REVIEW

The complement C3 protein family in invertebrates

M Nonaka

Department of Biological Sciences, Graduate School of Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

Abstract

Complement C3 plays a pivotal role in the innate immune system of mammals as the central component of the complement system essential for its activation mechanism and effecter function. C3 has a unique intra-chain thioester bond that is shared by some complement and non-complement proteins forming a thioester protein (TEP) family. Phylogenetic analysis of TEP family genes of vertebrates and invertebrates revealed that the TEP family is divided into two subfamilies, the C3 subfamily and the alpha-2-macroglobulin (A2M) subfamily. The establishment of the TEP genes and differentiation of them into the C3 and A2M subfamilies occurred prior to the divergence of Cnidaria and Bilateria, in a common ancestor of Eumetazoa more than 600 MYA. Since then the A2M subfamily has been retained by all metazoan lineages analyzed thus far. In contrast, the C3 subfamily has been retained only by deuterostomes and some protostomes, and has been lost in multiple protostome lineages. Although the direct functional analysis of the most invertebrate TEPs is still to be performed, conservation of the basic domain structure and functionally important residues for each molecule suggests that the basic function is also conserved. Functional analyses performed on a few invertebrate C3 support this conclusion. The gene duplication events that generated C4 and C5 from C3 occurred in a common ancestor of jawed vertebrates, indicating that invertebrate and cyclostome C3s represent the pre-duplication state. In addition to C3, complement Bf and MASP involved in the activation of C3 are also identified in Cnidaria and some invertebrates, indicating that the complement system is one of the most ancient innate immune systems of Eumetazoa.

**Key Words**: thioester-containing protein (TEP); α-2 macroglobulin (A2M); evolution; Cnidaria

Introduction

The human complement system is composed of about 30 plasma and cell surface proteins and has three physiological functions, host defense against infection, interface between innate and adaptive immunity and disposal of immune complex and apoptotic cells (Volanakis, 1998; Walport, 2001). Since the latter two functions are intimately connected with the canonical adaptive immune system unique to the jawed vertebrates (Kasahara *et al.*, 1997), the original function of the complement system, found in both vertebrates and invertebrates, is believed to be the host defense against infection. In the human complement system, this function is attained by three major effector mechanisms, opsonization, induction of inflammation by chemotaxis and activation of leucocytes, and lysis of bacteria and cells.

There are three activation pathways for the mammalian complement system (Volanakis, 1998). The classical pathway, termed so because it was discovered first among three pathways, is initiated by binding of C1 to antigen-bound antibody, and results in the formation of the classical pathway C3 convertase composed of C4 and C2, which is responsible for the proteolytic activation of the central component C3. The next found alternative pathway consists of factor D (Df), factor B (Bf) and C3, and the alternative pathway C3 convertase is composed of C3 itself and Bf (Pangburn and Muller-Eberhard, 1984). The activation mechanism of the alternative pathway is still not totally clear, although regulatory factors such as factor H and properdin are reported to play a critical role in initiation of the pathway. The lastly found lectin pathway is initiated by recognition of PAMP (pathogen-associated molecular pattern) by the lectins, MBL (mannose-binding lectin) and ficolin, and the lectin-associated serine proteases, MASP (MBL-associated serine protease) activate C4 and C2, merging to the classical pathway (Matsushita and Fujita, 1992).

Corresponding author:
Masaru Nonaka
Department of Biological Sciences
Graduate School of Science, University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
E-mail: mnonaka@biol.s.u-tokyo.ac.jp
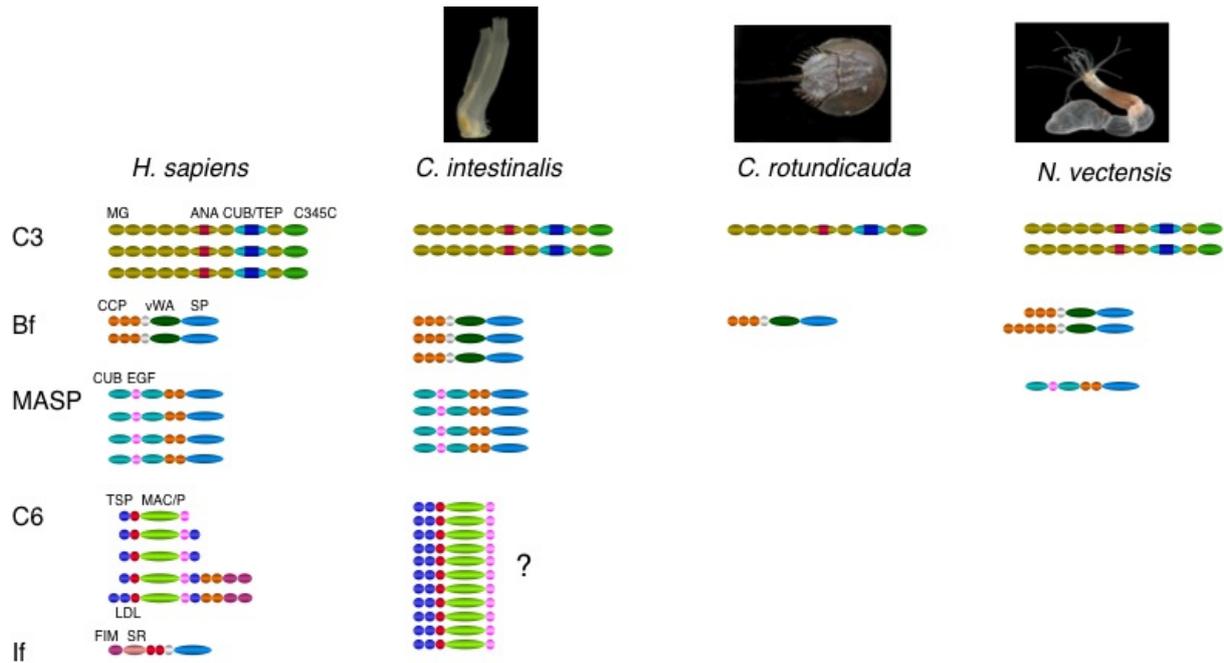
Fig. 1 Distribution of the complement genes with characteristic domain structure in invertebrate deuterostome, protostome and Cnidaria. Most key components of the complement system possess unique domain structure, and are classified into five mosaic protein families, C3, Bf, MASP, C6 and If families. The presence of these family genes is schematically shown for the representative species of Vertebrate (human, *H. sapiens*), invertebrate deuterostome (sea squirt, *C. intestinalis*), protostome (horseshoe crab, *C. rotundicauda*) and Cnidaria (sea anemone, *N. vectensis*). Since the genome sequence information is not yet available for *C. rotundicauda,* the absence of the MASP, C6 and If family member is still tentative. The question mark near the C6 family members of *C. intestinalis* indicates that these molecules are probably not involved in the sea squirt complement system in spite of close structural similarity to mammalian C6 family members (see text). Abbreviations of domain names are: MG, macroglobulin; ANA, anaphylatoxin; CUB/TEP, CUB domain inserted with thioester region; C345C, C-terminal of C3, C4 and C5; CCP, complement control protein; vWA, von Willebrand factor type A; SP, serine protease; CUB, C1r, C1s, uEGF, and bone morphogenetic protein; EGF, epidermal growth factor-like; TSP, thrombospondin type 1 repeats; MAC/P, membrane-attack complex/perforin; LDL, Low-density lipoprotein receptor domain class A; FIM, factor I/membrane attack complex; and SR, scavenger receptor Cys-rich.

Upon activation by C3 convertases, C3 is cleaved into the larger C3b and the smaller C3a fragments. C3b has ability to bind covalently to acceptor molecules on cell surfaces via ester or amide linkages (Law *et al.*, 1979), and cell-bound C3b is recognized by CR1 (complement receptor 1) on phagocytic cells, resulting in opsonic function (Ehlenberger and Nussenzweig, 1977). C3b can also react with C4b or C3b of the C3 convertases, leading to loss of C3 convertase activity and acquisition of C5 convertase activity (Takata *et al.*, 1987; Kinoshita *et al.*, 1988). Like C3 convertase, C5 convertase cleaves C5 into C5b and C5a. C5b initiates the assembly of the MAC (membrane attack complex) by sequential binding of C6, C7, C8 and C9. During this assembly process, hydrophobic domains of the participating proteins become exposed on the surface of the complex, and the complex becomes gradually inserted into the lipid bilayer and eventually forms a transmembrane channel (Podack *et al.*, 1981), leading to killing of the susceptible cells. C3a, C5a and the equivalent peptide derived from C4, C4a are termed as anaphylatoxin, since these peptides have prominent pro-inflammatory activity to induce chemotaxis and degranulation of leukocytes (Hugli, 1984).

