

Lachesis muta (Viperidae) cDNAs Reveal Diverging Pit Viper Molecules and Scaffolds Typical of Cobra (Elapidae) Venoms: Implications for Snake Toxin Repertoire Evolution

Inácio L. M. Junqueira-de-Azevedo,^{*,†,‡,1} Ana T. C. Ching,^{*,†} Eneas Carvalho,^{*,‡,§} Fernanda Faria,^{*,*} Milton Y. Nishiyama, Jr.,^{††} Paulo L. Ho,^{*,†,‡,††} and Marcelo R. V. Diniz^{††}

^{*}Centro de Biotecnologia, Instituto Butantan, São Paulo 05503-900, Brazil, [†]Interunidades em Biotecnologia, Universidade de São Paulo, São Paulo 05508-900, Brazil, [‡]Instituto de Biociências, Universidade de São Paulo, São Paulo 05508-900, Brazil, [§]Laboratório de Parasitologia, Instituto Butantan, São Paulo 05503-900, Brazil, ^{**}Laboratório de Bioquímica e Biofísica, Instituto Butantan, São Paulo 05503-900, Brazil, ^{††}Instituto de Química, Universidade de São Paulo, São Paulo 05508-900, Brazil and ^{†††}Centro de Pesquisa e Desenvolvimento, Fundação Ezequiel Dias—Belo Horizonte 30510-010, Brazil

Manuscript received January 30, 2006
Accepted for publication March 31, 2006

ABSTRACT

Efforts to describe toxins from the two major families of venomous snakes (Viperidae and Elapidae) usually reveal proteins belonging to few structural types, particular of each family. Here we carried on an effort to determine uncommon cDNAs that represent possible new toxins from *Lachesis muta* (Viperidae). In addition to nine classes of typical toxins, atypical molecules never observed in the hundreds of Viperidae snakes studied so far are highly expressed: a diverging C-type lectin that is related to Viperidae toxins but appears to be independently originated; an ohanin-like toxin, which would be the third member of the most recently described class of Elapidae toxins, related to human butyrophilin and B30.2 proteins; and a 3FTx-like toxin, a new member of the widely studied three-finger family of proteins, which includes major Elapidae neurotoxins and CD59 antigen. The presence of these common and uncommon molecules suggests that the repertoire of toxins could be more conserved between families than has been considered, and their features indicate a dynamic process of venom evolution through molecular mechanisms, such as multiple recruitments of important scaffolds and domain exchange between paralogs, always keeping a minimalist nature in most toxin structures in opposition to their nontoxin counterparts.

THE venomous snakes are classified in four main families: Colubridae, Viperidae, Elapidae (including Hydrophiinae), and Atractaspidae (CAMPBELL and LAMAR 2004). The first three families are present in the New World, where Viperidae (Crotalinae subfamily: pit vipers) is the predominant group. The general composition of snake venoms varies between each family, from genus to genus, and even between species, with toxin shapes restricted to some groups and absent in others. For instance, postsynaptic neurotoxins belonging to the three-finger scaffold group are widely distributed in Elapidae snakes (ENDO and TAMIYA 1991), reported in Colubridae (FRY *et al.* 2003a), but, until now, not indubitably found in Viperidae. For this reason, Elapidae envenoming has adopted a highly neurotoxic strategy, whereas Viperidae evokes complex hemorrhagic and inflammatory effects.

Among the pit vipers (the Viperinae subfamily of the Viperidae), the genus *Lachesis* is particularly interesting in containing the longest vipers of the world, up to 3.5 m, and also the largest venomous snakes of the Americas (CAMPBELL and LAMAR 2004). *Lachesis muta* (popular names: Surucucu, Bushmaster) is the one with the widest geographical distribution (ZAMUDIO and GREENE 1997). Nausea, hypotension, bradycardia, shock, and even death due to hemorrhagic, coagulant, and neurotoxic activities comprise the envenoming features (JORGE *et al.* 1997) and are probably a consequence of the direct action of the few molecules already characterized from *L. muta*. They are common pit viper toxins, such as serine proteases (WEINBERG *et al.* 2004), snake venom metalloproteases (SVMPs) (SANCHEZ *et al.* 1991), phospholipases A2 (PLA2s) (DAMICO *et al.* 2005), and a C-type lectin (CASTRO *et al.* 1999).

An extensive knowledge of the compositions of venoms is important not only for the comprehension of the envenoming but also for the understanding of the possible origins and evolutionary tracks that specialized molecules (toxins) are able to follow during differentiation from their nontoxin ancestor counterparts. For instance, did some existing polypeptide scaffolds become allocated to other functions, and, if so, how did

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under the following full-length confirmed cDNA accession nos.: DQ396473–DQ396477. EST clusters in GenBank dbEST are from DY403159 to DY404320.

¹Corresponding author: Centro de Biotecnologia, Instituto Butantan, Av. Vital Brazil, 1500, São Paulo, SP 05503-900, Brazil.
E-mail: ijuncaze@butantan.gov.br

they diverge and become exchanged? This question is being examined with the new and the available databases, and new descriptions of novel venom gland components will be especially helpful (ALAPE-GIRÓN *et al.* 1999; FRY 2005; FRY *et al.* 2006). In addition, toxins are powerful tools for understanding physiological processes and for generating drug developments. But technically, the efforts to describe venom constituents are usually directed toward isolating toxins responsible for largely known activities, which reduces the chances of finding unusual constituents. Transcriptomic or proteomic studies could provide an opportunity to find unresearched molecules, providing insights into the real diversity of venom composition. For these reasons, we generate and analyzed an expressed sequence tags (ESTs) database from *L. muta* venom gland. This first set of cDNAs from this animal, and one of the few from a reptile, allowed the identification of new and very unexpected molecules, some common to other snake families and paradigmatically supposed to be absent from Viperidae. In this article we describe the characteristics of these molecules and provide some clues about the possible mechanisms involved in their origins under a general overview of the transcriptome of the *L. muta* venom gland.

MATERIALS AND METHODS

cDNA library construction and EST generation: A specimen of *L. muta* had been kept for years at the serpentarium of Fundação Ezequiel Dias (Belo Horizonte, Brazil) for the production of antivenom serum. The pair of venom glands was obtained a few moments after its natural death and total RNA was isolated following the procedure described by CHIRGWIN *et al.* (1979). mRNA purification was performed on a column of oligo(dT) cellulose (Amersham Biosciences) and its integrity was evaluated by *in vitro* translation in a rabbit reticulocyte lysate system (PELHAM and JACKSON 1976). The cDNAs were synthesized from 5 µg of mRNA using the Superscript plasmid system for cDNA synthesis and cloning (Invitrogen, San Diego), directionally cloned in pGEM11Zf+ plasmid (Promega, Madison, WI) and transformed in *Escherichia coli* DH5α cells, as described in JUNQUEIRA-DE-AZEVEDO and HO (2002). For large-scale DNA sequencing (EST generation), random clones were grown in antibiotic selective medium for 22 hr and plasmid DNA was isolated using alkaline lysis, as described in JUNQUEIRA-DE-AZEVEDO and HO (2002). The DNA was sequenced on an ABI 3100 sequencer using BigDye2 kit (Applied Biosystems, Foster City, CA) and the standard M13 forward primer.

Clusters assemble and identification: The chromatogram files of sequences were exported to a database in a LINUX-based workstation running a homemade pipeline of sequence analysis software: The Phred program (LAZO *et al.* 2001) was used to remove poor quality sequences by trimming sequences with quality below a Phred value of 20 in 75% of a window of 75 bases. The CrossMatch program was used to remove specific vector sequences, 5' adapters, and 3' primers containing poly(A/T) sequences >15. Sequences <150 bp (after trimming) were also discarded. A final manual examination and base editing of sequences were also carried out, intending to fix eventual software mistakes. ESTs were then assembled in clusters of contiguous sequences using the CAP3 program

(HUANG and MADAN 1999), set to join only sequences with at least 98% of base identity. Therefore, each cluster sequence basically derives from the alignment consensus of >98% identical ESTs and was considered as a virtual "transcript" for the following annotation and analysis steps. A second round of grouping, using the lowest stringent parameter on CAP3 (66%), was also performed over the cluster to facilitate the identification of groups of related clusters (here referenced as *GRC*) that may represent sets of paralogous cDNAs. Each cluster sequence (not the *GRC* consensus) was searched against GenBank NCBI databases by network-client BLAST package (<http://www.ncbi.nlm.nih.gov/blast/download.shtml>) using BLASTX and BLASTN algorithms to identify similar products with an *e*-value cutoff <10⁻⁵. No-hit sequences were also checked for the presence of signal peptide using SignalP 3.0 (BENDTSEN *et al.* 2004). A final annotation table in Microsoft Excel format was generated containing all the relevant information about clusters.

