Immunogenetics (2004) 56:89–106
DOI 10.1007/s00251-004-0665-2

ORIGINAL PAPER

Keri A. Multerer · L. Courtney Smith

Two cDNAs from the purple sea urchin, *Strongylocentrotus purpuratus*, encoding mosaic proteins with domains found in factor H, factor I, and complement components C6 and C7

Received: 10 December 2003 / Revised: 26 February 2004 / Published online: 16 April 2004

Abstract The vertebrate complement system is composed of about 30 serum and cell surface proteins that make up three activation pathways, a lytic pathway, and a set of proteins that regulate complement. Regulatory proteins are required for host protection against autologous complement attack and to control the amplification feedback loop of the alternative pathway. Purple sea urchin, *Strongylocentrotus purpuratus*, homologues of complement C3 (SpC3) and factor B (SpBf) have been identified, suggesting the presence of an alternative complement pathway. This implies that echinoderms require a complement regulatory system for the same reasons that it is required in higher vertebrates. Two cDNAs, *Sp5* and *Sp5013*, have been characterized from coelomocytes and the deduced structures of the encoded mosaic proteins, SpCRL (*S. purpuratus* complement related protein, long form) and SpCRS (short form), have domains that are also found in regulatory proteins such as factor H and factor I and the terminal pathway components C6 and C7. These domains include multiple short consensus repeats, a fucolectin domain, Ser/Thr/Pro-rich regions, a Cys-rich region, and a factor I-membrane attack complex domain. The genes are constitutively expressed in all tissues of the sea urchin and are not induced in response to immune challenge. Multiple bands of varying intensity on both genome blots and RNA blots suggest that *Sp5* and *Sp5013* are members of a small gene family and that they might undergo alternative splicing. Based on the domains present in SpCRL and SpCRS, they might be either examples of complement regulatory proteins or members of the terminal pathway of complement.

Keywords Evolution · Innate · Echinoderm · Complement

K. A. Multerer · L. C. Smith (✉)
Graduate Program in Genetics,
The Institute of Biomedical Sciences,
George Washington University,
Washington, DC 20052, USA
e-mail: csmith@gwu.edu
Tel.: +1-202-9949211
Fax: +1-202-994-6100

L. C. Smith
Department of Biological Sciences,
George Washington University,
340 Lisner Hall, 2023 G Street NW, Washington, DC 20052, USA

Present address:
K. A. Multerer, Human Biology Division,
Fred Hutchinson Cancer Center,
Seattle, WA 98109, USA

Introduction

All multicellular organisms have some form of innate immunity that functions in the identification of, and protection against, invading pathogens and parasites. The complement system, which is composed of about 35 serum and cell surface proteins (Volanakis 1998), is an important component of the more ancient innate immune system. It has been identified throughout the lineage of deuterostome animals (Smith et al. 1999; Nonaka 2001; Azumi et al. 2003) and is being discovered in the rest of the animal kingdom as well, such as in a gorgonian (accession no. AAN86548) and a squid (M. McFall-Ngai, personal communication). There are three activation pathways—classical, alternative, and lectin—that are activated by different types of molecules and lead to the activation of the terminal or lytic pathway. The activation pathways culminate in the formation of two distinct C3-convertase enzymes: the alternative pathway convertase (C3bBb) and the classical pathway convertase (C4bC2b) (reviewed by Nonaka et al. 1998; Xu et al. 2001). The formation of C3-convertases leads to the formation of C5-convertases (Pangburn and Rawal 2002) that activate the lytic pathway. C3-convertases function to cleave and activate additional C3, creating a positive feedback loop within the alternative pathway that accelerates the activation of the entire complement system (Liszewski et al. 1996), resulting in quick and efficient opsonization and lysis of foreign cells (Lambris 1988; Becherer et al. 1989).

Because complement proteins C3 and C4 can form covalent thioester bonds with amines and hydroxyls on any molecule, they have the ability to bind any surface, including self, which would lead to inappropriate cell lysis and inflammatory reactions. Covalent bond formation between C3 or C4 and non-self is an important mechanism for identifying pathogens. Therefore, a complement regulatory system is necessary to both protect self-cells against autologous complement attack and, consequently, to direct the attack towards foreign pathogens. Tight regulation of active C3-convertases is also essential to prevent unnecessary depletion of complement proteins from uncontrolled activation, as well as to inhibit complement activities after a pathogen has been cleared (Liszewski et al. 1996). One important mechanism of regulation is based on dissociating the C3-convertase complexes and by degrading C3. Some of the proteins involved are membrane cofactor protein (MCP), complement receptors 1 and 2 (CR1 and CR2), decay accelerating factor (DAF), C4 binding protein (C4BP), factor H, and factor I (Medof et al. 1987; Krych et al. 1991; Liszewski et al. 1996; Arlaud et al. 1998; Kirkitadze and Barlow 2001; Barilla-LaBarca et al. 2002).

Sea urchin immune response

The sea urchin defends itself against pathogens with its innate immune system. Coelomocytes are the immune effector cells (Boolotian and Geise 1958; Johnson 1969; Gross et al. 2000) that carry out many protective functions including encapsulation, phagocytosis, chemotaxis, and expression of putative cytotoxic agents in response to invasion of foreign cells (Smith et al. 1992; Smith and Davidson 1994). There are four morphologically distinct classes of coelomocytes (Johnson 1969; Edds 1993) of which the phagocytes appear to be the major immune effector cells. Phagocytes express *Sp064*, which encodes SpC3, a homologue of complement component C3 (Al-Sharif et al. 1998; Gross et al. 2000), and *Sp152*, which encodes SpBf, a homologue of complement component factor B (Bf) (Smith et al. 1998; unpublished data). True thioester activity, similar to that characterized for C3 in higher vertebrates, has been demonstrated chemically for SpC3 through autolytic fragmentation and thioester inactivation with methylamine (Smith 2002). In biological assays, SpC3 functions as an opsonin and augments phagocytosis of yeast by coelomocytes (Smith 2001; Clow et al. 2004). The complement system in the sea urchin has been proposed to function like an alternative pathway with the formation of a C3-convertase complex from SpC3 and SpBf, resulting in a feedback loop to augment the rate of opsonization (Smith et al. 1999; Smith 2001). This suggests that the thioester binding activity and the convertase functions would require a mechanism for protection against autologous attack by controlling or limiting the convertase activity that would avoid depletion of the components. Initial evidence for a complement regulatory system in the sea urchin was obtained from the sequence of SpC3, which has two conserved cleavage sites for factor I-like activity that are located in conserved positions (Al-Sharif et al. 1998). This predicts that additional complement components function in the sea urchin.

In the present study, we show the analysis of two cDNAs, *Sp5* (accession no. AY494840) and *Sp5013* (accession no. AY494841), which encode proteins SpCRL (*S. purpuratus* complement related protein, long form) and SpCRS (short form), respectively. Deduced amino acid sequences from both proteins show multiple short consensus repeats (SCRs) (18 in SpCRL and four in SpCRS), two Ser/Thr/Pro (S/T/P)-rich regions, and a factor I-membrane attack complex (FIMAC) domain. In addition, SpCRL has a fucolectin domain and a Cys-rich region. Both genes are expressed in coelomocytes, gut, gonad, pharynx, esophagus, and axial organ, and the level of message accumulation in coelomocytes before and after immune challenge suggests that expression is constitutive. Northern blots for both messages show multiple bands of varying intensity, suggesting the presence of a small family of genes with similar sequences in addition to possible alternative splicing. Genome blots for both genes also show multiple bands of varying intensity consistent with gene structure of multiple exons and perhaps with cross-hybridization with other members of the gene family. Structural and phylogenetic analyses of the deduced amino acid sequences for these two proteins indicate that they share domains with a number of complement proteins from higher vertebrates including factor H, factor I, C6 and C7. These results imply functions within the complement system, possibly as complement regulatory proteins and/or within a primitive terminal pathway in the sea urchin.

Materials and methods

Animals

Purple sea urchins, *Strongylocentrotus purpuratus*, were obtained and housed as previously described (Gross et al. 2000; Shah et al. 2003).

RNA isolation

Total RNA from gut, gonad, esophagus, pharynx, axial organ, and coelomocytes was isolated using the RNAeasy Midi kit (Qiagen, Valencia, Calif.). Animals were sacrificed by removing Aristotle's lantern (mouth parts), and the coelomic fluid was poured from the body cavity through sterile cheese cloth and mixed in a 1:1 ratio with ice-cold, Ca^{2+}- and Mg^{2+}-free sea water containing 70 mM EDTA and 50 mM imidazole (pH 7.4) (CMFSW-EI) according to Gross et al. (1999). Coelomocytes were pelleted by centrifugation at 6,500 *g* for 5 min at 4°C. To harvest the internal organs, the test was cut open and tissues were removed from the coelomic cavity and placed in RNAlater (Ambion, Austin, Tex.). Pelleted coelomocytes and approximately 0.24 g of each solid tissue were lysed in guanidine isothiocyanate lysis buffer (Qiagen). The lysate was bound to a silica gel-based column membrane, digested with 27 U DNaseI (Qiagen), washed, eluted in RNase-free water, and quantified with a DU 640 spectrophotometer (Beckman Instruments).

RNAsin (Promega, Madison, Wis.) was added to each sample (0.8 U/μl) prior to storage at −70°C.

Reverse transcriptase polymerase chain reaction

Reverse transcriptase (RT) reactions were performed with 1–3 μg total RNA and 5 μM random hexamer primer with Superscript II reverse transcriptase (Invitrogen, Carlsbad, Calif.) according to the manufacturer's instructions. To identify samples with contaminating genomic DNA, approximately 150 ng of each RNA sample was used directly as a template for PCR. All primer sets employed in RT-PCR reactions (see below) were used to identify amplification from genomic DNA. Samples showing bands larger than expected for cDNA amplification were assumed to contain genomic DNA and were discarded.

The cDNA generated from the RT reactions (1 μl) was mixed with 0.5 U *Taq* DNA polymerase (Invitrogen), 3 mM of each deoxynucleotide, 1 μM each primer, 1× company-supplied buffer (Invitrogen), and 1.5 mM $MgCl_2$ in a volume of 20 μl. Primers included *Sp5*: 5for: 5′ CCC TGG ACA GTA TGT GTT GCA TGG TAG, 5rev: 5′ TAT CCC TGG TTG CAT CCT ATG AGC ACA); *Sp5013*: 5013for: 5′ TCG ATG GGT GTT CCG AGT GGG TCT, 5013rev: 5′ TCT ACA TCT AGC AAC TAG CAG GGT GCC; *Sp056*: 056for: GCA CAG CCA GCA ACC AGC ACT ACA AT, 056rev: ACG CCG ATG GGT TCT ACA GTG AAG GT; and *SpL8*: L8for: CAG CGT AAG GGA GCG GGA AGC GTC TT, L8rev: GTT TGC CGC AGA AGA TGA ACT GTC CCG TGT A. Reactions were heated to 95°C for 5 min; followed by 25 cycles of 94°C for 30 s, 52°C for 30 s, and 72°C for 10 min; followed by 72°C for 2 min and 4°C hold. The amplified fragments were electrophoresed on a 0.8% agarose/0.4% NuSieve gel (BioWhittaker Molecular Application, Rockland, Md.) containing 0.5 μg/ml ethidium bromide in TAE buffer (40 mM Tris base, 20 mM glacial acetic acid, 1 mM EDTA, pH 8.3). Gels were imaged with a DC120 digital camera and 1D digital software (Eastman Kodak, Rochester, N.Y.) followed by image optimization in Photoshop (Adobe Systems, Seattle, Wash.).

Cloning PCR fragments

Fragments amplified by PCR were cloned into the pCRII-TOPO vector according to the manufacturer's instructions (Invitrogen) and transformed into TOP10 bacteria (Invitrogen).

