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 †Molecular 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 α' -chain of the activated third (C3b) and fourth (C4b) components of the human complement system. A specific protease and a 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 a cDNA clone of barred sand bass was isolated that represents a 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 3397 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 I), thereby mediating the proteolytic degradation of C3b and C4b and the formation of haemolytically inactive C3bi and C4bi respectively.

Human [9] and mouse [10,11] factor H have been shown to be glycoproteins composed of a 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 (α -chains) and one 45 kDa β -chain which are linked by disulphide bonds. The α -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 α 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 16 exhibit significant homology to SCRs 2, 15 and 19 of human factor H, and SCRs 11, 12 and 13 have homology to SCRs 1, 2 and 3 of human C4b-binding protein. For the first time a complete cDNA representing a 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 I, 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 × SET, 100 mM NaCI+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 110 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 a 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 4 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 $\lambda gt11$ (Promega) and λZAP (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 $\lambda gt11$ cDNA library. The isolated cDNA fragment (c1) was used for screening, and a 5' 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 $[\alpha-[^{35}S]$ 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 amplification

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 3' 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 s at 95 °C, 1 min at 50 °C, 1 min at 72 °C), the amplified product was cloned into the pCRII vector (Invitrogen).

Northern-blot analysis

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

Southern-blot analysis

Genomic DNA was prepared from sand bass liver DNA; for analysis 10 μ g was digested to completion with the restriction enzyme *Eco*RI or *Hind*III. 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 5' end (position 166-1023) and P6 the 3' end (position 1023-3088) of the complete cDNA termed SB1. For Northern-blot analysis three different probes were used: P1, P3 and P4. The last one is a fragment representing the 5' end of a unique cDNA clone, clone 29. The nucleotide sequence of this clone (2.1 kb) differed from the sequence of SB1 cDNA at the 5' end, but was identical with approximately 300 bp at the distal 3' 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 ($4 \times SET$, $10 \times Denhardt's$, 0.1 % sodium pyrophosphate, 0.1 ° SDS, 50 μ g/ml salmon sperm DNA) for 14 h. The filters were washed at 52 °C in 1 × SET (100 mM NaCl, 20 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 1 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 c120). Fragments used for Northern- (P1, P3) and Southern- (P5, P6) blot and primer-extension (PE) analysis are also indicated.

16 aatgaagaaaaa ATGACGTCATAGCTCGAAGCTGTGTCCTATTTTTGTGGCTGCAAACA S V K S Q V P C T L Q Q F G 36 L T S V K S Q V P C T L Q Q F I D G E H 121 TIGACTETGIGANAGICAGTACCTIGALACTICALACITTAITGAIGGGAAACT Y D S N F D T G G M E A S Y P G G R Q V 181 TAIGATICAATITIGACAACTGGCATGGAAGCCAGCTACCTGGIGGAAGACAAATG R V G C N V G Y S G F F K L V C V E G K 241 AGAGTGGGCTGCAATGITGGCTACAGTGGGCTTTTTTAAGCTGGITGTGGIGGAGGAAAA W E T R G A K C Q P R S C G H P G D A Q 301 TGGGAGACTCGAGGGGCTAAATGTCAACCGAGATCATGTGGTCATCCTGGTGATGCCCAG 96 A D F H L A E G N D F V F G 116 361 TTTGCAGATTTTCACCTGGCAGAAGGGAATGATTTTGTCTTTGGATCAAAAGTTGTGTAC G QMV 136 0 5 R I N 421 ACTTGCCAAAAAGGTTATCAAATGGTCAGTCGCATAAACTACCGGCGTTGCGTGGCTGAA DGV PVC ESQQC t. 156 481 GGCTGGGATGGTGTGGTTCCTGTTTGTGAATCCCAACAGTGCCCATTGATTCATGTGGAC N N V Q V I G G P E E A T F G N V V R F 541 AATAATGTGCAGGTGATTGGGGGGCCCAGAGGAAGCAATGTAGTICGATTC 176 KSRSEILDGSP 196 ΕL 601 AGCTGCAAGTCCAGAAGTGAAATCCTGGATGGGTCACCAGAGCTGTATTGTGATGAACGT G D W S G P V P K C K A I T C A I P P I 216 661 GGAGATTGGAGTGGCCCAGTTCCAAAATGCAAAGCTATAACATGTGCCATACCTCCCATT PGAIR 721 GAAAATGGTAATGTTCCTGGCGCTATCCGAGAATACAAAGAAAATGATGTGCTGCACTAT E C D R A F K H I D R P S T C I K Q G I 781 GAGTGTGATCGAGCATTTAAACAACATAGACAGACCTTCAACATGCATTAAACAAGGAATA 276 YEP A Y R N LESPG 296 901 GATGGAACCCGGTACGAACCTGCTTATAGAAATTTGTTTTCACCTGGTGAAACACTGAAA 316 ΤS W I s P 0 E R 961 GTCATTTGTGCACGCACATCCTGGATTTCTACCCCTCAGGAAACCTCCGTAGTAACAACA 336 NGEW S I тсуе 1021 TGCCAAGACAATGGAGAGTGGTCCATCAGACCGACATGCCAAGAGGTGAGATGCAGCAAT R R P E H V D S W D V R S W E R Y T L D R P E H V D S W D V R S W E R Y T L D 1081 CGAAGACCAGAACATGTGGATTCCTGGGATGTGAGATCTTGGGAACGATACACATTGGAT 356 D N T R Y W C K R G Y K R T G G V T W A 1141 GACAATACAAGATATTGGTGTAAGAGAGGGCTACAAGAGGACAGGTGGTSTCACCTGGGCC 376 G w M P N P I 396 1201 ACATGCGGCAGAAATGGGTGGATGCCAAATCCACTTTGTGAAGTAAAAACATGCAGCAAA E N I Q D A V I V G T D K Q I Y N L N Q 1261 GAGAATATCCAGGATGCAGTTATTGTTGGCACTGACAAGCAGATATACAACTTAAATCAA 416 YACG EGN RGRI т 1. 436 1321 AAAGCCATTTATGCCTGCGGTGAAGGTAATCGGGGAAGAATTACTCTAACCTGTGGAGAA GWSGDRKC ΤΥΚΡΟΡ L P P 456 1381 AATGGTTGGTCCGGGGATAGAAAATGTACAGTTAAACCATGTCCACTTCCACCTAAAGAT PNGFFRGFVJJGTCGGGGATAGAAAATGTACAGTTAAACCATGTCCACTTCCACCTAAAGAT 476 1441 CCCAATGGATTTTTTCGTGGTCCATACACTGGCAGAGTACTGTACTACACCTGCAAAGAC G Y K L F T E G W W A E A K C V D G V W 496 516 ае V I R R Y Р Е **V Q T V Q V I с N Q G** 536 1621 GCGGAAGTGATACGTAGATATCCAGAGGTCCAGACGGTCCAGGTTATATGTAATCAAGGA

