Structural and Functional Diversity of Peptide Toxins from Tarantula *Haplopelma hainanum* (*Ornithoctonus hainana*) Venom Revealed by Transcriptomic, Peptidomic, and Patch Clamp Approaches*

Received for publication, December 29, 2014, and in revised form, February 24, 2015. Published, JBC Papers in Press, March 13, 2015, DOI 10.1074/jbc.M114.635458

Yi-Ya Zhang‡1, Yong Huang‡1, Quan-Ze He‡, Ji Luo‡, Li Zhu‡, Shan-Shan Lu‡, Jin-Yan Liu‡, Peng-Fei Huang‡, Xiong-Zhi Zeng‡2, and Song-Ping Liang‡3

From the ‡1Key Laboratory of Protein Chemistry and Developmental Biology of Ministry of Education, College of Life Sciences, Hunan Normal University, Changsha 410081, China and the ‡2State Key Laboratory of Pathogen and Biosecurity, Beijing Institute of Microbiology and Epidemiology, 20 Dong-Da Street, Fengtai District, Beijing 100071, China

**Background:** Spider venom is a complex mixture and contains over 400 venom peptides/species.

**Results:** Highly abundant hypermutation, fragment insertion/deletion, and variable post-translational modifications were observed in venom gland, and the highly diverse toxins exhibit diverse functions.

**Conclusion:** Spider toxins are abundant and variable, factors that work together to generate diverse bioactivities.

**Significance:** Our study reveals the mechanism of structural and functional diversity of peptide toxins from *Ornithoctonus hainana*.

Spider venom is a complex mixture of bioactive peptides to subdue their prey. Early estimates suggested that over 400 venom peptides are produced per species. In order to investigate the mechanisms responsible for this impressive diversity, transcriptomics based on second generation high throughput sequencing was combined with peptidomic assays to characterize the venom of the tarantula *Haplopelma hainanum*. The genes expressed in the venom glands were identified, and the bioactivity of their protein products was analyzed using the patch clamp technique. A total of 1,136 potential toxin precursors were identified that clustered into 90 toxin groups, of which 72 were novel. The toxin peptides clustered into 20 cysteine scaffolds that included between 4 and 12 cysteines, and 14 of these groups were newly identified in this spider. Highly abundant toxin peptide transcripts were present and resulted from hypermutation and/or fragment insertion/deletion. In combination with variable post-translational modifications, this genetic variability explained how a limited set of genes can generate hundreds of toxin peptides in venom glands. Furthermore, the intraspecies venom variability illustrated the dynamic nature of spider venom and revealed how complex components work together to generate diverse bioactivities that facilitate adaptation to changing environments, types of prey, and milking regimes in captivity.

Spiders are some of the most ancient venomous animals and have been roaming the earth for 300 million years, numbering nearly 40,000 species. Tarantulas are a group of often hairy and large arachnids that comprise more than 860 species. Like other venomous animals, these predators use a complex arsenal of venom to paralyze and kill their prey, and many of these toxins have proved to be invaluable tools for pharmacological studies of voltage-sensitive and ligand-gated ion channels (1–4). The vast majority of spider toxins are cysteine-rich polypeptides, and their properties and structures have been reviewed in detail (4–7). The molecular diversity of spider toxins has also been investigated, and they appear to be based on a limited set of structural scaffolds, such as the inhibitor cysteine knot (ICK) and disulfide-directed β-hairpin, whose cysteines form 3–5 characteristic disulfide bonds, which is a much smaller number than toxins from other species, such as marine cone snails. The lack of a complete genome sequence may be responsible for our currently limited knowledge of the cysteine pattern diversity present in tarantula toxins, and improving this knowledge will be challenging.

Expressed sequence tags (ESTs) are short sequence reads derived from cDNA libraries that are useful tools for the identification of transcripts in species without a fully sequenced genome (8, 9). In the past few years, all reports on tarantula toxin transcriptomics have utilized classical cloning and Sanger sequencing strategies (9). In previous work, 420 peptides were detected by mass spectrometry, but few could be paired with peptide precursors identified from cDNA and genomic DNA sequencing (10). Furthermore, this limited data focused mainly on highly abundant and smaller toxin precursors, whereas less

---

*This work was supported by National Basic Research Program of China (973) Grant 2010CB529801 and by Cooperative Innovation Center of Engineering and New Products for Developmental Biology of Hunan Province Grant 20134486.

1 Both authors contributed equally to this work.

2 To whom correspondence may be addressed. Tel.: 86-731-88872556; Fax: 86 731 88861304; E-mail: xiongzhizeng@gmail.com.

3 To whom correspondence may be addressed. Tel.: 86-731-88872556; Fax: 86 731-88861304; E-mail: liangsp@hunnu.edu.cn.

4 The abbreviations used are: ICK, inhibitor cysteine knot; TTX, tetrodotoxin; TTX-S, tetrodotoxin-sensitive; DRG, dorsal root ganglion; EST, expressed sequence tag; contig, group of overlapping clones; GO, gene ontology; VGSC, voltage-gated sodium channel.
prevalent and longer gene sequences were largely ignored, which has proved a barrier to research on the molecular diversity and genetic mechanisms of toxin evolution (11). The relatively new 454 Life Sciences pyrosequencing technology has been successfully implemented in a number of species, including spiders (12–16). This approach provides a more comprehensive landscape of the transcriptomic content of venom glands and has improved technical capabilities that identify longer sequences, a wider range of sensitivity, and greater accuracy than traditional Sanger sequencing (17, 18). The longer reads (>300 bp on average) generated with 454 pyrosequencing allow coverage of full-length (60–120 amino acids) precursors, which affords the direct identification of spider toxin precursors and avoids the errors inherent in assembling reads into contigs as typically required for other second generation technologies that generate shorter read lengths (19). In this study, 1,136 toxin precursors were identified using 454 Life Sciences pyrosequencing and classified into 15 classes (class A–O). Sequence analysis revealed that extensive hypermutation and fragment insertion/deletion dramatically increased the molecular diversity of toxin transcripts, and this diversity was further enhanced by highly variable post-translational modifications.

The high intraspecies variability may explain the transcriptome differences between this work and previous work. Analysis of the bioactivity of 14 of the identified toxins against DRG neurons and were ground to fine powder in liquid nitrogen. Total RNA was extracted with TRIzol milked by electrical stimulation and were ground to fine pow-

**Experimental Procedures**

**Materials**—Dithiothreitol (DTT), iodoacetamide, and trifluoroacetic acid (TFA) were obtained from Sigma. Acetonitrile was a domestic product (chromatogram grade). Milli-Q H₂O was used for the preparation of all buffers. Other chemicals were of analytical grade.

cDNA Library Construction and 454 Sequencing—The tarantula spiders were collected from Hainan, China. Venom glands of *Haplopelma hainanum* were obtained 2 days after being milked by electrical stimulation and were ground to fine powder in liquid nitrogen. Total RNA was extracted with TRIzol (Invitrogen) and used to construct a cDNA library. 5 μg of full-length double-stranded cDNA was then processed by the standard genome sequencer library preparation method using the 454 DNA library preparation kit (titantium chemistry) to generate single-stranded DNA ready for emulsion PCR (emPCR™). The cDNA library was sequenced according to GS FLX technology (454/Roche Applied Science).

Sequence Assembly and Alignment—Sequence reads were trimmed by excluding adapters and low quality regions using the NGen module of the DNAStar Lasergene software suite. Subsequently, assembly was performed with SeqMan Pro (DNASTAR) using high stringency *de novo* transcriptome assembly (100% identity between reads with 50-nucleotide sequence overlap). Similar reads were assembled into contigs using CLC Genomics Workbench 3 with its default parameters. All assembled contigs and rest reads were searched against the NCBI subset of the EST database with BLASTx. All of these sequences were also searched with BLASTx in all six reading frames to the non-redundant protein database to determine their correct translation products. For both searches, Blast version 2.2.25+ was used, and an e value threshold of e = 10⁻³ with a bit score of >40 was considered and recorded as a significant match for each query sequence. We analyzed the Blast results and used a homemade PERL script to classify representative sequences into five categories ("toxin-like," "putative toxins," "cellular proteins," "unknown function," and "no hit").

