



Review

Were arachnids the first to use combinatorial peptide libraries?

Brianna L. Sollod^a, David Wilson^b, Olga Zhaxybayeva^c, J. Peter Gogarten^c,
Roger Drinkwater^b, Glenn F. King^{a,*}

^a Department of Molecular, Microbial, and Structural Biology, University of Connecticut Health Center,
263 Farmington Avenue, Farmington, CT 06032-3305, USA

^b Xenome Ltd., 50 Meiers Road, Brisbane, Qld 4068, Australia

^c Department of Molecular and Cell Biology, University of Connecticut, 91 North Eagleville Road, Storrs, CT 06269-3125, USA

Received 21 May 2004; accepted 19 July 2004

Abstract

Spiders, scorpions, and cone snails are remarkable for the extent and diversity of gene-encoded peptide neurotoxins that are expressed in their venom glands. These toxins are produced in the form of structurally constrained combinatorial peptide libraries in which there is hypermutation of essentially all residues in the mature-toxin sequence with the exception of a handful of strictly conserved cysteines that direct the three-dimensional fold of the toxin. This gene-based combinatorial peptide library strategy appears to have been first implemented by arachnids almost 400 million years ago, long before cone snails evolved a similar mechanism for generating peptide diversity.

© 2004 Elsevier Inc. All rights reserved.

Keywords: Peptide toxin; Toxin evolution; Combinatorial peptide library; Spider; Scorpion; Arachnid; Cone snail; Atracotoxin; Conotoxin; Prepropeptide; Toxin precursor; Cysteine scaffold

Contents

1. Introduction	132
2. Conotoxin evolution: focal hypermutation and diversifying selection	132
3. Peptide diversity in scorpion venoms: a tale of toxic telsons	134
4. Evolution of spider toxins: a web of intrigue	135
5. Were arachnids the first combinatorial peptide chemists?	137
Acknowledgements	137
References	137

* Corresponding author. Tel.: +1 860 679 8364; fax: +1 860 679 1652.

E-mail address: glenn@psel.uconn.edu (G.F. King).

1. Introduction

The range of organisms that produce toxins for prey capture, defense, or competitor deterrence is surprisingly diverse and includes eubacteria, arthropods, molluscs, cnidarians, plants, and numerous vertebrates. The venoms of spiders, scorpions, cone snails and, to a lesser extent, snakes are particularly interesting from a pharmaceutical and agrochemical perspective because of the remarkable extent and diversity of gene-encoded peptide toxins that are expressed in the venom glands of these animals [1,16,22,32,36,53,54]. Notably, each of these animals possesses a specialized envenomation apparatus that enables injection of venom into the soft tissue of prey, thus obviating the need for the toxins to be orally active [32]. Indeed, the vast majority of these peptides are orally inactive neurotoxins that target specific subtypes of neuronal voltage- or ligand-gated ion channels [32,53].

The pharmacological inventory encoded in the genomes of these animals is extraordinary. It has been estimated that cone snails express ~50,000 different toxic peptides based on a species count of ~500 (see <http://faculty.washington.edu/kohn/>) and a toxic repertoire of ~100 peptides per venom [38]. Scorpions have been estimated to encode a similar diversity of peptide neurotoxins (~100,000) based on a pharmacological repertoire of ~70 peptides per venom and a species count of ~1500 [44]. However, in evolutionary terms, spiders (class: Arachnida, order: Araneae) are the most successful predatory animals and they maintain by far the largest pool of toxic peptides. There are ~38,000 described species [43], with an even greater number awaiting characterization [5]. Using a conservative estimate of ~50 peptides per venom obtained from mass spectrometric analysis of 55 tarantula venoms [16] and a conservative species count of 80,000 leads to an estimated total of 4 million spider-venom polypeptides.

Remarkably, despite their much larger suite of toxic peptides, spider venoms have been studied much less extensively than venoms from scorpions and cone snails. As far as we are aware, no studies have attempted to examine the underlying genetic mechanism by which spiders generate peptide diversity in their venoms. In this report we review what is known about the molecular mechanisms used by scorpions and cone snails to generate peptide toxin diversity and compare this with new data obtained from Australian funnel-web spiders. The data suggest that spiders, like scorpions and cone snails, have diversified their repertoire of toxic polypeptides by duplication of ancestral toxin gene(s) followed by induction of a hypervariability-generating mechanism that increases the rate of mutation of the mature-toxin loci. The pool of resulting paralogous genes were then subject to adaptive evolution. The chemical outcome of this underlying genetic mechanism is essentially a collection of combinatorial peptide toxin libraries that are conformationally constrained by the presence of conserved, disulfide-forming cysteine residues in the mature toxins. This gene-based combinatorial peptide library strategy appears to have been first implemented by arach-

nids long before marine cone snails independently evolved a similar mechanism for generating peptide diversity.

2. Conotoxin evolution: focal hypermutation and diversifying selection

Cone snails (order: Neogastropoda, genus: *Conus*) are a group of predatory marine gastropods that first appeared following the massive marine extinction at the Cretaceous/Tertiary boundary [54]. The extinction of the previously dominant ammonites, a group of predatory cephalopod molluscs, may have provided the ecological opportunity for subsequent rapid speciation of cone snails. The oldest *Conus* fossils date to the early Eocene ~50 million years ago (Mya). *Conus* has subsequently diversified more rapidly than any other marine gastropod, resulting in >500 extant species [54].

Most cone snails are specialist predators, dining predominantly on worms (vermivorous), fish (piscivorous), or other marine gastropods (molluscivorous) [38]. All *Conus* species use venom for prey capture; the method of envenomation has been outlined by Olivera [37,38]. Venom components are produced in epithelial cells lining a tubular venom duct. When a snail forages for prey, a single hollow tooth with a barbed end like a harpoon is moved into the lumen of the distensible proboscis. Once the proboscis makes contact with prey, the tooth is propelled into the victim and grasped by muscles at the anterior tip of the proboscis. Venom is then injected through the hollow tooth by contraction of a muscular bulb at the posterior end of the venom duct.

