

Coelomocytes Express SpBf, a Homologue of Factor B, the Second Component in the Sea Urchin Complement System^{1,2}

L. Courtney Smith,³ Chi-Schin Shih,⁴ and Sebastian G. Dachsenhausen⁵

A homologue of factor B, SpBf, has been cloned and sequenced from an LPS-activated coelomocyte cDNA library from the purple sea urchin, *Strongylocentrotus purpuratus*. The deduced amino acid sequence and domain structure show significant similarity to the vertebrate Bf/C2 family proteins. SpBf is a mosaic protein, composed of five short consensus repeats, a von Willebrand Factor domain, and a serine protease domain. It has a deduced molecular mass of 91 kDa, with a conserved cleavage site for a putative factor D protease. It has ten consensus recognition sites for N-linked glycosylation. Amino acids involved in both Mg²⁺ binding and in serine protease activity in the vertebrate C2/Bf proteins are conserved in SpBf. Phylogenetic analysis of SpBf indicates that it is the most ancient member of the vertebrate Bf/C2 family. Additional phylogenetic analysis of the SCRs indicates that five SCRs in SpBf may be ancestral to three SCRs, which is the typical pattern in the vertebrate Bf/C2 proteins. RNA gel blots show that SpBf transcripts are 5.5 kb and are specifically expressed in coelomocytes. Genome blots suggest that the SpBf gene (*Sp152*) is single copy gene per haploid genome. This is the second complement component to be identified from the sea urchin, and, with the sea urchin C3 homologue, these two components may be part of a simple complement system that is homologous to the alternative pathway in higher vertebrates. *The Journal of Immunology*, 1998, 161: 6784–6793.

Echinoderms have a nonspecific, nonadaptive immune response that shows similarities to higher vertebrate innate immunity. This has been demonstrated with a number of experimental approaches. Sea stars and sea cucumbers were first used to establish that echinoderms are capable of differentiating between self and nonself tissues, in that they can reject allogeneic skin grafts but would not reject autografts (1, 2). However, immune memory could not be demonstrated because clearance rates of bacteria, xenogeneic cells, and bacteriophage from the coelomic cavities of sea urchins and sea stars did not accelerate with multiple injections (3–5). Furthermore, recognition specificity could not be demonstrated (6) because rejection rates of second set and third party allografts in the sea urchin, *Lytechinus pictus*, were identical even though they were accelerated relative to the primary rejections (7, 8). Together, these reports defined immunity in echinoderms as a nonadaptive or innate system functioning in the absence of adaptive and specific nonself recognition capabilities typical of vertebrates.

At the molecular level, quantitation of profilin transcripts in coelomocytes was used to characterize the sea urchin immune response as extremely sensitive to minimal injury (9) and to injections of small amounts of LPS (10). Profilin is a key actin

regulatory protein that couples signal transduction with cytoskeletal alterations (11). Increases in profilin transcripts in coelomocytes were interpreted to imply changes in cell shape that, in turn, indicated the activation of these amoeboid, phagocytic cells responding to injury or to LPS (9, 10). A subset of sea urchin coelomocytes, the phagocytes, are known for their dramatic cytoskeletal shape changes in response to minor perturbations (12). Although these reports helped to characterize immune reactivity in this species, they did not indicate that the sea urchin system was anything other than a typical, albeit very sensitive, invertebrate immune system (for review, see Ref. 13).

Homology of the innate immune response within the deuterostome lineage of animals that includes the echinoderm (sea urchins) and chordate (vertebrates) phyla was established with the identification of an expressed sequence tag (EST)⁶ 064 from an LPS-activated sea urchin coelomocyte cDNA library (14). GenBank search results indicated that EST064 encoded a new member of the thioester family of complement components. Further characterization of this cDNA revealed the sea urchin protein, SpC3, was a homologue of the complement component C3 and was the most ancient member of the thioester family of complement proteins (15)⁷. This conclusion was based on sequence similarities, overall protein structure, and phylogenetic analysis. Identification of a simple complement system as a part of the sea urchin immune response established that echinoderms and, by inference, all deuterostome invertebrates share innate immune system homologies with vertebrates. Furthermore, characterizing the simpler immune response exhibited by sea urchins is important for understanding the ancestral deuterostome defense system and for reconstructing the evolutionary changes that occurred during the process of assembling the higher vertebrate immune system.

Department of Biological Sciences and the Graduate Program in Genetics, Institute of Biomedical Sciences, George Washington University, Washington, DC 20052

Received for publication May 4, 1998. Accepted for publication August 21, 1998.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¹ This work was supported by National Science Foundation (NSF) Grants MCB9219330, MCB9596251, and MCB9603086 to L.S.C. and by an NSF Research Experience for Undergraduates Award to C.-S.S.

² GenBank accession number for the sea urchin factor B sequence is AF059284.

³ Address correspondence and reprint requests to Dr. L. Courtney Smith, Department of Biological Sciences, 334 Lisner Hall, George Washington University, 2023 G St. NW, Washington, DC 20052. E-mail address: csmith@gwu.edu

⁴ Current address: School of Medicine, George Washington University, Washington, DC 20052.

⁵ Current address: Naval Nuclear Power Propulsion Program, United States Navy.

⁶ Abbreviations used in this paper: EST, expressed sequence tag; Bf, factor B; SCR, short consensus repeats; vWF, von Willebrand Factor; UT, untranslated; nt, nucleotide; ORF, open reading frame; MASP, mannose-binding lectin-associated serine protease; MBL, mannose-binding lectin.

⁷ In our report of the SpC3 sequence (see Ref. 15), figure 2 incorrectly showed the histidine in position 1090 as being putatively associated with the thioester. The correct histidine is in position 1145.

The suggestion that the alternative complement cascade was the foundation on which aspects of higher vertebrate adaptive immunity that are dependent on complement effector functions may have been built and expanded (14) is reinforced with the characterization of a second sea urchin EST. We report here the complete sequence of EST152, hereafter called *Sp152*, which encodes a homologue of vertebrate factor B (Bf), called SpBf, which is a new member of the Bf/C2 protein family. In the alternative pathway in higher vertebrates, Bf is the second complement component to function in the cascade and binds to activated C3b and C3(H₂O) (for a review of the alternative complement cascade, see Ref. 16). The SpBf domain structure is typical of the Bf/C2 complement family of proteins in vertebrates. It is a mosaic protein composed of five short consensus repeats (SCRs) (which are sometimes referred to as complement control protein (CCP) modules), a von Willebrand Factor (vWF) domain, and a serine protease domain. Alignments with other Bf/C2 proteins show that SpBf has conserved amino acids for binding Mg²⁺ and a conserved cleavage site for a putative factor D. Results from phylogenetic analyses indicate that SpBf is the most ancient Bf/C2 family member and that five SCRs may be the ancestral condition rather than three SCRs, which is typical for the vertebrate proteins. Transcripts from *Sp152* are specifically expressed in coelomocytes and appear to be generated from a single copy gene. The two sea urchin proteins, SpC3 and SpBf, appear to be homologous to the two-component complement system that has opsonin functions in agnathans (17–20) and to the alternative cascade in higher vertebrates. We hypothesize that these two proteins act together and that this sea urchin complement system also functions to opsonize foreign cells and particles, augmenting their phagocytosis and subsequent destruction by the coelomocytes.

Materials and Methods

RNA isolation

Total RNA was isolated from coelomocytes and other adult tissues as previously described (9, 15). Briefly, coelomic fluid (40 ml) was poured through sterile cheese cloth and mixed into 10 ml of cold Ca²⁺ and Mg²⁺ free sea water (21) 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 NaOAc, 50 mM EDTA, 50 mM Tris (pH 7.4), 5% 2-ME) 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 NaOAc and 50 mM EDTA at 10⁵ × *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 (1:1) phenol/sevag (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) according to the manufacturer's instructions.

cDNA library construction

The sea urchin immune response was activated by injections of LPS, and activated coelomocyte RNA was isolated (10). An activated coelomocyte cDNA library was constructed from poly(A)⁺ RNA using the Time-Saver cDNA kit and directionally cloned into the λExCell phage (Pharmacia, Piscataway, NJ) as previously described (14, 15). The library was screened using ³²P-labeled RNA probes that were generated according to technical information from Promega (Madison, WI) and as previously reported (9, 15).