**Gene Ontology (GO) Annotations**—Functional characteristics of the transcriptome were predicted using BLAST2GO software (20) with the NCBI non-redundant protein database (cut-off e value of ≤10⁻⁵) using contigs. Each contig with GI accession (NCBI) of the significant hits retrieved was assigned GO terms according to molecular function, biological process, and cellular component ontologies at a level that provided the most abundant category numbers (21).

**Toxin Identification and Evolutionary Analyses**—The identification of toxin peptide sequences was carried out from the raw data using tBLASTx, and a signal peptide can be predicted with the SignalP version 3.0 program (available from the SignalP server). The propeptide cleavage site was ascertained from the known start site of previously characterized mature toxins. As mentioned previously, such long sequence reads are likely to contain the full nucleic sequences of toxin precursors. The toxin-identifying sequences, the sequences representing no hits, and those with an abundance of cysteine residues may encode new toxin peptides. Toxin precursors were selected out according to the four following parameters: the proteins came from a full open reading frame (ORF) translated by Geneious software (22); the proteins contained more than 4 Cys residues (23); the proteins contained more than 45 amino acids; the toxin precursors were clustered into toxin groups according to their sequence similarity. All precursor sequences were aligned using ClustalX. The resulting alignment was imported into MEGA software to construct phylogenetic trees by the neighbor-joining method (24), and bootstrap values were estimated from 500 replicates.

**Electrophysiological Assays**—Toxins were purified and identified as described previously (10). Rat DRG neurons were acutely dissociated from 30-day-old Sprague-Dawley rats and maintained in short term primary culture according to the method described by Xiao and Liang (25). Ionic currents were recorded from DRG cells under whole-cell patch clamp techniques using an Axon 700B patch clamp amplifier (Axon Instruments, Irvine, CA) at room temperature (20–25°C). Patch pipettes with DC resistance of 2–3 megaohms were fabricated from borosilicate glass tubing (VWR micropipettes, VWR Co., West Chester, PA) using a two-stage vertical microelectrode puller (PC-10, Narishige, Tokyo, Japan) and fire-polished by a heater (Narishige). Experimental data were acquired and analyzed using the programs Clampfit version 10.0 (Axon Instruments) and Sigmaplot version 9.0 (Sigma). For Na⁺ current recording, the pipette solution contained 145 mM CsCl, 4 mM MgCl₂, 10 mM HEPES, 10 mM EGTA, 10 glucose mM, 2 mM ATP (PH 7.2), and the external solution contained 145 mM NaCl, 2.5 mM KCl, 1.5 mM CaCl₂, 2 mM MgCl₂, 10 mM HEPES,
Molecular Diversification of Peptide Toxins from the Tarantula H. hainanum

10 mM d-glucose (pH 7.4). For K+ current recording, the pipette solution contained 135 mM KCl, 25 mM KF, 9 mM NaCl, 0.1 mM CaCl2, 1 mM MgCl2, 1 mM EGTA, 10 mM HEPES, and 3 mM ATP-Na2, adjusted to pH 7.4 with 1 M KOH, and the external bath solution contained 150 mM NaCl, 30 mM KCl, 5 mM CaCl2, 4 mM MgCl2, 0.3 mM tetrodotoxin (TTX), 10 mM HEPES, and 10 mM d-glucose, adjusted to pH 7.4 with 1 M NaOH. For Ca2+ current recording, the internal solution contained 110 mM Cs-methane sulfonate, 14 mM phosphocreatine, 10 mM HEPES, 10 mM EGTA, 5 mM ATP-Mg, adjusted to pH 7.3 with CsOH, and the external solution contained 10 mM BaCl2, 125 mM tetraethylammonium chloride, 0.3 mM TTX, and 10 mM HEPES, adjusted to pH 7.4 with tetroethylammonium hydroxide.

Results

454 Sequencing Statistics and Transcriptome Assembly—The mRNAs of six venom glands from the tarantula Ornithoctonus hainana were extracted and sequenced using GS FLX technology (454/Roche Applied Science) following the manufacturer’s protocol. Sequencing revealed a total of 249,549 reads (amounting to ~757 Mb) with an average length of ~328 bases/read (maximum 830 bp, minimum 40 bp, 6.72% of reads <100 bp, 84.6% of reads 100–500 bp, 8.63% of reads >500 bp). The raw sequencing data can be downloaded from SRA (NCBI) using accession number SRP040123. After removing sequences of low quality, a total of 215,640 reads were assembled into 65,432 contiguous DNA sequences (contigs) with an average length of 625 bp (36.3 reads/contig), with the rest remaining as single reads. Whereas this study focused mainly on toxin peptides, numerous other protein sequences were identified and will be described elsewhere. As outlined under “Experimental Procedures,” we searched for toxin peptide sequences directly from the sequencing reads, because the average read length of >200 bp allowed the identification of full-length toxin precursors. Toxin peptides were also searched for in the contigs, and no additional toxin peptide sequences were found. In total, 52,570 reads displayed similarity to known toxin peptides or toxin-like sequences; the category of putative toxins includes sequences rich in cysteine residues and sharing sequence identity with toxins or proteins including the ICK motif (5%) that were not identified by a BLAST search; the category of cellular proteins includes transcripts coding for proteins involved in cellular processes (44%); the unknown function category includes reads that shared sequence identity with previously described sequences with no functional assessment or hypothetical genes; and the no hit category indicates no match with currently known sequences. The results are summarized in Fig. 1.

Functional Annotation—A search against publicly available databases (nr/NCBI, Swiss-Prot + TREMBL/EMBL) revealed that 8,773 high confidence proteins were associated with GO terms and further grouped into the categories of molecular functions, biological processes, and cellular components at the second level according to standard gene ontology terms. Based on annotations from GO analysis (Fig. 2), transcripts were categorized into 2,610 groups with 80% identity threshold. These included functional annotations for 816 biological process (BP), 798 molecular function (MF), and 996 cellular component (CC) categories (Fig. 2). Highly expressed transcripts were enriched in metabolism and translation processes, indicating that venom glands are metabolically active and engaging in the intensive protein synthesis and processing required for venom production. Transcripts associated with binding, catalysis, and channel regulation were also highly represented, indicating that venom peptides may play an important role in prey inactivation by binding to ion channels. GO terms related to redox homeostasis and proteolysis were also enriched, which may be related to the extensive post-translational modification of spider toxins that was reported previously (9).

Spider Toxin Transcript Identification and Classification—Although spider toxins contain diverse disulfide bridge patterns and fold into a variety of three-dimensional structures, cysteine-rich domains are a common feature shared by many toxin sequences (26, 27). Overall, 1,136 toxin precursors were identified in the venom gland transcriptome, and 65.8% of the mature peptides included two adjacent cysteine residues. From previous proteomics results (10), we were able to predict precursor endoproteolytic and amidation sites with high confidence, and 90 toxins and variants accounted for 95% of all toxin precursors. These were divided into 14 classes (classes A–N) based on the number of cysteine residues present (Table 1), and 70 sequences did not fall into these categories in class O (Fig. 3). Of the 18 H. hainanum peptide toxins characterized previously, 11 belonged to class G along with 252 variants, four belonged to class E along with 39 variants, and the other three belonged to classes C, L, and M along with 19, 25, and 4 variants, respectively.

Class A—Class A contained seven novel transcripts, including HN-Aa and its variants (Fig. 4). The signal peptide and propeptide of these precursors are identical, and their cleavage sites are LFA/ED and LESEK, respectively. The mature peptide has 12 Cys residues, which is the highest number in a toxin peptide scaffold in H. hainanum reported to date, although two other spider toxins share this cysteine pattern (28, 29). TX-L precursors from Lycosa singoriensis superfAMILY VI have the cysteines arranged C-CC-C-C-C-C-C-C-C-C-C-C, whereas HN-Aa has the double cysteine shifted to the left. The double knot toxin from the earth tiger tarantula (Selenocosmia huwena) has the pattern C-C-CC-C-C-CC-C-C-C-C, and this sequence shares 87% similarity with HN-Aa. Double knot toxin selec-
tively and irreversibly activates the capsaicin- and heat-sensitive channel TRPV1 (29). This cysteine pattern therefore indicates that HN-Aa may be a TRPV1 channel activator. The six variants identified showed high similarity with the mature peptide, and the presence of different lengths of mature peptide results in great potential for toxin variability.

