

Article

The Dual α -Amidation System in Scorpion Venom Glands

Gustavo Delgado-Prudencio, Lourival D. Possani , Baltazar Becerril and Ernesto Ortiz * 

Departamento de Medicina Molecular y Bioprocesos, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Avenida Universidad 2001, Colonia Chamilpa, Cuernavaca, Morelos 62210, Mexico

* Correspondence: erne@ibt.unam.mx

Received: 4 June 2019; Accepted: 18 July 2019; Published: 20 July 2019



Abstract: Many peptides in scorpion venoms are amidated at their C-termini. This post-translational modification is paramount for the correct biological function of ion channel toxins and antimicrobial peptides, among others. The discovery of canonical amidation sequences in transcriptome-derived scorpion proproteins suggests that a conserved enzymatic α -amidation system must be responsible for this modification of scorpion peptides. A transcriptomic approach was employed to identify sequences putatively encoding enzymes of the α -amidation pathway. A dual enzymatic α -amidation system was found, consisting of the membrane-anchored, bifunctional, peptidylglycine α -amidating monooxygenase (PAM) and its paralogs, soluble monofunctional peptidylglycine α -hydroxylating monooxygenase (PHM*m*) and peptidyl- α -hydroxyglycine α -amidating lyase (PAL*m*). Independent genes encode these three enzymes. Amino acid residues responsible for ion coordination and enzymatic activity are conserved in these sequences, suggesting that the enzymes are functional. Potential endoproteolytic recognition sites for proprotein convertases in the PAM sequence indicate that PAM-derived soluble isoforms may also be expressed. Sequences potentially encoding proprotein convertases (PC1 and PC2), carboxypeptidase E (CPE), and other enzymes of the α -amidation pathway, were also found, confirming the presence of this pathway in scorpions.

Keywords: amidation; evolution; posttranslational modifications; scorpion; transcriptomics

Key Contribution: A dual enzymatic system responsible for α -amidation of scorpion venom peptides is described. Independent genes encode a bifunctional PAM enzyme and the monofunctional PHM*m* and PAL*m* enzymes.

1. Introduction

The order Scorpiones constitutes one of the most ancient lineages within the phylum Arthropoda [1,2]. The key to the ecological success of these arachnids resides in the production of potent venoms used for feeding, defense, and deterring competitors [3,4]. Scorpion venoms are complex mixtures of components, including bioactive peptides with potential therapeutic applications [4], enzymes, metabolites, and most importantly, an arsenal of toxins active on Na⁺, K⁺, Ca²⁺, and Cl⁻ channels [5–10]. By altering the normal mechanics of these channels, scorpion toxins unleash systemic havoc in their victims, which can lead to severe envenomation symptoms, including death [11,12]. The venom is produced and secreted by two symmetrical glands located in the last segment of the metasoma, the telson [12]. In these glands, the peptidyl venom components undergo synthesis and maturation, a complex process involving a series of post-translational modifications (PTMs) that result in the biologically active molecules [3]. The most common PTMs found in scorpion venom peptides are the formation of disulfide bridges from pairs of cysteines, proteolytic cleavage, and C-terminal amidation (α -amidation). Amidated toxins and peptides without disulfide bonds (NDBP) are well known in scorpion venoms (Table 1).

Table 1. Diversity of amidated peptides from scorpion venoms. Signal peptides are underlined. Mature peptides are in bold upper-case letters. Propeptides are italicized and amidation signals are shown in red.

Type	Peptide	Uniprot	Precursor Sequence	C-terminus
Sodium toxins	AaH2	P01484	<u>MNYLVMISLALLFVTGVESVKDGYIVDDVNCTYFCGRNAYCNEECTKLKGESGYCQWASPYGNAC</u> YCYKLPD H V R T K G P R C H GR	H-NH ₂
	LqhIT2	Q26292	<u>MKLLLLLIVSASMLIESLVNADGYIKRRDGGCKVACLIGNEGCDKECKAYGGSYGYCWTWGLACWCE</u> GLPDDKTWKSETNTCG GKK	G-NH ₂
	BmKITa	Q9XY87	<u>MKLFLLLLISASMLIDGLVNADGYIRGSNGCKVSLWGNENECNKECRAYGASYGYCWTWGLACWC</u> QGLPDDKTWKSESNTCG GKK	G-NH ₂
	Cn2	P01495	<u>LLIITACLALIGTVWAKEGYLVDKNTGCKYECLKLGDNDYCLRECKQQYGGKAGGYCYAFACWC</u> THLYEQAI V W P L P N K R C S GK	S-NH ₂
	Css4	P60266	<u>MNSLLMITACLALVGTVWAKEGYLVNSYTGCKFEFCKLGDNDYCLRECRQQYGGKSGGYCYAFG</u> CWCTHLYEQAVVW P L P N K T C N GK	N-NH ₂
	CsEI	P01491	<u>MNSLLMITACLVLIGTVWAKDGYLVEKTGCKKTCYKLGENDFCNRECKWKHIGGSYGYCYGFGC</u> YCEGLPDSTQ T W P L P N K T C GKK	C-NH ₂
	CsEv3	P01494	<u>MNSLLMITACLFLIGTVWAKEGYLVNKSTGCKYGCCLKLGENEGCDKECKAKNQGGSYGYCYAFA</u> CWCEGLPESTPT Y P L P N K S GKK	C-NH ₂
	Ts1	P15226	<u>MKGMILFISCLLLIGIVVECKEGYLMDEHGCKLSCFIRPSGYCGREGCIKKGSSGYCAWPACYCY</u> GLPNWVKVWDRATN KCGK	C-NH ₂
	Ts3	P01496	<u>LVVVCLLTAGTEGKKDGYPVEYDNCAYICWNYDNAYCDKLCCKDKKADSGYCYWVHILCYCY</u> GLPDSEPTKTNGK CSGKK	S-NH ₂
Potassium toxins	NTx	P08815	<u>MKAFYGILILLFCSMFNLNESTIINVKCTSPKQCSKPCKELYGSSAGAKCMNGKCKCYNNG</u>	N-NH ₂
	BmKTX	Q9NII7	<u>MKVFFAVLITLFICSMIIGHGVGINVKCHSGQCLKPCDAGMRFGKCINGKCDCTPKG</u>	K-NH ₂
	CoTx1	O46028	<u>MEGIAKITLILLFLVTMHTFANWNTEAAVCVYRTCDKDCKRRGYRSGKCINNACKCYPYGK</u>	Y-NH ₂
	OcKTx5	Q6XLL5	<u>MNAKFILLLVLTTMMLLPDTKGAEVIRCSGSKQCYGPCKQTGCTNSKCMNKVCKCYCG</u>	C-NH ₂
	OcKTx1	Q6XLL9	<u>MNAKFILLLVVATTMLLPDTQGAEVIKRTPKDCAGPCRKQTGCPHGKCMNRTCRCNRC</u>	C-NH ₂
Non disulfide bridged peptides	IsCT	Q8MMJ7	<u>MKTQFAILLVALVLFQMFAQSDAILGKIWEGIKSLFGKRGLSDLDGLDELFDGEISKADRDFLRELMR</u>	F-NH ₂
	BmKb1	Q718F4	<u>MEIKYLLTVFLVLLIVSDHCQAFLFSLIPSAISGLISAFKGRRKRDLNGYIDHFKNFRKRDAELEEELLSKLPY</u>	K-NH ₂
	Hp1090	P0DJ02	<u>MKTQFAIFLITLVLVLFQMFSQSDAIFKAIWSGIKSLFGKRGLSDLDDDESFDGEVSDADIDFLKELMQ</u>	F-NH ₂
	IsCT2	Q8MTX2	<u>MKTQFAILLVALVLFQMFAQSEAIFGAIWNGIKSLFGRRALNNDLDDGLDELFDGEISQADVDFLKELMR</u>	F-NH ₂
	VAMP-2	E4VP07	<u>MKSQTFFLLFLVFLLAITQSEAIFGAIAGLLKNIFGKRSLRDMDTMKYLYDPSLSAADLKTLQKLMENY</u>	F-NH ₂

Post-translational α -amidation is characteristic of bioactive peptides from many eukaryotic taxa [13]. C-terminal amidation confers on peptides enhanced resilience to degradation by carboxypeptidases, thus increasing their half-lives and decreasing their turnover rates [14]. Moreover, amidation is essential for correct functioning of many mammalian neuropeptides and hormones [15]. Several scorpion toxins have also been shown to require C-terminal amidation for full biological activity, without which, potency is severely reduced [16,17]. C-terminal amidation results in a change with two possible functional implications: the amidated terminal amino acid could be directly involved in molecular recognition events, or the amidation could simply reduce the negative charge of the carboxyl moiety and the peptide as a whole [18].

