

Larry J. Dishaw · Sylvia L. Smith · Charles H. Bigger

Characterization of a C3-like cDNA in a coral: phylogenetic implications

Received: 16 December 2004 / Accepted: 19 May 2005 / Published online: 22 July 2005
© Springer-Verlag 2005

Abstract C3, C4, and C5 are thiolester-containing proteins (TEPs) of vertebrate complement. The identification of the molecular origin of the TEP family, and more specifically the ancestor protein of complement components C3, C4, and C5, remains a fundamental question. The prevailing paradigm suggests that duplication and divergence of these proteins occurred after the deuterostome split in phylogeny. It is believed that the ancestor of thiolester-containing complement proteins was alpha-2-macroglobulin (A2M)-like, a noncomplement-related protein. Here we describe a C3-like cDNA from a gorgonian coral, *Swiftia exserta*. This study provides evidence for the origins of a complement-related C3-like gene in the Precambrian period, predating both protostomes and deuterostomes. Furthermore, one may speculate that complement-like opsonic reactions were evolving at the earliest stages of metazoan evolution. This calls for a reassessment of the present concepts concerning the origins and evolution of TEPs.

Keywords Coral · Thiolester proteins · TEP · A2M · Complement evolution

Introduction

The complement system is a major component of vertebrate innate immunity. It opsonizes foreign cells for elimination by phagocytosis, efficiently lyses target cells,

facilitates removal of antigen/antibody complexes from circulation, and can be activated via specific and nonspecific pathways (reviewed in Carroll 1998; Sahu and Lambris 2001). The system is composed of four pathways: three activation pathways, each involving the C3 protein, that lead into the fourth terminal lytic pathway, which results in a membrane attack complex (MAC). Vertebrate complement components, and their activation products, modulate both innate and acquired immunity (Barrington et al. 2001; Dempsey et al. 1996; Fearon and Locksley 1996; Fujita 2002). Components C3, C4, and C5 belong to the so-called alpha-2-macroglobulin (A2M) family of paralogous genes (Fig. 1a), which currently are believed to have diverged from a single A2M-like ancestor (Armstrong and Quigley 1999; Dodds and Law 1998; Sottrup-Jensen et al. 1985) after the protostome–deuterostome (P–D) split (see Fig. 1b), suggesting that the C3, C4, and C5 genes are exclusive to the deuterostome lineage. This belief is countered by our report of the full-length cloning of a C3-like cDNA (SeC3, accession no. AY186744) from the endosymbiont-free soft coral, *Swiftia exserta* (Cnidaria; Anthozoa; Gorgonacea).

Cnidarians are diploblastic (two germ layers) organisms that are neither protostome nor deuterostome. This phylum diverged prior to the phylogenetic split creating the two descendent lineages (Fig. 1b), which include most extant organisms today (Adoutte et al. 2000). Fossil and molecular evidence suggests that Cnidarians may have existed as early as 700 mya (Adoutte et al. 2000; Ayala et al. 1998; McMenamin and McMenamin 1990), with recent molecular-clock estimates placing the P–D divergence at 670 mya (Doolittle et al. 1996). Our finding indicates that the ancestor to modern complement components C3, C4, and C5 originated sometime within the Precambrian. Furthermore, this finding presents a case for reevaluating present concepts of thiolester-containing protein (TEP) evolution and suggests a much earlier establishment of complement-related function.

In vertebrates, a primary function of the thiolester-containing complement proteins, C3 and C4, is the opsonization of microorganisms or immune complexes for

L. J. Dishaw · S. L. Smith (✉) · C. H. Bigger
Department of Biological Sciences,
Florida International University,
University Park Campus, 11200 SW 8th Street,
Miami, FL, 33199, USA
e-mail: smiths@fiu.edu
Tel.: +1-305-3483183
Fax: +1-305-3481083

L. J. Dishaw
Department of Molecular Genetics,
All Children's Hospital and Children's Research Institute,
University of South Florida,
St. Petersburg, FL, 33701, USA

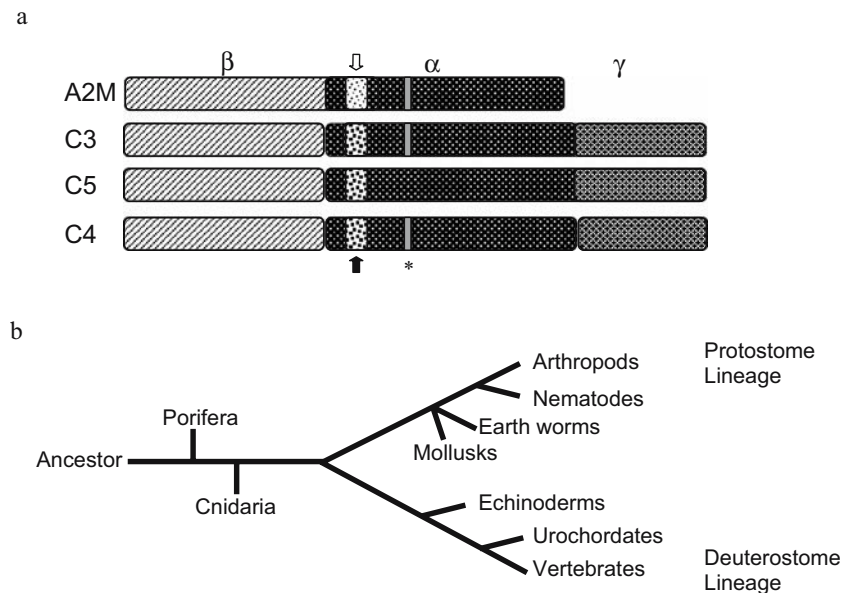


Fig. 1 Simplified models of TEP structure and animal phylogeny. **a** Schematic representation of primary structural relationships between human TEP proteins. Similar shadings correspond to homologous regions. Space between horizontal bars indicates post-translational cleavage site. A2M is a polymer of single-chain polypeptides; C3 and C5 are two-chain proteins, and C4 is a three-chain protein (see text, and note that lamprey C3 and cobra venom factor are also three-chain molecules). The three chains are labeled above as beta, alpha, and gamma. The black arrow indicates

the anaphylatoxin peptide region, whereas the open arrow indicates the corresponding position of the A2M-specific “bait” region. The solid bar (*) represents the homologous thiolester site, lost in C5 due to a substitution event early in its evolution. **b** Simplified schematic representation of animal evolution depicting the two major lineages of evolution, the protostomes and deuterostomes. Branch lengths are not to scale, and only a few representative taxa are displayed for simplicity

clearance by complement-receptor (CR) bearing phagocytes (Dempsey et al. 1996; Law and Levine 1977). Opsonic activity has been detected in invertebrate thiolester-containing A2M-like proteins, suggesting that this may have been an important function of the ancestral protein(s) (Levashina et al. 2001; Smith et al. 1999). Binding occurs primarily through intermolecular covalent interactions involving the thiolester (TE) site of the activated protein and its target (Gadjeva et al. 1998; Law and Levine 1977; Law et al. 1979; Law and Dodds 1997). C5, a related complement protein, is an exception since it has lost the TE site. The finding of C3-like genes (and not C4 or C5) in deuterostome invertebrates (Smith et al. 1999) and jawless fish (agnathans, Nonaka 1994) has suggested that the precursor to vertebrate C3, C4, and C5 was a C3-like gene (Nonaka et al. 1999).

Various types of divergent A2M-like TEPs, whose structural organization and function do not resemble C3, C4, or C5, have been shown to exist in the protostome lineage (i.e., nematodes and arthropods) (Levashina et al. 2003). A2M-like TEPs are paralogs of C3, C4, and C5, which, in addition to other properties, are single polypeptide proteins that are shorter by approximately 200 amino acids (aa) at the C-terminal end. They contain a polymorphic bait region in place of the anaphylatoxin region of the related complement proteins, C3, C4, and C5, and functionally are mostly nonspecific protease inhibitors (Armstrong and Quigley 1999; Sottrup-Jensen et al. 1985). A paralogous copy of A2M diverged early on, and a derived member found in some protostome invertebrates has been shown to

serve an opsonic function (Levashina et al. 2001). Because of their apparently distinct functional differences, the multiple paralogous types found are commonly referred to as “invertebrate TEPs.”