Class B—Class B contains only four members, the fewest of all classes (Fig. 4). HN-Ba and its three variants include 11 Cys residues, and HN-Ba precursors contain over 100 residues, which includes the consensus VIAYA cleavage signal, whereas the propeptide numbered only four residues. A large number of Gln residues were found in the mature peptide, and the C terminus is rich in Thr and Ser residues, which is strikingly different from other toxins present in this species. The cysteine pattern shares homology with the U1-hexatoxin-Hsp201a from the funnel web spider (30). Other than this highly conserved cysteine pattern, HN-Ba showed no obvious sequence similarity with other peptide toxins.

Class C—Class C contains 29 transcripts that clustered into the known toxin HN-Ca (HNTX-XIV) and two novel toxins (HN-Cb and HN-Cc; Fig. 4), all of which included 10 Cys residues, but class C toxins were further divided into two clades based on cysteine pattern. HN-Ca and HN-Cb shared the pattern -C-C-CC-C-C-C-C-C-C- and formed the first clade, and these toxins do not have a propeptide but do share a signal peptide containing the cleavage site ELVSC. The mature HN-Cb peptide is rich in Val residues and has a C terminus that is ~20 residues longer than HN-Ca (HNTX-XIV), indicating diverse functions. Further research is needed to investigate whether these additional residues are cleaved during post-translational modification or remain and have some additional functional significance. HN-Cc exhibited the cysteine pattern -C-C-C-CC-C-C-C-C-C-, and the signal peptide, propeptide, and mature peptide were distinct from HN-C[a~b], although HN-Cc peptides shared sequence similarity and included an extended cysteine variant not present in HNTX-XVIII.

Class D—The four toxin precursors in class D contain 9 cysteine residues and exhibited two different cysteine patterns (Fig. 4). HN-Da shares significant similarity with HNTX-IV (31) but has the cysteine pattern -C-C-CC-C-C-C-C- with an additional 3 cysteines at the C-terminal region that also included a particularly large number of Ser residues. HN-Db has the cysteine pattern -C-C-CC-C-C-C-C-C- and shared an identical signal peptide and propeptide with HNTX-XVIII, but the mature peptide was different, in particular due to the additional cysteine. The additional C-terminal residues could be removed to produce a more canonical spider toxin during maturation.

Class E—The mature peptides of class E displayed two distinct cysteine patterns, and the precursors were classified into nine groups (Fig. 4). HN-E[a~c, d, e~i] shared the cysteine pattern -C-C-CC-C-C-C-, and the other two toxins shared the cysteine pattern -C-C-CC-C-C-C- (Table 1). HN-E[a~c] share high sequence similarity with members of groups HNTX-I, HNTX-III, and HNTX-IV, respectively, and include a 30-residue C-terminal extension. Interestingly, the CIC motif was present in the majority of these sequences, which is typical of μ-AGTXs from the American funnel web spider (Agelesopsis aperta) and PnTx3-3 from the Brazilian armed spider (32–35). We speculated that the cysteines in this motif are not involved in disulfide bond formation because PnTx3-3 only has two disulfide bonds, with the fourth and fifth Cys residues excluded (32). Members of HN-Ed share sequence similarity with HNTX-VII, with the propeptide lost and a longer mature pep-
Molecular Diversification of Peptide Toxins from the Tarantula H. hainanum

TABLE 1
The main features of H. hainanum venom gland toxins

| Class | cysteine | precursors | Name          | similar sequences                                      | scaffold(s)                  |
|-------|----------|------------|---------------|--------------------------------------------------------|-----------------------------|
| A     | 12 (CC)  | 7          | HN-Aa         | double-knot toxin                                       | C-C-CC-C-C-CC-C-CC-C-C [1]  |
| B     | 11 (CC)  | 4          | HN-Ba         | U34-TRTX                                               | C-C-CC-C-CC-C-CC-C-CC-C-CC-C [2] |
| C     | 10 (CC)  | 29         | HN-C[a−c]     | HNTX-XIV, HNTX-XVIII                                   | C-C-CC-C-CC-C-CC-C-CC-C-CC-C-C [3], C-C-CC-C-CC-C-CC-C-C [4] |
| D     | 9 (CC)   | 4          | HN-D[a−b]     | HNTX-XVIII, HNTX-IV                                    | C-C-CC-C-CC-C-CC-C-CC-C-C [5], C-C-CC-C-CC-C-CC-C-C [6] |
| E     | 8 (CC)   | 84         | HN-E[a−i]     | HNTX-I, HNTX-III, HNTX-IV, HNTX-XII, HNTX-XV, HNTX-XVII, HNTX-XVIII, HWTX-XV, | C-C-CC-C-CC-C-CC-C-CC-C-C [7], C-C-CC-C-CC-C-C [8] |
| F     | 7 (CC)   | 150        | HN-F[a−I]     | U1-TRTX-Pc1a, HNTX-I, HNTX-III, HNTX-IV, HNTX-VIII, HNTX-XIV, HNTX-XV, HNTX-XVIII, HNTX-XVIII, HWTX-XVIII, | C-C-CC-C-CC-C-CC-C-CC-C-C [9], C-C-CC-C-C-CC-C-C [10] |
| G     | 6 (CC)   | 397        | HN-G[a−y]     | HNTX-XIII, HNTX-XIX, HNTX-XV, HNTX-XVI, HNTX-XVIII, HWTX-Xa, HWTX-XIII, HWTX-XVI, | C-C-CC-C-CC-C [11], C-C-CC-C-CC-C [12] |
| H     | 5 (CC)   | 2          | HN-Ha         | non                                                    | CC-CC-C [13]                |
| I     | 4 (CC)   | 28         | HN-I[a−f]     | HWTX-XV, HNTX-III                                      | C-C-CC [14], C-C-CC-C [15]  |
| J     | 8        | 8          | HN-Ja         | HNTX-II                                               | C-C-CC-C-CC-C-CC-C-CC-C-C [16] |
| K     | 7        | 6          | HN-Ka         | HNTX-II                                               | C-C-CC-C-CC-C-CC-C [17]     |
| L     | 6        | 94         | HN-L[a−i]     | HNTX-II, HWTX-VII                                     | C-C-CC-C-CC-C [18]          |
| M     | 5        | 132        | HN-M[a−h]     | HNTX-II, HNTX-VIII, HNTX-XI, HWTX-VII                 | C-C-CC-C-CC-C [19]          |
| N     | 4        | 134        | HN-N[a−i]     | HNTX-II, HNTX-VII, HNTX-VIII, HWTX-VII                | C-C-CC-C [20]               |

tide. Toxin precursors lacking propeptides are commonly observed in scorpions but rarely in spiders. Four known toxins, HN-Ef (HNTX-XVII), HN-Eg (HNTX-XVIII), HN-Ee (HNTX-XV), and HN-Eh (HNTX-XX), have been chemically and functionally characterized in our previous work (10). HN-Ef (HNTX-XVII) and HN-Eh (HNTX-XX) exhibited the shortest mature peptides in this class. The typical motif C\(_{v}\)X\(_{1}\)C\(_{v}\)X\(_{2}\)C\(_{v}\)X\(_{3}\)C\(_{v}\)X\(_{4}\)C\(_{v}\)X\(_{5}\)C\(_{v}\)X\(_{6}\)X\(_{7}\), where X\(_{i}\) is any residue except cysteine, was present in precursors and was first identified in Mu-agatoxins from A. aperta (32). Numerous variants differed by only one or two amino acids, suggesting that these may be ideal for studying structure-function relationships. Moreover, the HN-Ei mature peptide is identical with HWTX-XV (Table 2). Although interesting, it is not unprecedented for toxins from different venomous animals to share identical sequences.

