reported by Tosi et al. (29), the protease domain of huC1s is uniquely encoded by a single exon, as is huC1r (28, 29). Interestingly, the position of the intron in the N-terminal region of the protease domain is also conserved between the two genes, where both Val^{435} of huMASP1 and the homologous Val^{409} of C1s are interrupted by the introns at phase 1. These results clearly suggest that the MASP/C1r/C1s family was essentially generated from a common ancestor, regardless of the structural difference in their protease domains.

Exon organization of the region encoding the protease domain of MASPs

As shown in Figure 4, PCR amplification of the genomic regions encoding the protease domains of huMASP2, xeMASP2, caMASP, shMASP, and laMASP resulted in products with the same sizes as those obtained by PCR with the respective cDNAs as templates. This suggests that their protease domains are encoded by an intronless exon. In contrast, PCR amplification of the region encoding the protease domain of xeMASP1 resulted in no product with the same size as that from a similar PCR with the cDNA as a template, suggesting that the protease domain of xeMASP1 is encoded by multiple exons. These results clearly demonstrate that two lineages of the MASP genes are present in vertebrates.

A similar amplification of the genomic DNA encoding the protease domains of two invertebrate MASPs (asMASPa and -b) resulted in PCR products with larger sizes (2 kbp for asMASPa and 1.7 kbp for asMASPb) than those generated by PCR with the cDNA as templates (Fig. 4), suggesting that the protease domains of these MASPs are encoded by multiple exons. By a preliminary sequencing of the PCR products, the protease domains of asMASPa and -b were found to be encoded by at least seven and five exons, respectively (data not shown). Two of all of the positions of

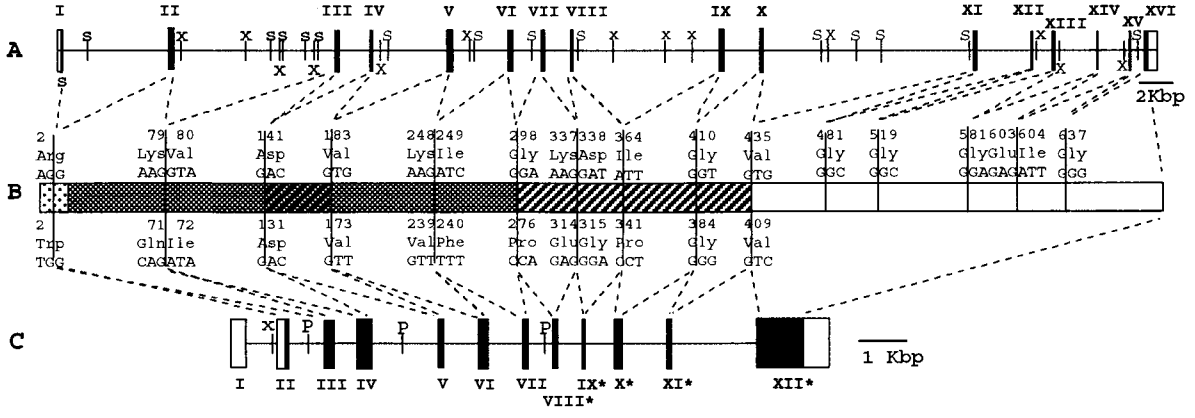

FIGURE 3. Structures of the genes encoding huMASP1 and huC1s. *A*, Restriction map and exon-intron structure of the huMASP1 gene. The exons for the protease domain from XI to XVI were reported previously (28). *B*, Schematic representation of the domain structure of huMASP1 and huC1s with the nucleotide and amino acid sequences of exon-intron boundaries. The relationship between the genome structure and protein domain structure is indicated with dotted lines. *C*, Restriction map and exon-intron structure of the huC1s gene. Exon numbers with asterisks indicate the exons reported by Tosi et al. (29). Restriction enzyme sites are indicated by: S, *Sac*I; X, *Xba*I; P, *Pst*I; X, *Xba*I. Shaded and open boxes indicate translated and untranslated regions of exon, respectively.