	Y	s	т	Q	A	N	s	F	s	с	Ε	N	G	N	W	L	L	Y	G	L	556
1681	TACT	cci	ACT	CAG	GCT	AAC	AGC	TTT	TCT	TGT	GAG	AT	GGA	AAC	TGG	стт	TTA	TAT	GGA	TTG	
	S	P	D	0	I	с	т	L	R	A	D	v	С	G	P	P	P	E	A	E	576
1741	TCAC	cc	GAT	CĀA	ATC	TGT		CTA	AGG	GCT	GAT	STT	TGT	GGC	ccc	CCA	- CCT	GĀG	GCT	GAG	• • •
	N	A	v	v	ĸ	т	S	Y	0	R	E	Y	L	s	G	s	E	v	T	Ŷ	596
1801	AATG		бтт	GTT	AAG	ACT	TCC	TAT	CAG	AGGO	GAA	PAC	TTA	TCT	GC	TCT.	GĂA	GTA	ACT	TAC	
	L	с	R	D	к	Ŷ	1	P	L	E	G	v	D	т	ī	R	c	R	N	G	616
1861	CTGT	GC	CGT	GAT	AAG	TAT	ATT	CCA	TTG	GĀGO	GGA	TG	GAC	ACG.	ATT	CGA	тĞТ	AGA	AAC	GGA	
	0	w	n	ĸ	E	т	ĸ	c	т	s	s	c	D	ĸ	τ.	v	- D	v	т	M	636
1921	CAAT	ŝ	GAC	AAG	GÃA	ATT	AAG	TGC	ACA	TCA	TCC	ГGТ(sac.	A A G	CTC4	с т а	сăт		ACC	ATC	
	D	F	т	A	D	ĸ	E	1	v	T	E	G	0	т	7	•	v	0	°.	L	656
1981	GATT	τc	аст	 ຄົງຄ	GÃO		GÃG	ATA	TAT	атт(5 Å Å	ວັດເ	- 44	ACC.	ата.		тат	~×.	тст		000
	T	n	G.	A	E	G	T	Δ	т	c	N	N	т	ĸ	ພ	v	к	5	P	0	676
2041	ATAG	ATO	сст	бст	د آم م م	າລົລ	ATT	672	ACG	тот	AC	AAT I		220	TCC	ТТ	• • •	тсь	റ്റം	. C	0.0
		r	v.	ĸ	D	с С		1	D	Δ.	n	T 1	D	M		~	v		7	T	606
21.01	TOTA		ста	220	~~»	TOT	~~~	CTT	~~~~	с СТТ	CAT	тт/	r.		сст.	~~~	т. т. т. т.	~~~		ATC	030
	E	с. С	517	- E	1		E	6	T	~	о л (1	~	~~~~		~	слл. м		040	~ <u>`</u> `	~	716
2161	0,000	G	сл. х	ດ້ຳດ	стт	т.т.т.	г ттт	~~~	×~×	***	ATC:		1	• • • •	т.Ст.	, N		~~~	1 737	CAN.	110
2101	0400		0 n.n	v	- -	2110	·	507		~~~	MIC/ M		IACI	511	101	AAC C	GAA	000	UM1	CAA D	776
2221	ATCA	÷.,	хот	220	с ^ь	~~~	• • • • •		ACT	тст	п лтс/	-70	- 20	-C A	TCC		19 2 2 T		~~~~	P CCN	130
2221	7107	~	- C	- nno	UAU T	occ c	ACC C	AGG.	ACI	1011	- D - D	-100	JAC	GGA	100	HGC	~~1	CAC	GIG	CCA	766
2201	3037	с. т.	- 2 2		ь ттс	200		C	r CCT	r CCN	r		- 2 7/	- 		~~~~		~~ .		CCT	/50
2201	ACAI	510	544	N N		AGC	160	GAA				4CAI	JAI	GGA	500	211	ACA	GIA		GGI	776
		r Ch		N	6	N	r cc.		L	P	0	ĸ	t .	-	D - D	. F	.5		_D	G	//6
2341	TIAC	CAI	UAI D	AAI	GGC	AAI	CCA	ATA	cru	LC I	JAC	JGC:	rru	CIU	GAC	TTC	AGC	TGT	GAC	GGT	
	5	6	. к	т - т	С.	N	6		A	ĸ	ь - т	1		6	. К	.U.,	6	2	_w	U	/96
2401	TCTG	66,	AGA	TAT	CIG	AAC	GGA	AGT	GCA	CGT	CTG/	ATT:	rGre	GGA	AAA	GAC	GGA	CAG	TGG	GAT	
	к	P	F	P	s	с	Е	D	I	Т	С	Е	I	G	v	М	₽	T	H	v	816
2461	3330	~ > >	-	COT		-	~ * *	~ ~ ~	ATC					~~~~					~	-	
2401	nnnc	CA.	111	CUI	101	160	GAA	GAC	AIC	MCI.	reit	JAAJ	MI 19		910	AIG	ccc	ACT	LAI	GIU	
2401	S	v	v	G	τ.	P	D D	GAC	m	K K	rGit T	JAA	~ ~	ں وی م	616	AIG	000 v	ACT		GIC T	074
2401	S	V	V	G	L	P	P	G	T	K	T	I	K	A	G	Q	ĸ	L	Q	L	836
2521	S AGTG	V	V	GGA				GGA	T ACA	K		I ATCI	K	A	G		K AAA	ACT L CTA		L	836
2521	S AGTG Q		V STT D	GGA N			P CCA P	GGA V	T ACAL	K AAG/ R		I ATC/ A	K AAGO	A GCT(I	G GGA E		K AAA L				836 856
2521 2581	S AGTG Q CAAT	V TAC C GTC	V STT D GAC	GGA N AAT				GGA GGA V GTG	T ACAJ D GAC	K AAG	T T ACAJ P CCTO	I ATC/ A SCAC	K AAGO E GAAJ	A GCT I ATT	G G G G G A A	Q CAA C TGT	K AAA L TTA			L CTT G GGG	836 856
2521 2581	S AGTG Q CAAT E	V TAC C GTC W	V STT D GAC N	GGA N AAT A	L CTT A GCA	P CCA Q CAG		GGA GGA GTGI A	T ACA D GAC	K AAG R AGGG	T ACAJ P CCTC G	I ATC/ A BCAC D	K AAGO E GAAJ T	A GCT I ATT	G GGA E GAA T	Q CAA C IGT L	K AAA L TTA T	L CTA Q CAA A		L CTT G GGG P	836 856 876
2521 2581 2641	S AGTG Q CAAT E GAGT	V TAC GTC W GGJ	V GTT D GAC N AAT	GGA N AAT ACA	CTT A GCA A GCC	P CCA Q CAG F TTT		GGA GGA GTGI A GCT	T ACA D GAC	K AAG R AGG G GGT	T T ACAJ P CCT(G GGA(I ATC/ A SCAC D SAC/	K AAGO E GAAJ T ACA	A GCT(I ATT(C IGC)	G GGA E GAA T ACC	Q CAA C IGT L ITG	K AAA L TTA T ACA	L CTA Q CAA CAA GCT		CTT GGGG P CCA	836 856 876
2521 2581 2641	S AGTG Q CAAT E GAGT D	V TAC GTC W GGJ	STT D GAC N AAT	GGA N AAT A GCA	CTT A GCA A GCC M	P CCA Q CAG F TTT T		GGA GGA V GTG A GCT R	T ACA D GAC C TGC A	AAGA R AGGG G GGT(A	T T ACAJ P CCTC G GGAC L	I ATC/ A BCAC D GAC/ K	K AAGO E GAAJ T ACA' N	A GCTO I ATTO C TGCO Q	GGA GGA E GAA T ACC	Q CAA C IGT L ITG R	K AAA L TTA T ACA K	L CTA Q CAA CAA GCT G	CAG CAG T ACT V GTG Q	L GGG GGG P CCA K	836 856 876 896
2521 2581 2641 2701	S AGTG Q CAAT E GAGT D GACA	V TAC GTC W GGJ N ATC	V GAC N AAT V GTA	GGA SGA N AAT A GCA R CGC	L CTT A GCA A GCC M ATG	P CCA Q CAG F TTT T ACT	P CCA P CCA P CCA P CCA	GGA GGA GTG GTG A GCT R CGT	ACA D GACA C TGCC A GCTC	AAGA R AGGG G GGT(A GCA(T ACAJ P CCTC G G GGAC L CTAJ	I ATC/ A GCAC D GAC/ K AAG/	K AAGO E GAAJ T ACA' N AACO	A GCT(I ATT) C TGC, Q CAA	G GGAI E GAA T ACC M ATG	Q CAA C IGT L ITG R AGA	K AAA L TTA T ACA K AAA	L CTA Q CAA GCT G GGA		L GGGG P CCA K AAG	836 856 876 896
2521 2581 2641 2701	S AGTG Q CAAT E GAGT D GACA L	V TAC C GTC GG GG N ATC K	STT D GAC N AAT V GTA	GGA AAT A GCA R CGC	CTT A GCA GCC M ATG C	P CCA Q CAG F TTT T ACT T	P CCA P CCA P CCA P CCA	GGA GGA V GTG GTG A GCT R CGT R	TGCI ACAJ D GACJ C TGCI A GCTI R	K AAG R AGG G GGT A GCA D	T ACAJ P CCTC G GGAC L CTAJ T	I ATC/ A GCAC D GAC/ K AAG/ L	K AAGO E GAAJ T ACA N AACO L	A GCTO I ATTO C TGC Q CAAL G	G GGA E GAA T ACC M ATG K	Q CAA C TGT L TTG R AGA A	K AAA L TTA T ACA K AAA E	L CTA Q CAA GCT GGA V	Q CAG T ACT V GTG Q CAG E	L CTT GGGG P CCA K AAG C	836 856 876 896 916
