

Article

Unveiling the Protein Components of the Secretory-Venom Gland and Venom of the Scorpion *Centruroides possanii* (Buthidae) through Omic Technologies

Patricia Elizabeth García-Villalvazo ^{1,†}, Juana María Jiménez-Vargas ^{1,2,†}, Gisela Jareth Lino-López ³, Erika Patricia Meneses ⁴, Manuel de Jesús Bermúdez-Guzmán ⁵, Carlos Eduardo Barajas-Saucedo ¹ , Iván Delgado Enciso ⁶ , Lourival Domingos Possani ^{4,*}  and Laura Leticia Valdez-Velazquez ^{1,*} 

- ¹ Facultad de Ciencias Químicas, Universidad de Colima, Colima 28400, Mexico; pgarcia29@uclm.mx (P.E.G.-V.); jjimenez45@uclm.mx (J.M.J.-V.); carlos_barajas24@uclm.mx (C.E.B.-S.)
- ² Consejo Nacional de Humanidades, Ciencia y Tecnología (CONAHCYT), Mexico City 03940, Mexico
- ³ Centro Nacional de Referencia de Control Biológico, Dirección General de Sanidad Vegetal SENASICAADER, Colima 28110, Mexico; giselalino01@gmail.com
- ⁴ Departamento de Medicina Molecular y Bioprocesos, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca 62210, Mexico; erika.meneses@ibt.unam.mx
- ⁵ Agrícolas y Pecuarias (INIFAP), Instituto Nacional de Investigaciones Forestales, Colima 28110, Mexico; bermudez.manuel@inifap.gob.mx
- ⁶ Facultad de Medicina, Universidad de Colima, Colima 28040, Mexico; ivan_delgado_enciso@uclm.mx
- * Correspondence: lourival.possani@ibt.unam.mx (L.D.P.); lauravaldez@uclm.mx (L.L.V.-V.)
- † These authors contributed equally to this work.



Citation: García-Villalvazo, P.E.; Jiménez-Vargas, J.M.; Lino-López, G.J.; Meneses, E.P.; Bermúdez-Guzmán, M.d.J.; Barajas-Saucedo, C.E.; Delgado Enciso, I.; Possani, L.D.; Valdez-Velazquez, L.L. Unveiling the Protein Components of the Secretory-Venom Gland and Venom of the Scorpion *Centruroides possanii* (Buthidae) through Omic Technologies. *Toxins* **2023**, *15*, 498. <https://doi.org/10.3390/toxins15080498>

Received: 29 June 2023

Revised: 28 July 2023

Accepted: 7 August 2023

Published: 9 August 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: *Centruroides possanii* is a recently discovered species of “striped scorpion” found in Mexico. Certain species of *Centruroides* are known to be toxic to mammals, leading to numerous cases of human intoxications in the country. Venom components are thought to possess therapeutic potential and/or biotechnological applications. Hence, obtaining and analyzing the secretory gland transcriptome and venom proteome of *C. possanii* is relevant, and that is what is described in this communication. Since this is a newly described species, first, its LD₅₀ to mice was determined and estimated to be 659 ng/g mouse weight. Using RNA extracted from this species and preparing their corresponding cDNA fragments, a transcriptome analysis was obtained on a Genome Analyzer (Illumina) using the 76-base pair-end sequencing protocol. Via high-throughput sequencing, 19,158,736 reads were obtained and assembled in 835,204 sequences. Of them, 28,399 transcripts were annotated with Pfam. A total of 244 complete transcripts were identified in the transcriptome of *C. possanii*. Of these, 109 sequences showed identity to toxins that act on ion channels, 47 enzymes, 17 protease inhibitors (PINs), 11 defense peptides (HDPs), and 60 in other components. In addition, a sample of the soluble venom obtained from this scorpion was analyzed using an Orbitrap Velos apparatus, which allowed for identification by liquid chromatography followed by mass spectrometry (LC-MS/MS) of 70 peptides and proteins: 23 toxins, 27 enzymes, 6 PINs, 3 HDPs, and 11 other components. Until now, this work has the highest number of scorpion venom components identified through omics technologies. The main novel findings described here were analyzed in comparison with the known data from the literature, and this process permitted some new insights in this field.

Keywords: *Centruroides possanii*; next generation sequencing; omics technologies; proteome; scorpion; transcriptome; venom; venom gland

Key Contribution: This work reports the largest number of coding transcripts identified for any scorpion species, and phospholipase A1 and phosphodiesterase enzymes were found in the venom in addition to an NDBP3-type HDP. These coding transcripts, peptides, and proteins represent potential medical and biotechnological applications.

1. Introduction

Scorpions are evolutionarily successful arachnids, having remained on the planet throughout environmental changes, mass extinctions, and recent coexistence with humans since their appearance approximately 450 million years ago. Since then, they have adapted to physiological changes and numerous ecological, biochemical, and ethological situations [1,2].

Throughout their evolution, they have developed the ability to produce venom, allowing them to hunt down their prey and defend themselves from predators. Human health could be affected by their venom. Of the 2762 described species [3], only 50 species are suspected of causing severe envenoming and deaths to humans [2].

Ancient civilizations have used scorpion venom to treat some illnesses and diseases, even for the stings themselves [4,5]. In recent decades, novel technological developments have allowed for the analysis of venom components, as well as the determination of their chemical structures and functions. In particular, the protein components of scorpion venom have been classified into five families according to their structural domain: toxins, enzymes, host defense peptides (HDPs), protease inhibitors (PINs), and other components. Of these, neurotoxins are the most relevant because they act on sodium channels (NaTx), generating severe symptoms of envenoming and even death to mammals [6]. These compounds have been identified through studies of the venom-producing gland by transcriptomic analysis and proteomics studies performed on the soluble venom.

The scorpion *Centruroides possanii* (now and then abbreviated *C. possanii*) was recently described by González-Santillán et al. [7]. It is endemic to the ecologic reserve El Terrero, Minatitlán, Colima, Mexico, and ranges in color from straw yellow to ochraceous orange. It is the smallest described *Centruroides* species [7] (Figure 1). The venomous components of this scorpion have not previously been investigated, and we intend to present such findings. The number of cases of scorpionism causing morbidity in Colima State was 9015 in 2022 [8], and these were caused by only five species, including *C. possanii*. The symptoms can be mild, moderate, or severe, and start with local discomfort in the sting area which progresses to dyspnea, which may cause the victim's death [9].

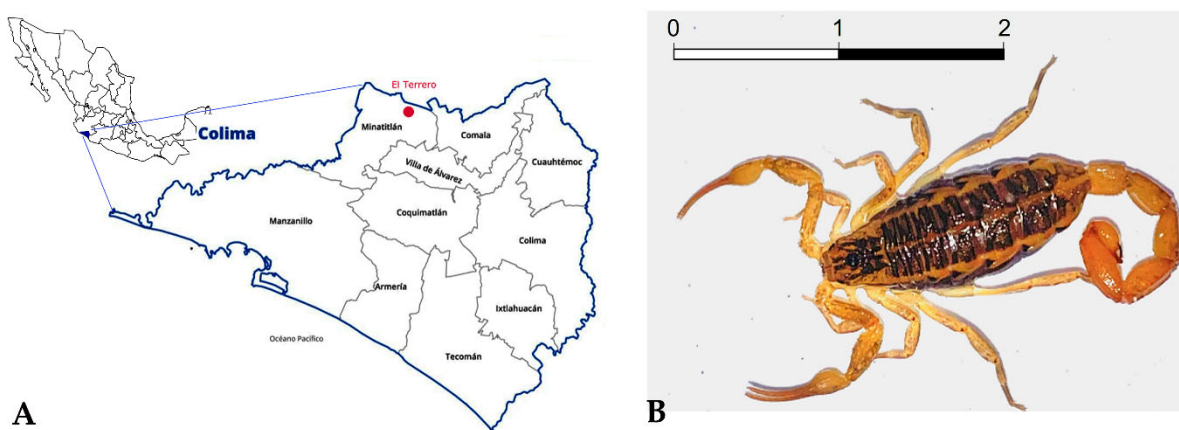


Figure 1. Distribution and morphology of the scorpion *Centruroides possanii*. (A) The orange dot indicates the geographical distribution and collection area of the scorpion *C. possanii* in Minatitlán, Colima, Mexico. (B) A female specimen of scorpion *C. possanii*. The scale bar 1:1 represents that 1 cm in the image equals 1 cm in reality.

Recent studies have permitted researchers to correlate the structures of the peptides and proteins with their functions, enriching the information contained in the database, improving antivenom, and identifying new therapeutic agents and compounds with biotechnological applications. Furthermore, the medium lethal dose of venom was determined. All of this evidence suggests that the scorpion *C. possanii* could be considered one of the most toxic Mexican species.

2. Results and Discussion

2.1. Determination of LD₅₀ of *C. possanii* Venom

The LD₅₀ of *C. possanii* venom has been estimated to be 13.18 ± 0.67 µg/20 g mouse, or 659 ± 33.5 ng/g of mouse weight. The dose was validated using six mice, and death was observed in 50%. Mice injected with this venom presented intoxication signs, such as squealing, jumping, goosebumps, difficulty breathing, and death in a few minutes. The signs which were observed may have been related to neurotoxins present in *C. possanii* venom, which affect excitable cells, as has been described for other scorpions of the same genus, like *C. tecomanus* (10.2 µg/20 g mouse) and *C. limpidus* (15.0 µg/20 g mouse) [10,11]. The scorpion *C. possanii* should be considered of medical importance.

2.2. RNA Integrity and Sequenciatiion Quality

The results obtained by the Bioanalyzer 2100 (Agilent Technologies, Waldbronn, Germany) showed little RNA degradation (Figure 2). The RNA integrity number (RIN) determines the RNA integrity through 18S and 28S peaks observed in the electropherogram. However, in scorpions, there is a dissociation of the 28S rRNA subunits, causing these to migrate on a gel at the same rate as the 18S rRNA subunit, which does not permit the determination of real RIN [12]. Nonetheless, a clearly defined 18S rRNA band in the gel and its peak in the electropherogram were observed, indicating the RNA's quality and integrity. The RNA concentration of the sample was 104 ng/µL. The Phred quality level of the reads was 30, where the probability of finding an incorrect base is 1 in 1000, with an accuracy of 99.9%.

