Cell. Mol. Life Sci. (2010) 67:2799–2813
DOI 10.1007/s00018-010-0359-x

Cellular and Molecular Life Sciences

RESEARCH ARTICLE

Venom components from *Citharischius crawshayi* spider (Family Theraphosidae): exploring transcriptome, venomics, and function

Elia Diego-García · Steve Peigneur · Etienne Waelkens · Sarah Debaveye · Jan Tytgat

Received: 10 November 2009 / Revised: 4 March 2010 / Accepted: 18 March 2010 / Published online: 6 April 2010

Abstract Despite strong efforts, knowledge about the composition of the venom of many spider species remains very limited. This work is the first report of transcriptome and venom analysis of the African spider *Citharischius crawshayi.* We used combined protocols of transcriptomics, venomics, and biological assays to characterize the venom and genes expressed in venom glands. A cDNA library of the venom glands was constructed and used to generate expressed sequence tags (ESTs). Sequence comparisons from 236 ESTs revealed interesting and unique sequences, corresponding to toxin-like and other components. Mass spectrometrical analysis of venom fractions showed more than 600 molecular masses, some of which showed toxic activity on crickets and modulated sodium currents in $DmNa_v1$ and $Na_v1.6$ channels as expressed in *Xenopus* oocytes. Taken together, our results may contribute to a better understanding of the cellular processes involved in the transcriptome and help us to discover new components from spider venom glands with therapeutic potential.

Keywords *Citharischius crawshayi* · Transcriptome · Spider toxin · Theraposidae · King Baboon spider

Abbreviations

Clone Cic	Nucleotide or amino acid sequence deduced from cDNA library from venom glands of spider *C. crawshayi*
Fraction Cic	Fraction from HPLC separation of venom from spider *C. crawshayi*
EST	Expressed sequence tags
ICK	Inhibitor cystine knot-motif
TFA	Trifluoroacetic acid
MALDI-TOF/TOF	Matrix-assisted laser desorption/ ionization time-of-flight/time-of-flight mass spectrometer

Introduction

Spiders belong to a very diversified group of arthropod predators with 40,700 described species in approximately 109 families [1]. They are the largest of venomous animals with specialized organs for the synthesis of venom (venom gland). Spider venom contains a vast diversity of molecules with different biological activities, some of which have evolved into highly selective and potent pharmacological tools for the characterization of cell receptors (e.g., Ca^{2+}, Na^+, or K^+ voltage-gated ion channels). This mix of several types of cell proteins and toxin peptides underlies and is responsible for the noxious effects. Over the last few years, it has become clear that studying the mechanism of action of spider toxins can lead to a better understanding of the structure, function, and pharmacology of specific ion channels, as well as their role in channelopathies (e.g., myotonias,

The gene sequences reported in this paper have been submitted to GenBank, the accession numbers corresponding GU170865 to GU170901.

E. Diego-García · S. Peigneur · S. Debaveye · J. Tytgat (✉)
Laboratory of Toxicology, University of Leuven, Campus Gasthuisberg, O&N2, Herestraat 49, PO Box 922, 3000 Louvain, Belgium
e-mail: Jan.Tytgat@pharm.kuleuven.be

E. Waelkens
Laboratory of Protein Phosphorylation and Proteomics, University of Leuven, Campus Gasthuisberg, O&N1, Herestraat 49, PO Box 901, 3000 Louvain, Belgium

E. Waelkens
Prometa and BioMacs, Louvain, Belgium

epilepsies, migraine, arrhythmias, and neurodegenerative disorders including Alzheimer and Parkinson's disease). Furthermore, advances in venomics and transcriptomics via, respectively, mass spectrometry and molecular biology methods, have allowed the characterization of genes related to such peptide toxins and shed light on the molecular diversity of spider venoms [2, 3].

The molecular megadiversity of spider venom is estimated at over 12 million biological active peptides [4]. Although several published works report on different components isolated from spider venoms, only around 600 different peptides have been described so far [5]. Additionally, transcriptomic and proteomic analysis information remains very limited. Some mass fingerprinting and proteomic analyses have been reported from spider venoms: *Pterinochilus murinus* [6]; *Loxosceles gaucho* [7]; *Lasiodora parahybana* [8]; *Chilobrachys jinzhao* [9]; *Ornithoctonus huwena* [10] and the recent comparison of peptidomes of the theraphosid *Coremiocnemis tropix*, *Selenocosmia crassipes*, *Selenotholus foelschei*, *Brachipelma albiceps*, *B. hamorri* and the Australian funnel-web spider *Hadronyche infesa* [11]. Nevertheless, the number of transcriptome references is still limited to: *Chilobrachys jingzhao* [12, 13]; *Loxosceles laeta* [14]; *Ornithoctonus huwena* [15]; *Aphonopelma* sp. [16] and *Lycosa singoriensis* [17]. Combinatorial strategies of transcriptome, mass spectrometry, and electrophysiology analyses, may help us to explore and identify several venom components from spider as possible new peptides and toxins.

The African spider *Citharischius crawshayi* Pocock, 1900 (King Baboon spider), is a tarantula considered as an extremely aggressive species. They can inflict a painful bite, but the species is considered not toxic to human and the animals are frequently kept as pets. Although the King Baboon spider is widely known, there are no studies available about its venom composition.

The aim of this work was to explore the genes and venom composition from *C. crawshayi*, particularly to search novel high-affinity ligands (toxins) targeting ion channels and to provide a general characterization of the venom. This study is to the best of our knowledge the first combinatorial strategy report of exploring the transcriptome and venom analysis from the African spider *C. crawshayi* (King Baboon).

Materials and methods

Expressed sequence tags generation and analysis of sequences

To obtain the expressed sequence tags (ESTs) and transcriptome analysis, we constructed a cDNA library of the venom gland from *C. crawshayi*. Total RNA was obtained from venom glands dissected from one spider; 2 days before, the spider had been milked by electrical stimulation; we used the TRI reagent from Sigma (Sigma, Belgium) for RNA isolation. With this material a cDNA library was constructed as described by protocols of Creator SMART cDNA Library Construction Kit (Clontech, USA). Random sequencing strategy was used to screen the cDNA libraries.

We generated a random screening of 282 colonies using a non-amplified cDNA library from the venom gland. We used colony polymerase chain reaction (PCR) to select the positive clones. Primer T7 in conjunction with M13 reverse primer used for libraries screening, were used to the PCR reaction, following the protocol: hot start 4 min denaturalization at 94°C, 1 min at 94°C, 1 min at 50°C and 1 min at 72°C, for 32 cycles and final step of 5 min at 72°C. The plasmid DNA of the colonies selected was obtained by mini-prep kit preparations (Roche, Germany) and were sequenced from both ends by Genetic Service (VIB, Belgium).

ESTs were analyzed by electropherogram quality analysis, process of uploading PHRED package via web and assembling in clusters using CAP3 (PHPH: (https://www.biomol.unb.br/). Quality values of the DNA sequencer trace data produced by PHRED are used in CAP sequence assembly program for overlaps between reads, removal of false overlaps and construction of contigs, generating multiple sequence alignments and consensus sequences. Results of CAP3 allow the generation of contigs, singlets and quality files. Each contig and singlet sequence was searched against the GenBank database with algorithms BLASTX, Protein BLAST (http://www.ncbi.nlm.nih.gov/blast/), FASTA and FASTA Fastx/Fasty (http://fasta.bioch.virginia.edu/fasta) to identify similar sequences to find genes and homologous proteins; an expectation value (e value) $< 10^{-5}$ was used as cut-off for homologue detection. All DNA sequences were analyzed with DNA Strider 1.4f6 and 4Peaks 1.7.2 tools. In addition, we used the ArachnoServer database (http://www.arachnoserver.org/) for a final search of sequence identity. The EST sequences were submitted to GenBank. SignalP 3.0 (http://www.cbs.dtu.dk/services/SignalP/) and ProP 1.0 (http://www.cbs.dtu.dk/services/ProP/) server predicts the presence and location of signal peptide and pro-peptide cleavage sites, respectively.

Biological materials and high-performance liquid chromatography separation

Total venom from *C. crawshayi* was obtained by electrical stimulation of chelicera by applying 15–22 Volts. The venom samples were collected, dissolved in distilled water

and centrifuged at 10,000 × *g* for 10 min. Soluble venom (1 mg protein) was fractionated by high-performance liquid chromatography (HPLC) via C18 reverse-phase column and eluted with a linear gradient from 2% solution A (0.12% TFA in water) to 60% solution B (0.10% TFA in acetonitrile) run at a flow rate of 0.5 ml/min for 70 min using a liquid chromatography system (Gilson, USA). Samples selected for a second HPLC fractionation used the same instrument and column with a linear gradient of 15% solution A to 50% solution B, run at a flow rate of 0.2 ml/min (60 min). After the second separation, the dry samples were suspended in water and the total protein amount in each fraction was estimated via Nanodrop (ND-1000 Spectrophotometer, Isogen-Life Sciences); the fractions with the highest protein levels were selected for further mass spectrometric characterization via ABI 4800 MALDI-TOF/TOF (Applied Biosystems, USA) and biological assays.

Biological assay of toxicity

The venom toxicity was assayed by injecting crickets (Insecta: Orthoptera: *Gryllus* spp.) with 10–30 μg of total venom and venom fractions obtained from HPLC fractionation. There were 3–4 crickets per fractions tested (each insect weighed approximately 20 mg). The qualitative assay included total venom from *Pandinus cavimanus* scorpion used as reference of positive control for paralysis or lethal activity; injection of BSA buffer or ND96 solution (in mM: NaCl 96, KCl 2, $MgCl_2$ 1, $CaCl_2$ 1.8, HEPES 5 adjusted to pH 7.5), were used as a control and reference for the negative effects on the crickets.

Electrophysiological analyses

The electrophysiology bioassays were analyzed by heterologous expression of ion channels in *Xenopus laevis* oocytes. cRNAs transcripts were synthesized from the linearized plasmids using the T7 or SP6 mMESSAGE mMACHINE transcription kit (Ambion, USA). Stage-V and -VI *X. laevis* oocytes were harvested by partial ovariectomy under anesthesia with 3-aminobenzoic acid ethyl ester methanesulfonate, 0.5 g/l (Sigma, Belgium). Anaesthetized animals were kept on ice during dissection. The oocytes were defolliculated by treatment with 2 mg/ml collagenase (Sigma, Belgium) in Ca^{2+}-free ND96 solution. Between 1 and 24 h after defolliculation, oocytes were injected with 50 nl of 50–100 ng/μl potassium channels cRNA or 50–96 nl of 1–3 μg/μl sodium channels cRNA using the micro-injector (Drummond Scientific, USA). The oocytes were then incubated in ND96 solution (supplemented with 50 mg/l gentamycin sulfate) at 16°C for 1–5 days to the expression of ion channels.

