

## GOPEN ACCESS

**Citation:** Myamoto DT, Pidde-Queiroz G, Gonçalvesde-Andrade RM, Pedroso A, van den Berg CW, Tambourgi DV (2016) Characterization of a Gene Coding for the Complement System Component FB from *Loxosceles laeta* Spider Venom Glands. PLoS ONE 11(1): e0146992. doi:10.1371/journal. pone.0146992

Editor: Paulo Lee Ho, Instituto Butantan, BRAZIL

Received: August 27, 2015

Accepted: December 25, 2015

Published: January 15, 2016

**Copyright:** © 2016 Myamoto et al. This is an open access article distributed under the terms of the <u>Creative Commons Attribution License</u>, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper.

**Funding:** This work was supported by funds from Fundação de Amparo à Pesquisa do Estado de São Paulo, Brazil (FAPESP), CeTICS Program FAPESP (2013/07467-1) and Conselho Nacional de Desenvolvimento Científico e Tecnológico, Brazil (CNPq). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. **RESEARCH ARTICLE** 

# Characterization of a Gene Coding for the Complement System Component FB from *Loxosceles laeta* Spider Venom Glands

Daniela Tiemi Myamoto<sup>1</sup>, Giselle Pidde-Queiroz<sup>1</sup>, Rute Maria Gonçalves-de-Andrade<sup>1</sup>, Aurélio Pedroso<sup>1</sup>, Carmen W. van den Berg<sup>2</sup>, Denise V. Tambourgi<sup>1</sup>\*

1 Immunochemistry Laboratory, Butantan Institute, São Paulo, Brazil, 2 Institute of Molecular and Experimental Medicine, School of Medicine, Cardiff University, Cardiff, United Kingdom

\* denise.tambourgi@butantan.gov.br

## Abstract

The human complement system is composed of more than 30 proteins and many of these have conserved domains that allow tracing the phylogenetic evolution. The complement system seems to be initiated with the appearance of C3 and factor B (FB), the only components found in some protostomes and cnidarians, suggesting that the alternative pathway is the most ancient. Here, we present the characterization of an arachnid homologue of the human complement component FB from the spider Loxosceles laeta. This homologue, named Lox-FB, was identified from a total RNA L. laeta spider venom gland library and was amplified using RACE-PCR techniques and specific primers. Analysis of the deduced amino acid sequence and the domain structure showed significant similarity to the vertebrate and invertebrate FB/C2 family proteins. Lox-FB has a classical domain organization composed of a control complement protein domain (CCP), a von Willebrand Factor domain (vWFA), and a serine protease domain (SP). The amino acids involved in  $Mg^{2+}$  metal ion dependent adhesion site (MIDAS) found in the vWFA domain in the vertebrate C2/FB proteins are well conserved; however, the classic catalytic triad present in the serine protease domain is not conserved in Lox-FB. Similarity and phylogenetic analyses indicated that Lox-FB shares a major identity (43%) and has a close evolutionary relationship with the third isoform of FB-like protein (FB-3) from the jumping spider Hasarius adansoni belonging to the Family Salcitidae.

## Introduction

During evolution, two systems of immunity have arisen: innate and adaptive. The innate immune system is the oldest and found in all multicellular organisms, while the adaptive immune system, which emerged about 450 million years ago, is present only in vertebrates, except for the Agnatha [1,2]. The complement system, in mammals, plays an important role in both, innate and adaptive immune system and is composed of more than 30 serum and cell-surface components that participate in the recognition and clearance of invading pathogens. The activation of the complement system can occur by three pathways: classical, lectin and



**Competing Interests:** The authors have declared that no competing interests exist.

alternative that converge at the cleavage of the central complement component C3, by the C3 convertases [3].

In the alternative pathway, FB acts as the catalytic subunit of the C3 convertase; in the classical and lectin pathways, this role is played by C2. In mammals, FB and C2 share the same domain and genomic organization, with a significant amino acid similarity and, possibly, they diverged at the jawed vertebrate lineage by gene duplication [1,4]. Human FB is a modular chymotrypsin-like serine protease comprised of N-terminal region, composed of three complement control protein (CCP) domains, a linker region, a vWFA (von Willebrand factor type A) domain, and a C-terminal serine protease (SP) domain, which contains the catalytic site. The vWFA and SP domains form the fragment Bb, while the CCP1-3 and the linker form the fragment Ba. Following binding of FB to C3b, FB is cleaved by factor D into fragments Ba and Bb. FB binding to C3b depends on the CCP elements in fragment Ba and on the Mg<sup>2+</sup>-metal iondependent adhesion site (MIDAS) motif, in the vWFA domain of fragment Bb [5].

The CCP module is a domain commonly present in many mammalian complement proteins that is responsible for mediating protein-protein interactions of complement proteins or, as in factor H, to bind to self-cells. Among the three CCPs present in human FB, the third one has structural elements that are crucial for the interaction with C3b fragment.

The studies of vertebrate and invertebrate genomes revealed that many domains of mammalian complement components are found in both deuterostomes and protostomes. According to Nonaka and Kimura (2006) [2], the origin of the complement system probably occurred with the appearance of C3 and FB, the only components found in some protostomes and in cnidarians, suggesting that the alternative pathway represents the most ancient complement pathway. Whereas C3 and FB were maintained in all deuterostomes, they were lost many times, independently, in the protostome lineage, which explains the absence of these components in the insect *Drosophila melanogaster* [6] and in the worm *Caenorhabditis elegans* [7].

Since *MBL* (mannose-binding lectin), *MASPs* (MBL-associated serine proteases) and *ficolins* genes, that play a role in the lectin pathway activation, have not been identified in protostomes and echinoderms, it was suggested that these components were recruited after the emergence of chordates, about 900 million years ago. However, the recent finding of a *MASP* gene in cnidarians [8] suggests that the primitive lectin pathway could operate, besides the alternative pathway, in those animals. The agnates that are jawless vertebrates have developed only the alternative and lectin pathways of the complement system, probably due to absence of immunoglobulin genes [9]. Finally, the gene duplication events that happened in *C3/C4/C5*, *FB/C2* and *MASPs/C1r/C1s*, before the emergence of cartilaginous fishes about 600 million years ago, were important steps in the establishment of the classic pathway [2].

Genes coding for FB-like proteins have been characterized in different organisms belonging to previous lineage of chordates. FB-like sequences from the limulus *C. rotundicauda* (CrC2/Bf) [10] and the sea urchin *S. purpuratus* (SpBf) [11] have some particularities, such as the presence of five CCP domains instead of three as found in mammals, suggesting that the presence of additional CCP domains is a primitive characteristic of C2/FB proteins. In contrast, Kimura et al. (2009) [8] characterized two FB-like protein isoforms from the anemone *Nematostela vectensis* (NvBf-1 and NvBf-2): the first one contained three CCPs while the second had five CCP domains as found in vertebrates. Recently, some FB-like gene with the same composition of CCP domains as found in vertebrates. Recently, some FB-like sequences from organisms belonging to the Arthropoda phylum were identified and multiple isoforms, containing between two and seven CCPs were found in these organisms [12]. It is possible that the composition of three CCP domains is an ancestral characteristic and that this region has undergone gene duplication or deletion in invertebrates FB. Nonetheless, it seems likely that predicted proteins presenting the same structural domains, found in anemone and limulus, are

orthologues of C2 and FB genes, and probably have arisen before divergence of Cnidaria/Bilateralia [2].

When we elucidated the transcriptome of the *Loxosceles laeta* spider venom gland, in addition to finding the expression profile of the Sphingomyelinases D, the major proteins responsible for the envenomation, other EST sequences with similarity to C3 and FB-like genes, from invertebrate organisms, were identified [13]. These findings suggest that the central components of the complement system could be expressed in the venom gland of the *Loxosceles* spiders. Thus, the present work aimed to clone and characterize the FB complement component from *Loxosceles laeta* venom gland and phylogenetically analyze its deduced amino acid sequence.

### **Material and Methods**

#### Loxosceles spiders and isolation of RNA

*Loxosceles laeta* spiders were collected in Campo Alegre, Santa Catarina, Brazil and kept at Immunochemistry Laboratory of Butantan Institute, São Paulo, Brazil. Eighty *L. laeta* female spiders were subjected to food restriction to stimulate the production of mRNA in the venom glands. After 5 days, the venom glands were collected and frozen at -80°C until use. For total RNA extraction, the Trizol reagent was used following the manufacturer's instructions (Gibco-BRL Life Technologies, MD, USA). The authorization to access the *L. laeta* (permission no. 01/ 2009) was provided by the Brazilian Institute of Environment and Renewable Natural Resources (IBAMA), an enforcement agency of the Brazilian Ministry of the Environment.

## RT-PCR and Rapid amplification of cDNA ends (RACE)

Based on the EST sequence LLAE0889S, which we previously identified in the transcriptome of *L. laeta* to be similar to the complement factor B [13], specific sense and antisense primers were designed to amplify the complete gene sequence (FB sense – 5' CGAAGCAGCTCAAG GACCAC 3' and FB antisense – 5' CCTTCCATCCATGCGACCAC 3'). The SMARTer RACE cDNA Amplification kit (Clontech, CA, USA), was used to amplify the *Loxosceles* FB (Lox-FB) RNA. The PCR reactions were performed using the following conditions: 2 cycles of 94°C for 30 sec, 72°C for 3 min; 5 cycles of 94°C for 30 sec, 70°C for 30 sec and 72°C for 3 min and 27 cycles of 94°C for 30 sec and 72°C for 3 min.

### Cloning and sequencing of Lox-FB

Resulting products from the RACE reactions were separated in 1% agarose gel and the positive PCR products were purified using PureLink<sup>™</sup> PCR Purification kit (Invitrogen, CA, USA) and directly cloned into pGEM-T-Easy Vector (Promega, WI, USA), at 16°C overnight, and transformed into *E. coli* competent Cells XL1Blue. Positively transformed cells were grown overnight at 37°C in LB (Luria Bertani) broth supplemented with 100 µg/mL ampicillin. Plasmids were isolated by Boiling Plasmid Mini Prep method [14], digested with restriction endonuclease enzyme EcoRI (New England Biolabs, MA, USA) and purified using phenol and chloroform [15].

The positive clones were sequenced from both ends with T7 (5' TAATACGACTCACTATA GGG 3'), Sp6 (3' TAAATCCACTGTGATATCTT 5') and specific primers for FB (FB sense—5' CGAAGCAGCTCAAGGGACAC 3'-and FB antisense—5' CCTTCCATCCATGCGACC AC 3') using Big Dye terminator Cycle Sequencing Ready Reaction Kit and automated DNA sequencer, model ABI 3100 capillary electrophoresis (Applied Biosystems, CA, USA).

## Sequence analysis and molecular modeling

All sequences were analyzed both at the nucleotide and amino acid levels using the Basic Local Alignment Search Tool (BLAST) from the National Center for Biotechnology Information (NCBI: <u>http://www.ncbi.mlm.nih.gov/blast/BLAST.cgi</u>). Translation and protein analyses were performed using ExPaSy tools (<u>www.expasy.org</u>). The deduced amino acid sequence of Lox-FB was aligned with the corresponding sequences of various animals using MEGA 6 software. On the basis of the human FB structure (PDB ID code: 20k5.1), Swiss-model website (<u>http://</u><u>swissmodel.expasy.org</u>) was used to create a comparative homology model of Lox-FB [16].

## Phylogenetic analysis

All FB-like sequences used for phylogenetic analysis were downloaded from the GenBank database. Multiple sequence alignments were performed with full length open reading frame sequences using MUSCLE (Multiple Sequence Comparison by Log-Expectation) and the phylogenetic tree was constructed based on this alignment using the Maximum Likelihood (ML) algorithm available in MEGA 6 software [17]. Statistical confidence of the evolutionary analysis was assessed by bootstraps of 1000 replicates.

## Results

### Characterization of L. laeta FB

The 5' and 3' RACE fragments yielded a complete reading frame (ORF) of *Loxosceles laeta* factor B-like composed of 1953 base pairs (bp) that encodes for a protein of 651 amino acids (Fig 1). The NCBI's conserved domain database (CDD) program identified that Lox-FB has a classical domain organization, composed of two CCPs, a vWFA domain and a SP domain (Fig 2). The leader peptide signal is composed of 25 amino acids producing a mature protein of 626 amino acids. Eighteen cysteines were found, eight of them present in the CCP domains, one cysteine in the vWFA domain and nine cysteines in the SP domain (Fig 1). The deduced molecular weight of Lox-FB was predicted as 72.38 kDa and the isoeletric point as 5.73, without considering the eight putative N-glycosylation sites.

## Multiple Alignment of Lox-FB with other FB/C2-like proteins

Although the C2/FB proteins described in the literature, until now, have the same architecture of domains composition, they are different in some aspects like the quantity of CCP domains. For example, the FB-like found in some species belonging to Echinodermata phylum [11], horseshoe crab [10] and one of two isoforms present in a sea anemone [8] contains five CCP domains instead of three found in vertebrates, in which they were the first to be characterized (Table 1). Lox-FB has only two CCP domains, as FB found in bivalves [18], FB-2 centipede, FB-2 sea spider and FB-3 jumping spider [12]. Because of these differences, alignments at this position tended to be out of register in C2/FB sequences that contain more than two CCP domains. Each CCP module from Lox-FB has approximately 60 amino acids of length and there are some highly conserved residues as proline (P), glycine (G), tryptophan (W) and four cysteines (forming two disulfide-bridges; I-III and II-IV) (Fig.3).

