JOURNAL OF MASS SPECTROMETRY
J. Mass Spectrom. 2007; **42**: 1405–1414
Published online 10 July 2007 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/jms.1242

JMS

Review

Snake venomics. Strategy and applications

Juan J. Calvete,* Paula Juárez and Libia Sanz

Instituto de Biomedicina de Valencia, C.S.I.C., Jaume Roig 11, 46010 Valencia, Spain

Received 31 March 2007; Accepted 26 April 2007

Snake bites can be deadly, but the venoms also contain components of medical and biotechnological value. The proteomic characterization of snake venom proteomes, snake venomics, has thus a number of potential benefits for basic research, clinical diagnosis, and development of new research tools and drugs of potential clinical use. Snake venomics is also relevant for a deep understanding of the evolution and the biological effects of the venoms, and to generate immunization protocols to elicit toxin-specific antibodies with greater specificity and effectiveness than conventional systems. Our snake venomics approach starts with the fractionation of the crude venom by reverse-phase HPLC, followed by the initial characterization of each protein fraction by combination of *N*-terminal sequencing, SDS-PAGE, and mass spectrometric determination of the molecular masses and the cysteine (SH and S–S) content. Protein fractions showing a single electrophoretic band, molecular mass, and *N*-terminal sequence can be straightforwardly assigned by BLAST analysis to a known protein family. On the other hand, protein fractions showing heterogeneous or blocked *N*-termini are analyzed by SDS-PAGE and the bands of interest subjected to automated reduction, carbamidomethylation, and in-gel tryptic digestion. The resulting tryptic peptides are then analyzed by MALDI-TOF mass fingerprinting followed by amino acid sequence determination of selected doubly and triply charged peptide ions by collision-induced dissociation tandem mass spectrometry. The combined strategy allows us to assign unambiguously all the isolated venom toxins representing over 0.05% of the total venom proteins to known protein families. Protocols and applications of snake venomics are reviewed and discussed. Copyright © 2007 John Wiley & Sons, Ltd.

KEYWORDS: snake venom; toxins; snake venomics; mass spectrometry; proteomics; cysteine residue; disulfide bond

WHY SNAKE VENOMICS?

Snakes of the family *Viperidae* (vipers and pitvipers) produce a complex mixture of a large number of distinct proteins[1,2] in paired specialized venom glands located ventral and posterior to the eyes. These venoms contain numerous proteins that interfere with the coagulation cascade, the normal haemostatic system, and tissue repair, and human envenomations are often characterized by clotting disorders, hypofibrinogenemia, and local tissue necrosis.[2,3]

In spite of the fact that viperid venoms may contain well over 100 protein components,[4] venom proteins belong to only a few major protein families, including enzymes (serine proteinases, Zn^{2+}-metalloproteases, L-amino acid oxidase, group II PLA_2) and proteins without enzymatic activity (ohanin, disintegrins, C-type lectins, natriuretic peptides, myotoxins, cysteine-rich secretory protein (CRISP) toxins, nerve and vascular endothelium growth factors, cystatin, and Kunitz-type protease inhibitors). This situation may reflect the fact that toxins were likely to be evolved from a restricted set of protein families with normal physiological functions that were recruited into the venom proteome before the diversification of the advanced snakes, at the base of the Colubroidea radiation.[5–8]

Given the central role that diet has played in the adaptive radiation of snakes,[9] venoms thus represent a key adaptation that has played an important role in the diversification of these animals. Venoms also represent the critical innovation in ophidian evolution that allowed advanced snakes to transition from a mechanical (constriction) to a chemical (venom) means of subduing and digesting prey larger than themselves, and as such, venom proteins have multiple functions including immobilizing, paralyzing, killing, and digesting prey. The existence in the same venom of a diversity of proteins of the same family which differ from each other in their pharmacological effects reflects an accelerated positive Darwinian evolution. Gene duplication followed by functional divergence is the main source of molecular novelty. Gene duplication creates redundancy and allows a gene copy to be selectively expressed in the venom gland, escaping the pressure of negative selection and evolving a new function through positive selection and adaptative molecular evolution.[10] The occurrence of multiple isoforms within each major toxin family evidences the emergence of paralogous groups of multigene families across taxonomic

*Correspondence to: Juan J. Calvete, Instituto de Biomedicina de Valencia, C.S.I.C., Jaume Roig 11, 46010 Valencia, Spain.
E-mail: jcalvete@ibv.csic.es

JMS

lineages where gene duplication events occurred prior to their divergence, and suggests an important role for balancing selection[11] in maintaining high levels of functional variation in venom proteins within populations.

In addition to understanding how venoms evolve, characterizing the large molecular variability within all the major toxin families may contribute to a deeper understanding of the biological effects of the venoms, and poses exciting challenges for delineating structure–function correlations and for designing *a la carte* antivenom production strategies.[12] Snake bite is still a serious threat in both developed and developing countries. Snake envenomation accidents represent a socio-medical problem of considerable magnitude with about 2.5 million people bitten by snakes annually around the world, of whom more than 100 000 lose their lives. The only effective treatment for systemic envenomation is the intravenous administration of an antivenom. Although antivenoms have gone a long way to reduce mortality, many of them do not achieve optimal protective effects. This is in part due to the fact that conventional antivenoms are prepared from sera of animals hyperimmunized with whole venom. The resulting polyclonal antisera include numerous antibodies with specificities not confined to the toxic target molecules. Hence, knowledge of the toxin composition of venoms could be devised to generate immunization protocols to elicit toxin-specific antibodies with greater specificity and effectiveness than conventional systems.

Paradoxically, although bites can be deadly, snake venoms also contain components of theraupeutic value. On relatively rare occasions, toxins represent potential therapeutic agents that have been used for the treatment of pathophysiological conditions in homeopathy, in folk remedies, and in Western and Chinese traditional medicine.[13,14]

To explore the putative venom components, several laboratories have recently carried out transcriptomic analyses of the venom glands of viperid (*Bothrops insularis*,[15] *Bitis gabonica*,[16] *Bothrops jararacussu*,[17] *Bothrops jararaca*,[18] *Agkistrodon acutus*,[19] *Echis ocellatus*,[20] and *Lachesis muta*[21]) and colubrid (*Philodryas olfersii*[22]) snake species. Transcriptomic investigations provide catalogues of partial and full-length transcripts that are synthesized by the venom gland. However, transcriptomes include translated and nontranslated mRNAs, as well as transcripts encoding nonsecreted, housekeeping, and cellular proteins, in addition to toxin precursor genes. Moreover, toxins may undergo post-translational processing and this event will not be evident in a transcriptomic analysis. Thus, outlining the full map of native toxins that actually constitute the venom requires a combined biochemical and proteomic approach. To address the need for detailed proteomic studies of snake venoms, we have initiated a 'snake venomics' project whose long-term goal is a detailed analysis of viperid venomes.