Electrophysiological measurements were conduced using a two-electrode voltage-clamp; the recordings were performed at room temperature (18–22°C) using a GeneClamp 500 amplifier (Axon Instruments, USA) controlled by a pClamp data acquisition system (Axon Instruments, USA). Whole-cell currents from oocytes were recorded 3–5 days after injection. Bath solution composition was (in mM): NaCl 96; KCl 2; $CaCl_2$ 1.8; $MgCl_2$ 2 and HEPES 5, pH 7.5. Voltage and current electrodes were filled with 3 M KCl. Resistances of both electrodes were kept low ($\leq$1.0 MΩ). Current traces were evoked in an oocyte expressing the cloned sodium channels by 5-mV depolarization steps from a holding potential of −90 mV. Potassium channel currents were evoked by depolarizations to 0 mV from holding potential of −90 mV. Currents were filtered at 500 Hz and sampled at 2 kHz using a four-pole low-pass Bessel filter. Leak subtraction was performed using a –P/4 protocol. The pClamp program was used for data acquisition and data files (Molecular Devices, Sunnyvale, CA) were directly imported, analyzed and visualized with Origin software (Originlab, USA).

The biological effect of total venom and fractions from HPLC separation from total venom were analyzed on sodium ($rNa_v1.4$, $hNa_v1.5$, $mNa_v1.6$, and $DmNa_v1$) and potassium channels ($hK_v1.3$, $rK_v1.5$ and *Shaker* IR). All assays included a minimum of three oocytes evaluated on each ion channel tested.

Mass spectrometry and data analyses

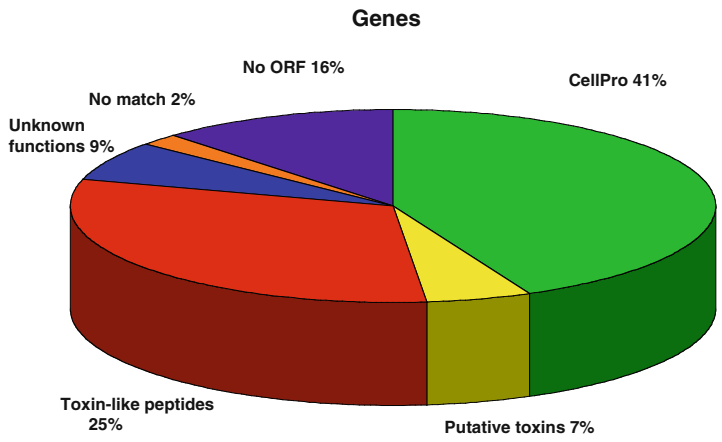
Analysis by mass spectrometry was performed using an ABI 4800 MALDI-TOF/TOF (Applied Biosystems, USA). The instrument was run in reflector mode (mass range 0.9–10 kDa) or linear high mass mode (range 4–20 kDa). Instrument calibration was achieved using Peptide Calibration Mixture 1 (Applied Biosystems, USA). All recordings were made in the positive ion mode with alpha-cyano-4-hydroxycinnamic acid (3.5 mg/ml in 50% CH_3CN/0.1% TFA) as matrix solution.

Results

Transcriptome analyses and identification of new genes

We constructed the cDNA library of venom glands from *C. crawshayi*. The library size obtained was of 1×10^5 with 99% recombinant clones in the non-amplified cDNA library. We randomly generated and analyzed 282 expressed sequence tags (ESTs) from the cDNA library. These ESTs correspond to 0.3% of the cDNA library and were analyzed via PHRED, CAP3, and algorithms previously described in methods. A total of 236 ESTs were

A



B

Category of genes from ESTs	No. of Contigs or Singlets (%)	No. of clones (%)	No. of ESTs
"Toxin-like peptides"	11 (25)	36 (23)	77
"Putative toxins"	3 (7)	14 (9)	30
"CellPro"	18 (41)	28 (18)	44
Cell Structure/Motility	3 (7)	3 (2)	3
Metabolism	6 (13)	9 (6)	16
Gene/Protein Processing	4 (9)	4 (2)	9
Peptidases	3 (7)	10 (6)	13
Cell Signaling / Communication	1 (2)	1 (0.6)	2
Viral Protein	1 (2)	1 (0.6)	1
"Unknown functions"	4 (9)	5 (3)	11
"No Match"	1 (2)	39 (25)	41
"No ORF"	7 (16)	32 (21)	33
Total	44 (100)	154 (100)	236

Fig. 1 Relative proportion of the unique sequences (14 contigs and 30 singlets) of the transcripts from cDNA venom gland library from *C. crawshayi*. **a** Frequency distribution of the different EST categories and the corresponding gene clusters. A total of 282 ESTs were analyzed in the current study, of which 236 ESTs according to their specific category are shown. "CellPro" includes transcripts coding for proteins involved in cellular processes (41%). "Toxin-like peptides" includes sequences with high identity to toxin families (25%). "Putative toxins" includes sequences rich in cysteine with identity to toxins or to the ICK-motif (7%). "Unknown function" includes ESTs that presented identity with already described sequences with no functional assessment or hypothetical genes (9%). "No ORF" includes sequences with non-identified open reading frame (16%). "No match" includes ESTs that did not match with currently known sequences (2%). **b** Proportion of categories and their corresponding number of genes, clones and EST values. "CellPro" category has been divided in different subcategories of functional genes

analyzed in the current study. In total, 14 contigs (i.e., clusters of nearly identical genes encoding the same toxin) and 30 singlet sequences were identified. The gene categories from ESTs are shown in Fig. 1, the relative proportion of each category and their corresponding number of genes, clones, and ESTs are shown in Fig. 1b. "CellPro" includes transcripts coding for proteins involved in cellular processes (41%): peptidases, cell signaling, cell structure and motility, metabolism, gene and protein processing, and viral protein. "Toxin-like peptides" includes sequences with high identity to toxin families (25%). "Putative toxins" includes sequences rich in cysteine with identity to toxins or ICK motive (7%). "Unknown function" includes ESTs that presented identity with already described sequences with no functional assessment or hypothetical genes (9%). "No ORF" includes sequences with non-identified open reading frame (16%). "No match" includes ESTs that did not match with currently known sequences (2%).

The relevant information and sequences were deposited in the GenBank database (accession numbers GU170865-GU170901). Table 1 shows the precursors of different categories of peptides and peptide sequences deduced from cDNA library sequences.

Toxin-like and putative toxin genes

The clusters clearly matching spider toxins or related genes were considered as toxin-like. Others that would not be classified as toxins at a first glance or showed low identity

Table 1 List of precursor sequences deduced from ESTs obtained from the cDNA library of *C. crawshayi*

