Available online at www.sciencedirect.com

SCIENCE DIRECT

Gene 337 (2004) 55–69

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

www.elsevier.com/locate/gene

Bitis gabonica (Gaboon viper) snake venom gland: toward a catalog for the full-length transcripts (cDNA) and proteins ☆

Ivo M.B. Francischetti[a,*], Van My-Pham[a], Jim Harrison[b], Mark K. Garfield[c], José M.C. Ribeiro[a]

[a] *Medical Entomology Section, Laboratory of Malaria and Vector Research, National Institutes of Allergy and Infectious Diseases, National Institutes of Health, 12735 Twinbrook Parkway, Building Twinbrook III, Room 2E-28, Rockville, MD 20852, USA*

[b] *Kentucky Reptile Zoo, 200 L and E Railroad, Slade, KY 40376, USA*

[c] *Research Technologies Branch, National Institutes of Allergy and Infectious Diseases, National Institutes of Health, 12441 Parklawn Dr., Building Twinbrook II, Room 225, Rockville, MD 20852, USA*

Received 17 November 2003; received in revised form 31 January 2004; accepted 18 March 2004

Received by J.A. Engler

Abstract

The venom gland of the snake *Bitis gabonica* (Gaboon viper) was used for the first time to construct a unidirectional cDNA phage library followed by high-throughput sequencing and bioinformatic analysis. Hundreds of cDNAs were obtained and clustered into contigs. We found mostly novel full-length cDNA coding for metalloproteases (P-II and P-III classes), Lys49-phospholipase A2, serine proteases with essential mutations in the active site, Kunitz protease inhibitors, several C-type lectins, bradykinin-potentiating peptide, vascular endothelial growth factor, nucleotidases and nucleases, nerve growth factor, and L-amino acid oxidases. Two new members of the recently described short coding region family of disintegrin, displaying RGD and MLD motifs are reported. In addition, we have identified for the first time a cytokine-like molecule and a multi-Kunitz protease inhibitor in snake venoms. The CLUSTAL alignment and the unrooted cladograms for selected families of *B. gabonica* venom proteins are also presented. A significant number of sequences were devoid of database matches, suggesting that their biologic function remains to be identified. This paper also reports the N-terminus of the 15 most abundant venom proteins and the sequences matching their corresponding transcripts. The electronic version of this manuscript, available on request, contains spreadsheets with hyperlinks to FASTA-formatted files for each contig and the best match to the GenBank and Conserved Domain Databases, in addition to CLUSTAL alignments of each contig. We have thus generated a comprehensive catalog of the *B. gabonica* venom gland, containing for each secreted protein: (i) the predicted molecular weight, (ii) the predicted isoelectric point, (iii) the accession number, and (iv) the putative function. The role of these molecules is discussed in the context of the envenomation caused by the Gaboon viper.

Keywords: Snake venom; *Bitis gabonica*; Metalloprotease; cDNA library; Viperidae; Toxins

1. Introduction

Snake venoms are complex mixtures of proteins, including enzymes and other biologically active components (Aird, 2002). These components are responsible for the envenomation caused by snake bites and display mostly neurotoxic (Harvey, 2001) or proteolytic (Bjarnason and Fox, 1995) activities. The Gaboon viper, *Bitis gabonica*, is a large viper widely distributed over West, Central, and East Africa. It produces the largest amounts of venom of all poisonous snakes, yielding in excess of 2 g of dried venom per milking. Bites from Gaboon vipers appear to be rare, however, due at least in part to the animal's extremely placid nature. Actually,

Abbreviations: BG-HP, *B. gabonica* hypothetical protein; bp, base pair; BPP, bradykinin-potentiating peptide; ds, double stranded; EtdBr, ethidium bromide; LAO, L-amino acid oxidases; NGF, nerve growth factor; NR, nonredundant; ORF, open reading frame; pI, isoelectric point; PLA2, phospholipase A2; PVDF, polyvinylidene difluoride; SDS/PAGE, sodium dodecyl sulfate/polyacrylamide gel electrophoresis; VEGF, vascular endothelial growth factor.

☆ Supplementary data associated with this article can be found, in the online version, at doi: 10.1016/j.gene.2004.03.024

* Corresponding author. Tel.: +1-301-402-62-00; fax: +1-301-402-49-41.

E-mail address: ifrancischetti@niaid.nih.gov (I.M.B. Francischetti).

0378-1119/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2004.03.024

the majority of reported bites have occurred from handling specimens in captivity (Marsh et al., 1997). In these cases, unstable circulation, a severe coagulation disorder, and tissue damage followed by necrosis are the most life-threatening conditions associated with the envenomation (Marsh et al., 1997).

As far as the biochemical composition of *B. gabonica* venom is concerned, several activities have been reported including arginine esterases (Viljoen et al., 1979), phospholipase A2 (PLA2) (Botes and Viljoen, 1974), thrombin-like enzyme (gabonase) (Pirkle et al., 1986), anti-platelet (gabonin) (Huang et al., 1992), and metalloprotease (Marsh et al., 1997) activities. Remarkably, information about the *B. gabonica* snake venom gland at the molecular level is almost nonexistent. In fact, only the N-terminus of gabonase (Pirkle et al., 1986) in addition to the amino acid sequence of a *B. gabonica* PLA2 has been reported (Botes and Viljoen, 1974). Furthermore, a GenBank search with the term "*B. gabonica*" displays in September 2003 only a PLA2 and cytochrome *b* sequence at the protein level and two housekeeping sequences at the nucleotide level. The striking lack of information on the molecular constituents of *B. gabonica* venom led us to choose this snake to perform a venom gland cDNA library followed by sequencing of the clones. Edman degradation of the most abundant protein was performed in parallel, allowing us to generate for the first time a comprehensive catalog containing *B. gabonica* transcripts (cDNA) and proteins. Roles of the components of Gaboon viper venom are discussed in the context of both envenomation and in vitro activities previously described for this venom.

2. Materials and methods

2.1. Reagents

All water used was of 18 MΩ quality and was produced using a MilliQ apparatus (Millipore, Bedford, MA, USA). Organic compounds were obtained from Sigma (St. Louis, MO, USA) or as stated otherwise.

2.2. Snake venom gland and snake venom

B. gabonica venom and venom gland were obtained from the same snake held in captivity at the Kentucky Reptile Zoo (Slade, KY). Three days after milking the head was cut and the gland immediately dissected and frozen in dry ice under the supervision of Jim Harrison and Kristen L. Wiley of Kentucky Reptile Zoo (http://www.geocities.com/Kentuckyreptilezoo).

2.3. Sodium dodecyl sulfate/polyacrylamide gel electrophoresis (SDS/PAGE)

Approximately 30 μg of venom was treated with LDS sample buffer (Invitrogen, San Diego, CA, USA) containing SDS without reducing conditions and applied to a NU-PAGE 4–12% Bis-Tris gel (MES buffer) (Invitrogen) 1 mm thick. The supplemental data of this paper contains detailed information on *B. gabonica* SDS/PAGE and Edman degradation.

2.4. Snake venom gland cDNA library construction and sequencing

A fragment was rapidly obtained from the center part of the gland. Fragments were transferred to a sterile plastic Petri dish located on the top of dry ice to avoid melting. *B. gabonica* salivary gland mRNA was obtained using Micro-Fast Track mRNA isolation kit (Invitrogen) according to the manufacturer's instructions. The PCR-based cDNA library was made following the instructions for the SMART cDNA library construction kit (Clontech, Palo Alto, CA, USA) as described (Francischetti et al., 2002). Cycle sequencing reactions using DTCS labeling kit from Beckman Coulter. (Fullerton, CA, USA) was performed as reported (Francischetti et al., 2002). The supplemental data of this paper contains detailed information on *B. gabonica* cDNA library construction and sequencing of *B. gabonica* cDNA library.

2.5. cDNA sequence clustering and bioinformatics

Other procedures were as in Francischetti et al. (2002) except that clustering of the cDNA sequences was accomplished using the CAP program (see Supplemental data). The electronic version of the complete tables (Microsoft Excel format), with hyperlinks to web-based databases and to BLAST results is available on request (ifrancischetti@niaid.nih.gov). The supplemental data of this paper contains detailed information on cDNA sequence clustering, sequencing information cleaning, blast search and other bioinformatic analysis.

3. Results and discussion

In an attempt to improve our understanding of the complexity of the proteins and transcripts expressed in *B. gabonica* venom glands, we have performed SDS/PAGE and a cDNA library using, respectively, the secreted proteins and mRNA from this same tissue.