The dominant components of cone snail venoms are small peptide toxins typically comprising 12–30 residues and two–three disulfide bonds [54]. Each *Conus* species is thought to express a unique profile of 50–200 conotoxins [40]. The toxins are initially expressed as prepropeptide precursors that are post-translationally processed to yield the mature toxin (Fig. 1A). Each precursor results from transcription of a single genetic locus that encodes a single mature toxin. The canonical precursor architecture comprises an N-terminal signal sequence (~20 residues) followed by a propeptide region of variable length that precedes a single downstream copy of the mature-toxin sequence. This one gene–one toxin paradigm contrasts with the tandem genetic architecture that is employed by arthropods to produce families of endogenous neuropeptides [2] and serine protease inhibitors [50]; in these latter cases the C-terminal region of the precursor often encodes multiple variants of the mature neuropeptide/protease inhibitor that are released as individual units during post-translational processing.

The majority of conotoxins are derived from a small number of gene superfamilies. Each gene superfamily is defined by a highly conserved signal sequence in the prepropeptide precursor and a strictly conserved disulfide framework in the mature toxin. In contrast with the well-conserved signal sequence and cysteine scaffold, there is remarkable intra- and

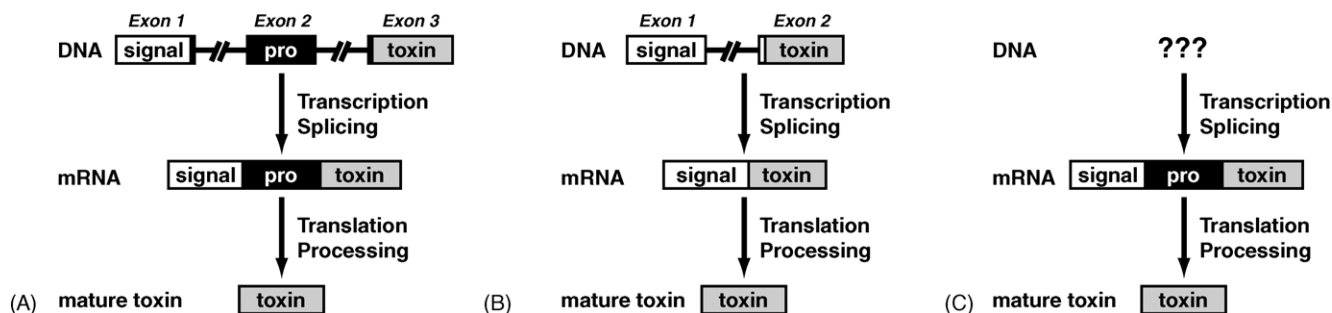


Fig. 1. Cartoon summarizing the mechanism of peptide toxin expression in (A) cone snails, (B) scorpions, and (C) spiders. Signal, propeptide, and mature-toxin regions are indicated by white, black, and gray boxes, respectively. Note that the coding sequence for the signal peptide in scorpion toxins is interrupted by a single intron [44]. The propeptide region in cone snails is largely encoded in exon 2, but a few N- and C-terminal residues are encoded in the 3' end of exon 1 and the 5' end of exon 3, respectively [40]. No DNA sequence information is currently available for genes encoding spider toxins with ICK motifs.

interspecific divergence in the mature-toxin sequences within a gene superfamily [39,40,54]. A comparison of members of the O-superfamily in *Conus abbreviatus* revealed that nearly every amino acid in the mature-toxin sequence, with the exception of the cysteine residues, had been substituted at least once, and generally with nonconservative substitutions [13]. This extraordinary variability in the mature-toxin sequence contrasts with most secreted polypeptides in which the signal sequences are less conserved than other regions of the precursor [40].

Recent studies have provided insight into the underlying molecular mechanism that has generated this remarkable peptide toxin diversity in cone snail venoms [6,7,13,14,17,40]. The initiating event in toxin diversification appears to be the rapid, repeated duplication of toxin-encoding genetic loci [13]. Analysis of complete mRNA transcripts from five different *Conus* species revealed that the 5' and 3' untranslated regions (UTRs) are almost as well conserved as the signal sequence, and much better conserved than both the mature toxin and propeptide regions of the mRNA [6]. There are no reports examining the level of conservation of intronic sequences in conotoxin genetic loci, but certain features are preserved in the single intron of duplicated genetic loci that encode scorpion sodium channel toxins [44]. Thus, the cone snail duplication events most likely involved complete genetic loci including introns and exons. This contrasts with the disulfide-reticulated arthropod defensins that evolved by simple exon shuffling [19].

Gene duplication was followed by focal hypermutation of the mature-toxin sequences. This process was presumably aided by the architecture of the genetic loci in which the signal, pro-, and mature-toxin sequences are encoded by discrete exons that are separated by much longer introns (Fig. 1A) [40]. The abnormally high bias for transversional versus transitional mutations¹ in the mature-toxin region of conotoxin

transcripts is the diagnostic molecular fingerprint that provides evidence for targeted hypermutation of the mature-toxin exon [6]. Remarkably, not only are the cysteine residues preserved, but also there is a position-specific conservation of cysteine codons in the conotoxin precursors [7], even in frameshifted pseudotranscripts that are not under selection pressure [6]. This suggests that cone snails have evolved a positive mechanism for protecting the cysteine codons in the face of the mutational onslaught on surrounding nucleotides in the mature toxin-encoding exon [6,7].