The lack of C3, C4, or C5 genes in protostomes had previously been concluded from extensive screening of the *Drosophila* and *Caenorhabditis elegans* genome databases (Dishaw et al., personal observations and unpublished data). This conclusion was further supported by data generated from the recent completion of the genomic sequence of *Anopheles gambiae* (Christophides et al. 2002; Holt et al. 2002). This, along with recent data from deuterostome invertebrates, sea urchin (Smith et al. 1999), tunicate (Marino et al. 1999; Nonaka et al. 2002), and lancelet (Suzuki et al. 2002), supports the prevailing paradigm that duplication and expansion of the TEPs, including complement components, occurred only after the deuterostome divergence in phylogeny (Dodds and Law 1998; Levashina et al. 2003; Sottrup-Jensen et al. 1985) (see Fig. 1b). This presumption implies that a TEP present in an organism that predates the P–D split is A2M-like in primary structure rather than C3-, C4-, or C5-like. However, this argument is based on data generated primarily from only three types of protostomes, two of which share subphyletic positions. We suggest that any conclusion based on such a small sample is premature; and indeed, the recent characterization of a C3-like protein from a protostome, the horseshoe crab (Zhu et al. 2005), as well as the full-length C3-like cDNA from a coral, described here, provides further evidence that this premise should now be reconsidered.

Materials and methods

Collection and maintenance of animals

S. exserta (Phylum Cnidaria, Class Anthozoa) was collected off the coast of southeast Florida in approximately 20–30 m of water. The live animals were transferred to FIU where they were maintained in seawater aquaria (35–37 °C; 21–23°C) with alternating light–dark cycles (14 and 10 h, respectively). The animals were fed with freshly hatched brine shrimp (*Artemia* sp.) larvae at regular intervals.

Control reactions for external contamination

Most PCR products were confirmed and recloned using *Swiftia* that were starved (not fed with brine shrimp) for 7–10 days while housed in nongravel and filtered artificial seawater-containing aquaria. In addition, all PCR primers were tested against brine shrimp cDNAs and total genomic DNA and against nucleic acids extracted from seawater. All PCR reactions against brine shrimp and seawater-derived nucleic acids did not yield amplified product. Nucleic acids were isolated from unfiltered aquarium seawater by high-speed centrifugation of 150 ml of water, followed by RNA isolation as described below using TriReagent.

Isolation of RNA

Total RNA was isolated from whole-body tissue preparations (coral coenchyme/colonial tissue) using TriReagent (Molecular Research Center, Cincinnati, OH, USA) with high salt precipitation as suggested by the manufacturer. Traces of genomic DNA were removed from the RNA using DNase I (Promega, Madison, WI, USA) treatment.

cDNA synthesis and degenerate PCR

cDNA synthesis was performed with Superscript II or Thermoscript (in 5'-RACE reactions) reverse transcriptases (RT) (Invitrogen, Carlsbad, CA, USA). For degenerate PCR, cDNAs were created in a degenerate primed RT reaction using 5–10 µg of total RNA in a 20-µl reaction with 400 µM of dNTP and Superscript II enzyme. The RNA was initially melted in the presence of 250 pmol of degenerate antisense primer (see below) at 80°C for 3 min and quenched in an ice-water bath for 2 min before the addition of the RT reaction mix. The RT reaction was incubated for 1 h at 42°C. Five microliters of the RT reaction was used as template along with 250 pmol of each degenerate primer (AS-5'-ACRTANGCNGTNAGCCANGT and S-5'-GNTGYGGNGARCARAAYATG) in a 50-µl degenerate PCR reaction as follows: 95°C for 5 min and 45 cycles of 1 min at 95°C, 1 min at 42°C, and 1 min at 72°C, followed by a 10-min final extension at 72°C. Comple-

mentary DNAs for 3'- and 5'-RACE were synthesized according to classic reverse transcription—RACE-PCR (Zhang and Frohman 1997). Taq polymerase and associated reagents were purchased from Qiagen (Valencia, CA, USA).

RACE-PCR and cloning of products

Rapid amplification of cDNA ends (RACE) was performed according to the classic RACE procedures (Zhang and Frohman 1997). For 5'-RACE, Thermoscript RT-polymerase (Invitrogen) with gene-specific antisense primers was utilized to prime the cDNA synthesis reaction. To facilitate the PCR amplification of some of the more difficult regions of the gene, 1–2% DMSO was used. All RACE products overlapped each other by at least 100 bp, and all were confirmed with nested PCR reactions. The products were gel purified (Qiagen gel extraction kit) and cloned into TOPO-TA cloning vectors (Invitrogen). The sequences of all gene-specific and RACE primers used in this study can be obtained from the authors upon request.

Northern and Southern blot analysis

Total RNA was extracted and separated on a 1% formaldehyde gel and transferred to a positively charged nylon membrane (Hybond XL, Amersham Biosciences). Probes were generated as riboprobes in runoff transcription reactions (with ³²P α-ATP) directly from the TOPO vectors using T7/SP6 polymerases (Roche Biochemical). To verify gene expression, Northern hybridization using riboprobes followed previously described methods (Krumlauf 1996).

Five micrograms of coral genomic DNA from a single animal was digested with *PvuI*, *KpnI*, and *SalI* (Promega) for 24 h. The DNA was run on a 0.7% TAE-agarose gel and transferred to a nylon membrane (Hybond XL) under alkaline conditions (Sambrook and Russell 2001). DNA probes were generated using RACE-PCR products corresponding to the gamma chain region of SeC3. Random priming reactions were performed with the Mega Prime Labeling kit (Amersham Biosciences) using ³²P α-dCTP. To estimate size and complexity of the gene, Southern blotting was performed using high stringency phosphate buffers (Sambrook and Russell 2001) at 60–65°C overnight. At this time, we do not have data on the existence of individual or population-based polymorphism for SeC3.

Assembly and analysis of cloned sequences

The full-length sequence of SeC3 was initially derived by assembling overlapping RACE clones. Sequences were aligned in Clustal X (Thompson et al. 1997), manipulated using Sequence Manipulation Suite (Stothard 2000), BioEdit (Hall 1999), and GeneDoc (v2.5) (Nicholas and Nicholas 1997) and assembled by eye using Microsoft Word. All RACE and other PCR product sequences were

confirmed by sequencing at least 10 randomly selected clones.

Multiple protein sequence alignments were performed in Clustal X using available, deduced TEP sequence data. Pairwise comparisons were produced in calculating distance scores, percent identity, and percent similarity using Mega2 (ver. 2.0) (Kumar et al. 2001), GeneDoc, and Sequence Manipulation Suite.

Amplification and cloning of full-length SeC3

The full-length cDNA for SeC3 was then amplified using primers designed to the 5' and 3' UTR of the assembled RACE-generated sequence. Using the following primers: SeC3-5'UTR-5' CAACTTCCGCACTCTGTGAA and SeC3-3'UTR-5' CTCGTGGTAACCAAGACAGA, the full-length cDNA was amplified with the proof-reading enzyme, Takara LA-Taq Polymerase (Fisher Scientific, USA), and the following amplification conditions: 95°C for 5 min and 30 cycles of 95°C for 1 min, 55°C for 1 min, and 68°C for 6 min, and terminated with a 15-min extension at 72°C. The PCR product was cloned into a TA-cloning vector using the TOPO kit (Invitrogen).

Screening of databases

Sequences used in this study were downloaded from the following databases and resources—NCBI-Genbank, using BLAST, as Blastx, Blastn, and PHI-BLAST: <http://www.ncbi.nlm.nih.gov/BLAST>; Drosophila Genome Project: <http://www.fruitfly.org/>; Flybase: <http://www.flybase.bio.indiana.edu/>; Sanger Center Project: <http://www.sanger.ac.uk/Projects/Celegans/>; and Washington University Genome Project: <http://www.genome.wustl.edu/>.

Phylogenetic analysis of SeC3

All TEP sequences were downloaded from the Genbank database. Phylogenetic analysis was performed using multiple sequence alignments of full-length protein sequences ($N=52$) and the minimum evolution distance method (Kumar 1996; Rzhetsky and Nei 1993) using both the Mega2 program (Kumar et al. 2001) and PAUP* beta version 10 (Swofford 1998). Phylogenetic trees were displayed in the TreeView program (Page 2001).

Multiple sequence alignments (MSA) were constructed using the Gonnet matrix (Gonnet et al. 1992) and gap open and extension penalties of 20 (initial pairwise parameters) and, for the subsequent multiple alignment parameters, gap open and extension penalties of 20 and 0.40, respectively.

