J Mol Evol (2002) 54:671–679
DOI: 10.1007/s00239-001-0068-z

JOURNAL OF MOLECULAR EVOLUTION

C6-Like and C3-Like Molecules from the Cephalochordate, Amphioxus, Suggest a Cytolytic Complement System in Invertebrates

Miho M. Suzuki,[1] Nori Satoh,[1] Masaru Nonaka[2]

[1] Department of Zoology, Graduate School of Science, Kyoto University, Sakyo-ku, Kyoto 606-8502, Japan
[2] Department of Biological Sciences, Graduate School of Science, University of Tokyo, Hongo, Tokyo 113-0033, Japan

Received: 24 August 2001 / Accepted: 12 November 2001

Abstract. The mammalian immune system has cytotoxic mechanisms, both cellular and humoral, that destroy the membrane integrity of target cells. The main effector molecules of these cytolytic mechanisms—perforin, used by killer lymphocytes, and the membrane attack complex (MAC) components of the complement system—share a unique module called the MAC/perforin module. Until now, both immunological cytotoxicity and the MAC/perforin module have been reported only in jawed vertebrates. Here, we report the identification of a protein containing the MAC/perforin module from the invertebrate cephalochordate, amphioxus (*Branchiostoma belcheri*), using expressed sequence tag (EST) analysis of the notochord. The deduced amino acid sequence of this molecule is most similar to the primary structure of human complement component C6 and is designated AmphiC6. AmphiC6 shares a unique modular structure, including the MAC/perforin module, with human C6 and other MAC components. Another EST clone predicts the presence of a thioester-containing protein with the closest structural similarity to vertebrate C3 (therefore designated AmphiC3). AmphiC3 retains most of the functionally important residues of vertebrate C3 and is shown by phylogenetic analysis to be derived directly from the common ancestor of vertebrate C3, C4, and C5. Only opsonic activity has been assigned to the invertebrate complement system until now. Therefore, this is the first molecular evidence for complement-mediated immunological cytotoxicity in invertebrates.

Key words: Immunological cytotoxicity — Modular structure — Complement — Amphioxus — Membrane attack complex

Introduction

The mammalian immune system is equipped with both cell-mediated and humoral cytotoxic mechanisms. Cell-mediated cytotoxicity is mediated by cytotoxic T cells and natural killer (NK) cells, and perforin acts as an effector molecule (Lichtenheld et al. 1988; Shinkai et al. 1988). Upon activation of these cells, perforin stored in their lytic granules is released and polymerizes to generate transmembrane pores in target cell membranes. Humoral cytotoxicity is instigated by the assembly of the terminal components of complement, leading to the formation of a similar pore called the "membrane attack complex" (MAC) (Plumb and Sodetz 1998). The complement components involved in MAC formation are C5, C6, C7, C8α–γ, and C9, and several of these (C6, C7, C8α, C8β, and C9) share a characteristic modular structure that includes the MAC/perforin module, which is also found in perforin. Therefore, all pivotal molecules involved in immunological cytotoxicity seem to have originated from a common ancestral molecule containing the MAC/perforin module. These molecules have been identified only in mammalian species so far, except for the identification of MAC (Nonaka et al. 1981) and

Correspondence to: M. Nonaka; *email:* mnonaka@biol.s.u-tokyo.ac.jp

cDNA clones for C8 and C9 (Katagiri et al. 1999; Stanley et al. 1985; Yeo et al. 1997) in bony fish. Therefore, the evolutionary origins of the immunological cytotoxic molecules are largely unknown. Recent progress in genome science has cast some light on the evolution of the immunological cytotoxic molecules containing the MAC/perforin module. There are only six copies of this module in the human genome (Venter et al. 2001), indicating that only perforin and the MAC components contain this module. Moreover, there is no MAC/perforin module in the *Drosophila* genome (Adams et al. 2000). Therefore, this module seems to have been established in the deuterostome lineage, although its presence in invertebrate deuterostomes is yet to be demonstrated.

Two major parts of the vertebrate immune system, adaptive immunity and innate immunity, have quite different evolutionary origins. The former originated in the jawed vertebrate lineage, whereas the latter has been traced back to the protostomes (Hoffmann et al. 1999). Although the mammalian complement system is intrinsically involved in both innate and adaptive immunity, accumulating evidence indicates that this system was established as one of the major components of innate immunity in the deuterostome lineage, well before the emergence of adaptive immunity (Nonaka 2001). The central component C3 and/or its activating enzyme factor B (Bf) have been identified in sea urchins (echinoderms) (Al-Sharif et al. 1998; Smith et al. 1998), ascidians (urochordates) (Nonaka et al. 1999), and lamprey (Nonaka and Takahashi 1992; Nonaka et al. 1994) and hagfish (Ishiguro et al. 1992) (both cyclostomes). In contrast, the recently published *Drosophila* (Adams et al. 2000) and *Caenorhabditis elegans* (*C. elegans* Sequencing Consortium 1998) genomes do not contain identifiable C3 and Bf homologues, suggesting that the complement system is unique to the deuterostomes. The mammalian complement system has three activation pathways, referred to as the classical, alternative, and lectin pathways, and one cytolytic effector pathway that leads to membrane damage (Volanakis 1998). All jawed vertebrates so far analyzed possess all four pathways, whereas only components of the alternative and lectin pathways have been identified in the apparently primitive complement systems of the sea urchin, ascidian, and cyclostomes (Nonaka 2001). Therefore, the evolution of the complement system seems to involve an increase in the number of components and activation pathways, probably via gene duplications, with the emergence of jawed vertebrates. Moreover, from a functional viewpoint, only opsonic activity that enhances phagocytosis has been identified in the ascidian and lamprey complement systems (Nonaka et al. 1984, 1999). It is still unclear whether these primitive complement systems also have cytolytic activity.

Cephalochordata is classified as a subphylum of Chordata, together with Urochordata and Vertebrata. Chordata, Hemichordata, and Echinodermata are considered the only deuterostome phyla. Although the phylogeny of the deuterostomes is still in dispute, recent molecular data indicate that Vertebrata is monophyletic and a sister clade of Cephalochordata. Urochordata is also monophyletic and constitutes a sister group to Vertebrata and Cephalochordata (Cameron et al. 2000; Wada and Satoh 1994). The presence of myotomes in vertebrates and cephalochordates and the presence of the tunic in the urochordates further support this phylogenetic relationship. Therefore, Cephalochordata occupies an important position in the deuterostome phylogeny as the closest relative to Vertebrata and provides a unique opportunity to analyze the evolution of the vertebrate immune system. However, there is still no report on the cephalochordate immune system at the protein or cDNA level.

Materials and Methods

Animals and cDNA Cloning. Adult amphioxus, *Branchiostoma belcheri,* were collected near the National Research Institute of Aquaculture, Mie, Japan. Partial cDNA clones for possible complement genes of amphioxus were originally derived from a *B. belcheri* notochord cDNA library by EST analysis (Suzuki and Satoh 2000). Full-length clones were isolated from the same Uni-ZAP XR library (Stratagene, La Jolla, CA) using the partial cDNAs as probes and were converted into pBluescript SK(−) phagemids using a rapid excision kit (Stratagene). The complete sequences of the inserts in these clones were determined using a Big-Dye Terminator Cycle Sequencing Ready Reaction kit and an ABI PRISM 377 DNA sequencer (Perkin Elmer, Norwalk, CT). Each sequence was determined at least twice from both strands.

