

Minireview

# Evolution of the initiating enzymes of the complement system

## Masaru Nonaka and Seita Miyazawa

Address: Department of Biological Sciences, Graduate School of Science, University of Tokyo, Hongo, Tokyo 113-0033, Japan.

Correspondence: Masaru Nonaka. E-mail: mnonaka@biol.s.u-tokyo.ac.jp

Published: 14 December 2001

Genome **Biology** 2001, **3(1)**:reviews1001.1–1001.5

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2001/3/1/reviews/1001>

© BioMed Central Ltd (Print ISSN 1465-6906; Online ISSN 1465-6914)

### Abstract

Analysis of the human *MASP-1/3* gene, which encodes two proteases of the lectin-triggered complement cascade, has revealed alternatively used serine-protease-encoding regions for the gene's two protein products. Phylogenetic studies indicate that one arose by retrotransposition early in vertebrate evolution, supporting the idea that the lectin branch of the complement cascade arose earlier than the 'classical' pathway.

The mammalian complement system, as one of the major constituents of innate immunity, plays a pivotal role in defending the body against pathogenic microorganisms. It has more than 30 components, organized into three major pathways: the 'classical' pathway, triggered by antibodies; the 'lectin' pathway, triggered by engagement of mannan-binding protein; and the 'alternative' pathway, triggered directly by pathogens (see Table 1). Of the 30 complement components, about one-fourth are serine proteases that are important for the activation or regulation of the whole system. Among these, C1r and C1s of the classical pathway and the MASPs (mannan-binding-protein-associated serine proteases) of the lectin pathway are crucial for starting the proteolytic activation cascades of their respective pathways. These components constitute a serine-protease family of the chymotrypsin type but characterized by a unique modular structure (Figure 1) [1-3].

Initial structural analysis of the *C1s* gene revealed a surprising exon-intron organization, in which the serine-protease domain is encoded by a single uninterrupted exon, although there are several introns throughout the remainder of the gene (Figure 1) [4]. Intronless genes or pseudogenes, thought to be generated by retrotransposition, are not rare, but it was curious to imagine that retrotransposition could have led to the removal of introns from only one part of the gene. Later, the *C1r* and *MASP-2* genes were shown to have essentially the same exon-intron organization as the *C1s*