Using the PAUP* package set to distance criteria, phylogenetic trees were generated to determine the relationship of SeC3, complement proteins, and other TEPs and to predict the evolutionary history of complement-related TEPs. This was done by bootstrapping (Felsenstein 1985) the heuristic search (100 replicates), with optimality criteria

set to *minimum evolution* (ME). The following criteria were implemented in the bootstrap search using the branch swapping method of *tree-bisection-reconnection* (TBR): the starting tree was determined by the stepwise addition method (not estimated by the neighbor joining method); distance measures were calculated by *mean character difference* and the branch lengths constrained to *nonnegative* numbers; *random trees* were generated as the starting point with the *steepest decent* option in effect; each bootstrap replicate reflected multiple resampling of the data where the *starting tree* in each round was estimated by *stepwise addition* with *random addition* of sequences (this was done three times per replicate). The bootstrap consensus tree was estimated by the 50% majority rule option.

Although the PAUP* ME analysis allows for more extensive analyses and manipulation of the search options, the Mega2 package (ver. 2.1) allows for correction of multiple substitutions (using Poisson correction), allows for treatment of MSA gaps in a pairwise deletion fashion, and can display branch lengths reflecting pairwise distances. These criteria are sometimes important in MSAs of very ancient and large sequence families and when comparing long sequences which also tend to cluster in groups of diverging sequences (usually obvious orthologs). Using the Mega2 program, the phylogeny was estimated with the bootstrap test (10,000 replicates) and the minimum evolution (with neighbor joining to estimate the starting tree) algorithm. Poisson correction was implemented to account for multiple amino acid substitutions over time. Branch swapping was done with close neighbor interchange (CNI), and gaps were removed/ignored in a pairwise manner.

Results

Cloning of SeC3 cDNA using RT-PCR

We have cloned the full-length cDNA for a C3-like gene from a coral. Corals are Cnidarians, which are acelomate diploblastic metazoans mostly displaying radial symmetry and whose ancestors diverged prior to the split creating the protostome and deuterostome lineages. Using degenerate RT-PCR, a partial cDNA product (Fig. 2a) was initially cloned and sequenced. Initial database searches, followed by protein sequence alignment and phylogenetic analysis, suggested that this cDNA encoded a protein that belonged to the TEP family. Classic RACE-PCR was utilized to clone overlapping PCR products that assembled into a 5.5 kb cDNA with a deduced amino acid sequence of 1,728 aa in one open reading frame (Genbank accession no. AY186744). Subsequently, primers were designed to the 5' and 3' UTR of the sequence, and the full-length cDNA was amplified using a high fidelity system.

Northern and Southern blotting

Northern blot analysis suggests expression of SeC3 at low concentrations in normal, healthy, and unchallenged tissue

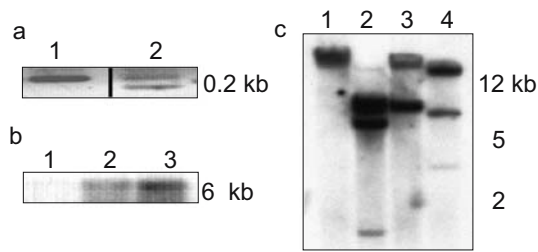


Fig. 2 PCR product cloning and Northern and Southern blotting data. **a** Degenerate PCR results for TEP-like sequences in *Swiftia*, right lane. Left lane is a rat positive control using the same degenerate primers. The top band, right lane, was excised, gel purified, and cloned. Sequence analysis indicated it to be very similar to cDNAs of A2M-like proteins. **b** Northern blotting suggested that the transcribed gene was about 6 kb. Northern blotting also indicated that the gene was constitutively expressed at low levels, since lane 3 consists of about 30 μ g of total RNA (lanes 1 and 2 are 5 and 20 μ g, respectively). **c** Southern blotting confirms presence of SeC3 in coral genomic DNA. Five micrograms of genomic DNA from a single animal was digested with *PvuI*, *KpnI*, and *SalI* (lanes 2, 3, and 4, respectively). Lane 1 is undigested genomic DNA

(Fig. 2b). High-stringency Southern blotting (Fig. 2c) confirms the presence of the gene in the coral genome. Based on known restriction sites from the PCR products used as probes, the expected banding pattern was attained.

In a few instances (with different enzymes), extra bands were identifiable which may suggest a complex genomic organization (similar to vertebrate complement genes) (Morley and Walport 2000; Vik et al. 1991) and/or the existence of a second paralogous TEP gene which may share sequence similarity.

Deduced amino acid sequence analysis using protein sequence alignments

Analysis of the full-length deduced amino acid sequence indicates that SeC3 shares 24 and 45% identity and similarity (allowing for conservative substitutions), respectively, with human C3 (HuC3). Similar values are shared with C4 and C5 (Table 1). Conservation of multiple, functionally critical sites are found in SeC3 (Fig. 3). These include the thiolester site (common to vertebrate C3 and C4), the C3- and C4-specific catalytic histidine (pos 1140 in SeC3; 1126 in HuC3), the β - α cleavage site (specific to C3, C4, and C5), a putative α - γ cleavage site (specific to mammalian C4 and lamprey C3), and the C3a peptide region (a protease-attracting site which releases the C3a anaphylatoxin upon cleavage) that is analogous to the “bait region” of A2M. In addition, the distinctive feature of the

Table 1 Pairwise comparisons of SeC3 vs other TEPs

Organism–gene	Accession no.	<i>I</i> (%) ^a	<i>S</i> (%) ^b	Distance \pm SE ^c	<i>p</i> -Distance (\pm 0.01)
PWM: C3 (<i>n</i> =16) ^d	–	24	44	1.300 \pm 0.041	0.731
Urchin C3	AF025526	26	45	1.270 \pm 0.040	0.719
Tunicate C3	AB006964	18	40	1.635 \pm 0.050	0.805
<i>Amphioxus</i> C3	AB050668	32	52	1.084 \pm 0.034	0.662
Lamprey C3	D10087	25	45	1.296 \pm 0.041	0.726
Carp C3-H1	AB016211	25	45	1.256 \pm 0.040	0.715
Chicken C3	I50711	25	45	1.288 \pm 0.041	0.724
Mouse C3	P01027	24	44	1.313 \pm 0.041	0.731
Human C3	NM_000064	24	43	1.322 \pm 0.041	0.733
PWM: C4 (<i>n</i> =5) ^d	–	23	44	1.350 \pm 0.042	0.760
<i>Xenopus</i> C4	D78003	24	45	1.361 \pm 0.042	0.744
Mouse C4	P01029	23	43	1.320 \pm 0.042	0.748
Human C4A	K02403	24	43	1.320 \pm 0.041	0.733
PWM: C5 (<i>n</i> =2)	–	22	44	1.404 \pm 0.044	0.754
Mouse C5	P06684	22	44	1.415 \pm 0.044	0.757
Human C5	M57729	22	44	1.392 \pm 0.043	0.751
PWM: A2M(<i>n</i> =22) ^d	–	20	39	1.376 \pm 0.046	0.780
<i>Drosophila</i> TEP1	AAF53490	17	34	1.379 \pm 0.050	0.748
<i>Limulus</i> A2M	D83196	21	39	1.365 \pm 0.045	0.745
Lamprey A2M	D13567	20	41	1.432 \pm 0.047	0.726
Mouse A2M	Q61838	21	40	1.373 \pm 0.045	0.747
Human A2M	NM_000014	21	39	1.380 \pm 0.045	0.741

PWM Pairwise mean

^aIdentities (*I*) are calculated as the percentage of identical amino acids per column/position in the alignments

^bSimilarity (*S*) was calculated as the percentage of identical plus similar residues, which are conservative substitutions (maintaining physicochemical properties)

^cPoisson-corrected distance scores (Mega2; Kumar et al. 2001) \pm standard error were calculated for all pairwise comparisons. Proportion of difference (*p*-distance) calculations, uncorrected for multiple substitutions, were performed by the analytical method (Mega2)

^dPairwise means were derived from averaging results from pairwise alignments of *n* number of sequences. Only a subsample of results are shown; table of all sequence calculations can be obtained from the authors

~200-aa C-terminal region of C3, C4, and C5 (i.e., corresponding to the γ -chain sequence of C4 and included in the extended length of the α -chain of C3 and C5) that is absent from A2M-related proteins (Fig. 1a) is found and conserved in the coral SeC3-deduced polypeptide sequence (Fig. 3).