3.1. SDS/PAGE of B. gabonica snake venom gland

Fig. 1 shows the pattern of separation of *B. gabonica* venom proteins by SDS-PAGE that have been stained by Coomassie Blue. The gel shows 15 clearly visible stained bands and many other slightly stained. The protein bands were numbered from 1 to 15 according to their decreasing apparent molecular weight, starting with the letter BG that stands for *B. gabonica*. To identify these proteins, they were transferred to PVDF membranes and the bands cut from the membrane and submitted to Edman degradation. Amino-

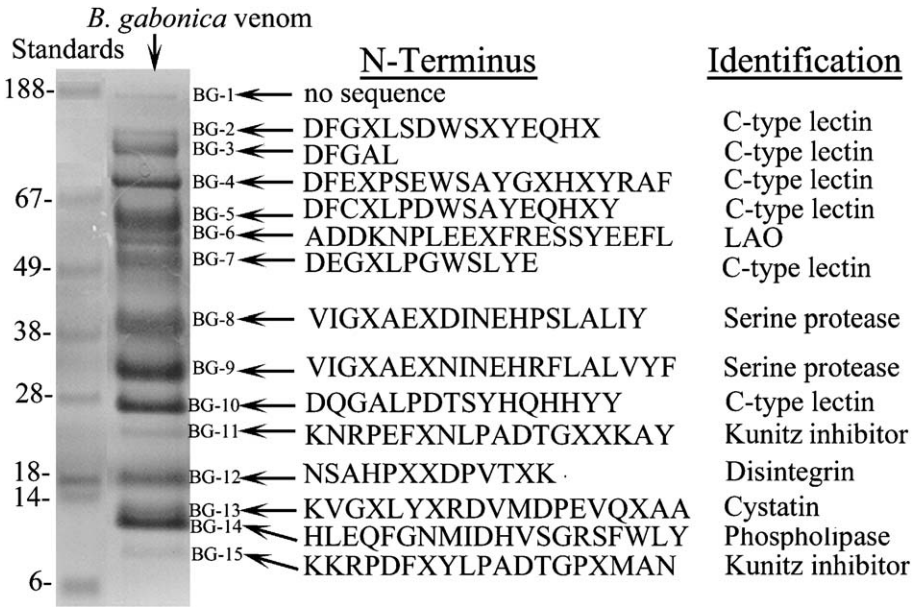

Fig. 1. SDS-PAGE of *B. gabonica* venom gland proteins under denaturing non-reducing conditions. Thirty μg of venom was applied to a 4–12% NU-PAGE pre-cast gel, MES buffer. Standard molecular mass is shown on the left. The match found is shown on the right.

terminal information was successfully obtained for all bands BG-2 to BG-15. To find matches to known proteins, the sequences were blasted against the NR GenBank database and to each cDNA sequence obtained in the mass-sequencing project of the *B. gabonica* venom gland described in this paper (see Materials and methods).

3.2. cDNA library of the venom gland of B. gabonica

A cDNA library was constructed using the venom gland of *B. gabonica* and about 600 of independent clones randomly 5′ sequenced. When a cluster analysis of all sequences from this library was performed, 300 independent contigs were organized. Subsequently, contigs were blasted against the NR nucleotide database, and the presence of signal peptides was predicted by submission of the sequences to the SignalP server. Our analysis shows that ~75% of all sequences have database hits; ~46% of all sequences code for protein with a putative signal peptide, 38% code for proteins with housekeeping function, and the remaining sequences could not be assigned as housekeeping or secretory (unknown). It is thus clear that cDNA for secretory proteins are highly represented in our library, suggesting that in vivo these molecules are preferentially expressed over housekeeping and unknown-function proteins. Also, because the cDNA have been obtained from a single animal, these variations do not represent populational diversity, as the maximal number of alleles would be 2.

Among the individual cDNA sequences containing putative signal peptides, 90% have hits in the GenBank database. Fig. 2A shows the relative proportion of the number of individual cDNA over the total number. These

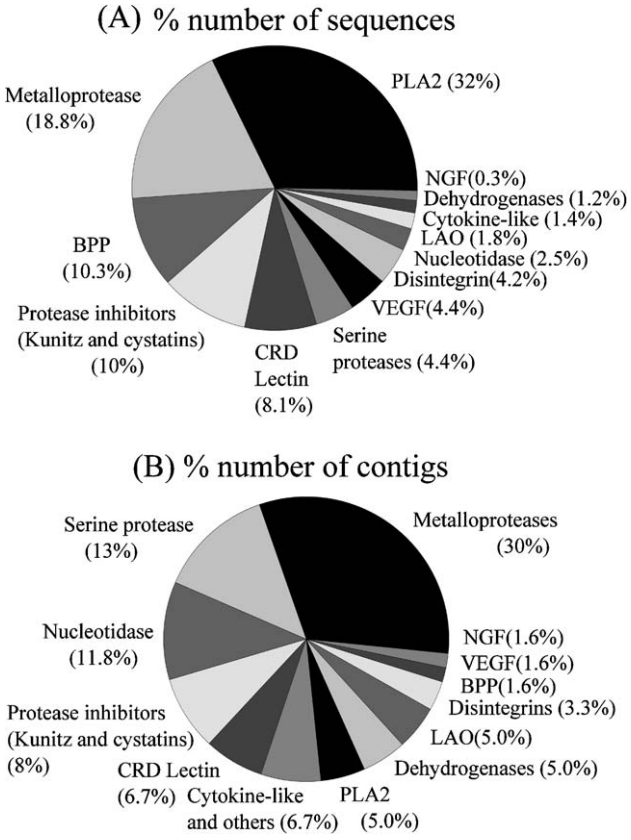

Fig. 2. The composition of *B. gabonica* cDNA coding for putative secreted proteins. (A) Number of sequences (%) for a given contig. (B) Number of contigs (%) for a given venom biologic function (e.g. phospholipase A2). The total number of sequences in (A) or contigs in (B) includes only the ones coding for putative secreted proteins. PLA2, phospholipase A2; BPP, bradykinin-potentiating peptide; CRD lectin, carbohydrate-recognition domain lectin, carbohydrate recognition domain containing lectin.

Table 1
Bitis gabonica venom gland cDNA clusters associated with putative secreted proteins

Contig no.[a]	Number of sequence[b]	Best match to NR protein database[c]	gi accession number	*E* value[d]	Best match to CDD database[e]	*E* value[d]	Comments	Gel Position[f]	Edman product[g]
ADAM family metalloproteases									
126	13	atrolysin A (EC 3.4.24.1) -western diamon	542663	4e – 067	ACR	9e – 041	similar to Western diamondback atrolysin		
239	7	hemorrhagic metalloproteinase HR1b [	20530121	1e – 112	DISIN	2e – 031	similar to *Trimeresurus flavoviridis* metalloprotease		
253	7	fibrinolytic metalloproteinase (EC 3.4.24	2118144	1e – 036	disintegrin	2e – 023	disintegrin similar to fibrinolytic metalloproteinase		
125	5	hemorrhagic metalloproteinase HR1b [	20530121	1e – 029	ACR	7e – 021	similar to *Trimeresurus flavoviridis* metalloprotease		
93	3	hemorrhagic metalloproteinase HR1b [	20530121	3e – 040	Pep_M12B	6e – 012	similar to *Trimeresurus flavoviridis* metalloprotease		
26	3	metalloproteinase (EC 3.4.24. –) H–I precu	1086019	1e – 015	ACR	4e – 010	similar to carpet viper enzyme - truncated?		
240	2	hemorrhagic metalloproteinase HF3 [Bo	18379369	6e – 094	DISIN	5e – 030	similar to *Bothrops jararaca* metalloproteinase		
34	1	fibrinolytic metalloproteinase (EC 3.4.24	2118144	8e – 086	Reprolysin	6e – 032	similar to Vipera lebetina enzyme		
80	1	hemorrhagic metalloproteinase HF3 [Bo	18379369	2e – 059	Reprolysin	2e – 036	similar to *Bothrops jararaca* metalloproteinase		
212	1	hemorrhagic metalloproteinase HR1a [	20530119	2e – 035	Reprolysin	4e – 018	similar to *Trimeresurus flavoviridis* metalloprotease		
217	1	atrolysin A (EC 3.4.24.1) - western diamon	542663	1e – 078	DISIN	2e – 033	similar to Western diamondback atrolysin		
245	1	hemorrhagic metalloproteinase HF3 [Bo	18379369	6e – 094	DISIN	5e – 030	similar to *Bothrops jararaca* metalloproteinase		
16	1	berythractivase [Bothrops e	17865171	5e – 034	DISIN	3e – 022	similar to Bothropus berythractivase		
172	1	hemorrhagic metalloproteinase HR1a [	20530119	3e – 015			similar to *Trimeresurus flavoviridis* metalloprotease		
30	1	hemorrhagic metalloproteinase HF3 [Bo	18379369	2e – 011	ACR	8e – 007	similar to *B. jararaca enzyme* - truncated?		
32	1	metalloproteinase (EC 3.4.24.-) H-II prec	1364104	2e – 015	ACR	2e – 008	similar to carpet viper enzyme - truncated?		
11	1	metalloproteinase (EC 3.4.24.-) H-I precu	1086019	4e – 010	ACR	1e – 006	similar to factor X activating enzyme - truncated?		
194	1	hemorrhagic metalloproteinase HR1b [	20530121	1e – 029	ACR	7e – 021	similar to Trimeresurus flavoviridis metalloprotease		
Serine proteases									
71	3	Platelet-aggregating proteinase	6093643	3e – 051	Tryp_SPc	7e – 024	serine protease		
60	3	serine beta-fibrinogenase precursor [	22417221	2e – 068	Tryp_SPc	5e – 031	serine protease	BG-7 and	VIGXAE*X*DINEHPSLALIY
90	1	beta-fibrinogenase [Agkistrodon blomh	6706013	2e – 017	Tryp_SPc	8e – 016	serine protease	BG-8	VI*X*AE*X*NINEHRFLALVYF
259	1	Flavoxobin precursor >gnl|BL_ORD	3915685	3e – 032	Tryp_SPc	1e – 008	serine protease		
63	1	serine alpha-fibrinogenase precursor	22417112	9e – 028	Tryp_SPc	1e – 016	serine protease		