Analyzing the other two domains, many conserved sites were detected, such as the five amino acids involved in the binding to C3b dependent on  $Mg^{2+}$  ions present in the vWFA domain; seven cysteines and those regions close to the active site also appeared in the same position in the SP domain (Fig 4). However, none of the three amino acids residues important in the serine protease activity (catalytic triad) were conserved (Fig 4). Apparently, the classic catalytic triad of serine peptidases belonging to trypsin-like family composed of histidine (H),

1	atgo	catc	cata	qaq	taa	aca	atto	stte	qqa	ttee	ete	ttt	taga	ttt	caact	atta	ataco	igcaa	agca	aat	age	tee	gta	tge	cat	cectto	90
1		A S				A								Ι.		-	I		-		s		v	C	Ř	PL	30
91	cegt	atte	ttet	ttg	egga	aac	ggga	aca	ate	atco	jete	gaaa	agto	gag	gaact	ttt	gett	cgaa	atge	aac	aaa	ggti	taci	tt	ctc	ggagga	180
31	P	Y S	s	L	R	Ν	G	Ν	I	I	Α	Ε	s	G (	3 Т	F	RE	E	С	Ν	K	G	Y	F	L	GG	60
181	ccct	caaa	aata	cgt	tget	tata	aagg	igaa	aaci	tgga	icgt	tegg	jaaa	gag	cacco	gtgt	tgeat	cgaa	atet	gga	gaat	ttt	tgtg	gate	geta	cctcaa	270
61	P	S K	I	Ř	C	Y	K	Ğ	N	W	т	s	E	Ř I	A P	v	C 1	ΕĒ	s	Ğ	E	F	С	D	A	P Q	90
271	aaga	tege	aaac	gca	aaa	ctg	cace	gto	ceci	acco	jt gi	tate	jaaa	ttga	attet	gtea	ataca	gtad	atg	tge	cgto	cca	ggat	cate	gtgo	ctggtg	360
91	K	IA	N	А	К	L	H	G	P	Т	v	Y	E	II	D S	v	IQ	2 Y	Μ	С	Ř	P	G	Y	V	ΓV	120
361	ggca	acgg	acgg	agg	atat	tge	ctca	igga	agto	gget	ati	tgga	ageg	gta	cact	ceca	acato	cate	ggat	gaa	tee	gag	ceto	ctgo	caga	aatgtt	450
121	G	N G	R	R	I	С	L	R	s	G	Y	W	s	G :	ΙТ	P	Т	M	D	Ε	s	Ε	P	L	Q	N V	150
451	getg	Jagag	attg	aac	aaaq	gag	tttç	ytta	acga	aaaa	atge	gcat	tege	act	caaat	gaca	agtto	aata	acaa	cag	geti	tati	tat	ctto	gaca	agtget	540
151	A	E R	L	Ν	K	E	F	v	т	К	М	А	s	H S	5 N	D	S S	5 I	Q	Q	А	Y	Y	L	D	S A	180
	tcta	iggegi	aaga	ggt	ctcq	gat	ctgg	ytta	atte	ctgo	jtte	gaca	agat	cca	gcago	gttg	gated	tttt	gac	ttg	aacq	gtt	gee	aaga	aaat	ttcata	630
181	s	R R	R	G	L	D	L	v	I	L	v	D	R	S :	5 S	v	DI	? F	D	L	Ν	v	А	К	К	FΙ	210
631	aaat	attt	gett	caa	gaat	tte	ggeg	gtaa	aaga	aaca	igea	aaca	agea	acca	aaato	iggea	acteg	iggea	iget	gtt	atte	gegi	tttg	gga <i>a</i>	acga	accatt	720
211	K	ΥL	L	Q	E	F	G	v	K	N	s	Ν	s	N (	2 M	G	ΤF	A S	A	v	I	А	F	G	Т	ТΙ	240
	gata	tcat	tttc	aac	ctga	aat	gaca	acco	caca	atto									gac	gaa	ttad	ctga	aago	2Cta	aatç	ggtggg	810
241	D	ΙI	F	Ν	L	N	D	т	н	I	D	s	P	D	A A	G	VZ	A L	D	Е	L	L	К	Ρ	N	GG	270
	ggaa	cage	aatg	gtt	cegg	gca	ctaa	acct	tead	gtgt	tti	ttga	aaag	itte	ageed	tte	tteg	Itgaa	aagg	tcc	aago	cat	geet	tgt	ttt	ctcatt	900
271	G	т А	М	v	P	А	L	т	s	v	F	L	К	V (	2 P	F	L F	ξ E	R	s	К	н	А	L	F	LI	300
	ageg	Jacgg	agaa	cca	aaca	atc	ggeg	làcă	gate	gaca	itca				acata	tcga	aggea	gete	aag	gac	cac	get	geti	ttt	gaaa	atettt	990
301	s	DG	Е	P	Ν	I	G	G	D	D	I	т	Р	ΕI	DI	s	RÇ	) L	К	D	н	А	А	F	Е	I F	330
						-		-	-				-		-	•	-	-	-			-		-		gataaa	
331		VG	_	_	Р	_	-	-		К	_	_			7 A		ΕI				н	-	_	М	L	DK	360
	ttca	ccga	cete	gag	gaca	atga	atgo	jeca		-			-	-	-	-	-	-	-	ctg	gata	-			ogea	atggat	
361	F		L	Е	_		М		_					-	ΓE		_	e v			D	R	С	G	R	M D	390
	ggaa	ggaa		-				-	_			-		-	-						acgo	caa	ctaa	-	ctat	tgegee	
391	G		E	s	E	G	W	P	W	v	A	А	v	V 1	7 5	Y	E A			s	т	_	-			C A	420
	ggcg							_								_	_		N		_		L	s	Г	_	
421			tttc	tgt	ageo	_	tggg	jtgi	ttga	acga	icto			get	tettt	tata	attga	ggaa	aac	aga	gtta	agti	ttta	aata		gaggaa	1350
	-	A L	F	tgt C	ageo S	Q	tggg W	ytgi V	ttga L	acga T	T	Α	s	get C 1	tettt F F	tata Y	attga I E	iggaa E E	aaac N	aga R	gtta V	agt: S	ttta F	aata N	I	gaggaa E E	1350 450
1351	-	A L	F	tgt C	ageo S ggao	Q gaa	tggg W tatg	ytyt V gact	ttga L ttga	acga T atgg	T T ggag	A gagg	S gato	get C 1 cga	tetti F F aacaa	tata Y agga	attga I E aattt	iggaa E E ctao	aaac N cgct	aga R agt	gtta V aaa	agt S gta	F gtca	aata N	I	gaggaa E E ccggac	1350 450 1440
1351 451	gtgt V	A L ttgt F V	F tgtt V	tgt C atg M	ageo S ggao G	Q gaar E	tggg W tatg Y	ytgi V gaci D	ttga L ttga L	acga T atgg M	T ggag G	A gago E	S gato D	get C 1 cgai P 1	tetti F F aacaa K Q	tata Y agga R	attga I E aattt N E	ggaa E E ctao Y	aac N cgct A	aga R agt S	gtta V aaa K	agti S gta V	gtca V	aata N atco I	I caco H	gaggaa E E ccggac P D	1350 450 1440 480
1351 451 1441	gtgt V tatg	A L ttgt F V jatga	F tgtt V atct	tgt atg M gac	ageo S ggao G acao	Q gaa E ctg	tggg W tatg Y ctga	yact V D act	ttga L ttga L aatt	acga T atgg M ttgg	T Jgag G Jcao	A gagg E ctgg	S gato D gtca	eget C 1 cgai P 1 age	tetti F F aacaa K Q ttead	tata Y agga R aago	attga I E aattt N E cctgt	ggaa E E Ctao Y Geca	aaac N cgct A actt	aga R agt S gag	gtta V aaa K acgt	gta S V tte	gtca V	atco N I CCC2	I caco H	gaggaa E E ccggac P D tgttta	1350 450 1440 480 1530
1351 451 1441 481	gtgt V tatg Y	A L ttgt F V gatga D E	F tgtt V atct S	tgt atg M gac D	ageo S ggao G acao T	Q gaa E ctg L	tggg W tatg Y ctga L	gtgt V gact D aaca N	ttga L ttga L aatt N	acga T atgg M ttgg L	ggag G gcao A	A gagg E ctgg L	S gato D gtca V	eget C 1 ecgai P 1 aget K 1	tetti F F aacaa K Q tteac L H	tata Y agga R aago K	attga I B aattt N B cctgt P \	iggaa E E Setao F Y Sgeea I P	aaac N cget A aett L	aga R agt S gag E	gtta V aaa( K acgt T	agti S gta V ttea F	gtca V aaa K	atco I CCC4 P	I caco H attt I	gaggaa E E coggao P D tgttta	1350 450 1440 480 1530 510
1351 451 1441 481 1531	gtgt V tatg Y cctc	A L ttgt F V gatga D E ccgac	F tgtt V atct S cgat	tgt atg M gac D agg	agco S ggao G acao T acao	Q gaar E ctgo L gtto	tggg W tatg Y ctga L ccat	gtgt V gact D aaca N tgt	ttga L ttga L aatt N ttti	acga T atgg M ttgg L ttgg	ggag G gcao A gato	A gago E ctgo L cata	S D gtca V Aaca	P I Rage K I	tetti F F aacaa K Q tteac L H acget	tata Y agga R aago K teeo	Attga I E AAttt N E CCtgt P V gttgo	Aggaa E E Seta C Y Sgeca 7 P Segga	Aaac N cget A Actt L Atgg	aga R agt S gag E ggg	gtta V aaao K acgt T attt	agti S gta V tto F	ttta F gtca V aaao K gcgo	aata N atco I ccca P ccga	I Caco H attt I aaaa	gaggaa E E coggac P D tgttta L aatggo	1350 450 1440 480 1530 510 1620
1351 451 1441 481 1531 511	gtgt V tatg Y cctc P	A L F V gatga D E ccgac P T	F tgtt V atct S cgat D	tgt atg M gac D agg R	agco S ggao G acao T acao T	Q E ctg L gtt V	tggg W tatg Y ctga L ccat P	gtgt V gact D acca N tgt L	ttga L ttga L aati N ttti F	acga T atgg M ttgg L ttgg L	T ggag G gcad A gato D	A gagg E ctgg L cata H	S gato D gtca V aaca N	get cga P I age K I taa I	tetti F F aacaa K Q ttead L H aeget A	tata Y R aagga K toog S	Attga I E AAttt N E Ctgt P V gttgo V E	Aggaa C E Setao C Y Sgeca J P Segga A G	Aaac N cget A Actt L Atgg W	aga R agt S gag gag ggg G	gtta V aaao K acgt T attt I	agti S gta V tte F tce S	ttta F gtca V aaao K gcgo A	aata N atco I ccca P ccga P	I caco H attt I aaaa K	gaggaa E E coggac P D tgttta C L aatggo N G	1350 450 1440 480 1530 510 1620 540
1351 451 1441 481 1531 511 1621	gtgt V tatg Y cctc P tttg	A L F V gatga D E cogac P T gaaac	F tgtt V atct S cgat D	tgt atg gac D agg R	agco S ggao G acao T acao T	Q gaar E ctgo L gtto V agto	tggg W tatg ctga L ccat P gcca	gtgi V gact D aaca N tgi L acga	ttga L ttga L aatt N ttti F aggi	acga T Atgg M ttgg L ttgg L tccc	ggag G gcao A gato D caat	A gagg E ctgg L cata H ttga	S gato D gtca V aaca N actg	regat C i cegai P i agen K i taai I i tge	tetti F F aacaa X Q ttead L H aeget A geaad	tata Y agga R aago K teeo S steeo	attga I E aattt N E cetgt P V gttgo V P atact	Aggaa E E Setao Y Y Sgeea V P Seegga A G Sggea	Aaac N Actt L Atgg W Agaa	aga R agt S gag gag G gaa	gtta V aaaq K acgt T attt I cttq	gta S V tte S gaa	ttta F gtca V aaao K gcgo A	aata N atco I ccca P ccga P	I caco H attt I aaaa K	gaggaa E E coggao P D tgttta C L aatggo N G agaato	1350 450 1440 480 1530 510 1620 540 1710
1351 451 1441 481 1531 511 1621 541	gtgt V tatg Y cctc P tttg F	A L ttgt F V gatga D E ccgac P T gaaac E T	F tgtt V atct S cgat D aatc I	atg atg gac D agg R agc S	ageo S G acao T acao T cega P	Q gaa E ctgo L gtto V agto S	tggg W tatg Ctga L ccat P gcca A	gtgt V gact D acca N tgt L acga T	ttga L ttga L aatt N tttt F aggt R	acga T atgg M ttgg L ttgg L ccc S	ggaq G gcaq A gato D caat	A gago E ctgo L cata H ttga L	S gato D gtca V aaca N acto T	get Cgai P I age K I ataa I Jtge V I	tettt F F aacaa X Q ttead L H aeget A geaad X A	tata Y agga R aago K teeo S cgtta V	Attga I E Aattt N E Sectgt P V Sttgo V P Atact I I	Aggaa E E Setao Y P Seegga A G Sggca A A	Aaac N eget A Actt L Atgg W Agaa E	aga R agt S gag gag G gaa E E	gtta V aaaq K acgt T attt I cttg L	gta S gta V tto F S gaa E	ttta F gtca V aaao K gcgo A tgco	aata N atco I ccca P ccga R	I caco H attt I aaaa K agga R	gaggaa E E coggac P D tgttta tgttta L aatggc N G agaatc R I	1350 450 1440 480 1530 510 1620 540 1710 570
1351 451 1441 481 1531 511 1621 541 1711	gtgt V tatg Y cctc P tttg F	A L ttgt F V gatga D E ccgac P T gaaac E T	F tgtt V atct S cgat D aatc I	atg atg gac D agg R agc S	ageo S G acao T acao T cega P	Q gaa E ctgo L gtto V agto S	tggg W tatg Ctga L ccat P gcca A atag	U gact D acca N tgt L acga T	ttga L ttga L aatt N tttt F aggt R	acga T atgg M ttgg L ttgg L tccc S tgcg	ggag G G JCa D Caat Q JCa JCa	A gago E ctgo L cata H ttga L gaaa	S gato D gtca V acto Acto Atca	get get P age K taa taa tge V I agtge	tettt F F aacaa X Q ttead L H acget A geaad R N geaad	tata Y agga R aago K tooo S sgtta V acto	Attga I E Aattt N E Cetgt P V gttgo V F Atact I I gacaa	Aggaa E E Setao Y P Sgeca A G Sggca A G Sggca A A	Aaac N eget A Actt L Atgg W Agaa E	aga R agt S gag gag G gaa E E	gtta V aaaq K acgt T attt I cttg L	agti S gta V ttc S gaat E cct	ttta F V aaao K gego A tgeo	atco I coca P coga R toco	I caco H attt I aaaa K agga R ccaa	gaggaa E E coggao P D tgttta C L aatggo N G agaato	1350 450 1440 480 1530 510 1620 540 1710 570 1800
1351 451 1441 481 1531 511 1621 541 1711 571	gtgt V tatg Y cete P tttg F aacc N	A L sttgt F V jatga D E scgac P T jaaac E T sgaag R R	F tgtt V atct S cgat D aatc I gtta L	atg atg gac D agg R agc S agc	ageo S G acao T acao T cega gaaa E	Q E ctgo L gtto V agto S acco T	tggg W tatg ctga L ccat P gcca A atag I	yact D act D act Stgt L acga T ytt V	ttga L ttga L aatt K tttt R aggt R tttt F	acga T Atgg L ttgg L tccc S tgcg	T ggag G gcad A gato D caat Q gcad A	A gagg E ctgg L cata H ttga L gaaa E	S gato D gtca V acta N acto T atca I	get P P tage K taa taa J tge S ( S	tettt F F aacaa K Q ttead L H aeget A geaac R N geaet G T	tata Y agga R aago K tooo S cgtta V acto T	Attga I E Aattt P V Sttgo V F Atact I I Sacaa D N	Aggaa E Setao Y Sgeca Y P Sggca A G Sggca A G Sggca A C A T C	Aaac N eget A actt L atgg W agaa E tgta	aga R S gag gag gaa ggg G G G G	gtta V aaao K acgt T attt L aato N	agti S gta V tto F tco S gaat E C C P	tti F V aaao K gego A tgeo G ggti G	ata N atco P coga P coga R tcco S	I caco H attt I aaaa K agga R ccaa P	gaggaa E E coggac P D tgttta L aatggo N G agaato R I ataata I I	1350 450 1440 1530 1620 510 1620 540 1710 570 1800 600
1351 451 1441 481 1531 511 1621 541 1711 571 1801	gtgt V tatg Y cctc P tttg F aacc N gcca	A L sttgt F V jatga D E scgac P T jaaac E T scgaag R R aaaga	F tgtt V atct S cgat D aatc I gtta L ccc	atg M gac D agg R agc S agg S agg	ageo ggao G acao T cega P gaaa E acco	Q gaar E ctgc L gtt V agt S acco T gga	tggg W tatg Ctga L ccat P gcca A atag I atca	yact D Aaca N tgt L Acga T yttt V Aaco	ttga L ttga L aatt F aggt R tttt F cace	acga T atgg L ttgg L tccc S tgcg gttt	ggag G gcag A gato D caat Q gcag A stgg	A gagg E ctgg L cata H ttga J gaaa E ggao	S gato D gtca V acta N acto T atca I	get egai P I age K I taa I tge S ( tga	tettt F F aacaa X Q ttead L H aeget A geaad R A geaad S T geeag	tata Y agga R agga tccc S sgtta V acto T agga	Attga I E Aattt N E Cotgt P N gttgo V P Atact I I gacaa D N A <u>a</u> taa	Aggaa E E Setao Y P Seegga A G Sggca A G Sggaa A G Sggaa Sgg	Aaac N cgct A actt L atgg W Agaa E cgta Cgac	aga R agt S gag gag gag gaa ggc G aaa	gtta V aaaq K acgt T atti L ctto L aato N ggtt	agti S gta V ttee S gaat E eete P tee	tti F gtca V aaao K gego A tgeo ggti G G	aata N atco I coca P coca P coca R toco S caat	I caco H attt aaaa K agga R ccaa P tatt	yaggaa E E coggac P D tgttta L tgttta L aatggo N G agaato R I ataata I I tttgag	1350 450 1440 480 1530 510 1620 540 1710 570 1800 600 1890
1351 451 1441 481 1531 511 1621 541 1711 571 1801 601	gtgt V tatg P tttg F aacc N gcca A	A L Sttgt F V gatga D E Scgac P T gaaac E T Sgaag R R saaga K D	F tgtt V atct S cgat D aatc I gtta L tccc P	atgo Atgo D aggo R aggo S aagc S aagc S aagc S aagc	ageo ggag G acao T eega P gaaa E aceo T	Q gaar E ctgo L gtto Agto S acco T ggaa G	tggg W tatg ctga L ccat P gcca A atag I atca I	ytgt V pact D access N ttgt L acgs T V access V V access N	ttga L ttga aatt N tttt R tttt F caco H	acga T atgg L ttgg L tccc S tgcg gttt V	Acto T ggag G gcao A gato D caat Q gcao A ttgo L	A gagg E ctgg L cata H ttga J gaas E ggao G	S gato D gtca V acca N acto T atca I ctct L	get cgai P I age K I taa I J tga S ( tga L	tettt F F aacaa X Q ttead L H aeget A geaad R N geaet 3 T geeag S Q	tata Y agga R agga K teco S sgtta V acto R agga R	Attga I F Aattt N F Sectgt P V Sttgo V F Atact I I gacaa D N Aataa	aggaa E E Setao Sgeca V P Sgeca A G Sggca A G Sggca A C Sggca A C Sggca C C	aac N eget A aett L atgg W agaa E tgta V cgac D	aga R agt S gag gag G gaa ggg G aaa K	gtta V aaaq K acgt T atti L ctto L aato N ggtt	agti S gta V tto F tco S gaat E C C P	tti F V aaao K gego A tgeo G ggti G	ata N atco P coga P coga R tcco S	I caco H attt aaaa K agga R ccaa P tatt	yaggaa E E coggac P D tgttta L tgttta L aatggo N G agaato R I ataata I I tttgag	1350 450 1440 1530 1620 510 1620 540 1710 570 1800 600
1351 451 1441 481 1531 511 1621 541 1711 571 1801 601 1891	gtgt V tatg Y cctc P tttg F aacc N gcca A ctca	A L sttgt F V Jatga D E scgac P T Jaaac E T Sgaage R R Laaga K D	F tgtt V atct D aatc gtta gtta gtta gtac	atg atg atg agg agg agg agg agg agg agg	ageo S ggag G acao T cega gaaa P gaaa E aceo T tett	Q gaar E ctg gtt V agt S acco T ggaa G G	tggg W tatg ctga L ccat P gcca A atag I atca I atca	ytgi V gaci D acca N ttgi L acga T V acca N V acca N N acca N	ttga L ttga L aatt F aggt R tttt F caco H aaao	acga T atgg L ttgg L tccc S tgcg gttt V gaat	T Jgaq G G Jgaq D Caat Q Jgaq A Caat	A gagg E ctgg L cata H ttga L gaaa E ggao G gaca	S yato D ytca V acca N acto T atca I etct L aaat	get cgaa P 1 age K 1 itge S 0 itga L 2 gte gte	tettt F F aacaa K Q ttead L H aeget A geaad R N geaad R N geaet G C geeag S Q acttt	A CONTRACTOR CONTRACTO	Attga I E Aattt P V Sttgo V F Atact I I Sacaa D N Aataa	aggaa E E Setao Sgeca V P Sgeca V P Sggca A G Sggca A G Sggca A G Sggca A G Sggca C C C C Sggaa C C C Sggaa C C C C Sggaa C C C C C C C C C C C C C C C C C C C	Aaac N actt A Actt A A A A A A A A A A A A A A A	aga R agt S gag ggg G gaa ggc G aaa K 53	gtta V aaaq K acgt T atti L ctto L aato N ggtt	agti S gta V ttee S gaat E eete P tee	tti F gtca V aaao K gego A tgeo ggti G G	aata N atco I coca P coca P coca R toco S caat	I caco H attt aaaa K agga R ccaa P tatt	yaggaa E E coggac P D tgttta L tgttta L aatggo N G agaato R I ataata I I tttgag	1350 450 1440 480 1530 510 1620 540 1710 570 1800 600 1890
1351 451 1441 481 1531 511 1621 541 1711 571 1801 601	gtgt V tatg Y cctc P tttg F aacc N gcca A ctca	A L Sttgt F V gatga D E Scgac P T gaaac E T Sgaag R R saaga K D	F tgtt V atct D aatc gtta gtta gtta gtac	atg atg atg agg agg agg agg agg agg agg	ageo S ggag G acao T cega gaaa P gaaa E aceo T tett	Q gaar E ctg gtt V agt S acco T ggaa G G	tggg W tatg ctga L ccat P gcca A atag I atca I	ytgi V gaci D acca N ttgi L acga T V acca N V acca N N acca N	ttga L ttga L aatt F aggt R tttt F caco H aaao	acga T atgg L ttgg L tccc S tgcg gttt V gaat	Acto T ggag G gcao A gato D caat Q gcao A ttgo L	A gagg E ctgg L cata H ttga L gaaa E ggao G gaca	S yato D ytca V acca N acto T atca I etct L aaat	get cgaa P 1 age K 1 itge S ( itgaa L 2 gte gte	tettt F F aacaa X Q ttead L H aeget A geaad X N geaet 3 T geeag S Q	tata Y agga R agga K teco S sgtta V acto R agga R	Attga I F Aattt N F Sectgt P V Sttgo V F Atact I I gacaa D N Aataa	aggaa E E Setao Sgeca V P Sgeca V P Sggca A G Sggca A G Sggca A G Sggca C C Sggca C C C C Sggca C C C Sggca C C C C Sggca C C C C C C C C C C C C C C C C C C C	Aaac N actt A Actt A A A A A A A A A A A A A A A	aga R agt S gag ggg G gaa ggc G aaa K 53	gtta V aaaq K acgt T atti L ctto L aato N ggtt	agti S gta V ttee S gaat E eete P tee	tti F gtca V aaao K gego A tgeo ggti G G	aata N atco I coca P coca P coca R toco S caat	I caco H attt aaaa K agga R ccaa P tatt	yaggaa E E coggac P D tgttta L tgttta L aatggo N G agaato R I ataata I I tttgag	1350 450 1440 480 1530 510 1620 540 1710 570 1800 600 1890

