Sea Urchin Coelomocytes Specifically Express a Homologue of the Complement Component C3[1]

Walid Z. Al-Sharif,* J. Oriol Sunyer,[†] John D. Lambris,[†] and L. Courtney Smith[2]*

A homologue of complement component C3 (SpC3) has been cloned and sequenced from the purple sea urchin, *Strongylocentrotus purpuratus*. The preprocessed, deduced protein size is estimated to be 186 kDa with a short leader and two chains, α and β. There are cysteines in conserved positions for interchain disulfide bonding, and there is a conserved thioester site in the α-chain with an associated histidine. There are five consensus *N*-linked glycosylation sites, and putative cleavage sites for factor I and C3 convertase. Partially purified SpC3 on protein gels shows a nonreduced size of 210 kDa and, under reducing conditions, reveals an α-chain of 130 kDa and a β-chain of 80 kDa. These sizes are larger than the deduced sizes, suggesting that the protein has carbohydrates added to most of the consensus *N*-linked glycosylation sites. Phylogenetic analysis of SpC3 compared with other members of the thioester protein family, which includes C3, C4, C5, and α_2-macroglobulin, shows that SpC3 is the first divergent complement protein, falling at the base of the complement protein clade. Transcripts from the SpC3 gene (Sp064) are 9 kb, and the gene is expressed specifically in coelomocytes, which are the immunocytes in the sea urchin. Genome blots suggest that SpC3 is encoded by a single copy gene per haploid genome. This is the first identification of a complement component in an invertebrate, and suggests homology of the innate immune system within the deuterostome lineage of animals. *The Journal of Immunology,* **1998, 160: 2983–2997.**

The immune response in higher vertebrates is composed of a complex of inter-regulated systems, including both adaptive and innate reactivities to foreign Ags, that are mediated by cellular and humoral systems. One of the major effector arms of the immune response is complement that is composed of about 30 distinct humoral and cell surface proteins. The complement system comprises three convergent pathways (classical, alternate, and lectin) that function to amplify the initiating signal through feedback systems of serine protease activities. The pathways converge and activate the terminal or lytic cascade, which results in the destruction of foreign cells through the formation of the membrane attack complex. The activities of these cascades are controlled by mechanisms that involve additional regulatory proteins. The classical cascade is initiated through specific Ag-Ab interactions and can be considered an effector system of adaptive immunity. The alternative cascade, part of the innate response, is initiated by the complement component C3 that undergoes a constant, low level spontaneous autoactivation reaction, enabling it to bind to molecules (reviewed in Ref. 1). The lectin pathway, also part of the innate system, is initiated through the mannan-binding lectin that interacts with mannose sugars on cell surfaces. It functions in place of C1q (2, 3), and the mannan-binding lectin-associated protease functions in place of C1s and C1r to activate either C4 of the classical pathway (4–6) or C3 of the alternate pathway (7).

Complement component C3, the central component in all cascades, is the most versatile and multifunctional molecule of the complement proteins (8). Not only does it function in the complement cascades, but it interacts with cell surface proteins on self and foreign cells to initiate and augment immune responsiveness, and it interacts with an array of proteins that regulate its activity (9–11). C3 also has opsonin functions and is important in immune surveillance and host protection against microbial infection by augmenting phagocytic removal and destruction of invading pathogens by phagocytic cells bearing receptors for C3b. Furthermore, the breakdown product of complement component C3, C3d, when bound to the Ag, is remarkably effective in augmenting the immunogenicity of foreign Ags to activate a specific immune response (12). Clearly, C3 functions effectively in both the adaptive and innate immune responses and acts to link them together. Yet because of the complex and essential activities of C3 in the higher vertebrate, this creates problems in studying each C3 function independent of the others. One means to understand the complexities and evolution of the higher vertebrate complement system and the immune system in general has been to investigate animals that are phylogenetically related, yet have simpler immune responses. The deuterostome lineage of animals is composed of two major phyla, the chordata that includes mammals, and the echinodermata that includes sea urchins, sea stars, sea cucumbers, and other groups. Because of this relationship, the echinoderms are an appropriate choice to understand innate immune responses without the added complexities of interactions between the innate and adaptive systems.

Coelomocytes are cells found in the coelomic cavity of the adult sea urchin that function as mediators of the immune system (for review, see Ref. 13 and references cited therein). Previously, we found that small doses of LPS would significantly activate the

*Department of Biological Sciences, George Washington University, Washington, D.C. 20052; and [†]Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA 19104

0022-1767/98/$02.00

coelomocytes from the purple sea urchin, *Strongylocentrotus purpuratus* (14). This study was then followed by an investigation of genes expressed in these LPS-activated coelomocytes (15). Of the 307 expressed sequence tags (ESTs)[3] that were reported, one (EST064) encoded an amino acid sequence that showed significant similarities to the thioester family of complement proteins. We report in this work the completed sequence of the EST064 cDNA, hereafter called Sp064, which shows that it encodes a homologue of complement component C3, called SpC3. The Sp064 gene appears to be single copy per haploid genome, and transcripts are present only in coelomocytes. The encoded protein, SpC3, is found in the coelomic fluid and is composed of two chains. This is the first identification of a complement component homologue expressed in an invertebrate. Because sea urchins are phylogenetically related to vertebrates, both groups being deuterostomes, these data indicate that a simple complement system was present in the deuterostome ancestor rather than the vertebrate ancestor and that it is a far more ancient system than had been previously assumed.

Materials and Methods

RNA isolation

Isolation of total RNA from coelomocytes and other adult tissues was done according to Smith et al. (16). Briefly, coelomic fluid (40 ml) was poured through sterile cheese cloth and mixed into 10 ml of cold Ca^{2+}- and Mg^{2+}-free sea water (17) containing 30 mM EDTA, pH 7.4 (CMFSW-E). Coelomocyte pellets and minced solid tissues were vortexed and homogenized using a dounce homogenizer in guanidinium thiocyanate extraction buffer (5 M guanidinium thiocyanate, 50 mM sodium acetate, 50 mM EDTA, 50 mM Tris, pH 7.4, and 5% β-mercaptoethanol), to which was added *N*-lauroyl sarcosine to a final concentration of 2%. Total RNA was pelleted through a cushion of 5.7 M CsCl containing 50 mM sodium acetate and 50 mM EDTA at $10^5 \times g$ in either a Ti60 fixed angle rotor (Beckman Instruments, Fullerton, CA) or a swinging bucket rotor (Sorvall, Newtown, CT) at 20°C for 20 h. Pellets were washed in 70% ethanol, resuspended in RNase-free water, extracted in phenol/sevag (1:1) (sevag is 24 parts chloroform, 1 part isoamyl alcohol), precipitated, and resuspended in RNase-free water. Poly(A)$^+$ RNA was isolated using oligo(dT) magnetic beads (Dynal, Great Neck, NY).

cDNA library construction

The LPS-activated coelomocyte cDNA library was constructed using the Time Saver cDNA kit (Pharmacia, Piscataway, NJ) according to kit instructions. The first strand of cDNA was made from Poly(A)$^+$ RNA with a random primer containing a *Not*I site. An *Eco*RI adapter was ligated to the 5′ end, and the cDNA was directionally cloned into λExCell phage arms (Pharmacia), as previously reported (15). An arrayed cDNA library was directionally constructed from nonactivated coelomocyte poly(A)$^+$ RNA using the same *Not*I/random primer, but with a *Sal*I adapter at the 5′ end. The cDNA was ligated into the pSPORT vector (Life Technologies, Grand Island, NY) and electroporated into DH10b bacteria. Individual colonies were inoculated into 384-well microtiter plates according to the methods of Maier et al. (18). The arrayed library consisted of approximately 92,160 clones. PCR products of clone inserts (18,432) were spotted in duplicate onto five Hybond N$^+$ nylon filters, 22 cm × 22 cm (Amersham, Arlington Heights, IL) for screening.

Screening cDNA libraries

The λExCell library was screened using ^{32}P-labeled RNA probes that were generated according to technical information from Promega Corp. (Madison, WI) and as previously reported (16). Filters were prehybridized for 2 h in hybridization solution (50% formamide, 250 mM phosphate buffer, pH 7.4, 1 mM EDTA, 0.1% BSA, and 7% SDS), and then incubated with the probe at 42°C overnight in a rotating oven (Robbins Scientific, Sunnyvale, CA). Final washes were conducted at 68°C in 1× SSC (0.15 M NaCl and 15 mM sodium citrate, pH 7) with 1% SDS. Filters were exposed overnight to X-OMAT XAR-5 x-ray film (Eastman Kodak, Rochester, NY), and positive plaques were rescreened once to purify the phage clones. Phagemids were released according to manufacturers' instructions (Pharmacia),

and in some cases, inserts were subcloned into Bluescript (Stratagene, La Jolla, CA).

The arrayed library was screened with a random primed ^{32}P-labeled probe generated from a deletion clone that was made for sequencing, and consisted of approximately 500 bp of the 5′ end of pExCell14 (Fig. 1). The deletion clone was first used as a PCR template with the M13-40 primer and a primer specific for Sp064 (TCTAAGCAGGTAGACAGC; see Fig. 2, nucleotides 1988 to 2001). After amplification of the insert, the 5′ polylinker was removed by *Eco*RI digestion, producing a 160-bp PCR fragment, which was isolated by gel electrophoresis and eletroelution. The fragment was then labeled with ^{32}P by random priming.

Filters spotted with DNA from the arrayed library were prehybridized for 1 h at 65°C in 5× SSPE (20× SSPE is 3 M NaCl, 0.2 M NaH$_2$PO$_4$, and 20 mM EDTA, pH 7.4), 5× Denhardt's solution (50× Denhardt's is 1% BSA (fraction V; Sigma Chemical Co., St. Louis, MO), 1% polyvinylpyrrolidone, and 1% Ficoll type 400), 0.5% SDS, and 20 µg/ml denatured salmon sperm DNA (type VI; Sigma Chemical Co.). Denatured probe was added and allowed to hybridize for 16 h at 65°C. Filters were washed twice for 10 min each at room temperature in 2× SSPE with 0.1% SDS, followed by three washes of 15 min each in 1× SSPE, 0.1% SDS at 65°C. Filters were then mounted in Saran wrap and exposed to film for 4 h. Plasmids were isolated from positive clones for subsequent characterization.

cDNA sequencing

Sequencing was conducted on plasmid DNA according to the dideoxynucleotide termination protocol (19) using the TaqTrack sequencing kit (Promega Corp.) and incorporating $[\alpha$-^{35}S]dATP (DuPont NEN, Boston, MA). Sequencing reactions were electrophoresed on a 6% acrylamide gel with 0.6× TBE (10× TBE is 0.9 M Tris, 0.9 M boric acid, and 20 mM EDTA, pH 8.3) running buffer, after which the gel was dried and exposed overnight to BioMax MR-1 x-ray film (Eastman Kodak). For clones longer than several hundred nucleotides, such as the *Eco*RI to *Not*I fragment of pExcell139 and all of pExCell14 (see Fig. 1), insert fragments were subcloned into Bluescript (Stratagene), and the Erase-a-Base kit (Promega Corp.) was used to create a nested set of deleted clones. Sequences were analyzed using the MacVector sequence analysis program (Eastman Kodak) and the AssemblyLIGN sequence assembly program (Eastman Kodak) on a Power Macintosh (Apple Computer, Cupertino, CA).

Protein alignment and phylogenetic analysis

The cDNA sequence was used to run a basic blast search of GenBank (20), and several of the matched sequences were used to construct protein alignments with the Clustal W program (21). Without altering the default parameters, a number of variations were used to obtain the best alignment. We defined "best" alignments to be those in which the thioester site and junctions between chains corresponded best between the various proteins. It was noticed that the order in which the proteins were listed slightly affected the alignment results using Clustal W. Therefore, the order was changed in several runs to obtain the best alignment. These variations included 1) listing SpC3 first, followed by the other proteins arranged from lower to higher phylogenetic order, 2) keeping multiple proteins from the same species together, and 3) listing the α_2-macroglobulin (α_2 M) protein family last. The best alignment on Clustal W was verified by using a second alignment program, DNASIS version 2.1 (Hitachi, San Bruno, CA). To optimize the alignment on DNASIS, various gap penalties were tried (5, 10, 25, 50) and then kept at 5. The fixed gap penalty (10, 25, 50, 100) and floating gap penalty (10, 25, 50, 100) were adjusted separately and in combinations. We found that alignments with DNASIS could not be improved over that generated by Clustal W, and results from the two programs agreed best when the gap penalties on DNASIS were maintained at the default settings.

The PAUP program (version 3.1.1) (22) was used with standard set parameters (character types were set as unordered, and character weights were set as 1) to compare and assemble multiple sequences from the Clustal W alignments into a phylogenetic tree. Slightly different alignments resulting from different ordering of the proteins in Clustal W were used in the PAUP program to identify the shortest tree. The heuristic search method was used with various search options to obtain the shortest tree. The general search options were set to keep minimal trees only and to collapse zero-length branches. The stepwise addition searches were either simple or random with seed numbers of 1, 50, and 100. The degree of support for internal branches was assessed using the bootstrapping method with 1000 bootstrap replications.

Preparation of anti-SpC3 antiserum

A rabbit antiserum was produced against the peptide, DNAKVQEEVD VSPSIGR (see Fig. 2), which was chosen from the amino acid sequence

---

[3] Abbreviations used in this paper: EST, expressed sequence tag; α_2 M, α_2-macroglobulin; nt, nucleotide; UT, untranslated.
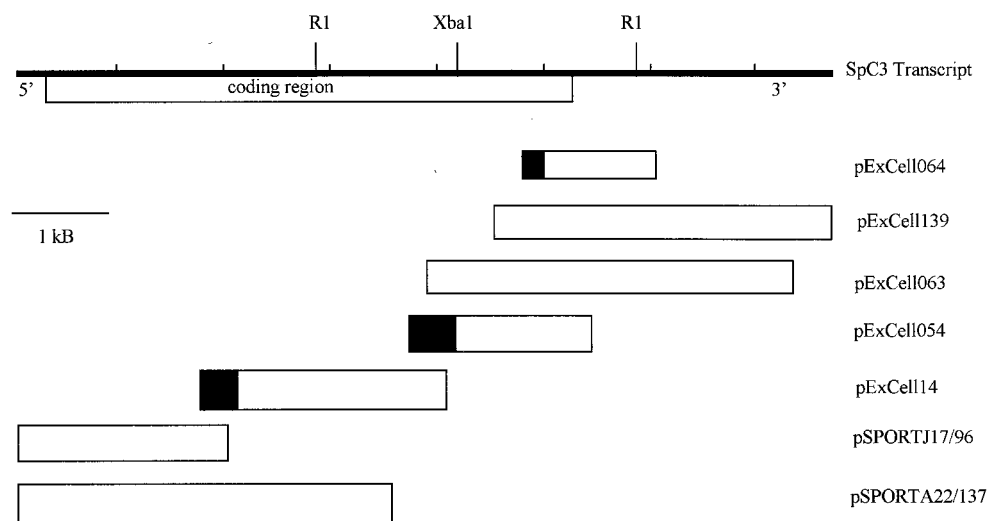
FIGURE 1. Map of cDNAs that span Sp064. This schematic drawing shows some of the clones that were isolated and sequenced in analyzing Sp064. pExCell clones were isolated from the LPS-activated coelomocyte λExCell library (15), and pSPORT clones were isolated from the nonactivated coelomocyte-arrayed cDNA library. Some of the 3′ UT region is missing from Sp064 since the poly(A)+ tail was not identified. In addition, much of the 5′ UT region is also missing. The EcoRI (R1) and XbaI sites were used for orienting clones relative to each other and for subcloning. Templates for generating RNA antisense probes and PCR-amplified DNA probes (denoted as black ends) were used in screening libraries, the Northern blot, and the genome blot.

deduced from Sp064 and was based on predictions from the human C3 structure that it would be located on an exposed region of the α-chain and would therefore produce a useful anti-peptide antiserum. The peptide was synthesized using a 430A peptide synthesizer (Applied Biosystems, Foster City, CA), conjugated to keyhole limpet hemocyanin using glutaraldehyde (23), mixed with complete (first injection) or in CFA (second and third boosts). The rabbit received three injections given weekly, after which it was bled weekly for 3 wk. The serum was collected and stored at −70°C.