108	1	thrombin-like enzyme ussurase [Gloydi	17933276	4e – 004	Tryp_SPc	4e – 005	serine protease		
136	1	Venom serine proteinase 2 precu	13959639	3e – 010	Tryp_SPc	0.001	serine protease		
236	1	Venom serine proteinase 3 precu	13959638	3e – 011	Tryp_SPc	8e – 004	serine protease		
Nucleotidases and nucleases									
40	1	5'-nucleotidase precursor (Ecto-n	112824	0.063			Ecto-nucleotidase		
166	1	adenine phosphoribosyltransferase; A	4502171	0.005			Adenine phosphoribosyltransferase		
42	1	Bovine Bcnt includes a region derive	22779307	6e – 006			endonuclease?		
113	1	Bcnt [Cervus nippon] 41 0.003	20196208	0.003			endonuclease?		
59	1	ORF2 [Platemys spixii] 87 2e – 029	6576738	2e – 029			reverse transcriptase-endonuclease		
52	1	Bcnt [Cervus nippon] 60 1e – 008	20196208	1e – 008	UPF0032	0.070	reverse transcriptase-endonuclease		
28	1	putative p150 [Homo sapiens] 60 4e – 012	2072964	4e – 012	rvt	1e – 004	reverse transcriptase-endonuclease?		
Protease inhibitors									
203	19	textilinin [Pseudonaja text	15321630	4e – 026	Kunitz_BPTI	3e – 015	Kunitz protease inhibitor	BG-11/	KNRPEF*C*NLPADTGX*C*KAY/
146	1	beta bungaratoxin B2b chain [Bungarus	24459200	3e – 004	Adeno_PV	0.047	Kunitz similar to beta bungaratoxin B2b chain	BG-15	KKRPDF*X*YLPADTGP*X*MANF
137	1	Kunitz-like protease inhibitor precur	22901764	3e – 030	KU	3e – 020	Kunitz-like protease inhibitor		
233	1	textilinin [Pseudonaja text	15321630	7e – 004	KU		similar to textilinin [Pseudonaja]		
264	5	Cystatin >gnl|BL_ORD_ID|509601 gi|	118194	3e – 046	cystatin	2e – 022	cystatin	BG-13	KVGXLYXRDVMDPEVQXAAA
C type lectins									
214	8	fibrinogen clotting inhibitor A chain	4337050	1e – 046	CLECT	2e – 024	C type lectin-fibrinogen clotting inhibitor A chain	BG-4	DFX*X*PSEWSAYGXH*X*YRAF
215	7	agglucetin-alpha 2 subunit precursor	23321261	6e – 049	CLECT	8e – 021	C type lectin similar to agglucetin-alpha 2 subunit precursor	BG-4	DFE*X*PSEWRPFDXH*X*YRAFD
15	5	Galactose-binding lectin precurs	7674107	2e – 055	CLECT	5e – 018	C type lectin-galactose binding	BG-7	DEG*X*LPGWSLYEXH*X*YKFF
216	2	Coagulation factor IX/factor X-bi	2851435	3e – 033	CLECT	2e – 015	C type lectin similar to factor IX/factor X binding protein A	BG-7	DEG*X*LPGWSLYE
Phospholipases									
1	63	Phospholipase A2 isozyme CM-II (P	129423	3e – 056	phoslip	3e – 046	phospholipase	BG-12/13	HLEQFGNMIDHVSGRSFWLY
3	23	Phospholipase A2 isozyme CM-II (P	129423	2e – 055	phoslip	2e – 046	phospholipase	BG-12/13	HLEQFGNMIDHVSGRSFWLY
222	1	Phospholipase A2 5 precursor (P	27734438	8e – 019	phoslip	5e – 012	phospholipase		
Disintegrins									
119	7	acostatin alpha chain [Agkistrodon c	27922986	2e – 034	disintegrin	7e – 020	Disintegrin similar to acostatin alpha chain		
127	4	Disintegrin EMF10A (Platelet agg	6225272	1e – 020	disintegrin	5e – 009	Disintegrin similar to eristocophin I	BG-11	NSAHP*XX*DPVT*X*K

(continued on next page)

Table 1 (*continued*)

Contig no.[a]	Number of sequence[b]	Best match to NR protein database[c]	gi accession number	*E* value[d]	Best match to CDD database[e]	*E* value[d]	Comments	Gel Position[f]	Edman product[g]
Growth factors									
227	12	vascular endothelial growth factor [G	27368068	1e – 087	PDGF	1e – 031	vascular endothelial growth factor		
219	1	putative neurotrophic growth factor p	15407254	3e – 035	NGF	8e – 018	nerve growth factor-similar to putative neurotrophic growth factor		
Amino acid oxidases									
165	2	FAD-containing L-amino acid oxidase Ap	5565692	4e – 017	Amino_oxidase	3e – 004	amino acid oxidase	BG-6	ADDKNPLEEXFRESSYEEFL
182	2	M-LAO [Agkistrodon blomhoffi] 264 e – 101	15887054	1e – 101	Amino_oxidase	6e – 015	amino acid oxidase		
181	1	M-LAO [Agkistrodon blomhoffi] 260 8e – 069	15887054	8e – 069	Amino_oxidase	1e – 018	amino acid oxidase		
Cytokine-like molecules									
13	1	cytokine-like protein 2–21 [Mus mus	22296880	4e – 026			similar to cytokine-like protein 2–21		
Dehydrogenases									
251	1	unnamed protein product [Mus musculus] 56 1e – 007	26345686	1e – 007			similar to 3-phosphoglycerate dehydrogenase		
175	1	unnamed protein product [Mus musculus] 192 2e – 068	26345686	2e to 068	2-Hacid_DH_C	2e – 017	similar to 3-phosphoglycerate dehydrogenase		
179	1	NADH dehydrogenase subunit 1 [Afronat	16225796	7e – 074	NADHdh	1e – 056	NADH dehydrogenase subunit		
Bradykinin-potentiating peptide									
188	28	BPP	427226	1.00E – 79			Bradykinin-potentiating peptide		

[a] Contigs were obtained using program CAP assembler.
[b] Number of sequenced clones in cluster.
[c] Best protein match by BlastX to the non redundant protein database of NCBI.
[d] Significance of the match.
[e] Best match by RPSblast to the Conserved Domain Database.
[f] Position of a given protein in the SDS/PAGE shown in Fig. 1.
[g] Edman degradation of a given protein.

sequences were organized into 60 contigs (70%). Fig. 2B shows the relative proportion of the number of contigs for each venom toxin family. Based on the distribution described above, cDNAs coding for PLA2 are the most abundant but organized in only three contigs. This indicates that PLA2 in this venom are rather similar and highly expressed. On the other hand, the metalloproteases are organized in 18 different contigs, suggesting that these enzymes may have evolved to perform other functions. Finally, among the 30 sequences (10% of all sequences) that do not have matches to any database, 25 contigs (30%) were organized.

Among the housekeeping cDNAs, we have found sequences involved in transcription and translation (ribosomal proteins, cAMP-dependent transcription factors, elongation factors), metabolism (ATP synthase, amine oxidase, glutathione *S*-transferase, guanine nucleotide-binding protein, cytochrome c oxidase, NADH-ubiquinone oxidoreductase chain I), processing (versican core protein precursor), cell regulation (lithostatine), structural functions (microtubule binding protein), storage (ferritin heavy chain), and retrotransposable (L-1) elements. The complete list of the sequences coding for proteins with secretory, housekeeping, or undetermined function, with or without database hits, can be obtained on request (ifrancischetti@niaid.nih.gov).

3.3. cDNA coding for putative secretory proteins

Table 1 describes the contigs we have found coding for putative secreted proteins and, when available, their corresponding N-terminus obtained by Edman degradation. Matches to the NR, snake DNA, and Conserved Domains Database in addition to accession numbers are also reported. A detailed discussion on the sequences assigned by each cluster and its participation in envenomation by *B. gabonica* is presented below.

3.3.1. Metalloproteases

The metalloproteases make up the most complex group of proteins, being composed of 18 contigs or 30% of all assembled cDNAs. These findings are consistent with the functional characterization of two hemorrhagic proteins (HTa and HTb) in *B. gabonica* venom that were previously shown to degrade collagen and affect endothelial cell morphology (Marsh et al., 1997). Metalloproteases, the primary proteins responsible for snake venom-induced hemorrhage, belong to the reprolysin family of venom metalloproteases (Bjarnason and Fox, 1995). These enzymes are capable of hydrolyzing various components of the extracellular matrix and have also been reported to affect endothelial cells leading to apoptosis. These enzymes are organized into four classes PI-PIV, according to size and domain composition (Bjarnason and Fox, 1995; Jia et al., 1997).

According to our cDNA library, we have found a number of contigs containing partial sequences with homology to the reprolysin, disintegrin or cysteine-rich domains of diferent venom metalloprotease. However, no matches to the C-type lectin domains of metalloproteases were found (Class P-IV). It appears that *B. gabonica* venom contains the P-II and P-III classes of metalloprotease although the presence of the P-I and P-IV classes cannot be excluded. Contig 34 has the longest cDNA we have found in our library coding for a metalloprotease and accordingly, it was extended with appropriated primers in an attempt to identify its functional domains. Although the pre- and pro-domains are not available, the regions coding for the metalloprotease and disintegrin domains were found, indicating that this enzyme belongs to the P-II class and named herein *B. gabonica* metalloprotease-4 (AY44228). The metalloprotease domain is typical and contains the zinc-binding motif, but is unusual in the sense that 5 instead of 6 cysteines are present. Since the nucleotide sequence coding for this regions was reproducible and unambiguous, and found in a reliable region of the chromatogram, it is concluded that this is a true substitution. Actually, it has been reported that the number of cysteine in the metalloprotease domain of these enzymes may differ (Kini et al., 2002). Of interest, atrolysin A, a class III enzyme from *Crotalus atrox* also has an additional cysteine residue in the proteinase domain. The oxidation state or potential disulfide bond partner of this residue in atrolysin is unknown (Bjarnason and Fox, 1995). At present, however, it cannot be completely excluded that a mutation may have occurred in this cDNA during the first or second strand synthesis; or unlikely, the mRNA used to generate the cDNA presents a mutation. As far as the disintegrin domain is concerned a typical RGD sequence was found, and the cysteine pattern was similar but again, not identical to the disintegrin domain of most metalloproteases. In fact, the *B. gabonica* metalloprotease-4 (AY442287) does not contain 8 aminoacids in the N-terminus including a cysteine that is conserved in most P-II enzymes. Likewise, these amino acids are missing in a metalloprotease from *Macrovipera lebetina* (gi 2118144) (Fig. 3) suggesting that the amino acid changes we have observed are consistent. Of note, the N-terminus NSAHPCCDPVTXK (BG-12) obtained by Edman degradation of *B. gabonica* venom proteins (Fig. 1) is identical to the putative N-terminus coded by the corresponding cDNA in contig 34. This finding strongly suggests that this P-II metalloprotease when processed may generate disintegrin peptides. We could not identify the N-terminus of the metalloprotease domain in our proteome study (Fig. 1), suggesting that these proteins may have their N-terminus blocked as previously reported (Fox et al., 2002). The CLUSTAL alignment of *B. gabonica* P-II metalloprotease (*B. gabonica* metalloprotease-4; AY442287) with other P-II class enzymes is shown in Fig. 3A. The unrooted cladogram is presented in Fig. 3B and shows that metalloproteases from *B. gabonica* and *M. lebetina* venoms are the most closely related enzymes.