In general, metazoan amidated peptides are translated as larger polypeptidyl precursors, which contain an amidation signal, a glycine typically followed by one or two basic residues (R-X-Gly-Basic or R-X-Gly-Basic-Basic) and the rest of the propeptide sequence. This signal is first targeted by endoproteolytic proprotein-processing enzymes of the protein convertase family (PCs), resulting in peptides terminated with basic residues, which are substrates for carboxypeptidases that remove those residues from the processing intermediate and expose the C-terminal glycine. This glycine is then further subjected to sequential reactions that amidate the peptide [19,20] (Figure 1A). Two enzymatic activities catalyze these reactions. First, a peptidylglycine α -hydroxylating monooxygenase (PHM, EC 1.14.17.3) catalyzes the hydroxylation of the glycine residue, using ascorbate and molecular oxygen as co-substrates. Then, a peptidyl- α -hydroxyglycine α -amidating lyase (PAL, EC 4.3.2.5) cleaves the hydroxyglycine residue, yielding the amidated product and glyoxylate [21,22] (Figure 1B).

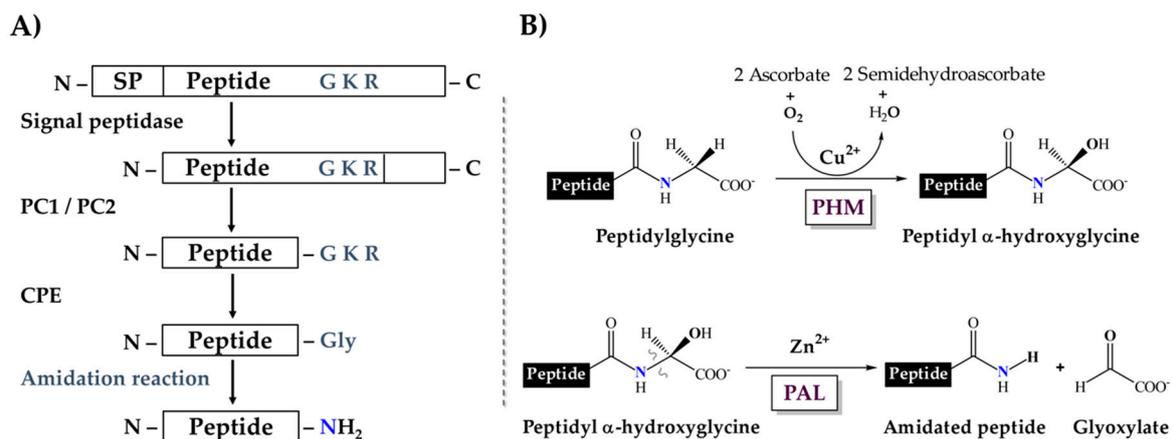


Figure 1. Enzymatic processing of the amidation signal. (A) The complete pathway leading to amidation. SP, Signal Peptide; PC1/PC2, proprotein convertases 1/2; CPE, carboxypeptidase E. (B) Sequential amidation reaction catalyzed by the PHM and PAL domains.

Peptide amidation seems to be common to all metazoans, and PHM and PAL are assumed to have monophyletic origins [23]. However, the way these activities are expressed differs among taxa. For example, in insects *Apis mellifera*, *Drosophila melanogaster* and others, PHM and PAL are encoded by independent genes [24–26]. In vertebrates, such as *Bos taurus*, *Rattus norvegicus*, *Xenopus laevis*, *Homo sapiens* and others, a single gene encodes both activities in a bifunctional enzyme comprising a single polypeptide, peptidylglycine α -amidating monooxygenase (PAM) [27–30]. The same two-domain PAM structure was reported for the gastropod *Aplysia californica* [31]. Curiously, another gastropod, *Lymnaea stagnalis*, produces a zymogen comprising four different PHM domains and a single PAL domain, which is endogenously converted to a mixture of monofunctional isoenzymes [32]. PAM isoforms have been reported in *R. norvegicus*, with up to seven isoforms generated by alternative splicing. These isoforms include configurations with and without internal proteolytic sites, resulting in both independent PHM and PAL, and the bifunctional PAM [27,33]. Among arthropods, independent PHM and PAL, but not the bifunctional PAM, are expressed in insects, as indicated above, whereas both

independent and bifunctional enzymes are expressed in crustaceans [26]. No information is available on the α -amidating system of other subphyla, e.g., the chelicerates, and in particular, the arachnids.

Amidated peptides are common in venoms produced by various animals. For example, marine snails of the genus *Conus*, produce a large array of peptidyl toxins (conotoxins), a significant fraction of which are amidated. From the venom ducts of cone snails, cDNAs were cloned that encode bifunctional PAMs. Heterologously expressed PAMs were demonstrated to be active [34]. Although no amidating system has been described in arachnids, the abundance of amidated peptides in their venoms, in particular scorpion venoms, suggests that amidating enzymes are active in their venom glands. The correlation between canonical α -amidation signals in transcripts from different transcriptomic analyses and amidation of the mature encoded peptides, confirmed by biochemical and proteomic analyses [35–39], indicates that the classical PHM plus PAL and/or PAM systems must be present in scorpion venom glands. In this work, the enzymatic amidation system of Old and New World scorpions is assessed by transcriptomic analysis.

2. Results

2.1. The Dual Enzymatic System for α -Amidation in the Order Scorpiones

We investigated venom gland transcriptomes of 21 scorpion species and the genome of *Centruroides sculpturatus* and identified sequences encoding orthologs of the bifunctional PAM enzyme in 13 of them. Partial sequences for PAM were found in the remaining eight transcriptomes. A 14th complete PAM-coding sequence was recovered by RT-PCR from venom-gland total RNA from the scorpion *Centruroides noxius* (Table 2 and Supplementary Table S2). Complete coding sequences (CDS) from those transcripts translate into proteins of 861–887 amino acids (Supplementary Figure S1). The deduced topology of the scorpion PAM precursor is similar to that of the PAM-2 isoform described for *Rattus norvegicus* (Figure 2A,B). A signal peptide sequence (SP) for secretion is followed by a short propeptide (PP) region, a PHM domain, a linker sequence (Linker 1), a PAL domain, a second linker sequence (Linker 2), a membrane spanning domain (MSD), and a cytosolic domain (CD) (Figure 2A). The rat PAM-2 isoform lacks the Exon A-encoded linker region with respect to the rat PAM-1 isoform. This extra region contains an endoproteolytic site which, after processing, cleaves the PHM and PAL monofunctional enzymes into separate polypeptides. This Exon-A-encoded region has been described only for vertebrates [23], and has no equivalent sequence in the scorpion PAM (Supplementary Figure S2). It is notable that although the scorpion PAM lacks this region, two putative endoproteolytic sites are still present in the scorpion PAM sequence (Figure 2A). The first site, defined by a lysine dyad (KK), is located between the PHM and PAL domains, and is proposed to delimit the PHM domain. The second site, located between the PAL sequence and the MSD, is also defined by a KK dyad, and if subjected to post-translational processing, would liberate a soluble PAL enzyme from the MSD and CD domains. Thus, the scorpion bifunctional PAM enzyme could be post-translationally processed to generate independent, soluble PHM and PAL enzymes.

Shorter transcripts encoding the monofunctional PHM and PAL enzymes (PHM_m and PAL_m) were also identified in most of the analyzed scorpion transcriptomes/genome (Table 2 and Supplementary Table S2). The encoded proteins are 345–350 amino acids long (PHM_m) and 356–366 (PAL_m) (Supplementary Figures S3 and S4). Topologies of the monofunctional enzymes are similar to those of the PHM and PAL-2 isoforms from *D. melanogaster* (Figure 2C). The proproteins include a SP and the catalytic domain. No MSD and CD domains are detected; therefore, the monofunctional enzymes are predicted to be soluble.

Table 2. Enzymes of the α -amidation pathway detected in scorpions.

