# Cloning and characterization of a cDNA representing a putative complement-regulatory plasma protein from barred sand bass (Parablax neblifer)

Antje DAHMEN,\* Toshio KAIDOH,† Peter F. ZIPFEL† and Irma GIGLI§||

\*Department of Molecular Immunology and tMolecular Biology, Bernhard Nocht Institute for Tropical Medicine, Bernhard-Nocht-Strasse 74, D-20359 Hamburg, Germany, \$Faculty of Bioscience, Fukui Prefectural University, Yoshido-gun, Fukui-ken 910-11, Japan, and §Department of Medicine, Division of Dermatology, University of California at San Diego School of Medicine, 200 West Arbor Drive, San Diego, CA 92103-8420, U.S.A.

It has been demonstrated previously that plasma from a number of vertebrate species including the phylogenetically old barred sand bass possesses molecules that cleave the  $\alpha'$ -chain of the activated third (C3b) and fourth (C4b) components of the human complement system. A specific protease and <sup>a</sup> cofactor protein were identified to be responsible for this cleavage. The cofactor activity in sand bass correlated with a 110 kDa polypeptide chain of a 360 kDa plasma protein. The evolutionary conservation was probed at the cDNA level and subsequently <sup>a</sup> cDNA clone of barred sand bass was isolated that represents <sup>a</sup> protein with structural similarity to mammalian complementregulatory proteins. The cDNA (SB1) was identified by immunoscreening of a sand bass liver expression library using affinitypurified IgG antibodies raised against the isolated 110 kDa material. The cDNA is <sup>3397</sup> bp in size and the open reading

# INTRODUCTION

C4b-binding protein (C4bp) [1,2] and factor H [3,4] are complement-regulatory proteins of human serum and that of other mammals. They modulate the activation and function of the C3-converting enzymes of both the classical (C4b2a) and alternative (C3bBb) pathways. C4bp binds to C4b thereby blocking its association with C2 or displacing C2a from the C4b2a complex [5-7]. Similarly, factor H prevents binding of factor B to C3b and dissociates Bb from the C3bBb complex [4,8]. In addition, both C4bp and factor H function as cofactors for the serum protease C3b/C4b inactivator (factor 1), thereby mediating the proteolytic degradation of C3b and C4b and the formation of haemolytically inactive C3bi and C4bi respectively.

Human [9] and mouse [10,1 1] factor H have been shown to be glycoproteins composed of <sup>a</sup> 150 kDa polypeptide chain. Human factor H is an elongated molecule, as shown by its high frictional ratio [9]. A similar high frictional ratio has been determined for human C4bp [12]. C4bp is composed of seven identical 75 kDa subunits ( $\alpha$ -chains) and one 45 kDa  $\beta$ -chain which are linked by disulphide bonds. The  $\alpha$ -chain is the site of attachment of protein S of the coagulation system [13,14]. Murine [15] and guinea-pig [16] C4bps are also composed of multiple subunits which, however, are linked by non-covalent forces.

Analysis of the primary structure of factor H [17] and the  $\alpha$ chains of C4bp [14,18] has shown that these polypeptides contain frame represents a protein of 1053 amino acid residues with a hydrophobic signal peptide indicative of a secreted protein. The calculated mass of the mature protein (SBP1) is 115.2 kDa which is in good agreement with the molecular mass of 110 kDa determined for the sand bass serum protein. Similarly to mammalian complement-regulatory proteins, the protein deduced from the sand bass cDNA is organized into short consensus repeats (SCR). It consists of 17 SCRs, of which SCRs 2, 12 and <sup>16</sup> exhibit significant homology to SCRs 2, 15 and <sup>19</sup> of human factor H, and SCRs 11, <sup>12</sup> and <sup>13</sup> have homology to SCRs 1, 2 and <sup>3</sup> of human C4b-binding protein. For the first time a complete cDNA representing <sup>a</sup> putative complement-regulatory protein which is structurally related to mammalian complement proteins has been isolated from a bony fish.

tandem repetitive elements of approximately 60 amino acid residues, termed short consensus repeats (SCRs). Characteristic of the SCRs is the presence of four conserved cysteine residues, which are linked to each other by disulphide bonds. Highly conserved within SCRs are certain proline, tryptophan, tyrosine/ phenylalanine and glycine residues [19]. SCRs are found in other plasma and cell-membrane proteins with very diverse functions [20] suggesting that their corresponding genes may have evolved early in evolution from an ancestral SCR sequence [21,22].

The proteins of the complement system are phylogenetically old. Invertebrates possess a 'cytolytic system' that can be activated by cobra venom factor [23] and a protein homologous to C3 of mammalian complement has been isolated from plasma of primitive vertebrates [24-29]. We have previously shown that plasma and serum of species that preceded humans in evolution by 300 million years possess a proteolytic activity for C4b and C3b of human origin that is capable of generating fragments comparable with those obtained with the human regulatory proteins, factors 1, H and C4bp [30,31]. We have investigated the molecular basis of this phenomenon using barred sand bass plasma. After separation of the plasma proteins by molecularsieve chromatography, C4b and C3b cleavage was obtained only when two fractions of different molecular size were combined. One fraction contained 360 kDa proteins and displayed cofactor activity, whereas the other contained a 155 kDa protein which was functionally similar to human factor I. The high-molecular-

Abbreviations used: SCR, short consensus repeat; C4bp, C4b-binding protein; 1 x SET, 100 mM NaCl+20 mM EDTA+50 mM Tris/HCl, pH 8.0, +0.01 % SDS.