Specific anti-peptide Abs were obtained by affinity chromatography. The synthetic peptide was coupled to activated CH-Sepharose 4B (Pharmacia) according to the manufacturer's instructions. The peptide column was equilibrated with PBS containing 10 mM EDTA, and 2 ml of the rabbit antiserum containing 10 mM EDTA was passed over the column twice. Unbound protein was washed with PBS, and the specific Ab was eluted with 0.1 M glycine/HCl, pH 2.5. The pH of the eluted fractions was immediately neutralized by the addition of 1 M Tris-HCl, pH 8. Affinity-purified Ab was stored at −70°C.

Purification of SpC3 protein from coelomic fluid

Sea urchin C3 was partially purified from coelomic fluid using modifications to published methods (24). Coelomic fluid was pooled from several sea urchins to which had been added EDTA (3 to 15 mM, final concentration), PMSF (2 mM, final concentration; Sigma Chemical Co.), and pepstatin A (1 to 100 μM, final concentration; Sigma Chemical Co.). Coelomocytes were pelleted, and the cell-free fluid was stored at −70°C until further purification could be conducted. Forty milliliters of sea urchin coelomic fluid were concentrated with an Amicon filter (10-kDa cut-off) (Amicon, Bedford, MA) to 2 ml, and the sample was passed through a PD-10 gel filtration column (Pharmacia) to exchange the buffer to 10 mM phosphate, pH 7.5. The concentrated coelomic fluid was precipitated with 4% polyethylene glycol, by stirring for 30 min at 4°C, followed by centrifugation at 15,000 × g for 20 min. The supernatant was brought to 16% polyethylene glycol by stirring at 4°C for 30 min and centrifuging as before. The pellet was then resuspended in 10 mM phosphate buffer, pH 7.5, and applied to a Mono Q HR 5/5 anion exchange chromatography column (Pharmacia) that had been equilibrated with the same buffer. Bound proteins were eluted with a linear salt gradient (0–500 mM NaCl), and fractions containing SpC3 were identified by gel electrophoresis and Western blotting using the affinity-purified anti-SpC3 peptide Ab.

Determination of NH2-terminal amino acid sequence

Purified protein was subjected to SDS-PAGE under reducing conditions and electroblotted onto ProBlott membranes (Applied Biosystems), and the NH2-terminal sequences of the protein chains were obtained by using a modification of the method of Matsudaira (25), as previously described (26). The individual SpC3 chains were separated by SDS-PAGE, electro-

blotted, cut out of the filter, and subjected to Edman degradation, using an 473 protein sequencer (Applied Biosystems).

Results

Isolation and sequencing the SpC3 cDNAs

A cDNA, identified as encoding a putative sea urchin complement component, was one of 307 clones that were partially characterized from the LPS-activated coelomocyte λExCell library, and was reported as expressed sequence tag 064 (EST064) (15). In preliminary BLAST (20) searches of GenBank, EST064 matched to cobra C3 (27), human and mouse C3 (28, 29), mouse sex-limited protein (30), in addition to a few others. This result suggested that EST064 encoded a new member of the thioester gene family. The 5′ end of pExCell064 (Fig. 1) was used to rescreen the same library to obtain additional clones. The initial screen yielded 144 clones, 43 of which rescreened positive, and 12 of which were chosen for sequencing based on the size differences of their 5′ ends. Sizes were determined by PCR using Sp6, a primer that hybridizes to the 5′ polylinker of the phagemid, and an insert-specific primer that hybridized to the 5′ end of pExCell064. Because this set of clones had approximately 100- to 200-bp increment differences in size, this created a natural deletion set for sequencing. Three (pExCell139, 063, and 054) of the 12 clones from this set are shown on Figure 1. The sequence obtained from the 5′ ends of these 12 clones was published as the deduced protein, and was shown aligned with other complement protein family members (see Ref. 15, Fig. 1). However, at the time, it was not clear whether SpC3 was homologous to C3, C4, or C5, although it did not appear to be α2 M (15).

Because the total sequence length of these overlapping clones (shown in Fig. 1 as pExCell054, pExCell063, pExCell064, and pExCell139) covered only about one-half of the estimated open reading frame and only about one-half of the transcript, the λExCell library was screened a second time. The first 450 bp of pExCell054, the 5′-most clone of the set (Fig. 1), was used to make another riboprobe that yielded 84 clones, of which 18 rescreened positive. The clone with the longest 5′ end, pExCell14 (Fig. 1), was sequenced in its entirety using nested set deletion clones.

```
  1 TCGGTGCGGCATCTCTCCTGAATTCTAACCTACTGCATGATCAGTCAAGGAGCTGCAGATTCCTCCTAGGTTTGGTAAAGACGAAACTGCTTACCGAAGACGTGGTCAATGACTGGTGTC      120

121 TCGACCACCATGGGGTCGATTATTCTCATCTTCGTACTCGCGATTAGTGTGCATCACTTGTACCGGGAGGGGTGACTCAAGATCCACCTGCGACCTTTTTCGTGATCTCTCCTAACGTGTTT      240
  1            M  G  S  I  I  L  I  F  V  L  A  I  S  A  S  L  V  P  G  G  V  T  Q  D  P  P  A  T  F  F  V  I  S  P  N  V  F      37
            Kozak sequence        Leader

241 CGGGCAGGTGTAGAAGAGAAAGTCGTCGTAACCCTCATCAGATCCCCACCAAACGTTCAGAACATTGAAGTGAAAGTCTCATTGATGCGGCCCGGTTCGACTGTGACTTTTTCGGAAGAT      360
 38 R  A  G  V  E  E  K  V  V  V  T  L  I  R  S  P  P  N  V  Q  N  I  E  V  K  V  S  L  M  R  P  G  S  T  V  T  F  S  E  D      77

361 ACGAGACTTGTCAGCCCAGGTTCATCTCAATCCTTTTCGGTCGTAATTCAAGCTGGTGATTTGATGAGAAATGAGGGAGCTTTCCAACACATGGTCTTAAAGGCTGAATCTCTGAACCCT      480
 78 T  R  L  V  S  P  G  S  S  Q  S  F  S  V  V  I  Q  A  G  D  L  M  R  N  E  G  A  F  Q  H  M  V  L  K  A  E  S  L  N  P      117

481 ATCTACCCATTCGAAGAGCAAACAGATATCCTGGTCACTCTTCAATCTGGTTACGTCTTCGTACAGACGGACAAACCAATCTACACCCCAAACCAAGACGTGATGATCAAAGTCATGTCC      600
118 I  Y  P  F  E  E  Q  T  D  I  L  V  T  L  Q  S  G  Y  V  E  V  Q  T  D  K  P  I  Y  T  P  N  Q  D  V  M  I  K  V  M  S      157

601 TTGGATCAGGACATGCTTCCATCAGATCGAGAGCTCCACGTCGAAATAATGGATCCTAGTGGGACCAGTGTAAAAAGACTCCGAGAGACCCATGCCGACGTTCCGGCTTCTATCGGGCGA      720
158 L  D  Q  D  M  L  P  S  D  R  E  L  H  V  E  I  M  D  P  S  G  T  S  V  K  R  L  R  E  T  H  A  D  V  P  A  S  I  G  R      197

721 TTACAGGTCCATAATCATCCAGCTTCGGTGAATGGACCGTTTCGGCTTTCTATGGACCCGAGTTCCTTGTCAAGTCAGGTGACATTTGTGGTCAAGGAGTATGTTCTACCGACATTTTCA      840
198 L  Q  V  H  N  H  P  A  S  V  N  G  P  F  R  L  S  M  D  P  S  S  L  S  S  Q  V  T  F  V  V  K  E  Y  V  L  P  T  F  S      237

841 GTGAGCATCGAAACGCCTAAATACATTCTAGCCGGGAGCTGATTCAGTGGTCTCAACAATAATTGCTGAGTACGTATTTGGCAAGCCTGTGATTGGCCGCTATACCTTCAAGTATGGCGTC      960
238 V  S  I  E  T  P  K  Y  I  L  A  G  A  D  S  V  V  S  T  I  I  A  E  Y  V  F  G  K  P  V  I  G  R  Y  T  F  K  Y  G  V      277

961 GTGGATAATGGGACTGTTCAAGAGCTGGGAACTAAGCAAGGGCTGCTTGGTGATAGTGGAAGGGCTGAAGTTACTTTGGATGATCTCAGTACTGATTTTGGTGAAGACTGGTTTGTTCGG      1080
278 V  D  N  G  T  V  Q  E  L  G  T  K  Q  G  L  L  G  D  S  G  R  A  E  V  T  L  D  D  L  S  T  D  F  G  E  D  W  F  V  R      317
      ↑ CHO

1081 TTCAGAGGTCGCCACTTCCATGCAAATGCAATTGTACATGAATCAGCGACTCAGTTCAGCGAAGCAAGCATCAACACGAAGGCAATCTTCGTTGACTCACCATACAAATTCAGTACAAAG      1200
318 F  R  G  R  H  F  H  A  N  A  I  V  H  E  S  A  T  Q  F  S  E  A  S  I  N  T  K  A  I  F  V  D  S  P  Y  K  F  S  T  K      357

1201 CGAACTGTGCAGCACTTCAAGGCTGGACTCAATCTGCAAGTAAAGCTGGATTTAACATTTGCAAATGGAGATGTGGCCCCCAATGTTCCTGTAAATGTCATAGCAACTGCCACATTAAGA      1320
358 R  T  V  Q  H  F  K  A  G  L  N  L  Q  V  K  L  D  L  T  F  A  N  G  D  V  A  P  N  V  P  V  N  V  I  A  T  A  T  L  R      397

1321 GATGGTTCTGAGGAGCTACTACGGAGACCTGATATAGCTGGTGGTCAAGAGGAGATATCCAACACTGATGATCAAGGAGGAGTCAGTATGGTTCTTAACGTCCAATCAAGTATAACGAGC      1440
398 D  G  S  E  E  L  L  R  R  P  D  I  A  G  G  Q  E  E  I  S  N  T  D  D  Q  G  G  V  S  M  V  L  N  V  Q  S  S  I  T  S      437

1441 ATTCACATTGTGGCAGCAACTCATGATCCGCAATATCCAAACAACCAAGCTGATATCTCCTTTGAAGTAAGTCCAGGAGTCTCGGCCAACGGTCATGATTACTTAGTCATCAGACCTCAG      1560
438 I  H  I  V  A  A  T  H  D  P  Q  Y  P  N  N  Q  A  D  I  S  F  E  V  S  P  G  V  S  A  N  G  H  D  Y  L  V  I  R  P  Q      477

1561 AACACAGATGATCGTAACCTGAATGTTGATGTGAGCACTGATTTCTTGATTCAGAGGATTGGAACAACAGGAGATCAAGACATAGATCTCCACTTTCTTTGCATTACAGGTGGCAAGGTT      1680
478 N  T  D  D  R  N  L  N  V  D  V  S  T  D  F  L  I  Q  R  I  G  T  T  G  D  Q  D  I  D  L  H  F  L  C  I  T  G  G  K  V      517

1681 GTTCTGGAGGGAGTTCAGAGAAGCATAAGTGCAGCTGGCACAAACCTGGCCATAACAATCAGAGCATACATGGCTCCACAAATGCGACTGATCGTCTACTATGTCACCATGGATGGCAGC      1800
518 V  L  E  G  V  Q  R  S  I  S  A  A  G  T  N  L  A  I  T  I  R  A  Y  M  A  P  Q  M  R  L  I  V  Y  Y  V  T  M  D  G  S      557

1801 GTTATTGCCGACTCTCTCCTTTTAGGTGTGGAAGAAAAGTGCAGACAAGACCCACAGCTTTCTCTGGACATTCTTCCAAGGAATGTAGGACCTGAACGTGACGTTTATGAACCCAATGGG      1920
558 V  I  A  D  S  L  L  L  G  V  E  E  K  C  R  Q  D  P  Q  L  S  L  D  I  L  P  R  N  V  G  P  E  R  D  V  Y  E  P  N  G      597

1921 CAGATACAGGTGGAAGTTACTGCACCTATCGACTCCAATGTTGGCCTTCTAGCTGTTGATAAAGCTGTCTACCTGCTTAGAGACAAGGACCGCATGGGCAAACAAAAGATGTATGAACGG      2040
598 Q  I  Q  V  E  V  T  A  P  I  D  S  N  V  G  L  L  A  V  D  K  A  V  Y  L  L  R  D  K  D  R  M  G  K  Q  K  M  Y  E  R      637

2041 ATGAGGAGCTATGACACAGGATGTGGACCAGGAGGGGGACAAAACACAGCTCAAATTTTCAAGGATTGTGGAATGACTGTACTTACCAACGCAGGTCTGGATGTTCCTATTCGTGAAGAT      2160
638 M  R  S  Y  D  T  G  C  G  P  G  G  G  Q  N  T  A  Q  I  F  K  D  C  G  M  T  V  L  T  N  A  G  L  D  V  P  I  R  E  D      677

2161 GTTGAATGCATGGATGAAGATACAAGAAGGAAGAGATCGATTGACAGGGATCAACTGTGTCTGTATGATCCGACTTATCTAGCAGACTGTTTAGCTGATAAACCAAGGTTGCGAAGGGTG      2280
678 V  E  C  M  D  E  D  T  R  R  K  R  S  I  D  R  D  Q  L  C  L  Y  D  P  T  Y  L  A  D  C  L  A  D  K  P  R  L  R  R  V      717
             β chain ←  *  *  *  *  *  → α chain
                        ↑N-terminal sequence obtained by Edman Degradation

2281 CTTACAGAAGGCGGTTGGCTATGTCCCCGCCGTGCTCGCGTTTTGGAAGCCGAGTGTAACCTAGAACCGGACCAAACACAGGTATATCGAGATTGCTGTATAAATAGTATGAATCCAGTC      2400
718 L  T  E  G  G  W  L  C  P  R  R  A  R  V  L  E  A  E  C  N  L  E  P  D  Q  T  Q  V  Y  R  D  C  C  I  N  S  M  N  P  V      757

2401 GTGACTGCAACTTCCCGATCAGGTGGGGATGGTGGAGAGCAGAATGCTGCTGTGAAGGTGCGAGATGACTTCAGAGAGACGTGGTTTTTTGACGTTGTTTCCATGCCTGAAGATGGTAGT      2520
758 V  T  A  T  S  R  S  G  G  D  G  G  E  Q  N  A  A  V  K  V  R  D  D  F  R  E  T  W  F  F  D  V  V  S  M  P  E  D  G  S      797

2521 CCGTACTTATACCCGGTCTCCATTCCAAGTAGTATCACGGATTGGCATTTGACAGCTGTCAGTCTTTCACCTACTCAAGGAATGTGTGTTGAAGATGAAACCACAGTGAGTGTGTTCCAG      2640
798 P  Y  L  Y  P  V  S  I  P  S  S  I  T  D  W  H  L  T  A  V  S  L  S  P  T  Q  G  M  C  V  E  D  E  T  T  V  S  V  F  Q      837

2641 GATTTCTTCATCCAGCTTCCACCTTCCCTACTCGGTTGTCAGGCTCGAACAGACACAAGTCATCGCCACCATCTTCAACTATGGTTTCAGTGATTTTGAGGTCAGTGTTAATTTCACTGTT      2760
838 D  F  F  I  Q  L  H  L  P  Y  S  V  V  R  L  E  Q  T  Q  V  I  A  T  I  F  N  Y  G  F  S  D  F  E  V  S  V  N  F  T  V      877
                                                                                                      ↑ CHO

2761 GATCAGGGCTTGTGCACAGCGGAAAATCCTGCCATTCGCCACGTTTTTGTTGAATCTAAACGGGCAGCCTCTACAACCTTTCATTGTACGCCCTGTGGAAGTAGGGGAATTCCATATACT      2880
878 D  Q  G  L  C  T  A  E  N  P  A  I  R  H  V  F  V  E  S  K  R  A  A  S  T  T  F  H  C  T  P  C  G  S  R  G  I  P  Y  T      917

2881 GTGACTGCTGGTGGAAGCAAGAAGCGGGACAGCGTCAGGAATAACCTTCGAGTTGTGTCTCAAGGAGTTATGCAGAGGAAAAGTAGATCTTTGACCCTGAATCCTGGGAGAGTGATGTTC      3000
918 V  T  A  G  G  S  K  K  R  D  S  V  R  N  N  L  R  V  V  S  Q  G  V  M  Q  R  K  S  R  S  L  T  L  N  P  G  R  V  M  F      957

3001 AGTGATGATGTCACCACACCGTCCCCTAATAACAGCTTAGGCTCAGGCCCAGGAACTATATTTGGGGAGGGCTTTCAGCATGAGGAGATTGCCATTTCTTTGCCTGGATCTTCTATTCCA      3120
958 S  D  D  V  T  T  P  S  P  N  N  S  L  G  S  G  P  G  T  I  F  G  E  G  F  Q  H  E  E  I  A  I  S  L  P  G  S  S  I  P      997
                      ↑ CHO

3121 GACACTGAATCTTGCAGTGTCAAACTTATCGGTAACCTACTTGGAACTGCATCCACAGATCCCATAGGTGGGCTAGATCATCTCGGTGAGGCAACCTAGAGGCTGTGGTGAACAAACCATG      3240
998 D  T  E  S  C  S  V  K  L  I  G  N  L  L  G  T  A  S  T  D  P  I  G  G  L  D  H  L  V  R  Q  P  R  G  C  G  E  Q  T  M      1037
                                                                                                Thioester Site

3241 ATCTACCTGGCTCCAACACTCTTTGTCTATCAGTACCTAATAGCCGTAGGATCAGATACAGCAGAACAAGAGGCACGGATCTATGACTACATCGCTGATGGAGTTGCCGAGAGTTGACT      3360
1038 I  Y  L  A  P  T  L  F  V  Y  Q  Y  L  I  A  V  G  S  D  T  A  E  Q  E  A  R  I  Y  D  Y  I  A  D  G  V  A  R  E  L  T      1077
```