Family	Species	PAM	PHM	PAL	PC1	PC2	CPE
Buthidae	<i>Centruroides sculpturatus</i>						
	<i>Centruroides hentzi</i>						
	<i>Centruroides noxius</i> ^a						
	<i>Centruroides limpidus</i> ^b						
	<i>Centruroides orizaba</i>						
	<i>Centruroides ochraceus</i>						
	<i>Centruroides hirsutipalpus</i>	✓					
	<i>Tityus trivittatus</i>						
	<i>Leiurus abduhbayrami</i> [*]						
	<i>Mesobuthus martensii</i> [*]						
Vaejovidae	<i>Thorellius cristimanus</i>						
	<i>Paravaejovis schwenkmeyeri</i>	✓					
	<i>Chihuahuanus coahuilae</i>						
	<i>Serradigitus gertschi</i>	✓					
Caraboctonidae	<i>Hoffmannihadrurus aztecus</i>						
	<i>Hadrurus concolorus</i>						
Euscorpiidae	<i>Megacormus gertschi</i>						
Chactidae	<i>Anuroctonus pococki bajae</i>	✓					
Superstitionidae	<i>Superstitionia donensis</i>	✓					
Diplocentridae	<i>Diplocentrus melici</i>						
Urodacidae	<i>Urodacus yaschenkoii</i> [*]						
Scorpionidae	<i>Pandinus imperator</i> [*]						

(- , and): Complete PAM, PHM_m and PAL_m sequences; (-): Partial PAM sequences with 93% or more of the sequence determined; (, ,): Complete PC1, PC2 and CPE sequences; (, ,): PC1, PC2 and CPE sequences with more than 50% of the sequence determined; (✓): Partial sequences with less than 50% of the estimated total sequence determined; ^a PAM sequence amplified by PCR; ^b PAM sequence verified by DNA sequencing; ^{*} Old World scorpion. The tblast and blastn algorithms were used to identify sequences in the local scorpion transcriptomic databases, with an e-value of 1×10^{-6} . Empty spaces indicate that no sequences were identified in those transcriptomes.

Key residues involved in catalysis and metal coordination are conserved in both scorpion amidation systems (Figure 2A and Supplementary Figures S5 and S6), suggesting that those enzymes are probably functional. The percentage of sequence identity between homologous domains of the bifunctional and independent enzymes for each species are indicated in Supplementary Table S3. As an example, for *C. noxius*, the percentage of identity between the PAM subdomains and the PHM_m and PAL_m are 29.8% and 32.5%, respectively.

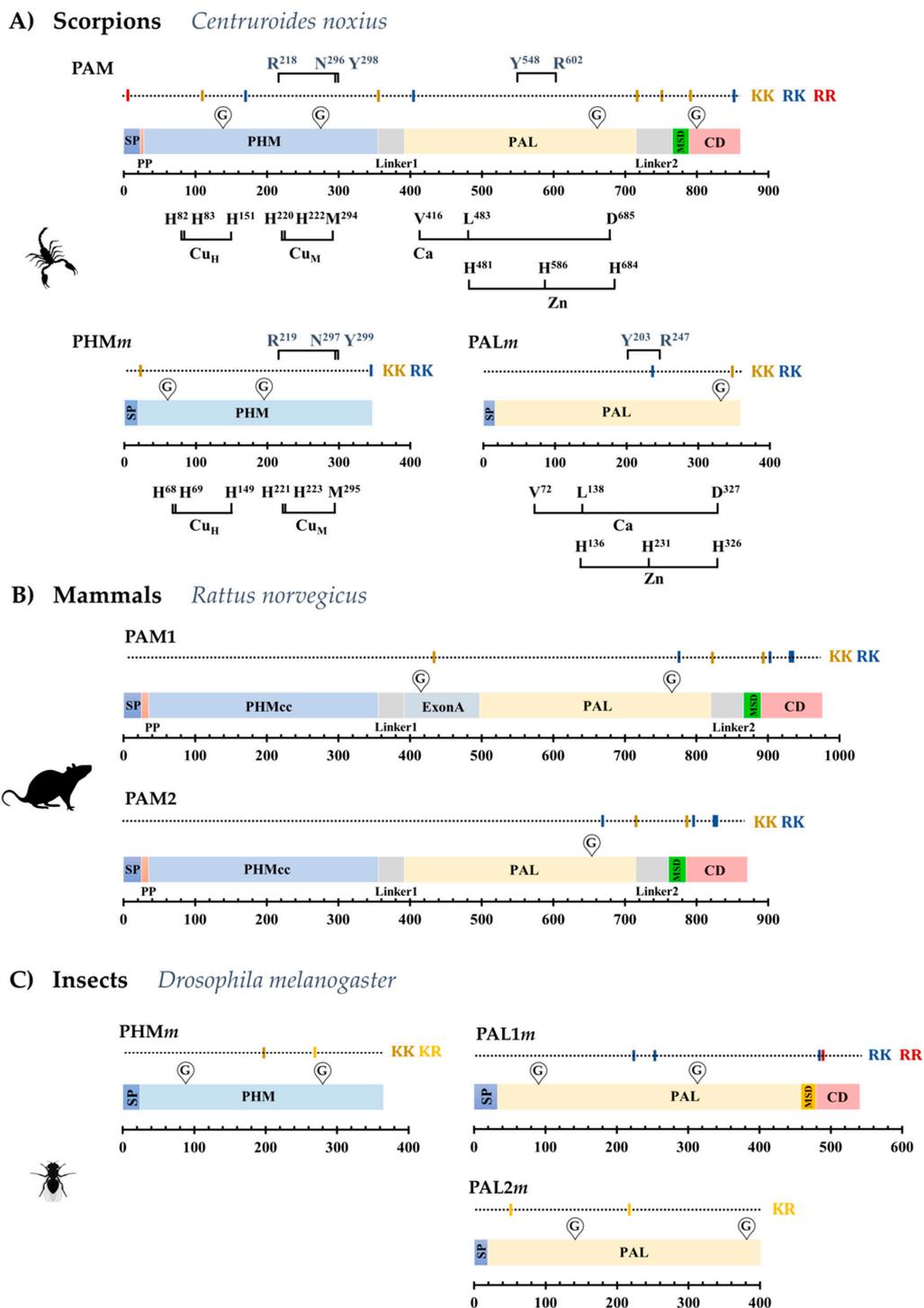


Figure 2. Representative structures of the precursors of (A) scorpion (*C. noxius*) bifunctional PAM and mono-functional PHM m and PAL m ; (B) mammalian (*R. norvegicus*) PAM-1 and PAM-2 isoforms; (C) monofunctional PHM m , PAL1 m and PAL2 m in *D. melanogaster*. Structural and sequence features are indicated as: SP, Signal peptide; PP, Propeptide; MSD, Membrane Spanning Domain; CD, Cytosolic Domain; KK, RK, KR and RR, putative proprotein convertase cleavage sites at basic dyads; G, predicted glycosylation site; PHMcc, catalytic core of the PHM domain.

Sequences encoding other components of the α -amidation pathway were also sought among available scorpion transcriptomic/genomic sequences. Transcripts encoding orthologs of proprotein convertases 1 and 2 (PC1 and PC2) and carboxypeptidase E (CPE), enzymes that operate upstream in the

α -amidation pathway (Figure 1A), were also found, as well as their genes in the *Centruroides sculpturatus* genome (Table 2, Supplementary Table S2), reinforcing the notion of a conserved α -amidation pathway in scorpions.

These results indicate that in scorpions, a dual enzymatic system for α -amidation is responsible for the amidation of venom peptides. Transcripts for both the bifunctional PAM and the monofunctional PHM and PAL are present in scorpion venom glands. Among arthropods, a similar dual system is present in crustaceans, but not in insects [26].

2.2. The PAM-, PHM- and PAL-Coding Genes

The search for genomic sequences in *C. sculpturatus* using blastn showed that separate genes encode the bifunctional and monofunctional enzymes, demonstrating that they are encoded by paralogous genes and are not the result of alternative splicing, a phenomenon reported in the generation of isoforms in *R. norvegicus* [27,33]. Figure 3A shows the structure of the genes for the PAM, PHM and PAL enzymes in *C. sculpturatus*, including their sizes, exon numbers and distributions. The structures of the rat PAM and fruit fly PHM and PAL genes are also shown for comparison (Figure 3B,C).