C3, C4 and *C5* are structurally homologous genes (Wetsel et al., 1987), arose from a common ancestor by gene duplications occurred in the early stage of vertebrate evolution (Nonaka *et al.*, 1984). They are similar in size (~200kDa) and are composed of α and β subunits (C3 and C5) or α, β and γ subunits (C4). During complement activation, all three proteins are proteolytically cleaved at near the N-terminal end of α chain, liberating the C3a, C4a and C5a anaphylatoxins. C3 and C4 have a unique intrachain thioester bond formed between Cys and Gln in α chain, which is hidden inside of the molecule (Janatova and Tack, 1981; Levine and Dodds, 1990). Upon structural change induced by proteolytic activation of these molecules, the highly reactive thioester bond is exposed at the molecular surface, and can form an ester bond with a hydroxyl group or amide bond with an amino group on the target surface. This covalent tagging of the foreign particles seems to be the central function of the complement system, enabling the following elimination or killing of the pathogens.
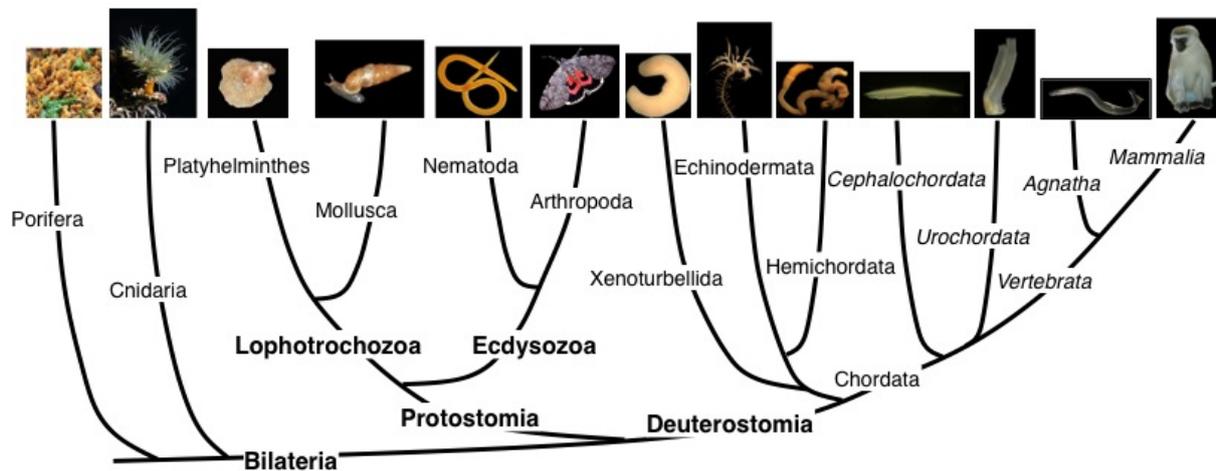
Fig. 2 Phylogenetic tree of animals. Out of more than 30 animal phyla, only those relevant to this review are shown with photographs of a belonging species. Taxonomic groups higher than phylum are shown in bold, whereas lower than phylum are shown in italic.

The 3-dimensional structure analysis of human C3 revealed the presence of the unpredicted macroglobulin (MG) domain, which repeats eight times and constitute the core of the TEP family proteins (Janssen *et al.*, 2005). Most functional sites of human C3 are present in the ANA, TED and C345C domains (see Fig.1 for the abbreviation and location of the domains) inserted within or between these eight MG domains. Further analysis of the 3-dimensional structure of C3b indicated that the dramatic conformational change actually occurs when C3 is proteolytically activated into C3b as predicted from biochemical analyses (Janssen *et al.*, 2006).

This review provides an overview of our current knowledge on the C3 genes or proteins of invertebrates, and will try to reconstitute the evolution of the complement system. Although there is no logical basis of an argument whether the common ancestral gene for *C3, C4* and *C5* should be called *C3, C4* or *C5*, here I refer it as *C3* since the role of C3 in the mammalian complement system is so pivotal that it is difficult to imagine a complement system without C3.

Animal phylogeny

Figure 2 shows the phylogenetic tree of the animal phyla relevant to this review based mainly on the recent multigene molecular phylogenetic analyses (Philippe *et al.*, 2005; Dunn *et al.*, 2008). Multi-cellular animals are divided into Porifera (sponges) without typical germ layers and Eumetazoa with typical germ layers. Eumetazoa is further divided into Cnidaria (sea anemone, hydra etc) possessing two germ layers and Bilateria possessing three germ layers as well as left-right symmetric body. Bilateria has two major groups, Protostomia and Deuterostomia. Protostomia is further divided into Ecdysozoa and Lophotrochozoa, each containing several phyla. Deuterostomia has four phyla, Xenoturbellida, Hemichordata,

Echinodermata and Chordata. Vertebrata is one of the three subphyla of Chordata. Invertebrate, all animals except for vertebrates, is thus not a proper taxonomic group. From the viewpoint of the complement evolution also, it is difficult to classify the invertebrate and vertebrate complement systems. Rather, the complement system of cyclostomes (Agnatha), the most basal extant vertebrates, shows more similarity to the invertebrate complement system than to the jawed vertebrate complement system. This is also true for the C3 family proteins as will be discussed in the following.

Origin of the TEP family genes

The thioester-containing protein (TEP) family members possess the unique intrachain thioester bond originally found in the human protease inhibitor, alpha-2-macroglobulin (A2M) and complement C3 (Dodds and Law, 1998). Although some members such as complement C5 (Wetsel *et al.*, 1987) and certain insect TEP (Lagueux *et al.*, 2000) secondary lost the thioester bond, they are identified as members of this family based on an overall sequence homology. Seven members of this family are encoded in the human genome: *C3*, *C4*, *C5*, *A2M* (Sottrup-Jensen *et al.*, 1985), pregnancy zone protein (*PZP*) (Sottrup-Jensen *et al.*, 1984), *CD109* (Lin et al. 2002) and the complement 3 and PZP-like A2M domain-containing 8 (*CPAMD8*) (Li *et al.*, 2004). C3, C4 and C5 are complement components derived from a C3-like common ancestor by gene duplications in the early stage of jawed vertebrate evolution (Nonaka and Takahashi, 1992; Terado *et al.*, 2003). Thus, whereas all C3, C4 and C5 are present in sharks (Terado *et al.*, 2003; Graham *et al.*, 2009) and higher vertebrates, only C3 has been identified from lamprey (Nonaka and Takahashi, 1992) and hagfish (Ishiguro *et al.*, 1992). Although C4 shows a close structural and functional similarity to C3, C5 lacks the thioester bond and plays a function that has diverged markedly from those of C3

23

and C4 in the human complement system (Lambris *et al.*, 1998). A2M is a serum protease inhibitor, inhibiting the diverse array of proteases by trapping proteases inside of the molecule rather than binding to the active site (Armstrong, 2010). The domain structure of A2M is essentially the same with that of C3 except that A2M has a bait domain instead of the ANA domain and lacks the C345C domain (Janssen *et al.*, 2005). PZP is a major pregnancy-associated plasma protein, with similar structure and function to A2M (Sottrup-Jensen *et al.*, 1984). CD109, unlike other members of the TEP family, is a GPI-linked glycoprotein originally found on endothelial cells, platelets and activated T-cells (Lin *et al.*, 2002). CD109 suppresses transforming growth factor (TGF)-β signaling in human keratinocytes by binding to TGF-β receptor I (Finnson *et al.*, 2006), and a high level of CD109 expression is detected in squamous cell carcinomas of the esophagus, lung, uterus and oral cavity (Hagiwara *et al.*, 2008). However, biochemical details of the CD109 function are still unknown except that the involvement of furin in processing is reported recently (Hagiwara *et al.*, 2010). CPAMD8, also termed KIAA1283, has a Kazal-type serine proteinase inhibitor-like domain at the C-terminus and is expressed mainly in the kidney, brain and testis, although its function is poorly characterized (Li *et al.*, 2004).

