Published in final edited form as:

J Proteome Res. 2014 February 07; 13(2): 817–835. doi:10.1021/pr400875s.

Spit and venom from *Scytodes* spiders: a diverse and distinct cocktail

Pamela A. Zobel-Thropp^{§,*}, Sandra M. Correa[‡], Jessica E. Garb[‡], Greta J. Binford[§]

Pamela A. Zobel-Thropp: pamela@lclark.edu; Sandra M. Correa: scor006@ucr.edu; Jessica E. Garb: jessica_garb@uml.edu; Greta J. Binford: binford@lclark.edu

[§]Department of Biology, Lewis & Clark College, Portland, OR 97219, USA

[‡]Department of Biological Sciences, University of Massachusetts Lowell, Lowell, MA 01854, USA

Abstract

Spiders from the family Scytodidae have a unique prey capturing technique – they spit a zig-zagged silken glue to tether prey to a surface. Effectiveness of this sticky mixture is based on a combination of contraction and adhesion trapping prey until the spider immobilizes it by envenomation and then feeds. We identify components expressed in *Scytodes thoracica* venom glands using combined transcriptomic and proteomic analyses. These include homologs of toxic proteins astacin metalloproteases and potentially toxic proteins including venom allergen, longistatin and translationally controlled tumor protein (TCTP). We classify 19 distinct groups of candidate peptide toxins; 13 of these were detected in the venom making up 35% of the proteome. Six have significant similarity to toxins from spider species spanning mygalomorph and non-haplogyne araneomorph lineages, suggesting their expression in venom is phylogenetically widespread. Twelve peptide toxin groups have homologs in venom gland transcriptomes of other haplogynes. Of the transcripts, approximately 50% encode glycine-rich peptides that may contribute to sticky fibers in *Scytodes* spit. Fifty-one percent of the identified venom proteome is a family of proteins that is homologous to sequences from *Drosophila sp.* and *Latrodectus hesperus* with uncharacterized function. Characterization of these components holds promise for discovering new functional activity.

Keywords

Scytodes; venome; toxin; adhesion; cDNA library; Orbitrap mass spectrometry

Introduction

Spiders are highly successful predators whose venoms contain rich stores of toxic proteins, peptides and small molecules that have evolved as tools to help capture and immobilize prey. Over 44,000 species of spiders have been identified¹, but only a handful of venoms from these species have been characterized with molecular techniques. Spider venoms have

*Corresponding author: pamela@lclark.edu; Tel: 503-768-7653; FAX: 503-768-7658.

Supporting Information Available: This material is available free of charge via the Internet at <http://pubs.acs.org>

gained recent interest because they contain combinatorial peptide libraries that can be mined for bioinsecticides and medicinal purposes^{2–13}.

Spiders with unusual feeding biologies are excellent candidates for discovering novel venom components, and *Scytodes* rank among the most unusual. These spiders spit glue to tether arthropod prey before immobilizing them with venom. The genus *Scytodes* has a worldwide distribution that contains nearly 200 species¹. These spiders have a domed cephalothorax that houses a large pair of glands, each of which can produce venom and glue¹⁴. The venom is produced in the gland's anterior and the glue substance is produced in the posterior portion of the gland, but both are extruded through the same duct¹⁴. They have diminutive chelicerae and fangs, from which a viscous, gluey substance is rapidly sprayed independently from each fang in a zigzag fashion. The substance is adhesive, contractile, and immobilizes prey while tethering them to a surface^{15,16}. High-resolution microscopy of *Scytodes* spit shows long, continuous fibrous strands connecting sticky glue droplets. These strands shrink a remarkable 40–60% upon ejection from the fangs¹⁶. Direct application of the mixture on prey has no toxic effects¹⁷.

Once a prey item is tethered, the spider approaches and envenomates, with the chemical action of the venom further immobilizing the prey. Nothing is known about the components in *Scytodes* venom that contribute to the adhesive and contractile properties of the spit, or the toxins that paralyze their prey. *Scytodes*, however are in the same superfamily, Scytodoidae, as brown recluse spiders in the genus *Loxosceles* whose venoms have been more extensively studied^{18–20}.

The goal of this work was to identify the composition of peptides and proteins expressed in the venom/spit glands of *Scytodes*. This is the first analysis of *Scytodes* venom at the molecular level to our knowledge. Using transcriptomics and proteomics, we have identified components that make up a complex set of proteins and peptides expressed in the venom/spit glands. Taken together, this venome (venom transcriptome and proteome) analysis gives new insight into the evolution of the chemical arsenal involved in the unique prey capture technique of *Scytodes*. It also fills in an important piece of the phylogenetic puzzle that allows us to better understand evolutionary dynamics of venoms.

Materials and Methods

Our data combine transcriptome analyses from Jessica Garb's lab at the University of Massachusetts Lowell (UML) and Greta Binford's lab at Lewis & Clark College (LC). Two venom gland cDNA libraries were independently constructed from North American *Scytodes thoracica*. The UML library was optimized for capturing larger transcripts, such as potential silk (spidroin) transcripts whereas the LC library was constructed for screening a broader range of transcripts including small venom peptides <10 kDa in size. We pooled the data into a single analysis because 16S ribosomal RNA nucleotide transcript sequences in both libraries share 99% identity, suggesting these spiders are from closely related populations.

Spiders

Jessica Garb received *Scytodes thoracica* from Gail Stratton in Oxford, Mississippi where spiders were collected in July 2010. Greta Binford and colleagues collected *Scytodes thoracica* spiders in Del Rio, Texas in May 2010. Detailed collecting information is available on request and vouchers are retained in the personal collections of JG and GJB.

Proteomic analysis

We obtained venom/spit from seven females, five of which were then used for the Binford lab venom gland cDNA library. The extruded venom/spit was sticky and viscous so we wrapped it around a capillary tube while twisting and stored the aliquot at -70°C . In preliminary analyses we were able to dissolve the venom, but not all of the viscous material, in a standard buffer we use for venom protein analysis (5 mM CaCl_2 /50 mM Tris, pH 8) after incubating the capillary tips for 2 days (4°C). We visualized the proteins with SDS-PAGE (Supplemental Figure 1).

The venom-expressed proteins from the individuals used for the transcriptome were analyzed from an in-gel tryptic digest with LC-LC-MS/MS (MudPIT) and also with LC-MS/MS using an LTQ Orbitrap Velos mass spectrometer for detection at the Arizona Proteomics Consortium (University of Arizona). For MudPIT, the digested peptide mixtures were loaded by pressure packing onto a fused silica capillary cation exchange column (100 micron), packed with PolySulfoethyl-Asp-coated 5-micron beads (3.5 cm) (SCX, PolyLC Inc.) packed against a microfilter assembly (Upchurch Scientific). This was then connected to a fused silica capillary column (100 micron), packed with Zorbax Eclipse XDB-C18 5-micron beads (8 cm) (Agilent) that was also the nanospray emitter tip by pulling the fused silica to a 5-micron tip using a laser puller (Sutter Instrument). This set-up was connected to a split-flow microbore HPLC system (Surveyor, Thermo Electron) and peptides were separated in a gradient using Buffer A (0.1% formic acid), Buffer B (acetonitrile/0.1% formic acid), Buffer C (250 mM ammonium acetate) and Buffer D (1.5 M ammonium acetate). The analysis was performed at a flow rate of 400 nL/min involving multiple steps of salt bumps (with Buffers C and D) followed by a gradient (Buffers A and B), each of which was followed by a column equilibration of 5 min with 5% B before continuing: step (1) 0% C with a gradient of 5–50% B over 90 min 50–98% B over 20 min, 98% B for 5 min; steps (2–11) 10–100% C with an increase in increments of 10% each over 4 min, buffer rinse of 5% B for 7 min, 5–50% B over 60 min, 50–98% B over 20 min, 98% B for 5 min; step (12) 50% D loaded over 4 min, buffer rinse of 5% B for 7 min, 5–50% B over 60 min, 50–98% B over 20 min, 98% B for 5 min. The HPLC eluent was directed into the ESI source of an LTQ linear ion trap mass spectrometer (Thermo Fisher Scientific). Electrospray voltage of 2.0 kV was applied using a gold electrode *via* a liquid junction upstream of the column, and spectra were scanned over a range of 400–1500 amu. Automated peak recognition, dynamic exclusion and daughter ion scanning of the top seven most intense ions were performed using the Xcalibur data system (Thermo Fisher Scientific).

For detection in Orbitrap, the peptide mixtures were analyzed using an EASY n-LC (Proxeon, Odense) coupled to an Orbitrap LTQ Velos mass spectrometer (Thermo Fisher Scientific). First, samples were loaded onto a C18 pre-column (100 micron (2 cm), Thermo

Fisher Scientific) at a pressure of 280 Bar in 0.1% formic acid; the column was then switched online with the analytical column (75-micron (10 cm), C18, Thermo Fisher Scientific) and equilibrated for 10 min in 95% solvent A (1% formic acid) and 5% solvent B (acetonitrile/0.1% formic acid). Peptides were eluted at a flow rate of 300 nl/min using a gradient of 5–40% B over 145 min followed by 40–85% B over 15 min. The eluate was directed through the nanospray source (Nanomate, Advion) with a voltage of 1.85 kV into the Orbitrap. The Orbitrap LTQ Velos was operated in data-dependent acquisition mode with Xcalibur software. Survey scans were acquired at 60,000 resolution scanning m/z 400–1600. The 14 most intense ions observed in the survey were selected for CID fragmentation, with detection in the linear ion trap. Precursor ions were selected by the monoisotopic precursor selection (MIPS) setting with selection or rejection of ions held to a ± 10 ppm window. Once observed, any selected m/z was placed on an exclusion list for 45 seconds.

Peptide masses were analyzed using Sequest (version 1.3.0.339, Thermo Fisher Scientific) against masses of theoretical fragments of a database that included our transcriptome sequences (all SCV sequences translated in 6 frames along with all SCY sequences translated in forward 3 frames (see “cDNA library construction” below)), all chelicerate sequences in NCBI (downloaded 11/17/13), and unpublished sequences from multiple Haplogyne venom gland tissue cDNA libraries constructed in the Binford lab (93,243 sequences in all). For MudPIT data, Sequest was searched with a fragment ion mass tolerance of 0.80 Da and a parent ion tolerance of 1.5 Da; for Orbitrap data, Sequest was searched with a fragment ion mass tolerance of 0.80 Da and a parent ion tolerance of 10 ppm. Oxidation of methionine and carbamidomethyl of cysteine were specified as variable modifications for both methods. All data for both methods including protein and peptide thresholds, FDR and peptides detected are available in Supplemental Table 2. We obtained more data from Orbitrap analysis (683 peptides) than MudPIT analysis (353 peptides), with an overlap of 43%; however, the Orbitrap method is sensitive enough to detect small sized venom peptides, so we did not have the MudPIT repeated for a total of three runs, which is suggested for improved coverage of a complex protein mixture (Linda Brexi, University of Arizona Proteomics Facility, pers. comm.).

Venom gland and RNA isolation, cDNA library construction and screening

At LC, venom/spit was extracted from seven female spiders using electrical stimulation as in Binford and Wells²¹, and stored at -70° C. After two days, five of the seven spiders were anesthetized using CO₂, and the venom and glue glands were removed and flash-frozen immediately in liquid nitrogen. Two-three days post electrostimulation is a window of time when venom-expressed genes are known to be transcribed while replenishing venom in the glands in other Haplogyne spiders²². Total RNA was isolated using the ChargeSwitch[®] Total RNA Cell Kit (Invitrogen).

The LC venom gland cDNA library was constructed using the SMART[™] cDNA library construction kit (Clontech), which is optimized for making complete, full-length cDNAs. We used 0.4 μ g of total RNA for first strand synthesis of cDNA and followed the manufacturer's protocol for library construction. Upon ligation into the λ TriplEx2 vector, the cDNAs were packaged into MaxPlax[™] lambda packaging extracts (Epicentre) according

the manufacturer's protocol and infected *E. coli* XL1-Blue bacterial cells (Clontech). The titer for this library was 3×10^6 pfu/ml. Phage particles were eluted from plaques by poking a pipet tip through each plaque and stirring it into 500 μ l of a buffered solution (100 mM NaCl/10 mM MgSO₄/50 mM Tris-Cl, pH 7.5/0.01% gelatin) with 20 μ l of chloroform. These mixtures were incubated for 2 hrs at room temperature and stored at 4° C.