Phylogenetic Analysis. The deduced amino acid sequences of the amphioxus complement genes were aligned with the corresponding sequences of various animals using the CLUSTAL W software (Thompson et al. 1994). Based on this alignment, a phylogenetic tree was constructed using the neighbor joining method (Saitou and Nei 1987).

Results

Identification of Two Possible Complement Component Genes from Amphioxus. Expressed sequence tag (EST) analysis of the amphioxus notochord (Suzuki and Satoh 2000) resulted in the identification of two possible complement genes that, according to a BlastX homology search, are most similar to vertebrate complement components C6 and C3. The complete primary structures of these proteins, designated AmphiC6 and AmphiC3, respectively, were deduced from the nucleotide sequences of full-length cDNA clones.

Structure of AmphiC6. One of two AmphiC6 EST clones contained a full-length protein-coding region within a 3.5-kb-long insert. The deduced complete primary structure of AmphiC6 predicts a characteristic

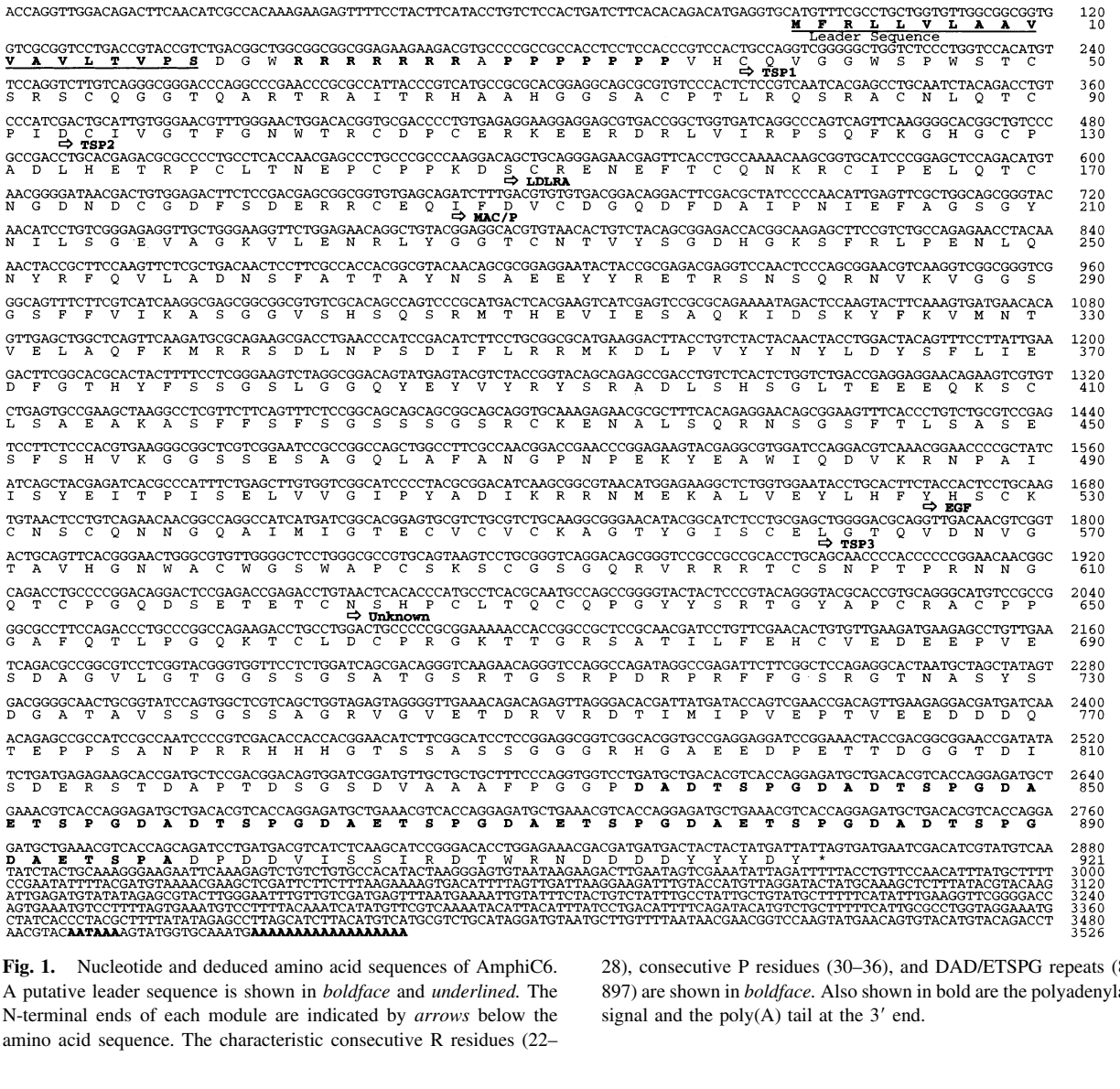

Fig. 1. Nucleotide and deduced amino acid sequences of AmphiC6. A putative leader sequence is shown in *boldface* and *underlined.* The N-terminal ends of each module are indicated by *arrows* below the amino acid sequence. The characteristic consecutive R residues (22–28), consecutive P residues (30–36), and DAD/ETSPG repeats (835–897) are shown in *boldface.* Also shown in bold are the polyadenylation signal and the poly(A) tail at the 3′ end.

module structure in its central portion (Figs. 1 and 2). This module structure, reading from the N-terminal end, comprises two TSP (thrombospondin type 1) modules, one LDLRA (low-density lipoprotein-receptor class A) module, one MAC/perforin (membrane attack complex/perforin) module, one EGF (epidermal growth factor) module, and one TSP module and is identical to the six-N-terminal module structure of human C6 (DiScipio and Hugli 1989; Haefliger et al. 1989) (Fig. 2). The four C-terminal modules of human C6, two SCR (short consensus repeats) and two FIM (factor I modules), are missing from AmphiC6. Compared with the module structures of human C8α (Rao et al. 1987) and C8β (Haefliger et al. 1987; Howard et al. 1987), the central module structure of AmphiC6 has one extra TSP module at its N terminus. The module structure of other complement MAC components, C7 (DiScipio et al. 1988) and C9 (Stanley et al. 1985), and perforin (Lichtenheld et al. 1988; Shinkai et al. 1988) are less similar to that of AmphiC6 (Fig. 2). From the shared module structure, it is apparent that AmphiC6 is evolutionarily related to perforin and the MAC components of the mammalian complement system.