Class F—Class F includes 15 toxins (149 precursors) and is divided into three clades based on the seven-cysteine pattern present in these peptides (Fig. 4). Ten (HN-F[a−g, k, n−o]) share the -C-C-CC-C-CC-C- pattern, and four (HN-F[f−j, m]) exhibited the -C-C-CC-C-CC-C- pattern, whereas only HN-Fl had the pattern -C-C-CC-C-CC-C-CC-C-. The pattern in HN-Fa is shared with that in U1-theraphotoxin-Pc1a, a toxin that possesses strong in vitro anti-plasmodial activity against the intra-erythrocyte stage of Plasmodium falciparum (33, 36). HN-F[b−c] share high homology with HNTX-I and HNTX-III, including the conservative sequence NEINACSPVF in the C-terminal region. HN-Fd shares high sequence identity with HNTX-IV but includes the much longer conserved C-terminal sequence “MRSMYPVQFSNMYLANGGIMSSTSSACQL-MSINK” that is enriched with Ser. Although HN-Fe has the same signal peptide as HNTX-VIII, the C terminus is much longer. The mature peptide of HN-Ff shares minimal sequence identity with HNTX-XIV, and the key Cys residues are changed into other amino acids. HN-F[g−j] shares high sequence similarity with HNTX-XV and HNTX-XVIII but cannot form four disulfide bonds because there is a cysteine missing in the C-terminal region. We speculated that despite the high identity with known toxins, these peptides must display a novel cysteine pat-
Molecular Diversification of Peptide Toxins from the Tarantula H. hainanum

Patterns included a signal–peptide cleavage site similar to the CYASE sequence.

Classes H and I—Although most of the known spider toxins contain more than six cysteine residues, the members of classes H and I may include four to five cysteines. Class H included five transcripts with five cysteines in a distinctive -CC-CC-C-pattern comprising two contiguous Cys pairs (Fig. 5). This cysteine pattern is novel, and homology between HN-Ha and other toxins is very low. HN-Ha groups only accounted for about 0.2% of the toxin transcripts and are therefore unlikely to play a key role in the venom function of H. hainanum.

All of the 28 transcripts were classified into six toxin groups according to sequence homology. Three toxins, accounting for a quarter of the class I sequences, did not generate hits in a BLAST search (Fig. 8). HN-Id shares homology with HWTX-XV from O. huwena, which is a TTX-sensitive voltage-gated sodium channel (VGSC) inhibitor. The signal peptide and propeptide of HN-Id are highly variable, but HN-Id has a highly conserved mature peptide and shares the -C-CC-C- cysteine arrangement with HN-ic. HN-[a–b] also share this pattern but also did not generate hits in a BLAST search. HN-1a differs from HNTX-III, having a C-terminal mutation and lacking Cys-5 and Cys-6. The first and second Cys are replaced by other residues in the mature peptide of HN-If.

Class J—Class J contains a set of eight homologous transcripts (HN-Ja) with the cysteine pattern X_2X_3X_4X_5X_6X_7X_8X_9, where X is any amino acid, and n is an undefined number (Fig. 6). In contrast to some other classes, these peptides do not contain the double-cysteine (-CC-) motif, and this cysteine pattern is consistent with that of LSTX-M1 and ω-agatoxin-1A (28, 37), in which 10 Cys residues form four intrachain disulfide bonds and one interchain disulfide bond. In contrast, HN-Ja did not show sequence similarity with LSTX-M1 or ω-agatoxin-1A but is similar to HNTX-II. Class K contains a set of six transcripts that share the cysteine pattern -C-CC-C-CC-C-C-. The C-terminal region is elongated compared with HNTX-II family peptides and includes an additional Cys residue (Fig. 9). The signal peptide and propeptide were the same for all members of these classes, but the mature peptides were variable in C-terminal regions.

Class L—Class L comprises 100 transcripts that share the cysteine pattern -C-CC-C-CC-C- (Fig. 6). HN-La (HNTX-II) was the most significant component in this class. The signal peptide and propeptide of HN-[b–d] were identical with that of HNTX-II, and the mature peptides were also very similar, and all included the conserved NHHDKIRNKRV sequence motif in the C-terminal region. HN-Le shares high sequence similarity with HWTX-VII, but the third Cys in the mature peptide is replaced by Leu, and the conservative sequence VLKCR is added to the C-terminal region. HN-Lf differs from HNTX-III, having a C-terminal mutation and lacking Cys-5 and Cys-6. The first and second Cys are replaced by other residues in the mature peptide of HN-Lf.

Classes M and N—Mature peptides of class M shared the cysteine pattern -C-CC-C-C-C-, which is characteristic of

-tern. HN-F[k–m] share sequence similarity with toxins from Ornithoctonus huwena, and HN-Fk shares sequence similarity with HWTX-VIII, including a high proportion of basic residues in the mature peptides. HN-Fm shares similarity with HWTX-XVIII, and HN-Fl also includes a large number of basic residues, but no homologs were identified from a BLAST search; therefore, this appears to be a novel toxin.

Class G—Class G includes the greatest number of peptides, accounting for 36% of all toxins, and 10 of these (HN-Ga (HNTX-I), HN-Gc (HNTX-III), HN-Ge (HNTX-IV), HN-Gg (HNTX-IX), HN-Gj (HNTX-VII), HN-Gm (HNTX-XIII), HN-Gn (HNTX-X), HN-Gp (HNTX-XIII), HN-Gs (HNTX-XIII), HN-Gu (HNTX-XVI), and HN-Go (HNTX-XII)) have been chemically and functionally characterized in our previous work. These 10 toxins contain 6 cysteine residues forming three disulfide bonds. Based on sequence similarity, class G peptides were further divided into 25 distinct toxin groups (Fig. 5). HN-G[b, d, f, i, l, q] are similar to the known toxins HNTX-I, HNTX-III, HNTX-IV, HNTX-IX, HNTX-IX, HNTX-VII, and HNTX-XII but with a longer C-terminal region. The three disulfides have a 1–4, 2–5, and 3–6 connectivity that forms the distinctive ICK motif. The majority of members in this class are variants that differ by only one or two amino acids, implying that they are natural mutants. In contrast to the traditional ICK motif-containing spider toxins, HN-Gt and HN-Gv are similar to HNTX-XV and HNTX-XVIII and adopt a -C-CC-C-C- cysteine pattern. Toxins HN-G[w–y] share high sequence similarity with HWTX-Xa, HWTX-XIII, and HWTX-XVI, respectively, whereas HN-Gw, HN-Gx, and HN-Gy differ from HWTX-Xa, HWTX-XIII, and HWTX-XVI, respectively by only two amino acid residues. The signal peptides and propeptides of these toxins included a signal–peptide cleavage site similar to the CYASE sequence.

-classes H and I—Although most of the known spider toxins contain more than six cysteine residues, the members of classes H and I may include four to five cysteines. Class H included five transcripts with five cysteines in a distinctive -CC-CC-C-pattern comprising two contiguous Cys pairs (Fig. 5). This cysteine pattern is novel, and homology between HN-Ha and other toxins is very low. HN-Ha groups only accounted for about 0.2% of the toxin transcripts and are therefore unlikely to play a key role in the venom function of H. hainanum.

All of the 28 transcripts were classified into six toxin groups according to sequence homology. Three toxins, accounting for a quarter of the class I sequences, did not generate hits in a BLAST search (Fig. 8). HN-Id shares homology with HWTX-XV from O. huwena, which is a TTX-sensitive voltage-gated sodium channel (VGSC) inhibitor. The signal peptide and propeptide of HN-Id are highly variable, but HN-Id has a highly conserved mature peptide and shares the -C-CC-C- cysteine arrangement with HN-ic. HN-[a–b] also share this pattern but also did not generate hits in a BLAST search. HN-1a differs from HNTX-III, having a C-terminal mutation and lacking Cys-5 and Cys-6. The first and second Cys are replaced by other residues in the mature peptide of HN-If.

Class J—Class J contains a set of eight homologous transcripts (HN-Ja) with the cysteine pattern X_2X_3X_4X_5X_6X_7X_8X_9, where X is any amino acid, and n is an undefined number (Fig. 6). In contrast to some other classes, these peptides do not contain the double-cysteine (-CC-) motif, and this cysteine pattern is consistent with that of LSTX-M1 and ω-agatoxin-1A (28, 37), in which 10 Cys residues form four intrachain disulfide bonds and one interchain disulfide bond. In contrast, HN-Ja did not show sequence similarity with LSTX-M1 or ω-agatoxin-1A but is similar to HNTX-II. Class K contains a set of six transcripts that share the cysteine pattern -C-CC-C-CC-C-. The C-terminal region is elongated compared with HNTX-II family peptides and includes an additional Cys residue (Fig. 9). The signal peptide and propeptide were the same for all members of these classes, but the mature peptides were variable in C-terminal regions.