Sequencing

DNA sequencing was conducted on plasmid DNA according to the dideoxynucleotide termination protocol (22) using the TaqTrack sequencing kit (Promega) incorporating [α -³⁵S]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, 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, Rochester, NY). For

clones 152L and 152-69X2, which were longer than several hundred nucleotides (see Fig. 1), inserts were subcloned into the Bluescript vector (Stratagene, La Jolla, CA), and the Erase-a-Base kit (Promega) was employed to create a nested set of deleted clones. Sets of overlapping insert deletions were sized by PCR, and sequences were assembled using the DNASIS sequence analysis program (Hitachi Software, San Francisco, CA) on a pentium personal computer.

RNA gel blots

Poly(A)⁺ RNA (0.4 μg), isolated with oligo(dT) magnetic beads (Dynal), was electrophoresed through a 0.8% agarose gel containing 2.2 M formaldehyde in 1× MOPS buffer (20 mM 3-[*N*-morpholino] propanesulfonic acid, 5 mM NaOAc, 1 mM EDTA, pH 7) and capillary blotted onto GeneScreen Plus (DuPont/NEN). The blot was probed with a ³²P-labeled PCR-amplified DNA insert of clone 152S (500 bp; Fig. 1) or the control clone EST1219 (900 bp), which encodes a homologue of human L8 ribosomal protein (14). The PCR reaction was performed on a 9600 PCR machine (Perkin-Elmer, Norwalk, CT) and contained 200 ng template, 1.0 μM each primer (Sp6 or T3 and T7), 7.5 μM dNTPs, 2.5 mM MgSO₄, 10 μCi [α -³²P]dCTP (Dupont/NEN), and 1 U *Taq* polymerase (Promega), with buffer supplied by the company. The thermocycler was programmed as follows: 94°C for 5 min followed by 20 to 30 cycles of 30 sec at 94°C, 30 sec at 57°C, 1 min at 72°C, finishing with 72°C for 2 min. The entire PCR reaction was passed through a G-50 Sephadex (Pharmacia) spin column to remove unincorporated nucleotides. The probe was denatured at 100°C for 2 min before being added to the hybridization solution.

Filters were prehybridized for 2 h in hybridization solution (50% formamide, 250 mM phosphate buffer (pH 7.4), 1 mM EDTA, 0.1% BSA, 7% SDS) and then hybridized 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, 15 mM Na Citrate, pH 7) with 1% SDS. Filters were exposed overnight to X-OMAT XAR-5 x-ray film (Eastman Kodak). Transcript sizes were estimated from RNA standards (Bio-Rad, Hercules, CA). Reprobing was conducted after the blots were stripped in 0.1× SSC at 100°C for 15 min.

Protein alignments and phylogenetic analysis

A basic BLAST search of GenBank (23) was done using the deduced amino acid sequence of SpBf to identify sequence matches to other proteins. The BLAST list included the Bf/C2 protein family members in addition to other mosaic proteins containing vWF domains, serine protease domains, and SCRs. All of the Bf/C2 members and several additional matched sequences were used to construct protein alignments with the CLUSTAL W program, using default parameters (24).

To identify phylogenetic relationships among the sea urchin and vertebrate Bf/C2 proteins, sequences were first aligned with the CLUSTAL W program (24) and were then imported into the PAUP program (version 3.1.1) (25). Outgroups were identified from BLAST results, and the heuristic search method was used to obtain the shortest tree. The heuristic search in PAUP was set for tree-bisection-reconnection branch-swapping with an initial MAXTREES setting of 100, with all data weighted equally. The general search options were set to keep minimal trees only and to collapse zero-length branches. When multiple trees were obtained, a strict consensus tree was calculated. A number of analyses were done on full-length sequences of the Bf/C2 proteins, on SCR-deleted sequences, and on independent SCR sequences. Different outgroups were chosen to root individual trees, which were based on either sequence similarities to the vWF domain or the serine protease domain, or were SCRs with known binding function or lack thereof. In some SCR analyses, additional SCRs with known binding function were added to the Bf/C2 ingroup.

Results

Isolation and sequencing *Sp152*

One of the partially sequenced cDNAs that was reported as EST152 from the purple sea urchin matched to SCR domains from complement receptors and regulatory proteins (14). Although our analysis indicated that the EST152 BLAST matches were below significance (see Q value, Table II in Ref. 14), matches were mostly restricted to the consensus amino acids in the SCR domains. Consequently, we identified two complete and two partial SCRs in the EST152 protein sequence (see Fig. 2 in Ref. 14). However, since the sequence of the EST152 clone began within the ORF, the library was rescreened with a riboprobe made from the 152S subclone (Fig. 1). We picked 74 positives, which were

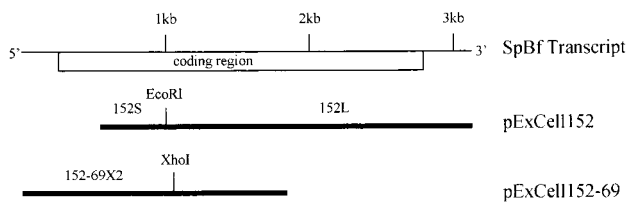


FIGURE 1. Map of cDNAs that span the ORF within the *Sp152* transcript. The entire pExCell152 clone and the subclone 152-69X2 were sequenced to obtain the coding region from the *Sp152* gene. Both clones were obtained from the λ ExCell library constructed from LPS-activated coelomocytes (10) and released as phagemids according to the manufacturer's instructions (Pharmacia). The fragment 152S was subcloned into the Bluescript vector (Stratagene) and used as a template to produce an antisense riboprobe to rescreen the library and obtain the second clone, pExCell152-69, to screen the RNA gel blot and to screen the genome blot.

analyzed by PCR to identify the clone with the longest 5' end. A PCR primer was designed that would hybridize to the 5' end of pExCell152 (5'TGTTTGATCCCAGAGTTTGC3') and that could be used under the same annealing conditions as the Sp6 primer that hybridizes to the polylinker at the 5' end of the insert. The clone with the longest amplified band, pExCell152-69, was chosen for further characterization (Fig. 1). The 5' end of pExCell152-69 (152-69X2) and the 3' end of pExCell152 (152L) were subcloned into Bluescript (Stratagene) to create a nested set of insert deletions for sequencing. The overlapping sequence of these two clones spanned the entire ORF (Fig. 1).

The cDNA sequence of *Sp152* and the deduced protein of SpBf are shown in Fig. 2. The total length of the two overlapping cDNAs is 3163 nt. This length is significantly shorter than the transcript size as seen by RNA gel blot (see results below), and we have assumed that these clones are missing parts of both UT regions since the 5' UT region is 334 nt and the 3' UT region is only 303 nt. Typical UT regions in sea urchin transcripts are usually significantly longer. No consensus polyadenylation signal or poly(A)⁺ tail were identified in the 3' UT region, but, because the library was constructed with a random primer (see Ref. 14), the missing poly(A)⁺ tail was expected. Following the stop codon that defines the end of the ORF, the 3' UT region has 15 additional stops located in all three reading frames. There is one AU-rich repeat, which is typical of transcripts from inducible genes and is thought to function in stabilizing transcripts (26). Six AU-rich repeats were found in the 3' UT region of the sea urchin C3 homologue (15), which appears to be inducible in response to challenge with LPS (L. Clow, P. Gross, and L. C. Smith, unpublished observations).

In the 5' UT region, there are four start-translation codons (ATGs), none of which are in the same reading frame as that of the coding region. Also, there are nine stop codons, two of which are in the correct reading frame, with one being located only 60 nt upstream from the putative start site. Because there is no Kozak sequence to aid in identifying the correct ATG, we have deduced that the fifth ATG from the 5' end is probably the correct start site for translation. This is based on the positioning of the stop codons, because the fifth ATG is in-frame, and also because it is followed by a short hydrophobic region ending with a serine. This is typical of a leader region as defined by the "[−3, −1]-rule" of von Heijne (27). Although this leader region appears rather short, it is followed, within three amino acids, by the first cysteine of the first SCR.

