Published in final edited form as:
Mol Immunol. 2009 July ; 46(11-12): 2299–2308. doi:10.1016/j.molimm.2009.04.002.

Characterization of shark complement factor I gene(s): genomic analysis of a novel shark-specific sequence

Dong-Ho Shin[1], **Barbara M. Webb**[1], **Miki Nakao**[2], and **Sylvia L. Smith**[1,3]

1 *Department of Biological Sciences, Florida International University, Miami, Florida 33199*

2 *Laboratory of Marine Biochemistry, Graduate School of Bioresource and Bioenvironmental Sciences, Kyushu University, Hakozaki, 812-8581 Fukuoka, Japan*

3 *Comparative Immunology Institute, Florida International University, Miami, Florida 33199*

Abstract

Complement factor I is a crucial regulator of mammalian complement activity. Very little is known of complement regulators in non-mammalian species. We isolated and sequenced four highly similar complement factor I cDNAs from the liver of the nurse shark (*Ginglymostoma cirratum*), designated as GcIf-1, GcIf-2, GcIf-3 and GcIf-4 (previously referred to as nsFI-a, -b, -c and –d) which encode 689, 673, 673 and 657 amino acid residues, respectively. They share 95% (≤) amino acid identities with each other, 35.4 ~ 39.6% and 62.8 ~ 65.9% with factor I of mammals and banded houndshark (*Triakis scyllium*), respectively. The modular structure of the GcIf is similar to that of mammals with one notable exception, the presence of a novel shark-specific sequence between the leader peptide (LP) and the factor I membrane attack complex (FIMAC) domain. The cDNA sequences differ only in the size and composition of the shark-specific region (SSR). Sequence analysis of each SSR has identified within the region two novel short sequences (SS1 and SS2) and three repeat sequences (RS1, 2 and 3). Genomic analysis has revealed the existence of three introns between the leader peptide and the FIMAC domain, tentatively designated intron 1, intron 2, and intron 3 which span 4067, 2293 and 2082 bp, respectively. Southern blot analysis suggests the presence of a single gene copy for each cDNA type. Phylogenetic analysis suggests that complement factor I of cartilaginous fish diverged prior to the emergence of mammals. All four GcIf cDNA species are expressed in four different tissues and the liver is the main tissue in which expression level of all four is high. This suggests that the expression of GcIf isotypes is tissue-dependent.

Keywords

Shark; complement; factor I gene; shark specific sequence; genomic analysis

Address correspondence and reprint requests to Sylvia L. Smith, Department of Biological Sciences, Florida International University, 11200 S.W 8th Street, Miami Florida 33199, USA, Email address: E-mail: smiths@fiu.edu, or E-mail: shindong@fiu.edu Tel.: 1-305-348-3183 (or 3421), Fax.: 1-305-348-1083.

The nucleotide data reported in this paper have been submitted to the DDBJ, EMBL, and the GenBank nucleotide sequence databases with the following accession number(s): GcIf-1, **EF647997**; GcIf-2, **EF647998**; GcIf-3, **EF647999**; GcIf-4, **EF648000**. The accession numbers of amino acid sequences cited in this paper are as follows: human If, **NM_000204**; mouse If, NM_007686; rat If, **NM_024157**; bovine If, **NM_001038096**; carp If-A, AB072912; carp If-B, **AB072913**; Trsc If (banded houndshark, *Triakis scyllia*), **AB064368**; xenopus If, **X59958**; chicken If, **XM_426329**; dog If, **XM_858413**

1. Introduction

The complement system is an integral part of mammalian innate immunity and involves a complex net work of interacting soluble and membrane-associated proteins that function in a defined manner according to the immune stimulous initiating complement activation. Depending upon the nature of the activating substance, complement activation can occur via three distinct pathways, the lectin, the classical, and the alternative pathways, each leading to the terminal sequential assembly of the membrane attack complex, which when inserted into membranes brings about target cell lysis. In addition, complement activation also results in the generation of opsonins and anaphylatoxins and other biologically active peptides which play an important role in phagocytosis, chemotaxis, mast cell degranulation, removal of immune complexes, and modulation of adaptive immunity by enhancing B and T cell responses (Reid and Porter, 1981; Schifferli *et al*, 1986; Fearon and Carroll, 2000; Sim and Laich, 2000; Carroll, 2004; Cunnion, 2004; Saevarsdottie *et al*, 2004; Thangam *et al*, 2005).

Complement activation is tightly controlled by several control and regulator proteins which limit the extent of activation. In mammals, factor I, is an essential regulatory serine protease that functions to limit complement activity by cleaving sequentially several peptide bonds in the α′-chain of C3b and C4b (cleavage fragments of C3 and C4) (Davis and Harrison, 1982; Hsiung *et al*, 1982; Discipio, 1992; Fremeaux-bacchi *et al*, 2004). The inactivation of C3b and C4b by factor I requires certain cofactors, such as, factor H, C4b-binding protein, complement receptor 1 (CR1/CD35) and membrane cofactor protein (MCP/CD46) (Pangburn *et al*, 1977; Nagasawa and Stroud, 1977; Blom *et al*, 2003). Recent research has shown that human factor I can also cleave synthetic substrates in the absence of cofactor(s) (Tsiftsoglou and Sim, 2004).

The molecular structure of the mature human factor I protein consists of 317 amino acid residues in a N-terminal heavy chain and 244 amino acid residues in the C-terminal light chain. The two chains are covalently linked via a disulfide bond (Catterall et al, 1987; Goldberger *et al*, 1987]). Each chain contains three occupied N-linked glycosylation sites contributing 20~25% (w/w) of the apparent protein molecular weight (Goldberger *et al*, 1984; Ritchie *et al*, 2002). The factor I gene is located on chromosome 4q25 and spans 63 kb (Goldberger *et al*, 1987; Shiang *et al*, 1989). It comprises 13 exons and there is a strong correlation between the exonic organization of the gene and the modular structure of the protein (Shiang *et al*, 1989; Vyse *et al*, 1994). The complete primary structure of factor I has been determined by cDNA sequencing of human (Catteral *et al*, 1987; Goldberger *et al*, 1987), mouse (Mints *et al*, 1996), rat (Schlaf *et al*, 1999), Xenopus (Kunnath-Muglia *et al*, 1993), carp (Nakao *et al*, 2003), and banded houndshark (Terado *et al*, 2002). In human factor I protein modules have been arranged, starting from N-terminal end, a LP, the FIMAC domain, CD5 domain (also known as SRCR, scavenger receptor cysteine rich domain), two LDLRA domains, and the SP domain (Morley and Walport, 2000). Nakao *et al* (2003) and Terado *et al* (2002) have shown in the common carp and houndshark, respectively, the presence of an extended cDNA sequence between LP and FIMAC sequence, which has not been reported for other vertebrates.

In this study, four factor I cDNA sequences were isolated from the nurse shark, all of which are identical in sequence except for an additional novel shark-specific region, hitherto not described in mammals, that is formed by insertion, deletion, and/or duplication of small defined sequences. The SSR has been shown to be derived from differential organization in the shark genome. This is the first report to describe and show the presence of a shark-specific region within the GcIf gene(s) that is absent from mammalian factor I gene. The presence of four highly similar factor I isotypes differing only over a short span of sequence suggests that the shark proteins have different domain structure which might indicate unique functional roles when expressed in different tissues.

2. Materials and methods

2.1. Materials

Restriction enzymes, PCR Supermix High Fidelity, Oligo(dT)$_{12–18}$ primer, TOPO Cloning Kit, TriZol reagent, and Superscript II Reverse Transcriptase were purchased from Invitrogen/Life Technologies (Carlsbad, CA, USA). Big-Dye Terminator Cycle Sequencing Kit (V.2.0) was obtained from PE Biosystems (Poster city, CA, USA). Wizard PlusSV Minipreps DNA Purification System was purchased from Promega (Madison, WI, USA). PCR DIG (digoxigenine) Probe Synthesis Kit was obtained from Boehringer Mannheim (Indianapolis, IN, USA). Hybond N^+ nylon membrane was purchased from Amersham Biosciences (Piscataway, NJ, USA). The 5′/3′-Smart RACE amplification and Advantage 2 PCR Kits were obtained from Clontech (Palo Alto, CA, USA), while 5′/3′ RACE system was purchased from GibcoBRL (Gaithersburg, MD, USA). Lumiphos Plus was purchased from Whatman Biosciences (MA, USA.

