Genomic organization of *Echis ocellatus* PI- and PII-SVMP genes

Libia Sanz[1], Juan J. Calvete[1]

[1]Instituto de Biomedicina de Valencia, Jaume Roig 11, 46010 Valencia, Spain.

E-mail: libia.sanz@ibv.csic.es, jcalvete@ibv.csic.es

INTRODUCTION

The evolutionary pathway and the mechanism of selective expression of toxins in the snake venom gland remain poorly understood. The snake venom metalloproteinases are a large multigene toxin family that encode differing multidomain proteins. SVMP are typically characterized by the presence of a catalytic " H-box" amino motif (HEX2HX2GX2HD) in the metalloproteinase domain, and categorized into classes (PI, PII, PIII) based upon the presence or absence of additional nonproteinase domains. The origin of SVMPs has been inferred to have occurred following the split of the Pareatidae from the remaining Caenophidians, approximately 54-64 Mya, through recruitment, duplication, and neofunctionalization by positive Darwinian selection of a closely related body cellular ADAM 7 or 28 ancestor gene. However, details on the molecular events underlying the origin and evolution of the SVMP multigene family remain elusive. The genomic structure of the 15652 bp *Echis ocellatus* pre-pro-EOC00089-like PIII-SVMP gene (Sanz et al., Toxicon 2012; 60: 455–469) pointed to introns as possible key elements in the recruitment and amplification process of SVMPs into the venom gland of extant snakes. In order to understand the structure and evolution of SVMP genes, we have determined the genomic structures of the 17768 bp *Echis ocellatus* long ocellatusin precursor PII-SVMP EOC00006, and the 21643 bp *Echis ocellatus* PI-snake venom metalloproteinase EOC00028.

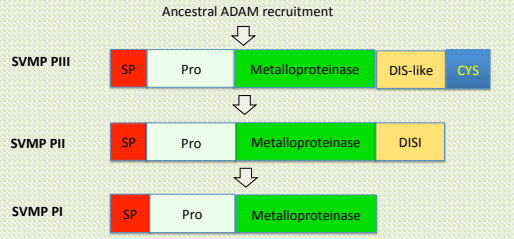

Figure 1. Diagram of the different domains present in PIII-, PII- and PI SVMP. SP-signal peptide, Pro-prodomain, Metalloproteinase-metalloproteinase domain, DIS-disintegrin domain, DIS-like-disintegrin-like domain and CYS-cysteine-rich domain.

Results

We employed an iterative process for sequencing the genes coding for PII-SVMP EOC00006 and PI-SVMP EOC00028 . We took advantage of knowledge of the genomic organization of pre-pro-EOC00089-like PIII-SVMP gene [Sanz *et al,* Toxicon 2012; 60: 455-69], the organization of disintegrin domains of SVMPs from *E. ocellatus* [Bazaa *et al.*, J. Mol. Evol 2007; 64: 261-71], and the cDNAs sequences of contigs EOC00006 (GenBank accession code AM039693) and EOC000028 (GenBank accession code AM039698), to design primers for the PCR amplification of genomic sequences. Using an iterative process of PCR-amplification and sequencing we have determined the genomic structures of the 17768 bp Echis ocellatus long ocellatusin precursor PII-SVMP EOC00006, and the 20941 bp Echis ocellatus PI-snake venom metalloproteinase EOC00028.

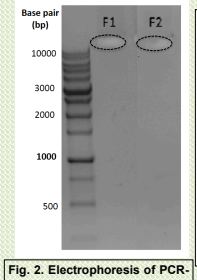

Fig. 2. Electrophoresis of PCR-amplified F1 and F2 PI_SVMP fragments.F1 and F2 were cut, purified, and cloned into the vector pJET1.2 .

Fig.3 Overlapping PCR-amplified genomic fragments used to assemble the full- length PII long ocellatusin

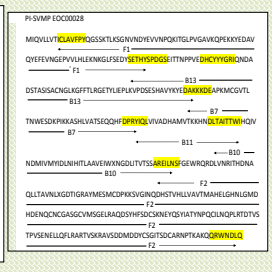

Fig.4 Set of PCR-amplified overlapping genomic fragments used to assemble the PI EOC00028 gene.

Pairwise amino acid sequence comparisons between Ac_ADAM28 and PII- EOC00006 and PI-SVMP EOC 00028 indicated that the ORFs of the proteins Ac_ADAM28/ PII and Ac_ADAM28/ PI share 35% and 41% identity repectively.