The ORF, 2502 nt, encodes the SpBf protein, which is composed of 834 amino acids (Fig. 2). SpBf has a deduced molecular mass of 91 kDa, although this estimate does not take into consideration putative glycosylations or removal of the leader. Like the other C2/Bf family members, SpBf is a mosaic protein with SCRs,

a vWF domain, and a serine protease domain. It is curious that SpBf has five SCRs when all vertebrate Bf/C2 proteins sequenced to date have three (however, see SCR analysis below). These two extra SCRs give SpBf a deduced size that is significantly larger than vertebrate Bf/C2 proteins. SpBf has a conserved cleavage site for a putative factor D (Arg³⁷⁸-Lys³⁷⁹) that is located at the beginning of the vWF domain and corresponds to cleavage sites in other Bf/C2 proteins (Fig. 2). Furthermore, the serine protease domain contains a conserved histidine, aspartic acid, and serine in conserved positions expected for protease function (Fig. 2). There are ten consensus recognition sequences for N-linked glycosylation located throughout the SpBf sequence (Fig. 2). Five are located in the SCRs, one is found in the region between the SCRs and the vWF domain, and two each are located in the vWF domain and in the serine protease domain.

SpBf protein alignments to other Bf/C2 family proteins

Preliminary sequence comparisons between SpBf and other Bf/C2 members indicated that the vWF domains and the serine protease domains aligned well. However, since SpBf has five SCRs while the vertebrate proteins have three, alignments in this region of the proteins tended to be out of register relative to the four cysteines that define the SCR domains. We determined that, when alignments were performed without the SCRs, the results for the non-homologous linker region, the vWF, and serine protease domains did not change. Therefore, the alignment shown in Fig. 3 uses sequences with the SCRs deleted and begins with the tyrosine or phenylalanine located five amino acids from the last cysteine in the last SCR for each protein. The alignment reveals a number of highly conserved amino acids. These include a factor D cleavage site at the beginning of the vWF domain, the five amino acids that are involved in binding Mg²⁺ (28), and the amino acids involved in the serine protease activity (Fig. 3).

There are 30 cysteines in SpBf, 20 of which are located in the SCR domains (Fig. 4), and the remaining 10 are shown in the alignment (Fig. 3). SpBf has one cysteine located in the vWF domain, which is a region in the vertebrate Bf/C2 proteins where no cysteines are found. In the serine protease domain, SpBf has nine cysteines. Eight align perfectly with cysteines in the Bf/C2 proteins (Fig. 3) and with cysteines in other serine proteases chosen from the BLAST results (data not shown). In addition, there is one cysteine in SpBf that does not align with the vertebrate Bf/C2 proteins, and there are two positions where cysteines align in all the vertebrate proteins but are missing in SpBf (Fig. 3).

An SCR domain is typically about 60 amino acids long with a number of conserved residues. These amino acids include four cysteines, three glycines, two prolines, two tyrosines (or phenylalanine), and one tryptophan, all of which are located at specific positions within the domain (29). These consensus positions are shown in Fig. 4 where the five SCR domains from SpBf are aligned to each other. Two disulfide bonds are formed between the four cysteines and maintain the topology of the domain (Fig. 4). Although some of the consensus amino acids are missing from SpBf, SCR 4 (two tyrosines) and SCR 5 (the third glycine), all four cysteines are present in each SCR, suggesting that these domains fold as expected. In our previous report (14), we suggested that one SCR was missing a cysteine; however, this was due to sequencing errors.

Phylogenetic relationship between SpBf and other Bf/C2 family proteins

Since the protein encoded by *Sp152* appears to be a new member of the Bf/C2 family of proteins, we were interested to know whether it was more similar to Bf or C2. Pairwise alignments between SpBf and all the other Bf/C2 proteins, with and without

1 GGGACGTGTTTTGTTCGCTGTCTGGTCTACCATTCGCTTCTGGACTCGTGGCTAGAATCACTTCAGTTTGCCTGAGAAGGGATCCACGAGGTGTGAAATGAAGTTTTGGCGGA 120
 121 TCATCGTTGCGGCCATTTAGGATACATCTGTTCCTGTGCGAGGGTACTAGCTCGCTGACCAAGGAGTTTGGCGAAAGGACGTTACGAATATCAGAACAGTGCAGACTATGGATTGAC 240
 241 GCATCTCCCTGTGAATTACATCGTGGCGCAGTGTGTGAGCCACAGCAGCCTTGTAGGAGATGAATTCATCATCTGCCGAGACATGGACAAATGGACAAATGGAGGATTACCCATATGCCATCT 360
 M E R L P I L P S 9
 leader

361 CCGCTAGTATGTTTTCGTCTTCTGAGAGCCCTGCAACGGGTATGTGGCGCATGCGTGGTAAATCTGCAAGAAGCATTGCTGTGTATCGCTGTAACCCCTGGTTTCGAGTCCATGGACAA 480
 10 P L V C F R L P E S P A N G Y V R M R G K S A R S I A V Y R C N P G F E L H G Q 49
 | SCR 1→

481 AGAACGGTGTCTGCAATGAGGCTACGGGTAGTTGGCAGGAGATGTACCTCGATGTTTCAGACCTCCATCATGCCCTGACCCAAACGAACCATCCGGGATGCTAGCCTTGGTCTTGTAT 600
 50 R T V V C N E A T G S W Q G D V P R C F R P P S C P D P N E P S G M L A L V L D 89
 ←SCR 1 | | SCR 2→

601 GAAACTCTGGAATTAACACGGTATCTTACACCGAAGCAGTGTCAACTATACCTGCCAAACAGGATACCACCAACCGGGGAGCTCACAATCACGTGTGAGAATGGCTCATGG 720
 90 E N S G I K H G D L Y T E R T V F N Y T C Q T G Y H P T G D V T I T C Q N G S W 129
 ^CHO ^CHO

721 GAAGCGGATCGGATGTTCCGGAATGCAACAAGATCATCTGTGCCCCACCCTGGAACCCATAGAGAATGGTACACACAACATTGAGAACAGGACATACAATGTTGGAGAAATGGT 840
 130 E A D A D V P E C N K I I C A P T L E P I E N G T H N I E N R T Y N V G E M V 169
 ←SCR 2 | | SCR 3→ ^CHO ^CHO

841 CAATACGAATTAACGAAGGATATCTCGTCGATGGAATAAGTGGCAAGTATGTGAGAGTACAGTGGACCAATGGAATATGGTCCGATAACGTACCAGCTGCAAAGTCGAAAGTGT 960
 170 Q Y E C N E G Y L V D G I S W Q V C Q S H D G P N G I W S D N V P R C K V A R C 209
 ←SCR 3 | | SCR 4→

961 CCCATCCGGGTTCCCGTCCACCTTGACACCGGCGAGATTTCTTCTCGCCGAGAAACGGACAGGCCACGAATTCCTGGCCCGTAGGAGCGGTATGACCCCTCAGTGTCCGTCACAT 1080
 210 P H P G S R H L D N G E I S F S P R N G Q G A T N S W P V G A V M T L R C R S H 249

1081 CTAGAGTCTGGTCAAGTGAGAGGTTGTCAACCGATGGTCACTGGAATGGTACTCTGACGTCATGCCACAATGAAGAGTACGACTGCCAGCCTTGGGAAACACCATACACGGC 1200
 250 L E L L G S S E R V C Q P D G H W N G T L T S C H N E E Y D C P A L G N T I H G 289
 ^CHO ←SCR 4 | | SCR 5→

1201 TCTAAAGTCCGCTCTCCGAGTACAGAGTAAACACCAAGTCAATTTTCGATGCAACCATGGTTATCAGATGGTGGGTTCCGAAGAGAGGATGTGACGTTTCTGAAAGTGGACTGGG 1320
 290 S K V G S P Q Y R V N T K V N F R C N H G Y Q M V G S E E R M C T F S E E W T G 329

1321 GAGATGCCAAGATGCATGGTCCGTCGATGAGTTCGACGAACTTTCTCGAGTGAATGGGCTAGGGAAGACCTTCGACGAAATCGTCATCGACGGCGGCTCTGGGTGGTGGAAATGAC 1440
 330 E M P R C M G R D E F D E L S R V M N G L G K T F D E I V I D G G S W V G R N D 369
 ←SCR 5 | ^CHO

1441 ACCAGTCCGAAGTGGGTTGCGAAGGAAACGTACAATTAACCTCTCGATGGTATGGATATCTACTTTCGCTTGGTGGTGGTCCGATGATGTTGGACTTAAACCTTCGAGATAGGCAAG 1560
 370 T S S N S G L R R K R T I K L S D G M D I Y F A F D A S N S V G L K N F E I G K 409
 ^ putative Factor D cleavage site
 | von Willebrand Factor Domain→