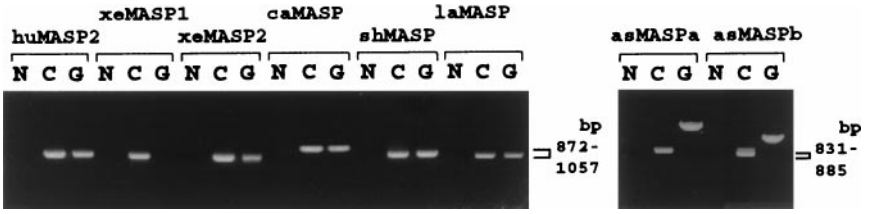

FIGURE 4. Amplification of the region encoding the protease domain of MASP. Genomic DNA was amplified by PCR (*G lanes*) using the primers that flank the protease domains of huMASP2, xeMASP1 and -2, caMASP, shMASP, laMASP, and asMASPa and -b. The same regions were also amplified starting from uncloned liver cDNA (*C lanes*). Control PCR without template DNA (*N lanes*).

introns were conserved between asMASPa and -b genes. One of the two was quite similar to the most downstream position of intron of the huMASP1 gene.

Phylogeny of the MASP/C1r/C1s family

A phylogenetic tree constructed based on the alignment of the entire amino acid sequences shows that caMASP and shMASP formed a branch together with the MASP1 group (Fig. 5*A*). The high percentages of bootstrap analysis support this branching. A phylogenetic tree based on the regional sequences of the heavy chains show essentially the same branching pattern (data not shown). As shown in Figure 5*B*, however, another phylogenetic tree based on the regional amino acid sequences of the protease domains show that these two MASPs form a unique branch together with the laMASP, MASP2, and C1r/C1s groups. Although bootstrap values supporting this branching pattern are not fully convincing, these results may suggest that the heavy and light chain domains of MASP followed separate evolutionary pathways (see below).

Figure 5*B* suggests that a single evolutionary event before the emergence of primitive vertebrates may have generated two lineages. One is the MASP1 group, termed the TCN type, which possesses a set of structural features such as the presence of the histidine loop in the protease domain, TCN codon at active site serine, a proline residue at position 3 relative to the active site serine, and split exons for the protease domain. Another group could include MASP2, caMASP, shMASP, laMASP, and huC1r/C1s. This group, termed the AGY type, is characterized by the absence of the histidine loop, an AGY codon at the active site serine, an alanine or valine residue at position 3 relative to the serine, and a single exon for the protease domain. Accumulating evidence seems to support the notion that the TCN type is a prototype of the MASP/C1r/C1s family. TCN-type MASP possesses a histidine loop, which is seen in most serine proteases, regardless of whether they are TCN type or AGY type, including digestive enzymes, blood coagulation factors, and complement serine proteases such as factor B (16, 17, 30). Furthermore, the AGY codon for the active site serine was reportedly found only in the serine protease genes of vertebrates (31, 32). Actually, the active site serines of asMASPa and -b are encoded by the TCG and TCT codon (13), respectively. The proline residue at position 3 relative to the active site serine is conserved in most of the above-mentioned serine proteases. Proline in laMASP is the only exception, suggesting that it contains an ancient form of the AGY type.

The question arises as to whether the lower vertebrates such as carp, shark, and lamprey have a TCN-type MASP. To date, our repeated attempts using PCR with a variety of primers have failed to isolate such a clone from liver cDNA in these species. A plausible explanation is that although the expected genes are present in the genomes of these species, they are silent or at least not expressed in liver.

Figure 6 shows a possible mechanism for generating an AGY-type MASP/C1r/C1s from a prototype of TCN-type MASP. Before the emergence of vertebrates, retroposition of the TCN-type MASP gene and base changes from TCN to AGY at the active site serine would generate a prototype of the AGY-type MASP gene. A partially processed mRNA (33) consisting of an unprocessed sequence for heavy chain and a precisely processed sequence for light chain (protease domain) would be transpositioned. A similar retroposition event was reported in the generation of the preproinsulin I gene from the preproinsulin II gene (34). Owing to generation of the AGY-type MASP gene, the TCN-type MASP gene might have become a silent gene in primitive vertebrates. At the evolutionary stage between cyclostomes and cartilaginous fish, the