To whom correspondence should be addressed.

The nucleotide sequence reported in this paper will appear in the EMBL/GenBank/DDBJ Nucleotide Sequence Databases under the accession number L21703.

mass material was purified further and found to consist of two identical polypeptide chains of approximately 110 kDa and a 42 kDa chain which was antigenically distinct. The functional activity of the molecule resides in the <sup>110</sup> kDa chains [31]. On the basis of the functional data and amino acid analysis it was concluded that the 360 kDa protein isolated from sand bass plasma may have structural similarities to either C4bp or factor H or both.

In this paper we describe the structure of <sup>a</sup> cDNA clone, which was identified using an affinity-purified IgG antibody raised against the 110 kDa polypeptide chain of the purified sand bass cofactor protein. The cDNA sequence revealed that the encoded protein, SBP1, is composed of 17 SCRs. Individual SCRs display amino acid sequence homology to human and mouse factor H and to a lesser degree to human C4bp.

#### MATERIALS AND METHODS

### Preparation of DNA and RNA

Parablax neblifer (class: Osteichthyes; order: Perciformis; common name: barred sand bass) were captured in the Pacific ocean and immediately frozen. The liver was isolated and homogenized in <sup>4</sup> M guanidine thiocyanate. DNA and RNA were isolated by CsCl-gradient centrifugation as described [32]. mRNA was purified using oligo(dT) coupled to magnetic beads as described by the manufacturer (Dynal, Hamburg, Germany).

# Construction of cDNA libraries

Synthesis of two oligo(dT)-primed sand bass liver cDNA expression libraries was performed according to the manufacturer's protocol using  $\lambda$ gtl 1 (Promega) and  $\lambda ZAP$  (Stratagene, La Jolla, CA, U.S.A.) bacteriophage DNA as vectors.

# cDNA screening

Affinity-purified rabbit IgG antibody raised against the 110 kDa chain of the sand bass cofactor protein isolated from plasma [31] was used to screen the  $\lambda$ gtl 1 cDNA library. The isolated cDNA fragment (cl) was used for screening, and a <sup>5</sup>' probe of the isolated fragment (c90) for a second screen to yield fragment c120.

## Sequence analysis of the cDNA clones

The purified cDNA inserts and the primer-extension products were sequenced by the dideoxy chain-termination method [33], using  $[\alpha$ -[35S]thio]dATP and Sequenase II (U.S. Biochemical Corp.). Various oligonucleotides were synthesized (gene assembler-1; Pharmacia) and used as primers to sequence the cDNA inserts in both orientations.

## Primer extension and PCR ampliffication

An oligonucleotide complementary to nucleotides 340-358 of the full-length sequence was used as a primer for the reverse transcription of sand bass liver mRNA using Superscript reverse transcriptase (BRL). The cDNA was then adenylated at the <sup>3</sup>' end using a poly(A) tailing kit (Boehringer), and the resulting fragment was amplified by PCR, using  $Taq$  polymerase (Perkin-

Elmer/Cetus), oligo(dT) and the extension primer [34]. After 30 cycles of amplification (Perkin-Elmer/Cetus thermocycler: 40 <sup>s</sup> at 95 °C, <sup>1</sup> min at 50 °C, <sup>1</sup> min at 72 °C), the amplified product was cloned into the pCRII vector (Invitrogen).

## Northern-blot analysis

Total sand bass RNA (10  $\mu$ g) was separated by electrophoresis in a formaldehyde/agarose gel and subsequently transferred to a nitrocellulose membrane (Schleicher und Schiill) [35].

## Southern-blot analysis

Genomic DNA was prepared from sand bass liver DNA; for analysis 10  $\mu$ g was digested to completion with the restriction enzyme EcoRI or HindIII. The fragments were separated by electrophoresis in a  $1\%$  agarose gel and transferred to a nitrocellulose membrane [35].