1561 ACATTTGCCAAGCAACTAGTGGGAAAGCTTCAAGTGAATACCTCGCTGGTGGAAACGGTTCGCGGGCAGTCTCCTACTCGTCTGAAGCAGCAGATTGTTCAACGTCACAGATTTTACG 1680
 410 T F A K Q L V G K L Q V N T S P G G T R V G A V S Y S S E A R R L F N V N D F T 449
 ^CHO

1681 TCTACAGTGTATGATTAAGCCATCGAGGCAATGTCAATTACACAATAAAGGAACAATTTGCCCGCTCTCGAGACAATCGGGTAATGATTACCGAAACAGCGGATGAATCT 1800
 450 S T V D V V K A I E A N V N Y T N K G T N L P A A L E T I G V M I T E T A D E S 489
 ^CHO

1801 GGTACAGCAGCAGGAAACGCATTTCTATTATCATAACAGAAGGTTTCAGTAACTGGTGGTGGCCCATCGAAAGCGCTCAACCACTAAAGGAAGATGCGGCCCTCAAGATACACTGC 1920
 490 G Y S S R K R I L F I I T D G F S N V G G A P S K S A Q P L K E D A A L K I H C 529

1921 ATAGGAATTAGCAGGAACACGGCAAGACAGCAGTTCGCCAATCGCCTCTCCGCTGTGACGGAAACATGTATTTATCTCTCGGATTATAATGAGCTGGAAAGCGCGCTGGAGCCATT 2040
 530 I G I S R N T D K T A L A E I A S P P V S E H V F Y L S D Y N E L R A V E A I 569

2041 ACCTCTACAACAGATCTTACGAGGAATGTGGAGAGTCAAACATCCGTCGCGCAGCTCTCGCATGTTGGTGGTGTGATGCCATTCCGGGGATTGGCGGTGGCAGGCGCCCTTAC 2160
 570 T S T N R S Y E E C G E S K H P S A T S R I V G G S E S H S G D W P W Q A A L Y 609
 ^CHO
 | Serine Protease Domain→

2161 GATGAAGACTCAAATCAACTTTTGTGGGGGCTCTCTCATTTGAGAAAACTGGATTTTGACCGCAGCGCACTGTTTCAGTGGAGAAAATACGCTGAGCCAGAATGGAACAAGTGTAT 2280
 610 D E D S N Q L L C G G S L I E K N W I L T A A H C F S G E N T L S Q N G T T V Y 649
 ^ protease active site ^CHO

2281 CTGGACTGACACAGAGTGAATGATTTGAATAGGCCAGTTCGATGTGAAGGAATGATTACGCACCCGACTACTCCAAGGACTCGAGGAGGAGACAATGACATCGCCCTC 2400
 650 L G L T H R V N D L N R P S V R C E G I D Y A P G L L Q G L D G G E H N D I A L 689
 protease active site ^

2401 TTGGTTTGGATCGTGAAGCAGAGTGAATTCGATGCTGCTGCTTCCACCAAGTACCCCAAAAAGTTAATGGTACGTGAATCTAGGAGGACGGTTCGTCACAGGG 2520
 690 L R L D R E A E L S P F V R T V C L P P S D P Q K V N W Y V N P R R T A F V T G 729

2521 TGGGGTACATCCCTCAAAGGACAGACCTCGCCTGCTCATGGAATCATGATACCCAGGTTCTAGATAGTTCGTTGATGATGAGCATGAGGATGAGGATGAGTGGACAGGACT 2640
 730 W G H T L K G Q T S P A L M E I M I P P V L D S S C S I A M S A H G I A V D T T 769

2641 ACTGAATGTGTGACGGAATAGAGAGGAAGGATTATGCAAGGCGATTTCAGGTGGCCCTTGGTCTGCAACGGAATAACAAATACAGGACAGTCCGGATCGTGAAGTACGGCATCGGA 2760
 770 T E L C A G I E R K D S C Q G D S G G P L V V Q R N N K Y R Q I G I V S Y G I G 809
 ^ protease active site

2761 TGTGGTGAACCTATGGTGTATACACCAGAGTCCACACTATGTGACTGGATTAATGGAATCATTAAATAAGTAGAAAATACAAACATTATCGTGATCTAACAACAAAAACATAGTA 2880
 810 C G V T Y G V Y T R V P H Y V D W I N G I I N K stop 833

2881 TCACCAATCTCTCTTACTACCGCGCTATAGTTCTAAGCTCAAGAAAGATTACACTATCTTCTACTGACTACCAACAGTCTTTTCAGTACTCCCAATCCCCCTCTCACATGATAT 3000
 3001 AACCGCAAGCCTGTGATATATACATACCATAGAGAGCGGATCTATAAGCAGCATCTAAACTATGGTATGAATCAACATATAGCAAACTATACAATCACATCGCCTGATCAC 3120
 3121 ATTCATATTAAAGG 3136

FIGURE 2. cDNA sequence and deduced protein sequence of SpBf. The sequence was obtained from the two overlapping cDNA clones and the set of nested deletion clones that were obtained from 152L and 152-69X2 using the sequence alignment program from DNASIS (Hitachi Software) on a pentium personal computer. The deduced amino acid sequence is shown below the DNA sequence. These sequences have been deposited in GenBank with the accession number AF059284. Stop codons are underlined in both UT regions. Four ATGs are shown in bold in the 5' UT region. The start site is not identified by a Kozak sequence; however, a hydrophobic region, the putative leader, is underlined at the N terminus of the protein. The domains are labeled in the figure including the consensus positions for serine protease activity and a conserved factor D cleavage site. The consensus N-linked glycosylation sites are noted with CHO. One AU-rich element, ATTTA, is denoted in bold and is present in the 3' UT region.