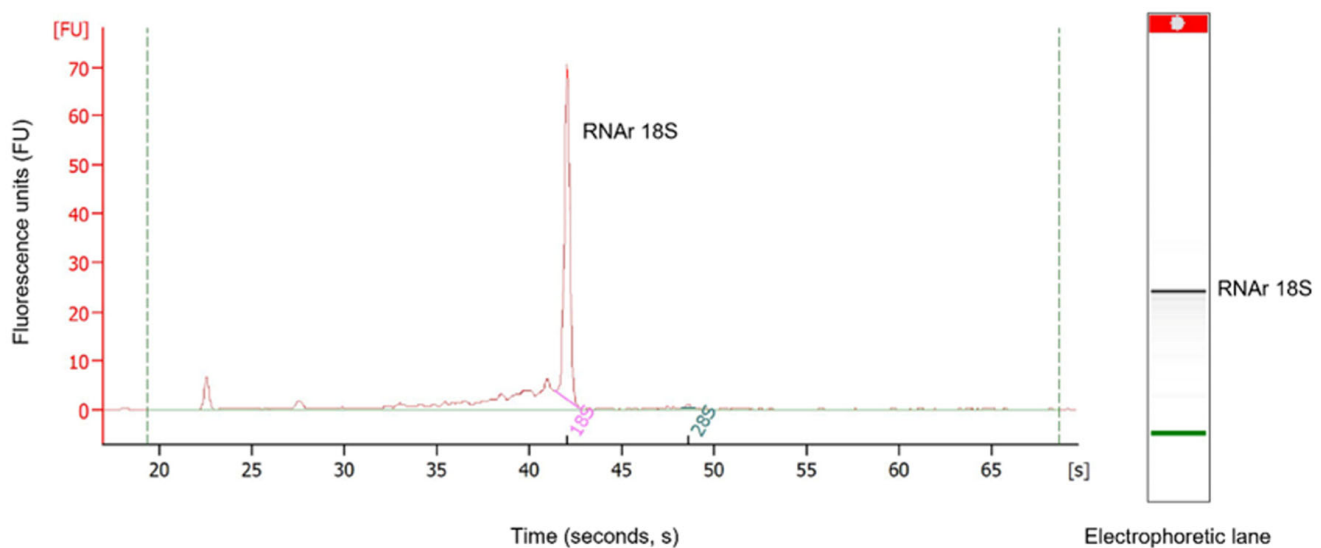


Figure 2. RNA analysis using Bioanalyzer 2100 (Agilent). The left image shows the representative electropherogram of the total RNA extracted from the venom-producing gland of the scorpion *C. possanii*. The separation of the 18S rRNA fragment in the gel is shown in the right image. The dotter points indicated in green refer to the internal standard of the equipment.

2.3. Bioinformatics

In the transcriptome of *C. possanii*, 19,158,736 reads were obtained with 76 base pairs (bp), as expected. The reads were assembled de novo, and this resulted in 835,204 transcripts. From them, 720,463 were annotated with some biological function. Of this, 434,492 were putative venom transcripts of the scorpion. The Pfam annotation resulted in 28,399 transcripts. These were classified into 56 Pfam scorpion domains, of which 50 were specific to Pfam scorpion venom domains. Table 1 shows the results of the assembly and annotation.

Table 1. Scorpion transcriptome assembly and annotation for *C. possanii*.

Entry	Quantity
Sequence length	76 base pairs (bp)
Reads	19,158,736
Total transcripts	835,204
Annotated transcripts	720,463
Putative transcripts	598,252
Venom transcripts	434,492
No venom transcripts	163,760
Pfam annotation	28,399
Pfam scorpion domain	56
Pfam scorpion venom domain	50
GO terms	18.4% cell components 30.8% biological process 50.7% molecular function

2.4. Venom Components Found in the Secretory Venom Gland of *C. possanii*

A database was generated for identification of the components which were found. We downloaded another database of scorpion venom components and other venomous species, such as spiders; sea cone snail anemones; and snakes, from UniProt, Animal Toxin Annotation Project, ArachnoServer, PDB, NCBI, and ScorpKTx. With all of the above information, a single local database was developed and 434,606 unique venom sequences were obtained, of which 162,285 were toxins, 129,176 enzymes, 7757 host defense peptides, 39,895 protease inhibitors, and 48,316 classified as other peptides. The curing of the sequences was performed by means of the BLASTp algorithm, using a transdecoder file with the local database as the first step for obtaining the list of transcripts to cure. Through transcriptomic analysis, the sequences with identity to venom components were identified, and the enzymes were the most diverse group (Table 2). However, the sodium channel-acting toxins (NaTx) were the most abundant. The TMP values indicate the relative abundance of the transcripts, shown in Supplementary Table S1.

Table 2. Results in summary of the *C. possanii* scorpion transcriptome.

Family	Subfamily	Number of Coding Transcripts
Toxins	α -NaTx	21
	β -NaTx	50
	α -KTx	20
	β -KTx	1
	γ -KTx	5
	δ -KTx	12
Enzymes	PA1	4
	PA2	4
	PLB	1
	MtP	16
	SeP	14
	Hya	1
	CiP	2
	PDE	3
	Mon	2

Table 2. Cont.

Family	Subfamily	Number of Coding Transcripts
Inhibitory peptides	SERPIN	5
	TIL	11
	Kunitz type	1
Defense peptides	Defensin	10
	NDBP-3	1
Other components	CAP	7
	IGFBP	27
	La1-like	7
	Und	19
Total		244

Abbreviations: α -NaTx = alpha-type sodium toxins; β -NaTx = beta-type sodium toxins; α -KTx = alpha-type potassium toxins; β -KTx = beta-type potassium toxins; γ -KTx = gamma-type potassium toxins; δ -KTx = delta-type potassium toxins; PA1 = phospholipase type 1; PA2 = phospholipase type 2; PLB = phospholipase type B; MtP = metalloproteinase; SeP = serine protease; Hya = hyaluronidase; CiP = cysteine protease; PDE = phosphodiesterase; Mon = monooxygenase; SERPIN = serine protease inhibitor; TIL = trypsin inhibitor; Kunitz type = kunitz-type inhibitor peptide; NDBP-3 = non-disulfide-bridged peptide type 3; CAP = cysteine-rich secretory proteins, antigen 5, and pathogenesis-related 1 proteins (CAP superfamily); IGFBP = insulin-like growth factor-binding proteins; Und = undefined proteins.

2.4.1. Ion Channel-Acting Toxins

Sodium Channel-Acting Toxins

Previous reports of transcriptomes in scorpions with different levels of toxicity to mammalian species have shown that toxins acting on sodium channels are the most abundant peptides in the venom of these arachnids, especially those of the *Buthidae* family [10]. This is also the case for *C. possanii*, in which 218 (110 complete) different coding transcripts surmised to affect NaTxS were identified (Supplementary Table S1). The transcripts were named CpoNaTx.

α -NaTx. Analysis of the results showed 39 coding transcripts for α -NaTx, 21 of which were complete. The CpoNaTalp10 sequence showed 69% identity with the peptide CsE5 (UniProt P46066) from the venom of *C. sculpturatus*, which is toxic to vertebrates and invertebrates, but more toxic to vertebrates. The toxicity of this protein to mammals may be due to the presence of lysine 32 and 75 (Figure 3), as well as in beta toxins AaH2 from the scorpion *Androctonus australis* and Csx2 from *C. suffusus*. The capacity of CsE5 to block neuronal transmission and induce spontaneous rhythmic muscle contractions is due to structure–function relationships. In insects, the toxicity is related to the additional three C-terminal residues (N⁸¹, P⁸², A⁸³) [13]. CpoNaTalp10 also has 67% identity with the toxin Pg8 (UniProt B7SNV8), which is the most toxic component of the venom of the South African scorpion *Parabuthus granulatus* [14] (Figure 3). Toxin Pg8 acts on site 3 of sodium channels, slowing the inactivation process [15].

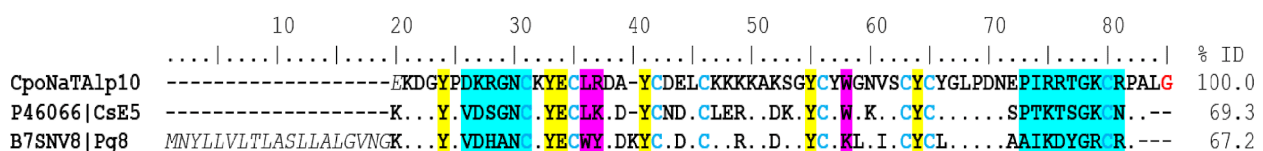


Figure 3. Alignment of some alpha toxins that act on sodium channels. The signal peptide is shown in gray and italics, and the mature peptide is in bold. Dots indicate identical amino acids. In blue is the identical pattern of the eight cysteines forming the disulfide bridges (positions 31–83, 35–56, 42–66, and 46–68). In yellow are the hydrophobic residues essential for their function on sodium channels; in cyan the variable region of the domain functional amino-carboxyl “NC”; and in pink, the amino acids of the functional “core” domain. Percent identity (% ID) is shown for mature peptides only. The amidation signal is represented in red color (position 85).

The sequence CpoNaTAlp05 shows an identity of 57% and 35% with the CvIV4-like toxin from *C. sculpturatus* and Cn12 from the *C. noxius* scorpion, respectively. These are alpha toxins, but they show tridimensional conformation similar to beta toxins. Experimental data show that it behaves like an alpha-type toxin by inducing inactivation of the sodium channel [6]. The presence of CpoNaTAlp05 was also identified in the soluble venom, from which three unique peptides were found (Supplementary Table S2).

β -NaTx. Fifty complete sequences of coding transcripts in the transcriptome and fourteen proteins in proteome analyses were identified. These peptides are the most abundant in the venom of scorpions of the *Centruroides* genus, and are the main toxins responsible for causing the symptoms and signs of envenoming in mammals [16]. *C. hirsutipalpus*, with 77 β -NaTxs, is thus far the scorpion described as having the highest number of transcripts assumed to code for sodium channel toxins affecting mammals [17], whereas *C. limpidus* contains 59 transcripts [10], *Tityus obscurus* 48 [18], and *C. hentzi* 36 transcripts [19].

CpoNatBet09, CpoNatBet14, and CpoNaTBet35 were the most abundant both in the transcriptome and the proteome (Supplementary Tables S1 and S2). The sequence CpoNatBet09 was 100% identical to the Cll2b toxin (UniProt P59899) from the *C. limpidus*; it presented identities of 89% and 86% with the toxins Css4 (UniProt P60266; *C. suffusus*) and Cn2 (UniProt P01495; *C. noxius*), respectively (Figure 4). These toxins are highly represented in their venoms, and present regions that are conserved in their amino acid sequences and are essential for their activity on sodium channels (L³⁹, N⁴², Y⁴⁴, R⁴⁷, E⁴⁸, Q⁵², W⁷⁸) [20,21]. In the case of the Cn2 toxin, it represents 6.8% of the total venom, with an LD₅₀ of 0.25 μ g/20 g of the mouse to which it was administered intraperitoneally [22,23]. It also showed affinity to Nav 1.6 channels [24]. The transcript CpoNaTBet09, which codes for a peptide showing high sequence similarity or even complete identity to the *C. limpidus* toxin Cll2b, is certainly a component that plays an important role in the intoxication process.

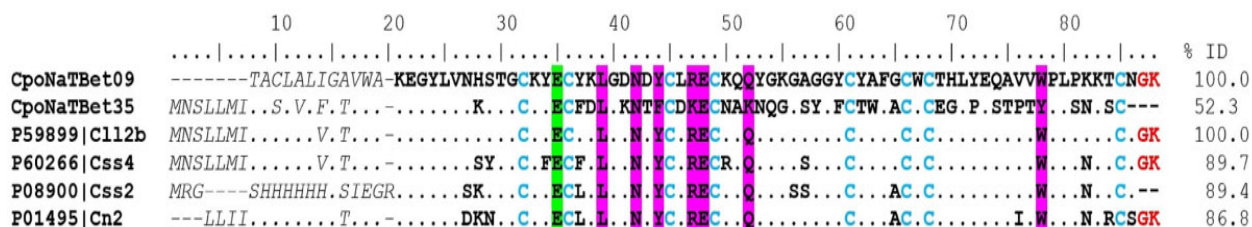


Figure 4. Alignment of beta toxins that act on sodium channels. The signal peptide shown is in italics and a gray color, and the mature peptide is in bold. Dots indicate identical amino acids. The identity percentage (% ID) refers only to the region of mature peptides. Cysteines (blue) present a conserved pattern formed by disulfide bridges at positions 31–84, 35–60, 44–65, and 48–67. Glutamic acid of position 35 (E35; green) is essential for binding β -NaTxs with the Nav channel. The crucial residues for the activity are shown in magenta, and the amidation and cut signals are in red.