# Labelling and hybridization

For the isolation of full-length cDNA clones, three cDNA fragments were used: P1 is 916 bp long (position 2172-3088), and P2 and P3 represent 142 bp (position 166-308) and 495 bp (position 1012-1507) respectively. Two different probes were labelled and used for Southern-blot analysis: fragment P5 represents the <sup>5</sup>' end (position 166-1023) and P6 the <sup>3</sup>' end (position 1023-3088) of the complete cDNA termed SBl. For Northern-blot analysis three different probes were used: P1, P3 and P4. The last one is a fragment representing the <sup>5</sup>' end of a unique cDNA clone, clone 29. The nucleotide sequence of this clone (2.1 kb) differed from the sequence of SB<sup>1</sup> cDNA at the <sup>5</sup>' end, but was identical with approximately 300 bp at the distal <sup>3</sup>' end. The various cDNA fragments were excised, separated on low-melting-point agarose gels and labelled using the multiprime labelling system (Promega). Hybridization was performed at 50 °C (4 × SET,  $10 \times$  Denhardt's, 0.1% sodium pyrophosphate, 0.1 $\degree$ <sub>0</sub> SDS, 50  $\mu$ g/ml salmon sperm DNA) for 14 h. The filters were washed at <sup>52</sup> °C in <sup>1</sup> <sup>x</sup> SET (100 mM NaCl, <sup>20</sup> mM EDTA, 50 mM Tris/HCl, pH 8.0, 0.01  $\%$  SDS) for 2 h, and were exposed at  $-70$  °C using intensifying screens (Quanta III, Dupont-New England Nuclear).



#### Figure <sup>1</sup> Isolation of the barred sand bass cDNA (SB1)

Clone 1 (c1) was isolated using an affinity-purified antibody against the 110 kDa chain of barred sand bass cofactor protein. cDNA probes P1, P2 and P3 were used for the isolation of additional cDNA clones (c90 and cl 20). Fragments used for Northern- (P1, P3) and Southern- (P5, P6) blot and primer-extension (PE) analysis are also indicated.

1 tctgagtctataatagagtttcagtctctaggtttctacttctgtttggcggttcagggt<br>M D V I A R S C V L F L W L Q T 16 aatgaagaaaaa ATGACGTCATAGCTCGAAGCTGTGTCCTATTTTTGTGGCTGCAAACA<br>L T S V K S Q V P C T L Q Q F I D G E H L T <sup>S</sup> V K <sup>S</sup> 0 V <sup>P</sup> C T I, Q Q <sup>F</sup> <sup>I</sup> D G E H 36 121 TTGACCTCTGTGAAAAGTCAGGTACCCTGTACACTTCAACAGTTTATTGATGGGGAACAT Y D <sup>S</sup> N <sup>F</sup> D <sup>T</sup> T G M E A <sup>S</sup> Y P G G R Q V 56 181 TATGATTCAAATTTTGACACAACTGGCATGGAAGCCAGCTACCCTGGTGGAAGACAAGTG R V G C N V G Y <sup>S</sup> G F <sup>F</sup> K L V C V <sup>E</sup> G K 76 241 AGAGTGGGCTGCAATGTTGGCTACAGTGGCTTTTTTAAGCTGGTCTGTGTCGAGGGAAAA W E T R G A K C Q P R S C G H P G D A Q 96 F A D F H L A E G N D F V F G S K V V Y 116 361 TTTGCAGATTTTCACCTGGCAGAAGGGAATGATTTTGTCTTTGGATCAAAAGTTGTGTAC T C Q K G Y Q M V <sup>S</sup> R <sup>I</sup> N Y R R C V A E 136 421 ACTTGCCAAAAAGGTTATCAAATGGTCAGTCGCATAAACTACCGGCGTTGCGTGGCTGAA G W D G V V P V C E S Q Q C P L I H V D 156<br>481 GGCTGGGATGGTGTGGTTCCTGTTTGTGAATCCCAACAGTGCCCATTGATTCATGTGGAC N N V Q V I G G P E E A T F G N V V R F 176 <sup>S</sup> C K <sup>S</sup> <sup>R</sup> <sup>S</sup> <sup>E</sup> <sup>I</sup> L D G <sup>S</sup> <sup>P</sup> <sup>E</sup> L <sup>Y</sup> C D <sup>E</sup> R 196 601 AGCTGCAAGTCCAGAAGTGAAATCCTGGATGGGTCACCAGAGCTGTATTGTGATGAACGT<br>G D W S G P V P K C K A T T C A T P P I 216 661 GGAGATTGGAGTGGCCCAGTICCAAAATGCAAAGCTATAACATGTGCCATACCTCCCATT<br>E N G N V P G A I R E Y K E N D V L H Y E N G N V P G A <sup>I</sup> R E Y K E N D V L H Y 236 721 GAAAATGGTAATGTTCCTGGCGCTATCCGAGAATACAAAGAAATATGTGCTGCACTAT E C D R A F K H <sup>I</sup> D R <sup>P</sup> <sup>S</sup> T C <sup>I</sup> K Q G <sup>I</sup> 256 781 GAGTGTGATCGAGCATTTAAACACATAGACAGACCTTCAACATGCATTAAACAAGGAATA K A <sup>E</sup> W <sup>S</sup> <sup>P</sup> <sup>T</sup> <sup>P</sup> <sup>L</sup> C <sup>E</sup> <sup>S</sup> <sup>I</sup> <sup>K</sup> C R L T <sup>I</sup> M 276 841 AAAGCAGAGTGGAGCCCGACTCCTCTGTGTGAATCAATAAAATGCAGACTGACTATAATG D G T R Y E P A Y R N L F S P G E T L K 296<br>901 GATGGAACCCGGTACGAACCTGCTTATAGAAATTTGTTTTCACCTGGTGAAACACTGAAA V I C A R T S W I S T P Q E T S V V T T 316<br>961 GTCATTTGTGCACGCACATCCTGGATTTCTACCCCTCAGGAAACCTCCGTAGTAACAACA C Q D N G E W S I R P T C Q E V R C S N 336<br>1021 TGCCAAGACAATGGAGAGTGGTCGATGGTCAGACCGACATGCCAAGAGGTGAGATGCAGCAAT R R P E H V D S W D V R S W E R Y T L D 356<br>1381 CGAAGACCAGAACATGTGGATTCCTGGGATGTGAGATCTTGGGAACGATACACATTGGAT D N T R Y W C K R G Y K R T G G V T W A 376<br>1141 GACAATACAAGATATTGGTGTAAGAGAGGCTACAAGAGAGACAGGTGGTGTCACCTGGGCC T C G R N G W M P N P L C E V K T C S K 396<br>1201 ACATGCGGCAGAAATGGGTGGATGCCAAATCCACTTTGTGAAGTAAAAACATGCAGCAAA E N <sup>I</sup> Q D A V <sup>I</sup> V G T D K Q <sup>I</sup> Y N L N Q 416 1261 GAGAATATCCAGGATGCAGTTATTGTTGGCACTGACMGCAGATATACMCT TAAATCAA K A <sup>I</sup> Y A C G E G N R G R <sup>I</sup> T L T C G E 436 1321 AAAGCCATTTATGCCTGCGGTGAAGGTAATCGGGGAAGAATTACTCTAACCTGTGGAGAA N G W S G D R K C T V K P C P L P P K D 456<br>1381 AATGGTTGGTGGGATAGAAAATGTACAGTTAAACCATGTCCACTTCCACTTAAAACT<br>P N G F F R G P Y T G R V L Y Y T C K D 476 1441 CCCAATGGATTTTTTCGTGGTCCATACACTGGCAGAGTACTGTACTACTCGCAAGACCTGCAAAGACCGTACTGTACTACTGCAAAGACCGTACTGCAGAGTACTGTACTACTGCAGAGTACTGTACTACTGCAGAGTACTGTACTACTGCAGAGTACTGCAGAGTACTGCAGAGTACTGCAGAGTACTGCAGGTACTGCAGGTACTGCAGGTACTG G Y K L <sup>F</sup> T <sup>E</sup> G W W A <sup>E</sup> A K C V D G V W 496 1501 GGTTATAACCTTTTCACTGAAGGCTGGTGGGCTGAGGCCAAATGTGTCGATGGCGTGTGG P E L T T C <sup>I</sup> S N T T C G K F P E <sup>I</sup> <sup>P</sup> N 516 1561 CCTGAACTGACAACATGCATCAGTAATACAACGTGTGGTAAATTTCCTGAGATTCCTAAT <sup>A</sup> <sup>E</sup> <sup>V</sup> <sup>I</sup> <sup>R</sup> <sup>R</sup> <sup>Y</sup> <sup>P</sup> <sup>E</sup> V Q T V Q V I <sup>c</sup> N Q G <sup>536</sup> 1621 GCGGAAGTGATACGTAGATATCCAGAGGTCCAGACGGTCCAGGTTATATGTAATCAAGGA