Table 1 shows that a number of other contigs (e.g. contig 16, AY430411 and 30, AY430412) contains the partial-length

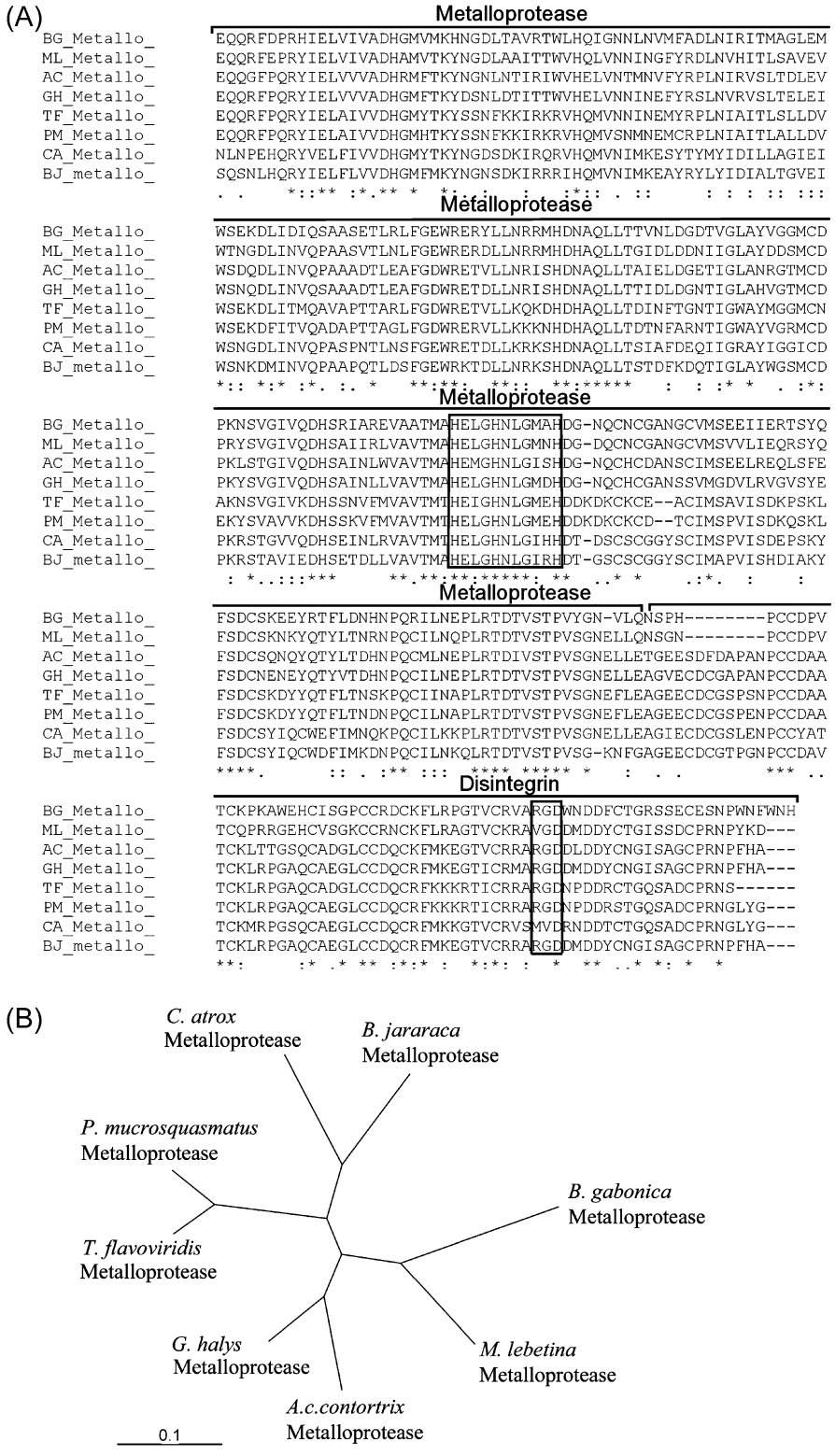

Fig. 3. (A) Alignment of *B. gabonica* P-II metalloprotease. Proteins were deduced from a *B. gabonica* venom gland cDNA library. Asterisks, colons, and stops below the sequences indicate identity, high conservation, and conservation of the amino acids, respectively. (B) Unrooted cladogram indicating the families of P-II snake venom metalloprotease. The bar represents the degree of divergence among sequences. BG_metallo (*B. gabonica* metalloprotease-4; AY442287); ML-metallo (*M. lebetina* metalloprotease; gi 2118144); AC_metallo (*A. c. contortix* metalloprotease; gi 7630286); GH_metallo (*G. halys* metalloprotease; 4106005); TF_metallo (*T. flavoviridis* metalloprotease; gi 14595995); PM_metallo (*P. mucrosquamatus* metalloprotease; gi 995748); CA_metallo (*C. atrox* metalloprotease; gi 462320); and BJ_metallo (*B. jararaca* metalloprotease; gi 13194760). The conserved zinc-binding domain is boxed.

sequences homologous to P-III metalloproteases from other venoms (Jia et al., 1997). In addition, contig 172 (AY430412) contains the full-coding region for the disintegrin-like and cysteine- rich domains of a typical P-III venom metalloproteases. In fact, the protein coded by this cDNA has a SECD motif in the disintegrin-like domain in addition to a conserved pattern of cysteines commonly found in the cysteine-rich region. This cDNA codes for a protein that can be aligned (not shown) with the disintegrin/cysteine-rich domains of other P-III metalloproteases including berythractivase, a prothrombin activator from *Bothrops erythromelas* and other related molecules (Silva et al., 2003). These enzymes may participate together with other venom components in the pathogenesis of *B. gabonica* envenomation (Marsh et al., 1997).

3.3.2. Serine proteases

In addition to metalloproteases, crotalid and viperid species contain large amounts of serine proteases (Markland, 1998). In most cases, theses enzymes have 12 cysteines strongly conserved in addition to a catalytic triad characteristic to serine proteases-His57-Asp102-Ser195 (Castro et al., 2001). These enzymes are frequently blocked by serine protease inhibitors and preferentially hydrolyze the α chain of fibrinogen over the β chain and/or induce platelet aggregation (Markland, 1998).

Among the serine proteinases found in our cDNA library, contig 71 encodes a protein similar to platelet pro-aggregatory PA-BJ from *Bothrops jararaca* (Serrano et al., 1995) and thrombocytin from *B. atrox* (Niewiarowski et al., 1979). Consistent with these contigs, we have found the sequence VIGXAEXDINEHPSLALIY for BG-8 and VIGXAEXNINEHRFLALVYF for BG-9 to be similar to the N-terminus of PA-BJ (VVGGRPCKINVHPSLVLL). It also resembles the sequence VVGGAGECKIDGHRCLA LLY described for gabonase, a pro-coagulant thrombin-like enzyme from *B. gabonica* (Pirkle et al., 1986). We could assign contigs with matches to enzymes that cleave fibrinogen (like gabonase) but are devoid of platelet aggregatory properties. In addition, contigs 60 and 90 have matches to β-fibrinogenases from *Vipera lebetina* (gi 2241722) and *Agkistrodon blomhoffi* (gi 6706013) snake venoms, respectively. The cDNAs identified in contig 60 (*B. gabonica* serine protease-1; AY430410) was completely sequenced and the CLUSTAL alignment with other venom serine proteases is shown in the Fig. 6 of the supplemental data. Interestingly, in this protein the catalytic triad His 57 is replaced by Arg 57, and Ser 195 is replaced by Asp195. Identical substitutions have been found in the serine protease VLP2 from *V. lebetina* venom and it has been suggested that generation of such clones occurs via trans-splicing of the primary gene transcript, by exon shuffling or by unequal crossing-over on the genome level (Siigur et al., 2001). Since these proteins have not been expressed as recombinant protein it is a matter of debate whether it behave as serine proteases.

Finally, it is noteworthy that contigs 60, 90, 71, 108, and 259 match kallikrein-like enzymes in the gene ontology database, indicating that these serine proteases may act on kininogen to release bradykinin. This conclusion is consistent with reports showing that *B. gabonica* venom serine protease activities can be separated into kinin-releasing, clotting, and fibrinolytic activities (Viljoen et al., 1979).