2.2. Animal

A 2 kg female nurse shark, *Ginglymostoma cirratum*, was obtained from the Keys Marine Laboratory (Long Key, Florida Keys, FL). The animal was transported to Florida International University (FIU) in a 124L tank in aerated seawater. It was anesthetized in 3-aminobenzoic acid ethylester, bled from the caudal vein and then sacrificed for meticulous dissection of tissues while avoiding cross contamination. All dissected tissues were immediately frozen in liquid nitrogen and stored at −80°C for further use.

2.3. Extraction of genomic DNA

One part whole shark blood was mixed with 9 parts Queen's lysis buffer (0.01 M tris, 0.01 M NaCl, 0.01 M Na-EDTA, and 1.0 % n-laurylsarcosine, pH 8.0) and stored at 4°C until used for the extraction of total genomic DNA. A 20 ml aliquot of the mixture was used to prepare the extract.. Genomic DNA was extracted according to the method described by Sambrook *et al* (1989) and stored at 4 °C until used.

2.4. Degenerate PCR

Degenerate PCR amplification of shark factor I cDNA was performed with a sense-directional degenerate primer according to the method described by Shin et al (2007) with an antisense primer, AUAP (see manufacturer's instructions for 3′ RACE system, GibcoBRL) PCR amplification was carried out under the following conditions: thirty five cycles at 94°C for 30 sec, 58°C for 30 sec, and 72°C for 2 min.

2.5. cDNA sequence analysis

Complete cDNA sequences spanning the full coding region of four isolates (GcIf-1, 2, 3, and 4) were subcloned into the TOPO-TA cloning vector. Clones containing expected-sized insert DNA were obtained by direct colony PCR from white colonies and used for plasmid DNA purification using the Wizard Plus SV Minipreps, according to the manufacturer's instructions. Nucleotide sequences were determined for both strands of more than 10 independent clones to correct for any errors that occurred in PCR amplification. The dideoxy chain termination method (Sanger *et al*, 1977) was employed using the PE Applied Biosystems 377 DNA sequencer (Foster City, CA, USA). The M13 forward/reverse primers (supported with TOPO cloning kit) and designed gene specific primers (P1-P6, Table 1) were used for the sequence analysis.

2.6. Genomic analysis: Identification and sequencing of introns

To rule out the possibility that the SSR might have resulted from PCR errors and to confirm their derivation from the nurse shark genome, PCR amplification was employed using six primers (P12–17) designed to identify introns (Table 1). Primers sets, P12–13, P14–15, and P16–17 were utilized for amplifying each intron corresponding to Intron 1, Intron 2, and Intron 3, respectively. PCR amplifications were performed under the following conditions: for Intron 1, 43 cycles at 94°C for 30 sec, 54°C for 30 sec, and 70°C for 6 min; for Intron 2, 43 cycles at 94°C for 30 sec, 53°C for 30 sec, and 70°C for 5 min; and for Intron 3, 43 cycles at 94°C for 30 sec, 53°C for 30 sec, and 68°C for 5 min.

2.7. Phylogenetic analysis

Amino acid sequences for complement factor I were downloaded from the Genbank database and aligned using the multiple sequence alignment feature of the Clustal X program (Tompson *et al*, 1997), using default gap open and extension penalties. The exported alignment file was imported into the Mega 3.1 program for phylogenetics analysis. Phylogenetic bootstrap analysis (using 10,000 bootstrap replicates) was performed using the Neighbor-joining method and Poisson-corrected distances (Saitou and Nei, 1987). Gaps were handled by pairwise deletion.

2.8. Southern blot analysis

DIG-labeled cDNA probes covering 301-nucleotides (nt) (spanning 1995 to 2255-nt) of GcIf-1 were prepared using the PCR DIG Probe Synthesis Kit and primer sets P8/P9 (Table 1), according to the manufacturer's instructions. Five micrograms of shark genomic DNA were digested to completion with 15 units of *Bam*HI, *Eco*RI, *Hind*III, and *Pst* I at 37°C for 20 hours. The digests were electrophoresed on 0.8% agarose gels, transferred onto Hybond-N^+ membranes, and fixed by cross linking at optimal conditions (1200 × 100 μJ/cm^2 employing the XL-1000 UV crosslinker (Spectrolinker, Westbury, New York, USA)). After pre-hybridization at 42°C for 2 h in a solution containing 50% formamide, 5 × SSC, 7% SDS, 2% DIG blocking reagent, 50 mM sodium phosphate buffer (pH 7.0), 0.1% sodium N-lauroylsarcosinate, and 50 μg/ml salmon sperm DNA extracts, the membranes were hybridized at 42°C for 17 h with the DIG-labeled cDNA probe in the same solution as described above. The hybridized membranes were washed twice with 0.1 × SSC containing 0.1% SDS at 68°C for 30 min and then the hybridized bands were detected with alkaline phosphatase (AP)-conjugated anti-DIG antibody (Fab fragments) and Lumiphos plus chemiluminescent detection reagent.

2.9. Gene Expression in tissue

Nine tissues (kidney, spleen, brain, liver, intestine, ovary, muscle, heart, and pancreas,) obtained from a single animal were used for gene expression analysis by RT-PCR. Total RNA from each tissue was isolated with TRISOL reagent according to manufacturer's instructions. First strand cDNA was synthesized from 4 μg of total RNA using the Superscript II reverse transcriptase and the oligo-dT primer. This was used as the template for RT-PCR. Thirty one PCR amplification cycles (for beta actin control) and 42 cycles (for GcIf-1 ~ 4) were performed using sense and antisense primers (P10/5 for GcIf-1 ~ 2 and P11/5 for GcIf-1 ~ 4) described in Table 1 in PCR Supermix high fidelity (with 1.5 mM Mg^{2+}) using the Geneamp PCR System 2400 (Perkin Elmer). PCR amplification was carried out under the following conditions: for the beta actin control, 31 cycles at 94°C for 30 sec, 54°C for 30 sec, and 72°C for 50 sec; and for GcIf-1~4; 42 cycles at 94°C for 30 sec, 53°C for 30 sec, and 72°C for 40 sec. A negative control reaction was run in parallel consisting of the template without the reverse transcriptase. A 682 bp-PCR fragment of shark β-actin was amplified as a positive control, using the specifically-designed primers described by Shin et al (2007).

3. Results

3.1. The nurse shark complement factor I cDNA sequences

To obtain the complete shark factor I cDNA sequence, 5′ universal primer (provided by Clontech) and a gene specific primer (P7, Table 1) were used for an extended PCR amplification. Four full-length shark factor I cDNA sequences designated GcIf-1, GcIf-2, GcIf-3, and GcIf-4, were isolated from the liver (Fig. 1). GcIf-1, GcIf-2, GcIf-3, and GcIf-4 were of 2357, 2309, 2309, and 2261 nucleotides (nt) in length and encode 689, 673, 673, and 657 amino acid residues, respectively. Each of the four nurse shark sequences included a 39-nt 5′UTR and 251-nt 3′UTR before and after the coding region which ranged from 1971-nt to 2067-nt. Like factor I of other vertebrates, GcIfs contain, in order from the N-terminal, the LP, FIMAC domain, SRCR (or CD5) domain, two LDLRA domains, and the SP domain. GcIfs have an atypical cleavage site RSKR between the heavy chain and light chain. Each of the nurse shark factor I pre-proteins has 13 N-linked glycosylation sites, with 12 sites on the heavy chain and only one site on the light chain (Fig. 1). Six of the twelve sites in the heavy chain are located all within the SS2.

As shown in Figure 1, the clones are identical to each other along the entire coding sequence except for a region designated as the SSR between the LP and the FIMAC domain. This region is composed of SS1, repeat sequence RS1, RS2, and RS3, and SS2. The four GcIf clones differ from each other in primary structure only in the composition of the SSR: GcIf-1 SSR consists of SS1-RS1-RS2-RS3-SS2; GcIf-2 SSR consists of SS1-RS1-RS2-SS2; GcIf-3 SSR consists of RS1-RS2-RS3- SS2; and GcIf-4 SSR consists of RS1-RS2-SS2 (Fig. 1). The novel sequences common to all four GcIf isoforms are RS1, RS2, and SS2, while the SS1 sequence exists only in GcIf-1 and GcIf-2 and the RS3 repeat is present only in GcIf-1 and GcIf-3. The variation in total length of the four GcIf cDNA sequences is primarily due to a difference in the composition of the SSR, particularly due to insertion or repeat of the SS1 or RS3. For the SSR of each GcIf isoforms, the RS and the SS1 (not SS2) consist of 48 nucleotides and encode 16 amino acid residues. The composition of amino acid residues of the SS1, however, is distinctly different from that of the three repeat sequences (1, 2 or 3). RS1 ~ RS3 are identical except for three nucleotides (indicated as 1, 2, 3 and 1′, 2′, 3′ in the RS in Fig. 1) which in RS1 are different from those at corresponding positions in RS2 or RS3. RS2 and RS3 are exactly identical repeat sequences with each other. Changes at the three nucleotide positions, 18^{th}-nt, 37^{th}-nt, and 41^{st}-nt in RS1 converts the amino acid residues, Asp, Cys, and Asn to Glu, Ser, and Ser at corresponding positions within the RS2 and/or RS3. Furthermore, analysis of the deduced amino acid sequence of RS1 and SS2 shows the presence of a relatively high number of polar amino acid residues that most likely contribute to the high polarity of the molecule at this region.