From the LC library 1,588 clones were PCR screened to identify cDNA inserts 500 bp. Each reaction contained 1 μ l of eluted cDNA, the λ TriplEx2 primers that flank the insert (Clontech, 0.2 μ M final concentration), 2X MasterAmpTM F PCR Premix (Epicentre, 1x final concentration) and *Taq* polymerase (New England Biolabs, 0.05 U/ μ l final concentration) with cycling parameters of 95° C for 1 min 30 sec followed by 30 cycles of 95° C for 1 min/66° C for 1 min/72° C for 1 min, and a final extension at 72° C for 7 min. PCR products were checked for size on a 1% agarose gel compared to a 1kb DNA mass ladder (Invitrogen) and stained with SYBR[®] Safe (Invitrogen). Clones with PCR products >500 bp were converted from the λ vector to plasmids (pTriplEx2) using *E. coli* BM25.8 cells provided with the library kit (Clontech). A total of 1,062 plasmids were sequenced with the 5' - and 3' -sequencing primers specified by the TriplEx2 vector map. We trimmed vector sequence and assembled all sequences using Sequencher 4.7 (Gene Codes Corp.), resulting in 765 sequences for transcriptomic analysis ("SCY" sequences).

At UML, two total RNA extractions were made from the combined venom glands of eight females and eleven males by homogenization in TRIzol (Invitrogen) followed by purification using the RNeasy Mini-kit (Qiagen). mRNA was isolated from the combined RNA extractions using oligo-(dT)₂₅-tagged magnetic beads (Invitrogen) and SuperScript III (Invitrogen) used to make cDNA with an oligo-(dT)₁₈V primer as an anchor. The cDNA was size fractionated using ChromaSpin 1000 columns (Clontech) to enrich for transcripts 1000 bp. cDNAs were then cloned into the pZER0-2 plasmid using blunt-end ligation at EcoRV-digested sites and recombinant plasmids transformed into *Escherichia coli* TOP10 competent cells (Invitrogen) by electroporation. In total, 1,728 colonies were screened for cDNAs 700 bp by extracting DNA with phenol:chloroform, and visualizing recombinant plasmids on an agarose gel, along with supercoiled DNA ladder (NEB). Size-selected plasmids were then purified and cDNA inserts were sequenced with either T7 or Sp6 universal primers. Chromatograms were edited using Sequencher 4.8 (Gene Codes Corp.), removing vector and poor quality sequence. In total, 448 edited sequences were used for transcriptome analysis ("SCV" sequences).

Bioinformatics

Our analysis pipeline is illustrated in Figure 1B. To identify transcripts, all nucleotide sequences (1,213 total) were submitted to tBLASTx searches in the ArachnoServer spider toxin database (<http://www.arachnoserver.org>) and the NCBI nucleotide database (<http://www.ncbi.nlm.nih.gov/blast/Blast.cgi>). We also used tBLASTx to compare our transcriptome to several Haplogyne venom gland cDNA libraries that we have constructed in the Binford lab (unpublished). Hits were considered significant if e-values were 10^{-5} . Seventy-one sequences <100 bp were too small for subsequent translation and analysis, so they were deleted from the dataset. Using Sequencher 5.0 (Gene Codes Corp.), we

assembled the nucleotide sequences from our combined libraries into sequence clusters using 99% and 85% threshold identities with a minimum 20 bp overlap. The 99% threshold determined redundant vs. nonredundant sequences, whereas the 85% threshold identified homologous sets of sequences. The clustering from the 85% assembly is presented for transcriptome analysis.

All sequences were translated using EMBOSS Transeq (http://www.ebi.ac.uk/Tools/st/emboss_transeq/); the sequences from the Garb lab were translated in all six frames and those from the Binford lab were translated in the forward three frames (the Clontech SMARTTM cDNA library construction kit is optimized for forward directional cloning). All translated sequences were submitted to the CDD (Conserved Domain Database) to search for conserved protein structures (<http://www.ncbi.nlm.nih.gov/Structure/bwrpsb/bwrpsb.cgi>). All nonredundant amino acid sequences were submitted to the KOG (eukaryotic orthologous groups) database using KOGnitor (<http://www.ncbi.nlm.nih.gov/COG/grace/kognitor.html>).

Sequences that were identified as metabolism/homeostasis (housekeeping) proteins were not analyzed further. For all sequences that did not hit matches in KOG or CDD we did a variety of analyses to identify the candidate general function of the transcripts. First, we considered full-length protein coding sequences “open reading frames” (ORFs) if they contained a methionine, a stop codon and consisted of 30 aa. Ten clusters and 40 singleton sequences were deemed “no ORF” (Fig. 1B). We visually looked for conserved characteristics consistent with known venom toxins (e.g. cysteine-rich) and characteristics consistent with silk proteins and known adhesive peptides (e.g. glycine- and alanine-rich, diglutamine (QQ) and dityrosine (YY) repeat motifs, and repetitive architecture)^{23–26}. Cysteine-rich sequences with putative toxin folding patterns (*i.e.*, -C-C-CC-C-C-) were submitted to the Knott1d program (http://knottin.cbs.cnrs.fr/Tools_1D.php) to predict if they fold into knottin/ICK structures; only a few of the sequences were predicted to be knottins, which is the case for a majority of spider venom sequences with the same cysteine patterns (personal observation). We also used DISULFIND (<http://disulfind.dsi.unifi.it>), which resulted in the same few sequences bonding with the same cysteine patterns. We looked for signal peptide cleavage sites within all ORF sequences using SignalP (<http://www.cbs.dtu.dk/services/SignalP/>). We used compute pI/MW (http://web.expasy.org/compute_pi/) to predict size (MW) and overall charge (pI) of mature protein and peptide sequences. Distance estimates were calculated (total number of amino acid differences) using MEGA 5.1²⁷.

To identify conserved sequence features we aligned the translated amino acid sequences within each cluster produced by the 85% threshold using MUSCLE (<http://www.ebi.ac.uk/Tools/msa/muscle/>) and then refined the alignment manually using Mesquite version 2.75 (<http://www.mesquiteproject.org>). We used Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) and Jalview 2.8 (<http://www.jalview.org>) to refine alignments for presentation in figures. Along with the *Scytodes* transcripts, we included in our alignments any hits in databases that were $<10^{-5}$ from searches in ArachnoServer, NCBI nucleotide, NCBI EST, and our unpublished database of Haplogyne spider venom gland transcripts.

Accession numbers KF860345-KF860787 have been assigned to nonredundant nucleotide sequences by NCBI.

Phylogenetic analyses

For select gene families, we estimated phylogenetic relationships of amino acid sequences using Bayesian analyses sampling across fixed amino acid rate matrices using the aamodelpr=mixed option (MrBayes v3.2.1)²⁸. Each analysis consisted of two runs with four simultaneous Markov Chain Monte Carlo (MCMC) chains. All analyses reached convergence by 1 million generations, in all cases with average standard deviation of split frequencies reaching <0.01 ²⁹. We determined the burnin using the Tracer 1.3 which was always 10% (<http://evol.zoo.ox.ac.uk/software.html?id=tracer>). There were no confident outgroups for any phylogenies so most phylogenies are midpoint rooted. Datasets that included genes from divergent taxa were rooted with the most divergent outgroup in the dataset based on species relationships.

Results and Discussion

We analyzed 1,142 transcriptome sequences combined from two independently constructed *Scytodes thoracica* venom and glue gland cDNA libraries. Within this dataset, 563 sequences are nonredundant at the nucleotide level (at the 99% identity threshold), constituting 49% of the transcriptome. Clustering the nucleotide sequences with the 85% identity threshold resulted in 74 clusters and 194 singletons. Of the 74 clusters, 54 contained only 2–5 sequences, 17 contained 6–100 sequences and three contained more than 100 sequences. All clusters at 85% identity contained sequences that were easily aligned to one another, suggesting that clusters group sets of paralogs or allelic variants. For analyses to infer putative function or functional category, we assume function is consistent within clusters and refer to the total number of “entities” as 268, the combined total of singletons and clusters that form at the 85% threshold (Fig. 1B). However, all sequences were searched against all databases. Figure 1 is a flow chart overview of our strategy for estimating functional categories of sequences.

Before clustering, 37% of all 1,142 sequences hit something in GenBank (nucleotide or EST databases) or Arachnoserver at e^{-5} . Hits included proteins involved in nonvenomous “housekeeping” (metabolism/homeostasis) functions, putative venom peptides and toxic proteins and proteins or peptides for which we were not able to predict general functional category. Overall transcript abundance among these categories is presented in Figure 1A. A search in the Conserved Domain Database identified 135 entities with e-values ranging from 10^{-5} to 0.0. KOGnitor identified 88 distinct KOG groups (Supplemental Table 1), 85 of which matched supported hits from CDD searches. We did not further analyze proteins that are likely involved in housekeeping/metabolism/homeostasis. Of the remaining 124 entities that hit nothing in any database, we were able to predict the function of 74 that were full-length (initiating M and stop) based on sequence motifs (Fig. 1B).

Proteomic analyses help determine the subset of our transcripts from venom gland tissues that are expressed as functional components in the venom. We analyzed the proteome using LC-LC-MS/MS (MudPIT) once and LC-MS/MS (Orbitrap) twice with a goal of capturing a wide breadth of protein/peptide sizes (Supplemental Table 2). We are confident in the identification of components detected by proteomics, however lack of detection may be false negatives and cannot be interpreted with confidence that a transcript was not present in crude

venom extract. The data from MudPIT and Orbitrap resulted in 43% overlap and each detected a group of proteins that was not found with the other method (Fig. 2A). LC-MS/MS analyses of crude venom and dissolved viscous spit from *Scytodes* produced mass spectra with significant matches (95% probability) to 48 different peptides and proteins translated from our transcriptome cDNAs that represent 29 clusters and 7 singletons. These sequences were found with a combined total of 1,036 tryptic peptide matches; numbers of peptide matches directly correlate with relative abundance of proteins in the venom/spit mixture^{30,31} (also Linda Brei, University of Arizona Proteomics Facility, pers. comm.) (Fig. 2B). Only 5% of the peptides hit chelicerate sequences; 4% represents various subunits that make up hemocyanin and the remaining 1% match other proteins such as kinesin, papilin and uncharacterized proteins (Fig. 2B). Below we describe each of the types of protein components found in the *Scytodes* venom gland transcriptome and discuss the evidence of their presence in crude venoms. We organize this by putative functional category and size with proteins > 10 kDa.

Candidate Venom Toxin Proteins

Transcripts that we characterize as potentially toxic proteins (> 10 kDa) make up 2% of all 1,142 sequences in the *Scytodes* transcriptome. We identified a homolog of longistatin (1 cDNA sequence) ($e10^{-25}$), a toxic salivary protein in ticks^{32,33}. Nine sequences are homologous to proteins that are expressed in venom glands of other spider species and proposed to be toxic, including astacins³⁴ (5) ($e10^{-32}$), translationally controlled tumor protein or TCTP³⁵ (3) ($e10^{-87}$) and venom allergen^{18,20,36} (1) ($e10^{-17}$). Of these, astacin and venom allergen were detected in our proteome (Fig. 2).

We found a single clone in the *Scytodes* transcriptome (SCY1053) that shares equally strong similarity ($\sim e10^{-28}$) to sequences annotated as **longistatin** from *Haemaphysalis longicornis* (tick), multiple coagulation factor deficiency 2-like proteins from *Ixodes scapularis* (tick), *Haplopelma huwenum* (tarantula) and additional hypothetical proteins annotated from various organisms including *Scorpiops jendeki* (scorpion) (Fig. 3). The protein contains a signal cleavage site and the mature sequence is predicted to be 13.8 kDa with a pI of 4.7. Longistatins are fibrinolytic and degradative proteins, anticoagulants secreted by ticks that are essential for the formation of blood pools and feeding^{32,33}. They are predicted to bind Ca^{2+} ions *via* the EF-hand domain pair, which is a helix-loop-helix structural domain³⁷ (black bar, Fig. 3). Recent studies by Anisuzzaman *et al.* using rabbits have shown that longistatin from the tick *H. longicornis* functions as a unique type of serine protease³³ hydrolyzing fibrinogen, an essential coagulant, and activating the zymogen plasminogen into its active form, plasmin, which dissolves fibrin clots in blood pools formed³².