Fig 1. cDNA sequence and deduced amino acid sequence of Lox-FB. The ORF predicts a protein of 651 amino acids with domains of known C2 and Bf proteins. The signal peptide is in italics and comprised of the initial 25 amino acids. The cysteines and the putative N-glycosylated sites are marked in black and grey, respectively.

doi:10.1371/journal.pone.0146992.g001

LOS ONE

aspartic acid (D) and serine (S) residues [19,20] was replaced by other amino acids: serine (S), asparagine (N) and proline (P). Interestingly, Lox-FB, FB-like isoform 3 from the spider *Hasarius adansoni* and FB-like molecule of bivalve *Ruditapes decussatus* are similar, since they have only two CCP domains, two extra cysteines (highlighted in grey) and did not preserve the catalytic triad. Considering only the spider species, both of them share the same amino acids at the first and second position of the protease active site, but not the last one. Because of these





Species	Protein	N° of CCPs	Extra Domains	Length (aa)
N. vectensis	FB-1	3	-	708
	FB-2	5	-	858
Ammothea sp	FB-1	7	-	1023
	FB-2	2	-	658
H. adansoni	FB-1	7	-	1006
	FB-2	4	-	847
	FB-3	2	-	636
Loxosceles laeta	FB	2	-	651
S. subspinipes	FB-1	7	-	1076
	FB-2	2	-	658
T. tridentatus	C2/FB-1	5	-	889
	C2/FB-2	7	-	972
C. rotundicauda	C2/FB-1	5	-	889
	C2/FB-2	5	-	889
R. decussatus	FB	2	-	697
S. purpuratus	FB	5	-	833
A. japonicus	FB-1	5	-	913
	FB-2	5	-	865
B. belcheri	FB/C2	3	EGF-CA	752
C. intestinalis	FB-1	4	-	999
	FB-2	4	-	998
	FB-3	4	-	963
H. roretzi	FB	5	LDL_A	1084
L.camtschaticum	FB-1	3	-	763
	FB-2	3	-	749
Homo sapiens	FB	3	-	764
	C2	3	-	752

#### Table 1. Structural features of invertebrates FB.

doi:10.1371/journal.pone.0146992.t001

differences, it could be that these proteins have lost their proteolytic activity or they have a different mechanism of activation. Despite the difference in the amino acids considered to be of importance to the enzymatic function of human FB, Lox-FB had conserved amino acids residues placed surrounding the triad, particularly Thr<sup>54</sup>, Ala<sup>56</sup> and Ser<sup>214</sup> (chymotrypsin numbering) that play an important role in stabilization of catalytic triad [20].

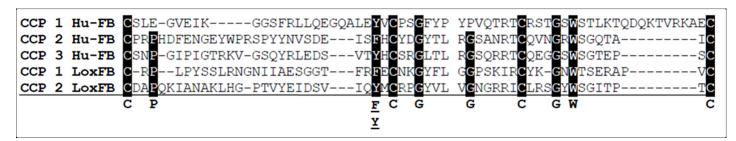


Fig 3. Alignment of Hu-FB and Lox-FB CCPs. The five CCPs are aligned to each other with the consensus amino acids shown in bold and at the bottom. Two disulfide bonds sustain the CCP domain and are formed between the first and third cysteine, and the second and the fourth cysteine. The alignment was done with CLUSTAL W using Bioedit v. 7.0.9.0 software.