3.2. Genomic analysis: Exon-intron structure between the LP and the FIMAC domain

cDNA sequence analysis of the GcIf isoforms identified a SSR, that has not been reported in mammals. The SSR is located between the LP and the FIMAC domain (Fig. 1) and is composed of two short sequences, SS1 and SS2, and three repeat sequences, RS1, 2 and 3. In order to investigate whether those members of the SSR were derived from the nurse shark's genome or artifacts of PCR, we performed two PCR amplifications: a specific regular PCR amplification using cDNA-mixture synthesized from mRNA and genomic PCR amplification. PCR amplification based on the GcIf-mRNA with a sense primer positioned from 53 to 73 and an antisense primer from 327 to 308 on the nucleotide sequence of GcIf-1 gave one thick band located between 200-base pairs (bp) and 400-bp (Fig. 3). Sequence analysis of the band resulted in the same sequences as the four composition of the SSR sequences (SS1, RS1~RS3) shown in GcIf-1~-4 in Figure 1. The genomic PCR amplifications were performed using several sets of gene specific primers positioned on the largest GcIf-1 (P12/13, P14/15, and P16/17) (Table

1). The three genomic PCR amplification results gave about 4.6-kb, 2.3-kb, and 2.1-kb PCR band, respectively for the corresponding primer sets P12/13, P14/15, and P16/17. As shown in Figure 4-A, three introns (Intron 1, 2 and 3) were identified at three exon-intron boundaries located between the LP and the SS1, between RS1 and RS2, and between RS3 and SS2 spanning 4067 bp (Intron 1), 2293 bp (Intron 2), and 2082 bp (Intron 3), respectively. The locations of splice donor/acceptor sites in all three introns follow the consensus "GT/AG" rule. The three introns were designated intron 1, intron2, and intron3. Since there exists the possibility that an additional intron(s) might be present between the SSR and FIMAC intron 4 is listed with a query. Intron 1 shares 50% (≥) nucleotide identity with both intron 2 and intron 3 but shows higher nt identity from the middle portion of the sequence. Intron 2 and intron3, however, share 83.7% nt identity with each other, suggesting that intron 1 has diverged at a different time scale from intron 2 and intron 3 that diverged within a closer time scheme although they originated from a single ancestral gene. Figure 4-B illustrates the structurally different genome organization between the leader peptide and the FIMAC domain in human and the nurse shark. There is no exon between the leader peptide and the FIMAC in human, while there are at least three exons in the corresponding shark regions.

3.3. Gene expression in tissues

The level of expression of GcIf isotypes in tissues was determined by RT-PCR analyses. As shown in Figure 5, RT-PCR for GcIf-1 ~ -2 (Fig. 5A) and for GcIf-1 ~ -4 (Fig. 5B) were performed using primers sets P10/P5 and P11/5 (Table 1), respectively. GcIf-1 and GcIf-2 were expressed only in the kidney and the liver. GcIf-1 ~ -4 were expressed from four nurse shark tissues, the kidney, the brain, the liver, and the muscle. Results showed that the liver expressed all four GcIf isotypes at high levels. Comparison of expression between figures 5A and 5B shows that in addition to the kidney and liver, GcIf-3 and/or GcIf-4 are also expressed in the brain and muscle. Taken together the nurse shark factor I isotypes may exhibit tissue-dependent gene expression, suggesting that their function has been specified according to the nature of target substrate molecule in the evolutionary process.

3.4. Southern blotting and phylogenetic analysis

To investigate the number of copies of GcIf gene(s), Southern blot analysis was performed under stringent hybridization and washing conditions (68°C) using total genomic DNA extracted from whole peripheral blood cells. As shown in Figure 6, the PCR-DIG probe for GcIf gave one strong band and zero to two additional weaker bands in each restriction enzyme-digested DNA sample. The cDNA-DIG probe was designed such that the sequence corresponded to the area of GcIf cDNA that did not contain cut sites (except for one *Pst* I site) for the restriction enzymes used to prepare the digests. Over the entire GcIf cDNA sequence there were no *Bam*H I, *Eco*R I, or *Hind* III cut sites. Southern blot analysis using a genomic DNA probe prepared based on intron 2 produced a hybridized band pattern (data not shown) similar to that shown in Figure 6. A search for restriction enzyme cut sites in the three introns (Fig. 3A) revealed *Bam* HI has a single cut site in intron 1, whereas the other three enzymes have at least one cut site in each intron. Taken together, the Southern blot results suggest that the nurse shark may have single gene copy for each of the GcIf genes in its genome and that GcIf spans at least 25-kb and contains 15 exons.

To clarify the location of GcIf in the phylogenetic tree, the neighbor-joining method unrooted was employed. Phylogenetic analysis revealed that GcIf belongs to the cluster that includes Trsc FI (Fig. 7), and also indicates that bony fish and cartilaginous fish diverged from a common ancestor with cartilaginous fish sharing a remote branch giving rise to mammals and amphibia. This result is largely in coincident with the phylogenetic results shown for the banded houndshark (Terado et al, 2002) and carp (Nakao *et al*, 2003). It is speculated that *Xenopus* and human complement factor I diverged around ~350 million years ago (Nakao *et al*, 2003),

indicating that cartilaginous fish and a sub-ancester of amphibia and mammals diverged much earlier than the emergence of amphibia and mammals, and that shark factor I might have recently evolved over a relative short period of evolutionary history.

4. Discussion

Earlier reports have identified additional sequences present in the heavy chain of non-mammalian species (when compared to mammalian factor I). In *Xenopus*, an inserted atypical divergent sequence has been reported for the heavy chain of factor I following the second LDLRA domain (Kunnath-Muglia *et al*, 1993). The 29 amino acid residues of *Xenopus* factor I insert are highly charged portion and could be made by a repeat because each half of the first twenty four amino acid residues within the *Xenopus* insert only shows differences at two positions of each twelve amino acid residues. Unlike human factor I the common carp has a large cDNA sequence inserted between the LP and the FIMAC domain. From partial genomic data Nakao et al (2003) showed the presence of a single 3-kb intron between the LP and the inserted sequence in carp Factor I (FI-B) and concluded that the additional sequence present in the carp represented an extension of and was part of the N-terminal of the FIMAC domain. Since genomic analysis of carp FI-B was performed on the inserted cDNA segment spanning from LP up to the FIMAC domain one can speculate that genomic sequence extending further downstream into the FIMAC domain might have revealed another intron. Unlike that of the carp and human factor I the genomic organization of GcIf is distinctly different, our studies have shown there are at least 2–3 introns present in the SSR between the LP and the FIMAC domain. This type of genomic organization more closely resembles that of Fugu factor I where in addition to the intron following LP, two introns are present in what is considered the putative FIMAC domain (Nakao et al, 2003). In the banded houndshark Terado et al (2002) described two sequence repeats present between LP and FIMAC domain and considered them the N-terminus of FIMAC. Since genomic data is presently not available we cannot rule out the possibility that the genomic organization of *Triakis* factor I will reveal the presence of introns making the sharks similar in organization. Taken together, the data suggests that an ancestral factor I molecule might have contained species specific repeat sequences encoded by one or more exons. The presence and position of these additional exons and introns between the LP and the FIMAC domain of GcIfs further suggest that the area close to the heavy chain N-terminus, particularly between the LP and FIMAC, might play a key role in contributing to structural and functional diversity of complement factor I in lower vertebrates, such as the shark. One can further speculate and suggest that the N-terminal residues of the FIMAC domain of human factor I are residual vestige of ancestral sequence that became incorporated into the FIMAC exon through loss of an intervening intron.