No TEP gene is present in the published genome information of a sponge, *Amphimedon queenslandica* and a choanoflagellate, *Monosiga brevicollis* (King *et al.*, 2008; Kimura *et al.*, 2009; Srivastava *et al.*, 2010). Thus, it is suggested that the TEP gene arose in the Eumetazoa lineage. Comprehensive cloning of TEP genes of a Cnidarian, a sea anemone, *Haliplanella lineate* resulted in identification of four TEP genes (Fujito *et al.*, 2010). The genome analysis performed in another sea anemone species, *Nematostella vectensis*, also identified the same set of TEP genes (Kimura *et al.*, 2009), indicating that Anthozoan Cnidaria has these four TEP genes. Phylogenetic analysis of the four identified Cnidarian TEP genes and various TEP genes of many Eumetazoa resulted in a NJ tree shown in Fig. 3. Although this is an unrooted tree, the root most probably resides at the branch marked by the black triangle, since genes of various animals from Cnidaria to Vertebrata are present on both sides of this branch indicating that this branch represents an ancient diversifying event, and the ANA and C345C domains are present in all members on one side of this branch but none of the members on the other side. Thus, the TEP gene family is divided by this branch into two subfamilies: the A2M subfamily including A2M, CD109, CPAMD8 and insect TEPs, and the C3 subfamily including complement C3, C4 and C5 (Fig. 3).

Insect TEPs were originally found in *Drosophila melanogaster*, and six members, TEP1-TEP6, have been identified in this species (Lagueux *et al.*, 2000). Drosophila TEPs contain a hypervariable region at the position corresponding to the bait region of A2M and the ANA domain of C3. TEP2, TEP3 and TEP6 bind to *Escherichia coli, Staphylococcus aureus* and *Candida albicans*, respectively, and promote their phagocytosis by cultured S2 cells (Stroschein-Stevenson *et al.*, 2006). In a mosquito,

Anopheles gambiaeare, the TEP1 gene was shown to promote phagocytosis of some Gram-negative bacteria (Levashina *et al.*, 2001), and to bind to the surface of *Plasmodium* ookinetes and promote their lysis and melanization (Blandin *et al.*, 2004). Therefore, the insect TEP genes were referred to as complement-like genes. However, the phylogenetic analysis clearly indicates that they belong to the A2M subfamily (Fig. 3), suggesting that the reported functional similarity of these insect TEPs and mammalian C3 was caused by convergent molecular evolution.

Two of the four Cnidarian TEP genes belonged to the A2M subfamily, showing a close similarity to human *A2M* and *CD109*, respectively, and thus were termed *HaliA2M* and *HaliCD109* (Fujito *et al.*, 2010). The other two genes belonged to the C3 subfamily, and were termed *HaliC3-1* and *HaliC3-2* (Fujito *et al.*, 2010). Cnidarian TEPs retained the basic domain structure and functionally important residues for each molecule, and their mRNA were detected at different parts of the sea anemone body. Thus a strong signal for *HaliC3-1* was detected in the endoderm of tentacles, *HaliA2M* was detected in endoderm of the mesentery as strong granular signals, and *HaliCD109* showed strong granular expressions in ectoderm of tentacles and in endoderm of the mesentery (Fujito *et al.*, 2010). Different expression pattern implies functional differentiation among these three genes. Therefore it is suggested that gene duplication and subsequent functional differentiation among *C3*, *A2M* and *CD109* were very ancient events predating the divergence of the Cnidaria and Bilateria more than 600 MYA. In contrast, the genome of *Hydra magnipapillata,* belonging to Hydrozoa, contained only A2M subfamily member (Miller *et al.*, 2007). These results indicate that the creation of the TEP genes and subsequent gene duplication and functional differentiation into *C3*, *A2M* and *CD109* have occurred in a relatively short evolutionary period after the divergence of sponges and before the divergence of Cnidaria from the Bilateria lineage, and that the C3 subfamily was lost by some Cnidaria lineages.

Since both C3 and A2M subfamily members of the TEP family are present in Cnidaria and Bilateria, it is not clear which subfamily arose first. The recent genome analysis in Placozoa (Srivastava *et al.*, 2008) identified at least two TEP genes, and both of them belong to the A2M subfamily, one at the basal position of A2M cluster and the other in the CD109 cluster (Fig. 3). Although the phylogenetic relationship among Porifera, Placozoa, Cnidaria and Bilateria is not conclusively resolved (Philippe *et al.*, 2009; Schierwater *et al.*, 2009), if Placozoa diverged prior to the divergence between Cnidaria and Bilateria, it is suggested that the A2M subfamily is more ancient than the C3 subfamily. Structural features of the common ancestor of the C3 and A2M subfamilies can be deduced from the comparison of the human and sea anemone TEP structures (Fig. 4). Comparison of four human TEPs and three sea anemone TEPs showed that the signal peptide and the thioester domain are present in all members, suggesting that these features were present in the common ancestor of the C3 and A2M subfamilies. In addition, the α–β processing site and the His residue
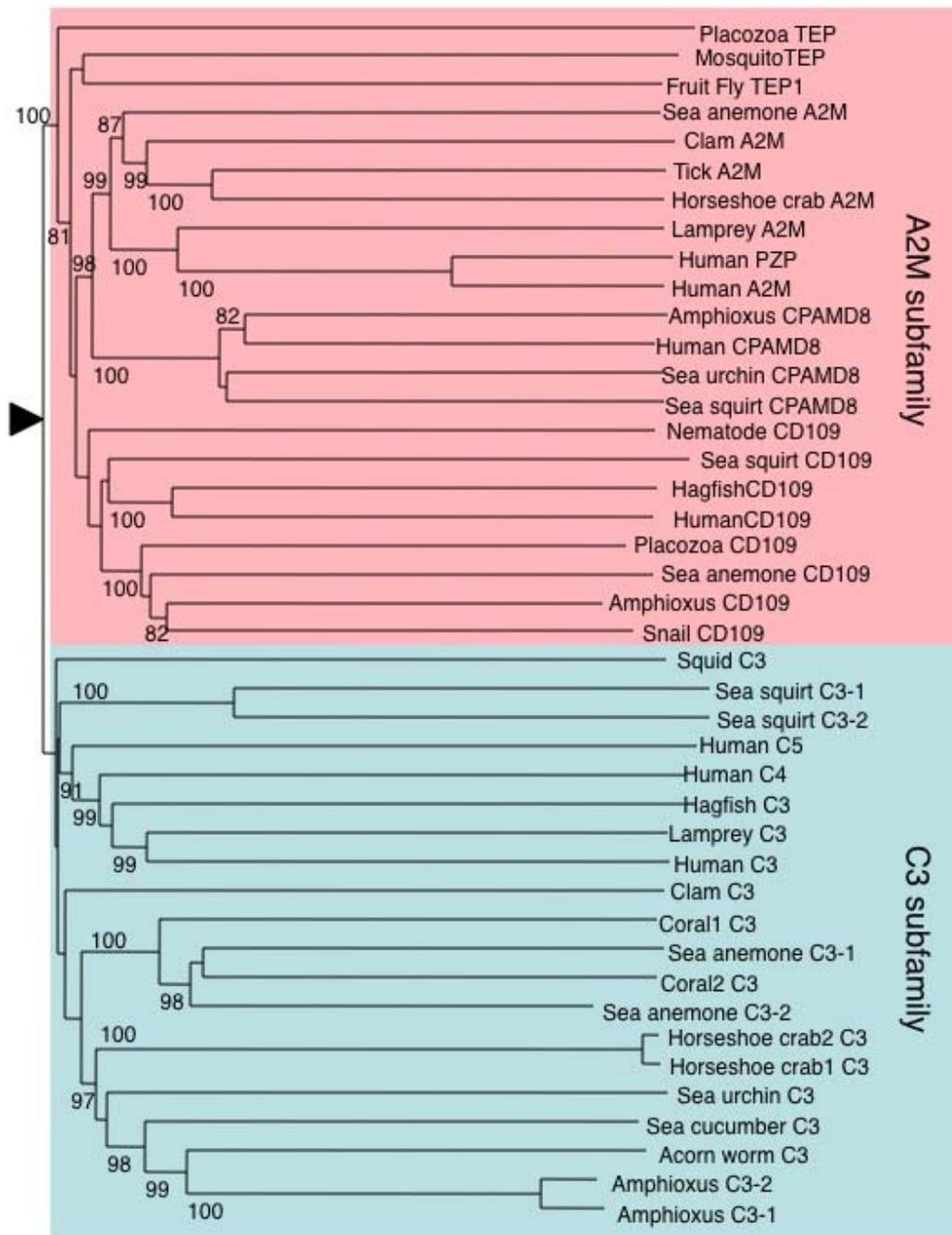
Fig. 3 Phylogenetic tree of TEP genes. The tree was constructed based on the alignment of the full length amino acid sequences of the TEP family genes, using the Neighbor-joining method. Gaps were not excluded. Bootstrap percentages of more than 50 with 1000 replicates are given. Accession numbers of the used sequences and scientific names of animals are; human (*Homo sapiens*) C3, C4, C5, A2M, CD109, CPAMD8, and PZP (NP_000055, P0C0L4, AAA51925, P01023, NP_598000, NP_056507 and CAA38255), squid (*Euprymna scolope*) C3 (ACF04700), sea urchin (*Strongylocentrotus purpuratus*) C3 and CPAMD8 (NP_999686 and XP_785018), sea squirt (*Ciona intestinalis*) C3-1, C3-2, CD109, and CPAMD8 (NP_001027684, CAC85958, NP_001027688 and XP_002124325), lamprey (*Lethenteron japonicum*) C3 and A2M (Q00685 and BAA02762), hagfish (*Eptatretus burgeri*) C3 and CD109 (P98094 and BAD12264), horseshoe crab 1 (*Cainoscorpius rotundicauda*) C3 (AAQ08323), horseshoe crab 2 (*Tachypleus tridentatus*) C3 and A2M (BAH02276 and BAA19844), amphioxus (*Branchiostoma floridae*) C3-1, C3-2, CD109 and CPAMD8 (AAM18874, XP_002612866, XP_002586872 and XP_002612485), coral 1 (*Swiftia exserta*) C3 (AAN86548), coral 2 (*Acropora millepora*) C3 (ABK78771), sea anemone (*Haliplanella lineate)* C3-1, C3-2, A2M, and CD109 (AB481383, AB481384, AB481385 and AB481386), fruit fly (*Drosophila melanogaster*) TEP1 (NP_523578), mosquito (*Anopheles gambiae*) TEP (AAG00600), nematode (*Caenorhabditis elegans)* CD109 (NP_493614), clam (*Hyriopsis cumingii*) A2M (ABJ89824), clam (*Venerupis decussatus*) C3 (FJ392025), ticks (*Ornithodoros moubata*) A2M (AAN10129), snail (*Euphaedusa tau*) CD109 (BAE44110), sea cucumber (*Apostichopus japonicus*) C3 (ADN97000), acorn worm (*Saccoglossus kowalevskii*) C3 (XP_002732077), placozoa (*Trichoplax adhaerens*) CD109 and TEP (XP_002111588 and XP_002111589).