SEPARATION AND INITIAL CHARACTERIZATION OF ISOLATED TOXINS

For a detailed characterization of the toxin content of snake venoms ('venome'), the approach that we have coined '*snake venomics*'[23] (schematically outlined in Fig. 1) starts with the fractionation of the crude venom by reverse-phase high performance liquid chromatography (HPLC), followed by the initial characterization of each protein fraction by a combination of *N*-terminal sequencing, SDS-PAGE analysis, and mass spectrometric determination of the molecular masses and the cysteine (SH and S–S) content of the isolated toxins (Fig. 1(a–d)).

For the fractionation of venom proteins, we use a reverse-phase HPLC C_{18} column (250 × 4 mm, 5 μm particle size) eluting at 1 ml/min with a linear gradient of 0.1% TFA in water (solution A) and acetonitrile (solution B) (typically, isocratic elution with 5%B for 5 min, followed by linear gradients of 5–15%B over 20 min, 15–45%B over 120 min, and 45–70%B over 20 min). In our experience, this procedure allows the quantitative recovery of all venom components comprised in the apparent molecular mass range of 7–150 kDa that can be separated by conventional 2D-SDS-PAGE. On the other hand, the initial part of the acetonitrile gradient of the reverse-phase chromatography resolves peptides and small proteins (0.4–7 kDa), which would not be recovered from a two dimensional-electrophoretic separation. Moreover, for the accurate determination of toxin-specific features, such as the native molecular mass, the quaternary structural arrangement and the number of sulfydryl groups and disulfide bonds (see below), toxins need to be available in the solution. In addition, given that the wavelength of absorbance for a peptide bond is 190–230 nm, protein detection at 215 nm allows the estimation of the relative abundances (expressed as percentage of the total venom proteins) of the different protein families from the relation of the sum of the areas of the reverse-phase chromatographic peaks containing proteins from the same family to the total area of venom protein peaks in the reverse-phase chromatogram[24–26] (Fig. 1, panel (i), also see below). In a strict sense, the calculated relative amounts correspond to the '% of the total peptide bonds in the sample', which is a good estimate of the percentage by weight (g/100g) of a particular venom component.

Protein fractions showing single electrophoretic band, molecular mass, and *N*-terminal sequence (i.e. fractions 8–16 in panel c of Fig. 1) can be straightforwardly assigned by BLAST analysis (http://www.ncbi.nlm.nih.gov/BLAST) to a known protein family, indicating that representative members of most snake venom toxin families are present among the 923 viperid protein sequences deposited to date in the SwissProt/TrEMBL database (Knowledgebase Release 10.0 of March 2007; http://us.expasy.org/sprot/).

COUNTING CYSTEINE RESIDUES

Besides the *N*-terminal sequence, most snake venom proteins are characterized by a high and protein family-specific cysteine content (Table 1). This is particularly evident for small proteins whose global folds are stabilized primarily by the formation of disulfide bonds. Hence, determining the number of sulfydryl groups and disulfide bonds per molecule represents a useful tool for the preliminary classification of toxins into protein families. Mass spectrometry is perhaps the best-suited technique for counting cysteine residues. For quantitation of free cysteine residues and disulfide bonds, the

DOI: 10.1002/jms

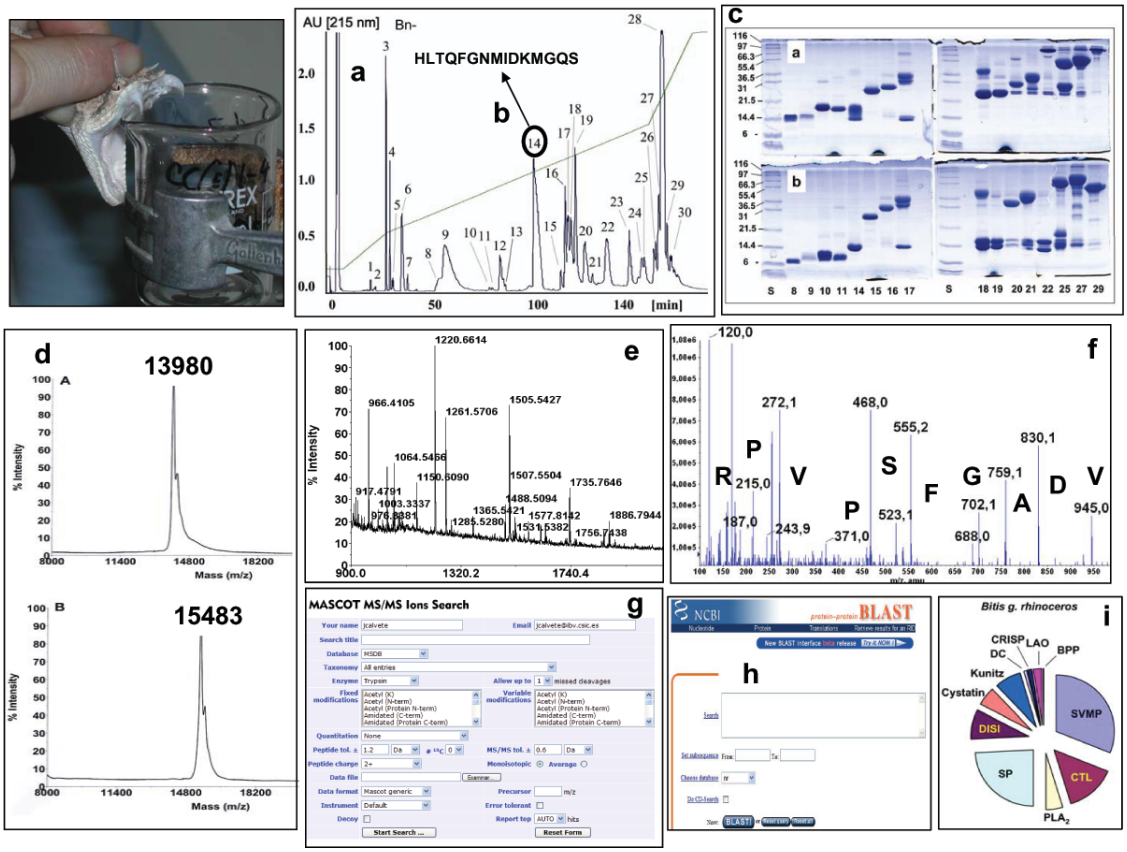


Figure 1. Snake venomics. Schematic representation of the steps typically followed in a snake venomics project. (a) Reverse-phase chromatographic separation of the venom proteins; (b) *N*-terminal sequencing of the isolated protein fractions; (c) SDS-PAGE of the RP-HPLC isolated proteins run under nonreduced (upper panel) and reduced (lower panels) conditions; (d) determination of the molecular masses of the proteins isolated in (a), and quantitation of their sulfydryl group and disulfide bond contents; (e) MALDI-TOF mass fingerprints of in-gel digested protein bands excised from SDS-polyacrylamide gels run as in (c); (f) amino acid sequence determination by nanospray-ionization CID-MS/MS of doubly and triply charged tryptic peptide ions obtained as in (e); (g) database searches using MS/MS spectra in MASCOT or (h)MS/MS-derived amino acid sequence through basic local alignment search tool (BLAST); (i) summary of the relative amounts of toxin families using a pie chart representation.