Potassium Channel-Acting Toxins (KTx)

In the transcriptome of *C. possanii*, we identified 38 complete sequences which coded for toxins that act on potassium channels. Of these, 20 sequences were classified as α -KTx, one β -KTx, five γ -KTx, and twelve δ -KTx (Supplementary Table S1). The proteomic analysis permitted the identification of five of these sequences: one β -KTx, three α -KTx, and one γ -KTx (Supplementary Table S2). This indicates that the transcriptome of *C. possanii* has the highest number of KTx-similar transcripts reported thus far, when compared with 26 of *C. limpidus* [10] and 21 of *C. hirsutipalpus* [17].

α -KTx. In the transcriptome of *C. possanii*, one coding transcript with an identity to α -KTx, called CpoKTxAlp04, was identified, which showed 100% identity with cobatoxin 1 (UniProt O46028) of the scorpion *C. noxius* and 96.9% identity with the KTx 10.4 toxin (UniProt C0HJW2) from *C. tecomanus* (Figure 5). These toxins consist of a polypeptide chain of 32 amino acids stabilized with 3 disulfide bridges. In their mature sequences,

five conserved residues are present: K⁴⁹, I⁵¹, N⁵², K⁵⁶, and Y⁵⁸. These are believed to be important to the interaction with the channel Kv1.1 channels [25].

	10	20	30	40	50	60	% ID
CpoKTxAlp04	MAAIAKITLILLFLFVTMDTFANWNAEA	AVCVYRTCDKDCKRRGYRSGK	CINNACKCYPYGK	100.0			
O46028 Cobatoxin-1	.EG.....H.....T.....C.....C.....C.....K.....C.....C.....GK	100.0					
C0HJW2 Alpha-KTx 10.4	-----C.....C.....C.....K.....C.....C.....A-	96.9					

Figure 5. Alignment of alpha toxins that act on potassium ion channels. The signal peptide is shown in gray and italics. The underlined region represents the propeptide, the mature peptide is shown in bold, and in blue are the cysteines that form the disulfide bridges (positions 31–50, 35–55, and 40–57). Dots indicate identical amino acids. Amidation (G) and cleavage (K) signals are shown in red. The identity percentage (% ID) takes into account only mature peptides. The green shaded regions indicate the amino acids that form the functional dyad (K and Y) and interact with the potassium channel.

β -KTx. The CpoKTxBet01 sequence was identified in the proteomic analysis, and two unique peptides were found (Supplementary Table S2). It showed similarity with the TdiKIK-like toxin (NCBI XP_023211330) of *C. sculpturatus*.

γ -KTx. Twenty-two coding transcripts were found with identity to gamma toxins. Five coding transcripts were complete (Supplementary Table S1). The CpoKTxGam01 coding transcript showed identities of 55% and 52% with CnErg1 (UniProt Q86QT3, *C. noxius*), CeErg1 (UniProt Q86QV6, *C. elegans*), and CmERG1 (UniProt C0HLM3, *C. margaritatus*) toxins, respectively (Figure 6). These toxins contain residue of lysine in position 34 (K³⁴) in their sequences, which could be the principal interaction mode with the potassium channel [26]. The CnErg1 toxin blocks the Kv11.1 and 11.3 channels in a closed state in humans and rats [27], while the CmERG1 toxin only blocks the Kv11.1 channel [28]. The CpoKTxGam05_01 sequence was identified in the proteome, where two unique peptides were identified (Supplementary Table S2). This peptide showed 100% identity with the CllErgTx2 toxin (UniProt Q86QU9) from *C. limpidus*.

	10	20	30	40	50	60	% ID
							
CpoKTxGam01	MKNILLIVIIIFSAFILNGVELAK	DSCKHLKRCGKYGFYKNC	SECCREFGHEGGYCTMFCKCKT				100.0
Q86QT3.1 KGX11	..VLI..M..A.LM.M-...MDR..	CVDKS.CAK..Y.QECQDCKKNA..N..TCMF..C.C-A					55.0
Q86QV6.1 KGX12	-----DR..	CVDKS.CAK..Y.QECTDCKKYY..N..TCMF..C.C-A					55.0
C0HLM3.1 KGX1A	-----DR..	CVDKS.CAK..YFQECTDCKKYY..N..TCMF..C.C-A					52.5

Figure 6. Alignment of gamma toxins acting on potassium ion channels. The signal peptide is shown in a gray color and italics. The mature peptide is shown in bold, and the cysteines forming the disulfide bridges are blue. Dots indicate identical amino acids. Highlighted in magenta is the amino acid involved with the binding site of the channel. The identity percentage (% ID) was estimated only for mature peptides.

δ -KTx. The peptides of this subfamily have a Kunitz-type structural folding with protease activity and Kv1.3 potassium channel blockers [29]. According to this classification, ninety-four coding transcripts were identified. Still, only 12 sequences were complete. Of these, the coding transcripts (CpoKTxDel01–10) showed identity with the BmKTT-2 toxin (UniProt P0DJ50) from the scorpion *Mesobuthus martensii*, presenting an identical pattern of eight cysteines in their sequence, and were classified in group 1. Two coding transcripts (CpoKTxDel11 and 12) with sequences like the BmKTT-1 toxin (UniProt P0DJ49) from *Mesobuthus martensii* were found and categorized in group 2. The CpoKTxDel03 coding transcript shows 86% identity with the BmKTT-2 toxin and 55% and 50% identity with the Kunitz-type carboxypeptidase inhibitor Kci-1 (GenBank AIX8708, *Androctonus bicolor*) and Conotoxin precursor conkunitzin, partial (GenBank Uma82726, *Conus ebraeus*), respectively (Figure 7). The BmKTT-2 toxin inhibits trypsin and chymotrypsin and is a

potent inhibitor of human plasmin, suggesting it may have good stability for oral drug delivery [30]. In contrast, the putative peptides of group 2, i.e., the CpoKTxDel11 and 12 showed 86% and 56% identity with peptides of Kunitz-type serine protease inhibitor LmKTT-1a-like from *C. sculpturatus* (NCBI XP_023242637) and *Lychnis mucronatus* (UniProt P0DJ46), respectively, as well as 50% identity with venom protein VP5 (GenBank ALX72370) of *Odontobuthus doriae* [31] (Figure 8). Although the LmKTT-1a toxin completely inhibits trypsin activity, it does not affect chymotrypsin or elastase and blocks potassium currents in the Kv1.3 channel [29,32].

	10	20	30	40	50	60	70	80	% ID
CpoKTxDe103		MKLIINILVIL-LFDSFCTSEDVDCTLPDPSGLCLALFIRYYNPKTKGEGFSYGGCGGNSNSFKSRECCKACGDDGCLK	100.0						
P0DJ50.2 BmKTT-2		.MNV.TVVG...SVVCTISDA.G...C...S.T.RCKAY...F..Q.A.ECQK.V...CE...N.LTKSDCC.QCSPGKC--	86.4						
AIX87608.1 Koi-1		.MNV.IVIG...PVLC.ISNA.E...C...ET.RCK.NIF..F..QSAAECQT.I...CE...N.ETKPDCCRHCSPGNC--	55.9						
UMA82726.1 Conkunitzin		L-----ANADRC...VT.VCM.Y.P.WH.DSTS.QCQQ.VW...C...A.N.L.KD.CVQ.C..H----	50.8						

Figure 7. Alignment of some group 1 delta toxins acting on potassium channels. The signal peptide is shown in gray and italics, the mature peptide in bold, and the cysteines forming the disulfide bridges in blue. Dots indicate identical amino acids. In yellow, the amino acids Lys³⁶ and Ala³⁷ of the BmKTT-2 toxin are essential for function. The identity percentage (% ID) took into consideration only mature peptides.

	10	20	30	40	50	60	70	80	% ID
CpoKTxDel11		MKSILLIVAILCCLFNSFADGKKDCFAKPDVGRGRALLKRYYNTETSTCHMFYGGAGGNGNFLSKSGCCQVCIHGS							100.0
CpoKTxDel12		C.....F.....CE.....CC..C...C							96.6
XP_023242637.1 LmKTT-1a-like		VV.....A...C...A...S...CDI...T...E...CC..ICK...C							86.4
P0DJ46 LmKTT-1a		...F...ALV.FF...V.Y.SA..KCQLPS...K.K.SFT...E.SGKCET.I...V...S...T.EDCCRECAQ...C							55.9
P0DJ45.1 LmKTT-1b		...F...ALV.FF...V.Y.SA..NKCKLPS...K.K.SFT...E.GGKCET.I...V...S...T.EDCCRECAQ...C							54.2
P0DJ48.1 LmKTT-1c		-----NKCKLPS...K.K.SFT...F..E.SGKCET.I...M...S...T.EACCRECAQ...C							52.5
ALX72370.1 VP5		...AVIVAIT...C...LY.NA...CSLPV.T.K.KGWFL...KNSK.CET.I...I...K...N.ENCCIKCNAENC							50.4

Figure 8. Alignment of group 2 delta toxins acting on potassium ion channels. The signal peptide is shown in gray and italics, the mature peptide in bold, and the cysteines forming the disulfide bridges in blue. Dots indicate identical amino acids. In yellow is the amino acid Lys³⁵ of the LmKTT-1a toxin, which interacts directly with trypsin. The identity percentage (% ID) considered only the region of mature peptides.

2.4.2. Enzymes

Enzymes are a group of diverse components identified in the transcriptome of *C. possanii* and other species, like *C. hirsutipalpus* [17] and *C. limpidus* [10]. This study reports the presence of 915 coding transcripts showing identity to enzymes. They can be classified into 12 subfamilies: cysteine proteases (CiP), serine proteases (SeP), metalloproteases (MtP), monooxygenases (Mon), hyaluronidases (Hya), phosphodiesterases (PDE), phospholipase A1 (PLA1), phospholipase A2 (PLA2), phospholipase B (PLB), 5'-nucleotidases (5'-NT), phospholipase C (PLC), and phospholipase D (PLD). Here, 47 transcripts were identified, coding for 9 of the 12 subfamilies (Supplementary Table S1).

PLA1. Four complete coding transcripts (CpoEnzPA101-04) with identity to phospholipases A1 were identified in the transcriptome, and three sequences (CpoEnzPA101, 03 and 04) in the proteomic analysis (Supplementary Tables S1 and S2). Enzymes from this group have been reported in venom gland transcriptomes of non-toxic scorpions, such as *Hadrurus spadix* [33] and *C. hentzi* [19]. Unfortunately, none of these have experimental evidence of their function or the presence of the protein in the venom. This is the first report that has determined the presence of PLA1 in a transcriptomic and proteomic analysis of scorpion species known to be toxic to mammals.