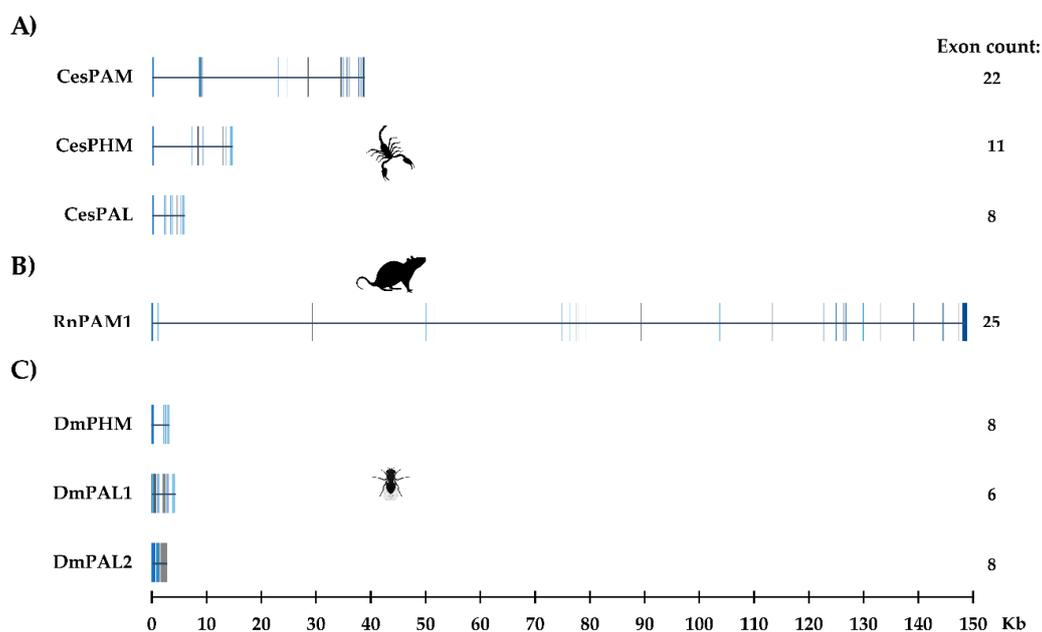


Figure 3. The structure of genes encoding the amidating enzymes of (A) *C. sculpturatus* (CesPAM, CesPHM and CesPAL); (B) *R. norvegicus* (PAM1); (C) *D. melanogaster* (DmPHM, DmPAL1 and DmPAL2). Exons are indicated as vertical blocks.

2.3. Phylogenetic Reconstruction of Amidating Enzymes of Arachnids

Phylogenomic analyses have proposed two basal branches from which all scorpions have descended (parvorders Buthida and Iurida) [1,40,41]. Maximum likelihood analyses with the nucleotide sequences of the PHM and PAL domains from the PAM (designated as *phm*-PAM and *pal*-PAM in these analyses, respectively) and the PHM and PAL enzymes, show a correlation between the phylogeny of these enzymes and the phylogeny of the scorpion families from which they originate [40]. Figure 4; Figure 5 show the two main clades in which the sequences of the *phm*-PAM and *pal*-PAM are separated from the sequences PHM and PAL, respectively. Within those clades, there is a clear divergence between sequences from species belonging to the family Buthidae (parvorder Buthida) and sequences from species belonging to families of the parvorder Iurida (Vaejovidae, Caraboctonidae, Euscorpidae, Chactidae, Superstitionidae, Diplocentridae, Urodacidae, Scorpionidae). Within the family Buthidae, sequences from Old World scorpions *Leiurus abduallahbayrami* (Turkey) and *Mesobuthus martensii* (Eastern Asian countries) are placed in an independent, supported clade that precedes the

clade of New World species. The sequences from *Tityus trivittatus* (from the south-central part of South America, Argentina, and Brazil) are in independent supported clades with respect to those of the genus *Centruroides* (distributed in Central America, the Caribbean, and North America). The variable numbers of sequences recovered from different transcriptomes, limits comparative analyses of the catalytic domains, though a consistent topology for the phylogenetic trees is observed. Sequences putatively encoding a dual amidation system, as in scorpions, were also found in other arachnids, including members of the orders Araneae (*Liphistius malayanus*, *Frontinella communis*, *Parasteatoda tepidariorum*, *Leucauge venusta*), Opiliones (*Siro boyerae*, *Trogulus martensi*) Ricinulei (*Ricinoides atewa*), and the xiphosuran, *Limulus polyphemus* (recently placed within the class Arachnida [42]), among others (Supplementary Table S1). This indicates that the same dual α -amidation system is also employed by other arachnids.

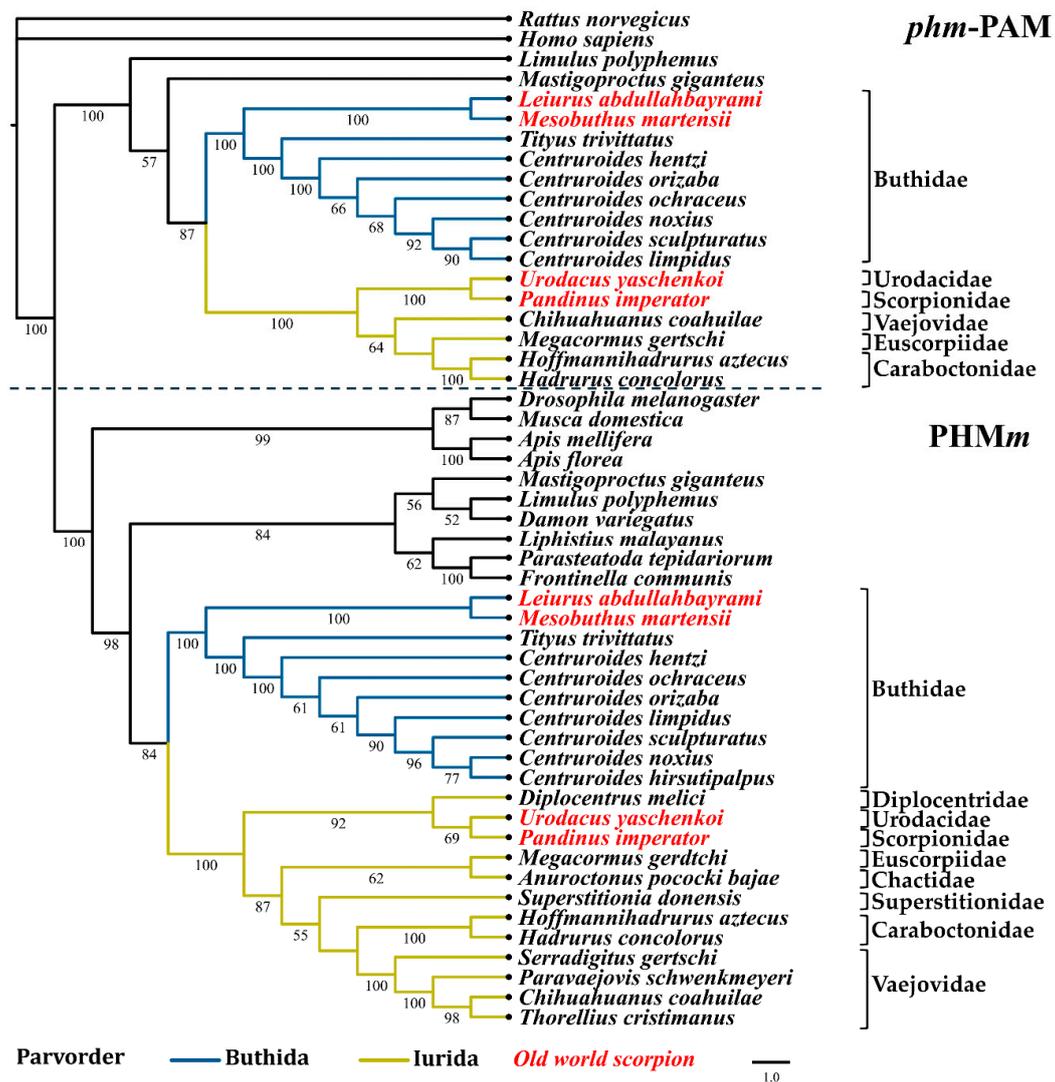


Figure 4. Reconstructed evolutionary history of the phm-PAM and PHMm domains. Maximum likelihood analyses were performed with nucleotide sequences corresponding to the respective domains. Numbers under the nodes indicate the values of ultrafast bootstrap (UFBoot) (only branches with values higher than 50 are shown).