Accession no.|Gene	Amino acid deduced sequence
	Signal peptide *pre-peptide* *mature peptide*
	1 10 20 30 40 50 60 70 80 90 100 110
"Toxin-like"	
GU170865|CicContig12	MNSVRIIFLVVLVLTATIGQADERE YKDETLEKTLSEAEENYLKDADGNLKDVVDILLGTDMEESRNSRPKR**CIGESVPCGENGELGCCPGLECLKPTGYGWWYKDYWCYRKS**g*
GU170866|CicContig4	MKTSMLAVFVALPLAFVLTAATEERAHPNELVNSLVELVKLDAER**GVDKEGCKYMFGSCGKSDDCCPKLACKRTFNYCAWDGSV***
GU170867|CicCContig5	MKTSMLAVFVALPLAFVLTAATEERAHPNELVNSLVELVKLDAER**GVDKEGCKYMFGSCGKSDDCCPKLACKRTFNYCAWDGSV***
GU170868|Cic684	MKTSMLAVFVALPLAFVLTAATEERAHPNELVNSLVELVKLDAER**GVDKEGCKYMFGSCGKSDDCCPKLACKRTFNYCAWDGSV***
GU170869|Cic283	MKTSMLAVFVALPLAFVLTAATEERAHPNELVNSLAELVKLDAER**GVDKEGCKYMFGSCGKSDDCCPKLACKRTFNYCAWDGSV***
GU170870|CicContig7 and GU170871|CicContig2	MKTLTLIAILSFALLVATAGEELEALEGYPIEPEDDTAVAPLEDER**LFECVISCDIEKEGKPCKPKGEKECKPKGGWKCKFNFCLKV***
GU170872|Cic667	MKTLTLIAVLAFALLVVFHATSIEGLEALEGYTIKPDDDTAVTPLDEE**RFECAISCDIQKEGKPCKPNGEKECKAKGGWKCKFNFCVKI**g*
GU170873|Cic279	MKLSSVIIIAIFVVVVVGLPSRKTASETSKLLEKLGVSREAIPQEMAR**ACSKQIGEKCEHDCQCCGATVVCGTIYVGGNAVEQCMSKTSNNAVLNTMGHGMNAVQNAFTSVMCWG***
GU170874|Cic67 and GU170875|Cic657	MFSLKFVLLVTILCFAAVATVSDAQR**YGGGNRCRNIQCGVPRCPPGSRIMTNRAVSCCPFCS***
"Putative toxin"	
GU170876|Cic277	MNVILLVAAVCILTVHLTVGQDDQTTLRPRPGCRFDHCHLPKIPGCCEFFEAFRKEMDALKQDPAQSCKRFPNRVQQEDCKMQKILTARSRVQPSEECKARMEAFVQGTPPSSTTVAA*
GU170877|CicContig1	MNSILIAALLCVLAVEMTVSQPGTTDSTCRTEHAHLPKIPGCCEFLGAFRIEMDKLKEDPEESCKRFPDRQTQQECKMQKILTARAKIEPSPQCKDNMLAFMQGKPPSR*
GU170878|CicContig10	MNVILVLAFCCILALQSTVSHTDPHHHHHHATEPGHEPCVPLDRCRSPKIPGCCEFFEAFGKEMAELKRDKTQSCKRFPPDETEKREQCKMQKMFIARSKVEPSPECKERMEAFIQGTPP STTTSP*
"CellPro"	
GU170879|Cic31|PEP	MITRAVSSFSIILCMVLAGISVPSAAAPKYRQDTEECTTEACQKTAQYLLESMDTSANPCQDFYQYACGGWVRRHPVPEEKSRYSAFDVLNDEVLDIVTGILKNASSEVHPRAIVDAAKF FDGCIDTEARESAGLESLKNLLNELGGWPMANPDWTGDGYEWQTEVATVARILSLPVIITVMVEGDSNQTTRNILYIDQQWLFGIGRNQLKDKDLESNRDIVTAYKNYVIATGKL…
GU170880|CicContig9|PEP	…CFPCSLCVHSVFDRVSAHSSPRGGSENQ*
GU170881|Cic41|PEP	RLMEDLRIIGICLGIPPATFTWEYYNKNKYCSIGPITPLEFYNEYVKRHCSIENKVCLVNDPRPDNPYGHTYTVDCLGNMVGGRKTIYLNAPAEQLLKVSVESLKKNEPVWFGCEIDKK FYGKAGIQDLQVLDFKLLFGVEVTTNLSKAERLIYGDSMMNHAMVLTAVSLDGHDNPTKWRVENSWGEDRGDKGYLIMTTEWFQEFVFEVVVDKKLISPEMLQALEKEPKVLPAWDPMGS LA*
GU170882|Cic605|CSC	MSEALEKLRQQCRARGANGIIGLGRAFRLMDDNRDKKLNLEEMKLGLQEYGADIEPSEVDSLFKELDRDGSGTVNFDEFLRAVRPPLPQERLDMIAKAFAKMDKTGDEVVNLPDLKVVYN AKEHPDVIGGKKSEDEVLVEFPQNFETPNNPDGKVTKQEFIDYYTGVSSSIDLDEYFVTMMQKAWQM*
GU170883|Cic37|CSM	LSAQVANKLAGKRDPHLEGEVLQWVEAILGQRLPAGPYEEVLRDGVVLCNLMNVLMPGCIPKINTSGGQFKMMENINRFQEACKKWGVPEIDVFQTVDLWERRNIPQVTQCLMALGRACY QHPDYNGPCLGPKPSEENKRQFTEEQLRASEGIINLQYG…
GU170884|Cic666|CSM	MPFKPVEQAKCPKCGKSVYAAEEMLAAGAKWHKTCFKCGLCHKRLNSTNATEHAGDLFCKQCYGRKFGPKGYGFGGGAGCLSMDKGEQSGNTECAGNKPTLDPTYG*
GU170885|Cic253|CSM	MCDDDIAALVVDNGSGMCKAGFAGDDAPRAVFPSIVGRPRHQGVMVGMGQKDSYVGDEAQSKRGILSLKYPIEHGIITNWDDMEKIWHHTFYNELRVAPEEHPILLTEAPLNPKANREKM TQIMFETFNAPAMYVAIQAVLSLYASGRTTGIVLDSGDSVSHTVPIYEGYALPHAILRLDLAGRDLTDNLMKILTERGYSFVTTAEREIVRDIKEKLCYVALDFEQEMAIAASSSTVEKS YESPDGQVITIGNERFRCPETLF…
GU170886|Cic22|GPP	MSLGVPVKVFHEAEGHVVTLETNTGEIYRGRLVEAEDNMNCQMADVTVTYRDGRVGQLENIYIRGSKIRFLILPDTLKNAPMFKRTGPRGGGGSTGRGKSAILRAQAARGGGRARALFQR RK*
GU170887|Cic49|GPP	MTQFRDWNLECKVYIGNLGSRATKDDIEAVFSRYGPIRNVWVARNPPGFAFVEFEDSRDAEDAVKALDGTRICGARVRVEMSHGRTRNDRGRRRNPPPPPPPQYHRRRYSRSRSRSLSPR YHHQRGRSDSRERR*
GU170888|Cic630|GPP	MSGRGKGGKGLGKGGAKRHRKVLRDNIQGITKPAIRRLARRGGVKRISGLIYEETRGVLKVFLENVIRDAVTYTEHAKRKTVTAMDVVYALKRQGRTLYGFGG*
GU170889|Cic275|GPP	VVLESGKHEMEKAGGSGDLSQKWKKLRELHARRNEARTLNHQEVIEEDRRKKLPPNYETKSKWAAYLLEEAEKRQEASARGEDYDRLKLLDIGAEDA
GU170890|Cic243|MeS	MRNMFKYWPAIVFSLVVAVNSNDECDPRQLDSCLSNGELTIEGLEQLPGEEFFNERCGPDIEALKCVKSVVERCPEHNELNARIETIESVINFLNDICNESSTVRKGYLENVECYNNALP NIFTCVINGATILRNFALSLRNKTGDVRDIRPVLCLSPVLITSCLIKESTELRGSENTAPLVIISTDLTSIAKEILCPDEDVFRELGDQFAEYLDGNNLELMKGGKWESVRIPDWSII*
GU170891|Cic643|MeS	…TVSILPTPNHPKTELRPSPPLRHPPACLPLRNSAAPGSKPQTKKVNPXCPRLKPNTGHPFPTKIWPQRRKTRVPKKSPLLWEQRLKIFQPGEKPPPPWGIGSCK…
GU170892|CicContig3|MeR	MGKPRGIRTARKHRNHRRDQRWHDKDYKKAHLGTRWKANPFGGASHAKGIVLEKVGVEAKQPNSAIRKRVRVQLIKNGKKITAFVPRDGCLNFIEENDEVLVAGFGRKGHAVGDIPGVRF KVVKVANVPLLALYKEKKERPRS*
GU170893|CicContig11|MeC	…VGKREPFGTLGMIYAMVRIGGIGFVVXAHHIFSVGMDVDTRAYFTAATMIIAVPTGIRVFSWMATLYGSFFRIEVSLMWCVGFVFLFTLGGLTGVVLANSSLDIVLHDTYYVVAHFHYV LRMGAVFAIMGGLVYWFPLFYGIVMNLRXVKLQFIVMFLGVNMTFFPQHFLGLNGIPRRYSDYPDSFYVXNLVSSLGSILSLMAVVVLVILVXEALVSRNEIVVEYYVGGSLEXQHDNPP LEHTYGQIVRVRR*
GU170894|Cic73|MeC	…ELESVXTLLPGVFLLAIAFPSLKLLYLMEEMDDPGLTLKSSGRQWYWVYEYADLGVKIYESYIIPSGEEDNLRLVEVDNSLVVPMNVGIRMVVTSGDVIHSWTVPALGIRADGIPGRLN QVLFECNRPGIFIGQCSEICGSNHRFIPIVMEVIDERDFLINW…
GU170895|Cic20|MeA	MIPLRGAFCVRAFRKMITFHRPSVFSRGIATCVDPSVGLSDEQVQMQRMATEFAQKEMAPHMAKWDEEEIFPVEEMRKAAYLGFGAIYIPEKYGGSGMTRVDASII…
GU170896|Cic13|VP	…LVATCDSSDRPRVYTITAADDYQPSSQYQSGGVTITLFSANTDAITSLSVGGELVFXTSVLALILGAPMCLIRLVGTTVITRARSA…
"Unknown"	
GU170897|Cic69	MALPSTASSARNLFTESKARLADRIQVNVNSAGSLARQIIRGSKSNETLMYTARNFALQEYAIDNSEMNLKRMQTLTNHLQLQLDNIQKSTLLIERVKEQLYHLHC*
GU170898|CicContig14	MKTFLAAALLFLIVLHYAQYGNDDDFAVNGGGNGLMKAALEERDDFDWTWQKDGEEPEKAFVAVKGGHEDKGKGEKAAKFAKDFFNFYDGAKWSK*
GU170899|Cic29	MTVGQRSLNVMSHPNWFFRRGGIAKINLLKTVDKSQITSLLWPG*
GU170901|Cic72	MALKLFALFLLFQITQAINPFEGGFIRSKCQKGRPLKGCNQTLLEAYGNTELSKTDKNLNNATTYMMLF…
"No match"	
GU170900|CicContig13	MNHNSDTTEKPLKCTSITHCYI*

Complete and partial ESTs of different categories are shown. Precursor sequences shown putative mature sequence in *bold*; *lowercase g* representing presumed modification or elimination of residue in mature toxin; *asterisks* indicate stop codon; partial sequences include three points in the starting or ending of the sequence as reference of sense of precursor in the end terminal

General category of the genes is shown in the top of each block. Subcategories of cellular procedures genes are indicated as: *PEP* peptidase, *CSC* cell signaling/communication, *CSM* cell structure/motility, *GPP* gene/protein processing; *Me* metabolism (*MeS* secretory protein, *MeR* ribosomal protein, *MeC* Cytocrome C, *MeA* Acyl-CoA dehydrogenase); *VP* viral protein

with well-known toxins, were considered putative toxin genes. We obtained 11 genes that encode six new toxin-like peptides. In addition, three putative toxin genes have been found. Figure 2 shows the putative mature sequences deduced from toxin-like genes and the multiple sequences alignment with peptides of different, but related toxins. We

assume that toxin-like mature peptides with the inhibitor cystine knot motif (ICK-motif) and sequences with six or eight cysteine residues are involved in the formation of disulphide bonds. Presumably some of them play also a role in the mechanism of post-translational modifications to process the precursor sequence.

Toxin-like GVDKE

Four different genes encode a new toxin-like peptide. This group of genes includes two contigs (clusters) and two singlet sequences that correspond to: CicContig4 (with four clones), CicContig5 (with seven clones), and the singlet sequences from clones Cic684 and Cic283. These genes encode a very similar precursor of 94 residues with 98% identity between the gene sequences. The deduced sequence shows the similar precursor sequence with one different residue in the pro-peptide region: amino acid V46 or A46 in the primary sequence (see Table 1). As shown in Table 1, the four genes encode the same mature sequence of 39 residues that corresponds to amino acid 46–84 of the precursor sequence. This mature sequence (GVDKE…GSV) shows identity values of 66 and 52% (expect value = $6e^{-10}$ and $5e^{-07}$) to toxin SNX-482 from *Hysterocrates gigas* (GenBank: P56854) and Ceratotoxin-3 from *Ceratogyrus cornuatus* (GenBank: P84509), respectively. We named this sequence "toxin-like GVDKE" on the basis of the N-terminal sequence of the putative toxin sequence. Figure 2 shows the alignment of toxin-like GVDKE exemplified by CicContig5 and the toxins related by identity.

Toxin-like RFEC and toxin-like LFEC

We obtained precursor sequences with identity to insect or mammals toxin. Clone Cic667 showed a precursor sequence of 90 residues and CicContig7 (14 clones) and CicContig2 (3 clones) 87 residues, respectively. The mature sequences are identical to a group of spider toxin sequences toxic to crickets and mice; they contain a conserved motif IFEC or LFEC in the N-terminus sequence (Fig. 2). As shown in the alignment of the toxins in Fig. 2b, the sequences of this group are very similar and show several conserved regions.

Toxin-like ACSKQ

Clone Cic279 encodes a precursor of 115 residues with a mature sequence of 67 residues. The motive ACSKQ is conserved in long insect toxins Magi15 and 16 from the spider *Macrothele gigas*. Cic279 showed identity of 54 and 64% to Magi-15 (GenBank: P0C2V1) and Magi-16 toxins (GenBank: P0C2V2), respectively (Fig. 2).