Arrayed cDNA libraries

An arrayed cDNA library was constructed in the pBK-CMV vector (Stratagene, La Jolla, Calif.) using activated coelomocytes from five sea urchins 24 h post-injection of 1 ml of heat-killed bacteria (for details of bacterial isolation and library construction, see Pancer et al. 1999; Cameron et al. 2000; Rast et al. 2000). A similar arrayed cDNA library was constructed using non-induced coelomocytes in the pSPORT vector (Life Technologies, Rockville, Md.) using coelomocytes from non-activated sea urchins (for details, see Smith et al. 1996; Al-Sharif et al. 1998; Cameron et al. 2000). Each library was arrayed into 240 plates of 384 wells each for a total of 92,160 clones per library (Al-Sharif et al. 1998; Cameron et al. 2000; Rast et al. 2000). The insert from every clone was amplified by PCR and spotted in duplicate onto five 22×22-cm Hybond-N^+ (Amersham) filters for each library, which were used for screening.

cDNA library screens

Filters were prehybridized with 10 ml of hybridization solution [0.1% BSA (w/v), 1 mM ethylenediaminetetraacetic acid (EDTA), 0.25 M phosphate buffer (pH 7.4), 7% sodium dodecylsulfate (SDS) (w/v), and 50% formamide (v/v)] at 42°C for 2 h by rotation in a hybridization oven (Robbins Scientific, Sunnyvale, Calif.). The hybridization solution was replaced, a riboprobe (see below) was added to the filters and then rotated overnight at 42°C. Filters were washed twice in 4× SSC (20× SSC is 0.3 M sodium citrate, 3 M NaCl, pH 7.0) with 1% SDS, twice in 2× SSC with 1% SDS, and twice in 1× SSC with 1% SDS at 65°C for 30 min. Each wet filter was sealed into a plastic bag and exposed to X-OMAT AR film (Eastman Kodak) without an intensifying screen. After exposure, filters were stripped by washing once in 0.4 M NaOH at 45°C for 3 min, twice in stripping buffer [0.1× SSC, 0.1% SDS, 0.2 M Tris (pH 7.5)] at 65°C for 30 min, and twice in stripping buffer with 0.5 M EDTA at room temperature for 10 min. After stripping, filters were dried and stored at −20°C.

Riboprobe synthesis

Clones that served as templates for riboprobe synthesis (see Fig. 1) were linearized with either *Bam*H1 at the 5′ end or *Xba*I at the 3′ end (Promega), and 170 ng to 515 ng was labeled using 50 μCi to 70 μCi ^{32}P rUTP (ICN, Irvine, Calif.); 0.5 mM each of rATP, rCTP, and rGTP; 15 U of T7 RNA polymerase (Promega); 1× company-supplied transcription buffer (Promega); 20 mM dithiothreitol (DTT); and 24 U of RNAsin (Promega) in a total volume of 20 μl. Samples were incubated at 37° for 1 h, followed by the addition of 46 μg yeast tRNA, 24 U of RNAsin, and digested with 1 U RQ1 DNAse (Promega) at 37°C for 15 min. Unincorporated nucleotides were removed by passing the probe through a G-50 fine Sephadex column (Amersham Pharmacia Biotech AB, Piscataway, N.J.) spun at 2,000 *g* for 1 min. Incorporation of ^{32}P was analyzed with an LS6500 liquid scintillation counter (Beckman Instruments).

Clone blots

Individual colonies were dispersed in sterile water (50 μl) and 5 μl was used as the template in PCR reactions using a 9600 thermal cycler (Perkin Elmer, Wellesley, Mass). Reactions included 0.5 U *Taq* DNA polymerase (Invitrogen), 0.25 mM each deoxynucleotide, 1 μM T3 and T7 primers (Qiagen), 1× company-supplied buffer (Invitrogen), and 1.5 mM $MgCl_2$ in a volume of 20 μl. Reactions were heated to 95°C for 5 min, followed by 25 cycles of 94°C for 30 s, 52°C for 30 s, and 72°C for 10 min, with a final step of 72°C for 10 min. Amplified products were electrophoresed on an agarose gel (as above), and images were captured using UV illumination with a digital camera (Eastman Kodak). Gels were double blotted by capillary action onto two Genescreen Plus membranes (NEN Life Science Products, Boston, Mass.) by standard procedure (Sambrook et al. 1989).

Non-radioactive DNA probes

Templates for *Sp5* and *Sp5013* were generated using clones that were amplified by PCR using T3 and T7 primers for clones matching to *Sp5013* and T7 and Sp6 primers for clones matching to *Sp5*. Amplified products were purified using the Geneclean Turbo kit for PCR (Bio101, Carlsbad, Calif.) according to the manufacturer's instructions, and DNA fragments (100 ng) were labeled using alkaline phosphatase non-radioactive GeneImages Alkphos Direct kit (Amersham) according to the manufacturer's instructions. The blots were hybridized with the probe and washed according to the manufacturer's instructions.

Clone maps

Analysis by PCR

Plasmid insert sizes were analyzed from single colonies by PCR as described above, using a combination of a gene-specific primer with either T3 or T7 primers. A touch-down thermal cycling program was used because of the significant difference in annealing

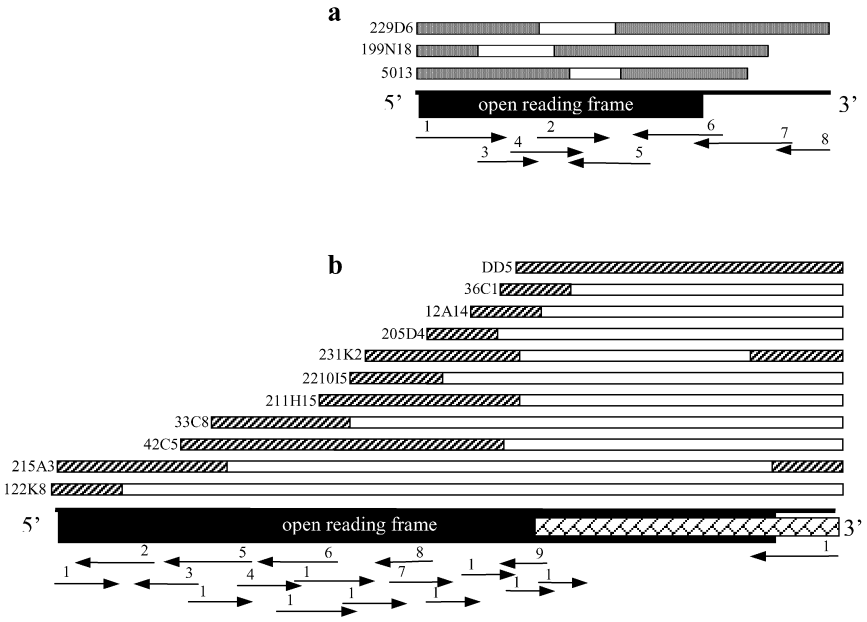


Fig. 1a, b Maps and sequencing passes for *Sp5013* and *Sp5*. **a** *Sp5013*. An overlapping series of 12 clones was analyzed by touch-down PCR reactions (see Materials and methods) using the 5013for primer and the T7 primer in addition to restriction digests. Three of the six clones that were used to generate the sequence for *Sp5013* are shown. The sequencing passes are indicated as *arrows*. The number associated with each arrow corresponds with a primer listed in Tables 1 and 2. Clone 199N18 served as the template to generate the riboprobe for Northern-blot analysis (Fig. 9). **b** *Sp5*. An overlapping series of 39 *Sp5* clones was analyzed using 5for (5′ CCC TGG ACA GTA TGT GTT GCA TGG TAG) with T7 and reverse RceR1 (5′ AAC ACA GCT TGG TTG CAC TCC TGT CC) with T3 by touch-down PCR reactions (see Materials and methods). Eleven of 17 clones that were used to generate 5.65 kb of sequence are shown. The length and overlap of the sequencing passes are indicated as *arrows* and the number associated with each arrow corresponds with primers listed in Tables 1 and 2. The area of the figure shown as a *tiled pattern* denotes the sequence provided by Z. Pancer. Clone 231K2 served as the template to generate the riboprobe for Northern-blot analysis (Fig. 9)

temperatures between the two primers (gene specific = 62–64°C; T3 and T7 = 55°C). The PCR program (described above) with a total of 30 cycles was changed so that the initial annealing temperature of 62°C was decreased by 1°C each cycle for eight cycles to 55°C, where it was maintained for 22 cycles, followed by 72°C for 10 min and 4°C hold. Amplified products were separated on an agarose gel and imaged as above.

Restriction-enzyme digests

Digests were done with *Eco*R1, *Xho*I, and *Bam*HI, and fragments were separated on agarose gels.

Northern blots

Poly(A)$^+$ RNA was isolated from 100 μg total RNA from coelomocytes using the Oligotex mRNA mini kit (Qiagen) according to the manufacturer's instructions. Poly(A)$^+$ RNA was electrophoresed through a 1% agarose gel containing 2.2 M formaldehyde in 1× MOPS buffer [10× MOPS is 20 mM 3-(*N*-morpholino) propanesulfonic acid, 5 mM NaOAc, 1 mM EDTA, pH 7] and blotted by capillary action onto Genescreen Plus (NEN Life Science Products) with 10× SSC. Filters were hybridized and washed as described above for library screens, and exposed to X-OMAT AR film (Eastman Kodak). Transcript sizes were estimated from RNA standards (Ambion) as well as coelomocyte rRNA. Filters were stripped using the protocol described above for library filters.

Sequencing

Plasmids were isolated using the Wizard Plus SV Miniprep kit (Promega) according to the manufacturer's instructions, quantified with the DU 640 spectrophotometer (Beckman Instruments), and sequenced with BigDye terminator cycle sequencing ready reaction kits (Applied Biosystems, Foster City, Calif.) using either T3, T7, or internal primers. Cycle sequencing reactions were done in a 9600 thermal cycler (Perkin Elmer) with 25 cycles of 96°C for 30 s, 50°C for 30 s, and 60°C for 4 min. Unincorporated nucleotides were removed by passing the samples through a G-50 fine Sephadex (Pharmacia) spin column and dried in a centrivap concentrator (Labconco, Kansas City, Mo.). Dried samples were either dissolved in 15 μl template suppression buffer and loaded onto an ABI prism 310 Capillary Sequencer or 3 μl sequencing loading buffer [5:1 ratio of deionized formamide to 25 mM EDTA (pH 8.0), plus bromophenol blue] and loaded onto a 377 Automated Sequencer (Applied Biosystems).

Results and discussion

Sp5 and *Sp5013* clones

Two clones, *Sp5* and *Sp5013*, were identified during a library screen using a probe designed to find sequences specific for scavenger receptors with Cys-rich (SRCR) domains (Pancer et al. 1999; Pancer 2000). However, the initial analysis of these two clones only revealed regions that encoded SCRs, and therefore they were kindly provided to us by Z. Pancer. Further analysis was pursued

Table 1 Sequencing primers for *Sp5013* listed in Fig. 1a. All annealing temperatures were between and 55.3°C and 57.2°C

Primer number in Fig. 1a	Primer	Sequence
3	199N18T3F	5′ TTGCTAGATGTAGAGACC
2	199N18T3F3	5′ GGGAGGATACTAGATGG
4	199N18T3F4	5′ GCAATGGGTCGTCATCA
5	199N18T7R2	5′ TACACAATTTTGCCCACC
6	199N18T7R3	5′ TCAGGAATTAAAGCCTCC
7	199N18T7R	5′ TTCCTTCAATCCCCACC
1	T3	5′ ATTAACCCTCACTAAAGGGA
8	T7	5′ TAATACGACTCACTATAGGG

because many complement proteins have SCRs, and the sequences obtained did not match to *Sp152*, which encodes SpBf with five SCRs (Smith et al. 1998).

Messages and deduced proteins

Two arrayed cDNA libraries were screened using riboprobes made from clones generated by RT-PCR with primers specific for *Sp5* and *Sp5013*. Sixty *Sp5* clones and 30 *Sp5013* clones were characterized by mixed-primer PCR analysis and restriction digests to identify overlapping regions and to optimize sequencing strategy (Fig. 1). Both standard and internal primers were used to sequence 17 *Sp5* clones and 6 *Sp5013* clones (Tables 1, 2). Some of the sequence near the 3′ end of *Sp5* was provided by Z. Pancer. The sequenced region of the *Sp5* transcript was 5.96 kb and was composed of 5′UTR [36 nucleotides (nt)], 3′UTR (432 nt), and 5,472 nt of open reading frame (ORF) (Fig. 2). The sequenced region of the *Sp5013* transcript was significantly shorter, 2.751 kb, and composed of 3′UTR (42 nt), and 5′UTR (924 nt), with an ORF of 1,785 nt (Fig. 3). Although the 5′UTRs for both of these cDNAs were quite short, the entire 5′UTR for *Sp5013* may have been obtained because six clones had identical sequences at the 5′ end. However, it is possible that significant secondary structure in this region of the message may have blocked complete reverse transcriptase activity.