### Figure 2 Nucleotide and derived protein sequence of SB1

Untranslated regions are indicated by lower-case letters; the open reading frame is marked in upper-case letters. Start and stop condons are shown in bold letters. Nucleotide numbering is on the left side of each column, amino acid positions are above the cDNA sequence and amino acid numbering is on the right side. The putative N-linked glycosylation sites (505-507, 667-669) and the RGD motif (196-198) are underlined. Italic bold letters correspond to the N-terminal sequence of a 73 kDa (VQTVQVIxNQGY) and a 35 kDa (IGVMPTHVSxVGLPPTK) fragment of the purified plasma protein obtained by digestion with Staphylococcus aureus V8 protease (50  $\mu$ g of barred sand bass purified cofactor protein/0.1 unit V8 protease).

# RESULTS

#### Isolation of SB1 cDNA

In order to identify <sup>a</sup> cDNA clone representing the <sup>110</sup> kDa polypeptide chain of the cofactor protein isolated from sand bass plasma [31], an oligo(dT)-primed sand bass liver cDNA expression library was synthesized and screened with an affinitypurified IgG antibody specific for the purified 110 kDa polypeptide chain. A cDNA clone (clone 1) was isolated, and sequence analysis revealed that it was not <sup>a</sup> full-length clone. A <sup>5</sup>'-end fragment of this clone (P1) was used to screen <sup>a</sup> second cDNA library constructed in  $\lambda ZAP$ . Two additional clones (clone 90) and 120) were isolated and the nucleotide sequences determined. Clone 90 was 3234 bp long (position 163-3397) and clone 120 was 1164 bp long (position 1-1164). Fragment c120 represents a partial cDNA clone, as it includes the <sup>5</sup>' linker sequence used for cDNA cloning, but lacks <sup>a</sup> poly(A) tail (results not shown). In the region that overlaps with c90 the nucleotide sequences of the two clones are identical. An alignment of the three clones (Figure 1) shows that together they represent almost the full-length

cDNA. The complete cDNA sequence, termed SBI, is shown in Figure 2. The sequence consists of 3397 bp including a  $poly(A)$ signal, 'AATAAA,' (position 3359-3364).