Fig. 2 Multiple alignments of predicted mature sequences of toxin-like peptides from cDNA library from venom gland of *Citharischius crawshayi* (Cic). Accession number and name of the putative ICK peptides motif with three disulfide bridges or toxin-like with four disulfide bridges are included on the *left* with targets from experimental data reported (i.e., ion channels –ex vivo work–, or animal studies –in vivo work–); amino acid number and the percentage of identity with related toxins are shown on the *right* of the alignment. *CaCh* effect on calcium channel, *KCh* effect on potassium channel, *TRPV* effect on vanilloid receptors, *NaCh* effect on sodium channel, *nAChR* effect on nicotinic acetyl choline receptors, *toxcric*, peptide toxic to crickets; *toxmice*, peptide toxic to mice; *toxfly*, peptide toxic to house fly. Abbreviation *pre*, precursor sequence or amino acid sequence deduced from cDNA or gene. Sequence in *bold* corresponding to toxins or putative mature sequences. *Lower cases* in the amino acid sequences show sequence of pro-peptide that did not correspond to the mature peptide; sequences of mature peptides predicted are in *bold*; *filled circles* indicate post-translational modification of precursor for amidation of the last residue. Mature peptides were used to obtain theoretical mass data and compared with our mass results of Table 2. ICK and DDH-motif toxins group are indicated in *red*, conforming of cysteine residues are designed by roman numerals [32]

Toxin-like CIGES

CicContig12 encodes a precursor of 114 residues; the mature sequence shows 64% identity with the insect toxin Bs1 (GenBank: B3FIV1.1) from *Brachipelma smithi* spider (Fig. 2). Our precursor sequence shows a glycine residue at the 3′ end similar to the Bs1 precursor.

Peptide resemblance with secapin

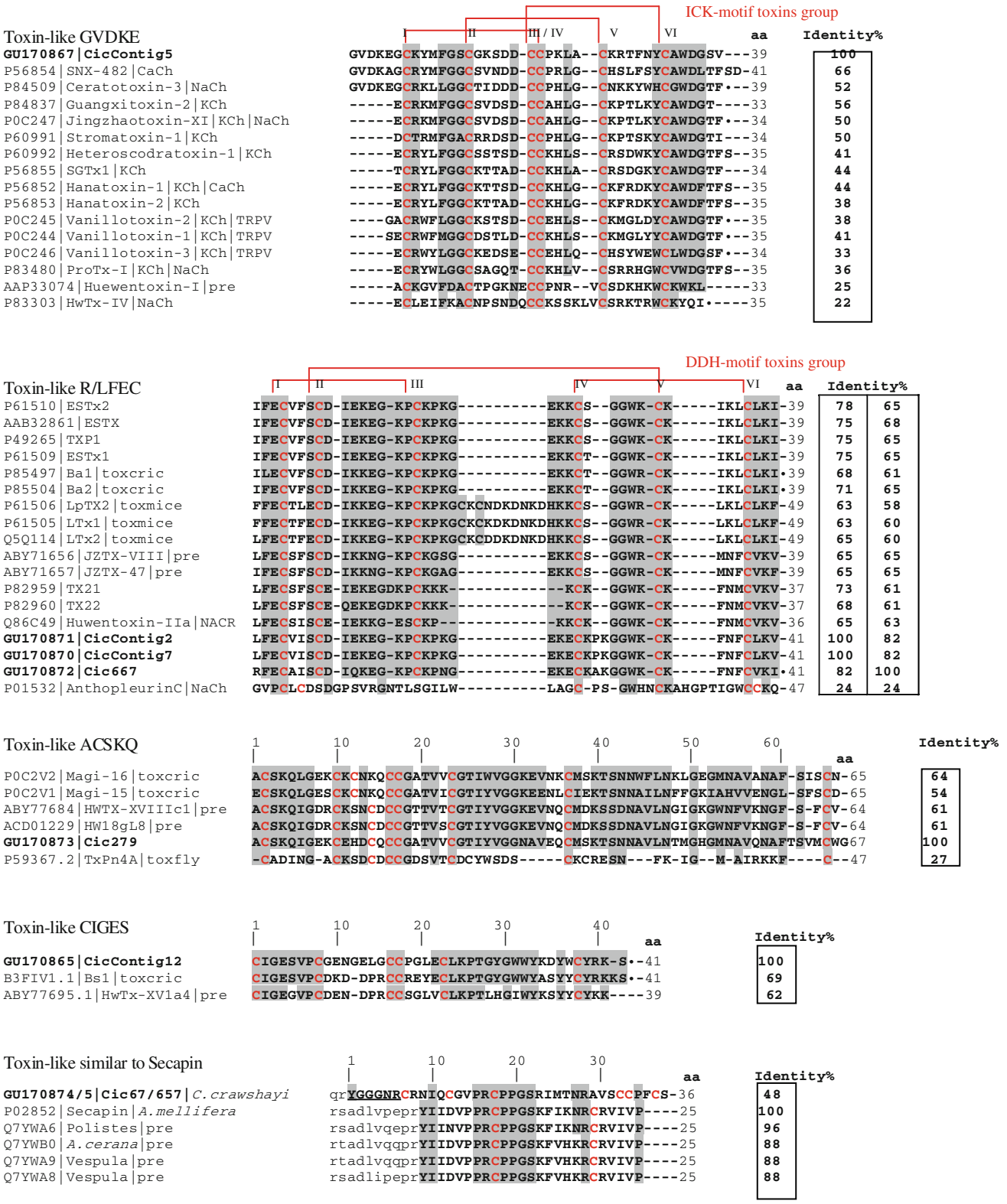
Toxin-like sequences of clones Cic67 and Cic657 encode a precursor with 62 residues and 52% identity to the precursor of secapin from *Apis mellifera* venom (GenBank: P02852). Clones Cic67 and Cic657 encode the same precursor sequence but showed some differences in the UTR-3′ regions of the genes. Figure 2e shows the alignment of the putative mature sequence similar to secapin and related genes.

Putative toxin genes

This category includes three precursor sequences of CicContig1 (11 clones), CicContig10 (two clones), and Clone Cic277. The sequences show an ICK-motif or low identity to venom component-related toxins. These precursor sequences are shown in Table 1.

Orphan and cell functions genes

After comparative analysis of sequences, we identified genes possibly related to different cellular functions, uncharacterized and new genes. ESTs from *C. crawshayi* show a high similarity with genes or peptides from spiders



and proteins families of other arthropods. We obtained 18 genes (28 clones) that correspond to different proteins with cellular functions in the venom gland, representing 41% of the total genes obtained from cDNA library ("CellPro"). Sequencing and comparative analyses of these sequences indicates that several components correspond to known functions. The partial distribution of contigs and singlets is represented by CellPro genes related to: (1) peptidases:

three genes (7%); (2) cell structure and motility elements: three genes (7%); (3) cell signaling and communication: one gene (2%); (4) metabolism: six genes (13%); (5) RNA and protein processing: four genes (9%); and (6) viral proteins: one gene (2%). The abundance and distribution of genes and the number of ESTs are shown in Fig. 1. Additionally, we obtained four genes that are unrelated with cell functions and that are considered "unknown" functions/hypothetical proteins. The unknown gene Cic69 shows 67% identity with the hypothetical spider protein from *Ixodes scapularis* (GenBank: EEC17710) and 55% identity with the *A. mellifera* protein (GenBank: XP_001120522), with conserved domains in the amino end region. The categories "no match" and "No ORF" represent the number of clones with an elevated proportion of ESTs that encode the same gene (39 and 32 clones, respectively). As shown in Table 1, CicContig13 is an ORF from 39 clones that encodes a small gene with no match in database and with two cysteines.

Table 1 shows precursor sequences of different categories and cellular functions. In some cases, we could not determine the residues from the nucleotide sequences obtained. These residues are indicated as "X" in the precursor sequence. Complete sequences include a putative start codon (methionine) and a stop codon (asterisk symbol). We included ORFs of "incomplete" genes obtained from the cDNA library by similarity of sequence size in homologous genes.

Characterization of components from venom and biological effects

Figure 3 shows the HPLC profile from the whole venom of *C. crawshayi.* The fractions are numbered and those indicated with an asterisk were selected for a second separation procedure. All fractions obtained were collected and finally vacuum-dried. Dry HPLC fractions were solubilized in distilled water for protein quantification. Thirty of these fractions were analyzed mass spectrometrically with MALDI-TOF/TOF. Fractions with most abundant values of protein were dried and used to biological assays and mass spectrometry analysis (Fig. 3e).

Total venom and fractions were suspended in ND96 solution and injected to insecticidal activity toward crickets or electrophysiology assays (Fig. 4; Table 2). Spider total venom from *C. crawshayi* was toxic to crickets at 10–30 μg. Paralysis effect on crickets was observed from 0.5 μg/g of cricket (10 μg) and complete paralysis at 1.5 μg/g (30 μg). Additionally, we observed that total venom modulated sodium currents in ion channels expressed in oocytes. Figure 4 shows selected current traces of the effects of total venom: modulation of $DmNa_v1$ and $Na_v1.6$ channels was observed. Cell disruption (cytolytic effect) was also observed at 30 μg. Table 3 summarizes and compares the effect of total venom from *C. crawshayi* and of *P. cavimanus* scorpion, as obtained in *Lepidoptera* and in *Xenopus* oocytes. We tested the effect of 16 fractions from venom in oocytes and some also in crickets. The fractions were tested on $DmNa_v1$ and $K_v1.3$ channels, but did not show any blocking activity at a concentration of 5 μg. The numbers of fractions used in these biological assays are indicated in bold in Table 2.

Venom analysis by MALDI-TOF/TOF

We obtained more than 600 values of molecular mass from the different venom fractions analyzed in the range of 902.43 to 15,685.31 Da (Fig. 3e; Table 2). The mass range of components corresponds to the following categories: (1) 0–3,000 Da, 180 different mass values (29% of total values of mass); (2) 3,001–6,000 Da, 357 molecular masses (58%); (3) 6,001–9,000 Da, 52 mass values (9%); (4) 9,001–12,000 Da, 18 mass values (3%) and (5) 12,001–20,000 Da, three mass values (0.5%). As shown in Table 2, we obtained some identical mass values in different fractions and neighboring fractions. We discarded similar values with a discrepancy of mass value of <0.05–0.90 Da by considering the possibility that they belong to the same component. In total, our result shows 610 different molecular masses that correspond to 30 different HPLC fractions from the crude venom.