	↓ /vWFA **	✓ Mg <sup>2+</sup> Binding site	
Lox-FB	YYLDSASRRRGLDLVILVDRSS	SVDPFDLNVAKKFIKYLLQEFGV-KNSNSNQMGTRAAVIAFGTTIDIIFNLNDTHIDSPDAAGVALDELLK	PNGGGTAMVPALTSVFLKVQPFLRERSKHALFLISDGEPNIGG 306
		SIDPVHLEDAKNFVKFLLRRFGV-NNKPNNNNGTRASVLAFGTEVQIVFNIDDTNISNPRIAAAAVDD-IH	
		SVGSYYLRKSIEFAKAIVRKIGI-SRDGSRASAVVFSSKAETVFYPMQIETEEEVLKYLDN-IT	
		SVGQHNFDKSIEFAKAIVKRVGI-SEAGARAGALIFGSKSENMFLPLSYTTTEEVLDALDK-IN	
		SIGRKYFNSAIKFAKGLVTRMGV-KEFGTRFGAVSFSSTVSASFLPQDYTTEEEVLNGLDQ-FD	
		SVGEKYFYSAIKFAKALVKRMGV-KEQGTRFGAVSFSSTVSNSFLPQDYKTVEEVHSALDK-FN	
Rd-FB Ny FB	SRLSPGKSGLDVVLLVDVSS	SIGDRSMESAKKFMKLLVDIFGVSNETSGGKNGTRFALLTFSNEADIVFNLNDGTARSKEEVKRRIDE-IQ SVGEDNFRKGIQFARTIIDEFGI-SATPSGTRVAVIVFSDAAQVIFNLKSNRIVDKEEAVRRLEN-LQ	
	GPI SPGS3GLDL VFVFD55A	SVGLDMERKGIQEARTIIDEFGI-SATESGTRVAVIVESDAAQVIENDKSNRIVDKLEAVKKLEN-DQ SIDAVDENQAIQETRSIINEFGV-TNKEGGTQVALVIFGSQAQLEWNLGQLDSKRKVERQLRQ-LQ	
	RRKRTIKLSDGMDIYFAFDASN	SVGLKNFEIGKTFAKOLVGKLOV-NTSPGGTRVGAVSYSSEARRLENVNDFTSTVDVVKAIEANVN	
		SVGEDNFNMAKKFAKELVKEIGV-TDRPNSLRVGALVFNSEAEIGFHTVAFDSTADVLDAIDS-ME	
		SIGASNFTGAKKCLVNLIEKVASYGV-KPRYGLVTYATYPKIWVKVSEADSSNADWVTKQLNE-IN	
Eu-C2	GRKIQIQRSGHLNLYLLLDCSQ	SVSENDFLIFKESASLMVDRIFSFEI-NVSVAIITFASEPKVLMSVLNDNSRDMTEVISSLEN-AN	
		*	/SP * _
	DDITPEDISRQLKD	HAA-FEIFAVGIGP-DVQRKTLASVASEPVLHHVFMLDKFTDLEDMMAIMKNRPTEEPPVSLDRCG	
	NPEMEAKILK	SPNDFEIFTVGIGK-GIKMNLLNELASEPPLSBVFILENYPDLNEVMKIIEDSKPPPPPISKDQOG ERN-IDIYCIGITG-DPRLETLYKIASTSKYGNVERSNVFILONYATLSHLIOEITNGTLDFSEOG	
	SPEVEADLLK		
	DPKOVAKELKA		
	RLPRIRQAANDLKN		
Nv FB	SPDRPAKVLRA	GFN-FEIFAIGVSD-SVDKDELKSIASEPFRTHVYQIKDYATLVKLKELITTKGTDYDECG	VAGDTQLRDS-SDKRFRIVGGREAKAGAWPWLAAIYVKGSFRCGGALIARN 491
		EEA-FEVFAVGVGA-NIDKNELNSVASQPFTSHVFLINDFSNFDTLVNTIAEKEIDYEQCG	
		DAA-LKIHCIGISR-NTDKTALAEIASPPVSEHVFYLSDYNELERAVEAITSTNRSYEECG	
		DQD-VTIHCIGISE-NATRRTLSGMASEPLSEHLFFLKDYSTLEEFIQIVTNQTIDYSECG	
Eu-FB Eu-C2	DPITVIDEIRDLLYIGKDR	<pre>KNPREDYLDVYVFGVGP-LVNQVNINALASKKDNEQHVFKVKDMENLEDVFYQMIDESQSLSLOG RNDYLDIYAIGVGKLDVDWRELNELGSKKDGERHAFILQDTKALKQVFEHMLDVSKLTDTICG</pre>	MVWEHRKGTDYHKQPWQAKISVIRPSKGHES-CMGAVVSEY 519
Hu-C2	# #₩* Protease activ		VGNMSANASDCRGALISDQ 500
Lox-FB		Ve site Protease active site V NIEEVFVVMGEYDLMGEDPKQRNFYASKVVIHPDYDESDTLLNNLALVKLHKPVPLETF-KP	# ICLPPTDRTVPLFLDHNINASVAGWGISAPKNGFETISPSATRSOLT 554
		DMKDVFVVLGERHLLKAEQGQTNFYVTDV0IH0NYNPKONSIENNLAVLKLNLPASRYRP	
	FILTAAECMFRKDKRRQILHLR	PKEEIIVKIGLFSL-NDESKVQEFGVQRIFIHEKYDPGNYSVQLYDYDIAILELDGSIVYDRRIRP	ICLPPRDLAENSHLYSFKEFGWAGGWGHNGVLIPYNPYSLSMQLNNSPNLK 746
	FILTAAECMYHREGKKR	SKKDIIVKLGLTDV-KNETYVQESEVSEMFIHPDYRPAGSYDYDIALLLLDKPIEYNPFVRP	
Tt-FB1	FILTAAHCMYHREGKKR WILTAAHCVQNKDPQRKKNQNL	SKKDI IVKLGLTDV-KNETYVQESEVSEMFIHPDYRPAGSYDYDIALLLLDKPIEYNPFVRP VPADI IVKLGVLNV-LNSSDLEEFEVAE IHRNENYNFTTYDHDIALLKLDRPVTYKPFVRP	ICLPPTELPENTPLYSSDEFGWATGWGHEGVVSAAVNERLKSSQILK 899 ICLPPFNVPENSILYKPGQSAFATGWGYDQRVAIDEIVPFKRVDQLK 791
Tt-FB1 Tt-FB2	FILTAAHCMYHREGKKR WILTAAHCVQNKDPQRKKNQNL WILTAAHCFVLKNDDPKKVEYL	SKKDIVKLGLTDV-KNETYVÖESEVŠEMFTH PYRPAG3YDYDIALLLLDKPIGYNPFVRP VFADIVKLGVLNV-LNSSDLEEFEVAEIHRNENYNFTTVDHDIALLKLDRPVTYNFFVRP VFANVTVKLGLLNV-RNSBUKEFEVTDISLHEKKYYTTYDHDIALLKLORPITYEFTRP	ICLPETELPENTELYSSDEFGWATGMGHEGVVSAAVNERLSSQILK 899 ICLPEFNVPENSILYKPGQ&AFATGMGYDQXVALDELVFFKRVDQLK 791 VCLPFAVIPENSILYQAGQ&AFVTGMGSDKRVELGHEGVLKGIDHLK 874
Tt-FB1 Tt-FB2 Rd-FB	FILTAAECMYHREGKKR WILTAAECVQNKDPQRKKNQNL WILTAAECFVLKNDDPKKVEYL WFLTTASCVSKKEGDRVVPY	SKKDIIVKLGLTDV-KNETYVQESEVSEMFIHPDYRPAGSYDYDIALLLLDKPIEYNPFVRP VFADIIVKLGVLNV-LNSSDLEFEFVAEIKENENYNFTIYDHDIALLKLDRFVIYNFFVRP VFANVTVKLGLINV-RNSSDLKEFEVDIRLHEKKNYTIYDHDIALLKLGRFIYEFFIRP NTSMLNVFLSDFNIFEKDTDEIWPVVQDILIHDNYTGEEIGDLGIQENDVALLDLGKNETERFIFDKYLRP	ICLPFTELPENTFLYSSDEFGWATGWGHEGVVSAAVNERLKSSQILK 899 ICLPFFNVPENSILYKPGQ&FATGWGYDQRVAIDEIVFFKRVDQLK 791 VCLPFAVIPPINTIVQAGQ&FYTGWGNAPDRVELGHEGVLKGIDHLK 874 ICVAADNVSLSLAKNLLKMLLSSFNQYDTYTAGWGNAPDNFEHMDILEEKLM 589
Tt-FB1 Tt-FB2 Rd-FB Nv FB	FILTAAECMYHREGKKR WILTAAECVQNKDPQRKKNQNL WILTAAECFVLKNDDPKKVEYL WFLTTASCVSKKEGDRVVPY WVVTAAECFYYDGKI	SKKDIVKLGLIDV-KNEIVVÖSSUVŠENFIH DVRPAGSVDVDIALLLLOMPIEVNPVAR VPADIVKLGVINV-LNSSDLEPEVAEIHRNENYNPTTYDHDIALLKLGRPUTYNPVAR VFANVTVKLGLINV-RNSSDLKEPEVTDIRLHEKKYYTTYDHDIALLKLGRPITYERFIRP NTSMLNVFLGENNIERKOTDEINFVVQDILHENYTGEEIGDLGIQENDVALLDLGGNETREPITPKYLRP VSSDLVRLGEHSRILEGSEQNUVASNUVLHPIANKNGLDEVALLDLGGNETREPITPKYLRP	ICLPFTELPENTFLYSSDEFGMATGMGGEGVVSAAVMERLSSQILK 899 ICLPFPNVPENSILYKPGQSAFATGMGYDQRVALDEIVPFKRUVQLK 791 ICLPFANIPENSILYQAGQAFVIGMGSDRNVELGHGVLKGIDELK 874 ICVAADNVSLSLANLLMMLSSENQYDTYTAGMGSAPDNFEHMDILEEKLM 589 ICLPGPTDA
Tt-FB1 Tt-FB2 Rd-FB Nv_FB Bb-FB	FILTAAHCMYHREGKKR WILTAAHCVQNKDPQRKKNQNL WILTAAHCCVLKNDDPKKVEYL WFLTASCVSKKEGDRVVPY WVVTAAHCFYYDGKI WVLTAAHCLERRGERI	SKKDIIVKLGLTDV-KNETYVQESEVŠEMFIHPDYRPAGSYDYDIALLLLDKPIEYNPFVRP VEADIIVKLGVLNV-LNSSDLEEFEVAEIKRNENYNFTIYDHDIALLKLDRPITYNPFVRP VEANUTVKLGLLNV-RNSSDLEEFEVDEIKHEKKNYTIYDHDIALLKLORPITYEFTRP NTSMLNVFLSDFNIFEKDTDEIKFVVQDILHENYTGEEIGDLGIQZNDVALLDLGKNETERFIFDKYLRP VFSDILVRLGEHRRKEEGTEQUVVQDILHENYTKNGLDFDVALLDLGKNETERFIFDKYLRP JSDOLVVAGEHRRKEEGTEQUVVQUYHVPEYRFRGLDFDVALLDLGKNETERFIFDKYLRP	ICLPPTELPENTPLYSSDEFGWATGWGHEGVVSAAVNERLKSSQILK 899 ICLPPTNVPENSILYKRGQ&AFATGWGYDQXVSIDEIVPFKRVDQLK 791 ICUPPANPENSILYKRGQ&AFATGWGYDQNDELGHGVLKGIDHK 874 ICVAADNVSISLAKNLLKMLISSPNQYDTYTAGWGXAPDNFEHKDILEEKLM 589 VCLPQPTDAIVVRGSVGIVAGWGSTQKGDASVRSGPPYFVLK 610 ICLPKGDLRDSYQTRAGSTMLFAGWGATDPVRGEVFGIPTUFQ 660
Tt-FB1 Tt-FB2 Rd-FB Nv_FB Bb-FB Sp_FB	FILTAAHCMYHREGKKR WILTAAHCVQNKDPQRKKNQNL WILTAAHCCVLKNDDPKKVEYL WFLTASCVSKKEGDRVVPY WVVTAAHCFYYDGKI WVLTAAHCLERRGERI	SKKDIIVKLGLTDV-KNETYVQESEVSEMFIHPDYRPAGSYDYDIALLLLDKPIEYNPFVRP VFADIIVKLGVINV-LNSSDLEFEFVAEIKRNENYNFTIYDHDIALLKLDRFVIYKPFVRP VFANVTVKLGLINV-RNSSDLKEFEVDIALHEKKNYTIYDHDIALLKLGRPIIYEFFIRP NTSMLNVFLSDFNIFEKDTDEIWPVVQDILIHDNYIGEEIGDLGIQENDVALLDLGKNETERPIFDKYLRP VFSDLIVRLGEHBRRLEGSEQUVARASHLVLHFLANKNGLDFDVALIQLKGGVKLTAYVAT SPDQLYVVAGEHARDKEEGTEQVVVVQEYHVPFYRFDRLDYDIALLKLSTPSQLGFVRT SQNGTTVYLGLTHR-WDLNRFSVRCEGIDYAFGLIQGLGGEHNDIALLKLSTPSQLGFVRT	ICLPFTELPENTFLYSSDEFGWATGWGHEGVVSAAVNERLKSSQILK 899 ICLPFFNVPENSILYKRGQ&AFATGWGYDQRVAIDEIVFFKRVDQLK 791 VCLPFAVIPENSILYKRGQ&AFYTGWGBKAVEVELGHEGVKGIDHK 874 ICVAADNVSLSLAKNLLKMLLSSPNQYDTYTAGWGNAPDNFEHMDILEEKLM 589 VCLPGYDDADSYQTRAGSTMLFAGWGSTQKGDASVRSGFPYFVLK 610 LCLPKGDLRDSYQTRAGSTMLFAGWGRTDFVRFGEVFGIPIVFQ 660 VCLPFSDPQCKNWYVNPRRTAFVTGWGRTLKGQTSFALM 743
Tt-FB1 Tt-FB2 Rd-FB Nv FB Bb-FB Sp_FB Aj FB	FILTRABCMYRREGKKR WILTRABCVQNKDPQRKKNQNL WILTRABCVQNKBCDRVVPY WVVTRABCFYDGK	SKKDIIVKLGLTDV-KNETYVQESEVSEMFIHPDYRPAGSYDYDIALLLLDKPIEYNPFVRP VFADIIVKLGVINV-LNSSDLEFEFVAEIKRNENYNFTIYDHDIALLKLDRFVIYKPFVRP VFANVTVKLGLINV-RNSSDLKEFEVDIRLHEKKNYTIYDHDIALLKLGRPITYEFFIRP NTSMLNVFLSDFNIFEKDTDEIWPVVQDILIHDNYTGEEIGDLGIQENDVALLDLGKNETERPIFDKYLRP VFSDLIVRLGEHRRLEEGSEQNVRASNLVLHPLANKNGLDFDVALIQLKGGVKLTAYVRT SPDQLYVVAGEHARDKEEGTEQYVVVQEYHVPEYRPDRLDFDVALIQLKGSVKLTAYVRT SSNGTTVYLGLTHR-VNDLNRFSVXCEGIDYAFGLIQGLDGGEHNDIALLKLSFRSQLSPVRT	ICLPETELPENTFLYSBOEFGWATGWGEHEGVVSAAVNEPRKSSQILK 899 ICLPEFNVPENSILYKBGQSAFATGWGYDQRVALDEIVFFKRVDQLK 791 ICUPADNVSLSLANLLMALSSPNQYDTYTAGWGRAPDNFEHNDILEEKLM 589 ICLPGYDA
Tt-FB1 Tt-FB2 Rd-FB Nv FB Bb-FB Sp_FB Aj_FB Eu-FB	FILTAABC(WYREGKK)	SIKHDIIVKLGLIDV-KNETYVÖSSUVŠENTIH PDYRPAGSVDYDIALLLLOKPIGVPFVPP VPADIIVKLGULNV-KNETYVÖSSUKEPEVDAEIHRNENYNPTTYDHDIALLKLORPITYNFPVPP VPANVTVKLGLLNV-RNSSUKEPEVDILHENKYYTTYDHDIALLKLORPITYEPFIRP NTSMLNVFLSDRNIFEKOTDEINFVVQDILHENYTGEBIGDGIGENDVALLDIGNNTERFIFDYLRP VYSSULVALGEHDRTLEBGSEQUVASULVHPLANKNGLDFDVALIGLKGVKLGFVARI SROGTVYLGEHDRTLEBGSEQUVAVGEYHVPPYRPDRLDFDVALLDLGNSTERLGFVARI SQNGTVYLGLITRVDLNFPAVRCGIDYAPGLQGLDGGENDIALLKLSTPSQLGFVARI SQNGTVYLGLITR-VDLNFPAVRCGIDYAPGL	ICLPFTELPENTFLYSSDEFGMATGMGEGUVSAAVNEPLKSSQILK 899 ICLEPFNVPENSILYKEGQ&AFVTGMGEDGNVALDEIVFKRVDQLK 791 ICLEPANV9LSLANLLMALLSPNQYDTYTAGMGENRVELGHGVLKGIDHLR 874 ICUAADNV9LSLANLLMALLSPNQYDTYTAGMGENRVFELGHGVLKGIDHLR 874 ICUAEDPTDAIUTR9GSVCVAGMGSTORGDSYBGFPYFVLK 610 ICLEPKDGLRDSYQTRAGSTMLFAGMGETDVREGUVFGIPIUFQ 660 ICLEPSDA
Tt-FB1 Tt-FB2 Rd-FB Nv FB Bb-FB Sp FB Aj FB Eu-FB Eu-C2	FILTAABC(WYREGKK)	SIKHDIIVKLGLIDV-KNETYVÖSSUVŠENTIH PDYRPAGSVDYDIALLLLOKPIGVPFVPP VPADIIVKLGULNV-KNETYVÖSSUKEPEVDAEIHRNENYNPTTYDHDIALLKLORPITYNFPVPP VPANVTVKLGLLNV-RNSSUKEPEVDILHENKYYTTYDHDIALLKLORPITYEPFIRP NTSMLNVFLSDRNIFEKOTDEINFVVQDILHENYTGEBIGDGIGENDVALLDIGNNTERFIFDYLRP VYSSULVALGEHDRTLEBGSEQUVASULVHPLANKNGLDFDVALIGLKGVKLGFVARI SROGTVYLGEHDRTLEBGSEQUVAVGEYHVPPYRPDRLDFDVALLDLGNSTERLGFVARI SQNGTVYLGLITRVDLNFPAVRCGIDYAPGLQGLDGGENDIALLKLSTPSQLGFVARI SQNGTVYLGLITR-VDLNFPAVRCGIDYAPGL	ICLPFTELPSNTPFI/SSDEFGMATGMGHEGVVSAAVNEPLKSSQILK 899 ICLPFNVPENSILYKEGGAFATGMGYDQRVALDEIVFFKRVDQLK 791 ICLPFANPENSILYQAGGAFVTGMGKDRVVELGHGVINGIDHLK 814 ICUAADNVSISLANLLMALSSPMQYDYTYAGMGKAPDNFERNDILEEKLM 589 UCLPGPTDALVRPGSVCVAGMGSTQRVGDASVBSGPFYVLK 610 ICLPKGDLRDSYQTRAGSTMLFAGMGKTDPVRFGUFGIPTIVFQ 660 UCLPSDDQCVSWFVTGMGHIKGCTSBLM 743 ACLPDPAKERULTVG-AAVVIGMGHNGVVEAGADRSELIPEPHLQ 805 ICLPCTEGTTRALRIPFTTCQQNEBLIFAQDINALFVSBEENKIT 641 ICLPCTEGTTRALRIPFTTCRDENLINNGSVFHFVLANGSKIN 626
Tt-FB1 Tt-FB2 Rd-FB Nv FB Bb-FB Sp FB Aj FB Eu-FB Eu-C2 Lox-FB	FILTAABC(WYRRGKK) WILTAABC(VNKDQRKKNOYL WILTAABCVNKDGDRVVE WUTAABCLERRGER WUTAABCLERRGER WILTAABCLERRGER FVLTAABCLHGN FVLTAABCHGRUGKDH WUTAABCFRGKDH * VRNVILAEELEC	SIKHDIIVKLGLIDV-KNEIVVÖSSUVŠENFIH DYRPAGSVDYDIALLLLOKPIEVPFVPP VPADIIVKLGVINV-LNSSDLEEFEVAEIHRNENYNFTTYDHDIALLKLORPITYNFPVPP VFANVTVKLGULNV-RNSSDLKEFEVIDIRLEKKPVYTTYDHDIALLKLORPITYRFVRP NTSMLNVFLSDENIFEKOTDEINFVVQDILHENYTGEEIGDGIGENDGALLDIGGNTETRFI VSSDLIVALGEHORTLEEGSEQUVINQUILHENANNGGLDFDVALIQUGGNTLIAVVRT SPOQIVVVAGEHARRVEGEGQUVINVUHUHFLANKNGGLDFDVALIQUGGVKLGFVART SPOQITVVAGEHARRVEGEGQUVINVEYNVPEYRPORLDFDVALIQUKGGVKLGFVART SPOQITVVAGUNRFDGQRKLQVIQTIHPPVDDDTKSHDIALIQURFVMLGFVART LIFDVVAGUNRFDGQRKLQVIQTIHPPVDDDTKSHDIALIQURFVMNFHIRK SLRVNVADSCHRDLEIEVVLFHPVNINKKEAGIPFYDVDVALINLKNKKKVGQTIRP -SLRVNVADPKSQRKEFLIEKUISGEFUVFAKKNQGIEFYGDDIALLKLAKVKKMSTHARP RRINRRINFIFTAEISG-TINNCGRESPIIANDPFTGINVUGLŠORNKC	ICLPFTLPSALVNEPLNSSDEFGWATGWGHEGVVSALVNEPLNSSQLIK 899 ICLPFNVPENSILYKEGG&FATGWGYDQRVALDEIVPFKRVDQLK 791 ICLPFNVPENSILYQAGG&FVTGWGXDRNVELGHEGVLNGIDHLK 814 ICVAADNVSLSLANLLMALSSENQYDTYTAGWGXAPDNFEHNDILEEKLM 584 ICVAEQFTDA