In order to ascertain whether the SB1 cDNA sequence was nearly full length, a PCR-based primer-extension analysis was performed. The primer-extension product was cloned and sequenced. The sequence ends approximately 20 nucleotides upstream from the cDNA of clone 120. The experiment demonstrates that the cDNA clone represents the nearly complete <sup>5</sup>' end of the authentic mRNA

## Deduced protein sequence

The nucleotide sequence of SB<sup>1</sup> displayed an open reading frame of 1053 amino acids encoding a protein (SBP1) of 117.5 kDa. Amino acids in positions 526-537 and 809-826 are in agreement with the N-terminal sequence of <sup>a</sup> <sup>73</sup> and <sup>35</sup> kDa fragment obtained by proteolytic digestion of the plasma protein (VQTVQVIxNQGY and IGVMPTHVSxVGLPPGTK respectively) (Figure 2). Within the predicted amino acid sequence of SBP1, two potential N-linked glycosylation sites of the type Asn-X-Ser/Thr were found at positions 505-507 (Asn-Thr-Thr) and 667-669 (Asn-Asn-Thr). A tripeptide motif Arg-Gly-Asp was present at position 196-198. This motif is found in various adhesion proteins which are recognized by special cell-surface receptors [36]. Analysis of the N-terminal amino acid sequence revealed a hydrophobicity pattern indicative of a signal peptide, suggesting that SBP1 is secreted [37]. According to the criteria common for signal peptide cleavage, we suggest that the signal sequence is cleaved at position 21 (lysine-serine) [37]. The molecular mass of the secreted non-glycosylated product was calculated to be 115.2 kDa.

# Structural analysis and homology

The processed SBP1 is organized into repetitive elements (SCRs) found in a number of complement-regulatory proteins [19]. Alignment of these SCRs (Figure 3) showed the organization of SBP1 into 17 SCRs, each containing four typical cysteine residues (C). In addition, a proline (P), a phenylalanine (F) or tyrosine (Y), a glycine (G), a tryptophan (W) and another proline (P) can be found in most SCRs.

A homology search indicated structural similarity between SBPI and human and mouse factor H, human C4bp, CR1 and other SCR-containing proteins. Comparison of the SBP1 deduced amino acid sequence reveals an overall similarity to these proteins of  $20-25^\circ$ <sub>0</sub>, but individual SCRs display a higher degree of homology (Figure 4). SCR 2, <sup>12</sup> and <sup>16</sup> of SBP1 show 51, 43 and  $52\degree$ <sub>0</sub> homology to SCR 2, 15 and 19 of human factor H respectively, and SCR 11, <sup>12</sup> and <sup>13</sup> show homology of 33, <sup>40</sup> and  $38^\circ$ <sub>0</sub> to SCR 1, 2 and 3 of human C4bp respectively. A comparison of the conserved amino acids is shown in Figure 5.

# Expression analysis

The presence of the SBP1 mRNA in sand bass liver was demonstrated by Northern-blot analysis. A fragment representing the <sup>3</sup>' end of the SB1 cDNA (P1) detected four different RNAs of 1.4, 1.8, 2.4 and 3.5 kb (Figure 6). To gain <sup>a</sup> better understanding of this complex pattern, an additional probe specific for the <sup>5</sup>' region of SB1 (P3) was used. This probe hybridized to <sup>a</sup> single mRNA of 3.5 kb. The complex band pattern observed by Northern-blot analysis correlated with the findings at the cDNA level. Several incomplete cDNA clones were isolated; partial-sequence analysis in combination with restriction-fragment analysis revealed that some clones displayed distinct sequences and were only in part identical with SB1. Alignment of the nucleotide sequence of one of these clones (clone 29) showed that the <sup>3</sup>' end was identical with SB 1, whereas the <sup>5</sup>' end was distinct. Northern-blot analysis using the unique <sup>5</sup>' end fragment of clone 29 as probe (P4) revealed predominant signals with 2.4, 1.8 and 1.4 kb mRNAs that were also visualized with the P1 probe.