For higher vertebrates like human and mouse, any insertion between the LP and FIMAC has not been shown so far, indicating that the region showing diversity in cartilaginous fish has been lost at a specific point prior to the emergence of mammals. Further genomic analyses to determine whether additional introns exist between the LP and FIMAC domain of the GcIf will suggest how evolution of complement factor I has proceeded. The GcIf-4 may represent a more complex form than the GcIf-1 if the evolution of complement factor I has proceeded in the direction from more diverse exon/intron complex to more simpler exon/intron system, that is from fish, to *Xenopus* to human. If we interpret the boundaries between SS1 and RS1, and RS2 and RS3 (shown in figure 4A) as splicing acceptor sites, the different domain composition of the SSR of the four GcIf isotypes could be generated from a single gene by alternative splicing. However, we cannot emphatically rule out the possibility that further genomic analysis of SSR of all four GcIf cDNAs when available might show that the GcIf-1~4 isolates reflect four genes.

The amino acid sequence identities between the repeat sequence region of Trsc FI and the RS1~RS3 of the GcIf isotypes are not significantly high, however, the number of amino acid residues making up the repeat sequence is 16 in both species, suggesting that both GcIf and Trsc FI have evolved from an ancestor having a similar pattern of gene organization. Furthermore, the presence of these additional gene inserts may suggest a subtle role in factor I function in these species.

The amino acid sequence identities of the nurse shark factor I show more than 95% with each other and 25.8% with human, mouse and *xenopus*. The amino acid sequence identity with the hound shark factor I is distinctly higher (62.8%≤) than that with mammals (Table 2). Multiple alignment result of the deduced amino acid sequence shows that the primary structure of the nurse shark complement factor I is similar to those of mammals such as human, mouse and rat except for the SSR (Fig. 2). In most vertebrates a cleavage site of four amino acids, RRKR, has been reported except in Trsc FI and *Xenopus* FI, which have RSKR and RKKR, respectively, between the heavy and light chain (Kunnath-Muglia *et al*, 1993;Terado *et al*, 2002). The GcIfs have the same cleavage site as that in Trsc FI. The first, third, and forth amino acid residues within the cleavage site are conserved in all organisms reported so far, indicating that keeping the composition (RXKR; X denotes any aa) of amino acid residues within the cleavage site is essential for molecular cleavage forming heavy and light chains.

Each of the nurse shark factor I pre-proteins has 13 N-linked glycosylation sites: 12 sites on the heavy chain and only one site on the light chain (Fig. 1), and six sites of twelve in the heavy chain, interestingly, are all located only within the SS2. Thus, the nurse shark factor I isotypes have 6–7 putative N-linked glycosylation sites more than that of human, mouse, and rat factor I (Fig. 2). Human complement factor I (Catterall et *al*, 1987) contains three N-linked glycosylation sites in each chain and all are occupied, contributing 20–25% of the protein molecular weight (Goldberger *et al*, 1984;Ritchie *et al*, 2002), but in other mammals such as the dog (accession No, **XM_858413)**, bovine (Menger and Aston, 2003), mouse (Minta *et al*, 1996), and rat (Schlaf *et al*, 1999) have two, one, four, and five in the heavy chain and four, three, two, and two in the light chain, respectively, indicating that the number, position and occupation of N-linked glycosylation sites in the heavy and light chain may not be critical for factor I function as a highly specific serine protease in mammals. The complete deglycosylation of human factor I has been found to cause a large loss of solubility of the enzyme and that the native and partially deglycosylated forms of complement factor I have shown very similar proteolytic activities against C3(NH_3), indicating that for complement factor I the glycans play an important role in the hydration of the molecule in solution and that the charged glycans of complement factor I do not serve any functional role(s) in the factor I-cofactor-substrate interactions.

The question of why the nurse shark has retained half of its N-linked glycosylation sites within the SSR in its heavy chain remains to be addressed and leads us to speculate that the distribution of N-linked glycosylation sites, if occupied could affect the “fit” of the GcIf molecule to its substrate i.e., affinity, and/or the substrate specificity of the SP domain. Given that glycosylation of proteins does affect function, it is reasonable to suggest that these sites, when occupied, might play a role in the molecule’s binding affinity. Whether the SSR per se with its unusually high potential N-linked glycosylation sites affects interspecies compatibility remains undetermined. Uniqueness of structural features can affect (a) the binding affinity of GcIf for its substrate, (b) its requirement for cofactor(s), and (c) the enzymatic specificity of its SP domain. Another cartilaginous fish, the banded houndshark (Terado *et al*, 2002) also contains a high number of glycosylations sites within its factor I heavy chain, however, not necessarily corresponding to the SSR of the nurse shark. Further structural and functional studies on shark factor I will shed light on whether all these potential sites are glycosylated and why so many are present only within the factor I heavy chain. As stated above, the number

and position of glycosylation may have a role in functional specificity of the different isotypes, perhaps affecting binding to substrate.

The liver has been reported to be the main site of synthesis of factor I in human (Morris *et al*, 1982) and other species (Kunnath-Muglia *et al*, 1993; Nakao *et al*, 2003; Terado *et al*, 2002). In the rat factor I was expressed in the liver, the small intestine, and the uterus, with the latter being an atypical site of high expression of factor I (Schlaf *et al*, 1999). In another species of cartilaginous fish, the banded houndshark (Terado *et al*, 2002) factor I gene expression was shown only in the liver. However, in the carp, a bony fish, Nakao et al (2003) have shown differential tissue expression of two factor I isotypes, FI-A and FI-B. FI-A was expressed in the hepatopancreas (known to function as the liver in carp) and the ovary, whereas FI-B was expressed in a variety of tissues, the renal kidney, the head kidney, the heart, the hepatopancreas, and the spleen. Furthermore, FI-B was expressed at significantly higher levels in a wider range of tissues. Taken together, the results suggest that the liver is not essentially the major site of expression of factor I in all organisms and further indicate that the expression of factor I depends on the factor I isotype and the animal species. The only structural difference between GcIf-1/-2 and GcIf-3/-4 is the presence of SS1, which might suggest that the SS1 may have a role in factor I expression. It is plausible that in the nurse shark site specific expression is based on the regulatory response of factor I in complement activities.

Currently it is not known which of the GcIf isotypes is functionally active in nurse shark complement activation and regulation. Also, it is unclear why the nurse shark possesses four highly-similar factor I molecules. Since C3 and C4 homologues have been isolated from shark serum and that there are two distinct shark C3 genes (GcC3-1, GcC3-3, Smith unpublished) it is also possible that the four factor I isotypes differ in their affinity, co-factor requirement and substrate specificity for C3 and/or C4. In mammals a single factor I protein cleaves the α′ chain of C3b at two sites generating C3d and C3dg, the two site specific cleavages may be the function of separate factor I molecules in the shark, with cleavage of C4b being the function of yet another factor I molecule.

Acknowledgments

The authors thank the Keys Marine Laboratory for maintaining and bleeding the sharks, and the F.I.U. DNA sequencing core facility for their services. The work was carried out with the approval of FIU's IACUC and in accordance with State and Federal guidelines. The author BMW is deceased. This research was supported by NIH grant GM08205 to S.L.S. This is publication number FIU-CII-005 from the Comparative Immunology Institute.

References

Blom AM, Kask L, Dahlback B. CCP1-4 of the C4b-binding protein alpha-chain are required for factorI mediated cleavage of complement factor C3b. Mol Immunol 2003;39:547–556. [PubMed: 12431388]

Carroll MC. The complement system in regulation of adaptive immunity. Nat Immunol 2004;5:981–986. [PubMed: 15454921]

Catterall CF, Lyons A, Sim RD, Day AJ, Harris TJ. Characterization of the primary amino acid sequence of human complement factor I from an analysis of cDNA clones. Biochem J 1987;242:849–856. [PubMed: 2954545]

Cunnion KM, Benjamin DK, Hester CG, Frank MM. Role of complement receptors 1 and 2 (CD35 and CD21), C3, C4 and C5 in survival by mice of *Staphylococcus aureus* bacteremia. J Lab Clin Med 2004;143:358–365. [PubMed: 15192652]

Davis AE, Harrison RA. Structural characterization of factor I mediated cleavage of the third component of complement. Biochemistry 1982;21:5745–5749. [PubMed: 6217832]

Discipio RG. Ultrastructures and interactions of complement factors H and I. J Immunol 1992;149:2592–2599. [PubMed: 1401896]

Fearon DT, Carroll MC. Regulation of B lymphocyte responses to foreign and self-antigens by the CD19/CD21 complex. Annu Rev Immunol 2000;18:393–422. [PubMed: 10837064]

Fremeaux-Bacchi V, Dragon-Durey MA, Blouin J, Vigneau C, Kuypers D, Boudailliez B, Loirat C, Rondeau E, Fridman WH. Complement factor I: a susceptibility gene for atypical haemolytic uraemic syndrome. J Med Genet 2004;41:1–5. [PubMed: 14729816]

Goldberger G, Bruns GAP, Rits M, Edge MD, Kwiatkowski DJ. Human complement factor I: Analysis of cDNA-derived primary structure and assignment of its gene to chromosome 4. J Biol Chem 1987;262:10065–10071. [PubMed: 2956252]

Goldberger G, Arnaout MA, Aden D, Kay R, Rits M, Colten HR. Biosynthesis and postsynthetic processing of human C3b/C4b inactivator (factor I) in three hepatoma cell lines. J Biol Chem 1984;259:6492–6497. [PubMed: 6327681]

Hsiung LM, Barclay AN, Gagnon J, Brandon MR, Porter RR. The C3b inactivator is a serine protease. Mol Immunol 1982;19:1376.