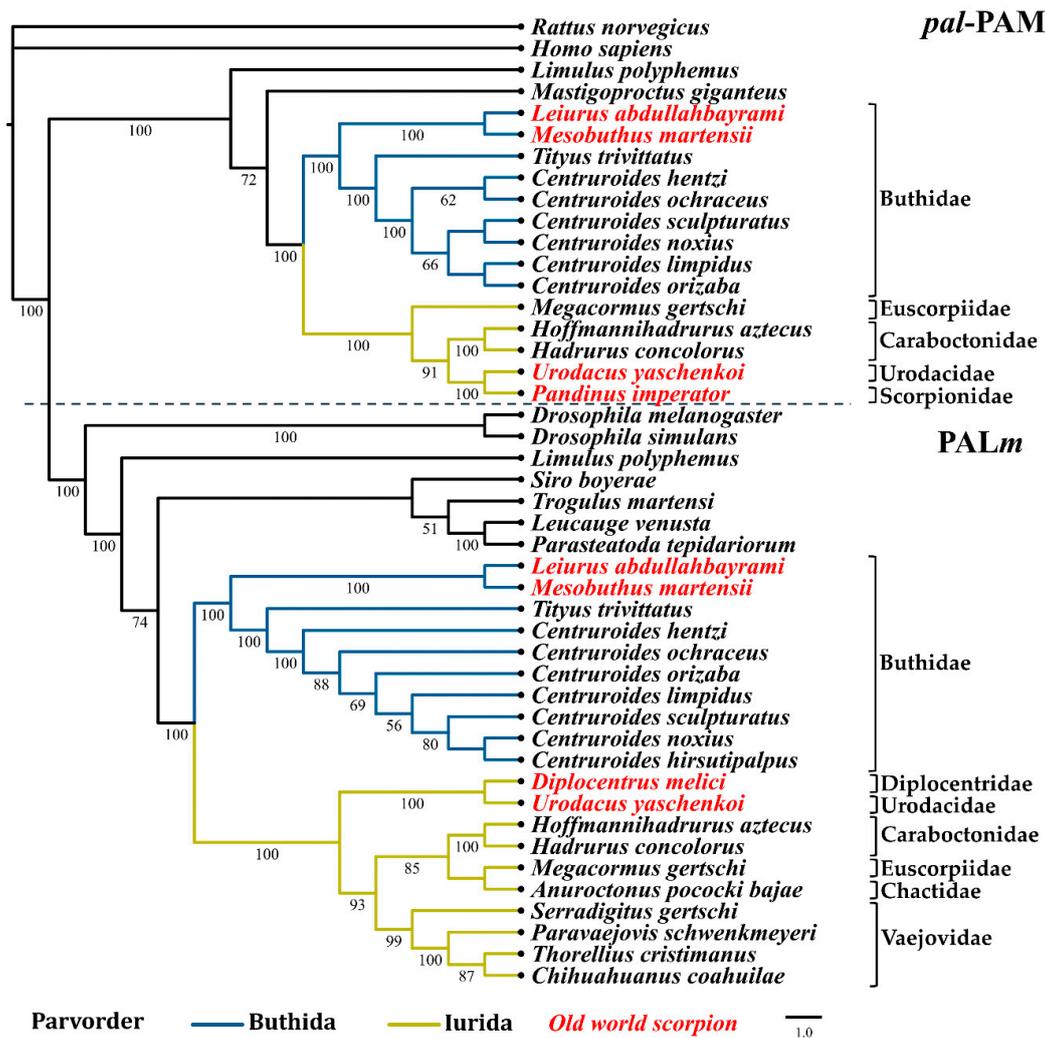


Figure 5. Reconstructed evolutionary history of the pal-PAM and PAL_m domains. Maximum likelihood analyses were performed with the nucleotide sequences corresponding to the respective domains. Number under the nodes indicate the values of ultrafast bootstrap (UFBoot) (only branches with values higher than 50 are shown).

3. Discussion

Venom gland transcriptomic analyses performed with representative scorpion families from both the Old and New Worlds have shown the enormous diversity of compounds that comprise these important biofluids [4]. Together with available biochemical information on scorpion venom components, sequences of many transcripts indicate that amidation is one of the most common PTMs of scorpion venom peptides. The discovery of canonical amidation signals in the translated sequences suggested that a conserved α -amidation system might be present in scorpion venom glands to convert propeptides into shorter, amidated, mature peptides. In this work, transcripts encoding the relevant components of this pathway are described, confirming that a dual amidation system, including a bifunctional PAM enzyme and individual non-membrane bound PHM_m and PAL_m is employed. Genes for this dual system were found in the genome of *C. sculpturatus*, demonstrating that the bifunctional and the monofunctional enzymes are encoded by independent genes and are not the result of alternative splicing. Paralogs involved in various developmental processes and cellular functions within the orders Scorpiones and Araneae arose as a consequence of a genome duplication in the common ancestor of scorpions and spiders [43,44]. Given the importance of amidation in peptide signaling and the functionality of toxins and other amidated venom peptides, it is not surprising that both amidation

enzyme systems were retained in this lineage of venomous arachnids, where they evolved to target specific substrates, or to be expressed in particular cell types or physiological conditions.

Together with conserved functional residues for cation coordination and enzymatic activity, the scorpion PAM sequence contains all the structural elements for generation of a membrane-anchored protein. However, the sequence of the bifunctional PAM contains putative endoprotease cleavage sites (dyads of basic amino acids), which are normally targeted by proprotein convertases, flanking the catalytic domains. This means that the PAM proprotein could in principle be processed to the complete membrane-bound two-domain enzyme or it could be post-translationally cleaved by convertases to render soluble monofunctional domains. The presence of transcripts encoding convertases in the scorpion venom glands, also described in this work, reinforces this possibility. Whether both the two-domain PAM and the PAM-derived monofunctional enzymes coexist in the venom gland remains to be established. We expect that the soluble PHM m and PAL m , as well as the putative PAM-derived soluble isoforms, are secreted by the venom glands into the venom. This has been confirmed, at least for PHM m with liquid chromatography-mass spectrometry (LC-MS/MS) in scorpion venom proteomic analyses. Although it is not clear what additional functions they might have in scorpion venom, it is known that the bovine PAM enzyme is capable of catalyzing three alternative reactions: sulfoxidation, N-dealkylation of amines and O-dealkylation [45]. This raises the possibility of finding new natural substrates for this set of enzymes and taking advantage of their catalytic capacities for synthesis or chemical modification of molecules of biotechnological interest.

Other proteomic analyses have confirmed the presence of putative amidating enzymes in arachnid venoms. One of these sequences was reported as a PAM from the spider *Cupiennius salei* (annotated as PAM_CUPSA [MH766628]) [46]. However, a rigorous sequence analysis demonstrates that this sequence is not from a PAM ortholog, but a monofunctional PHM m . Similarly, for the scorpion *Tityus obscurus*, a sequence reported as a PAM (GenBank: JAT91064) [38], shares 87% sequence identity with the PHM m from *T. trivittatus*, as reported here, and is therefore also a PHM m . A third report found a PHM m sequence in transcriptomic and proteomic analyses of the scorpion *C. hentzi* (annotated as GFWZ01000197.1 TSA: *Centruroides hentzi* Chent_MonoO transcribed RNA sequence) [47]. Sequences encoding orthologs of PHM m were also identified in venoms of *Centruroides limpidus*, *Centruroides hirsutipalpus* and *Superstitionia donensis* (data not shown). Therefore, this constitutes the first report of the monofunctional PAL m and the bifunctional PAM enzymes from any arachnid, and demonstrates that a conserved, functional dual α -amidation system is present in scorpion venom glands, as well as in other arachnids.

4. Materials and Methods

4.1. Sequence Data and Transcriptome Assembly

Previously reported transcriptomic analyses from venom glands of the scorpion species *C. limpidus*, *Paravaejovis schwenkmeyeri*, *Urodacus yaschenkoi*, *Thorellius cristimanus* (reported as *T. atrox*), *Serradigitus gertschi*, *S. donensis*, and *Megacormus gertschi* [37,48–53] were used to obtain relevant sequence information. Complementary sequence information was obtained from other unpublished transcriptomes for the species *Centruroides noxius*, *C. orizaba*, *C. ochraceus*, *C. hirsutipalpus*, *T. trivittatus*, *L. abduallahbayrami*, *Hoffmannihadrurus aztecus*, *Hadrurus concolorus*, *Anuroctonus pococki bajae*, *Chihuahuanus coahuilae* and *Diplocentrus melici*. Publicly available reads from massive transcriptome analyses of other species were assembled de novo and also used (*M. martensii* SRR3061379, *Pandinus imperator* SRR1721600, *C. hentzi* SRR6041834/SRR6041835; external groups are shown in Supplementary Table S1). Assembly was performed using Trinity 2.0.3 [54] with previously reported parameters [37]. Genomic sequences from *C. sculpturatus* (BioProject: PRJNA168116) were obtained from NCBI. Sequence information from 22 different scorpion species, belonging to nine of the 20 recognized scorpion families [55] was used in this work.