2521 2581 2641 2701 2761	S AGTG Q CAAT E GAGT D GACA L TTGA	V TAC GTC W GGJ N ATC K AA	STT D GAC N AAT V GTA F TTT	GGA GGA AAT A GCA R CGC A GCT	L CTT A GCA A GCC M ATG C TGC	P CCA Q CAG F TTT T ACT T ACT	CCA P CCA P CCA P CCA D CCA	GGA GGA GTG GTG GCT R CGT R CGT	TGCI GACA GACA GACA C TGCI A GCTI R CGAI	K AAG R AGG G G G G G C A G C A G C A G A C A	T ACAJ P CCTO G GGAO L CTAJ T ACTO	I ATC/ A GCAC D GAC/ K AAG/ L CTTC	K AAGO E GAAJ T ACA N AACO L CTAO	A GCTI I ATTI C IGC. Q CAAJ G GGGJ	G GGA E GAA T ACC M ATG K AAGO	Q CAA C TGT L TTG R AGA A GCA	K AAA L TTA T ACA K AAA E GAG	L CTA Q CAA GCT G GGA V GTT	CAG T ACT GTG Q CAG E GAA	L CTT G GGG P CCA K AAG C TGT	836 856 876 896 916
2521 2581 2641 2701 2761	S AGTG Q CAAT E GAGT D GACA L TTGA L	V TAC C GTC W GGG N ATC K AAC D	STT D GAC N AAT V GTA F TTT N	GGA NAAT AAT A GCA R CGC A GCT G	L CTT A GCA GCC M ATG C TGC	P CCA Q CAG F TTT T ACT T ACT	P CCA P CCA P CCA P CCA D CCA	GGA GGA GTG GTG A GCT R CGT R CGT Y	TGCI GACA GACA GACA C TGCI A GCTI R CGAI P	K AAGG G GGT GGT A GCA D GCA F	T ACAJ P CCTC G GGAC L CTAJ T ACTC P	I ATC/ A SCAC D SAC/ K AAG/ L TTC T	K AAGO E GAAJ T ACA' N AACO L CTAO C	A GCTI I ATTI C IGC Q CAAL G G G G G	G GGAI E GAA T ACC ⁴ M ATGJ K AAGO D	Q CAA C TGT L TTG R AGA A GCA P	K AAA L TTA T ACA K AAA E GAG L	L CTA Q CAA GCT GGA GGA V GTT D	CAG CAG T ACT CAG CAG CAG CAG	L GGG GGG P CCA K AAG C TGT G	836 856 876 896 916 936
2521 2581 2641 2701 2761 2821	S AGTG Q CAAT E GAGT D GACA L TTGA L CTAG	V TAC C GTC W GG N ATC K AA C AC	V GTT D GAC N AAT V GTA F TTT N AAC	GGA AAT A GCA CGC A GCT GGA	L CTT A GCA A GCC M ATG C TGC K AAG	P CCA Q CAG F TTT ACT T ACT T C ACC W TGG	P CCA P CCA P CCA P CCA D CCA S AGT	GGA GGA GTG GTG GTG CGT R CGT Y TAT	T ACA D GAC C TGC A GCT C R C GA C CA	K AAGG AGG G GGT GGT A GGT G A GGT G A G G G C A G C A G C A G C A G C A G G C A G G C A G C A G C A G C C C C	T ACAJ P CCTC G GGAC L CTAJ T ACTC P CCAJ	I ATC/ A SCAC D SAC/ K AAG/ L CTTC T ACA1	K AAGO E BAAJ T ACA' N AACO L C C C IGTO	A GCTI ATTI C IGC. Q CAAL G GGGJ G GGA0	G G G G G G A C C M A C C M A C C M A C C M A C C M A C C M A C C M A C C M A C C M A C C M A C C M A C C C M A C C M A C C C M A C C C C	Q CAA C TGT L TTG R AGA A GCA P CCT	K AAA L TTA T ACA K AAA E GAG L TTG	L CTA Q CAA GCT GGA GGA V GTT D GAT	Q CAG T ACT V GTG Q CAG GAA C TGT	L GGG GGG P CCA K AAG C TGT GGA	836 856 876 896 916 936
2521 2581 2641 2701 2761 2821	S AGTG Q CAAT E GAGT D GACA L TTGA L CTAG N	V TAC GTC GG W GG N ATC K AA C D AC P	V GTT D GAC N AAT V GTA F TTT N AAC P	GGA GGA AAT AAT GCA R CGC A GCT GGA S	L CTT A GCA A GCC M AATG C TGC K AAG L	P CCA Q CAG F TTT T ACT T ACC W TGG	P CCA P CCA P CCA P CCA D CCA D CCA D CCA D CCA	GGA GGA GTG GTG GCT R CGT R CGT Y TAT G	ACA D GACA C TGCO A GCTO R CGAO P CCA	AAGG R AGGG GGT GGT A GCA GCA GCA GCA C C C C C C C C C C C	T ACAJ P CCTC G GGAC L CTAJ T ACTC P CCAJ T	I ATC/ A GCAC D GAC/ K AAG/ L CTTC T ACA1 T	K AAGG E GAAJ T AACA N AACC L C TAC C C R	A GCTI I C TGC. Q CAA. G GGG G GGGA R	G GGAN E GAA T ACC M AAGO D GATO Q	Q CAA C TGT L TTG R AGA A GCA C C C T S	K AAAA L TTA T ACA K AAAA E GAG L TTG V	ACT L CTA Q CAA GCT GGA V GTT D GAT Y	Q CAG T ACT V GTG Q CAG E GAA C TGT R	L GGG GGG P CCA K AAG C TGT GGA H	836 856 876 896 916 936 956
2521 2581 2641 2701 2761 2821 2881	S AGTG Q CAAT E GAGT D GACA L TTGA L CTAG N AATC	V TAC GTC GTC W GGG N ATC K AA C AC P CAC	STT D GAC N AAT V GTA F TTT N AAC P CCA	GGA NAAT A AAT GCA RCGC A GCT GGA S GGA S AGC	L CTT A GCA A GCC M ATG C TGC K AAG L CTT	P CCA Q CAG F TTT T ACT T ACC W TGG GAT	P CCA P CCA P CCA P CCA P CCA D CCA CCA P CCA CCA CCA CCA CCA CCA CCA C	GGA GGA GGTG GGTG R GGTG R CGTG R CGTG Y TATG GGAG	T ACAL D GACAL C GACA C GATO C GATO GATO	K AAGG AAGG G GGT A GGT A GGT D GAC F F TTC V GTT/	T ACAJ P CCTC G GGAC L CTAJ T ACTC P CCAJ T ACGJ	I ATC/ A SCAC D SAC/ K AAG/ L CTTC T T AACA1 T	K AAGG E BAAA T AACA N AACC C C TAG C R AGAG	A GCT(I ATT(C CAAJ G GGGG G GGGG R CGT(G GGAA E GAA T ACC' M ATGJ K AAGO D GAAJ	Q CAA C TGT L TTG R AGA A GCA P CCT S AGT	KAAA L TTA TAACA KAAA E GAG L TTG V GTA	ACT L CTA Q CAA A GCT GGA C GGA V GTT D GAT Y TAC	Q CAG T ACT V GTG Q CAG GAA C TGT R AGA	L GGG GGG P CCA K AAG C TGT GGA H CAT	836 856 876 896 916 936 956
2521 2581 2641 2701 2761 2821 2881	S AGTG Q CAAT E GAGT D GACA L TTGA L CTAG N AATC G	V TAC GTC GTC W GGJ N ATC K A A C A C A	STT D GAC N AAT V GTA F TTT N AAC P CCA S	GGA NAAT AAT A GCA R CGC A GCT GGA S AGC V	L CTT A GCA A GCC M A C TGC K A A G C T G C T C C T C C T C C T C C T C C T C C T A C A C	P CCA Q CAG F TTT T ACT T ACT T ACT T G G AT Y	CCA P CCA P CCA P CCA P CCA D CCA D CCA T S AGT S AGT R	GGA GGA V GTG A GGT R CGT R CGT Y TAT G GGA C	T ACAL D GACAL C TGCC TGCC A GCTC A GCTC C CA' D GATC Q	K AAGG AAGG G GGT A GGCA D GACA F TTCC V STT/ A	T T ACAJ P CCTC G GGAC L CTAJ T ACCJ L	I ATC/ A SCAC D SAC/ CTTC T A CTTC T A CC/ Y	K AAGG E BAAA T AACA N AACA C C C T AGAG T	A GCT(I ATT(C IGC) Q CAAL G GGG(G GGG(R CGT(M	G GGAA E GAA T ACC ^{**} M ATGJ K AAGO D GAA C AAG E	Q CAA C TGT L TTG R AGA SCA C C C T S AGT G	K AAA L TTA T ACA K AAAA E GAG L TTG V GTA G	ACT L CTA Q CAA A GCT GGA V GTT D GAT Y TAC P	CAI Q CAG T ACT V GTG Q CAG C AGA C TGT R AGA R	L CTT GGG P CCA K AAG C TGT GGA H CAT K	836 856 876 916 936 956 976