Alignments and phylogeny inference: The sequences used for the phylogenetic analyses were obtained through Blast or Entrez searches on GenBank or SWISS-PROT or kindly provided prealigned by B. G. Fry (FRY 2005) and their accession numbers are referred to in Figures 3–6. Multidomain proteins were trimmed only to the domain of interest. Detailed analyses of individual sequences were done using Vector NTI 9.1 suite (Invitrogen) and alignments of the sequence sets were carried out by the CLUSTAL-W algorithm (HIGGINS *et al.* 1994) followed by manual editing. For the Bayesian inferences of phylogenies based upon the posterior probability distribution of the trees, we used the MrBayes software version 3.1.1-p1 (RONQUIST and HUELSENBECK 2003), utilizing 2 × 10⁶ numbers of cycles for the Markov chain Monte Carlo algorithm.

RESULTS

EST sequencing and clustering: Since many snake species are endangered, rare, or difficult to keep in captivity, as is the case in the genus *Lachesis*, we obtained a pair of venom glands after the natural death of a specimen in captivity. The purified mRNA was of good quality, as assessed by *in vitro* translation assay (supplemental Figure 1 at <http://www.genetics.org/supplemental/>). The taxonomic classification of the genus was redefined some years ago, but we confirmed this rare specimen as *L. muta* by sequencing its mitochondrial NADH dehydrogenase and cytochrome B cDNAs and comparing them to sequences of the specimens used to define the genus by ZAMUDIO and GREENE (1997) (supplemental Table 1 at <http://www.genetics.org/supplemental/>). The random sequencing of this library gave readable sequences for a total of 2095 clones that were grouped with stringent parameters to form 1162 clusters of unique sequences (including 941 singlets and 221 of two or more ESTs). These clusters were also compared against each other with weak parameters (66% identity on CAP3) and arranged in 1029 *GRCs*, from which 70 have two or more clusters that may represent paralogous transcripts, isoforms, alleles, etc., and the remainders are clusters with no similar sequences. Since toxin families were shown to be highly expressed in snake transcriptomes (JUNQUEIRA-DE-AZEVEDO and HO 2002), the

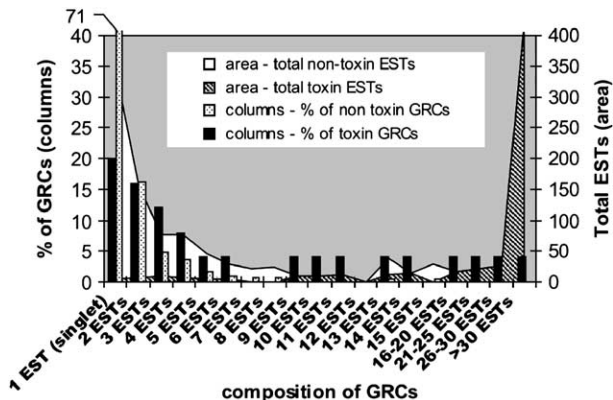


FIGURE 1.—Frequencies of the *GRCs* according to the number of ESTs. All ESTs from all *GRCs* containing 1, 2, 3, or *n* ESTs from toxin or nontoxin clusters were summed and presented absolutely (area) or in relation (columns) to the total number of ESTs in each category.

analysis of the abundance of these *GRCs* is also useful in observing the expression patterns of transcripts coding “toxin” or “nontoxin” proteins and thus in defining thresholds that could help to mine for new toxins, supporting a given transcript as a toxin or not. Therefore, the frequencies of *GRCs* according to the number of ESTs that they contained are shown in Figure 1, tabulated separately for *GRCs* further identified by similarity (see below) as coding “toxin” or “nontoxin” proteins. The lowly expressed *GRCs* (those with a low number of ESTs) are much more frequent than the highly expressed ones (those with many ESTs on a *GRC*), especially for the nontoxin *GRCs* (Figure 1, dotted bars). Nevertheless, the frequencies of medium to highly expressed “toxin” *GRCs* (solid bars) are unusually high, holding a large set of ESTs (shaded area). This allows the approximation of a threshold of abundance (approximately more than three ESTs per *GRC*) beyond which the sequences in a *GRC* are more often expected to be toxins, as shown below.

Consensual cluster sequences were compared against GenBank: 429 (25.1%) were not identified, but 717 (74.9% of clones) produced significant hits (*e*-value <10⁻⁰⁵), being separated in those matching known toxin sequences (toxins) and those matching any other products (nontoxins) (Table 1 and Figure 2, center).

The complete list with putative identifications can be provided upon request.

The typical Viperidae toxins: The overall distribution of toxins (Figure 2, right) shows that the most common types of pit viper toxins may be present in *L. muta*. Their relevant properties are described below.

The Bradykinin-potentiating peptides/C-type natriuretic precursor is the most expressed toxin transcript: The Bradykinin-potentiating peptides (BPPs)/C-type natriuretic precursor (CNP) was identified by similarity to that from many species and represents 72% of all toxin transcripts and 20% of venom gland mRNAs. There were 27 clusters (418 ESTs), all but one included in the *GRC* 21. In fact, there are two almost identical cDNAs differing by a few nucleotides, each one containing half of the total ESTs, suggesting two expressed alleles. The prototypical BPP precursor found here contains five BPPs (*Lm*-BPP 1–5) interspersed with spacers, plus the 22-amino-acid-long CNP at the C terminus, and was further studied by means of resequencing the cDNAs characterized from the venom and is described elsewhere (SOARES *et al.* 2005). The remarkable feature is that *Lm*-BPP 1 is unique because it was shown by matrix-assisted laser desorption ionization time-of-flight mass spectrometry to be processed at the N terminus, being three residues shorter than that expected from the cDNA, beginning with a Trp at the N terminus (SOARES *et al.* 2005), whereas all other purified BPPs have a N-terminal Gln residue that is circularized to pyroglutamate (HIGUCHI *et al.* 1999).

The different types of SVMPs reveal a domain-exchanging mechanism: SVMP precursor cDNAs are classified in three confirmed structural groups according to the presence of extra domains (Figure 3A) (HITE *et al.* 1994). Two *L. muta* SVMPs have been characterized so far: mutalysin I, a highly hemorrhagic protease, and mutalysin IIa, a low hemorrhagic but potent fibrin(ogen)olytic protease, probably from the N-I type (SANCHEZ *et al.* 2003). This mutalysin IIa was almost perfectly matched with cluster LMUT0061C, and the four-residue differences suggest that it could be the mutalysin IIB isoform (SANCHEZ *et al.* 2003) or a new one.

It was also possible to identify some precursors passing through the disintegrin region, including five of the N-II group (Figure 3B). The only disintegrin characterized so far from *Lachesis* is lachesin (P31990) and

TABLE 1
Representation of the 1146 clusters assembled from the *L. muta* ESTs

Category	Clusters	Clones	Clones/ cluster	% of total clones	% of matching clones
Matching sequences	717	1519	2.1	74.90	—
Similar to toxin transcripts/proteins	68	571	8.4	28.15	37.6
Similar to nontoxin transcripts/proteins	649	948	1.5	46.75	62.4
No database match	429	509	1.2	25.10	—

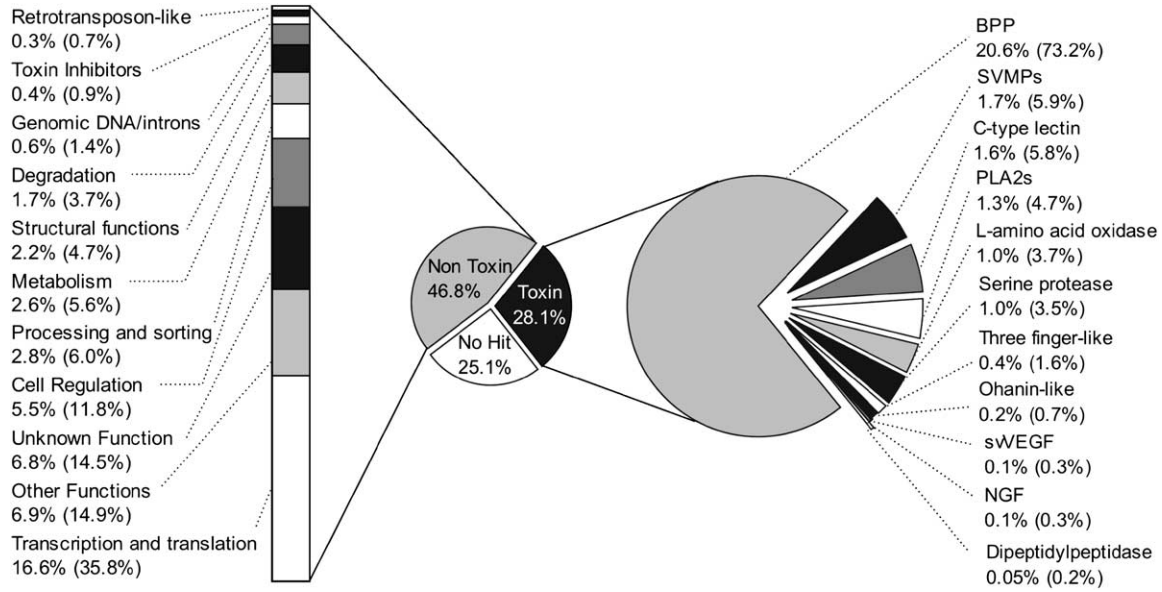


FIGURE 2.—Relative proportion of each category of product found. Toxin, nontoxin, and “no hit” categories are, respectively, clusters matching GenBank snake toxin sequences, nonsnake venom toxin sequences, or no sequence with an e -value $< 10^{-05}$ on Blastx or Blastn searches. Percentages of total ESTs and percentages of ESTs over the category are presented.