25

which catalyze the cleavage of the thioester bond are present in most members, although they are missing from human A2M. Thus, the common ancestor molecule likely had these characters, showing a closer similarity to human C3 than to human A2M. For the position for the ANA or bait domains and the C-terminus where C345C, Kazal and GPI attachment domains appear, it is difficult to deduce the ancestral state, since none of them is in majority. Therefore, the common ancestor of the C3 and A2M subfamilies most probably was, secreted protein synthesized with the signal peptide, two subunit (α and β) chain protein, and endowed with the thioester bond with the catalytic histidine.

Cnidarian C3 and complement system

In addition to C3-1 and C3-2 genes identified from sea anemone species, *Haliplanella lineate* (Fujito *et al.*, 2010) and *Nematostella vectensis* (Kimura *et al.*, 2009)*,* the C3 genes have been reported from two coral species, *Swifta exserta* (Dishaw *et al.*, 2005) and *Acropora millepora* (Miller *et al.*, 2007)*.* The basic domain structure of the C3 family and the signal peptide for secretion were completely conserved in these Cnidarian C3 proteins (Table 1). The conservation of the C3a anaphylatoxin region, the thioester site (GCGEQ), and the catalytic His residue for cleavage of thioester suggest that the Cnidaria C3 proteins retain the inflammatory and opsonic functions. The C345c domain was also identified and five Cys residues characteristic to this domain are perfectly conserved. Since the mammalian C345C domain of C3 is involved in the interaction with the vWA domain of Bf, in the alternative pathway C3 convertase, C3bBb (Rooijakkers *et al.*, 2009; Torreira et al. 2009), presence of this domain in Cnidarian C3 proteins imply that the Cnidarian C3 proteins are capable to form C3 convertase with Bf. The conservation of the α/β chains processing site 'RXXR' suggested that Cnidarian C3 proteins are processed into the two-subunit chain structure. However, both of Cys residues involved in the disulfide linkage between the α and β chains of mammalian C3 (Dolmer and Sottrup-Jensen, 1993) were substituted into another residue, and there is no other pair of Cys residues, which is possibly involved in the inter-chain linkage. The most unique feature of Cnidarian C3s is the presence of about 50 residue-long, highly Lys/Arg-rich insertion in the MG8 domain. Although insertion into the MG8 domain was also found in the horshoe crab C3 (Zhu *et al.*, 2005), ascidian C3 (Nonaka *et al.*, 1999; Marino *et al.*, 2002), lamprey C3 (Nonaka and Takahashi, 1992) and vertebrate C4, the Lys/Arg content of Cnidarian C3 is extremely high (~60 %), likely providing a unique, extremely positive charge to this region of Cnidarian C3. The α/γ chain-processing motif 'RXXR' has been reported in the jawed vertebrate C4, lamprey C3, and horseshoe crab C3 at this region. It is possible, therefore, that the Lys/Arg-rich insertion was present in the common ancestor of C3 subfamily proteins, and the α/γ chain-processing motif is its remnant.

In addition to two *C3*, two *Bf* and one *MASP* genes were identified in the draft genome sequence of *Nematostella vectensis* (Kimura *et al.*, 2009). In contrast, no *C6* and *factor I* (*If*) family genes were identified. The deduced primary structures of the cnidarian Bf and MASP shared the unique domain structures and most functionally critical amino acid residues with their mammalian counterparts, suggesting the conservation of basic biochemical functions throughout the metazoan evolution. *In situ* hybridization analysis indicated that all five Cnidarian complement genes are co-expressed at the tentacles, the pharynx and the mesentery in an endoderm-specific manner (Kimura *et al.*, 2009). These results indicated that the multi-component complement system composed of at least C3, Bf and MASP was established in a common ancestor of Cnidaria and Bilateria more than 600 MYA to protect coelenterons, the primitive gut cavity with putative circulatory functions.

Protostome C3 and complement system

The presence of C3 in Cnidaria indicated that the C3 gene has been established prior to the divergence of Cnidaria and Bilateria. However, the firstly-elucidated protostome genomes of *Drosophila melanogaster* (Adams *et al.*, 2000) and *Caenorhabditis elegans* (Consortium 1998) did not contain the C3 gene. Moreover, our attempt to RT-PCR amplify TEP cDNAs using universal primers compatible with both the C3 and A2M subfamily members in four Lophotrochozoa species, *Planocera multitentaculata* (Platyhelminthes), *Siphonosoma cumanense* (Sipuncula), *Hesione reticulata* (Annelida) and *Euphaedusa tau* (Mollusca), resulted in isolation of only A2M family cDNA, suggesting the absence of C3 in these species (Kim, Fujito and Nonaka, unpublished data). On the other hand, the C3 subfamily members have been reported from Ecdysozoan horseshoe crab, *Carcinoscorpius rotundicauda* and *Tachypleus tridentatus* (Zhu *et al.*, 2005; Ariki *et al.*, 2008) and Lophotrochozoan squid, *Euprymna scolope* (Castillo *et al.*, 2009) and clam, *Hyriopsis cumingii* (Prado-Alvarez *et al.*, 2009). These results indicate that whereas A2M has been retained by all protostomes, C3 has been lost secondarily multiple times during the protostome evolution. The deduced primary structure of these protostome C3 proteins did not show any large insertion or deletion compared to vertebrate C3, and retained all primary structure motifs reported to have functional importance, such as the thioester, anaphylatoxin, and $\alpha-\beta$ processing regions (Table 1). For squid and clam, the composition of the complement system is still unknown except that the Bf-like gene was identified from clam (Prado-Alvarez *et al.*, 2009). However, the active center Ser is substituted into Ile in this Bf-like molecule, making it unlikely that this molecule functions as the catalytic subunit of a C3 convertase. In contrast, horseshoe crab C3 is the best-analyzed C3 of invertebrates at the biochemical level. Molecular cloning of the C3 has been reported from two distantly-related horseshoe crab species, *Carcinoscorpius rotundicauda* (Cr) (Zhu *et al.*, 2005) and *Tachypleus tridentatus* (Tt) (Ariki *et al.*, 2008). Interestingly, the primary structures of CrC3 and TtC3 show quite high degree of identity (Fig. 3), in spite of a remote phylogenetic relationship of these two species. CrC3