Kunnath-Muglia LM, Chang GH, Sim RG, Day AJ, Ezekowitz RAB. Characterization of Xenopus laevis complement factor I structure-conservation of modular structure except for an unusual insert not present in human factor I. Mol Immunol 1993;30:1249–1256. [PubMed: 7692240]

Menger M, Aston WP. Isolation and characterization of factor I of the bovine complement system. Am J Vet Res 2003;64:989–993. [PubMed: 12926590]

Minta JO, Wong MJ, Kozak CH, Kunnath-Muglia LM, Goldberger G. 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 1996;33:101–112. [PubMed: 8604219]

Morley, BJ.; Walport, MJ., editors. Complement Facts book. Academic Press; London: 2000.

Morris KM, Aden DP, Knowles BB, Colten HR. Complement biosynthesis by the human hepatoma-derived cell line HepG2. J Clin Invest 1982;70:906–913. [PubMed: 6288774]

Nakao M, Hisamatsu S, Nakahara M, Kato Y, Smith SL, Yano T. Molecular cloning of the complement regulatory factor I isotypes from the common carp (*Cyprinus carpio*). Immunogenetics 2003;54:801–806. [PubMed: 12618913]

Nagasawa S, Stroud RM. Mechanism of action of the C3b inactivator: requirement for a high molecular weight cofactor (C3b-C4bINA cofactor) and production of a new C3b derivative (C3b′). Immunochemistry 1977;14:749–756. [PubMed: 146022]

Pangburn MK, Schreiber RD, Muller-Eberhard HJ. Human complement C3b inactivator: isolation, characterization, and demonstration of an absolute requirement for the serum protein beta 1H for cleavage of C3b and C4b in solution. J Exp Med 1977;146:257–270. [PubMed: 301546]

Reid KBM, Porter RR. The proteolytic activation systems of complement. Annu Rev Biochem 1981;50:433–464. [PubMed: 7023363]

Ritchie GE, Moffatt BE, Sim RB, Morgan BP, Dwek RA, Rudd PM. Glycosylation and the complement system. Chem Rev 2002;102:305–319. [PubMed: 11841245]

Saevarsdottie S, Vikingsdottir T, Valdimarsson H. The potential role of mannan-binding lectin in the clearance of self-components including immune complexes. Scand J Immunol 2004;60:23–29. [PubMed: 15238070]Review

Saitou S, Nei M. The Neighbor-joining Method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 1987;4:406–425. [PubMed: 3447015]

Sambrook, J.; Fritsch, EF.; Maniatis, T. Molecular cloning: a labolatory manual. Vol. 2. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 1989.

Sanger F, Nicklen A, Coulson R. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 1977;74:5463–5467. [PubMed: 271968]

Schifferli JA, Ng YC, Peters DK. The role of complement and its receptor in the elimination of immune complexes. N Engl J Med 1986;315:488–495. [PubMed: 2942776]

Schlaf G, Rothermel E, Oppermann M, Schieferdecker HL, Jungermann K, Götze O. Rat complement factor I: molecular cloning, sequencing and expression in tissues and isolated cells. Immunology 1999;98:464–474. [PubMed: 10583609]

Shiang R, Murray JC, Morton CC, Buetow KH, Wasmuth JJ, Olney AH, Sanger WG, Goldberger G. Mapping of the human complement factor I gene to 4q25. Genomics 1989;4:82–86. [PubMed: 2563353]

Shin DH, Webb B, Nakao M, Smith SL. Molecular cloning, structural analysis and expression of complement component Bf/C2 genes in the nurse shark, *Ginglymostoma cirratum*. Dev Comp Immunol 2007;31:1168–1182. [PubMed: 17482263]

Sim RB, Laich A. Innate immunity: serine proteases of the complement system. Biochem Soc Transactions 2000;28:545–550.

Terado T, Nonaka MI, Nonaka M, Kimura H. Conservation of the modular structure of complement factor I through vertebrate evolution. Dev Comp Immunol 2002;26:403–413. [PubMed: 11906721]

Thangam EB, Venkatesha RT, Zaidi AK, Jordan-Sciutto KL, Goncharov DA, Krymskaya VP, Amrani Y, Panettieri RA, Ali H. Airway smooth muscle cells enhance C3a-induced mast cell degranulation following cell-cell contact. FASEB J 2005;19:798–800. [PubMed: 15758041]

Tompson JD, Gibson TJ, Plewniak F, Higgins DG. The Clustal X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Research 1997;24:4876–4882.

Tsiftsoglou SA, Sim RB. Human complement factor I does not require cofactors for cleavage of synthetic substrates. J Immunol 2004;173:367–375. [PubMed: 15210795]

Tsiftsoglou SA, Arnold JN, Roversi P, Crispin MD, Radcliffe C, Lea SM, Dwek RA, Rudd PM, Sim RB. Human complement factor I glycosylation: structural and functional characterization of the N-linked oligosaccharides. Biochi Biophy Acta 2006;1764:1757–1766.

Vyse TJ, Bates GP, Walport MJ, Marley BJ. The organization of the human complement factor I gene (IF): a member of the serine proteases gene family. Genomics 1994;24:90–98. [PubMed: 7896293]

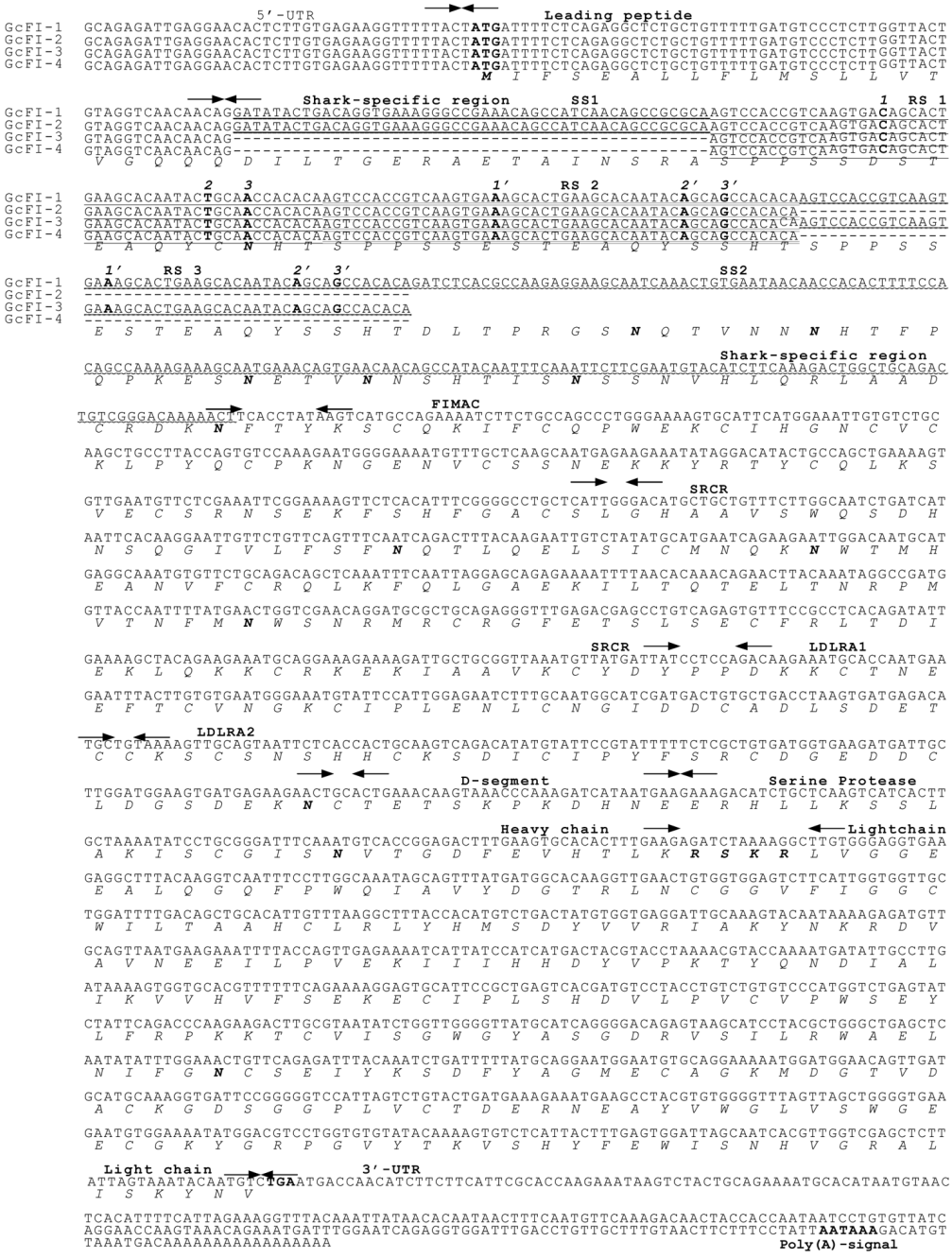

Figure 1.