4.2. Identification and Annotation of Amidating Enzymes in Scorpions and Related Organisms

Sequences putatively encoding PAM, PHM and PAL homologs were identified in transcriptomes using tBLASTn, with the sequence of the *R. norvegicus* PAM (Uniprot, P14925) as query. Recovered nucleotide sequences were translated with the ExPASy server [56]. The presence and organization of characteristic domains was evaluated with NCBI-CDART [57] in accordance with [23]. Other sequence hallmarks were identified: the signal peptide (SP) with SignalP 4.1 and Phobius [58,59], the propeptide region (Pp) with ArachnoServer v. 3.0 [60] and the transmembrane domain with the TMHMM server v. 2.0 [61]. Identification and delimitation of the catalytic domains and the residues involved in metal coordination and disulfide formation was manually performed by sequence alignment with the reference *R. norvegicus* PAM (Uniport P14925). Potential glycosylation sites were predicted with the NetNGlyc 1.0 Server (<http://www.cbs.dtu.dk/services/NetNGlyc/>). The annotation of each determined sequence can be found in Supplementary Table S1. The sequences were submitted to the European Nucleotide Archive (ENA) under project PRJEB32831.

4.3. Amplification and Cloning of the PAM Sequence from *Centruroides noxius*

Total RNA was extracted from the telson of a single female *C. noxius* using an SV Total RNA Isolation System kit (Promega Corporation, Madison, WI, USA). cDNA was amplified with a First Strand cDNA Synthesis Kit for RT-PCR (AMV) (Roche, Basel, Switzerland). Primers, Cen-Fw3 (5'-GAT CTT GTA AAC GGC GTA TTT CCC TT-3') and Cen-Rv4 (5'-CCG ATA TCC TCC CAA CCA TCC TTT C-3'), were designed from the consensus of the PAM sequences from two scorpions of the genus *Centruroides* (*C. limpidus* and *C. orizaba*). Amplification conditions were 3 min at 96 °C, followed by 30 cycles of 3 sec at 96 °C, 1 min at 56 °C and 2 min at 68 °C, plus a final step of 5 min at 68 °C. A recombinant Pfu polymerase produced in-house was used. The PCR product was purified with the QIAQuick Gel extraction Kit (QIAGEN GmbH, Hilden, Germany), ligated into an EcoRV-digested pBluescript II KS(+) vector, and electroporated to electrocompetent DH5 α *Escherichia coli* cells. Positive clones were selected with the blue/white system by growing the cells in X-Gal/IPTG-complemented LB/ampicillin medium. Plasmids were prepared by alkaline lysis and submitted to sequencing with the primers T7-Like (5'-GCG TAA TAC GAC TCA CTA TA-3'), T3-Like (5'-CTC ACT AAA GGG AAC AAA AGC-3'), Cen-In1 (5'-CTC GTT GCT TAG ATA TAG AGA-3'), Cen-In2 (5'-ACA TCA GTC AAC CAA ACA-3') and Cnox-In3 (5'-ATT GAT GCT GAT GAT GCC TA-3').

4.4. Multiple Alignments and Phylogeny Reconstruction of PAM, PHM, and PAL

Phylogenetic reconstruction of the PAM enzyme and its two catalytic domains *phm*-PAM and *pal*-PAM (with the suffix '-PAM' used to differentiate them from those of the monofunctional enzymes), and of the independent enzymes PHM m and PAL m (with the suffix ' m ', for 'monofunctional') was performed using the maximum likelihood (ML) method with nucleotide sequences. Additional sequences from phylogenetically related organisms (external groups) were obtained from NCBI or assembled from transcriptome raw reads deposited at SRA-NCBI. All sequences were aligned with MAFFT v7.407 [62]. The best substitution model (GTR+F+I+G4) and the ML analysis were evaluated with IQ-TREE v1.6.9 [63,64], using the ultrafast bootstrap method (UFBoot2) [65] with 10,000 replicates.

4.5. Genomic Organization of Scorpion PAM, PHM, and PAL

Genome sequences of *C. sculpturatus* (NCBI:txid218467) corresponding to the amidating enzymes were recovered from NCBI using BLASTn, with the nucleotide sequences for PAM, PHM, and PAL from *C. limpidus* as queries. Identification of introns and exons was performed with the Splign utility [66].

Supplementary Materials: The following are available online at <http://www.mdpi.com/2072-6651/11/7/425/s1>: Figure S1: Schematic alignment of PAM sequences with >90% of the estimated sequence determined, Figure S2: Schematic alignment of PAM1 and PAM2 isoforms from *R. norvegicus* and the completely sequenced scorpion PAM, Figure S3: Schematic alignment of the PHM sequences found in 22 analyzed scorpion transcriptomes, Figure S4: Schematic alignment of the 20 PAL sequences found in 22 analyzed scorpion transcriptomes, Figure S5: Sequence alignment of PHM domains, Figure S6: Sequence alignment of PAL domains, Table S1: Nomenclature of transcripts in various scorpion species, Table S2: Sequence conservation between catalytic domains of the bifunctional and monofunctional enzymes by species (% of identity), Table S3: External groups used for phylogenetic reconstruction of the evolutionary history of the functional domains.

Author Contributions: Conceptualization: E.O.; methodology: G.D.-P. Validation: E.O., L.D.P., B.B. Formal analysis: G.D.-P. Investigation: G.D.-P. Resources: L.D.P. Data curation: G.D.-P. Writing—original draft preparation: G.D.-P., E.O. Writing—review and editing: L.D.P., B.B. Visualization: G.D.-P., E.O. Supervision: E.O. Project administration: E.O. Funding acquisition: L.D.P.

Funding: This research was funded by grant IN202619 from Dirección General del Personal Académico, UNAM, awarded to L.D.P., G.D.-P. is the recipient of a PhD fellowship from CONACyT (540036).

Acknowledgments: We are indebted to Jorge L. Folch-Mallol and Marcela Ayala-Aceves for insightful discussions. We appreciate the support given by María T. Romero-Gutiérrez with transcriptomic analyses, and by Erika P. Meneses with proteomic analyses. The help of Edmundo González-Santillán with phylogenetic analysis is greatly appreciated. We thank Jérôme Verleyen for the bioinformatics tools in the cluster of the Instituto de Biotecnología-UNAM. The technical assistance of Arturo Ocadiz-Ramírez, David S. Castañeda-Carreón, Roberto P. Rodríguez-Bahena, Jesús O. Arriaga-Pérez, Juan M. Hurtado-Ramírez, and Servando Aguirre-Cruz is also greatly acknowledged.

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. Sharma, P.P.; Fernandez, R.; Esposito, L.A.; Gonzalez-Santillan, E.; Monod, L. Phylogenomic resolution of scorpions reveals multilevel discordance with morphological phylogenetic signal. *Proc. Biol. Sci.* **2015**, *282*, 20142953. [[CrossRef](#)] [[PubMed](#)]
2. Waddington, J.; Rudkin, D.M.; Dunlop, J.A. A new mid-Silurian aquatic scorpion—one step closer to land? *Biol. Lett.* **2015**, *11*, 20140815. [[CrossRef](#)] [[PubMed](#)]
3. Oldrati, V.; Arrell, M.; Violette, A.; Perret, F.; Sprungli, X.; Wolfender, J.L.; Stocklin, R. Advances in venomics. *Mol. Biosyst.* **2016**, *12*, 3530–3543. [[CrossRef](#)] [[PubMed](#)]
4. Ortiz, E.; Gurrola, G.B.; Schwartz, E.F.; Possani, L.D. Scorpion venom components as potential candidates for drug development. *Toxicon* **2015**, *93*, 125–135. [[CrossRef](#)] [[PubMed](#)]
5. Xiao, L.; Gurrola, G.B.; Zhang, J.; Valdivia, C.R.; SanMartin, M.; Zamudio, F.Z.; Zhang, L.; Possani, L.D.; Valdivia, H.H. Structure-function relationships of peptides forming the calcin family of ryanodine receptor ligands. *J. Gen. Physiol.* **2016**, *147*, 375–394. [[CrossRef](#)]
6. Fuller, M.D.; Thompson, C.H.; Zhang, Z.R.; Freeman, C.S.; Schay, E.; Szakacs, G.; Bakos, E.; Sarkadi, B.; McMaster, D.; French, R.J.; et al. State-dependent inhibition of cystic fibrosis transmembrane conductance regulator chloride channels by a novel peptide toxin. *J. Biol. Chem.* **2007**, *282*, 37545–37555. [[CrossRef](#)] [[PubMed](#)]
7. Gurevitz, M. Mapping of scorpion toxin receptor sites at voltage-gated sodium channels. *Toxicon* **2012**, *60*, 502–511. [[CrossRef](#)] [[PubMed](#)]
8. Possani, L.D.; Becerril, B.; Delepierre, M.; Tytgat, J. Scorpion toxins specific for Na⁺-channels. *Eur. J. Biochem.* **1999**, *264*, 287–300. [[CrossRef](#)]
9. Rodríguez de la Vega, R.C.; Possani, L.D. Overview of scorpion toxins specific for Na⁺ channels and related peptides: Biodiversity, structure-function relationships and evolution. *Toxicon* **2005**, *46*, 831–844. [[CrossRef](#)]
10. Rodríguez de la Vega, R.C.; Possani, L.D. Current views on scorpion toxins specific for K⁺-channels. *Toxicon* **2004**, *43*, 865–875. [[CrossRef](#)]
11. Chippaux, J.P. Emerging options for the management of scorpion stings. *Drug Des. Dev. Ther.* **2012**, *6*, 165–173. [[CrossRef](#)] [[PubMed](#)]
12. Chippaux, J.P.; Goyffon, M. Epidemiology of scorpionism: A global appraisal. *Acta Trop.* **2008**, *107*, 71–79. [[CrossRef](#)] [[PubMed](#)]