2521 2581 2641 2701 2761 2821 2881 2941	S AGTG Q CAAT E GAGT D GACA L TTGA L CTAG N AATC G GGAG	V TAC GTC GTC W GGJ N ATC K AA D CAC A C C C	STT D GAC N AAT STA F TTT N AAC P CCA S ICG	GGA AAT A GCA R GCA R CGC A GCA GGA S AGC V GTG	L CTT A GCA A GCC M A C T GCA L CTT E GAA	P CCA Q CAG F TTT T ACT T ACT T ACC W TGG D GAT Y TAC	CCA P CCA P CCA P CCA CCA D GAT S AGT R AGG	GGA GGA V GTG A GCT R CGT R CGT Y TAT G GGA C C TGT	ATC. ACA D GAC. C TGCI A GCTI A CCA C GATI Q CAG	K AAG/ R AGGC G GGTC A GCAC D GAC/ F ITCC V GTT// A GCTT	T ACAA P CCTC G GGAC L CTAA T ACCA T T CCAA T CCAA T L ITGT	I ATC/ A GGAC/ D GGAC/ K AAG/ L CTTC T T ACC/ Y I ACC/ Y	K AAGA E BAAA T AAACA N AAACA C T C TAACA C R AGA C T ACTA	A GCTI I ATTI C IGCI Q CAAJ G GGG G GGG R CGTI M ATGI	G G G G G G A T T A A C C A A C A A C C A A C C A A C C A A C C A A C C A A C C C A C	Q CAA C TGT L TTG R AGA C T C C T S G G G G G G G G	K AAA L TTA T ACA K AAAA E GAG U TTG GTA GGG	L CTA Q CAA GCT GGA GGA V GTT D GAT TAC P CCT	CAI QCAG TACT VGTG QCAG GAA CAG CAG CAG CAG CAG CAG CAG CAG	L GGG GGG P CCA K AAG C TGT GGA H CAT K AAA	836 856 876 916 936 956 976
2521 2581 2641 2701 2761 2821 2881 2941	S AGTG Q CAAT E GAGT D GACA L TTGA L CTAG N AATC G GGAG T	V TAC GTC GTC GGJ N ATC K AA C C C C	V GTT D GAC N AAT V GTA F TTT N AAC P CCA S ICG D	GGA NAAT A GCA R GCA R GCA GCA S GGA S AGC V GTG N	L CTT A GCA GCC M ATG C TGC K AAG C TTC E GAA G	P CCA Q CAG F TTT T ACT T ACT T ACC W TGG D GAT Y TAC E	CCA P CCA P CCA P CCA P CCA CCA CCA CCA	GGA GGA V GTG A GCT R CGT R CGT Y TAT G GGA C T TT T	ATCA ACA D GACA TGCC A GGAC R CGAC Q CCA G GATC G	AAGG R AAGGG G GGT(A GGT(A GGT(V SACI SACI SACI SACI SACI SACI SACI SACI	T T ACAJ P CCT(G GGAG CTAJ T ACT(P CCAJ T ACGJ L ITG7 I	I ATC/ A GCAC D GAC/ K AAG/ L CTTC T T ACC/ T ACC/ R	K AAGCA E BAAJ T ACA N AACCA C C TAC C T AGAC T C	A GCTI I ATTI C IGCI Q CAAJ G GGGG G GGGG R CGTI M ATGC I	GGAN EGAA T ACC ^M AATGJ SATG CAAJ E SAAG K	Q CAA C IGT L ITG. R AGA C C C C S G T G G T C G T C G T C G T C C C C C C	K AAAA L TTA T ACA K AAAA E GAG U TTG GTA G GGG C	ACT L CTA Q CAA GCT GGA V GTT D GAT Y TAC P CCT T	CAI Q CAG T ACT V GTG Q CAG CAG C GAA C TGT R AGA R CGC V	L CTT GGG P CCA K AAG C T G G A H C AT K AAA D	836 856 876 916 936 956 976 996
2521 2581 2641 2701 2761 2821 2881 2941 3001	S AGTG Q CAAT E GAGT D GACA L TTGA L CTAG GGAG GGAG T ACCT	V TAC GTC GTC GTC GTC GTC N GTC N A TC A C C C C C C C C C C C C C C C C	V GTT D GAC N AAT V GTA F TTN N AAC P CCA S ICG D GAC	GGA NAAT A AAT GCA RCGC A GCT GGA S GCT GGA S GTG NAAT	L CTT A GCA A GCC A A C T GCA A C T C C T C C T C C T C C T C C T C C T C C T C C T C C T C C C T C	P CCAG CAG F TTT T ACT T ACC W TGG GAT TAC E GAA	CCA P CCA P CCA P CCA P CCA D GAT R GAT R GAT R GAT R GAT R GAT	GGA GGA V GTG A GCT R CGT R CGT Y TAT G GGA C T T T T T T T T T T T T	ATCA ACA D GACA C TGCC A GGAC R CGAC C GGAT GGAT	K AAGG R AAGG G G GGT A GGT A GGT V S GTT I S C G S C T S S C T S S C T S S C T S S C T S S C T S S C T S S C T S S S T C S S S T C S S S T C S S S S	T ACAJ P CCTC G GGAC T CCAJ T CCAJ T CCAJ T T CCAJ T T CCAJ T T CCAJ T T CCAJ T T CCAJ T T CCAJ T T CCAJ T T C T C T C T C T C T C T C T C T C	I ATC/ A BGCAC D BAC/ C T T C T T C C T T C C T T T C C T T T C C T T T C C T T T C C A C A	K AAGK E BAAJ T ACA' N AACA' L CTAC C IGTC R AGAC T C C IGC/	A GCTI I GCTI C CAAL G GGGG G GGGG G GGGG R CGTI M TGC I ATC/	G G G G G G A T T A A C C A A C C A A C C A A C C A A C C A A C C A A C C C A A C C C A C C C A C C C C A C	A I G CAA C TGT L TTG AGA C T S G T C S G T C C C	K AAAA L TTA T ACA K AAAA G G G G G G G G G G G G G G G G	ACT L CTA Q CAA A GCT GGA C GGA V D GTT TAC C C T TAC T ACT	CAG CAG TACT GTG QCAG GAA C TGT R AGA R CGC V GTT	L CTT GGG P CCA K AAG C T G GGA H CAT K AAA D GAT	836 856 876 916 936 956 976 996