The cDNA sequences, predicted amino acid sequences and primary structure of four nurse shark complement factor I pre-proteins. Each coding region is flanked by 5′- and 3′-UTR spanning 39 nt and 25 nt, respectively. Within the SSR the short sequences SS1 and SS2 and the repeat sequences RS1, 2 and 3 are identified as follows: SS1, bold wave-underlined; RS1, single-underlined; RS2, double-underlined; RS3, double wave-underlined; and RS3, single wave-underlined, respectively. Nucleotide substitutions in RS1, RS2, and/or RS3 are indicated by bold -italicized numerals (1~3 in the RS1 and 1′~3′ in the RS2 and RS3). Capital letters and italic capital letters show nucleotide sequence and deduced amino acid sequence, respectively. Bold nucleotides and amino acis residues indicate initiation codon methionine (M), N-linked

glycosylation sites (N), the cleavage site (RSKR) between heavy chain and light chain, the stop codon (TGA), and the polyadenylation signal (AATAAA).

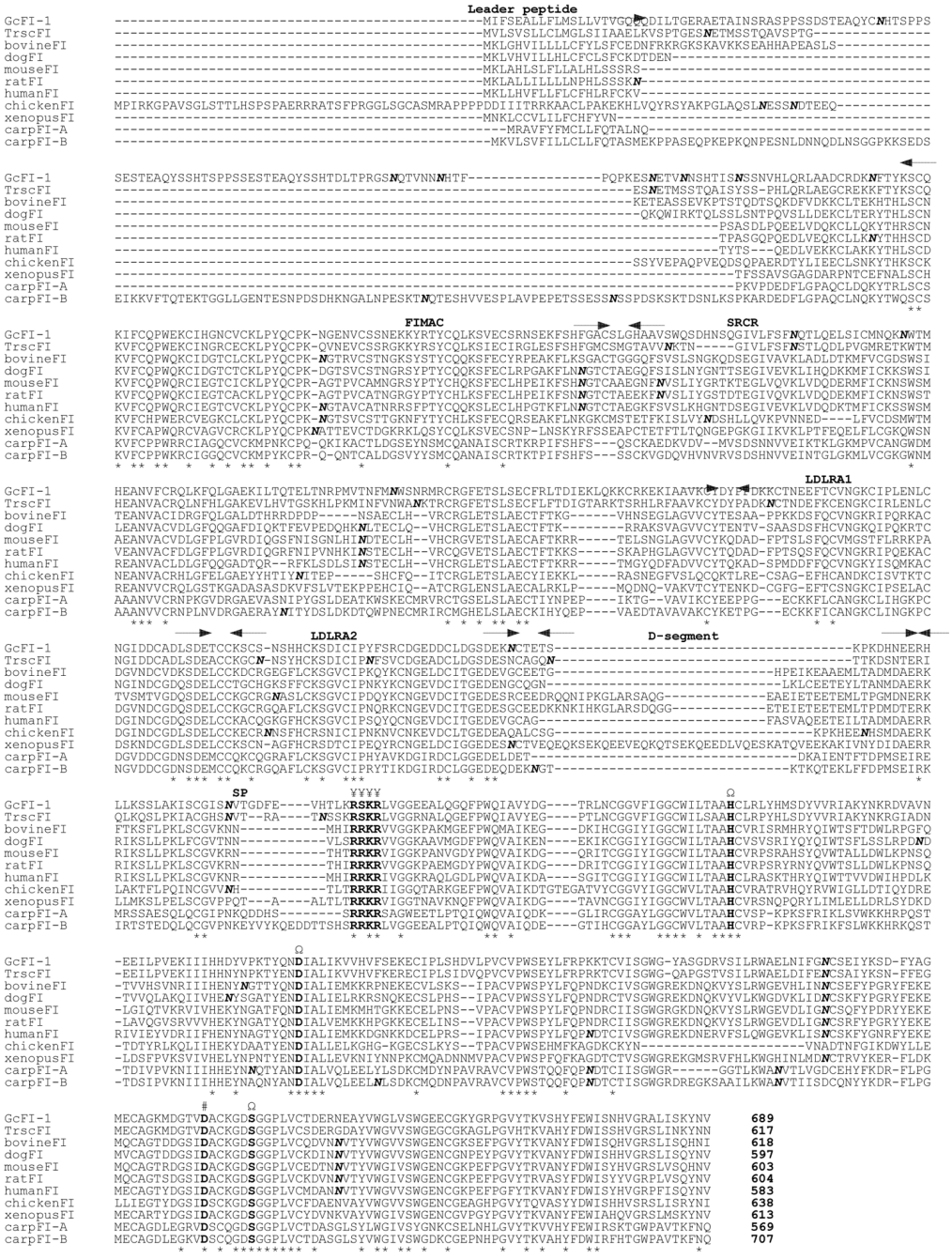

Figure 2.

Multiple alignment of the deduced amino acid sequences of GcFI-1~4 genes with other factor I sequences (human, bovine, mouse, dog, rat, chicken, carp, and banded houndshark Trsc). Dashes indicate gaps which are introduced for maximum matching alignment. Identical amino acid residues in identity are denoted by asterisks. The number of total amino acid residues for each organism is represented at the end of the sequence. Bold italic letters indicate N-linked glycosylation sites. The cleavage site between heavy chain and light chain is indicated by quad ¥. The catalytic triad (His, Asp, and Ser) in the SP domain are emphasized by Ω. The key amino acid residue (Asp) that determines S1-pocket specificity is denoted by #. The domain nomenclature was established by comparison with the primary structure of human complement

factor I. Amino acid sequences employed in here were retrieved from the DDBJ, EMBL, and the GenBank nucleotide sequence databases (see Footnotes).