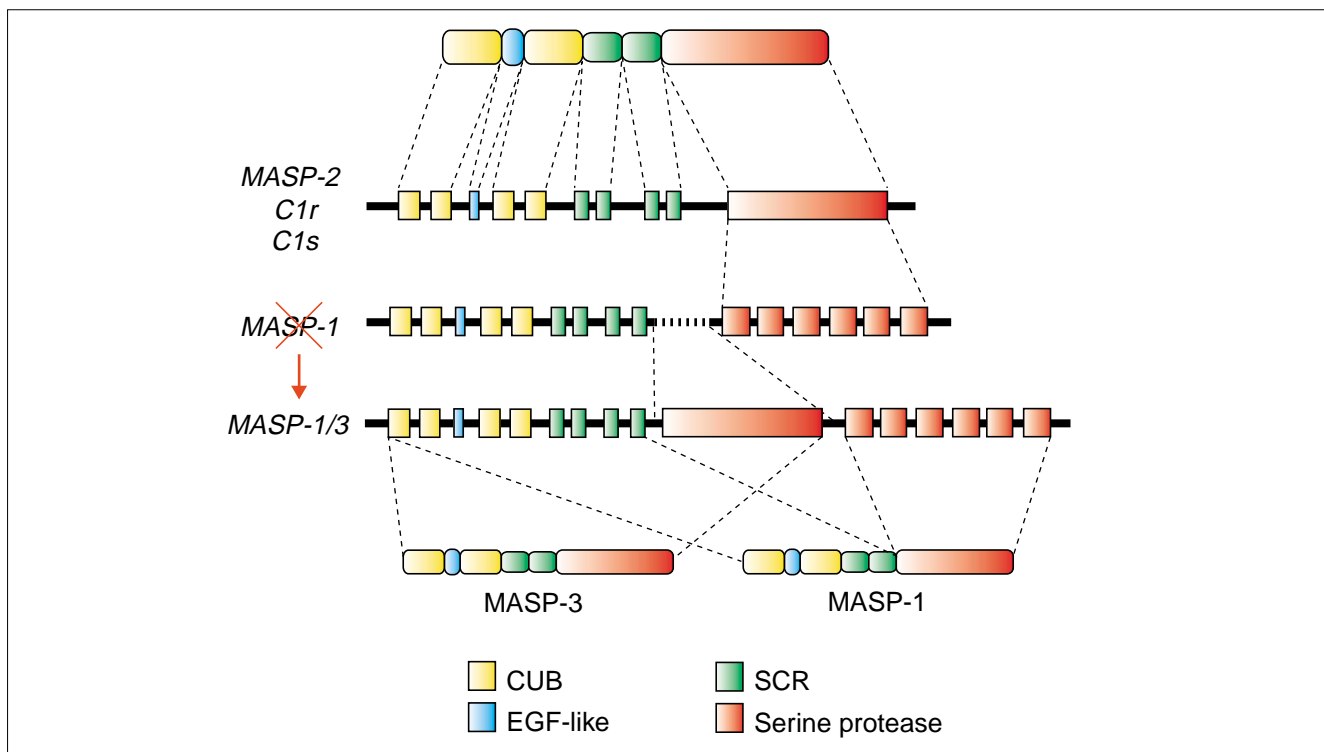
gene; surprisingly, though, *MASP-1* proved to have introns within its serine-protease-encoding region (Figure 1) [5].

The recent identification of *MASP-3* [3] provided an explanation of how this situation might have arisen. A new intronless serine-protease-encoding region was found in the *MASP-1* gene, upstream of the *MASP-1* serine-protease-encoding region. Furthermore, the two serine-protease-encoding regions were differentially selected by alternative splicing to generate *MASP-1* and *MASP-3*, which differ only in their serine protease domains (Figure 1); the gene encoding them was therefore renamed *MASP-1/3*. In addition to the absence of introns, the upstream protease domain has characteristics that distinguish it from most of the chymotrypsin-like serine-protease family: it lacks the disulfide bond that is believed to stabilize the structure around the active-site histidine (the histidine loop), and its active-site

**Table 1**

**A summary of the three branches of the complement activation pathway**

	Classical pathway	Lectin pathway	Alternative pathway
Triggering factor	Antigen-antibody complexes	Mannan-binding lectin	Pathogen (directly)
Proteases involved	C1r, C1s and C2	MASP-1, MASP-2 and MASP-3	Factor D and Factor B