For phylogenetic analysis, the amino acid sequences of the TSP, LDLRA, MAC/perforin, and EGF modules shared by AmphiC6 and the human MAC components and the corresponding region of perforin were aligned using the CLUSTAL W software. Part of this alignment, corresponding to the entire MAC/perforin module, is shown in Fig. 3. In the MAC/perforin module, AmphiC6 shares 36, 33, 30, 27, 25, and 27% amino acid identity with human C6, C7, C8α, C8β, C9, and perforin, respectively. Four cysteine residues in the MAC/perforin module of AmphiC6 are perfectly conserved in all human MAC components and perforin. The membrane-interaction site identified in human C9 (Peitsch et al. 1990) is marked by a line over the alignment (Fig. 3). The degree of sequence similarity between AmphiC6

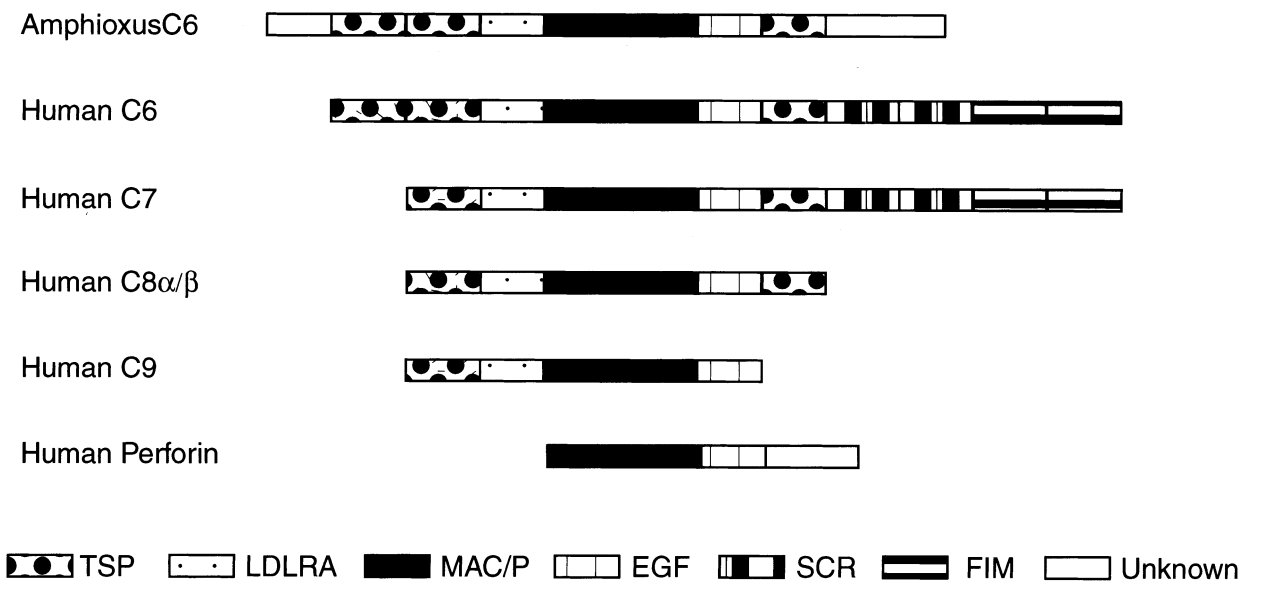

Fig. 2. Module structures of AmphiC6, human MAC components, and perforin. Each module is shown by a different shading pattern, with its name given at the bottom.

and the human MAC components and perforin is highest in this region, implicating it in a membrane-interacting function.

The phylogenetic tree based on this alignment of the four modules was drawn using the neighbor-joining method. As shown in Fig. 4A, AmphiC6 forms a cluster with human C6, and this clustering is supported by a high bootstrap value (97%). This result suggests that gene duplications among the MAC components predated the divergence of the cephalochordates from the vertebrates. However, this is an unrooted tree and further studies, which may include a search for C7, C8, or C9 genes in amphioxus, are required to clarify this point. A phylogenetic tree constructed using only the MAC/perforin module had the same topology (data not shown). In addition to the central module structure, AmphiC6 has unique sequences in its N- and C-terminal regions. These unique sequences show no significant similarity to any known sequence when analyzed by a BLAST search of GenBank. The N-terminal region is short and contains Arg and Pro stretches. The C-terminal region is about 300 amino acid residues long and contains nine consecutive repeats of the heptapeptide DAD/ETSPG (Fig. 1; boldface letters) at its C-terminal end.

Structure of AmphiC3. The predicted prepro-AmphiC3 is 1792 amino acids long (Fig. 5) and shows sequence homology with C3, C4, and C5 of other species throughout its entire length. The amino acid sequence identity is 29% with human C3 (de Bruijn and Fey 1985), 29% with human C4 (Belt et al. 1984), 26% with human C5 (Haviland et al. 1991), and 25% with human α2-macroglobulin (α2M) (Kan et al. 1985). There is a possible β–α processing site, and the typical thioester site containing a catalytic His residue (Dodds et al. 1996). The possible C3a region contains six Cys residues, perfectly conserved in vertebrate C3a, C4a, and C5a but not conserved in the ascidian (Nonaka et al. 1999) or sea urchin (Al-Sharif et al. 1998) C3 sequences. Furthermore, the C-terminal region of AmphiC3 shows the characteristic distribution of eight cysteine residues that are shared by all the C3, C4, and C5 sequences of vertebrates and invertebrates so far determined, but are not found in α2M, including the *C. elegans* and *Drosophila* thioester proteins (*C. elegans* Sequencing Consortium 1998; Lagueux et al. 2000). Therefore, AmphiC3 is identified conclusively as a member of the C3/C4/C5 family, and not as α2M. Phylogenetic analysis using the neighbor-joining method with human α2M as an outgroup shows AmphiC3 located outside the vertebrate C3, C4, and C5 cluster (Fig. 4B). This result strongly suggests that the gene for AmphiC3 was derived directly from the common ancestor of the vertebrate C3, C4, and C5 genes and that the gene duplications giving rise to the latter occurred in the vertebrate lineage after the divergence of Cephalochordata.

Discussion

This is the first report of an invertebrate gene encoding the MAC/perforin module, suggesting an ancient origin for the cytotoxic pore-forming mechanism that uses this module. All other modules found in AmphiC6 (TSP, LDLRA, and EGF) are present in the *Drosophila* or *C. elegans* genomes, whereas the MAC/perforin module does not occur in these protostome genomes (Venter et al. 2001). Therefore, it is probable that the MAC/perforin module arose in the deuterostome lineage and that the TSP, LDLRA, and EGF modules were recruited to es-

Fig. 3. Amino acid sequence alignment of the MAC/perforin modules of AmphiC6, the human MAC components, and perforin. Amino acid residue numbers of the rightmost residues of each lane of AmphiC6 are shown at the *right*. Cysteine residues are shown in *boldface* and *asterisks* indicate the completely conserved residues. The *line* above the alignment indicates the human C9 region identified as a membrane-interacting region.