The deduced amino acid sequence encoded by *Sp5* and called "SpCRL" consisted of 1,829 amino acids (Fig. 2). The deduced amino acid sequence encoded by *Sp5013* and called "SpCRS" consisted of 595 amino acids (Fig. 3). The absence of a Kozak sequence surrounding the start codon for *Sp5* and the presence of a very short 5′UTR made a reliable identification of the correct start codon difficult. However, the choice of the probable start codon was based on (1) the presence of one in-frame stop codon 21 nt upstream in the 5′UTR in addition to a second stop that was out-of-frame and (2) an ORF following the ATG that began with a leader region (underlined, Fig. 2) as predicted by the pSignal program (Nielson et al. 1997). The cleavage site for the removal of the leader was predicted to occur after Ser27, which was directly followed by Cys30 of the first SCR. *Sp5013* had an imperfect Kozak sequence [GCC (A/G)CC ATG G] (Kozak 1987) (Fig. 3, underlined), two stop codons in the 5′UTR, and a hydrophobic leader putatively cleaved after Ser27, which was followed by Cys30 of the first SCR (Fig. 3). Neither sequence revealed the presence of a transmembrane region (TMPredict database, Hofmann and Stoffel 1993). Both messages had numerous stop codons in all reading frames in the 3′ UTR and both appeared to have polyadenylated stretches at the 3′ end of the sequence (however, see below). *Sp5* did not have a conserved polyadenylation signal sequence, but four were identified in the 3′UTR of *Sp5013*. Two of these were located just 5′ of the poly(A)$^+$ tail (Fig. 3, underlined). One AU-rich element (ATTTA, Asson-Batres et al. 1994) was identified in the 3′UTR of *Sp5013* (Fig. 3, underlined).

SpCRS and SpCRL domain structure

Analysis of the deduced amino acid sequences of SpCRL and SpCRS indicated that both were mosaic proteins with a variety of domains, many of which were typically found in complement regulatory proteins and some complement components. Overall, both proteins contained SCRs (18 in SpCRL and four in SpCRS), a FIMAC domain, two S/T/P-rich regions, and numerous N-linked and O-linked glycosylation sites (Figs. 2, 3, 4). In addition, SpCRL contained a fucolectin domain and a Cys-rich region. The S/T/P-rich regions were present in similar locations of both SpCRL and SpCRS (Fig. 4). The composition of Ser and Thr plus Pro in the N-terminal region was 17% in SpCRL

Table 2 Sequencing primers for *Sp5* listed in Fig. 1b. All annealing temperatures were between and 55.3°C and 57.2°C

Primer number in Fig. 1b	Primer	Sequence
2	215A3w42C5R-2	5′ CCATAGTCGTCGCC
3	42C5RT3	5′ ACGGCTATCCACATTCC
4	42C5T3F	5′ AGGAGAGTGGCGATAC
5	33C8RT3	5′ ACAGTCATAATGTACTCTGG
6	211H15RT3	5′ GATTGTGTGCACGTAGG
7	229H19T3F	5′ CACCCGACGCCAAAT
8	210I5R1	5′ CCTGCTGTGTGGCAA
9	36C1RT3	5′ GGAGTGAGCAACGAGAA
10	CR.R1	5′ TCTACGTGCAGTTTGCTGAG
1	T3	5′ ATTAACCCTCACTAAAGGGA
11	T7	5′ TAATACGACTCACTATAGGG

Sp5 and SpCRL

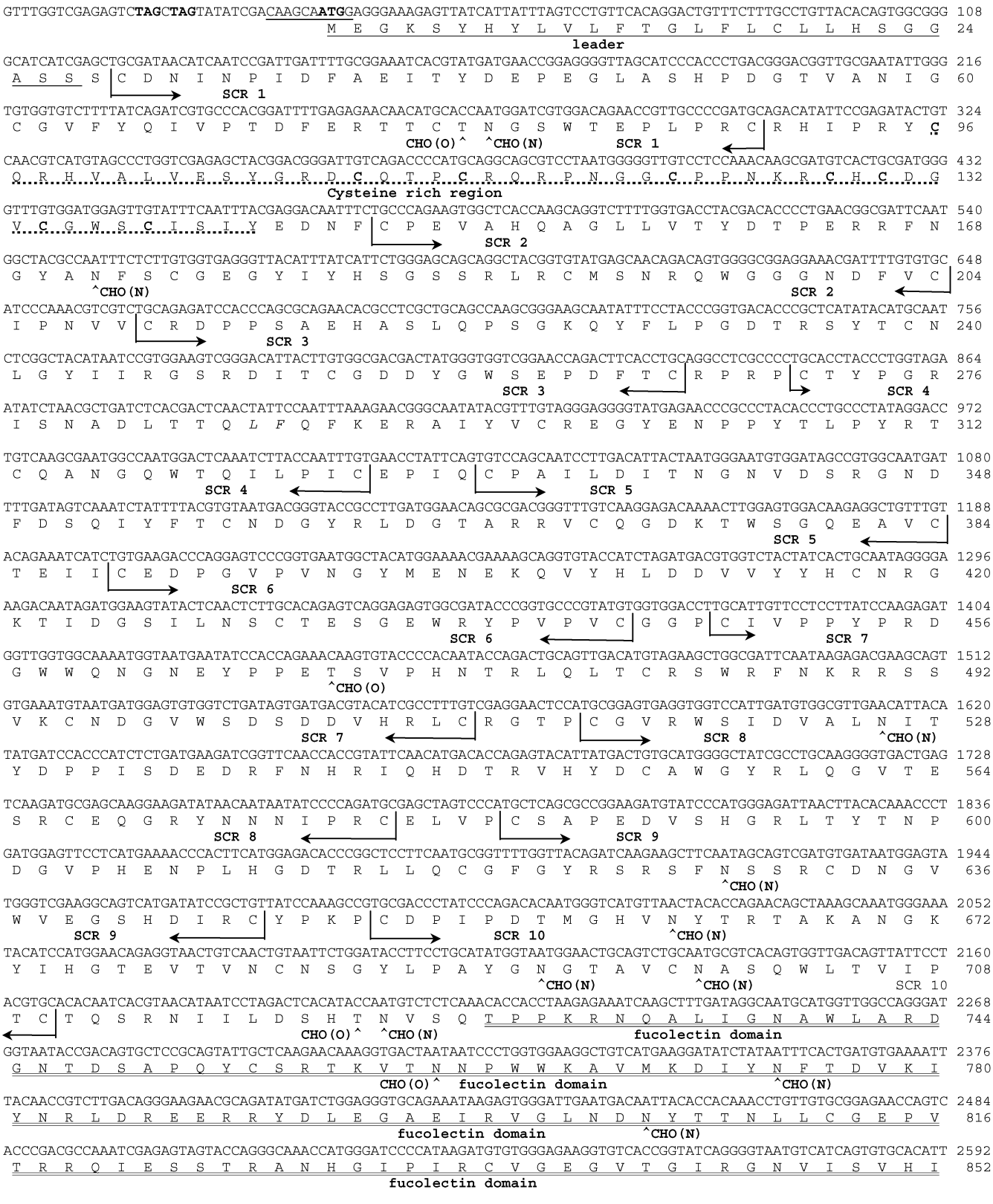


Fig. 2 cDNA sequence of *Sp5* and the deduced protein sequence of SpCRL (*Strongylocentrotus purpuratus* complement related protein, long form). Overlapping sequencing passes were used to generate the cDNA sequence and the deduced amino acid sequence was translated using DNASIS (Hitachi, Cascade, Colo.). The signal sequence was predicted using the Expasy SignalP V1.1 program. The start codon and stop codons, including those in the 5′UTR and 3′UTR, are indicated in *boldface*. The Cys in the Cys-rich region and the Ser, Thr, Pro in the Ser/Thr/Pro (S/T/P)-rich regions are also indicated in *boldface*. All domains are labeled and/or numbered, and the limits of each SCR is indicated with an *arrow*. The N- and O-linked glycosylation sites are noted with *CHO(N)* and *CHO(O)*, respectively

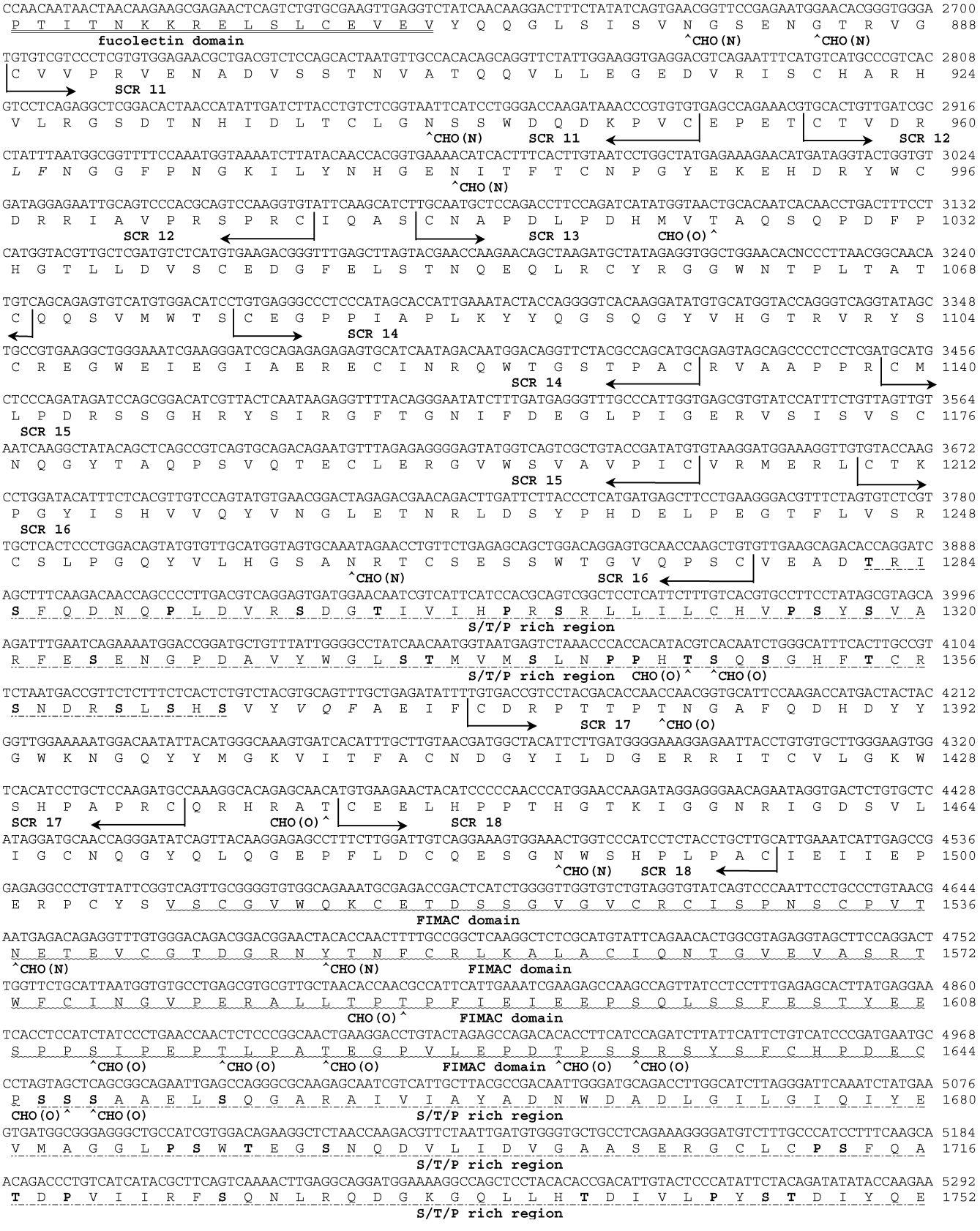


Fig. 2 (continued)

```
ATAAGACTGCTATTAGAAGAGCCTGTTCGTGTACAGACCCAAGCACCCCAGGCACCTCACCCACAAGATGAATCGAGCGACAGCTTTAGCAGCCTTAACAGCCGTAAT 5400
 I  R  L  L  L  E  E  P  V  R  V  Q  T  Q  A  P  Q  A  P  H  P  Q  D  E  S  S  D  S  F  S  S  L  N  S  R  N  1787
                                  CHO(O)^           S/T/P rich region CHO(O)^  ^CHO(O)
GGTCCAGAGGGACAAAATTCATGGTTTCAATCTGATTCAGACATCATTACTACCACGGCCCGCACCAACACAGAGATCGTTGATTCGCCGCGTTCATCAGAATCCTCT 5508
 G  P  E  G  Q  N  S  W  F  Q  S  D  S  D  I  I  T  T  T  A  R  T  N  T  E  I  V  D  S  P  R  S  S  E  S  S  1824
                                                 CHO(O)^  ^CHO(O)  S/T/P rich region       CHO(O)^  CHO(O)^
TTTGACACATAATCTTGTCTAAGAGTGGATTGTACAGGTGGCTCATACAGACTTTTTTTTGTTGTTATTTTACAATGCCATCAAATTGCTTTCTCTTCTTCCACTTCT 5616
 F  D  T  STOP                                                                                                 1829

TTTTGTCTTCCTCTTCTTCTTCCTTCCCTTCTTCACCTTTCTCCTACTTATCATTATTATATTATTTTCTCTACTCTTTATTTTAATTTTTCTTTCCTTCTCAAATTC 5724
ACTCTCCTTGTACGTCTCTAGAATCTCACATTATACTACTCCATTCTCGCCTCCTATAATACCTTTTAAATTCAGACCTACTATATTGTGTTAATCAGTACGGGAGCC 5832
TATTCACATACAAGCTATGCTTTTTATTTAGGCTCCCCACATTTTACATCGTTTTATCTACATATTTCAAAATTTTGTACTTATTTTTCGTACATTATACCTGATTTG 5940
AAATCTTTAATTAAAAAAAAAAAAAAAAAA 5965 3'
```

Fig. 2 (continued)

and 27% in SpCRS, while the same composition in the C-terminal region was 24% and 38%, respectively. In SpCRS the first S/T/P-region was located between SCRs 2 and 3, the FIMAC domain was located C-terminal of the SCR domains, and the second S/T/P-rich region was located C-terminal to the FIMAC domain. The 18 SCRs in SpCRL were spread throughout the protein and the Cys-rich region, which had eight Cys in a span of 57 amino acids, was located between SCRs 1 and 2 (Fig. 4). A fucolectin domain in SpCRL was positioned between SCRs 10 and 11, with the first S/T/P-rich region between SCRs 16 and 17. The FIMAC domain in SpCRL was located between SCR 18 and the second S/T/P-rich region.