Downloaded from <http://www.jimmunol.org/> by guest on July 16, 2019

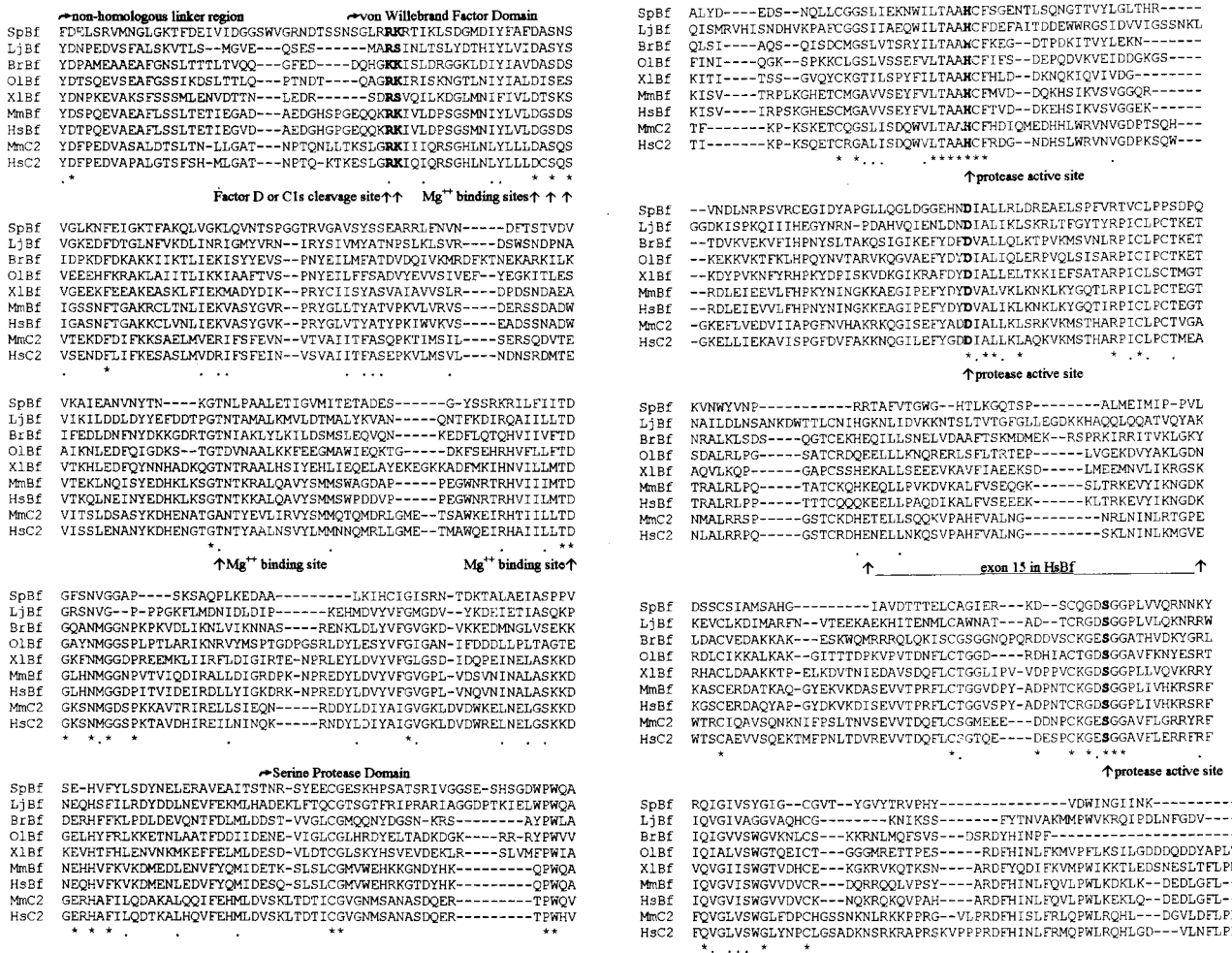


FIGURE 3. Alignment of the vWF and serine protease domains from SpBf and the vertebrate Bf/C2 family proteins. To simplify the analysis, the SCR domains plus five amino acids were deleted, and the remaining domains were aligned with CLUSTAL W program (24), with some minor adjustments by hand. Positions of identity and similarity are identified below the alignment with stars and dots respectively. Domains and amino acids with functional significance are labeled. Location of exon 15 from HsBf was identified from Volanakis (16). Accession numbers for sequences used in this alignment can be found in the legend for Fig. 5. Bf, factor B; C2, complement component C2; Sp, *Strongylocentrotus purpuratus*; Lj, *Lampetra japonica*, lamprey; Br, *Brachydanio rerio*, zebra fish; Ol, *Oryzias latipes*, medaka fish; Xl, *Xenopus laevis* (A gene); Mm, *Mus musculus*; Hs, *Homo sapien*.

the SCR domains, were used to calculate percentage of amino acid similarities and identities between the proteins (Table I). Results of this analysis show that SpBf is about equally similar to all Bf/C2 protein family members. Differences in the number of charged amino acids in exon 15 from human Bf and C2 genes have been suggested as a means to differentiate these two genes (30), and this approach has been used to characterize the Bf/C2 homologue from the medaka fish (31). There are very few charged amino acids in the region of SpBf that align with the diagnostic exon 15, but this may have been due to the overall similarity between SpBf and human Bf and C2 in this region being very poor (Fig. 3). Consequently, to assess relationships among the Bf/C2

protein sequences, alignments of full length and SCR-deleted sequences of the sea urchin and vertebrate proteins were done for phylogenetic analysis. Outgroups consisted of either the three A domains from vWF or three serine proteases chosen from the BLAST results. All trees generated by these methods were similar, and a representative tree, using full-length sequences and the three vWF A domains as the outgroup, is shown in Fig. 5. The alignment that was used for this analysis is available by e-mail (see legend to Fig. 5). Although some bootstrap numbers are low for some branches, which correspond to minor differences between trees, the position of SpBf at the base of the Bf/C2 clade was consistent in all trees. This suggests that SpBf predates the Bf/C2 duplication

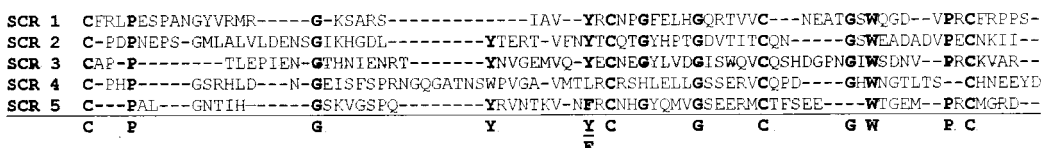


FIGURE 4. Alignment of the SpBf SCRs. The five SCRs from SpBf are aligned to each other with the consensus amino acids shown in bold and at the bottom. Two disulfide bonds maintain the SCR domain structure and are formed between the first and third cysteine, and the second and fourth cysteine. The alignment was done with CLUSTAL W (24) with significant adjustments by hand.

Table I. Amino acid comparisons between SpBf and other Bf/C2 family proteins^a

Vertebrate Bf/C2 Proteins	Full-Length Sequences		Sequences Excluding SCRs	
	% Identical	% Similar plus identical	% Identical	% Similar plus identical
LjBf	24.1	36.7	27.2	43.1
BrBf	20.4	35.9	21.7	40.1
OIBf/C2	19.9	32.9	21.4	35.5
XIBfB	22.6	34.1	25.8	39.9
MmBf	22.5	36.1	23.6	39.9
HsBf	21.3	35.6	23.0	40.1
MmC2	22.7	37.7	23.0	38.9
HsC2	23.2	36.5	26.2	45.3

^a See legend to Fig. 5 for accession numbers.

event that appears to have occurred at some point during the evolution of vertebrates (32) and, therefore, should be considered a Bf homologue. Furthermore, since the thioester-containing complement component that has been identified in the sea urchin is a homologue of complement C3, being less similar to C4 (15), this implies the presence of a simple alternative pathway and, in turn, this also indicates that SpBf is a homologue of vertebrate Bf.

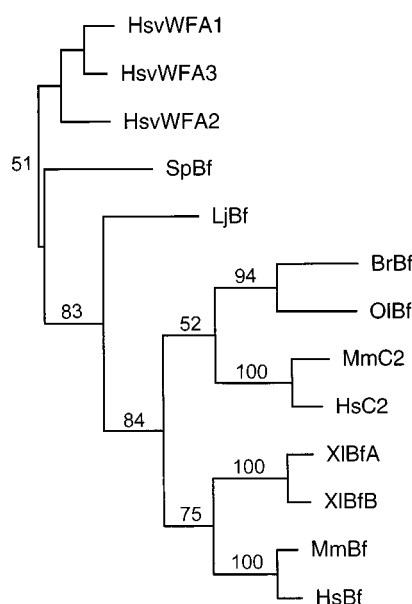


FIGURE 5. Phylogenetic relationships between members of the Bf/C2 family proteins. The tree was constructed using the PAUP program (25) from an alignment done with CLUSTAL W (24). The alignment used to generate this tree can be obtained by contacting the EMBL server by e-mail. Send the request to Netserv@ebi.ac.uk and include in the message GET ALIGN:DS33817.DAT. Two analyses were done using different outgroups. The figure shows the result when three A domains from the human vWF protein were used as the outgroup. The outgroup for the second tree included three serine proteases that were chosen from the BLAST search results (tree not shown). No significant differences between these two trees were seen with respect to the position of SpBf within the vertebrate Bf/C2 clade. Minimum tree length was 3107. The degree of support for internal branches was assessed using the bootstrapping method with 1000 replications, and bootstrap numbers are shown. Accession numbers: SpBf, AF059284; LjBf, D13568; BrBf, U34662; OIBf, D84063; XIBfA, D29796; XIBfB, D49373; MmBf, M57890; MmC2, M57891; HsBf, X72875; HsC2, X04481; HsvWF, X04385.

SCR domain similarities between SpBf and the vertebrate Bf/C2 family

Although SpBf shows significant sequence similarities to the Bf/C2 family proteins, it was important to determine whether the five SCR domains were a result of domain duplications in echinoderms. To understand the SpBf protein structure in more detail, sequence similarities (or differences) between the five SCRs from SpBf and the three SCRs from the vertebrate Bf/C2 proteins (i.e., the ingroup) were assessed through amino acid alignments and phylogenetic analyses. The SCR domains were employed as independent sequences beginning with the first consensus cysteine and ending with the linker region between the SCR domains (four to seven amino acids). Because the SCR domain is short, about 60 amino acids, the degree of support for the branches within an individual tree was low. Consequently, a large number of trees were generated from 18 runs that employed additional SCRs from other SCR-containing proteins. These additional SCRs were either used in the outgroup or were added to the ingroup (Table II). For the 18 phylogenetic analyses that were done, the choice of SCRs for the outgroup was based on a) distant phylogenetic relationship to deuterostomes (Table II, run no. 18; Ref. 33), b) close phylogenetic relationship to deuterostomes (Table II, run nos. 15–17; Refs. 34–38), c) SCRs with known protein binding function (Table II, run nos. 7–14; Refs. 39–42), d) SCRs with putative spacer function (Table II, run nos. 1–3; Ref. 43), and e) SCRs with known three-dimensional structure (Table II, run nos. 4–6; Refs. 44 and 45). In some cases, SCRs with documented C3 binding function were included with the ingroup (Table II, run nos. 2–6) to increase the ingroup size. The number of trees generated from individual runs ranged from 1 to 20 (Table II), and, when more than one tree was obtained, strict consensus trees were calculated for subsequent analysis.