In humans, target-bound C3b is cleaved into bound C3d, which is recognized by receptors on phagocytes and C3c that is released from the target. The major residues involved in assembling the helical structure of C3d are conserved in the corresponding region of SeC3, as confirmed by aligned comparison to the crystallized model of human C3d (Nagar et al. 1998). Comparative modeling (Guex and Peitsch 1997; Peitsch et al. 2000) and secondary structural prediction (Jones 1999; Karplus et al. 1998; McGuffin et al. 2000; McGuffin and Jones 2003; Rost 1996) suggest that the corresponding region of SeC3 may share a similar complex helical backbone to other TEP family members. Thus, the entire length of sequence for SeC3 shares conserved secondary structure patterns with other complement-related proteins (data not shown) rather than A2M-related proteins.

Structurally, mammalian C3 is a two-chain protein, consisting of α - and β -chains. The deduced chain structure of SeC3 predicts a three-chain molecule resembling mammalian C4 (Karp et al. 1981; Morley and Walport 2000), lamprey C3 (Nonaka 1994), and cobra venom factor (Vogel et al. 1996). SeC3 contains two putative cleavage locations, which would permit processing of the promolecule into a three-chain structure. The predicted (unglycosylated) sizes of the individual chains are 74, 86, and 32 kDa for β , α , and γ , respectively, which are conserved with human C4 (Karp et al. 1981; Morley and Walport 2000). In human C3, the α -chain is longer than that of C4 as it includes the homologous sequence of the γ -chain that remains because the α - γ cleavage site is absent (Figs. 1a, 3).

A novel structural aspect of SeC3 in the putative α - γ cleavage region (Fig. 3) is the presence of one to two cleavage sites in what is predicted (Liu and Rost 2003) to be a NORS region (regions that have no regular secondary structure; Liu et al. 2002). Cleavage at both sites would generate an unusual 74-aa NORS region peptide that is lysine- and arginine-rich. The presence of a cysteine residue within the second cleavage site may actually interfere with cleavage and release of the NORS peptide. Nonetheless, the highly hydrophilic K-R-rich region may represent a relic of an ancient processing event in the evolution of the cleavage site/region (R-x-x-R).

While most of the cysteine residues are conserved between SeC3 and mammalian C3, C4, and C5, those corresponding to positions 559 and 816 of HuC3 (Fig. 3) are a special exception. These two cysteines are responsible for the joining of the β to the α -chain in all known C3 proteins. Both cysteines are missing from SeC3, presenting a novel and interesting puzzle for the β - α -chain interaction. The two chains either associate in a different manner or the β -chain is released and is not an integral part of the processed or functional protein. The latter case seems unlikely since the coral β -chain region shares sequence con-

servation in the corresponding regions of C3, C4, and C5, whose β -chains coevolved within the structural constraints associated with their function. Hypothetically, though, the lack of the associated β -chain in the processed molecule would leave the anaphylatoxin C3a region highly exposed and susceptible to rapid protease cleavage and activation of SeC3.

C3-convertase cleavage of mammalian C3 releases the C3a anaphylatoxin peptide and causes a conformational change in the resulting C3b protein, bringing the catalytic histidine in direct contact with the thiolester site. The highly reactive C3b interacts in an immediate covalent manner with the target. The C3a peptides usually span 65–70 amino acids and contain six cysteine residues that are organized in a conserved fashion (Huber et al. 1980). This organization of the cysteines is well conserved in the coral (Fig. 3). The signature cleavage motif for vertebrate C3a is –LAR/S and also serves as a receptor-binding site for the released peptide (Huber et al. 1980; Muto et al. 1985; Sahu and Lambris 2001). A putative SeC3 cleavage site, –RTR/S, is found in the corresponding region (Fig. 3).

Phylogenetic analysis of SeC3 and related complement proteins

To investigate the evolutionary history of C3, C4, and C5 complement proteins and to determine how the coral sequence fits into this picture, 52 sequences were downloaded from the Genbank database and subjected to phylogenetic analysis. Although the TEP family consists of divergent, homologous sequences, the phylogenetic relationships are well resolved (Fig. 4). All methods of analyses used in this work produced almost identical tree topologies, and in all cases, SeC3 clusters with the deuterostome invertebrate C3-like proteins (sister group to the vertebrate C3, C4, and C5 components). The phylogeny indicates three major groups, all sharing a single node (ancestor).

A2M-like sequences make up two major groups (group A and B), diverging from one very early common ancestor (Fig. 4). In at least some protostomes (such as the fruit fly and mosquito), group A-type A2M-like genes have duplicated multiple times in a lineage-specific manner apparently to increase diversity of recognition (Levashina et al. 2003). The third group (group C) consists of the complement-related proteins, which are subdivided into the invertebrate C3/C4/C5-like genes (conventionally referred to

Fig. 3 Full-length sequence alignment of SeC3 and human C4A, C3, C5, and A2M. All major reactive sites, receptor binding sites, and cysteines are boxed or highlighted. Shading represents shared identity per column or sequence position. Where applicable, the different chains and the respective cleavage sites have been labeled as well. Labeling of sites is based on what is known from human C3 functional and biochemical studies (Sahu and Lambris 2001; Morley and Walport 2000). Note that SeC3 shares similarity in the anaphylatoxin region and contains a catalytic histidine, a conserved C-terminal end with C3, C4, and C5 (in contrast to A2M), and two putative cleavage sites that could generate a three-chain-type molecule. All major cysteines are labeled with an *, and additional non-conserved cysteines unique to SeC3 are labeled with ^

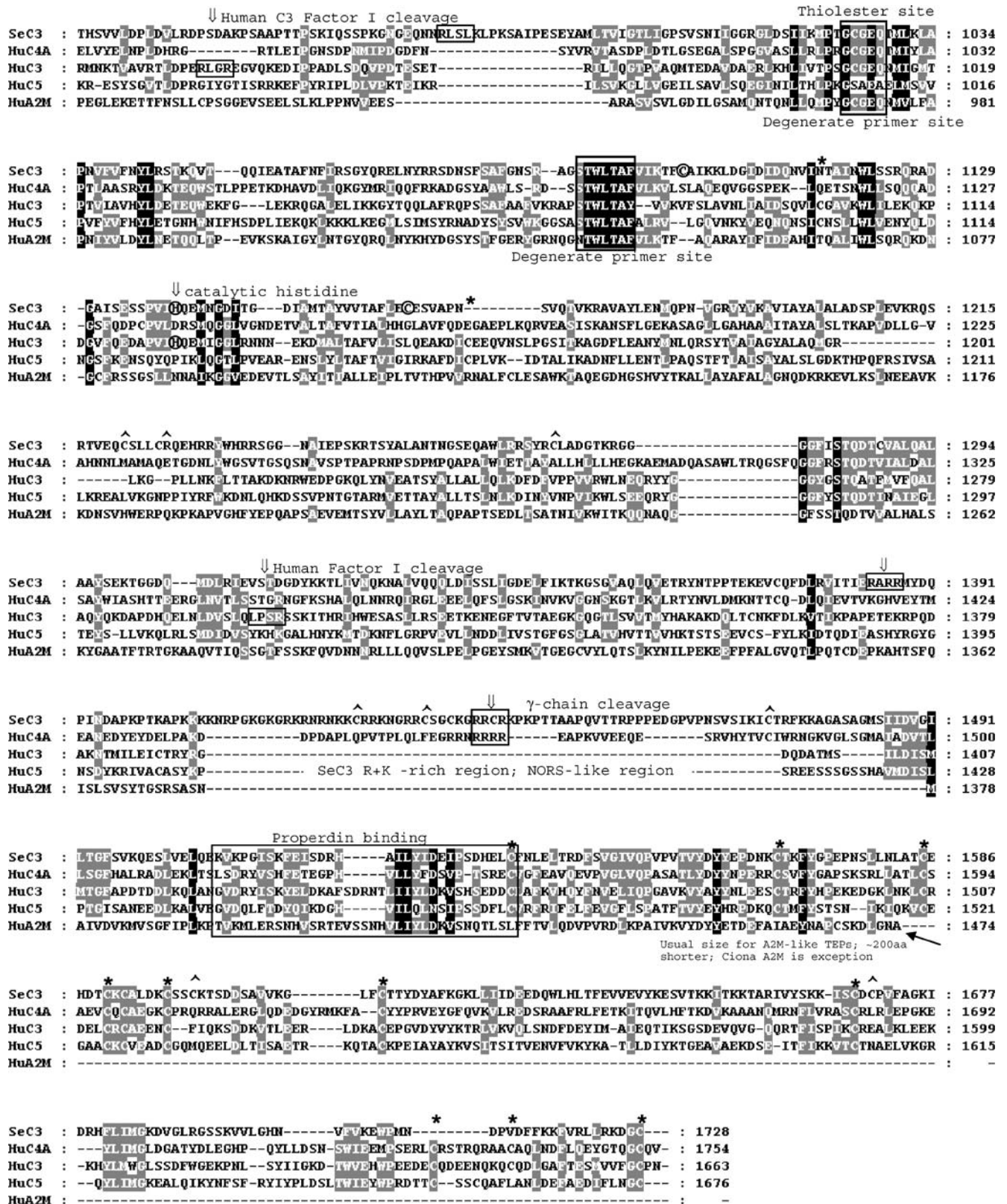


Fig. 3 (continued)