# Southern-blot analysis

Southern-blot analysis was performed to determine the genomic structure of the SBI locus. The sand bass genomic DNA was probed with two different fragments of the SB1 cDNA: one probe represented SCRs 1-5 (P5) and the second SCRs 6-17 (P6). Although with varying intensity, both probes hybridized to several bands of EcoRI- and HindIII-digested sand bass DNA (Figure 7). The <sup>5</sup>'- and 3'-end probes hybridized to distinct fragments. The identified genomic DNA bands represent at least



tnU)

 $\sim$ :s  $\epsilon$ 

i.  $\ddot{\phantom{1}}$ --

cn E  $\simeq$ **E** ∙ , **.** 0 :5 W ្ថ E°  $\alpha$ ູ n E **Co** .ह ۲. . ত x r<br>E ă Ì E <u>ছে</u> ই ខ្ពុ <sup>o</sup> - ទ្ធ<br>ខ 5 e **≧** ឌូ i<br>E<br>a ج<br>چ  $\mathbf{g}$   $\mathbf{g}$ **C** ضا∞<br>C  $\frac{5}{9}$  –  $\frac{1}{2}$   $\frac{1}{2}$  **ទី** ೭៖



## Figure 4 Schematic comparison of SBP1, human factor H and human C4bp

SCRs are numbered consecutively starting with number 1. Related SCRs are aligned and the numbers between the aligned SCRs indicate the similarity (%). SCRs that display homology of 40% or more are shaded.



# Figure <sup>5</sup> Comparison of amino acid sequence of SCRs with more than 40% identity of SBP1, human and mouse factor H and human C4bp

The conserved residues are aligned and the tour structurally important cysteine residues (C) are boxed with double lines.

50 kb, indicating either a complex genomic organization of the SB1 gene or the existence of additional SB1-related genes.

# **DISCUSSION**

It has previously been demonstrated that serum of a number of vertebrate species possesses molecules that cleave the  $\alpha'$ -chain of the activated human third (C3b) and fourth (C4b) components of the complement system [30]. To understand such an evolutionary conservation at the molecular level, we have purified and characterized the plasma proteins that participate in the cleavage reaction from the most primitive species in which specific cleaving activity has been demonstrated (barred sand bass). Analogously to the human and mouse systems, <sup>a</sup> specific protease and a cofactor protein were identified. The cofactor activity correlated with <sup>a</sup> 110 kDa polypeptide chain of <sup>a</sup> 360 kDa plasma protein [3 1]. Affinity-purified IgG antibodies raised against the isolated 110 kDa material were used to immunoscreen <sup>a</sup> sand bass liver cDNA library, and <sup>a</sup> near-fulllength cDNA clone (SBl) of <sup>3397</sup> bp was isolated which represents a putative complement-regulatory protein. The predicted amino acid sequence (SBPI) contains <sup>a</sup> hydrophobic signal peptide. The calculated molecular mass of the predicted mature SBPI polypeptide chain is 115.2 kDa (1023 amino acid residues), which is close to the 110 kDa molecular mass estimated for the serum cofactor protein by SDS/PAGE under nonreducing conditions [31]. Two putative N-linked glycosylation sites were identified (NTT and NNT) in the deduced amino acid sequence of SBPI at position 505-507 and 667-669 respectively. This finding is in agreement with the observation that the cofactor protein isolated from plasma is glycosylated (I. Gigli, unpublished work). Of interest is the finding of an RGD motif at position 196-198. Amino acids 526-537 (VQTVQ-VICNQGY) correspond to the sequence (QVTVQVIxNQGY) of the N-terminus of <sup>a</sup> <sup>73</sup> kDa peptide derived from proteolytic cleavage of the purified barred sand bass plasma protein (Figure 2). This sequence is represented by c90, which shows  $100^\circ$ <sub>o</sub> homology to c1 and also shares an overlapping sequence of 1 kb with  $100\degree$  homology to c120. Amino acids 809-826 (IGVMPTHVSxVGLPPGTK), represented by <sup>c</sup><sup>I</sup> and c90, correspond to the N-terminus sequence obtained from <sup>a</sup> <sup>35</sup> kDa fragment of the plasma protein. The amino acid composition of



Figure 6 Northern-blot analysis of sand bass liver RNA

Total RNA separated on denaturing agarose gels was blotted and the filters were probed with three distinct cDNA probes (P1, P3 and P4). The regions represented by P1 and P3 are shown in Figure 1. P4 represents the unique sequence of <sup>a</sup> cDNA clone, with <sup>a</sup> <sup>3</sup>' end identical with SB1 but with a distinct <sup>5</sup>' end. The positions of the RNA standard markers are shown in (kb) on the left.

SBPI deduced from the cDNA is very similar to that of the 110 kDa polypeptide chain isolated from the 360 kDa sand bass plasma protein [31]. Taken together these data support the idea that SB1 represents the complement-regulatory protein previously identified in barred sand bass.

The isolation of several distinct cDNA clones and the results of the Northern- and Southern-blot analyses indicate that SB1 is a member of a multigene family. Similarly in mouse and man, several factor-H-related genes have been demonstrated at both the cDNA [38-40] and protein [41] level. As observed with SBl, the multiple hybridization bands reflect either the presence of several related genes or a complex organization of a single gene.