Glycosylation sites in each protein were identified by searching for the conserved amino acid sequences NXS or NXT, and two prediction programs for O-linked glycosylation sites were used to identify O-linked sites (Hansen et al. 1997, 1998). Both proteins had many conserved sites for both O-linked and N-linked oligosaccharides (Figs. 2, 3). There were 22 O-linked glycosylation sites in SpCRL and 17 in SpCRS, while SpCRL had 17 N-linked glycosylation sites and SpCRS had seven (Fig. 4). In both proteins, many (40% for SpCRL and 64% for SpCRS) of the conserved glycosylation sites were located in the C-terminal S/T/P-rich region (Fig. 4). The rest of the sites in both proteins were located throughout the sequence. In general, both proteins showed similar distributions of the conserved glycosylation sites, and both may be highly glycosylated, particularly at the C-terminal end.

Phylogenetic analysis of domains

An approach for predicting functions of domains and proteins as a whole when the only available data are sequences is to align the amino acids with other sequences from known proteins and to use phylogenetic analysis to identify similarities which can be used to infer function. For the SCR, FIMAC, and fucolectin domains, there were enough sequences available from other proteins that have been characterized previously, some with known function, to make this approach feasible. On the other hand, this approach was not feasible for the S/T/P-rich regions and the Cys-rich region. It is noteworthy that an S/T/P-rich region is present at the C-terminus of MCP, which is involved in protecting self-cells from autologous complement attack; however, the functions mediated by the S/T/P region of MCP are not clear (Liszewski and Atkinson 1992).

Short consensus repeats

SCRs are found in many proteins that function in the complement system; however, many are also present in immune-related proteins that are not complement components, and others that are present in proteins that are not involved in immune responses. The conserved amino acids within SCRs include four Cys, three Gly, two Pro, two Tyr or Phe, and one Trp (Chou and Heinrickson 1997). An alignment of the SCRs from both SpCRL and SpCRS was done using ClustalX (Thompson et al. 1997) and demonstrated that they were typical SCRs, with four Cys and at least 9 of the 12 conserved amino acids present in each domain (Fig. 5). When the alignment was used to generate an unrooted neighbor-joining tree (Saitou and Nei 1987), results showed that the four C-terminal SCRs in SpCRL (SCRs 15–18) were most similar to the four SCRs in SpCRS, and that they were positioned in the same relative order within the two proteins. (This similarity in SCR sequence and domain organization is indicated in Fig. 4 by fill patterns.) This region, which included the four C-terminal SCRs from SpCRL, an N-terminal S/T/P-region, and the FIMAC domain but excluded the C terminal S/T/P-region (which was quite different in sequence and length between the two deduced proteins), was 37% identical and 56% similar between the two proteins (Fig. 4). For the remaining SCRs in SpCRL, four (SCRs 1, 8, 9, and 13) clustered together, with SCR 8 being most similar to SCR 13, and SCR 1 being most similar to SCR 9. Two additional pairs of similar SCRs were identified in SpCRL; SCR 4 was most similar to SCR 14, and SCR 7 was most similar to SCR 12 (denoted with similar patterns in Fig. 4). Although the relative orders of the SCRs at the C-terminus of SpCRL were the same as those in SpCRS, the two sets of SCRs within SpCRL that show sequence similarities were not positioned in the same order (1, 4, 7, 8 vs 9, 14, 12, 13; see Fig. 4).

The phylogenetic analysis was expanded to identify sequence similarities between the SCRs from SpCRS

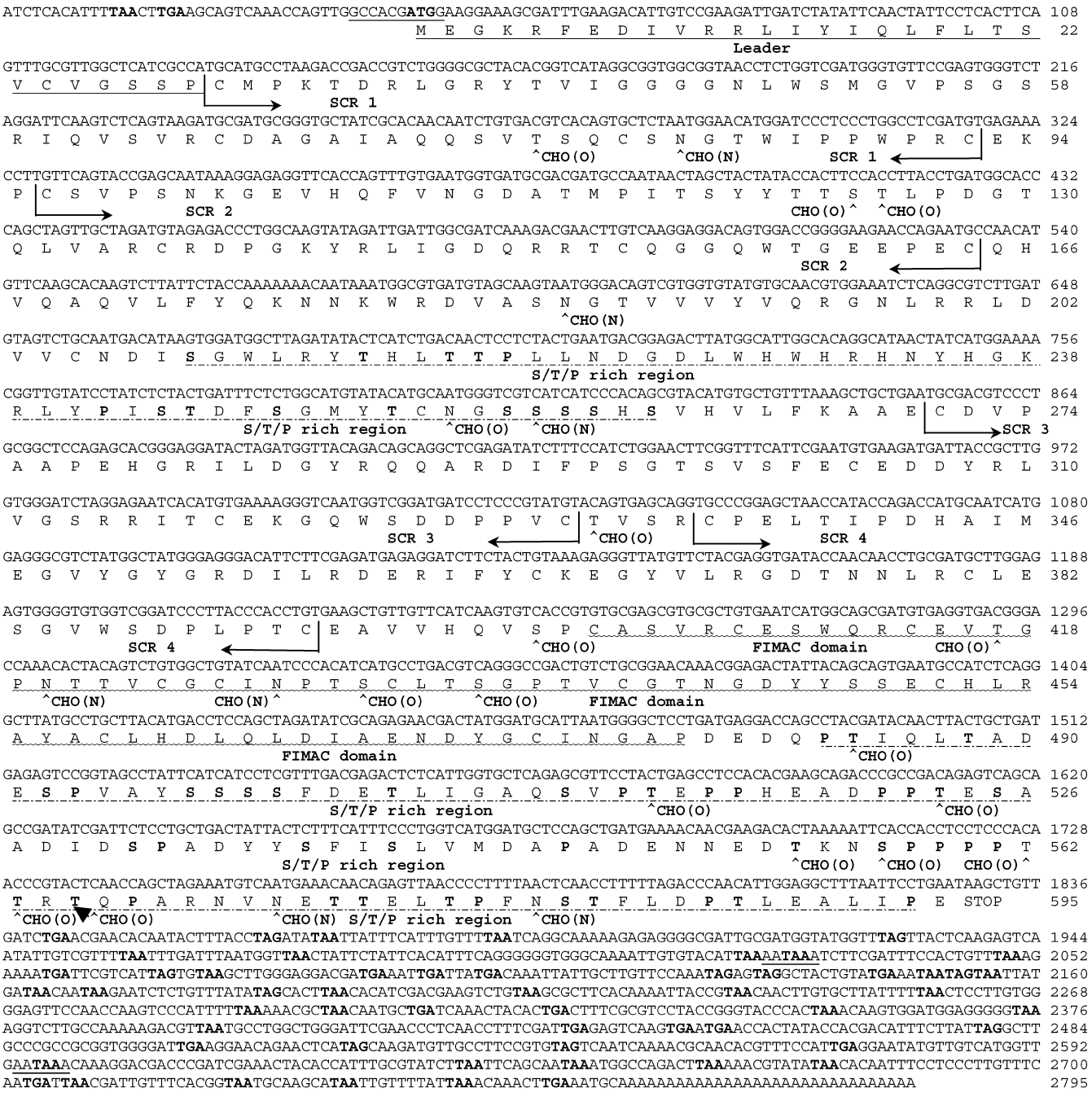


Fig. 3 cDNA sequence of *Sp5013* and the deduced protein sequence of SpCRS (*S. purpuratus* complement related protein, short form). The nucleotide and amino acid sequences were generated as described for *Sp5* (see legend for Fig. 2). The leader, domains, and conserved sites are labeled as in Fig. 2. An almost-perfect Kozak sequence (GCC(A/G)CCATGG) surrounds the start codon and is *underlined* (Kozak 1987). Three poly(A)$^+$ signal sequences (AATAAA) and one non-consensus-type poly(A)$^+$ signal sequence (ATTAAA) plus one AU-rich element (ATTTA) are present in the 3′UTR and are *underlined*

and SpCRL with SCRs from both complement and non-complement proteins. GenBank was queried using BLASTX with the entire SpCRL and SpCRS protein sequences to identify proteins that contained SCRs that were similar to those in SpCRL and/or SpCRS. This search identified SCRs from both complement and non-complement proteins (see footnote to Table 3) and all SCRs (n=549 from 87 proteins) were aligned and used to generate a neighbor-joining tree (not shown). Clustering of SCRs implied sequence similarities, which were used to infer putative functional similarities (Table 3) and have been indicated with symbols in Fig. 4. Many SCRs from proteins not involved in complement or immune function that were included in this analysis did not cluster with the

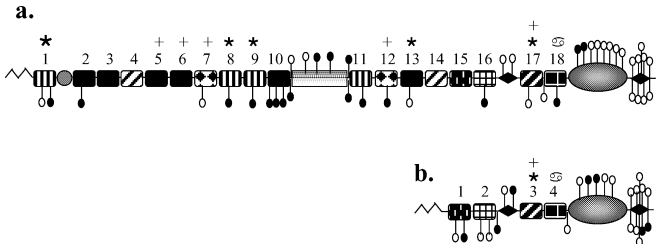


Fig. 4 Overview of the domain structures of SpCRL (**a**) and SpCRS (**b**). The structure for SpCRL and SpCRS are positioned to illustrate the similarities between the C-terminal end of SpCRL and the structure of the entire SpCRL. Domains are indicated as follows: The leader is shown as a *jagged line* at the N-terminal end of both proteins. Short consensus repeats (SCRs) are numbered and indicated with a *black rectangle*, and similar *fill patterns* indicate SCRs with similar sequences. The Cys-rich region in SpCRL is shown with a *dark gray circle*, and the fucolectin domain is indicated with a *variegated gray rectangle*. The S/T/P domains are shown in both proteins as a *black diamond* and the factor I-membrane attack complex (FIMAC) domain as a *variegated gray oval*. Inferred functions of some SCRs, based on phylogenetic analysis, are shown above the SCR as a *star* to indicate putative binding to SpC3, a *plus sign* to indicate putative binding to polyanionic structures on cell surfaces such as sialic acid or heparin, and a *(Zodiac) Cancer symbol* to indicate cell adhesion function. Conserved N-linked glycosylation sites are shown with a *black circle/black line* and O-linked sites are shown with a *white circle/black line*

SCRs from SpCRL and SpCRS (see footnote to Table 3). The most significant result of this analysis suggested that SCRs 5, 6, 7, and 12 from SpCRL have sequence similarities with SCRs in vertebrate Bf or C4BP, which are known to bind C3b or C4b, respectively (Hourcade et al. 1995; Blom et al. 2001), and are also known to have cofactor activity (Gordon et al. 1995; Kuhn et al. 1995) (Table 3). SCRs 1, 8, 9, 13, and 17 from SpCRL clustered with SCRs in factor H that bind sialic acid (Gordon et al. 1995) and/or C3b (Zipfel et al. 1999; Jokiranta et al. 2000). Because the C-terminal SCRs in both SpCRL and SpCRS were similar to each other and were positioned in the same relative order within the proteins (see above), it was not surprising that they clustered with each other and with the same SCRs from other proteins. The similarities identified for the C-terminal SCRs in both sea urchin proteins implied functions such as maintenance of structural conformation (SCR 1 from SpCRS, SCR 15 from SpCRL), cofactor activity and binding functions for sialic acid and C3b (SCR 3 from SpCRS, SCR 17 from SpCRL; Zipfel et al. 1999), and putative cell-adhesion activities (SCR 4 from SpCRS, SCR 18 from SpCRL; Fries et al. 1993). Overall, these results suggested the possibility that some of the SCRs in SpCRL and SpCRS may be involved in functioning within the complement system of the sea urchin and perhaps through interactions with SpC3. Furthermore, because cell surface sialic acid in mammals is important in blocking the alternative and lectin pathways of complement through binding factor H (Jarvis et al. 1987; Jack et al. 2001), SCRs in the sea urchin proteins that show sequence similarities to SCRs with sialic acid-binding activities suggest that SpCRL may also associate with cell surfaces and perhaps function in complement regulation.