13. Jekely, G. Global view of the evolution and diversity of metazoan neuropeptide signaling. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 8702–8707. [[CrossRef](#)] [[PubMed](#)]
14. Gutte, B. *Peptides: Synthesis, Structures, and Applications*, 1st ed.; Academic Press: San Diego, CA, USA, 1995; pp. 288–289.
15. Merkler, D.J. C-terminal amidated peptides: Production by the in vitro enzymatic amidation of glycine-extended peptides and the importance of the amide to bioactivity. *Enz. Microb. Technol.* **1994**, *16*, 450–456. [[CrossRef](#)]
16. Coelho, V.A.; Cremonez, C.M.; Anjolette, F.A.; Aguiar, J.F.; Varanda, W.A.; Arantes, E.C. Functional and structural study comparing the C-terminal amidated beta-neurotoxin Ts1 with its isoform Ts1-G isolated from *Tityus serrulatus* venom. *Toxicon* **2014**, *83*, 15–21. [[CrossRef](#)] [[PubMed](#)]
17. Liu, Z.; Yang, G.; Li, B.; Chi, C.; Wu, X. Cloning, co-expression with an amidating enzyme, and activity of the scorpion toxin BmK ITa1 cDNA in insect cells. *Mol. Biotechnol.* **2003**, *24*, 21–26. [[CrossRef](#)]
18. Estrada, G.; Restano-Cassulini, R.; Ortiz, E.; Possani, L.D.; Corzo, G. Addition of positive charges at the C-terminal peptide region of CsxII, a mammalian scorpion peptide toxin, improves its affinity for sodium channels Nav1.6. *Peptides* **2011**, *32*, 75–79. [[CrossRef](#)] [[PubMed](#)]
19. Eipper, B.A.; Stoffers, D.A.; Mains, R.E. The biosynthesis of neuropeptides: Peptide alpha-amidation. *Annu. Rev. Neurosci.* **1992**, *15*, 57–85. [[CrossRef](#)] [[PubMed](#)]
20. Fricker, L.D. Neuropeptide-processing enzymes: Applications for drug discovery. *AAPS J.* **2005**, *7*, E449–E455. [[CrossRef](#)] [[PubMed](#)]
21. Eipper, B.A.; Perkins, S.N.; Husten, E.J.; Johnson, R.C.; Keutmann, H.T.; Mains, R.E. Peptidyl-alpha-hydroxyglycine alpha-amidating lyase. Purification, characterization, and expression. *J. Biol. Chem.* **1991**, *266*, 7827–7833.
22. Perkins, S.N.; Husten, E.J.; Eipper, B.A. The 108-kDa peptidylglycine alpha-amidating monooxygenase precursor contains two separable enzymatic activities involved in peptide amidation. *Biochem. Biophys. Res. Commun.* **1990**, *171*, 926–932. [[CrossRef](#)]
23. Attenborough, R.M.; Hayward, D.C.; Kitahara, M.V.; Miller, D.J.; Ball, E.E. A “neural” enzyme in nonbilaterian animals and algae: Preneuronal origins for peptidylglycine alpha-amidating monooxygenase. *Mol. Biol. Evol.* **2012**, *29*, 3095–3109. [[CrossRef](#)] [[PubMed](#)]
24. Han, M.; Park, D.; Vanderzalm, P.J.; Mains, R.E.; Eipper, B.A.; Taghert, P.H. *Drosophila* uses two distinct neuropeptide amidating enzymes, dPAL1 and dPAL2. *J. Neurochem.* **2004**, *90*, 129–141. [[CrossRef](#)] [[PubMed](#)]
25. Kolhekar, A.S.; Roberts, M.S.; Jiang, N.; Johnson, R.C.; Mains, R.E.; Eipper, B.A.; Taghert, P.H. Neuropeptide amidation in *Drosophila*: Separate genes encode the two enzymes catalyzing amidation. *J. Neurosci.* **1997**, *17*, 1363–1376. [[CrossRef](#)] [[PubMed](#)]
26. Zabriskie, T.M.; Klinge, M.; Szymanski, C.M.; Cheng, H.; Vederas, J.C. Peptide amidation in an invertebrate: Purification, characterization, and inhibition of peptidylglycine alpha-hydroxylating monooxygenase from the heads of honeybees (*Apis mellifera*). *Arch. Insect Biochem. Physiol.* **1994**, *26*, 27–48. [[CrossRef](#)] [[PubMed](#)]
27. Eipper, B.A.; Green, C.B.; Campbell, T.A.; Stoffers, D.A.; Keutmann, H.T.; Mains, R.E.; Ouafik, L. Alternative splicing and endoproteolytic processing generate tissue-specific forms of pituitary peptidylglycine alpha-amidating monooxygenase (PAM). *J. Biol. Chem.* **1992**, *267*, 4008–4015. [[PubMed](#)]
28. Eipper, B.A.; Park, L.P.; Dickerson, I.M.; Keutmann, H.T.; Thiele, E.A.; Rodriguez, H.; Schofield, P.R.; Mains, R.E. Structure of the precursor to an enzyme mediating COOH-terminal amidation in peptide biosynthesis. *Mol. Endocrinol.* **1987**, *1*, 777–790. [[CrossRef](#)]
29. Glauder, J.; Ragg, H.; Rauch, J.; Engels, J.W. Human peptidylglycine alpha-amidating monooxygenase: cDNA, cloning and functional expression of a truncated form in COS cells. *Biochem. Biophys. Res. Commun.* **1990**, *169*, 551–558. [[CrossRef](#)]
30. Mizuno, K.; Ohsuye, K.; Wada, Y.; Fuchimura, K.; Tanaka, S.; Matsuo, H. Cloning and sequence of cDNA encoding a peptide C-terminal alpha-amidating enzyme from *Xenopus laevis*. *Biochem. Biophys. Res. Commun.* **1987**, *148*, 546–552. [[CrossRef](#)]
31. Fan, X.; Spijker, S.; Akalal, D.B.; Nagle, G.T. Neuropeptide amidation: Cloning of a bifunctional alpha-amidating enzyme from *Aplysia*. *Brain Res. Mol. Brain Res.* **2000**, *82*, 25–34. [[CrossRef](#)]
32. Spijker, S.; Smit, A.B.; Eipper, B.A.; Malik, A.; Mains, R.E.; Geraerts, W.P. A molluscan peptide alpha-amidating enzyme precursor that generates five distinct enzymes. *FASEB J.* **1999**, *13*, 735–748. [[CrossRef](#)]