Interestingly, the transversion:transition (T_v/T_s) ratio of ~ 2 observed for the mature-toxin region of conotoxin transcripts is similar to the in vitro transversion bias measured for mutagenic replication by DNA polymerase V, and the low processivity of this polymerase (6–8 nucleotides) is similar to the typical intercysteine spacings in conotoxins [6]. This led Fainzilber and colleagues to propose a hypermutation mechanism in which the cysteine codons recruit a macromolecule that interrupts normal DNA replication and creates a lesion that acts as an attractor for a weakly processive mutagenic polymerase that inserts a mutated tract downstream of the cysteine codon [6]. At this stage, however, the molecular process that causes hypermutation of the mature-toxin exon remains to be determined. Regardless of mechanism, the rate of mutation is extraordinary. Duda and Palumbi calculated that the rate of nonsynonymous substitutions in O-superfamily toxins in *C. abbreviatus* (≥ 17 per site per 10^9 years) is at least five times the highest nonsynonymous rate reported for mammals, and nearly three times the highest rate reported for *Drosophila* [13].

There is strong evidence for positive Darwinian selection, also known as adaptive evolution or diversifying selection, following conotoxin gene duplication and focal hypermutation. Positive selection fixes nonsynonymous substitutions² at a higher rate than synonymous (silent) substitutions, since

¹ A transitional mutation is a substitution in which a purine nucleotide is replaced with another purine, or a pyrimidine is substituted with a pyrimidine (i.e., C \leftrightarrow T, A \leftrightarrow G). A transversional mutation occurs when a purine is replaced with a pyrimidine, or vice versa (i.e., C/T \leftrightarrow A/G).

² Synonymous, or silent, substitutions are nucleotide changes that do not alter the encoded amino acid. In contrast, nonsynonymous substitutions cause either a missense mutation that alters the encoded amino acid or a nonsense mutation that results in production of a truncated protein.

nonsynonymous substitutions lead to amino acid changes that are more likely to be adaptive [61]. The characteristic molecular fingerprint of adaptive evolution, namely a higher proportion of nonsynonymous substitutions (d_N) than synonymous substitutions (d_S) have been reported by several groups for conotoxin multigene superfamilies [6,13,14,40].

As pointed out by Olivera, the initial conotoxin gene duplication and focal hypermutation events can be likened to a generation of a combinatorial peptide library, while the subsequent positive selection is akin to ‘biopanning’ of the library to select useful variants [38]. This combinatorial peptide library approach has enabled cone snails to perform a comprehensive search of pharmacophore space; this is exemplified by the fact that the sequences of the mature-toxin region of paralogous conotoxin genes in extant snails are so divergent that it would be impossible to recognize them as paralogs if the highly conserved signal sequence were absent.

3. Peptide diversity in scorpion venoms: a tale of toxic telsons

Scorpions comprise the order: Scorpiones within the class: Arachnida, which also includes spiders, pseudoscorpions, ticks, and mites [28]. Arachnids differ from insects in lacking antennae and wings, and having eight legs rather than six. Moreover, whereas insects have three distinct body segments, most arachnids have two segments—the prosoma or cephalothorax (combined head and thorax) and the opisthosoma or abdomen. The cephalothorax supports the eight legs and a pair of leg-like pedipalps that are located between the chelicerae and the first pair of legs. Scorpions are readily recognized by the pincer-like terminal segments of the pedipalps, which are used for grasping prey and for defense. In scorpions, the abdomen is subdivided into a broad pre-abdomen (mesosoma) and narrow tail-like post-abdomen (metasoma).

Like spiders, scorpions have a complete reliance on predation as trophic strategy. Venom for prey capture is produced in two glands located in the bulb-shaped telson at the distal end of the metasoma. The prey is usually held by the pincers, and then (if necessary) envenomated with the hollow needle-like stinger (the aculeus) located at the tip of the tail. Venom is delivered voluntarily via two ducts that run from the venom glands to the tip of the stinger, which acts like a hypodermic needle. Venom is used economically; small or non-threatening prey can either be injected with a pre-venom that is rich in K^+ salts but deficient in metabolically expensive peptide toxins [26] or simply crushed with the pedipalps and chelicerae without the need for envenomation [46].

While disulfide-reticulated peptide toxins are the major components of scorpion venoms, these peptides are generally larger than those found in cone snails and spiders. The so-called long-chain toxins typically comprise 60–76 residues with four conserved disulfide bridges and they mostly target voltage-gated Na^+ channels [22,45]. The short-chain toxins, which contain 29–41 residues with three–four disulfide

bridges, are less abundant and act predominantly on K^+ and Cl^- channels [22,45]. It was recently suggested that the short-chain toxins originated from a long-chain precursor gene via a position-specific deletion [4], but further research is required to validate this hypothesis. While many conotoxins and most spider toxins contain an inhibitory cysteine knot (ICK) motif, in which a central disulfide is threaded through a ring of covalently bonded atoms formed by two other disulfides and the intervening sections of polypeptide backbone [9,35], this motif is rare in scorpion toxins [63].

Molecular evolution studies have revealed some similarities in the overall process used by cone snails and scorpions to evolve peptide toxin diversity. Like cone snails, each scorpion toxin superfamily seems to have evolved by duplication of a common ancestral gene followed by positive Darwinian selection [62]. However, there are significant differences at the molecular level. First, the precursors of both short- and long-chain toxins typically comprise an 18–22-residue signal sequence that precedes a single downstream copy of the mature toxin (Fig. 1B) [20]. The conspicuous absence of a pro-region contrasts with the canonical prepropeptide architecture seen in conotoxin precursors (Fig. 1A). Second, the genetic loci for both short- and long-chain toxins have a conserved architecture in which a single intron interrupts the coding sequence for the signal peptide (Fig. 1B) [20]. This differs from cone snails where the signal peptide, propeptide, and mature toxin are essentially encoded by three separate exons that are neatly separated by two larger introns (Fig. 1A). The single intron is poorly conserved between different pharmacological classes of scorpion toxin, but highly similar within each gene superfamily [20].