FIMAC domains

Two types of proteins are known to contain FIMAC domains: the heavy chain of factor I and complement components C6 and C7 (Minta et al. 1996). The modular structure of factor I heavy chain consists of a FIMAC domain, a SRCR domain, and two LDL-receptor domains, while the light chain is a serine protease domain (reviewed by Arlaud et al. 1998). The FIMAC domain in factor I is also thought to be involved in protein-protein interactions (DiScipio 1992). The domain organization of complement components C6 and C7 have one or two thrombospondin domains, a type A LDL-receptor domain, a perforin-like domain, another thrombospondin domain, and two SCRs followed by two FIMAC domains (DiScipio and Hugli 1989; Haefliger et al. 1989). The SCR domains and the FIMAC domains in C6 are known to bind C5b (DiScipio 1992; DiScipio et al. 1999). Initial analysis of the C-terminal region in both SpCRL and SpCRS identified a single Kazal or FIMAC domain based on BLASTX searches of GenBank. This sequence has been identified previously as sea urchin coelomocyte expressed sequence tags (ESTs) [EST132 (accession no. R61984), EST351 (accession no. R62091), and EST202 (accession no. R62023), Smith et al. 1996]. Kazal domains are part of the follistatin serine protease family and function as serine protease inhibitors (Schlott et al. 2002). FIMAC domains are a subset of that family; however, they function as complement cofactors by binding C3 convertase, which promotes the cleavage of C3b by factor I (Terado et al. 2002), or by binding C5b in the terminal pathway (DiScipio 1992; DiScipio et al. 1999).

To determine whether the domains from SpCRL and SpCRS were similar to FIMAC domains from factor I and C6/C7 or if they were more like Kazal domains in protease inhibitors, they were used in a phylogenetic analysis to determine with which set of domains they clustered. The data set included 79 Kazal and FIMAC domains from 44 proteins as identified by BLASTX searches of GenBank. The sequences were analyzed by parsimony using the ratchet algorithm (Nixon 1999) in PAUP (Sikes and Lewis 2001; Swofford 2002), and a portion of the 50% strict consensus tree is shown in Fig. 6a. All of the FIMAC domains, plus those from SpCRL and SpCRS and two Kazal domains, clustered together in a clade defined by node A. Within that clade, a subclade defined by node B included all factor I FIMAC domains plus the sea urchin domains. The FIMAC domains from complement components C6 and C7, plus the zebrafish FIMAC domains (from a protein with a structure consistent with a C6 homologue, accession no. AAH57429.1), clustered in two clades defined by C (Fig. 6a). This result suggested that the SpCRL and SpCRS domains were more similar to FIMAC domains than Kazal domains, and perhaps more similar to factor I FIMAC domains than to C6/C7 FIMAC

Fig. 5 SCR alignment. The SCRs from SpCRS and SpCRL were aligned using ClustalX (Higgins et al. 1996) and edited in Word (Microsoft). The consensus amino acids that have been defined for SCRs in general, and most of which are found in the SCRs shown here (between 33% and 100%), are indicated in *boldface* and on the *bottom line*. Some linkers between SCRs were truncated for the purposes of improving this alignment and are indicated with an *asterisk*

domains. Bremer support (Bremer 1988) was calculated for several of the internal nodes, and is shown below the branch near the node (for details, see legend to Fig. 6). Bremer support indicates the number of extra steps that is required in a longer tree before a clade is lost from the consensus tree (Kitching et al. 1998). Consequently, support was quite good for nodes A and B, within which SpCRL and SpCRS clustered. Bootstrapping analysis was conducted, but results were uninformative. This was due to the short length of the domain (84 informative positions) and the large number of sequences (79) that were analyzed. Bootstrapping is optimal when the data matrix has at least 1,000 informative characters (Kitching et al. 1998). Bootstrapping results may also have been confounded by the presence of several sequences, such as the FIMAC domain from carp factor I-A, which did not align well (see Fig. 6b). The carp sequence did not cluster repeatedly within a given clade but "jumped" to a variety of locations within clade A in different, but equally parsimonious, trees. An alignment of the sequences that clustered in the clades defined by nodes B and C, plus the carp sequence, is presented in Fig. 6b. For most sequences, including those from SpCRS and SpCRL, the conserved amino acids are present including 10 Cys, which are involved in forming the disulfide bonds within the domain (Terado et al. 2002), in addition to Trp, Gln/Glu, Arg/Lys, Pro, Val and Ala/Ser. The carp sequence does not align well, particularly in the C-terminal end of the domain.

Although the domains of both SpCRL and SpCRS cluster with FIMAC domains from factor I proteins, the domain structure of the sea urchin proteins, in which the FIMAC domains were preceded by two SCRs (Figs. 2, 3, 4), was similar to the domain structure of complement proteins C6 and C7 (DiScipio 1992). Factor I, on the other hand, has the FIMAC domain located at the N-terminal end of the protein and is not associated with adjacent SCRs (Arlaud et al. 1998; Terado et al. 2002). Together, the combination of sequence similarity to factor I and the structural similarity to C6 and C7 implies that the FIMAC domains in SpCRS and SpCRL may have protein-protein binding functions similar to that known for FIMAC domains in factor I (DiScipio 1992) or for C6 (DiScipio et al. 1999), both of which interact with C3 or C5—members of the thioester-containing protein family. It is interesting that the C6 homologues identified in *Amphioxus*

Table 3 Sequence similarities between short consensus repeats (SCRs) from *Strongylocentrotus purpuratus* complement related protein, long form (SpCRL) and *S. purpuratus* complement related protein, short form (SpCRS) and SCRs from other proteins

	Protein source of SCR	Similar SCR	Protein function, SCR function	Species	Accession nos.
SpCRL SCRs[a]					
1, 9, 8, 13	Factor H-related protein-5, factor H	SCR 6, SCR 13	Complement regulation, SCR13 binds sialic acid	Human, pig, mouse, rat	XP_037279.1, CAC81999.1, NP_034018.1, NP_569093.1
3	IL2 receptor	SCR 2	Binds IL-2, SCR function unknown	Mouse, rat	P01590, P26897
5	Factor B	SCR 1, SCR 2	Alternative pathway, SCRs bind C3b	Pig, mouse	Q03710, 67613
6	C4 binding protein	SCR 6, SCR 8	Complement regulation, SCR6 has cofactor activity, SCR8 binds C4b	Rat, human, mouse	Q63514, P04003, P08607
7, 12	Factor B	SCR 1	Alternative pathway, SCR1 binds C3b	Mouse	67613
10	MASP 2, C1r	SCR 1	Lectin and classical complement pathways, SCR function unknown	Human, mouse	NP_006601.2, NP_075632.1
15	Complement receptor type 2	SCR 8, SCR 11	Complement receptor type 2, binds C3d, SCRs function in structural conformation?	Mouse, sheep	XP_129684.1, AAB92375
17	Factor H, factor H-like protein 1	SCR 3	Complement regulation, SCR3 binds sialic acid, C3b	Mouse, rat, pig, human	NP_034018.1, NP_569093.1, CAC81999.1, NP_002104.1
18	P selectin	SCR 2, SCR 3	Cell adhesion, SCR function unknown	Mouse, rat, human	NP_035476, NP_620234.1, NP_002996.1
SpCRS, SCRs[a]					
1	Complement receptor type 2	SCR 8, SCR 11	Receptor for C3d, SCR function structural conformation?	Mouse, sheep	CAB03143.2, XP_129684.1, AAB92375
2	Complement receptor type 1	SCR 6, SCR 13, SCR 20, SCR 27, SCR 34	Complement regulation, SCR function structural conformation?	Human	I73012
3	Factor H, factor H-like protein 1	SCR 3	Complement regulation, SCRs bind sialic acid and C3b	Mouse, rat, pig, human	NP_150094, NP_444401, NP_034018.1, NP_569093.1, CAC81999.1, NP_002104.1.1, XP_037279.1
4	P selectin	SCR 2, SCR 3	Cell adhesion, SCR function unknown	Mouse, rat, human	NP_035476, NP_620234.1, NP_002996.1

[a] SCRs used in this analysis were obtained from the following proteins: factor B (accession nos. NP_571413, AAA31021, P81475, NP_032224.1), mouse complement component C2 (XP_123064.1), C4 binding protein (NP_000706.1, NP_000707.1, P08607NP_036648.1, NP_058691.1, NP_031602.1), factor H (NP_569093.1, NP_034018.1, CAC81999, gil2135094), factor H-related proteins (XP_037279.1, CAA66980.1, NP_002104.1, CAA48639.1), MASP (NP_006601.2, XP_029605.1, XP_148328.1), C1s (13787045), C1r (NP_075632.1), selectins (NP_002996.1, NP_035477.1, NP_000646.1, NP_037246.1), type 1 and 2 complement receptors (21536276, A46458, XP_002008.7, AAB92375, XP_129684.1), chondroitin sulfate proteoglycan (NP_113841.1), versican (NP_004376.2, XP_127448.1), aggrecan (NP_071526.1), neurocan (XP_125051.1), brevican (NP_031555.1, NP_068767.2), MCP (NP_002380.2, NP_062063.1, NP_034908.1), DAF (NP_000565.1, NP_034146.1, NP_031853.1, P49457), haptoglobin (XP_042621.1, NP_059066.1), scavenger receptor cysteine-rich protein (AAB40715.1, T17405), apolipoprotein receptor (AAA30994.1), apolipoprotein H (NP_038503.1), seizure gene 6 (XP_114203.1, XP_126232.1), pregnancy protein A (NP_002572.1, XP_131437.1), polydomain protein (NP_078776.2, NP_073725.1), sushi domain protein (NP_081114, NP_150094.1, NP_071969.1, NP_058607.1), thyroid peroxidase (NP_062226.1, NP_033443.1, NP_000538.2), interleukin receptors (NP_037295.1, NP_032384.1, NP_032393.1), complement component C6 (XP_122801.1), complement component C7 (NP_000578), C3b/C4b receptor (I73012), protein X (NP_006298.1), human KIAA1884 (XP_055539.6), *Drosophila* scavenger receptor (NP_477102, NP_524747), *Drosophila* hikaru genki (Q09101), β-2 glycoprotein (AB20668), pox virus complement inhibitor (gil10120606), GABA B (XP_165689.1), p100 serine protease (BAA03944.1), factor XIII β (NP_112441.1), zona pellucida 3 receptor (NP_033607.1), *Caenorhabditis elegans* K07E12 (AAA50715.1), *C. elegans* LDL receptor (CAB03143.2), *C. elegans* SCRs (NP_509052.1), *C. elegans* AF304125 (AAG50238.1), and breast cancer antigen (AF308289_1)

(Suzuki et al. 2002) and *Ciona* (Azumi et al. 2003) have a domain organization that is similar to the N-terminal half of vertebrate C6 and C7, including the thrombospondin domains, a type A LDL-receptor domain, and a perforin-like domain, but lack the SCR and FIMAC domains. Consequently, the C6-like sequences in these two lower chordates share no domains with SpCRL and SpCRS, which also appear to be C6-like. Determination of whether the chordate or echinoderm C6-like proteins might actually function in a terminal pathway will require additional analysis of complement activities from these invertebrates.