**Table 1** Structural features of invertebrate C3s

	$\alpha-\beta$	ANA	Thioester	Catalytic His	$\alpha-\gamma$
Haliplanella lineate 1	RKKR	CC,C,C,CC[#1]	GCGEQ	H	KR-rich
Haliplanella lineate 2	?	?,CC	GCGEQ	H	KR-rich
Swifta exserta	RKRR	CC,C,C,CC	GCGEQ	H	KR-rich
Acropora millepora	RKKR	CC,C,C,CC	GCGEQ	H	KR-rich
Euprymna scolopes	RYKR	CC,C,CC	GCGEQ	H	
Venerupis decussatus	RKRR	C,CC,C,C,C	GCGEQ	H	RKKR
Carcinoscorpius rotundicauda	RKKR	CC,C,C,CC	GCGEQ	H	LR-rich
Tachypleus tridentatus	RKKR	CC,C,C,CC	GCGEQ	H	LR-rich
Apostichopus japonicus	RRRR	C,C,CC	GCGEQ	H	DR-rich
Strongylocentrotus purpuratus	RRKR	C,C	GCGEQ	H	
Saccoglossus kowalevskii	[#2]	CC	GCGEQ	H	DR-rich
Branchiostoma belcheri 1	[#2]	C,CC	GCGEQ	H	EDSR-rich
Branchiostoma belcheri 2	[#2]	C,CC	GCGEQ	H	EDSR-rich
Ciona intestinalis 1	RKKR	C,C	GCGEQ	H	
Ciona intestinalis 2	RNKR	C,C	GCGEQ	H	
Eptatretus burgeri	RRKR	CC,C,C,CC	GCGEQ	H	
Lethenteron japonicum	RKPR	CC,C,C,CC	GCGEQ	H	RRRR
Homo sapiens C3	RRRR	CC,C,C,CC	GCGEQ	H	
Homo sapiens C4	RKKR	CC,C,C,CC	GCGEQ	H	RRRR
Homo sapiens C5	RPRR	CC,C,C,C,CC	GSAEA	P	

[#1]Distrition of the Cys residues in the ANA domain is shown. "," represents multiple amino acid residues.
[#2]Blank indicates the absence of characteristic residues. However, the absence of the $\alpha-\beta$ processing signal and some Cys residues in the ANA domain of these three C3s could be due to inaccurate gene prediction.

showed binding activity to *Staphylococcus aureus* and other bacteria and this binding activity was inhibited by hydroxylamine, suggesting the involvement of the thioester bond of CrC3 in binding. In addition, divalent cation-dependent induction of trypsin-like proteolytic activity in horseshoe crab plasma was observed, although it is still to be demonstrated directly if this activity is due to CrBf, and if CrBf directly activates CrC3 (Zhu *et al.*, 2005). For TtC3, factor C originally identified as an LPS-sensitive initiator of hemolymph coagulation stored within hemocytes was identified as an activating enzyme (Ariki *et al.*, 2008). Thus, upon invasion of Gram-negative bacteria, an LPS-responsive factor C plays the central role in the initiation of the horseshoe crab complement activation. However, factor C seems not to be involved in TtC3 binding to *S. aureus*, suggesting the presence of other C3 activating enzymes in horseshoe crab, possibly the C3 convertase comprising Bf. In support of this view, direct binding of factor C and Bf to PRRs was reported in *C. rotundicauda* (Le Saux *et al.*, 2008). The apparent functional substitution by factor C for MASP suggest a significant deviation of the horseshoe crab complement system from the other complement systems. In this context, the reported primary structure of horseshoe crab Bf shows curious characteristics (Zhu *et al.*, 2005). The serine protease domains of mammalian Bf and C2 are known to be unique in the following points (Volanakis and Arlaud, 1998). Firstly, although they cleave peptide bonds following a positively-charged amino acid residue, they lack Asp189 (chymotrypsinogen numbering) which in trypsin positions the positively charged P1 residue. Instead, Asp226 is responsible for substrate specificity of human Bf (Jing *et al.*, 2000). Secondary, they lack the highly conserved and functionally important N-terminal sequences of serine proteases. Thus, following activation their serine
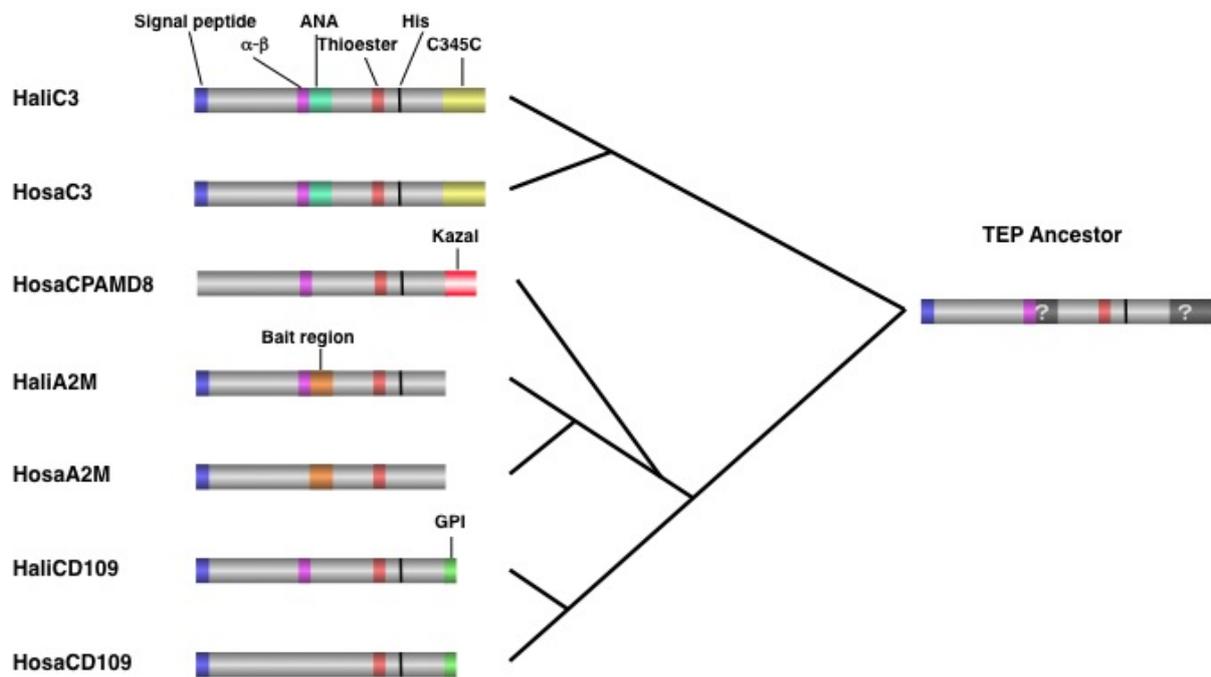
Fig. 4 Structural features of human and sea anemone TEP molecules. Structural features of human (*H. sapiens*) and sea anemone (*H. lineate*) TEP molecules are compared, and the structure of the common ancestor of the TEP molecules is deduced based on maximum parsimony principle. Question marks in the ancestor molecule indicate that there are multiple candidates with equal probability.

protease domains remain attached to a vWA domain. Comparison of amino acid sequences of the serine protease domain of Bfs from various animals shown in Figure 5 suggests that both structural specializations occurred in a common ancestor of jawed vertebrates (Terado *et al.*, 2001). Thus, both structural specializations seem to have occurred simultaneously leading to a drastic change in structure and activation mechanism of Bf. The horseshoe crab Bf is exceptional in that it retains the original specificity-determining Asp189 without retaining the conserved N-terminal sequence of the serine protease domain. Thus the activation mechanism of the horseshoe crab Bf, and whole complement system also, could be a highly deviated one, although the details are still to be clarified.

