Toxicon 129 (2017) 81–88

Contents lists available at ScienceDirect

Toxicon

journal homepage: www.elsevier.com/locate/toxicon

Identifying different transcribed proteins in the newly described Theraphosidae *Pamphobeteus verdolaga*

Sebastian Estrada-Gómez [a, b, *], Leidy Johana Vargas-Muñoz [c], Mónica Saldarriaga-Córdoba [d, e], Yeimy Cifuentes [f, g], Carlos Perafan [h]

[a] Programa de Ofidismo/Escorpionismo – Serpentario, Universidad de Antioquia UdeA, Carrera 53 No 61-30, Medellín, 050010, Antioquia, Colombia
[b] Facultad de Ciencias Farmacéuticas y Alimentarias, Universidad de Antioquia UdeA, calle 70 No 52-21, Medellín, 050010, Antioquia, Colombia
[c] Facultad de Medicina, Universidad Cooperativa de Colombia, Calle 50 A No 41-20, Medellín, 050012, Antioquia, Colombia
[d] Departamento de Ciencias, Universidad Iberoamericana de Ciencias y Tecnología, Padre Miguel de Olivares, 1620, Santiago, Chile
[e] Universidad Bernardo O'Higgins, Centro de Investigación en Recursos Naturales y Sustentabilidad, Fábrica, 1990, Segundo Piso, Santiago, Chile
[f] Laboratório Especial de Ecologia e Evolução, Instituto Butantan, Avenida Vital Brazil 1500, 05503-900, São Paulo, SP, Brazil
[g] Programa de Pós-Graduação em Biologia Animal, Universidade Estadual Paulista "Júlio de Mesquita Filho", São José do Rio Preto, SP, 15054-000, Brazil
[h] Sección Entomología, Facultad de Ciencias, Universidad de La República, Iguá 4225, 11400, Montevideo, Uruguay

ARTICLE INFO

Article history:
Received 23 September 2016
Received in revised form
8 February 2017
Accepted 11 February 2017
Available online 13 February 2017

Keywords:
Pamphobeteus verdolaga
Theraphosidae
Transcriptomic
Proteomic

ABSTRACT

Theraphosidae spider venoms are well known for possess a complex mixture of protein and non-protein compounds in their venom. The objective of this study was to report and identify different proteins translated from the venom gland DNA information of the recently described Theraphosidae spider *Pamphobeteus verdolaga*. Using a venom gland transcriptomic analysis, we reported a set of the first complete sequences of seven different proteins of the recenlty described Theraphosidae spider *P. verdolaga*. Protein analysis indicates the presence of different proteins on the venom composition of this new spider, some of them uncommon in the Theraphosidae family. MS/MS analysis of *P. verdolaga* showed different fragments matching sphingomyelinases (sicaritoxin), barytoxins, hexatoxins, latroinsectotoxins, and linear (zadotoxins) peptides. Only four of the MS/MS fragments showed 100% sequence similarity with one of the transcribed proteins. Transcriptomic analysis showed the presence of different groups of proteins like phospholipases, hyaluronidases, inhibitory cysteine knots (ICK) peptides among others. The three database of protein domains used in this study (Pfam, SMART and CDD) showed congruency in the search of unique conserved protein domain for only four of the translated proteins. Those proteins matched with EF-hand proteins, cysteine rich secretory proteins, jingzhaotoxins, theraphotoxins and hexatoxins, from different Mygalomorphae spiders belonging to the families Theraphosidae, Barychelidae and Hexathelidae. None of the analyzed sequences showed a complete 100% similarity.

1. Introduction

Theraphosidae tarantulas comprises the most speciose family of Mygalomorphae, mainly distributed in the tropical and subtropical regions (WorldSpiderCatalog., 2016). *Pamphobeteus* Pocock, 1901 is a wide spread genus of the Theraphosinae sub-family, with 13 species described worldwide and six reported in Colombia (Cifuentes et al., 2016; WorldSpiderCatalog., 2016).

Only the venom from 21 Theraphosidae genus (18% of the total reported genus) have been studied with the description of 448 different components, which only 98 (21%) of the reported sequences corresponds to Theraphosidae genus distributed in the American continent (*Theraphosa*, *Thrixopelma*, *Psalmopoeus*, *Lasiodora*, *Brachypelma*, *Grammostola* and *Aphonopelma*) (Herzig et al., 2011). Although Colombia possess 10% of the worldwide reported genera of the Theraphosidae family and the half of the *Pamphobeteus* species reported today (WorldSpiderCatalog., 2016), there is

* Corresponding author. Programa de Ofidismo/Escorpionismo – Serpentario, Universidad de Antioquia UdeA, Carrera 53 No 61-30, Medellín, 050010, Antioquia, Colombia.
E-mail addresses: sebastian.estrada@udea.edu.co (S. Estrada-Gómez), leidy.vargasmu@campusucc.edu.co (L.J. Vargas-Muñoz), monica.saldarriaga@uibero.cl, monica.saldarriaga.cordoba@gmail.com (M. Saldarriaga-Córdoba), yelicigi@gmail.com (Y. Cifuentes), caperafanl@gmail.com (C. Perafan).

http://dx.doi.org/10.1016/j.toxicon.2017.02.004
0041-0101/

not any major research focused in the description of the venom of this individuals. In Colombia, there is only two studies available describing the venom biochemical and molecular content of two Theraphosidae species, *Aguapanela arvi* Perafán, Cifuentes & Estrada, 2015 (Perafan et al., 2015) and *Pamphobeteus verdolaga* Cifuentes, Perafán & Estrada-Gomez, 2016 (Cifuentes et al., 2016). The latter initially determined as *Pamphobeteus aff. nigricolor* (Estrada-Gomez et al., 2013; Perafan et al., 2015). *P. verdolaga* is a recently described species distributed from the Aburra Valley (Medellín) at 1400 m above the sea level (m.a.s.l.), to the south west region of the Antioquia, in the municipality Jardin at 2100 m.a.s.l. (Cifuentes et al., 2016). The male of *P. verdolaga* is distinguished by having palpal bulb with broad embolus, poorly developed apical and prolateral inferior keels, prolateral accessory keels present and retrolateral keel length similar to apical keel. Females is distinguished by having a spermatheca with a wide base and very short seminal receptacles, oval and curved toward the center. This is the thirteenth specie described of Pamphobeteus and the sixth for Colombia (Cifuentes et al., 2016).