**Figure 1**

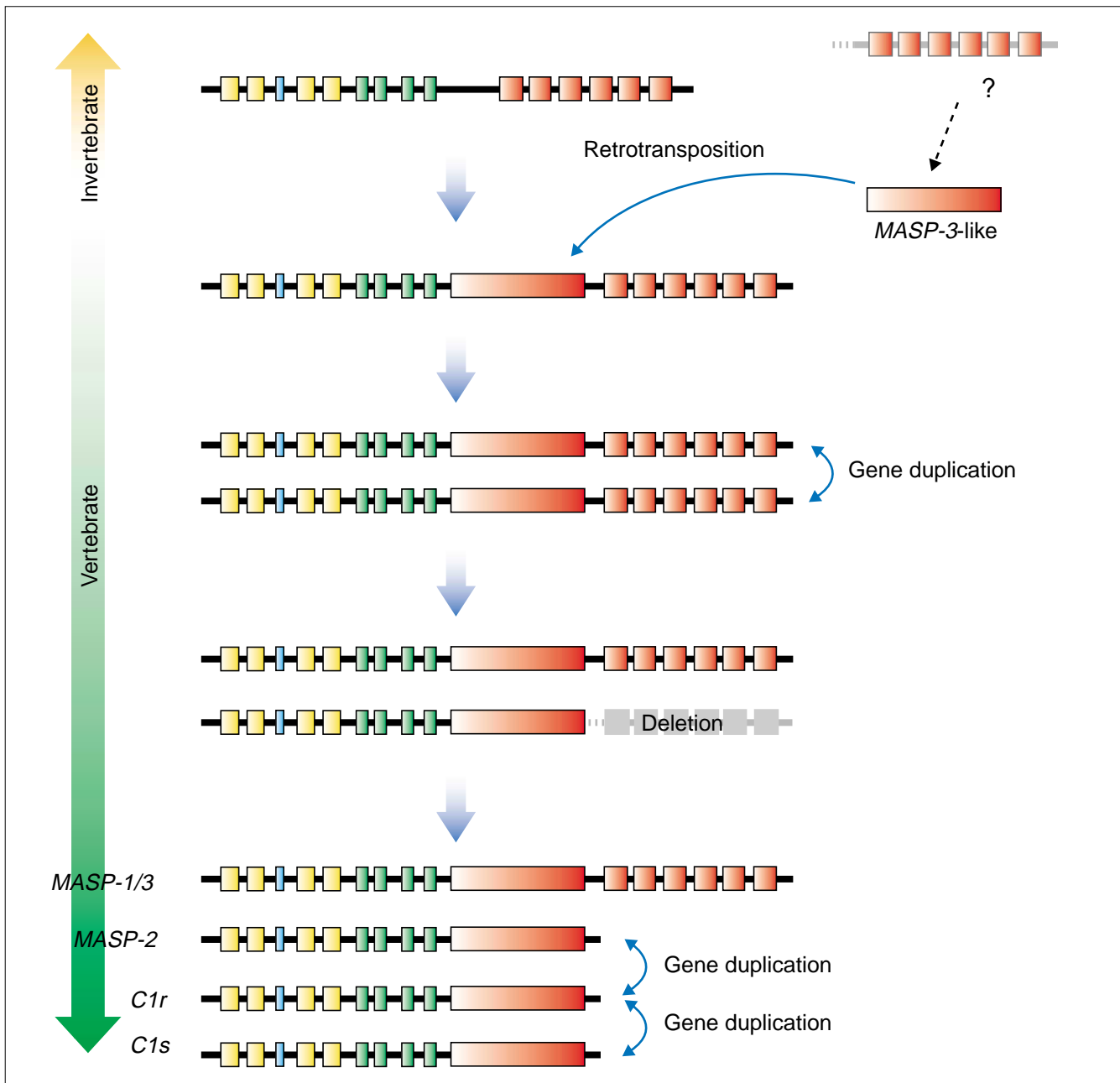
The domain structure and gene organization of the MASP/C1r/C1s serine protease family. The first line shows the domain structure of an archetypical member of the family; the central three lines are representations of the exon-intron organization of the genes encoding the family members. The *MASP-1* gene was recently found to contain another serine-protease-encoding region within it, encoding MASP-3, and was therefore renamed *MASP-1/3*. The bottom line shows the alternative splicing pattern of *MASP-1* and *MASP-3*. Corresponding domains and the exons encoding them are shown in the same color. Abbreviations: CUB, C1r/C1s, sea urchin Uegf, bone morphogenetic protein domain; EGF-like, similar to epidermal growth factor; SCR, short consensus repeat.