2521 2581 2641 2701 2761 2821 2881 2941 3001	S AGTG Q CAAT E GAGT D GACA L TTGA L CTAG N AATC GGAG T ACCT T	V TAC GTC GTC W GGJ N GGJ N ATC K A A C C C C C C C C C C C C C C C C	GTT DGAC NAAT GTA GTA GTA V GTA V GTA V GTA V GTA V GTA V GTA V GTA V GTA V GTT D GAC N V GTT D GAC N STT D GAC STT D GAC STT STT STT STT STT STT STT STT STT ST	GGA AAT A AAT A CA CGC A CGC A CGC GGA S CC GGA S AGC V GTG N AAT M	L CTT A GCA A GCC A A GCC A A GCC A A GCC A A GCC A A GCC A A GCC A A GCC A A GCC A A GCC A A GCC A A GCC A A GCC A A GCC A A GCC A A GCC A A GCC A A A GCC A A A A	P CCA Q CAG F TTT T ACT T ACC TGG GAT TAC E GAA R	CCA P CCA P CCA P CCA P CCA P CCA D GAT R GAT R GAT R GAT R GAT R GAT R GAT	GGA GGA V GTG GGTG R CGT Y TAT GGA C C T T T T T T T T T T T T T T T T T	ATCA ACA D GACA C TGCC A GGAC R CGAC C GGAC C AGA GGA I	AAGG R AAGGG G GGTC A GGTC D GACA D GACA D GACA TTCC V S GTTJ A S GCTT S S CCGJ E	T ACAJ P CCTCC G GGAC L CTAJ T ACTO P CCAJ T ACGJ I T T G T G T G T G T G T G T G T G T G	I ATC/ A GCAC D GAC/ C T C T T C C T T C C T T C C T T C C T T C C T T C C T T C C A C A	K K AAGA E GAAJ T N ACA C I GTA C C I GTA C R AGA C T C I GCJ Y	A GCTI I GCTI C CAAJ G GGGG G GGGG G GGGG M CGTI M TGC I S	G G G G G G G A T C A A C C A A C C A A C C A A C C A A C C A A C C A A C C C A A C C C A C C C A C C C A C C C C A C	A I G CAA C TGT L ITG. AGA A G CCT S GGT P CCC E	KAAA LTTA TTA ACA KAAAA E GAG GTA GGG C C TGC A	ACT L CTA Q CAA A GCT GGA CAA A GTT GGA T ACT T ACT K	CAG CAG TACT GTG CAG GAA CGC CAG CAG CAG CAG CAG CAG CA	L GGG P CCA CCA C CCA C CCA C CCA C CCA C C CCA C C CCA C C CCA C C CCA C	836 856 876 916 936 956 976 996
2521 2581 2641 2701 2761 2821 2881 2941 3001 3061	S AGTG Q CAATT E GAGT D GAGT L CTAG G GGAG G GGAG T ACCT T ACAG	CA TAC CGTC GGTC W GGJ AC C CA CA C CA CA CA CA CA CA CA CA CA	TT GAC SAC NAAT GAC SAC SAC CCA SAC CCA SAC ICG SAC ICG SAC	GGA AAT AAT GCA AGC AGC GGA SCC AGC GGA SCC AGC AAT AAT	CTT A GCA GCA A GCC M ATG TGC K AAG GGT R AGG R AGG	P CCA CAG F TTT T ACT T ACT T ACT T ACT T ACT C ACA R AGA	CCA P CCA P CCA P CCA D GAT S GAT R GAT R GAT R GAT R GAT R GAT CAC	GGA GGA V GTG GGTG R CGTC R CGTC T TATC GGA C C T GGA C C T T T T T T T T T T T T T T T T T	ATCA ACAJ GACA GACI TGCI A GGATI CCGAU CCA GGATI CAGG GGATI I ATTI	AAG AAG R AAG G G G G G G G G G C A G G T C C S T C C S T C C S C A C C C C C C C C C C C C C C C	T ACAJ P CCTC G GGAC L T T ACTC P CCAJ T T CCAJ I T G CCAJ T T I T CCAJ T T I T CCAJ T T T T T T T T T T T T T T T T T T T	I ATC/ A GCAC D GAC/ CTCC T T CTTCC T T A CC/ T T A CC/ T R A GAT	K K AAGO E SAAJ T ACA' N AACO' L CTAO C CTAO C R AGAO T C C IGC/ Y TATT	A GCT(I ATT(C TGC, Q CAAJ G G G G G G G G G G G G G G G G G G	G G G G G G A A C C A A C C A A C A A A C C A A A C C A A A C C A A C C A A C C C A A C C C A C C C A C	A I G CAA C C T L T G A G A C C T C T C C C C C C C C C C C C C C C	KAAA LTTA TTA ACA KAAA E GAG GGG C C C C C C C C C C C C C C C C	ACT L CTA Q CAA A GCT GGA CAA A GCT GGA V TAC T TAC T ACT K AAG	CAI CAG T ACT G T G CAG C G C A G A C A C A C A C A C A C	L GGG P GGG P CCA K AAG C GGA H CAT K AAA D GAT Y TAT	836 856 876 916 936 956 976 996
2521 2581 2641 2701 2761 2821 2881 2941 3001 3061	SAGTG Q CAAT E GAGT D GACA L CTAG GGAG T TTGA AATC G GGAG T T ACCT T ACCAG A	V GTAC GTC GTC GGC N GGC N ATC A C C A C C C C C C C C C C C C C C	STT GAC NAAT GTA GTA FTT NAAC STA STA STA SAC SAC H	GGA AAT AAT GCA AGC AGC GGA AGC GGA AAT AAT AAT A	CTT A GCA GCA A GCC M ATG TGC K AAG GGT R GGA GGT R AGG D	P CCA CAG F TTT T ACT T ACT T ACT T ACT T ACT C ACA R ACA R ACA V	CCA P CCA P CCA P CCA C C CCA C CCA C CCA C CCA C CCA C CCA C CCA C CCA C CCA C CCA C CCA C C CCA C C CCA C	GGA GGA V GTGI A GGTG R CGTI R CGTI R CGTI GGA C C TTTT G GGA C T TTT T A A TT T A A TT T	ATCA ACAJ GACA GACI TGCI A GGACI R CGAC C GGAC CAGA GGAC I F	AAG/ R AAGG G GGT GGT GGT C SAC/ S GTT/ A S C C S A S C C S A S C C S A S C A S C A C A	T ACAJ P CCT(C G GGAC L T ACT(C P CCAJ T ACT(C P CCAJ I T T ACC(J F I T C ACC(J C C C C C C C C C C C C C C C C C C	I ATC/ A GCAC D GAC/ CTCC T T CTTCC T T A CCA T A CCA T A CCA T R A GA T	K K AAGO E BAAJ T ACA' N AACO' L CTAO C CTAO C R AGAO T C IGC/ Y TATI S	A GCT(I ATT(C ATT(Q CAAJ G G G G G G G G G G G G G G G G G G	G G G G G G A A C C A A G C A A C A A C A C	A I G CAA C C T L T G A G A C C T C T C C C C C C C C C C C C C C C	KAAA LTTA TTAAT ACA KAAAA EGG GGG CCA GGG CCA GCG V	ACT L CTA GCT GCA A GCT GGA V D GAT TAC T CCT T ACT K AAG G	CAI QCAG TACT GTG QCAG GAA CGC CAG CAG CAG CAG CAG CAC CAC	L GGGG GGGG P CCA K GGGA H CAT K AAA C AT GAT Y I TAT Y	836 856 876 916 936 956 976 996 016
2521 2581 2641 2701 2761 2881 2881 2941 3001 3061 3121	SAGTG AGTG Q CAAT E GAGT D GACA L CTAG GGAG T CTAG GGAG T AATCC GGAG C A ACAG A CCAA	V TAC GTC GTC W GGJ N GGJ N ATC A C C C C C C C C C C C C C C C C C	STT GAC NAAT STA STA STA STA STA STA STA SAC SAC SAC SAC SAC SAC	GGA NAAT AAAT CGCA CGCA CGCA CGCA CGCA CGCA C	L CTT GCA GCC M GCC M GCC M GCC M GCC M CTT E A G G G C T C C T C C T C C M G C M C C M C C M C C M C M C M C	P CCA Q CAG F TTT T ACT T ACT T ACT T G GAT R AGA V G TA	CCA PCCA PCCA CCA CCA CCA CCA CCA CCA CC	GGA GGA V GTGI A GGTG R CGTC R CGTC R CGTC T T T T GGGA C T GGA C T T T T T T T T T T T T T T T T T T	ATC ACAJ GACAJ C GACJ C GGAC C GGAC C GGAC C GGAC C GGAC C GGAC C GGAC C GGAC C C GGAC C C GGAC C C C	AAG R AAGG G GGT A GGT A GGT A GGT A GGT A GGT A GGT A GGT A AGG C	T ACAJ P CCTC G G GGAC L T T ACCJ T T T CCAJ T T T C T C T C T C T C T C T C T C T	I ATC/ A GCAC D GAC/ C T C T T C T T C C T T C C T T C C T T C C T T C C T T C C T T C C T T C C T T C	K AAGO E BAAJ T AACO L CTAO C IGTO R AGAO T C IGCJ Y S AGTO	A GCTI I ATTI C IGCI Q CAAL G G G G G G G G G G G G G G G G G G	G G G G G G G A A G G A A G G A A G A C A A A G A G	A I G CAA C C T G T C T C T C T C T C T C C C C C C C C	KAAA TTA TTA ACA KAAA EGG GGG GGG C C A GCG V GTC	ACT L CTA GCT GCA GCT GCT GCT CCT TAC CCT K AAG GCC GCC	CAI QCAG TACT GTG QCAG GAA CAG CAG CAG CAG CAG CAG CAG CAG	L GGG GGG P CCA K GGA H CAT K AAA C GGA H CAT K AAA D GAT TAT V GGG	836 856 876 916 936 956 956 996 016