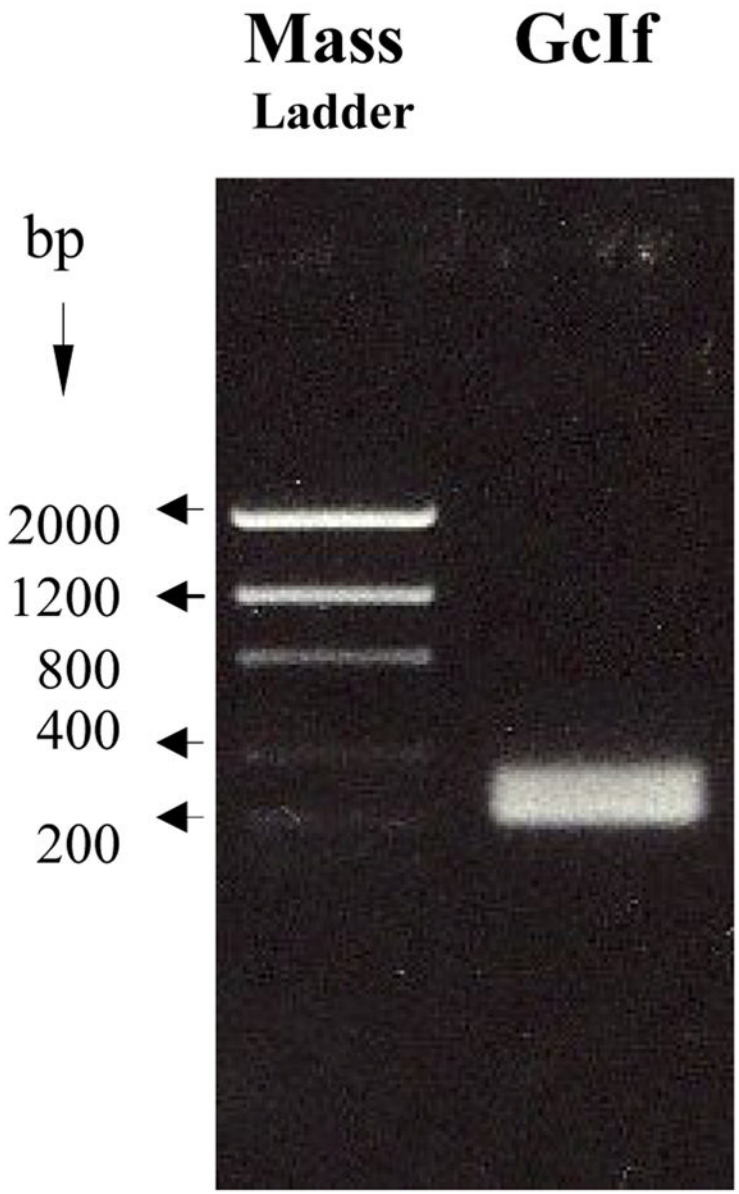

Figure 3.
Result of PCR amplification of SSR site-specific sequence. Lanes designated Mass ladder and GcIf indicate a standard size marker (low DNA mass ladder, Invitrogen) and the sample (GcfI). A sense primer positioned from 53 to 73 and an antisense primer from 327 to 308 in the nucleotide sequence of GcfI-1 were used in this experiment. PCR amplification was performed under the following condition: forty cycles of 94°C for 30 sec, 53°C for 30 sec, and 72°C for 40 sec. The mass ladder and sample were loaded on a 0.8% agarose/EtBr gel in 1 X TAE buffer. The size of the PCR amplified product ranged from ~179-bp to 275-bp.

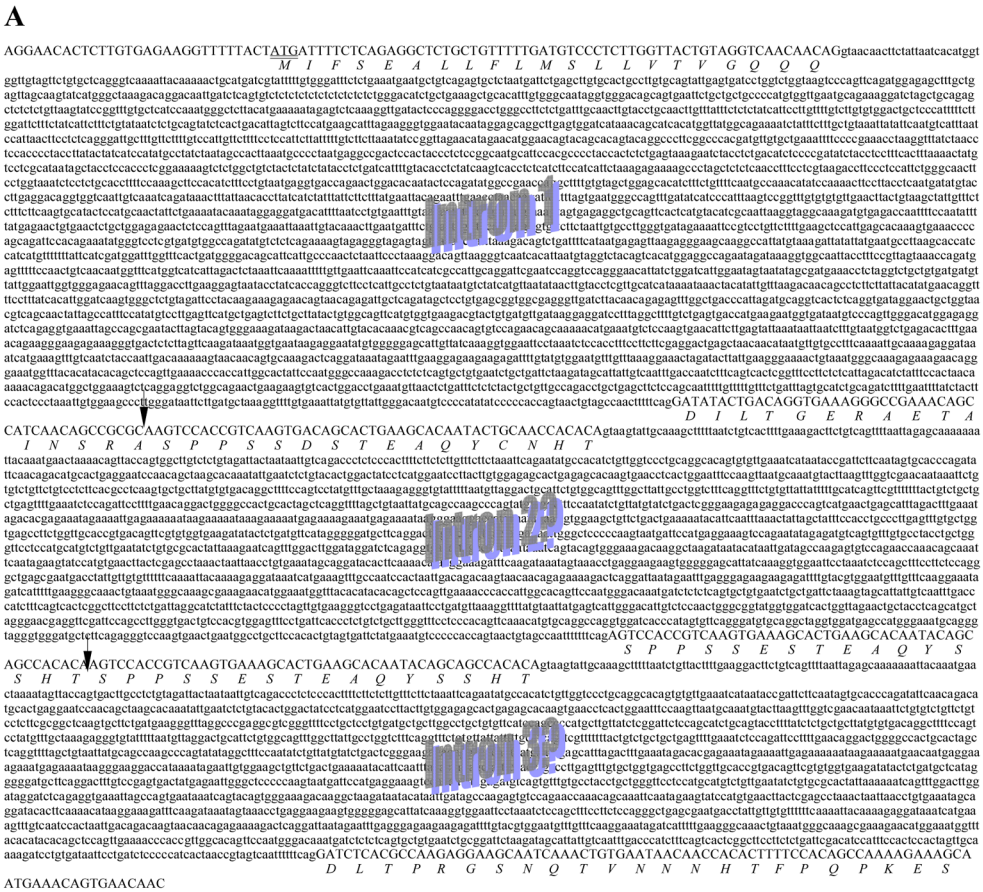

Figure 4.

Partial genomic DNA sequence and structural comparison between the LP and FIMAC domain of GcIf with human complement factor I. 4A indicates genomic DNA sequence between the LP and FIMAC domain of GcIf. The DNA sequence of the exons are written in small capital letters with the corresponding deduced amino acid sequence written in large capital letters. Intron sequences appear as small lower case letters. Arrows indicate the boundaries between SS1and RS1, and between RS2 and RS3. 4B compares the genomic arrangement of human factor I with that of two shark GcIf isotypes and shows the differences in genomic organization of SSR between GcIf-1 and -4 which represent the largest and smallest of the inserts, respectively. Genomic organization of GcIf-2 and -3 are not shown since they differ minimally from GcIf -1 and -4 (see figure 1).

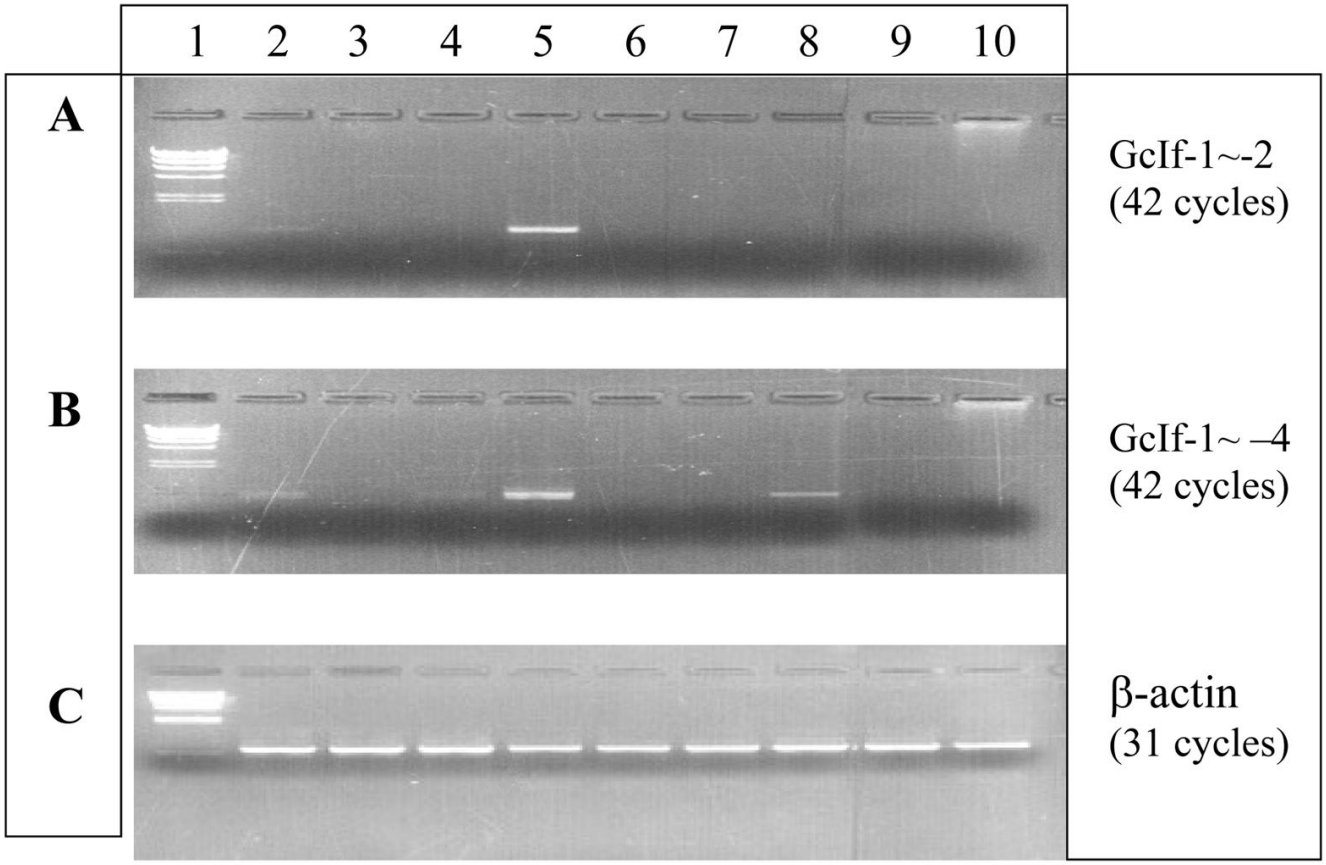

Figure 5.