Deuterostome C3 and complement system

C3 sequence of invertebrate deuterostome has been reported from sea squirt (Urochordata) (Nonaka *et al.*, 1999; Raftos *et al.*, 2002; Azumi *et al.*, 2003; Pinto *et al.*, 2003), amphioxus (Cephalochordata) (Suzuki *et al.*, 2002; Holland and Gibson-Brown, 2003), acorn worm (Hemichordata) (XP_002732077), sea urchin (Al-Sharif *et al.*, 1998) and sea cucumber (Echinodermata) (ADN97000) thus far. Since gene duplications among *C3, C4* and *C5*, and subsequent functional differentiation among them likely occurred at the early stage of jawed vertebrate evolution (Nonaka and Takahashi, 1992; Nonaka and Kimura, 2006), these C3 genes

together with cyclostome C3 genes are considered to represent the ancestral state before the gene duplications. All these invertebrate deuterostome C3 proteins retain the thioester site, the His residue which catalyzes the cleavage of the thioester bond, and the α–β processing site. In contrast, the Cys residues in the ANA domain and the Leu-X-Arg sequence at the cleavage site of the C3 convertase are poorly conserved by these C3 proteins (Table 1).

In addition to C3, Bf gene has been identified from sea urchin (Smith *et al.*, 1998), and several complement genes have been identified from sea squirt, mainly from two species, *Halocynthia roretzi* and *Ciona intestinalis*. Those genes are; Bf (Azumi *et al.*, 2003; Yoshizaki *et al.*, 2005), MASPs (Ji *et al.*, 1997; Azumi *et al.*, 2003), mannan-binding lectin (MBL) (Bonura *et al.*, 2009), ficolin (Kenjo *et al.*, 2001), complement receptor 3 (CR3) alpha (Miyazawa *et al.*, 2001) and beta (Miyazawa *et al.*, 2001) and glucose binding lectin (GBL) lacking the collagen domain as a possible functional substitute for MBL (Sekine *et al.*, 2001). Thus, the sea squirt complement system is the best-analyzed invertebrate complement system from the viewpoint of the component genes. For the functional aspect, opsonic function of the sea squirt complement system has been demonstrated in *H. roretzi,* where C3, ficolin and GBL proteins isolated from the body fluid work together as opsonin. Moreover, the C3a fragment was shown to have a chemotactic activity (Pinto *et al.*, 2003; Raftos *et al.* 2003), indicating that the role of the complement system in inflammation is also conserved between mammals and urochordates.

```
              656                                    #                                746
Human Trypsin  CLDAPV--LSQAKCEASYPGKITSNMFCVGFLE--GGKDSCGGDSGGPVVCNG---------QLQ-GVVSWGDG--CA----QKNKPGVYTK-------VYNYVK
Human Bf       CERDAQYAP-GYDKVKDISEVVTPRFLCTGGVSPYADPNTCRGDSGGPLIVHKRS------RFIQVGVISWGVVDVCK---NQKRQKQVPA---HARDFHINLF
Xenopus Bf     CLEAAKKAP-ELKNVTNIEDAVTDQFLCTGGIVPVADPPVCKGDSGGPLLIQVKR------RYVQVGIISWGTVDHCD----KGTRIKQTQK---NARDFYQDIF
Oryzias Bf     CIKKAL--KAKGITTTDDPKVPVTDNFLCTGGD---RDHIACTGDSGGAVFKNYES------RTIQIALVSWGTQEICTGGG---MRETTPES---RDFHINLF
Shark Bf       CEANALQAP-EYKNMSHVHLVVTDRFLCTGGHEPVVEKVSCKGDSGGALFIQKKR------RYIQVGVLSWGVINSCDN-----VALKEKYA-----RDFHINLF
Human C2       CAEVVSQEKTMFPNLTDVREVVTDQFLCSGTQE---DESPCKGESGGAVFLERRF------RFFQVGLVSWGLYNPCLGSADKNSRKRAPRSKVPPPRDFHINLF
Lamprey Bf     CLKDIM--ARFNVTEEKAEKHITENMLCAWNAT----ADTCRGDSGGPLVLQKNR------RWIQVGIVAGGVAQHCG---------KNIKS-----SFYTNVA
Sea Squirt Bf  CQDAIK----YINEKQRIKYSYNTTLFCCLDPH-DNGVDTCQGDSGGPVVREVLNQTSGVSCWVQIGLVSFGWG--CG----------ALRDD------VHAYVP
Sea Urchin Bf  CSIAMS---------AHGIAVDTTTELCAGIER----KDSCQGDSGGPLVVQRNN------KYRQIGIVSYGIG--CG---------VTYG------VYTRVP
Horseshoe crab Bf CVQSLE---------NTKDPMTDFMICAGDGR--GVADTCQGDSGGPLAQSLLDE--SGMNYWIQVGIISWGRG-CKN---------RGQYG--FYTHVAKLRP
Sea anemone Bf CQVNHT-----------NAITKRMRCAGDVM--GERDACKGDSGSPIVVKRTDG-----SWSAVGLSSWGEG--CAQ---------KGKFG------VYADLL


              478             495
Human Bf       CGMV--------------WEHRKGTDYHKQPW
Xenopus Bf     CGLS---------KYHSVELDPKKKATVMFPW
Oryzias Bf     CGLH---------RDYELTADKDGKRRRYPW
Shark Bf       CGFA--------------NKSLIESDQSSYPW
Human C2       CGVG--------------NMSANASDQERTPW
Lamprey Bf     CGTS-----GTFRIPRARIAGGDPTKIELWPW
Sea Squirt Bf  CGQAGEVGLK--HMALKRVVGGEESVENAWPW
Sea Urchin Bf  CGES------KHPSATSRIVGGSESHSGDWPW
Horseshoe crab Bf CGYG------MDEVVEEVNKPGGTRAEKPWPW
Sea anemone Bf CGVAGDTQLRDSSDKRFRIVGGREAKAGAWPW
```

Fig. 5 Evolution of Bf primary structure. The primary structures of Bf and C2 of various animals were compared at two regions of the serine protease domain. The upper panel: the region around the active center Ser corresponding to 656th-746th residues of human Bf. The active center Ser is marked by #. The Asp residue at the S1 pocket essential for the trypsin-like specificity is shown in red. The lower panel: the N-terminal region of the serine protease domain orresponding to 478 th - 495 th residues of human Bf. The possible proteolytic activation site is shown in red.

In contrast, the third activity of the mammalian complement system, cytolytic activity, has not been recognized in the urochordate complement system.

Evolution of the complement system

Figure 1 summarize the distribution of the complement genes with the characteristic domain structure in the representative species of Cnidaria (*N. vectensis*), Protostomia (*C. rotundicauda*), Invertebrate Deuterostomia (*C. intestinalis*) and Vertebrata (*Homo sapiens*). C3 arose in a common ancestor of Cnidaria and Bilateria more than 600 MYA from an ancestral TEP gene with unknown function. Bf and MASP seem to be generated at the same time, establishing a primitive complement system consisting of C3, Bf and MASP. Conservation of the C3 C345C domain and the Bf vWA domain suggests that they are able to assemble to form the alternative pathway C3 convertase, C3bBb. MASP is likely an initiating enzyme to induce C3bBb formation, and covalent binding of C3b to the surface of foreign molecules most probably enhanced phagocytosis. In addition to this opsonic function, the conservation of the ANA domain between Cnidaria and Vertebrata C3 suggests that the primitive complement system also had another function to induce inflammation.

By unknown reason, the complement system has been lost multiple times in the protostome lineages. Even in some protostome species which retains C3, MASP has not been identified, and Bfs has highly deviated structure, either proteolitically inactive or lacking the conserved N-terminal sequence of the serine protease domain. Thus, the activation mechanism of protostome complement system should be very different from that of Cnidaria or Vertebrata. Interestingly, some insect species lacking the C3 gene have TEP molecules functioning as opsonin like vertebrate C3. However, molecular phylogenetic analyses indicate that these insect TEPs belong to the A2M subfamily, suggesting that the similar functions were attained by convergent evolution.