Class L—Class L comprises 100 transcripts that share the cysteine pattern -C-CC-C-CC-C- (Fig. 6). HN-La (HNTX-II) was the most significant component in this class. The signal peptide and propeptide of HN-[b–d] were identical with that of HNTX-II, and the mature peptides were also very similar, and all included the conserved NHHDKIRNKRV sequence motif in the C-terminal region. HN-Le shares high sequence similarity with HWTX-VII, but the third Cys in the mature peptide is replaced by Leu, and the conservative sequence VLKCR is added to the C-terminal region. HN-Lf differs from HNTX-III, having a C-terminal mutation and lacking Cys-5 and Cys-6. The first and second Cys are replaced by other residues in the mature peptide of HN-Lf.

Classes M and N—Mature peptides of class M shared the cysteine pattern -C-CC-C-C-C-, which is characteristic of...
Kunitz-type sequences and is found in HWTX-XI and HWTX-VII from *O. hainana* (38). HN-M[a~b] share high sequence similarity with HNTX-II, with the third Cys replaced by Asp, Leu, or Ile in HN-Ma. Similarly, the fifth cysteine in mature HN-Mb peptides is replaced by other residues, and there are several additional Cys residues in the C-terminal region. HN-M[c, h] share sequence homology with HNTX-VIII, with a double cysteine missing from the cysteine pattern (Fig. 10), whereas HN-Md shares high sequence similarity with HNTX-XI, except the HN-Md3 variant has an extended C-terminal region enriched with basic residues. The three toxins HN-M[e~g] were novel and generated no hits in a BLAST search (Fig. 6).

Class N toxins exhibited a -C-C-C-C- cysteine pattern, and HN-N[a~h] shared some similarity with HNTX-II, huwentoxin II, and HWTX-VII. HN-Ni exhibited no obvious sequence similarity with any known sequences (Fig. 6).

**Toxin Abundance among Classes**—The abundance of toxin ESTs is reflective of venom composition and provides information on their diversity and evolution. Major differences in transcription levels between different toxin groups and between toxin classes were investigated (Fig. 7). Unsurprisingly, most venom peptides were expressed at relatively high levels and were present in a variety of forms at the transcriptional level. This suggests that the more abundant toxin transcripts are mirrored by a greater number of peptide toxins. Moreover, gene classes containing larger numbers of precursors also tended to produce the highest number of total reads. Overall, Class G produced the highest number of reads and the largest number of precursors, while classes E, F, L, M and N were of intermediate abundance (Fig. 7A). Analysis of the data showed that the 30 most abundant toxin peptides accounted for 84.7% of the total toxin transcripts (Fig. 7B). HN-Gc (HNTX-III) had the highest number of transcripts, which were translated into 82 different HN-Gc variants. A BLAST search revealed that HN-Gc shares considerable sequence identity with 20 toxins from spiders and cone snails (Fig. 8), indicating that HN-Gc may be an ancestral toxin from which others evolved. Most toxin precursors included a number of variants that enriches toxin diversity, although there exist only three mutations between HN-Ee, HN-Md and HN-Gn, suggesting that spiders could further improve venom selectively through greater sequence diversification.

**Variation in Toxin Precursor Sequence and Expression**—Classes A–N included multiple variants for each toxin sequence, due to random mutations in both prepro and mature
toxin regions. This led to broad distributions in the chromatographic fractions due to the close molecular weights of the variants and is reflected in the diversity of structures and physicochemical properties. Rapid gene duplication accompanied by focal hypermutation of residues encoded in the mature peptide sequence, and the high conservation of cysteine residues due to high codon bias can explain the diversity of the variants (10, 11, 39). Alignment of the amino acid sequences of HN-Ge and 18 homologous peptides identified several mutations (Fig. 9). All variants except for HN-Ge18 and Ge-15 share the same pro-

peptide sequence, and some sequences are highly similar. HN-Gb and HN-Gd have an extended C-terminal region compared with HN-Gc (Fig. 5). Alteration of the termination codon, point mutations, and insertions and deletions and shuffling of sections of oligonucleotide sequence are responsible for the diversification of peptides (Fig. 10) and are common to the transcriptome of both snakes and spiders (40–42). Mutations appear to occur randomly and display evident sequence diversity without disrupting the ICK scaffold. Unsurprisingly, 321 precursor variants resulted from 18 known toxins, but only 23

### Table 2

| Amino acid sequences | aa | Peptide toxin | species | Reference |
|----------------------|----|--------------|---------|-----------|
| CIGEGVPCDENDPRCCSGVLCLKPTLHGIWYKSYYYCYKK | 39 | HWTX-XVIa1 | Haploperla schmidi | (38) |
| HNTX-XVI | Haploperla hainanum | (10) |
| CIGEGVPRDENDPRCCSGVLCLKPTLHGIWYKSYYYCYKK | 39 | HWTX-XVIa5 | Haploperla schmidi | (38) |
| HNTX-XVI-5 | Haploperla hainanum | (10) |
| CIGEGVPCDENDPRCCSGVLCLKPALHGIWYKSYYYCYKK | 39 | HWTX-XVIa10 | Haploperla schmidi | (38) |
| HNTX-XVI-16 | Haploperla hainanum | (10) |
| ECRWYLGGCSQGDCCKHLQCHSNVEWCVWDGDFS | 36 | HWTX-V | Haploperla schmidt | (57–59) |
| HNTX-IX-2 | Haploperla hainanum | (10) |
| DCAGYMRECKKLCCSGYVCSSRWKWCVLPAWP | 33 | HWTX-III | Haploperla schmidt | (53) |
| HNTX-VIII | Haploperla hainanum | (10) |
| SACKPGEMCMRIKCDGQCQGNCNRTGCRCFK | 31 | HWTX-XVIB1 | Haploperla schmidt | (38) |
| HNTX-XX | Haploperla hainanum | (10) |
| IFECVSCDIEKEKPKCKPKGKEKKCSGKWCKIKLCLKI | 39 | TXP1 | Brachypelma smithi | (60) |
| ESTx2 | Aphonopelma californicum | (61) |
| YCQKWMWTCDDEERKCEGELVCRWLCKRIJNM | 31 | MTX2 | Grammostola rosea | (62) |
| PCTX2 | Paraphysa scrofa | (63) |
| CGGWMACDSDCCETFHCTRNFVCGK | 28 | HWTX | Haploperla schmidt | (43) |
| CM3-42.8 | JZTX F3-11.69 | Chilobrachys guangxiensis | (64) |
| ADCVGDGQRCDWAPYCCSYYCCRSMPYRCRSDS | 38 | CT-II | Hololema curta | (65) |
| Mu-uGatoxin-3 | Agelenopsis aperta | (66) |
| EYCIPPRRCVTTEQCKPYDTVNNFAACGMAWPDDKKRKNKCYICD | 57 | HN-Ei4 | Haploperla hainanum | This work |
| NELTLCTR | HWTX-XV | Haploperla schmidt | (38) |
| GCFKEGKWCPSAPCCAPLCKGPSIKQQKCVRE | 35 | HN-Gw2 | Haploperla hainanum | This work |
| HWTX-Xa | Haploperla schmidt | (38) |
| KCKPNGQKCESPTECCSMACSSDKSCEVEHTLHFG | 38 | HN-Gx4 | Haploperla hainanum | This work |
| HWTX-XIII | Haploperla schmidt | (38) |
| KCKPNGQKCESPTECCSMACSSDKSCEVEHTLHFG | 38 | HN-Gx4 | Haploperla hainanum | This work |
| HWTX-XIII | Haploperla schmidt | (38) |
| CNGDVPDPDPKANRRCCGLECLKPYLHIGIYQDYCYVEKSGR | HN-Gy3 | Haploperla hainanum | This work |
| CNGDVPDPDPKANRRCCGLECLKPYLHIGIYQDYCYVEKSGR | HWTX-XVI | Haploperla schmidt | (38) |
Molecular Diversification of Peptide Toxins from the Tarantula H. hainanum

FIGURE 5. Alignment of *O. hainana* venom gland toxin precursors from classes G–I. Signal peptides and propeptides are shown in framed boxes. Identical residues are shaded. Homology with other spider toxins is shown with black dots.