PLA2. Phospholipases A2 (PLA2) found in scorpion venom presented histidine (H) and aspartic acid (D) residues conserved in the active site, according to the following proteins, the function of which was already experimentally proven: Plaiodactylipin (UniProt Q6PXP0) from *Anuroctonus phaiodactylus*, HgPLA2 (UniProt P0C8L9) from *Hoffmanniadrurus gertschi*, HfPLA2 (UniProt Q3YAU5) from *Chersonesometrus fulvipes*, Imperatoxin-1 (UniProt P59888) from

Pandinus imperator, Heteromtoxin (UniProt P0DMI6) *Heterometrus laoticus*, and MtsPLA2 (UniProt Q6T178) from *Hottentotta tamulus*. Four coding transcripts with identity to PLA2 (CpoEnzPA201-04) were identified in this transcriptome (Supplementary Table S1) and the CpoEnzPA203 sequence in the proteomic analysis (Supplementary Table S2). PLA2 hydrolyzes phospholipids, modifying the structure of the membranes and performing pre-digestive functions, and could be considered a spreading factor. They are also the most widely distributed enzyme in animal venoms [34].

PLB. The complete transcript was named CpoEnzPLB01 and presented 98.2% identity with the phospholipase B-like 1 (UniProt Q6P4A8) from the scorpion *C. sculpturatus*. The CpoEnzPLB01 coding transcript showed identity with the Phospholipases B-like (UniProt A0A1W7R9X8) from *Hadrurus spadix* and (UniProt A0A2I9LPG8) from *C. hentzi*. The CpoEnzPLB01 complete coding transcript presented the Pfam PF04916 (Phospholipase B) domain and the conserved Phospholipase B domain, according to CDD/Sparcle [35]. Six unique peptides of CpoEnzPLB01 coding transcripts were found in the proteome (Supplementary Table S2). The exact role of this enzyme in scorpion venom has not yet been described, but in snake venom, it is known to act as a potent hemolytic agent [36].

Hya. One coding transcript (CpoEnzHya01) with an identity to hyaluronidase was identified in the transcriptome and in the venom from *C. possanii*. Its sequence presented 98.9%, 75.9%, and 73.3% identity with the hyaluronidase-1-like (NCBI XP_023226974.1) from *C. sculpturatus*, hyaluronidase 1 (UniProt P85841) from *Tityus serrulatus*, and (UniProt P86100) from the scorpion *Mesobuthus martensii*, respectively. These sequences contain an aspartic acid at position 131 (D¹³¹), corresponding to the active site. Its function in the venom is well documented. It degrades the hyaluronic acid of the extracellular matrix, thus allowing for the diffusion of other venom components, such as toxins, increasing its potency and damaging the local tissue at the sting site [37].

Mtp. Sixteen complete transcripts were identified as coding for metalloproteinases, including antarease-like, astacin-like, and angiotensin-converting enzyme-like. The CpoEnzMtp06 coding transcript has 91.7% identity with the antarease-like TtrivMP_A (NCBI XP_023228637.1) from *C. sculpturatus*. These two enzymes contain 505 amino acids and present the same pattern of cysteines, forming nine disulfide bridges. The region with metalloprotease activity comprises amino acids 181 to 405; it has been proposed that residues 338 to 348 constitute the active site of the enzyme. Eleven metalloproteases of the transcriptome also were found in the proteome (Supplementary Table S2).

On the other hand, a sequence of 606 amino acids with identity to angiotensin-converting enzyme (ACE)-type protease was found. In mammals, this protein plays a role in the renin-angiotensin system, which regulates blood pressure and chloride retention via the kidney. The ACE protein converts angiotensin I to angiotensin II, its active form, resulting in increased vasoconstrictor activity and causing an elevation of peripheral vascular resistance, therefore leading to an increase in blood pressure [38]. ACE-type proteins have been found in some sea cone and wasp species. In scorpions, they have been reported in transcriptomic analyses of *Hottentotta judaicus* [39], *Tityus bahiensis* [40], *T. stigmurus* [41], and *T. serrulatus* [42]. ACE-like protein, one of the components of scorpion venom, may contribute to the arterial hypertension observed in human victims during envenomation caused by the sting of this arachnid [40]. The CpoEnzMtp08 coding transcript showed identity mainly with two ACE-like proteins from *C. sculpturatus* (NCBI XP_023209358.1) and *Tityus serrulatus* (UniProt JAW07033.1). These sequences present the conserved HHExXH motif of metallopeptidases as well as acetylated and deamidated regions.

CiP. Two coding transcripts with identity to cysteine proteases were identified in the transcriptome (Supplementary Table S1), and one sequence in the proteome (Supplementary Table S2). Cysteine proteases have not been reported in species of the *Centruroides* genus. But they have been identified in the transcriptome of *Hadrurus spadix* [33], *T. obscurus*, *T. serrulatus* [18], and *T. bahiensis* [40]. The coding transcript CpoEnzCiP01 has 97.2% identity with cathepsin L-like precursor (NCBI XP_023238299.1) from the scorpion *C. sculpturatus*. This sequence presents its cysteine protease domain in the 236–453 position

of the amino acid sequence. CpoEnzCiP01 also showed 83.4% identity with the cathepsin F-like (UniProt U6JPB2) from the scorpion *Tityus serrulatus*, and its cysteine protease domain was found in the 228–441 positions of the sequence. The cysteine proteases are large proteins, and the sequences comprise 446 (CpoEnzCiP01 and *T. serrulatus*) and 454 (*C. sculpturatus*) amino acids. The cysteine protease activity is thought to be at the carboxyl terminal. This family of proteins has the cysteine–histidine–asparagine catalytic triad at the active site (C³⁸⁹, H³⁹⁶, and N⁴²²) (Supplementary Figure S1).

Mon. Two coding transcripts with identity to monooxygenase were identified in the transcriptome and the proteome: CpoEnzMon01 and CpoEnzMon02 (Supplementary Tables S1 and S2). The CpoEnzMon01 transcript showed identity with the monooxygenase sequences (NCBI XP_023239442.1) from *C. sculpturatus* (97.5%) and *C. hentzi* (NCBI A0A2I9LP74; 92.3%). One of the main activities of the monooxygenases is to carry out post-translational modifications to some toxins, particularly through amidation of the C-terminus (α -amidation), which is necessary for the toxins to be functional [43].

SeP. Fourteen transcripts coding for serine proteases were identified in the transcriptome, and four in the proteome. The CpoEnzSeP04 coding transcript had 99.4% identity with the serine proteinase stubble-like (NCBI XP_023209481.1) from *C. sculpturatus*. Both sequences contained 360 amino acids and the same cysteine pattern (Supplementary Figure S2). These enzymes presented a region with trypsin-like activity in the 116–355 positions of the amino acids. Their active sites were formed by the H¹⁶³, D²¹³, and S³¹¹, while the substrate binding sites comprised D³⁰⁵, S³³¹, and G³³³ residues. Serin proteases are abundant enzymes of the venom of viperids [44]. Although their functions are diverse, they generally have hemotoxic activity, acting as anticoagulants or procoagulants [44].

PDE. Twenty-three sequences showed identity with phosphodiesterases; three of them were complete sequences and identified as CpoEnzPDE01–CpoEnzPDE03 in the transcriptome and proteome results. Phosphodiesterases have been reported at the transcriptomic level of the scorpion venom gland [45], in the venom of snakes [46], and in *Loxosceles* spiders, where it is the component responsible for causing dermo-necrosis and systemic effects [47]. The CpoEnzPDE01 coding transcript shows 97.6% identity with the phosphodiesterase-type sphingomyelin (NCBI XP_023216777.1) from the scorpion *C. sculpturatus* (Supplementary Figure S3). This was the first report demonstrating the presence of phosphodiesterases in scorpion venom. The great diversity and abundance of enzymes in the venom could contribute to the effects of intoxication observed in their prey and act as spreading factors, allowing for the diffusion of the venom components throughout the body of the victim. They also could be activators of other venom components due to their preservative or multifunctional qualities [45].

2.4.3. Host Defense Peptides (HDPs)

Defensins

Ten complete coding transcripts were identified with identity to defensins (Supplementary Table S1), and two of them were found in the proteome (CpoHDPDef07 and CpoHDPDef10; Supplementary Table S2). Defensins are small peptides (less than 50 amino acids) containing cysteines. They were found in insects, mollusks, and arachnids [48]. The transcriptomic analysis of venom secretory glands from diverse scorpion families has allowed for the description of coding transcripts with identity to defensin-like sequences, which are more abundant in species of scorpions that are non-toxic to mammals [10].

The CpoHDPDef01 coding transcript has 83.3%, 75%, and 75% identity with dual-function peptides such as Bmkdfsin3 (UniProt A0A384E0Y8) from scorpions *Buthus martensii*; 4 kDa defensin (UniProt P56686) from *Androctonus australis*; and 4 kDa defensin (UniProt P41965) from *Leiurus hebraeus*, respectively (Figure 9). These sequences present a conserved cysteine pattern, forming three disulfide bridges (4–25, 11–33, 15–35).

	10	20	30	40	50	60	% ID
CpoHDPDef01	IRVAVIYIV	ETATVEAGFG	CPFNQGA	CHKHCQS	IRKRGGY	CDGFLKHR	100.0
BmkDfsin3	MKTI..LF.LALVFCTL.MGM.	C.....	KC.R.CR.	R...R...	C.....	Q.CVC.RK	83.3
P56686 Aa	-----	C.....	C.R.CR.	R...R...	CA.LF.QTCTC	R-	75.0
P41965 Lh	-----	C.L....	C.R.CR.	R...R...	CA..F.QTCTC	RN	75.0

Figure 9. Alignment of Host Defense Peptides of the Defensin type. In italics are the amino acids of the signal peptide, in bold are the mature peptides, and dots indicate identical amino acids. The conserved cysteine pattern is in blue, and the amino acids indicate cleavage (RK) in red. The percentages of identity (% ID) between sequences correspond to the mature peptide.

The Bmkdfs3 toxin acts on the Kv1.1, Kv1.2, and Kv1.3 channels, inhibiting their current, and has potent antibacterial activity on Gram-positive bacteria such as *Staphylococcus aureus* (2.5 µM), *Micrococcus luteus* (2.5 µM), and *Staphylococcus epidermidis* (1.25 µM) [49]. The 4 KDa defensin from *L. hebraeus* also shows activity against *Bacillus subtilis*, although it is not considered a human pathogen; it could contaminate food. None of the above peptides show hemolytic activity [50]. Therefore, they could be potential candidates for the development of new antimicrobial agents.

Non-Disulfide-Bridged Peptide (NDBP)

These peptides do not contain cysteines, but have a high diversity of bioactivities, such as bradykinin-potentiating [51], antimicrobial [52], anticancer [53], and cytolytic and hemolytic activity [54]. The NDBPs were classified into five groups according to their pharmacological action, sequence similarity, and length [54,55]. This is the first report demonstrating the presence of NDBP-3 (CpoHDPND301) in the venom of scorpions of the *Centruroides* genus. This peptide presented a 74% identity with the NDBP 4.23-like (UniProt Q5G8B5) coding transcript of the scorpion *Tityus costatus*.

2.4.4. Peptides Inhibitors (PINs)

Seventeen complete coding transcripts with identity to protease-inhibitory peptides (PINs) were identified in the transcriptome. These were classified into three subfamilies: the cysteine-rich trypsin inhibitor domain (TIL domain), the serine protease inhibitor domain (SERPIN), and those with a Kunitz-type domain. The TIL protease members include anticoagulant proteins, elastase inhibitor, thrombin, trypsin, and chymotrypsin [56]. This domain contains ten cysteines forming five disulfide bridges (1–7, 2–6, 3–5, 4–10, and 8–9) in a sequence of 60 amino acids. The active site of these inhibitors is in the loop connecting beta sheets 1 and 2, located between two disulfide bridges [57]. In the transcriptome of *C. possanii*, eleven coding transcripts were identified with identity to TIL proteases. The CpoPinTIL07 coding transcript shows 76.3% identity with the VP5.1 peptide (GenBank ABY26681) from the scorpion *Lyctas mucronatus*, with cysteines in the same positions and a chain length of 60 amino acids in the mature peptide.