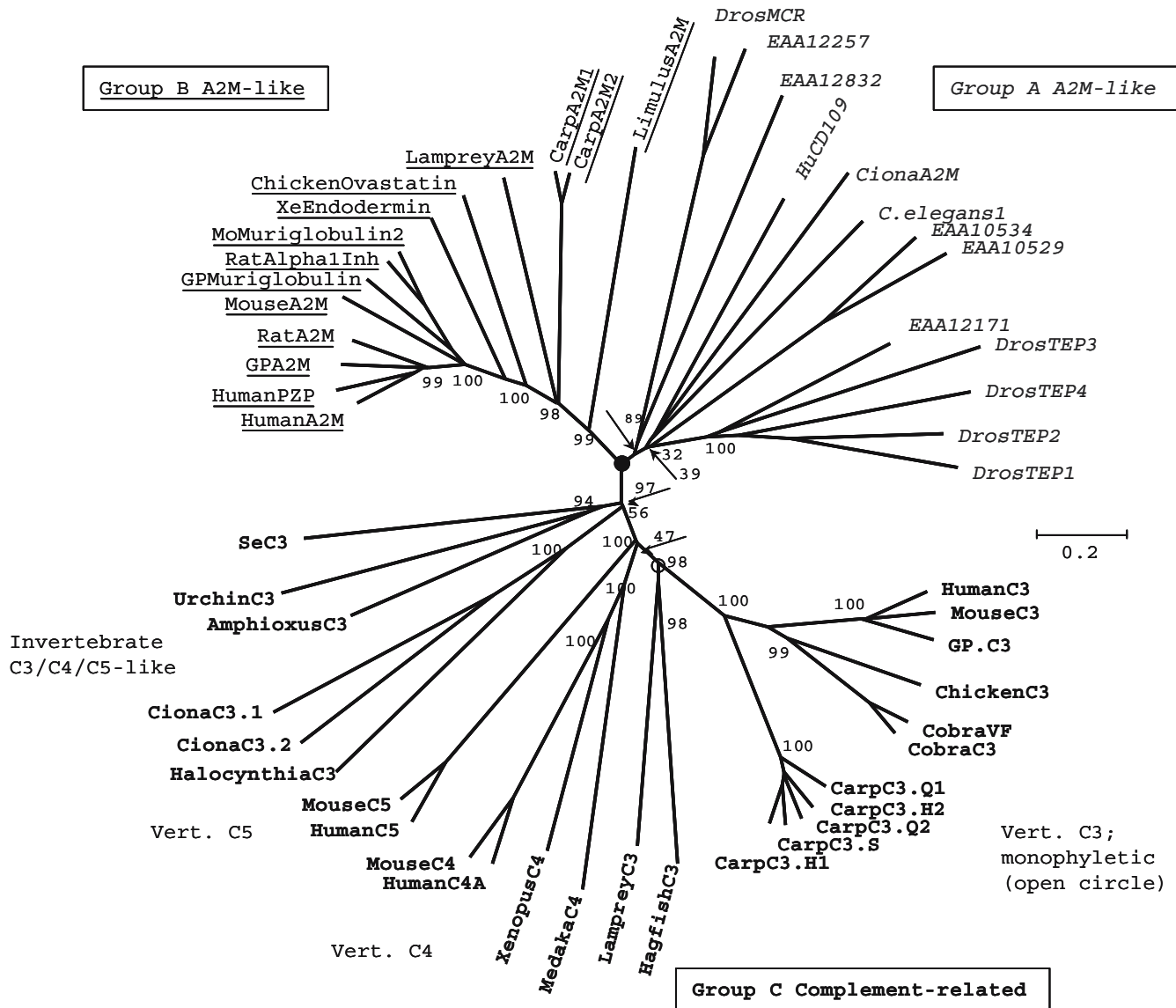


Fig. 4 Unrooted minimum evolution (ME) phylogenetic tree (bootstrapped with 10,000 replicates) of various members ($N=52$) of the TEP family. Branch lengths shown are equal to corrected distances (Poisson correction) between the sequences in that clade. The phylogenetic history of the TEP family can be grouped into the evolution and divergence of three groups of sequences sharing a monophyletic origin. The *closed circle* represents the predicted point of origins (ancestry) for the TEP family (see text). Group a sequences are in *italics*, group b are *underlined*, and group c are in *bold*. SeC3 forms an orthologous monophyletic group with urchin and amphioxus C3-like TEPs with 94% support. The invertebrate C3-like sequences represent paralogous extant forms of the C3/C4/C5-like ancestor and, as a group, are orthologous to vertebrate group c complement component TEPs, existing prior to the divergence of the vertebrate complement forms. Vertebrate C3 has monophyletic origins (*open circle*) beginning in jawless fish (lamprey and hagfish). The large branch distance to the other vertebrate C3s represents missing data or taxa (C3 sequence representatives) in the transition to jawed fish divergence. Groups a and b A2M-like sequences represent two different groups of protease inhibitors sharing deep monophyletic origins, at the beginning of the origins and divergence of the TEP family (see text). Accession numbers of the sequences used to construct this tree are as follows. Group a: mosquito and fruit fly have various paralogous forms of the group a sequences, 15 and 6, respectively, and only a few are used here since the others are species-specific expansions; *Drosophila* TEP 1–4 and DrosMCR, AAF53490, CAB87808, CAB87809, AAF53826, and NM_079949,

respectively; *Caenorhabditis elegans*, Z82090; *Anopheles gambiae* TEPs, EAA12171, EAA10529, EAA10534, EAA12832, and EAA12257; human CD-109, NM_133493; *Ciona intestinales* A2M, AJ431688; group b: horseshoe crab, *Limulus* sp. A2M, D83196; carp A2M 1&2, *Cyprinus carpio*, AB026128 and AB026129, respectively; lamprey A2M, *Lampreta japonica*, D13567; chicken ovastatin (A2M-like paralog), *Gallus gallus*, X78801; Xenopus endodermin (A2M-like paralog), *Xenopus laevis*, AAB51432; mouse murinoglobulin 2, *Mus musculus*, NM_008646; rat alpha1 inhibitor III, *Rattus norvegicus*, J03552; guinea pig (GP) murinoglobulin, *Cavia procella*, D84339; mouse A2M, Q61838; GP A2M, D84338; human pregnancy zone protein (PZP), NM_002864; and human A2M, NM_000014; group c: coral C3-like (SeC3), *Swiftia exserta*, AY186744; urchin C3-like, *Strongylocentrotus purpuratus*, AF025526; cephalochordate—amphioxus C3-like, *Branchiostoma belcheri*, AB050668; urochordate—tunicate/ascidian C3-like, *Ciona* C3-1 and C3-2, AJ320542 and AJ320543, respectively; tunicate C3-like, *Halocynthia roretzi*, AB006864; mouse C5, P06684; human C5, M57729; human C4A, K02403; mouse C4, P01029; Xenopus C4, D78003; fish—*Medaka* C4, *Oryzias latipes*, BAA92287; jawless fish—lamprey C3, *L. japonica*, D10087; hagfish C3, *Eptatretus burgeri*, Z11595; jawed fish—carp C3 (5 paralogous forms), Q1, Q2, H1, H2, and S, AB016214, AB016215, AB016211, AB016212, and AB016213, respectively; cobra C3 and venom factor, *Naja naja*, Q01833, U09969, respectively; chicken C3, I50711; guinea pig C3, P123887; mouse C3, P01027; and human C3, NM_000064

as C3-like) and vertebrate C3, C4, and C5 genes. The clade (s) representing the invertebrate C3-like genes shares a closer ancestor with the vertebrate complement components than it does with group A or B members.