All trees were inspected, and the frequencies with which the vertebrate SCRs clustered into independent clades are shown in Table IV and the frequencies with which SpBf SCRs clustered with vertebrate SCR clades are shown in Table III. Because no individual tree demonstrated all the results shown in Tables III and IV, no tree is shown. This analysis revealed several interesting points. First, the three SCRs from the vertebrate Bf/C2 proteins tend to cluster into separate clades rather than to form multiple or mixed clades, indicating that this approach for amino acid sequence comparisons is sensitive enough to identify sequence differences between the SCRs in the Bf/C2 proteins (Table IV). Second, the five SpBf SCRs tend to cluster with certain vertebrate SCR clades (Table III). This result suggests that a) SpBf SCRs 1 and 2 are most similar to vertebrate SCR 1, b) SpBf SCR 3 is most similar to vertebrate SCR 2, c) SpBf SCR 4 is similar to both vertebrate SCRs 2 and 3, and d) SpBf SCR 5 is most similar to vertebrate SCR 3. In addition, in almost all cases, when a sea urchin SCR clustered with a vertebrate SCR clade, it almost always fell at the base of the clade. The one exception was that, when SpBf SCR 5 clustered with vertebrate SCR 3 clade, in three of twelve cases it was positioned terminally (data not shown). In general, this analysis shows that the SCRs in SpBf are in the same relative order in the protein as the SCRs in the vertebrate Bf/C2 proteins in terms of sequence similarity. Furthermore, this analysis indicates that a structural condition of five SCRs may be ancestral for the Bf/C2 protein family and that, therefore, the vertebrate proteins may have lost two. The information presented here on phylogenetic relationships that includes sequence similarities between the SCRs (Tables II, III, and IV) and the overall comparisons of the SpBf and the vertebrate Bf/C2 proteins (Fig. 5) suggest

Table II. SCRs used as outgroups and additions to the ingroup for amino acid sequence analysis of SCRs from *SpBf* and the vertebrate *Bf/C2* family

Run No.	SCR Source	SCRs Used as the Outgroup	SCRs Added to the Ingroup ^a	Trees Generated
1	HsFactorH	15th		10
2	HsFactorH	15th	HsDAF, 1–4	19
3	HsFactorH	15th	HsFactorH, 1–3 HsDAF, 1–4	2
4	HsFactorH	16th		4
5	HsFactorH	16th	HsDAF, 1–4	1
6	HsFactorH	16th	HsFactorH, 1–3 HsDAF, 1–4	4
7	HsFactorH	1st		12
8	HsFactorH	2nd		8
9	HsFactorH	3rd		1
10	HsFactorH	2nd and 3rd		6
11	HsDAF	1st		3
12	HsDAF	2nd		2
13	HsDAF	3rd		13
14	HsDAF	4th		14
15	HrMASPa HrMASPb HsMASP1 HsMASP2	2 SCRs each		8
16	HsClr & HsCls	2 SCRs each		20
17	HrMASPa & b HsMASP1 & 2 HsClr and Cls	2 SCRs each		6
18	TtFactorC	5 SCRs		10

^a The ingroup consisted of the five SCRs from *SpBf* and the three SCRs from each of the vertebrate *Bf/C2* proteins shown in Table I. Protein alignments were done on CLUSTAL W (24), and phylogenetic relationships were analyzed with the PAUP program (25). DAF, decay-accelerating factor; Hr, *Halocynthia roretzi*, tunicate; Tt, *Tachypleus tridentatus*, horseshoe crab. Accession numbers: HsFactorH, Y00716; HsDAF, M30142; HrMASPa, D88204; HrMASPb, D88205; HsMASP1, D28593; HsMASP2, Y09926; HsClr, M14058; HsCls, X06596; TtFactorC, D90271.

that *SpBf* should be considered a new and ancestral member of the *Bf/C2* protein family.

Sp152 gene expression in sea urchin tissues

We have previously shown that transcripts from the *SpC3* gene, *Sp064*, are specifically expressed in coelomocytes (15). Since the original clone, EST152, was obtained from the an LPS-activated coelomocyte cDNA library (14), we were interested to know whether expression of the *Sp152* gene was also coelomocyte specific. The poly(A)⁺ RNA gel blot of major sea urchin tissues (coelomocytes, ovary, testis, and gut), originally probed for *Sp064* transcripts (see Fig. 6 in Ref. 15), was stripped and reanalyzed with a PCR probe generated from the 152S subclone (Fig. 1). The blot shows that a single transcript of 5.5 kb is present in coelomocytes and that this band is either very weak or absent in the other major adult tissues (Fig. 6). The weak bands seen in the gonad tissues may be due to low level expression in these sea urchin tissues or may be due to expression by coelomocytes that were present in or on these organs at the time of tissue collection and RNA isolation.

Table III. Frequency of *SpBf* SCRs to cluster with certain vertebrate *Bf/C2* SCRs^a

SpBF SCRs	Vertebrate Bf/C2 SCR Clusters (%)		
	SCR 1	SCR 2	SCR 3
SCR 1	67	15	18
SCR 2	60	18	21
SCR 3	20	50	30
SCR 4	13	48	43
SCR 5	16	32	51

^a The clustering of the *SpBf* SCRs with vertebrate SCR clades was assessed from 18 phylogenetic analyses using outgroups with some additions to the ingroup as shown in Table II.

Note that the transcript size of 5.5 kb is significantly longer than 3.1 kb obtained from the cDNAs. As discussed above, this difference is probably based on sequences missing from both the 5' and 3' UT regions. In the absence of a liver equivalent in sea urchins (or a hepatopancreas as is found in sea stars), the expression patterns of both *SpBf* and *SpC3* (15) indicate that the coelomocytes are the major source of complement components.

Sp152 gene copy number

In most vertebrates, *Bf* is a single copy gene, although *Bf* in *Xenopus* and trout appears recently duplicated (46, 47) and *Bf* and *C2* are generally considered to have arisen from an ancient duplication in a common ancestor of mammals (48). To determine whether *Sp152* is a single copy gene per haploid genome, we probed a genome blot of sea urchin DNA isolated from three individuals and digested as previously described (15, 49). The genome blot was analyzed at high stringency with a PCR-generated DNA probe that corresponded to the SCR region of the ORF (500 bp, 152S; see Fig. 1). One or two large bands (≥ 12 kb) were seen for all three individuals when the DNA was digested with *KpnI* or *BamHI*, suggesting that *Sp152* is a single copy gene (data not shown).

Table IV. Frequency of SCRs from vertebrate *Bf/C2* proteins to cluster in single vs multiple clades^a

Vertebrate Bf/C2 SCRs	Clusters in a Single Clade (%)	Clusters in Multiple Clades (%)
SCR 1	100	0
SCR 2	83	17
SCR 3	67	33

^a The formation of a single or multiple clades of vertebrate SCRs was assessed from 18 phylogenetic analyses using various outgroups and some additions to the ingroup as shown in Table II.

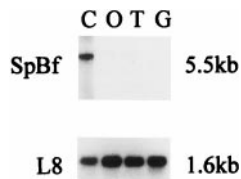


FIGURE 6. *Sp152* gene expression in adult sea urchin tissues. The 152S subclone (Fig. 1) was used to probe a poly(A)⁺ RNA gel blot as described in *Materials and Methods*. C = coelomocyte, O = ovary, T = testis, G = gut.

However, the *EcoRI* digests in all lanes showed a pattern of eight smaller bands ranging in size from 0.5 to 8 kb. This result is more difficult to understand without data on the intron/exon structure of the *Sp152* gene, but at present our interpretation of the multiple bands in the *EcoRI* digest for the three sea urchins is as follows. The probe sequence spans the region of the cDNA from nt 522 to nt 1032 (Fig. 2), which includes part of two and all of two more SCRs. Because these domains are generally known to be contained within separate exons (50, 51), this suggests that the probe may bind to four exons in the *Sp152* gene. If each of the three introns between the SCR exons contains an *EcoRI* site, this would result in eight bands on an *EcoRI* digest for a diploid animal. In general, the genome blot results are consistent with 1) *KpnI* and *BamHI* sites located outside of the SCR regions from two alleles resulting in one or two large fragments hybridizing with the probe, and 2) *EcoRI* sites located between each of the SCR exons, which would separate them into smaller fragments to which the probe would hybridize. These results, as interpreted, are consistent with a single copy gene per haploid genome.