two clusters match it: LMUT0050C with 100% and LMUT0067C with an Ala \rightarrow Thr polymorphism (at position 22 of the alignment). LMUT0023C has an important distinctive feature in lachesin: in addition to two position changes, there is a three-amino-acid insertion (position 25–27), including an extra cysteine residue, proposed in jerdonitin (AAQ63966) from *Trimeresurus*

jerdonii to form an extra disulfide bond responsible for the linking of disintegrin with the metalloproteinase domains, which hinders the release of the first from the latter after processing (CHEN *et al.* 2003). Two related clusters, LMUT0836S and LMUT1111S, are structurally interesting because indubitably they are N-II precursors, since they have the stop codon after the disintegrin

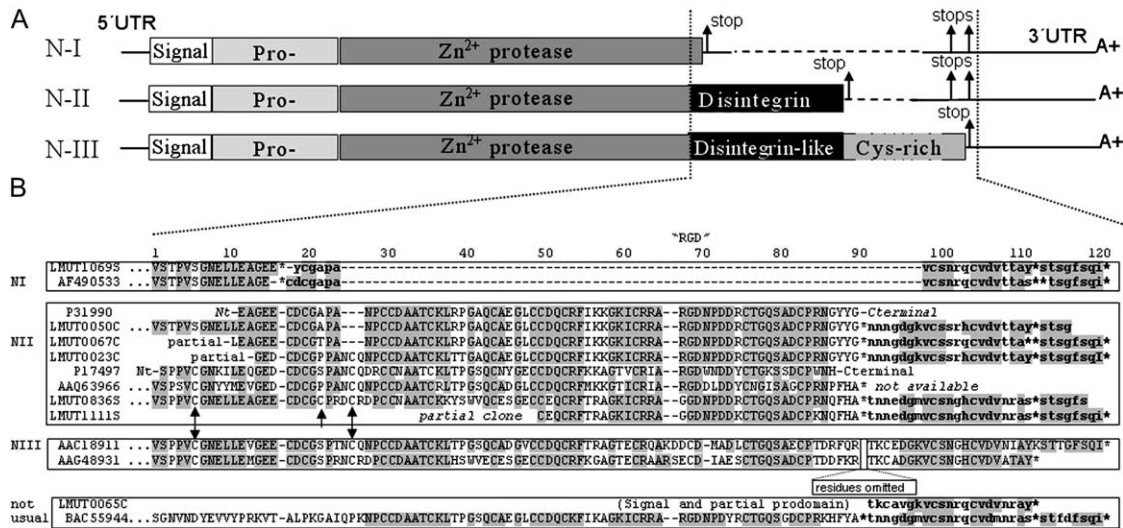


FIGURE 3.—Complexity of the coding pattern of SVMP domains from cDNAs. (A) Schematic of SVMP precursors based on HITE *et al.* (1994). (B) Alignment of the translated amino acid sequences of N-I, N-II, and N-III precursors passing through the disintegrin and Cys-rich domains (part of the N-III Cys-rich sequence was omitted to fit the scheme). If the stop codons (*) were ignored and frameshifted in one base, the 3'-UTR could be translated in the amino acid sequences shown in boldface lowercase type, which match coding regions from other precursor types. LMUTs are clusters from this work, P31990 is lachesin from *L. muta* venom, and the others are AF490533 (*B. insularis*), P17497 (bitistatin; *Bitis arietans*), AAQ63966 (jerdonin; *T. jerdonii*), AAC18911 (*Agkistrodon contortrix*), AAG48931 (HF3; *Bothrops jararaca*), BAC55944 (acostatin; *A. contortrix*). “RGD” marks the integrin bind site and arrows indicate the unusual cysteines of LMU0836S.

domain, but they show more sequence similarities to N-III at the metalloprotease (not shown) and at the beginning of the disintegrin region (Figure 3B). For instance, the 14-residue N-terminal linker sequence of the longer one (LMUT0836S) is identical to that of N-III metalloproteases, including a cysteine residue present in N-III and in some longer N-II-derived disintegrins, such as bitistatin (P17497). In these clusters, there is an alternative GGD sequence in the place of the canonical RGD integrin-binding site. The derived polypeptide of these clusters, if present in the venom, is likely to represent a possible atypical novel P-II long disintegrin.

Assuming that N-I, N-II, and N-III precursors are present in hundreds of species and thus are supposed to have diverged early, the “mixed” molecules found here could possibly result from a convergent shift of a given type to another after gene duplication, either by accumulating punctual changes along the molecule or by gaining, losing, or exchanging parts of their genes. In fact, examining cDNAs from various precursor types, we found three pieces of evidence pointing to the existence of this second possible molecular evolutionary mechanism. First, if the stop codons of N-I mRNAs (after the metalloprotease domain) are ignored and the immediate beginning of the 3'-UTR is forcedly translated, we will obtain an amino acid sequence matching the N terminus of the N-II disintegrin domain, followed by the C terminus of the N-III Cys-rich domain (boldface lowercase residues of LMUT1069S and AF490533 N-I precursors in Figure 3B). In the same way, if the stop codons of N-II mRNAs (after disintegrin) are ignored, the following codons reveal the amino acid sequence of the C terminus of the N-III Cys-rich domain (N-II boldface lowercase residues in Figure 3B). This observation is not restricted to the *L. muta* mRNAs but could be extended to other species. Second, a very unusual cluster, LMUT0065C, codes a molecule with a signal peptide, followed by a frameshifting deletion of the first 33 amino acids from the prodomain, after which only the C-terminal Cys-rich region is present, being absent in the important catalytic and disintegrin domains (Figure 3B). This unusual cDNA is not simply an artifact since there are six independent clones with this improbable organization. And third, a disintegrin cDNA precursor coding only for the signal peptide, the prodomain, and the disintegrin (without the metalloprotease) was described from *Agkistrodon* species (BAC55944; OKUDA *et al.* 2002) (Figure 3B). These observations suggest that the distinct domains of SVMPs, in fact parts of them, may be coded by different gene segments (probably exons) that could be lost or gained in each precursor-type gene, possibly by exon-shuffling mechanisms.

Wide range of serine protease functions: Three subgroups of serine proteases, named according to their main (but not exclusive) activity, were proposed and associated with typical residues: CL, for coagulating enzymes; PA, for plasminogen activators; and KN, for kininogenases

(WANG *et al.* 2001). Our database revealed eight clusters of matching serine proteases, six within the coding region that could be putatively assigned to the three groups (supplemental Figure 2A at <http://www.genetics.org/supplemental/>). LMUT0131C, although partial, was very similar to venombin (S35689), a fibrinogenase characterized from *L. muta*. Accordingly, it possesses typical basic residues, R₉₁ and R₁₉₃ (numbering based on chymotrypsinogen), in addition to F₉₆ and T₉₇, both usual in CL (supplemental Figure 2A at <http://www.genetics.org/supplemental/>). LMUT0128C was full length, possessing the V₈₄, S₉₄, and S₉₅ typical of the KN group. Interestingly, it is more similar to halystase (P81176), a KN from *Agkistrodon halys*, than to LV-Ka, a KN already described from *L. muta* (WEINBERG *et al.* 2004), suggesting that it is another KN from this species. Finally, LMUT0402S, which was full length sequenced with internal primers (data not shown), is similar but not identical to LV-PA from *L. muta* venom (P84036) and shows the P₉₃, N₉₄, R₉₅, and the acid residue at 97, suggesting another plasminogen activator. Since various other partial clusters were found, a high diversity of this component in the venom should be expected, as previously suggested (DINIZ and OLIVEIRA 1992).

The common C-type lectins: Common C-type lectins (CTLs) are dimeric proteins containing a carbohydrate recognition domain (CRD) and are ubiquitously expressed in almost all snake venoms. Six different CTL clusters were observed. LMUT0078 and LMUT0074 correspond, respectively, to α - and β -chains of factor Xa inhibitors from various sources (supplemental Figure 2B at <http://www.genetics.org/supplemental/> and Figure 4). Interestingly, both are within the equally most expressed GRCs of CTLs, reflecting the equal stoichiometric ratio expected for the two chains of a dimeric protein. LMUT0112C is not a complete cluster but its sequence is 97% identical to mutina, the galactose binding “true” lectin of *Lachesis stenophrys* (ARAGON-ORTIZ *et al.* 1996). Another possible CTL with a very unusual sequence was also found and is described later.