FIGURE 2. Sequence of the SpC3 cDNA and deduced protein. The sequence of the entire open reading frame encoded by Sp064 was compiled from overlapping regions of several cDNA clones and the sets of nested deletions using the MacVector sequence analysis program (Oxford Molecular Group, Campbell, CA) on a PowerMac computer. The deduced amino acid sequence is shown below the DNA sequence. The complete cDNA and amino acid sequences have been deposited in GenBank with the accession number AF025526. Several regions of the protein are labeled. CHO denotes a consensus site for *N*-linked carbohydrates. Three ATG codons, putative start sites, and two stop codons are underlined within the 5′ UT region. Six AU-rich elements of the consensus sequence ATTTA are located in the 3′ UT region and are denoted in bold.

```
3361 TATCGGCAAGACAATGGTGCATATGCAGCATGGAAGCATCGACCTGGAAGTACATGGTTGACGGCATTTGTTGTGAAGGTGTTCAGCCAGGCTAATAGGTTCACTAGGGTGGAGCCAGGC 3480
1078  Y  R  Q  D  N  G  A  Y  A  A  W  K  H  R  P  G  S  T  W  L  T  A  F  V  V  K  V  F  S  Q  A  N  R  F  T  R  V  E  P  G  1117
                      Thioester associated Histidine ↑

3481 CATGTAGAAGGTAGTATAAACTGGCTCATTGACAACAACCAGCTACCGTCAGGGGCTTTCCAAGAAAGTCAACAAGTAATTCACCAGGAAATGATTGGAGCTGTAAAGGGAGAAACTTCT 3600
1118  H  V  E  G  S  I  N  W  L  I  D  N  N  Q  L  P  S  G  A  F  Q  E  S  Q  Q  V  I  H  Q  E  M  I  G  A  V  K  G  E  T  S  1157

3601 ATGACAGCATTTGTTCTCATCTCTCTCCTGGAGAGCCGGAATTTACTTGTCCCTGCAAACCAGAAAATTGATGAGGCAATTGGGAAAGCGACCGAGTACCTGGTTACCCAAGTTGAAAAT 3720
1158  M  T  A  F  V  L  I  S  L  L  E  S  R  N  L  L  V  P  A  N  Q  K  I  D  E  A  I  G  K  A  T  E  Y  L  V  T  Q  V  E  N  1197

3721 ATAGACAGAGTCTATGACAAGGCTCTCGTCACTTCTGCGCTGCGTGAAAGCGCAAGTGTGGGGGACTGCAAATGGCAAACTGTGGGAAGATCGAAATGAAGATGGTACTGGGGCTGTTAGT 3840
1198  I  D  R  V  Y  D  K  A  L  V  T  S  A  L  R  E  S  A  S  V  G  T  A  N  G  K  L  W  E  D  R  N  E  D  G  T  G  A  V  S  1237

3841 TTTACACCAGATGATGCCAACTATGAAGATGGCAGTCAGCCATTTTGGCTACAGAGGAAGCCTTCTGCTATTGAGATAGAAACCTCCGGCTACGCTCTCCTGGCACAACTTGCCTTACTT 3960
1238  F  T  P  D  D  A  N  Y  E  D  G  S  Q  P  F  W  L  Q  R  K  P  S  A  I  E  I  E  T  S  G  Y  A  L  L  A  Q  L  A  L  L  1277

3961 GATTATCAAAAGGCAGGCAAAATTGCTCTCTGGCTTAGCAAGCAACAGAACGATGGTGGAGGATTTGTTTCACCTCAGGATACCGTAGTAGCACTACAGGCACTTGCCAAGTATACTGAG 4080
1278  D  Y  Q  K  A  G  K  I  A  L  W  L  S  K  Q  Q  N  D  G  G  G  F  V  S  P  Q  D  T  V  V  A  L  Q  A  L  A  K  Y  T  E  1317

4081 CGTCCGGAGTTCAATACTATCGAAATGAACTGTGACGTGGCAACTGAGCAAATCCCTCTTCATAGGTATCACATTGGGAATGACAATGCCAAAGTACAAGAAGAAGTTGATGTGTCACCG 4200
1318  R  P  E  F  N  T  I  E  M  N  C  D  V  A  T  E  Q  I  P  L  H  R  Y  H  I  G  N  D  N  A  K  V  Q  E  E  V  D  V  S  P  1357
                                                                                 peptide employed for raising the rabbit

4201 TCAATTGGACGCAGCTTAACATTTGATTCTAGAGGAACTGGGGTTGCCAAAGCAAATGTGGAGCTGCGTTACAACACAGAAAAATCTGATATTGACACCTGTCCATTCCATCTTAACATC 4320
1358  S  I  G  R  S  L  T  F  D  S  R  G  T  G  V  A  K  A  N  V  E  L  R  Y  N  T  E  K  S  D  I  D  T  C  P  F  H  L  N  I  1397
      antiserum                                                                                                        ↑ CHO

4321 AGTGCTGTAGAAGTCCCCATCAGATGAAGTCAGTGGTCAGAATGTGAAGGGGCTGATGATAACAGTCTGCACAAGCTACAATGGTGATGGAACAACCCATATGATCATTGTTGACGTAGGA 4440
1398  S  A  V  E  V  P  S  D  E  V  S  G  Q  N  V  K  G  L  M  I  T  V  C  T  S  Y  N  G  D  G  T  T  H  M  I  I  V  D  V  G  1437

4441 TTGTACTCTGGATTTAAGGCTGTTGAAGAGGGACTTACAGGGCTCAAACAGGTGGTGAACGGCAGTAGCTTAATTTCTTCGTATGAAGCCAGCTCACGCTCTGTTATCTTTTATTTAGAT 4560
1438  L  Y  S  G  F  K  A  V  E  E  G  L  T  G  L  K  Q  V  V  N  G  S  S  L  I  S  S  Y  E  A  S  S  R  S  V  I  F  Y  L  D  1477
                                          ↑ CHO

4561 ACGATTCCGAGTGATGAAGATCTGTGTTTTTACTTTTTCTGCCGAATCTGACGTGGTTGTAGGAAATGTGCAAGCAGCTGCTGTACATGTCTACGACTATTATGACCCAGAGAAATCTTGC 4680
1478  T  I  P  S  D  E  D  L  C  F  T  F  S  A  E  S  D  V  V  V  G  N  V  Q  A  A  A  V  H  V  Y  D  Y  Y  D  P  E  K  S  C  1517

4681 ACGATATTCTACAAGCCTGGTGATGGCAGTGCCCTCTTGTCGACGCTTTGCTCTGAAAATGAATGTATCTGCTCTGGTGGATCCTTAGAATACTGCAACATTGATCCATGTCCTGGGCCA 4800
1518  T  I  F  Y  K  P  G  D  G  S  A  L  L  S  T  L  C  S  E  N  E  C  I  C  S  G  G  S  L  E  Y  C  N  I  D  P  C  P  G  P  1557

4801 TATACAAGAATTGACCTGGAAGGAACTGCCTGTGCAAGCCACAGTAGTTATGCACTCAAGATCCGGATCGACGAGGTAGAAATAAAAGAAGGCTTCAGGATCTGCAAATTTACCGTCTTG 4920
1558  Y  T  R  I  D  L  E  G  T  A  C  A  S  H  S  S  Y  A  L  K  I  R  I  D  E  V  E  I  K  E  G  F  R  I  C  K  F  T  V  L  1597

4921 AACCCGATAAAAACGGGAGATGAAGATGTACCCCATCAGGCACAGAGACAATTGTTCATCAATGAAGGTTGTGACTGTCCTAAGGTCAAGGGTCGTAATATTGGTGGAACCTTTTTGCTT 5040
1598  N  P  I  K  T  G  D  E  D  V  P  H  Q  A  Q  R  Q  L  F  I  N  E  G  C  D  C  P  K  V  K  G  R  N  I  G  G  T  F  L  L  1637

5041 GTTGGCCAGAAGTCACTCAAGTATACCACTGAACAAGGAGAGGAAAGATACCGATACGTGTACGGCCCAACTTCTAAGCTGGAATTTTGGCCAGCAACAAGAAAAGCTAGAAATGCAGCG 5160
1638  V  G  Q  K  S  L  K  Y  T  T  E  Q  G  E  E  R  Y  R  Y  V  Y  G  P  T  S  K  L  E  F  W  P  A  T  R  K  A  R  N  A  A  1677

5161 GTATTTGATAAACTGGTAGCATTTGAAGCCCAAATGGCACCTGCAAGCGCTTGTTCCAATGAGAATTAAGTCATCAGTGAAAAGAAAGAGCAAATCATGAGCTTCCTTTTTTTCTTCTTT 5280
1678  V  F  D  K  L  V  A  F  E  A  Q  M  A  P  A  S  A  C  S  N  E  N  *  1694

5281 TTTTTTTTGTCGATGTAGAGGTCGATGTAGAGGTGTCTTGGTCTAGTGGATATTAAAGTCACCAAACTTTGAATTGCATGGTTCACAGTTCAAGTCCCGCCTTGGCACTAATGTCCTTTG 5400
5401 GCCAGACATTTATCTACATTTGCCACTTCTACCCTAGTTGATGTAAATGGGTACCCGGTAGAATTTTTTCTTTATATGCTTGAGCACCTTAGCATCGTGGTGGCTCAGCTACAGCCGGGG 5520
5521 TAATAATGTGATACTTTGTATTAAGGGCAGTAGAGTATATTCATATACATGTAGTAGCTGTTCTATATAAATGGACATGTCATTATTATTTTCTTAATTATTATGTTCCATGCCATTATG 5640
5641 ACTTTACTCATGCAGATTCGTAAACATTAGTTGGTAGATTATATACTTGTTTTGTATAGTAAGGCCACAGTCAGATGTGAGATGGTCAACCACGACATTTGTTTAAAAAAGGACATATAA 5760
5761 AAGGTATTACGTTTCACATTTTCGTATCTTACAACACAGAGGTCAGTGAATATACAACATGTAGGATCTTTATTGCGTGCAGTTGTCATTAATGTATATTTTAATAAACTGTGGCAGCTT 5880
5881 TGTTGAGTTGGTTTACCATGGCATATCTGAACAGGCCTTAAGAAATTATGTGTAACTCTGACGACAATCACTTCAAGAGGAAATATTGAATTCTTCTTATCTTTTTTCGTGATAAAAAGG 6000
6001 ACAAAGAGAAACAAAATAAGGAACGAGCATACAGGTGTACATGTGTTTGAAGGACTTAGTACATATTGATTCATTTTTATTAATCAAAGTAGAAAGAGACTAGGCAGTGAAATGTTTGTT 6120
6121 TCATCGTACAGACATACATCAGTTTTGGAGGTTTGAGTTTTTACACTTGTTTGACATGTATCTCACAAACCTCTCCATTAGAACCTCCTTTCAATATTACAAAAATATGCAAGAAAATTA 6240
6241 AAATTCCTGAAGCACAGGAATACATGTGGAGCTGCAGAGAGCTGCAGATGATTTCAACTGTAACCATGAAACAACTACAAAGTCACATAATTTACCAGTACTTTTGATAATTTCCTGTTT 6360
6361 GACATTAAATCTCCACATTTTTAGCCTTGCATGTATTGCAGAGTACGATAGTATGATGTACATGTATGGAAATAGCATGTATGTAAACCAATAAAATATGCTGTGTTTTACATGTAGTTC 6480
6481 AATCTGTTATGTACTTGTAATTATTGCAGTTCTGACAATGACAATTGTAGATTTTTGGTTGAGGTATAATATTGATAATCGCTTTAATAATCAGGGCCGCTATTCAATTGAATGCAGTAT 6600
6601 TTTACTCAGTTTTCAGCTATAAAGACAAAATACTAAGCCAAATTTGGTATGCAGTTGAAGACTTCGGCTATTAGTATTTTACCTTTAAGACAGCTCCCTAGGAGACTTGGGCCCTATTAT 6720
6721 TTTGGAGGTCTTGATTGGCAGACATGATGCTATACCGTCTTCATTAGCTGTGCACAATTGTATTACGTAGCGCGCATTTAATGGCGACTACAATGCATGGAGTTAATGGCGTATTGTTGTT 6840
6841 GATTGACAATGTGATGTAACAATGCTTATCAGCAAGTACCTTAAAGTCGGCTTACTAAAGTAAAAGTGCTCAGATAAGATCTAAATACGAGTTAAGACATGTTCCATAGCCAAGCAAACTA 6960
6961 CTAAGTAAACTACTGACTTAATCCACCAAATGAATTGAATACGGCCCCTGCCTGGTACTTTGAAAATGTTACTTTTGTATGGATTGTTTCTAATGACGAAAAATATGTATCATGTGTATT 7080
7081 GCAATGCATGTTCGTGTACACAGTATATGGACAATTGTACATGTATTTAGTAGAAGAGGATATAAAGTTTATGATTGCTTTGAGCGATATTTACATGTAGTAGCTACAGTAGGTTAAAAA 7200
7201 ATACTTGTATATGCAACATTGTTATTAGATCTTCTGCACATGGAAGATTCAATCATCTGTTGAGAATTGTTATGCAATGAATAAATACAATACCAGTAGTAGCGCCTTGGCAGAATTACT 7320
7321 GCTGGTTCAGTTGCTTCTCTTTTTACCAGCATTATATGATCACCACATTTATATATCATTACATAAACTTTGTTTTGGTGCACTATACTTCAGAACATACCGTCGGTTGGGTGACAAAAC 7440
7441 CTTGTATCTCAATTTTTTGAGATTTTGATTTTTTTGCATAATTTTGTACCTAATGGCCTACAACACCACCTGATAAATCCACGGAACGCCTCCTTCTCGGTTTTTTCACAAATTCCACAC 7560
                                                                              polyadenylation signal

7561 AAACACCCTTAAACTAAACCTTGTAACTTTAACCCCCAATTGTAGCGGCCG 7611
```

FIGURE 2. (continued)

However, the beginning of the open reading frame was not obtained. The third library screen used the arrayed coelomocyte library constructed by Jonathan Rast and Eric Davidson at the California Institute of Technology (Pasadena, CA). The 32 positive clones obtained from this screen were analyzed by PCR using the Sp6 and T7 primer sites in the vector to determine insert sizes, and the Sp6 primer in combination with an internal primer specific for the 5′ end of pExCell14 to determine the sizes of the 5′ ends. These PCR-amplified fragments were also checked by Southern blots using the same probe as that used to screen the library. Analysis of the 5′ end of pSPORTA22/137, pSPORTG11/211, and pSPORTJ17/96 showed an additional 2 kb 5′ of pExCell14. Sequence analysis of two clones, pSPORTA22/137 and J17/96 (Fig.