Comparative data from transcriptomics and partial venomics analyses

We used the prediction of signal peptide, information of cleavage site of peptidases to describe the signal and propeptide sequence. The additional information of post-translational modifications in spider toxin was also used to obtain the mature sequence and theoretical mass value. After these analyses, we obtained the theoretical mass of putative new peptide sequences from *C. crawshayi.* Comparative data showed "matches" that correspond to the experimental and theoretical mass values from four of our toxin-like peptides. Mass values of toxin-like peptides are shown in Table 2. The matches correspond to the mature sequence (see sequences shown in bold, Fig. 2): (1) toxin-like GVDKE (CicContig5, precursor sequence GVDKE…GSV) with a theoretical mass of 4,267.89 Da and corresponding to the experimental molecular weight of 4,267.85 Da (venom fraction Cic96); (2) toxin-like LFEC (CicContig7 and Contig2, precursor sequence LFEC…LKV) with a theoretical mono-isotopic mass and experimental values of 4,669.33 Da and 4,669.08 Da, respectively, found in fraction Cic59.74); (3) toxin-like RFEC (Clone Cic667) with a theoretical mono-isotopic mass 4,643.27 and experimental mass 4,643.68, found in fraction Cic62.87. The molecular weight

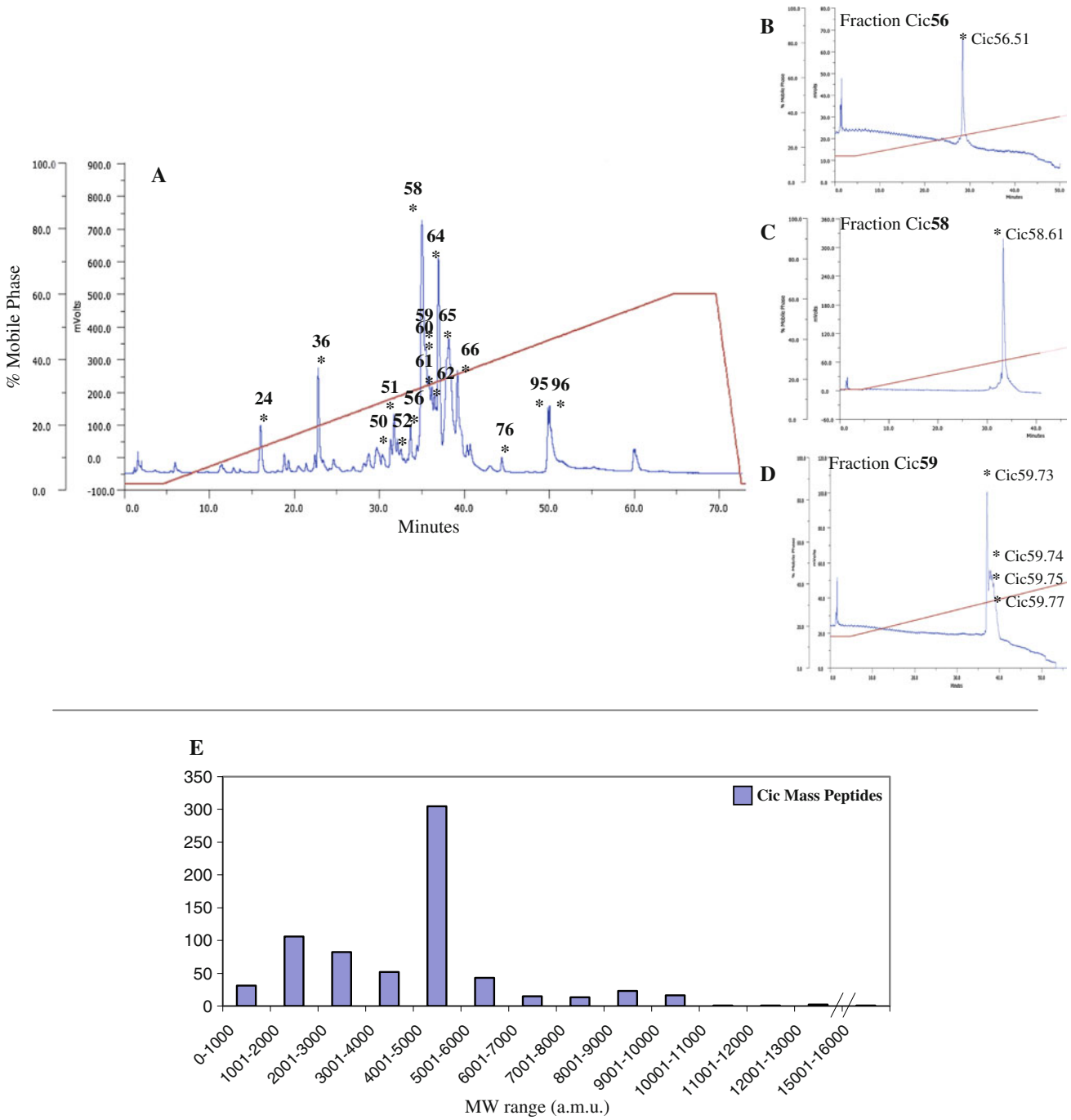


Fig. 3 Chromatogram of venom fractionation of *C. crawshayi* by HPLC. **a** Soluble venom (1 mg protein) was separated in a C18 analytic reverse-phase column, eluted with a linear gradient from 2% solution A (0.12% TFA in water) to 60% solution B (0.10% TFA in acetonitrile) run at a flow rate of 0.5 ml/min, for 70 min. Components labeled with an *asterisk* were selected for a second separation of fractionation. *Panels* **b**, **c**, and **d** show examples of some fractions selected and their second fractionation. **e** Relative abundance of venom mass peptides (frequency/mass)

theoretically expected for the first 41 amino acids of the mature peptide (corresponded region 49–89 of precursor sequence RFEC…VKI) and excluded the amino acid glycine before the stop codon, known to be a donor of amides for amidation of peptides at the C-terminal region. However, similar mass values of around 4,643.9 were found in some neighboring fractions (underlined in Table 2); and (4) toxin-like CIGES (CicContig12) whose mature sequence (residues 73–113 that corresponding CIGES…RKS sequence of precursor) show a theoretical molecular mass of 4,646.30 Da and an experimental mass of 4,646.41 Da, corresponding to fraction 65.122. The cDNA of CIGES toxin-like, contain a

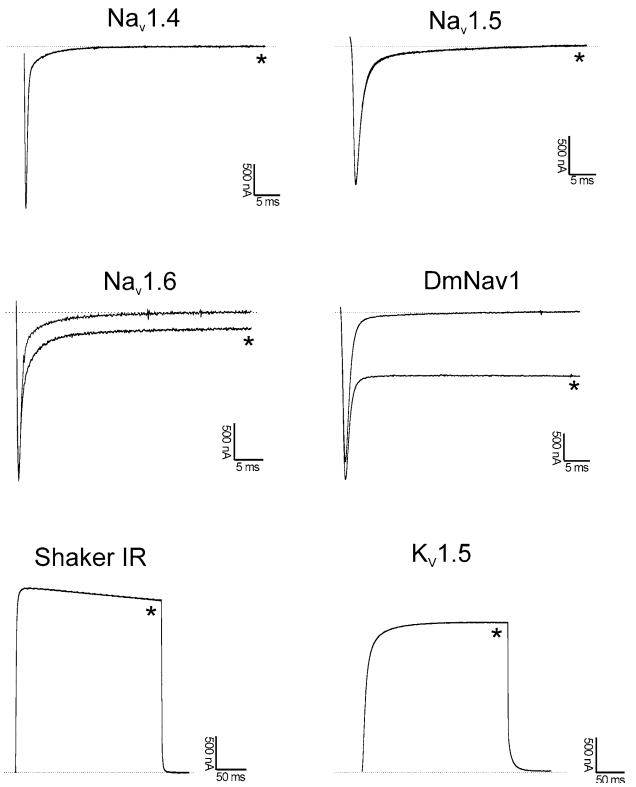


Fig. 4 Electrophysiological effects of total venom on different voltage-gated ion channels expressed in *Xenopus* oocytes. *Panels* show the traces of the effect of total venom in currents of oocytes expressing the $Na_v1.4$, $Na_v1.5$ $Na_v1.6$, $DmNa_v1$ (*Drosophila melanogaster* voltage-gated sodium channel), *Shaker* IR and $K_v1.5$. The *dotted line* indicates the zero-current level. The *asterisks* represent the effect of venom at 20 µg

C-terminal glycine alone, which presumable is cleaved to yield a C-terminal amidated serine as a result of post-translational modification.

Discussion

The generation of several EST sequences, combined with the analyses of venom by biological and electrophysiology assays, highlights the potential of venom compounds from *C. crawshayi* spider with potential therapeutic applications. Our results indeed show new but also very similar gene sequences as the (few) ones described in other spider venoms in general thus far. In particular, we found some genes encoding peptides related to toxins and different cell function proteins, of which the primary structures and/or functions have been shown in Fig. 1, 2 and Table 1. Typically, the precursor of spider toxins contains an N-terminal signal of 15–25 residues (signal peptide), followed by a region of variable length that contains a pro-peptide (of 15–60 residues) that includes the toxin or mature sequence of approximately 35–45 amino acids. Very often, an ICK-motif with multiple disulfide bridges is found in the mature sequence [18]. Normally, it can be anticipated that a precursor sequence containing a C-terminal amino acid sequence with a glycine residue followed by a single or double basic amino acid residue, which is the endoproteolytic site for the amidation signal, undergoes a post-translational modification to become amidated [19]. It can be noted that some of the precursor sequences from *C. crawshayi* possess the characteristics previously described for the precursors of toxins and conserved sequences similar or identical to genes and proteins from other arachnids or arthropods.

Toxin like-peptides

Eleven new toxin-like genes encode six new peptides. These sequences were analyzed and compared to similar sequences of toxins peptides. Toxin-like GVDKE (CicContig5 and gene-related) from *C. crawshayi* show the same mature sequences that correspond to toxin SNX-482 from *Hysterocrates gigas*, a selective blocker of the class E Ca^{2+} channel [20] and Ceratotoxin-3 from *Ceratogyrus cornuatus* that acts on voltage-gated Na^+ channels [21]. We found 13 genes (included in four different genes from two contigs and two singlets) that encode the same mature sequences of toxin-like GVDKE; this group of genes has the most abundant number of ESTs (27 sequences) in *C. crawshayi* transcriptome of toxin-like peptides. Figure 2a shows the alignment of mature sequences deduced from the four genes (exemplified by CicContig5) and related toxins. Cheng et al. [11] reported several cystine knot toxin (CKT) precursors from spider *C. jinzhao* and they described groups of several gene families, with family E being the largest family of CKT in the venom from *C. jinzhao*. Toxin-like GVDKE sequences posses 43–50% identity with precursors of family E. Furthermore, jingzhaotoxin-XI is a peptide included in this family E and it has been described both as a modulator of Kv2.1 channels as well as a gating modifier of rat cardiac sodium channels [22]. The alignment of toxin-like GVDKE sequences in Fig. 2 also clarifies the physiological effect described for the related toxin. We observed that this group includes toxins known to be gating modifiers albeit targeting different ion channel types. As such, they constitute a fine example of the promiscuity present in spider toxins.