PLA2s: Two clusters of PLA2s, LMUT0119C and LMUT0188C, were identified—the first with 25 ESTs assembling a full-length cDNA and the second, with 2 ESTs, generating a partial sequence. Both are typical acid (Asp49) PLA2s. Interestingly, LMUT0119C did not show a perfect match with any of the four *L. muta* PLA2 isoforms previously characterized, and the other, LMUT0188C, could yield only an inconclusive match with LM-PLA-I and -II (DAMICO *et al.* 2005) due to the partial coverage. It should be noted that we were unable to find (Lys49) PLA2.

L-amino acid oxidases: Two almost identical clusters (within GRC66) code for L-amino acid oxidases (LAOs). LMUT0069C is the longest cDNA assembled on the library (2705 pb), matches LAO from Viperidae (~90%) over the entire extension, and should correspond to the *L. muta* LAO identified earlier by SANCHEZ

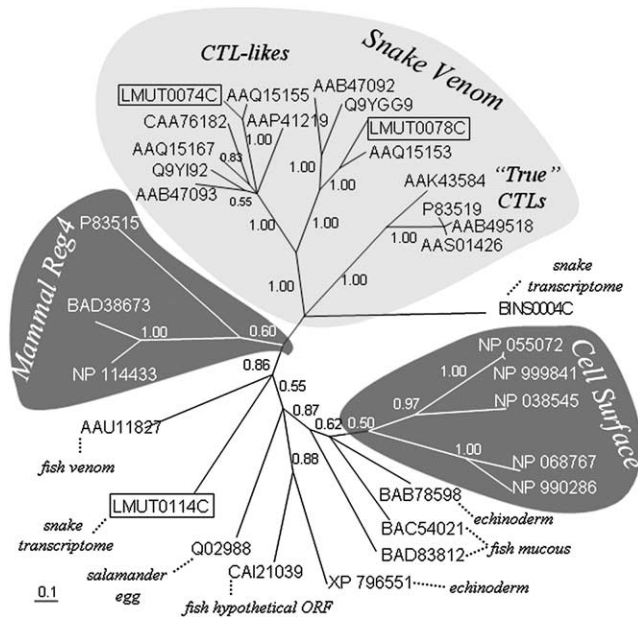


FIGURE 4.—Positioning of *L. muta* clusters (LMUTs) among CTL phylogeny. Bayesian analysis was carried out on the strict CRD of the proteins indicated by their GenBank accession numbers on the tree. Note that LMUT0074C and LMUT0078C are typically CTL like, whereas LMUT0114C is from the snake venom CTLs, being in a basal branch, as are the other secreted CTLs.

and MAGALHÃES (1991). Although LAOs have been studied for a long time regarding some activities, they have an unclear role in venoms.

Growth factors: Some growth-factor-like proteins are now known to be constituents of many snake venoms. Nerve growth factor (NGF) was first isolated (HOGUE-ANGELETTI *et al.* 1976), but its role is still controversial. More recently, KOMORI *et al.* (1999), GASMI *et al.* (2000), JUNQUEIRA-DE-AZEVEDO *et al.* (2001), and others have described snake venom vascular endothelial growth factors (svVEGFs) from Viperidae species. Subsequently, several studies have extended these observations, helping to establish it as a hypotensive and vascular-permeability-increasing factor. Here we found clusters coding for single forms of both NGF and svVEGF, helping to establish their wide occurrence.

Other remarks: A singlet cluster matching only the long 3'-UTR of a dipeptidylpeptidase 4b cDNA, an unpublished sequence from *Gloydus blomhoffi*, was also found and, surprisingly, no cluster coding for the recently described cysteine-rich secretory proteins, shown to be present in many snakes from Elpidae to Viperidae (YAMAZAKI *et al.* 2003).

The atypical possible toxins: In addition to the presence of common Viperidae toxins, to our surprise, one cluster codes a very divergent form of CTL and two other code sequences that are similar to toxins described only in Elapidae venoms.

A divergent CTL from a new recruitment: The descriptions of new venom CTL sequences usually report mole-

cules not very divergent from previously known ones. Conversely, LMUT0114C is a medium-abundant cluster that unusually shows identities at the same low level (~34%) to snake and nonsnake lectins, including a different Cys pattern at the C terminus (supplement Figure 2B at <http://www.genetics.org/supplement/>). In agreement, Bayesian phylogenetic analysis of the CRD region places it in an undefined branch with many metazoan CTLs involved in several roles, including that of a fish toxin (AAU11827), but separated out with a strong branch support from the typical snake CTLs, which appear to be derived from a single ancestor sequence related to the nontoxin reg4 protein (Figure 4). If we assume the wide distribution of CTLs in the most distant snakes, and, despite this, its grouping under a single common ancestor, this atypical LMUT0114C would suggest a different event of CRD recruitment to the venom, possibly from a diverse source of CTL scaffold. While it cannot be definitely classified as a venom constituent, our data suggest this, since its expression level (five clones) is above the proposed "toxin cutoff" and its ORF has the organization of the typically secreted CTLs rather than that of many physiological proteins, which are proteoglycans or membrane-associated proteins that contain other extensions in addition to the CRD. Interestingly, in the *Bothrops insularis* transcriptome (JUNQUEIRA-DE-AZEVEDO and HO 2002) we also found a distinctive CTL (BINS0004C), but in this case it could be grouped with snake CTLs although it should have diverged earlier (Figure 4). These CTLs are then likely to be very recent with an unexpected function remaining to be investigated.

Ohanin-like is also a Viperidae toxin: Ohanin was first isolated from *Ophiophagus hannah* (king cobra) and shown to cause hypolocomotion and hyperalgesia in mice (PUNG *et al.* 2005). Thai cobrin (P82885), another related protein from *Naja kaouthia*, was deposited in databases without further publication. A related molecule was also recently identified in the incipient maxillary venom glands of the lizard *Pogona barbata* from the family referenced as Vespryn (FRY *et al.* 2006). The snake molecules are 93% similar to each other and much less (~54%) similar to the PRY-SPRY domain containing proteins. These in tandem subdomains are part of a larger B30.2-like domain conserved in the C-terminal region of many proteins showing distinct functions, including butyrophilin (an integral membrane protein secreted in milk fat droplets), RING-finger proteins (cytoplasmatic), stonustoxin (a fish toxin), and many others. The cluster LMUT0120C, composed of four clones, shows 93 and 85% identity, respectively, to the mature ohanin and to Thai cobrin, thus becoming the third member of this new group of toxins (Figure 5A). The complete cDNA sequence of 1500 bp was resolved with internal primers (data not shown), which code a precursor molecule of 218 amino acids, with the probable starting methionine 48 residues upstream from the position corresponding to the mature N