3.3.3. Kunitz- and cystatin-type protease inhibitors

Kunitz domains are about 60 residues and contain six specifically spaced cysteines that form disulfide bonds. In most cases, they are reversible inhibitors of serine proteases that bind the active site. In our library, we have found contig 203 with sequence homology to textilin, a Kunitz-type protease inhibitor that tightly inhibits plasmin and is supposed to have anti-hemorrhage or pro-thrombotic activity (Aird, 2002). Consistent with these data, we have found by Edman degradation that BG-11 and BG-15 protein bands share a similar sequence: KNRPEFX NLPADTGXXKAY and KKRPDFXYLPADTGPXMAN, respectively. These sequences match the N-terminus KDRPKFCELPADIG reported for textilin (gi 15321630). The full-length clones of two Kunitz-protease inhibitors from *B. gabonica* venom gland have been obtained and called Bitisilin-1 (AY430402) and Bitisilin-2 (AY430413). The CLUSTAL alignment of both sequences with other venom Kunitz inhibitors and unrooted cladogram of all sequences is presented in the supplemental data. Of interest, a third cDNA (contig 137) codes for a molecule containing at least two Kunitz domains organized in tandem and called herein Bitisilin-3 (AY442289). Although multi-Kunitz molecules from exogenous sources have been identified in the salivary gland of ticks (gi 15077001), this is the first description of a multi-headed Kunitz in snake venoms. Both heads are highly homologous, and are most likely the result of gene duplication from a common ancestor (Zupunski et al., 2003). Consistent with these data, we have found two protein bands with clearly distinct molecular weights (BG-11 and BG-15) with N-terminus that matches Kunitz inhibitors (Fig. 1). Finally contig 146 has sequence homology to α-bungarotoxin from *Bungarus candidus* (gi 24459200), a well-studied Kunitz type K^+ channel blocker from *Bungarus* spp. (Harvey, 2001). The CLUSTAL alignment of *B. gabonica* and other venom Kunitz-like protein and the unrooted cladogram of all sequences are shown in the Fig. 7 of the supplemental data.

We have also found that contig 264 assigns for cystatin-like molecules. Cystatins are tight and reversible inhibitors of the cysteine proteinases and are present in a variety of mammalian and non-mammalian tissues including snake venoms (Aird, 2002). According to our library, *B. gabonica* also contains the full-length clone coding for a cystatin-like protein called herein Bitiscystatin (AY430403). In addition, BG-13 sequence KVGXLYXRDVMDPEVQXAA is similar to the N-terminus of *B. arietans* cystatin (gi 118194). The CLUSTAL alignment of *B. gabonica* and other cystatins and the unrooted cladogram of all sequences are shown in the Fig. 8 of the supplemental data.

3.3.4. C-type lectins

C-type lectins are molecules containing a carbohydrate-recognition domain (CRD). Most C-type lectins are Ca^{2+} dependent; however, many of them have lost their sugar-binding properties and have evolved to interact with platelet receptors and/or blood coagulation factors (Markland, 1998). Notably, snake venoms are a rich source of C-type lectins, and not surprisingly our library also contains a large amount of cDNA coding for this family of proteins. Among the cDNA we have sequenced, contig 218 assigns for a fibrinogen-clotting inhibitor from *Gloydius halys brevicaudus* (gi 4337050) and contig 216 assigns for Factor IX/X-Binding protein from *Trimeresurus flavoviridis*. Also, contig 215 assigns for a GPIb agonist from *Agkistrodon acutus* and may affect platelet function either by direct agglutination of platelets, through binding to von Willebrandt factor (Matsui et al., 2002). Finally, contig 15 assigns for a galactose-binding lectin from the venom of *Trimeresurus stejnegeri* (gi 7674107). Consistent with these results described above, we have found for BG-4 the sequence DFEXPSEWSAYGXHXYRAF in addition to BG-5 (DQGXLPDWSAYE QHXY), BG-7 (DEGXLPGWSLYE), and BG-2 (DFGXLSDWSXYEQH) that resembles the N-terminus of a C-type lectin from *B. arietans* DFQCPSEWSAYGQHCYR (Harrison et al., 2003). BG-3 (DFGA) and BG-10 (DQGALPDTSYHQHHYYP) are also similar to *B. arietans* C-type lectin in addition to the DQDCLPDWSSHERHCY N-terminus of *Echis pyramidum leakeyi* C-type lectin (gi 33243102). The full-length clones of three *B. gabonica* C-type lectin have been obtained and named *B. gabonica* C-type lectin-1 (AY439477), *B. gabonica* C-type lectin-2 (AY429478), and *B. gabonica* C-type lectin-3 (AY429479). The CLUSTAL alignment of *B. gabonica* and other venom C-type lectins and the unrooted cladogram of all sequences are shown in the Fig. 9 of the supplemental data.

3.3.5. PLA2

Snake and other venoms are rich sources of PLA2 (E.C.3.1.1.4), a family of enzymes known to have edematogenic, antiplatelet, anticoagulant, mast cell degranulating, or neurotoxic properties (Bon et al., 1994). On the basis of primary structure and disulfide bond pairings, snake venom PLA2s were classified as type I (Elapidae) or class II PLA2s (Viperidae/Crotalidae). The catalytic site of class II $PLAs_2$ contains a highly conserved aspartic acid or lysine at position 49 (Ownby et al., 1999).

In our library, contigs 1 and 3 assign to PLA2, similar to one described in *B. nasicornius* venom (gi 67204), whereas contig 222 matches a PLA2 described in *E. pyramidum leakeyi* (gi 27734438). The cDNA sequenced in this library code for Lys49-PLA2; no Asp49-PLA2 has been sequenced. The presence of a PLA2 protein in this venom was confirmed by the BG-14 sequence HLEQFGNMIDHVSGRSFWLY that is similar to the N-terminus DLTQFGNMIN previously reported for *B. gabonica* PLA2 (Botes and Viljoen, 1974). The full-length clone of *B. gabonica* PLA2 has been obtained and named *B. gabonica* PLA2-1 (AY430410). The CLUSTAL alignment of *B. gabonica* and Lys49-PLA2 and the unrooted cladogram of all sequences are shown in the Fig. 10 of the supplemental data.

3.3.6. Disintegrins

Disintegrins are cysteine-rich, low-molecular-weight platelet aggregation inhibitor polypeptides that usually contain an RGD sequences or other motifs that are recognized by integrins in different cell types (McLane et al., 1998). In most cases, venom disintegrin are encoded with a signal peptide, pre-peptide (pro-domain), metalloprotease, and disintegrin region on their common precursors (P-II class metalloproteases). It is suggested that the metalloprotease/disintegrin precursor is cleaved by protease(s), resulting in production of metalloprotease and disintegrin (Bjarnason and Fox, 1995; McLane et al., 1998). More recently, a new gene structure of the disintegrin family was identified in *Agkistrodon contortrix contortix* and *A. p. piscivourus* venoms and it consists of signal peptide, pre-peptide (pro-domain), a disintegrin domain and lacking the protease domain (Okuda et al., 2002).

In our library, contigs 119 and 127 assign, respectively, for disintegrin similar to acostatin from the venom of *A. contortrix contortrix* (Okuda et al., 2002), and eristochophin I from *Eriscocophis macmahonii* (gi 6225272). Interestingly, cluster 119 contains sequences that code for two proteins respectively called herein *B. gabonica* disintegrin-1 (gabonin-1, AY430904) and *B. gabonica* disintegrin-2 (gabonin-2, AY430505). Remarkably, these two protein sequences were identical except for nine amino acids that occurs between the cysteine residues that form the putative acidic hairpin loop where the disintegrin domain is found. One of these sequences contains a typical RGD sequence known to bind to β3 integrins (McLane et al., 1998), whereas the second sequence contains the motif MLDG, known to interact with integrin α9β1 and to affect neutrophil function (McLane et al., 1998). Since a typical signal peptide and a pre-peptide region were found for both gabonin-1 and -2, it is clear that these proteins together with acustatin and piscivostatin α chains are new members of the short coding region family of disintegrins (Okuda et al., 2002). The CLUSTAL alignment of gabonin-1 and -2 with acostatin and piscivostatin is shown in Fig. 4A. The schematic domain structure of this family of protein is shown in Fig. 4B (Okuda et al., 2002).

Consistent with these contigs, Edman degradation of the protein band BG-12 yields the sequence NSAHPXXDPV TXK that matches the N-terminus NSANPCCDPITCK of eristocophin (gi 265034). Since BG-12 N-terminus also matches the N-terminus found for the disintegrin domain of *B. gabonica* metalloprotease-4 (Fig. 3A), it is unclear whether the protein we have identified as a disintegrin is a processed form of a *B. gabonica* venom P-II metallopro-

(A)

```
                    Signal peptide          Pre-peptide
BG_Disintegrin_1_   MIQVLLVIICLAVFPYQGSSIILESGNVNDYEIVYPKKVTVLPTGAMNSAHPCCDPVTCK
BG_Disintegrin_2_   MIQVLLVIICLAVFPYQGSSIILESGNVNDYEIVYPKKVTVLPTGAMNSAHPCCDPVTCK
AC_Disintegrin_     MIQVLLVTLCLAVFPYQGSSIILESGNVNDYEVVYPRKVTALPKGAIQPKNPCCDAATCK
AP_Disintegrin_     MIQVLLVTICLAVFPYQGSSIILESGNVNDYEVVYPRKITPLPKGAVQPKNPCCDAATCK
                    ******* :**********************:***:*:* **.**::. :****..***
                                        Disintegrin
BG_Disintegrin_1_   PKRGEHCISGPCCRNCKFLNAGTICKRGRGDSLHDY-CTGVTPDCPRNPNKGESDELEWS
BG_Disintegrin_2_   PKRGEHCISGPCCRNCKFLNAGTICKKTMLDGLNDY-CTGVTPDCPRNPNKGESDELEWS
AC_Disintegrin_     LTPGSQCAEGLCCDQCKFIKAGKICRRARGDNP-DYRCTGQSGDCPR---K------HFY
AP_Disintegrin_     LTPGSQCAEGLCCDQCKFIKAGKICRRARGDNP-DYRCTGQSGDCPR---K------HFY
                     . *.:* .* ** :***::**.**::   *.  ** *** : ****   *      .:

                    ┌──────┐
BG_Disintegrin_1_   AAATGSVLM
BG_Disintegrin_2_   AAATGSVLM
AC_Disintegrin_     A--------
AP_Disintegrin_     A--------
                    *
```