serine is encoded by the unusual AGY codon, instead of the more common TCN codon. These characteristics are all conserved in *MASP-2*, *C1r*, and *C1s* - we refer to this type of domain as 'MASP-3-like' - whereas *MASP-1*, which uses the downstream serine-protease-encoding region of the *MASP-1/3* gene, has a more 'typical' serine-protease domain. These data provide us with a hypothesis to explain the evolutionary history of this gene family, as depicted in Figure 2. The ancestral gene, viewed here as *MASP-1*-type gene, had a classical serine-protease-encoding region interrupted by several introns (Figure 2, top). Then, the second MASP-3-like serine-protease-encoding region was inserted by retrotransposition upstream of the original serine-protease-encoding region, generating a *MASP-1/3*-type gene. After gene duplication, one of the duplicates lost the downstream serine-protease-encoding region, resulting in a gene of the *MASP-2* type. Finally, further gene duplications generated the *C1r* and *C1s* genes (Figure 2, bottom), resulting in the four genes found in the human.

The order of appearance of the *MASP-2*, *C1r* and *C1s* genes during evolution is not obvious from their gene structure alone, but there are other indications as to which came first.

The phylogenetic relationships among these genes are such that *C1r* and *C1s* are closest to each other, and the genomic organization is such that only *C1r* and *C1s* are closely linked in a tail-to-tail orientation [6], supporting a scenario in which *C1r* and *C1s* were the last to be generated, with the final evolutionary step being an *in situ* duplication. This scenario gives further support to the generally accepted idea that the lectin pathway, in which the MASPs operate, predated the classical pathway, which relies on *C1r* and *C1s*.

So, our next question is when and how these steps actually occurred during the evolution of deuterostomes (the branch of animals that includes echinoderms, hemichordates and chordates). To date, we do not have enough data to answer this question definitively. Nevertheless, accumulating observations from our group and others about the complement system of lower vertebrates and invertebrates seem to provide some indications. The two *MASP* genes identified from an ascidian, *Halocynthia roretzi*, are *MASP-1*-like [7] and lack the intronless serine-protease-encoding region (Y. Endo, M.N. and T. Fujita, unpublished observations), while the lamprey *MASP* gene seems to be *MASP-1/3*-like [8]. Retrotransposition of the MASP-3-like serine-protease



**Figure 2**

A proposal for the evolution of the *MASP-1/3*, *MASP-2*, *C1r* and *C1s* genes. Colors are as in Figure 1. The top line represents the ascidian gene and the bottom lines the human genes. The intermediate lines illustrate hypothetical intermediate species. Although two *MASP* genes have been reported from ascidians, they are considered to be products of a recent gene duplication in the ascidian lineage, not corresponding to the gene duplications in the vertebrate lineage.

domain therefore most probably occurred in the main line of vertebrate evolution after the divergence of urochordates (ascidians), but before the divergence of cyclostomes (jawless vertebrates such as the lamprey and hagfish). Furthermore, the serine-protease domain of the *Halocynthia roretzi* complement factor B gene (encoding a protease of the alternative pathway; see Table 1) is MASP-3-like (X. Ji,

M. Sasaki and M.N., unpublished observations), except that it is interrupted by introns. All genes that have a MASP-3-like serine-protease domain may therefore prove to be on a single evolutionary lineage, as of all the known human chymotrypsin-family protease genes, only the *C1r*, *C1s*, *MASP-1/3* and *MASP-2* genes encode a MASP-3-like domain [9]. Thus, it is possible that ascidian factor B was the donor

gene, which was the source of the MASP-3-like domain acquired by *MASP-1/3* through retrotransposition. The presence of the classical complement pathway in bony and cartilaginous fish has been demonstrated at the functional level [10,11], and cDNA clones for *C1s/C1r* have been reported in bony fish [12]. Thus, all the steps illustrated for the vertebrate pathway in Figure 2 seem to have been completed before the emergence of jawed vertebrates.