Fucolectin domain

A BLASTX search of the region between SCR 10 and SCR 11 in SpCRL revealed a significant sequence similarity with a family of fucolectins that have been characterized from the serum of the European eel, *Anguilla anguilla*, and have been denoted the *A. anguilla* agglutinins (AAA) (Bianchet et al. 2002). Fucolectins are a class of lectins that recognize fucose on the surface of cells and play an important role in the innate immune functions of both vertebrates and invertebrates as pattern-recognition receptors (Bianchet et al. 2002). An alignment

between the SpCRL fucolectin domain and the seven eel fucolectin proteins revealed a number of conserved amino acids in addition to several positions where the eel and sea urchin sequences differed (Fig. 7). The fucolectin structure, which is diagnostic of the F-lectin family, is a β barrel with jellyroll topology of eight antiparallel β strands oriented in two sheets consisting of five and three strands (Bianchet et al. 2002). The domain is locked into its conformation with two disulfide bonds and two salt bridges. The disulfide bonds were conserved in both the eel fucolectins and the SpCRL fucolectin domain (Fig. 7). The salt bridges were formed between Arg41-Glu140 and Asp64-Arg131 in the eel sequences that interact with a cation—either Na^+ or Ca^{2+}. In SpCRL, neither Arg involved in salt bridges were present in the conserved positions, but were located within two to four positions of the site conserved in the eel sequences. In the AAA fucolectins, there are five loops connecting the β strands of the β barrel that protrude like complementarity-determining regions (CDRs) and surround the fucose-binding site. The CDRs in SpCRL showed conserved amino acids that defined the borders of the regions and the sizes of the regions were similar to that in AAA. The fucose-binding site, composed of the motif $H(X)_{24}$RGDCC(G/E)ER, showed conservation of the significant amino acids in SpCRL except for His52, which was replaced with an Arg, and the double-Cys motif, which was missing in SpCRL. Fucolectins have been shown to have hemagglutination activity mediated by the conserved residues AIDGN located between CDR 1 and CDR 2, a motif that was partially conserved in SpCRL, ARDGN. The detailed structural analysis of fucolectins reported by Bianchet et al. (2002) has enabled a detailed analysis of the fucolectin domain of SpCRL and overall, the alignment suggests that the sea urchin protein has a fucolectin domain, but comparisons revealed a divergent structure from that of the AAA proteins (G. Vasta, personal communication). Variations in amino acids required for salt-bridge formation to stabilize the barrel structure and the missing His52 which is involved in fucose binding in AAA, make it possible that the sea urchin domain may bind a different ligand than that characterized for AAA. The location of the ligand, perhaps on self-cells or on the pathogen, may provide a clue as to how SpCRL might be involved in the complement system of the sea urchin.

Transcript sizes and alternative splicing

Sp5

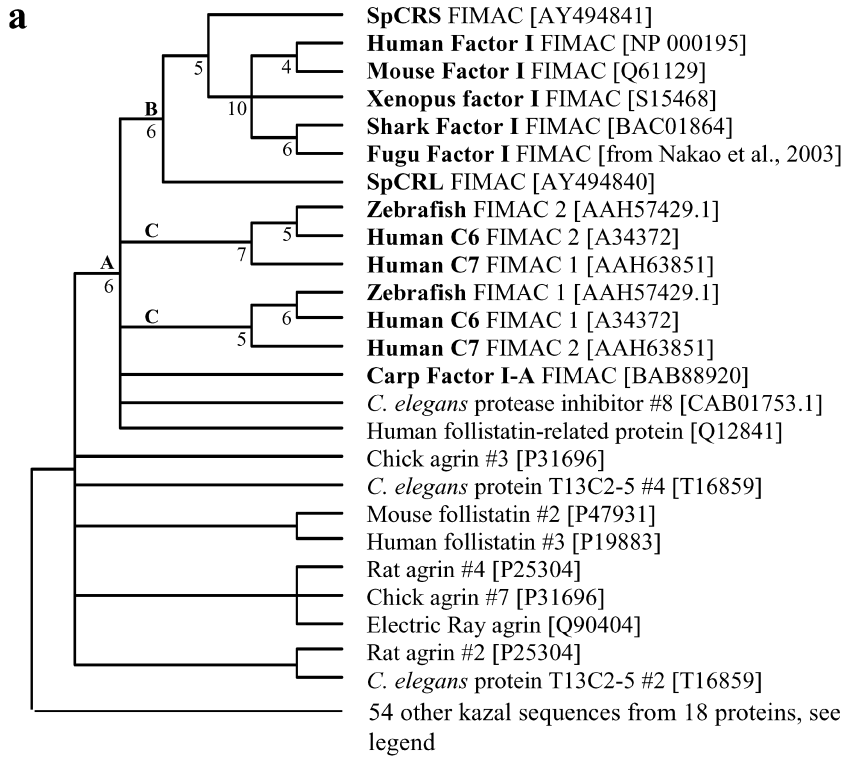
The mRNA length for both cDNAs were determined by poly(A)$^+$ Northern blots (Fig. 8). Results for *Sp5* (Fig. 8a) showed two pairs of bands (7.4 kb and 7.9 kb, 11.9 kb and 12.7 kb) with the smaller of each pair being more intense. All major bands were larger than the sequenced *Sp5* cDNA (5.965 kb) suggesting that parts of the 5′UTR and 3′UTR were missing. The 5′UTR was very short, and it is feasible that much of this region was missing. However, the blot and the cDNA sequence both suggested that the short stretch of As at the 3′ end was not the actual poly(A)$^+$ tail, indicating that some of the 3′UTR was also missing. This was supported by the absence of a polyadenylation signal sequence located 5′ of the poly(A)$^+$ stretch (Fig. 2). Taking this into consideration, the known sequence for *Sp5* may correspond to either the 7.4-kb or 7.9-kb band. Longer exposures of the same blot showed five less prominent bands ranging in size from 1.2 kb to 3.8 kb (Fig. 8b). These bands may not correspond directly to the *Sp5* sequence because they were significantly shorter than the known sequence of the cDNA and did not hybridize well to the probe (see below for a discussion of small gene family and alternative splicing).

Sp5013

The Northern blot for *Sp5013* revealed three major bands of 6.0 kb, 4.6 kb, and 3.8 kb after a short exposure (Fig. 8c), plus six weakly hybridizing bands of 12.7 kb, 11.4 kb, 9.3 kb, 7.4 kb, 2.9 kb, and 2.2 kb that appeared after a longer exposure (Fig. 8d). The known sequence of *Sp5013* (2.795 kb) was not exactly the same size as any of the bands on the Northern; however, the cDNA might correspond with either the 2.9-kb or 3.8-kb band if parts of both the 5′UTR and 3′UTR were missing from the cDNA sequence. The 3′UTR was probably complete because the poly(A)$^+$ tail was preceded by two polyadenylation signal sequences located 100 nt to 150 nt upstream. Although we argue above that the 5′UTR was complete based on duplicate cDNAs with the same sequence at the 5′ end, the shortness of this region is not typical for sea urchin mRNA.

Similar band sizes and alternative splicing

Size comparisons between the array of bands for *Sp5* and *Sp5013* in Fig. 8 indicate that a few may be the same. However, bands of the same size never appeared as major bands for both probes. The best example of message sizes present on both blots was the 12.7-kb band, which was one of the major *Sp5* doublet bands and was a minor band for *Sp5013* (Fig. 8b, d). Another example was the 3.8-kb band, which was a major band on the blot for *Sp5013* but was a minor band for *Sp5*. Finally, the 7.4-kb band, which appeared on both blots, was a major band for *Sp5* and a minor band for *Sp5013*. In addition, there was a set of minor bands of 2.8 kb/2.9 kb that was present on both blots. The probes used on the Northern blots were produced from clones that included the entire *Sp5013* sequence, and that spanned the 3′ half of the *Sp5* sequence (the fucolectin domain to the 3′ end, see legend to Fig. 1). These were the regions of the two messages that showed the greatest similarity and may have resulted in some cross-hybridization, even at high stringency. Because the small bands did not correspond to sizes estimated from the known sequences of the cDNAs, this suggests that the



b

```
SpCRS FIMAC       -------CASVRCESWQRCEVTGPNTTVCGCINPTS-C-LTSGPTVCG
SpCRL FIMAC       IIEPERPCYSVSCGVWQKCETDSSGVGVCRCISPNS-CPVTNETEVCG
Human Factor I    ------SCDKVFCQPWQRCIE-----GTCVCKLPYQ-CPKNG-TAVCA
Mouse Factor I    ------SCNKVFCQPWQRCIE-----GTCICKLPYQ-CPRAG-TPVCA
Xenopus Factor I  ------SCHKVFCAPWQRCVA-----GVCRCKLPYQ-CPKNATTEVCT
Shark Factor I    -----KSCQKVFCQPWEKCIN-----GRCECKLPYQ-CPKQV-NEVCS
Fugu Factor I     RRYTRQSCDLVFCPPWERCLD-----GQCLCKVPYQ-CPSENVTAVCG
Human C6 FIMAC 1  ------TKLKGHCQLGQKQSG-----SECICMSPEEDCSHHS-EDLCV
Human C6 FIMAC 2  ------SCGYDTCYDWEKCSAS---TSKCVCLLPPQ-CFKGGNQLYCV
Human C7 FIMAC 1  ------TQAVPKCQRWEKLQN-----SRCVCKMPYE-CGPS--LDVCA
Human C7 FIMAC 2  PASAEKACG--ACPLWGKCDAE---SSKCVCREASE-CEEEG-FSICV
Zebrafish FIMAC 1 --------PDSSCKPGEINDG-----TKCVCMTKER-CRGYR-EDLCV
Zebrafish FIMAC 2 -----EPCGSDTCYEWETCSVS----KTCECKMPRE-CPKDGKKIYCL
Carp Factor I-A   ----MRAVFYFMCLLFQTALN--------QPKVPDE--DFLG-PAQCL
conserved positions      C  V C  WQKC         C C  P   C       VC
                                 ER

SpCRS FIMAC       T-NGDY---YS-SECHLRAYACLHDLQ--LDIAENDYGCINGAP---
SpCRL FIMAC       T-DGRN---YT-NFCRLKALACIQNTG--VEVASRTWFCINGVP---
Human Factor I    T-NRRS---FP-TYCQQKSLECLHP----------GT----------
Mouse Factor I    M-NGRS---YP-TYCHQKSFECLHPE---IKFSHNGT-CAA------
Xenopus Factor I  D-GKRK---LQ-SYCQLKSVECSNPLNSKYRFSSEAP-CTE------
Shark Factor I    S-RGKK---YR-SYCQLKSIECIRGLES---FSHFGM-CSMGT----
Fugu Factor I     R-DGRN---YR-SYCQVMAVSCRTKSP---KFSHFGQNCAVVRVFVP
Human C6 FIMAC 1  FDTDSND-YFTSPACKFLAEKCLNNQQ--LHFLHIGS-CQD------
Human C6 FIMAC 2  K-MGSSTSEKTLNICEVGTIRCANRK---MEILHPGK-CLA------
Human C7 FIMAC 1  Q-DERSKRILPLTVCKMHVLHCQGRN---YTLTGRDS-CTL------
Human C7 FIMAC 2  EVNGKE---QTMSECEAGALRCRGQS---ISVTSIRP-CAA------
Zebrafish FIMAC 1 YDAGKET-AIMMSLCAFHADRCHGDRL---YFMNNGP-CKSD-----
Zebrafish FIMAC 2 KIVRTQTTRSMN-LCFMAAMKCSSIEF---ELQHEGP-CAGS-----
Carp Factor I-A   D---QK---YT------R-LSCSKVF--------------------
conserved positions            C   A  C               C
                                   S
```

Fig. 6a, b Phylogenetic analysis and alignment of the FIMAC domains from SpCRL and SpCRS. **a** An alignment using 79 FIMAC and Kazal domains from 44 proteins was done using ClustalX (Thompson et al. 1997), edited in BioEdit sequence alignment editor (Hall 1999), formatted in Winclada (obtained from http://www.cladistics.com), and analyzed in PAUP (Sikes and Lewis 2001; Swofford 2002) to generate a phylogenetic tree. Multiple FIMAC and Kazal domains from the same protein are *numbered* and treated as independent sequences. Accession numbers are indicated in *brackets*. The most parsimonious tree was calculated using the parsimony ratchet algorithm, PAUPRat, version 1 (Nixon 1999), with 1,000 iterations, each with 20 random additions of taxa and a limit of 20 trees saved and swapped for each random addition. A portion of the 50% strict consensus tree is shown that was