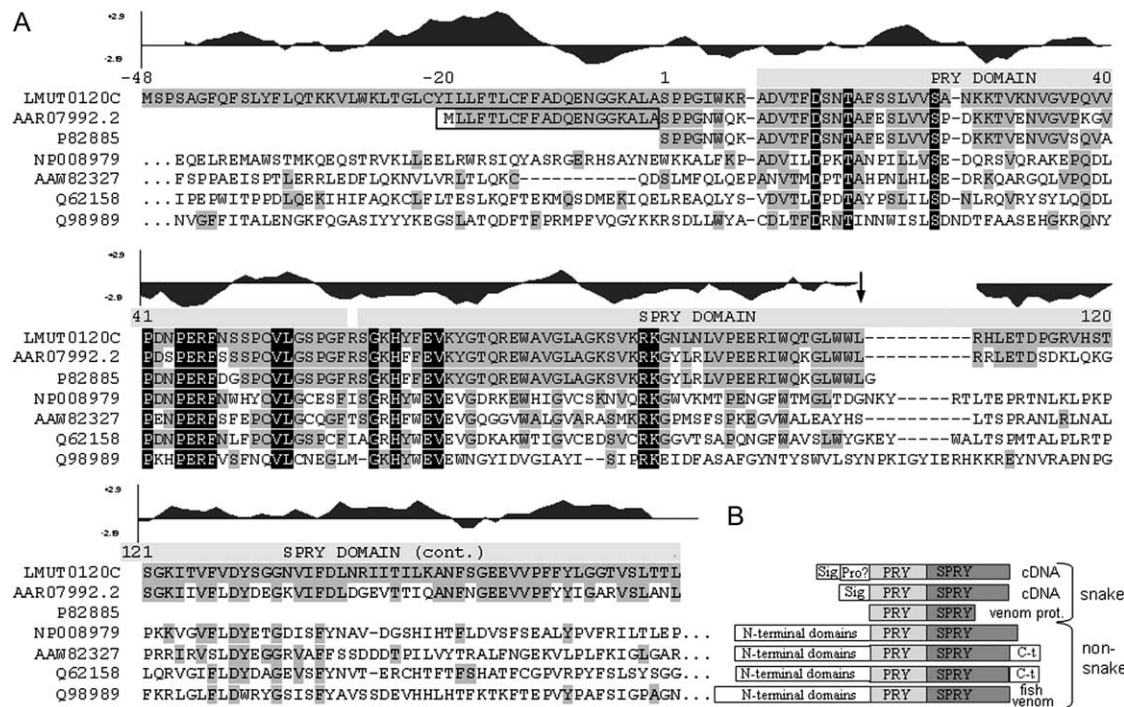


FIGURE 5.—Primary structure of the *L. muta* ohanin-like precursor. (A) Cluster LMUT0120C is aligned with the following B30.2-containing proteins: *Ophiophagus hannah* ohanin precursor (AAR07992.2), *Naja kaouthia* Thai cobrin venom (P82885), human butyrophilin (NP008979), chicken TRIM27 (AAW82327), mouse ret finger protein (Q62158), and the *Synanceia horrida* fish stonustoxin (Q98989), numbered from the mature ohanin-like N terminus. Putative ohanin signal peptide is boxed, and the C terminus of mature ohanin/Thai cobrin is indicated by an arrow. The PRY and the SPRY domains predicted by SMART (<http://smart.embl-heidelberg.de>) and a hydropathicity plot (KYTE and DOOLITTLE 1982) of *L. muta* ohanin-like precursor (LMUT0120C) are marked above the sequences. (B) Schematic of whole-sequence alignment of the above proteins, showing that nonsnake toxins possess other domains.

terminus of ohanin (Figure 5A). Conversely, the unpublished ohanin precursor available on databases (AAR07992.2) has its starting codon only 20 positions upstream from the mature N terminus (Figure 5A). The precursor indicates also a C-terminal processing, since the molecules purified from the venom end 63 residues before the stop codon (PUNG *et al.* 2005). This processed C terminus is different from other B30.2 proteins (Figure 5B); it is shorter and predominantly hydrophobic (see plot in Figure 5A), perhaps suggesting a membrane association during the maturation process. These ohanin-like proteins would be the shortest members of the B30.2 family, and, whatever their main activity is, it is probably due only to the PRY and the half-SPRY domains, whereas other members also have longer and distinctive N-terminal domains responsible for their various activities (Figure 5B). In addition, unlike most of the other B30.2-containing proteins that are cytoplasmatic or nuclear, these are secreted, and unlike other toxins, have a low Cys content. The finding of this *L. muta* ohanin-like toxin would show the general occurrence of this new type of B30.2 protein in different venoms and may provide some insights into the structure and function of this widely distributed domain.

The three-finger-toxin-like scaffold is available in Viperidae: Perhaps the most interesting observation of this work is the finding of an abundant cluster (LMUT0205C), composed of nine clones, assembling a complete cDNA of 561 bp with an ORF of 95 codons, including a signal peptide that, when translated, matches Elapidae three-finger-shaped toxins (3FTx). These are short proteins, 60–75 residues long, involved in several pharmacological effects, which are most related to their ability to bind and block nicotinic acetylcholine receptors, resulting in the famous postsynaptic neurotoxicity of oriental and marine snake venoms (NIRTHANAN and GWEE 2004). Different from the above case of ohanin-like toxin, which is a recently described class of toxin, during the last 50 years there have been up to 350 sequences of snake 3FTx found exclusively in Elapidae (including Hydrophiinae) venoms and only recently in Colubridae (FRY *et al.* 2003a). The most important conserved feature of their sequences is the cysteine arrangement, which involves four or five disulfide bonds responsible for the characteristic structure of three β -sheet looped domains, known as the “three-finger” shape. The distinctive features of their sequences have been largely correlated to the particular activities, making these short toxins a good target for structure-function and

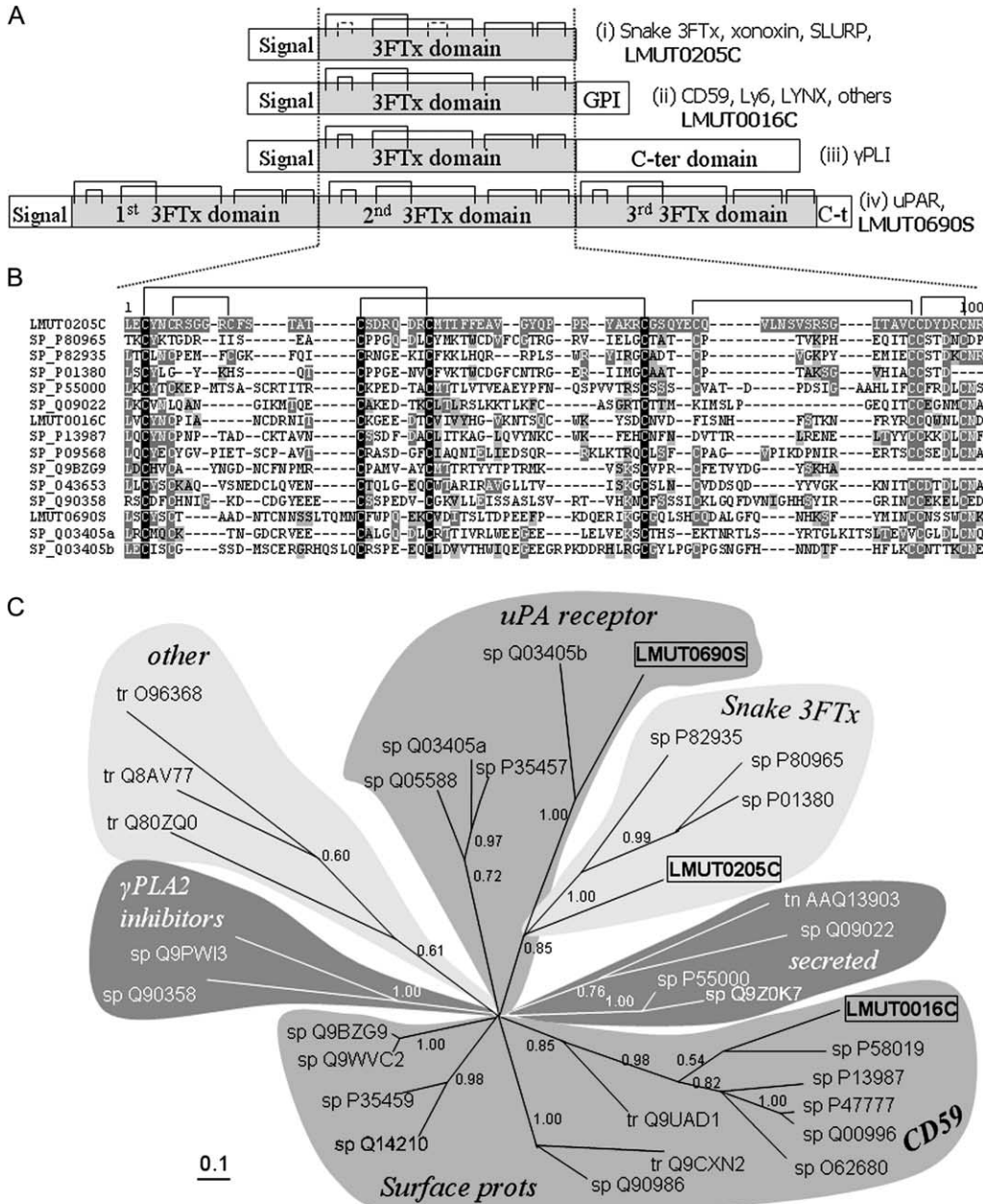


FIGURE 6.—Primary structure of the *L. muta* 3FTx-like precursor. (A) Schematic of the four arrangements (i–iv) of 3FTx-domain-containing precursors. (B) The strict 3FTx domains of some representative proteins are aligned with LMUTs on the basis of a larger prealigned data set kindly provided by B. G. Fry (FRY 2005). (C) Bayesian analysis was carried out on the proteins indicated by their database and accession numbers on the tree (some branches deriving from the central node were omitted).

drug development studies (MENEZ 1998; NIRTHANAN and GWEE 2004). This shape also appears in some other groups of proteins, as follows (see Figure 6A): (i) as the whole molecule of some secreted proteins, such as the frog toxin xenoxin and the mammal blood protein SLURP1; (ii) as part of physiological surface proteins containing a hydrophobic GPI consensus at the C terminus, such as CD59 antigen, Ly6, and Lynx; (iii) as the N-terminal half of the snake plasma gamma PLA2 inhibitor (γ PLI); and (iv) three times in tandem repeated in the urokinase plasminogen activator receptor (uPAR). Clusters coding this later (LMUT0690S) and CD59 (LMUT0016C) were also found, but the presence of the C-terminal or in tandem domains leaves no doubt that they are not other 3FTx-like toxins.