1), indicated that the 5′ end of the open reading frame and a short stretch of the 5′ untranslated (UT) region were included.

The complete sequence of Sp064 and the deduced protein are shown in Figure 2. There are 7611 nucleotides (nt) in the overlapping cDNAs, which are missing much of the 5′ UT region and part of the 3′ UT region of the transcript since, by RNA gel blot, the transcript is 9 kb (see below). Because a random primer was used to construct the cDNA library, the poly(A)+ tail is not expected to be identified in any clone. However, a consensus polyadenylation signal, GATAAA, is located 93 nt from the 3′ end (Fig. 2), which suggests that most of the 3′ UT region may be present. The sequence shows 129 nt in the 5′ UT region (based on the most probable start site), 5097 nt in the

```
C3S.URCHIN   ----M-GSIILIFVLAISASLVPG--GVTQDPPATFFVISPNVFRAGVEEKVVVTLIRSP    C3S.URCHIN   PS-SLSSQVTFVVKEYVLPTFSVSIET-PKYILAG---ADSVVSTIIAEYVFGKPV-IGR
C3HAGFISH    ------------------------------------VLVIAPAATSSYDDLAVAILMVDQK    C3HAGFISH    DKPQFNYTSGFKVEEYVLPTFDVSITSEQPYLHVY---DKAFTIHIKAMHIYGKPV-MGR
C3LAMPREY    ------------VLLLMSVGTSV-------TQDPMVLLSVPSVILIGSDVNVLVDHAAST    C3LAMPREY    GAPNINSSAEFDVREYILPTFEVKINPKQRVFHIN---DEEFVVDITANYFNQELV-SGT
C3COBRA      ----MEGMALYLVAALLIGFP-------GSSHGALYTLITPAVLRTDTEEQILVEAHGDS    C3COBRA      HSPEESYTAYFDVREYVLPSFEVRLQPSDKFLYIDG--NKNFHVSITARYLYGKKV-EGV
C3CHICKEN    ----M-GLLLLPLLLGVLLL------HAVPTPAQMVTMVTPAVLRLDTDEKVVLEAPGLS    C3CHICKEN    DSQDQVFSTQFEVKEYVLPSFEVTLDPQEKFLYIDP--AEDFRVTITARYLYGKNL-QGT
C3MOUSE      ----M-GPASGSQLLVLLLLLASS---PLALGIPMYSIITPNVLRLESEETIVLEAHDAQ    C3MOUSE      HAPKQIFSAEFEVKEYVLPSFEVRVEPTETFYYIDD--PNGLEVSIIAKFLYGKNV-DGT
C3HUMAN      ----M-GPTSGPSLLLLLLTHLPL-----ALGSPMYSIITPNILRLESEETMVLEAHDAQ    C3HUMAN      NSPQQVFSTEFEVKEYVLPSFEVIVEPTEKFYYIYN--EKGLEVTITARFLYGKKV-EGT
C4MOUSE      ----M-RLLWGLAWVFSFC--------ASSLQKPRLLLFSPSVVNLGTPLSVGVQLLDAP    C4MOUSE      DGLESNRSTHFEVKKYVLPNFEVKITPWKPYILMVPSNSDEIQLDIQARYIYGKPV-QGV
C4AHUMAN     TRSAPRAASWLEDPREVRSVCLSATFFTLSLQKPRLLLFSPSVVHLGVPLSVGVQLQDVP    C4AHUMAN     DGLESNSSTQFEVKKYVLPNFEVKITPGKPYILTVPGHLDEMQLDIQARYIYGKPV-QGV
C5MOUSE      ----M-GL-WGILCLLIFLD-------KTWGQEQTYVISAPKILRVGSSENVVIQVHGYT    C5MOUSE      KDFTTTGTAYFEIKEYVLPRFSVSIELERTFIGYKN--FKNFEITVKARYFYNKVVPDAE
C5HUMAN      ----M-GLL-GILCFLIFL-------GKTWGQEQTYVISAPKIFRVGASENIVIQVYGYT    C5HUMAN      EDFSTTGTAYFEVKEYVLPHFSVSIEPEYNFIGYKN--FKNFEITIKARYFYNKVVTEAD
                                .     *             .  .                               *  .   *.** * * .                        . *


C3S.URCHIN   PNVQNIEVKVSLMRPGSTVTFSEDTRLVSPGSS--QSFSVVIQAGDLMR----NEGAFQH    C3S.URCHIN   YTFKYGVVDNG----TVQELGTKQGLLGDSGRAEVTLDDLSTDFG----EDWFVRFRGRH
C3HAGFISH    KITEVHVLLVNPH-TGATLDEKK--VKLQWDNKFIAFTKLQVTPKEVEK----WKEDFVR    C3HAGFISH    AYVRYGVKHQS---KRTLLSTSSALARFEQGEAMHTLRQKHILEQY---PDPKLLLGQS
C3LAMPREY    EDVSVVVRAEEFLTKKQLATQTITLTQLDP-----AIATLKLGFDIENPDKTNSASTKHH    C3LAMPREY    AYVRYFLENGD---VPKLVDSS--STTLVAGEGLSILKKEKLLKLF----PNAKDLLAFS
C3COBRA      TPKSLDIFVHDFPRKQKTLFQSR--VDMNQAGSMFVTPTIKVPAKELNK----DSKQNQY    C3COBRA      AFVVFGVKIDD---AKKSIPDSLTRIPIIDGDGEATLKRDTLRSRF----QDLNQLVGHT
C3CHICKEN    APTEANILVQDFPQKRKVLFQVR--KQLNPAEGMMAIATVKVPVKLLPP-----VVGKHF    C3CHICKEN    AFVLFGVVVDD---EKKTIPQSLQRVKVTDGDGQAVLPMAMLRQPF----PSNADALVGHS
C3MOUSE      GDIPVTVTVQDFL-KRQVLTSEK--TVLTGASGHLRSVSIKIPASKEFN----SDKEGHKY    C3MOUSE      AFVIFGVQDGD---KKISLAHSLTRVVIEDGVGDAVLTRKVLMEGVR--PSNADALVGKS
C3HUMAN      GDVPVTVTVHDFPGKKLVLSSEK--TVLTPATNHMGNVTFTIPANREFK---SEKGRNKF    C3HUMAN      AFVIFGIQDGE---QRISLPESLKRIPIEDGSGEVVLSRKVLLDGVQ--NLRAEDLVGKS
C4MOUSE      PGQEVKGSVFLRNPKGGSCSPKKD-FKLSSGDD-FVLLSLEVPLEDVRSCGLFDLRRAPH    C4MOUSE      AYTRFALMDEQG--KRTFLRGLETQAKLVEGRTHISISKDQFQAALDKINIEVRDLEGLR
C4AHUMAN     RGQVVKGSVFLRNPSRNNVPCSPKVDFTLSSERDFALLSLQVPLKDAKSCGLHQLLRGPE    C4AHUMAN     AYVRFGLLDEDG--KKTFFRGLESQTKLVNGQSHISLSKAEFQDALEKLNMGITDLQGLR
C5MOUSE      EAFDATLSLKSYPDKKVTFSSGY--VNLSPENKFQNAALLTLQPNQVPR------EESPV    C5MOUSE      VYAFFGLREDIKDEEKQMMHKATQAAKLVDGVAQISFDSETAVKELS--YNSLEDLNNKY
C5HUMAN      EAFDATISIKSYPDKKFSYSSGH--VHLSSENKFQNSAILTLQPKQLPG------GQNPV    C5HUMAN      VYITFGIREDLKDDQKEMMQTAMQNTMLINGIAQVTFDSETAVKELS--YYSLEDLNNKY
                                                            .                         . .                                  *


C3S.URCHIN   MVLKAES-LNPIYPFEEQT----DILVTLQSGYVFVQTDKPIYTPNQDVMIKVMSLDQ--    C3S.URCHIN   FHANAIVHESATQFSEASINTKAIFVDSPYKFSTKRTVQHFKAGLNLQVKLDLTFANGDV
C3HAGFISH    LMVKWD---GGQHMEIDIP-------LTSRRGLVFAQTDQPIYTPNNDVNIRLFPVTR--    C3HAGFISH    LYVEASVISSDAGEIENSILDDIPIVASPYSIKSKWTVPFFKPGVPVIYKVLVLNPDGSP
C3LAMPREY    VRLVAK---VESKSFNKEITA--HALLSYRSGHVVVQTDKPIYTPDEKVKYRMFPMNRED    C3LAMPREY    LTIKTTVLSSQAAETEEAELVGIKIVESRYQITATKTSRYFKPELPYFIQVEVRNADGSP
C3COBRA      VVVKVT---GPQVALEKVV------LLSYQSGFVFIQTDKGIYTPGSPVRYRVFSVDH--    C3COBRA      LYVSVTVITESGSDMVVTEQGGIHIVTSPYQIYFTKTPKYFKPGMPYELTVYVTNPDGSP
C3CHICKEN    VSVVAR---VGQVTLEKVL------LVSLQSGHIFLQTDKPIYTPGSTVLSRLFALSH--    C3CHICKEN    LYVTVTVLTESGSDMVEAQRSGIRIVTSPYTIHFTHTPKYFKPGMPFDPTVYVTNPDNSP
C3MOUSE      VTVVAN---FGETVVEKAV------MVSFQSGYLFIQTDKTIYTPGSTVLYRIFTVDN--    C3MOUSE      LYVSVTVILHSGSDMVAERSGIPIVTSPYQIHFTKTPKFFKPAMPFDLMVFVTNPDGSP
C3HUMAN      VTVQAT---FGTQVVEKVV------LVSLQSGYLFIQTDKTIYTPGSTVLYRIFTVNH--    C3HUMAN      LYVSATVILHSGSDMVQAERSGIPIVTSPYQIHFTKTPKYFKPGMPFDLMVFVTNPDGSP
C4MOUSE      IQLVAQSPWLRNTAFKATETQGVNLLYSSRRGHIFVQTDQPIYNPGQRVRYRVFALDQ--    C4MOUSE      LYAATAVIESPGGEMEEAELTSWRFVSSAFSLDLSRTKRHLVPGAHFLLQALVQEMSGSE
C4AHUMAN     VQLVAHSPWLKDSLSRTTNIQGINLLFSSRRGHLFLQTDQPIYNPGQRVRYRVFALDQ--    C4AHUMAN     LYVAAAIIESPGGEMEEAELTSWYFVSSPFSLDLSKTKRHLVPGAPFLLQALVREMSGSP
C5MOUSE      SHVYLE---VVSKHFSKSK----KIPITYNNGILFIHTDKPVYTPDQSVKIRVYSLGD--    C5MOUSE      LYIAVTVTESSGGFSEEAEIPGVKYVLSPYTLNLVATPLFVKPGIPFSIKAQVKDSLEQA
C5HUMAN      SYVYLE---VVSKHFSKSK----RMPITYDNGFLFIHTDKPVYTPDQSVKVRVYSLND--    C5HUMAN      LYIAVTVIESTGGFSEEAEIPGIKYVLSPYKLNLVATPLFLKPGIPYPIKVQVKDSLDQL
                           . .  *  .**. .*.*    *   .                           .          .       * *.     .                 .


C3S.URCHIN   -DMLPSDRELHVEIMDPSGTSVKRLRETHADVPASIGRLQVHNHPASVN-GPFRLSMD--    C3S.URCHIN   --APNVPVNVIATATLRDGSEELLRRPDIAGGQEEISNTDDQGGVSMVLN--VQSSITSI
C3HAGFISH    -QLNPILSSLVVDIMNPDGVVVD-RIEKNAFEVEKVMELRPFHVPAITSLGDWKIVSWMK    C3HAGFISH    --ASGVPIKVSFSFDS--SGNWITQK----------RKTMDNGIAMQTIN--TARNSKKL
C3LAMPREY    VHRIPVRQSMTVDIVNADGVIVERQIKTIKATDEGIVDGTSFTIPAISKHGTWKIFARMS    C3LAMPREY    SKEVDVVAKVQVG-----SATINPQK----------MRTDSNG---LTSFTVTPPNVNQL
C3COBRA      -NMHRMDKTVIVEFQTPEGIVVS-SKPVNPSG--SIRP---YNLPELVSFGTWKAVAKYE    C3COBRA      --AAHVPVVSEAIH----SEGTTLS---------------DGTAKLILN--TPLNIQSL
C3CHICKEN    -FMQPLLKTVIVEVKTPDNVIIK-QVPVSSPMRNGIFS-INHNLPEVVSLGTWTITAKFE    C3CHICKEN    -AAAGIPVKADNF-----QGLVSTQ--------------RDGTAKLVLN--MPANKNSV
C3MOUSE      -NLLPVGRTVVILIETPDGIPVK-RDILSSNNQHGILP-LSWNIPELVNMGQWKIRAFYE    C3MOUSE      --ASKVLVVTQGS-----NAKALTQ--------------DDGVAKLSIN--TPNSRQPL
C3HUMAN      -KLLPVGRTVMVNIENPEGIPVK-QDSLSSQNQLGVLP-LSWDIPELVNMGQWKIRAYYE    C3HUMAN      --AYRVPVAVQGED----TVQSLTQ--------------GDGVAKLSIN--THPSQKPL
C4MOUSE      -KMRPSTDFLTITVENSHGLRVLKKEIFTSTS--IFQD--AFTIPDISEPGTWKISARFS    C4MOUSE      --ASNVPVKVSATLVS--GSDSQVLD----------IQQSTNGIGQVSISFPIPPTVTEL
C4AHUMAN     -KMRPSTDTITVMVENSHGLRVRKKEVYMPSS--IFQD--DFVIPDISEPGTWKISARFS    C4AHUMAN     --ASGIPVKVSATVSSP-GSVPEAQD----------IQQNTDGSGQVSIPIIIPQTISEL
C5MOUSE      -DLKPAKRETVLTFIDPEGSEVDIVEENDYTGIISFPD---FKIPSNPKYGVWTIKANYK    C5MOUSE      --VGGVPVTLMAQTVDV-NQETSDLET------KRSITHDTDGVAVFVLN--LPSNVTVL
C5HUMAN      -DLKPAKRETVLTFIDPEGSEVDMVEEIDHIGIISFPD---FKIPSNPRYGMWTIKAKYK    C5HUMAN      --VGGVPVILNAQTIDV-NQETSDLDP------SKSVTRVDDGVASFVLN--LPSGVTVL
                     .        .            .          .*   .  *                              .                                    *           .
```
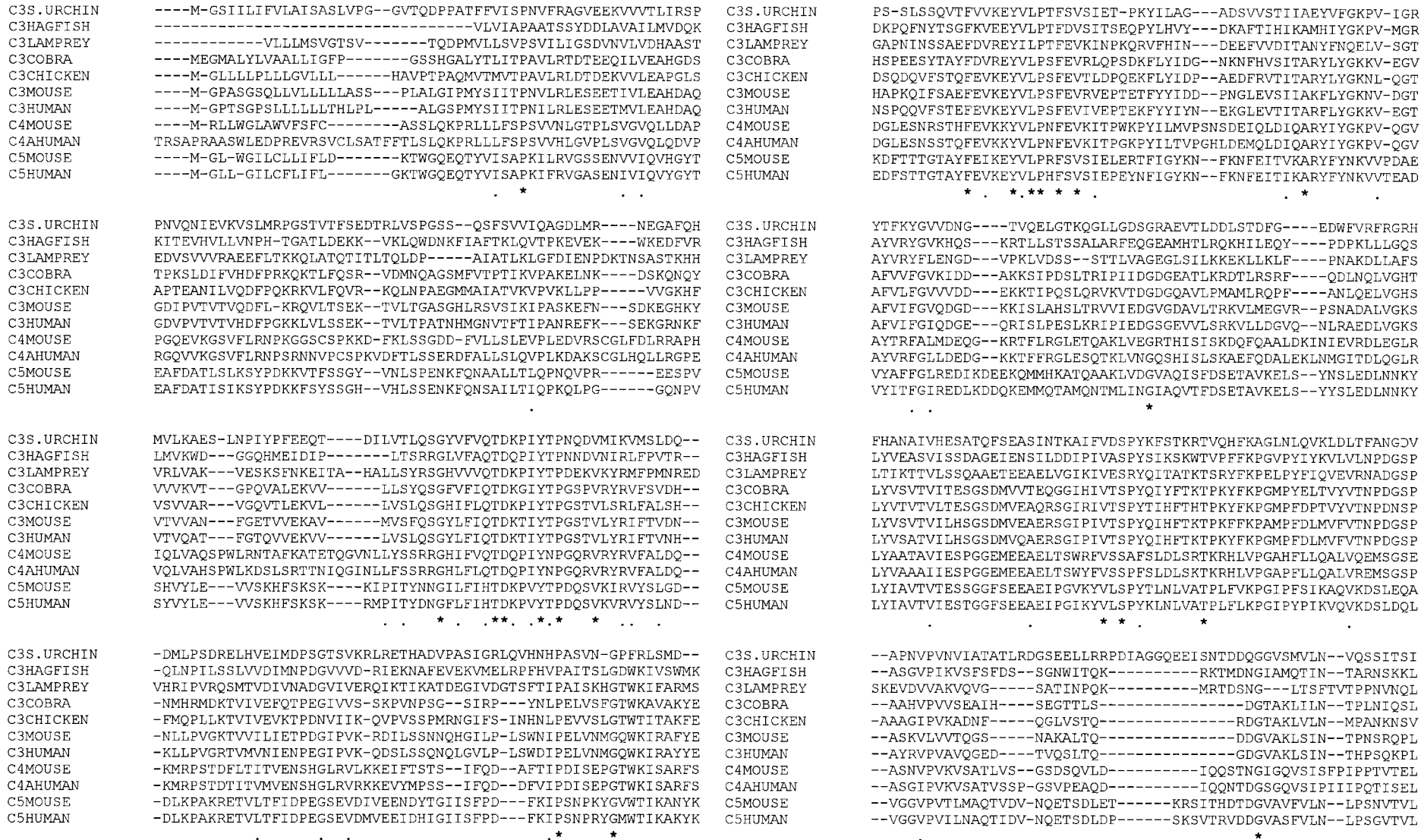
FIGURE 3. Alignment between SpC3 and other complement components. SpC3 was aligned to other thioester family complement components using Clustal W (21) with a few adjustments made by hand. Accession numbers for sequences used in this alignment can be found in Table I. Several areas of functional significance are labeled. Boxed areas have been taken from Figure 4 in Lambris et al. (36) and correspond to functional sites in human C3. C3aR, C3a receptor; Bf, factor B; H, factor H.

```
                                β chain ←        → α chain
C3S.URCHIN   AGLDVPIREDVECMDED-TRRKRS--IDRDQLCLYDPYLADCLADKPRLRRVLTEGGWL
C3HAGFISH    EISGMDDYVKDPGCPNGH-TRRKRE--LVLEIAIEKASTYP-AELRKCCRDAAIESPLRLS
C3LAMPREY    KGLNTTDRSEIGCPKVP-SRKPRQ-LSMLQIR--REAKYTQEFRKCCVDGLKMSPTGQG
C3COBRA      TNLNTKQRSAAKCPQPA-NRRRRSSVLLDSKASKAAQPDQGLRKCCEDGMHENPMGYT
C3CHICKEN    VNINTEQRSEVQCAKPA-KRKRRS-VRLIKHKGTKMAEYSDKNLRKCCEDGIRKNLMGYS
C3MOUSE      QGLQTEQRADLECTKPA-ARRRRS-VQLMERRMDKAGQYTDKGLRKCCEDGMRDIPMRYS
C3HUMAN      SGQTAQRAELQCEQPA-ARRRRS-VQLTEKRMDKVGKYP-KELRKCCEDGMRENPMRFS
C4MOUSE      D-RLTQTREDLSCPKEKKSRQKRN-VNFQKAVSEKLGQYSSPDAKRCCQDGMTKLPMKRT
C4AHUMAN     D-QWTLSRKRLSCPKEKTTRKKRN-VNFQKAINEKLGQYASPTAKRCCQDGVTRLPMMRS
C5MOUSE      ANADDSHYRDDSCKEIL--RSKRNLHLLRQKIEQAAKYKHSVPKKCCYDGARVN-FYET
C5HUMAN      ANADDSQENDEPCKEIL--RPRRT--LQKKIE-ELAAKYKHSVVKKCCYDGACVN-NDET
                            *        *   *

                      Cysteines in C3a fragment↑↑
```