Five complete and eight partial coding transcripts with identity to SERPIN-type proteases were identified (Supplementary Table S1). This family is involved in inflammatory responses, complementary activation, blood coagulation, and mutations associated with emphysema, cirrhosis, and dementia [58].

The CpoPinKun01 coding transcript shows 80% identity with isoinhibitor K-like (NCBI XP_0232117496) from the scorpion *C. sculpturatus*. Both sequences contain seven cysteines. Finally, fragments of two TIL and four SERPIN-type proteases were identified in the proteomic analysis (Supplementary Tables S1 and S2).

2.4.5. Other Components

This category comprises various proteins, including insulin-like growth-factor-binding proteins (IGFBPs); members of the cysteine-rich secretory protein group, antigen 5, and pathogenesis-related 1 proteins (CAP superfamily); La1-like peptides with von Willebrand factor-containing proteins; and orphan or undefined proteins. The common characteristic

among these groups is that their functions as either transcripts or venom components remain unclear. Seventy-four coding transcripts were classified within this category, with sixty complete sequences (Supplementary Table S1).

Insulin-like Growth Factor-Binding Proteins (IGFBP)

Twenty-seven coding transcripts were identified in which a conserved IGFBP domain was found (Supplementary Table S1). One of them was also identified in the proteome (Supplementary Table S2). In this report, IGFBPs were the most abundant group in the “Other components” category. Although their function in venom is unknown, IGFBPs have been reported in previous transcriptomic studies [17,19,32,33,40,59], as well as in proteomes [17,33,40]. The CpoOthIGF02 coding transcript had an identity of 98% with the venom protein 302-like (NCBI XP_023234366) of *C. sculpturatus* and a 57% identity with the AbOp-5 peptide (GenBank AIX87724) from *Androctonus bicolor*.

Cysteine-Rich Secretory Proteins, Antigen 5, and Pathogenesis-Related 1 Proteins (CAP)

Protein groups belonging to this superfamily have been documented in arthropods, mammals [60], and reptiles [61]. These secreted proteins have a generalized extracellular function, acting in an endocrine or paracrine manner. They are involved in various processes, including regulating branching morphogenesis and the extracellular matrix, potentially as proteases or protease inhibitors; modulating ion channels in fertility as tumor suppressors; and during fertilization in cell–cell adhesion [62]. In scorpions, their function in the venom remains unknown. However, about 119 coding transcripts have been reported in transcriptomic analyses, and their presence in venom has been verified in species of the Buthidae, Vaejovidae, Caraboctonidae, Superstitionidae, and Euscorpidae families [18,33,63]. Seven coding transcripts with identity to CAP proteins were identified in this transcriptomic analysis, and three of them were found in the proteome (Supplementary Tables S1 and S2).

Orphan or Undefined Coding Transcripts

These components with no defined structural domain and no known function have been found in *Centruroides* scorpions: four coding transcripts and three peptides in *C. hentzi* [19], and three peptides in *C. limpidus* [10]. We identified nineteen coding transcripts with identity to orphan peptides, and five peptides were also found. This is the highest number of coding transcripts with identity to orphan peptides which has been reported in scorpions of the *Centruroides* genus to date, and the second most abundant group within this category according to transcriptome analysis. Given the information above, this group of peptides will certainly be studied further in the near future.

La1-like Peptides with Single-Domain von Willebrand Factor

The La1-like peptides are named after the first peptide to be identified in the venom of the *Liocheles australiase* scorpion of the family Hormuridae. These peptides present the structural domain of von Willebrand factor type C (SVWC) and contain eight cysteines within their amino acid sequences. Generally, their function is to respond to environmental challenges such as bacterial infections or antiviral immunity [64]. In the analysis of the transcriptome of venom-secreting glands of the scorpion *C. possanii*, seven coding transcripts were found with identity to proteins with the SVWC domain (Supplementary Table S1). In the proteome, one protein was identified (Supplementary Table S2).

2.5. Venom Proteomic Components of *C. possanii* Scorpion

2.5.1. Peptide Mass Fingerprinting

A fraction of the soluble venom from the scorpion *C. possanii* was analyzed using a mass spectrometer for molecular mass identification. The system yielded 180 unique individual masses with molecular weights of 800 to 17,000 Daltons (Da), as shown in Supplementary Table S3. The most abundant components (39 peptides) presented molec-

ular weights between 7001 and 8000 Da, attributed to toxins acting on sodium channels. This was related to the number of transcripts and peptides identified in the transcriptome and proteome, respectively, of the *C. possanii* scorpion. The lowest number of compounds (8) was found in the 800–1000 Da range. Additionally, 30 components were identified in the range of 1001 to 2000 Da, 23 components in 2001–3000 Da, 18 in the range of 3001–4000 Da, 24 in 4001–5000 Da, 10 in 5001–6000 Da, 19 in the range of 6001–7000, and only 9 in the range of 8001 to 17000 Da. Table 3 shows a summary of the results regarding the number of peptides and types of proteins identified in the venom of the *C. possanii* scorpion.

Table 3. Results and summary of the *C. possanii* scorpion proteome.

Family	Subfamily	Number of Peptides and Proteins
Toxins	α -NaTx	4
	β -NaTx	14
	α -KTx	3
	β -KTx	1
	γ -KTx	1
Enzymes	PA1	3
	PA2	1
	PLB	1
	MtP	11
	SeP	4
	Hya	1
	CiP	1
	PDE	3
	Mon	2
	SERPIN	4
Inhibitory peptides	TIL	2
Defense peptides	Defensin	2
	NDBP-3	1
Other components	CAP	4
	IGFBP	1
	La1-like	1
	Und	5
Total		70

2.5.2. Proteomics: Peptides and Proteins Identified in the *C. possanii* Scorpion Venom

In the LC-MS/MS, 50,873 spectra were analyzed, identifying 70 peptides and proteins. Between them were 23 toxins assumed to act on ionic channels. These consisted of four α -NaTx, fourteen β -NaTx, three α -KTx, one β -KTx, and one γ -KTx (Supplementary Table S2). Also, 27 enzymes were identified. These were classified into nine subfamilies: three phospholipases A1, one phospholipase A2, one phospholipase B, eleven metalloproteases, four serine proteases, one hyaluronidase, one cysteine protease, three phosphodiesterases, and two monooxygenases. Six protease-inhibitor peptides were identified: four of the SERPIN type and two of the TIL domain. The host defense peptide family was the least abundant, represented by two defensins and one NDBP-3. Finally, in the category of other components, four CAPs, one IGFBP, one La1-like, and five undefined peptides were identified (Table 3).

2.6. Correlation between Transcriptomics and Proteomics Results

The transcriptomic and proteomic analysis allowed for the identification of the five families of venom components, classified according to their structural domain: toxins, enzymes, PINs, HDPs, and other components. The most diverse family identified in the transcriptome was the toxins, with 109 members, followed by 60 in the group of other components, 47 enzymes, 17 PINs, and 11 HDPs. Instead, in the proteome, the enzymes were the most diverse, with 27 identified proteins, then the toxins with 23, followed by 11 peptides and proteins in the group of other components, 6 PINs, and 3 HDPs. The diversity pattern by family was very similar in the transcriptome and the proteome, the toxin family being the most diverse. This also supports the idea of recognizing *C. possanii* as a toxic species for mammals. Enzymes were greatly diverse in the proteome and induce post-translational modifications of the toxins, allowing for their activation. In addition, enzymes function as dispersal factors, allowing other venom components, such as toxins, to reach their sites of action. As in other species of the genus *Centruroides*, toxins and enzymes are well represented in the transcriptome and proteome of *C. possanii*, while the PIN and HDP families are less diverse in scorpions which are toxic to mammals.

The last published transcriptomes and proteomes for a scorpion of the genus *Centruroides* were released by Cid-Urbe et al. [10] and Valdez-Velázquez et al. [17]. In these, 192 coding transcripts and 46 identified proteins and peptides were reported in *C. limpidus* [10], in addition to 147 identified coding transcripts and 77 proteins and peptides in the venom of *C. hirsutipalpus* [17]. The same sequencing technology and methodology were used in both studies and for *C. possanii*. The successful identification of a large number of coding transcripts was due to the use of the reference database. This local database integrates identified coding transcripts, peptides, and protein sequences downloaded from various public databases, and includes articles on both terrestrial and marine venomous species, not just scorpions. Therefore, because a large amount of reference data was available, it was possible to identify a greater quantity of coding transcripts, peptides, and proteins. In addition to the above, the quality of the RNA, as well as the assembly and precise annotation of the transcriptome, allowed for the systematic curation of sequences that culminated in their analysis and successful identification.

3. Conclusions

A total of 244 complete transcripts were identified, coding for the five families of scorpion venom components: 109 toxins, 47 enzymes, 11 host defense peptides (HDPs), 17 protease inhibitors (PINs), and 60 in the category of other components. Transcriptomic and proteomic analyses of the *C. possanii* scorpion concluded that the peptides expected to act on sodium channels were the most abundant components of the soluble venom. The relative abundance of the CpoNaTBet35 peptide was 11.63%, whereas the peptide CpoNaTBeT09 component constituted 3.5% in the proteome. The venomous glands and the soluble venom of *C. possanii* showed the presence of enzymes such as phospholipase A1, phospholipase A2, phospholipase B, metalloprotease, serine protease, hyaluronidase, cysteine protease, phosphodiesterase, and monooxygenase. In this paper, three components (phospholipases A1, NDBP3, and phosphodiesterases) are described for the first time in the literature regarding scorpion species of medical importance. The *C. possanii* scorpion should be considered as a medically importance species due to the large number of toxic components in the venom and the confirmation of the LD₅₀.

4. Materials and Methods

4.1. Collection of Scorpions

C. possanii were collected at night using UV light in El Terrero, an area of Cerro Grande Mountain, under the collection permit SEMARNAT SGPA/DGVS/02139/22, which was granted to our group. This geographical elevation is located within the Sierra de Manantlán Biosphere Reserve in the municipality of Minatitlán, Colima (latitude 19°26′40.0″ N; longitude 103°57′02.5″ W), where the discovery of this species occurred and it was described by

González-Santillán et al. (2019) [7]. Ten scorpions were collected, carried to the Biological Products Laboratory of the University, and milked, applying an electric stimulation of 10 mA. They were kept in captivity at room temperature with permanent water access and fed with mealworms (*Tenebrio molitor*) or crickets (*Acheta domesticus*).

4.2. Determination of the Medium Lethal Dose (LD₅₀) of the Scorpion *C. possanii*

Twelve male and female mice of the BALB/c strain (25–30 g) of three months of age, obtained from the animal house of the School of Medicine of the University of Colima, were injected intraperitoneally with soluble venom of the *C. possanii* scorpion, adjusting the dosage according to the weight of each mouse. The dose injected into the first mouse was 15 µg/20 g mouse. Six mice were used to determine the dose, and another six mice were used to confirm it. The mice were maintained with humidity and a constant room temperature, with a 12 h dark–light cycle and food and water ad libitum, and were monitored for 24 h after the venom-injected assay. The study was performed with authorization of the University of Colima Ethics Committee (ethical approval code: FCQUCOL-CEIUEA-1). The determination of the LD₅₀ was performed using the Up and Down method reported by Dixon (1965) [65].