FIGURE 6. Alignment of *O. hainana* venom gland toxin precursors from classes J–N. Signal peptides and propeptides are shown in framed boxes. Identical residues are shaded. Homology with other spider toxins is shown with black dots.
of these overlapped. The significance of the molecular evolution process is 2-fold; most residues mutate randomly, which generates numerous variants, but at the same time, the molecular mechanisms of transcription preserve cysteine residues, resulting in a high conservation of the molecular scaffold (43). Although a high number of toxin variants were present in the venom gland transcriptome of H. hainanum, many of the gene products may not be functionally active but may play a role in the evolution of future toxin variants. Comparison of proteomic and transcriptomic data revealed that only 15 fully sequenced and 9 partially sequenced venom peptides were identified using the 454 transcriptome approach (Fig. 11), and the vast majority of transcripts coding for cysteine knot toxins did not appear as translated peptides in the venom proteome. The combination of a targeted mutagenic mechanism to generate high variability with the subsequent action of diversifying selection on highly expressed variants might explain the hyper-variability (44). Most transcripts are perhaps unlikely to be translated, apart from critically important components. This mechanism might prevent the venom gland from going to the expense of producing toxins unnecessarily, which makes the process more systematic and efficient (45). Studying toxin diversity is of importance for the discovery of novel toxins and will help to explain toxin evolution.

Modification and Processing of Multiple Peptide Products—
The variety of toxin variants in spider venom results partly from the diversity of toxin gene products and partly from post-translational modifications, sequence homologues, and protein degradation (38, 41, 46, 47). For example, F1-29.54-4123 and F1-30.82-4112.8 share the same precursor, but several residues are modified in the C-terminal region of the mature peptides (Fig. 11), and amidation provides further diversity. A single glycine is a common amidation signal, as is Gly followed by a single or dibasic endoproteolytic site at the C terminus (48). Amidation signals were identified at the C terminus of three toxins (G-K in HN-Ga, -Gc, and -Ge). During post-translational processing, precursors are processed to yield mature truncated toxins. A single peptide (K) or tripeptide (RMD) was removed from the C terminus of the deduced peptides by the precursor-processing enzyme in HNTX-IX and HNTX-Go (HNTX-IX-2). F3-30.36-3538.0 and F3-25.85-3351.9 are truncated forms of HN-Gm, with a dipeptide (RR) or tripeptide (WRR) deleted from the C terminus, whereas F6-25.12-3998.8 and F3-24.71-4057.6 are truncated forms of HN-Go with a tetrapeptide (GRMD) or tripeptide (RMD) deleted from the C terminus. We speculate that precursors were alternatively spliced and modified during posttranslational processing, as was observed in venoms from O. huwena (41). HNTX-IX shares the same amino acid sequence with HNTX-IX-2, but the molecular mass is higher by 10 Da, and preliminary sequence analysis showed that this may be due to an unknown modification, as observed in
O. huwena (49). Despite the very low expression levels of most of these variants, they contributed a great deal to the overall diversity of spider toxins.

Effect of Functional Diversity on Rat DRG Neurons—To explore the functions of these novel spider toxins, we investigated the activities of 14 representative toxins on Nav (voltage-activated sodium), Kv (voltage-activated potassium), and Cav (voltage-activated calcium) channels in rat DRG neurons using the whole-cell patch clamp technique. In total, 11 six-cysteine-containing toxins were chosen from class G. HN-Ef (HNTX-XVII) has four disulfide bonds, HN-La (F8_17.06) includes the DHH motif, and HN-O38 has three cysteine residues that can only form one disulfide. All 14 tested toxins show very weak inhibition or lack inhibition on voltage-gated potassium channels in rat DRG neurons. The Figure 9 and Figure 10 illustrate the alignment of HN-Ge group toxin precursors and the comparison of cDNA and protein sequences of HN-Gc and HN-Gd. A C-terminal extension in HN-Gd caused by a cytosine insertion in the cDNA sequence leads to the delayed termination of the coding region. The Figure 11 and Figure 12 display the full and partial sequences in the O. hainana venom transcriptome and proteome.
nels, although they exhibit variable inhibitory effects on sodium or calcium channels. Some toxins with high sequence similarity exhibited similar bioactivity against the ion channels (Fig. 12). HN-Gc (HNTX-III), HN-Ge (HNTX-IV), and HNTX-V all contain the ICK motif, and all inhibited TTX-sensitive (TTX-S) VGSCs by >70%. However, HN-Ga (HNTX-I) also shares this motif but did not affect TTX-S VGSCs. The hydrophilic Asn-19 and basic His-27 are conserved among the first three toxins but are replaced by Gly and acidic Asp in HN-Ga, which may be responsible for the different channel inhibition activity. This result indicates that the mutation of key residues could profoundly affect toxin activity. HN-Gu (HN-XVI) showed high inhibition against TTX-resistant VGSCs, and this toxin has an ICK motif, but there are 14 residues between the fourth and fifth cysteines, which is much more than in other ICK-containing toxins. Moreover, toxins with diverse cysteine patterns could have similar functions. For example, at a concentration of 100 nM, HN-Ge (HNTX-IV) inhibited calcium channels by 50% and TTX-S sodium channels by 30%. Similarly, HN-O38 has only three cysteine residues, whereas F8_17.06 has a complete disulfide-directed β-hairpin motif, and both of these toxins have a cysteine pattern distinct from...
that of ICK motif toxins, yet all three inhibited calcium currents by >30%.

In summary, the activity of spider toxins on ion channels was highly variable. Toxins with conserved sequence patterns can share similar functions or behave very differently. Toxins exhibiting low activity against DRG ion channels may be important for inactivating prey, but further functional studies are needed.

**Discussion**

Spider toxins form a huge reservoir of chromosomally encoded short peptides that exhibit remarkable structural diversity. Many of the more highly abundant toxins have been studied, but the mechanisms of toxin diversity and evolution are far from being well understood. In this work, a high throughput 454 pyrosequencing approach was deployed
Molecular Diversification of Peptide Toxins from the Tarantula H. hainanum

because it offers the high throughput, accuracy, and relatively long reads (>300 bp on average) required to cover the full length of spider toxin precursors (60–120 amino acids). A total of 1,136 toxin precursors were detected that clustered into 90 toxin peptides, 18 of which have been previously reported (Fig. 13), 44 of which were novel but shared sequence similarity with previously reported H. hainanum toxins, 14 of which shared sequence similarity with known toxins from other spiders, and 14 of which were novel and did not retrieve any hits following a BLAST search (Fig. 14). A total of 20 different cysteine scaffolds were present that contained 4–12 Cys residues. Six of these cysteine scaffolds have been presented in previous work, but 14 were novel. The cysteine pattern -C-C-CC-C-C- that folds into the highly stable ICK motif was the most abundant. This toxin scaffold is found in a wide range of bioactive peptides in spiders, snakes, or other venomous animals (4, 50–54), suggesting that this structure evolved in an ancient venomous ancestor.

At the transcriptional level, many toxin variants were identified that arose from a limited set of genes through hypermutation and fragment insertion/deletion. Although the mutations occurred randomly, in general, sequences diverged, but the common cysteine scaffold was conserved. Moreover, alternative cleavage sites, heterogeneous post-translational modifications, and highly variable N- and C-terminal truncations further increase peptide diversity (10).

The abundance of different toxin transcripts was highly variable and was reflected in the composition of venom peptides, and this is important for understanding of toxin diversity and evolution. In general, there was a positive correlation between mRNA levels and peptide abundance. Class G contained both the highest number of transcripts and the largest number of peptide precursors. This important toxin class may target channels conserved in various types of prey animal. Toxins of intermediate abundance also may also be related to hunting. A surprisingly large number of transcript sequences were expressed at relatively low levels, many of which differed from more abundant peptides by single amino acid changes, deletions, and frame and stop codon shifts. Some toxin variants were generated from alternative cleavage sites, interrupted or elongated cysteine patterns. The toxins in low abundance might be considered as the track of toxin evolution; namely, natural selection might drive the spider to retain abundant fractions of the venom, which are applied by the spider in prey capture and/or defense.

Only 18 of the identified toxin genes have been described in previous work. We hypothesize that toxin variants arose by random mutations and focal hypermutations from rapid gene duplication, and post-translational modifications provide further structural diversity. These processes allow the venom transcriptome to rapidly adapt to changes in the environment or type of prey (55, 56). It has been suggested that it is the cumulative effect of individual spider toxins that is responsible for the lethality, and different toxin groups exist in different spiders.