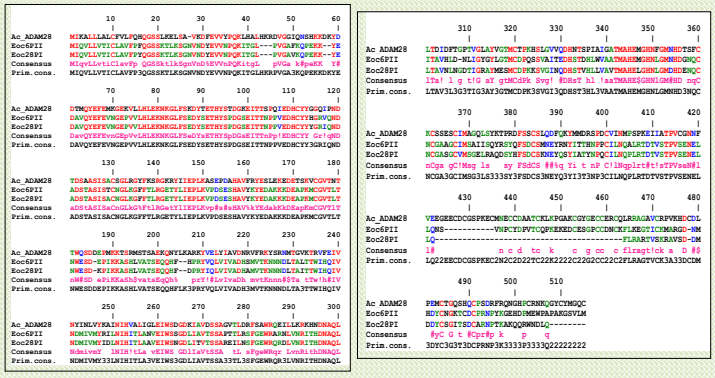

Fig 5. Alignment of the amino acid sequences of EOC00006 PII- , EOC00028 PI-SVMP and *A. carolinensis* (Ac) ADAM28.

The PII-SVMP EOC00006 gen consists of 15 exons interrupted by 14 introns, while the PI- EOC00028 comprises 13 exons interrupted by 12 introns. In both of them the signal peptide is encoded by exon 1, the pro-peptide comprises exons 2-6 , the metalloproteinase domain of PII is code ford by exons 7-11 and part of exon 12, and by exons 7-13 in the case of PI EOC00028. The PII gene contains a C-terminal disintegrin domain that is encoded by part of exon 12 and by the 13-15 exons. The distribution and length of the introns are schematized in Figures 6 and 7. Most introns belong to phase 0, followed by phase 2, and only the first intron is classified as phase 1.

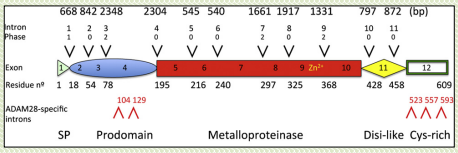

Figure 6. Scheme of the genomic organization of EOC00089-like PIII-SVMP gene. The distribution, phase, and size of the introns and the boundaries of the protein-coding regions are also highlighted.The insertion positions of 5 introns present in human and *A. carolinensis* ADAM 28 but absent in the *E.ocellatus* are shown in red. SP, signal peptide; Disi-like, disintegrin-like domain; Cys-rich, Cysteine-rich domain

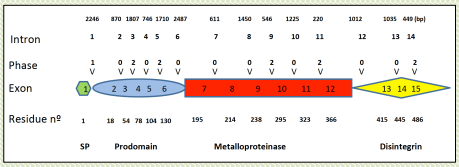

Figure 7. Scheme of the genomic organization of EOC00006 PII-SVMP gene. The distribution, phase, and size of the introns and the boundaries of the protein-coding regions are also highlighted. SP, signal peptide.

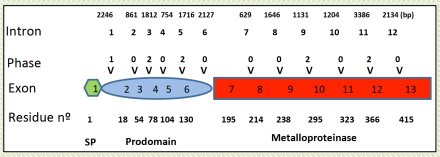

Figure 8. Scheme of the genomic organization of EOC00028 PI-SVMP gene. The distribution, phase, and size of the introns and the boundaries of the protein-coding regions are also highlighted. SP, signal peptide.

Transposable elements occupy a large fraction of eukaryotic genomes.Their mobility and amplification represent a major source of genomic variation. RepeatMasker identified sequences of long interspersed nuclear retroelements (LINE) beloging to the L2/ CR1 (chicken repeat 1), short interspersed nuclear retroelements SINE/ Sauria, and hobo-activator DNA (Charlie, hobo AT) transposon in some introns of both genes (Table 1). Table 2 displays the degree of sequence identity between the introns of the PII and the PI genes.

Table 1. Comparison of intron size and retroelements identified in *E. ocellatus* Eoc00089-like PIII-SVMP, Eoc00006 PII-SVMP, Eoc00028 PI-SVMP and *A. carolinensis* ADAM28 genes.