Discussion

This is the first identification of a factor B homologue from an invertebrate. Like vertebrate Bf/C2 proteins, SpBf is a mosaic protein composed of SCRs, a vWF domain, and a serine protease domain. Highly conserved regions in the SpBf protein include the SCR consensus positions, a putative factor D cleavage site and Mg²⁺ binding sites within the vWF domain, and conserved positions within the serine protease domain required for protease activity (Figs. 2 and 3). The homology to Bf rather than C2 is based on sequence comparisons from alignments (Fig. 4), phylogenetic trees (Fig. 5), and BLAST searches (lamprey Bf was consistently the best match, with the top 12 matches being to other Bf or C2 proteins, data not shown). The Bf homology is also supported by the fact that C2 proteins have not been found in animals other than mammals (52–54), indicating that, in nonmammalian vertebrates, Bf must function in both classical and alternative pathways (32). This has, in fact, been shown to occur in trout where two Bf proteins function in both pathways (47). Finally, the homology to Bf is implied from the predicted function of SpBf, which is the interaction between SpBf and the recently identified sea urchin C3 homologue SpC3 (15). This appears similar to the simple complement system in lamprey, which is also composed of C3 and Bf (19, 20). Together, these two sea urchin complement components may be part of a simple complement system that is homologous to the alternative pathway in vertebrates.

The SCR domains in complement components are commonly involved in protein-protein interactions. Three-dimensional structure analysis and electron microscopy of SCRs have indicated that they are small globular domains like “beads on a string” (55) and that binding pockets are formed between two adjacent SCRs (42, 44, 45). SCR deletions, order swapping, and site-directed mutagenesis in a number of proteins have indicated the importance of

each SCR in complement binding. Although all three SCRs in human C2 and Bf are involved in binding C4b and C3b, respectively, the third is most important and the first is least important in this interaction (56–58). While the four N-terminal SCRs in factor H are involved in binding C3b, the second and third are the most important (39, 43). Similarly, for human decay-accelerating factor, SCRs two through four are essential for binding C3b and inhibiting the alternative pathway (41), with the third being most important and the first being least important (40). Together, these studies show that the first SCR in the vertebrate proteins is not very important for binding function, while the third SCR is essential. These functional analyses of the three vertebrate Bf/C2 SCRs and their respective importance in complement binding is reflected in our phylogenetic analyses by the clustering of these domains into independent clades (Table IV).

Phylogenetic analyses were initially used to identify sequence similarities between sea urchin and vertebrate SCRs, but, because our results show that the three vertebrate SCRs cluster into separate clades (Table IV), we have assumed that sequence similarity can be used to infer functional similarity. Consequently, our data suggest that, since SCRs 1 and 2 from SpBf are similar to SCR 1 from the vertebrate Bf/C2 proteins, these domains may be less important in protein interactions. Similarly, since SCRs 3, 4, and 5 from SpBf are more similar to SCRs 2 and 3 from the vertebrate proteins, these SpBf SCRs may be more important in protein interactions.

Model of the sea urchin complement cascade

We hypothesize that the SpBf and SpC3 proteins function together as part of a simple complement system to opsonize foreign cells, particles, and molecules that augments their removal and destruction by phagocytic coelomocytes. This simple sea urchin complement pathway might function like the “archo-complement system” that was first proposed by Lachmann (59), which would make it essentially homologous to the alternative pathway in vertebrates. Opsonization would begin with spontaneously activated SpC3 in solution, in the form of SpC3(H₂O), which would become bound as SpC3b to a foreign surface by its thioester site and would subsequently be bound by SpBf. The change in conformation of bound SpBf would result in its cleavage by a putative factor D. The possible interaction between SpC3 and SpBf is supported by the conserved Mg²⁺ binding sites and the conserved cleavage site for factor D in the SpBf sequence (Figs. 2 and 3). The involvement of Mg²⁺ and the Mg²⁺ binding sites within the vWF domain has been implicated in Bb binding to C3b after factor D cleavage of vertebrate Bf (28). The SpC3b-SpBb complex would then function as a C3 convertase through the activation of the SpBf serine protease domain, creating an amplification feedback loop to cleave and activate more SpC3 for deposition onto the foreign surface. The feedback loop is inferred from the conserved C3 convertase cleavage site in the deduced SpC3 sequence (see Fig. 3 in Ref. 15) and from the sizes of SpC3 fragments in activated coelomic fluid (W. Al-Sharif and L. C. Smith, unpublished observations). Fragments indicate that SpC3 convertase cleavage site is functional (W. Al-Sharif and L. C. Smith, unpublished observations). The feedback loop of the archo-complement system would result in efficient opsonization of foreign cells or particles that, in turn, would act to augment phagocytosis by coelomocytes bearing a putative complement receptor. The efficiency, speed, and extent of opsonization that occurs as a result of this feedback loop would be advantageous over simple opsonins and might make this simple complement system an important defense mechanism for the sea urchin.

The identification of two MASP proteins in the tunicate *Halocynthia roretzi* is evidence for the presence of a lectin complement pathway in this species (38), and this information is important for understanding the early evolution of the complement system in invertebrate deuterostomes. The authors suggest that the lectin pathway was the first complement activation mechanism to evolve (38). Our data, on the other hand, indicate that the sea urchin complement system could function in the absence of a lectin pathway because of the conserved thioester and C3-convertase sites in SpC3 and the conserved factor D site in SpBf. Although homologues for MBL or MASP have not been identified in the sea urchin, this does not suggest that a lectin pathway is absent, and, conversely, because a Bf homologue has not been identified in tunicates, this does not suggest that an alternative pathway is absent from ascidians. Indeed, the presence of an alternative pathway in the sea urchin infers its presence in tunicates. Based on the available data on complement proteins in sea urchins and tunicates, it is not possible to discern which mechanism for activating C3 (C3 convertase or MBL-MASP) is more ancient, but we predict that both will be found to function in both groups of animals.

Immune homology within the deuterostomes

The echinoderm immune system has previously been characterized as a nonspecific, nonadaptive response (6, 8), and there was nothing about this system that characterized it as anything other than a typical invertebrate immune system. However, this has now changed. With the identification of a simple complement system in the sea urchin, this establishes immune homology within the entire lineage of the deuterostomes and differentiates immune functions in deuterostome invertebrates from those in protostomes. Continued analysis of the sea urchin complement system not only will clarify our characterization of the echinoderm immune response but also, through comparisons with homologous systems in the vertebrates, will advance our understanding of changes that have occurred during evolution in the deuterostome lineage that culminated in the complex and multitiered immune system that functions in mammals.

Acknowledgments

We acknowledge the support and encouragement from Eric Davidson, in whose laboratory this work was initiated. We also express our gratitude to W. Al-Sharif for his time and assistance during the course of this work. M. Allard, J. Miller, M. Daly, and P. Herendeen gave their time and advice on the phylogenetic analyses. We thank O. Sunyer for significant input on improving the final draft of the manuscript. We are also grateful to P. Gross, M. Allard, L. Heiman, W. Al-Sharif, L. Clow, and R. Packer for additional manuscript improvements.