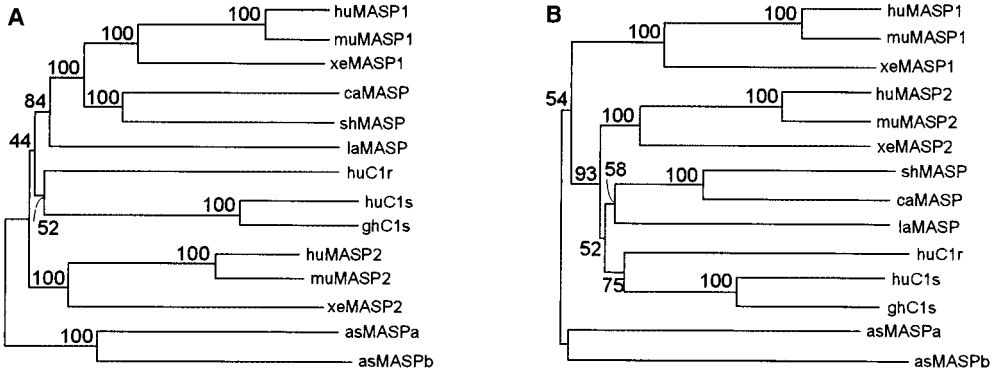

FIGURE 5. Phylogenetic trees of the MASP/C1r/C1s family. Fourteen members of the MASP/C1r/C1s family were aligned by Clustal w software using their entire amino acid sequences (*A*) and the regional sequences of protease domains (*B*); the trees were constructed using neighbor-joining method. Numbers on branches are bootstrap percentages supporting a given partitioning.

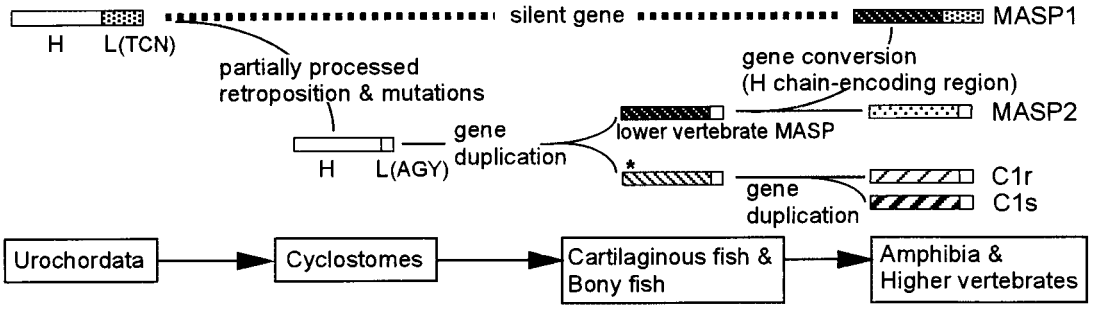

FIGURE 6. A model for the evolution of the MASP/C1r/C1s family gene. The MASP gene is schematically presented by marked boxes. The H and L under the schematic box of the MASP gene represent the regions directed for the heavy and light chains, respectively. The asterisk represents a putative C1r/C1s gene in cartilaginous and bony fish.

AGY-type MASP gene might have doubled by gene duplication to produce a prototype of the C1r/C1s genes. Jensen et al. have presented evidence for the classical pathway in nurse shark, a cartilaginous fish, comprising six complement components (35). One of the six, termed C1n, is thought to be equivalent to mammalian C1 complex, the structure of which remains to be elucidated. Before the divergence of amphibians, a gene conversion event could have replaced the region of the AGY-type MASP gene containing the promoter and the region for the heavy chain with the homologous region of a silent TCN-type gene, bringing about a MASP1 gene. At this stage, the AGY-type MASP gene could have changed rapidly, in turn, to MASP2 to share the function with MASP1.

Lectin pathway in vertebrates

The presence of MASPs in all of the species analyzed in this study strongly suggests that the lectin pathway functions extensively in the vertebrate line. The structural similarity between shMASP and caMASP suggests that cartilaginous fish, bony fish, and probably cyclostomes may have lectin pathway similar to each other, which is slightly different from that in mammals and amphibians. At the stage of amphibian evolution, the lectin pathway may have developed to such a point that two the MASPs, MASP1 and MASP2, function cooperatively in a complex. The presence of the two types of MASP in mammals, amphibians, and ascidians, regardless of the structural difference in the protease domains, leads us to speculate that these two MASPs have distinct functions. It was reported that huMASP1 cleaves C3, whereas huMASP2 cleaves C4 (3). The difference in the specificity of the substrate molecule between huMASP1 and -2 seems to be due to structural differences in the heavy chains, especially in the SCR domains (36) of both proteins. It is likely that the MASPs in lower vertebrates activate C3-like molecules as does MASP1, because their heavy chains are closely related to those of MASP1. Further studies of MASP proteins will clarify the mechanisms underlying the lectin pathway in these vertebrates.