Previous research by Estrada-Gómez et al, (Estrada-Gomez et al., 2013) indicated that venom from *P. verdolaga* displays an indirect hemolytic activity with a minimum hemolytic dose (MHeD) calculated of 370.1 μg indicating the possible presence of phospholipases A_2. The electrophoretic profile showed compounds with masses close to 14 kDa, indicating the possible presence of molecules like phospholipases A_2, but these enzymes had been never described in this genus or family. Other compounds with a molecular mass between 40 kDa and 50 kDa, and low molecular mass compounds below 6 kDa were detected. Three different internal peptides fragments were identified, two of them showing homology with toxins exhibiting a modulatory effect over calcium channels (RCLPAGKPCAGVTQKI) and one other that did not match any known spider toxin (KDSAFKHPAPTFGDLSKL) (Estrada-Gomez et al., 2013). Post-translational modifications (PTM) play a key role in functional proteins like the mentioned above, giving higher complexity and peptide diversity of venoms. Despite the importance of this feature, there is not enough literature of these relevant modifications described for scorpion venoms.

The objective of this study was to report and identify different proteins translated from the venom gland DNA information and detect similar sequences to find homologous proteins within Mygalomorphae spiders sequences published in data bases. We reported the first complete sequences of seven different Theraphosidae proteins, with the respective DNA of this newly described spider *Pamphobeteus verdolaga* (Theraphosidae). These transcribed proteins were used to perform a Non-metric Multidimensional Scaling (NMDS) analysis to visualize in two dimension the homology grade among these proteins based on distance matrix.

2. Material and methods

2.1. Transcriptomic analysis

2.1.1. RNA and data collection

RNA-seq experiment was carried out using the Illumina HiSeq technology. Briefly, total RNA was purified from the spider glands and surrounding tissue using TRIZOL reagent. RNA quality was assessed using the bioanalyzer capillary system that showed a RIN over seven. mRNA purification and library preparation was carried out with the illumina mRNA TRUSEQ kits. Sequencing was performed with Hiseq 2500 instrument with paired-end reads of 100 bases. Reads were cleaned using PRINSEQ-LITE and the transcriptome was assembled using TRINITY package and protein orthologs was selected using TBLASTX and TBLASTN programs. Multiple sequence alignment was completed by the Clustal Omega program only of the mature proteins, signal peptides were removed. The signal peptide was predicted by the SignalP 4.01 server available at http://www.cbs.dtu.dk/services/SignalP/ (Petersen et al., 2011).

2.2. MS/MS analysis

2.2.1. Sample digestions

Sequence grade Lys-C/Trypsin (Promega) was used to enzymatically digest the RP-HPLC venom samples. The samples were reduced and alkylated. All digestions were carried out in the Barocycler NEP2320 (PBI) at 50 °C under 20 kpsi for 2 h. Digested samples were cleaned over C18 spin columns (Nest Group) and dried. Resulting pellets were resuspended in 15 μL of 97% purified H2O/3% ACN/0.1% formic acid (FA). Five μL of volume is used for nano LC-MS/MS analysis.

2.2.2. LC-MS/MS

RP-HPLC fractions were run on a nano Eksigent 425 HPLC system coupled to the Triple TOF 5600 plus (Sciex, Framingham, MA). The method used for analysis was 120 min at 300 nL/min over the cHiPLC nanoflex system. The trap column was a Nano cHiPLC 200 μm × 0.5 mm ChromXP C18-CL 3 μm 120 Å followed by the analytical column, the Nano cHiPLC 75 μm × 15 cm ChromXP C18-CL 5 μm 120 Å. The sample was injected into the Triple TOF 5600 plus through the Nanospray III source equipped with emission tip from New Objective. Peptides from the digestion were eluted from the columns using a mobile phase A of purified H_2O/0.1% formic acid (FA) and a mobile phase B of ACN/0.1% FA. With a flow rate of 0.3 μl/min, the method started at 95% A for 1 min followed by a gradient of 5% B to 35% B in 90 min and from 35% B to 80% B in 2 min. Eighty percent B was held for 5 min before being brought to 5% B and held for 20 min.

2.2.3. Data analysis

The data acquisition was performed monitoring 50 precursor ions at 250 ms/scan. Mascot Daemon v.2.4.0 (Matrix Science) was used for database searches against the different databases. Some were downloaded from UniProt, NCBI or the ArachnoServer website (Herzig et al., 2011). Samples were run in the Bindley Bioscience Center at Purdue University.

2.3. Bioninformatic analysis

All seven transcribed sequences were used to perform Bioinformatic analysis. The search of similar sequences proteins translated from *P. verdolaga* with other arachnids toxins (including Theraphosidae, Hexatelidae and Barychelidae) was done using two protein data base, BLAST (BLASTX: search protein databases using a translated nucleotide query and BLASTp http://blast.ncbi.nlm.nih.gov/Blast.cgi) and Arachnoserver (http://www.arachnoserver.org/blastForm.html) (Herzig et al., 2011). For each query sequence (contig), the sequence with the lowest E-value which presents the highest percentage of identity was selected. Alignment of amino acid sequences of mature chains were performed PRALINE (Heringa, 1999) using default parameters and the conservative sites of each alignment were estimated in MEGA version 7 (Kumar et al., 2016).

2.3.1. Protein domain searching

The search of conserved domain was performed in three database of protein domains, including Pfam (Bateman et al., 2002), SMART (Letunic et al., 2002), and CDD (Marchler-Bauer et al., 2002). In order to identify the domain location within the query sequence, a multiple sequence alignment was performed in PRALINE (Heringa, 1999) using default settings. The alignment was

performed among query sequence, homology sequence and domain sequence identified by protein domain database.

2.3.2. NMDS analysis based on the distance matrix

In order to visualize in two dimensions the degree of homology between the sequences obtained in this study with those published in databases, a NMDS analysis was performed in PAST (Hammer et al., 2001). NMDS was based on a similarity matrix computed from amino acid substitution matrices with JTT (Jones-Taylor-Thornton) substitution model obtained in MEGA version 7 (Hammer et al., 2001).