(B)

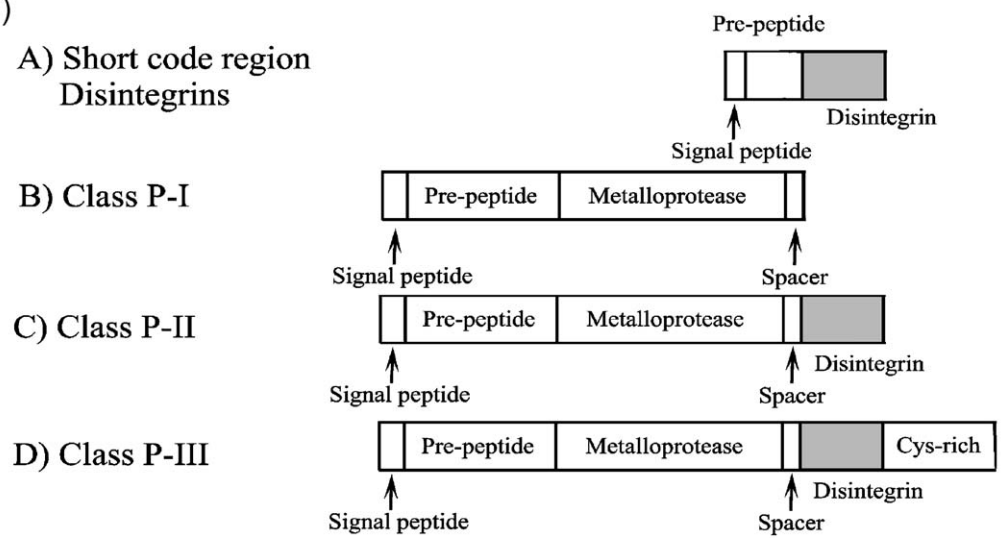

Fig. 4. (A) Alignment of *B. gabonica* short code region disintegrins. See Fig. 3 legend for an explanation of the layout. (B) Schematic domain structure of short code region disintegrins, P-1, P-II and P-III snake venom metalloproteases. BG_Disintegrin_1 (*Bitis gabonica* disintegrin 1 or Gabonin-1, AY430404); BG_Disintegrin_2 (*Bitis gabonica* disintegrin 2 or Gabonin-2, AY430405); AC_Disintegrin (Acostatin α chain from *A. c. contortrix*; gi 27922986); and AP_Disintegrin (Piscivostatin α chain from *A. p. piscivorus*; gi 27922990).

tease, a short code region disintegrin, or both. Finally, the finding that *B. gabonica* contains disintegrins reinforces the notion that *B. gabonica* venom targets hemostasis and may also indicate that the hemostatic disturbance found after *B. gabonica* envenomation is mediated, at least in part, by these molecules.

3.3.7. LAO

LAO are widely found in snake venoms and are thought to contribute to toxicity upon envenomation. It has been shown that these enzymes affect platelets, induce apoptosis, and have hemorrhagic effects (Aird, 2002). In our library, contig 165 (AY434453) assigns for a truncated clone coding for proteins with sequence homology to apoxin-1, a LAO and apoptosis inducer from *C. atrox* venom (gi 5565692). Contigs 181 and 182 are similar to LAO with platelet and coagulation inhibitory properties isolated from *Agkistrodon halys blomhoffii* (gi 15887054). In addition, the N-terminus of BG-6, ADDKNPLEEXFRESSYEEFL is almost identical to the N-terminus ADDRNPLEECFRETDYEEFL of LAO from *A. halys blomhoffii* venom (gi 15887054), confirming the presence of this family of enzymes in *B. gabonica* venom. It remains to be demonstrated how LAO from *B. gabonica* venom affect hemostasis.

3.3.8. Nucleotidases and nucleases

Snake venoms are a rich source of nucleotidases, and their participation in envenomation has been reviewed

recently (Aird, 2002). In our library, contig 40 has a truncated clone whose sequence is similar to the sequence coding for an ectonucleotidase from the electric ray electric lobe (gi 112824). Nucleotidases inhibit platelet aggregation, and it appears that *B. gabonica* may affect platelet function by removal of ADP. We have also identified cDNA coding for endonucleases, a family of enzymes ubiquitously found in snake venoms. Venom endonucleases work together with venom and endogenous phosphodiesterase degrading nucleic acids to free nucleotides, which serve as substrate for 5′ nucleotidases, which, in turn, liberate free nucleosides. Adenosine, in particular, is a potent vasodilator and inhibitor of platelet aggregation (Aird, 2002).

3.3.9. Growth factors

In our library, we have found cluster 227 that match VEGF from *Gallus gallus* (gi 27368068). The VEGF are the most potent vascular permeability factors known and characteristically cause reversible increase in permeability and have been described in venoms (Aird, 2002). The full-length clone of *B. gabonica* VEGF has been obtained and named *B. gabonica* VEGF (AY429481). This protein may be involved in edema induced by *B. gabonica* bite. The CLUSTAL alignment of *B. gabonica* and other venom VEGF and the unrooted cladogram of all sequences are shown in the Fig. 11 of the supplemental data. In addition, cluster 219 (AY430406) in our library is similar to NGF from *B. jararacussu* venom (gi 15407254). NGF is ubiquitous in snake venoms and exhibit non-neuronal effects such as the induction of plasma extravasation and histamine release from whole blood cells. Although we could not find the N-terminus of growth factors in any of the bands shown in Fig. 1, this paper describes for the first time transcripts for this family of proteins in *B. gabonica* venom gland.

3.3.10. BPP

We have found an abundant contig 188 containing 28 truncated cDNA (AY434452) whose sequence has matches to the 3′ untranslated region of BPP from *A. halys blomhoffi* (gi 427226). BPP were first isolated in the venom of *B. jararaca* snake and shown to display intense hypotensive properties (Aird, 2002). We suggest that this family of peptides is involved in hypotension associated with *B. gabonica* envenomation.

3.3.11. Cytokine-like and unknown proteins

We have sequenced other cDNA (contig 28) whose sequences are similar to a cytokine-like protein that inhibits insulin secretion (Zhu et al., 2002). This is the first description of this family of proteins in snake venoms. By immunohistochemistry it was shown that a member of this cytokine family was expressed prominently in the vascular endothelium, particularly in capillaries (Zhu et al., 2002). The function of this cytokine-like protein in snake venom remains to be determined, but it may be that it somehow affects vascular biology. The full-length clone of *B. gabonica* cytokine-like protein has been obtained and named *B. gabonica* cytokine-like protein-1 (AY429480). The CLUSTAL alignment of *B. gabonica* and other cytokine-like proteins and the unrooted cladogram of all sequences are shown in Fig. 5.

Finally, no matches have been found for some clusters, and these were assigned as unknowns (not shown, available on request). In some selected cases, we have named hypothetical proteins (HP) when a sequence without database hits has an open-reading frame (ORF) containing metionine, a stop codon and a putative signal peptide (Table 2).

3.4. A catalog toward the full-length cDNA and proteins from the snake venom gland of B. gabonica

To gather the maximum amount of information about the putative secreted proteins from the *B. gabonica* venom gland, selected sequences presented in Table 1 were re-sequenced and extended to obtain, when applicable and possible, their full-length cDNA. The full coding sequences with database hits were then blasted again to the NR protein database and SignalP server to confirm, respectively, sequence similarity and the presence of a signal peptide. In the event a signal peptide was predicted to exist, the molecular weight and the pI of the mature protein were also calculated and the putative function annotated. Most of the sequences displayed in Table 2 are full-length clones, with the exception of the metalloproteases, LAO, and BPP (see below). It may be that the cDNA coding for these proteins have an *Sfi*I site that is purposely cleaved during the cDNA library construction (see Materials and methods). Although our library is PCR-amplified, it is clear that the base changes observed in different contigs, including contig 34 or 60 and others are not artefactual. In fact, similar base changes have been found for all individual sequences of a given contig; actually, this diversity can be explained by accelerated evolution that has been well-documented in snake venom glands (Deshimaru et al., 1996). It is also known that PCR based libraries which cDNAs have not been size-fractioned may be enriched with small cDNAs. However, our libraries have been constructed using low, medium and high molecular weight cDNAs that have been separated by gel-filtration (see Material and methods in the Supplemental data). This separation minimizes the preferential amplification of small transcripts over larger ones, and the preferential ligation of small-sized cDNAs over larger ones, in the TripleX2 vector. Of note, the putative proteins coded by the most abundant clusters have been identified in the SDS/PAGE (e.g. PLA2, protease inhibitors, C-type lectins, serine protease, and disintegrins) with the exception of the metalloproteases, which N-terminus are found to be frequently blocked (Fox et al., 2002), and the BPP that are nor appropriately separated by 4–12% PAGE due to its low molecular weight. Accordingly, PCR-based libraries appear to provide a reasonable qualitative estimate of the transcripts expressed in a given tissue. Alternatively, construction of a