Five sequences have significant similarity to **astacins**, which are metalloproteases known to be toxic venom proteins^{34,38–40}. They contain the conserved Zn-binding and protease domain **HEXXHXXGFXHE** along with the Met-turn motif **SXMX^Y**⁴¹ and share up to 33% identity with metalloproteases from *C. quinquefasciatus* and *A. aegypti* ($e10^{-27}$). These proteins are abundant in the transcriptomes of *Loxosceles* species^{18,20}, close relatives of *Scytodes*, and have been shown to degrade matrix proteins (gelatin, fibronectin and enactin), suggesting a key role in the spreading of venom^{34,38–40}. Astacins have also been identified

in venom gland transcriptomes of *Austrarchaea mainae*, *Physocylus mexicanus*, *Plectrurus tristis*, *Sicarius dolichocephalus* and various *Loxosceles* species including *spinulosa*, *rufescens*, *reclusa* and *arizonica*; they are among the most abundant transcripts found in *P. mexicanus* and *S. dolichocephalus* (Binford *et al.*, unpublished).

Three cDNAs we identified as **SthTCTP** (*S. thoracica* TCTP) are 86% identical to LiTCTP from *L. intermedia*. LiTCTP was recently shown to directly cause edema and increase vascular permeability in rabbits, providing evidence for an inflammatory role in loxoscelism³⁵. The molecular weight of the *Scytodes* amino acid sequence is predicted to be 19.3 kDa with a pI of 4.63. Phylogenetic analysis of SthTCTP (SCY273) and several significant hits (e^{-5}) from NCBI support a close relationship between LiTCTP and the *Scytodes* sequence (Fig. 4). TCTP, also known as HRF (histamine-releasing factor), is widely expressed in various tissues and cell types, and expression levels have been found to respond to a wide range of signals, both extracellular and intracellular in origin^{37,42}. TCTP has recently been identified and cloned from venom or saliva libraries of a few phylogenetically widespread parasitic and venomous organisms including the lamprey *Lampetra japonica*⁴³, ixodid ticks⁴⁴ and the tarantula *Grammostola rosea*⁴⁵. Thus, it is likely that activity as a histamine releasing factor plays a role in antagonistic processes, but its primary role as a venom component in prey capture is unknown.

We identified a single cDNA (SCY1199) with high similarity to “**venom allergen/antigen 5**” with top hits to sequences from the spiders *G. rosea* ($e10^{-27}$), *L. laeta* ($e10^{-25}$) and the tick *Isonomus maculatus* ($e10^{-20}$). Venom allergens are thought to cause sensitization, inducing immune responses at the cellular level²⁰, and have been identified as potential toxins in transcriptomic¹⁸ and proteomic¹⁹ analyses of venom glands of *L. laeta*, and *L. intermedia*²⁰. Duan *et al.*³⁶ detected a “venom allergen antigen 5-like” protein in a proteomic analysis of *Latrodectus tredecimguttatus* crude venom. We also detected the SCY1199 peptide sequence in the *Scytodes* proteome using MudPIT (Fig. 2). Together, this is strong evidence that the venom allergen is secreted as a venom component and may be phylogenetically widespread, but it may not be an abundantly transcribed toxic venom protein in *S. thoracica*.

Finally, a noteworthy protein was represented by a cluster of three sequences with strong homology to “angiotensin converting enzyme-like protein” ($e10^{-68}$). We detected this protein in the proteome (Fig. 2B) and it also has strong matches with *Sicarius* ($e10^{-108}$) and *Loxosceles* ($e10^{-73}$) sequences in our Haplogyne database. It has homologs in current databases in various species including ticks, but no other spiders outside of Haplogynes. While we find no direct evidence in the literature confirming this protein family as an active venom component in arachnids, the fact that it is a homolog of vasoconstrictive enzymes is provocative, and may prove interesting with more detailed analyses of function.

Candidate Venom Peptides

Of the total candidate venom toxin sequences, 138 translate into small peptides (< 10 kDa) and contain posttranslational processing sites for signal and mature cleavage. The mature peptide sequences are rich in cysteine residues and have characteristic patterns for forming disulfide bonds to fold into compact configurations similar to known venom peptide

neurotoxins. The sizes of the predicted mature peptides range from ~3–9 kDa (Table 1). These peptides constitute 35% of the proteome (Fig. 1C). We were conservative in inclusion of peptides in this category and likely exclude some that are currently placed in the “unknown” category. Three sequences – SCV214, SCV259, and SCV318 – that were not full length or had an unresolved amino acid because of sequence error were left in the unknowns (Supplemental Table 3).

Scytotoxins (SYTX) is the reference nomenclature suggested for *Scytodes* toxins⁴⁶. Based on amino acid sequence alignments, we assigned 19 groups of peptide sequences “unknown” functional notations (U₁–U₁₉) and a “SYTX-Sth” name. Each sequence in a group was given a number (1) and an ascending letter starting with “a” if more than one different sequence was in that group (i.e., 1a, 1b, 1c, etc.) (Table 1, Fig. 5). These higher-level designations (U₁–U₁₉) are based on sequences having sufficient divergence that they likely differ in biological function and/or molecular targets⁴⁶. We predict our delineations are conservative and these designations will be refined as function is determined. Scytotoxins suffixed with an underscore followed by a number are those for which the signal sequence is the only region where the sequences vary within a particular group (Fig. 2B). In cases where cysteine patterns were identical, we combined alignments of clusters and assessed our ability to impose alignments with confidence. In one case (U₁₉-SYTX) we merged three clusters and three singletons that were easily alignable, however groups U₁₁ and U₁₂, while clearly homologous, were sufficiently divergent (mean distance between groups 82±3.3% aa difference) that we retain them as separate entities. Proteomics confirms the presence in crude venom of 23 different transcripts that represent 13 out of the 19 groups that we call venom peptides (Fig. 2B).

As illustrated above, interpreting patterns of homology among candidate venom peptides can be difficult because short cysteine-rich sequences with high variability between cysteines lend themselves to convergence, and standard e-values for homology thresholds may be missing homologous hits and hitting sequences that are sufficiently divergent to easily align. With this caveat, of the 19 distinct categories of candidate peptide toxins, three have significant similarity with non-spider arachnids, U₁₁-SYTX and U₁₂-SYTX with ticks, and U₁₃-SYTX with scorpions (Table 1). U₁₂-SYTX and U₁₃-SYTX are found in the *Scytodes* venom proteome and should be considered candidates for widespread peptides, repeatedly recruited for venom expression among arachnids. Two of our peptide toxin categories (U₃-SYTX and U₉-SYTX) hit mygalomorph sequences, but no non-spider arachnids with high similarity, suggesting these are phylogenetically widespread within spiders with sufficient conservation and/or convergence for homology detection. Three other sets hit only peptides from other araneomorphs (U₅-, U₁₂- and U₁₉-SYTX) (Table 1).

Only one of the six sequences from arachnid venom that had significant similarity to *Scytodes* candidate venom peptides has a defined neurotoxic function (μ-ctenitoxin-Pn1a)⁴⁷. Sequences from U₉-SYTX and U₁₉-SYTX both hit μ-ctenitoxin-Pn1a from *Phoneutria nigriventer*, a voltage-gated sodium channel (Na_v1.2) blocking peptide⁴⁸. However U₉-SYTX also hit U₂-theraphotoxin-Pc1a from *Psalmopoeus cambridgei* with equally high e-values (e10⁻⁵). Despite the common highest hit in databases, these two groups are sufficiently different to not be easily aligned to one another (Fig. 5B and 5D), and so

different from μ -ctenitoxin-Pn1a that we do not argue that the homologous *Scytodes* peptides share the same function.

Three Scytotoxin groups had significant hits to unannotated sequences in NCBI's EST database. The sequences in U₁₂-SYTX share significant similarity with arachnid sequences in both ArachnoServer and EST databases, including *I. scapularis* (tick) sequences ($e10^{-10}$) and *Dysdera erythrina* (CV178049) ($e10^{-6}$) and *L. laeta* (EY189664) ($e10^{-6}$) transcripts from venom gland libraries. The U₁₁-SYTX sequence has significant similarity with a diverse set of arthropod ESTs including *Daphnia pulex* (water flea) ($e10^{-8}$) and *Rhipicephalus microplus* (tick) ($e10^{-10}$) and a secreted protein from *I. scapularis* ($e10^{-7}$). U₁₁-SYTX-Sth1a also hit several serine protease precursors (HTRA3) ($e10^{-8}$), but the hits are large proteins (490 amino acids) and the Scytotoxin sequence aligns only with the N-terminus. The U₁₃-SYTX alignment includes a diverse set of sequences and hit sequences from *L. laeta* and *intermedia* venom gland ESTs ($e10^{-10}$), hypothetical proteins from *Latrodectus hesperus* and *Amblyomma maculatum* (tick), and TXLP4 protein from *Mesobuthus martensii* ($e10^{-5}$) (Fig. 5C). All of our putative toxins hit sequences from other haplogynes in our unpublished datasets except U₁-SYTX-Sth1b and U₂-SYTX-Sth1a. Finally, U₁-SYTX-Sth1a, U₃-SYTX-Sth1a-1e, U₄-, U₆-, U₇-, U₈-, U₁₄- and U₁₇-SYTX, U₁₀-SYTX-Sth1b-1d and U₁₉-SYTX-Sth1b-1e hit sequences only within sicariids, the most closely related family to *Scytodes* included in our broad comparative analyses (Table 1).

The cysteine skeleton patterns of the small peptides in groups U₁-, U₃₋₁₂-, U₁₆- and U₁₈-SYTX (-C-C-CC-C-C-) are characteristic of knottin or inhibitor cysteine knot (ICK) peptides, including the six amino acid linker between the 1st and 2nd cysteines of the mature peptide⁴⁹; however, only sequences in groups U₁-, U₅-, U₆- and U₉-SYTX are predicted to be knottins by the Knoter1d program (cysteine skeleton pattern, Table 1 and Fig. 5B). Knottin or ICK peptides are abundant in spider venoms and many have been identified and characterized in recent years. The compact structure of these peptides provides stability and proteolytic resistance. Their high affinity and nuanced selectivity for manipulating ion channels makes them attractive targets for discovery of toxins with potential utility as drugs or insecticides^{9,10,50}. The LiTx peptides 1–4 from *L. intermedia*⁵¹ are ICK peptides. Matsubara *et al.*⁵² have cloned and expressed LiTx3 (U₂-SCRTX-Li1b) and showed evidence of immunological cross-reactivity against *Loxosceles* crude venoms. Other homologs of LiTx have been identified in venom gland transcriptomes of *L. laeta*¹⁸, *L. intermedia*²⁰, and we have detected them in *S. dolichocephalus*, *L. spinulosa*, *L. rufescens*, *L. reclusa*, and *L. arizonica* (Binford *et al.*, unpublished) suggesting widespread expression in sicariids at least. We have identified several ICK venom gland transcripts in all of our venom gland Haplogyne transcriptomes that are predicted to be knottins; in addition to the sicariid libraries mentioned above, we have identified several in pseudoscorpion (*Synsphronus apimelus*), archaeid (*A. mainae*), pholcid (*P. mexicanus*) and plectreuid (*P. tristis*) cDNA libraries (Binford *et al.*, unpublished). Loening *et al.*⁵³ recently solved the first NMR structures of two ICK peptides from *S. dolichocephalus*. The clones from U₆-SYTX share up to 67% amino acid identity with one of these.

U₈-SYTX peptides have an amphipathic stretch of 17 amino acids following the last cysteine residue in the mature sequence, which are predicted to form an alpha helix (Table 1). Recent

work by Kuhn-Nentwig *et al.*⁵⁴ shows that a toxin (CsTx-1) expressed by *Cupiennius salei*, a venom peptide with similar properties, acts alone to carry out dual functions. CsTx-1 has an ICK structure and also exhibits cytolytic activity through the alpha helix at its C-terminus^{54,55}. The alpha helix attaches to the phospholipid membrane and acts as an anchor so the ICK peptide can carry out its function, which is to inhibit Ca²⁺-channels. Kuzmenkov *et al.*⁵⁶ observed similar sequence patterns and functional characteristics by latartoxins in the venom of the “ant spider” *Lachesana tarabaevi*; these highly expressed peptides were also shown to be both cytolytic and neurotoxic. Sequence characteristics of Ug-SYTX make them excellent candidates of toxins with comparable function.

Other Proteins

Thirteen percent of the transcriptome and 60% of the proteome are classified as “unknown proteins” (Fig. 1C, Fig. 2A, Supplemental Table 3), including sequences that have open reading frames. Some in this category confidently hit cDNA sequences in NCBI nr or EST databases ($e^{-10^{-5}}$) that are annotated as “hypothetical proteins” or “proteins of unknown function”. One particularly notable cluster of two sequences hits only “conserved hypothetical proteins” and is found in the proteome. These sequences are cysteine-rich (12% cysteine), >300 aa with predicted MW of 33 kDa after signal peptide cleavage, and slightly acidic with a pI of 5.7. They have significant matches (up to $e^{-10^{-28}}$) to unannotated hypothetical proteins from a broad set of insect pests, but no hits within arachnids, including our Haplogyne venom gland cDNA library database. These are included in a phylogeny (Fig. 6) that pairs the *Scytodes* sequences with a sequence isolated from *Aedes aegypti*.