4.3. RNA Isolation, Sequencing and Reads Quality

First, the scorpions were milked by electric stimulation to obtain the venom, and after fourteen days of collection, the telsons from two males and two females of *C. possanii* were dissected under RNase-free conditions. The SV Total RNA Isolation System kit (Promega, Madison, WI, USA) was used for RNA extraction. The telsons were placed in a microtube with lysis buffer for maceration with a Kontes pestle. The sample was diluted with RNA dilution buffer and incubated at 70 °C for three minutes. Cell debris was separated by centrifugation, and the supernatant was mixed with ethanol at 95%. The product was centrifuged in a column provided by the kit. The column contents were washed with RNA wash solution, and the RNA was eluted in nuclease-free water. RNA was quantified on a Nanodrop, and its integrity was confirmed on the Agilent 2100 Bioanalyzer. Using the Illumina kit, the cDNA library was constructed starting from the total RNA.

Automated DNA sequencing was performed at the Bioinformatics and Sequencing Unit at the Biotechnology Institute of UNAM in Cuernavaca, Morelos, Mexico. The base pairs of the cDNA fragments were sequenced on a Genome Analyzer (Illumina) using the 76-base pair-end sequencing protocol. After trimming the adapter, the quality of the raw reads was evaluated with the FastQC program, v0.11.8.

4.4. Bioinformatics: Assembly, Annotation and Data Mining

The transcriptome assembly was performed with Trinity version 20.3 software [66,67], and the functional annotation of transcripts was performed using the Trinotate v3.2.1 pipeline (<https://github.com/Trinotate/Trinotate/wiki> (accessed on 1 July 2020) [68]. The BLAST program was used for the identification of the sequences, and thousands of sequences of proteins contained in the venoms database of UniProt's Animal Toxin Annotation Project (<https://www.uniprot.org/help/Toxins#:~:text=The%20animal%20toxin%20annotation%20project,a%20few%20fish%20and%20platypuses> (accessed on 1 July 2020)), ArachnoServer (<http://www.arachnoserver.org/> (accessed on 1 July 2020)), PDB, and NCBI were downloaded. Then, a local base made up of sequences extracted from publications of transcriptomes and proteomes from 2007 to 2020 was generated. SignalP Server version 5.0 or 6.0 (<https://services.healthtech.dtu.dk/services/SignalP> (accessed on 1 July 2020)) allowed for the identification of signal peptides, ProP 1.0 Server (<http://www.cbs.dtu.dk/services/ProP/> (accessed on 1 July 2020)) for the identification of propeptides, and DeepLoc 1.0 for determining the subcellular location (<http://www.cbs.dtu.dk/services/DeepLoc/> (accessed on 1 July 2020)). The conserved domains were determined with CDD/SPARCLE (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi> (accessed on 1 July 2020 to February 2023)). Multiple sequence alignments were

generated with MAFFT, version 7.0 (<https://www.ebi.ac.uk/Tools/msa/mafft/> (accessed on 1 July 2020 to 1 February 2023)). The coding transcripts reported in the following section were classified as complete sequences when they contained all amino acids corresponding to the mature peptide. Instead, those sequences with at least 50% of the amino acids of the mature peptide were considered to be partial. The Salmon tool, v1.9.0, (<https://github.com/COMBINE-lab/Salmon> (accessed on 1 February 2023)) was employed to estimate transcript-level abundance from RNA-seq data using the database of coding transcripts and the raw transcriptome database [69]. The relative abundance of the transcripts was expressed in units of transcripts per million (TPM).

4.5. Soluble Venom Collection and Quantification

Scorpions of the species *C. possanii* were milked by electric stimulation to obtain the venom. After adding distilled water, the sample was centrifugated and the supernatant, called soluble venom, was collected. The determination of the protein concentration in the venom was estimated by absorbance at 280 nm using an Evolution 300 UV-VIS spectrophotometer (Thermo Scientific, Madison, WA, USA), assuming that one unit of absorbance was equal to 1 mg/mL of protein.

4.6. Fingerprinting Mass Analysis of Soluble Venom

Five micrograms of soluble venom were applied in an LC-MS system composed of an HPLC nanoflow (Dyonex 3000, Thermo Scientific, San Jose, CA, USA) and a mass spectrometer (Velos-Orbitrap, from Thermo Scientific, San Jose, CA, USA). The sample was fractionated online and ionized by a nanospray ion source. For venom fractionation, the liquid chromatography system was stabilized by 95% solvent A (0.01% formic acid in water) using a homemade reverse phase column (10 cm in length, filled with Jupiter 4 μ m and Proteo 90 Å resin (Phenomenex, Torrance, CA, USA) for 20 min [14]. Then, the linear gradient was set to pumping solvent B (0.01% formic acid in acetonitrile) from 5 to 80% for 280 min at a 300 nL/min flow rate. All data on the spectra were acquired, and the raw files were extracted. Components with molecular masses lower and larger than 3 kDa were reported as monoisotopic and average masses, respectively. The allowed error between the theoretical and experimental molecular weight was 1 Da due to the sensitivity of the equipment.

4.7. Proteomic Analysis of Venom

One hundred micrograms of soluble venom were reduced with 50 mM dithiothreitol (Sigma Aldrich, Saint Louis, MO, USA) at 56 °C for 30 min, then alkylated with 10 mM iodoacetamide (Sigma Aldrich, Saint Louis, MO, USA) and incubated at room temperature for 30 min. After this treatment, the sample was desalted with C₁₈ ZipTips (Millipore; Billerica, MA, USA) and digested with trypsin (Promega, Madison, WI, USA) in 50 mM ammonium bicarbonate buffer. Two micrograms of the digested product were applied into the LC-MS system, which was composed of a Dionex UltiMate 300 nanoflow HPLC (Thermo Scientific, San Jose, CA, USA) with a nanoelectrospray ion source. The compounds were separated in the C18 column (Jupiter 4 μ m Proteo 90 Å resin), employing a gradient from 10% to 80% solvent B in 250 min. The peptidic fragments were analyzed in positive ion mode. Mass spectra were acquired in a full-scan *m/z* range between 350 and 1500 with a resolution of 60,000, using a collision-induced dissociation (CID) fragmentation method. The raw files were analyzed using Sequest HT in the Proteome Discoverer v1.4.1.14 software (Thermo Fisher Scientific, San Jose, CA, USA). In addition, static side chain modification of the cysteines (Carbamidomethylation), as well as two dynamic modifications, were considered: the oxidation in methionine and glutamine and asparagine deamidation. Sequest was searched with a fragment ion mass tolerance of 0.6 Da and a parent ion tolerance of 20 ppm. Scaffold (version Scaffold_5.1.2, Proteome Software Inc., Portland, OR, USA) was used to validate the MS/MS-based peptide and protein identifications. The protein sequences stored in the UniProt and the coding transcripts obtained from the transcriptome were used

to identify the peptides and proteins in the venom. Peptide identifications were accepted if they achieved greater than 99.0% probability and contained at least two identified peptides. Protein probabilities were assigned using the Protein Prophet algorithm [70]. Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/toxins15080498/s1>, Figure S1: Alignment of cysteine proteases; Figure S2: Alignment of Serine Protease Enzymes; Figure S3: Alignment of Phosphodiesterase Enzymes; Table S1: Coding transcripts identified in the venom-secreting gland transcriptome of the scorpion *Centruroides possanii*; Table S2: Peptides and proteins identified in the soluble venom of the scorpion *Centruroides possanii*. Table S3: Molecular weights (MW) of the venom components from *C. possani*.

Author Contributions: Conceptualization, L.L.V.-V. and L.D.P., methodology, P.E.G.-V., J.M.J.-V., G.J.L.-L., E.P.M., M.d.J.B.-G., C.E.B.-S. and I.D.E., formal analysis, P.E.G.-V., J.M.J.-V., G.J.L.-L. and E.P.M.; investigation, P.E.G.-V.; writing—original draft preparation, P.E.G.-V. and J.M.J.-V., writing—review and editing, L.L.V.-V., L.D.P., P.E.G.-V. and J.M.J.-V.; funding acquisition, L.L.V.-V. and L.D.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by a grant, partially supported by CONACYT Mexico, grant PRONAH number 303045 and Universidad Nacional Autónoma de México DGAPA-IN200323. Additional support was provided by Fortalecimiento de Cuerpos Académicos PRODEP 2021- UCOL-CA 117 “Bioquímica aplicada” to L.L.V.-V., J.M.J.-V., and C.E.B.-S.

Institutional Review Board Statement: *Centruroides posanii* scorpion specimens were collected in the El Terrero community in Minatitlan (Colima, Mexico), supported with official permission from the Secretaria de Medio Ambiente y Recursos Naturales, Mexico (SEMARNAT): SGPA/DGVS/02139/22. The LD₅₀ study of *C. possanii* venom was conducted in accordance with the Guide for Care and Use of Laboratory Animals of the US National Institutes of Health (NIH Publication No. 85–23, revised 1996), and was approved by the Institutional Animal Care and Use Committee of the University of Colima (approval code: FCQUCOL-CEIUEA-1) on June 1 2023.

Informed Consent Statement: Not applicable.

Data Availability Statement: A total of 19,158,736 reads were sequenced and are available in the European Nucleotide Archive (ENA), under accession number PRJEB60470.

Acknowledgments: We acknowledge CONACYT for the scholarship granted to P.E.G.-V. to obtain her doctoral degree (scholarship 747702). We acknowledge Alan Galvan Hernandez for technical support. We also acknowledge the students of the Faculty of Chemical Sciences of the University of Colima for their support in the collection of the *Centruroides possanii* scorpions.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

References

1. Lourenço, W.R. Scorpion Diversity and Distribution: Past and Present Patterns. In *Scorpion Venoms*; Springer: Dordrecht, The Netherlands, 2015; pp. 3–23.
2. Lourenço, W.R. The Evolution and Distribution of Noxious Species of Scorpions (Arachnida: Scorpiones). *J. Venom. Anim. Toxins Incl. Trop. Dis.* **2018**, *24*, 1. [CrossRef] [PubMed]
3. Rein, J. The Scorpion Files. Norwegian University of Science and Technology: Trondheim, Norway, 2022. Available online: <https://www.ntnu.no/ub/scorpion-files/> (accessed on 29 May 2023).
4. González, J.A.; Vallejo, J.R. The Scorpion in Spanish Folk Medicine: A Review of Traditional Remedies for Stings and Its Use as a Therapeutic Resource. *J. Ethnopharmacol.* **2013**, *146*, 62–74. [CrossRef] [PubMed]
5. Shao, J.; Zhang, R.; Ge, X.; Yang, B.; Zhang, J. Analgesic Peptides in *Buthus martensii* Karsch: A Traditional Chinese Animal Medicine. *Asian J. Tradit. Med.* **2007**, *2*, 45–50.
6. Bosmans, F.; Tytgat, J. Voltage-Gated Sodium Channel Modulation by Scorpion α -Toxins. *Toxicon* **2007**, *49*, 142–158. [CrossRef]
7. González-Santillán, E.; Galán-Sánchez, M.A.; Valdez-Velázquez, L.L. A New Species of *Centruroides* (Scorpiones, Buthidae) from Colima, Mexico. *C. R. Biol.* **2019**, *342*, 331–344. [CrossRef] [PubMed]