Similarly to other complement proteins, SBP1 is organized into SCRs. A structural comparison demonstrates significant homology to the complement-regulatory protein factor H [17,42,43]. Both proteins are secreted, but SBPI is composed of <sup>17</sup> SCRs, whereas human and mouse factor H are composed of 20 SCRs. In addition to the identical structural organization, individual SCRs of SBPI and human factor H display an amino acid sequence homology that exceeds that due to conserved SCR motifs. Homologies of 51, 43 and  $52\degree$ <sub>0</sub> were observed between the amino acid sequence in SCR 2, <sup>12</sup> and <sup>16</sup> of SBPI and SCR 2, <sup>15</sup> and <sup>19</sup> ofhuman factor H respectively. A similar comparison with human C4bp shows homologies of 30, 40 and 38 $\degree$ <sub>0</sub> in SCR 11, <sup>12</sup> and <sup>13</sup> of SBPI and SCR 1, <sup>2</sup> and <sup>3</sup> of C4bp. Because similar functions have been demonstrated for the purified sand bass cofactor protein, human factor H and C4bp, it is suggested that there are analogous functional domains conserved in these molecules. The high degree of homology in SCR <sup>2</sup> between SBP <sup>I</sup> and human factor H suggests that the N-terminal region of SBP1 may correspond to the region of factor H responsible for complement-regulatory activity [44]. Similarly the conserved domains SCR <sup>12</sup> in SBPI and SCR <sup>2</sup> human C4bp might contribute to the C4b-binding activity [45,46].

For the first time <sup>a</sup> complete cDNA representing <sup>a</sup> protein that is structurally related to mammalian complement-regulatory proteins has been isolated from <sup>a</sup> bony fish. On the basis of its genomic organization as well as the functional observations on its putative gene product, the sand bass cofactor protein is an



#### Figure 7 Southern-blot analysis

Sand-bass genomic DNA was restricted with EcoRI (E) or HindIII (H). cDNA probes P5 and P6 representing the <sup>5</sup>' and <sup>3</sup>' ends of SB1 (see Figure 1) were used for hybridization. Positions of DNA standards in kb are shown on the left.

interesting candidate for investigation of the phylogeny of complement-regulatory proteins.

This work was supported by NIH grant A120067 and a grant from the Deutsche Forschungsgemeinschaft (Zi432/1). We thank Sea World of San Diego for their generous donation of barred sand bass.

# REFERENCES

- <sup>1</sup> Scharstein, J., Ferreira, A., Gigli, I. and Nussenzweig, V. (1976) J. Exp. Med. 148, 1207-1211
- 2 Nagasawa, S., Ichihara, C. and Stroud, R. M. (1980) J. Immunol. 125, 578-582
- 3 Whaley, K. and Ruddy, S. (1976) J. Exp. Med. 144, 1147-1163
- 4 Weiler, J. M., Daha, M. R., Austen, K. F. and Fearon, D. T. (1976) Proc. Natl. Acad. Sci. U.S.A. 734, 3268-3272
- 5 Gigli, I., Fujita, T. and Nussenzweig, V. (1979) Proc. Natl. Acad. Sci. U.S.A. 76, 6596-6600
- 6 Fujita, T., Gigli, I. and Nussenzweig, V. (1978) J. Exp. Med. 148, 1044-1051
- 7 Nagasawa, S. and Stroud, R. M. (1977) Immunochemistry 14, 749-756
- 8 Pangburn, M. K. and Müller-Eberhard, H. J. (1978) Proc. Natl. Acad. Sci. U.S.A. 75, 2416-2419