In contrast, the complement system of invertebrate deuterostome represented by *C. intestinalis* shows much closer similarity to the mammalian complement system. However it lacks If, the main component of the regulatory mechanism. In addition, although there are several C6-like genes with the membrane attack complex/perforin (MACP) domain in the *C. intestinalis* genome, all of them lack the C-terminal short consensus repeat (SCR) and factor I/membrane attack complex (FIM) domains reported to be essential for interaction with other complement components. Thus, it is unlikely that these C6-like molecules are integrated in the urochordate complement system. These results indicate that the lytic activity and regulatory mechanism of mammalian complement system arose in the vertebrate lineage, probably after the divergence of cyclostomes since lamprey also lacks the lytic activity. In addition to the regulatory and lytic pathways, the jawed vertebrates seem to have gained the classical pathway by the C3/C4/C5 and Bf/C2 gene duplications. The modern complement system equipped with all these pathways was most probably established in a common ancestor of the jawed vertebrates.

References

Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, *et al.* The genome sequence of *Drosophila melanogaster*. Science 287: 2185-2195, 2000.

Al-Sharif WZ, Sunyer JO, Lambris JD, Smith LC. Sea urchin coelomocytes specifically express a homologue of the complement component C3. J. Immunol. 160: 2983-2997, 1998.

Ariki S, Takahara S, Shibata T, Fukuoka T, Ozaki A,

Endo Y, *et al.*, Factor C acts as a lipopolysaccharide-responsive C3 convertase in horseshoe crab complement activation. J. Immunol. 181: 7994-8001, 2008.

Armstrong PB. Role of α2-macroglobulin in the immune responses of invertebrates. Inv. Surv. J. 7: 165-180, 2010.

Azumi K, De Santis R, De Tomaso A, Rigoutsos I, Yoshizaki F, Pinto MR, *et al.* Genomic analysis of immunity in a Urochordate and the emergence of the vertebrate immune system: "waiting for Godot". Immunogenetics 55: 570-581, 2003.

Blandin S, Shiao SH, Moita LF, Janse CJ, Waters AP, Kafatos FC, *et al.* Complement-like protein TEP1 is a determinant of vectorial capacity in the malaria vector *Anopheles gambiae.* Cell 116: 661-670, 2004.

Bonura A, Vizzini A, Salerno G, Parrinello N, Longo V, Colombo P. Isolation and expression of a novel MBL-like collectin cDNA enhanced by LPS injection in the body wall of the ascidian *Ciona intestinalis.* Mol. Immunol. 46: 2389-2394, 2009.

Castillo MG, Goodson MS, McFall-Ngai M. Identification and molecular characterization of a complement C3 molecule in a lophotrochozoan, the Hawaiian bobtail squid *Euprymna scolopes.* Dev. Comp. Immunol. 33: 69-76, 2009.

C. elegans Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science 282: 2012-2018, 1998.

Dishaw LJ, Smith SL, Bigger CH. Characterization of a C3-like cDNA in a coral: phylogenetic implications. Immunogenetics 57: 535-548, 2005.

Dodds AW, Law SK. The phylogeny and evolution of the thioester bond-containing proteins C3, C4 and alpha 2-macroglobulin. Immunol. Rev. 166: 15-26, 1998.

Dolmer K, Sottrup-Jensen L. Disulfide bridges in human complement component C3b. FEBS Lett. 315: 85-90, 1993.

Dunn CW, Hejnol A, Matus DQ, Pang K, Browne WE, Smith SA, *et al.* Broad phylogenomic sampling improves resolution of the animal tree of life. Nature 452: 745-749, 2008.

Ehlenberger AG, Nussenzweig V. The role of membrane receptors for C3b and C3d in phagocytosis. J. Exp. Med. 145: 357-371, 1977.

Finnson KW, Tam BY, Liu K, Marcoux A, Lepage P, Roy S, *et al.* Identification of CD109 as part of the TGF-beta receptor system in human keratinocytes. FASEB J. 20: 1525-1527, 2006.

Fujito NT, Sugimoto S, Nonaka M. Evolution of thioester-containing proteins revealed by cloning and characterization of their genes from a cnidarian sea anemone, *Haliplanella lineate.* Dev. Comp. Immunol. 34: 775-784, 2010.

Graham M, Shin DH, Smith SL. Molecular and expression analysis of complement component C5 in the nurse shark (*Ginglymostoma cirratum*) and its predicted functional role. Fish Shellfish Immunol. 27: 40-49, 2009.

Hagiwara S, Murakumo Y, Mii S, Shigetomi T, Yamamoto N, Furue H, *et al.* Processing of CD109 by furin and its role in the regulation of TGF-beta signaling. Oncogene 29: 2181-2191, 2010.

Hagiwara S, Murakumo Y, Sato T, Shigetomi T, Mitsudo K, Tohnai I, *et al.* Up-regulation of CD109 expression is associated with carcinogenesis of the squamous epithelium of the oral cavity. Cancer Sci. 99: 1916-1923, 2008.

Holland LZ, Gibson-Brown JJ. The *Ciona intestinalis* genome: when the constraints are off. Bioessays 25: 529-532, 2003.

Hugli TE. Structure and function of the anaphylatoxins. Springer Semin. Immunopathol. 7:193-219, 1984.

Ishiguro H, Kobayashi K, Suzuki M, Titani K, Tomonaga S, Kurosawa Y. Isolation of a hagfish gene that encodes a complement component. EMBO J. 11: 829-837, 1992.

Janatova J, Tack BF. Fourth component of human complement: studies of an amine-sensitive site comprised of a thiol component. Biochemistry 20: 2394-2402, 1981.

Janssen BJ, Christodoulidou A, McCarthy A, Lambris JD, Gros P. Structure of C3b reveals conformational changes that underlie complement activity. Nature 444: 213-216, 2006.

Janssen BJ, Huizinga EG, Raaijmakers HC, Roos A, Daha MR, Nilsson-Ekdahl K, *et al.* Structures of complement component C3 provide insights into the function and evolution of immunity. Nature 437: 505-511, 2005.

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 94: 6340-6345, 1997.

Jing H, Xu Y, Carson M, Moore D, Macon KJ, Volanakis JE, *et al.* New structural motifs on the chymotrypsin fold and their potential roles in complement factor B. EMBO J. 19: 164-173, 2000.

Kasahara M, Nakaya J, Satta Y, Takahata N. Chromosomal duplication and the emergence of the adaptive immune system. Trends Genet. 13: 90-92, 1997.

Kenjo A, Takahashi M, Matsushita M, Endo Y, Nakata M, Mizuochi T, *et al.* Cloning and characterization of novel ficolins from the solitary ascidian, *Halocynthia roretzi.* J. Biol. Chem. 276: 19959-19965, 2001.

Kimura A, Sakaguchi E, Nonaka M. Multi-component complement system of Cnidaria: C3, Bf, and MASP genes expressed in the endodermal tissues of a sea anemone, *Nematostella vectensis.* Immunobiology 214: 165-178, 2009.

King N, Westbrook MJ, Young SL, Kuo A, Abedin M, Chapman J, *et al.* The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. Nature 451: 783-788, 2008.

Kinoshita T, Takata Y, Kozono H, Takeda J, Hong KS, Inoue K. C5 convertase of the alternative complement pathway: covalent linkage between two C3b molecules within the trimolecular complex enzyme. J. Immunol. 141: 3895-3901,

1988.

Lagueux M, Perrodou E, Levashina EA, Capovilla M, Hoffmann JA. Constitutive expression of a complement-like protein in toll and JAK gain-of-function mutants of *Drosophila*. Proc. Natl. Acad. Sci. USA 97: 11427-11432, 2000.

Lambris JD, Sahu A, Wetsel RA. The chemistry and biology of C3, C4, and C5. In: Volanakis JE, Frank MM (eds), The human complement system in health and disease. Marcel Dekker, Inc., New York, 1998.

Law SK, Lichtenberg NA, Levine RP. Evidence for an ester linkage between the labile binding site of C3b and receptive surfaces. J. Immunol. 123: 1388-1394, 1979.