(A)

```
BG_Cytokine_like_      ----------------------------------------MRTNCKNFMKFGGFIFASVCA
MM_Cytokine_like_      ----------------------------------------MRPVATGIFKALVFIFSSLCA
RN_Cytokine_like_      ------------MGTAIGCRRDPRVS---IVKFDISRNSCHLSLGIFKALLFIFSSLCA
HS_Cytokine_like_      MRPLAGGSRWACWLTRCLISCFDINVQGRLLVKLRPKPTANTTCPGLLKVVFVVFASLCA
                                                          . .::*   .:*:*:**

BG_Cytokine_like_      WYLGTLLAYLVPQDPLEFDQL----LPEKPKIYMPAPKRQKCDHWSPCLPGNYAYRIVSG
MM_Cytokine_like_      WYSGYLLAELIPDVPLSSTLYNIRSIGERPVLKAPAPKRQKCDHWSPCPPDTYAYRLLSG
RN_Cytokine_like_      WYSGYLLAELIPDVPLSNTLYNIRSIGERPVLKAPVPKRQKCDHWSPCPPDTYAYRLLSG
HS_Cytokine_like_      WYSGYLLAELIPDAPLSSAAYSIRSIGERPVLKAPVPKRQKCDHWTPCPSDTYAYRLLSG
                       ** * *** *:*: **.        : *:* :  *.*********:** ...****::**

BG_Cytokine_like_      GGKERFPKICFEDKLLLSKQEGNAGRGINIAIVDYKTGKVAAANYFDMWTGEFSIPMTDF
MM_Cytokine_like_      GGRDKYAKICFEDEVLIGEKTGNVARGINIAVVNYETGKVIATKYFDMYEGDNSGPMAKF
RN_Cytokine_like_      GGLDKYAKICFEDEMLLGEKTGNVARGINIAVVDYETGKVISTKYFDMYEGDNSEPMTKF
HS_Cytokine_like_      GGRSKYAKICFEDNLLMGEQLGNVARGINIAIVNYVTGNVTATRCFDMYEGDNSGPMTKF
                       ** .::.******::*:.:: **..******:*:* **:* ::. ***: *: * **:.*

BG_Cytokine_like_      LNKTPQKSLIFIVTHDDGSTKLKEDGKKSNSRMGSKEIWNLKFRSGWVFLAAEGFKLPEN
MM_Cytokine_like_      IQSTPSKSLLFMVTHDDGSSKLKAQAKDAIEALGSKEIKNMKFRSSWVFVAAKGFELPSE
RN_Cytokine_like_      IQSIPSKSLLFMVTHDDGSHKLQAQAKDAIEALGSKEIKNIKFRSSWVFVAAKGFELPSE
HS_Cytokine_like_      IQSAAPKSLLFMVTYDDGSTRLNNDAKNAIEALGSKEIRNMKFRSSWVFIAAKGLELPSE
                       ::. . ***:*:**:**** :*: :.*.: . :***** *:****.***:**:*::**.:

BG_Cytokine_like_      IQKEKIIRSDNSKNRYSGWPTEIQIEGCIPRNLTS
MM_Cytokine_like_      IEREKINHSDQSRNRYAGWPAEIQIEGCIPKGLR-
RN_Cytokine_like_      MEREKINHSDQSRNRYSGWPAEIQIEGCIPKGLR-
HS_Cytokine_like_      IQREKINHSDAKNNRYSGWPAEIQIEGCIPKERS-
                       :::*** :** ..***:***:*********:
```

(B)

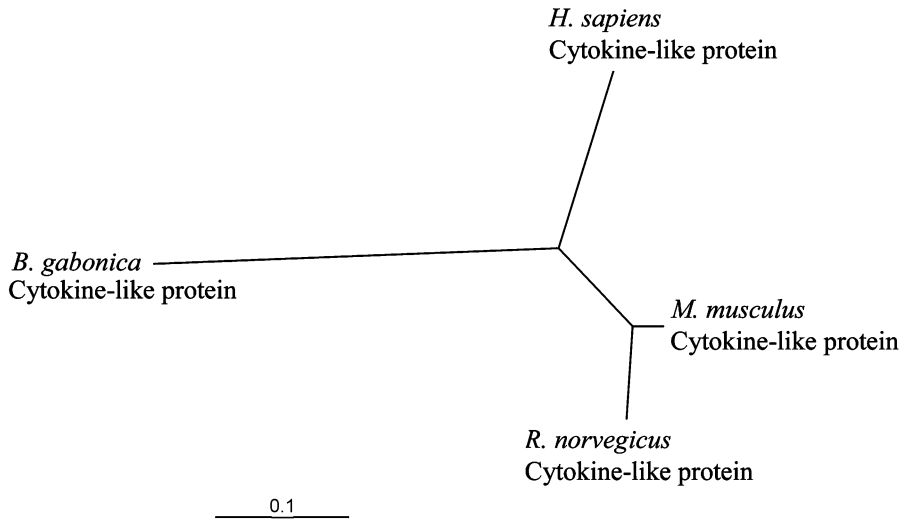

Fig. 5. (A) Alignment of *B. gabonica* cytokine-like protein and (B) unrooted cladogram. See Fig. 3 legend for an explanation of the layout. The bar represents the degree of divergence among sequences. BG-Cytokine-like (*Bitis gabonica* cytokine-like protein, AY429480); MM_cytokine_like (*Mus musculus* cytokine-like protein, gi 22296880); HS_cytokine_like_a (*Homo sapiens* cytokine-like protein, gi 21361798); RN_cytokine_like protein (*Rattus norvegicus* cytokine-like protein, gi 34867988); and HS_cytokine_like_b (*Homo sapiens* cytokine-like protein, gi 23821535).

normalized *B. gabonica* cDNA library could be a useful strategy to follow in an attempt to identify rare transcripts that have been eventually missed in our library. Likewise, separation of venom proteins by 2-D gel followed by Edman degradation may well complement the data obtained herein using one-dimensional PAGE (Fox et al., 2002).

The summary of our findings is presented in Table 2. To our knowledge, this table is the first attempt to create a comprehensive catalog of the cDNA from the *B. gabonica* snake gland. Eventually, such a catalog will contain a non-redundant set of full-coding cDNA sequences covering every *B. gabonica* venom gland cDNA and possibly each venom protein function. Thus, this transcript and protein catalog for *B. gabonica* and other snakes could form part of a large-scale and comprehensive functional analysis of snake venom genes and cDNA. Together with information derived from the venom gland genome, proteome (Fox et al., 2002), and microarrays (Gallagher et al., 2003), infor-

Table 2
A catalog of *Bitis gabonica* venom gland cDNAs for secretory proteins

Sequence name[a]	Contig[b]	Clone[c]	GenBank[d]	N[e]	MW[f]	SP[g]	MW[h]	pI[i]	Function[j]
Bradykinin potent.peptide	188	partial	AY43452	Y	NA	NA	NA	NA	Hypotension
C-type lectin-1	15	full-length	AY439477	Y	18624	23	16088	8.22	Anti-hemostatic
C-type lectin-2	215	full-length	AY429478	Y	18156	21	15959	5.17	Anti-hemostatic
C-type lectin-3	214	full-length	AY429479	Y	18094	23	15640	5.98	Anti-hemostatic
Cystatin (Bitiscystatin)	264	full-length	AY430403	Y	15899	24	13325	7.01	Protease inhibitor
Ctokine-like protein	28	full-length	AY429480	Y	26411	29	23138	9.07	Unknown
Disintegrin-1 (Gabonin-1)	119	full-length	AY430404	Y	13791	20	11614	7.06	Plate inhibitor
Disintegrin-2 (Gabonin-2)	119	full-length	AY430405	Y	13785	20	11608	6.31	Plate inhibitor
VEGF[l]	222	full-length	AY429481	Y	22357	26	19303	7.8	Edema inducer
Hypothetical protein-1	99	full-length	AY430407	Y	9740	13	8312	9.55	Unknown
Hypothetical protein-2	103	full-length	AY430408	Y	7602	15	5719	7.24	Unknown
Kunitz inhibitor-1 (Bitisilin-1)	203	full-length	AY430402	Y	9922	24	7498	8.26	Protease inhibitor
Kunitz inhibitor-2 (Bitisilin-2)	146	full-length	AY43013	Y	10005	24	7581	9.3	Protease inhibitor
Kunitz inhibitor-3 (Bitisilin-3)	137	partial	AY442289	Y	NA[k]	NA	NA	NA	Protease inhibitor
L-amino acid oxidase	147	partial	AY434453	Y	NA	NA	NA	NA	Anti-hemostatic
Metalloprotease-1	16	partial	AY430411	Y	NA	NA	NA	NA	Hemmorhagic
Metalloprotease-2	30	partial	AY430412	Y	NA	NA	NA	NA	Hemmorhagic
Metalloprotease-3	172	partial	AY442288	Y	NA	NA	NA	NA	Hemmorhagic
Metalloprotease-4	34	partial	AY442287	Y	NA	NA	32506	5.76	Hemmorhagic
Nerve Growth Factor	219	partial	AY430406	Y	NA	NA	NA	NA	Edema inducer
Phospholipase A2	1	full-length	AY429476	Y	15745	16	13932	4.99	Edema inducer
Serine-protease-1	60	full-length	AY430410	Y	16803	24	14104	9.07	Pro-hemostatic?

[a] Including putative secretory proteins only.
[b] Contig, contig number.
[c] Clone (partial of full-length).
[d] Genbank, NR database accession number.
[e] N, novel cDNA (Y = yes).
[f] MW.; molecular mass before signal peptide removed.
[g] SP-signal peptide.
[h] MW.; molecular mass of the mature protein.
[i] pI, Isoelectric point.
[j] Function, putative function or possible biological activity.
[k] NA, Not available.
[l] VEGF, Vascular Endothelial Growth Factor.

mation provided by this catalog could be an essential tool to understand snakes physiology (Perales and Domont, 2002), the molecular basis of envenomation, as well as to find potential candidates for serum production (Theakston et al., 2003) and/or tools to study cell biology and biochemistry (Ménez, 1998).