purified protein at 2–5 mg/ml in 10 µl of 50 mM HEPES, pH 9.0, 5 M guanidinine hydrochloride containing 1 mM EDTA is heat-denatured at 85 °C for 15 min, allowed to cool at room temperature, and incubated with either 10 mM of an alkylating agent (4-vinylpyridine or iodoacetamide) for 30 min at room temperature, or with 10 mM 1,4-dithioerythritol (DTE) for 15 min at 80 °C, followed by addition of the alhylating reagent (4-vinylpyridine or iodoacetamide) at 25 mM final concentration and incubation for 1 h at room temperature. Alkylated proteins are then freed from reagents using a C18 Zip-Tip pipette tip (Millipore) after activation with 70% ACN and equilibration in 0.1% TFA. Following protein adsorption and washing with 0.1% TFA, the PE-proteins are eluted with 1–5 µl of 70% ACN and 0.1% TFA and subjected to mass spectrometric (MALDI-TOF or ESI-MS) analysis.

The number of free cysteine residues (N_{SH}) can be determined using Eqn (1):

$$N_{SH} = (M_{ALK} - M_{NAT})/M_R \quad (1)$$

where M_{ALK} is the mass of the denatured but nonreduced protein incubated in the presence of the alkylating reagent; M_{NAT} is the mass of the native, HPLC-isolated protein; and M_R is the mass increment due to the alkylation of one thiol group (105.3 Da for S-pyridylethylation with 4-vinylpyridine; 57.1 Da for carbamidomethylation with iodoacetamide).

The number of total cysteine residues (N_{Cys}) can be calculated from Eqn (2):

$$N_{Cys} = [(M_{CM-} - M_{ALK})/(M_R + 1)] + N_{SH}, \quad (2)$$

where M_{CM} is the mass (in Da) of the reduced and alkylated protein, and $(M_R + 1)$ is the mass increment due to the alkylation of a cysteine residue, which prior to reduction was involved in the formation of a disulfide bond (106.3 and 58.1 Da, respectively, if 4-vinylpyridine or iodoacetamide were employed).

Finally, the number of disulfide bonds N_{S-S} can be calculated from Eqn (3):

$$N_{S-S} = (N_{Cys} - N_{SH})/2. \quad (3)$$

All mass values in Eqns (1–3) are in Daltons. Figure 1, panel d, shows an example of the quantitation by MALDI-TOF mass spectrometry of the cysteine residues of a

JMS

Table 1. Classification of snake venom toxins to protein families according to their cysteine content; [a] intersubunit disulfide bonds; [b] intrasubunit disulfide linkages; C-type natriuretic peptide; cysteine-rich secretory protein; snake venom metalloproteinase (SVMP); disintegrin-like/cysteine-rich domains (DC) of PIII-SVMP; snake venom vascular endothelial growth factor (svVEGF); L-amino acid oxidase (LAO)

Molecular mass range (kDa)	Total cysteine residues -SH	Total cysteine residues S–S	Protein family
1.6–2	–	1	C-NP
4–5	–	3	Myotoxin
	–	4	Short disintegrin
6–8	–	3	Kunitz-type inhibitor
	–	5	Dimeric disintegrin subunit
	–	6	Medium-sized disintegrin
10–12	1	–	Ohanin
13–15	–	2	Cystatin
	–	($2^a + 4^b$)	Dimeric disintegrin
	–	7	PLA_2
23–33	–	8	CRISP
	1	4	PI-SVMP
	–	6	Serine proteinase
	–	($1^a + 3^b$)	$\alpha\beta$ C-type lectin-like
	–	($1^a + 4^b$)	svVEGF
	–	13	DC-fragment
46–58	–	3	LAO
	1	18	PIII-SVMP

reverse-phase purified *Sistrurus barbouri* venom protein.[23] The mass of the native protein (13 980 Da) did not change upon incubation with 4-vinylpyridine under nonreducing conditions indicating that the protein did not contain sulfydryl groups. On the other hand, the molecular mass of the protein after reduction and *S*-pyridylethylation was 15 483 Da. The number of cysteine residues was derived from Eqn (2), $N_{Cys} = (15\,483 - 13\,980)/106.3 = 14.1$, and the number of disulfide bonds was calculated using Eqn (3), $N_{S-S} = (14.1 - 0)/2 = 7.05$ (7). The data strongly suggested that the protein may correspond to a PLA_2 molecule (Table 1).

This approach can also be directly applied for quantitating the cysteine content of a homodimeric protein (A_2) (Fig. 2) with just substituting M_{CM} by ($2 \times M_{CM}A$) in Eqn (2).

$$N_{Cys}A_2 = [(2 \times M_{CM}A - M_{ALK})/(M_R + 1)] + N_{SH}, \quad (4)$$

where '$M_{CM}A$' is the molecular mass of the reduced and alkylated subunit 'A'. Similarly, for counting the total number of cysteine residues of a heterodimeric protein (AB) (Fig. 3), the term M_{CM} in Eqn (2) must be replaced by the sum of the masses of the reduced and alkylated subunits, 'L' (large) and 'S' (small):

$$N_{Cys}LS = [((M_{CM}L + M_{CM}S) - M_{ALK})/(M_R + 1)] + N_{SH} \quad (5)$$

Determination of the number of cysteine residues in each subunit requires a combination of two sets of data, such as the molecular masses of the reduced and *S*-pyridylethylated subunits (DTT + 4VP, Fig. 3, upper panels) and the masses of the reduced and carbamidomethylated subunits (DTT + IA, Fig. 3, lower panels). The number of cysteines in subunits 'L' and 'S' can be calculated by Eqn (6):

$$N_{Cys}L = (M_{VP}L - M_{IA}L)/48.2 \text{ or } N_{Cys}S = (M_{VP}S - M_{IA}S)/48.2 \quad (6)$$

where 48.2 corresponds to the difference between the mass increments due to the alkylation of one thiol group with 4-vinylpyridine (106.3) and with iodoacetamide (58.1). It should be noticed that independentl of the alkylating reagent used, the larger molecular mass will always correspond to the same (L or S) alkylated subunit. In other words, in the example displayed in Fig. 3, ions at m/z 8687.1 and m/z 8204.6 (and those at m/z 8162.8 and 7680.8) belong to the same subunit though *S*-pyridylethylated and carbamidomethylated, respectively. This proportion stands always if (1) the large subunit contains equal or a higher number of cysteine residues than the small subunit ($N_{Cys}L \geq N_{Cys}S$); and (2) if $N_{Cys}S > N_{Cys}L$, then the difference between