EOC00089-like PIII-SVMP			EOC00006 PII-SVMP			EOC00028 PI-SVMP			Ac_ADAM28		
Intron	bp	Inserted retroelements		bp	Inserted retroelements	Intron	bp	Inserted retroelements			
Eo1	668		Eo1	2246	SINE/ Sauria	2888	2 SINE/ Sauria 1 hobo AT		Ac1	8815	SINE/ Sauria, LINE/ Penelope, L2/ CR1, 4 hobo AT,4 Tc1- IS630-Pogo
Eo2	842		Eo2	870		864			Ac2	347	
Eo3	2348	LINE/ L2/ CR1,RTE/ Bov-B, SINE/ Sauria	Eo3	1807	LINE/ L2/ CR1	1812	LINE/ L2/ CR1		Ac3	3579	SINE/ Sauria
			Eo4	746		755			Ac4	6791	7 SINE/ Sauria
			Eo5	1710	LINE/ L2/ CR1	1716	LINE/ L2/ CR1		Ac5	2935	SINE/ Sauria
Eo4	2304	RTE/ Bov-B	Eo6	2487		2127			Ac6	1892	SINE/ Sauria
Eo5	545		Eo7	611		629			Ac7	4009	SINE/ Sauria
Eo6	540		Eo8	1450	2 LINE/ L2/ CR1	1646			Ac8	1064	SINE/ Sauria
Eo7	1661	SINE/ Sauria, SINE	Eo9	546		1131	2 SINE/ Sauria		Ac9	236	
Eo8	1917	hobo AT	Eo10	1143	hobo AT	1204	hobo AT		Ac10	2840	2 LINE/ Dong-4,2 hobo AT, 3 Tc1 IS630-Pogo
Eo9	1331		Eo11	220		3386			Ac11	2135	
Eo10	797		Eo12	1012		2134			Ac12	278	
Eo11	872		Eo13	975					Ac13	2386	2 SINE/ Sauria, hobo AT, 2 Tc1- IS630-Pogo
			Eo14	449					Ac14	1071	SINE/ Sauria
									Ac15	183	
									Ac16	1032	SINE/ Sauria

Table 2. Percentage of nucleotide sequence identity between PII and PI introns

	PII													
PI	Percentage Identities													
	Intr_1	Intr_2	Intr_3	Intr_4	Intr_5	Intr_6	Intr_7	Intr_8	Intr_9	Intr_10	Intr_11	Intr_12	Intr_13	Intr_14
Intr_1	69%	31%	49%	27%	48%	49%	24%	42%	21%	39%	9%	35%	33%	18%
Intr_2	31%	96%	36%	50%	37%	28%	47%	40%	42%	47%	22%	50%	50%	36%
Intr_3	48%	36%	95%	33%	53%	46%	28%	48%	26%	45%	11%	41%	39%	22%
Intr_4	28%	50%	33%	92%	34%	26%	49%	38%	46%	43%	26%	46%	48%	42%
Intr_5	47%	37%	53%	33%	99%	46%	29%	49%	26%	46%	12%	41%	39%	22%
Intr_6	49%	33%	50%	29%	48%	82%	24%	44%	22%	40%	10%	37%	35%	18%
Intr_7	24%	46%	29%	48%	29%	22%	88%	32%	48%	39%	29%	43%	45%	45%
Intr_8	45%	38%	49%	34%	49%	42%	29%	66%	27%	45%	12%	42%	41%	23%
Intr_9	37%	47%	43%	44%	44%	35%	40%	46%	44%	49%	18%	49%	49%	32%
Intr_10	39%	48%	45%	44%	45%	37%	38%	47%	33%	95%	17%	49%	47%	30%
Intr_11	43%	22%	39%	20%	38%	46%	16%	33%	15%	28%	6%	25%	25%	12%
Intr_12	50%	32%	50%	29%	48%	48%	24%	44%	22%	40%	10%	45%	39%	19%

CONCLUDING REMARKS

E. ocellatus EOC00006 (PII-SVMP) and EOC00028 (PI-SVMP) genes have identical genomic organization regarding the number and position of the 13 superimposable exon sequences, as well as the number, phase, and boundaries of the 12 introns that interrupt these homologous coding sequences. In addition, topologically equivalent introns contain the same class of retroelements. However, the size and degree of sequence similarity is highly conserved in some (introns 2-5), but varies among other (1, 6-12), topologically equivalent introns. The degree of structural conservation of homologous exons between Eoc00006 PII-SVMP and Eoc00028 PI-SVMP also shows a pattern of decreasing similarity from 5'-3'. Hence, PII and PI exons 1, 3, 4, 6 and 8 share 94-100% amino acid sequence identity, whereas this figure is 11- 85% for exons 5, 7, and 9-13. This disparity in the distinct diversification of various structural elements of homologous PI and PII SVMPs is puzzling. Clearly, the sequence of larger genomic regions encoding clusters of homologous genes are eagerly required to distinguish between a "whole-gene-duplication" mechanism and a mechanism involving the combinatorial arrangement of duplicated structural domains.