397

- 9 Sim, R. B. and Discipio, R. B. (1982) Biochem. J. **205**, 285–293<br>10 Kaidoh, T., Fuilta, T., Takata, Y., Natsuume-Sakai, S. and Takahas
- 10 Kaidoh, T., Fujita, T., Takata, Y., Natsuume-Sakai, S. and Takahashi, M. (1984) Complement 1, 44-51
- 11 Kristensen, T., Wetsel, R. A. and Tack, B. F. (1986) J. Immunol. 136, 3407-3411
- 12 Dahlbäck, B. (1984) Biochem. J. 209, 847-851<br>13 Dahlbäck B. Smith C. A. and Müller-Eherhard
- Dahlbäck, B., Smith, C. A. and Müller-Eberhard, H. J. (1983) Proc. Natl. Acad. Sci. U.S.A. 80, 3461-3465
- 14 Hillarp, A. and Dahlbäck, B. (1990) Proc. Natl. Acad. Sci. U.S.A. 87, 1183-1187<br>15 Kai. S. Fuilta, T. Gigli, L. and Nussenzweig, V. (1980) J. Immunol, 125, 2409-2
- 15 Kai, S., Fujita, T., Gigli, I. and Nussenzweig, V. (1980) J. Immunol. 125, 2409-2413
- 16 Burge, J., Nicholson-Weller, A. and Austen, K. F. (1981) J. Immunol. **126**, 232–235<br>17 Ripoche, J., Dav. A. J., Harris. T. J. and Sim. R. B. (1988) Biochem. J. **249**
- 17 Ripoche, J., Day, A. J., Harris, T. J. and Sim, R. B. (1988) Biochem. J. 249, 593-602
- 18 Chung, L. P., Bentley, D. R. and Reid, K. B. M. (1985) Biochem. J. 230, 133–136<br>19 Kristensen, T., D'Eustachio, P., Ogata, R. T., Chung, L. P., Reid, K. B. M. and Tack
- Kristensen, T., D'Eustachio, P., Ogata, R. T., Chung, L. P., Reid, K. B. M. and Tack, B. F. (1987) Fed. Proc. 46, 2463-2469
- 20 Reid, K. B. M. and Day, A. J. (1989) Immunol. Today 10, 177–180<br>21 Rodriguez de Cordoba S. Lublin, D. M. Rubinstein, P. and Atkinsor
- Rodriguez de Cordoba, S., Lublin, D. M., Rubinstein, P. and Atkinson, J. P. (1985) J. Exp. Med. 161, 1189-1195
- 22 Farries, T. and Atkinson, J. P. (1991) Immunol. Today **12**, 295–300<br>23 Day N. K. B. Gewurz, H. Johannsen, B. Finstad, J. and Good, B. A.
- Day, N. K. B., Gewurz, H., Johannsen, R., Finstad, J. and Good, R. A. (1970) J. Exp. Med. 132, 941-950
- 24 Alsenz, J., Avila, D., Huemer, H. P., Esparza, I., Bechereer, J. D., Kinoshita, T., Wang, Y., Oppermann, S. and Lambris, J. D. (1992) Dev. Comp. Immunol. 16, 63-76
- 25 Nonaka, M., Fujii, T., Kaidoh, T., Natsuume-Sakai, S., Nonaka, M., Yamaguchi, N. and Takahashi, M. (1984) J. Immunol. 133, 3242-3249
- 26 Fujii, T., Nakamura, T., Sekizawa, A. and Tomonaga, S. (1992) J. Immunol. 148, 117-123
- 27 Ishiguro, H., Kobayashi, K., Suzuki, M., Titani, K., Tomonaga, S. and Kurosawa, Y. (1992) EMBO J. 11, 829-837
- 28 Nonaka, M., Iwaki, M., Nakai, C., Nozaki, M., Kaidoh, T., Nonaka, M., Natsuume-Sakai, S. and Takahashi, M. (1984) J. Biol. Chem. 259, 6327-6333

Received 29 December 1993; accepted 9 February 1994

- 29 Grossberger, D., Marcuz, A., Du Pascier, L. and Lambris, J. D. (1989) Proc. Natl. Acad. Sci. U.S.A. 86, 1323-1327
- 30 Kaidoh, T. and Gigli, I. (1987) J. Immunol. **139**, 194–201<br>31 Kaidob T. and Gigli J. (1989) J. Immunol. **142**, 1605–16
- 31 Kaidoh, T. and Gigli, I. (1989) J. Immunol. **142**, 1605-1613<br>32 Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. and Rutter
- 32 Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. and Rutter, W. J. (1979) Biochemistry 18, 5294-5299
- 33 Sanger, F., Nicklen, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 5463-5467 5463-5467
- 34 Frohman, M. A. (1990) in PCR Protocols: A Guide to Methods and Applications (Innis, M. A., Gelfand, D. H., Sninsky, J. J. and White, T. J., eds.), pp. 28-38, Academic Press, San Diego
- 35 Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- 36 Ruoslahti, E. and Pierschbacher, M. D. (1986) Cell 44, 517–518<br>37 von Heilne. T. (1986) Nucleic Acids Res. 14, 4683–4692
- 37 von Heijne, T. (1986) Nucleic Acids Res. 14, 4683-4692
- Skerka, C., Horstmann, R. D. and Zipfel, P. F. (1991) J. Biol. Chem. 266, 12015-12020
- 39 Skerka, C., Kuhn, S., Gunther, K., Lingelbach, K. and Zipfel, P. F. (1993) J. Biol. Chem. 268, 2904-2908
- 40 Vik, D. P., Munoz-Canoves, P., Kozono, H., Martin, L. G., Tack, B. F. and Chaplin, D. D. (1990) J. Biol. Chem. 265, 3193-3201
- 41 Skerka, C., Timmann, C., Horstmann, R. D. and Zipfel, P. F. (1992) J. Immunol. 148, 331 3-3318
- 42 Lintin, S. J., Lewih, A. R. and Reid, K. B. M. (1988) FEBS Lett. 232, 328-332
- 43 Kristensen, T. and Tack, B. F. (1986) Proc. Natl. Acad. Sci. U.S.A. 83, 3963–3967<br>44 Alsenz J. Tambris J. D. Schulz T. F. and Dierich P. (1984) Biochem J. 224
- Alsenz, J., Lambris, J. D., Schulz, T. F. and Dierich, P. (1984) Biochem. J. 224, 389-398
- 45 Nagasawa, S., Mizuguchi, K., Ichihara, C. and Koyama, J. (1982) J. Biochem. (Tokyo) 92, 1329-1332
- 46 Ogata, R. T., Mathias, P., Bradt, B. M. and Cooper, N. R. (1993) J. Immunol. 150, 2273-2280