2521 2521 2581 2641 2701 2701 2821 2821 2881 2941 3001 3061 3121	SAGTG Q Q GAGT D GACAA L CTAG G GACAA L CTAG G GAGG A AATCC G GAGG A CCAA T ACAG G GAGA A Q CAAT	CA CGA GGI GGI GGI NAT CA CA CA CA CA CA CA CA CA CA CA CA CA	STT GAC GAC GAC GAC GAC GAC GAC GAC CAT R	GGA NAAT AAAT CGCA CGCA CGCA CGCA CGCA CGCA C	LCTT GCA GCA GCC MAG CTC KAAG CTT EAAG GGT R AGG GAT K	CCA Q CAG FTTT T ACT T ACT T ACT T G GAT R AGA V G TA C C	CCA P CCA P CCA P CCA P CCA P CCA P CCA C C CCA C	GGA GGA V GTG GGA CGT R CGT GGA C T T T T T T T T T T T T T T T T T T	ATC. ACAJ D GACJ C GACJ C G G G G G G G G G G G G G	AAG AAG AAG GGT GGT A GGT A GGT A GGT A GGT A GGT A GGT A GGT A C GAA A GGT V V A C C C C C C C C C C C C C C C C C	T ACAJ P CCTC G G GGAC L T T ACCJ T T C CAJ T T T C T C T C T T C T C T M	I ATC/ A GCAC D GAC/ C T C T T C T T C C T T C C T T C C T T C C T T C C T T C C T T C C T T C C T T C C A C C A C A	K K AAGO E BAAJ T AACO L CTAO C IGTO R AGAO T C IGCJ Y S AGTO F	A GCTI I ATTI C IGCI Q G G G G G G G G G G G G G G G G G G	G G G G G G G A T G G A A G G A C A A G A C A A A C A T C A T C A T C A T	A I G CAA C C I T G T C C T C C C C C C C C C C C C C C C C	KAAA L TTA T T ACA K AAA E GAG G GGG G GGG G C G G C G G C G C Q	ACT L CTA Q CAA GCT GGA V GGT GGA V CT TAC. P CT TAC. CT K AAG GGC •	CAI Q CAG T ACT GTG Q GTG CAG CAG CAG CAG CAG CAG CAG CAG CAG CA	L GGG GG P CCA K AAG C TGT G GGA H CAT K AAA D GAT TAT V GTG 1	836 856 876 916 936 956 976 996 016 036
2521 2521 2581 2641 2701 2701 2821 2881 2941 3001 3001 3121 3181	S AGTG Q CAAT E GACA L CTAG GGACA L CTAGG GGAG A AATC G GGAG A CCAGA	CA V TAC GTO GGI W GGJ NATO ACC CA CA CA CA CA CA CA CA CA CA CA CA	STT GAC GAC GAC GAC GAC GAC GAC GAC CAT CAT CGT	GGA AAT A GCA CGC A GGA S GGA S GGA CAG CAG	LCIT A GCA GCCA A GCC A A GCC A A G G C T C C T C C T C C T C C T C A A G C C T C A A C C T C A A C C A A C C A C A	CCA Q CAG F TTT T T ACT T ACT T G G AT C G AA C G TA C G TA C G TA C G TA C G TA C G AG C AG F T T T T T C CA G F T T T T T T T T T T T T T T T T T T	P CCA P CCA P CCA P CCA C CCA C CCA C CCA C C CCA C C C C	GGA GGA V GTG A GGT R CGT R CGT T GGA C T GGA C T GGA C T GGA C T GGA C T GGA C T GGA C T GGA C T GGA C T GGA C T T T T T T T T T T T T T T T T T T	TCA GACA D GACA C GGAC C C GGAC C C GGAC C C GGAC C C C	AAGG RAAGG GGC GGC A GGC A GGC V A GC SACI A GC SACI R C GC SACI R C GC SACI C SACI C SACI C SC SC SC SC SC SC SC SC SC SC SC SC S	T ACAJ P CCTC G G G G G G G G G C C A C T T T C C A C G J C C A C G J C C A C T T C C A C T J C C A C T A C T A C T C T C C C C C C C	I AATC/ AATC/ AAGCAC D BAC/ C T AAGC/ T T C C T T C C T T C C T T C C T T C C T T C C T T C C T T C C T T C C T T C C T T C	K AAGG E BAAJ T ACA' N ACA' C C TAC C C C C C C C C C C C C C C C C	A GCTI I ATTI C GCAAJ G G G G G G G G G G G G G G G G G G	G G G G G G G A A C C A A C A C A C A C	A I G CAA C C T L T R A A A A A A A A A A A A A A A A A A A	KAAA L TTA T TA KAAA E GAG G G G G G G G G G G C G C G	ACT L CTA Q CAA GCT GGA CTT GGA CTT CT TAC CTT K AAG GGC C TAA	CAI QCAG TACT GTG QCAG GAA CAG CAG CAG CAG CAG CAG CAG CAG	GIC GGG GGG CCA CCA CAT GGA CAT TAT VI GTG GGA D CAT TAT VI GTG GGA D CAT	836 856 876 916 936 956 976 976 976 1016
2521 2521 2581 2641 2701 2761 2821 2881 2941 3001 3061 3121 3181 3241	SAGTG Q QCAAT E GACA L CTAG GGACA L CTAG GGAG T T ACCCT T ACCCT T ACCAG A QCAGA	CA V TAC GTO GGI W GGI NATO ACC CA CA CA CA CA CA CA CA CA CA CA CA	GTT GAC NATV GTA STA STA STA STA STA STA STA STA STA S	GGA AAT A GCA CGC GGA GCT GGA S GCT GGA AAT A AAT GAT Q CAG Q CAG Q CAG	L CIT A CIT	CCA CAG TTT TT ACT TGG GAT TAC GAA RA C GAA C TGT TGT T T GAA	P CCA P CCA P CCA CCA C CCA C CCA C CCA C CAC C CAC C CAC C CAC C CAC C CCA C C CCA C CCA C CCA C	GGA GGA GGT R GGT R CGT C R CGT C T T T T T T T T T T T T T T T T T	TCA GACA GACA C GAC C GGA C GGA C GGA GGA	AAGG R AAGG G GGTC A GGTC A GGTC A GGTC A GGTC A C GTC A C C C C C C C C C C C C C C C C C C	T ACAJ P CCTC G G G G G CTAJ T T C CCAJ T T C CCAJ T T C CCAJ T T C CCAJ T T C CCAJ T T C CCAJ T T C CCAJ T T C CCAJ C T A C T C T C T C T C T C T C T C T C	I AATC/ AATC/ AATC/ AGAC/ BGC/ C T C C T C C C T C C C C C C C C C C	K AAGG E BAAJ T ACA' N AAGC C C C C C C C C C C C C C C C C C	A GCTI I AATTI C CAAJ G G G G G G G G G G G G G G G G G G	G G G G G G G A T G G A A C C A C A C A C A C A C A C A C	A I G CAA C C T L TTG. AGA A GA C S T C S AGT C S AGT C S AA C C T C S AA C C T C T C C T C C T C C C T C C C C C	KAAA LTTA TAACA KAAA EGAG LTG GGG C GGG C GGG C GGG C GGC C A GCC C A GCC C A GCC C A GCC C A GCC C A GCC C A GCC C C C	ACT L CTA Q CAA A GCT GGA CTA C GGA TAA C C T A C C T A C C T A C C A A C G G C T A C A A C C A A C C A A C C A A C C A A C C A A C C C A A C C C A A C C C A C C C A C C C C A C C C C A C	CAI QCAG TACT GTG QCAG GAA CCAG CAG CAG CAG CAG CAG CAG CAG	GIC GGG GGG CCA CCA CTGT GGA CAT CAT CAT CAT CAT CAT CAT CAT CAT CA	836 856 876 916 936 956 976 976 976 1036