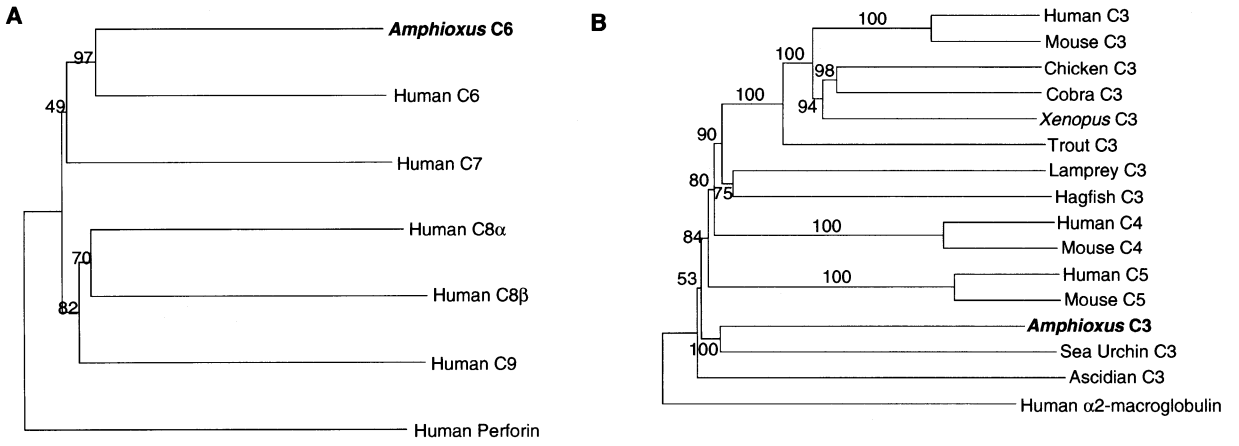

Fig. 4. Phylogenetic tree calculated from the sequences of MAC/perforin module-containing proteins and C3 family proteins. A MAC/perforin module-containing proteins. Only the four domains shared by all mammalian MAC components and AmphiC6 (TSP, LDLRA, MAC/P, and EGF) were used in the calculation. For perforin, which lacks the TSP and LDLRA modules, the same number of amino acid residues was used for comparison. **B** C3 family. The complete amino acid sequences of C3, C4, and C5 of various animals were aligned using the CLUSTAL W software, and a phylogenetic tree was constructed by the neighbor joining method. Human α2-macroglobulin was used as the outgroup. Numbers indicate the proportion of bootstrap replicates supporting the given branching pattern.

tablish the ancestral immunological cytolytic gene. AmphiC6 shows a closer similarity in modular structure to the vertebrate MAC components, especially C6, than to vertebrate perforin. Although further searches are required in amphioxus for proteins containing the MAC/perforin module, our results suggest that the complement-based cytotoxic mechanisms have a more ancient evolutionary origin than the lymphocyte-based cytotoxic mechanisms. Accumulating evidence indicates that adaptive immunity was established in the jawed vertebrate lineage (Kasahara et al. 1997), whereas the complement system is more ancient (Nonaka 2001). The present results give further support to this evolutionary scenario of the immune system.

Recently, C3-like thioester-containing proteins (TEP) have been reported in insects (Lagueux et al. 2000; Levashina et al. 2001), suggesting a protostome origin for the complement system. However, insect TEPs and vertebrate C3, C4, and C5 share no derived structural characteristics, indicating that they arose independently from their common ancestor, α2M (Sottrup-Jensen et al. 1985). In contrast, authentic complement genes have been identified in deuterostome invertebrates, such as sea urchins and ascidians, indicating that the complement system was established at an early stage of deuterostome evolution (Nonaka 2001). Therefore, it is highly probable that AmphiC3 and AmphiC6, and possibly additional components such as Bf and MASP, found in other invertebrate deuterostomes, constitute a multicomponent complement system in amphioxus. Another intriguing possibility, however, is that AmphiC3 and AmphiC6 act independently. In this context, it is worth noting that there are unique sequences at the N and C termini of AmphiC6. At the N terminus, the unique RRRRRRRA-PPPPPPP sequence immediately follows a possible leader sequence. The unique C-terminal sequence contains nine consecutive repeats of the DAD/ETSPG motif. A similar heptapeptide repeat is present in the C-terminal region of the largest subunit of RNA polymerase II and is a known phosphorylation site that is important for transcriptional regulation (Uptain et al. 1997). Although the functional significance of these unique sequences in AmphiC6 is unknown, they may be involved in an activation process of AmphiC6. If this is the case, the activation mechanism for AmphiC6 is quite different from that of the terminal pathway of the vertebrate complement system, even if AmphiC6 is a complement component of amphioxus. The third possibility, that AmphiC6 has a novel function different from that of complement terminal components or perforin, has not yet been ruled out. Further studies of invertebrates are required, especially at the protein level, to clarify the original function of the MAC/perforin-module-containing proteins.

The cDNA clones for AmphiC3 and AmphiC6 were originally isolated from the notochord. However, a preliminary in situ hybridization analysis indicates that both