The PXCC gene family—One of the most highly expressed gene families makes up 51% of the *Scytodes* proteome. This apparently undescribed gene family contains 35 nonredundant sequences from our dataset, 15 of which are full-length. These are small peptides, ~50 aa after signal cleavage, and have seven cysteine residues which implies that they may fold into compact tertiary structures (Fig. 7A). The PXCC motif is 100% conserved throughout all sequences, thus we propose the name “PXCC gene family”. Transcripts can be combined into one alignment that includes 64 sequences from ten clusters and two singletons. We anchored the alignment with their highly conserved cysteine pattern that is not unlike those found in typical venom peptides. We separated them from the venom peptide group because only four are predicted to have a propeptide recognition sequence characteristic of these small potential toxins, and because the strongest hits in databases of some were to cDNAs from a *L. hesperus* silk gland cDNA library, to proteins with unknown function from several insects, and to a venom gland cDNA sequence from *Sicarius dolichocephalus* ($e^{-10^{-8}}$ to $e^{-10^{-5}}$) (Supplemental Table 3). Although the *L. hesperus* clone (AV24) was isolated from a silk gland library, we do not assume that homology with this sequence defines a silk-like function of these peptides in *Scytodes* venom; however, their high levels of expression suggest they play an important role in the feeding biology of these spiders. Phylogenetic analysis of nonredundant full-length sequences from this set resolves six distinct clades with poor resolution of relationships among them (Fig. 7B). However, clades III–VI are resolved as monophyletic and more closely related to the AV24 clone from *L. hesperus* than they are to *Scytodes* sequences in clades I & II (Fig. 7B). Remarkably,

peptides from all six clades are represented in the proteome (filled circles, Fig. 7B, Fig. 2B) with clade V being the most abundant (188 peptides), followed by clade III (99 peptides).

The PXCC motif has been found in small proteins from several *Drosophila* species, insects (*Aedes*) and a scorpion (*Buthus*) by Sheldon *et al.*⁵⁷. They were named as members of the “SVC protein family” – single VWC domain-like proteins, where VWC stands for the von Willebrand factor C-domain, a component of the multi-subunit molecule hemolymph. These proteins, however, do not resemble VWC in that they are small, secreted single polypeptide proteins that contain eight cysteine residues instead of ten that VWC contains⁵⁷, and VWC domains do not have the highly conserved PXCC motif at the C-terminus⁵⁸. The SVC protein sequences were identified using a Pfam search and were aligned using the C-terminal PXCC motif as an anchor⁵⁷. None of these sequences have significant homology with *Scytodes* sequences or sequences in our Haploglyne database ($e > 10^{-4}$), so they are not included in Figure 7, however, their abundance in both the proteome and transcriptome are strong evidence of recruitment for venom function.

The RP23-like cluster—Two sequences form a distinct cluster of proteins (the RP23-like cluster) (>200 aa) that hit a second sequence cloned from the same *L. hesperus* silk gland library mentioned above; *L. hesperus* clone RP23 ($e10^{-29}$) aligns with our sequences along with a sequence from *Rhipicephalus pulchellus* (tick) and EST sequences from *D. erythrina* (Fig. 8). No homologues were found in our Haploglyne database with $e > 10^{-5}$. The RP23-like protein is predicted to contain a signal recognition sequence and 31 residues are 100% conserved across these lineages (Fig. 8). It was also found as 3% of the venom proteome (Fig. 2B).

Candidate contractile/adhesive proteins and peptides

The Glycine-rich Gene Family—Almost half of our 1,142 sequences (483) are small (3.5–7 kDa) glycine-rich peptides that were detected in the proteome and have characteristics that we consider sufficient to categorize as having potential function in the contraction and adhesion properties of *Scytodes* spit (Fig. 1A). All peptides in this group are distinct from any sequences in databases including our own. These peptides group into six clusters that are similar enough to combine into a single alignment that we used to analyze phylogenetic relationships (Fig. 9A). This gene family (the “glycine-rich” gene family) is represented in our dataset by 68 nonredundant sequences that are resolved into at least six distinct clades (Fig. 9B). All contain a signal peptide sequence, which suggests that the peptides are processed for secretion (Fig. 9A).

We chose a clone from each branch of the tree (labeled with * in Fig. 9B) and used ProtParam (<http://web.expasy.org/protparam>) to calculate the amino acid composition of each mature peptide. Glycine is the most prevalent amino acid in all of the sequences, 15–34%, except in the largest branch (clade II), where it is the 2nd most frequent amino acid following glutamine (17%). Leucine and proline are next in abundance in all clades, making up 10–15% and 8–11%, respectively among the peptides. When we combine the glycine and alanine contents of each peptide, the percentage range increases from 15–34% to 19–40% across the family. Certain silk proteins are rich in glycine and alanine and serine content, but

are usually much larger and contain multiple iterations of varied amino acid sequence repeats^{23–25,59,60}. Maruyama *et al.*²⁶ showed that heterogeneous glycine-rich proteins are abundant in ticks, that are similar to spider silk proteins and have adhesive and tensile properties that aid in attachment for prolonged feeding²⁶. The glycine-rich peptides we have found each contain at least two GGX motifs that are commonly found in elastic silk proteins^{61,62}, however, GGX motifs in silks are typically consecutive which is not the case in the *Scytodes* peptides (underline, Fig. 9A). Shrinkage of silk fibers up to 50% due to water absorption which disrupts hydrogen bonds (supercontraction) has been observed in silks, including those containing GGX motifs, from various spider species, ultimately affecting the silk's extensibility⁶¹. A similar type of shrinkage was measured from *Scytodes* spit by Suter and Stratton¹⁶. These characteristics make this set of small peptides good candidates for contributing to contraction/adhesion properties of *Scytodes* spit.

Sequences with diglutamine (QQ) and dityrosine (YY) motifs—We identified a cluster of three sequences in the transcriptome that contains three GGX, three diglutamine (QQ) and three dityrosine (YY) motifs (Fig. 10). They are identical and encode a 118 aa protein (12.8 kDa) that is predicted to have a signal sequence which yields a 97 aa (10.6 kDa) mature protein. These do not hit anything in databases with significant similarity, including our Haploglyne database. Diglutamines (QQ) are found embedded in repetitive motifs of tick glycine-rich proteins (GRPs)²⁶, silk proteins such as MaSp2²³, piriform silk proteins⁶³, and in the elastic protein gluten from wheat grain^{64,65}. The side chains of glutamine residues have been shown to interact with each other to form β -sheets from interchain hydrogen bonding networks in gluten^{64,65}. Dityrosine (YY) crosslinking, resulting from tyrosine-tyrosine side chain interactions, has been reported in *B. mori* silk⁶⁶ and is suggested to drive the intermolecular cross-linking among silk aggregate proteins in *L. hesperus*⁶². Vasanthavada *et al.*⁶² recently reported that some components of the protein mixture extruded from the aggregate gland during web construction can self-assemble into fibrous structures; AgSF1 is large (1223 aa) and is rich in GGX motifs and YY residues, the latter of which is alluded to be involved in intermolecular interactions that may allow for aggregation. Half of the glycine-rich peptides in the glycine-rich gene family contain QQ and YY motifs at the C-terminus (clades I and II, Fig. 9).

Incidentally, we had a difficult time completely dissolving our venom/spit mixture for proteomic analysis; after days (and up to three weeks) of sitting in a buffer used to dissolve venom, we observed a considerable amount of protein that remained in the stacking region of a polyacrylamide gel (Supplemental Figure 1), suggesting that something in this mixture remained aggregated. We detected a much smaller amount of candidate contractile/adhesive peptides (i.e., Glycine-rich gene family (Fig. 9) and no sequences with dityrosine motifs (Fig. 10)) in the proteome than expected (0.1%), given their abundance in the transcriptome (42%) and suspect that these peptides are far more abundant in the proteome than suggested by our analyses. Perhaps the molecular interactions are so strong that a conventional buffer is not sufficient for dissolving them and a more directed effort to break up these aggregates would allow us to analyze them apart from all the other components in the mix.

Unknown open reading frames (ORFs)

We have 68 remaining entities from Figure 1B composed of four clusters of sequences and 64 singleton sequences that contain protein-coding regions (>30 aa) which vary in size, secretion probability and what they hit – or don't – in the databases (Supplementary Table 3). Although most were not detected in the proteome analyses, their presence in the transcriptome suggests a venom function that is yet undescribed (Fig. 1B, Fig. 2B). The sequences that hit a “hypothetical protein” or a “conserved hypothetical protein” with $e < 10^{-5}$ retain that label, whereas those that didn't hit any sequences in public databases are labeled “unknown ORF” (Supplemental Table 3). Of the 23 entities that are homologous to “(conserved) hypothetical proteins”, eight hit sequences in our Haplogyne database. A closer look at the remaining 42 unknown ORF sequences shows that 11 hit sequences in our Haplogyne database, suggesting a broad phylogenetic set of Haplogyne taxa express these proteins (Supplemental Table 3). It will be interesting to investigate the function of the remaining 32 protein sequences, particularly since they have no close matches and may have evolved for a functional role in the feeding behavior of *Scytodes*.

Scytodes venom/spit has highly divergent elements

The diverse set of peptides and proteins that make up the venom of *Scytodes* includes some components that appear to be widespread across spider venoms, but a very interesting subset is distinct from sequences in published databases. In general, peptides and proteins that are candidates for functioning in prey immobilization and classic venomous function have detectable homologs outside of *Scytodes*. The exception is the PXCC gene family, which has several features consistent with toxins (cysteine rich) but is only found in *Scytodes*. However 42% are unique peptides that have features that make us suspect they contribute to adhesion and contraction and are distinct from anything in databases.

To elaborate on the phylogenetic distribution of venom proteins and peptides, we more easily detect widespread homology among candidate venom proteins than we do with venom peptides. For example, we detected homologs of “longistatins”, “TCTPs” and “venom allergen” in tarantulas. Tarantulas are mygalomorphs, and descendants of the most recent common ancestor of tarantulas and *Scytodes* include nearly all spiders. We predict these will be commonly found in spider venoms with more thorough sampling across the breadth of spider diversity. Notably, all *Scytodes* candidate venom proteins except astacins had hits with Acari, and some hit scorpion venoms. Unless we are mistaking ubiquitous housekeeping genes with venom proteins, this pattern is best explained by independent recruitment of these gene families for venomous function.

Among venom proteins, it is worth noting well-known toxins that are missing. *Scytodes* is a close relative of sicariids, including *Loxosceles* and *Sicarius* spiders whose venoms are toxic to mammals (*Loxosceles* reviews^{67,68}; *Sicarius*^{69,70}). We searched the *Scytodes* venom gland transcripts specifically for sequences with homology to sphingomyelinase D (SMase D), a highly abundant 30–35 kDa venom protein that is neurotoxic to insects⁷¹ and causes cell death *in vivo* in mammals^{38,39,72–81}. Sicariid venoms include multiple paralogs of the SMase D gene family (*SicTox*)^{76,82–90}. While the SMase D enzyme plays an important role in prey capture in sicariids, we have not found it in the venom transcriptome or proteome of

Scytodes or any other haplogyne venom gland transcriptome so far. While missing toxins may be false negatives (not captured with our methods), the fact that we found homologs of metabolic proteins such as actin, histones and ribosomal proteins in *Scytodes* and our other haplogyne venom gland cDNAs suggests our transcriptome was capturing a broad set of large proteins. The absence in the proteome indicates that the absence in the transcriptome was not likely a matter of transcriptional timing of tissue extraction. Thus, we interpret this lack of detection of *SicTox* homologs in *Scytodes* as further evidence that this toxin was uniquely recruited for venom function in the most recent common ancestor of sicariids^{21,90}.