2.4. Secondary structure prediction and three dimensional structure modeling

The secondary analysis prediction and three dimensional structure modeling of the proteins was made using I-TASSER server (Roy et al., 2010; Yang et al., 2015; Zhang, 2008). Briefly as described by Zhang, in this on-line protein modeling tool, the target sequences are threaded through a representative PDB structure library (with a pair wise sequence identity cut-off of 70%) to search for the possible folds using four variants of Profile-Profile Threading Alignment (PPA) methods with different combinations of the hidden Markov model (HMM) and PSI-BLAST profiles and the Needleman-Wunsch and Smith-Waterman alignment algorithms. The continuous fragments are then excised from the threading aligned regions which are used to reassemble full-length models while the threading unaligned regions (mainly loops) are built by *ab initio* modeling (Zhang, 2008). For I-TASSER, the C-score, which lies in the (−5,2) range, was calculated for each model. Only proteins showing an inhibitory cysteine knot (ICK) were analyzed to predict disulfide bonds. Disulfide bounds were predicted using the Cysteines Disulfide Bonding State and Connectivity Predictor, a web-based tool for disulfide engineering in proteins (Ceroni et al., 2006). The 3D structure were visualized in the UCSF Chimera software, a free program for molecular graphics and analysis (Pettersen et al., 2004).

3. Results

3.1. Transcriptomic analysis

Transcriptomic analysis showed the presence of different groups of proteins like phospholipases, hyaluronidases, inhibitory cysteine knots (ICK) peptides among others. The analysis showed a group of seven different expressed sequence tags (ESTs) encoding different compounds showing sequence similarity with toxins like: EF-hand proteins, jingzhaotoxins, theraphotoxins, cysteine rich secretory proteins (CRISP) and hexatoxins, described in the families like Theraphosidae, Hexatelidae and Barychelidae, all Mygalomorphae. Best hits of the traduced proteins, with an e-value above $1e^{-30}$, correspond to EF-hand protein-like proteins, theraphotoxins and CRISP from different Theraphosidae individuals. Two of this different EST (Contigs c18710 and c9919) showed sequence similarity with two different CRISP proteins identified in the Theraphosidae species from Chile *Grammostola rosea*. Fig. 1 shows all amino acid translated sequences of each protein. Supplementary material shows the respective DNA sequence encoding each protein. Table 1 shows each translated protein displaying similarity with the respective proteins. All translated sequences showed E-values above $1e^{-10}$. Signal peptide was found for only five sequences (c4691, c4418, c18219, c14008 and c9919). The three database of protein domains used in this study (Pfam, SMART and CDD) showed congruency in the search of unique conserved protein domain for c51827, c18710, c18219 and c9919 contig and no conserved domains for c4691, c4418 and 14008 coting were identified. In addition, Contig c51827 shows a conserved domain with EF hand calcium binding, inducing conformational changes and the activation or inactivation of target proteins. Contigs c18710 and c9919 domain is found on CRISP proteins which contain pfam00188 and has been termed the CRISP domain recognized to regulate ryanodine receptor Ca2+ signalling. It contains 10 conserved cysteines which are all involved in disulphide bonds. Contig c18219 domain corresponds to a peptidase family M13 of metallopeptidases including neprilysin. These proteins consist of a short N-terminal cytoplasmic domain, a single transmembrane helix, and a larger C-terminal extracellular domain containing the active site. Supplementary material shows the multiple alignment among each query sequences (contig) with its homologous sequence and protein domain sequence identify by protein domains database.

3.2. MS/MS analysis

MS/MS analysis of *P. verdolaga* showed different fragments matching sphingomyelinase (sicaritoxin), barytoxins, hexatoxins, latroinsectotoxins, and linear (zadotoxins) peptides. Only four different fragments matched one of the transcribed toxins (Theraphotoxin-Pv_Cysteine rich-like) that showed sequence similarity with GTx VA1-CRISP2 (Uniprot- M5AWW7), a cysteine rich protein found in the venom from the Chilean tarantula *Grammostola rosea*. All fragments showed 100% amino acid sequence similarity with the GTxVA1-CRISP2 toxin. Table 2 summarize all the fragments with the respective score. Fig. 2 shows the pairwise sequence alignment (Clustal omega) of all Theraphotoxin-Pv_Cysteine rich-like fragments vs GTxVA1-CRISP2 (Uniprot- M5AWW7) toxin from *Grammostola rosea*. We found 6 fragments showing a PTM and the most common were (K) acetylations and (M) oxidations.

3.3. Bioinformatic analysis

Analysis indicated that all translated sequences showed common evolutionary ancestry (homology) with other Mygalomorphae individuals. Specifically, *P. verdolaga* analyzed sequences showed a relationship with toxins from Theraphosidae, Barychelidae and Hexathelidae showing the closer relationship with Theraphosidae toxins than the other spider family toxins. Despite this, none of the analyzed sequences showed a total identity or similarity; only c51827 (Theraphotoxin-Pv_EF-hand protein-like) showed the closest distance with the protein of two Theraposidae spider, *Haplopelma schmiditi* with a ID of 91% (genetic distance with JTT model of 15.2%) and *Pelinobius muticus* with a ID of 90% (genetic distance with JTT model of 15.2%) (Fig. 3). All other translated sequences showed IDs below 70%. Contig c9919 showed homology with CRISP toxin. The greater genetic distance (ID 70% - genetic distance with JTT model of 40.5%) is observed with the toxin GTx-VA1 CRISP2 from *Grammostola rosea* (Theraphosidae). Contig c18219 showed an identity of 65% (genetic distance with JTT model of 43.1%) with the toxin Neprilysin-1 from *Trittame loki* (Barychelidae). Contig c4418 (Theraphotoxin-Pv_2) showed a close similitude (ID 58% - genetic distance with JTT model of 63%) with proteins from *Grammostola rosea* (Theraphosidae). Contig number c18710 showed a relationship with a species from the Theraphosidae family and is one of the CRISP toxins. The greater genetic distance (ID 58% - genetic distance with JTT model of 64.5%) is observed with the toxin GTx-CRISP1 from *Grammostola rosea* (Theraphosidae) (Fig. 3). Contig 14008 showed a close relationship (ID 48% - genetic distance with JTT model of 84.6%) with U1-hexatoxin-Iw1b and U1-hexatoxin-Iw1d from *Illawarra wisharti*, a Hexatelidae spider. Contig c4691 (Theraphotoxin-Pv_1), showed a low similitude with a theraphotoxin from *Chilobrachys guangxiensis*