A third major difference is that the venom of a single scorpion often contains numerous paralogous toxins with very similar amino acid sequences (>80% sequence identity) [20,22]. Simple allelic variation cannot account for this polymorphism. In contrast, cone snails often express two closely related conotoxin sequences that are most likely allelic variants of a single locus [14], but they rarely express more than two paralogs with high sequence identity. Scorpions may not have developed a mechanism for focal hypermutation of toxin loci with the result that paralogous genes evolve more slowly than in cone snails, and thus resemble one another more closely. Dramatically different scorpion toxin families may only arise from rare recombinational events [17] or radical episodic events, such as the proposed position-specific deletion that converted a long-chain toxin to a short-chain toxin [4]. It will be interesting in future studies to examine whether there has been a targeted hypermutation of scorpion toxin loci as evidenced by a transversional mutation bias.

In summary, scorpions have employed gene duplication followed by positive Darwinian selection to evolve venoms that contain optimized combinatorial peptide libraries. However, these peptide libraries do not appear to be as functionally and structurally diverse as the conotoxins, despite their much longer evolutionary history. As discussed below, the maintenance in scorpion venom of a pool of highly

similar paralogous toxins might result from environmental pressures that do not impact significantly on conotoxin evolution.

4. Evolution of spider toxins: a web of intrigue

Spiders evolved from an arachnid ancestor in the Devonian period around 400 Mya and currently comprise over 100,000 different species in more than 3500 genera [5,43]. They produce venom in two glands that are located in each of the two fang bases in mygalomorphs, or just posterior to each of the two chelicerae in araneomorphs. Venom is injected into prey via a duct that runs from each gland to the tip of the proximal fang. Venom is used sparingly by optimizing the bite location and varying the amount of venom injected according to the susceptibility of the prey [31].

The major venom components are typically small peptide toxins of mass 2–9 kDa that can be grouped into three categories [34]: (i) cytolytic peptides, typically of mass 2–4 kDa, with no disulfide bonds; (ii) medium-sized neurotoxic peptides, typically comprising 35–45 residues with three–five disulfide bonds; (iii) larger polypeptides (5–9 kDa), mostly of unknown function, with four or more disulfide bonds. Large polypeptides (>9 kDa) and enzymes are typically absent [28]. However, there are some notable exceptions to this rule: the venom of black widow spiders (*Latrodectus* spp.) contains a high proportion of >100 kDa proteins that are life-threatening to humans due to their ability to stimulate massive neurotransmitter release at vertebrate nerve terminals [56], while sphingomyelinase D is a major component of the venom from recluse spiders (*Loxosceles* spp.) and is responsible for the necrotic arachnidism caused by these arachnids [3].

The medium-sized neuropeptides are the most extensively studied spider toxins and they appear to predominate in most spider venoms [34]. These toxins, almost without exception, contain an ICK motif and their precursors generally conform to the conotoxin prepropeptide paradigm [8,21,57]. The precursor typically comprises an N-terminal signal sequence of 15–25 residues, followed by a propeptide region of highly variable length (15–60 residues), which precedes a single downstream copy of the 35–45-residue mature-toxin sequence (Fig. 1C). The propeptide region of most spider-toxin precursors is rich in acidic residues [12,30,33,41,48,57]. Although they have not been studied in as much detail, the less abundant 5–9 kDa toxins in spider venom, like scorpion toxins, appear to be produced from precursors that lack a pro-region. Examples include the TaITX toxins from *Tegegnaria agrestis* [27] and the MIT-like toxins from Australian funnel-web spiders [58].

The results of two preliminary studies [11,53] indicate that spider toxins, like conotoxins, are probably derived from a small number of gene superfamilies. However, in contrast with scorpion and cone snail toxins, there has been no systematic study of intra- and interspecific divergence within a spider toxin superfamily with a view to understanding the evolution-

ary history of these toxins. Thus, we compared the prepropeptide precursor sequences of 30 paralogs of the insect-specific calcium channel blocker ω -atracotoxin(ACTX)-Hv1a [18] that we obtained from analysis of venom-gland cDNA libraries prepared from three species of Australian funnel-web spider: *Atrax robustus*, *Hadronyche infensa*, and *Hadronyche versuta* (see legend to Fig. 2 for experimental details).

Logo analysis [10,49] of the 30 atracotoxin precursor sequences (Fig. 2A) reveals a striking similarity with conotoxin superfamilies: the signal sequence is highly conserved, whereas there is extensive sequence variation within both the propeptide and mature-toxin regions. The only strongly conserved portion of the propeptide region is the Arg–Arg sequence that immediately precedes the mature-toxin sequence at positions 41–42. This dibasic site is better conserved than the signal peptide cleavage site at positions 22/23 and more strongly preserved than the propeptide cleavage site in conotoxins [6,7,13], which implies that there has been strong selection pressure to preserve this protease recognition site. Within the mature-toxin sequence, there is enormous sequence variation although it is striking that the cysteine residues that form the disulfide bridges of the ICK motif are strictly conserved. Interestingly, there are a small number of non-cysteine positions that are reasonably well conserved, and some of these correspond to functionally important residues in the ω -ACTX-1 toxins as determined previously using alanine scanning mutagenesis [52].

There is striking codon bias at four of the six cysteine positions in the mature-toxin region of the ω -ACTX-1 precursor sequences (Fig. 2B). These six cysteines form the ICK structural scaffold in the mature toxin [18,29]. In the absence of some underlying mechanism that induces global codon bias, one might expect there to be no preference for either of the two possible cysteine codons (TGC and TGT). However, we find 100% use of TGT at cysteine positions I, IV, and VI and ~95% at cysteine III (Fig. 2B). Intriguingly, the positions with codon bias correspond to the I–IV and III–VI disulfide bridges in this family of toxins. There is no codon bias for the other two cysteines which form the II–V disulfide bridge. The observed position-specific codon bias is not simply a manifestation of global codon bias in these spiders as we have observed a strong preference for TGC as opposed to TGT for cysteine residues in other atracotoxin families (data not shown).