Sequence comparisons of LMUT0205C-coded polypeptide show few similarities with 3FTx, not allowing even a speculative prediction of what kind of activity it could have if present in the venom. However, it possesses the short-distance Cys bond in loop I, present in some venom 3FTx and in almost all nontoxins (Figure 6B). Bayesian phylogenetic analysis (Figure 6C) based on the data set previously used by FRY (2005), in fact, placed it in the basal but not the exclusive node with 3FTx, together with domain 2 of uPAR. But this analysis was carried out on the strict 3FTx domain, and if we consider the surrounding regions outside the 3FTx domain, it becomes clear that this is a protein of the type (i) cited above (Figure 6A), since its ORF matches the beginning and the ending positions that are characteristic

TABLE 2
Clusters not matching venom components that may be associated with venom functions

Product	No. of ESTs	Reason
5' nucleotidase	3	5' nucleotide degradation activity is widely described in snake venoms but the enzyme responsible for it has never been characterized. The molecule found has a signal peptide.
ADAM (a disintegrin and metalloprotease)	1	Metalloproteases closely related to the SVMPs, the important toxins from Viperidae. The sequence found showed a strong match to ADAM 4 at the C-terminal spacer portion.
Cathepsins L and S	2	Although cysteine proteases have never been described in snake venoms, they could be involved in venom processing. Due to the importance of proteases in envenoming, they may also be involved in toxicity.
uPAR (urokinase type plasminogen activator receptor)	1	Tissue-type plasminogen activator is present in most Viperidae venoms. This uPAR may point to some involvement of uPAR in coagulation disturbances. Its second 3FTx domain is aligned in Figure 6.
S-lectin	1	Galectins show cell-surface interaction functions that may be related to the roles of CTL. These proteins are secreted without a signal sequence and were never described in snake venoms.
β -PLIs (β -PLA2 inhibitors)	10	β -PLIs are trimeric proteins of 50-kDa glycosylated subunits, known for inhibiting venom PLA2s. They were isolated from plasma of only two snake species and thought to be expressed in the liver and then secreted.
TIMP (tissue inhibitors of metalloproteases)	2	May be an inhibitor of SVMPs. TIMP1 has never been described as a natural protective molecule against SVMPs, although several other types of inhibitors were described in snake sera.

of those from the venom 3FTx, the frog toxin, and the SLURPs. However, the phylogenetic data clustered these short secreted proteins out from LMUT0205C. Although the correct assignment of it as a toxin or not depends on biochemical analysis, the phylogenetic information, the fact that it was cloned from a snake venom gland, and the expression level (nine ESTs) that is three times higher than the proposed "toxin cutoff" suggest that it is a possible new toxin.

Other sequences related to venom functioning: In addition to those clusters matching snake toxins, either from Viperidae or Elapidae, some sequences do not match any snake protein and would be classified at first glance as cellular proteins, but due to particular reasons we could associate them with possible venom functions. There are also toxin inhibitors never observed in the venom glands, possibly associated with a protective effect on their cells, although a possible participation in the envenoming through inhibition of host enzymes should not be discounted (see Table 2).

Panorama of clusters related to venom gland physiology: The large universe of nontoxin matching transcripts, 47% of ESTs (649 clusters), was categorized according to the major physiologic functions of cells.

A low redundancy was observed, 1.5% (*vs.* 8.4% for toxins), indicating that the universe of cellular transcripts is very diversified and that this medium-throughput EST sequencing will give only a survey of them rather than a complete description. Nevertheless, the relative abundance of each category could be estimated (Figure 2, left). As previously observed for venom glands and other secretory tissues, DNA transcription and translation is the most expressed set, indicating protein synthesis specialization of this tissue. Accordingly, protein processing and sorting is also intense and the individually most expressed nontoxin transcript is within this category: the protein disulfide isomerase (2 clusters containing 15 and 8 ESTs each). Several (117 clusters) conserved cDNAs with unknown function were found. Six different (retro)transposable-like elements were identified in our dbEST, as previously observed for *B. insularis* (JUNQUEIRA-DE-AZEVEDO and HO 2002). Most are truncated forms of ORF2 from several sources but one of them corresponds to the Bov-B LINE, reported inside the PLA2 and other toxin introns and is highly widespread on Viperidae genomes (KORDIS and GUBENSEK 1997). The close relationship of these sequences with toxin gene duplication is becoming evident and the

biological significance of their expression in the venom glands is an interesting matter to be explored.

There were 429 clusters (25% of total) without significant matches to the database (e -value $>10^{-05}$) that are generally in low abundance (approximately one clone per cluster). But there was an extreme exception: *GRC 08* grouped 18 clusters (62 ESTs) with absolutely no hit found against any database over its >2 -kb sequence extension. Intriguingly, no ORF was found, with all the frames showing several stop codons. So this could indicate an unusually long 3'-UTR of some incomplete highly expressed molecule or possibly a regulatory RNA that could be important in this highly expressing tissue. From the rest of the no-match set, the longest ORFs from each cluster were predicted and screened for possible signal peptides. Sixty-four were positive, 24 of which were highly probable, which might represent new venom components.

DISCUSSION

The diversity of toxin components makes venom a valuable source of bioactive molecules that frequently interfere with the homeostasis of prey by pathways related to common physiopathological processes. The information given by a pool of transcripts would be a first step in cataloguing this diversity but should be considered under the notion that not all of them would be translated in proteins and those translated could not arise in proteins at the same level of their mRNAs and should pass through several post-translational changes. Nevertheless, transcriptomic approaches based on ESTs applied to venom study are proving to be valuable tools in describing sequences of previously isolated toxins and in identifying new unexpected ones, which would not be feasible through direct biochemical approach (JUNQUEIRA-DE-AZEVEDO *et al.* 2001; JUNQUEIRA-DE-AZEVEDO and HO 2002; FRANCISCHETTI *et al.* 2004; KASHIMA *et al.* 2004; MAGALHÃES *et al.* 2006; QINGHUA *et al.* 2006).

In this work, we have attempted to investigate possible toxins from the Viperidae snake *L. muta* through a transcriptomic approach. This fragile equatorial species is difficult to keep in captivity and the few animals that have thrived *ex situ* are used in antivenom production. The specimen used here was obtained from such a source and its transcriptome represents the catalog of messages from this animal at one moment, perhaps not representing the large variability of venom commonly observed between individuals and the possible variation within a single individual during aging (CHIPPAUX *et al.* 1991). Considering this, we observed that the transcripts related to known toxins make up more than a quarter of the gene expression in *L. muta* venom glands, which is much lower than that observed for *B. insularis* (56%) (JUNQUEIRA-DE-AZEVEDO and HO 2002), for *Bothrops*

jararacussu (60%) (KASHIMA *et al.* 2004), for *Bitis gabonica* (46%) (FRANCISCHETTI *et al.* 2004), or for *Agkistrodon acutus* (40%) (QINGHUA *et al.* 2006). This could be related to the fact that the animal used was not stimulated to produce venom by milking prior to the RNA extraction and may represent the physiological steady-state transcription of this tissue. The striking observation of those toxins is the unexpected exaggerated abundance of BPP precursor mRNA. Most probably there is a differential regulation of BPP mRNA expression and/or turnover in the steady state, since no other studies reported such an expression level of BPP mRNA in stimulated venom glands. The peptides secreted to the venom may or may not be at the same abundance, but these data and the presence of an unusually processed BPP at the venom (SOARES *et al.* 2005) suggest an important contribution of this component to the known hypotensive effect of the venom (JORGE *et al.* 1997). The overall occurrence of the other toxins is consistent with that observed in the Viperidae transcriptomes cited above.

In addition to the typical toxins expected for a Viperidae species, we also identified unusual possible toxins represented by (i) diverging members of those common classes; (ii) molecules related to toxins occurring in distant snake groups; and (iii) proteins that, although not matching toxin sequences, could be new toxin candidates. As an example of case (i) of diverging members from a common toxin class, we found a distinctive CTL transcript that possessed a diverse evolutionary history. As examples of case (ii), for the first time in a Viperidae snake a representative sequence of the most important group of Elapidae/Colubridae toxins (3FTx), which possesses a versatile scaffold exploited in evolutionary history for many physiological purposes or artificially in drug development, was observed (MENEZ 1998). In addition, the ohanin-like molecule, being found in a Viperidae, reshaped the notion of an exclusive Elapidae toxin. In case (iii), there are some sequences such as the 5' nucleotidase, some proteases, etc., whose reported activities resemble venom toxicities or also toxin inhibitors that should be reanalyzed as possible new toxins.