C3 convertase cleavage site (human)↓   ↓Binding sites

```
C3S.URCHIN   CPRRARVLEA--E-CNLEPDQTQVYRDCCIN-SMNPVTATSRSG|----DGGEQNAA
C3HAGFISH    CEERTKHIHDEGEGCQETFLECCKHVEEELLIAMEEEDEDLGRSQ|----EDFMIQESQ-
C3LAMPREY    CEERLKRVTG-PKECVDAFLQCCKKAEYRKSESLGARTVLRRND|----MELDLMNEDE
C3COBRA      CEKRAKYTVD-GDACKAAFLECCHYIKGIRDENQRESELFLARSD|----EDELFGDDN-
C3CHICKEN    CEKRATYVLD-AKSCTEAFLSCCLYIKGIRD-EREQLQYELARSEV|----DDAFLSDED-
C3MOUSE      CQRRARLITQ-GENCIKAFIDCCNHITKLREQHRRDHVLGLARSE|----EEDIIPEED-
C3HUMAN      CQRTRFISL-GEACKKVFLDCCNYITELRQHARASHIGLARSN|----DEDIIAEEN-
C4MOUSE      CEQRAARVPQ--QACREPFLSCCKFAEDLRR-NQTRSQAHLARNNHNMLQEDLIDEDD-
C4AHUMAN     CEQRAARVQQ--PDCREPFLSCCQFAESLRKKSRDKQQAGLQRALEILQ-EEDLIDEDD-
C5MOUSE      CEERVARVTI-GPLCIRAFNECCTIANKIRK-ESPHKFVQLGRIHI|----KTLLFVMK-
C5HUMAN      CEQRAAARISL-GPRCIKAFTECCVVASQLRA-NISHKDMQLGRLHM|----KTLLPVSK-
          *  *            *    **                   *

                                             C3aR
Cysteines in C3a fragment↑↑
```

for Bf, CR1, CR3, H ↓

```
C3S.URCHIN   VRVRDDFRETWFFDVVSMPEDGS-----PYI|YPVSIPSSITDWHLTAVSLSPTQGMCVED
C3HAGFISH    VVIRSHFPESFMWEIIKLSRSAEN-G--KSFITKKMPDSITTWDIQAVEVSQSKGLCVGP
C3LAMPREY    VNMMAYFPQSWGWNKYKNSCKYGR---HPQIRLQLPDTITTWMQAVSISKTRGVCLAD
C3COBRA      IISRSDFPESWLWLIELTGEPNNQGISSKTVPFYLRDSITTWELLAVGLSPTKGICVAE
C3CHICKEN    ITSRSLFPESWLWQVEELTEPPNEQGISMKILPIYLKDSITTWEVLAVSISENKGLCVAD
C3MOUSE      ILSRSHFPQSWIWTIEELKEPKN-GISTKVMNIFLKDSITTWEILAVSLDKKGICVAD
C3HUMAN      IVSRSEFPESWLWNVEDLKEPPKN-GISTKIMNIFLKDSITTWEILAVSMSDKKGICVAD
C4MOUSE      ILVRTSFPENWLWRVEPVDS------SKILTVWLPDSMTTWEIHGVSLSKSKGLCVAK
C4AHUMAN     IPVRSFFPENWLWRVETVDR------FQILTIWLPDSLTTWEIQGIGISDN-GICVAD
C5MOUSE      ADIRSYFPESWLWEIHRVPK------RKQLQVTLPDSLTTWEIQGISDN-GICVAD
C5HUMAN      PEIRSYFPESWLWEVHLVPR------RKQLQFALPDSLTTWEIQGIGISNT-GICVAD
             *  *          *        **                  .*  *.
```

forms disulfide bond with β chain (human)↑

```
C3S.URCHIN   HIVAATHDPQYP-NNQADISFEVSPGVSANGHDYLVIRPQNTDDRNLNVDVSTDFLIQRI
C3HAGFISH    NIKVQTEDERIEQSQQAEEASFTIASYSSP--SGSFIHLNAHREVKSPGEHIVFDVFIKSA
C3LAMPREY    TVTVRTDER-HPSNEQGELVTPAQKYAS----ASYMHIDV-TRIMRLGETLNVFLITAKTT
C3COBRA      PITVRTNHGDLPRERQAIKSMTATAYCTQGGSENYLHVAITSTEIKPGDNLPVNFNVRGN
C3CHICKEN    PITVRTDQKDLPPERQASRQIVAEAYQSQGNSGNYLHLAVGASQVQPGDNLPINFHLKSN
C3MOUSE      TITVRTKKDTLPESRQATKTMEAHPYSTMHNSNNYLHLSVSRMELKPGDNLNVNFHLRTD
C3HUMAN      SIIVRTKKQELSEAEQATRTMQALPYSTVGNSNNYLHLSVLRTELRPGETLNVNFLLRMD
C4MOUSE      RLLVSAGSL-YPAIAR--LTVQAPPSRG---TGFLSIEPLDPRSPSVGDTFILNLQPVGI
C4AHUMAN     QLSVSAGSP-HPAIAR--LTVAAPSGG---PGFLSIERDSRPPRVGDTLNLNLRAVGS
C5MOUSE      KFELRTDDPELPEENQASKEYEAVAYSSLSQSYIYIAWTENYKPMLVGEYLNIMVTPKSP
C5HUMAN      EFNVKTDAPDLPEENQAREGYRAIAYSSLSQSYLYIDWTDNHKALLVGEHLNIIVTPKSP
                    .                                               *
```

```
C3S.URCHIN   G-TTGDQDIDHFLCITGGKVVLEGVQRSIS-AAGTNLAITIRAYMAPQMRLIVYVTMD
C3HAGFISH    A---KDHVLHENYLMISNGKIHNFLQEGRK-G-DTTSVSLLLTPELVPQFRLVAFFLLPS
C3LAMPREY    Q---LNAVTHFTYMVLTRGVIVKINRKTKESGGGPSNVRLPITPDMAPRFRFLAYYILPG
C3COBRA      AN-SLNQIKYFTYLIINKGKIFKVGRQPREGDQSLVTMSLPVTANLIPSFRIVAYYHVKP
C3CHICKEN    RDDVRKSVSYFTYLILSKGHIVHVGRQPREGDQIVVLSLPITPEFIPSFRLVAYYTLIG
C3MOUSE      PG-HEAKIRYTYLVMNKGKLLKAGRQVREPGQDLVVLPLSITTDFIPSFRLVAYYTLIG
C3HUMAN      R-AHEAKIRYTYLIMNKGRLLKAGRQVREPGQDLVVLPLSITTDFIPSFRLVAYYHQG
C4MOUSE      P---APTFSHYYMISRGQIMAMGREPRK----TVTSVSVLVDHQLAPSFYFVAFYHQG
C4AHUMAN     G---ATFSHYYMILSRGQIVFMNREPRK---TLTSVSVFVDHHLAPSFYFVAFYHGD
C5MOUSE      Y---IDKITHYNYLILSKGKIVQYGTREKLFSSTYQNINIPVTQNMVPSARLLVYYIVTG
C5HUMAN      Y---IDKITHYNYLILSKGKIIHFGTREKFSDASYQSINIPVTQNMVPSSRLLVYYIVTG
               .            .                *
```

↑forms disulfide bond with α chain (human)

```
C3S.URCHIN   G----SVIADSLLGVEE-KCRQDPQLSLDILPRNVGPERDVYEPNGQIQVEVTAPIDSN
C3HAGFISH    G----ELVADSIIIDVKD-SCHAKLSLDVAGGKR-------LFSSPRDNVNFDLSGESDSW
C3LAMPREY    G----EIVADSVTEVTE-LCMSLQVSLS-LKGRP-----TLEPKAMLTLDLIGEPDAR
C3COBRA      N----EIVADSVWVDVKD-TCMGTLVVKGASSRDD----RIQKPGAAMKIKEGDPGAR
C3CHICKEN    G----EIIADSVWVDVKD-TCMGSLVVR-GASEADN----RVHEPRTPMRLHIEGDHKAH
C3MOUSE      ASGQREVVADSVWVDVKD-SCIGTLVVK-GDPRDN------HLAPGQQTTLRIEGNQGAR
C3HUMAN      ASGQREVVADSVWVDVKD-SCVGSLVVKSGQSEDR------QPVPGQQMTLKIEGDHGAR
C4MOUSE      H----PVANSLLINIQSRDCEGKLELKVDGAKQ------YRNADMMKLRIQTDSKAL
C4AHUMAN     H----PVANSLRVDVQAGACEGKLELSVDGAKQ------YRNGESVKLHLETDSLAL
C5MOUSE      EQT-AELIVADAVWINIEE-KCGNQLQVHLSPDEY------VYSPGTVSLDMVTEADSW
C5HUMAN      EQT-AELVSDSVWLNIEE-KCGNQLQVHLSPDAD------AYSPGQTVSLNMATGMDSW
             .   .                                         .         .
```

```
C3S.URCHIN   VGLLAVDKAVVLLRD--KDRMGKGKQKMYERMR-SYDTGCPGGGQNTAQIFKDCGMTVLTN
C3HAGFISH    VAVGVVDKAAYVLDK--KNKLTANKVYKAME-ASDLGCSVSGSKTGELVFRDAGLAIMAK
C3LAMPREY    VGLLAVDQAVYAVNR--KHRLTQDRVWKAME-TFDTGCTABGGAGRPGVFSDAGLALITS
C3COBRA      VGLVAVDKAVVVLND--KYKISQAKIWDTIE-KSDFGCTAGSGQNNLGVFEDAGLALTTS
C3CHICKEN    VGLVAVDKAVYVLN---KNKLTQSKVWDTVE-NSDIGCTPGSGRNQVFADAGLSLTSN
C3MOUSE      VGLVAVDKGVFVLNK--KNKLTQSKIWDVVE-KADIGCTPGSGKNYAGVFMDAGLAFKTS
C3HUMAN      VVLVAVDKGVFVLNK--KNKLTQSKIWDVVE-KADIGCTPGSGKDYAGVFSDAGLTFTSS
C4MOUSE      VALGAVDTALYAVGGRSHKPLDMSKVFEVIN-SYNVGCGPGGGDDALQVFQDAGLAFSDG
C4AHUMAN     VALGALDTALYAAGSKSHKPLNMGKVFEANN-SYDLGCGPGGGDSALQVFQAAGLAFSDG
C5MOUSE      VALSAVDRAVYKVQG--NAKRAMQRVFQALDEKSDLGCGAGGGHDNADVFHLAGLTFLTN
C5HUMAN      VALAAVDSAVYGVQR--GAKKPLERVEQFLE-KSDLGCGAGGGLNNANVFHLAGLTFLTN
             *   *       *.    . . .    *. *      *
```

Figure 3. (continued)

Histidine

```
C3S.URCHIN  ANRFTR-VEPGHVEGSINWLIDNNQLPSGAFQESQQVIHQEMIGAVKGE-----TSMTAF
C3HAGFISH   ARQLVF-IPVSELCGSVRWLMR-KQDKDGSFLESKPVVHLNMMGQVTGK----VVLTSF
C3LAMPREY   AKRVI-VDNQELCGPVFMIIKNRQNSDGSYREDGPVIHREMGGGVGGT-EGHVSMTAF
C3COBRA     ASNMVKDISHEIICGGVKWLILNRQQPDGVFKENAPVIHGEMLGGTKGA-EPEASLTAF
C3CHICKEN   AINMVD-IKPEVVCGAIKWLILEKQQPDGLFQEDAPVIHKEMVGGYHGA-EPSVSLTAF
C3MOUSE     AANLIA-IDSHVLCGAVKWLILEKQKPDGVFQEDGPVIHQEMIGGFRNAK-EADVSLTAF
C3HUMAN     AVNLIA-IDSQVLCGAVKWIILERQKPDGVFQEDAPVIHQEMIGGLRNN--EKDMALTAF
C4MOUSE     AQEQVG-NSPEKLQETASWLLA-QGLGDGSHDPCPVIHRAMQGGLVGS--DETVALTAF
C4AHUMAN    AQEQVG-GSPEKIQETSNWLLS-QQQADGSFQDPCPVLDRSMQGGLVGN--DETVALTAF
C5MOUSE     VAKYVK-QDENSICNSLLWLVEKCQLENGSFKENSQYLPIKLQGTLPAEAQEKTLYLTAF
C5HUMAN     VNKYVE-QNQNSICNSLLWIVENYQIDNGSFKENSQYQPIKLQGTLPVRARENSLYLTAF
                              *.  .              .         .  . .*.*
```