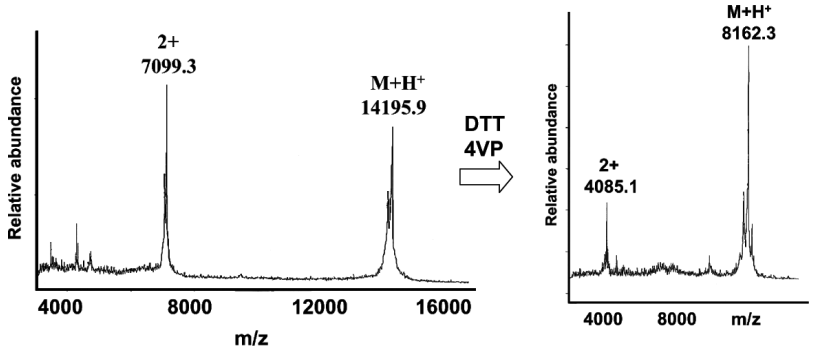


Figure 2. Counting the cysteine residues of a homodimeric protein. MALDI-TOF mass spectra of native (left) and reduced and *S*-pyridylethylated (right) homodimeric protein (A_2) isolated in fraction 11 shown in panel (c) of Fig. 1. The total number of cysteines was determined using Eqn (4): $N_{Cys} = [(8162.3 \times 2) - 14\,195.9]/106.3 = 20.0$. The presence of 10 cysteine residues per subunit is consistent with the preliminary classification of the 14.2 kDa protein as a homodimeric disintegrin.

DOI: 10.1002/jms

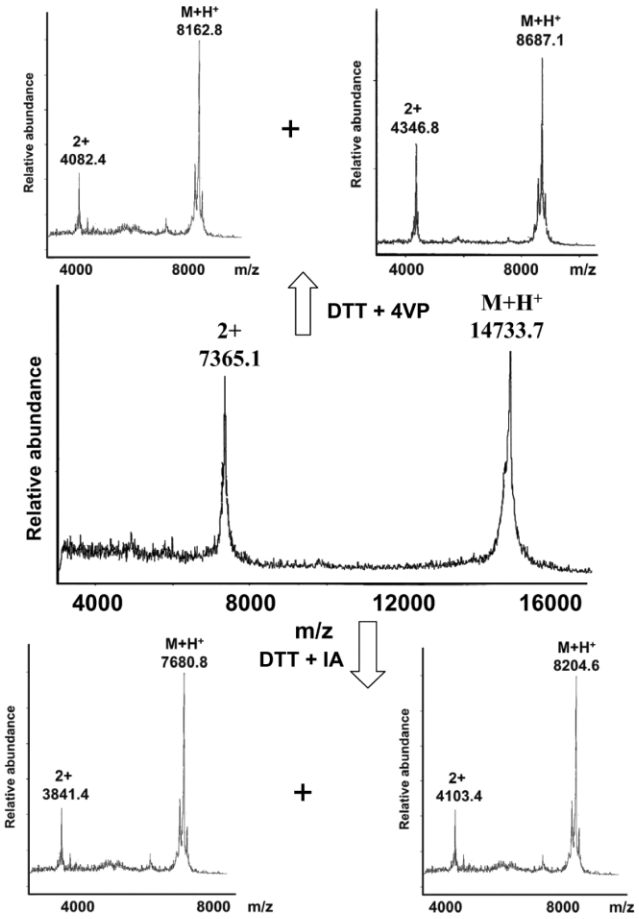


Figure 3. Counting the total number of cysteine residues of a heterodimeric protein and of its isolated subunits. MALDI-TOF mass spectra of (center panel) native heterodimeric protein (LS) isolated in fraction 10 shown in panel (c) of Fig. 1, its reduced and S-pyridylethylated subunits (DTT + VP, upper panels), and the reduced and carbamidomethylated subunits (DTT + IA, lower panels). In both cases, the total number of cysteine residues in LS can be calculated using Eqn (5) using either set of data: $N_{Cys}LS = [((8162.6 + 8687.1) - 14\,733.7)/106.3] = [((7680.8 + 8204.6) - 14\,733.7)/58.1] = 19.9(20)$. The number of cysteine residues in each subunit was determined applying Eqn (6): $N_{Cys}L = (8687.1 - 8204.6)/48.2 = 10$, and $N_{Cys}S = (8162.8 - 7680.8)/48.2 = 10$. The presence of 10 cysteine residues per subunit is consistent with the preliminary classification of the 14.7 kDa protein as a heterodimeric disintegrin.

the native masses of the large (M_L) and the small (M_S) subunits is higher than the difference between the number of cysteine residues of the small and the large subunits times the molecular mass increment due to alkylation of a thiol group with the alkylating reagent which produces the highest mass increment (M_{RH}) (Eqn (7)):

$$(M_L - M_S) > (N_{Cys}S - N_{Cys}L) \times M_{RH} \quad (7)$$

Finally, although information on the molecular masses of the subunits is insufficient for disclosing the number of inter and intrasubunit disulfide bonds, it can be easily demonstrated that subunits of homo- or heterodimeric proteins bearing an even or uneven number of cysteine residues, all of them engaged in disulfide bonds, must also be joined together, respectively, through an even or uneven number of intersubunit linkages.