Several insecticidal peptides from *Brachipelma* spp. have been described; toxin Bs1 from *B. smithi* [23] and the peptides Ba1 and Ba2 from *B. albiceps* [24] are peptides which have been shown to be toxic for crickets, but not for mice. An obvious difference after alignment of the peptides toxin-like R/LFEC with toxins LpTx2, LTx1 and LTx2, can be found in the additional ten amino acids in the sequence of the latter three toxins. Interestingly, LpTx2,

Table 2 Mass spectrometry analysis of HPLC fractions from *C. crawshayi* venom

Fraction	Experimental Molecular mass (a.m.u)	Fraction	Experimental Molecular mass (a.m.u)
24.00	1234.85, 1262.95, 1306.94, 1363.04, 1391.08, 1412.99, 1435.03, 1450.94, 1462.05, 1704.09, 1844.29, 3000.84, 4193.53, 4223.60, 4302.48, 4596.73, 4615.05, 4671.89, 4694.08, 4808.35, 4998.17, 5237.54, 5277.38, 6024.71, 6501.18, 7437.51, 8500.76	61.112•	908.87, 2323.43, 2336.43, 2351.93, 2362.43, 4108.52, 4615.06, 4627.06, 4644.00, 4650.95, 4661.95, 4670.00, 4676.91, 4685.97, 4692.87, 4701.92, 4715.92, 4735.90, 4758.89, 5047.56, 6158.89, 8126.63,
36.00	1012.34, 1034.35, 1056.39, 1100.38, 1209.44, 1554.38, 2297.68, 2308.72, 2314.67, 2330.68, 2336.68, 2354.59, 2374.63, 2380.66, 2390.61, 2396.60, 2402.63, 2412.61, 2418.59, 2424.59, 2434.64, 2447.62, 9402.83	61.113°•	1029.51, 2323.44, 2336.44, 2352.45, 3467.23, 4606.95, 4638.08, 4643.99, 4676.94, 4650.98, 4661.95, 4686.93, 4702.92, 4719.89, 4726.92, 4832.02, 5048.67, 7186.94
51.00	1208.41, 1283.41, 1448.42	62.83•	927.48, 983.50, 999.51, 1029.50, 1185.59, 1225.62, 1308.65, 1441.70, 1814.93, 2323.38, 2336.89, 4578.04, 4601.66, 4626.98, 4643.83, 4650.82, 4658.86, 4669.85, 4674.83, 4680.84, 4708.77, 4721.85, 5048.52, 5707.43
52.65	1744.77, 2889.2, 3019.28, 3147.37, 3174.45, 3194.5, 3198.53, 3215.44, 3217.42, 3398.55, 3415.49, 3453.61, 3471.59, 3478.59, 3486.56, 3491.53, 3540.46, 3915.93, 4393.08, 5058.74, 5064.78, 5354.14, 5471.57, 5524.30, 5589.63, 7194.14, 9336.19	62.84•	902.43, 2323.33, 2336.35, 2351.82, 4110.23, 4546.78, 4575.68, 4615.78, 4637.92, 4643.77, 4650.73, 4657.74, 4662.72, 4676.69, 4692.67, 4733.58, 5048.78, 5629.78, 6410.66, 8674.86, 9295.67, 9302.99, 9520.57, 9644.23
56.51°•	909.01, 924.99, 1497.43, 1923.89, 2245.16, 2261.15, 2896.64, 3849.8, 3830.9, 3810.9, 4437.37, 4487.4	**62.87•**	972.44, 999.50, 2323.31, 2336.33, 2348.33, 2418.90, 4094.16, 4109.22, 4417.53, 4622.72, 4626.75, 4628.73, 4643.68, 4650.66, 4669.69, 4675.67, 4786.70, 5873.2, 7708.21
58.61•	948.47, 1037.48, 2148.97, 2156.48, 3415.6, 3766.57, 3864.7, 3928.2, 4015.8, 4168.07, 4246.2, 4262.24, 4277.2, 4282.08, 4299.08, 4313.06, 4322.03, 4348.5, 4408.6, 4449.3, 4497.5, 4575.1, 4647.8, 4771.0, 5337.1, 5389.1, 5427.9, 5576.9, 5815.0	64.67	908.86, 2323.33, 2336.34, 3998.18, 4106.34, 4109.28, 4179.23, 4571.67, 4617.66, 4627.82, 4643.76, 4651.70, 4659.73, 4670.71, 4096.26, 4699.58, 4732.56, 6622.06, 8069.57, 8269.18, 8714.95, 9322.46
59.73	1037.48, 1054.57, 2126.49, 2148.97, 2156.97, 2164.97, 3058.2, 3771.67, 4201.17, 4244.28, 4263.2, 4281.16, 4273.22, 4288.07, 4299.11, 4344.05	64.69°•	909.00, 1897.33, 2055.44, 2176.98, 2825.61, 3729.18, 3750.76, 3758.84, 3780.78, 3791.8, 3797.77, 4049.0, 4100.0
59.74	908.93, 2148.80, 2164.79, 3539.29, 4076.55, 4151.72, 4161.73, 4165.57, 4237.57, 4252.82, 4259.83, 4278.72, 4292.66, 4301.63, 4306.53, 4312.63, 4319.78, 4338.65, 4349.55, 4546.29, 4644.07, 4650.02, 4662.04, 4669.08, 4676.02, 4689.02, 4702.99, 4720.95, 5056.60, 5406.48, 8565.74, 8882.23, 12884.84	65.118°•	909.0, 1019.13, 1044.13, 1098.04, 1300.08, 1897.31, 1914.33, 1924.28, 3136.27, 3714.36, 3730.15, 3776.72, 3792.71, 3797.71, 3807.69, 3850.66, 4064.7
59.75	908.87, 2323.40, 4145.47, 4181.45, 4196.44, 4294.46, 4297.44, 4308.53, 4312.42, 4574.83, 4599.87, 4641.01, 4643.91, 4659.84, 4660.86, 4670.86, 4698.76, 4677.77, 4676.79, 4685.80, 4727.74, 5720.55, 5049.87, 5550.10, 8914.13, 8079.27, 8560.25, 9257.48	**65.122•**	908.84, 943.39, 994.0, 1007.43, 1065.91, 1093.44, 1103.86, 1224.99, 2075.67, 2322.70, 2339.18, 2440.81, 3777.28, 3825.08, 3884.43, 3948.05, 4080.18, 4102.16, 4151.17, 4282.25, 4646.41, 4661.4, 4675.41, 4662.42, 4674.43, 4682.43, 4687.39, 4692.38, 4705.38, 4734.39, 4759.31, 4625.57, 4655.35, 4720.33, 4848.48, 6354.0, 8730.41
59.77•	916.43, 2323.37, 2336.37, 2352.37, 4612.95, 4621.83, 4637.84, 4643.89, 4651.83, 4652.86, 4656.82, 4657.85, 4669.83, 4729.69, 4738.58, 5048.37, 6188.87, 7501.11	65.123•	902.56, 915.55, 959.58, 972.60, 1003.60, 1091.65, 1161.67, 1205.7, 1249.72, 1311.75, 1482.55, 1901.62, 1918.63, 2345.69, 2595.88, 4544.60, 4574.54, 4604.45, 4630.48, 4642.48, 4647.46, 4661.45, 4670.45, 4679.46, 4681.44, 4698.41, 4704.41, 4711.42, 4733.35, 4741.40, 4790.62, 4845.55, 5598.20, 8120.93
60.74	908.87, 918.83, 2148.69, 2164.68, 2175.64, 3687.19, 4108.33, 4160.49, 4181.34, 4206.64, 4230.66, 4250.54, 4257.49, 4261.57, 4279.54, 4292.53, 4298.48, 4299.43, 4310.54, 4313.45, 4321.48, 4326.48, 4331.39, 4352.30, 4422.42, 4650.79, 4672.82, 5048.34, 5552.92, 7005.74, 8628.91, 8653.49	67.00	1050.31, 1088.26, 1072.26, 1132.27, 1149.30, 1293.30, 1342.40, 3591.97, 4229.91
60.75•	908.87, 2323.41, 2148.70, 2352.92, 3884.39, 4178.40, 4294.45, 4301.47, 4572.84, 4627.03, 4643.91, 4649.88, 4656.92, 4660.87, 4669.90, 4676.84, 4687.84, 4627.95, 4700.76, 4730.81, 5045.76, 5505.66, 5632.6, 7833.32, 8067.95, 9264.51, 9286.72, 9382.07, 9553.0	95.00	1051.85, 1165.91, 1449.13, 1700.34, 1857.47, 3163.00, 4187.00, 4200.99, 4306.99, 4699.75, 4722.29, 4699.75, 4768.70, 4952.19, 5000.61, 5441.67, 5489.83, 5823.71, 6075.85, 6472.22, 6585.11, 6939.89, 8063.39, 8232.40, 8612.33, 9926.22, 11455.18, 15685.31
60.76•	941.44, 2323.43, 2336.43, 2346.92, 4280.64, 4609.15, 4628.07, 4637.02, 4643.98, 4650.93, 4651.92, 4669.92, 4676,87, 4677.83, 4703.81, 4744.88, 5043.54, 6103.78, 7685.27, 9250.87	**96.00**	1083.81, 1111.85, 1199.93, 1259.95, 1303.98, 1317.98, 1332.03, 1348.02, 1362.01, 1376.07, 1406.05, 1420.10, 1494.12, 1714.28, 1758.30, 1862.37, 1962.45, 2282.66, 2493.83, 2681.98, 2947.16, 4267.85, 4268.72, 4301.52, 4535.05, 4642.14, 4645.29, 4651.09, 4664.37, 4675.85, 4676.26, 4671.36, 4716.65, 4821.39, 5018.97, 5268.69, 5435.05, 5470.52, 5814.85, 6524.56, 7434.50, 7584.97, 9087.32, 10283.02, 12896.60
60.77	908.88, 2323.44, 2351.92, 4162.49, 4597.27, 4614.04, 4640.05, 4643.99, 4648.89, 4654.79, 4669.90, 4660.90, 4672.79, 4685.79, 4698.72, 4724.74, 5045.04, 6862.21, 8498.77, 9331.18	97.00	1034.34, 1078.32, 1094.31, 1110.28, 1143.36, 1305.36, 1321.34, 3447.79, 4881.52, 7233.73, 7383.12, 8366.15
61.111•	908.87, 924.84, 999.52, 1061.47, 1119.86, 1532.66, 1737.79, 2148.72, 2164.72, 2325.92, 2339,96, 2356.78, 3969.48, 4100.27, 4110.41, 4196.25, 4230.68, 4247.70, 4253.68, 4265.49, 4282.60, 4283.59, 4291.62, 4293.73, 4295.55, 4301.50, 4310.55, 4326.52, 4327.49, 4381.51, 4440.68, 4563.77, 4644.92, 4720.94, 5045.15, 5352.16, 7011.37, 8463.34	137.00	1012.31, 1089.94, 1106.51, 1177.98, 1222.01, 1283.49, 1357.52, 1398.16, 1487.19, 1663.39, 1802.42, 2506.96, 2971.77, 4914.34, 6598.02, 8498.72