It is not possible to define whether the "Elapidae proteins" (3FTx-like and ohanin-like) are present in other Viperidae species and simply not yet observed or are characteristic of Lachesis. The possibility of other Viperidae orthologous yet undiscovered seems more reasonable, since only a few large cDNA data sets are available. The recent transcriptomic analysis of *A. acutus* (Viperidae) report clusters matching, by BlastN searches, nucleotide sequences of 3FTx from Elapidae, but noncoding parts of toxin genes are used to be widespread in snake genomes (KORDIS and GUBENSEK 1997). In fact we also observed other clusters in our database matching the same 3FTx gene segment from *Bungarus multicinctus*, but they seem not to code a

polypeptide. Other reports of postsynaptic 3FTx neurotoxins in Viperidae are based only on immuno or pharmacological evidences (JIANG *et al.* 1987) and do not describe sequences, excepted by a N terminus of a possible 3FTx from *Daboia russelli* (SHELKE *et al.* 2002).

In any case, one can ask if these unusual proteins derive from a plesiomorphic character present in a common Elapidae/Viperidae ancestor or derive from a new recruitment event (an apomorphic character). For the ohanin-like, since the similarity is high and this class of protein was only recently described (PUNG *et al.* 2005), being reported also in lizards (FRY *et al.* 2006), we believe that a common origin is very probable. For the diverging 3FTx, the phylogenetic tree also suggests a common and ancestral origin with Colubridae and Elapidae toxins, but in this case, the evidence is not very strong, since the uPAR (domain 2) also was in the group. The short length and the rapid evolution of the 3FTx group make a through conclusion on the basis of this single Viperidae sequence difficult. The Colubridae 3FTx (α -colubritoxin), for instance, was more closely grouped within Elapidae toxins here and elsewhere and has been proposed to evolve before the split of the Colubroidea superfamily (FRY *et al.* 2003a). In fact, the recently proposed phylogenies of snake 3FTx are strong in grouping most functional types with common ancestors but not very efficient in showing their order of appearance; meanwhile, their evolutionary relationship with the nontoxin proteins is also not well demonstrated (FRY *et al.* 2003b; PHUI-YEE *et al.* 2004; FRY 2005), frequently resulting in polytomic trees, as we obtained. Because it commonly or independently originates with snake venom toxins, the possible presence of a 3FTx protein in a Viperidae venom will support the 3FTx domain as a versatile scaffold for toxic purposes, independent of the pharmacological action, which is conceivable if we consider the well-known versatility of this domain in binding many diverse things (ALAPE-GIRON *et al.* 1999). If not a toxin, it will become a singular new member of the proteins possessing the 3FTx shape. Independently of the steps in the evolutionary process, the idea of recruiting or upholding important scaffolds for diverse functions (ALAPE-GIRÓN *et al.* 1999) will be sustained once more, as in other cases discussed here. And primarily, Viperidae venoms should be reanalyzed in the search of such unexpected molecules.

The relationship between the evolutionary origins of the main taxa of venomous reptiles and the time at which the toxin classes appeared is only now becoming of interest (FRY *et al.* 2006), but several particular issues still need to be addressed. In any case, it is possible to speculate, on the basis of the molecules observed here and from the literature, that the repertoire of toxin scaffolds available for recruitment in snake families is not very distinctive, despite the different abundance that they present. This means that one could require

more molecules of a given type whereas the other would discard them for adaptation purposes, although a common scaffold could still be present.

The bioinformatics analysis of large sets of toxin cDNAs also allows some insights into the mechanisms of recruitment and evolution of some protein domains. For instance, the SVMP cDNAs show stretches of sequences presented as 3'-UTRs in some transcripts and as parts of translated domains in others, suggesting that exon shuffling between paralogous genes followed by stop codon mutations may allow a recruitment/disposal of particular domains. This could eventually happen more than once during SVMP evolution in a particular species, thus allowing the generation, for instance, of N-II precursors with an N-III resembling metalloprotease domain, as observed for LMUT0836S or for the intriguing LMUT0065C, which may result from a "mis-assembled" allele. Of course, the ultimate resolution of this hypothesis will depend on the cloning of SVMP genes, an important but until now neglected issue in the literature. Another feature of the domain organization of toxins that we could observe from the different classes of proteins in the database is that the snake toxins frequently seem to correspond to a minimally functional domain of some large multidomain proteins, which act as a provision for innovations. The role played in concert by a single domain within a multidomain protein could be changed if it is expressed alone and thus altered to a new function. There are many examples of this, as for the SVMPs that lack some domains present in a disintegrin and metalloprotease (ADAM) counterparts and in the case of CTLs that are represented in the venom only by the CRD domain. And it is even more evident for the two "Elapidic" molecules noted here: the 3FTx-like molecules, whose single domain could be represented as a part of larger nontoxin molecules such as CD59, Ly6, PLIs, or uPAR and the ohanin-like molecules containing a partial B30.2 domain, the first reported to be secreted without the extra domains found in their nontoxin counterparts. Of course, all the inferences about the ancestry of these domains would be speculative here, but further investigative efforts in light of these new molecules uncovered by transcriptome could help in understanding the origins of venom toxin.

Independently of the evolutionary aspects of those molecules, several complete and partial *L. muta* mRNAs were described, hopefully contributing to future understanding of venom functioning, toxin biochemistry, and immunogenic aspects relevant for antivenom production and other pharmacological interests. In addition to these interests in toxins, if we consider the hundreds of animals from the most variable taxa being scarcely studied by means of sequencing, the reptile quota is clearly underestimated, leaving a large gap of knowledge on this Chordate, which the nontoxin-like transcripts available here and from similar works would help to fill.

We thank Michael Richardson for manuscript review and Bryan G. Fry for his support in phylogenetic analysis and data sets. This work was supported by grants from the Brazilian agencies Fundação de Amparo à Pesquisa do Estado de São Paulo, Conselho Nacional de Desenvolvimento Científico e Tecnológico, Fundação Butantan, and Fundação de Amparo à Pesquisa de Minas Gerais.