Tt-FB1 Tt-FB2 Rd-FB Nv FB Bb-FB Sp FB Aj_FB Hu-FB Hu-FB Eu-C2 Lox-FB Ead-FB3	FILFABC(WYREGKKR WILFABCV(NKDGQKKNOYL WFLFABCV(NKDGDKVEY WVLFABCUSKKGGDKVFY WVJFABCJERRGER	SKKDITVKLGLTDV-KNETVYÖZSEVŠEMFIH PDYRPAGSVDYDIALLLDKPIGVNPFVAP VFADITVKLGVINV-LNSSDLEEFEVADIHRNENYNFTTYDHDIALLKDRPVTVNPFVAP VFANUTVKLGLINV-RNSSDLEEFEVADIHRNENYNFTTYDHDIALLKDRPTTYNPFVAP WTSMLNVFLBDFNIFEKDTDEIHFVYQDILHENYTGEEIGDLGIQENDVALLDLGKNTERFFIFDKILAP VSBDLYFLGEHDRTLEEGESQUIVASULVHPLANKGGLDEPONLIDLGKGVKLIAYVAT SEDQLYVVAGEHARDKEEGEQUVYVQUYHVPEYRPDRLDEPONLIDLKLSTPSQLGPFVRT SQNGTTVYLGLHR-VNDLNRSSVRCEGIDYAFGLLQGLGGEHNDIALLKDFRSQLGPFVRT -LTFDVFAGUNDRFDGQRKLQVIGTIHPDYDDDTSBDIALIGLREFVKLSPFVRT SGKRDLDIEVVLHFNNNKKKAGIFFFVDYDVALIKLKNKKK	ICLPFTELPSANVEPELNSQLFGWATGWGHEGVVSANVEPELNSQLIK 899 ICLPFTNVPENSILYKPGQSAFATGWGYDQRVALDEIVFFKRUDQLK 791 ICLPFAUPENSILYQAGQSAFVTGWGKDQRVALDEIVFFKRUDQLK 791 ICUPADNVSLSLANLLMALSSPNQYDTYTAGWGKAPDNFEHNDLEEKLM 589 VCLPGYDA
Tt-FB1 Tt-FB2 Rd-FB Bb-FB Sp_FB Aj_FB Hu-FB Hu-C2 Lox-FB Had-FB3 Had-FB3	FILTAAECMYHREGKKR WILTAAECVONKDGQKKKQUL WFLATASCUSKNEGDAVVPY WVVTAAECFYVDGKU WVLTAAECFSGENTL WILTAAECFSGENTL WILTAAECFROGNDH	SKKDIIVKLGLTDV-KNETYVQESEVŠEMFIHPDYRPAGSYDYDIALLLLDKPIEYNPFVRP VEADIIVKLGVINV-LNSSDLEEFEVAEIKANENVNFTIYDHDIALLKLDKPIEYNPFVRP VFANUTVKLGUINV-RNSDLEFEVDISLKEFEVTDISLKEKKIYTHOYDHDIALLKLORPTIYEFTRP NTSMLNVFLSDFNIFEKDTDEINFVVQDILHENVTGEEIGDLGIQENDVALDLGKNTTERFIFDKYLRP SPDILVVLAEHARDKEEGTEQUVVQUEYNVPFYRPDRLDYDIALLKLSTF3QLGFVRT SQNGTTVYLGLTRR-VNDLNRPSVRCEGIDVAFGLQCLDGEKNDIALLKLDREAELSFVRT LTFDVAAGUNDRF-DGQRKKLQUIQTIHPDYDDDTKBHDIALGLKREVKLSFVRT LTFDVAAGUNDRF-DGQRKKLQUIQTIHPDYDDDTKBHDIALGLREVKLSFVRT SLNKVNVGDDK3QKKLQUIQTIHPDYDDDTKBHDIALGLKREVKKSTRAF 	ICLPFTELPENTFLYSSDEFGWATGWGHEGVVSAAVNERLKSSQILK 899 ICLPFTNVPENSILYKBGQAFATGWGHEGVVAAIDEIVFFKRUDQLK 791 ICLPFNVPENSILYKBGQAFATGWGKDQNVFERKDILEKLW 874 ICVAADNVSLSLAKNLLKMLISSPNQYDTYTAGWGKAPNFERKDILEEKLM 589 VCLPQPTDA
Tt-FB1 Tt-FB2 Rd-FB Bb-FB Sp_FB Aj_FB Eu-FB Eu-FB Eu-C2 Lox-FB Ead-FB3 Ead-FB3	FILTAABC(WYREGKK)	SKKDIVKLGLTDV-KNETVVÖSSUŠENTEHPDYRPAGSVDVDIALLLLDKPIGVPFVPP VFADIVKLGVINV-LNSSDLEEFEVAEIHRNENYNPTTVDHDIALLLKDRPVTVPFVPP VFANUTVKLGUINV-RNSDLKEFEVTDILHENKYNTTTVDHDIALLKGRFTTVPFVRP VTSMUTVKLGEHNRTLEEGESQUVASSULVHEFLANKNGDEPONALLDLGKNETERFI SVOGTVVJAGEHRRNEEGESQUVASULVHFLANNGGDEPONALLDLGKNETERLGFVAF SVOGTVVJAGEHRRVEDGESQUVASULVHFLANPGRDEPONALLGKAGVKLFAVVAF SVOGTVVJAGENARDKEEGESQUVVQEYHVPPEYRPGRDEPONALLGLKSTPSOLGFVAF 	ICLPFTELPSNATUREJESSDEFGWATGWGEHEGVVSAAVNEPENSSQLIK 899 ICLPFTVPENSILYKBGQSAFATGWGYDQRVALDEIVFFKRVDQLK 791 ICLPFAUPENSILYQAGQSAFVTGWGKADDNFEHNDILZEKIM 584 ICUAADNVSISLANULIMILSSPNQYDTYTAGWGKADDNFEHNDILZEKIM 584 VCLPGPTDA
Tt-FB1 Tt-FB2 Rd-FB Bb-FB Sp_FB Aj_FB Eu-FB Eu-FB Eu-C2 Lox-FB Ead-FB3 Ead-FB3	FILTAAECMYHREGKKR WILTAAECVNKDGQKNKOUL WILTAAECVLKNDGKKVEL WVITAAECERGERL WVITAAECERGERL WILTAAECERGERL WILTAAECFNGNDE	SKKDIIVKLGLTDV-KNETYVQESEVSEMFIHPDYRPAGSYDYDIALLLLDKPIEYNPFVRP VFADIIVKLGVLNV-LNSSDLEFEVDISLKENKVYTTYDHDIALLKLDRPITYNPFVRP VFANUTVKLGULNV-RNSDLEFEVDISLKENKVYTTYDHDIALLKLGPITYEFTRP NTSMLNVFLBDFNIFENDTDEINFVVQDILHENNYGEEIGDLGIQENDVALLDLGKNTERFYIFDKYLRP VFSDLIVLGEHRRFLEGESGEQUVASUULHELANKIGLDYDIALLKLSTPSQLGFFVRT SQNGTTVILGLTHR-VNDLNRPSVRCEGIDYAFGLLQGLGGEHNDIALLKLDPEAELSFTVAT GEKRDLZIEVVLHENANNSHDIALGKRFVKLSFTVAT 	ICLPPTELPENTPLYSBDEFGWATGWGHEGVVSAAVNEPLKSSQILK 899 ICLPPTELPENBILYKBGQ&AFATGWGHEGVVAIDEIVPFKRUDQLK 791 ICLPPTVPENBILYKBGQ&AFATGWGHQDRVAIDEIVPFKRUDQLK 791 ICVADADVSLSLANLLANLISBPNQYDTYTAGWGNAPDNFEHNDILEEKLM 589 VCLPQPTDA
Tt-FB1 Tt-FB2 Rd-FB Nv FB Bb-FB Sp_FB Sp_FB Eu-FB Eu-FB Eu-FB Ead-FB3 Ead-FB1 Tt-FB1	FILTABECWYRREGKK WILTABECVNKDEQRKKNOL WILTABECVNKDEQRKKOL WUTABECENKEEL WUTABECERGER WULTABECERGER	SKKDIVKLGLTDV-KNETVVÖSSUŠENTEHPDYRPAGSVDVDIALLLLDKPIGVPFVPP VFADIVKLGVINV-LNSSDLEEFEVAEIHRNENYNPTTVDHDIALLLKDRPVTVPFVPP VFANUTVKLGUINV-RNSDLKEFEVTDILHENKYNTTTVDHDIALLKGRFTTVPFVRP VTSMUTVKLGEHNRTLEEGESQUVASSULVHEFLANKNGDEPONALLDLGKNETERFI SVOGTVVJAGEHRRNEEGESQUVASULVHFLANNGGDEPONALLDLGKNETERLGFVAF SVOGTVVJAGEHRRVEDGESQUVASULVHFLANPGRDEPONALLGKAGVKLFAVVAF SVOGTVVJAGENARDKEEGESQUVVQEYHVPPEYRPGRDEPONALLGLKSTPSOLGFVAF 	ICLEPETELPENTFLYSSDEFGWATGWGHEGUVSAAVWEPENSSQLIK 899   ICLEPETELPENSTLYKEGGAFATGWGYDQRVALDEIVFFKRUDQLK 791   ICLEPATNPENSTLYGAGGAFATGWGYDQRVBLCHGVINGTDHLK 817   ICUAEDANVSLSLANLLMALSSPWQYDTYTAGWGKAPDNFENNDILEEKLM 589   ICUAEDANVSLSLANLLMALSSPWQYDTYTAGWGKAPDNFENNDILEEKLM 580   ICUAEDANVSLSLANLLMALSSPWQYDTYTAGWGKAPDNFENNDILEEKLM 580   ICUAEQATDA
Tt-FB1 Tt-FB2 Rd-FB Bb-FB Bb-FB Sp_FB Aj_FB Eu-FB Eu-C2 Lox-FB Ead-FB3 Ead-FB3 Ead-FB1 Tt-FB1 Tt-FB2 Rd-FB	FILTAABC(WYREGKKR WILTAABCVQNKDGQKKKQWL WFLDTASCUSKKGGDKVFY WVLTAABCFYDGK WVLTAABCFRGER	SIKHDIVKLGLTDV-KNETVVÖSSUŠENTHEDYRPAGSVDYDIALLLLOKPIGVPFVPP VPADIVKLGVINV-LNSSDLEEFEVAEIHRNENYNFTTYDHDIALLKLORPITYNFVPP VPADIVKLGVINV-LNSSDLEEFEVAEIHRNENYNFTTYDHDIALLKLORPITYNFVPP VPANVTVKLGELNV-RNSSDLKEFEVTDIRLEKKYTTTYDHDIALLKLORPITYNFVPF VPSDLVRLGEHORTLEEGESQUVASULUHEIANNGGLDFDVAILDLGGNTTRFJFDVLAP SPOLTVVLGEHARVNCBINFSVRCEGIVDASULUHEIANQGLOGGENDIALLKLORPITLGFVART SPOLTVVLGUTRFDGQRKKLQVIQTIHPPYDDGTKSHDIALIKLERFSQLGFVART LTFDVAGUDRFDGQRKKLQVIQTIHPPYDDDTKSHDIALIKLERFSQ	ICLEPTELPENTFLYSSDEFGWATGWGHEGUVSAAVNEPLNSSQLIK 899     ICLEPSTNPENSILYKEGGAFVTGWGHEGUVSAAVNEPLNSSQLIK 899     ICLEPSTNPENSILYKEGGAFVTGWGHEGUNSLDEUVEFKRUDQLK 791     ICUEPSANVSLSAKNLIMMLSSENGYPTTGWGHEGUNNVELGHEGVLNGIDHLK 874     ICUADADNVSLSAKNLIMMLSSENGYPTTGWGHEGUNNVELGHEGVLNGIDHLK 874     ICUADADNVSLSAKNLIMMLSSENGYPTTGWGHEGUNNELGHEGVLNGIDHLK 810     ICUEPSDP
Tt-FB1 Tt-FB2 Rd-FB Bb-FB Bb-FB Aj_FB Bu-FB Bu-C2 Lox-FB Had-FB3 Bad-FB2 Ead-FB2 Rd-FB2 Rd-FB Bb-FB	FILTAABC(WYRRGKK)	SKKDIVKLGLIDV-KNETVVÖSSUŠENTHPDYRPAGSVDYDIALLLDKPIGVPFVP VFADIVKLGVINV-LNSSDLEFEVAEIHRNENYNPTTYDHDIALLKDRPVTVPFVP VFANUTVKLGUINV-RNSDLKEFEVTDILHENKYNYTTYDHDIALLKDRPVTVPFVP VFSDUFVLGEHROTLESGESQUVASULVHEFLANKKGDEPOVALIDLGKNETERFY SFDQIVVAGEHROTLESGESQUVASULVHFLANKKGDEPOVALIDLGKNETERF SQNGTTVVLGLTR-VNDLNRSVRCEGIYAPGULVHFUR SUGTVVLGLTRDGQRKKLQVIGTIHPDYDDDTKSHDIALIGLRKPVKLNFVRT 	ICLPFTELPSNATUREJESSDEFGWATGWGHEGUVSAAVNEPLKSSQILK 899 ICLPFTNVPENSIIYKBGQSAFATGWGYDQRVALDEIVFFKRVDQLK 791 ICLPFNVPENSIIYKBGQSAFATGWGYDQRVBLCHGVLKGIDHLK 874 ICUADDNVSISLANLLAMLISSPNQYDTYTAGWGRAPDNFEHRDILEEKLM 589 VCLPGPDDAUNVRGSTLFAGWGRAPDVRPGEVFGIPTIVPG 660 UCLPSDPQNUNTYNPRRTAFVIGWGHILKGRPGEVFGIPTIVPG 660 UCLPSDPQRDLITVG-AAVVTGWGRNGVBAGADSSELIPEPHUQ 805 ICLPCTEGTRALLTVG-AAVVTGWGRNGVBAGADSSELIPEPHUQ 805 ICLPCTEGTRLITVG-AAVVTGWGRNGVBAGADSSELIPEPHUQ 805 ICLPCTEGTRLITVG-AAVVTGWGRNGVBAGADSSELIPEPHUQ 805 ICLPCTEGTRLITVG-AAVVTGWGRNGVBAGADSSELIPEPHUQ 805 ICLPCTEGA
Tt-FB1 Tt-FB2 Rd-FB Bb-FB Sp_FB Aj_FB Hu-FB Hu-FB Hu-C2 Lox-FB Had-FB2 Tt-FB1 Tt-FB1 Tt-FB2 Rd-FB Nv_FB Bb-FB Sp_FB	FILTAAECMYHREGKKR WILTAAECVNKDGQKKKQUL WILTAAECVLKNDGKKVEL WVITAAECSUKKGDRVVFY WVVTAAECLERKGERL WVUTAAECLERKGER	SKKDI IVKLGLTDV-KNETYVÖESEVŠEMFIH PDYRPAGSYDYDIALLLLDKPIEYNPFVRP VRADI IVKLGVLNV-LNSSDLEEFEVAEIRNENYNFTTDHDIALLKLDRPTYNPFVRP VRANUTKLGULNV-RNSDLEFEVTDIALHENKYYTTYDHDIALLKLGRPTTYEFTRP NTSMLNVFLBDFNIFEKDTDEINFVVQDILHENYTGEEIGDLGIQENDVALLDLGKDTERFYFTDYLAP VYSDLIVLGEHRRTLEGESQUWXASULVHPLANNKGLDFDVALIDLGKOTKALTAVVRT SPDQLYVVAGEHARDKEEGTEQUVYVQEYHVPPEYRPDRLDFDVALIDLGKOTKALTAVVRT SPDQLYVVAGEHARDKEEGTEQUVYVQEYHVPPEYRPDRLDFDVALIDLGKOTKALSFTWAT GEKRDLIEUVLFHENNINGKEAGIFFFYDYDVALILLKAKKK	ICLPTELPENTFLY3SDEFGWATGWGHEGVVSAAVNEPLKSSQILK 899 ICLPFTELPENSILYKPGQ&AFATGWGYDQNVAIDELVFFRKNDQLK 791 ICLPFTNVPENSILYKPGQ&FATGWGYDQNVELGHGVKGUDLK 791 ICVADADVSLSLANLLMALLSFPNQYDTYTAGWGADDNFEHNDLEEKLM 589 VCLPGPTDADSYQTRAGSTMLFAGWGATDFVFRENGESFPYLVFK 610 ICLPKGDLRDSYQTRAGSTMLFAGWGATDFVRGEVFGIPTLVFQ 660 VCLPGSDPQNLAUFSTAGVGWGATDFVRGEVFGIPTLVFQ 660 VCLPGSDPQNLAUFSTAGVGWGATDFVRGEVFGIPTLVFQ 660 ICLPCTGGTRLITVG-AAVVTGWGMNGVVEAGADSSELFPFHQ 805 ICLPCTGGTRLITVG-AAVVTGWGMNGVVEAGADSSELFPFHQ 805 ICLPCTGGTRLITVG-AAVVTGWGMNGDU
Tt-FB1 Tt-FB2 Rd-FB Bb-FB Bb-FB Bu-FB Bu-FB Bu-C2 Lox-FB Ed-FB3 Tt-FB2 Rd-FB1 Tt-FB2 Rd-FB Bb-FB Bb-FB Sp_FB Sp_FB	FILTABECWYRREGKK	SIKKDIVKLGLTDV-KNETYVÖSSUŠENTHHDYRPAGSVDYDIALLLLÖKPIEVPFVPP VPADIVKLGVINV-LNSSDLEEFEVAEIHRNENYNFTTYDHDIALLLKÖRPTVPFVPP VPADIVKLGVINV-LNSSDLEEFEVAEIHRNENYNFTTYDHDIALLLKÖRPTTVPFVPP VPANVTVKLGUNV-SDLKEFEVTDIRLHENKYVTTYDHDIALLKÖRPTTVPFVPP VPSDLVPLGEHDRTLEEGESQLVIVASILVHPLANKIGDDFDVALGLKGVKLATVAT SPODIVVAGEHARNEEGFEQVVVVQEYHVPPEYRPGRDUFDIALLKSTPSQGFVRT- SQNGTTVVLGLTRE-VNDLMPBAVRCGIDVAFGL	ICLPFTELPSALVNEPLYSSDEFGWATGWGHEGUVSALVNEPLYSSQILK 899 ICLPFTUPENSILYKBGQ&FATGWGYDQRVALDEIVFFKRVDQLK 791 ICLPFNVPENSILYQAGQ&FVTGWGKDQRVELGHGVLKKIDHLK 814 ICUADNVSLSLANLLMALSSPWQYDTYTAGWGKAPDNFEHNDILEEKLW 588 ICUPCGDPA
Tt-FB1 Tt-FB2 Rd-FB Bb-FB Bb-FB Bu-FB Hu-FB Hu-FB Had-FB3 Had-FB3 Tt-FB1 Tt-FB1 Tt-FB1 Tt-FB1 Tt-FB1 Sb-FB Sb-FB Sb-FB Su-FB Su-FB	FILTAABC(WYREGKKR WILTAABCVQNKDGQKKKQWL WFLTASCUSKKGGDRUFSY WULTAABCFYDCK WULTAABCERGER	SKKDIVKLGLIDV-KNETYVÖSSUŠENTEHPYRPAGSVDYDIALLLDKPIGVPFVRP VFADIVKLGVINV-LNSSDLEFEVAEIHRNENYNPTTYDHDIALLLKDRPVTVPFVRP VFADUTVKLGULNV-RNSSDLEFEVDEIHRNENYNPTTYDHDIALLKDRPVTVPFVRP VFANUTVKLGULNV-RNSSDLEFEVDILHENYNVTTYDHDIALLKDRPVTVPFVRP VFSDULVLGEHROFLEGESQUWASSULVHEFLANKNGDEPVALIDLGKOFKLIAVVRT SPDULVWAGEHARDKEEGFEQYVYVQEYHVPFYRPDRLDFDVALIDLGKOFKLIAVVRT SPDULVWAGEHARDKEEGFEQYVYVQEYHVPFYRPDRLDFDVALIDLGKOFKLIAVVRT SPDULVWAGEHARDKEEGFEQYVYVQEYHVPFYRPDRGLDGGEHNDIALLKDREAELBFVRT SUNGTVYLGLYRVNDLNRFSVRCEGIDYAGGLQGLDGGEHNDIALLKLSTPSQLGFVRT 	ICLPETELPSNATUREJESSDEFGWATGMGEHEGUVSAAVNEPELSSQILK 899 ICLPEPENVPENSILYKBGQSAFATGMGYDQRVALDEIVPFKRVDQLK 791 ICLPEPNVPENSILYKBGQSAFATGMGYDQRVBLCHGOVLKGIDHLK 814 ICUAPADNVSLSLANLLMALSSENQYDTYTAGMGRAPDNFEHNDILEEKIM 589 VCLPEPDDA
Tt-FB1 Tt-FB2 Rd-FB Nv FB Bb-FB Sp FB Aj_FB Hu-FB Hu-FB Had-FB2 Had-FB2 Had-FB2 Had-FB2 Had-FB2 Bad-FB3 Tt-FB1 Tt-FB2 Rd-FB Sp FB Sp FB	FILTAABC(WYREGKKR WILTAABCVQNKDGQKKKQWL WFLTASCUSKKGGDRUFSY WULTAABCFYDCK WULTAABCERGER	SIKKDIVKLGLTDV-KNETYVÖSSUŠENTHHDYRPAGSVDYDIALLLLÖKPIEVPFVPP VPADIVKLGVINV-LNSSDLEEFEVAEIHRNENYNFTTYDHDIALLLKÖRPTVPFVPP VPADIVKLGVINV-LNSSDLEEFEVAEIHRNENYNFTTYDHDIALLLKÖRPTTVPFVPP VPANVTVKLGUNV-SDLKEFEVTDIRLHENKYVTTYDHDIALLKÖRPTTVPFVPP VPSDLVPLGEHDRTLEEGESQLVIVASILVHPLANKIGDDFDVALGLKGVKLATVAT SPODIVVAGEHARNEEGFEQVVVVQEYHVPPEYRPGRDUFDIALLKSTPSQGFVRT- SQNGTTVVLGLTRE-VNDLMPBAVRCGIDVAFGL	ICLPETELPSNATUREJESSDEFGWATGMGEHEGUVSAAVNEPELSSQILK 899 ICLPEPENVPENSILYKBGQSAFATGMGYDQRVALDEIVPFKRVDQLK 791 ICLPEPNVPENSILYKBGQSAFATGMGYDQRVBLCHGOVLKGIDHLK 814 ICUAPADNVSLSLANLLMALSSENQYDTYTAGMGRAPDNFEHNDILEEKIM 589 VCLPEPDDA