Additionally, the groups A and B A2M-like sequences are separated by an ancient common ancestor (labeled node). While many references to these sequences in the literature incorrectly regard them all as A2M, the phylogenetic evidence (Fig. 4) clearly shows otherwise. A putative ancestral node is marked, and the data suggest that very early in phylogeny (before the separation of most metazoans), a common ancestral gene underwent two duplication events, creating three lineages that began to diverge and radiate into a complex superfamily of complement proteins (opsonins) and protease inhibitors.

Discussion

Studies pursuing the origins of the complement system, including complement-like functional activities, parallel many early investigations of phagocytosis in deuterostome invertebrates (i.e., echinoids) (Bertheussen 1979; Bertheussen and Sejelid 1978; Lachmann 1979) such that early models of complement phylogeny hypothesized invertebrate origins (Lachmann 1979). Later work demonstrated that phagocytosis of red blood cells could be enhanced by first reacting the RBCs with human C3 (opsonization) (Bertheussen 1982; Bertheussen and Seljelid 1982). These data provided evidence for the presence of a component related to a complement-like system in invertebrates, in this case, specifically the presence of C3-like receptors (which suggested the existence of a C3-like homologue) (Bertheussen 1983).

These very significant findings were initially limited to deuterostomes (e.g., echinoderms, urochordates, and vertebrates) and the evolution of complement in the deuterostome lineage. Early functional similarities had previously been demonstrated in protostomes (e.g., horseshoe crab, mosquitoes, worms), notably the horseshoe crab (Armstrong and Quigley 1999), which also hinted at a system involving some sort of opsonization and protease inhibition (characteristic of A2M-like TEPs). Although opsonization via a TEP-like protein(s) does occur in some protostomes (Levashina et al. 2001), it apparently did not involve a C3-like protein (in the mosquito model). Subsequent sequencing and genomic analyses (Ainscough et al. 1998; Christophides et al. 2002; Holt et al. 2002; Levashina et al. 2003; Saravanan et al. 2003; Valenzuela et al. 2002) further verified the absence of bona fide C3-like sequences in these species. Since the invertebrate TEPs, whether C3-like or A2M, possess and share some complex functions, the original functions of the ancient and evolving complement-related proteins are difficult to predict (Armstrong and Quigley 1999; Barrington et al. 2001; Dempsey et al. 1996; Fearon and Locksley 1996; Fujita 2002; Levashina et al. 2001, 2003; Mastellos and Lambris 2002; Saravanan et al. 2003). While complement-like opsonic activity has been demonstrated in some invertebrate species, from data cur-

rently available, it is premature to conclude that a complement “system” operates in invertebrates.

As a first step to determining the molecular origins of a primitive complement-like system, identifying and understanding the nature of the ancestral TEP are necessary in order that a more comprehensive view of the original functional purpose be obtained. Consequently, we pursued a C3-like sequence homologue in a phylum that predates the protostome and deuterostome (P–D) separation. Cnidaria is an extant taxa whose ancestor branched (see Fig. 1b) prior to the major evolutionary split creating two independently evolving lineages (P–D). The presence of diverse TEPs in both lineages suggested to us that a functionally significant C3/A2M-like ancestor existed very early on in the evolution of animals. Here, we have described the existence of a highly conserved C3-like (complement-like) cDNA sequence from a coral, providing the initial evidence for that significant complement-related ancestor. Since submission of this manuscript, additional work by an independent group has established, for the first time, the existence of a bona fide C3-like gene in a protostome, the horseshoe crab (see Zhu et al. 2005). This lends support to our viewpoint that the origin and evolution of complement-like and related proteins, and their possible functional interactions as a “system,” should be reconsidered.

C3, C4, C5, and A2M: proteins of the TEP family

The TEP family, which includes complement proteins C3, C4, and C5, is composed of multiple paralogous glycoproteins, which are essential for not only immune-related functions but also for homeostasis in general and are partly responsible for defining the evolution of innate and adaptive immunity. Early findings of functional and structural similarity, and later with partial sequence data, afforded supporting evidence for homology among the thiolester-containing (or related) proteins, such as C3, C4, C5 (in the latter, the TE site is missing), and alpha-2 macroglobulin (Bokischi et al. 1975; Campbell et al. 1988; Dodds and Law 1998; Law and Levine 1977; Sottrup-Jensen et al. 1985). As A2M was the first TEP to be found in an invertebrate (Armstrong and Quigley 1999), it was proposed to be the ancestor of complement-like TEPs (Dodds and Law 1998; Sottrup-Jensen et al. 1985). It is only recently, following accumulation of vast sequence data from multiple phyla, including the realization of the critical immune and developmental functions served by these related proteins (Barrington et al. 2001; Carroll 1998; Fearon and Locksley 1996; Mastellos and Lambris 2002), that it has become apparent that this evolutionary story is not so simple as first conceived.

While most invertebrates appear to have more than one gene encoding TEPs, some deuterostome invertebrate species have been shown to possess a single C3-like complement-related gene (Al-Sharif et al. 1998; Marino et al. 2002; Nonaka et al. 1999; Smith et al. 1999; Suzuki et al. 2002) that may encode for certain opsonic functions (Nonaka et al. 1999; Smith et al. 1999), in the absence of a

complete, vertebrate-type complement system. Until very recently (Zhu et al. 2005), protostome invertebrates did not appear to possess complement-like genes, but instead only possessed divergent A2M-like genes (Fig. 4; Christophides et al. 2002; Holt et al. 2002; Levashina et al. 2003, Dishaw et al., personal observations). Although the A2M-like molecules typically serve as universal protease inhibitors, at least in some invertebrates, the proteins have been shown to serve opsonic functions (Levashina et al. 2001, 2003). The conventional argument, therefore, has been that complement-like genes (and a complement system) are a deuterostome-exclusive characteristic; consequently, related thiolester-containing complement-like proteins existing prior to the P–D split in phylogeny must be A2M-like (Fig. 4, group A or B or both). Our data from the coral, and that obtained from the horseshoe crab (Zhu et al. 2005), suggest otherwise.

S. *exserta* expression of a C3-like cDNA

In this report, we describe the existence of a C3-like gene from an animal whose phylum origin predates the origins of protostomes and deuterostomes. Furthermore, here we show that the expressed coral cDNA sequence shares specific characteristics with mammalian C3, C4, and C5 complement proteins. This finding has significant phylogenetic implications, which are discussed below. Analysis of the deduced translation of the SeC3 cDNA sequence, along with comparison of the primary structure to related proteins (TEPs), suggests that the newly discovered coral cDNA is C3/C4/C5-like rather than A2M-like. Multiple sequence alignment shows a considerably high conservation of amino acids with vertebrate C3 and overall with mammalian C3, C4, and C5. The anaphylatoxin region is a unique attribute of C3, C4, and C5 and is not present in A2M, which contains the so-called bait region instead. The coral SeC3 sequence does not possess a bait region (like A2M), but instead has an anaphylatoxin-like sequence with all six cysteine residues arranged in their characteristic/conserved fashion. In addition, the presence of a putative α – γ cleavage site, which is characteristic of mammalian C4 and some vertebrate C3s, would give the protein a three-chain structure. In SeC3, two putative cleavage sites exist in the same corresponding region (the α – γ cleavage region) as in vertebrate C4, in addition to the cleavage site separating the α - and β -chains (see Fig. 3). Therefore, posttranslational cleavage at either (or both) of the sites within the α – γ region can produce a three-chain structure (i.e., β + α + γ and Fig. 1a).

Phylogenetic analysis and prediction reveals a highly resolved evolutionary history for the TEP family of proteins and implicates the newly characterized coral cDNA, SeC3, as a descendent of the ancestor to vertebrate C3, C4, and C5 (Fig. 4). This finding therefore suggests that the ancestor to thiolester-containing complement components predates the divergence of bilaterian animals. In addition, the last common ancestor (LCA) of complement components and A2M-like TEPs appears to have existed in a very

ancient time, most likely during the early diversification and radiation of all metazoans.