LITERATURE CITED

- ALAPE-GIRÓN, A., B. PERSSON, E. CEDERLUND, M. FLORES-DIAZ, J. M. GUTIERREZ *et al.*, 1999 Elapid venom toxins: multiple recruitments of ancient scaffolds. *Eur. J. Biochem.* **259**: 225–234.
- ARAGON-ORTIZ, F., R. MENTELE and E. A. AUERSWALD, 1996 Amino acid sequence of a lectin-like protein from *Lachesis muta stenophytis* venom. *Toxicon* **34**: 763–769.
- BENDTSEN, J. D., H. NIELSEN, G. VON HEIJNE and S. BRUNAK, 2004 Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* **340**: 783–795.
- CAMPBELL, J. A., and W. W. LAMAR, 2004 *The Venomous Reptiles of the Western Hemisphere*. Cornell University Press, Ithaca, NY.
- CASTRO, H. C., M. FERNANDES and R. B. ZINGALI, 1999 Identification of bothrojaracin-like proteins in snake venoms from *Bothrops* species and *Lachesis muta*. *Toxicon* **37**: 1403–1416.
- CHEN, R. Q., Y. JIN, J. B. WU, X. D. ZHOU, Q. M. LU *et al.*, 2003 A new protein structure of P-II class snake venom metalloproteinases: it comprises metalloproteinase and disintegrin domains. *Biochem. Biophys. Res. Commun.* **310**: 182–187.
- CHIPPAUX, J. P., V. WILLIAMS and J. WHITE, 1991 Snake venom variability: methods of study, results and interpretation. *Toxicon* **29**: 1279–1303.
- CHIRGWIN, J. M., A. E. PRZYBYLA, R. J. MACDONALD and W. J. RUTTER, 1979 Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* **18**: 5294–5299.
- DAMICO, D. C., S. LILLA, G. NUCCI, L. A. PONCE-SOTO, F. V. WINCK *et al.*, 2005 Biochemical and enzymatic characterization of two basic Asp49 phospholipase A2 isoforms from *Lachesis muta muta* (*Surucucu*) venom. *Biochim. Biophys. Acta* **1726**: 75–86.
- DINIZ, M. R. V., and E. B. OLIVEIRA, 1992 Purification and properties of a kininogenin from the venom of *Lachesis muta* (bushmaster). *Toxicon* **30**: 247–258.
- ENDO, T., and N. TAMIYA, 1991 Structure-function relationships of postsynaptic neurotoxins from snake venoms, pp. 165–222 in *Snake Toxins*, edited by A. L. HARVEY. Pergamon Press, New York.
- FRANCISCHETTI, I. M. B., V. MY-PHAM, J. HARRISON, M. K. GARFIELD and J. M. RIBEIRO, 2004 *Bitis gabonica* (Gaboon viper) snake venom gland: toward a catalog for the full-length transcripts (cDNA) and proteins. *Gene* **337**: 55–69.
- FRY, B. G., 2005 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 Res.* **15**: 403–420.
- FRY, B. G., N. G. LUMSDEN, W. WUSTER, J. C. WICKRAMARATNA, W. C. HODGSON *et al.*, 2003a Isolation of a neurotoxin (alpha-colubritoxin) from a nonvenomous colubrid: evidence for early origin of venom in snakes. *J. Mol. Evol.* **57**: 446–452.
- FRY, B. G., W. WUSTER, R. M. KINI, V. BRUSIC, A. KHAN *et al.*, 2003b Molecular evolution and phylogeny of elapid snake venom three-finger toxins. *J. Mol. Evol.* **57**: 110–129.
- FRY, B. G., N. VIDAL, J. A. NORMAN, F. J. VONK, H. SCHEIB *et al.*, 2006 Early evolution of the venom system in lizards and snakes. *Nature* **439**: 584–588.
- GASMI, A., F. ABIDI, N. SRAIRI, A. OIJATAYER, H. KAROUÏ *et al.*, 2000 Purification and characterization of a growth factor-like which increases capillary permeability from *Vipera lebetina* venom. *Biochem. Biophys. Res. Commun.* **268**: 69–72.
- HIGGINS, D., J. THOMPSON, T. GIBSON, J. D. THOMPSON, D. G. HIGGINS *et al.*, 1994 CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**: 4673–4680.
- HIGUCHI, S., N. MURAYAMA, K. SAGUCHI, H. OHI, Y. FUJITA *et al.*, 1999 Bradykinin-potentiating peptides and C-type natriuretic peptides from snake venom. *Immunopharmacology* **44**: 129–135.
- HITE, L. A., L. G. JIA, J. B. BJARNASON and J. W. FOX, 1994 cDNA sequences for four snake venom metalloproteinases: structure, classification, and their relationship to mammalian reproductive proteins. *Arch. Biochem. Biophys.* **308**: 182–191.
- HOGUE-ANGELETTI, R. A., W. A. FRAZIER, J. W. JACOBS, H. D. NIALI and R. A. BRADSHAW, 1976 Purification, characterization, and partial amino acid sequence of nerve growth factor from cobra venom. *Biochemistry* **15**: 26–34.
- HUANG, X., and A. MADAN, 1999 CAP3: a DNA sequence assembly program. *Genome Res.* **9**: 868–877.
- JIANG, M., J. HAGGBLAD and E. HEILBRONN, 1987 Isolation and pharmacological characterization of a new alpha-neurotoxin (alpha-AgTx) from venom of the viper *Agkistrodon halys* (*Pallas*). *Toxicon* **25**: 1019–1022.
- JORGE, M. T., I. S. SANO-MARTINS, S. C. TOMY, S. C. CASTRO, R. A. FERRARI *et al.*, 1997 Snakebite by the bushmaster (*Lachesis muta*) in Brazil: case report and review of the literature. *Toxicon* **35**: 545–554.
- JUNQUEIRA-DE-AZEVEDO, I. L. M., and P. L. HO, 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.
- JUNQUEIRA-DE-AZEVEDO, I. L. M., S. H. FARSKY, M. L. OLIVEIRA and P. L. HO, 2001 Molecular cloning and expression of a functional snake venom vascular endothelium growth factor (VEGF) from the *Bothrops insularis* pit viper. A new member of the VEGF family of proteins. *J. Biol. Chem.* **276**: 39836–39842.
- KASHIMA, S., P. G. ROBERTO, A. M. SOARES, S. ASTOLFI-FILHO, J. O. PEREIRA *et al.*, 2004 Analysis of *Bothrops jararacussu* venomous gland transcriptome focusing on structural and functional aspects. I. Gene expression profile of highly expressed phospholipases A2. *Biochimie* **86**: 211–219.
- KOMORI, Y., T. NIKAI, K. TANIGUCHI, K. MASUDA and H. SUGIHARA, 1999 Vascular endothelial growth factor VEGF-like heparin-binding protein from the venom of *Vipera aspis aspis* (*Aspic viper*). *Biochemistry* **38**: 11796–11803.
- KORDIS, D., and F. GUBENSEK, 1997 Bov-B long interspersed repeated DNA (LINE) sequences are present in *Vipera ammodytes* phospholipase A2 genes and in genomes of *Viperidae* snakes. *Eur. J. Biochem.* **246**: 772–779.
- KYTE, J., and R. F. DOOLITTLE, 1982 A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**: 105–132.
- LAZO, G. R., J. TONG, R. MILLER, C. HSIA, C. RAUSCH *et al.*, 2001 Software scripts for quality checking of high-throughput nucleic acid sequencers. *Biotechniques* **30**: 1300–1305.
- MAGALHÃES, G. S., I. L. M. JUNQUEIRA-DE-AZEVEDO, M. LOPES-FERREIRA, D. M. LORENZINI, P. L. HO *et al.*, 2006 Transcriptome analysis of expressed sequence tags from the venom glands of the fish *Thalassophryne nattereri*. *Biochimie* (in press).
- MENEZ, A, 1998 Functional architectures of animal toxins: A clue to drug design? *Toxicon* **36**: 1557–1572.
- NIRTHANAN, S., and M. C. GWEE, 2004 Three-finger alpha-neurotoxins and the nicotinic acetylcholine receptor, forty years on. *J. Pharmacol. Sci.* **94**: 1–17.
- OKUDA, D., H. KOIKE and T. MORITA, 2002 A new gene structure of the disintegrin family: a subunit of dimeric disintegrin has a short coding region. *Biochemistry* **41**: 14248–14254.
- PELHAM, H. R., and R. J. JACKSON, 1976 An efficient mRNA-dependent translation system from reticulocyte lysates. *Eur. J. Biochem.* **67**: 247–256.
- PHUI-YEE, J. S., G. NANLING, F. AFIFIYAN, M. DONGHUI, P. SIEW-LAY *et al.*, 2004 Snake postsynaptic neurotoxins: gene structure, phylogeny and applications in research and therapy. *Biochimie* **86**: 137–149.
- PUNG, Y. F., P. T. WONG, P. P. KUMAR, W. C. HODGSON and R. M. KINI, 2005 Ohanin, a novel protein from king cobra venom, induces hypolocomotion and hyperalgesia in mice. *J. Biol. Chem.* **280**: 13137–13147.
- QINGHUA, L., Z. XIAOWEI, Y. WEI, L. CHENJI, H. YIJUN *et al.*, 2006 A catalog for transcripts in the venom gland of the *Agkistrodon acutus*: identification of the toxins potentially involved in coagulopathy. *Biochem. Biophys. Res. Commun.* **341**: 522–531.
- RONQUIST, F., and J. P. HUELSENBECK, 2003 MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**: 1572–1574.

- SANCHEZ, E. O., and A. MAGALHÃES, 1991 Purification and partial characterization of an L-amino acid oxidase from bushmaster snake (Surucucu Pico de Jaca) *Lachesis muta muta* venom. *Braz. J. Med. Biol. Res.* **24**: 249–260.
- SANCHEZ, E. F., A. MAGALHÃES, F. R. MANDELBAUM and C. R. DINIZ, 1991 Purification and characterization of the hemorrhagic factor II from the venom of the bushmaster snake (*Lachesis muta muta*). *Biochim. Biophys. Acta* **1074**: 347–356.
- SANCHEZ, E. F., C. T. SOUZA, C. A. BELLO, M. RICHARDSON, E. B. OLIVEIRA *et al.*, 2003 Resolution of isoforms of mutalysin II, the metalloproteinase from bushmaster snake venom. *Toxicon* **41**: 1021–1031.
- SOARES, M. R., A. L. OLIVEIRA-CARVALHO, L. S. WERMELINGER, R. B. ZINGALI, P. L. HO *et al.*, 2005 Identification of novel bradykinin-potentiating peptides and C-type natriuretic peptide from *Lachesis muta* venom. *Toxicon* **46**: 31–38.
- SHELKE, R. R., S. SATHISH and T. V. GOWDA, 2002 Isolation and characterization of a novel postsynaptic/cytotoxic neurotoxin from *Daboia russelli russelli* venom. *J. Pept. Res.* **59**: 257–263.
- WANG, Y. M., S. R. WANG and I. H. TSAI, 2001 Serine protease isoforms of *Deinagkistrodon acutus* venom: cloning, sequencing and phylogenetic analysis. *Biochem. J.* **354**: 161–168.
- WEINBERG, M. L., L. F. FELICORI, C. A. BELLO, H. P. MAGALHÃES, A. P. ALMEIDA *et al.*, 2004 Biochemical properties of a bushmaster snake venom serine proteinase (LV-Ka), and its kinin releasing activity evaluated in rat mesenteric arterial rings. *J. Pharmacol. Sci.* **96**: 333–342.
- YAMAZAKI, Y., F. HYODO and T. MORITA, 2003 Wide distribution of cysteine-rich secretory proteins in snake venoms: isolation and cloning of novel snake venom cysteine-rich secretory proteins. *Arch. Biochem. Biophys.* **412**: 133–141.
- ZAMUDIO, K. R., and H. W. GREENE, 1997 Phylogeography of the bushmaster (*Lachesis muta*: Viperidae): implications for neotropical biogeography, systematics, and conservation. *Biol. J. Linn. Soc.* **62**: 421–442.

Communicating editor: P. J. OEFNER