**Fig 4. Multiple alignment of the vWFA and serine protease domains from Lox-FB with sequences of FB/C2 proteins from other organisms.** The alignment was performed using MUSCLE algorithm available in MEGA software. Amino acids that are highlighted in bold indicate identical regions. The amino acids residues that are functionally important at Factor D or C1s cleavage site; the metal ion dependent binding site (MIDAS) and the protease active sites are indicated by dark arrows; the three amino acids residues (T<sup>431</sup>, A<sup>433</sup>, S<sup>616</sup>) that are important on stabilization of catalytic triad are indicated by sign #; the conserved cysteines residues are indicated by asterisks and the two extra cysteines present in Lox-FB, Hd-FB3 and Rd-FB are highlighted in grey. *Loxosceles laeta* (Lox-FB), *Hasarius adansoni* (Hd-FB1; HD-FB2; Hd-FB3), *Tachypleus tridentatus* (Tt-FB1;Tt-FB-2), *Ruditapes decussatus* (Rd-FB), *Nematostella vectensis* (Nv-FB), *Branchiostoma belcheri* (Bb-FB), *Strongylocentrotus purpuratus* (Sp-FB), *Apostichopus japonicus* (Aj-FB), *Homo sapiens* (Hu-C2;Hu-FB).

doi:10.1371/journal.pone.0146992.g004

#### 3D Structure of Lox-FB

The resulting predictive structure of Lox-FB, obtained after computational modeling using the available crystal structure of human factor B (2ok5.1), revealed 23.02% of identity and, despite differences in the number of CCP domains, a remarkable structural similarity between the CCP2-CCP3 domains of human FB and CCP1-CCP2 from Lox-FB was observed (Fig 5). Furthermore, the vWFA domain fits perfectly with human FB, since the six major  $\alpha$ -helices surrounding a central twisted  $\beta$ -sheet are present and conserved at the same positions with minor differences on the conformation of the loop folds. As well as linear alignment, the amino acids that represent the metal ion-dependent binding site (MIDAS) are conserved and occupy the same positions on the 3D structure (Fig 6A). With respect to the serine protease domain (SP), the overlap was not as perfect as that observed for the vWFA domain; nevertheless, the regions that constitute the secondary structures of the  $\beta$  sheet are overlapping (Fig 6B). The three amino acids that comprise the catalytic triad of human factor B (H, D, S) are aligned with the same three amino acids (S, N, P) from Lox-FB observed in the alignment of the primary sequences (Fig 6B).