We constructed a maximum likelihood tree for the 30 atracotoxin nucleotide sequences and used this tree to estimate the ratio of nonsynonymous/synonymous substitutions (d_N/d_S) for the signal peptide ($d_N/d_S = 0.34$), propeptide ($d_N/d_S = 1.22$), and mature-toxin ($d_N/d_S = 0.65$) regions separately (see legend to Fig. 3 for details). Only in case of the signal peptide is the d_N/d_S ratio significantly different from unity ($P < 0.001$, calculated using a maximum likelihood ratio test); the fact that $d_N/d_S \ll 1$ indicates that the signal peptide is under strong purifying selection (i.e., there is strong selection against deleterious mutations). We cannot make any firm conclusions about the propeptide region as the analyses

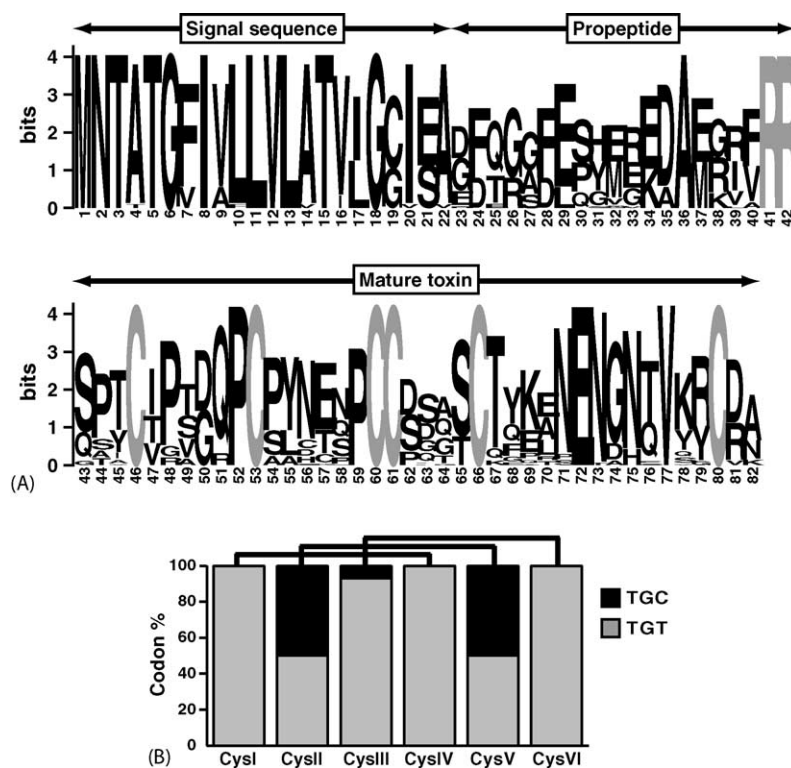


Fig. 2. Analysis of ω -ACTX-1 paralogs. (A) Sequence logo [10,49] from alignment of prepropeptide precursor sequences for 30 ω -ACTX-1 paralogs from three species of Australian funnel-web spider (*Atrax robustus*, *Hadronyche versuta*, and *Hadronyche infensa*). Sequences were aligned with ClustalW [55], and then optimized manually. The signal sequence is much better conserved than the propeptide and mature-toxin regions. However, note the strict conservation of the cysteine residues in the mature-toxin sequence and the Arg–Arg protease recognition site that terminates the propeptide region (both highlighted in gray). Construction and RACE analyses of venom-gland cDNA libraries were performed essentially as described previously [57]. (B) Codon usage for the six cysteine residues in the mature toxin. Note the strong bias for TGT at positions I, III, IV, and VI. Shown above the histogram is the disulfide bridge arrangement for the six cysteines as inferred from the 3D structure of ω -ACTX-Hv1a [18].

were hampered by the short stretch of available sequences and by the many gaps that had to be inserted to produce an alignment.

The overall d_N/d_S ratio for the mature-toxin region does not reveal significant purifying or diversifying selection. However, spider venoms, like those of scorpions, usually contain numerous paralogous toxins with very similar amino acid sequences (>80% sequence identity) that presumably have arisen from lineage-specific duplication events (e.g. note the clusters of closely related atracotoxin paralogs in Fig. 3). If one considers only the branches that lead to the most populated clusters of related paralogs, then d_N/d_S for the mature toxin is significantly larger than one (see Fig. 3), suggesting that speciation was associated with adaptive evolution of the mature-toxin sequence.

In summary, the mechanism employed by spiders to generate peptide toxin diversity appears to be at least superficially similar to that employed by cone snails. Extensive gene duplication was probably followed by focal hypermutation of the propeptide and mature-toxin regions. At present, insufficient spider toxin precursor sequences are available to unequivocally determine whether there is a bias for transversional substitutions in the propeptide and mature-toxin regions of the precursor. However, the extensive intra- and interspecific

sequence variation within these regions (Fig. 2A), combined with the position-specific codon bias observed for the cysteine residues in the mature toxin (Fig. 2B), hints at a hypermutator mechanism as proposed for cone snails [6,7]. Finally, as indicated by the analysis of atracotoxin precursor sequences, speciation appears to be associated with positive Darwinian selection from the pool of hypermutated toxin sequences.