>c51827_*P.verdolaga*
MSEALEKLRQQCRERGANGIIGLGRAFRLMDDNRDKKLNLEEMKFGLQEYGADMAPGEVESLFKELDRDKSGSVNFDEFLRAVRPPLSQERL
DMIAKAFAKMDKSGDGVVNLQDLKGVYNATEHPDVIGGKKSEDQVLVEFLQNFETPDNPDGKVTKQEFIDYYTGVSSSIDLDEYFXDYDAKG
LAAVKTECRCSFMYCKAKPFFKKLPLALCSVALHVVSCNS

>c4691_*P.verdolaga*
MRMLVIFLAVFAVLVKNGDSTCIVDFQNAMQVCTNKMQDATKEASELSSTDAQNQKMCCAFKEFESCAVDAGQKCGRTTEDLMKQAVEARKS
MLKAPCSECSGGNGVVPANAVTVYLLLFLGYCLYANKF

>c4418_*P. verdolaga*
MKLILSITAVCLVIAAVALPLRNLRDAFDPSEVLGQPMEKIQTENARACSKQVGQTCVRNCECCGATNICGYITVGSKVVQQCMNKTSNNLI
LNTLGHGWNAVNKFFSLCWG

>c18710_*P.verdolaga*
MLEMSWHKEAQESAQRWAEECQILTYDGILGKYVEDYGSCGQNIFVSNEKVPWTFVGEAWFAERYDFSYGSYNNSVSEVGHYTQMVWNGSHR
LGCGFHYCAKDVKKPFYNYVCNYCPMGNDPKRLGMPYSSGKPCSECIKHCKYKKLCTNTCPYADLWVNCAQLNVTWNDWLCSNPEHQR
HRGCRATCLCPGKVI

>c18219_*P.verdolaga*
MRAIHGFSITLFLVIAKITVPSAAAPKNRESTEVCTTPVCQKTAQYLLESMNMSVDPCQDFYQYACGGWISNHPIPDEKYRYATFDVLFDEL
RDTIAGILRNASSEDHTRAVLDSAMFYEACADKEARESAGLESLKSFLNELGGWPMANPSWNGDGYEWQTQVATVDRNLDITAIISVSVDAD
ANHTSQHIIYLDQPEAFGIGRNELKDRDASQRNREIVTAYKNFMVATGKLLNPNADETQLRNEADEIIEFESKLAQFSKKPEHRRDASSWYH
KMTLAEVQRVTENTAFQWTDFLNVVLADFPKVTSETEVILSEKDYIKNTLKLIQETYSTNPRVLANYIGWIVLKKIGHHTTKAFRENKFAFD
RVEMGIEKEEELLRTCADDAVSTMGHAVGRSYVDKYFPQEDKQDVDELVGSIWSAYKSTLNSNTWMDEATKTKALDKVDAMISKMGYPEWIK
NDTRLNEYYKSVANTIRRDDFFSSYVRVRKAVNAIELKKWNQPTDRVEDWDTTAAVVNAFYNPSSNSISFPAGILSIPFYQYGRTSALNYGG
IGAVIGHEVSHGFDDQGSQFDREGNLINWWSEETKERFDEKANCFRAQYASIVEPTTNMSLNGNNTVGENIADNGAVRNAYIAYAADVLQSR
GIVNKRLPGVPATPKQLFFIGFSSLWCGSERKESLEWSIQYDPHTLNQFRATVPLMNSDEFAEAFNCPRGSPMNPQDKCFLW

>c14008_*P.verdolaga*
XDLGATNNKVLEIYIFRRMALIAFIFLSFLASTFAQQQCGRRTCGKGECCRISLIFFQKSCRRLSREWQFCRRENEGNSNKSYFLFCPCEEG
LDCRHFLCLRGRRTTTTTGPTTEPTTSTSEGTSTIIEQSSTEETTSTTTVSSPATTTTAVPPTTETETEPAA

>c9919_g1_*P.verdolaga*
MRIRVILMLSWLWLGVSDGSCPALYRRYSKAHTYCLPPKSSSTILKSGISKSDIETIVRVHNELRSKVATGEETTYSMPKASNMRQMVWDSE
LAAVAQKHANQCLFKHDCNNCRKVKNFDVGQNLFKRNPFSVPPQPTWAQAVTDWYSEVNVFDQGQIDGFIDGEGPPQTGHFTQDIWAESWRV
GCGYSVCEEGNVLLELYTCNYGPAGNGENDPIYERGDPCTNCPLNSCCGSSCSGGTSYPGLCRISGENAPQYNRPEGLTFYCSFNNEHDCAK
TVTGHNKWQVSKTLSGSYIGIVLNGGESSTLSFKNPAKVPQKPLCFIINYRTGPQVDGEEVSGTANVIFKAGGSTFPSELNSNGFQSFTKFS
ITLGWNMPTMIDISISVPEGGSSRYLEIKDMSATKESARTMALSSLLSGHYARTMELSSLLSGHYARTMELSSLLSGHYARTMX

Fig. 1. Amino acid translated sequences of *P. verdolaga*. Translated sequences from *P. verdolaga* for each contig. Gray highlighted residues indicate signal peptide of the protein.

Table 1
Sequence similarity of each contig from *P. verdolaga* with the best hit of each different toxin. All accession numbers correspond to UniProt database.

Contig number	Givren name	Homology with	Accession	Organism	Score	E-value
C51827	Theraphotoxin-Pv_EF-hand protein-like	EF-hand protein-like protein	D5J6Y7	*Pelinobius muticus*	327.0	$2e^{-110}$
			B5M6G5	*Haplopelma schmiditi*	354	$1e^{-120}$
c4691	Theraphotoxin-Pv_1	U35-theraphotoxin-Cg1a	B1P1J4	*Chilobrachys guangxiensis*	94.0	$2e^{-21}$
c4418	Theraphotoxin-Pv_2	U7-theraphotoxin-Gr1b	I2FKH7	*Grammostola rosea*	128.0	$4e^{-35}$
c18710	Theraphotoxin-Pv_Cysteine rich-like	GTx-CRISP1	M5AYF1	*Grammostola rosea*	262.0	$3e^{-84}$
c18219	Theraphotoxin-Pv_4	Neprilysin-1	W4VS99	*Trittame loki*	1001.5	$0.00e^{+00}$
c14008	Theraphotoxin-Pv_3	U1-hexatoxin-Iw1d	Q5D229	*Illawarra wisharti*	68.9	$5e^{-12}$
C9919	Theraphotoxin-Pv_Cysteine rich-like1	GTx-VA1 CRISP2	M5AWW7	*Grammostola rosea*	566	$0.00e^{+00}$

(Mygalomorphae), with an ID of 39% (genetic distance with JTT model of 106%) (Fig. 3).