While we detect the above patterns of phylogenetic breadth of components within *Scytodes* venom/spit, 638 sequences are distinct to this transcriptome, 33 of which are singletons in clustering analyses at 85%, indicating that the unusual feeding biology of *Scytodes* involves a distinctive set of venom toxins and unique peptides. In particular, the notably divergent and abundant glycine rich and diglutamine/dityrosine-containing peptides (483 sequences) are unlike anything we are aware of in databases, and are good candidates for contributing to the contractile/adhesive properties of *Scytodes* glue. Functional testing may reveal that they have high biomimetic potential.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding: NIH R15-GM-097696-01 and the Biology Department at Lewis & Clark College (GB), seed grant and start up funds from UMass Lowell (JG), a Grant in Aid of Research from Sigma Xi and from the Society for Integrative and Comparative Biology (SC). Andrew Wood helped collect spiders with GB and PZT. Gail Stratton, Bob Suter provided spiders to JG. We thank Linda Brei and Cynthia David at the Proteomics Core at the University of Arizona for help with designing proteomic experiments and analysis.

References

1. Platnick, NI. The world spider catalog, version 14.0. <http://research.amnh.org/iz/spiders/catalog/>
2. Rash LD, Hodgson WC. Pharmacology and biochemistry of spider venoms. *Toxicon*. 2002; 40:225–254. [PubMed: 11711120]
3. Escoubas P, Rash L. Tarantulas: eight-legged pharmacists and combinatorial chemists. *Toxicon*. 2004; 43:555–574. [PubMed: 15066413]
4. Tedford HW, Sollod BL, Maggio F, King GF. Australian funnel-web spiders: master insecticide chemists. *Toxicon*. 2004; 43:601–618. [PubMed: 15066416]
5. Sollod BL, Wilson D, Zhaxybayeva O, Gogarten JP, Drinkwater R, King GF. Were spiders the first to use combinatorial peptide libraries? *Peptides*. 2005; 26:131–139. [PubMed: 15626513]
6. Escoubas P, Sollod B, King GF. Venom landscapes: Mining the complexity of spider venoms via a combined cDNA and mass spectrometric approach. *Toxicon*. 2006; 47(6):650–663. [PubMed: 16574177]
7. Escoubas P, Quinton L, Nicholson GM. Venomics: unravelling the complexity of animal venoms with mass spectrometry. *J Mass Spectrom*. 2008; 43(3):279–295. [PubMed: 18302316]
8. Escoubas P, King GF. Venomics as a drug discovery platform. *Exp Rev Proteomics*. 2009; 6(3):221–224.
9. Vassilevski AA, Kozlov SA, Grishin EV. Molecular diversity of spider venom. *Biochemistry*. 2009; 48(13):1505–1534. [PubMed: 20210706]

10. King GF. Venoms as a platform for human drugs: translating toxins into therapeutics. *Expert Opin Biol Ther.* 2011; 11(11):1469–1484. [PubMed: 21939428]
11. Kuhn-Nentwig, L, Stocklin, R, Nentwig, W. Venom composition and strategies in spiders: is everything possible?. In: Casas, J, editor. *Adv Ins Physiol.* Vol. 40. 2011. 1–86.
12. Windley MJ, Herzig V, Dziemborowicz SA, Hardy MC, King GF, Nicholson GM. Spider-Venom Peptides as Bioinsecticides. *Toxins.* 2012; 4(3):191–227. [PubMed: 22741062]
13. King, GF, Hardy, MC. Spider-Venom peptides: structure, pharmacology, and potential for control of insect pests. In: Berenbaum, MR, editor. *Annu Rev Entomol.* Vol. 58. 2013. 475–496.
14. Foelix, RF. *Biology of Spiders.* 2. Oxford University Press; New York: 1996.
15. Suter RB, Stratton GE. *Scytodes* vs. *Schizocosa*: predatory techniques and their morphological correlates. *J Arachnol.* 2005; 33:7–15.
16. Suter RB, Stratton GE. Spitting performance parameters and their biomechanical implications in the spitting spider, *Scytodes thoracica*. *J Insect Sci.* 2009; 9
17. Clements R, Li DQ. Regulation and non-toxicity of the spit from the pale spitting spider *Scytodes pallida* (Araneae : Scytodidae). *Ethology.* 2005; 111(3):311–321.
18. Fernandes-Pedrosa, MdF; Junqueira-de-Azevedo, IdLM; Gonçalves-de-Andrade, RM; Kobashi, LS; Almeida, DD; Ho, PL; Tambourgi, DV. Transcriptome analysis of *Loxosceles laeta* (Araneae, Sicariidae) spider venomous gland using expressed sequence tags. *BMC Genomics.* 2008; 9(1): 279. [PubMed: 18547439]
19. dos Santos LD, Dias NB, Pinto JRAS, Palma MS. Brown recluse spider venom: proteomic analysis and proposal of a putative mechanism of action. *Prot Pept Lett.* 2009; 16:933–943.
20. Gremski LH, da Silveira RB, Chaim OM, Probst CM, Ferrer VP, Nowatzki J, Weinschutz HC, Madeira HM, Gremski W, Nader HB, Senff-Ribeiro A, Veiga SS. A novel expression profile of the *Loxosceles intermedia* spider venomous gland revealed by transcriptome analysis. *Molecular Biosystems.* 2010; 6:2403–2416. [PubMed: 20644878]
21. 9Binford GJ, Wells MA. The phylogenetic distribution of sphingomyelinase D activity in venoms of Haplogyne spiders. *Comp Biochem Physiol, Part B: Biochem Mol Biol.* 2003; 135(1):25–33.
22. 8Binford GJ, Cordes MHJ, Wells MA. Sphingomyelinase D from venoms of *Loxosceles* spiders: evolutionary insights from cDNA sequences and gene structure. *Toxicon.* 2005; 45(5):547–560. [PubMed: 15777950]
23. Ayoub NA, Garb JE, Tinghitella RM, Collin MA, Hayashi CY. Blueprint for a high-performance biomaterial: full-length spider dragline silk genes. *PLoS ONE.* 2007; 2(6):e514–526. [PubMed: 17565367]
24. Römer L, Scheibel T. The elaborate structure of spider silk. *Prion.* 2008; 2(4):154–161. [PubMed: 19221522]
25. Blasingame E, Tuton-Blasingame T, Larkin L, Falick AM, Zhao L, Fong J, Vaidyanathan V, Visperas A, Geurts P, Hu X, La Mattina C, Vierra C. Pyriform spidroin 1, a novel member of the silk gene family that anchors dragline silk fibers in attachment discs of the black widow spider, *Latrodectus hesperus*. *J Biol Chem.* 2009; 284(42):29097–29108. [PubMed: 19666476]
26. Maruyama SR, Anatriello E, Anderson JM, Ribeiro JM, Brandão LG, Valenzuela JG, Ferreira BR, Garcia GR, Szabó MPJ, Patel S, Bishop R, de Miranda-Santos IKF. The expression of genes coding for distinct types of glycine-rich proteins varies according to the biology of three metastriate ticks, *Rhipicephalus* (Boophilus) *microplus*, *Rhipicephalus sanguineus* and *Amblyomma cajannense*. *BMC Genomics.* 2010; 11:363–379. [PubMed: 20529354]
27. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* 2011; 28(10):2731–2739. [PubMed: 21546353]
28. Ronquist F, Huelsenbeck JP. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics.* 2003; 19:1572–1574. [PubMed: 12912839]
29. Ronquist, F; Huelsenbeck, JP; van der Mark, P. MrBayes 3.2 manual. 2005. <http://mrbayes.sourceforge.net/manual.php>
30. Hu Q, Noll RJ, Li H, Makarov A, Hardman M, Graham Cooks R. The Orbitrap: a new mass spectrometer. *J Mass Spectrom.* 2005; 40(4):430–443. [PubMed: 15838939]

31. Paoletti AC, Zybaylov B, Washburn MP. Principles and applications of multidimensional protein identification technology. *Exp Rev Proteomics*. 2004; 1(3):275–282.
32. Anisuzzaman, Islam MK, Alim MA, Miyoshi T, Hatta T, Yamaji K, Matsumoto Y, Fujisaki K, Tsuji N. Longistatin, a plasminogen activator, is key to the availability of blood-meals for ixodid ticks. *PLoS Pathogens*. 2011; 7(3):e1001312. [PubMed: 21423674]
33. Anisuzzaman, Islam MK, Alim MA, Miyoshi T, Hatta T, Yamaji K, Matsumoto Y, Fujisaki K, Tsuji N. Longistatin is an unconventional serine protease and induces protective immunity against tick infestation. *Mol Biochem Parasitol*. 2012; 182(1–2):45–53. [PubMed: 22206819]
34. da Silveira RB, Wille ACM, Chaim OM, Appel MH, Silva DT, Franco CRC, Toma L, Mangili OC, Gremski W, Dietrich CP, Nader HB, Veiga SS. Identification, cloning, expression and functional characterization of an astacin-like metalloprotease toxin from *Loxosceles intermedia* (brown spider) venom. *Biochemical Journal*. 2007; 406:355–363. [PubMed: 17535156]
35. Sade YB, Bóia-Ferreira M, Gremski LH, da Silveira RB, Gremski W, Senff-Ribeiro A, Chaim OM, Veiga SS. Molecular cloning, heterologous expression and functional characterization of a novel translationally-controlled tumor protein (TCTP) family member from *Loxosceles intermedia* (brown spider) venom. *Int J Biochem Cell Biol*. 2012; 44:170–177. [PubMed: 22051631]
36. Duan Z, Yan X, Cao R, Liu Z, Wang X, Liang S. Proteomic analysis of *Latrodectus tredecimguttatus* venom for uncovering potential latrodectism-related proteins. *J Biochem Mol Toxicol*. 2008; 22(5):328–336. [PubMed: 18972397]
37. Anisuzzaman, Islam MK, Miyoshi T, Alim MD, Hatta T, Yamaji K, Matsumoto Y, Fujisaki K, Tsuji N. Longistatin, a novel EF-hand protein from the ixodid tick *Haemaphysalis longicornis*, is required for acquisition of host blood-meals. *Int J Parasitol*. 2010; 40:721–729. [PubMed: 19968997]
38. Feitosa L, Gremski W, Veiga SS, Elias M, Graner E, Mangili OC, Brentani RR. Detection and characterization of metalloproteinases with gelatinolytic, fibronectinolytic and fibrinogenolytic activities in Brown spider (*Loxosceles intermedia*) venom. *Toxicon*. 1998; 36(7):1039–1051. [PubMed: 9690796]
39. Veiga SS, Feitosa L, dos Santos VL, de Souza GA, Ribeiro AS, Mangili OC, Porcionatto MA, Nader HB, Dietrich CP, Brentani RR, Gremski W. Effect of brown spider venom on basement membrane structures. *Histochemical J*. 2000; 32:397–408.
40. da Silveira RB, dos Santos Filho JF, Mangili OC, Veiga SS, Gremski W, Nader HB, von Dietrich CP. Identification of proteases in the extract of venom glands from brown spiders. *Toxicon*. 2002; 40:815–822. [PubMed: 12175619]
41. Guevara T, Yiallourous I, Kappelhoff R, Bissdorf S, Stocker W, Gomis-Ruth FX. Proenzyme structure and activation of astacin metallopeptidase. *J Biol Chem*. 2010; 285(18):13958–13965. [PubMed: 20202938]
42. Chaim OM, Trevisan-Silva D, Chaves-Moreira D, Wille ACM, Ferrer VP, Matsubara FH, Mangili OC, Silveira RBd, Gremski LH, Gremski W, Senff-Ribeiro A, Veiga SS. Brown spider (*Loxosceles* genus) venom toxins: tools for biological purposes. *Toxins*. 2011; 3(12):309–344. [PubMed: 22069711]
43. Sun J, Wu Y, Wang J, Ma F, Liu X, Li Q. Novel translationally controlled tumor protein homologue in the buccal gland secretion of *Lampetra japonica*. *Biochimie*. 2008; 90:1760–1768. [PubMed: 18822340]
44. Mulenga A, Azad AF. The molecular and biological analysis of ixodid ticks histamine release factors. *Exp Appl Acarol*. 2005; 37:215–229. [PubMed: 16323052]
45. Kimura T, Ono S, Kubo T. Molecular cloning and sequence analysis of the cDNAs encoding toxin-like peptides from the venom glands of tarantula *Grammostola rosea*. *Int J Peptides*. 2012; doi: 10.1155/2012/731293
46. King GF, Gentz MC, Escoubas P, Nicholson GM. A rational nomenclature for naming peptide toxins from spiders and other venomous animals. *Toxicon*. 2008; 52(2):264–276. [PubMed: 18619481]
47. Martin-Moutot N, Mansuelle P, Alcaraz G, dos Santos RG, Cordeiro MN, de Lima ME, Seagar M, Van Renterghem C. *Phoneutria nigriventer* Toxin 1: A novel, state-dependent inhibitor of neuronal