One intriguing difference between the arachnids and cone snails is that spiders and scorpions usually maintain a small pool of closely related paralogs in their venom (e.g. see Fig. 3), suggestive of lineage-specific gene duplication events, whereas the paralogous toxins in cone snails tend to be highly divergent. This may be a consequence of different predatory habits. While cone snails are typically specialist predators, spiders and scorpions are mostly generalists that prey on a wide range of invertebrates and, sometimes, small vertebrates. Successful predation by most cone snails may require only a single toxin for each specific ligand- and voltage-activated ion channel within the chosen prey phylum. In contrast, there may have been selection pressure in arachnids to maintain a suite of different, but functionally homologous, toxins capable of targeting a specific receptor across a wide range of invertebrate and vertebrate phyla. The

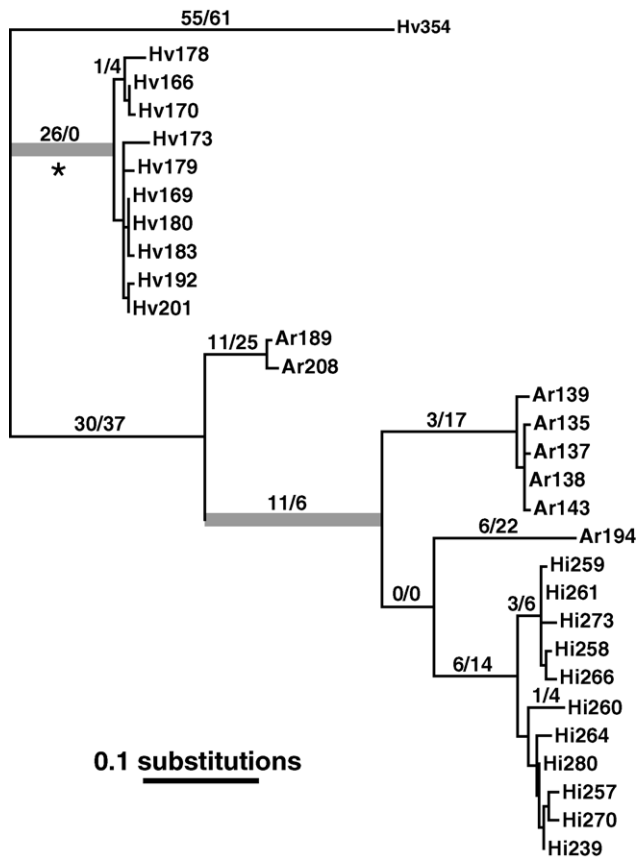


Fig. 3. Maximum likelihood phylogeny calculated from alignment of the nucleotide sequences of atracotoxin precursors using the HKY85 model [24] with four discrete categories for among site rate variation [59] and invariant sites; 10 independent replicates were performed. The tree was calculated using PAUP* Version 4.0 beta10 [51]. The nucleotide alignment was calculated from the amino acid sequence alignment using the tranalign program in the EMBOSS package [47]. Toxin precursors are identified by initials indicating the genus and species of spider (Hv, Ar, or Hi) followed by clone number. The tree topology was used to estimate the nonsynonymous/synonymous substitution ratio (d_N/d_S) for the signal peptide, propeptide, and mature-toxin regions separately; the number of nonsynonymous/synonymous substitutions for the mature-toxin region is indicated above each branch. The analyses were performed using PAML version 3.14 [60]. A d_N/d_S ratio >1 was inferred for the mature-toxin region for the two branches indicated by thick gray lines. A maximum likelihood ratio test (with a significance level of 0.025) indicated that the branch marked with an asterisk is under positive selection. The scale bar indicates the number of substitutions per site along each branch of the tree.

intrinsic variation in receptor architecture between phyla is further amplified by the high degree to which ligand- and voltage-activated ion channels appear subject to alternative splicing [23] and RNA editing [25], making this a fascinating ‘moving target’ problem for arachnids.

5. Were arachnids the first combinatorial peptide chemists?

There is little doubt that spiders and scorpions, which originated in the Palaeozoic, developed genetic strategies for

producing combinatorial peptide libraries long before cone snails, which first appeared in the Eocene over 300 million years later. Although spiders are grouped with scorpions in the class Arachnida, their evolutionary relationship remains uncertain. The first scorpions were at least partially aquatic and these arachnids were amongst the earliest terrestrial colonizers, first appearing on land in the late Silurian ~ 420 Mya [42]. However, there is little evidence for derivation of other arachnids from the scorpion clade, making it likely that spiders and scorpions diverged before independent terrestrializations by their aquatic ancestors in the late Silurian/early Devonian [15]. The evolution of mechanisms for generating combinatorial peptide libraries in scorpion and spider venoms may therefore have been independent adaptations to terrestrial life. This hypothesis is supported by several other considerations. First, spider and scorpion toxins are generally dissimilar, whether one considers their precursor architecture, protein/DNA sequence, three-dimensional structure or molecular target. Second, these arachnids produce venoms and evenomate their prey using morphologically disparate apparatuses (the posteriorly located telson/aculeus in scorpions and the anteriorly located chelicerae/fangs in spiders).

It appears, therefore, that spiders, scorpions, and cone snails independently evolved genetic mechanisms for generating combinatorial peptide libraries in their venoms. While these genetic mechanisms appear to be superficially similar, much further study will be required to determine whether they are comparable at the molecular level. A molecular understanding of the mechanisms used by these venomous creatures to generate structurally constrained combinatorial peptide libraries might have implications for the design of pharmacophore-rich peptide libraries.

Acknowledgements

This work was supported by grants from the National Science Foundation to G.F.K. (MCB9983242) and J.P.G. (MCB0237197) and a grant from the National Institute of Allergy and Infectious Diseases to G.F.K. (1-R41-AI51791).

References

- [1] Becerril B, Marangoni S, Possani LD. Toxins and genes isolated from scorpions of the genus *Tityus*. *Toxicon* 1997;35:821–35.
- [2] Bendena WG, Donly BC, Tobe SS. Allatostatins: a growing family of neuropeptides with structural and functional diversity. *Ann NY Acad Sci* 1999;897:311–29.
- [3] Binford GJ, Wells MA. The phylogenetic distribution of sphingomyelinase D activity in venoms of Haploglyne spiders. *Comp Biochem Physiol B* 2003;135:25–33.
- [4] Céard B, Martin-Eauclaire M, Bougis PE. Evidence for a position-specific deletion as an evolutionary link between long- and short-chain scorpion toxins. *FEBS Lett* 2001;494:246–8.
- [5] Coddington JA, Levi HW. Systematics and evolution of spiders (Araneae). *Annu Rev Ecol Syst* 1991;22:565–92.