3.4. Predicted structure

Fig. 4 shows the most probable structure and three-dimensional models of mature toxins of those that was found to contain a domain and shows a C-score of high confidence, between −1.5 and 2 (Yang et al., 2015; Zhang, 2008) (c18219, c18710). Contig c18219 (Fig. 4A) is a peptidase-like protein with a theoretical molecular weight of 79761.98 Da. Mature toxin structure is formed by a complex systems of α-helix with at least 3 parallel or antiparallel β-sheets. The C-score of the predicted three-dimensional structure of c18219 is 0.55. Contig c18710 (Fig. 4B) is a CRISP protein with a theoretical molecular weight of 22532.53 Da. Complete toxin shows a three-dimensional predicted structure characterized by presence of 4 α-helix region, 3 antiparallel β-sheets. The C-score of the predicted three-dimensional structure of c18710 is 0.97. Contig c4418 was included due to the similar structure with the ICK motif showed by the mature toxin despite its low level of confidence (C-score = −3.26). These motifs are essential to perform most of the neurotoxic effects of spider venoms (King, 2007; Nicholson, 2006, 2007). Contig c4418 (Fig. 4A) mature toxin, is a 94 amino acid compound with a theoretical molecular weight of 10338.8 Da.

Table 2
Pamphobeteus verdolaga summary of the identified peptides. All MS/MS derived sequence corresponds to internal peptides. All database numbers are from UniProtKB. Z: charge. Asterisk (*) indicates fragments with (K) acetylations and pilcrow (¶) indicates fragments with (M) oxidations as PTM.

Fragment given name	MS/MS derived sequence	Score	Matching peptide acc. number	Matched peptide family	Expected peptide m/z	Z	Calculated peptide mass	Matched organism
Theraphotoxin-Pv_Cysteine rich-like1	¶ K-VATGKETQYSMPK-A	50.3	M5AWW7	GTx-VA1 (CRISP2)	728.339	+2	1454.708	*Grammostola rosea*
Theraphotoxin-Pv_Cysteine rich-like2	* K-DWYKEIK-D	35.4			533.253	+2	1064.518	
Theraphotoxin-Pv_Cysteine rich-like3	* R-TGPQVKGEK-S	32.9			493.247	+2	984.524	
Theraphotoxin-Pv_Cysteine rich-like4	¶ K-SFPTVLTSSSMSFTKFTK-K	13			503.758	+4	2010.997	
Theraphotoxin-Pv_PLD-like	R-GVPQIEVTFDLDANGILQVSAQDKSTGK-S	20	C0JB91	Sphingomyelinase D-like	733.371	+4	929.503	*Sicarius albospinosus*
Theraphotoxin-Pv_PLD-like2	K-VDFSLSSDRDFILLNVPIPQAQAR-I	19		toxin	901.151	+3	2700.423	
Theraphotoxin-Pv_Barytoxin-like	K- VINECGKPK-I	18	W4VSB2	U5-barytoxin-Tl1b	516.305	+2	1030.548	*Trittame loki*
Theraphotoxin-Pv_Barytoxin-like2	* R- TTGIVLDSGDGVSHTVPIYEGYALPHAIL WLDLAGRDLTDYLMK - I	18	W4VRV1	U6-barytoxin-Tl1a	1211.870	+4	4843.441	
Theraphotoxin-Pv_Zadotoxin-like	* ¶ R- YDAMKMQTSLGQMKK-R	67	Q1ELU8	M-zodatoxin-Lt6a/b	611.969	+2	1832.847	*Lachesana tarabaevi*
Theraphotoxin-Pv_ Zadotoxin -like2	V-VNNLRKV-Q	21	Q1ELU5	M-zodatoxin-Lt4a	421.758	+2	841.513	
Theraphotoxin-Pv_ Zadotoxin -like3	E-LEEAERGGI-D	26			487.268	+2	972.487	
Theraphotoxin-Pv_ Zadotoxin -like4	R-KLMEMVNN-L	23			326.845	+3	977.467	
Theraphotoxin-Pv_ Zadotoxin -like5	E-EEGYDVSEEI-Q	26			585.223	+2	1168.477	
Theraphotoxin-Pv_ Zadotoxin -like5	N-LRKVQGREDTEAR-G	36			422.481	+4	1685.880	
Theraphotoxin-Pv_ Latroinsectotoxins -like	K-NYSPYYNTIDDLKDQVDLTVGNNK-T	21	Q02989	Alpha-Latroinsectotoxin-Lt1a	968.139	+3	2901.403	*Latrodectus tredecimguttatus*
Theraphotoxin-Pv_ ICK -like	K-TLNLAQNLLAYLPEELFHPLGR-L	19	D2Y283	U11-theraphotoxin-Hhn1m	841.468	+3	2521.369	*Haplopelma hainanum*

Shows a three-dimensional predicted structure characterized by presence of 4 α-helix region, 2 antiparallel β-sheets (Fig. 4A) and 8 different cysteine residues are present at the positions Cys31, Cys39, Cys43, Cys45, Cys46, Cys52, Cys65, Cys92 (Fig. 4A). The predicted cysteine disulfide bonding indicates that four different disulfide bridges are present between Cys31-Cys52/Cys39-Cys65/Cys43-Cys45/Cys46-Cys92 (CysI-CysVI/CysII-CysVII/CysIII-CysIV/CysV-CysVIII).

4. Discussion

Pamphobeteus verdolaga is a recently described spider from Therpahosidae family with a widely spread distribution in the high mountain across Antioquia-Colombia (Cifuentes et al., 2016). The venom of this spider have been poorly analyzed with only one report available where the authors described the presence of different internal peptides fragments, two of them matching ICK compounds (Estrada-Gomez et al., 2013). No report of any complete protein have been carried out. We report the firs 7 complete sequences including signal peptides, of this new *Pamphobeteus* species.