CR2 binding site

```
C3S.URCHIN  VLISLLESRNLLVP-------ANQKIDEAIGKATEYLVTQVEN-IDRVYDKALVTSALRESA
C3HAGFISH   VFIALLEARESCIN----EVEGFTVVVEKAHGYLITSQAMN-GLEDFPLAITAYALSLWK    SOPFWLQRKP-SA
C3LAMPREY   ILIGIQQAQEYCGV--------SVPNYKQSMNRAVQFLASKVSD-LKRMYTIAITRYALALQD  SNKGKAAA--
C3COBRA     IVTALLESRVSVC-------QINLDDSINKATDYLLKKYEK-LQRPYTTALTAYALAAAD     AEETSHVLRM-SA
C3CHICKEN   VLSALQESQKICKN------YVKSLDGSIAKASDYLSRKYQS-LTRPYTVALTSYALALTG   EEYNARTHN---
C3MOUSE     VLIALQEARDICEG------QVNSLPGSINRAGEYIEASYMN-LQRPYTVAIAGYALALMN   AEFNAHTY---
C3HUMAN     VLISLQEAKDICEE------QVNSLPGSITRAGDFLEANYMN-LQRSYTVAIAGYALAQMG   EEPDQQLYN---
C4MOUSE     VVIALHHGLVIFQDDDAKQLKNRVEASITKANSFLGKASAGLLGAHAAAITAYALTLTK     EDPGKQLYN---
C4AHUMAN    VTIALHHGLAVFQDEGAEPLKQRVEASISKANSFLGEKASAGLLGAHAAAITAYALSLTK
C5MOUSE     SVIGIRKAVDICPT-------MKIHTALDKADSFLLENTLP-SKSTFTLAIVAYALSLGD    KRPDSSVPSSGTA
C5HUMAN     TVIGIRKAFDICPL--------VK-IDTALIKADNFLLENTLP-AQSTFTLAISAYALSLGD  LQHKDSSVPNTGTA
                .   .       .       .          ..    .    .      **
```

Factor I cleavage sites (human)
↓1 ↓2 ↓3

Factor H binding site

```
C3S.URCHIN  VSQGVMQRKS--RSLTLNPGRVMFSDDVTTPSPNNSLGSGPGTIFGEGFQHEEIAISLPG
C3HAGFISH   MPEGNYEEMS--RSWSVQPRRHG--GQQVVENEFPQN------------------VVPG
C3LAMPREY   APEGVRDIRSESRSVHVEER-------ETFFIKNEISPD---------------VVPN
C3COBRA     VPEGERKN--IVTIIELDPSVKGVGGTQELTVIANKLDD--------------KVPD
C3CHICKEN   VPEGMRLEKT-VKIVELDPKTLGNNGVQEVKVKAANLSD--------------IVPN
C3MOUSE     VPEGMRINKT-VAIHTLDPEKLGQKVQVQVKPVPAADLSD-------------QVPD
C3HUMAN     VPEGIRMNKT-VAVRTLDPERLGREGVQKEDIPPADLSD--------------QVPD
C4MOUSE     EKEGAIHREE--LVYNLDPLNN---LGRTLEIPGSSDPN-------------IVPD
C4AHUMAN    EKEGAIHREE--LVYELNPLDHR---GRTLEIPGNSDPN------------MIPD
C5MOUSE     VPEGVKRES--YAGVLLDPKGIRGIVNRREFPYRIPLD--------------LVPL
C5HUMAN     VPEGVKRES--YSGVTLDPRGIYGTISRRKEFPYRIPLD------------LVPK
                *                                                  .*
```

Thioester site

```
C3S.URCHIN  SSIPDTESCSVKLIGNLLGTASTDPIGGLDHLVRQPRGCGEQTMIYLAPTLFVVQYLIAV
C3HAGFISH   TEMSAFLSAQGNIVAETI--QNTIKGSKISNLLRLPRGCGEQNMMYTSITVMVARYLNRS
C3LAMPREY   SDVLTFISVKGDELAETM--VNCLDAKSISNLIQIPTGCGEQNMIKMAPTTLTLIYLDSV
C3COBRA     TEVETRISVLGDPVAQII--ENSIDGSKLNHLIITPSGCGEQNMITMTPSVIATYYLDAT
C3CHICKEN   TESETKVSIQGNPVSIIV--EKATDGTKLKHLIVTPAGCGEQNMIGMTPTVIAVHYLDST
C3MOUSE     TDSETRIILQGSPVVQMA--EDAVDGERLKHLIVTPAGCGEQNMIGMPTVIAVHYLDQT
C3HUMAN     TESETRIILQGTPVAQMT--EDAVDAERLKHLIVTPSGCGEQNMIGMTPTVIAVHYLDET
C4MOUSE     GDFSSLVRVTASEPLETMGSEGALSPGGVASLLRLPQGCAEQTMIYLAPTLAASRYLDRT
C4AHUMAN    GDFNSYVRVTASDPLDTLGSEGALSPGGVASLLRLPRGCGEQTMIYLAPTLAASRYLDKT
C5MOUSE     TKVERILSVKGLLVGEFL--STVLSKEGINILTHLPKGSAEAELMSIAPVFYFYPHYLEAG
C5HUMAN     TEIKRILSVKGLLVGEIL--SAVLSQEGINILTHLPKGSAEAELMSVVPVFYFYPHYLETG
                                           .  *. *   .    *  **
```

```
C3S.URCHIN  G---SDTAEQEAR---IYDYIADGVARELTYRQDNGAYAAWKHRPGSTWLTAFVVKVFSQ    IEIETSGYALIAQLALL-DYQKAGKIALWLSKQQNDGGGFVSFQDTVVALQALAKYTE-R
C3HAGFISH   DQWNKMGDPQLKKR--SFDFITSGFASQLTYRKPDYSYAAWLHRASSTWLTAFVAKVFSQ   --VESTAYGLLAAIQHE-EGEIAEKATNWLSQSATFGGYFQSTQDTVMALQALTGFESCQ
C3LAMPREY   QEWEKIGLHREE---AIGFLKQGYSRELSYRKADHSYAAFIKRPSSTWLTAFVVKVYSL    ISVEATAYGLITYLRKK-DYESAREIVDWLTEQRNYGGGFQSTQDTILALQAMAQYKMDS
C3COBRA     GQWENLGVDRRTE---AIKQIMTGYAQQMVYKKADHSYAAFTNRASSSWLTAYVVKVLAM   N--IEGTSYALLALLQME-KFAEVGPVVRWLLIDQKYIGGTYGQTQATVMVFQALAEYEIQM
C3CHICKEN   MQWETFGINRRTE---AIELIKKGYTQQLAYRKEDGSFAAFITRPSSTWLTAYAVKVFAM   --VEATSYALLALLLLK-DFDSVPPVVRWLNEQRYGGGYGSTQATFMVFQALAQYTDV
C3MOUSE     EQWEKFGIEKRQE---ALELIKKGYTQQLAFRQPSSAYAAFNRPPSTWLTAYVVKVFSL    --VEATSYALLALLQLK-DFDFVPPVVRWLNEQRYGGGYGSTQATFMVFQALAQYQKDA
C3HUMAN     EQWEKFGLEKRQG---ALELIKKGYTQQLAFRQPSSAFAAFVKRAPSTWLTAYVVKVFSL   LWIETTAYALLHLLREGKGHADKAASWLTRQGSFHGAFRSTQDTVTILDALSAYWIAS
C4MOUSE     EQWSKLSPETKDH---AVDLIQKGYMRIQQFRKNDGSFGAWLHRDSSTWLTAFVLKILSL   LWIETTAYALLHLLHEGKAEMADQASAWLTRQGSFQGGFRSTQDTVIALDALSAYWIAS
C4AHUMAN    EQWSTLPPETKDH---AVDLIQKGYMRIQQFRKADGSYAAWLSRDSSTWLTAFVLKVYSL   GMVEATAYALLASLKLK-DMNYANPIIKWLSEEQRYGGGFYSTQDTINAIEGLTEYSLLL
C5MOUSE     NHWNIFYPDTISKRQSLEKKIKQGVVVSMSYRNADYSYSMWKGASASTWLTAFALRVLGQ   RMVETTAYALLTSINLK-DINYVNPVIKWLSEEQRYGGGFYSTQDTINAIEGLTEYSLLV
C5HUMAN     NHWNIFHSDPLIEKQKLKKKLKEGMLSIMSYRNADYSYSVWKGGSASTWLTAFALRVLGQ
                              .  *****.                  .    .       .
```

Figure 3. *(continued)*

Factor I cleavage sites (human)
↓4   ↓5

α chain ←   → γ chain

Properdin binding site

Species labels (for each alignment block):
C3S.URCHIN
C3HAGFISH
C3LAMPREY
C3COBRA
C3CHICKEN
C3MOUSE
C3HUMAN
C4MOUSE
C4AHUMAN
C4HUMAN
C5MOUSE
C5HUMAN

Figure 3. (continued)

open reading frame, and 2385 nt in the 3′ UT region. There are three possible start sites near the beginning of the sequence shown in Figure 2. The first ATG is followed by two in-frame stop codons, while the second and third are in-frame and are not followed by stops. Because the third ATG, located at nt 130, is surrounded by a Kozak sequence, ACC<u>ATG</u>G, this suggests that this is the most probable start site for translation. Following the start site, there is a hydrophobic region of 12 amino acids plus a serine, which is followed by a hydrophilic region that includes serine, proline, and glycine (31). This combination of a short hydrophobic and hydrophilic regions is typical of a leader or signal sequence of 13 amino acids and is based on the "[-3,-1]-rule" (32). A leader region is expected to be present since SpC3 is produced in the coelomocytes and appears to be secreted into the coelomic fluid (see Fig. 5; Gross and Smith, unpulished). There are six ATTTA repeats in the 3′ UT region (Fig. 2). These AU-rich repeats are typical of transcripts that encode inducible genes and may function to stabilize the transcript after induction, which results in increased translation (33).

The open reading frame encodes a protein of 1699 amino acids. The cleavage site for processing SpC3 into two chains before being secreted is located at Arg686 and consists of RRKR (Fig. 2). The β-chain has 672 amino acids (without the leader), and the α-chain has 1010 amino acids after removing the RRKR at the $\beta\alpha$ junction. The deduced m.w. for the preprocessed SpC3 is 186 kDa, and after processing the deduced α-chain is predicted to be 110 kDa and the β-chain 73.5 kDa. These predicted m.w. do not take into consideration the possibility of N-linked glycosylation, which is known for human C3. There are five consensus recognition sequences for N-linked glycosylation in SpC3, four of which are located in the α-chain (Fig. 2); however, none are conserved in position compared with glycosylation sites in human C3 (34). The deduced isoelectric point for SpC3, as calculated by DNASIS, is 4.76. This is substantially lower than the range of isoelectric points 5.87 to 7.99, which was calculated for other C3, C4, and C5 proteins that are listed in Table I.

SpC3 protein alignments to other thioester family proteins

Regions of the deduced SpC3 protein that are similar to other complement proteins can best be estimated through amino acid alignments. The BLAST search provided a list of proteins to which SpC3 matched best, and we chose a subset of these to construct an alignment that included some of the complement proteins (Fig. 3). Inspection of Figure 3 reveals a number of regions in SpC3 that are conserved when compared with the other complement sequences and other regions that are not conserved. All proteins in Figure 3 show a $\beta\alpha$ junction, including SpC3, whereas only the C4 proteins and the cyclostome C3 components show an $\alpha\gamma$ junction. A conserved thioester site (GCGEQ) is located in the SpC3 α-chain, identical to that seen in vertebrate C3 and C4 proteins, but that is not present in C5. The histidine involved in thioester binding to hydroxyl groups (35), located about 100 amino acids (H1090 in Fig. 2) toward the C terminus, is conserved in SpC3 (Fig. 3). The hydrophobic region surrounding the thioester site is also conserved in SpC3, the function of which is thought to shield the thioester from the aqueous environment and nucleophilic attack (36). In vitro mutagenesis experiments have shown that the two prolines surrounding the thioester in human C3 are necessary for stable formation of the activated thioester (37), and these positions are conserved in SpC3 (Fig. 3). This analysis of SpC3 compared with vertebrate complement components indicates that it is a two-chain structure with a conserved thioester site similar to other C3 proteins.

Complement proteins are folded and held together by the disulfide bonds. In human C3, there are 27 cysteines that align with the other cysteines in the thioester complement protein family. However, SpC3 has 32 cysteines, and we were interested to understand how this amino acid aligned with cysteines in the other members of the thioester protein family. Alignments between SpC3 and the other complement components using Clustal W (Fig. 3) showed that five of the SpC3 cysteines are located in the β-chain, three of which align with the three conserved cysteines in the β-chain of all of the other components. In the α-chain, 18 of the 27 cysteines align with conserved cysteines in other components. The two cysteines involved in interchain disulfide bonding in the human sequence (38) are aligned between the sea urchin sequence and all of the other proteins (Fig. 3). To understand more fully whether the nonaligned cysteines in SpC3 were due to differences between the protein sequences or due to errors made by the Clustal W algorithm, additional alignments were performed to inspect the positioning of the cysteines in SpC3. Both Clustal W and DNASIS were used to align the entire thioester protein family and a subset that included just the complement proteins. The Clustal W alignment of the complement protein subset is shown in Figure 3, and the Clustal W alignment for the entire thioester family can be obtained by E-mail (see legend to Fig. 4; DNASIS alignments are not shown). Results of these four alignments revealed similar, but not identical, positioning of the cysteines in SpC3 relative to the other proteins. Most (22 to 24) of the SpC3 cysteines align with or near cysteines that are generally conserved in the complement protein family. There are four to five positions in which cysteines are missing in SpC3, where the other proteins showed a conserved cysteine, and SpC3 has seven to nine extra cysteines that do not align with other proteins. In no case did any of the extra cysteines in SpC3 align with cysteines in the α_2 M group rather than the complement group. Interpretation of these data may suggest that SpC3 has a similar, but perhaps not identical folding pattern compared with the other complement proteins.

The human C3 protein is the most fully characterized of all of the complement proteins, and functional regions have been mapped using a variety of techniques (reviewed in Ref. 10). These functionally mapped regions have been taken from Figure 4 in Lambris et al. (36), and are indicated in Figure 3 for comparisons between SpC3 and human C3. There are nine matches between SpC3 and human C3 in a span of 43 amino acids (20.9%) within the region in which human C3 interacts with the C3a receptor, CR1 and CR3, factor H and factor B. This region is generally not well conserved in vertebrate C3 proteins (8), except for the C3 convertase cleavage site (Arg, Ser), which is conserved in most of the C3 proteins including SpC3. The second factor H binding site located farther down in the α-chain, within which is located the CR2 binding site, reveals 18 amino acids in SpC3 that match to human C3 in a span of 76 (23.7%). Eleven of thirty-two amino acids (34.3%) match between SpC3 and human C3 within the properdin binding site that is known as a region of high conservation (8). There are five factor I cleavage sites labeled in Figure 3 that are positioned relative to the human C3 sequence (10). Of those sites, none match perfectly to SpC3; however, there are two Arg/Ser sequences in the SpC3 that align near factor I sites 1 and 5 (Fig. 3). In general, the alignment with other complement proteins and the comparison with functional sites in human C3 suggest that the sea urchin protein may have a different folding pattern and fewer functions, and may not interact with as many other proteins as is known for human C3.