2521 2521 2581 2641 2761 2761 2881 2881 2941 3001 3061 3121 3181 3241 3301	S AGTG Q CAAT E GAGT D GACA L CTAG G GGAG T TTGA A ACCT T CAGG A GCAA Q CAGA A taaa	V TAC GTA GGJ N ATC KAA D AC C AC C C C C C C C C C C C C C C	GTT GAC STA	GGA AAT A A GGA GGA GGA GGA AAT A A GGA GGA CAG CAG CAG CAG CAG CAG CAG	L CTT A GCA A CTT A A C C T C C T C C T C C T C C T C C T C C T C C T C C T C C T C C T C C T C C A C C M G C A C C M G C A C C M G C A C C M G C A C C M G C A C C M G C C A C C M C C M C C C M C C C M C C C M C C C M C C C M C C C M C C C M C C C M C C C M C C C M C C C M C C C M C C C M C C C M C C C C M C C C M C C C C M C C C C M C	CCA CAG TTT TT ACT TGG GAT TAC GAA RA C GAA C TGT tga cag	CCA PCCA PCCA CCA CCA CCA CCA CCA CCA CC	GGA GGA GGTG GGTG CGTC R CGTC CGTC CGTC CGTC CG	T ACA D GACA C C C C C C C C C C C C C	K AAGG G GGT(A GGT(A GGT(A GGT) A GGT S A GGT R A GGT R A GGT R A GGT R A GGT C G A G G G G G G G G G G G G G G G G G	T ACAJ P CCTC G G G G CTAJ T T C CAJ T T C CAJ T T C C CAJ T T C C CAJ T T C C C T T C C C T C C T C T C T C	I ATC/ A A GCAC D GACJ/ K A GACJ/ T C T T C C T T C C T T C C T T C C T T C C T T C C T T C C T T C C T T C C T T C	K K AAGC E BAAT T NACC C C C C C C C C C C C C C C C C C	A GCT I ATT C CAA G G G G G G G G G G G G G G G G G	GIG, GGAU EGAA T T AACC' M AACC' AACC' CAAJ EGAA CAAC CAAJ CAAA CATC R AGAC T ACT' ACT' CATC	A I G CAA C TGT L G TGT RAGA A G C T S TG G G T C C C C C C C C C C C C C C C C C C C	CCC KAAA ITTA TAACA KAAA EGAG GGG CGC AGGC GGC GGC GCC GCC GCC GCC	ACT L CTA Q CAA A GCT GGA CTT D CCT T ACT K AAG GGC TAA AG GGC TAA AG CCT T ACT K AAG CCT A CTA A CTA A CTA A CTA A CTA A CTA A CTA A CTA A CTA A CTA A CTA A CTA A CTA A CTA A CTA CT	CAI QCAG TACT GTG QCAG GAA CCAG CAG CAG CAG CAG CAG CAG CAG	L CTT GGGG CCA K AAG C TGT GGA K AAA C TGT GGA TAT V J GTG GGT U TAT V J GTG GTC C T TAT V C T TAT C T C T C T C T C C C C C C C C	836 856 876 916 936 956 976 996 016 036
2521 2521 2581 2641 2761 2761 2881 2881 2941 3001 3121 3121 3121 3131 3361	S AGTG Q CAAT E GAGT D GACA L CTAG GGAG T T CAGA A ACCG GGAG T T ACAG A CCAGA C CAGA T T ACAG T T ACAG T T ACAG C CAAT C CAAT C CAAT C CAAT C CAAT C CAAT C C CAAT C C CAAT C C CAAT C C CAAT C C CAAT C C CAAT C C CAAT C C CAAT C C CAAT C C C C	V TAC GTA GGJ W GGJ N ATC A C C C C C C C C C C C C C C C C C	GTT DCN AAT V GTA CAT FTT NAC PCA S CC DC CAT CAT CAT CAT CAT CAT CAT CAT CAT CA	GGA NAATA AAATA GGA GGA GGA GGA GGA GGA AATA AATG GAATA QAATA A A ATA A ATA A A ATA A ATA A A ATA A A ATA A ATA A ATA A ATA A ATA A	L CTT AGCA GCM ACC TGC KAG GGT R GGT R GGT R GGT AAG GAT AAG CTT EAA GGT R GGT AAG CTT EAA CTT EAA CTT A ACC A CTT A A CTT A A C A C A C	CCA CAG FTT T ACT T GAT T C GAT T C GAA C C T G C C A C C C C C C C C C C C C C C C C	CCA PCCA PCCA CCA CCA CCA CCA CCA CCA CC	GGA GGA GGTG GGTG CGTC R CGTC C GGA C GGA C C T T A CTC N N A A A T T G G G A C C T T T C T G G G A C C T C T C T C T C T C T C T C T C C C T C C T C C T C C T C C T C C T C C T C C T C C T C C T C C T C C T C C T C C T C C C C C C T C	ATC ACA D GACA D GGAC C GGAC C C GGAC C C GGAC C C GGAC C C GGAC C C GGAC C C GGAC C C GGAC C C C	AAGGA RAAGGGG GGTG GGTG GGTG GGACJ FC GGACJ FC SGTCJ A GGCT SGTCJ FC SGAAC R SGTCJ S	IGIU T ACCAJ P CCCTC G G GGAC L CTAJ T T C CCAJ L I I GT J M T G J M T G J M M G G C C A C T J L C T C T C T A C C T C C T C C T C C T C C T C C T C C T C C T C C T C C T C C T C C T C C T C	I ATC/ A A GCA(D B A CA T C T T C C T T C C T T C C T T C C T T C C T T C C T T C C T T C C T T C C T T C C C T T C	K K AAGC E BAAT T NACC C C C C C C C C C C C C C C C C C	A GCTI I ATTI C CAAJ G G G G G G G G G G G G G G G G G G	GIG, GIG, EGAA TT ACC' M AAGC' AAGA CATG CAAJ EGAA CATC T AGAC T ACT' ACT'	A I G CAA C TGT L I TG AGA A A A CCT SGT C C C C C C C C C C C C C C C C C C C	KAAA L TTA T ACA KAAA E GAG G GGG C C GTA GGG C C A G GC Q C A G G C C A G G C C A G G C C A G G C C C A G G C C C C	ACT L CTA Q CAA A GCT GGA V GGA V GGA V TAC AAG GGC C TAA AAG GGC C TAA AAG	CAI Q CAG T ACT V GTG Q CAG GAA C CAG C CAG C CAG C C C G C A C C C C C C C C C C C C C C C C C C	GIC CTT GGGG CCA K AAG CTGT GGGA H CAT K AAA D GAT TAT V I GTG I GTG I GTG I CCA K AAG C C C C C C C C C C C C C	836 856 876 916 936 956 976 996 1016 1036