From the Thraphosidae family, only genus *Chilobrachys*, *Haplopelma* and *Grammostola* have transcriptomic information available. The analysis of the seven different translated proteins from *P. verdolaga* showed a sequence similarity with toxins like EF-hand proteins, cysteine rich secretory proteins, jingzhaotoxins, theraphotoxins and hexatoxins, all from different Mygalomorphae spiders belonging to the families Theraphosidae, Barychelidae and Hexathelidae. Although some of the proteins showed a high degree of similarity, none of the compounds are 100% identical indicating a specific venom content of *P. verdolaga*. Additionally, 3 of the 7 translated proteins did not match any known conserved protein domain. Only c51827 (Theraphotoxin-Pv_EF-hand protein-like) showed the closest genetic distance with the protein of two Theraposidae spider, *Haplopelma schmiditi* with a ID of 91% (genetic distance with JTT model of 15.2%) and *Pelinobius muticus* with a ID of 90% (genetic distance with JTT model of 15.2%). All this indicates that *P. verdolaga* venom show some protein homologous with excess of similarity and structure with other Theraphosidae spider and other proteins more distant due to the database fail to annotate a domain indicating that *P. verdolaga* may synthetize a venom with a specific content to this newly described species from Colombia. These differences with other tarantula proteins clearly sets it apart from each other. The same situation was reported with the reported peptide (K) DSAFKHPAPTFGDLSK (L) that did not match any other toxin reported in the Mygalomorphae order (Estrada-Gomez et al., 2013). Although the molecular target of all of these toxins still unknown, a previous report indicated that in this species two different sequences displays high sequence similarity with the toxin U2-theraphotoxin-Asp1a (from *Aphonopelma* spp) and u-theraphotoxin-Hh1a (from *Haplopelma schmidti*) both affecting voltage-gated calcium channels, indicating that this venom may enhance the same activity (Estrada-Gomez et al., 2013).

MS/MS analysis corroborated the presence of GTx-VA1 CRISP2 protein of the translated sequences found in the transcriptome, indicating that this protein is present not only in DNA information but in the venom as well as a secreted protein. As well, we detected different fragments matching sphingomyelinase (sicaritoxin) from *Sicarius albospinosus* (Siicaridae), barytoxins from *Trittame loki* (Barychelidae), hexatoxins from *Haplopelma hainanum* (Theraphosidae), latroinsectotoxins from *Latrodectus tredecimguttatus* (Theridiidae), and linear (zadotoxins) peptides from *Lachesana tarabaevi* (Zodariidae), all uncommon toxins reported in the Theraphosidae family excepting the hexatoxins.

The model of the 3D predicted secondary structure of the contig c4418 shows a molecule with a predominance of α-helical regions

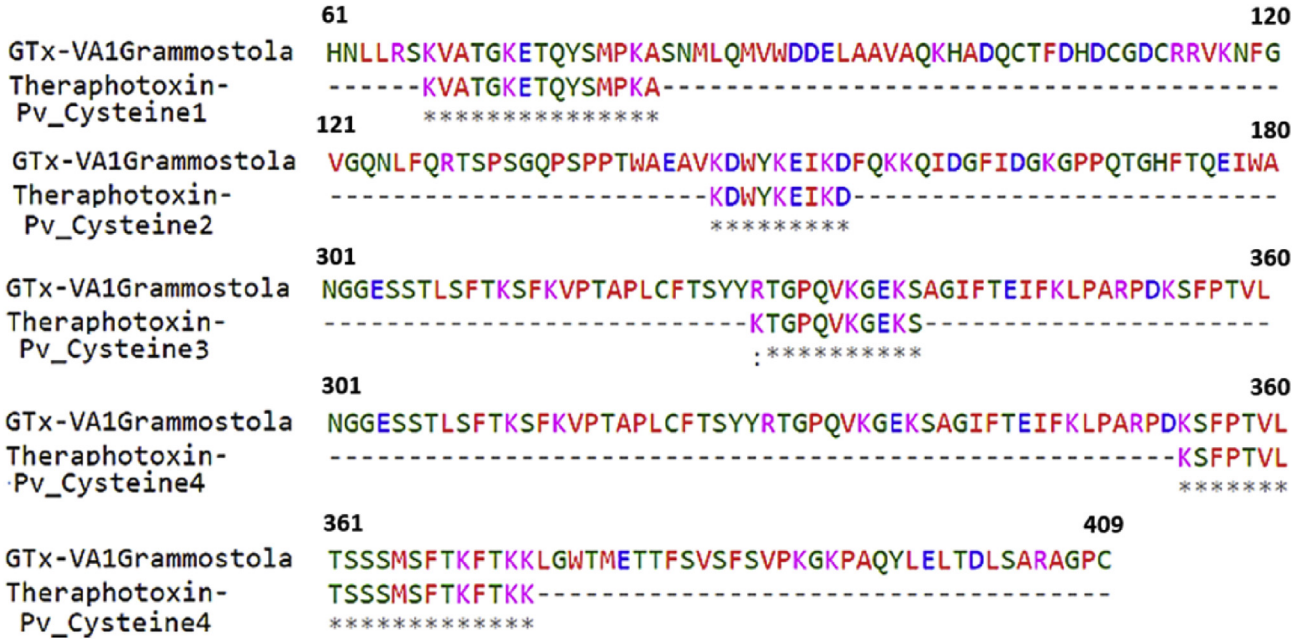

Fig. 2. Pairwise sequence alignment (Clustal omega) of the different Theraphotoxin-Pv_Cysteine rich-like fragments found in the venom of *Pamphobeteus verdolaga* with the toxin GTxVA1-CRISP2 (Uniprot- M5AWW7). Asterisks indicate identical amino acids from the compared sequence.