Table I. *Amino acid comparisons between SpC3 and other thioester family proteins[a]*

Protein, Species	Accession No.	% Identity	% Similarity Plus Identity
C3, hagfish	gb|Z11595	26.4	43.3
C3, lamprey	sp|Q00685	25.5	43.1
C3, trout	pir|I51339	25.1	42.7
C3, cobra	sp|Q01833	24.8	43.0
C3, chicken	pir|I50711	26.7	44.4
C3, guinea pig	sp|P12387	27.5	41.1
C3, rat	gb|X52477	26.8	43.9
C3, mouse	sp|P01027	26.9	41.7
C3, human	sp|P01024	27.9	45.2
C4, xenopus	gb|D78003	26.7	44.1
C4, mouse	sp|P01029	25.5	44.8
SLP, mouse	gb|M21576	26.6	44.4
C4A, human	gb|K02403	25.9	44.9
C4B, human	gb|U24578	25.7	43.1
C5, mouse	sp|P06684	24.1	42.8
C5, human	sp|P01031	23.1	43.9
α2M, limulus	gb|D83196	21.3	36.6
α2M, lamprey	gb|D13567	21.1	37.1
α2M, guinea pig	pir|JC5143	20.4	37.6
α1M, rat	gb|M77183	21.3	37.4
α1 inhibitor, rat	sp|P14046	21.4	37.3
α2M, rat	sp|P06238	20.7	37.4
α2M, mouse	sp|Q61838	20.8	37.1
α2M, human	sp|P01023	20.8	39.0
PZP, human	gb|X54380	20.5	39.2

[a] The Clustal W alignment program (21) was used to align SpC3 to each of the thioester proteins listed. The percent identical and percent similar plus identical amino acids were calculated based on the SpC3 length of 1699 amino acids. sp, Swiss Protein database; pir, PIR database; gb, GenBank database.

Phylogenetic relationships between SpC3 and other thioester family proteins

Our first approach to understanding the phylogenetic relationships between SpC3 and other thioester family proteins was to generate pairwise alignments between SpC3 and 25 sequences that included several C3, C4, C5, and α_2 M proteins from a number of different species. We then calculated the percentages of amino acids that showed identical matches and the percentages that were identical plus similar between the two proteins. The results, shown in Table I, indicate that SpC3 showed greater similarity to the complement proteins (23.1–27.9% identical; 41.1–44.9% similar) than to the α_2 M family (20.4–21.4% identical; 36.6–39.2% similar) in general agreement with our previous report (15). However, these results could have been due to the shorter length of α_2 M proteins and not due to fewer amino acid identities and similarities between SpC3 and the α_2 M proteins. Consequently, we generated a large alignment using all of the sequences listed in Table I with the Clustal W program (21) (available by E-mail, see legend to Fig. 4). The alignment was inspected at each amino acid position, and SpC3 was scored as identical to each group of proteins based on the number of sequences to which it matched. SpC3 was scored as similar to C3 when it matched to six of nine of the vertebrate C3 sequences; similar to C4 when it matched to four of five of the C4 sequences; similar to C5 when it matched to both of the C5 sequences; and similar to α_2 M when it matched to six of nine of the α_2 M sequences. Results indicated that SpC3 was equally similar to all of the complement components (19%) and was less similar to the α_2 M proteins (15%) for this alignment set. A similar result was found when the matches were calculated for a shorter region that corresponded only to the α_2 M protein length (20% similar to complement compo-
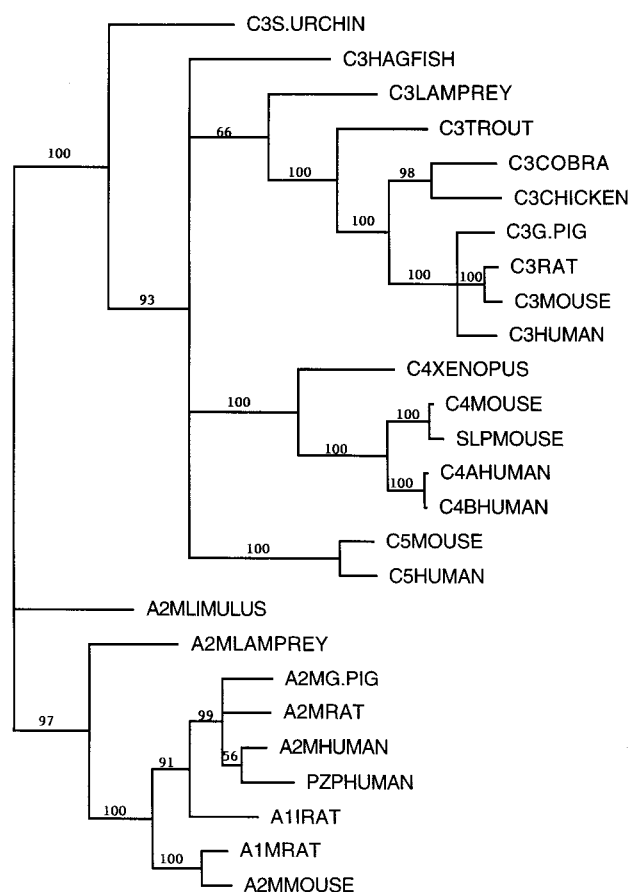
FIGURE 4. Phylogenetic relationships between members of the thioester protein family. The phylogenetic tree was constructed using the PAUP program (22), with the α_2 M protein family defined as the outgroup to root the tree. See Table I for accession numbers for all sequences used in this analysis. The alignment used to generate this tree can be obtained by contacting the EMBL server by E-mail. Send the request to Netserv@ebi.ac.uk and include in the message GET ALIGN: DS31395.DAT. The results from a number of variations for generating trees gave the same minimal tree length of 13614. The tree is plotted as a phylogram. Reliability of branch lengths were analyzed by bootstrapping, and the bootstrap numbers are shown.

nents and 17% similar to the α_2 M group). These alignment analyses suggest that SpC3 is a complement protein and is not an α_2 M protein.

Since the paired alignments (Table I) were not informative as to the relationships between SpC3 and the other complement protein family members, a phylogenetic analysis was done using the PAUP program (22). This program is designed to compare and assemble related sequences into phylogenetic trees that can then be used to infer evolutionary relationships. We used the same alignment between SpC3 and the 25 other thioester proteins listed in Table I in a phylogenetic analysis using PAUP. We were able to repeatedly identify the shortest tree (length = 13614) when we used a number of variations in generating the alignment with Clustal W, which were then used in the PAUP program. Based on the lower percentage identity between SpC3 and the α_2 M group (Table I), these proteins were selected as the outgroup. The phylogenetic tree (Fig. 4) shows that SpC3 is positioned basal to the complement clade that includes C3, C4, and C5 proteins. Furthermore, the branch arrangement within the vertebrate complement clade generally agrees with other published trees (26, 39, 40); however, the internal details of the positioning of some of the proteins differed in our trees depending on the protocols used. For
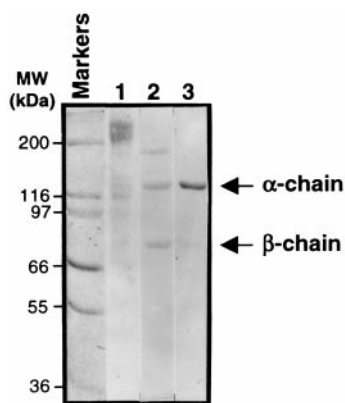
FIGURE 5. Protein gel analysis of SpC3. SpC3 was partially purified from coelomic fluid and analyzed by SDS-PAGE. *Lane 1*, SpC3 under nonreducing conditions; *lane 2*, SpC3 under reducing conditions; and *lane 3*, Western blot of reduced gel analyzed with the anti-peptide antiserum. Protein standards are shown on the *left*.

FIGURE 6. Sp064 gene expression in adult sea urchin tissues. Poly(A)$^+$ RNA (0.4 μg) was electrophoresed through a 0.8% agarose gel containing 2.2 M formaldehyde in 1× MOPS buffer (20 mM 3-[*N*-morpholino] pro-panesulfonic acid, 5 mM sodium acetate, and 1 mM EDTA, pH 7), capillary blotted onto GeneScreen *Plus* (DuPont NEN), and probed with a ^{32}P-labeled PCR-amplified DNA fragment that spanned approximately 500 bp of the 5′ end of clone pExCell14 (see Fig. 1). Filters were hybridized and washed, as described in *Materials and Methods*, for phage library screening. Filters were exposed to BioMax MR-1 x-ray film. Transcript sizes were estimated from RNA standards (Bio-Rad, Hercules, CA). RNA from the major tissue types; C = coelomocyte, O = ovary, T = testis, G = gut. To show that approximately equal amounts of poly(A)$^+$ RNA were loaded onto each lane, the blot was stripped and reprobed with a sea urchin homologue of the human L8 ribosomal gene (EST219, see Ref. 15).

example, the hagfish component clustered with either the C3 or the C4 clade in different analyses. It should be noted that the hagfish C3 position in Figure 4 is not supported by bootstrapping and the position of the lamprey C3 is poorly supported. The important result revealed in Figure 4 is that SpC3 appears as the first diverging member of the complement protein family.

We also investigated the α-chains of the thioester proteins because this chain includes many important functional regions. We chose 490 amino acids that started at the beginning of the α-chain, included the thioester site, and terminated with the end of the α_2 M proteins. We used this region in another alignment and phylogenetic analysis, again employing the α_2 M proteins as the outgroup. The resulting tree had a different appearance from that seen in Figure 4. The complement clade had six unresolved groups that included 1) sea urchin C3, 2) hagfish C3, 3) lamprey C3, 4) higher vertebrate C3, 5) vertebrate C4, and 6) mammalian C5 (results not shown). This decreased resolution perhaps reflects the similarity of the α-chains that contain significant sequence conservation in all of the complement proteins.

Analysis of SpC3 protein by SDS-PAGE and N-terminal sequencing

The presence of a single junction in the deduced, unprocessed protein predicted that SpC3 would have two chains. To test this and to characterize the size of the protein, SpC3 was partially purified from coelomic fluid and was separated by SDS-PAGE under reducing and nonreducing conditions (Fig. 5). The nonreduced protein is 210 kDa (Fig. 5, *lane 1*), and under reducing conditions (*lane 2*), two chains are resolved as the α-chain (130 kDa) and the β-chain (80 kDa). These observed sizes are larger than the sizes deduced from the cDNA sequence, suggesting that some or all of the consensus *N*-linked glycosylation sites are filled during SpC3 processing by the coelomocyte.

A rabbit antiserum was raised against a peptide designed from the deduced sequence in the α-chain (see Fig. 2). On Western blots of reducing gels, the antiserum bound to the larger α-chain (Fig. 5, *lane 3*). To ensure that the protein isolated from the coelomic fluid was encoded by Sp064, the N terminus of both chains was sequenced. The β-chain was found to be blocked; however, the α-chain gave the peptide, SIDRDQLXLYDP. In sequencing the

cDNA encoding the N terminus of the α-chain, we found the following peptide: SIDRDQLCLYDP (see Fig. 2). This identical match is evidence that the protein isolated from the coelomic fluid is the same as that encoded by Sp064.

Sp064 gene expression in sea urchin tissues

To determine whether coelomocytes were the only tissue to express the Sp064 gene, we probed poly(A)$^+$ RNA isolated from the major adult tissues (coelomocyte, ovary, testis, gut) and found that Sp064 is expressed exclusively in the coelomocytes (Fig. 6). The transcript size is approximately 9 kb, which is longer than the total length of the overlapping cDNAs that we have sequenced (9-kb transcript, 7.6-kb cDNA sequence). Very weak bands appeared in all lanes after very long exposures (data not shown), with expression in gonads being higher than that in the gut. Although the other tissues appear to have low expression of the Sp064 gene, it is likely that coelomocytes were present in or on these organs at the time of dissection and total RNA isolation. Since it is not possible to wash or remove all coelomocytes from other sea urchin tissues during RNA isolation, these cells may account for the weak bands (Fig. 6). These data indicate that coelomocytes are the major or perhaps the only source of Sp064 gene expression in the adult sea urchin.

Sp064 gene copy number

In most animals, C3 is a single copy gene; however, gene duplication events are known to have occurred in some organisms. Examples include C4A and C4B in humans (41), C4 and sex-limited protein in mice (30, 42), and multiple gene copies of C3 in trout (24, 43) and cobra (27, 44). To determine whether Sp064 is a single or multiple copy gene per haploid genome, we analyzed a genome blot. Three male sea urchins were treated to 15 V (direct current) and shaken to induce spawning. Sperm was collected and DNA was isolated according to Lee et al. (45). Each sample was digested with three endonucleases (*Eco*RI, *Kpn*I, *Bam*HI), and the genome blot was analyzed with a riboprobe that corresponded to the α-chain region of the message (500-bp fragment from the 5′ end of pExCell054). Only one or two bands were seen in each lane, indicating that Sp064 is a single copy gene (data not shown).

Discussion

Sea urchins display a nonspecific, innate type of immune response. This was first determined from an extensive series of allograft

rejection experiments, in which the kinetics of the second set rejections were found to be the same as that of the third party rejections, and both second set and third party rejection rates were significantly faster than that for first set rejections (46; reviewed in Ref. 16). Furthermore, general activation of the sea urchin immune response has been inferred from increases in profilin transcripts in coelomocytes responding either to injury (47) or to LPS (14). This inference is based on the fact that 1) profilin is a key regulatory protein involved in modifying the actin cytoskeleton (48), and 2) amoeboid, phagocytic cells, such as the major coelomocyte cell type, readily change their shape, i.e., their cytoskeleton, to increase motility, phagocytosis, and encapsulation upon being activated by immune challenge or injury.

The nonspecific, activatible immune response in the sea urchin is very effective at maintaining a healthy animal; however, little is known about the molecular mechanisms of gene expression or protein function through which this system functions. Our preliminary study of sea urchin ESTs was the first molecular evidence that a simple complement system exists in the sea urchin (15), and we report in this work the molecular characterization of SpC3. This is the first identification of a C3 homologue that is expressed in a sea urchin, and furthermore, it is the first complement component to be identified in an invertebrate. Homology of SpC3 with vertebrate C3 proteins is based on several pieces of evidence. 1) Sequence analysis reveals a $\beta\alpha$ junction, the absence of an $\alpha\gamma$ junction (which is present in C4), and the conserved thioester site (which is absent in C5). 2) Protein gel analysis shows that SpC3 is composed of two chains. The combination of a two-chain molecule with a conserved thioester site suggests that SpC3 is a C3 complement homologue. Calculations of identities and similarities of paired alignments shown on Table I plus the analysis of the alignment and the phylogenetic tree all show that SpC3 is not an α_2M homologue and that it is the first diverging member of the thioester complement protein family.

In mammals, the primary site of C3 biosynthesis and secretion is the liver, with more than 90% of all C3 being produced by hepatocytes (10). However, several other cell types appear to produce and secrete C3 besides the liver, including macrophages, monocytes, fibroblasts, B lymphocytes, polymorphonuclear leukocytes, type II pneumocytes, astrocytes, and microglial cells (10, 49–53). These extrahepatic sites of complement production are very important in local inflammatory reactions. Since sea urchins do not have an equivalent of a liver, or a hepatopancreas that is found in sea stars, the specific expression of the Sp064 gene in coelomocytes suggests that these cells are involved in producing SpC3 for both local and systemic function. Although we cannot rule out the possibility that all major tissues express Sp064 based on the minor bands present in all lanes in Figure 6, one of the functions of coelomocytes may involve patrolling and invasive activities within organs when responding to microinjuries or focal sites of inflammation. This possibility has been suggested previously and was based on several ESTs encoding putative proteases that may be capable of degrading the extracellular matrix (15). Consequently, the presence of coelomocytes in all tissues of the adult sea urchin would be revealed by low levels of Sp064 expression.

Simple complement systems have been identified in agnathan fishes and consist of a C3-like component (54–57), factor B (58), and a putative complement receptor on circulating leukocytes (59). Complement in hagfish has been shown to function as an opsonin (60, 61). Although little is known about complement in tunicates, a mannan-binding lectin-associated protease homologue has been characterized from a tunicate, suggesting the presence of a lectin activation pathway (62) and other complement components may

be present in these animals. Preliminary sequence data from a PCR fragment have indicated that a thioester protein may be present in the compound tunicate, *Botryllus schlosseri* (63). Based on what is known about complement function in lower vertebrates, SpC3 may function as an opsonin and be a very important mechanism for host protection against pathogens.