SeC3 and phylogenetic implications for complement evolution (rooting the TEP family)

Rooting the TEP family tree can provide evolutionary *direction*, which, in turn, illustrates routes taken by the ancestral complement-related thiolester protein (TEP) in response to variations of selective pressure. It is now known that innate immunity has been instrumental in shaping the evolutionary history of adaptive immunity, and complement components have been essential elements of that bridge (Barrington et al. 2001; Carroll 1998; Fearon and Locksley 1996; Fujita 2002; Mastellos and Lambris 2002; Sahu and Lambris 2001). The genes encoding A2M- and complement-like proteins can be grouped into three major groups which diverged from a common ancestor. These groups include A2M-like (groups A and B) and complement-like (group C) genes. Initially, based on genomic sequencing data obtained from nematodes (Ainscough et al. 1998), fruit flies (Adams et al. 2000), and mosquitoes (Holt et al. 2002; Levashina et al. 2003), it appeared that the group A A2M-like proteins were simply a divergent form of A2M that were unique to the protostome lineage of organisms. In fact, “conventional A2Ms” (and their vertebrate paralogous counterparts, i.e., ovastatin, muriglobulin, endodermin) appear to be a set of proteins that have diverged in the vertebrate lineage (Fig. 4). The discovery of human CD109 (Lin et al. 2002) and *Ciona* (Urochordate) A2M (Hammond et al. 2005) brought deuterostome counterparts into the group A proteins. This, therefore, suggests that the split of groups A and B A2M-like proteins was a very ancient event, and that the two lineages have evolved independently since that time.

The observation that *Limulus* (Chelicerata) A2M and A2M sequences from two different species of soft ticks (Uniramia) (Saravanna et al. 2003; Valenzuela et al. 2002) are of the group B type now places some protostome members in the group B lineage as well. Such data, along with the finding that *Ciona* has A2M of the group A form (see Fig. 4), lend support to the speculation that during animal phylogeny, various subphyletic lineages of organisms have randomly lost one (but never both) type of A2M-like genes from their genomes.

The group C-type genes include all the complement-like TEP genes. The phylogeny of this group indicates that the invertebrate C3/C4/C5-like genes (commonly referred to as C3-like) burst onto the scene rather quickly from the LCA. The various representatives [from urchin, amphioxus, tunicate, coral, and now the horseshoe crab (not shown)] appear to have been structurally modified or separated by several functions (functional divergence), as indicated by deep roots and/or long branch lengths. Therefore, one can say that as a group, they are orthologous to the vertebrate complement-like TEPs, but individually, it is difficult to determine if these sequences are true orthologs of C3 or if they represent various paralogous descendent relatives of

the ancestor to vertebrates C3, C4, and C5 (note how Urochordate C3-like sequences form a separate clade; Fig. 4).

Since the existence of an A2M-like gene in the coral genome remains to be confirmed, it is difficult to choose an appropriate root for phylogenetic analysis. Choosing a correct outgroup is further complicated by the observation that TEPs are known to not only duplicate numerous times and diverge in function (such as described for protostomes; Levashina et al. 2003), but appear to be lost from genomes as well. While we have partial cDNA sequence data to suggest the existence of a second TEP in the coral, we generated unrooted phylogenies (i.e., Fig. 4) and found that a common node (ancestor) results for the three major groups, A, B, and C, TEP sequences.

The ancestral form of the TEP family

The work described here presents a new paradigm in our understanding of the evolution of C3, C4, and C5 complement and other related proteins, where the ancestral form (s) of the family may have to be reconsidered. We present an alternative model, where an ancient TEP (with C3/A2M-like characteristics) undergoes two successive (tandem) duplication events prior to the divergence of bilaterian animals. One pair (probably linked) immediately diverges as A2M-like TEPs and becomes functionally separated into two lineages to create the group A and B types (the B type would later diverge into A2M in vertebrates). Then, sub-phyletic separation of animals leads to the random loss of one form (group A or B, but never both) of A2M-like TEPs in some lineages. The other paralogous form diverges (by specialization) into what becomes the ancestor to the vertebrate complement components (group C). Furthermore, this model suggests that in the protostomes, multiple sub-phyletic lineages lost the C3-like gene or the group C ancestor (e.g., via recombination and gene deletion in a lineage-specific manner) following its divergence from deuterostomes (a considerable amount of new data suggests extensive gene loss in many protostomes; see Kortschak et al. 2003 and Zdobnov et al. 2002). These events allowed for the recruitment of three similar TEPs into the developing complement system of deuterostomes. We anticipate that genomic sequencing endeavors in Cnidarians and Ctenophores, along with EST projects, and/or genomic data from a more diverse range of protostomes and invertebrate deuterostomes will support this model of TEP family evolution, which suggests that the ancestral protein of C3, C4, and C5 was C3-like rather than A2M-like. Future studies will address the role of SeC3 in the coral and determine whether the coral C3-like protein serves a function analogous to its vertebrate counterpart.

Acknowledgements We thank Dr. M. Nonaka and Dr. M. Nakao for critical comments on an early version of the manuscript. M.L. Herrera, A.S. Goyos, and M. Gomez provided excellent technical assistance. This is contribution FIU-CII-001 from the FIU Comparative Immunology Institute.

References

- Adams MD et al (2000) The genome sequence of *Drosophila melanogaster*. *Science* 287:2185–2195
- Adoutte A, Balavoine G, Lartillot N, Lespinet O, Prud'homme B, de Rosa R (2000) Special feature: the new animal phylogeny: reliability and implications. *Proc Natl Acad Sci U S A* 97:4453–4456
- Ainscough R et al (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. The *C. elegans* Sequencing Consortium. *Science* 282:2012–2018
- 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
- Armstrong PB, Quigley JP (1999) Alpha2-macroglobulin: an evolutionarily conserved arm of the innate immune system. *Dev Comp Immunol* 23:375–390
- Ayala FJ, Rzhetsky A, Ayala FJ (1998) Origin of the metazoan phyla: molecular clocks confirm paleontological estimates. *Proc Natl Acad Sci U S A* 95:606–611
- Barrington R, Zhang M, Fischer M, Carroll MC (2001) The role of complement in inflammation and adaptive immunity. *Immunol Rev* 180:5–15
- Bertheussen K (1979) The cytotoxic reaction in allogeneic mixtures of echinoid phagocytes. *Exp Cell Res* 120:373–381
- Bertheussen K (1982) Receptors for complement on echinoid phagocytes. II. Purified human complement mediates echinoid phagocytosis. *Dev Comp Immunol* 6:635–642
- Bertheussen K (1983) Complement-like activity in sea urchin coelomic fluid. *Dev Comp Immunol* 7:21–31
- Bertheussen K, Sejelid R (1978) Echinoid phagocytes in vitro. *Exp Cell Res* 111:401–412
- Bertheussen K, Sejelid R (1982) Receptors for complement on echinoid phagocytes. I. The opsonic effect of vertebrate sera on echinoid phagocytosis. *Dev Comp Immunol* 6:423–431
- Bokisch VA, Dierich MP, Muller-Eberhard HJ (1975) Third component of complement (C3): structural properties in relation to functions. *Proc Natl Acad Sci U S A* 72:1989–1993
- Campbell RD, Law SKA, Reid KBM, Sim RB (1988) Structure, organization, and regulation of the complement genes. *Annu Rev Immunol* 6:161–195
- Carroll MC (1998) The role of complement and complement receptors in induction and regulation of immunity. *Annu Rev Immunol* 16:545–568
- Christophides GK, Zdobnov E, Barillas-Mury C, Birney E, Blandin S, Blass C, Brey PT, Collins FH, Danielli A, Dimopoulos G, Hetru C, Hoa NT, Hoffmann JA, Kanzok SM, Letunic I, Levashina EA, Loukeris TG, Lycett G, Meister S, Michel K, Moita LF, Muller HM, Osta MA, Paskewitz SM, Reichhart JM, Rzhetsky A, Troxler L, Vernick KD, Vlachou D, Volz J, von Mering C, Xu J, Zheng L, Bork P, Kafatos FC (2002) Immunity-related genes and gene families in *Anopheles gambiae*. *Science* 298:159–165
- Dempsey PW, Allison MED, Akkaraju S, Goodnow CC, Fearon DT (1996) C3d of complement as a molecular adjuvant: bridging innate and acquired immunity. *Science* 271:348–350
- Dodds AW, Law SKA (1998) The phylogeny and evolution of the thioester bond-containing proteins C3, C4, and alpha2-macroglobulin. *Immunol Rev* 166:15–26
- Doolittle RF, Feng D-F, Tsang S, Cho G, Little E (1996) Determining divergence times of the major kingdoms of living organisms with a protein clock. *Science* 271:470–477
- Fearon DT, Locksley RM (1996) The instructive role of innate immunity in the acquired immune response. *Science* 272:50–54
- Felsenstein J (1985) Confidence limits on phylogenetics: an approach using the bootstrap. *Evolution* 39:783–791
- Fujita T (2002) Evolution of the lectin-complement pathway and its role in innate immunity. *Nat Rev Immunol* 2:346–353
- Gadjeva M, Dodds AW, Taniguchi-Sidle A, Willis AC, Isenman DE, Law SKA (1998) The covalent binding reaction of complement component C3. *J Immunol* 161:985–990