8. Secretaría de Salud Datos Abiertos—Dirección General de Epidemiología. Available online: https://epidemiologia.salud.gob.mx/anuario/html/morbilidad_estatal.html (accessed on 1 June 2023).
9. González-Santillán, E.; Possani, L.D. North American Scorpion Species of Public Health Importance with a Reappraisal of Historical Epidemiology. *Acta Trop.* **2018**, *187*, 264–274. [\[CrossRef\]](#)
10. Cid-Urbe, J.I.; Meneses, E.P.; Batista, C.V.F.; Ortiz, E.; Possani, L.D. Dissecting Toxicity: The Venom Gland Transcriptome and the Venom Proteome of the Highly Venomous Scorpion *Centruroides Limpidus* (Karsch, 1879). *Toxins* **2019**, *11*, 247. [\[CrossRef\]](#)
11. Riaño-Umbarila, L.; Rodríguez-Rodríguez, E.R.; Santibañez-López, C.E.; Güereca, L.; Uribe-Romero, S.J.; Gómez-Ramírez, I.V.; Cárcamo-Noriega, E.N.; Possani, L.D.; Becerril, B. Updating Knowledge on New Medically Important Scorpion Species in Mexico. *Toxicon* **2017**, *138*, 130–137. [\[CrossRef\]](#)
12. Natsidis, P.; Schiffer, P.H.; Salvador-Martínez, I.; Telford, M.J. Computational Discovery of Hidden Breaks in 28S Ribosomal RNAs across Eukaryotes and Consequences for RNA Integrity Numbers. *Sci. Rep.* **2019**, *9*, 19477. [\[CrossRef\]](#)
13. Jablonsky, M.J.; Watt, D.D.; Rama Krishna, N. Solution Structure of an Old World-like Neurotoxin from the Venom of the New World Scorpion *Centruroides sculpturatus* Ewing. *J. Mol. Biol.* **1995**, *248*, 449–458. [\[CrossRef\]](#)
14. Rojas-Azofeifa, D.; Sasa, M.; Lomonte, B.; Diego-García, E.; Ortiz, N.; Bonilla, F.; Murillo, R.; Tytgat, J.; Díaz, C. Biochemical Characterization of the Venom of Central American Scorpion *Didymocentrus krausi* Francke, 1978 (Diplocentridae) and Its Toxic Effects in vivo and in vitro. *Comp. Biochem. Physiol. Part C Toxicol. Pharmacol.* **2019**, *217*, 54–67. [\[CrossRef\]](#) [\[PubMed\]](#)
15. Stevens, M.; Peigneur, S.; Tytgat, J. Neurotoxins and Their Binding Areas on Voltage-Gated Sodium Channels. *Front. Pharmacol.* **2011**, *2*, 71. [\[CrossRef\]](#) [\[PubMed\]](#)
16. Mendoza-Tobar, L.L.; Meza-Cabrera, I.A.; Sepúlveda-Arias, J.C.; Guerrero-Vargas, J.A. Comparison of the Scorpionism Caused by *Centruroides margaritatus*, *Tityus pachyurus* and *Tityus* n. sp. aff. *metuendus* Scorpion Venoms in Colombia. *Toxins* **2021**, *13*, 757. [\[CrossRef\]](#)
17. Valdez-Velázquez, L.L.; Cid-Urbe, J.; Romero-Gutierrez, M.T.; Olamendi-Portugal, T.; Jimenez-Vargas, J.M.; Possani, L.D. Transcriptomic and Proteomic Analyses of the Venom and Venom Glands of *Centruroides hirsutipalpus*, a Dangerous Scorpion from Mexico. *Toxicon* **2020**, *179*, 21–32. [\[CrossRef\]](#)
18. de Oliveira, U.C.; Nishiyama, M.Y.; dos Santos, M.B.V.; Santos-da-Silva, A.d.P.; Chalkidis, H.d.M.; Souza-Imberg, A.; Candido, D.M.; Yamanouye, N.; Dorce, V.A.C.; Junqueira-de-Azevedo, I.d.L.M. Proteomic Endorsed Transcriptomic Profiles of Venom Glands from *Tityus obscurus* and *T. serrulatus* Scorpions. *PLoS ONE* **2018**, *13*, e0193739. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Ward, M.J.; Ellsworth, S.A.; Rokyt, D.R. Venom-Gland Transcriptomics and Venom Proteomics of the Hentz Striped Scorpion (*Centruroides hentzi*; Buthidae) Reveal High Toxin Diversity in a Harmless Member of a Lethal Family. *Toxicon* **2018**, *142*, 14–29. [\[CrossRef\]](#)
20. Martín, M.F.; García y Pérez, L.G.; el Ayeb, M.; Kopeyan, C.; Bechis, G.; Jover, E.; Rochat, H. Purification and Chemical and Biological Characterizations of Seven Toxins from the Mexican Scorpion, *Centruroides suffusus* Suffusus. *J. Biol. Chem.* **1987**, *262*, 4452–4459. [\[CrossRef\]](#)
21. Zamudio, F.; Saavedra, R.; Martín, B.M.; Gurrola-Briones, G.; Herion, P.; Possani, L.D. Amino Acid Sequence and Immunological Characterization with Monoclonal Antibodies of Two Toxins from the Venom of the Scorpion *Centruroides noxius* Hoffmann. *Eur. J. Biochem.* **1992**, *204*, 281–292. [\[CrossRef\]](#)
22. Riaño-Umbarila, L.; Ledezma-Candanoza, L.M.; Serrano-Posada, H.; Fernández-Taboada, G.; Olamendi-Portugal, T.; Rojas-Trejo, S.; Gómez-Ramírez, I.V.; Rudiño-Piñera, E.; Possani, L.D.; Becerril, B. Optimal Neutralization of *Centruroides noxius* Venom Is Understood through a Structural Complex between Two Antibody Fragments and the Cn2 Toxin. *J. Biol. Chem.* **2016**, *291*, 1619–1630. [\[CrossRef\]](#)
23. Selisko, B.; Licea, A.F.; Becerril, B.; Zamudio, F.; Possani, L.D.; Horjales, E. Antibody BCF2 against Scorpion Toxin Cn2 from *Centruroides noxius* Hoffmann: Primary Structure and Three-Dimensional Model as Free Fv Fragment and Complexed with Its Antigen. *Proteins* **1999**, *37*, 130–143. [\[CrossRef\]](#)
24. Schiavon, E.; Pedraza-Escalona, M.; Gurrola, G.B.; Olamendi-Portugal, T.; Corzo, G.; Wanke, E.; Possani, L.D. Negative-Shift Activation, Current Reduction and Resurgent Currents Induced by β -Toxins from *Centruroides* Scorpions in Sodium Channels. *Toxicon* **2012**, *59*, 283–293. [\[CrossRef\]](#) [\[PubMed\]](#)
25. Jouirou, B.; Mosbah, A.; Visan, V.; Grissmer, S.; M'Barek, S.; Fajloun, Z.; Van Rietschoten, J.; Devaux, C.; Rochat, H.; Lippens, G.; et al. Cobatoxin 1 from *Centruroides noxius* Scorpion Venom: Chemical Synthesis, Three-Dimensional Structure in Solution, Pharmacology and Docking on K⁺ Channels. *Biochem. J.* **2004**, *377*, 37–49. [\[CrossRef\]](#)
26. Torres, A.M.; Bansal, P.; Alewood, P.F.; Bursill, J.A.; Kuchel, P.W.; Vandenberg, J.I. Solution Structure of CnErg1 (Ergtoxin), a HERG Specific Scorpion Toxin. *FEBS Lett.* **2003**, *539*, 138–142. [\[CrossRef\]](#)
27. Arcangeli, A.; Becchetti, A. New Trends in Cancer Therapy: Targeting Ion Channels and Transporters. *Pharmaceuticals* **2010**, *3*, 1202–1224. [\[CrossRef\]](#)
28. Mendes, L.C.; Viana, G.M.M.; Nencioni, A.L.A.; Pimenta, D.C.; Beraldo-Neto, E. Scorpion Peptides and Ion Channels: An Insightful Review of Mechanisms and Drug Development. *Toxins* **2023**, *15*, 238. [\[CrossRef\]](#)
29. Chen, Z.; Luo, F.; Feng, J.; Yang, W.; Zeng, D.; Zhao, R.; Cao, Z.; Liu, M.; Li, W.; Jiang, L.; et al. Genomic and Structural Characterization of Kunitz-Type Peptide LmKTT-1a Highlights Diversity and Evolution of Scorpion Potassium Channel Toxins. *PLoS ONE* **2013**, *8*, e60201. [\[CrossRef\]](#)