Results of gene expression analysis of the nurse shark GcIf in tissues. The size of RT-PCR amplified product is ~977–1025 bp for GcIf-1/-2 (Panel A) and ~909–957 bp for GcIf-1-4 (Panel B). Panel C are results for the positive control of nurse shark beta-actin. Lane 1 is the standard DNA-size marker (λDNA/Hind III fragments, Invitogen), and lanes 2 through 10 are tissue samples; kidney, spleen, brain, liver, intestine, ovary, muscle, heart, and pancreas, respectively. For each panel, conditions of RT-PCR are provided in corresponding parenthesis.

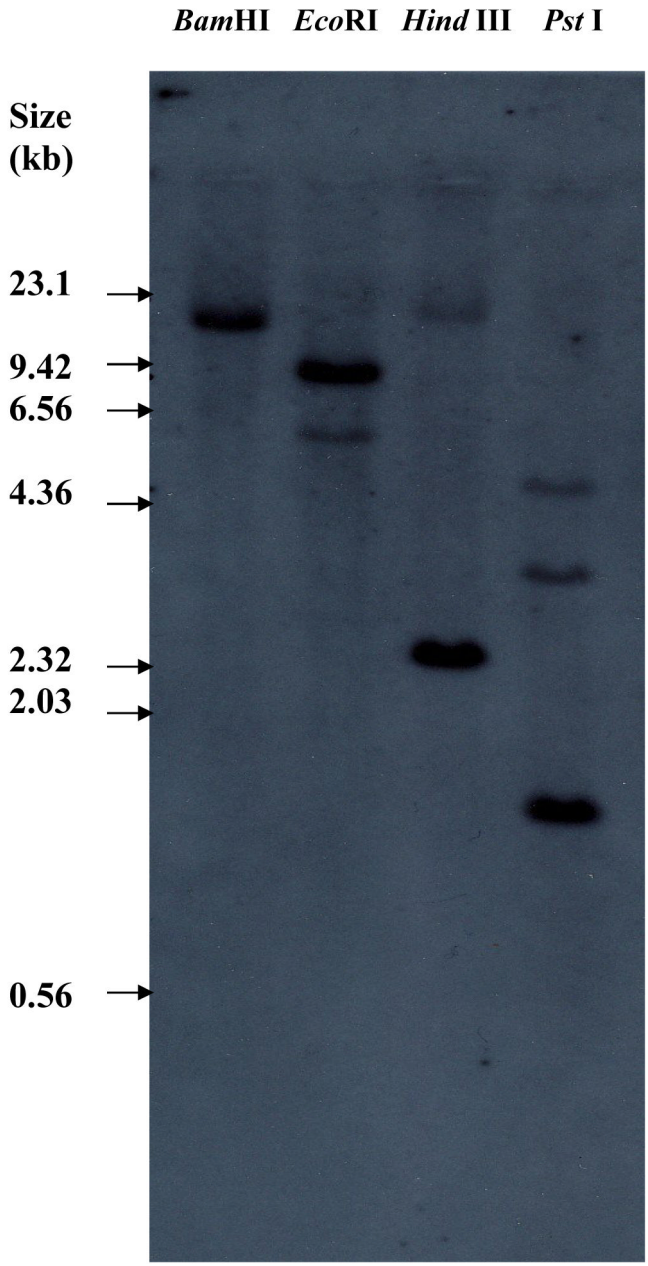

Figure 6.

Southern blot analysis of the nurse shark factor I. Five micrograms of the nurse shark genomic DNA was digested by the following restriction enzymes: lane 1, *Bam*H I; lane 2, *Eco*R I; lane 3, *Hind* III; lane 4, *Pst* I. Numerals on the left indicate size marker in a unit of kilobase (kb). The digested genomic DNA samples were run on 0.8% agarose gel in 1 X TAE buffer at 30 voltages. The probe corresponds to the last exon of GcIf.

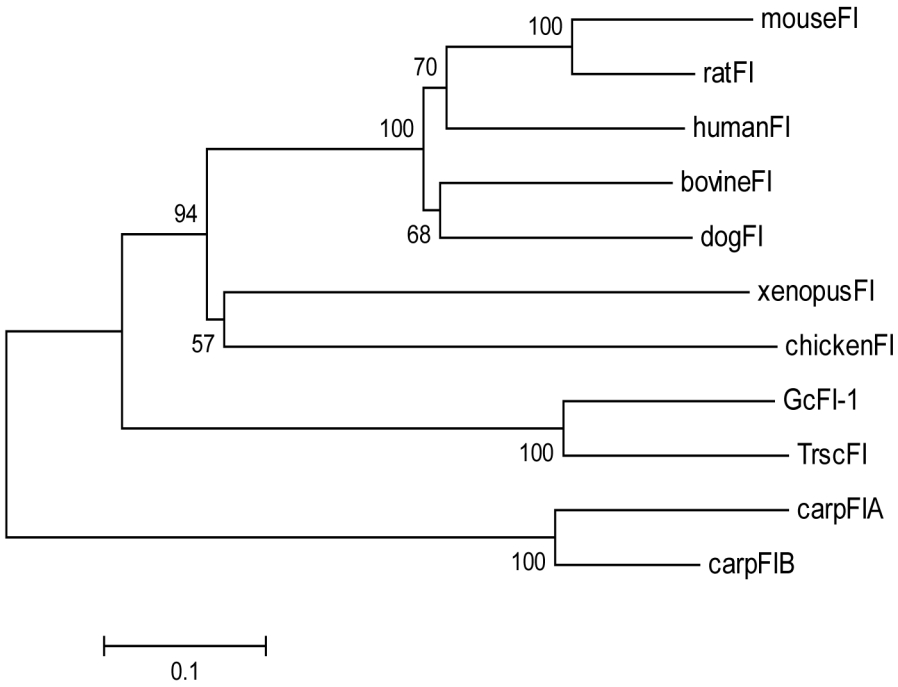

Figure 7.
Phylogenetic tree of the nurse shark complement factor I with that of other species, drawn by the neighbor-joining method. Number on a junction of each branch shows the bootstrap percentage supporting a given partition. A bar indicates a genetic distance of 0.1.

Table 1

Primers used in sequence analysis, synthesis of PCR-DIG probes, RT-PCR analysis, and identification of introns of nurse shark factor I (GcIf)

Name of primer	Sequence (5′→3′)	Position[#]
P1	ATGTTTGCTCAAGCAATGAGA	537–557
P2	TGCTTGGATGGAAGTGATG	1127–1146
P3	CTATTCAGACCCAAGAAGAC	1687–1707
P4	TAGGACATCGTGACTCAG	1698–1715
P5	TCACTTAGGTCAGCACAGT	1108–1126
P6	TGGCATGACTTATAGGTGA	511–530
P7	GATAACACAGGATTATTGGTGGTAG	2239–2259
P8	TGATGAAAGAAATGAAGCCTAC	1955–1976
P9	GGATTATTGGTGGTAGTTGTC	2235–2255
P10	CAGGATATACTGACAGGTGA	2731–2707
P11	CAGCACTGAAGCACAATACT	170–189
P12	AGGAACACTCTTGTGAGAAG	13–32
P13	GTTTCGGCCCTTTCACCTGT	113–132
P14	CAGCACTGAAGCACAATACT	170–189
P15	TGGACTTGTGTGGCTGCT	237–254
P16	AAGCACTGAAGCACAATACA	266–285
P17	CACAGTTTGATTGCTTCCTCT	309–329

[#] position of primers are representatively indicated on the largest sequence, GcIf-1, of the four nurse shark factor I cDNA sequences,

Table 2

Amino acid identity of nurse shark-factor I with known mammalian, amphibian, bird and fish factor I.

	Amino acid identity (%)			
	GcIf-1	**GcIf-2**	**GcIf-3**	**GcIf-4**
Human-FI	36.0	36.9	36.9	37.7
Mouse-FI	35.4	36.3	36.3	37.1
Rat-FI	36.4	37.3	37.3	38.2
Bovine-FI	38.8	38.8	38.8	39.6
Dog-FI	36.6	36.0	36.0	36.7
Xenopus-FI	35.7	36.6	36.6	37.4
Chickin-FI	35.7	36.1	36.0	37.8
Carp-FI-A	31.5	32.4	32.4	34.7
Carp-FI-B	36.4	37.4	38.0	38.2
* Trsc-FI	62.8	64.3	64.3	65.9

* Trsc indicates the banded houndshark ***Tr****iakis* ***sc****yllium*.