Total mass data of 30 fractions are shown; masses that correspond to the similar theoretical values of predicted peptides deduced from genes of the cDNA library are shown *underlined*. The experimental masses are listed as the $(M + H)^+$ values as measured in positive ion mode

Fractions (i.e., numbers) in *bold* corresponded to match of putative toxin-like, mass values are shown *highlighted*. Comparative theoretical and experimental values are the following: FracCic 59.74, toxin-like LFEC (4669.33/4669.08); FracCic 62.87, toxin-like RFEC (4643.27/4643.68), similar mass values were observed in adjacent fractions Cic61.113, 62.83, 62.84 and 64.67, they are shown *underlined*; FracCic 65.122, toxin-like CIGES (4646.30/4641.41); FracCic 96.00, toxin-like GVDKE (4267.89/4267.85). Additionally, *open circles* in the fractions numbers correspond to fractions tested at 10 μg of protein on crickets (20 mg/w); *closed circles*, fractions tested in oocytes at 5 μg ($DmNa_v1$ and $K_v1.3$ channels)

Table 3 Biological effects of total venom from *C. crawshayi* on crickets and ion channel expressed in oocytes of *Xenopus*

	Venom amount (μg)	Effect on crickets[a]	DmNa$_v$1.1[b]	mNa$_v$1.6	Shaker IR[b]	rK$_v$1.5[b]
C. crawshayi	10	++	N.T.	N.T.	N.T.	N.T.
C. crawshayi	20	++	+[c]	+[c]	–[c]	–[c]
C. crawshayi	30	+++	D	D	D	D
P. cavimanus	7	++++	D	D	N.T.	N.T.
P. cavimanus	30	++++, L				

Effect on crickets is indicated with increasing plus signs as follows: ++ incomplete paralysis at concentration tested within 60 min; +++ complete paralysis at any range of concentration tested within 30 min; ++++ toxicity and shaking of legs after 5–10 min

The groups tested included 3-4 crickets (approximate 20 mg/weight each organism)

L lethal effect at any range of concentration tested within 90 min

D disruption membrane effect in oocytes

– represents no effect

NT non-tested

[a] Qualitative assay, [b]Electrophysiological assays, [c]Plus and minus signs indicate positive or lack of activity, respectively

LTx1 and LTx2 are toxic to mice, whereas the other toxins are not. This observation can be interpreted to mean that motif "21-CKCXDKDNKD-32" plays an important role in the toxins' selectivity profile. Toxin-like sequence CIGES (CicContig12) and toxin-like RFEC (Clone Cic667) are sequences with homology to toxins Bs1, Ba1 and Ba2; experimental data of the precursor sequences of these toxins predicts post-translational modification at the C-terminal (i.e., amidation). Our results indeed show presumed sites of post-translational modification by amidation of glycine in precursors of toxin-like RFEC and CIGES (see alignments of Fig. 2, letter "g"). It is suggested that sequences at the C-terminus may be involved post-translational modifications as in the case of Bs1 and related toxins.

The precursor of a toxin-like peptide similar to secapin (spider secapin), shows 54% identity with peptide secapin (GenBank: P02852). Secapin has been estimated to be 1% of bee venom, not being toxic to mice at a concentration of 10–20 μg/g. The only effect reported is an apparent sedation in these animals at a concentration of 40 μg/g [25]. Very interestingly, the sequences of these two peptides and similar precursor sequences from venom gland of insects, share a very conserved motif of seven residues rich in proline: –PRCPPGS– (Fig. 2). Although perhaps speculative at present, we believe that our finding of such conserved motif in two peptides found in the venom of an insect and an arachnidae may be important in the light of evolutionary biology of the animals belonging to different (and often competing) phyla. Apart form that, secapin contains only two cysteines, while the precursor sequence from clone Cic67/657 contains six cysteines.

Antimicrobial activity has been described for peptides from spider venoms and other arthropods. N-terminal regions with Gly-rich domains are related with the antimicrobial activity [26]. CicContig14 and spider secapin sequences show regions enriched in Gly residues; which may suggest that these peptides also contain antimicrobial activity. However, for the moment the precise function remains to be proven, given the limitation of venom amount.

Putative toxin genes (Clone Cic277, CicContig1 and 10) encode precursors rich in cysteine and show a putative ICK-motif. We considered these sequences as putative toxin in spite of the low similarity with other precursors of toxins from GenBank and the ArachnoServer database.

Others genes

We obtained homologous sequences of genes that encode different proteins involved in cellular processes. Our result indeed shown a variety of components related to cellular functions; the high percentage of identity and conserved domains contribute to the different categories of proteins described in the literature. Here below, we try to emphasize the information that was obtained for genes encoding highly conserved proteins, related to gene/protein processing. Furthermore, we obtained interesting ESTs and complete genes to encode precursors of proteins with conserved domains associated to a calcium binding motif, muscular proteins and enzymes.

Peptidases

Peptidases are fundamental pieces in the processing of peptides and their activity in diverse cell functions. In the precursor sequences, peptidases remove the signal peptide and catalyze the maturation of peptides by post-translational processing of precursor peptides. We obtained three different sequences that correspond to peptidases and two

precursors that are similar to putative secretory proteins reported in ESTs from other spiders (Cic643 and Cic243). At present, these sequences have been considered hypothetical proteins and represent an orphan family of genes.

Histone

Histone H4 is one of the four histones, which form the eukaryotic nucleosome core. Modifications of the histone proteins by acetylation regulate numerous processes related of nucleosome assembly. Histone H4 acetylation regulates the degree of chromatin compactness, the formation of heterocromatin and acetylation, critical processes to gene transcription [27]. Clone Cic630 encodes histone H4, a protein extremely conserved in all organs over different phyla. The identity of Cic630 with several genes corresponding with histones from arthropods and others phyla is between 98–100% (expect = $3e^{-38}$, e.g., *Aedes aegypti* (GenBank: EAT44828.1); *Culex tarsalis* (GenBank: ACJ64450.1).

Actin

Actin is the main component protein involved in the formation of filaments of the cytoskeleton. The actin gene has been used in phylogenetic studies of spider in the past [28]. Clone Cic253 encodes an actin gene (1.2 kb), which was partially sequenced (790 bp). The fragment of 263 deduced residues shows 95% identity ($1e^{-143}$) with an actin gene from *Paraphidippus aurantius* (GenBank: ABZ91674) and *Lycosa singoriensis* spiders (GenBank: ABX75378.1) and the arachnid, *Ornithodoros moubata* (GenBank: AAS55945). Recently, a report of several ESTs obtained from cDNA of muscle from *Aphonopelma sp.*, showed a complete pro-peptide sequence of actin denominated ACT1_As (376 residues). This EST was an abundant sequence with a relative proportion of 14% [16]. Our Clone Cic253 shows 98% identity with ACT1_As.

Ribosomal and Cox genes

Information of ribosomal and cytochrome C oxidase (*COX*) genes have been an important contribution in the studies of phylogenetic relationships at several taxonomic levels. We obtained some ESTs that correspond to partial genes *Cox1* and *Cox2* (encoding the *COX* subunit I and II). Additionally, we found a cluster of genes to the 40S ribosomal protein S23 (CicContig3); it shows 95% identity ($1e^{-72}$) with ribosomal protein S23 from *A. mellifera* (GenBank: XP_624651.29 and *Ixodes scapularis* (GenBank: EEC16158.1). Mitochondrial and nuclear gene sequences have been used as phylogenetic markers; *Cox*, ribosomal, histone and actin genes from *C. crawshayi* may be provide information useful to phylogenetic studies of spiders and arthropods.

Biological activity

The venom analysis shows toxic activity on crickets, intra-abdomen injection of total venom causes paralysis 30 min after application at 30 μg, but was not lethal during the 7 h of the entire experiment. Electrophysiological assays show that the venom (20 μg) produces modulation of currents, in particular of insect sodium channels ($DmNa_v1$) but also $mNa_v1.6$ channels, expressed in *Xenopus* oocytes. However, the venom did not affect the potassium channels *Shaker* IR and $K_v1.5$. Disruption of the oocytes' membranes at 30 μg crude venom application was observed. Table 2 shows the biological effects of total venom. Unfortunately, selected samples from venom fractionation tested on crickets or $DmNa_v1$ and $K_v1.3$ channels, did not show effect in contrast to our bioassays with total venom. We presume that the absence of an effect can most likely be attributed to the (very) low concentration of toxic compounds available in the samples tested (5 μg of protein).

Mass analysis of venom fractions

Spider venoms are complex mixtures with hundred of components [4], proteomic and peptidomic analyses from spider venoms indicate that the peptides in tarantula spiders are 31–40 and 58–63 amino acids in length [29], are associated with mass values of 3–7 kDa, and also correspond to the toxic and insecticidal components in spider venom [3, 11]. MALDI-TOF/TOF is known as a very sensitive method that allows direct analyses of complex mixtures. Table 3 shows total mass values obtained from 30 venom fractions as analyzed by MALDI-TOF/TOF mass spectrometry. We found 694 molecular masses and redundancy in mass values for some HPLC peaks, where these values probably correspond to the same component as a result of incomplete separation of fractions. This effect may also concern partially folded proteins as a result of their non-homogeneous elution in the HPLC separation, as observed in mass fingerprint analyses of scorpion venom [30, 31]. Escoubas et al. (2006), detected more than 1,000 masses from spider venom. In our case, we describe a smaller number of masses, probably as a consequence of unexplored fractions from *C. crawshayi* venom. It should be kept in mind that the number of masses can be related to a differential processing of peptides and pro-peptides, like post-translational modifications. Post-translational modifications are indeed reported in several toxins and are also predicted to occur from the detailed sequences of the genes we obtain for arachnids. These elements show the existence

of even more complex toxin compounds than anticipated thus far, but at the moment it still is not clear which processes and modifications precisely are governing all these steps.