### The primitive complement system

As discussed here, the complement system seems to have a more ancient origin in evolution than adaptive immunity; the latter seems to have been established only from jawed vertebrates onwards. The central component of the complement system, the C3 protein on which the three activation pathways discussed here converge, has been identified in jawless vertebrates, the lamprey [13] and hagfish [14], as well as in deuterostome invertebrates, amphioxus (a cephalochordate) [15], ascidian (urochordate) [16] and sea urchin (echinoderm) [17]. Although C3- or complement-like molecules of insects show functional similarity with deuterostome C3, there is no shared derived structural character between them, indicating that the deuterostome C3 and insect C3-like molecules derived independently from their common ancestor,  $\alpha$ 2-macroglobulin. Thus, the authentic complement system seems to have been established in the deuterostome lineage; the classical pathway of complement activation was then acquired in the jawed vertebrate lineage, at the time adaptive immunity arose.

In contrast to the classical pathway, which uses immunoglobulins - the central components of adaptive immunity - as recognition molecules, the primitive complement system that emerged before adaptive immunity must have depended on innate recognition molecules. In this context, it is noteworthy that the human lectin pathway uses lectins - mannan-binding protein (MBP) [18] and ficolin [19] - as the molecules that recognize pathogen and trigger the pathway. These molecules share a collagen-like amino-terminal region believed to be involved in MASP binding, although their carboxy-terminal recognition domains are different - the C-type lectin domain and the fibrinogen domain, respectively. For ascidians, ficolin [20], MBP-like lectin [21] and a novel lectin termed GBL [22] have been described as the possible recognition molecules of the primitive complement system. No functional information is available for the first two of these, however, and GBL lacks the collagen domain, the common structure shared by MBP and the classical pathway factor C1q of vertebrates that provides the binding site for MASP, C1r and C1s. We believe that further analysis of the complement system of ascidians, in which the complexity of linkage to adaptive immunity can be eliminated, will bring us important insights into the as-yet mysterious mechanisms of the lectin pathway in mammals.

In addition to C3, MASPs and lectins, factor B has also been identified from ascidian (X. Ji, M. Sasaki and M.N., unpublished observations) and also from sea urchin [23]. Factor B is a key component of the mammalian alternative pathway, where it acts as a catalytic subunit of the alternative pathway's C3 convertase enzyme. Thus, the lectin and alternative pathways, recognized as independent pathways in the mammalian complement system, might in fact represent two parts of a single ancient pathway. Perhaps the lectin pathway has a recognition function and the alternative pathway an amplification function in the primitive complement system. Further analyses of the activation mechanisms of the lectin pathway in both mammals and more primitive animals will reveal how the complement system arose and how it has been built upon during evolution, in the constant war between hosts and pathogens.