In conclusion, the present phylogenetic analysis of the MASP/C1r/C1s family has revealed the presence of two different lineages of the MASP gene in vertebrates and suggests that one of the two, a unique lineage with a single exon-encoded protease domain, may have diverged from a common ancestor before the emergence of vertebrates. Compared with the origin of the classical complement pathway at the stage of cartilaginous fish (35, 37, 38), the lectin pathway seems to have a more ancient origin.

Acknowledgments

We thank Dr. H. Schulman for English editing and Ms K. Kanno for technical assistance.

References

1. Matsushita, M., and T. Fujita. 1992. Activation of the classical complement pathway by mannose-binding protein in association with a novel C1s-like serine protease. *J. Exp. Med. 176:1497.*
2. Sato, T., Y. Endo, M. Matsushita, and T. Fujita. 1994. Molecular characterization of a novel serine protease involved in activation of the complement system by mannose-binding protein. *Int. Immunol. 6:665.*
3. Thiel, S., T. Vorup-Jensen, C. M. Stover, W. Schwaeble, S. B. Laursen, K. Poulsen, A. C. Willis, P. Eggleton, S. Hansen, U. Holmskov, K. B. M. Reid, and J. C. Jensenius. 1997. A second serine protease associated with mannan-binding lectin that activates complement. *Nature 386:506.*
4. Ji, Y., T. Fujita, H. Hatsuse, A. Takahashi, M. Matsushita, and M. Kawakami. 1993. Activation of the C4 and C2 components of complement by a protease in serum bacterial factor, Ra reactive factor. *J. Immunol. 150:571.*
5. Matsushita, M., and T. Fujita. 1995. Cleavage of the third component of complement (C3) by mannose-binding protein-associated serine protease (MASP) with subsequent complement activation. *Immunobiology 194:443.*
6. Sastry, K., and R. A. Ezekowitz. 1993. Collectins: pattern recognition molecules involved in first line host defense. *Curr. Opin. Immunol. 5:59.*
7. Takada, F., H. Takayama, H. Hatsuse, and M. Kawakami. 1994. A new member of the C1s family of component protein found in a bactericidal factor, Ra-reactive factor, in human serum. *Biochem. Biophys. Res. Commun. 196:1003.*
8. Takayama,Y., F. Takada, A. Takahashi, and M. Kawakami. 1994. A 100-KDa protein in the C4-activating component of Ra-reactive factor is a new serine protease having module organization similar to C1r and C1s. *J. Immunol. 152: 2308.*
9. Hinds, K. R., and G. W. Litman. 1986. Major reorganization of immunoglobulin VH segmental elements during vertebrate evolution. *Nature 320:546.*
10. Rast, J. P., and G. W. Litman. 1994. T-cell receptor gene homologs are present in the most primitive jawed vertebrates. *Proc. Natl. Acad. Sci. USA 91:9248.*
11. Hashimoto, K., T. Nakanishi, and Y. Kurosawa. 1992. Identification of a shark sequence resembling the major histocompatibility complex class I a3 domain. *Proc. Natl. Acad. Sci. USA 89:2209.*
12. Kasahara, M., M. Vazqez, K. Sato, E. C. McKinney, and M. F. Flajnik. 1992. Evolution of the major histocompatibility complex: isolation of class II A cDNA clones from the catilaginous fish. *Proc. Natl. Acad. Sci. USA 89:6688.*
13. Ji, X., K. Azumi, M. Sasaki, and M. Nonaka. 1997. Ancient origin of the complement lectin pathway revealed by molecular cloning of mannan binding protein-associated serine protease from a urochordate, the Japanese ascidian, *Halocynthia roretzi. Proc. Natl. Acad. Sci. USA 94:6340.*
14. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA 74:5463.*
15. Chomczynski, P., and N. Sacchi. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem. 162: 156.*
16. Nonaka, M., M. Takahashi, and M. Sasaki. 1994. Molecular cloning of a lamprey homologue of the mammalian MHC class III gene, complement factor B. *J. Immunol. 152:2263.*
17. Kato, Y., L. Salter-Cid, M. F. Flajnik, M. Kasahara, C. Namikawa, M. Sasaki, and M. Nonaka. 1994. Isolation of the *Xenopus* complement factor B complementary DNA and linkage of the gene to the frog MHC. *J. Immunol. 153:4546.*
18. Gross-Bellard, M., P. Oudet, and P. Chambon. 1972. Isolation of high-molecular-weight DNA from mammalian cells. *Eur. J. Biochem. 36:32.*
19. Journet, A., and M. Tosi. 1986. Cloning and sequencing of full-length cDNA encoding the precursor of human complement C1r. *Biochem. J. 240:783.*
20. Lytus, S. P., K. Kurachi, K. S. Sakariassen, and E. W. Davie. 1986. Nucleotide sequence of the cDNA coding for human complement C1r. *Biochemistry 25: 4855.*
21. Mackinnon, C. M., P. E. Carter, S. J. Smyth, B. Dunbar, and J. E. Fothergill. 1987. Molecular cloning of cDNA for human complement C1s. *Eur. J. Biochem. 169:547.*