Our results show 610 distinct molecular masses from the venom of *C. crawshayi* with a range of mass compounds from 902.43 to 15,685.31 Da, which is similar to the mass compounds of venom spiders previously characterized. *C. crawshayi* venom contains a very high proportion of peptides in the 3–6 kDa (58%) range, a significantly proportion of lower molecular weight components in the range of 1–3 kDa (29%) and very few mass components higher than 12 kDa (0.5%). In this aspect, our results are similar to previous work using mass spectrometry in spider venom research [4, 6, 10, 11]. The analysis revealed different peptides with the potential of different levels of expression of mature peptides and non-processed peptides precursors of venom glands. Figure 3e shows molecular weight distribution of Cic peptides (frequency in 1,000 a.m.u.) revealing a distribution with peptides falling in the central class of 4,000–5,000 Da (305 mass values). Peptides with masses of 3,500–4,500 Da normally correspond to different structural motifs characterized by different cysteine arrangement and structural features: either the ICK-motif, or a disulfide bond pairing and DDH (disulfide-directed beta-harping)-motif lacking the cystine knot and comprising an antiparallel beta-harping stabilized by two mandatory disulfide bridges [29, 32]. The sequence of toxin-like GVDKE and LFEC can be clearly considered sequence members of the ICK-motif and DDH-motif, respectively (see Fig. 2), by similarities of known tarantula toxins [32].

Prediction of biological activity of putative toxins and unknown function genes remains a difficult task. However, the combinations of methods and tools in transcriptomics and proteomics may facilitate unraveling the venom composition and information of the molecular components. Exact mass measurements of venom components permit the confirmation of primary sequences obtained by Edman degradation (direct sequencing) or sequence deduced from cDNA. In this work we included correlations between (1) molecular weights and (2) theoretical masses from putative mature sequences of toxin-like peptides from ESTs and their correlation with experimental mass spectrometry results. We presented here, to the best of our knowledge, for the first time, the general venom characterization and transcriptome analysis of *C. crawshayi*. The generation of several EST sequences and the analyses of venom by biological and electrophysiology assays, highlight the potential of venom compounds and toxin genes. This study thus can help us to explore novel high-affinity ligands (toxins) targeting ion channels and receptors that could be valorized at a later stage in channelopathies or as diagnostic markers.

Acknowledgments We are grateful to Thomas Vandendriessche for the technical assistance on the milking of spiders. This work was supported in part by grants K.U. Leuven (OT/05/64), G.330.06 from the Fund for Scientific Research (FWO)-Flanders and Belgian Federal Science Policy IAP, P6/31 (Interuniversity attraction Poles Program-Belgian State- Belgian Science Policy) and Postdoctoral Research Fellowship from K.U. Leuven to EDG.

References

1. Platnick NI (2009) The world spider catalog, version 9.0. In: Merrett P, Cameron HD (eds) American Museum of Natural History, New York. Online at http://research.amnh.org/entomology/spiders/catalog/index.html
2. Corzo G, Escoubas P (2003) Pharmacologically active spider peptides toxins. Cell Mol Life Sci 60:2409–2426
3. Escoubas P, King GF (2009) Venomics as a drug discovery platform. Expert Rev Proteomics 6:221–224
4. Escoubas P, Sollod B, King GF (2006) Venom landscapes: mining the complexity of spider venoms via a combined cDNA and mass spectrometric approach. Toxicon 47:650–663
5. Wood DL, Miljenovic T, Cai S, Raven RJ, Kass Q, Escoubas P, Herzig V, Wilson D, King GF (2009) ArachnoServer: a database of protein toxin from spiders. BMC Genomics 10:375. doi: 10.1186/1471-2164-10-375
6. Escoubas P, Chamot-Rooke J, Stöcklin R, Whiteley BJ, Corzo G, Genet R, Nakajima T (1999) A comparison of matrix-assisted laser desorption/ionization time-of-flight and liquid chromatography electrospray ionization mass spectrometry methods for the analysis of crude tarantula venoms in the Pterinochilus group. Rapid Commun Mass Spectrom 13:1861–1868
7. Machado LF, Laugesen S, Botelho ED, Ricart CA, Fontes W, Barbaro KC, Roepstorff P, Sousa MV (2005) Proteome analysis of brown spider venom: identification of loxnecrogin isoforms in *Loxosceles gaucho* venom. Proteomics 5:2167–2176
8. Guette C, Legros C, Tournois G, Goyffon M, Célérier ML (2006) Peptide profiling by matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry of the *Lasiodora parahybana* tarantula venom gland. Toxicon 47:640–649
9. Liao Z, Cao J, Li S, Yan X, Hu W, He Q, Chen J, Tang J, Xie J, Liang S (2007) Proteomic and peptidomic analysis of the venom from Chinese tarantula *Chilobrachys jingzhao*. Proteomics 7:1892–1907
10. Yuan C, Jin Q, Tang X, Hu W, Cao R, Yang S, Xiong J, Xie C, Xie J, Liang S (2007) Proteomic and peptidomic characterization of the venom from the Chinese bird spider, *Ornithoctonus huwena* Wang. J Proteome Res 6:2792–2801
11. Gentz MC, Jones A, Clement H, King GF (2009) Comparison of the peptidome and insecticidal activity of venom from a taxonomically diverse group of theraphosid spiders. Toxicon 53:496–502
12. Chen J, Deng M, He Q, Meng E, Jiang L, Liao Z, Rong M, Liang (2008) Molecular diversity and evolution of cystine knot toxins of the tarantula *Chilobrachys jingzhao*. Cell Mol Life Sci 65:2431–44
13. Chen J, Zhao L, Jiang L, Meng E, Zhang Y, Xiong X, Liang S (2008) Transcriptome analysis revealed novel possible venom components and cellular processes of the tarantula *Chilobrachys jingzhao* venom gland. Toxicon 52:794–806
14. MdeF Fernandes-Pedrosa, IdeL Junqueira-de-Azevedo, Gonçalves-de-Andrade RM, Kobashi LS, Almeida DD, Ho PL, Tambourgi DV (2008) Transcriptome analysis of *Loxosceles laeta* (Araneae, Sicariidae) spider venomous gland using expressed sequence tags. BMC Genomics 12:279. doi:10.1186/1471-2164-9-279
15. Jiang L, Peng L, Chen J, Zhang Y, Xiong X, 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
16. Zhu J, Sun Y, Zhao FQ, Yu J, Craig R, Hu S (2009) Analysis of tarantula skeletal muscle protein sequences and identification of transcriptional isoforms. BMC Genomics 10:117. doi: 10.1186/1471-2164-10-117
17. Zhang Y, Chen J, Tang X, Wang F, Jiang L, Xiong X, Wang M, Rong M, Liu Z, Liang S (2009) Transcriptome analysis of the venom glands of the Chinese wolf spider *Lycosa singoriensis*. Zoology. doi:10.1016/j.zool.2009.04.001
18. Sollod BL, Wilson D, Zhaxybayeva O, Gogarten JP, Drinkwater R, King GF (2005) Were arachnids the first to use combinatorial peptide libraries? Peptides 26:131–139
19. Bradbury AF, Finnie MD, Smyth DG (1982) Mechanism of C-terminal amide formation by pituitary enzymes. Nature 298:686–688
20. Newcomb R, Szoke B, Palma A, Wang G, Chen X, Hopkins W, Cong R, Miller J, Urge L, Tarczy-Hornoch K, Loo JA, Dooley DJ, Nadasdi L, Tsien RW, Lemos J, Miljanich G (1998) Selective peptide antagonist of the class E calcium channel from the venom of the tarantula *Hysterocrates gigas*. Biochemistry 37:15353–15362
21. Bosmans F, Rash L, Zhu S, Diochot S, Lazdunski M, Escoubas P, Tytgat J (2006) Four novel tarantula toxins as selective modulators of voltage-gated sodium channel subtypes. Mol Pharmacol 69:419–429
22. Liao Z, Yuan C, Deng M, Li J, Chen J, Yang Y, Hu W, Liang S (2006) Solution structure and functional characterization of jingzhaotoxin-XI: a novel gating modifier of both potassium and sodium channels. Biochemistry 45:15591–15600
23. Corzo G, Diego-García E, Clement H, Peigneur S, Odell G, Tytgat J, Possani LD, Alagón A (2008) An insecticidal peptide from the theraposid *Brachypelma smithi* spider venom reveals common molecular features among spider species from different genera. Peptides 29:1901–1908
24. Corzo G, Bernard C, Clement H, Villegas E, Bosmans F, Tytgat J, Possani LD, Darbon H, Alagón A (2009) Insecticidal peptides from the theraposid spider *Brachypelma albiceps*: an NMR-based model of Ba2. Biochim Biophys Acta 1794:1190–1196
25. Gauldie J, Hanson JM, Rumjanek FD, Shipolini RA, Vernon CA (1976) The peptide components of bee venom. Eur J Biochem 61:369–376
26. Sperstad SV, Haug T, Vasskog T, Stensvåg K (2009) Hyastatin, a glycine-rich multidomain antimicrobial peptide isolated from the spider crab (*Hyas araneus*) hemocytes. Mol Immunol 46:2604–2612
27. Shahbazian MD, Grunstein M (2007) Functions of site-specific histone acetylation and deacetylation. Annu Rev Biochem 76:75–100
28. Vink CJ, Hedin M, Bodner MR, Maddison WP, Hayashi CY, Garb JE (2008) Actin 5C, a promising nuclear gene for spider phylogenetics. Mol Phylogenet Evol 48:377–382
29. Escoubas P (2006) Molecular diversification in spider venoms: a web of combinatorial peptide libraries. Mol Divers 4:545–554
30. Pimenta AM, Stöcklin R, Favreau P, Bougis PE, Martin-Eauclaire MF (2001) Moving pieces in a proteomic puzzle: mass fingerprinting of toxic fractions from the venom of *Tityus serrulatus* (Scorpiones, Buthidae). Rapid Commun Mass Spectrom 15:1562–1572
31. Batista CV, Román-González SA, Salas-Castillo SP, Zamudio FZ, Gómez-Lagunas F, Possani LD (2007) Proteomic analysis of the venom from the scorpion *Tityus stigmurus*: biochemical and physiological comparison with other *Tityus* species. Comp Biochem Physiol C Toxicol Pharmacol 146:147–157
32. Escoubas P, Rash L (2004) Tarantulas: eight-legged pharmacists and combinatorial chemists. Toxicon 43:555–574