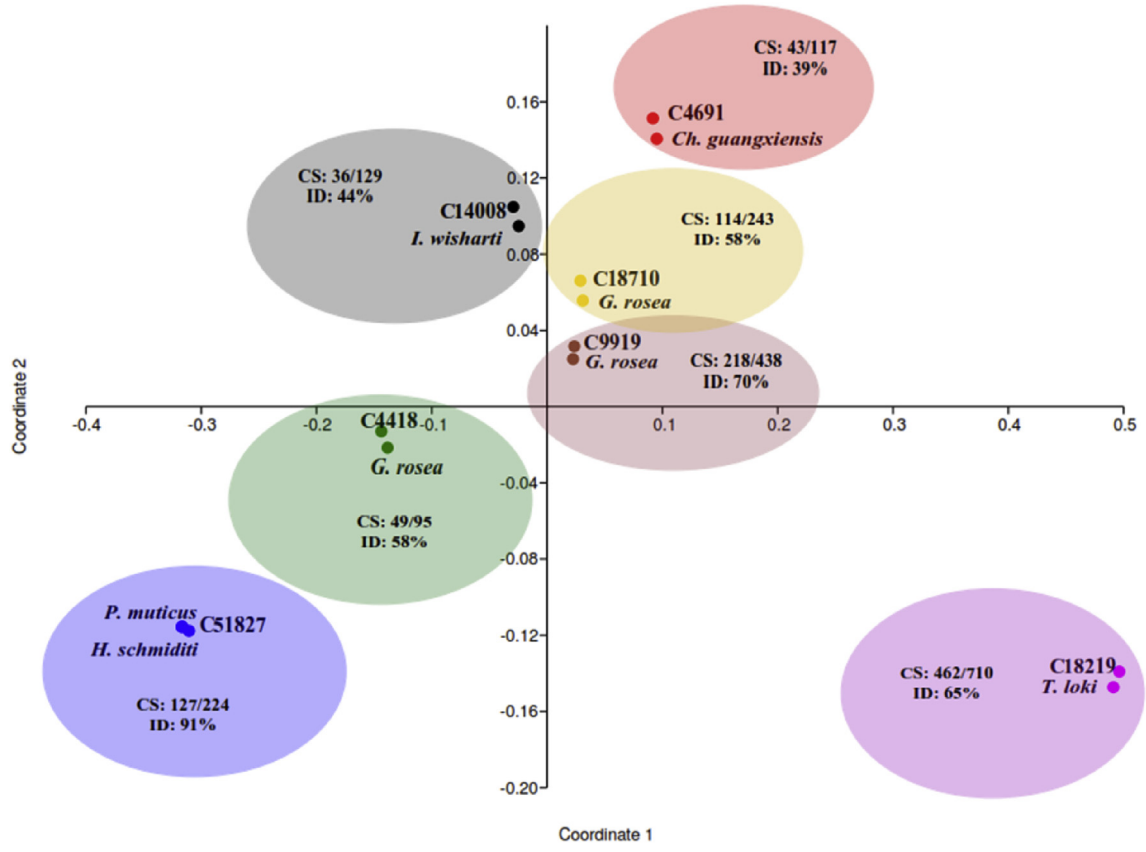

Fig. 3. No metric multidimensional scaling plots (NMDS). No metric multidimensional scaling plots (NMDS) with all query sequences with its respective homologue sequence obtain in the protein database.

and only two antiparallel β-sheets. The cysteines disulfide bonding prediction indicates that c4418 have four different disulfide bridges matching the characteristics of an ICK motif (derives of the primary sequence). The predictive software showed a pattern of CysI-CysVI/CysII-CysVII/CysIII-CysIV/CysV-CysVIII as seen in the ICK proteins in the toxins from the Theraphosidae venoms, were is very common