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 putified plasma protein obtained by digestion with *Staphylococcus aureus* V8 protease (50 µg of barred sand bass purified cofactor protein/0.1 unit V8 protease).

RESULTS

Isolation of SB1 cDNA

In order to identify a cDNA clone representing the 110 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 a full-length clone. A 5'-end fragment of this clone (P1) was used to screen a second cDNA library constructed in λ 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 5' linker sequence used for cDNA cloning, but lacks a 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 SB1, 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 5' end of the authentic mRNA

Deduced protein sequence

The nucleotide sequence of SB1 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 a 73 and 35 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 SBP1 and human and mouse factor H, human C4bp, CR1 and other SCR-containing proteins. Comparison of the SBP1deduced amino acid sequence reveals an overall similarity to these proteins of $20-25^{\circ}_{0}$, but individual SCRs display a higher degree of homology (Figure 4). SCR 2, 12 and 16 of SBP1 show 51, 43 and 52°_{0} homology to SCR 2, 15 and 19 of human factor H respectively, and SCR 11, 12 and 13 show homology of 33, 40 and 38°_{0} 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 3' 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 a better understanding of this complex pattern, an additional probe specific for the 5' region of SB1 (P3) was used. This probe hybridized to a 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 3' end was identical with SB1, whereas the 5' end was distinct. Northern-blot analysis using the unique 5' 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 SB1 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 *Eco*RI- and *Hind*III-digested sand bass DNA (Figure 7). The 5'- and 3'-end probes hybridized to distinct fragments. The identified genomic DNA bands represent at least

MDV I A R S C V L F L W L Q T L T S V K

SQVP	υ	TLQQ	FIDGEHYDSNFDTTGMEASYPGGRQ	VRVG	υ	2 N	GY SGFFKLV	υ	VΕ	צ ט	W ETRG	AK	υ	QPR
s	υ	GН	P GDAQFADFHLAEGNDFVFGSKVV	ΥT	υ	ок Х	G' QMVSRINYRR	υ	VAE	ჟ	W DGVV	P <	υ	ESQ
0	υ		P LIHVDNNVQVIGGPEEATFGNVVR	FS	υ	КS	RSEILDGSPELY	υ	DE	RGD	W SGPV	Р Х	υ	KAI
÷	υ	AI	P PIENGNVPGAIREYKENDVLH	ΥE	υ	DR	AFKHIDRPST	υ	IKQ	G IKAE	W SPT	ΡL	v	ESI
×	υ	RLTIM	TRYEPAYRNLFSPGETLK	١٧	υ	ARTS	WISTPQETSVVTT	υ	QDN	ш U	W SIR	P T	υ	QEV
ж	υ	SNRR	P EHVDSWDVRSWERYTLDDNTR	ХW	υ	КR	GY KRTGGVTWAT	υ	GRN	Ⴊ	M M PN	ΡL	υ	е < к
Ŀ	υ	SKENI	QDAVIVGTDKQIYNLNQKAI	ΥA	υ	GЕ	G NRGRITLT	υ	GEN	თ	W SGDR	×	υ	ТЧК
4	υ	ΡL	P PKDPNGFFRGPYTGRVLY	ΥT	υ	КD	GY KLFTEGWWAEAK	υ	۷D	> ט	3	P ELTT	υ	IS
1 I N	υ	G K F	F EIPNAEVIRRYPEVQTVQ	١٨	υ	ŎN	GY STQANSFS	υ	ы С	z უ	W LLYGLS	P DQI	υ	TLRAD
>	υ	GPF	F EAENAVVKTSYQREYLSGSEVT	ΥL	υ	RDK	YIPLEGVDTIR	υ	RN	ი ი	W DKEI	¥	υ	TS
S	υ	DKLVD	VTMDFTADKEIYIEGQTIR	ΥQ	υ	LID	G AEGIAT	υ	TNN	¥	W VKS	ΡQ	υ	K V K
പ	υ	ЕГ	P ADIPNGQYDIIEGEELVFGTKIK	ΥV	υ	ыN	GY QMISKEATRT	υ	MLD	ڻ	VHNS W	ЪТ	υ	EPL
თ	υ	ЕРР	P TDGGVTVIGLPDNGNPILPDRFLD	гs S	υ	DS	GRY LNGSARLI	υ	GKD	а U	W DKPF	P S	υ	EDI
E٩	υ	EIGVM	F THVSVVGLPPGTKTIKAGQKLQLQ		υ	ND	AQPVDRPAEIE	υ	LQT	ы U	W NAAF	P A	υ	GGD
Ч	υ	TLTAV	P DNVRMTPRAALKNQMRKGQKLK	FA	υ	TD	RRDTLLGKAEVE	υ	LDN	צ ט	W SYPF	г	υ	GDPL
D	υ	GNP	P SLDDGDVTTRRQSVYRHGASVE	ΥR	υ	QAL	Y TMEGGPRKT	υ	NQ	ы U	W TGS	IR	υ	IΚ
<u>م</u>	υ	TVDTD	IMRRHNIEFRYSHEAKLYATHDDVIE	FR	υ	Тs	G RHVGSVQMRQK	υ	ΓD	Ⴊ	VMNF	ЪТ	υ	o
	υ		Р	F/Y	υ			υ		G	W	Р	υ	
Figure 3	Amin	no acid alig.	nment of SCRs of SBP1 predicted from the nucle	otide sequen	ce of	SB1								

two potential glycosylation sites The four structurally important cysteine residues are boxed with double lines. The RGD motif is marked in italics and the sequences were aligned on the basis of conserved residues according to the SCR structure (box). underlined The are

										1	2	3	4	5	6	7	8	Human C4bp
										33	40	38	25	28	29	24	ł	lomology (%)
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17		SBP1
27	51	39	33	27	37	34	38	34	31	26	43	31	24	31	52	37	ł	lomology (%)
1	2	3	4	7	8	9	11	12	13	14	15	16	17	18	19	20]	Human H
					1			1										
			5	6			10											

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 5 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 four 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 α' -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, a specific protease and a cofactor protein were identified. The cofactor activity correlated with a 110 kDa polypeptide chain of a 360 kDa plasma protein [31]. Affinity-purified IgG antibodies raised against the isolated 110 kDa material were used to immunoscreen a sand bass liver cDNA library, and a near-fulllength cDNA clone (SB1) of 3397 bp was isolated which represents a putative complement-regulatory protein. The predicted amino acid sequence (SBP1) contains a hydrophobic signal peptide. The calculated molecular mass of the predicted mature SBP1 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 SBP1 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 a 73 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 ° o homology to c1 and also shares an overlapping sequence of 1 kb with 100 °, homology to c120. Amino acids 809-826 (IGVMPTHVSxVGLPPGTK), represented by c1 and c90, correspond to the N-terminus sequence obtained from a 35 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 a cDNA clone, with a 3' end identical with SB1 but with a distinct 5' end. The positions of the RNA standard markers are shown in (kb) on the left.

SBP1 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 SB1, 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 SBP1 is composed of 17 SCRs, whereas human and mouse factor H are composed of 20 SCRs. In addition to the identical structural organization, individual SCRs of SBP1 and human factor H display an amino acid sequence homology that exceeds that due to conserved SCR motifs. Homologies of 51, 43 and 52 °, were observed between the amino acid sequence in SCR 2, 12 and 16 of SBP1 and SCR 2, 15 and 19 of human factor H respectively. A similar comparison with human C4bp shows homologies of 30, 40 and 38°_{0} in SCR 11, 12 and 13 of SBP1 and SCR 1, 2 and 3 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 2 between SBP1 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 12 in SBP1 and SCR 2 human C4bp might contribute to the C4b-binding activity [45,46].

For the first time a complete cDNA representing a protein that is structurally related to mammalian complement-regulatory proteins has been isolated from a 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 *Eco*RI (E) or *Hind*III (H). cDNA probes P5 and P6 representing the 5' and 3' 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 Al20067 and a grant from the Deutsche Forschungsgeme
ünschaft (Zi432/1). We thank Sea World of San Diego for their generous donation of barred sand bass.

REFERENCES

- 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

- 9 Sim, R. B. and Discipio, R. B. (1982) Biochem. J. 205, 285-293
- 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
- 13 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
- 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
- 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
- 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
- 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
- 23 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
- 31 Kaidoh, T. and Gigli, I. (1989) J. Immunol. 142, 1605-1613
- 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
- 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
- 37 von Heijne, T. (1986) Nucleic Acids Res. 14, 4683-4692
- 38 Skerka, C., Horstmann, R. D. and Zipfel, P. F. (1991) J. Biol. Chem. 266, 12015–12020
- 39 Skerka, C., Kühn, S., Günther, 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, 3313–3318
- 42 Lintin, S. J., Lewin, 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
- 44 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