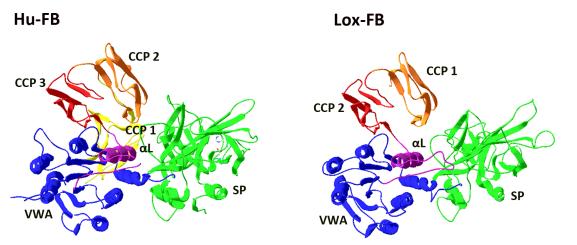


Fig 5. Molecular model of Lox-FB. Construction of molecular model based on structure of human factor B (PDB 20k5.1), using the tool SWISS-Model Workspace available on <a href="http://swissmodel.expasy.org/workspace/">http://swissmodel.expasy.org/workspace/</a>. The further analyses were performed using the software SwissPDBViewer.

doi:10.1371/journal.pone.0146992.g005

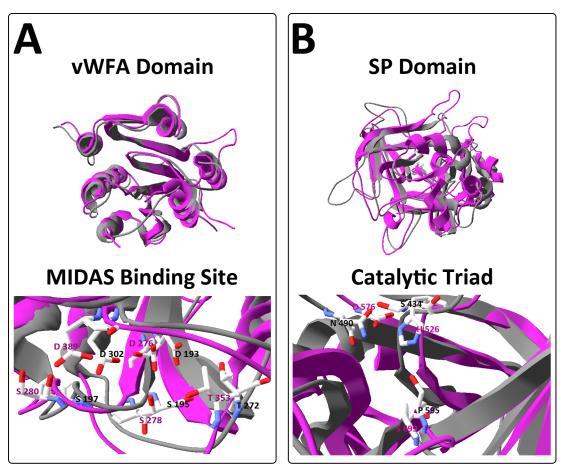


Fig 6. Overlap of human factor B (Hu-FB) (pink) and the Loxosceles factor B-like (Lox-FB) (grey). [A] vWA Domain [B] SP Domain [C] amino acids residues that comprise the catalytic triad of Hu-FB (H, D, S) and Lox-FB (S, N, P). The manipulation of models were performed using the software SwissPDBViewer.

#### Table 2. Pairwise comparisons of Lox-FB vs others FB and C2 proteins.

Organism gene	N° de Access	l (%)*	S (%)**	Score	E value
FB-3 H. adansoni	dbj  BAR45592.1	43	62	483	1e-158
FB-2 Ammothea sp	dbj BAR45607.1	35	52	349	1e-106
FB/C2 B. belcheri	gb ABY28382.1	30	48	264	3e-74
FB precursor N. vectensis	dbj BAH22727.1	30	46	237	1e-64
FB-2 S. subspinipes	dbj BAR45617.1	28	44	231	9e-63
FB precursor N. vectensis	dbj BAH22728.1	28	46	224	2e-59
FB-like R. decussatus	gb ACQ91095.1	27	47	206	7e-54
FB-1 Ammothea sp	dbj BAR45606.1	26	41	185	8e-46
FB-2 A. japonicus	gb AEP68015.1	27	42	180	1e-44
FB A. japonicus	gb ADX36428.1	27	44	180	1e-44
FB-1 H. adansoni	dbj BAR45590.1	25	45	171	2e-41
FB S. purpuratus	gb AAC79682.1	28	42	169	5e-41
FB L. japonicum	dbj BAA02763.1	23	40	154	3e-36
C2/FB-2 T. tridentatus	gb BAM15263.1	24	42	149	3e-34
FB-1 S. subspinipes	dbj BAR45616.1	24	42	142	1e-31
FB Danio rerio	gb AAH97100.1	25	43	140	2e-31
C2/FB-1 T. tridentatus	gb BAM15262.1	22	41	139	5e-31
FB/C2B G. cirratum	gb AAY56127.1	25	41	135	5e-30
C2/FB-2 C. rotundicauda	gb ABK30937.1	23	42	131	1e-28
C2/FB-1 C. rotundicauda	gb AAV65032.2	23	42	131	2e-28
FB/C2A C. carpio	gb BAA34706.1	25	41	126	5e-27
FB/C2A-3 C. carpio	gb BAB32650.1	25	44	125	8e-27
FB-2 precursor M. musculus	gb NP_001136178.1	24	42	125	8e-27
FB-1 precursor M. musculus	gb NP_032224.2	24	42	125	1e-26
C2 H. sapiens	gb AAB67975.1	26	41	119	4e-32
FB H. sapiens	gb CAA51389.1	25	41	121	8e-33

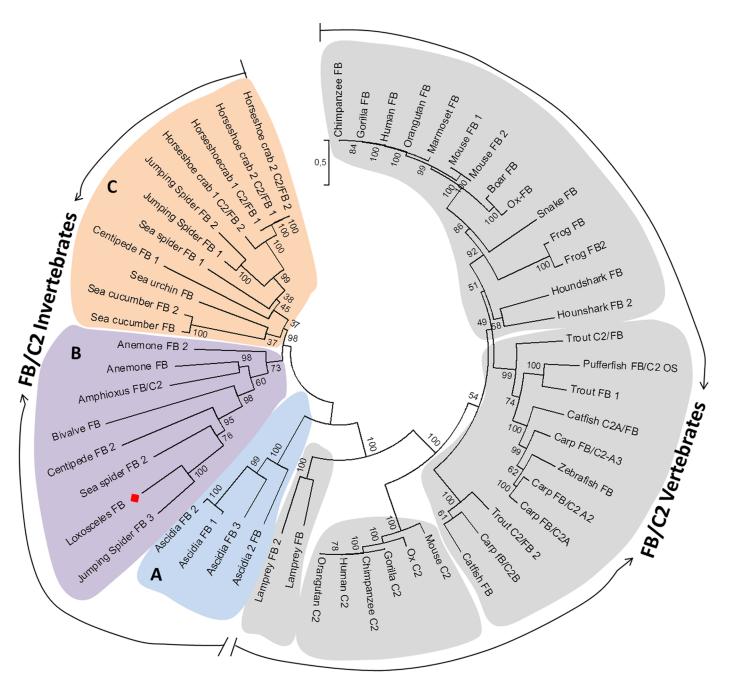
\*Identity (I) was calculated based on percentage of identical amino acids at per column/position in the alignments.

\*\*Similarity (S) was calculated as the percentage of identical plus similar residues, which are conservative substitutions.

doi:10.1371/journal.pone.0146992.t002

#### Phylogenetic analysis

Lox-FB showed similarity with complement proteins sharing 25% and 26% identity with human FB and C2, respectively (Table 2). Values obtained with invertebrates FB/C2 proteins exhibited a similarity ranging from 22% to 43% of identity. To investigate the evolutionary history of FB and C2 complement proteins and to determine how *Loxosceles* spiders fit into this picture, 56 FB and C2 sequences were used and subjected to phylogenetic analysis. An unrooted phylogenetic tree was constructed and resulted in two differentiated groups, vertebrate and invertebrate proteins (Fig 7). Considering the vertebrate proteins, there is a clear separation between C2 and FB proteins, except for some fishes FB/C2 sequences. The lamprey FB (jawless vertebrate) is positioned outside of the jawed vertebrate FB and C2 proteins. Considering the branch represented by invertebrates, there are three main groups: the first one is represented by ascidians FB-like sequences (group A) which are possibly the closest ancestor of FB/C2 vertebrate sequences. The second was called group B, which comprises Lox-FB and FB-like sequences from cnidarians, amphioxus and bivalve; Lox-FB is located at the same branch of FB proteins (isoform 2) as the centipede and sea spider, and at the same sub-branch of the



**Fig 7. Phylogenetic relationships among Lox-FB and reported C2 and FB proteins.** The tree was constructed based on alignment done with MUSCLE and Maximum Likelihood (ML) algorithm using MEGA 6 software. Statistical confidence of the evolutionary analysis was assembled by bootstraps of 1000 replicates. Sequences were obtained from GenBank Database. Gorilla FB (Q864V9), Chimpanzee FB (Q864W0), Human FB (P00751), Orangutan FB (Q864W1), Marmoset FB (JAB17376.1), Mouse FB-1 (P04186), Mouse FB-2 (B8JJM5), Ox-FB (P81187), Boar FB (ABX82825.1), Snake FB (AAR21601.1), Frog FB (BAA06179.1), Frog FB-2 (AAI70192.1), Trout C2/FB (ADY68777.1), Trout FB 1 (AAC83699.1), Pufferfish FB/C2 OS (CAD21938.1), Catfish FB/C2 A (AEW10545.1), zebrafish FB (AAB19093.1), Carp FB/C2 A3 (BAB32650.1), Carp FB/C2 A2 (BAA78416.1), Carp FB/C2 A (BAA34706.1), Houndshark FB (BAB63203.1), Houndshark FB 2 (BAF62177.1), Catfish FB (AEW38662.1), Carp FB/C2 B (BAA34707.1), Trout C2/FB 2 (BAB19788.1), Mouse C2 (P21180), Ox C2 (Q3SYW2), Orangutan C2 (Q8SQ74), Gorilla C2 (Q863A0), Human C2 (P06681), Chimpanzee C2 (Q8SQ74), Lamprey FB (BAA02763.1), Lamprey FB 2 (BAG66069.1), Ascida FB (AAK00631.1), Ciona FB-3 (BAD89301.1), Ciona FB-1 (BAD89299.1), Ciona FB-2 (BAD89300.1), Amphioxus FB/C2 (ABY28382.1), Anemone FB (BAH22726.1), Anemone FB 2 (BAH22728.1), Bivalve FB (ACQ91095.1), Sea cucumber FB 2 (BAF68015.1), Sea urchin FB (AAC79682.1), Horseshoe crab 1 C2/FB 1 (BAM15262.1), Horseshoe crab 1 C2/FB 2 (BAR45590-1), Jumping spider FB3 (BAR45592.1), sea spider FB1 (BAR45606.1), sea spider FB 2 (BAR45607.1), centipede FB 1 (BAR45616.1), centipede FB 2 (BAR45617.1).

jumping spider (isoform 3) (Bootstrap 100%). Finally, group C, consists of FB-like sequences from echinoderms and eight isoforms found in five species belonging to the Arthropoda phylum.

#### Discussion

The studies of the evolution of the complement system have progressed in the last years and due to technological advances, including the genomic and transcriptomic methodologies, resulting in the increased possibility of discovering genes related to the complement system in invertebrates. At the present moment, there is more information about the existence of C3 genes than of *FB* genes, however, it is known that the *FB* gene is missing from genome sequences of cnidarians hydra, the nematode *C. elegans* and several species of insects. These findings suggest that the origin of the *FB* genes probably has occurred before the divergence between Cnidaria and Bilateralia, more than one billion years ago, and the absence of these complement genes in cnidarian and in some protostome lineages can be explained due to a secondary loss [21,2]. The present work confirms this hypothesis, since a *FB*-like gene was also found expressed in *Loxosceles* spider venom gland. This finding represents the first identification of a factor B homologue from a *Loxosceles* spider (Arachnida, Sicaridae) and similar to the vertebrate FB/C2 proteins, Lox-FB is a mosaic protein composed of CCP, vWFA and serine protease domains.

Many studies have indicated the existence of a less complex complement system, named the "archeo-complement system" involving C3-like proteins associated to factor B-like proteins as it was described in lamprey [22], sea urchin [11], horseshoe crab [10], ascidians [23], amphioxus [24], sea anemones [8], bivalve [18] and sea cucumber [25]. Factor B is a component of the alternative pathway and since this gene was found in organisms belong to primitive lineage (cnidarians and protostomes), it is possible that these organisms are endowed with alternative pathway activity.

All invertebrate C2/FB proteins, characterized so far, preserve the same classic architecture domain found in related vertebrate proteins. Despite of some particularities in N-terminal region such as extra CCP domains, observed in sea urchin and horseshoe crab, or the additional domains, as low density lipoprotein receptor like domain (LDL\_A), found in ascidians or epidermal growth factor-like domain (EGF\_CA) in amphioxus, most of all these C2/FB proteins have conserved the regions that play important role for activation of this protein and because of that they are considered orthologues of the mammalian FB and C2.