- Gonnet GH, Cohen MA, Benner SA (1992) Exhaustive matching of the entire protein sequence database. *Science* 256:1443–1445
- Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modelling. *Electrophoresis* 18:2714–2723
- Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* 41:95–98
- Hammond JA, Nakao M, Smith VJ (2005) Cloning of a glycosylphosphatidylinositol-anchored alpha-2-macroglobulin cDNA from the ascidian, *Ciona intestinalis*, and its possible role in immunity. *Mol Immunol* 42:683–694
- Holt RA, Subramanian GM, Halpern A, Sutton GG, Charlab R, Nusskern DR, Wincher P, Clark AG et al (2002) The genome sequence of the malaria mosquito *Anopheles gambiae*. *Science* 298:129–159
- Huber R, Scholze H, Paques EP, Deisenhofer J (1980) Crystal structure analysis and molecular model of human C3a anaphylatoxin. *Hoppe-Seyler Z Physiol Chem* 361:1389–1399
- Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J Mol Biol* 292:195–202
- Karp DR, Parker KL, Shreffler DC, Capra JD (1981) Characterization of the murine C4 precursor (pro-C4): evidence that the carboxyterminal subunit is the C4 gamma-chain. *J Immunol* 126:2060–2061
- Karplus K, Barrett C, Hughey R (1998) Hidden Markov models for detecting remote protein homologies. *Bioinformatics* 14:846–856
- Kortschak RD, Samuel G, Saint R, Miller D (2003) EST analysis of the Cnidarian *Acropora millepora* reveals extensive gene loss and rapid sequence divergence in the model invertebrates. *Curr Biol* 13:2190–2195
- Krumlauf R (1996) Northern blot analysis. In: Harwood AJ (ed) *Basic DNA and RNA protocols*. Humana Press, Totowa, NJ
- Kumar S (1996) A stepwise algorithm for finding minimum evolution trees. *Mol Biol Evol* 13:584–593
- Kumar S, Tamura K, Jakobsen IB, Nei M (2001) Mega2: molecular evolutionary genetics analysis software. *Bioinformatics* 17:1244–1245
- Lachmann P (1979) An evolutionary view of the complement system. *Behring-Inst-Mitt* 63:25–37
- Law SK, Dodds AW (1997) The internal thioester and the covalent binding properties of the complement proteins C3 and C4. *Protein Sci* 6:263–274
- Law SK, Levine RP (1977) Interaction between the third complement protein and cell surface macromolecules. *Proc Natl Acad Sci U S A* 74:2701–2705
- Law SK, Lichtenberg NA, Levin RP (1979) Evidence for an ester linkage between the labile binding site of C3b and receptive surfaces. *J Immunol* 123:1388–1394
- 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
- Levashina EA, Blandin S, Moita LF, Lagueux M, Kafatos FC (2003) Thioester-containing proteins of protostomes. In: Ezekowitz RAB, Hoffmann JA (eds) *Innate immunity*. Humana Press Inc., Totowa, NJ, pp 155–174
- Lin M, Sutherland DR, Horsfall W, Totty N, Yeo E, Nayar R, Wu X-F, Schuh AC (2002) Cell surface antigen CD109 is a novel member of the alpha2-macroglobulin/C3, C4, C5 family of thioester-containing proteins. *Blood* 99:1683–1691
- Liu J, Rost B (2003) NORSp: predictions of long regions without regular secondary structure. *Nucleic Acids Res* 31:3833–3835
- Liu J, Tan H, Rost B (2002) Loopy proteins appear conserved in evolution. *J Mol Biol* 322:53–64
- Marino R, Kimura Y, De Santis R, Lambris JD, Pinto MR (2002) Complement in urochordates: cloning and characterization of two C3-like genes in the ascidian *Ciona intestinalis*. *Immunogenetics* 53:1055–1064
- Mastellos D, Lambris JD (2002) Complement: more than a ‘guard’ against invading pathogens? *Trends Immunol* 23:485–491
- McGuffin LJ, Jones DT (2003) GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. *Bioinformatics* 19:874–881
- McGuffin LJ, Bryson K, Jones DT (2000) The PSIPRED protein structure prediction server. *Bioinformatics* 16:404–405
- McMenamin MAS, McMenamin DLS (1990) *The emergence of animals: the Cambrian breakthrough*. Columbia University Press, New York
- Morley BJ, Walport MJ (2000) *The complement facts book*. Academic, New York
- Muto Y, Fukumoto Y, Arata Y (1985) Proton nuclear magnetic resonance study of the third component of complement: solution conformation of the carboxyl-terminal segment of C3a fragment. *Biochemistry* 24:6659–6665
- Nagar B, Jones RG, Diefenbach RJ, Isenman DE, Rini JM (1998) X-ray crystal structure of C3d: a C3 fragment and ligand for complement receptor 2. *Science* 280:1277–1281
- Nicholas KB, Nicholas HB Jr (1997) GeneDoc: a tool for annotating and editing multiple sequence alignments. Distributed by author
- Nonaka M (1994) Molecular analysis of the lamprey complement system. *Fish Shellfish Immunol* 4:437–446
- 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
- Page RDM (2001) TreeView ver 1.6.5. Distributed by author
- Peitsch MC, Schwede T, Guex N (2000) Automated protein modelling—the proteome in 3D. *Pharmacogenomics* 1:257–266
- Rost B (1996) PHD: predicting one-dimensional protein structure by profile based neural networks. *Methods Enzymol* 266:525–539
- Rzhetsky A, Nei M (1993) Theoretical foundation of the minimum-evolution method of phylogenetic inference. *Mol Biol Evol* 10:1073–1095
- Sahu A, Lambris JD (2001) Structure and biology of complement protein C3, a connecting link between innate and acquired immunity. *Immunol Rev* 180:35–48
- Sambrook J, Russell DW (2001) *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- Saravanna T, Wiese C, Sojka D, Kopacek P (2003) Molecular cloning, structure, and bait region splice variants of alpha2-macroglobulin from the soft tick *Ornithodoros moubata*. *Insect Biochem Mol Biol* 33:841–851
- Smith LC, Azumi K, Nonaka M (1999) Complement systems in invertebrates: the ancient alternative and lectin pathways. *Immunopharmacology* 42:107–120
- Sottrup-Jensen L, Stepanik TM, Kristensen T, Lonblad PB, Jones CM, Wierzbicki DM, Magnusson S, Domdey H, Wetsel RA, Lundwall A, Tack BF, Fey GH (1985) Common evolutionary origin of alpha2-macroglobulin and complement components C3 and C4. *Proc Natl Acad Sci U S A* 82:9–13
- Stothard P (2000) The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. *BioTechniques* 28:1102–1104
- Suzuki MM, Satoh N, Nonaka M (2002) C6-like and C3-like molecules from the cephalochordate, amphioxus, suggests a cytolytic complement system in invertebrates. *J Mol Evol* 54:671–679
- Swofford DL (1998) *PAUP*: phylogenetic analysis using parsimony (and other methods)*. Sinauer Associates, Sunderland, MA
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 24:4876–4882
- Valenzuela JG, Francischetti IM, Pham VM, Garfield MK, Mather TN, Ribeiro JM (2002) Exploring the sialome of the tick *Ixodes scapularis*. *J Exp Biol* 205:2843–2864

- Vik DP, Amiguet P, Moffat GJ, Fey M, Amiguet-Barras F, Wetsel RA, Tack BF (1991) Structural features of the Human C3 gene: intron/exon organization, transcriptional start site, and promoter region sequence. *Biochemistry* 30:1080–1085
- Vogel CW, Bredehorst R, Fritzing DC, Grunwald T, Ziegelmuller P, Kock MA (1996) Structure and function of cobra venom factor, the complement-activating protein in cobra venom. *Adv Exp Med Biol.* 391:97–114
- Zdobnov EM et al (2002) Comparative genome and proteome analysis of *Anopheles gambiae* and *Drosophila melanogaster*. *Science* 298:149–159
- Zhang Y, Frohman MA (1997) Using rapid amplification of cDNA ends (RACE) to obtain full-length cDNAs. In: Cowell IG, Austin CA (eds) *cDNA library protocols*. Humana Press Inc, Totowa, NJ, pp 61–88
- Zhu Y, Thangamani S, Ho B, Ding JL (2005) The ancient origin of the complement system. *EMBO J* 24:382–394