Analysis of toxin activity on DRG ion channels showed that some of the toxins with similar sequence patterns generally exhibited similar functions; however, with some exceptions, such toxins alternatively displayed distinct functions. Moreover, some of the peptides with more divergent sequences can also share similar functions if key residues are conserved. The diversity of venom toxins results in many peptides working together to improve the efficiency by which prey are incapacitated. In conclusion, a combination of transcriptomics, peptidomics, and electrophysiology revealed the impressive diversity of spider toxin peptides at the transcriptional and peptide structural and functional levels.

Acknowledgments—We thank Dr. Zhonghua Liu and Ying Wang for assistance with manuscript writing.

References

1. Binford, G. J., Bodner, M. R., Cordes, M. H., Baldwin, K. L., Rynerson, M. R., Burns, S. N., and Zobel-Thropp, P. A. (2009) Molecular evolution, functional variation, and proposed nomenclature of the gene family that includes spingomyelinase D in sciarid spider venoms. Mol. Biol. Evol. 26, 547–566
2. Liang, S. (2008) Proteome and peptidome profiling of spider venoms. Expert Rev. Proteomics 5, 731–746
3. Grishin, E. (1999) Polypeptide neurotoxins from spider venoms. Eur. J. Biochem. 264, 276–280
4. Liang, S. (2004) An overview of polypeptide neurotoxins from flower spider venoms. Eur. J. Pharmacol. 52, 794–806
5. Rash, L. D., and Hodgson, W. C. (2002) Pharmacology and biochemistry of spider venom. Toxicon 40, 225–254
6. Corzo, G., and Escoubas, P. (2003) Pharmacologically active spider peptide toxins. Cell Mol. Life Sci. 60, 2409–2426
7. Rash, L. D., and Hodgson, W. C. (2002) Pharmacology and biochemistry of spider venoms. Toxicon 40, 225–254
8. Fernandes-Pedrosa Mde, F., Junqueira-de-Azevedo Ide, L., Goncalves-de-Andrade, R. M., Kobashi, L. S., Almeida, D. D., Ho, P. L., and Tambourgi, D. V. (2008) Transcriptome analysis of Loxosceles laeta (Araneae, Sicariidae) spider venomous gland using expressed sequence tags. BMC Genomics 9, 279
9. Chen, J., Zhao, L., Jiang, L., Meng, E., Zhang, Y., Xiong, X., and Liang, S. (2008) Transcriptome analysis revealed novel venom components and cellular processes of the tarantula Chilobrachys jingzhao venom gland. Toxicon 52, 794–806
10. Tang, X., Zhang, Y., Hu, W., Xu, D., Tao, H., Yang, X., Li, Y., Jiang, L., and Liang, S. (2010) Molecular diversification of peptide toxins from the tarantula Haplopelma hainanum (Ornithoctonus hainana) venom based on transcriptomic, peptidomic, and genomic analyses. J. Proteome Res. 9, 2550–2564
11. Escoubas, P. (2006) Molecular diversification in spider venoms: a web of spider evolution. J. Proteome Res. 5, 54–55
12. Durban, J., Juárez, P., Angulo, Y., Lomonte, B., Flores-Diaz, M., Corzo, G., and Escoubas, P. (2003) Pharmacologically active spider peptide toxins. Cell Mol. Life Sci. 60, 2409–2426
13. Terrat, Y., Biass, D., Dutertre, S., Favreau, P., Remm, M., Stöcklin, R., Piquemal, D., and Ducancel, F. (2012) High-resolution picture of a venom gland transcriptome: case study with the marine snail Conus consors. Toxicon 59, 34–46
14. Rokyta, D. R., Lemmon, A. R., Margres, M. J., and Aronow, K. (2012) The spider venom-gland transcriptome: a case study with the marine snail Conus pulicarius discovered through next-generation sequencing of its venom duct transcriptome. Mar. Genomics 5, 43–51
15. Lluisma, A. O., Milash, B. A., Moore, B., Olivera, B. M., and Bandopadhyay, P. K. (2012) Novel venom peptides from the cone snail Conus pulicarius discovered through next-generation sequencing of its venom duct transcriptome. Mar. Genomics 5, 43–51
16. Prosdocimi, F., Bittencourt, D., da Silva, F. R., Kirst, M., Motta, P. C., and Rech, E. L. (2011) Spinning gland transcriptomics from two main clades of spiders (order: Aranaeae): insights on their molecular, anatomical and behavioral evolution. PLoS One 6, e21634
Molecular Diversification of Peptide Toxins from the Tarantula H. hainanum

17. Zhang, J., Chiodini, R., Badr, A., and Zhang, G. (2011) The impact of next-generation sequencing on genomics. *J. Genet. Genomics* **38**, 95–109

18. Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., and Law, M. (2012) Comparison of next-generation sequencing systems. *J. Biomed. Biotechnol.* **2012**, 251364

19. Zheng, Y., Zhao, L., Gao, J., and Fei, Z. (2011) *Assemblage: a package for de novo assembly of Roche-454/Sanger transcriptome sequences*. *BMC Bioinformatics* **12**, 453

20. Conesa, A., Götz, S., García-Gómez, J. M., Terol, J., Talón, M., and Robles, M. (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *BMC Bioinformatics* **21**, 3674–3676

21. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. (2000) Gene ontology: tool for the unification of biology. *The Gene Ontology Consortium. Nat. Genet.* **25**, 25–29

22. Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, M. V. (2001) Spider neurotoxins block the calcium uptake in rat brain cortical synaptosomes. *J. Biol. Chem.* **276**, 509–525

23. Duan, D., Rong, M., Zeng, Y., Teng, Y., Zhao, Z., Liu, B., Tao, X., Zhou, R., Fan, M., Peng, C., Chen, P., Liang, S., and Lu, M. (2011) Electrophysiological characterization of NSCs after differentiation induced by OEC conditioned medium. *Acta Neurochir.* **153**, 2085–2090

24. Kumar, S., Nei, M., Dudley, J., and Tamura, K. (2008) MEGA: a biologist-centric software for evolutionary analysis of DNA and protein sequences. *Brief. Bioinform.* **9**, 299–306

25. Xiao, Y., and Liang, S. (2003) Inhibition of neuronal tetrodotoxin-sensitive Na+ channels by two spider toxins: hainainoxin-III and hainainoxin-IV. *Eur. J. Pharmacol.* **477**, 1–7

26. Naamati, G., Askenazi, M., and Linial, M. (2009) ClanTox: a classifier of shared sequence homology but not pharmacological properties with AVIT family proteins. *Toxicon* **53**, 609–617

27. Bohlen, C. J., Priel, A., Zhou, S., King, D., Siemens, J., and Julius, D. (2010) A bivalent tarantula toxin activates the capsaicin receptor, TRPV1, by targeting the outer pore domain. *Cell* **141**, 834–845

28. Wen, S., Wilson, D. T., Kuruppu, S., Korsinczky, M. L., Hedrick, J., Pang, L., Szeto, T., Hodgson, W. C., Alewod, P. F., and Nicholson, G. M. (2005) Discovery of an M1-like arachatoxin family: spider venom peptides that share sequence homology but not pharmacological properties with AVIT family proteins. *Peptides* **26**, 2412–2426

29. Li, D., Xiao, Y., Xu, X., Xiong, X., Lu, S., Liu, Z., Zhu, Q., Wang, M., Gu, X., and Liang, S. (2004) Structure-activity relationships of hainainoxin-IV and structure determination of active and inactive sodium channel blockers. *J. Biol. Chem.* **279**, 37734–37740

30. Adams, M. E. (2004) Agatoxins: ion channel specific toxins from the American funnel web spider, *Agelenopsis aperta*. *Toxicon* **43**, 509–525

31. Pimentel, C., Choi, S. J., Chagot, B., Guette, C., Camadro, J. M., and Darzil, H. (2006) Solution structure of PcFK1, a spider peptide active against *Plasmodium falciparum* calcium channels. *FEBS Lett.* **572**, 109–117

32. Bindokas, V. P., Venema, V. J., and Adams, M. E. (1991) Differential antagonism of transmitter release by subtypes of 2,6-agatoxins. *J. Neurophysiol.* **66**, 590–601