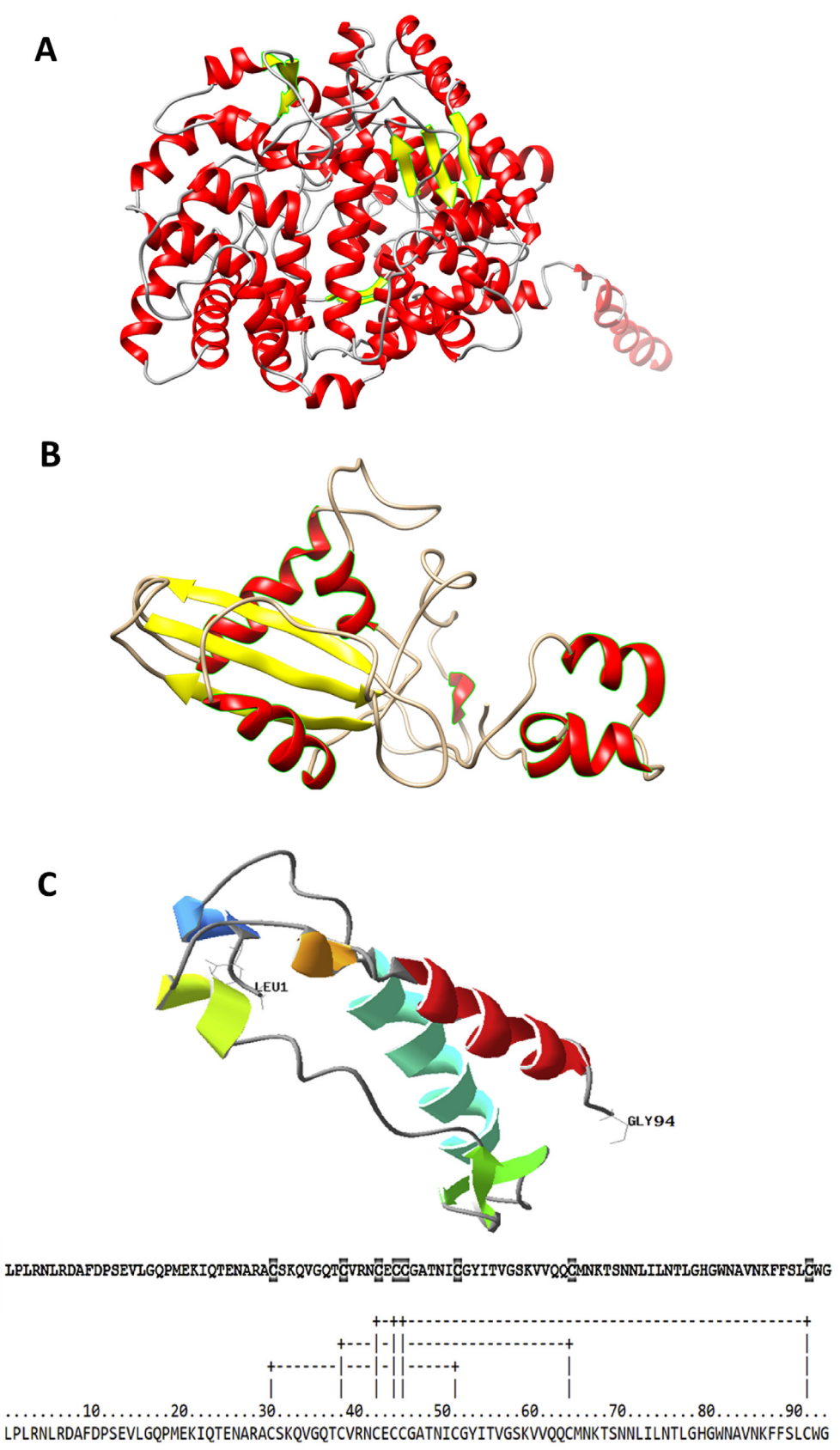

Fig. 4. I-TASSER structural models for the two contig that have the highest C-score (c18219 and c18710) and c4418 contig with its respective disulfide bridges. A. c18219 model of the 3-D predicted secondary **B.** c18710 model of the 3-D predicted secondary. **C.** c4418 model of the 3-D predicted secondary structure with the respective primary mature sequence and the predicted disulfide bridges.

to find molecules with a cysteine knot motif (Escoubas et al., 2000; King, 2007; Nicholson, 2006, 2007).

5. Conclusions

Pamphobeteus verdolaga is a recently described Theraphosidae spider distributed in Antioquia Colombia, synthetizing a venom with a specific content when compared with other Theraphosidae spiders. Besides the two previous reported fragments corresponding to low molecular compounds affecting ion channels, we report for first time 7 different complete transcribed proteins, some of them uncommon in the Theraphosidae family. MS/MS analysis showed the presence of different fragments matching, as well, uncommon proteins like sphingomyelinase (sicaritoxin), barytoxins, hexatoxins, latroinsectotoxins, and linear (zadotoxins) peptides.

Author contributions

Sebastian Estrada-Gómez contributed with the conceptualization, writing – review & editing, investigation, formal analysis, methodology and funding acquisition; Leidy Johana Vargas Muñoz contributed with writing – review & editing, investigation and funding acquisition; Monica Saldarriaga-Córdoba writing – review & editing, investigation, formal analysis and methodology; Yeimy Cifuentes and Carlos Perafan methodology and corroborated the taxonomic ID of the spider. All author contributed to the manuscript.

Conflicts of interest

The authors declare no conflict of interest.

Acknowledgments.

The authors are grateful with the Sostenibilidad program of the Universidad de Antiikoquia (UdeA), and Comité para el Desarrollo de la Investigación CONADI, Universidad Cooperativa de Colombia. This research was financed with the Project CIQF-211 supported by the Comité para el Desarrollo de la Investigación (CODI-UdeA).

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.toxicon.2017.02.004.

References

Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M., Sonnhammer, E.L., 2002. The Pfam protein families database. Nucleic Acids Res. 30, 276–280.
Ceroni, A., Passerini, A., Vullo, A., Frasconi, P., 2006. DISULFIND: a disulfide bonding state and cysteine connectivity prediction server. Nucleic Acids Res. 34, W177–W181.
Cifuentes, Y., Estrada-Gomez, S., Vargas Munoz, L.J., Perafan, C., 2016. Pamphobeteus Verdolaga a New Tarantula Species from Colombia, with Contributions to its Molecular Characterization (Araneae, Mygalomorphae, Theraphosidae) Zoologia e20160113.
Escoubas, P., Diochot, S., Corzo, G., 2000. Structure and pharmacology of spider venom neurotoxins. Biochimie 82, 893–907.
Estrada-Gomez, S., Vargas Munoz, L.J., Quintana Castillo, J.C., 2013. Extraction and partial characterization of venom from the Colombian spider Pamphobeteus aff. nigricolor (Aranae:Theraphosidae). Toxicon 76C, 301–309.
Hammer, O., Harper, D.A.T., Ryan, P.D., 2001. PAST: Paleontological Statistic Software Package for Education and Data analysis., 2.1 Ed.
Heringa, J., 1999. Two strategies for sequence comparison: profile-preprocessed and secondary structure-induced multiple alignment. Comput. Chem. 23, 341–364.
Herzig, V., Wood, D.L.A., Newell, F., Chaumeil, P.-A., Kaas, Q., Binford, G.J., Nicholson, G.M., Gorse, D., King, G.F., 2011. ArachnoServer 2.0, an updated online resource for spider toxin sequences and structures. Nucleic Acids Res. 39, D653–D657.
King, G.F., 2007. Modulation of insect Ca(v) channels by peptidic spider toxins. Toxicon 49, 513–530.
Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874.
Letunic, I., Goodstadt, L., Dickens, N.J., Doerks, T., Schultz, J., Mott, R., Ciccarelli, F., Copley, R.R., Ponting, C.P., Bork, P., 2002. Recent improvements to the SMART domain-based sequence annotation resource. Nucleic Acids Res. 30, 242–244.
Marchler-Bauer, A., Panchenko, A.R., Shoemaker, B.A., Thiessen, P.A., Geer, L.Y., Bryant, S.H., 2002. CDD: a database of conserved domain alignments with links to domain three-dimensional structure. Nucleic Acids Res. 30, 281–283.
Nicholson, G.M., 2006. Spider venom peptides. In: Hastin, A. (Ed.), Handbook of Biologically Active Peptides. ELSEVIER, Louisiana, p. 1640.
Nicholson, G.M., 2007. Insect-selective spider toxins targeting voltage-gated sodium channels. Toxicon 49, 490–512.
Perafan, C., Cifuentes, Y., Estrada-Gomez, S., 2015. Aguapanela, a new tarantula genus from the colombian andes (araneae, Theraphosidae). Zootaxa 4033, 529–542.
Petersen, T.N., Brunak, S., von Heijne, G., Nielsen, H., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Meth 8, 785–786.
Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera–a visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–1612.
Roy, A., Kucukural, A., Zhang, Y., 2010. I-TASSER: a unified platform for automated protein structure and function prediction. Nat. Protoc. 5, 725–738.
WorldSpiderCatalog, 2016. In: Bern, N.H.M. (Ed.), World Spider Catalog. http://wsc.nmbe.ch.
Yang, J., Yan, R., Roy, A., Xu, D., Poisson, J., Zhang, Y., 2015. The I-TASSER Suite: protein structure and function prediction. Nat. Meth 12, 7–8.
Zhang, Y., 2008. I-TASSER server for protein 3D structure prediction. BMC Bioinforma. 9, 40.