Fig. 5. Nucleotide and deduced amino acid sequences of AmphiC3. The nucleotide sequence of AmphiC3 is presented together with the deduced amino acid sequence. The initiation methionine codon was identified by comparison with vertebrate C3, and a putative leader sequence is shown in *boldface* and *underlined.* The nucleotide and amino acid numbers of the rightmost residues are shown for each lane. The β–α processing site (amino acid residues 660–663), the thioester site (1012–1016), and the His residue (1129) associated with the thioester are shown in *boldface* and *underlined.* Also shown in *boldface* and *underlined* are six cysteine residues perfectly conserved in vertebrate C3a, C4a, and C5a (683–719) and the conserved cysteine residues in the C-terminal C3/C4/C5-specific region (1554–1790).

```
CAGAGAAACTTGGAGGCGAAGCACGCGCCTAGAACTTGTCCCGGCCCGTTTTGCCGAAACTGTCCCAGAACAGCTACGATGATCCGTAGTTTGGTGTGTGTGTCTCCTGTTCGCCGCT  120
                                                                M  I  R  S  L  V  C  V  C  L  L  F  A  A                14
                                                                                                Leader Sequence
GTGCTGGGCCAGCAGGGGCCAGCATATTTCATCTCTGTTCCCAACCTCCTGAAAGTTGGTACAGAGGAGACTGTGTCAGTCAACGTCTTCAATGTTGCCAACCCTGTGCGGGTGAAGGTT  240
V  L  G  Q  Q  G  P  A  Y  F  I  S  V  P  N  L  L  K  V  G  T  E  E  T  V  S  V  N  V  F  N  V  A  N  P  V  R  V  K  V    54

TACCTTCAGGACTATCCTGACCGAAAAACCACCTTTTCGGAGGCAGAGGTTGATGTCAACCAAGATGAGCCCAGTCTTGTGACAGTTCGGGTGAATCCCGACAACCTGCCAGAGAGCAGG  360
Y  L  Q  D  Y  P  D  R  K  T  T  F  S  E  A  E  V  D  V  N  Q  D  E  P  S  L  V  T  V  R  V  N  P  D  N  L  P  E  S  R    94

GCGACCAAGCGCTATGTCTACGTGGTGGCCAAGTCCGATGATCCACAACTGACGTTCCAGAAAGAGGCACAAGTGCTGCTGAGTTACCAACAAGGCTATGTCTTTGTGCAGACTGACAAG  480
A  T  K  R  Y  V  Y  V  V  A  K  S  D  D  P  Q  L  T  F  Q  K  E  A  Q  V  L  L  S  Y  Q  Q  G  Y  V  F  V  Q  T  D  K  134

CCTATATACACTCCCAACCAGAAAGTGAAGATGAGAATCATGCCTCTGGACCAGGACATGACTCCAGCCAGCCAGCCAGTAAAACTGGAAATTTTGAATCCCCAGGGAATCATTGTGGAG  600
P  I  Y  T  P  N  Q  K  V  K  M  R  I  M  P  L  D  Q  D  M  T  P  A  S  Q  P  V  K  L  E  I  L  N  P  Q  G  I  I  V  E  174

CGGAAAACCTTCCCGGGGTCCGCAACGGGGTTCATCGCAGAAACCTTTGACTTCCCAGCATTCCCCCTGTTTGGAAATTGGACAGCTATTGCCCACTATGGTCCTGAGATGCAGCTGAAT  720
R  K  T  F  P  G  S  A  T  G  F  I  A  E  T  F  D  F  P  A  F  P  L  F  G  N  W  T  A  I  A  H  Y  G  P  E  M  Q  L  N  214

GTATCTACTCAGTTTGAAGTCAAGGAATATGTGTTGCCGACCTACGGTGTGAGAATCATCCCGTCCAATCCCTACATCTTACCACAGGATGATGTCATCAGTGGGGAGGTGGAGGCACTA  840
V  S  T  Q  F  E  V  K  E  Y  V  L  P  T  Y  G  V  R  I  I  P  S  N  P  Y  I  L  P  Q  D  D  V  I  S  G  E  V  E  A  L  254

TATACTTATGGAAAAGGTGTGGATGGTTTTCTGGACGTGAAGTTTGGTATAATTGACCTGGAGGGAAACAGACAGCTTTTTGCACAGCTTCAAACAGAGGTGAATGGCGGATTTGGTTTC  960
Y  T  Y  G  K  G  V  D  G  F  L  D  V  K  F  G  I  I  D  L  E  G  N  R  Q  L  F  A  Q  L  Q  T  E  V  N  G  G  F  G  F  294

TATGAAATCGACACACAGAGAATCAAAGACCTTGACCTGTGGTTTCCTGAAGGCAGTCGGTTGTACTTAGAGGCAGCCGTGACAGAAGAAGCGGGGGGGCTGCGAGAAATGGCAGTTCTC  1080
Y  E  I  D  T  Q  R  I  K  D  L  D  L  W  F  P  E  G  S  R  L  Y  L  E  A  A  V  T  E  E  A  G  G  L  R  E  M  A  V  L  334

ACATCTGTCAGGTTTGAGACATCGCCGTTCAGGATTGGCTATGACCTGACAGCAACGCATTTCAAGCCAGGCTTACCATTCCTTGTCAAGCTGACTCTGACATATCCAGATACCAAACCA  1200
T  S  V  R  F  E  T  S  P  F  R  I  G  Y  D  L  T  A  T  H  F  K  P  G  L  P  F  L  V  K  L  T  L  T  Y  P  D  T  K  P  374

GCTCAGGACATTCCAGTTCGGGTGTCCGCCACTGCGATTATTCCTGGTCAGGACCCCATCGTTATTCTGGGTCGTAACAACGAGCACAACAGTGACACAACCAATCAGTACGGACAAGCG  1320
A  Q  D  I  P  V  R  V  S  A  T  A  I  I  P  G  Q  D  P  I  V  I  L  G  R  N  N  E  H  N  S  D  T  T  N  Q  Y  G  Q  A  414

AGCTTCACGGTCGACGTTCCGCCCCGGAACACAGACCTTAACTGTCACGGCAAAGACAGAGCAGGTGGGTCTGCCTGTGGCTCACCAAGCCCAGGAGAATTTCGAGGCCACGCCGTACCAG  1440
S  F  T  V  D  V  P  P  G  T  Q  T  L  T  V  T  A  K  T  E  Q  V  G  L  P  V  A  H  Q  A  Q  E  N  F  E  A  T  P  Y  Q  454

TCTCCGTCTGGAAGCTACCTCCTGGTCCGCGTGCTTCAACGTGGTCCCGTTCCCGTGGACGAGGCGATTGACGTGGAGGCTGTCGTCACCAAACAGAACGACATCCAGAGCTATAACTAT  1560
S  P  S  G  S  Y  L  L  V  R  V  L  Q  R  G  P  V  P  V  D  E  A  I  D  V  E  A  V  V  T  K  Q  N  D  I  Q  S  Y  N  Y  494

ATGGTGGTGACACGAGGACAAGTGACACTGCAGGGGAAGATTGTGCGACAAGGGGGCGTGCTGAAGACGATAACGTTCCGTACTTCTGCGGTGATGGCGCCCATCTCGCGACTCATCGTT  1680
M  V  V  T  R  G  Q  V  T  L  Q  G  K  I  V  R  Q  G  G  V  L  K  T  I  T  F  R  T  S  A  V  M  A  P  I  S  R  L  I  V  534

TACTACATCAACCTTCAGGGGGAGGTGGTGGCAGACTCCACACTACTGGAGATCGAAAATGTCTGCAGAAATAAGGTTATGGTGTCCTCAACTGAAGATGTAGAACCACAGGAACAGGCC  1800