Le Saux A, Ng PM, Koh JJ, Low DH, Leong GE, Ho B, *et al.* The macromolecular assembly of pathogen-recognition receptors is impelled by serine proteases, via their complement control protein modules. J. Mol. Biol. 377: 902-913, 2008.

Levashina EA, Moita LF, Blandin S, Vriend G, Lagueux M, Kafatos FC. Conserved role of a complement-like protein in phagocytosis revealed by dsRNA knockout in cultured cells of the mosquito, *Anopheles gambiae*. Cell 104: 709-718, 2001.

Levine RP, Dodds AW. The thioester bond of C3. Curr. Top. Microbiol. Immunol. 153: 73-82, 1990.

Li ZF, Wu XH, Engvall E. Identification and characterization of CPAMD8, a novel member of the complement 3/alpha2-macroglobulin family with a C-terminal Kazal domain. Genomics 83: 1083-1093, 2004.

Lin M, Sutherland DR, Horsfall W, Totty N, Yeo E, Nayar R, *et al.* Cell surface antigen CD109 is a novel member of the alpha(2) macroglobulin/C3, C4, C5 family of thioester-containing proteins. Blood 99: 1683-1691, 2002.

Marino R, Kimura Y, De Santis R, Lambris JD, Pinto MR. Complement in urochordates: cloning and characterization of two C3-like genes in the ascidian *Ciona intestinalis*. Immunogenetics 53: 1055-1064, 2002.

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. 176: 1497-1502, 1992.

Miller DJ, Hemmrich G, Ball EE, Hayward DC, Khalturin K, Funayama N, *et al.* The innate immune repertoire in cnidaria-ancestral complexity and stochastic gene loss. Genome Biol. 8: R59, 2007.

Miyazawa S, Azumi K, Nonaka M. Cloning and characterization of integrin alpha subunits from the solitary ascidian, *Halocynthia roretzi*. J. Immunol. 166:1710-1715, 2001.

Nonaka M, Azumi K, Ji X, Namikawa-Yamada C, Sasaki M, Saiga H, *et al.* Opsonic complement component C3 in the solitary ascidian, *Halocynthia roretzi*. J. Immunol. 162: 387-391, 1999.

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. 133: 3242-3249, 1984.

Nonaka M, Kimura A. Genomic view of the evolution of the complement system. Immunogenetics 58: 701-713, 2006.

Nonaka M, Takahashi M. Complete complementary DNA sequence of the third component of complement of lamprey. Implication for the evolution of thioester containing proteins. J. Immunol. 148: 3290-3295, 1992.

Pangburn MK, Muller-Eberhard HJ. The alternative pathway of complement. Springer Semin. Immunopathol. 7: 163-192, 1984.

Philippe H, Derelle R, Lopez P, Pick K, Borchiellini C, Boury-Esnault N, *et al.* Phylogenomics revives traditional views on deep animal relationships. Curr. Biol. 19: 706-712, 2009.

Philippe H, Lartillot N, Brinkmann H. Multigene analyses of bilaterian animals corroborate the monophyly of Ecdysozoa, Lophotrochozoa, and Protostomia. Mol. Biol. Evol. 22: 1246-1253, 2005.

Pinto MR, Chinnici CM, Kimura Y, Melillo D, Marino DR, Spruce LA, *et al.* CiC3-1a-mediated chemotaxis in the deuterostome invertebrate *Ciona intestinalis* (Urochordata). J. Immunol. 171: 5521-5528, 2003.

Podack ER, Stoffel W, Esser AF, Muller-Eberhard, HJ. Membrane attack complex of complement: distribution of subunits between the hydrocarbon phase of target membranes and water. Proc. Natl. Acad. Sci. USA 78: 4544-4548, 1981.

Prado-Alvarez M, Rotllant J, Gestal C, Novoa B, Figueras A. Characterization of a C3 and a factor B-like in the carpet-shell clam, *Ruditapes decussatus*. Fish Shellfish Immunol. 26: 305-315, 2009.

Raftos DA, Nair SV, Robbins J, Newton RA, Peters R. A complement component C3-like protein from the tunicate, *Styela plicata*. Dev. Comp. Immunol. 26: 307-312, 2002.

Raftos DA, Robbins J, Newton RA, Nair SV. A complement component C3a-like peptide stimulates chemotaxis by hemocytes from an invertebrate chordate-the tunicate, *Pyura stolonifera*. Comp. Biochem. Physiol. 134A: 377-386, 2003.

Rooijakkers SH, Wu J, Ruyken M, van Domselaar R, Planken KL, Tzekou A, *et al.* Structural and functional implications of the alternative complement pathway C3 convertase stabilized by a staphylococcal inhibitor. Nat. Immunol. 10: 721-727, 2009.

Schierwater B, Eitel M, Jakob W, Osigus HJ, Hadrys H, Dellaporta SL, *et al.* Concatenated analysis sheds light on early metazoan evolution and fuels a modern "urmetazoon" hypothesis. PLoS Biol. 7:e20, 2009.

Sekine H, Kenjo A, Azumi K, Ohi G, Takahashi M, Kasukawa R, *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. 167: 4504-4510, 2001.

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. 161:

6784-6793, 1998.

Sottrup-Jensen L, Folkersen J, Kristensen T, Tack BF. Partial primary structure of human pregnancy zone protein: extensive sequence homology with human alpha 2-macroglobulin. Proc. Natl. Acad. Sci. USA 81: 7353-7357, 1984.

Sottrup-Jensen L, Stepanik TM, Kristensen T, Lønblad PB, Jones CM, Wierzbicki DM, *et al.* Common evolutionary origin of alpha 2-macroglobulin and complement components C3 and C4. Proc. Natl. Acad. Sci. USA 82: 9-13, 1985.

Srivastava M, Begovic E, Chapman J, Putnam NH, Hellsten U, Kawashima T, *et al.* The Trichoplax genome and the nature of placozoans. Nature 454: 955-960, 2008.

Srivastava M, Simakov O, Chapman J, Fahey B, Gauthier ME, Mitros T, *et al.* The *Amphimedon queenslandica* genome and the evolution of animal complexity. Nature 466: 720-726, 2010.

Stroschein-Stevenson SL, Foley E, O'Farrell PH, Johnson AD. Identification of *Drosophila* gene products required for phagocytosis of *Candida albicans*. PLoS Biol. 4:e4, 2006.

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. 54: 671-679, 2002.

Takata Y, Kinoshita T, Kozono TH, Takeda J, Tanaka E, Hong K, *et al.* Covalent association of C3b with C4b within C5 convertase of the classical complement pathway. J. Exp. Med. 165: 1494-1507, 1987.

Terado T, Okamura K, Ohta Y, Shin DH, Smith SL, Hashimoto K, *et al.* Molecular cloning of C4 gene and identification of the class III complement region in the shark MHC. J. Immunol. 171: 2461-2466, 2003.

Terado T, Smith SL, Nakanishi T, Nonaka MI, Kimura H, Nonaka M. Occurrence of structural specialization of the serine protease domain of complement factor B at the emergence of jawed vertebrates and adaptive immunity. Immunogenetics 53: 250-254, 2001.

Torreira E, Tortajada A, Montes T, Rodriguez de Cordoba S, Llorca O. 3D structure of the C3bB complex provides insights into the activation and regulation of the complement alternative pathway convertase. Proc. Natl. Acad. Sci. USA 106: 882-887, 2009.

Volanakis JE. Overview of the complement system. In: JE, Frank MM (eds), The human complement system in health and disease, Marcel Dekker, Inc., New York, pp 9-32, 1998.

Volanakis JE, Arlaud GJ. Complement enzymes. In: Volanakis JE, Frank MM (eds), The human complement system in health and disease, Marcel Dekker, Inc., New York, pp 49-81, 1998.

Walport MJ. Complement. First of two parts. N. Engl. J. Med. 344: 1058-1066, 2001.

Wetsel RA, Ogata RT, Tack BF. Primary structure of the fifth component of murine complement. Biochemistry 26: 737-743, 1987.

Yoshizaki FY, Ikawa S, Satake M, Satoh N, Nonaka M. Structure and the evolutionary implication of the triplicated complement factor B genes of a urochordate ascidian, *Ciona intestinalis*. Immunogenetics 56: 930-942, 2005.

Zhu Y, Thangamani S, Ho B, Ding JL. The ancient origin of the complement system. EMBO J. 24: 382-394, 2005.