- [6] Conticello SG, Gilad Y, Avidan N, Ben-Asher E, Levy Z, Fainzilber M. Mechanisms for evolving hypervariability: the case of conopeptides. *Mol Biol Evol* 2001;18:120–31.
- [7] Conticello SG, Pilpel Y, Glusman G, Fainzilber M. Position-specific codon conservation in hypervariable gene families. *Trends Genet* 2000;16:57–9.
- [8] Corzo G, Gilles N, Satake H, Villegas E, Dai L, Nakajima T, et al. Distinct primary structures of the major peptide toxins from the venom of the spider *Macrothele gigas* that bind to sites 3 and 4 in the sodium channel. *FEBS Lett* 2003;547:43–50.
- [9] Craik DJ, Daly NL, Waine C. The cystine knot motif in toxins and implications for drug design. *Toxicon* 2001;39:43–60.
- [10] Crooks GE, Hon G, Chandonia JM, Brenner SE. WebLogo: a sequence logo generator. *Genome Res* 2004;14:1188–90.
- [11] Diao J, Lin Y, Tang J, Liang S. cDNA sequence analysis of seven peptide toxins from the spider *Selenocosmia huwena*. *Toxicon* 2003;42:715–23.
- [12] Diniz MR, Paine MJ, Diniz CR, Theakston RD, Crampton JM. Sequence of the cDNA coding for the lethal neurotoxin Tx1 from the Brazilian “armed” spider *Phoneutria nigriventer* predicts the synthesis and processing of a preprotoxin. *J Biol Chem* 1993;268:15340–2.
- [13] Duda Jr TF, Palumbi SR. Molecular genetics of ecological diversification: duplication and rapid evolution of toxin genes of the venomous gastropod *Conus*. *Proc Natl Acad Sci USA* 1999;96:6820–3.
- [14] Duda Jr TF, Palumbi SR. Evolutionary diversification of multigene families: allelic selection of toxins in predatory cone snails. *Mol Biol Evol* 2000;17:1286–93.
- [15] Dunlop JA, Webster M. Fossil evidence, terrestriality and arachnid phylogeny. *J Arachnol* 1999;27:86–93.
- [16] Escoubas P, Rash L. Tarantulas: eight-legged pharmacists and combinatorial chemists. *Toxicon* 2004;43:555–74.
- [17] Espiritu DJ, Watkins M, Dia-Monje V, Cartie rGE, Cruz LJ, Olivera BM. Venomous cone snails: molecular phylogeny and the generation of toxin diversity. *Toxicon* 2001;39:1899–916.
- [18] Fletcher JI, Smith R, O’Donoghue SI, Nilges M, Connor M, Howden MEH, et al. The structure of a novel insecticidal neurotoxin, ω -atracotoxin-HV1, from the venom of an Australian funnel web spider. *Nat Struct Biol* 1997;4:559–66.
- [19] Froy O, Gurevitz M. Arthropod and mollusk defensins—evolution by exon-shuffling. *Trends Genet* 2003;19:684–7.
- [20] Froy O, Sagiv T, Poreh M, Urbach D, Zilberberg N, Gurevitz M. Dynamic diversification from a putative common ancestor of scorpion toxins affecting sodium, potassium, and chloride channels. *J Mol Evol* 1999;48:187–96.
- [21] Gomez MV, Kalapothakis E, Guatimosim C, Prado MAM. *Phoneutria nigriventer* venom: a cocktail of toxins that affect ion channels. *Cell Mol Neurobiol* 2002;22:579–88.
- [22] Goudet C, Chi CW, Tytgat J. An overview of toxins and genes from the venom of the Asian scorpion *Buthus martensi* Karsch. *Toxicon* 2002;40:1239–58.
- [23] Grabowski PJ, Black DL. Alternative RNA splicing in the nervous system. *Prog Neurobiol* 2001;65:289–308.
- [24] Hasegawa M, Kishino H, Yano T. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol* 1985;22:160–74.
- [25] Hoopengardner B, Bhalla T, Staber C, Reenan R. Nervous system targets of RNA editing identified by comparative genomics. *Science* 2003;301:832–6.
- [26] Inceoglu B, Lango J, Jing J, Chen L, Doymaz F, Pessah IN, et al. One scorpion, two venoms: prevenom of *Parabuthus transvaalicus* acts as an alternative type of venom with distinct mechanism of action. *Proc Natl Acad Sci USA* 2003;100:922–7.
- [27] Johnson JH, Bloomquist JR, Krapcho KJ, Kral RM, Trovalto R, Eppler KG, et al. Novel insecticidal peptides from *Teegenaria agrestis* spider venom may have a direct effect on the insect central nervous system. *Arch Insect Biochem Physiol* 1998;38:19–31.
- [28] King GF. The wonderful world of spiders: preface to the special Toxicon issue on spider venoms. *Toxicon* 2004;43:471–6.
- [29] King GF, Tedford HW, Maggio F. Structure and function of insecticidal neurotoxins from Australian funnel-web spiders. *J Toxicol-Toxin Rev* 2002;21:359–89.
- [30] Krapcho KJ, Kral Jr RM, VanWagenen BC, Eppler KG, Morgan TK. Characterization and cloning of insecticidal peptides from the primitive weaving spider *Diguetia canities*. *Insect Biochem Mol Biol* 1995;25:991–1000.
- [31] Kuhn-Nentwig L, Schaller J, Nentwig W. Biochemistry, toxicology and ecology of the venom of the spider *Cupiennius salei* (Ctenidae). *Toxicon* 2004;43:543–53.
- [32] Lewis RJ, Garcia ML. Therapeutic potential of venom peptides. *Nat Rev Drug Discov* 2003;2:790–802.
- [33] Liang S. An overview of peptide toxins from the venom of the Chinese bird spider *Selenocosmia huwena* Wang [*Ornithoctonus huwena* (Wang)]. *Toxicon* 2004;43:575–85.
- [34] Maggio F, Sollod BL, Tedford HW, King GF. Spider toxins and their potential for insect control. In: Gilbert, L.I., Iatrou, K., Gill, S.S., editors. *Comprehensive molecular insect science*, vol. 5. Oxford: Elsevier; 2004.
- [35] Norton RS, Pallaghy PK. The cystine knot structure of ion channel toxins and related polypeptides. *Toxicon* 1998;36:1573–83.
- [36] Ohno M, Menez R, Ogawa T, Danse JM, Shimohigashi Y, Fromen C, et al. 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;59:307–64.
- [37] Olivera BMEE. Just Lecture, 1996. *Conus* venom peptides, receptor and ion channel targets, and drug design: 50 million years of neuropharmacology. *Mol Biol Cell* 1997;8:101–9.
- [38] Olivera BM. *Conus* venom peptides: reflections from the biology of clades and species. *Annu Rev Ecol Syst* 2002;33:25–47.
- [39] Olivera BM, Hillyard DR, Marsh M, Yoshikami D. Combinatorial peptide libraries in drug design: lessons from venomous cone snails. *Trends Biotechnol* 1995;13:422–6.
- [40] Olivera BM, Walker C, Cartier GE, Hooper D, Santos AD, Schoenfeld R, et al. Speciation of cone snails and interspecific hyperdivergence of their venom peptides. Potential evolutionary significance of introns. *Ann NY Acad Sci* 1999;870:223–37.
- [41] Ostrow KL, Mammoser A, Suchyna T, Sachs F, Oswald R, Kubo S, et al. cDNA sequence and in vitro folding of GsMTx4, a specific peptide inhibitor of mechanosensitive channels. *Toxicon* 2003;42:263–74.
- [42] Pisani D, Poling LL, Lyons-Weiler M, Hedges SB. The colonization of land by animals: molecular phylogeny and divergence times among arthropods. *BMC Biol* 2004;2:1.
- [43] Platnick NI. *Advances in spider taxonomy, 1992–1995: with re-descriptions 1940–1980*. New York: New York Entomological Society & The American Museum of Natural History, 1997 (updated version available at <http://research.amnh.org/entomology/spiders/catalog>).
- [44] Possani LD, Becerril B, Delepierre M, Tytgat J. Scorpion toxins specific for Na⁺-channels. *Eur J Biochem* 1999;264:287–300.
- [45] Possani LD, Merino E, Corona M, Bolivar F, Becerril B. Peptides and genes coding for scorpion toxins that affect ion-channels. *Biochimie* 2000;82:861–8.
- [46] Rein JO. Sting use in two species of *Parabuthus* scorpions (Buthidae). *J Arachnol* 1993;21:60–3.
- [47] Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet* 2000;16:276–7.
- [48] Santos AD, Imperial JS, Chaudhary T, Beavis RC, Chait BT, Hunsperger JP, et al. Heterodimeric structure of the spider toxin ω -agatoxin IA revealed by precursor analysis and mass spectrometry. *J Biol Chem* 1992;267:20701–5.
- [49] Schneider TD, Stephens RM. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res* 1990;18:6097–100.