Y  Y  I  N  L  Q  G  E  V  V  A  D  S  T  L  L  E  I  E  N  V  C  R  N  K  V  M  V  S  S  T  E  D  V  E  P  Q  E  Q  A  574

AACATTGAGGTGAACGCCGACCCCAACTCACTGGTGGGACTGTTGGCTGTAGACCAGGCTGTCTACCTGCTCAATAACTACAACAGACTCACTTCTCAAAAGATGTTCCAGGCCATGGCC  1920
N  I  E  V  N  A  D  P  N  S  L  V  G  L  L  A  V  D  Q  A  V  Y  L  L  N  N  Y  N  R  L  T  S  Q  K  M  F  Q  A  M  A  614

AAGTATGATCAAGGATGTGGCCCAGGAGGGGGACAGGACAGTGCAAACGTCTTCAAGGACGCTGGTGTAACTGTGCTGACCAACACCCAGCTGTCGCCTGCAGTCCGCAGCAATGCTGGC  2040
K  Y  D  Q  G  C  G  P  G  G  G  Q  D  S  A  N  V  F  K  D  A  G  V  T  V  L  T  N  T  Q  L  S  P  A  V  R  S  N  A  G  654

TGCGGAGGCCAGTCACGCAGAAAAAGGGATCTGCAAAACTCACTGGAGGCAAAAGTTATGGAGTTCAATGAGACTCTCCAGCCCTGCTGTATGGACGGCCAACAGTGGGATCCTCTTGGC  2160
C  G  G  Q  S  R  R  K  R  D  L  Q  N  S  L  E  A  K  V  M  E  F  N  E  T  L  Q  P  C  C  M  D  G  Q  Q  W  D  P  L  G  694
                  Processing Site                                                                 Thioester Site
CGGTCTTGCCTCCAACGCGCCAAGCTGAACTCTACGTCCCAAGATGAGTGCTATTTCGCCTTCCTGACGTGCTGTAACCACGCACGCAGCCTGCGGAGGCTGGGACGCGGGCGCGGACGC  2280
R  S  C  L  Q  R  A  K  L  N  S  T  S  Q  D  E  C  Y  F  A  F  L  T  C  C  N  H  A  R  S  L  R  R  L  G  R  G  R  G  R  734

ATGGGAGGCGGCGGTGGCCTGCTCGATATCGATATCGATGAAGATGAGTCACAGCTCGTGGCACGTACGGAATTCCCAGAAACTTGGATCTTTGAGGATGTCCAAGTGGATGATAGAGGT  2400
M  G  G  G  G  G  L  L  D  I  D  I  D  E  D  E  S  Q  L  V  A  R  T  E  F  P  E  T  W  I  F  E  D  V  Q  V  D  D  R  G  774

CAGGCGGTTGTTCCTGTGACAGTTCCCGGCAGTATCACCACCTGGGTCATCCAAGCGGTGGGCATCTCCACTGCAAACGGCATGTGTGTGGCCAAACCCTTCCGTATGAAGTCCTTCAAG  2520
Q  A  V  V  P  V  T  V  P  G  S  I  T  T  W  V  I  Q  A  V  G  I  S  T  A  N  G  M  C  V  A  K  P  F  R  M  K  S  F  K  814

AAGTTCTTCATCCATCTGCAGCTGCCTTACTCTATCATCAGGGGAGAGCAAGTGGCCATCCGGGCAACGATATTCAACTACGACCAACAGGACCTCAGGGTGAACGTGTACATGCAAGGT  2640
K  F  F  I  H  L  Q  L  P  Y  S  I  I  R  G  E  Q  V  A  I  R  A  T  I  F  N  Y  D  Q  Q  D  L  R  V  N  V  Y  M  Q  G  854

GTGGAGGGTGTATGCTCAGGAGCGAGGGCAGGAGAGAAGCGAGAGGAAAACACTCTTTATCAAGGGCAACGACGCTGCCTCTGTCCTATTTCCCATCATCCCTCTGGAGGTCGGGACG  2760
V  E  G  V  C  S  G  A  R  A  G  E  R  S  E  R  K  T  L  F  I  K  G  N  D  A  A  S  V  L  F  P  I  I  P  L  E  V  G  T  894

TTTCCGATCCGTGTTGTCGCCTTCAGCACTGCCGCAGGAGGGGACATCATCGAAAAGTCGCTACAAGTTATCCCAGAGGGAGTTGAGAGGAGGTTGGTGAGATCCATCTTTGTGGATCCG  2880
F  P  I  R  V  V  A  F  S  T  A  A  G  G  D  I  I  E  K  S  L  Q  V  I  P  E  G  V  E  R  R  L  V  R  S  I  F  V  D  P  934

AAGGGGCGAGCACGAGACAGAAAACGGGAGGGGGAGGAAGAGGTGGCGCTTCCAACTGAACATGATGTTGACCCTGACAATGGACTACAGTTTGATGTTGTGGACGTACGCTTGCCTCCA  3000
K  G  R  A  R  D  R  K  R  E  G  E  E  E  V  A  L  P  T  E  H  D  V  D  P  D  N  G  L  Q  F  D  V  V  D  V  R  L  P  P  974

GAGACCATTGAAGGGTCAGAGCAATGTGCCGTCAGTATCATGGGTGACATCATGGGCCCAACTATCACTACCACCATCGGCGGCCTGGGAACCCCTTCTTCGCCTTCCAACGGGCTGCGGA  3120
E  T  I  E  G  S  E  Q  C  A  V  S  I  M  G  D  I  M  G  P  T  I  T  T  I  G  G  L  G  T  L  L  R  L  P  T  G  C  G  1014
                                                                                                Thioester Site
GAACAGACCATGATTAAGCTGGCGCCCAACGTCTACGTCCTCAGCTACCTGCACTGCACGGATCAGATCACTAAAGATGTGGAGGAGAAGGCATACGACTTTATCAGGCAAGGCTACAAC  3240
E  Q  T  M  I  K  L  A  P  N  V  Y  V  L  S  Y  L  H  C  T  D  Q  I  T  K  D  V  E  E  K  A  Y  D  F  I  R  Q  G  Y  N  1054

AAACAACTTAGCCACAGGCGACCGGAAGGCTGTTTTTCTGTGTGGGGCCAGAACAACAGATATCCCTGTAGCACTTGGCTGACAGCATTTGTGAACAAAGTGTTCTGCCAGGCTAAGAAG  3360
K  Q  L  S  H  R  R  P  E  G  C  F  S  V  W  G  Q  N  N  R  Y  P  C  S  T  W  L  T  A  F  V  N  K  V  F  C  Q  A  K  K  1094

TTTGTTACCAGTATCGATGAAGAAGCTGTGTGTAAGGCCACAGAGTGGCTCCTCAGCACACAACGGGAGGATGGAGCTTTCAAAGAGGTCTACAAAGTCCATCACAGAGAAATGACGGGT  3480
F  V  T  S  I  D  E  E  A  V  C  K  A  T  E  W  L  L  S  T  Q  R  E  D  G  A  F  K  E  V  Y  K  V  H  H  R  E  M  T  G  1134
                                                                                Thioester Catalytic
GGTGTACAGGGAGACGCATCCATGACGGCATTTGTGCTCATTTCTCTGCTGGAAAACTGTGAATGCCCCATAGCTGAACGCAGCATTGCCATTGAGAGAGCCACCTTGTTTCTTGAGAGA  3600
G  V  Q  G  D  A  S  M  T  A  F  V  L  I  S  L  L  E  N  C  E  C  P  I  A  E  R  S  I  A  I  E  R  A  T  L  F  L  E  R  1174

CAGCTGGAGCAGCTAAAGCGTCCGTACGTGATCGCCATAGTAACGTACGCGCTCCACTTAGCCGACAGTCCGCTCAAAGGTGCCGCCAATGAGAAGCTCAGAAGTATCGCCAAGTATGAC  3720
Q  L  E  Q  L  K  R  P  Y  V  I  A  I  V  T  Y  A  L  H  L  A  D  S  P  L  K  G  A  A  N  E  K  L  R  S  I  A  K  Y  D  1214