- sodium channels that interacts with μ Conotoxin binding sites. *Mol Pharmacol*. 2006; 69:1931–1937. [PubMed: 16505156]
48. Diniz MRV, Theakston RDG, Crampton JM, Cordeiro MD, Pimenta AMC, De Lima ME, Diniz CR. Functional expression and purification of recombinant Tx1, a sodium channel blocker neurotoxin from the venom of the Brazilian “armed” spider, *Phoneutria nigriventer*. *Protein Expression Purif*. 2006; 50(1):18–24.
 49. Kozlov S, Malyavka A, McCutchen B, Lu A, Schepers E, Herrmann R, Grishin E. A novel strategy for the identification of toxinlike structures in spider venom. *Proteins*. 2005; 59:131–140. [PubMed: 15688451]
 50. Saez NJ, Senff S, Jensen JE, Er SY, Herzig V, Rash LD, King GF. Spider-Venom Peptides as Therapeutics. *Toxins*. 2010; 2(12):2851–2871. [PubMed: 22069579]
 51. de Castro CS, Silvestre FG, Araújo SC, Yazbeck GDM, Mangili OC, Cruz I, Chávez-Olortegui C, Kalapothakis E. Identification and molecular cloning of insecticidal toxins from the venom of the brown spider *Loxosceles intermedia*. *Toxicon*. 2004; 44(3):273–280. [PubMed: 15302533]
 52. Matsubara FH, Gremski LH, Meissner GO, Constantino Lopes ES, Gremski W, Senff-Ribeiro A, Chaim OM, Veiga SS. A novel ICK peptide from the *Loxosceles intermedia* (brown spider) venom gland: Cloning, heterologous expression and immunological cross-reactivity approaches. *Toxicon*. 2013; 71:147–58. [PubMed: 23751278]
 53. Loening NM, Wilson ZN, Zobel-Thropp PA, Binford GJ. Solution structures of two homologous venom peptides from *Sicarius dolichocephalus*. *PLoS ONE*. 2013; 8(1)
 54. Kuhn-Nentwig L, Fedorova IM, Lüscher BP, Kopp LS, Trachsel C, Schaller J, Vu XL, Seebeck T, Streitberger K, Nentwig W, Sigel E, Magazanik LG. A venom-derived neurotoxin, CsTx-1, from the spider *Cupiennius salei* exhibits cytolytic activities. *J Biol Chem*. 2012; 287(30):25640–25649. [PubMed: 22613721]
 55. Trachsel C, Siegemund D, Kämpfer U, Kopp LS, Bühr C, Grossmann J, Lüthi C, Cunningham M, Nentwig W, Kuhn-Nentwig L, Schürch S, Schaller J. Multicomponent venom of the spider *Cupiennius salei*: a bioanalytical investigation applying different strategies. *FEBS J*. 2012; 279:2683–2694. [PubMed: 22672445]
 56. Kuzmenkov AI, Federova IM, Vassilevski AA, Grishin EV. Cysteine-rich toxins from *Lachesana tarabaei* spider venom with amphiphilic C-terminal segments. *Biochim Biophys Acta*. 2013; 1828:724–731. [PubMed: 23088912]
 57. Sheldon TJ, Miguel-Aliaga I, Gould AP, Taylor WR, Conklin D. A novel family of single VWC-domain proteins in invertebrates. *FEBS Lett*. 2007; 581(27):5268–5274. [PubMed: 18028914]
 58. Zhou YF, Eng ET, Zhu J, Lu C, Walz T, Springer TA. Sequence and structure relationships within von Willebrand factor. *Blood*. 2012; 120(2):449–458. [PubMed: 22490677]
 59. Garb JE, Ayoub NA, Hayashi CY. Untangling spider silk evolution with spidroin terminal domains. *BMC Evolutionary Biology*. 2010; 10:243–258. [PubMed: 20696068]
 60. Sutherland TD, Young JH, Weisman S, Hayashi CY, Merritt DJ. Insect Silk: One Name, Many Materials. *Annu Rev Entomol*. 2010; 55(1):171–188. [PubMed: 19728833]
 61. Boutry C, Blackledge TA. Evolution of supercontraction in spider silk: structure-function relationship from tarantulas to orb-weavers. *J Exp Biol*. 2010; 213:3505–3514. [PubMed: 20889831]
 62. Vasanthavada K, Hu X, Tuton-Blasingame T, Hsia Y, Sampath S, Pacheco R, Freeark J, Falick AM, Tang S, Fong J, Kohler K, La Mattina-Hawkins C, Vierra C. Spider glue proteins have distinct architectures compared with traditional spidroin family members. *J Biol Chem*. 2012; 287(43):35986–35999. [PubMed: 22927444]
 63. Perry DJ, Bittencourt D, Siltberg-Liberles J, Rech EL, Lewis RV. Piriform spider silk sequences reveal unique repetitive elements. *Biomacromolecules*. 2010; 11:3000–3006. [PubMed: 20954740]
 64. Shewry PR, Halford NG, Belton PS, Tatham AS. The structure and properties of gluten: an elastic protein from wheat grain. *Phil Trans R Soc Lond B*. 2002; 357:133–142. [PubMed: 11911770]
 65. Feeney KA, Wellner N, Gilbert SM, Halford NG, Tatham AS, Shewry PR, Belton PS. Molecular structures and interactions of repetitive peptides based on wheat glutenin subunits depend on chain length. *Biopolymers (Biospectroscopy)*. 2003; 72:123–131. [PubMed: 12583015]

66. Chen WQ, Priewalder H, John JPP, Lubec G. Silk cocoon of *Bombyx mori*: proteins and posttranslational modifications - heavy phosphorylation and evidence for lysine-mediated cross links. *Proteomics*. 2010; 10:369–379. [PubMed: 20029844]
67. da Silva PH, da Silveira RB, Helena Appel M, Mangili OC, Gremski W, Veiga SS. Brown spiders and loxoscelism. *Toxicon*. 2004; 44(7):693–709. [PubMed: 15500846]
68. Vetter RS. Spiders of the genus *Loxosceles* (Araneae, Sicariidae): a review of biological, medical and psychological aspects regarding envenomations. *J Arachnol*. 2008; 1:150–163.
69. Newlands G, Atkinson P. Review of southern African spiders of medical importance, with notes on the signs and symptoms of envenomation. *S Afr Med J*. 1988; 73:235–239. [PubMed: 3277297]
70. van Aswegen G, Van Rooyen JM, Van der Nest DG, Veldman FJ, de Villiers TH, Oberholzer G. Venom of a six-eyed crab spider, *Sicarius testaceus* (Purcell, 1908): causes necrotic and haemorrhagic lesions in the rabbit. *Toxicon*. 1997; 35(7):1149–1152. [PubMed: 9248014]
71. Zobel-Thropp PA, Kerins AE, Binford GJ. Sphingomyelinase D in sicariid spider venom is a potent insecticidal toxin. *Toxicon*. 2012; 60(3):265–271. [PubMed: 22561243]
72. Geren CR, Chan TK, Howell DE, Odell GV. Isolation and characterization of toxins from brown recluse spider venom (*Loxosceles reclusa*). *Arch Biochem Biophys*. 1976; 174:90–99. [PubMed: 820263]
73. Kurpiewski G, Forrester LJ, Barrett JT, Campbell BJ. Platelet aggregation and sphingomyelinase D activity of a purified toxin from the venom of *Loxosceles reclusa*. *Biochim Biophys Acta*. 1981; 678:467–476. [PubMed: 6274420]
74. Tambourgi DV, Magnoli FC, van den Berg CW, Morgan BP, de Araujo PS, Alves EW, da Silva WD. Sphingomyelinases in the venom of the spider *Loxosceles intermedia* are responsible for both dermonecrosis and complement-dependent hemolysis. *Biochem Biophys Res Commun*. 1998; 251:366–373. [PubMed: 9790962]
75. Veiga SS, Zanetti VC, Braz A, Mangili OC, Gremski W. Extracellular matrix molecules as targets for brown spider venom toxins. *Braz J Med Biol Res*. 2001; 34:843–850. [PubMed: 11449301]
76. Fernandes-Pedrosa MF, Junqueira-de-Azevedo ILM, Gonçalves de Andrade RM, van den Berg CW, Ramos CRR, Ho PL, Tambourgi DV. Molecular cloning and expression of a functional dermonecrotic and haemolytic factor from *Loxosceles laeta* venom. *Biochem Biophys Res Commun*. 2002; 298:638–645. [PubMed: 12419302]
77. Pretel F, Gonçalves-de-Andrade RM, Magnoli FC, da Silva MER, Ferreira JMC, van den Berg CW, Tambourgi DV. Analysis of the toxic potential of venom from *Loxosceles adalaida*, a Brazilian brown spider from karstic areas. *Toxicon*. 2005; 45(4):449–458. [PubMed: 15733566]
78. Chaim OM, Sade YB, da Silveira RB, Toma L, Kalapothakis E, Chávez-Olórtegui C, Mangili OC, Gremski W, von Dietrich CP, Nader HB, Veiga SS. Brown spider dermonecrotic toxin directly induces nephrotoxicity. *Toxicol Appl Pharmacol*. 2006; 211:64–77. [PubMed: 16005484]
79. Paixão-Cavalcante D, van den Berg CW, Fernandes-Pedrosa MF, de Andrade RMG, Tambourgi DV. Role of matrix metalloproteinases in HaCaT keratinocytes apoptosis induced by *Loxosceles* venom sphingomyelinase D. *J Invest Dermatol*. 2006; 126:61–68. [PubMed: 16417218]
80. Ribeiro ROS, Chaim OM, da Silveira RB, Gremski LH, Sade YB, Paludo KS, Senff-Ribeiro A, de Moura J, Chávez-Olórtegui C, Gremski W, Nader HB, Veiga SS. Biological and structural comparison of recombinant phospholipase D toxins from *Loxosceles intermedia* (brown spider) venom. *Toxicon*. 2007; 50(8):1162–1174. [PubMed: 17900646]
81. Paludo KS, Biscaia SMP, Chaim OM, Otuki MF, Naliwaiko K, Dombrowski PA, Franco CRC, Veiga SS. Inflammatory events induced by brown spider venom and its recombinant dermonecrotic toxin: a pharmacological investigation. *Comp Biochem Physiol Part C*. 2009; 149:323–333.
82. Babcock JL, Civello DJ, Geren CR. Purification and characterization of a toxin from brown recluse spider (*Loxosceles reclusa*) venom gland extracts. *Toxicon*. 1981; 19(5):677–689. [PubMed: 7302957]
83. Cunha RB, Barbaro KC, Muramatsu D, Portaro FCV, Fontes W, de Sousa MV. Purification and characterization of loxnecrogin, a dermonecrotic toxin from *Loxosceles gaucho* brown spider venom. *J Prot Chem*. 2003; 22:135–146.