### References

1. Volanakis JE, Arlaud GJ: **Complement enzymes**. In *The Human Complement System in Health and Disease*. Edited by Volanakis JE, Frank MM. Marcel Dekker; 1998: 49-81.
2. Thiel S, Vorup-Jensen T, Stover CM, Schwaebler W, Laursen SB, Poulsen K, Willis AC, Eggleton P, Hansen S, Holmskov U, et al.: **A second serine protease associated with mannan-binding lectin that activates complement**. *Nature* 1997, **386**:506-510.
3. Dahl MR, Thiel S, Matsushita M, Fujita T, Willis AC, Christensen T, Vorup-Jensen T, Jensenius JC: **MASP-3 and its association with distinct complexes of the mannan-binding lectin complement activation pathway**. *Immunity* 2001, **15**:127-135.
4. Tosi M, Duponchel C, Meo T, Couture-Tosi E: **Complement genes C1r and C1s feature an intronless serine protease domain closely related to haptoglobin**. *J Mol Biol* 1989, **208**:709-714.
5. Endo Y, Sato T, Matsushita M, Fujita T: **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* 1996, **8**:1355-1358.
6. Tosi M, Duponchel C, Meo T, Julier C: **Complete cDNA sequence of human complement C1s and close physical linkage of the homologous genes C1s and C1r**. *Biochemistry* 1987, **26**:8516-8524.
7. Ji X, Azumi K, Sasaki M, Nonaka M: **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* 1997, **94**:6340-6345.
8. Endo Y, Takahashi M, Nakao M, Saiga H, Sekine H, Matsushita M, Nonaka M, Fujita T: **Two lineages of mannose-binding lectin-associated serine protease (MASP) in vertebrates**. *J Immunol* 1998, **161**:4924-4930.
9. **MEROPS database** [<http://merops.iapc.bbsrc.ac.uk/>]
10. Nonaka M, Yamaguchi N, Natsuume-Sakai S, Takahashi M: **The complement system of rainbow trout (*Salmo gairdneri*). I. Identification of the serum lytic system homologous to mammalian complement**. *J Immunol* 1981, **126**:1489-1494.
11. Jensen JA, Festa E, Smith DS, Cayer M: **The complement system of the nurse shark: hemolytic and comparative characteristics**. *Science* 1981, **214**:566-569.
12. Nakao M, Osaka K, Kato Y, Fujiki K, Yano T: **Molecular cloning of the complement (C1r/C1s/MASP2-like serine proteases from the common carp (*Cyprinus carpio*))**. *Immunogenetics* 2001, **52**:255-263.
13. Nonaka M, Fujii T, Kaidoh T, Natsuume-Sakai S, Yamaguchi N, Takahashi M: **Purification of a lamprey complement protein homologous to the third component of the mammalian complement system**. *J Immunol* 1984, **133**:3242-3249.
14. Ishiguro H, Kobayashi K, Suzuki M, Titani K, Tomonaga S, Kurosawa Y: **Isolation of a hagfish gene that encodes a complement component**. *EMBO J* 1992, **11**:829-837.

15. Suzuki MM, Satoh N, Nonaka M: **C6-like and C3-like molecules from the cephalochordate, amphioxus, suggest a cytolytic complement system in invertebrates.** *J Mol Evol*, in press.
16. Nonaka M, Azumi K, Ji X, Namikawa-Yamada C, Sasaki M, Saiga H, Dodds AW, Sekine H, Homma MK, Matsushita M, et al.: **Opsonic complement component C3 in the solitary ascidian, *Halocynthia roretzi*.** *J Immunol* 1999, **162**:387-391.
17. Al-Sharif WZ, Sunyer JO, Lambris JD, Smith LC: **Sea urchin coelomocytes specifically express a homologue of the complement component C3.** *J Immunol* 1998, **160**:2983-2997.
18. Matsushita M, Fujita T: **Activation of the classical complement pathway by mannose-binding protein in association with a novel C1s-like serine protease.** *J Exp Med* 1992, **176**:1497-1502.
19. Matsushita M, Endo Y, Fujita T: **Cutting edge: complement-activating complex of ficolin and mannose-binding lectin-associated serine protease.** *J Immunol* 2000, **164**:2281-2284.
20. Kenjo A, Takahashi M, Matsushita M, Endo Y, Nakata M, Mizuochi T, Fujita T: **Cloning and characterization of novel ficolins from the solitary ascidian, *Halocynthia roretzi*.** *J Biol Chem* 2001, **276**:19959-19965.
21. Nair SV, Pearce S, Green PL, Mahajan D, Newton RA, Raftos DA: **A collectin-like protein from tunicates.** *Comp Biochem Physiol B Biochem Mol Biol* 2000, **125**:279-289.
22. Sekine H, Kenjo A, Azumi K, Ohi G, Takahashi M, Kasukawa R, Ichikawa N, Nakata M, Mizuochi T, Matsushita M, et al.: **An ancient lectin-dependent complement system in an ascidian: novel lectin isolated from the plasma of the solitary ascidian, *Halocynthia roretzi*.** *J Immunol* 2001, **167**:4504-4510.
23. Smith LC, Shih CS, Dachenhausen SG: **Coelomocytes express SpBf, a homologue of factor B, the second component in the sea urchin complement system.** *J Immunol* 1998, **161**:6784-6793.