GAAGGTACAAATTCCCGTTACTGGGAAGCTGATGCCTCCAGCATCGCTGATGGGCAGCAGCCCTACTGGTACACCCGGAAGCCCAGTGCCATCGCTGTAGAGACCACGGCGTATGCTTTA  3840
E  G  T  N  S  R  Y  W  E  A  D  A  S  S  I  A  D  G  Q  Q  P  Y  W  Y  T  R  K  P  S  A  I  A  V  E  T  T  A  Y  A  L  1254

CTGACACAGATGCACATTGGAGACATCCAGTACAGCAACCCCATCGTCGTGTGGCTCACACAGCAGCGCAACAGTGCAGGGGGCTTTGTGTCAACTCAGGACACAGTCGTGGCTCTCCAG  3960
L  T  Q  M  H  I  G  D  I  Q  Y  S  N  P  I  V  V  W  L  T  Q  Q  R  N  S  A  G  G  F  V  S  T  Q  D  T  V  V  A  L  Q  1294

GCTCTCTCCTCCTACTGTGGCAGCACCAAGGTCGACCCGACCCAGTTCACCTGCCAGATGACCTCCGATAACGACCTGGACTATAACGAGGAGATTCACGTGGATAAAGATAACGCCCTG  4080
A  L  S  S  Y  C  G  S  T  K  V  D  P  T  Q  F  T  C  Q  M  T  S  D  N  D  L  D  Y  N  E  E  I  H  V  D  K  D  N  A  L  1334

ATCAACCAAGAAAAAACTGCACCTGTTGGAGGGAAGTTGTTCCTGAGTACAAGTGGGACTGGCATCGGACAGATGCAGGTGGAGGTTCGTTATCATACCCCGGATGTTCACCGGGAGCGC  4200
I  N  Q  E  K  T  A  P  V  G  G  K  L  F  L  S  T  S  G  T  G  I  G  Q  M  Q  V  E  V  R  Y  H  T  P  D  V  H  R  E  R  1374

TGCCTCTTCGAAGTTGTCGTGACGACAGAGGAGGCGGAAGGTCCAGTGGAGCCTGACGAGGAACCAGAGGAAGGTCAGGAAGGAGACGAGGACTACCCAGACTACGATGGCGAACTCGAC  4320
C  L  F  E  V  V  V  T  T  E  E  A  E  G  P  V  E  P  D  E  E  P  E  E  G  Q  E  G  D  E  D  Y  P  D  Y  D  G  E  L  D  1414

GCCCGGGAAAGAAGTGCGTTCAGTTCACGGTCCCGCATCGGAGGTCTCCTGAGCTCGGGACGGTCGCGGTCTCGTATTGCACGTCAAGCAGAGGATGACCAAAGCCAGCACTTCTATTTG  4440
A  R  E  R  S  A  F  S  S  R  S  R  I  G  G  L  L  S  S  G  R  S  R  S  R  I  A  R  Q  A  E  D  D  Q  S  Q  H  F  Y  L  1454

GTTAGAGTCTGCACCAGCTACCGAGGCCAGAGGGGTGCGTCCAATATGGCCATCATGGACATCGGTATGTTCTCTGGTTTTGAGCCAGTCAAGCAAGACCTGGAACAGCTGTTGAACCGT  4560
V  R  V  C  T  S  Y  R  G  Q  R  G  A  S  N  M  A  I  M  D  I  G  M  F  S  G  F  E  P  V  K  Q  D  L  E  Q  L  L  N  R  1494

GCTGGGAAGAAAGCCGTGCAGCGTTATGAGACAACAAACCGTGCAGTTCTGTTGTACTTTGATGAGATAACCTCAGAGGAGATTTGTGTGAGCATACGTATGAAACGGGTGATGCTGGTA  4680
A  G  K  K  A  V  Q  R  Y  E  T  T  N  R  A  V  L  L  Y  F  D  E  I  T  S  E  E  I  C  V  S  I  R  M  K  R  V  M  L  V  1534

GGAGCCGTCCAGCCGGTACCAGTCAGCGTCTACGATTACTACCAACCAGACGAGGCCTGTACGACATTTTACCATCCCGGCCAGGGCAGCCCTCTGTTGGCCACACTGTGTGACGGCAGT  4800
G  A  V  Q  P  V  P  V  S  V  Y  D  Y  Y  Q  P  D  E  A  C  T  T  F  Y  H  P  G  Q  G  S  P  L  L  A  T  L  C  D  G  S  1574

CAGTGTGTCTGTGCTGAAGGGAAGTGTCCCCAAGGAGACCTCCCCTAAGAAACTGAGGGAGCTGACTCAACGAGACCGTGGAGGAGAAGCTTGCAACGACCTTGACTATGCTTTCAGTTC  4920
Q  C  V  C  A  E  G  K  C  P  K  E  T  S  P  K  K  L  R  E  L  T  Q  R  D  L  E  E  K  A  C  N  D  L  D  Y  A  F  R  V  1614

ATTGTGATGGAAGTGCGACAGGAGGGCAGTTTTGACCGGATCAGCATGATTGTGGAGAGCCCCATTAAGAAAGGACTTGACGATGTGTTTGATGGAGAGGAGCGTGTGTTTTGGAAACGA  5040
I  V  M  E  V  R  Q  E  G  S  F  D  R  I  S  M  I  V  E  S  P  I  K  K  G  L  D  D  V  F  D  G  E  E  R  V  F  W  K  R  1654

AAGACTTGCTCCGGCCTTCAGCTCGTAGAGGGGACGACCTACCTACTGATGGGGAAGGATGGGACCAAGTACACTGACGAGCAGGGCTTTGACAGCTTCAGGTATGTGATCACCGAGCAG  5160
K  T  C  S  G  L  Q  L  V  E  G  T  T  Y  L  L  M  G  K  D  G  T  K  Y  T  D  E  Q  G  F  D  S  F  R  Y  V  I  T  E  Q  1694

TCGTTTGTGGCCGAGTGGCCGACTGGGGCCAAGGCGGAGAGGAAGAAGTTCAAGCGAGTGGCCGACAAGTTCCAGCAACTGGCAGACAAAATCTTCCTGCAGGGCTGCACCACGTAGACC  5280
S  F  V  A  E  W  P  T  G  A  K  A  E  R  K  K  F  K  R  V  A  D  K  F  Q  Q  L  A  D  K  I  F  L  Q  G  C  T  T  *       1792

GACATCAGTTTGTTTCCCAGGTTGATTGCAGGGGCATTCGATTTTTGGACGTTCGTGTCCTGAATTCTCTTTGCTGCAGGTTGATCAACCTTTACTCATACTGTAACTTGATTAATTCTT  5400
GCAGTAGTTTTATTTTGTGGAAGGAGGAAAAATAAGTGTTTCGCATGCAGTGTTTTTTAAATTCACTGTTGAAACAATTCCGTAGTACAGTAGTAGTACATTGGAGCCATTTTCATCATA  5520
AATTTGCCCCAAAACCAAAAAAACTTAACCACGGCAAAAATGTGAAGATTTACAGTAGGCTATATCACCAACCTTCCTTTTTTATTCTTCAATTGAAGATTTTCATGTATATTGTTTTCT  5640
GACAGAAATGTTCGTTGACATGTGTCAATGATCAGTTGTATTTCTCGCTTTTTAGCTTCAACTATGTACTACATGTGTACATGTATGCTCATGTATTTTTGAGAGAGTGCCGAGTGAATG  5760
ATGTCAAGAACAGAAATTATTTTTTACCCAAAATGTGCAATTTTAGTAGCATCAAGGAGGTTTCAAATTCAGCCTCTTTGGTGGGTTGAAGTAACAGAACCATTGTGCAATGTTGAAACA  5880
TGCAGCTTTTTTTACAGTTGGCTGCAAAATTTTGCAGCAATGAAAATGTATTGAGAGTAGATAGTGTTGAACATAGTAGTTGTCCAGGACAGTTGGCAACAGACCATCTTTGATATTGAGG  6000
TCCTTGTATCTCACTTCAGAAGTTCAGAGAAAAGATGTTCTTGATAGCTTAACAGATGTCAGAATTTCATACTTACCCCGGTATTCTGTATAACAAACTGCACTTTGAACTCCCTCTGGC  6120
TAGCATTTGAGACCTAACTTTGAAATTCAAAATTTTATCGTGGAAATCTATGTTTTAGCATGCGCAGTTAGTTGCAGACTTGGTGTAGGGATAAACTGGAATTAAGGGTACAGCATTGAC  6240
TGGCCAATAACAAGTGCTCTTAACATTCAAGTTAAAAAGACACAATCATGCATAATTGATTCTTTGAAGGAAGTTGAATTGCCAAGTCTTTGTTGTCAATTTTTTATGTCTATGCGTTCA  6360
ATAAATACACTTTGGAC                                                                                                        6377
```