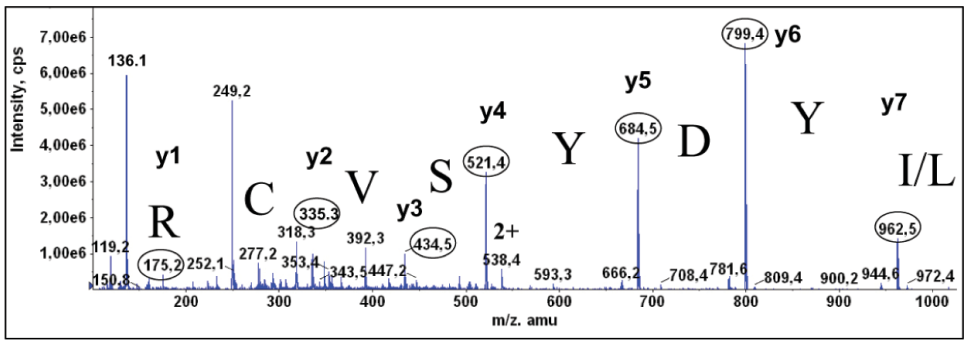
PROTEOMIC CHARACTERIZATION OF PROTEIN MIXTURES

Protein fractions showing heterogeneous or blocked *N*-termini (i.e. reverse-phase HPLC fractions 18–29 in Fig. 1, panel (c)) are analyzed by SDS-PAGE and the bands of interest are subjected to automated reduction, carbamidomethylation, and in-gel tryptic digestion in a ProGest digestor (Genomic Solutions). The resulting tryptic peptides are then analyzed by MALDI-TOF mass fingerprinting (in an Applied Biosystem's Voyager-DE Pro instrument) followed by amino acid sequence determination of selected doubly and triplycharged peptide ions by collision-induced dissociation tandem mass spectrometry (CID-MS/MS using an Applied Biosystem's QTrap 2000 mass spectrometer) (Fig. 1, panels e and f, and Fig. 4). As expected from the rapid amino acid sequence divergence of venom proteins evolving under accelerated evolution,[1,27–29] with a few exceptions, neither the tryptic mass fingerprints nor the product ion spectra match any known protein using the Protein Prospector (http://prospector.ucsf.edu) or the MASCOT (http://www.matrixscience.com) search programs (Fig. 1(g)). Furthermore, as illustrated in Fig. 4, it is not too unusual that a product ion spectrum matched with a high MASCOT score to a particular peptide sequence corresponds actually to a tryptic peptide of a homolog snake toxin containing one or more nearly isobaric amino acid substitutions. Hence, it is necessary to revise manually all the CID-MS/MS spectra (to confirm the assigned peptide sequence or for performing *de novo* sequencing), and submit the deduced peptide ion sequences to BLAST similarity searches (Figs 1(h); 4). Although the lack of any complete snake genome sequence is a serious drawback for the identification of venom proteins, high-quality MS/MS peptide ion fragmentation spectra usually yield sufficient amino acid sequence information derived from almost a complete series of sequence-specific b- and/or y-ions to unambiguously identify a homologue protein in the current databases. The outlined snake venomics approach allows us to assign unambiguously all the isolated venom toxins representing over 0.05% of the total venom proteins to known protein families (Table 2).

To date, we have applied snake venomics for elucidating the protein composition of the venoms from the North American rattlesnakes *Sistrurus miliarius barbouri*,[23,25] *Sitrurus catenatus* subspecies *catenatus, tergeminus* and *edwasdsii*,[25] *Crotalus atrox*, and *Agkistrodon contortrix contortrix* (manuscript in preparation); the Tunisian vipers *Cerastes cerastes cerastes, Cerastes vipera* and *Macrovipera lebetina transmediterranea*,[24] the African species *Bitis arietans* (Ghana),[30] *B. gabonica gabonica*,[26] *Bitis gabonica rhinoceros, Bitis nasicornis, Bitis caudalis*, and *E. ocellatus*, and the South American Bushmaster pit viper *L. muta* (manuscripts in preparation). As a whole, our results show that, in spite of the fact that venom proteins belong to only a few protein families, venoms depart from each other

DOI: 10.1002/jms

JMS




```
>gi|38570060|gb|AAR24534.1|  serine protease 1 [Bitis gabonica]
Length=260

 Score = 27.8 bits (58),  Expect =    10
 Identities = 7/8 (87%), Positives = 8/8 (100%), Gaps = 0/8 (0%)

Query  1    LYDYSVCR  8
            L+DYSVCR
Sbjct  170  LFDYSVCR  177


>gi|1351938|sp|P47797|VSP2_AGKRH  Ancrod precursor (Venombin A) (Protein C activator) (ACC-C)
 gi|403078|gb|AAA49195.1|  ancrod
Length=258

 Score = 26.1 bits (54),  Expect =    34
 Identities = 6/8 (75%), Positives = 8/8 (100%), Gaps = 0/8 (0%)

Query  1    LYDYSVCR  8
            LY+Y+VCR
Sbjct  170  LYNYTVCR  177


>gi|13959655|sp|Q9PT40|VSP2_VIPLE  Venom serine proteinase-like protein 2 precursor
 gi|6562944|emb|CAB62591.1|  serine proteinase, homolog [Macrovipera lebetina]
Length=260

 Score = 25.2 bits (52),  Expect =    61
 Identities = 6/8 (75%), Positives = 8/8 (100%), Gaps = 0/8 (0%)

Query  1    LYDYSVCR  8
            ++DYSVCR
Sbjct  170  MFDYSVCR  177
```

Figure 4. Product ion spectrum of the doubly-charged ion m/z 538.4. The fragmented ion corresponded to a tryptic peptide digest of a *Bitis gabonica gabonica* 29 kDa venom protein.[26] The CID-MS/MS spectrum was matched by MASCOT to the tryptic peptide sequence MFDYSVCR from a venom serine proteinase (29 559 Da) of *Macrovipera lebetina* (Q9PT40). Manual interpretation of the MS/MS spectrum clearly identified the sequence (L/I)YDYSVCR. Notice that MFDYSVCR and (L/I)YDYSVCR are identical except for the *N*-terminal dipeptide. BLAST analysis against the vertebrate protein database showed similarity to a number of snake venom serine proteinases, including *B. gabonica* serine proteinase-1 (highest score). The *M. lebetina* (Q9PT40) sequence matched by MASCOT shows up in third position.

in the composition and the relative abundance of their toxins (Table 2).

CHARACTERIZATION OF THE SUBUNIT COMPOSITION OF TOXINS

A comparison of the apparent molecular masses of proteins run under nonreducing and reducing SDS-PAGE conditions (Fig. 1(c)) provides valuable information regarding the aggregation state and subunit composition of the toxins. Fractions 8 and 9 (isoforms of a Kunitz-type inhibitor of serine proteinases) exhibited apparent molecular masses of 14–16 kDa and 7–8 kDa when analyzed in nonreduced and reduced SDS-PAGE, respectively (Fig. 1(c)). However, ESI-MS showed that these Kunitz-type inhibitors had native (nonreduced) molecular masses of about 7 kDa (Fig. 5(a)). Hence, the distinct electrophoretic behavior of the Kunitz-type inhibitors in nonreduced versus reduced gels may be regarded as artifactual or as indicating the existence of noncovalent dimers.

Reverse-phase HPLC fraction 29 (Fig. 1, panel (c)), identified as a PIII snake venom Zn^{2+}-metalloproteinase, had apparent molecular masses of 110 kDa (nonreduced) and 55 kDa (reduced), and may thus represent a dimeric metalloprotease. Fractions 18 and 19 (Fig. 1(c)) run in nonreduced gels as 28 kDa proteins, whereas upon reduction two 14–16 kDa subunits were released. These proteins are typical dimeric ($\alpha\beta$) C-type lectin-like molecules.[31]

The molecular mass of a native multimeric C-type lectin-like protein isolated from the venom of *B. caudalis* (Fig. 1, panel c, fraction 22) was accurately measured by ESI-MS (92 151 Da, Fig. 5(b)). The quaternary structure of the lectin, determined by mass spectrometry after reduction and separation of subunits by reverse-phase HPLC (Fig. 2(d)), was found to be built by the association of polypeptides of molecular masses (carbamidomethylated):

DOI: 10.1002/jms

Table 2. Overview of the relative occurrence of proteins (in percentage of the total HPLC-separated proteins) of the toxin families in the venoms of *Sistrurus catenatus catenatus* (SCC), *Sistrurus catenatus tergeminus* (SCT), *Sistrurus catenatus edwardsii* (SCE), *Sistrurus miliarius barbouri* (SMB), *Cerastes cerastes cerastes* (CCC), *Cerastes vipera* (CV) and *Macrovipera lebetina transmediterranea* (MLT), *Bitis arietans* (BA), *Bitis gabonica gabonica* (BGG), *Bitis gabonica rhinoceros* (BGR), *Bitis nasicornis* (BN), *Bitis caudalis* (BC), *Echis ocellatus* (EO), *Lachesis muta* (LM), *Crotalus atrox* (CA), and *Agkistrodon contortrix contortrix* (ACC)