The presence of a C3 protein in sea urchins may suggest that a number of accessory and regulatory proteins may also be present in this organism. If the sea urchin complement acts as an opsonin functioning to identify foreign cells for removal and destruction by phagocytic coelomocytes, this suggests that a C3 receptor should be present on these cells. Preliminary evidence for a receptor putatively associated with the sea urchin complement system has been reported previously, suggesting that it might be involved in augmented phagocytosis (64–67). In vertebrates, spontaneously activated C3 can be bound to self cells in the form of C3b, which is then inactivated by factor I that functions with a number of cofactors such as membrane cofactor protein, factor H, and CR1 (for review, see Refs. 68, 69). Decay-accelerating factor is an additional cell surface protein that may dissociate the C3b-factor B complex, called C3bBb, thereby deactivating its C3 convertase activity (70). Predictions of regulatory proteins in the sea urchin at present can only be based on the presence of a few conserved cleavage sites that have been identified in deduced SpC3 sequence. These include sites for factor I and C3 convertase, which is formed through an interaction between C3b and factor B. At present, none of these predicted proteins have been cloned and sequenced in the sea urchin, except one. We have characterized recently the complete open reading frame from EST152 (15), and it encodes a factor B-like protein (Smith, unpublished).

Lachmann (71) has proposed that the most primitive complement cascade or "archeo-complement" system would have resembled a simple alternate pathway consisting of a C3-like protein with a thioester site, a factor B-like protein containing short consensus repeats and a serine protease domain, and a complement receptor on phagocytic immune cells. Based on the sequence similarities among various complement components, it has been suggested that several of the complement protein families have been generated by gene duplication from a small number of primordial genes (35, 39, 72). Previous to the current work, the agnathans appeared to fulfill the prediction of an archeo-complement system, inferring not only that the alternate cascade was more ancient than the classical cascade, but that it was present in the common ancestor of the vertebrates. It is now clear, with this first characterization of a C3 homologue from the phylogenetically older deuterostome phylum, the echinodermata, that this protein represents the first diverging complement component. Furthermore, it is interesting to consider the possibility that it may still bear some similarities to the ancestral protein that functioned in the common ancestor of the deuterostomes that gave rise through gene duplication to the complement family of thioester proteins that functions in the higher vertebrates today.

Acknowledgments

References

1. Tomlinson, S. 1993. Complement defense mechanisms. *Curr. Opin. Immunol. 5:83.*

2. Super, M., and R. A. Ezekowits. 1992. The role of mannose-binding proteins in host defense. *Infect. Agents Dis. 1:194.*

3. Holmskov, U., R. Malhotra, R. B. Sim, and J. C. Jensenius. 1994. Collectins: collagenous C-type lectins of the innate immune defense system. *Immunol. Today 15:67.*

4. Lu, J. H., S. Thiel, H. Wiedmann, R. Timpl, and K. B. Reid. 1990. Binding of the pentamer/hexamer forms of mannan-binding protein to zymosan activates the proenzyme C1r2C1s2 complex of the classical pathway of complement without involvement of C1q. *J. Immunol. 144:2287.*

5. Ohta, M., M. Okada, I. Yamashina, and T. Kawasaki. 1990. The mechanisms of carbohydrate-mediated complement activation by the serum mannose-binding protein. *J. Biol. Chem. 265:1980.*

6. Sato, T., Y. Endo, M. Matsushita, and T. Fujita. 1994. Molecular characterization of a novel serine protease involved in activation of the complement system by mannose-binding protein. *Int. Immunol. 6:665.*

7. Matsushita, M., and T. Fujita. 1995. Cleavage of the third component of complement (C3) by mannose-binding protein-associated serine protease (MASP) with subsequent complement activation. *Immunobiology 194:443.*

8. Lambris, J. D., M. Mavroidis, and J. O. Sunyer. 1994. Phylogeny of third component of complement, C3. In *New Aspects of Complement Structure and Function.* A. Erdei, ed. R. G. Landes Co., Austin, TX, p. 15.

9. Becherer, J. D., J. Alsenz, C. Servis, B. L. Myones, and J. D. Lambris. 1989. Cell surface proteins reacting with activated complement components. *Complement Inflamm. 6:142.*

10. Lambris, J. D. 1988. The multifunctional role of C3, the third component of complement. *Immunol. Today 9:387.*

11. Lambris, J. D. 1989. *The Third Component of Complement: Chemistry and Biology.* Springer Verlag, Berlin.

12. Dempsey, P. W., M. E. D. Allison, S. Akkaraju, C. C. Goodnow, and D. T. Fearon. 1996. C3d of complement as a molecular adjuvant: bridging innate and acquired immunity. *Science 271:348.*

13. Smith, L. C., and E. H. Davidson. 1994. The echinoderm immune system: characters shared with vertebrate immune systems and characters arising later in deuterostome phylogeny. *Ann. NY Acad. Sci. 712:213.*

14. Smith, L. C., R. J. Britten, and E. H. Davidson. 1995. Lipopolysaccharide activates the sea urchin immune system. *Dev. Comp. Immunol. 19:217.*

15. Smith, L. C., L. Chang, R. J. Britten, and E. H. Davidson. 1996. Sea urchin genes expressed in activated coelomocytes are identified by expressed sequence tags (ESTs): complement homologues and other putative immune response genes suggest immune system homology within the deuterostomes. *J. Immunol. 156:593.*

16. Smith, L. C., R. J. Britten, and E. H. Davidson. 1992. SpCoel1, a sea urchin profilin gene expressed specifically in coelomocytes in response to injury. *Mol. Biol. Cell 3:403.*

17. Humphreys, T. 1963. Chemical dissolution and in vitro reconstruction of sponge cell adhesions. I. Isolation and functional demonstration of the components involved. *Dev. Biol. 8:27.*

18. Maier, E., S. Meier-Ewert, A. R. Ahmadi, J. Curtis, and H. Lehrach. 1994. Application of robotic technology and automated sequence fingerprint analysis by oligonucleotide hybridization. *J. Biotechnol. 35:191.*

19. Sanger, F., S. Nicklen, and A. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA 74:5463.*

20. Altschul, S., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol. 215:403.*

21. Thompson, J., D. Higgins, and T. Gibson. 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.*

22. Swofford, D. 1990. *PAUP: Phylogenetic Analysis Using Parsimony.* Version 3.1. Program documentation. Illinois Natural History Survey, Urbana.

23. Briand, J. P., S. Muller, and M. H. Van Regenmortal. 1985. Synthetic peptides as antigens: pitfalls of conjugation methods. *J. Immunol. Methods 78:59.*

24. Sunyer, J. O., L. Tort, and J. D. Lambris. 1997. Structural C3 diversity in fish: characterization of five forms of C3 in the diploid fish *Sparus aurata. J. Immunol. 158:2813.*

25. Matsudaira, P. 1987. Sequence from picomole quantities of proteins electroblotted onto polyvinylidene diflouride membranes. *J. Biol. Chem. 262:10035.*

26. Mavroidis, M., J. O. Sunyer, and J. D. Lambris. 1995. Isolation, primary structure, and evolution of the third component of chicken complement and evidence for a new member of the α_2-macroglobulin family. *J. Immunol. 154:2164.*

27. Fritzinger, D. C., E. C. Petrella, M. B. Connelly, R. Bredehorst, and C. W. Vogel. 1992. Primary structure of cobra complement component C3. *J. Immunol. 149:3554.*

28. deBruijn, M. H., and G. H. Fey. 1985. Human complement component C3: cDNA coding sequence and derived primary structure. *Proc. Natl. Acad. Sci. USA 82:708.*

29. Fey, G. H., A. Lundwall, R. A. Wetsel, B. F. Tack, M. H. deBruijn, and H. Domdey. 1984. Nucleotide sequence of complementary DNA and derived amino acid sequence of murine complement protein C3. *Philos. Trans. R. Soc. London B 306:333.*

30. Ogata, R. T., and D. S. Sepich. 1984. Genes for murine fourth complement component (C4) and sex-limited protein (Slp) identified by hybridization to C4- and Slp-specific cDNA. *Proc. Natl. Acad. Sci. USA 81:4908.*

31. Kyte, J., and R. F. Doolittle. 1982. A simple method for displaying the hydrophobic character of a protein. *J. Mol. Biol. 157:105.*

32. Von-Heijne, A. L. 1986. A new method for predicting signal sequence cleavage sites. *Nucleic Acids Res. 14:4683.*

33. Asson-Batres, M. A., S. L. Spurgeon, J. Diaz, T. G. DeLoughery, and G. C. Bagby. 1994. Evolutionary conservation of the AU-rich 3′ untranslated region of messenger RNA. *Proc. Natl. Acad. Sci. USA 91:1318.*

34. Hirani, S., J. D. Lambris, and H. J. Müller-Eberhard. 1986. Structural analysis of the asparagine-linked oligosaccharides of human complement component C3. *Biochem. J. 233:613.*

35. Dodds, A. W., and A. J. Day. 1993. The phylogeny and evolution of the complement system. In *Complement in Health and Disease*, Vol. 20. K. Whaley, M. Loos, and J. M. Weiler, eds. Kluwer Academic Pub., Boston, p. 39.

36. Lambris, J. D., Z. Lao, J. Pang, and J. Alsenz. 1993. Third component of trout complement: cDNA cloning and conservation of functional sites. *J. Immunol. 151:6123.*

37. Isaac, L., and D. E. Isenman. 1992. Structural requirements for thioester bond formation in human complement component C3: reassessment of the role of thioester bond integrity on the conformation of C3. *J. Biol. Chem. 267:10062.*

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

39. Hughes, A. L. 1994. Phylogeny of the C3/C4/C5 complement-component gene family indicates that C5 diverged first. *Mol. Biol. Evol. 11:417.*

40. Mo, R., Y. Kato, M. Nonaka, K. Nakayama, and M. Takahashi. 1996. Fourth component of *Xenopus laevis* complement: cDNA cloning and linkage analysis of the frog MHC. *Immunogenetics 43:360.*

41. Belt, K. T., M. D. Carroll, and R. R. Porter. 1984. The structural basis of the multiple forms of human complement component C4. *Cell 36:907.*

42. Nonaka, M., M. Takahashi, S. Natsuume-Sakai, M. Nonaka, S. Tanaka, A. Shimizu, and T. Honjo. 1984. Isolation of cDNA clones specifying the fourth component of mouse complement and its isotype, sex-limited protein. *Proc. Natl. Acad. Sci. USA 81:6822.*

43. Sunyer, J. O., I. K. Zarkadis, A. Sahu, and J. D. Lambris. 1996. Multiple forms of complement C3 in trout, that differ in binding to complement activators. *Proc. Natl. Acad. Sci. USA 91:8546.*

44. Fritzinger, D. C., R. Bredehorst, and C. W. Vogel. 1994. Molecular cloning and derived primary structure of cobra venom factor. *Proc. Natl. Acad. Sci. USA 91:12775.*

45. Lee, J. J., R. J. Shott, S. J. Rose, T. L. Thomas, R. J. Britten, and E. H. Davidson. 1984. Sea urchin actin gene subtypes; gene number, linkage and evolution. *J. Mol. Biol. 172:149.*

46. Coffaro, K. 1979. Transplantation immunity in the sea urchin. Doctoral dissertation, University of California, Santa Cruz, CA.

47. Smith, L. C., and E. H. Davidson. 1992. The echinoid immune system and the phylogenetic occurrence of immune mechanisms in deuterostomes. *Immunol. Today 13:356.*

48. Sohn, R. H., and P. J. Goldschmidt-Clermont. 1994. Profilin: at the crossroads of signal transduction and the actin cytoskeleton. *BioEssays 16:465.*

49. Strunk, R. C., A. S. Whitehead, and F. S. Cole. 1985. Pretranslational regulation of the synthesis of the third component of complement in human mononuclear phagocytes by the lipid A portion of lipopolysaccharide. *J. Clin. Invest. 76:985.*

50. Rothman, B. L., M. Merrow, M. Bamba, T. Kennedy, and D. L. Kreutzer. 1989. Biosynthesis of the third and fifth complement components by isolated human lung cells. *Am. Rev. Respir. Dis. 139:212.*

51. Botto, M., D. Lissandrini, C. Sorio, and M. J. Walport. 1992. Biosynthesis and secretion of complement component (C3) by activated human polymorphonuclear leukocytes. *J. Immunol. 149:1348.*

52. Strunk, R. C., J. A. Fleischer, Y. Katz, and F. S. Cole. 1994. Developmentally regulated effects of lipopolysaccharide on biosynthesis of the third component of complement and factor B in human fibroblasts and monocytes. *Immunology 82:314.*

53. Haga, S., T. Aizawa, T. Ishii, and K. Ikeda. 1996. Complement gene expression in mouse microglia and astrocytes in culture: comparisons with mouse peritoneal macrophages. *Neurosci. Lett. 216:191.*

54. Ishiguro, H., D. Kobayashi, M. Suzuki, K. Titani, S. Tomonaga, and Y. Kurosawa. 1992. Isolation of a hagfish gene that encodes a complement component. *EMBO J. 11:892.*

55. Hanley, P. J., J. W. Hook, D. A. Raftos, A. A. Gooley, R. Trent, and R. L. Raison. 1992. Hagfish humoral defense protein exhibits structural and functional homology with mammalian complement components. *Proc. Natl. Acad. Sci. USA 89:7910.*

56. Nonaka, M., and M. Takahashi. 1992. Complete complementary DNA sequence of the 3rd component of complement of Lamprey: implication for the evolution of thioester containing proteins. *J. Immunol. 148:3290.*

57. Fugii, T., T. Nakamura, and S. Tomonaga. 1995. Component C3 of hagfish complement has a unique structure: identification of native C3 and its degradation products. *Mol. Immunol. 32:633.*

58. Nonaka, M., M. Takahashi, and M. Sasaki. 1994. Molecular cloning of a lamprey homologue of the mammalian MHC class III gene, complement factor B. *J. Immunol. 152:2263.*

59. Raison, R. L., J. Coverley, J. W. Hook, P. Towns, K. M. Weston, and D. A. Raftos. 1994. A cell-surface opsonic receptor on leukocytes from the phylogenetically primitive vertebrate, *Eptatretus stouti. Immunol. Cell Biol. 74:326.*

60. Fugii, T., T. Nakamura, A. Sekizawa, and S. Tomonaga. 1992. Isolation and characterization of a protein from hagfish serum that is homologous to the third component of the mammalian complement system. *J. Immunol. 149:117.*

61. Raftos, D. A., J. W. Hook, and R. L. Raison. 1992. Complement-like protein from the phylogenetically primitive vertebrate *Eptatretus stouti*, is a humoral opsonin. *Comp. Biochem. Physiol. B 103:397.*

62. Ji, X., K. Azumi, M. Sasaki, and M. Nonaka. 1997. Ancient origin of the complement lectin pathway revealed by molecular cloning of mannan binding protein-associated serine protease from a urochordate, the Japanese ascidian, *Halocynthia roretzi. Proc. Natl. Acad. Sci. USA 94:6340.*

63. Baish, M. A., R. L. Lohr, and S. Bartl. 1997. Molecular evidence for complement and α_2-macroglobulin family members in the colonial ascidian, *Botryllus schlosseri. Dev. Comp. Immunol. 22:147.*

64. Bertheussen, K. 1981. Endocytosis by echinoid phagocytes in vitro. I. Recognition of foreign matter. *Dev. Comp. Immunol. 5:241.*

65. Bertheussen, K. 1982. Receptors for complement on echinoid phagocytes. II. Purified human complement mediates echinoid phagocytosis. *Dev. Comp. Immunol. 6:635.*

66. Bertheussen, K. 1983. Complement-like activity in sea urchin coelomic fluid. *Dev. Comp. Immunol. 7:21.*

67. Bertheussen, K., and R. Seljelid. 1982. Receptors for complement on echinoid phagocytes. I. The opsonic effect of vertebrate sera on echinoid phagocytosis. *Dev. Comp. Immunol. 6:423.*

68. Parker, C. J. 1992. Regulation of complement by membrane proteins: an overview. *Curr. Top. Microbiol. Immunol. 178:1.*

69. Weiler, J. M. 1993. Introduction: complement in health and disease. In *Immunology and Medicine*, Vol. 20. K. Whaley, M. Loos, and J. M. Weiler, eds. Kluwer Academic Publishers, Boston, p. 1.

70. Kuttner-Kondo, L., M. E. Medof, W. Brodbeck, and M. Shoham. 1996. Molecular modeling and mechanism of action of human decay-accelerating factor. *Protein Eng. 9:1143.*

71. Lachmann, P. J. 1979. An evolutionary view of the complement system. *Behring Inst. Mitt. 63:25.*

72. Bentley, D. R. 1988. Structural superfamilies of the complement system. *Exp. Clin. Immunogenet. 5:69.*