33. Eipper, B.A.; Milgram, S.L.; Husten, E.J.; Yun, H.Y.; Mains, R.E. Peptidylglycine alpha-amidating monooxygenase: A multifunctional protein with catalytic, processing, and routing domains. *Protein Sci.* **1993**, *2*, 489–497. [[CrossRef](#)]
34. Ul-Hasan, S.; Burgess, D.M.; Gajewiak, J.; Li, Q.; Hu, H.; Yandell, M.; Olivera, B.M.; Bandyopadhyay, P.K. Characterization of the peptidylglycine alpha-amidating monooxygenase (PAM) from the venom ducts of neogastropods, *Conus bullatus* and *Conus geographus*. *Toxicon* **2013**, *74*, 215–224. [[CrossRef](#)]
35. Becerril, B.; Corona, M.; Garcia, C.; Bolivar, F.; Possani, L.D. Cloning of Genes Encoding Scorpion Toxins—An Interpretative Review. *J. Tox. Tox. Rev.* **1995**, *14*, 339–357. [[CrossRef](#)]
36. Bougis, P.E.; Rochat, H.; Smith, L.A. Precursors of *Androctonus australis* scorpion neurotoxins. Structures of precursors, processing outcomes, and expression of a functional recombinant toxin II. *J. Biol. Chem.* **1989**, *264*, 19259–19265.
37. Romero-Gutierrez, T.; Peguero-Sanchez, E.; Cevallos, M.A.; Batista, C.V.F.; Ortiz, E.; Possani, L.D. A Deeper Examination of *Thorellius atrox* Scorpion Venom Components with Omic Technologies. *Toxins* **2017**, *9*, 3399. [[CrossRef](#)]
38. De Oliveira, U.C.; Nishiyama, M.Y., Jr.; Dos Santos, M.B.V.; Santos-da-Silva, A.P.; Chalkidis, H.M.; Souza-Imberg, A.; Candido, D.M.; Yamanouye, N.; Dorce, V.A.C.; Junqueira-de-Azevedo, I.L.M. Proteomic endorsed transcriptomic profiles of venom glands from *Tityus obscurus* and *T. serrulatus* scorpions. *PLoS ONE* **2018**, *13*, e0193739. [[CrossRef](#)]
39. Abdel-Rahman, M.A.; Quintero-Hernandez, V.; Possani, L.D. Venom proteomic and venomous glands transcriptomic analysis of the Egyptian scorpion *Scorpio maurus palmatus* (Arachnida: Scorpionidae). *Toxicon* **2013**, *74*, 193–207. [[CrossRef](#)]
40. Santibanez-Lopez, C.E.; Gonzalez-Santillan, E.; Monod, L.; Sharma, P.P. Phylogenomics facilitates stable scorpion systematics: Reassessing the relationships of Vaejovidae and a new higher-level classification of Scorpiones (Arachnida). *Mol. Phylogenet. Evol.* **2019**, *135*, 22–30. [[CrossRef](#)]
41. Sharma, P.P.; Baker, C.M.; Cosgrove, J.G.; Johnson, J.E.; Oberski, J.T.; Raven, R.J.; Harvey, M.S.; Boyer, S.L.; Giribet, G. A revised dated phylogeny of scorpions: Phylogenomic support for ancient divergence of the temperate Gondwanan family Bothriuridae. *Mol. Phylogenet. Evol.* **2018**, *122*, 37–45. [[CrossRef](#)]
42. Ballesteros, J.A.; Sharma, P.P. A Critical Appraisal of the Placement of Xiphosura (Chelicerata) with Account of Known Sources of Phylogenetic Error. *Syst. Biol.* **2019**, *10*. [[CrossRef](#)]
43. Leite, D.J.; Baudouin-Gonzalez, L.; Iwasaki-Yokozawa, S.; Lozano-Fernandez, J.; Turetzek, N.; Akiyama-Oda, Y.; Prpic, N.M.; Pisani, D.; Oda, H.; Sharma, P.P.; et al. Homeobox gene duplication and divergence in arachnids. *Mol. Biol. Evol.* **2018**, *35*. [[CrossRef](#)]
44. Schwager, E.E.; Sharma, P.P.; Clarke, T.; Leite, D.J.; Wierschin, T.; Pechmann, M.; Akiyama-Oda, Y.; Esposito, L.; Bechsgaard, J.; Bilde, T.; et al. The house spider genome reveals an ancient whole-genome duplication during arachnid evolution. *BMC Biol.* **2017**, *15*, 62. [[CrossRef](#)]
45. Katopodis, A.G.; May, S.W. Novel substrates and inhibitors of peptidylglycine alpha-amidating monooxygenase. *Biochemistry* **1990**, *29*, 4541–4548. [[CrossRef](#)]
46. Kuhn-Nentwig, L.; Langenegger, N.; Heller, M.; Koua, D.; Nentwig, W. The Dual Prey-Inactivation Strategy of Spiders-In-Depth Venom Analysis of *Cupiennius salei*. *Toxins* **2019**, *11*, 167. [[CrossRef](#)]
47. Ward, M.J.; Ellsworth, S.A.; Rokyta, 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](#)]
48. Cid-Uribe, 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](#)]
49. Cid-Uribe, J.I.; Santibanez-Lopez, C.E.; Meneses, E.P.; Batista, C.V.F.; Jimenez-Vargas, J.M.; Ortiz, E.; Possani, L.D. The diversity of venom components of the scorpion species *Paravaejovis schwenkmeyeri* (Scorpiones: Vaejovidae) revealed by transcriptome and proteome analyses. *Toxicon* **2018**, *151*, 47–62. [[CrossRef](#)]
50. Luna-Ramirez, K.; Quintero-Hernandez, V.; Juarez-Gonzalez, V.R.; Possani, L.D. Whole Transcriptome of the Venom Gland from *Urodoacus yaschenkoi* Scorpion. *PLoS ONE* **2015**, *10*, e0127883. [[CrossRef](#)]

51. Romero-Gutierrez, M.T.; Santibanez-Lopez, C.E.; Jimenez-Vargas, J.M.; Batista, C.V.F.; Ortiz, E.; Possani, L.D. Transcriptomic and Proteomic Analyses Reveal the Diversity of Venom Components from the Vaejovid Scorpion *Serradigitus gertschi*. *Toxins* **2018**, *10*, 359. [[CrossRef](#)]
52. Santibanez-Lopez, C.E.; Cid-Urbe, J.I.; Batista, C.V.; Ortiz, E.; Possani, L.D. Venom Gland Transcriptomic and Proteomic Analyses of the Enigmatic Scorpion *Superstitionia donensis* (Scorpiones: Superstitioniidae), with Insights on the Evolution of Its Venom Components. *Toxins* **2016**, *8*, 367. [[CrossRef](#)]
53. Santibanez-Lopez, C.E.; Cid-Urbe, J.I.; Zamudio, F.Z.; Batista, C.V.F.; Ortiz, E.; Possani, L.D. Venom gland transcriptomic and venom proteomic analyses of the scorpion *Megacormus gertschi* Diaz-Najera, 1966 (Scorpiones: Euscorpiidae: Megacorminae). *Toxicon* **2017**, *133*, 95–109. [[CrossRef](#)]
54. 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](#)]
55. Santibanez-Lopez, C.E.; Francke, O.F.; Ureta, C.; Possani, L.D. Scorpions from Mexico: From Species Diversity to Venom Complexity. *Toxins* **2016**, *8*, 2. [[CrossRef](#)]
56. Artimo, P.; Jonnalagedda, M.; Arnold, K.; Baratin, D.; Csardi, G.; de Castro, E.; Duvaud, S.; Flegel, V.; Fortier, A.; Gasteiger, E.; et al. ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.* **2012**, *40*, W597–W603. [[CrossRef](#)]
57. Geer, L.Y.; Domrachev, M.; Lipman, D.J.; Bryant, S.H. CDART: Protein homology by domain architecture. *Genome Res.* **2002**, *12*, 1619–1623. [[CrossRef](#)]
58. Petersen, T.N.; Brunak, S.; von Heijne, G.; Nielsen, H. SignalP 4.0: Discriminating signal peptides from transmembrane regions. *Nat. Methods* **2011**, *8*, 785–786. [[CrossRef](#)]
59. Kall, L.; Krogh, A.; Sonnhammer, E.L. A combined transmembrane topology and signal peptide prediction method. *J. Mol. Biol.* **2004**, *338*, 1027–1036. [[CrossRef](#)]
60. Pineda, S.S.; Chaumeil, P.A.; Kunert, A.; Kaas, Q.; Thang, M.W.C.; Le, L.; Nuhn, M.; Herzig, V.; Saez, N.J.; Cristofori-Armstrong, B.; et al. ArachnoServer 3.0: An online resource for automated discovery, analysis and annotation of spider toxins. *Bioinformatics* **2018**, *34*, 1074–1076. [[CrossRef](#)]
61. Moller, S.; Croning, M.D.; Apweiler, R. Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* **2001**, *17*, 646–653. [[CrossRef](#)]
62. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **2013**, *30*, 772–780. [[CrossRef](#)]
63. Trifinopoulos, J.; Nguyen, L.T.; von Haeseler, A.; Minh, B.Q. W-IQ-TREE: A fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* **2016**, *44*, W232–W235. [[CrossRef](#)]
64. Kalyaanamoorthy, S.; Minh, B.Q.; Wong, T.K.F.; von Haeseler, A.; Jermiin, L.S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nat. Methods* **2017**, *14*, 587–589. [[CrossRef](#)]
65. Hoang, D.T.; Chernomor, O.; von Haeseler, A.; Minh, B.Q.; Vinh, L.S. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* **2018**, *35*, 518–522. [[CrossRef](#)]
66. Kapustin, Y.; Souvorov, A.; Tatusova, T.; Lipman, D. Splign: Algorithms for computing spliced alignments with identification of paralogs. *Biol. Direct.* **2008**, *3*, 20. [[CrossRef](#)]