	Venom															
	SCC	SCT	SCE	SMB	CCC	CV	MLT	BA	BGG	BGR	BN	BC	EO	LM	CA	ACC
Protein family	% of total venom proteins															
Disintegrins																
Long	–	–	–	–	–	–	–	17.8	–	–	–	–	–	–	–	–
Medium	2.5	4.2	0.9	7.7	–	–	–	–	–	–	–	–	–	–	6.5	–
Dimeric	–	–	–	–	8.1	<1	6.0	–	3.4	8.5	3.5	–	4.2	–	–	1.5
Short	–	–	–	–	–	–	<1	–	–	–	–	–	2.6	–	–	–
Myotoxin	0.4	<0.1	–	–	–	–	–	–	–	–	–	–	–	–	–	–
C-type BPP/NP	–	–	<0.1	<0.1	–	–	<1	–	2.8	0.3	–	–	–	14.7	2.1	<0.1
Kunitz-type inhibitor	–	–	<0.1	<0.1	–	–	–	4.2	3.0	7.5	–	3.2	–	–	–	–
Cystatin	–	–	–	–	–	–	–	1.7	9.8	5.3	4.2	–	–	–	–	–
DC-fragment	<0.1	<0.1	<0.1	1.3	–	–	1.0	–	0.5	0.6	<0.1	–	1.7	–	–	<0.1
NGF/svVEGF	<0.1	<0.1	<0.1	<0.1	–	–	2.1	–	1.0	–	–	–	–	–	–	–
Ohanin-like	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	<0.1
CRISP	0.8	1.3	10.7	2.9	–	–	–	–	2.0	1.2	1.3	1.2	1.5	1.8	4.2	–
PLA_2	29.9	31.6	13.7	**32.5**	20.0	21.1	4.0	4.3	11.4	4.8	20.1	**59**.8	12.6	8.7	16.3	18.5
Serine proteinase	18.2	20.4	24.4	17.1	9.1	20.0	9.2	19.5	**26**.4	23.9	21.9	15.1	2.0	**31.2**	10.1	13.8
C-type lectin-like	<0.1	<0.1	<0.1	<0.1	24.0	0.9	10.1	13.2	14.3	14.1	4.2	4.9	7.0	8.1	1.6	–
L-amino acid oxidase	4.2	1.6	2.5	2.1	12.0	9.0	–	–	1.3	2.2	3.2	1.7	1.4	2.7	8.0	2.2
Zn^{2+}-metalloproteinase	**43.8**	**40.6**	**48.6**	**36.1**	**37.0**	**48.1**	**67.1**	**38.5**	**22.9**	**30.8**	**40.9**	11.5	**67.0**	**31.9**	**51.1**	**63.6**

16 479, 16 366, 15 190, 15 293, and 15 235 Da. Assuming that each subunit may contain 8 cysteine residues,[31] the best fitting subunit combination is $[16\,479 + (16\,366)_2 + 15\,190 + 15\,293 + 15\,235](94\,929 - (48 \times 58.1) = 92\,140.2$ Da), indicating that the multimeric C-type lectin may represent a trimeric arrangement of $\alpha\beta$ dimers. C-type lectin-like proteins isolated to date from a large number of viperid and crotalid venoms occur in a variety of oligomeric forms, including $\alpha\beta$, $(\alpha\beta)_2$, and $(\alpha\beta)_4$.[31] However, highlighting the potentiality of snake venomics as a 'discovery science', to date no $(\alpha\beta)_3$ C-type lectin-like structure has been reported.

APPLICATIONS OF SNAKE VENOMICS

The detailed proteomic characterization of snake venoms provides a comprehensive catalog of the toxins secreted into the venoms. Toxins represent valuable biotechnological tools for studying physiological processes. Characterization of the snake venom proteomes has also a number of potential benefits for basic research, clinical diagnosis, and development of new research tools and drugs of potential clinical use.[12–14] In addition, snake venom composition may also aid in understanding the biology and ecology of snakes. Thus, the availability of detailed proteomic information on individual proteins make possible detailed estimates of the similarity and differentiation of the venom proteomes of different taxa, which are then useful in revealing broad-scale evolutionary patterns. There is a small but increasing number of studies that strongly support the idea that venom composition variation reflects adaptation for differential utilization of distinct ecological niches or prey types.[32–34] Hence, the venoms of the Tunisian vipers *Cerastes cerastes*, *C. vipera* and *M. lebetina* contain distinct mixtures of proteins, which target the hemostatic system[24] (Table 2). *C. cerastes* and *C. vipera* secrete venoms with similar protein composition, probably reflecting their similar ecological niches, i.e. sandy desert, and prey habits (lizards, rodents, and small mammals). A rapid co-evolution between snakes and their prey in driving the evolution of venom proteins has been discussed.[32] The more complex protein composition of the venom of *M. lebetina* may relate to its adaptation to maquis terrain with bushes on gravel and rocky ground. Hunting in such an uneven habitat may require a much quicker immobilization and killing of the prey than in the sandy desert where the snake can easily follow the track of the envenomed animal. The high degree of differentiation in the venom proteome among recently-evolved congeneric *Sistrurus* taxa[25] also emphasizes the uniqueness of the venom composition of even closely related species which have different diets.

The venom composition appears to provide information on the evolutionary history of congeneric taxa.[35] Protein similarity coefficients used to estimate the similarity of venom proteins of *Bitis* taxa support the monophyly of the