References

- Hildemann, W. H., and T. G. Dix. 1972. Transplantation reactions of tropical Australian echinoderms. *Transplantation* 15:624.
- Karp, R. D., and W. H. Hildemann. 1976. Specific allograft reactivity in the sea star *Dermasterias imbricata*. *Transplantation* 22:434.
- Reinisch, C. L., and F. B. Bang. 1971. Cell recognition: reactions of the sea star (*Asterias vulgaris*) to the injection of amoebocytes of sea urchin (*Arbacia punctulata*). *Cell. Immunol.* 2:496.
- Coffaro, K. A. 1978. Clearance of bacteriophage T4 in the sea urchin *Lytechinus pictus*. *J. Invert. Pathol.* 32:384.
- Yui, M. A., and C. J. Bayne. 1983. Echinoderm immunology: bacterial clearance by the sea urchin *Strongylocentrotus purpuratus*. *Biol. Bull.* 165:473.
- 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.
- Coffaro, K. A., and R. T. Hinegardner. 1977. Immune response in the sea urchin *Lytechinus pictus*. *Science* 197:1389.
- Coffaro, K. 1979. Transplantation immunity in the sea urchin. Doctoral dissertation, University of California, Santa Cruz, CA.
- Smith, L. C., R. J. Britten, and E. H. Davidson. 1992. SpCoe11, a sea urchin profilin gene expressed specifically in coelomocytes in response to injury. *Mol. Biol. Cell* 3:403.
- Smith, L. C., R. J. Britten, and E. H. Davidson. 1995. Lipopolysaccharide activates the sea urchin immune system. *Dev. Comp. Immunol.* 19:217.
- Bubb, M. R., I. C. Baines, and E. D. Korn. 1998. Localization of actobindin, profilin I, profilin II, and phosphatidylinositol-4,5-bisphosphate (PIP2) in *Acanthamoeba castellanii*. *Cell Motil. Cytoskeleton* 39:134.
- Edds, K. T. 1977. Dynamic aspects of filopodial formation by reorganization of microfilaments. *J. Cell Biol.* 73:479.
- 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.
- 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.
- Al-Sharif, W. Z., J. O. Sunyer, J. D. Lambris, and L. C. Smith. 1998. Sea urchin coelomocytes specifically express a homologue of the complement component C3. *J. Immunol.* 160:2983.
- Volanakis, J. E. 1989. Participation of C3 and its ligands in complement activation. *Curr. Top. Microbiol. Immunol.* 153:1.
- 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.
- 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.
- Nonaka, M., and M. Takahashi. 1992. Complete complementary DNA sequence of the third component of complement of lamprey: implication for the evolution of thioester-containing proteins. *J. Immunol.* 148:3290.
- 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.
- 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.
- Sanger, R., S. Nicklen, and A. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74:5463.
- Altschul, S., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403.
- 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.
- Swofford, D. 1990. PAUP: phylogenetic analysis using parsimony. Version 3.1. Program documentation. Illinois Natural History Survey, Urbana, IL.
- 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.
- von-Heijne, A. L. 1986. A new method for predicting signal sequence cleavage sites. *Nucleic Acids Res.* 14:4683.
- Tuckwell, D. S., Y. Xu, P. Newham, M. J. Humphries, and J. E. Volanakis. 1997. Surface loops adjacent to the cation-binding site of the complement factor B von Willebrand factor type A module determine C3b binding specificity. *Biochemistry* 36:6605.
- Reid, K. B. M., and A. J. Day. 1989. Structure-function relationships of the complement components. *Immunol. Today* 10:177.
- Volanakis, J. E. 1989. C3 convertases of complement: molecular genetics, structure and function of the catalytic domains, C2 and B. In *The Year in Immunology 1988: Cellular, Molecular and Clinical Aspects*, Vol. 4. J. M. Cruse and R. E. Lewis, Jr., eds. Karger, Basel, p. 218.
- Kuroda, N., H. Wada, K. Naruse, A. Simada, A. Shima, M. Sasaki, and M. Nonaka. 1996. Molecular cloning and linkage analysis of the Japanese medaka fish complement Bf/C2 gene. *Immunogenetics* 44:459.
- Seeger, A., W. E. Mayer, and J. Klein. 1996. A complement factor B-like cDNA clone from the zebra fish (*Brachydanio rerio*). *Mol. Immunol.* 33:511.
- Muta, T., T. Miyata, Y. Misumi, F. Tokunaga, T. Nakamura, Y. Toh, Y. Ikehara, and S. Iwanaga. 1991. Limulus factor C: an endotoxin-sensitive serine protease zymogen with a mosaic structure of complement-like, epidermal growth factor-like, and lectin-like domains. *J. Biol. Chem.* 266:6554.
- Leytus, S. P., K. Kurachi, K. S. Sakariassen, and E. W. Davie. 1986. Nucleotide sequence of the cDNA coding for human complement C1r. *Biochemistry* 25:4855.
- Mackinnon, C. M., P. E. Carter, S. J. Smyth, B. Dunbar, and J. E. Fothergill. 1987. Molecular cloning of cDNA for human complement component C1s: the complete amino acid sequence. *Eur. J. Biochem.* 169:547.
- 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.
- Thiel, S., T. V. Jensen, C. M. Stover, W. Schwaible, S. B. Laursen, K. Poulsen, A. C. Willis, P. Eggleton, S. Hansen, U. Holmskov, K. B. M. Reid, and J. C. Jensenius. 1997. A second serine protease associated with mannan-binding lectin that activates complement. *Nature* 386:506.
- 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.

39. Gordon, D. L., R. M. Kaufman, T. K. Blackmore, J. Kwong, and D. M. Lublin. 1995. Identification of complement regulatory domains in human factor H. *J. Immunol.* 155:348.
40. Coyne, K. E., S. E. Hall, S. Thompson, M. A. Arce, T. Kinoshita, T. Fujita, D. J. Anstee, W. Rosse, and D. M. Lublin. 1992. Mapping of epitopes, glycosylation sites, and complement regulatory domains in human decay accelerating factor. *J. Immunol.* 149:2906.
41. Brodbeck, W. G., D. Liu, J. Sperry, C. Mold, and M. E. Medof. 1996. Localization of classical and alternative pathway regulatory activity within the decay-accelerating factor. *J. Immunol.* 156:2528.
42. 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.
43. Sharma, A. K., and M. K. Pangburn. 1996. Identification of three physically and functionally distinct binding sites for C3b in human complement factor H by deletion mutagenesis. *Immunology* 93:10996.
44. Norman, D. G., P. N. Barlow, M. Baron, A. J. Day, R. B. Sim, and I. D. Campbell. 1991. Three-dimensional structure of a complement control protein module in solution. *J. Mol. Biol.* 219:717.
45. Barlow, P. N., M. Baron, D. G. Norman, A. J. Day, A. C. Willis, R. B. Sim, and I. D. Campbell. 1991. Secondary structure of a complement control protein module by two-dimensional HNMR. *Biochemistry* 1991:997.
46. Kato, Y., L. Salter-Cid, M. F. Flajnik, C. Namikawa, M. Sasaki, and M. Nonaka. 1995. Duplication of the MHC-linked *Xenopus* complement factor B gene. *Immunogenetics* 42:196.
47. Sunyer, J. O., I. Zarkadis, and J. D. Lambris. 1997. Structural and functional characterization of B/C2 complement molecules in teleost fish. *Dev. Comp. Immunol.* 21:146.
48. Campbell, R. D., A. S. K. Law, K. B. M. Reid, and R. B. Sim. 1988. Structure, organization, and regulation of the complement genes. *Annu. Rev. Immunol.* 6:161.
49. 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.
50. Campbell, R. D., and D. R. Bentley. 1985. The structure and genetics of the C2 and factor B genes. *Immunol. Rev.* 87:19.
51. Ishii, Y., Z.-B. Zhy, K. J. Macon, and J. E. Volanakis. 1983. Structure of the human C2 gene. *J. Immunol.* 151:170.
52. Bentley, E. R. 1986. Primary structure of human complement component C2: homology to two unrelated protein families. *Biochem. J.* 238:339.
53. Horiuchi, T., K. J. Macon, V. J. Kidd, and J. E. Volanakis. 1989. cDNA cloning and expression of human complement component C2. *J. Immunol.* 142:2105.
54. Ishikawa, N., M. Nonaka, R. A. Wetsel, and H. R. Colten. 1990. Murine complement C2 and factor B genomic and cDNA cloning reveals different mechanisms for multiple transcripts of C2 and B. *J. Biol. Chem.* 265:19040.
55. DiScipio, R. G. 1992. Ultrastructure and interactions of complement factors H and I. *J. Immunol.* 149:2592.
56. Horiuchi, T., S. Kim, M. Matsumoto, I. Watanabe, S. Fujita, and J. E. Volanakis. 1993. Human complement factor B: cDNA cloning, nucleotide sequencing, phenotype conversion by site-directed mutagenesis and expression. *Mol. Immunol.* 30:1587.
57. Hourcade, D. E., L. M. Wagner, and T. J. Oglesby. 1995. Analysis of the short consensus repeats of human complement factor B by site-directed mutagenesis. *J. Biol. Chem.* 270:19716.
58. Xu, Y., and J. E. Volanakis. 1997. Contribution of the complement control protein modules of C2 in C4b binding assessed by analysis of C2/Bf chimeras. *J. Immunol.* 158:5958.
59. Lachmann, P. J. 1979. An evolutionary view of the complement system. *Behring. Inst. Mitt.* 63:25.