- [50] Simonet G, Claeys I, Franssens V, De Loof A, Broeck JV. Genomics, evolution and biological functions of the pacifastin peptide family: a conserved serine protease inhibitor family in arthropods. *Peptides* 2003;24:1633–44.
- [51] Swofford D, 1998. PAUP*: phylogenetic analysis using parsimony (and other methods). Sinauer Associates, Sunderland.
- [52] Tedford HW, Gilles N, Ménez A, Doering CJ, Zamponi GW, King GF. Scanning mutagenesis of ω -atracotoxin-Hv1a reveals a spatially restricted epitope that confers selective activity against invertebrate calcium channels. *J Biol Chem* 2004;279:44133–40.
- [53] Tedford HW, Sollod BL, Maggio F, King GF. Australian funnel-web spiders: master insecticide chemists. *Toxicon* 2004;43:601–18.
- [54] Terlau H, Olivera BM. *Conus* venoms: a rich source of novel ion channel-targeted peptides. *Physiol Rev* 2004;84:41–68.
- [55] Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994;22:4673–80.
- [56] Ushkaryov YA, Volynski KE, Ashton AC. The multiple actions of black widow spider toxins and their selective use in neurosecretion studies. *Toxicon* 2004;43:527–42.
- [57] Wang X-H, Connor M, Wilson D, Wilson HI, Nicholson GM, Smith R, et al. Discovery and structure of a potent and highly specific blocker of insect calcium channels. *J Biol Chem* 2001;276:40806–12.
- [58] Wilson DTR, 2001. The identification and characterization of Australian funnel-web spider venom. Ph.D Thesis. Institute for Molecular Biosciences, University of Queensland, Brisbane.
- [59] Yang Z. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J Mol Evol* 1994;39:306–14.
- [60] Yang Z. PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* 1997;13:555–6.
- [61] Yang Z. Inference of selection from multiple species alignments. *Curr Opin Genet Dev* 2002;12:688–94.
- [62] Zhu S, Bosmans F, Tytgat J. Adaptive evolution of scorpion sodium channel toxins. *J Mol Evol* 2004;58:145–53.
- [63] Zhu S, Darbon H, Dyason K, Verdonck F, Tytgat J. Evolutionary origin of inhibitor cystine knot peptides. *FASEB J* 2003;17:1765–7.