DOI: 10.1002/jms

JMS

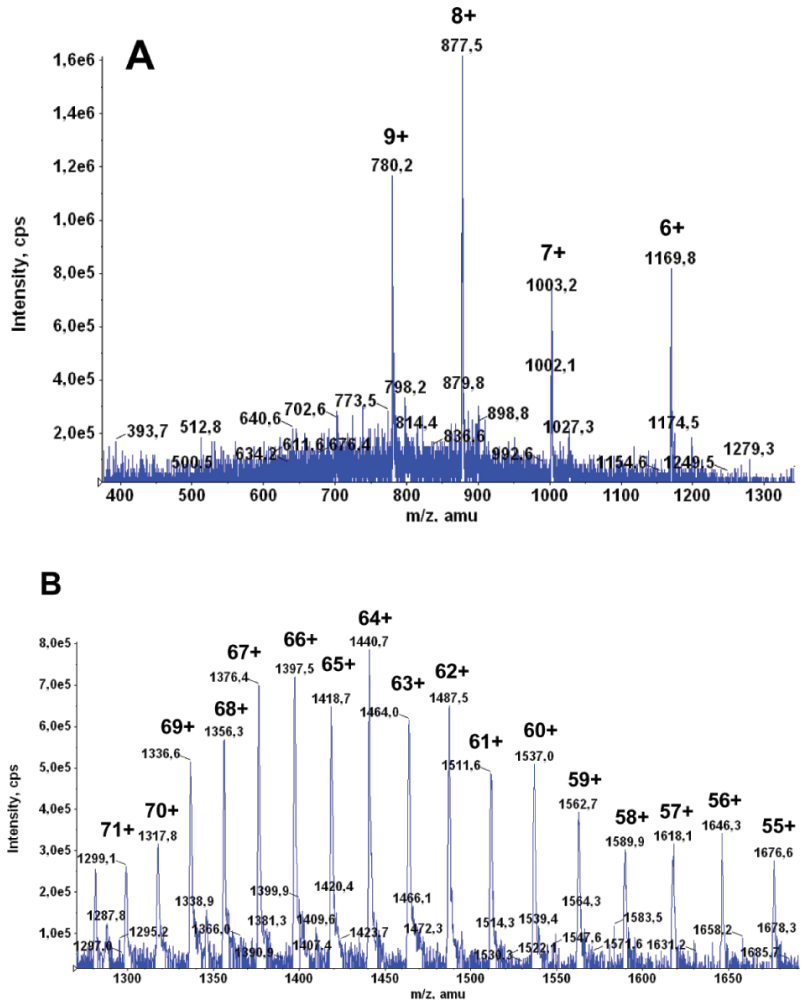


Figure 5. Molecular mass determination by ESI-MS. (a) Electrospray-ionization mass spectrometric characterization of native (nonreduced) Kunitz-type inhibitor isolated in fraction 8 of Fig. 1, panel c, from which a molecular mass of 7012.6 ± 0.3 Da was calculated. (b) Electrospray-ionization mass spectrum of a novel $(\alpha\beta)_3$ multimeric C-type lectin-like protein showing a molecular mass of 92 151 ± 18 Da. Spectra were recorded in a hybrid triple quacrupole/linear ion-trap instrument (QTrap 2000 from Applied Biosystems).

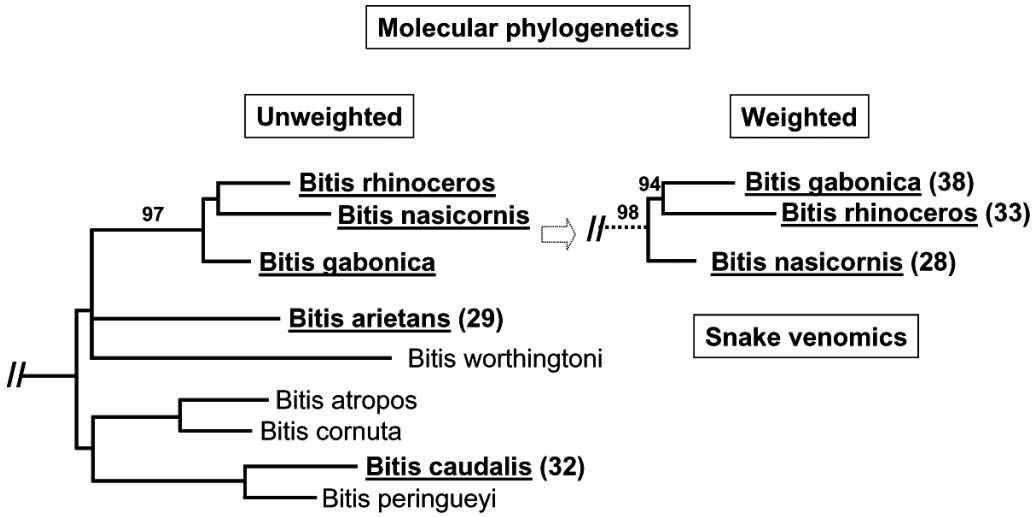


Figure 6. Phylogenetic relationships among Bitis taxa. Details of maximum-parsimony trees generated with *Causus* serving as the outgroup (adapted from Ref. 36). Left, the 75% majority-rule cladogram of the unweighted maximum-parsimony analysis. Right, weighted maximum-parsimony phylogram based on overall genetic distance-adjusted step matrix and the successive approximation approach. Bootstrap values of nodes with confidence values >90% are indicated. Bitis species for which the venom proteome has been analyzed are underlined. The number of distinct proteins identified in each taxa using a venomics approach are given in parentheses.

DOI: 10.1002/jms

three West African taxa (*B.g. gabonica, B.g. rhinoceros*, and *B. nasicornis*) based on genetic distance reconstructions, the lack of alliances between *B. arietans* and any other Bitis species, and are also consistent with the taxonomic association of *B. caudalis* within the differentiated group of small Bitis species (Fig. 6). The low level of venom toxin composition similarity between the two conventionally recognized subspecies of *B. gabonica, B. g. gabonica* and *B. g. rhinoceros*, support the consideration by some authors[36] of *B. g. rhinoceros* as a separate species, *Bitis rhinoceros*. Bitis snake venomic data fit better to a weighted phylogram based on overall genetic distances than to an unweighted maximum-parsimony tree (Fig. 6).

Acknowledgements

Research on snake venomics has been financed by grants from the Ministerio de Ciencia y Tecnología and Ministerio de Educación y Ciencia, Madrid, Spain. PJ and LS are recipients of a pre-doctoral fellowship (FPI, formación de personal investigador) from the Spanish Ministerio de Ciencia y Tecnología, and an I3P contract, respectively.