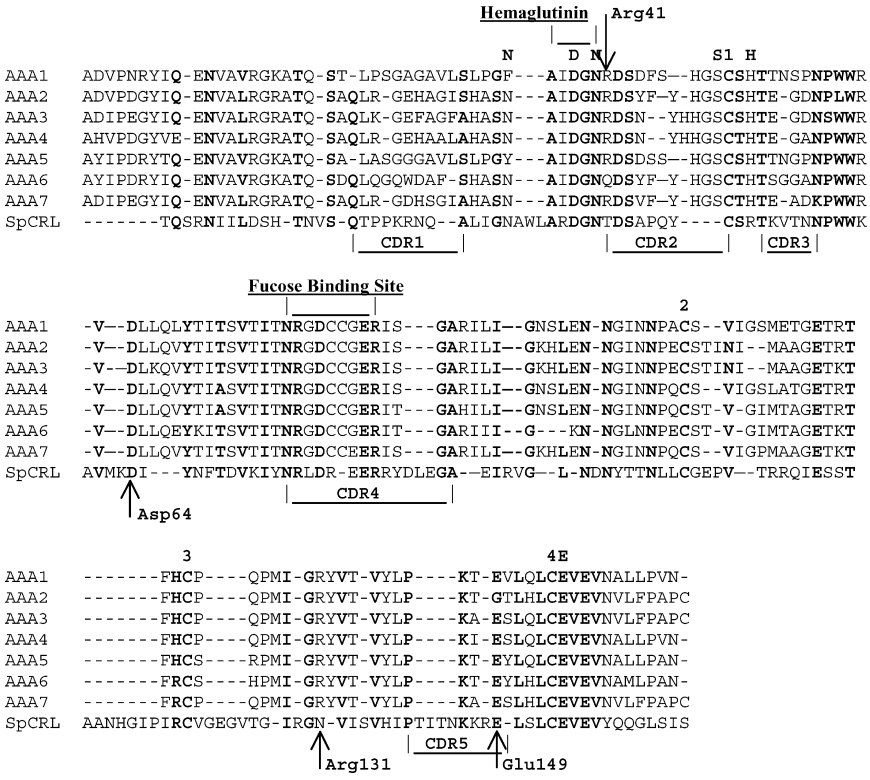


Fig. 7 Fucolectin alignment. The fucolectin domain from SpCRL was aligned with the seven fucolectins from the European eel, *Anguilla anguilla* [denoted *A. anguilla* agglutinins (AAA)], using ClustalX (Higgins et al. 1996) and edited in Word (Microsoft). The conserved amino acids are shown in *boldface* within the alignment. Amino acids directly involved in binding functions are shown in *boldface* in the *top line* in addition to the fucose-binding site and the hemagglutination site. The complementarity-determining regions (CDRs) are labeled and identified with *brackets*. The two Arg located within the fucose-binding site and the His located towards the N-terminus (in *boldface*, *top line*) are the triad of amino acids that are involved in the polar interactions with fucose. The first residue of this triad in SpCRL is Arg, which does not match the conserved His in the eel fucolectins. The Cys are identified with numbers on the top line and disulfide bonds are formed between Cys1–4 and Cys2–3 in AAA sequences (Bianchette et al. 2002). Four amino acids involved in stabilizing the domain through salt bridges in the AAA sequences are indicated with *arrows*. Salt bridges are formed between Arg41 and Glu149 and between Asp64 and Arg131 (numbering is based on Bianchette et al. 2002). Amino acids involved in cation binding are shown in *italics* on the *top line*

generated from 919 equally parsimonious trees. Bremer support (Bremer 1988) was calculated for several internal nodes using PAUP and TreeRot, version 2 (Sorneson 1999), and is indicated near the relevant nodes below the branches. Kazal and FIMAC domains (54) from 18 other proteins are not shown on the tree. They include bovine acrosin inhibitor [P01000], *C. elegans* protease inhibitor [CAB01753.1], chick agrin [P31696], chick follistatin [Q90844], chick ovoinhibitor [P10184], chick ovomucoid [P01005], crayfish proteinase inhibitor [CAA56043.1], *Dipetalogaster* thrombin inhibitor [CAA10384.1], dog double-headed protease inhibitor [P01002], eel pancreatic proteinase inhibitor [P11706], Eurasian badger double-headed protease inhibitor [P16226], herring sperm activating protein [BAA14008.1], human acrosin trypsin inhibitor [P20155], human agrin [AAC39776.1], human serine protease inhibitor [CAB40839.1], human prostacyclin stimulating factor [AAB32370.1], human testican 3 [NP-058646], human transmembrane protein [AAA64622.1], Japanese quail QR1 protein [P23499], lion double-headed protease inhibitor [P08481], leech protease inhibitor [AAK58688.1], mouse testican [CAA63448.1], mouse serine protease inhibitor Kazal type 4 [NP035593.1], pig sperm-associated acrosin inhibitor [P00999], rat pancreatic trypsin inhibitor [P09655_IPK1], sea anemone elastase inhibitor [P16895], turkey ovomucoid [P01004]. **b** An alignment of FIMAC domains from SpCRL and SpCRS with FIMAC domains that clustered in **a**. The alignment was done with ClustalX (Thompson et al. 1997) and edited in Word (Microsoft). Consensus amino acids are shown in *boldface* and on the *bottom line*

probes may have cross-hybridized to messages other than *Sp5013* and *Sp5*, implying the presence of a small gene family with similar sequences.

Multiple bands on the Northern blots may also be interpreted as evidence of alternative splicing. The pattern of size differences among the bands for *Sp5013* was present in multiples of 0.7–0.8 kb (Fig. 8e), which corresponds with regions of the message that would encode pairs of SCRs including the linker between them. For example, if the length of the message encoding SCR 1 and SCR 2 plus the linker were doubled, it would be 810 nt. Similarly, if the sequence encoding SCR 3 and SCR 4 plus the linker were doubled, it would be 738 nt (Fig. 8e). These sizes correspond with regular size differences noted between the bands for *Sp5013* (Fig. 8c, e) and could be interpreted as the result of alternative splicing of multiples of SCR exons. If this set of messages of varying lengths were transcribed, it would result in a set of proteins with many more SCRs and perhaps fewer SCRs than were identified from the deduced sequence of SpCRS. In support of this interpretation, several clones for *Sp5013* were identified with identical sequences at the 5′ and 3′

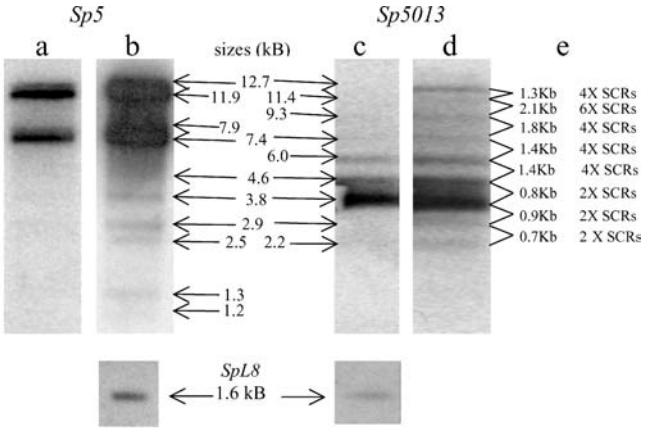


Fig. 8a–d Expression of *Sp5* and *Sp5013* in coelomocytes. Poly(A)$^+$ RNA was isolated from coelomocytes from an animal within 1 day of shipment from California and therefore assumed to be activated. Between 1 μg and 1.8 μg mRNA was loaded per lane on a Northern blot and analyzed with riboprobes made from *Sp5* and *Sp5013*. Clones used as riboprobe templates for this purpose included the entire sequence for *Sp5013* and approximately the 3′ half of *Sp5* (see legend to Fig. 1). The blot probed for *Sp5* was exposed for 30 min (**a**) and 5 h (**b**). The blot probed for *Sp5013* was also exposed for 30 min (**c**) and 5 hr (**d**). Expression of *SpL8*, the sea urchin homologue of the human ribosomal gene *L8*, was used as a control, and a single band of 1.6 kb is shown. Sizes of the bands are indicated *between* the two sets of blots. **e** Differences in observed sizes between bands for *Sp5013* (**d**) and the approximate number of deleted SCRs which might have resulted in the observed differences are indicated to the right (see text)

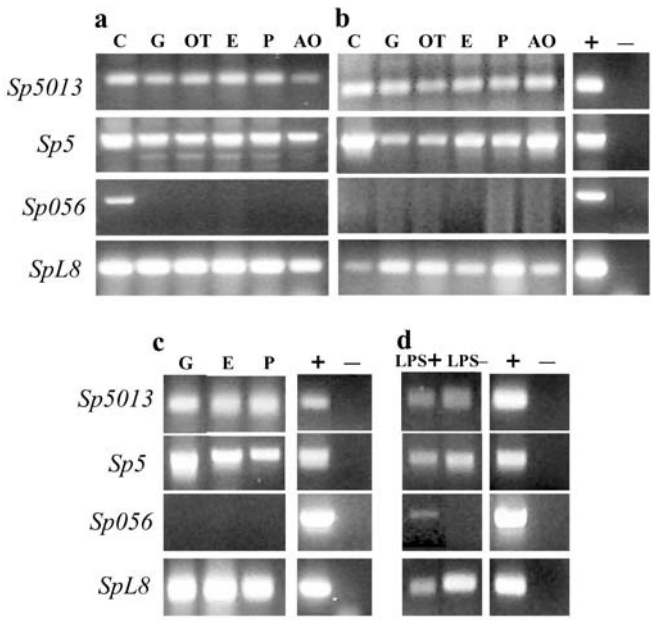


Fig. 9 Expression of *Sp5* and *Sp5013* in sea urchin tissues. Total RNA was isolated from coelomocytes, gut, gonad, pharynx, esophagus, and axial organ and used in reverse transcriptase (RT)-PCR. The RNA was analyzed for expression of *Sp5* and *Sp5013*, and *Sp056* (an LPS-inducible C-type lectin, accession no. AY336600, unpublished data). *SpL8*, a constitutively expressed homologue of the human ribosomal gene *L8*, was used as the control. **a** A sea urchin was injected with LPS for 2 days and coelomocytes and other tissues were collected for RNA isolation 24 h after the last injection. **b** The sea urchin used for analysis was assumed to be upregulated as a result of shipping stress. **c** Tissues were collected from an immunoquiescent sea urchin which had been kept in a closed system aquarium for 2 years prior to being used in this experiment (see Gross et al. 1999). **d** Coelomocytes were taken from an immunoquiescent sea urchin 15 min prior to a single LPS injection and 24 h post-injection. *C* Coelomocytes, *G* gut, *OT* ovary or testis, *E* esophagus, *P* pharynx, *AO* axial organ. Control PCR reactions employed cloned templates with the appropriate primers (positive) or reactions in which the cloned template was omitted (*negative*)

ends to the sequence provided here, but with protein-coding regions of variable sizes (data not shown). The deletion of SCRs by alternative splicing has recently been demonstrated for another sea urchin cDNA, *Sp152*, which encodes the homologue of complement factor B, SpBf (Terwilliger et al. 2000; unpublished data). Splice variants of *Sp152* were identified in which the first and/or the fourth SCR were/was deleted from the total of five.

Genome analysis

Genome blots using sperm DNA from two sea urchins that was digested with *Hin*dIII, *Eco*RI, and *Pst*I were analyzed with probes for each gene. Results showed multiple bands of variable intensity in both blots that would be consistent with gene structure of multiple exons plus cross-hybridization among members of a small gene family (data not shown).

Expression patterns of *Sp5* and *Sp5013* in sea urchin tissues

Expression patterns of *Sp5* and *Sp5013* were characterized by RT-PCR using several sets of gene-specific primers and total RNA from several sea urchin tissues (Fig. 9). The tissues examined included coelomocytes, gut, gonad, pharynx, esophagus, and axial organ and were isolated from sea urchins that were either activated by injection of LPS (Fig. 9a), were in an immune activated state (Fig. 9b), or were immunoquiescent (Fig. 9c). Results showed that all tissues expressed both *Sp5* and *Sp5013*, and that expression was not altered in coelomocytes from immunoquiescent animals after injection with LPS (Fig. 9d). In comparison, expression of *Sp056* (accession no. AY336600), which encodes a C-type lectin that is reliably inducible with LPS (unpublished data), was restricted to LPS-activated coelomocytes. A sea urchin homologue of the human ribosomal gene *L8*, *SpL8* (EST219, accession no. R62029; Smith et al. 1996) was used as the control to ensure that approximately equal amounts of cDNA template were used in the PCR reactions. Overall, the RT-PCR analysis indicated that *Sp5* and *Sp5013* were ubiquitously expressed and were not induced by immune challenge.