33. Ding, J., Peng, L., Chen, J., Zhang, Y., Xiong, X., and Liang, S. (2008) Molecular diversification based on analysis of expressed sequence tags from the venom glands of the Chinese bird spider *Ornithoctonus huwena*. *Toxicon* **51**, 1479–1489

34. Wang, E. S., and Belov, K. (2012) Venom evolution through gene duplications. *Gene* **496**, 1–7

35. Ingrao, F., Rauw, F., Lambrechts, B., and van den Berg, T. (2013) Infectious bursal disease: a complex host-pathogen interaction. *Dev. Comp. Immunol.* **41**, 429–438

36. Choi, S. J., Parent, R., Guillame, C., Deregnaucourt, C., Delarbre, C., Ojcius, D. M., Montagne, J. J., Célérie, M. L., Phelipot, A., Amiche, M., Molgo, J., Camadro, J. M., and Guette, C. (2004) Isolation and characterization of Psalmopoeus toxin I and II: two novel antimalarial peptides from the venom of the tarantula *Psalmopoeus cambridgei*. *FEBS Lett.* **572**, 109–117

37. Jiang, L., Zhao, L., Gao, J., and Fei, Z. (2011) iAssembler: a package for de novo assembly of Roche-454/Sanger transcriptome sequences. *BMC Bioinformatics* **12**, 453

38. Pimentel, C., Choi, S. J., Chagot, B., Guette, C., Camadro, J. M., and Darzil, H. (2006) Solution structure of PcFK1, a spider peptide active against *Plasmodium falciparum* calcium channels. *FEBS Lett.* **572**, 109–117

39. Wong, E. S., and Belov, K. (2012) Venom evolution through gene duplications. *Gene* **496**, 1–7

40. Conticello, S. G., Gilad, Y., Avidan, N., Ben-Asher, E., Levy, Z., and Painzilber, M. (2001) Mechanisms for evolving hypervariability: the case of conopeptides. *Mol. Biol. Evol.* **18**, 120–131

41. Zhang, Y., Huang, Y., He, Q., Liu, J., Luo, J., Zhu, L., Lu, S., Huang, P., Chen, X., Zeng, X., and Liang, S. (2014) Toxin diversity revealed by a transcriptomic study of *Ornithoctonus huwena*. *PLoS One* **9**, e100682

42. Dutertre, S., Jin, A. H., Kaas, Q., Jones, A., Alewod, P. F., and Lewis, R. J. (2013) Deep venomomics reveals the mechanism for expanded peptide diversity in cone snail venom. *Mol. Cell. Proteomics* **12**, 312–329

43. Violette, A., Biss, D., Dutertre, S., Koua, D., Piquemal, D., Pierrat, F., Stöcklin, R., and Favreau, P. (2012) Large-scale discovery of conopeptides and conoproteins in the injectable venom of a fish-hunting cone snail using a combined proteomic and transcriptomic approach. *J. Proteomics* **75**, 5215–5225

44. Urbanova, I., Karlsen, B. O., Okkenhaug, S., Seternes, O. M., Johansen, S. D., and Emblem, A. (2012) Digital marine bioprospecting: mining new neurotoxin drug candidates from the transcriptomes of cold-water sea anemones. *Mar. Drugs* **10**, 2265–2279

45. Rong, M., Duan, Z., Chen, J., Li, J., Xiao, Y., and Liang, S. (2013) Native pyroglutamation of huwentoxin-IV: a post-translational modification that increases the trapping ability to the sodium channel. *PLoS One* **8**, e65984

46. Feng, K., Shu, Q., Liu, Z., and Liang, S. (2002) Solution and structure determination of huwentoxin-IV, a potent neuronal tetrodotoxin (TTX)-sensitive sodium channel antagonist from Chinese bird spider *Selenocosmia huwena*. *J. Biol. Chem.* **277**, 47564–47571

47. Zhang, D., and Liang, S. (1999) Assignment of the three disulfide bridges of huwentoxin-I, a neurotoxin from the spider *Selenocosmia huwena*. *J. Proteome Res.* **12**, 735–740

48. Liu, S., Liang, S., and Gu, X. (1999) Three-dimensional structure of *Selenocosmia huwena* lectin-1 (SHL-1) from the spider venom of the spider *Selenocosmia huwena* by 2D-NMR. *J. Protein Chem.* **18**, 609–617

49. Huang, R. H., Liu, Z. H., and Liang, S. (2003) Purification and characterization of a neurotoxic peptide huwentoxin-III and a natural inactive mutant from the venom of the spider *Selenocosmia huwena* (Ornithoctonus huwena Wang). *Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue* **23**, 976–980

50. Liu, Z., Dai, J., Dai, L., Deng, M., Hu, Z., Hu, W., and Liang, S. (2006) Function and solution structure of Huwentoxin-X, a specific blocker of N-type calcium channels, from the Chinese bird spider *Ornithoctonus hainanum*.
55. Morgenstern, D., Rohde, B. H., King, G. F., Tal, T., Sher, D., and Zlotkin, E. (2011) The tale of a resting gland: transcriptome of a replete venom gland from the scorpion Hottentotta judaicus. *Toxicon* 57, 695–703

56. Zelanis, A., Andrade-Silva, D., Rocha, M. M., Furtado, M. F., Serrano, S. M., Junqueira-de-Azevedo, I. L., and Ho, P. L. (2012) A transcriptomic view of the proteome variability of newborn and adult *Bothrops jararaca* snake venoms. *PLoS Negl. Trop. Dis.* 6, e1554

57. Diao, J., Lin, Y., Tang, J., and Liang, S. (2003) cDNA sequence analysis of seven peptide toxins from the spider *Selenocosmia huwena*. *Toxicon* 42, 715–723

58. Zhang, P. F., Chen, P., Hu, W. J., and Liang, S. P. (2003) Huwentoxin-V, a novel insecticidal peptide toxin from the spider *Selenocosmia huwena*, and a natural mutant of the toxin: indicates the key amino acid residues related to the biological activity. *Toxicon* 42, 15–20

59. Deng, M., Luo, X., Meng, E., Xiao, Y., and Liang, S. (2008) Inhibition of insect calcium channels by huwentoxin-V, a neurotoxin from Chinese tarantula *Ornithoctonus huwena* venom. *Eur. J. Pharmacol.* 582, 12–16

60. Kaiser, I. I., Griffin, P. R., Aird, S. D., Hudiburg, S., Shabanowitz, J., Francis, B., John, T. R., Hunt, D. F., and Odell, G. V. (1994) Primary structures of two proteins from the venom of the Mexican red knee tarantula (*Brachypelma smithii*). *Toxicon* 32, 1083–1093

61. Savel-Nieman, A. (1989) Tarantula (*Eurypelma californicum*) venom, a multicomponent system. *Biol. Chem. Hoppe Seyler* 370, 485–498

62. Oswald, R. E., Suchyna, T. M., McFeeters, R., Gottlieb, P., and Sachs, F. (2002) Solution structure of peptide toxins that block mechanosensitive ion channels. *J. Biol. Chem.* 277, 34443–34450

63. Diochot, S., Drici, M. D., Moinier, D., Fink, M., and Lazdunski, M. (1999) Effects of phrixotoxins on the Kv4 family of potassium channels and implications for the role of Ito1 in cardiac electrogensis. *Br. J. Pharmacol.* 126, 251–263

64. Liao, Z., Cao, J., Li, S., Yan, X., Hu, W., He, Q., Chen, J., Tang, J., Xie, J., and Liang, S. (2007) Proteomic and peptidomic analysis of the venom from Chinese tarantula *Chilobrachys jingzhao*. *Proteomics* 7, 1892–1907

65. Stapleton, A., Blankenship, D. T., Ackermann, B. L., Chen, T. M., Gorder, G. W., Manley, G. D., Palfreyman, M. G., Coutant, J. E., and Cardin, A. D. (1990) Curtatoxins: neurotoxic insecticidal polypeptides isolated from the funnel-web spider *Hololena curta*. *J. Biol. Chem.* 265, 2054–2059

66. Skinner, W. S., Adams, M. E., Quistad, G. B., Kataoka, H., Cesarin, B. J., Enderlin, F. E., and Schooley, D. A. (1989) Purification and characterization of two classes of neurotoxins from the funnel web spider, *Agelenopsis aperta*. *J. Biol. Chem.* 264, 2150–2155