Highly conserved regions were identified in Lox-FB such as CCP consensus sequence, Mg<sup>2+</sup> binding sites in the vWFA domain and conserved positions near to the catalytic center within the SP domain. However, considering the domain architecture, Lox-FB has only two CCP domains, one less than observed in vertebrates, but is similar to FB-3 present in the jumping spider *Hasarius adansoni*, the centipede *Scolopendra subspinipes japonica* FB-2, the sea spider *Ammothea sp* FB-2 and the bivalve *Ruditapes decussatus* FB [12,18]. The CCP module is a domain commonly present in many mammalian complement proteins and is responsible for the interaction of complement proteins, with each other and their respective regulators and receptors, but also for the binding of e.g. factor H to human cells. Pathogens often mimic or capture CCP-like molecules to avoid detection and destruction by the complement system and can also use CCP-containing molecules to gain entry into the cell (*e.g* EBV binding to CR2) [26]. Not much of the roles of CCPs in pathogen evasion/protection in invertebrates is known, but recently, some studies demonstrated a role for CCPs in preventing lethal flaviviral infection of mosquitoes that are responsible for transmission of e.g. Dengue fever and Yellow fever. The mosquitoes are not affected by these viruses themselves, while they can transmit the disease to

humans. The mosquito *Aedes Aegypti* was found to contain a neural factor AaHig that contained 5 CCP domains and that functions as viral recognition factor interacting with surface proteins of dengue virus (DENV) or Japanese encephalitis virus (JEV), thereby preventing the flaviviral entry into the mosquitos neural cells [27]. Another study showed that a scavenger receptor binds to DENV via CCP modules and indirectly helps to control flavivirus infection by inducing antimicrobial peptides [28]. Thus, even if the main role of CCP domains in invertebrates FB is to interact with C3b fragment and cause pathogen opsonization, the possibility cannot be excluded that they may also function as a pathogen recognition factors in the spider. However, a major difference between the mosquitos and the spiders is that the mosquitos are vectors in the transmission of viruses and other pathogens, and thereby it is essential that they are protected against the pathogen themselves. Such a role for spiders has not been described yet and thus no evolutionary pressure may have been present to evolve such a role for the CCP-containing spider molecules.

Factor B belongs to family S1 of clan PA of peptidases, since it has a serine protease domain that bears the chymotrypsin fold and almost all representatives of this class utilize the canonical catalytic triad represented by Asp<sup>102</sup>, His<sup>57</sup> and Ser<sup>195</sup> (chymotrypsin numbering) [20]. Despite of many similarities in the secondary structures between human FB and Lox-FB, the classical catalytic triad was not found, however, the adjoining amino acids were conserved, but whether Lox-FB has proteolytic activity remains to be investigated. It is possible that the amino acids present at the putative catalytic site may form an active site or this protein corresponds to a Lox-FB inactive isoform. We only found one FB-like molecule in the *Loxosceles* venom gland, but we cannot exclude that the *Loxosceles* spiders have other FB isoforms that possess a conserved triad catalytic in their hemolymph.

Thus far, in invertebrates, there is no clear information about specific proteolytic activity of an alternative pathway (AP) convertase C3bBb-like and how it is assembled. A serine protease activity in horseshoe crab plasma is triggered by PAMP molecules, such as LTA and LPS in Mg<sup>2+</sup> and Ca<sup>2+</sup>-dependent manner, however, it is not clear if CrC2/Bf participates directly in CrC3 activation or if it has to be activated by other serine protease similar to FD vertebrate complement [10]. Le Saux and collaborators demonstrated a new role played by CrC2/Bf that is able to binding to the three members of PRRs: galactose-binding protein (GBP), Carcinolectin-5 (CL5) and C-reactive protein (CRP), promoting their assembly on pathogens and, consequently, activating the complement system [29]. These findings also suggest that CrC2/Bf could function as a MASP counterpart, participating in putative lectin pathway.

Although the horseshoe crab complement system is being studied with great depth considering proteolytic activities, most data derived from invertebrates FB/C2-like is represented by characteristics based on their putative structures. In other words, there is no experimental evidence that those isoforms that did not retain the classic catalytic triad actually had lost their proteolytic activity.

Furthermore, it is possible that in the *Loxosceles* complement system there is another mechanism of activation, independently of factor B, as observed in the horseshoe crab *Tachypleus tridentatus*. It was demonstrated that these organisms are endowed by a serine protease named Factor C, originally characterized as an LPS-sensitive initiator of hemolymph coagulation stored within the hemocytes, and that this factor could act as a C3 convertase on the surface of invading Gram-negative bacteria in the initial phase of complement activation [30]. Perhaps, there is also a component similar to the factor C in the hemolymph or the venom gland from *Loxosceles*, however, the activity of Lox-FB should be evaluated to understand if it has physiological roles in the *Loxosceles* complement system activation. Recently, Tagawa et al. (2012) [31] characterized two isoforms of factor B from *Tachypleus tridentatus* (TtC2/Bf-1 e TtC2-Bf2) and both of them were indispensable for TtC3b deposition on Gram-positive bacteria and fungi. Even though, they have not characterized a factor D-like serine protease in horseshoe crab, and because of this they suggested that other components, such as plasma lectins, which could be important for recruitment of the C3bBb-like on the surfaces of Gram positive bacteria and fungi. Then, it seems that the mechanism of activation of horseshoe crab complement system is different when compared to mammals and maybe the *Loxosceles* spider has the same pattern of activation.

According to phylogenetic analysis, there is a divergence between the proteins present in vertebrates and in invertebrates. At the branch represented by components factor B and C2 from vertebrates, the lamprey FB appears as sister group of vertebrates, indicating that the gene duplication events happened before the origin of jawed vertebrates. This configuration of phylogenetic tree is in agreement with the absence of classical pathway in jawless vertebrates, since they do not have immunoglobulins genes [9]. Almost all sequences from invertebrates have more than one isoform and some of them are grouped at the same clade (group C) as observed for the two isoforms (1 and 2) of the limulus *Tachypleus tridentatus* (horseshoe crab 2 C2/FB), sea cucumber *Apostichopus japonicus* and jumping spider *Hasarius adansoni*. However, in some species that expressed different FB-like isoforms, their isoforms did not locate to the same group; for instance, some isoforms of FB-like sequences from the centipede, sea spider and the jumping spider were grouped together in group B in which Lox-FB is also found, while other isoforms of the same species were grouped in group C.

The phylogenetic history of a protein does not always follow the evolution of the species because of different selective pressures. Along with the type of pathogen that the organism is infected with, other factors as adoption of different habitats, life histories and complexity will influence immune system design and mode of action and evolution [32]. Considering the whole organism, there are many types of selective pressures that influence the survival as, for instance, climate changes, predation, food availability and infections. Factor B is a protein related to the immune response and the selective pressures worked on it are mainly represented by recurrent infections, to which the organisms were exposed. Therefore, it is possible that the shell clam *Ruditapes decussatus*, amphioxus *B. belcheri* and the spider *Loxosceles laeta* have been infected by similar pathogens that express the same molecular patterns, which explains the distribution of these species at the same group on phylogenetic tree. Further studies will be necessary to understand the nature of the protein Lox-FB and to investigate how it interacts with other complement proteins possibly present in *Loxosceles* hemolymph. Knowing these aspects could contribute to better understand the defense mechanisms of *Loxosceles* spiders in the context of immunologic responses.

#### **Author Contributions**

Conceived and designed the experiments: DTM GPQ DVT. Performed the experiments: DTM GPQ. Analyzed the data: DTM GPQ AP CWvdB DVT. Contributed reagents/materials/analysis tools: RMGA DVT. Wrote the paper: DTM GPQ CWvdB DVT.

#### References

- 1. Zarkadis IK, Mastellos D, Lambris, J. Phylogenetic aspects of the complement system. Dev. Comp. Immunol 2001; 25: 745–762. PMID: <u>11602194</u>
- 2. Nonaka M, Kimura A. Genomic view of the evolution of the complement system. Immunogenetics 2006; 58: 701–713. PMID: <u>16896831</u>
- 3. Walport MJ. Complement. First of two parts. N. Engl. J. Med 2001; 344: 1058–1066 PMID: 11287977
- Bentley DR, Campbell RD. C2 and Factor B: structure and genetics. Biochem. Soc. Symp 1986; 51: 7– 18 PMID: <u>3101701</u>

- Milder FJ, Gomes L, Schouten A, Janssen BJC, Huizinga EG, Romijin RA, et al. Factor B structure provides insights into activation of the central protease of the complement system. Nat. Struct. Mol. Immunol 2007; 14: 224–228.
- 6. Adams MD, Celniker SE, Holt RA, Evanus CA, Gocaynej D, Amanatides PG, et al. The genome sequence of *Drosophila melanogaster*. Science 2000; 287: 85–95.
- The *C. elegans* Sequencing Consortium. Genome Sequence of the nematode *C. elegans*: a platform for investigating biology. Science 1998; 282:2012–2018. PMID: <u>9851916</u>
- Kimura A, Sakaguchi E, Nonaka M. Multi-component complement system of Cnidaria: C3, Bf, and MASP genes expressed in the endodermal tissues of a sea anemone, *Nematostela vectensis*. Immunobiology 2009; 214: 165–178. doi: <u>10.1016/j.imbio.2009.01.003</u> PMID: <u>19195737</u>
- Sunyer JO, Lambris JD. Evolution and diversity of the complement system of poikilothermic vertebrates. Immunol. Rev 1998; 27: 549–564.
- Zhu Y, Thangamani S, Ho B, Ding LD. The ancient origin of the complement system. EMBO J 2005; 24: 382–394. PMID: <u>15616573</u>
- Smith LC, Shih C, Dachenhausen SG. Coelomocytes express SpBf, a homologue of factor B, the second component in the sea urchin complement system. J. Immunol 1998; 161: 6784–6793. PMID: <u>9862709</u>
- Sekiguchi R, Nonaka M. Evolution of the complement system in protostomes revealed by de novo transcriptome analysis of six species of Arthropoda. Dev. Comp. Immunol 2015; 50:58–67. doi: <u>10.1016/j.</u> <u>dci.2014.12.008</u> PMID: <u>25530095</u>
- Fernandes-Pedrosa M, Junqueira de Azevedo ILM, Gonçalves-de-Andrade RM, Kobashi LS, Almeida DD, Lee Ho P, et al. Transcriptome analysis of *Loxosceles laeta* (Araneae, Sicariidae) spider venomous gland using expressed sequence tags. BMC Genomics 2008; 9: 1–12.
- Holmes DS, Quigley M. A rapid boiling method for the preparation of bacterial plasmids. Analytical Biochemistry 1981; 114: 193–197. PMID: <u>6269464</u>
- 15. Ausubel FA. Current protocols in molecular biology. 2nd ed. New York: John Wiley and Sons; 1995
- Bordoli L, Kiefer F, Arnold K, Benkert P, Battey J, Schwede T. Protein structure homology modeling using SWISS-MODEL workspace. Nature Protocols 2009; 4: 1–13. doi: <u>10.1038/nprot.2008.197</u> PMID: <u>19131951</u>
- Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Bio Evol 2013; 30: 2725–2729.
- Prado-Alvarez M, Rotlant J, Gestal C, Novoa, B, Figueras A. Characterization of an C3 and a factor Blike in the carpet-shell clam, *Ruditapes decussates*. Fish Shellfish Immunol 2009; 26: 305–315. doi: <u>10.</u> 1016/j.fsi.2008.11.015 PMID: 19073265
- 19. Pólgar L. The catalytic triad of serine peptidases. Cell Mol Life Sci 2005; 62:2161–2172 PMID: 16003488
- 20. Page M J, Cera ED. Serine peptidases: Classification, structure and function. Cellular and Molecular Life Sciences 2008; 65: 1220–1236. doi: 10.1007/s00018-008-7565-9 PMID: 18259688
- Nonaka M. Evolution of Complement System. In: Anderluh G, Gilbert R, editors. MACPF/CDC Proteins–Agents of Defense, Attack and Invasion Subcell Biochem; 2014; 80:31–43.
- Nonaka M, Takahashi M, Sasaki M. Molecular cloning of a lamprey homologue of the mammalian MHC class III gene, complement factor B. J. Immunol 1994; 152: 2263–2269. PMID: <u>7510741</u>
- Yoshizaki FY, Ikawa S, Satake M, Satoh N, Nonaka M. Structure and the evolutionary implication of the triplicated complement factor B genes of a urochordate ascidian, *Ciona intestinalis*. Immunogenetics 2005; 56: 930–942. PMID: <u>15778902</u>
- He Y.; Tang B.; Zhang S.; Liu Z.; Zhao B.; Chen L. Molecular and immunochemical demonstration of a novel member of Bf/C2 homolog in amphioxus *Branchiostoma belcheri*: Implications for involvement of hepatic cecum in acute phase response. Fish and Shellfish Immunol 2008; 24: 768–778.
- Zhong L, Zhang F, Chang Y. Gene cloning and function analysis of complement B factor -2 of Apostichopus japonicus. Fish and Shellfish Immunol 2012; 33: 504–513.
- Zipfel PF, Hallström T, Riesbeck K. Human complement control and complement evasion by pathogenic microbes—tipping the balance. Mol. Immunol 2013; 56(3):152–160. doi: <u>10.1016/j.molimm.</u> <u>2013.05.222</u> PMID: <u>23810413</u>
- Xiao X, Zhang R, Pang X, Liang G, Wang P, Cheng G. A neuron-specific antiviral mechanism prevents lethal flaviviral infection of mosquitoes. Plos Pathogens 2015; 11: e1004848. doi: <u>10.1371/journal.</u> <u>ppat.1004848</u> PMID: <u>25915054</u>

- Xiao X, Liu Y, Zhang X, Wang J, Li Z, Pang X, et al. Complement-related proteins control the flavivirus infection of *Aedes aegypti* by inducing antimicrobial peptides. Plos Pathogens 2014; 10: e1004027. doi: <u>10.1371/journal.ppat.1004027</u> PMID: <u>24722701</u>
- Le Saux A, Ng PM, Koh JJ, Low DH, Leong G, Ho B, et al. The macromolecular assembly of pathogenrecognition receptors is impelled by serine proteases, via their complement control protein modules. J. Mol. Biol 2008; 377: 902–913 doi: 10.1016/j.jmb.2008.01.045 PMID: 18279891
- Ariki S, Takahara S, Shibata T, Fukuoka T, Ozaki A, Endo Y, et al. Factor C acts a lipopolysaccharideresponsive C3 convertase in horseshoe crab complement activation. J. Immunol 2008; 181: 7994– 8001. PMID: <u>19017991</u>
- Tagawa K, Yoshihara T, Shibata T, Kitazaki K, Endo Y, Fujita T, et al. Microbe-specific C3b deposition in the horseshoe crab complement system in a C2/Factor B-dependent or-independent manner. Plos One 2012; 7: 1–9.
- **32.** Loker E.; Macroevolutionary immunology: a role for immunity in the diversification of animal life. Frontiers in Immunology 2012; 3: 1–21.