Summary

Analysis of the two mosaic proteins, SpCRL and SpCRS, indicated that they share domains with factor H, factor I, C6 and C7. Similarities among SCRs from a variety of complement proteins (Table 3) indicated that many of the SCRs in both SpCRL and SpCRS may function in interactions with other proteins such as SpC3, or with cell surfaces perhaps through binding carbohydrates (Fig. 4). The FIMAC domains in SpCRL and SpCRS show sequence similarities with FIMAC domains from complement proteins rather than with Kazal domains from protease inhibitors (Fig. 6). Functions of SpCRL might involve interactions with SpC3 as a regulatory protein to modulate complement activation, either in fluid phase or in association with self-cell surfaces, either through some of the SCRs and/or through the fucolectin domain. This is the first evidence of proteins in the sea urchin that might function in a putative complement regulatory system and/or perhaps in a putative terminal pathway. Knowledge of whether an expanded complement system exists in the sea urchin, as has been identified from the *Ciona* genome (Azumi et al. 2003), will also most likely come from the analysis of the sea urchin genome when it is completed. An understanding of all the components and pathways of the sea urchin complement system will provide a better understanding of the evolution of this system that has culminated in the complex complement system that is essential to immune functions in the higher vertebrates today.

Acknowledgements The authors would like to thank Dr. Sham V. Nair, David P. Terwilliger, and Dr. Zeev Pancer for their assistance with this research. We are greatly indebted to Fernando Alvarez-Padilla for his assistance in the phylogenetic analysis. We appreciated the helpful suggestions from the anonymous reviewers. This research was supported by funding from the National Science Foundation (MCB-0077970).

References

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

Arlaud GJ, Volanakis JE, Thielens NM, Narayana SV, Rossi V, Xu Y (1998) The atypical serine proteases of the complement system. Adv Immunol 69:249–307

Asson-Batres MA, Spurgeon LS, Diaz J, DeLougherty TG, Bagby GC (1994) Evolutionary conservation of the AU-rich 3′ untranslated region of messenger RNA. Proc Natl Acad Sci USA 91:1318–1322

Azumi K, De Santis R, De Tomaso A, Rigoutsos I, Yoshizaki F, Pinto MR, Marino R, Shida K, Ikeda M, Ikeda M, Arai M, Inoue Y, Shimizu T, Satoh N, Rokhsar DS, Du Pasquier L, Kasahara M, Satake M, Nonaka M (2003) Genomic analysis of immunity in a Urochordate and the emergence of the vertebrate immune system: "waiting for Godot." Immunogenetics 55:570–581

Barilla-LaBarca ML, Liszewski MK, Lambris JD, Hourcade D, Atkinson JP (2002) Role of membrane cofactor protein (CD46) in regulation of C4b and C3b deposited on cells. J Immunol 168:6298–6304

Becherer JD, Alsenz C, Servis B, Myones B, Lambris JD (1989) Cell surface proteins reacting with activated complement components. Complement Inflamm 6:142–165

Bianchet MA, Odom EW, Vasta GR, Amzel LM (2002) A novel fucose recognition fold involved in innate immunity. Nat Struct Biol 9:628–634

Blom AM, Kask L, Dahlback B (2001) Structural requirements for the complement regulatory activities of C4BP. J Biol Chem 276:27136–27144

Boolootian RA, Giese AC (1958) Coelomic corpuscles of echinoderms. Biol Bull 15:53–63

Bremer K (1988) The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. Evolution 42:795–803

Cameron RA, Mahairas G, Rast JP, Martinez P, Biondi TR, Swartzell S, Wallace JC, Poustka AJ, Livingston BT, Wray GA, Ettensohn CA, Lehrach H, Britten RJ, Davidson EH, Hood L (2000) A sea urchin genome project: sequence scan, virtual map, and additional resources. Proc Natl Acad Sci USA 97:9514–9518

Chou K-C, Heinrikson R (1997) Prediction of the tertiary structure of the complement control protein module. J Protein Chem 19:765–773

Clow LA, Raftos DA, Gross PA, Smith LC (2004) The sea urchin complement homologue, SpC3, functions as an opsonin. J Exp Biol (in press)

DiScipio RG (1992) Formation and structure of the C5b-7 complex of the lytic pathway of complement. J Biol Chem 267:78087–78094

DiScipio RG, Hugli TE (1989) The molecular architecture of human complement component C6. J Biol Chem 267:16197–16206

DiScipio RG, Linton SM, Rushmere NK (1999) Function of the factor I molecules (FIMs) of human complement component C6. J Biol Chem 274:31811–31818

Edds KT (1993) Cell biology of echinoid coelomocytes. I. Diversity and characterization of cell types. J Invert Pathol 61:173–178

Fries JW, Williams AJ, Atkins RC, Newman W, Lipscomb MF, Collins T (1993) Expression of VCAM-1 and E-selectin in an in vivo model of endothelial activation. Am J Pathol 143:725–737

Gordon DL, Kaufman RM, Blackmore TK, Kwong J, Lublin DM (1995) Identification of complement regulatory domains in factor H. J Immunol 155:348–356

Gross PS, Al-Sharif WZ, Clow LA, Smith LC (1999) Echinoderm immunity and the evolution of the complement system. Dev Comp Immun 23:429–442

Gross PS, Clow LA, Smith LC (2000) SpC3, the complement homologue from the purple sea urchin, *Strongylocentrotus purpuratus*, is expressed in two subpopulations of the phagocytic coelomocytes. Immunogenetics 51:1034–1044

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

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser 41:95–98

Hansen JE, Lund O, Tolstrup N, Gooley AA, Williams KL, Brunak S (1997) O-glycbase, ver 2.0—a revised database of O-glycosylated proteins. Nucleic Acids Res 25:278–282

Hansen JE, Lund O, Tolstrup N, Gooley AA, Williams KL, Brunak S (1998) NetOglyc: prediction of mucin-type O-glycosylation sites based on sequence context and surface accessibility. Glycoconj J 15:115–130

Higgins DG, Thompson JD, Gibson TJ (1996) Using Clustal for multiple sequence alignments. Methods Enzymol 266:383–402

Hofmann K, Stoffel W (1993) TMbase—a database of membrane spanning proteins segments. Biol Chem 374:166

Hourcade DE, Wagner LM, Oglesby TJ (1995) Analysis of the short consensus repeats of human complement factor B by site-directed mutagenesis. J Biol Chem 70:19716–19722

Jack PL, Klein NJ, Turner MW (2001) Mannose-binding lectin: targeting the microbial world for complement attack and opsonophagocytosis. Immunol Rev 180:86–99

Jarvis GA, Vedros NA (1987) Sialic acid of group B *Neisseria meningitidis* regulates alternative complement pathway activation. Infect Immun 55:174–180

Johnson PT (1969) The coelomic elements of sea urchins (*Strongylocentrotus*). I. The normal coelomocytes: their morphology and dynamics in hanging drops. J Invertebr Pathol 13:25–41

Jokiranta TS, Hellwage J, Koistinen V, Zipfel PF, Meri S (2000) Each of the three binding sites on complement factor H interacts with a distinct site on C3b. J Biol Chem 275:27657–27662

Kirkitadze MD, Barlow PN (2001) Structure and flexibility of the multiple domain proteins that regulate complement activation. Immunol Rev 180:146–161

Kitching IJ, Forey PL, Humphries CJ, Williams DM (1998) Cladistics: the theory and practice of parsimony analysis, 2nd edn. Oxford University Press, New York, N.Y.

Kozak, M (1987) An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res 15:8125–8148

Krych M, Hourcade D, Atkinson JP (1991) Regulator of complement activation interact with their ligands via SCRs. Proc Natl Acad Sci USA 88:4353–4357

Kuhn S, Zipfel PF (1995) Mapping of the complement regulatory domains in the human factor H-like protein 1 and in factor H. J Immunol 155:5663–5670

Lambris JD (1988) The multifunctional role of C3, the third component of complement. Immunol Today 9:387–393

Liszewski MK, Atkinson JP (1992) Membrane cofactor protein. Curr Top Microbiol Immunol 178:45–60

Liszewski MK, Farries TC, Lublin DM, Rooney IA, Atkinson JP (1996) Control of the complement system. Adv Immunol 61:201–283

Medof EM, Lublin DM, Holers M, Ayers DJ, Getty RR, Leykam JF, Atkinson JP, Tykocinski ML (1987) Cloning and characterization of cDNAs encoding the complete sequence of decay-accelerating factor of human complement. Proc Natl Acad Sci USA 84:2007–2011

Minta JO, Wong MJ, Kozak CA, Kunnath-Muglia LM, Goldberger G (1996) cDNA cloning, sequencing and chromosomal assignment of the gene for mouse complement factor I (C3b/C4b inactivator): identification of a species specific divergent segment in factor I. Mol Immunol 33:101–112

Nakao M, Hisamatsu S, Nakahara M, Kato Y, Smith SL, Yano T (2003) Molecular cloning of the complement regulatory factor I isotypes from the common carp (*Cyprinus carpio*). Immunogenetics 54:801–806

Nielsen H, Engelbrecht J, Brunak S, Von Heijne G (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng 1:101–106

Nixon KC (1999) The parsimony ratchet: a new method for rapid parsimony analysis. Cladistics 15:407–414

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

Nonaka M, Kuroda N, Naruse K, Shima A (1998) Molecular genetics of the complement C3 convertase in lower vertebrates. Immunol Rev 166:59–65

Pancer Z (2000) Dynamic expression of multiple scavenger receptor cysteine-rich genes in coelomocytes of the purple sea urchin. Proc Natl Acad Sci USA 97:13156–13161

Pancer Z, Rast JP, Davidson EH (1999) Origins of immunity: transcription factors and homologues of effector genes of the vertebrate immune system expressed in sea urchin coelomocytes. Immunogenetics 49:773–786

Pangburn MK, Rawal N (2002) Structure and function of complement C5 convertase enzymes. Biochem Soc Trans 30:1006–1010

Rast JP, Amore G, Calestani C, Livi CB, Ransick A, Davidson EH (2000) Recovery of developmentally defined gene sets from high-density cDNA macroarrays. Dev Biol 228:270–286

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

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning. A laboratory manual, 2nd edn. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Schlott B, Wohnert J, Icke C, Hartmann M, Ramachandran R, Guhrs KH, Glusa E, Flemming J, Gorlach M, Grosse F, Ohlenschlager O (2002) Interaction of Kazal-type inhibitor domains with serine proteinases: biochemical and structural studies. J Mol Biol 318:533–546

Shah M, Brown KM, Smith LC (2003) The gene encoding the sea urchin complement proteins, SpC3, is expressed in embryos and can be upregulated by bacteria. Dev Comp Immunol 27:529–538

Sikes DS, Lewis PO (2001) Beta software, ver 1. PAUPRat: PAUP* implementation of the parsimony ratchet. Distributed by the authors. Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, Conn.

Smith LC (2001) The complement system in sea urchins. Adv Exp Med Biol 484:363–372

Smith LC (2002) Thioester function is conserved in SpC3, the sea urchin homologue of the complement component C3. Dev Comp Immunol 26:603–614

Smith LC, Davidson EH (1994) The echinoderm immune system. Characters shared with vertebrate immune systems and characters arising later in deuterostome phylogeny. Ann NY Acad Sci 712:213–226

Smith LC, Britten RJ, Davidson EH (1992) *SpCoel1*: a sea urchin profilin gene expressed specifically in coelomocytes in response to injury. Mol Biol Cell 3:403–414

Smith LC, Chang L, Britten RJ, Davidson EH (1996) Sea urchin genes expressed in activated coelomocytes are identified by expressed sequence tags. Complement homologues and other putative immune response genes suggest immune system homology within the deuterostomes. J Immunol 156:593–602

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

Smith LC, Azumi K, Nonaka M (1999) Complement systems in invertebrates. The ancient alternative and lectin pathways. Immunopharmacology 42:107–120

Sorenson MD (1999) TreeRot, ver 2. Boston University, Boston, Mass.

Suzuki MM, Sato M, Nonaka M (2002) C6-like and C3-like molecules from the cephalochrodate, *Amphioxus*, suggest a cytolytic complement system in invertebrates. J Mol Evol 54:671–679

Swofford DL (2002) PAUP*. Phylogenetic analysis using parsimony (*and other methods), ver 4. Sinauer, Sunderland, Mass.

Terado T, Nonaka MI, Nonaka M, Kimura H (2002) Conservation of the modular structure of complement factor I through vertebrate evolution. Dev Comp Immunol 26:403–413

Terwilliger DP, Clow LA, Gross PS, Smith LC (2000) Expression of SpBf, the homologue of complement factor B, and evidence for alternative splicing in the purple sea urchin, *Strongylocentrotus purpuratus*. Dev Comp Immunol 24:S28

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 24:4876–4882

Volanakis JE (1998) Overview of the complement system. Immunol Med 20:9–32

Xu Y, Narayana SV, Volanakis JE (2001) Structural biology of the alternative pathway convertase. Immunol Rev 180:123–135

Zipfel PF, Jokiranta TS, Hellwage J, Koistinen V, Meri S (1999) The factor H protein family. Immunopharmacology 42:53–60