22. Tosi, M., C. Duponchel, T. Meo, and C. Julier. 1987. Complete cDNA sequence of human complement C1s and close physical linkage of the homologous genes C1s and C1r. *Biochemistry 26:8516.*
23. Kinoshita, H., H. Sakiyama, K. Tokunaga, S. Imajho-Ohmi, Y. Hamada, K. Isono, and S. Sakiyama. 1989. Complete primary structure of a calcium-dependent serine protease capable of degrading extracellular matrix proteins. *FEBS Lett. 250:411.*
24. Saitou, N., and M. Nei. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol. 4:406.*
25. Kraut, J. 1977. Serine proteases: structure and mechanism of catalysis. *Annu. Rev. Biochem. 46:331.*
26. Arlaud, G. J., and J. Gagnon. 1981. C1r and C1s subcomponents of human complement: two serine proteases lacking the "histidine-loop" disulphide bridge. *Biosci. Rep. 1:779.*
27. Nonaka, M., and M. Takahashi. 1992. Complete complementary DNA sequence of the third component of complement of lamprey. *J. Immunol. 148:3290.*
28. Endo, Y., T. Sato, M. Matsushita, and T. Fujita. 1996. Exon structure of the gene encoding the human mannose-binding protein-associated serine protease light chain: comparison with complement C1r and C1s genes. *Int. Immunol. 8:1355.*
29. Tosi, M., C. Duponchel, T. Meo, and E. Couture-Tosi. 1989. Complement genes C1r and C1s feature an intronless serine protease domain closely related to haptoglobin. *J. Mol. Biol. 208:709.*
30. Campbell, R. D., and R. R. Porter. 1983. Molecular cloning and characterization of the gene coding for human complement protein factor B. *Proc. Natl. Acad. Sci. USA 80:4464.*
31. Irwin, D. M. 1988. Evolution of modern proteins. *Nature 336:429.*
32. Brenner, S. 1988. The molecular evolution of genes and proteins: a tale of two serines. *Nature 344:528.*
33. Weiner, A. M., P. L. Deininger, and A. Efstratiadis. 1986. Nonviral retroposons: genes, pseudogenes, and transposable elements generated by reverse flow of genetic information. *Annu. Rev. Biochem. 55:631.*
34. Soares, M. B., E. Schon, A. Henderson, S. K. Karathanasis, R. Cate, S. Zeitlin, J. Chirgwin, and A. Efstratiadis. 1985. RNA-mediated gene duplication: the rat preproinsulin I gene is a functional retroposon. *Mol. Cell. Biol. 5:2090.*
35. Jensen, J. A., E. Festa, D. Smith, and M. Cayer. 1981. The complement system of the nurse shark: hemolytic and comparative characteristics. *Science 214:566.*
36. Arlaud, G., M. G. Colomb, and J. Gagnon. 1987. A functional model of the human C1 complex. *Immunol. Today 8:106.*
37. Farris, T. C., and J. P. Atkinson. 1991. Evolution of complement system. *Immunol. Today 12:295.*
38. Dodds, A. W., and A. J. Day. 1993. The phylogeny and evolution of the complement system. In *Complement in Health and Disease*. K. Whaley, M. Loos, and J. M. Weiler, eds. Kluwer, Boston, p. 39.