3.5. *B. gabonica* venom components and envenomation

Snake venom envenomation employs three well-integrated strategies including prey immobilization via hypotension, prey immobilization via paralysis, and prey digestion (Aird, 2002). Although the identification of the toxin clusters does not allow us to determine quantitatively the contribution of each protein cluster in the envenomation, it allow us to speculate about the mechanisms of envenomation by *B. gabonica* venom. It is remarkable that proteins such as metalloproteases, serine proteases, C-type lectins, PLA2, Kunitz inhibitors, growth factors, and LAO account for most of our sequences. As described above and reviewed elsewhere (Aird, 2002), these proteins act on the hemostatic system and/or affect vascular biology. In this respect, *B. gabonica* venom resembles an expressed sequence tag (EST) approach reported for *Bothrops insularis*, where a large number of cDNA code for metalloproteases, BPP, C-type lectins, serine protease, PLA2, and growth factors (Junqueira-de-Azevedo and Ho, 2002). We have also found an abundant cluster whose sequences match the 3′ untranslated region cDNA of *A. halys blomhoffi* BPP, a family of peptides also abundant in the *B. insularis* cDNA library (Junqueira-de-Azevedo and Ho, 2002) (see Table 2). Because we also found sequences coding for kallikrein-like enzymes, it is plausible that these enzymes and BPP are primarily responsible for the hypotension associated with *B. gabonica* and possibly *B. insularis* envenomation. The similarity in the cDNA composition between *B. gabonica* and *B. insularis* libraries is also consistent with the symptoms resulting from envenomation by *Bitis* and *Bothrops* spp. that is characterized by consumption coagulopathy, hypotension, and local damage (Aird, 2002).

4. Conclusion

It is worth noting that the description of the *B. gabonica* venom gland cDNA database match biologic activities

described before for this venom, including the molecules involved with hypotension, bleeding, digestion, and tissue damage (Marsh et al., 1997). This indicates that an approach combining cDNA library construction, massive sequencing, and bioinformatic analysis, in addition to Edman degradation of the main proteins, may be useful to study exogenous secretion from different venom glands, and to the development of recombinant antigens for antibody production.

Acknowledgements

We thank Drs. Thomas E. Wellems, Robert W. Gwadz and Thomas J. Kindt for encouragement and support, and Brenda Rae Marshall for editorial assistance.

References

Aird, S.D., 2002. Ophidian envenomation and the role of the purines. Toxicon 40, 335–393.

Bjarnason, J.B., Fox, J.W., 1995. Snake venom metalloendopeptidases: reprolysins. Methods Enzymol. 248, 345–368.

Bon, C., Choumet, V., Delot, E., Faure, G., Robbe-Vincent, A., Saliou, B., 1994. Different evolution of phospholipase A2 neurotoxins (beta-neurotoxins) from Elapidae and Viperidae snakes. Ann. N.Y. Acad. Sci. 710, 142–148.

Botes, D.P., Viljoen, C.C., 1974. *Bitis gabonica* venom. The amino acid sequence of phospholipase A. J. Biol. Chem. 249, 3827–3835.

Castro, H.C., Silva, D.M., Craik, C., Zingali, R.B., 2001. Structural features of a snake venom thrombin-like enzyme: thrombin and trypsin on a single catalytic platform? Biochim. Biophys. Acta 1547, 183–195.

Deshimaru, M., Ogawa, T., Nakashima, K., Nobuhisa, I., Chijiwa, T., Shimohigashi, Y., Fukumaki, Y., Niwa, M., Yamashina, I., Hattori, S., Ohno, M., 1996. Accelerated evolution of crotalinae snake venom gland serine proteases. FEBS Lett. 397, 83–88.

Fox, J.W., Shannon, J.D., Stefansson, B., Kamiguti, A.S., Theakston, R.D.G., Serrano, S.M.T., Camargo, A.C.M., Sherman, N., 2002. Role of discovery sciences in toxinology: examples in venom proteomics. In: Ménez, A. (Ed.), Perspectives in molecular toxinology. Wiley, West Sussex, UK, pp. 97–106.

Francischetti, I.M., Valenzuela, J.G., Pham, V.M., Garfield, M.K., Ribeiro, J.M.C., 2002. Toward a catalog for the transcripts and proteins (sialome) from the salivary gland of the malaria vector *Anopheles gambiae*. J. Exp. Biol. 205, 2429–2451.

Gallagher, P.G., Bao, Y., Serrano, M.T., Kamiguti, A.S., Theakston, R.D.G., Fox, J.W., 2003. Use of microarrays for investigating the subtoxic effects of snake venoms: insights into venom-induced apoptosis in human umbilical vein endothelial cells. Toxicon 41, 429–440.

Harrison, R.A., Oliver, J., Hasson, S.S., Bharati, K., Theakston, R.D., 2003. Novel sequences encoding venom C-type lectins are conserved in phylogenetically and geographically distinct Echis and Bitis viper species. Gene 315, 95–102.

Harvey, A.H., 2001. Twenty years of dendrotoxins. Toxicon 39, 15–26.

Huang, T.F., Peng, H.C., Peng, I.S., Teng, C.M., Ouyang, C., 1992. An antiplatelet peptide, gabonin, from *Bitis gabonica* snake venom. Arch. Biochem. Biophys. 298, 13–20.

Jia, L.G., Wang, X.M., Shannon, J.D., Bjarnason, J.B., Fox, J.W., 1997. Function of disintegrin-like/cysteine-rich domains of atrolysin A. Inhibition of platelet aggregation by recombinant protein and peptide antagonists. J. Biol. Chem. 272, 13094–13102.

Junqueira-de-Azevedo, I.L., Ho, P.L., 2002. A survey of gene expression and diversity in the venom glands of the pitviper snake *Bothrops insularis* through the generation of expressed sequence tags (ESTs). Gene 299, 279–291.

Kini, R.M., Joseph, J.S., Rao, V.S., 2002. Prothrombin activator from snake venoms. In: Ménez, A. (Ed.), Perspectives in Molecular Toxinology. Wiley, West Sussex, UK, pp. 341–355.

Markland, F.S., 1998. Snake venoms and the hemostatic system. Toxicon 36, 1749–1800.

Marsh, N., Gattullo, D., Pagliaro, P., Losano, G., 1997. The Gaboon viper, *Bitis gabonica*: hemorrhagic, metabolic, cardiovascular and clinical effects of the venom. Life Sci. 61, 763–769.

Matsui, T., Hamako, J., Matsushita, T., Nakayama, T., Fujimura, Y., Titani, K., 2002. Binding site on human von Willebrand factor of bitiscetin, a snake venom-derived platelet aggregation inducer. Biochemistry 41, 7939–7946.

McLane, M.A., Marcinkiewicz, C., Vijay-Kumar, S., Wierzbicka-Patynowski, I., Niewiarowski, S., 1998. Viper venom disintegrins and related molecules. Proc. Soc. Exp. Biol. Med. 219, 109–119.

Ménez, A., 1998. Functional architectures of animal toxins: a clue to drug design? Toxicon 36, 1557–1572.

Niewiarowski, S., Kirby, E.P., Brudzynsky, T.M., Stocker, K., 1979. Thrombocytin, a serine protease from *Bothrops atrox* venom. 2. Intercation with platelets and plasma clotting factors. Biochemistry 18, 3570–3577.

Okuda, D., Koike, H., Morita, T., 2002. A new gene structure of the disintegrin family: a subunit of dimeric disintegrin has a short coding region. Biochemistry 41, 14248–14254.

Ownby, C.L., Selistre de Araujo, H.S., White, S.P., Fletcher, J.E., 1999. Lysine 49 phospholipase A2 proteins. Toxicon 37, 411–445.

Perales, J., Domont, G.B., 2002. Are inhibitors of metallopoteases, phospholipases A2, and miotoxins members of the innate immune system? In: Ménez, A. (Ed.), Perspectives in Molecular Toxinology. Wiley, West Sussex, UK, pp. 435–456.

Pirkle, H., Theodor, I., Miyada, D., Simmons, G., 1986. Thrombin-like enzyme from the venom of *Bitis gabonica*. Purification, properties, and coagulant actions. J. Biol. Chem. 261, 8830–8835.

Serrano, S.M., Mentele, R., Sampaio, C.A., Fink, E., 1995. Purification, characterization, and amino acid sequence of a serine proteinase, PA-BJ, with platelet-aggregating activity from the venom of *Bothrops jararaca*. Biochemistry 34, 7186–7193.

Siigur, E., Aaspollu, A., Siigur, J., 2001. Sequence diversity of Vipera lebetina snake venom gland serine protease homologs—results of alternative-splicing or genome alteration. Gene 263, 199–203.

Silva, M.B., Schattner, M., Ramos, C.R., Junqueira-de-Azevedo, I.L., Guarnieri, M.C., Lazzari, M.A., Sampaio, C.A., Pozner, R.G., Ventura, J.S., Ho, P.L., Chudzinski-Tavassi, A.M., 2003. A prothrombin activator from *Bothrops erythromelas* (jararaca-da-seca) snake venom: characterization and molecular cloning. Biochem. J. 369, 129–139.

Theakston, R.D., Warrell, D.A., Griffiths, E., 2003. Report of a WHO workshop on the standardization and control of antivenoms. Toxicon 4, 541–557.

Viljoen, C.C., Meehan, C.M., Botes, D.P., 1979. Separation of *Bitis gabonica* (Gaboon adder) venom arginine esterases into kinin-releasing, clotting and fibrinolytic factors. Toxicon 17, 145–154.

Zupunski, V., Kordis, D., Gubensek, F., 2003. Adaptative evolution in the snake venom Kunitz/BPTI protein family. FEBS Lett. 547, 131–136.

Zhu, Y., Xu, G., Patel, A., McLaughlin, M.M., Silverman, C., Knecht, K., Sweitzer, S., Li, X., McDonnell, P., Mirabile, R., Zimmerman, D., Boyce, R., Tierney, L.A., Hu, E., Livi, G.P., Wolf, B., Abdel-Meguid, S.S., Rose, G.D., Aurora, R., Hensley, P., Briggs, M., Young, P.R., 2002. Cloning, expression, and initial characterization of a novel cytokine-like gene family. Genomics 80, 144–150.