AmphiC3 and AmphiC6 are expressed mainly in the epipharyngeal groove, which is located just below the notochord (data not shown). Since no significant signal for AmphiC3 and AmphiC6 was detected in the notochord, the original EST clones probably originated from contamination with mRNA from the epipharyngeal groove. Although the function of the epipharyngeal groove is still not clear, the present results imply a secretory function for this organ. Again, this point is still to be tested at the protein level.

Acknowledgments. M.M.S. was supported by a Predoctoral Fellowship from the Japan Society for the Promotion of Science for Japanese Junior Scientists, with research grant 8-6806. This work was supported by Grants-in-Aid from the Ministry of Education, Science, Sports, and Culture, Japan, to N.S. The nucleotide sequences reported in this paper have been submitted to the DDBJ/GenBank/EBI Data Base under accession numbers AB050668 and AB050669.

References

Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, et al. (2000) The genome sequence of *Drosophila melanogaster.* Science 287: 2185–2195

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

Belt KT, Carroll MC, Porter RR (1984) The structural basis of the multiple forms of human complement component C4. Cell 36:907–914

Cameron CB, Garey JR, Swalla BJ (2000) Evolution of the chordate body plan: New insights from phylogenetic analyses of deuterostome phyla. Proc Natl Acad Sci USA 97:4469–4474

C. elegans Sequencing Consortium (1998) Genome sequence of the nematode *C. elegans:* A platform for investigating biology. Science 282:2012–2018

de Bruijn MH, Fey GH (1985) Human complement component C3: cDNA coding sequence and derived primary structure. Proc Natl Acad Sci USA 82:708–712

DiScipio RG, Hugh TE (1989) The molecular architecture of human complement component C6. J Biol Chem 264:16197–16206

DiScipio RG, Chakravarti DN, Muller-Eberhard HJ, Fey GH (1988) The structure of human complement component C7 and the C5b-7 complex. J Biol Chem 263:549–560

Dodds AW, Ren XD, Willis AC, Law SK (1996) The reaction mechanism of the internal thioester in the human complement component C4. Nature 379:177–179

Haefliger JA, Tschopp J, Nardelli D, Wahli W, Kocher HP, Tosi M, Stanley KK (1987) Complementary DNA cloning of complement C8 beta and its sequence homology to C9. Biochemistry 26:3551–3556

Haefliger JA, Tschopp J, Vial N, Jenne DE (1989) Complete primary structure and functional characterization of the sixth component of the human complement system. Identification of the C5b-binding domain in complement C6. J Biol Chem 264:18041–18051

Haviland DL, Haviland JC, Fleischer DT, Hunt A, Wetsel RA (1991) Complete cDNA sequence of human complement pro-C5. Evidence of truncated transcripts derived from a single copy gene. J Immunol 146:362–368

Hoffmann JA, Kafatos FC, Janeway CA, Ezekowitz RA (1999) Phylogenetic perspectives in innate immunity. Science 284:1313–1318

Howard OM, Rao AG, Sodetz JM (1987) Complementary DNA and derived amino acid sequence of the beta subunit of human complement protein C8: Identification of a close structural and ancestral relationship to the alpha subunit and C9. Biochemistry 26:3565–3570

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

Kan CC, Solomon E, Belt KT, Chain AC, Hiorns LR, Fey G (1985) Nucleotide sequence of cDNA encoding human alpha 2-macroglobulin and assignment of the chromosomal locus. Proc Natl Acad Sci USA 82:2282–2286

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

Katagiri T, Hirono I, Aoki T (1999) Molecular analysis of complement component C8beta and C9 cDNAs of Japanese flounder, *Paralichthys olivaceus.* Immunogenetics 50:43–48

Lagueux M, Perrodou E, Levashina EA, Capovilla M, Hoffmann JA (2000) 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

Levashina EA, Moita LF, Blandin S, Vriend G, Lagueux M, Kafatos FC (2001) 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

Lichtenheld MG, Olsen KJ, Lu P, Lowrey DM, Hameed A, Hengartner H, Podack ER (1988) Structure and function of human perforin. Nature 335:448–451

Nonaka M (2001) Evolution of the complement system. Curr Opin Immunol 13:69–73

Nonaka M, Takahashi M (1992) 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

Nonaka M, Takahashi M, Sasaki M (1994) Molecular cloning of a lamprey homologue of the mammalian MHC class III gene, complement factor B. J Immunol 152:2263–2269

Nonaka M, Yamaguchi N, Natsuume-Sakai S, Takahashi M (1981) The complement system of rainbow trout (*Salmo gairdneri*). I. Identification of the serum lytic system homologous to mammalian complement. J Immunol 126:1489–1494

Nonaka M, Fujii T, Kaidoh T, Natsuume-Sakai S, Yamaguchi N, Takahashi M (1984) Purification of a lamprey complement protein homologous to the third component of the mammalian complement system. J Immunol 133:3242–3249

Nonaka M, Azumi K, Ji X, Namikawa-Yamada C, Sasaki M, Saiga H, Dodds AW, Sekine H, Homma MK, Matsushita M, Endo Y, Fujita T (1999) Opsonic complement component C3 in the solitary ascidian, *Halocynthia roretzi.* J Immunol 162:387–391

Peitsch MC, Amiguet P, Guy R, Brunner J, Maizel JV Jr., Tschopp J (1990) Localization and molecular modelling of the membrane-inserted domain of the ninth component of human complement and perforin. Mol Immunol 27:589–602

Plumb ME, Sodetz JM (1998) Proteins of the membrane attack complex. In: Volanakis JE, Frank MM (eds) The human complement system in health and disease. Marcel Dekker, New York, pp 119–148

Rao AG, Howard OM, Ng SC, Whitehead AS, Colten HR, Sodetz JM (1987) Complementary DNA and derived amino acid sequence of the alpha subunit of human complement protein C8: Evidence for the existence of a separate alpha subunit messenger RNA. Biochemistry 26:3556–3564

Saitou N, Nei M (1987) The neighbor joining method: A new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Shinkai Y, Takio K, Okumura K (1988) Homology of perforin to the ninth component of complement (C9). Nature 334:525–527

Smith LC, Shih CS, Dachenhausen SG (1998) Coelomocytes express SpBf, a homologue of factor B, the second component in the sea urchin complement system. J Immunol 161:6784–6793

Sottrup-Jensen L, Stepanik TM, Kristensen T, et al. (1985) Common evolutionary origin of alpha 2-macroglobulin and complement components C3 and C4. Proc Natl Acad Sci USA 82:9–13

Stanley KK, Kocher HP, Luzio JP, Jackson P, Tschopp J (1985) The sequence and topology of human complement component C9. EMBO J 4:375–382

Suzuki MM, Satoh N (2000) Genes expressed in the amphioxus notochord revealed by EST analysis. Dev Biol 224:168–177

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Uptain SM, Kane CM, Chamberlin MJ (1997) Basic mechanisms of transcript elongation and its regulation. Annu Rev Biochem 66: 117–172

Venter JC, Adams MD, Myers EW, et al. (2001) The sequence of the human genome. Science 291:1304–1351

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

Wada H, Satoh N (1994) Details of the evolutionary history from invertebrates to vertebrates, as deduced from the sequences of 18S rDNA. Proc Natl Acad Sci USA 91:1801–1804

Yeo GS, Elgar G, Sandford R, Brenner S (1997) Cloning and sequencing of complement component C9 and its linkage to DOC-2 in the pufferfish *Fugu rubripes.* Gene 200:203–211