REFERENCES

1. Ménez A (ed). *Perspectives in Molecular Toxinology*. John Wiley & Sons: Chichester 2002.
2. Fox JW, Serrano SMT. Snake toxins and hemostasis. *Toxicon* 2005; **45**: 951.
3. Markland FS. Snake venoms and the hemostatic system. *Toxicon* 1998; **36**: 1749.
4. Serrano SMT, Shannon JD, Wang D, Camargo AC, Fox JW. A multifaceted analysis of viperid snake venoms by two-dimensional gel electrophoresis: an approach to understanding venom proteomics. *Proteomics* 2005; **5**: 501.
5. Fry BG, Wüster W. Assembling an arsenal: origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences. *Molecular Biology and Evolution* 2004; **21**: 870.
6. Fry BG. From genome to ''venome'': molecular origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences and related body proteins. *Genome Research* 2005; **15**: 403.
7. Vidal N. Colubroid systematics: evidence for an early appearance of the venom apparatus followed by extensive evolutionary tinkering. *Journal of Toxicology Toxin Reviews* 2002; **21**: 21.
8. Fry BG, Vidal N, Norman JA, Vonk FJ, Scheib H, Ramjan SF, Kuruppu S, Fung K, Hedges SB, Richardson MK, Hodgson WC, Ignjatovic V, Summerhayes R, Kochva E. Early evolution of the venom system in lizards and snakes. *Nature* 2006; **439**: 584.
9. Greene HW. Dietary correlates of the origin and radiation of snakes. *American Zoologist* 1983; **23**: 431.
10. Lynch VJ. Inventing an arsenal: adaptative evolution and neofunctionalization of snake venom phospholipase A2 genes. *BMC Evolutionary Biology* 2007; **7**: 2.
11. Richman A. Evolution of balanced genetic polymorphism. *Molecular Ecology* 2000; **9**: 1953.
12. Ménez A, Stöcklin R, Mebs D. ''Venomics'' or: the venomous systems genome project. *Toxicon* 2006; **47**: 255.
13. Harvey AL, Bradley KN, Cochran SA, Rowan EG, Pratt JA, Quillfeldt JA, Jerusalinsky DA. What can toxins tell us for drug discovery? *Toxicon* 1998; **36**: 1635.
14. Koh DCI, Armugam A, Jeyaseelan K. Snake venom components and their applications in biomedicine. *Cellular and Molecular Life Sciences* 2006; **63**: 3030.
15. Junqueira-de-Azevedo IL, Ho PL. 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* 2002; **299**: 279.
16. Francischetti IM, My-Pham V, Harrison J, Garfield MK, Ribeiro JMC. *Bitis gabonica* (Gaboon viper) snake venom gland: towards a catalog of full-length transcripts (cDNA) and proteins. *Gene* 2004; **337**: 55.
17. Kashima S, Roberto PG, Soares AM, Astolfi-Filho S, Pereira JO, Giuliati S, Faria M Jr., Xavier MAS, Fontes MRM, Giglio JR, Franca SC. Analysis of *Bothrops jararacussu* venomous gland transcriptome focusing on structural and functional aspect: I-gene expression profile of highly expressed phospholipases A_2. *Biochimie* 2004; **86**: 211.
18. Cidade DAP, Simão TA, Dávila AMR, Wagner G, Junqueira-de-Azevedo ILM, Ho PL, Bon C, Zingali R, Albano RM. *Bothrops jararaca* venom transcriptome: analysis of the gene expression pattern. *Toxicon* 2006; **48**: 437.
19. Qinghua L, Xiaowei Z, Wei Y, Chenji L, Yijun H, Pengxin Q, Xingwen S, Songnian H, Guangmei Y. A catalog for transcripts in the venom gland of the *Agkistrodon acutus*: identification of the toxins potentially involved in coagulopathy. *Biochemical and Biophysical Research Communications* 2006; **341**: 522.
20. Wagstaff SC, Harrison RA. Venom gland EST analysis of the saw- scaled viper, *Echis ocellatus*, reveals novel $\alpha_9\beta_1$ integrin-binding motifs in venom metalloproteinases and a new group of putative toxins, renin-like proteases. *Gene* 2006; **377**: 21.
21. Junqueira-de-Azevedo ILM, Ching ATC, Carvalho E, Faria F, Nishiyama MY Jr., Ho PL, Diniz MRV. *Lachesis muta* (*Viperidae*) cDNAs reveal diverging pitviper molecules and scaffolds typical of cobra (*Elapidae*) venoms: implications in snake toxin repertoire evolution. *Genetics* 2006; **173**: 877.
22. Ching ATC, Rocha MMT, Leme AFP, Pimenta DC, Furtado MFD, Serrano SMT, Ho PL, Junqueira-de-Azevedo ILM. Some aspects of the venom proteome of the Colubridae snake *Philodryas olfersii* revealed from a Duvernoy's (venom) gland transcriptome. *FEBS Letters* 2006; **580**: 4417.
23. Juárez P, Sanz L, Calvete JJ. Snake venomics: characterization of protein families in *Sistrurus barbouri* venom by cysteine mapping, N-terminal sequencing, and tandem mass spectrometry analysis. *Proteomics* 2004; **4**: 327.
24. Bazaa A, Marrakchi N, El Ayeb M, Sanz L, Calvete JJ. Snake venomics: comparative analysis of the venom proteomes of the Tunisian snakes *Cerastes cerastes, Cerastes vipera* and *Macrovipera lebetina*. *Proteomics* 2005; **5**: 4223.
25. Sanz L, Gibbs HL, Mackessy SP, Calvete JJ. Venom proteomes of closely-related *Sistrurus* rattlesnakes with divergent diets. *Journal of Proteome Research* 2006; **5**: 2098.
26. Calvete JJ, Marcinkiewicz C, Sanz L. Snake venomics of *Bitis gabonica*. Protein family composition, subunit organization of venom toxins, and characterization of dimeric disintegrins bitisgabonin-1 and bitisgabonin- 2. *Journal of Proteome Research* 2007; **6**: 326.
27. Ohno M, Ménez R, Ogawa T, Danse JM, Shimohigashi Y, Fromen C, Ducancel F, Zinn-Justin S, Le Du MH, Boulain JC, Tamiya T, Ménez A. Molecular evolution of snake toxins: is the functional diversity of snake toxins associated with a mechanism of accelerated evolution? *Prog. Nucleic Acid Res. Mol. Biol.* 1998; **59**: 307.
28. Ogawa T, Kitajima M, Nakashima K, Sakaki Y, Ohno M. Molecular evolution of group II phospholipases A2. *J. Mol. Evol.* 1995; **41**: 867.
29. Ogawa T, Chijiwa T, Oda-Ueda N, Ohno M. Molecular diversity and accelerated evolution of C-type lectin-like proteins from snake venom. *Toxicon* 2005; **45**: 1.
30. Juárez P, Wagstaff SC, Oliver J, Sanz L, Harrison RA, Calvete JJ. Molecular cloning of disintegrin-like transcript BA-5A from *Bitis arietans* venom gland cDNA library: a putative intermediate in the evolution of the long chain disintegrin bitistatin. *J. Mol. Evol.* 2006; **63**: 142.
31. Morita T. Structure-function relationships of C-type lectin-related proteins. *Pathophysiology of Haemostasis and Thrombosis* 2005; **34**: 156.

32. Daltry JC, Wuster W, Thorpe RS. Diet and snake venom evolution. *Nature* 1996; **379**: 537.
33. Mackessy SP, Sixberry NM, Heyborne WH, Fritts T. Venom of the brown treesnake, *Boiga irregularis*: ontogenetic shifts and taxa-specific toxicity. *Toxicon* 2006; **47**: 537.
34. Li M, Fry BG, Kini RM. Eggs-only diet: its implications for the toxin profile changes and ecology of the marbled sea snake (*Aipysurus eydouxii*). *Journal of Molecular Evolution* 2005; **60**: 81.
35. Calvete JJ, Escolano J, Sanz L. Snake venomics of Bitis species reveals large intragenus venom toxin composition variation. application to taxonomy of congeneric taxa. *Journal of Proteome Research* (in press).
36. Lenk P, Herrmann HW, Joger U, Wink M. Phylogeny and systematics of *Bitis* (Reptilia: Viperidae) based on molecular evidences. *Kaupia* 1999; **8**: 31.

DOI: 10.1002/jms