30. Ding, L.; Wang, X.; Liu, H.; San, M.; Xu, Y.; Li, J.; Li, S.; Cao, Z.; Li, W.; Wu, Y.; et al. A New Kunitz-Type Plasmin Inhibitor from Scorpion Venom. *Toxicon* **2015**, *106*, 7–13. [\[CrossRef\]](#)
31. NaderiSoorki, M.; Galehdari, H.; Baradaran, M.; Jalali, A. First Venom Gland Transcriptomic Analysis of Iranian Yellow Scorpion “*Odonthubuthus doriae*” with Some New Findings. *Toxicon* **2016**, *120*, 69–77. [\[CrossRef\]](#) [\[PubMed\]](#)
32. Ruiming, Z.; Yibao, M.; Yawen, H.; Zhiyong, D.; Yingliang, W.; Zhijian, C.; Wenxin, L. Comparative Venom Gland Transcriptome Analysis of the Scorpion *Lychas mucronatus* Reveals Intraspecific Toxic Gene Diversity and New Venomous Components. *BMC Genom.* **2010**, *11*, 452. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Rokyta, D.R.; Ward, M.J. Venom-Gland Transcriptomics and Venom Proteomics of the Black-Back Scorpion (*Hadrurus spadix*) Reveal Detectability Challenges and an Unexplored Realm of Animal Toxin Diversity. *Toxicon* **2017**, *128*, 23–37. [\[CrossRef\]](#)
34. Oliveira, A.L.; Viegas, M.F.; da Silva, S.L.; Soares, A.M.; Ramos, M.J.; Fernandes, P.A. The Chemistry of Snake Venom and Its Medicinal Potential. *Nat. Rev. Chem.* **2022**, *6*, 451–469. [\[CrossRef\]](#) [\[PubMed\]](#)
35. Marchler-Bauer, A.; Bo, Y.; Han, L.; He, J.; Lanczycki, C.J.; Lu, S.; Chitsaz, F.; Derbyshire, M.K.; Geer, R.C.; Gonzales, N.R.; et al. CDD/SPARCLE: Functional Classification of Proteins via Subfamily Domain Architectures. *Nucleic Acids Res.* **2017**, *45*, D200–D203. [\[CrossRef\]](#) [\[PubMed\]](#)
36. Ullah, A.; Masood, R. The Sequence and Three-Dimensional Structure Characterization of Snake Venom Phospholipases B. *Front. Mol. Biosci.* **2020**, *7*, 175. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Kemparaju, K.; Girish, K.S. Snake Venom Hyaluronidase: A Therapeutic Target. *Cell Biochem. Funct.* **2006**, *24*, 7–12. [\[CrossRef\]](#) [\[PubMed\]](#)
38. Hubert, C.; Houot, A.M.; Corvol, P.; Soubrier, F. Structure of the Angiotensin I-Converting Enzyme Gene. Two Alternate Promoters Correspond to Evolutionary Steps of a Duplicated Gene. *J. Biol. Chem.* **1991**, *266*, 15377–15383. [\[CrossRef\]](#)
39. Morgenstern, D.; Rohde, B.H.; King, G.F.; Tal, T.; Sher, D.; Zlotkin, E. The Tale of a Resting Gland: Transcriptome of a Replete Venom Gland from the Scorpion *Hottentotta judaicus*. *Toxicon* **2011**, *57*, 695–703. [\[CrossRef\]](#)
40. de Oliveira, U.C.; Candido, D.M.; Coronado Dorce, V.A.; Junqueira-de-Azevedo, I.d.L.M. The Transcriptome Recipe for the Venom Cocktail of *Tityus bahiensis* Scorpion. *Toxicon* **2015**, *95*, 52–61. [\[CrossRef\]](#)
41. Almeida, D.D.; Scortecci, K.C.; Kobashi, L.S.; Agnez-Lima, L.F.; Medeiros, S.R.B.; Silva-Junior, A.A.; Junqueira-de-Azevedo, I.d.L.M.; Fernandes-Pedrosa, M.d.F. Profiling the Resting Venom Gland of the Scorpion *Tityus stigmurus* through a Transcriptomic Survey. *BMC Genom.* **2012**, *13*, 362. [\[CrossRef\]](#)
42. Alvarenga, E.R.; Mendes, T.M.; Magalhaes, B.F.; Siqueira, F.F.; Dantas, A.E.; Barroca, T.M.; Horta, C.C.; Kalapothakis, E. Transcriptome Analysis of the *Tityus serrulatus* Scorpion Venom Gland. *Open J. Genet.* **2012**, *2*, 210–220. [\[CrossRef\]](#)
43. Goudarzi, M.H.; Eagles, D.A.; Lim, J.; Biggs, K.A.; Kotze, A.C.; Ruffell, A.P.; Fairlie, D.P.; King, G.F.; Walker, A.A. Venom Composition and Bioactive RF-Amide Peptide Toxins of the *Saddleback caterpillar*, *Acharia stimulea* (Lepidoptera: Limacodidae). *Biochem. Pharmacol.* **2023**, *213*, 115598. [\[CrossRef\]](#)
44. Vaiyapuri, S.; Wagstaff, S.C.; Harrison, R.A.; Gibbins, J.M.; Hutchinson, E.G. Evolutionary Analysis of Novel Serine Proteases in the Venom Gland Transcriptome of *Bitis gabonica* Rhinoceros. *PLoS ONE* **2011**, *6*, e21532. [\[CrossRef\]](#)
45. Baradaran, M.; Salabi, F. Identification, Classification and Characterization of the Dermonecrotic Toxins in Venom Glands of *Hottentotta saulcyi*, *Androctonus crassicauda* and *Hemiscorpius lepturus* Using Transcriptome Analysis. *Res. Sq.* 01 February **2023**, preprint, 1–32. [\[CrossRef\]](#)
46. Fox, J.W. A Brief Review of the Scientific History of Several Lesser-Known Snake Venom Proteins: L-Amino Acid Oxidases, Hyaluronidases and Phosphodiesterases. *Toxicon* **2013**, *62*, 75–82. [\[CrossRef\]](#) [\[PubMed\]](#)
47. Pedrosa, A.; Matioli, S.R.; Murakami, M.T.; Pidde-Queiroz, G.; Tambourgi, D.V. Adaptive Evolution in the Toxicity of a Spider’s Venom Enzymes. *BMC Evol. Biol.* **2015**, *15*, 290. [\[CrossRef\]](#)
48. Boto, A.; Pérez de la Lastra, J.; González, C. The Road from Host-Defense Peptides to a New Generation of Antimicrobial Drugs. *Molecules* **2018**, *23*, 311. [\[CrossRef\]](#) [\[PubMed\]](#)
49. Meng, L.; Zhao, Y.; Qu, D.; Xie, Z.; Guo, X.; Zhu, Z.; Chen, Z.; Zhang, L.; Li, W.; Cao, Z.; et al. Ion Channel Modulation by Scorpion Hemolymph and Its Defensin Ingredients Highlights Origin of Neurotoxins in Telson Formed in Paleozoic Scorpions. *Int. J. Biol. Macromol.* **2020**, *148*, 351–363. [\[CrossRef\]](#)
50. Cociancich, S.; Goyffon, M.; Bontems, F.; Bulet, P.; Bouet, F.; Menez, A.; Hoffmann, J. Purification and Characterization of a Scorpion Defensin, a 4 kDa Antibacterial Peptide Presenting Structural Similarities with Insect Defensins and Scorpion Toxins. *Biochem. Biophys. Res. Commun.* **1993**, *194*, 17–22. [\[CrossRef\]](#)
51. Ferreira, L.A.F.; Alves, E.W.; Henriques, O.B. Peptide T, a Novel Bradykinin Potentiator Isolated from *Tityus serrulatus* Scorpion Venom. *Toxicon* **1993**, *31*, 941–947. [\[CrossRef\]](#)
52. Fan, Z.; Cao, L.; He, Y.; Hu, J.; Di, Z.; Wu, Y.; Li, W.; Cao, Z. Ctriporin, a New Anti-Methicillin-Resistant *Staphylococcus aureus* Peptide from the Venom of the Scorpion *Chaerilus tricoatus*. *Antimicrob. Agents Chemother.* **2011**, *55*, 5220–5229. [\[CrossRef\]](#)
53. Hoskin, D.W.; Ramamoorthy, A. Studies on Anticancer Activities of Antimicrobial Peptides. *Biochim. Biophys. Acta-Biomembr.* **2008**, *1778*, 357–375. [\[CrossRef\]](#)
54. Almaaytah, A.; Albalas, Q. Scorpion Venom Peptides with No Disulfide Bridges: A Review. *Peptides* **2014**, *51*, 35–45. [\[CrossRef\]](#) [\[PubMed\]](#)
55. Zeng, X.-C.; Corzo, G.; Hahin, R. Scorpion Venom Peptides without Disulfide Bridges. *IUBMB Life Int. Union Biochem. Mol. Biol. Life* **2005**, *57*, 13–21. [\[CrossRef\]](#)

56. Lu, S.; Wang, J.; Chitsaz, F.; Derbyshire, M.K.; Geer, R.C.; Gonzales, N.R.; Gwadz, M.; Hurwitz, D.I.; Marchler, G.H.; Song, J.S.; et al. CDD/SPARCLE: The Conserved Domain Database in 2020. *Nucleic Acids Res.* **2020**, *48*, D265–D268. [[CrossRef](#)] [[PubMed](#)]
57. Zeng, X.-C.; Nie, Y.; Luo, X.; Wu, S.; Shi, W.; Zhang, L.; Liu, Y.; Cao, H.; Yang, Y.; Zhou, J. Molecular and Bioinformatical Characterization of a Novel Superfamily of Cysteine-Rich Peptides from Arthropods. *Peptides* **2013**, *41*, 45–58. [[CrossRef](#)]
58. Law, R.H.P.; Zhang, Q.; McGowan, S.; Buckle, A.M.; Silverman, G.A.; Wong, W.; Rosado, C.J.; Langendorf, C.G.; Pike, R.N.; Bird, P.I.; et al. An Overview of the Serpin Superfamily. *Genome Biol.* **2006**, *7*, 216. [[CrossRef](#)] [[PubMed](#)]
59. Zhang, L.; Shi, W.; Zeng, X.-C.; Ge, F.; Yang, M.; Nie, Y.; Bao, A.; Wu, S.; E, G. Unique Diversity of the Venom Peptides from the Scorpion *Androctonus bicolor* Revealed by Transcriptomic and Proteomic Analysis. *J. Proteom.* **2015**, *128*, 231–250. [[CrossRef](#)]
60. Reddy, T.; Gibbs, G.M.; Merriner, D.J.; Kerr, J.B.; O'Bryan, M.K. Cysteine-Rich Secretory Proteins Are Not Exclusively Expressed in the Male Reproductive Tract. *Dev. Dyn.* **2008**, *237*, 3313–3323. [[CrossRef](#)]
61. Fry, B.G.; Vidal, N.; Norman, J.A.; Vonk, F.J.; Scheib, H.; Ramjan, S.F.R.; Kuruppu, S.; Fung, K.; Blair Hedges, S.; Richardson, M.K.; et al. Early Evolution of the Venom System in Lizards and Snakes. *Nature* **2006**, *439*, 584–588. [[CrossRef](#)]
62. Milne, T.J.; Abbenante, G.; Tyndall, J.D.A.; Halliday, J.; Lewis, R.J. Isolation and Characterization of a Cone Snail Protease with Homology to CRISP Proteins of the Pathogenesis-Related Protein Superfamily. *J. Biol. Chem.* **2003**, *278*, 31105–31110. [[CrossRef](#)]
63. Zhang, Q.; Xu, J.; Zhou, X.; Liu, Z. CAP Superfamily Proteins from Venomous Animals: Who We Are and What to Do? *Int. J. Biol. Macromol.* **2022**, *221*, 691–702. [[CrossRef](#)]
64. Qin, N.; Sun, H.; Lu, M.; Wang, J.; Tang, T.; Liu, F. A Single von Willebrand Factor C-Domain Protein Acts as an Extracellular Pattern-Recognition Receptor in the River Prawn *Macrobrachium nipponense*. *J. Biol. Chem.* **2020**, *295*, 10468–10477. [[CrossRef](#)] [[PubMed](#)]
65. Dixon, W.J. The Up-and-Down Method for Small Samples. *J. Am. Stat. Assoc.* **1965**, *60*, 967–978. [[CrossRef](#)]
66. Grabherr, M.G.; Haas, B.J.; Yassour, M.; Levin, J.Z.; Thompson, D.A.; Amit, I.; Adiconis, X.; Fan, L.; Raychowdhury, R.; Zeng, Q.; et al. Full-Length Transcriptome Assembly from RNA-Seq Data without a Reference Genome. *Nat. Biotechnol.* **2011**, *29*, 644–652. [[CrossRef](#)]
67. Haas, B.J.; Papanicolaou, A.; Yassour, M.; Grabherr, M.; Blood, P.D.; Bowden, J.; Couger, M.B.; Eccles, D.; Li, B.; Lieber, M.; et al. De Novo Transcript Sequence Reconstruction from RNA-Seq Using the Trinity Platform for Reference Generation and Analysis. *Nat. Protoc.* **2013**, *8*, 1494–1512. [[CrossRef](#)] [[PubMed](#)]
68. Bryant, D.M.; Johnson, K.; DiTommaso, T.; Tickle, T.; Couger, M.B.; Payzin-Dogru, D.; Lee, T.J.; Leigh, N.D.; Kuo, T.-H.; Davis, F.G.; et al. A Tissue-Mapped Axolotl De Novo Transcriptome Enables Identification of Limb Regeneration Factors. *Cell Rep.* **2017**, *18*, 762–776. [[CrossRef](#)]
69. Patro, R.; Duggal, G.; Love, M.I.; Irizarry, R.A.; Kingsford, C. Salmon Provides Fast and Bias-Aware Quantification of Transcript Expression. *Nat. Methods* **2017**, *14*, 417–419. [[CrossRef](#)] [[PubMed](#)]
70. Nesvizhskii, A.I.; Keller, A.; Kolker, E.; Aebersold, R. A Statistical Model for Identifying Proteins by Tandem Mass Spectrometry. *Anal. Chem.* **2003**, *75*, 4646–4658. [[CrossRef](#)] [[PubMed](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.