84. Machado LF, Laugesen S, Botelho ED, Ricart CAO, Fontes W, Barbaro KC, Roepstorff P, Sousa MV. Proteome analysis of brown spider venom: Identification of loxnecrogin isoforms in *Loxosceles gaucho* venom. *Proteomics*. 2005; 5(8):2167–2176. [PubMed: 15852345]
85. da Silveira RB, Pigozzo RB, Chaim OM, Appel MH, Dreyfuss JL, Toma L, Mangili OC, Gremski W, Dietrich CP, Nader HB, Veiga SS. Molecular cloning and functional characterization of two isoforms of dermonecrotic toxin from *Loxosceles intermedia* (brown spider) venom gland. *Biochimie*. 2006; 88:1241–1253. [PubMed: 16581177]
86. da Silveira RB, Pigozzo RB, Chaim OM, Appel MH, Silva DT, Dreyfuss JL, Toma L, Dietrich CP, Nader HB, Veiga SS. Two novel dermonecrotic toxins LiRecDT4 and LiRecDT5 from brown spider (*Loxosceles intermedia*) venom: from cloning to functional characterization. *Biochimie*. 2007b; 89:289–300. [PubMed: 17296256]
87. Kalapothakis E, Araujo SC, de Castro CS, Mendes TM, Gomez MV, Mangili OC, Gubert IC, Chavez-Olortegui C. Molecular cloning, expression and immunological properties of LiD1, a protein from the dermonecrotic family of, *Loxosceles intermedia* spider venom. *Toxicon*. 2002; 40(12):1691–1699. [PubMed: 12457881]
88. Kalapothakis E, Chatzaki M, Gonçalves-Dornelas H, de Castro CS, Silvestre FG, Laborne FV, de Moura JF, Veiga SS, Chávez-Olortegui C, Granier C, Barbaro KC. The Loxtox protein family in *Loxosceles intermedia* (Mello-Leitão) venom. *Toxicon*. 2007; 50(7):938–946. [PubMed: 17825864]
89. Appel MH, da Silveira RB, Chaim OM, Paludo KS, Silva DT, Chaves DM, da Silva PH, Mangili OC, Senff-Ribeiro A, Gremski W, Nader HB, Veiga SS. Identification, cloning and functional characterization of a novel dermonecrotic toxin (phospholipase D) from brown spider (*Loxosceles intermedia*) venom. *Biochim Biophys Acta*. 2008; 1780:167–178. [PubMed: 18082635]
90. Binford GJ, Bodner MR, Cordes MHJ, Baldwin KL, Rynerson MR, Burns SN, Zobel-Thropp PA. Molecular evolution, functional variation, and proposed nomenclature of the gene family that includes sphingomyelinase D in sicariid spider venoms. *Mol Biol Evol*. 2009; 26(3):547–566. [PubMed: 19042943]

Figure 1. Overview of *Scytodes* venom analysis

(A) cDNA sequences are divided based on overall transcript abundance into five general categories based on database hits ($e < 10^{-5}$) before clustering and categorizing functional prediction. (B) Table, the 99% and 85% identity threshold assemblies compare redundant *vs.* homologous sequences within the library; the latter is used for cDNA library analysis in the remainder of the manuscript. Flowchart, clusters and singleton sequences from the 85% threshold analysis were initially divided based on significant tBLASTx hits ($e < 10^{-5}$) in Arachnoserver or NCBI nt/nr databases. General function prediction was made based on sequences hit in databases or patterns within sequences without hits. Signal sequences were determined from predicted protein translations using SignalP. Sequences that did not hit anything in public database searches were searched against a Haplogyne database (Binford *et al.*, unpublished) to infer unique *vs.* shared homology among Haplogyne spiders. (C) The percentage of proteins detected by mass spectrometry from the crude venom was calculated based on the number of tryptic peptides that matched translated sequences of coding proteins divided by the total number of peptides detected in the digest.

Figure 3. Amino acid alignment of longistatin homologs

The gray bar signifies the predicted signal peptide for the *Scytodes* sequence and the black bar is the predicted EF-hand-5 domain pair predicted using the Pfam sequence search (<http://pfam.janelia.org/search/sequence>). Amino acids with an asterisk (*) below are 100% conserved across lineages shown. Homologous sequences are from *Scorpiops jendeki* (GH47935), *Haploelma huwenum* (EU979481), *O. haiana* (GT865932), *Ixodes scapularis* (EW835901, DQ066287 and DN970126) and *Haemaphysalis longicornis* (AB519820).

Figure 4. Bayesian analysis of TCTP homologs

Branch thickness corresponds with increasing confidence of posterior probabilities (pp) of descendent nodes. The widest branches are pp > 0.95 and the thinnest branches are pp < 0.5. This tree is rooted with a non-arthropod clade. GenBank IDs are listed along with genus names. A dashed box is drawn around the clade containing spider sequences.

Figure 5. Selected amino acid alignments of venom peptide sequences
SYTX predicted signal peptides are underlined in gray and mature sequences are underlined in black. Amino acids in the mature sequence that are 100% conserved across species have an asterisk (*). Peptides detected in the proteome are boxed. Black brackets above sequences in (B) indicate predicted disulfide bonds forming an ICK peptide. Homologous sequences in (C) are listed with GenBank IDs and genus names from *Mesobuthus martensii* (scorpion), *Latrodectus hesperus* (widow spider), *Acanthoscurria gomesiana* (tarantula), *Scorpiops magerisonae* (scorpion), *Amblyomma maculatum* (tick), *Loxosceles laeta* and *L. intermedia* (brown recluse spiders).

Figure 6. Bayesian analysis of cysteine-rich conserved sequences

GenBank IDs are listed along with genus names and common names are in parentheses.

Branch thickness corresponds with increasing confidence of posterior probabilities (pp) of descendent nodes. The widest branches are pp > 0.95 and the thinnest branches are pp < 0.5.

Figure 7. Alignment and Bayesian analysis of sequences from the PXCC gene family
 (A) Amino acid alignment of representative sequences (▲) from each clade in (B) shows predicted signal sequence (black underline), propeptide (gray underline) and amino acids that are 100% conserved (*). Peptides detected in the proteome are underlined in gray. (B) Branch thickness corresponds with increasing confidence of posterior probabilities (pp) of descendent nodes. The widest branches are pp > 0.95 and the thinnest branches are pp < 0.5. Clades are labeled I–VI; sequences listed with filled circles (●) were detected in the proteome. Tree is rooted with the non-arachnid clade.

Figure 8. Alignment of *Scytodes* RP23-like sequences and homologs

The gray bar indicates the signal sequence for processing and asterisks (*) indicate amino acids that are 100% conserved throughout all five sequences. Peptides detected in the proteome are boxed. Homologous sequences are from *Rhipicephalus pulchellus* (JAA61713), *L. hesperus* (HQ006063) and *Dysdera erythrina* (CV177902).

Figure 9. Alignment and Bayesian analysis from the glycine-rich gene family

(A) Amino acid alignment of representative sequences (▲) from each clade in (B) shows predicted signal sequence (gray underline) and GGX motifs (black underline). Peptides detected in the proteome are boxed. (B) Branch thickness corresponds with increasing confidence of posterior probabilities (pp) of descendent nodes. The widest branches are pp > 0.95 and the thinnest branches are pp < 0.5. The tree is midpoint rooted. Each distinct clade is labeled I–VI. A sequence with an asterisk (*) denotes which predicted mature sequences were submitted for theoretical amino acid analysis as described in the text. Sequences with a filled circle (●) were found in the proteome.

```

      10      20      30      40      50      60      70      80      90     100     110
SCV60  MRNFALVFLVASV I YATVAKPSYR I AKRDYAYDQNGNVRVGYQNGYGGEQVAYNGYAQQQGEPPVQYYDVNGGSAGQQYYNGGAGQQYYNGDAGRQHGVD I VRAP - - - - -
consensus MRNFALVFLVASV I YATVAKPSYR I AKRDYAYDQNGNVRVGYQNGYGGEQVAYNGYAQQQGEPPVQYYDVNGGSAGQQYYNGGAGQQYYNGDAGRQHGVD I VRAPMNL LSG I LGKL
SCV23  - - - - - KRDYAYDQNGNVRVGYQNGYGGEQVAYNGYAQQQGEPPVQYYDVNGGSAGQQYYNGGAGQQYYNGDAGRQHGVD I VRAPMNL LSG I LGKL
SCV429 - - - - - AKRDYAYDQNGNVRVGYQNGYGGEQVAYNGYAQQQGEPPVQYYDVNGGSAGQQYYNGGAGQQYYNGDAGRQHGVD I VRAPMNL LSG I LGKL

>consensus
MRNFALVFLVASVIYATVAKPSYRIAKRDYAYDQNGNVRVGYQNGYGGEQVAYNGYAQQQGEPPVQYYDVNGGSAGQQYYNGGAGQQYYNGDAGRQHGVDIVRAPMNL LSG I LGKL*

```

Figure 10. Alignment of sequences containing QQ and YY motifs

Three identical sequences in the cDNA library were aligned and combined into one consensus sequence. The predicted signal sequence underlined in gray and the GGX, QQ and YY motifs are underlined in black.

Scytodes venom peptide analysis

Small cysteine-rich venom peptide groups are listed in order of ascending predicted mature sizes. The Ux-SYTX nomenclature was assigned based on nonredundant full-length (FL) sequences from sequence alignments. All peptide lengths were determined using SignalP and SpiderP; mature sequences were determined to the best of our abilities using SpiderP and proteomic results.

Table 1

top hit in each category of ArachnoServer, NCBI (EST), and Haplogyme (cDNA) database if e 10 ⁻⁵	proposed <i>Scytodes</i> toxin name (SYTX)	#cDNA seqs	#non- redundant FL seqs	size of FL/pro/mature peptide (#aa)	avg.MW of peptides in group (Da)	detected in proteome	predicted sequence of mature peptide	cysteine skeleton pattern
<i>Loxosceles</i> cDNA	U ₁ -SYTX-Sth1a	3	2	60/37/31	3430.11		IVGGLMCMMPNIPCVENDDDCCSKWCEGLICH	-C6C-CC-C-C-
	U ₁ -SYTX-Sth1b	1	1	60/37/31			IVGSLMCMMPNIPCVENDDDCCSKWCEGLICH	
<i>Loxosceles</i> cDNA <i>Loxosceles</i> cDNA <i>Loxosceles</i> and <i>Sicarius</i> cDNA <i>Loxosceles</i> and <i>Sicarius</i> <i>Loxosceles</i> cDNA	U ₂ -SYTX-Sth1a	1	1	47/30/30	3494.02		CTPIDWHVDCPLVVEDMSKNYPPECCPRPNCK	-C-C-CC-C
	U ₃ -SYTX-Sth1b	1	1	60/37/31	3632.79		LIESIACIQKGLPCMEHSDDCCRGVCEALFCQ	-C6C-CC-C-C-
	U ₃ -SYTX-Sth1b	3	1	60/37/31		✓	LIESIACIQKGFP PCMEHS DDCGRGVCEALFCQ	
	U ₃ -SYTX-Sth1c	1	1	68/45/39			LIESIACIQKGWPCMEHSDDCHGVCEALFLSVLYEMRA	
	U ₃ -SYTX-Sth1d	6	1	61/38/32			LIESIACIQKGWPCMEHSDDCHGVCEALFCQY	
<i>Loxosceles</i> cDNA <i>Loxosceles</i> and <i>Sicarius</i> cDNA <i>Loxosceles</i> cDNA <i>Loxosceles</i> and <i>Sicarius</i> cDNA <i>Loxosceles</i> and <i>Sicarius</i> cDNA <i>Loxosceles</i> and <i>Sicarius</i> cDNA <i>Loxosceles</i> and <i>Sicarius</i> cDNA <i>Loxosceles</i> and <i>Sicarius</i> cDNA <i>Loxosceles</i> and <i>Sicarius</i> cDNA <i>Loxosceles</i> and <i>Sicarius</i> cDNA	U ₃ -SYTX-Sth1e	1	1	61/38/32			LIESVACMQQGWPCVVEHTDCCGRMCISMFCAQ	
	U ₃ -SYTX-Sth1f	1	1	71/38/32			LIESVACMQQGWPCVDHTDCCGRMCSSMFCAY	
	U ₃ -SYTX-Sth1g	1	1	71/38/32			LIESIACMQKGWPCMEHVDCCCHGVCDLSFLCLY	
	U ₃ -SYTX-Sth1h	14	3	71/38/32			LIESIACMQKGLPCMEHVDCCCHGVCDLSFLCLY	
	U ₃ -SYTX-Sth1i	3	1	71/38/32			LIESIACMQKGLPCMEHVDCCCHGVCDLSFLCLY	
	U ₃ -SYTX-Sth1j	1	1	71/38/32			LIESIACMQKGLPCMEHVDCCCHGVCDLSFLCLY	
	U ₄ -SYTX-Sth1a	1	1	62/39/34	3667.30		CGYCNPMCEPQVNCPCGAYMDACNCCALCRNCRG	C-C-C-C-C-C6C-CC-C-C-
	U ₅ -SVTX-Sth1a	1	1	71/51/36	3920.32	✓	DETPDECVTRG NFCAT PEVHG DWCC GS LKCVSNSCR	-C6C-CC-C-C
	U ₆ -SYTX-Sth1a	1	1	75/54/39	4448.50		DDSPTYCIQPGNWCDDPREGNWCQKCVKNGPKDFICEE	-C6C-CC-C-C
	U ₆ -SYTX-Sth1b	1	1	75/54/39		✓	DDSP TYCI QPG NWC DDP KEGNWC QK CVK NGPKDF IC KE	
<i>Loxosceles</i> cDNA <i>Loxosceles</i> cDNA <i>Loxosceles</i> cDNA <i>Loxosceles</i> cDNA	U ₇ -SYTX-Sth1a	4	1	64/41/41	4732.53	✓	SEPA PGEC NKFR CP VDW CQ CG LE MD RC GC RYCKPCDR	C-C-C-C-C-C6C-CC-C-C
	U ₈ -SYTX-Sth1a	10	2	83/60/45	4915.73	✓	CGS AL HOA CK CGD Y MD K CG C LY CK PC PGERE AE SL FE ELIS PK ^Δ	C-C-C6C-CC-C-C
	U ₈ -SYTX-Sth1b	1	1	83/60/45			CGSALHQACKCGDYMDKMDKCGLYCCKLYCKPCPGERE AE GL FE ELIS PK ^Δ	
	U ₉ -SYTX-Sth1a	4	1	92/76/61	6094.12		QMQYCMYPVGRCDQGGVPCCGNCVKSCGILGDIVGGIGGNSCTC RETT AGVACEILG	-C6C-CC-C-C-C-C-C-C-C-

top hit in each category of ArachnoServer, NCBI (EST), and Haplogyne (cDNA) database if e 10 ⁻⁵	proposed <i>Scydos</i> toxin name (SYTX)	#cDNA seqs	#non- redundant FL seqs	size of FL/pro/mature peptide (#aa)	avg. MW of peptides in group (Da)	detected in proteome	predicted sequence of mature peptide	cysteine skeleton pattern
μ-cientoxin-Pn1a, U ₂ -theraphotoxin-Pe1a, <i>Loxosceles</i> cDNA	U ₉ -SYTX-Sth1b	1	1	92/76/61			QMQYCMPVGSRCDDGGIPCCGNCVKSCGILGDVGGGIFGGNSCTCKETTIAGVACEILG	
	U ₉ -SYTX-Sth1c	3	1	92/76/61			QMQYCMPVGSRCDKGGIPCCGNCVKSCGILGDVGGGIFGGNSCTCRETTIAGVACEILG	
<i>Sciarus</i> and <i>Physocyclus</i> cDNA <i>Loxosceles</i> cDNA <i>Loxosceles</i> cDNA <i>Loxosceles</i> cDNA	U ₁₀ -SYTX-Sth1a	2	1	99/79/66	6876.15	✓	NIATCYPDDMDIESDMPGYQCCTPCHKCHCPTAGTFFGLFGSSSGPCRCSEGWGEKMLHIVTSC	-C6C-CC-C-C1C-C1C-C
	U ₁₀ -SYTX-Sth1b	1	1	98/65/65		✓	NIPATCYPDNVDCIESDMPGYQCCTPCHKCHCPTARLFLFGSSSGPCRCSEGLGAKILHVVTTGC	
	U ₁₀ -SYTX-Sth1c	2	1	98/65/65			NIPATCYPDNVDCIESDMPGYQCCTPCHKCHCPKAGFFGLFGSSSGPCRCSEGLGAKILHVVTTGC	
	U ₁₀ -SYTX-Sth1d	8	1	98/65/65			NIPATCYPDNVDCIESDMPGYQCCTPCHKCHCPTAGTFFGLFGSSSGPCRCSEGLGAKILHVVTTGC	
<i>Amblyomma</i> EST, <i>Loxosceles</i> and <i>Plectreurys</i> cDNA	U ₁₁ -SYTX-Sth1a	1	1	90/67/67	7037.87		SDYCNTECDRTTCGKVSCACGVHLDVCGCCSVCNLCPDGDKCNVYDSYSGSGSKCKFPSENSLDGVCE	-C-C-C-C-C-C6C-CC-C-C-C-C- C-C-C
U ₁₂ -lycotoxins-Ls1a,1b,1c,1d, <i>Ixodes</i> EST and <i>Loxosceles</i> cDNA	U ₁₂ -SYTX-Sth1a	3	2	91/67/67	7265.72	✓	SEYCDSPSCDLSKCPAVNCKCGTHLDPCECCCTRCSTCFGERCYLYGPGGNGSSCKLEKNEKYGTCTV	-C-C-C-C-C-C6C-CC-C-C-C- C-C-C
U ₁₂ -lycotoxins-Ls1a,1b,1c,1d, <i>Ixodes</i> EST and <i>Loxosceles</i> cDNA	U ₁₂ -SYTX-Sth1b	1	1	91/67/67		✓	SEYCDSPSCDLSKCPAVDCKCGTHLDPCECCCTRCSTCFGERCYLYGYPCGNGSSCKLEKNEKYGTCTV	
TXLP4, <i>Loxosceles</i> EST and cDNA <i>Loxosceles</i> cDNA <i>Loxosceles</i> cDNA <i>Loxosceles</i> cDNA <i>Loxosceles</i> cDNA <i>Loxosceles</i> cDNA <i>Loxosceles</i> cDNA	U ₁₃ -SYTX-Sth1a	2	1	109/88/65	7513.50	✓	CNEINGIKLANGEHYHDTREKYSKDDKDLTISRCDTLFVPSNCIRQKSGSYNCCPTILRCWTD	C-C-C-C-C-C-CC-C-
	U ₁₄ -SYTX-Sth1a	1	1	89/68/68	7538.91		GRCENSEIMGVAKIKPCVCGSYEDQELCQIWCYTCEKRPCVPGSKQKCEPG TACVTKNRPDVGCKKP	-C-C-C-C-C-C-C-C-C-C-C-
	U ₁₄ -SYTX-Sth1b	2	1	89/68/68		✓	GRCENSEIMGVAKIKPCVCGSYKDQEQCIWCYTCERRPCVPGSKEKCEP GTACTITKNRPDIGVCKKP	
	U ₁₄ -SYTX-Sth1c	1	1	89/68/68			GRCENSEIMGVAKIKPCVCGSYEDQEQCIWCYTCERRPCVPGSKGKCEPG TACITKNRPDIGVCKKP	
	U ₁₄ -SYTX-Sth1d	2	2	89/68/68			GRCENSEIMGVARIKPCSCGSYEDQEQCIWCYTCERRPCVPGSKEKCEPG TACVTKNRPDIGVCKKP	
	U ₁₄ -SYTX-Sth1e	7	1	89/68/68			GRCENSEIMGVARIKPCVCGSYEDQEQCIWCYTCERRPCVPGSKEKCEPG TACVTKNRPDIGVCKKP	
	U ₁₄ -SYTX-Sth1e	5	3	89/68/68		✓	GRCENSEIMGVARIKPCVCGSYEDQEQCIWCYTCERRPCVPGSKCEPGTACTITKNRPDIGVCKKP	
	U ₁₅ -SYTX-Sth1a	1	1	96/69/69	7774.98	✓*	EKCDERDGMLLKVGETIYDDIQCKMYVCTKEALIEDCGYVPPDEGTTCTKINKGSPKCCPDIDCMK	-C-C-C-C-C-C-CC-C
<i>Acanthoscurria</i> EST and <i>Sciarus</i> cDNA	U ₁₅ -SYTX-Sth1a	1	1	96/69/69	7774.98	✓*	EKCDERDGMLLKVGETIYDDIQCKMYVCTKEALIEDCGYVPPDEGTTCTKINKGSPKCCPDIDCMK	
<i>Loxosceles</i> , <i>Plectreurys</i> and <i>Physocyclus</i> cDNA	U ₁₆ -SYTX-Sth1a	1	1	100/80/69	7806.84		NEYSGQQLCWEVDNECWQTSEYGCCNECYKCHCDDEGHGMFGQILGKAKASHCACRKNLFDMMVCSVMHG	-C6C-CC-C-C1C-C1C-C
<i>Loxosceles</i> , <i>Plectreurys</i> and <i>Physocyclus</i> cDNA	U ₁₆ -SYTX-Sth1b	5	2	100/80/69			NEYSGQQLCWEVDNECWQTSEYGCCNECYKCHCDDEGHGMFGQILGKAKASHCTCRKNLFDMMVCSVMHG	
<i>Loxosceles</i> , <i>Plectreurys</i> and <i>Physocyclus</i> cDNA	U ₁₆ -SYTX-Sth1c	1	1	100/80/69			NGYSGQQLCWEVDNECWQTSEYGCCNECYKCHCDDEGHGMFGQILGKAKASHRTCRKNLFDMMVCSVMHG	
<i>Loxosceles</i> , <i>Plectreurys</i> and <i>Physocyclus</i> cDNA	U ₁₆ -SYTX-Sth1d	14	1	100/80/69		✓	NGYSGQQLCWEVDNECWQTSEYGCCNECYKCHCDDEGHGMFGQILGKAKASHCTCRKNLFDMMVCSVMHG	
<i>Sciarus</i> cDNA	U ₁₇ -SYTX-Sth1a	1	1	101/80/75	8088.13	✓	QSGCQGGDGKLYVPGEYYDDENCVKSVCMEDGTFNFNDVCTPIDTLTKLCKFIKPSGSGYPACGSAQLRL	-C-C-C-C-C-C-CC-C
<i>Loxosceles</i> and <i>Plectreurys</i> cDNA	U ₁₈ -SYTX-Sth1a	1	1	91/69/69	8167.59		KRVKPGCPLRCDRKKCPVISCEGGMKLDECKCPCYLCRGQYDCSESRPCDKRRCEWDYRWGRTHCV	-C-C-C-C-C-C6C-CC-C-C-C- C-C-C-

top hit in each category of ArachnoServer, NCBI (EST), and Haploglyne (cDNA) database if e 10 ⁻⁵	proposed <i>Scytodes</i> toxin name (SYTX)	#cDNA seqs	#non- redundant FL seqs	size of FL/pro/mature peptide (#aa)	avg. MW of peptides in group (Da)	detected in proteome	predicted sequence of mature peptide	cysteine skeleton pattern
μ-ententoxin-Pn1a, <i>Loxosceles</i> and <i>Sicarius</i> cDNA	U19-SYTX-Sth1a	1	1	121/101/80	8879.63	✓	DDEPTQCFPKGNRCDAQGADYCCDCSSDGGWCVTCPSGCKKSGYKGGCHSNSDCCSGKCTRVYDNEPASFVYRGHPGLGR	-C6C-CC-CC-CC-C-C-C-C-CC- C-C
<i>Loxosceles</i> and <i>Sicarius</i> cDNA	U19-SYTX-Sth1b	2	1	121/99/80		✓	DDEPTECFQRGHWCDSYGADYCCCECDLNDTCVACKSGCKRSGYKGGCHSGAECCSGVCTRTILESDDPYKYCVYRGHPGLGR	
<i>Loxosceles</i> and <i>Sicarius</i> cDNA	U19-SYTX-Sth1c	1	1	121/99/80		✓	DDEPTECFQRGSWCDSYGGDHCCDCCDLNNTCVACKAGCKKSNNNVCHSGEECCSGSGCSRVRDWDPKKYCVYRGHPGLGR	
<i>Loxosceles</i> and <i>Sicarius</i> cDNA	U19-SYTX-Sth1d	3	1	116/94/80		✓	LSDEEVPTECYPRGHRCDAYDADCDCDCLNNKCVACTPGCKKSGYKGRSNEECCSGVCYRTDRSSAYKYCVYRGYGG	
<i>Loxosceles</i> and <i>Sicarius</i> cDNA	U19-SYTX-Sth1e	3	1	119/97/83		✓	ASDEEPLHCFEKGHWCDSYGADYCCDCCDLNNTCVDCCKPCIRSGQGGQCYSGEECCSGDCSRVSRSDPKKYCVYRGHPGLGR	
<i>Loxosceles</i> , <i>Sicarius</i> and <i>Physocyclus</i> cDNA	U19-STYX-Sth1f	1	1	119/97/83			GGDEEPLHCFEKGHWCDSYGADYCCDCCDLNNTCVDCCKPCIRSGQGGECYSGEKCSCGHCSRVLRSDDPKTYCVYRGYSGLGR	
<i>Loxosceles</i> , <i>Sicarius</i> and <i>Physocyclus</i> cDNA	U19-SYTX-Sth1g	1	1	119/97/83		✓	GGDEEPTECFPRGHWCDSYGGDHCCDCCDLNNTCVACTPGCKRSGYKGGCHSSECCSGVCYRTDKSSPYKYCVYRGYGGYGR	
		138	60					

Cysteine skeleton patterns are listed to show characteristic features of known venom peptide neurotoxins: a dashed line (–) indicates the presence of more than one amino acid surrounding the cysteine residue, primary structural motifs –C6C-CC-, –CXC-CC (PSM) and –C1C- (ESM) are indicated if present. Sequences from the U8-SYTX group (*) are predicted to contain an α-helix at the C-terminus following the final cysteine residue (<http://bioinf.cs.ucl.ac.uk/psipred/>). Checkmarks (✓) indicate that one or more peptides from these groups were detected in the proteome, which are **bold** in the predicted sequence. The checkmark with an asterisk (✓*) indicates that a portion of the propeptide sequence was detected as well.