



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

Maestría y Doctorado en Ciencias Bioquímicas

El sistema enzimático dual de α -amidación en las glándulas de veneno de escorpión

TESIS

QUE PARA OPTAR POR EL GRADO DE:

Doctor en Ciencias

PRESENTA:

Gustavo Delgado Prudencio

TUTOR PRINCIPAL

Dr. Ernesto Ortiz Suri

[Instituto De Biotecnología, UNAM](#)

MIEMBROS DEL COMITÉ TUTOR

Dra. Marcela Ayala Aceves

[Instituto de Biotecnología, UNAM](#)

Dr. Jorge Luis Folch Mallol

[Centro de Investigación en Biotecnología \(CEIB\), UAEM \(Morelos\)](#)

Ciudad de México. septiembre, 2021



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

El presente trabajo se realizó en el Departamento de Medicina Molecular y Bioprocesos del Instituto de Biotecnología de la Universidad Nacional Autónoma de México bajo la dirección del Dr. Ernesto Ortiz Suri.

Los estudios de doctorado se realizaron con apoyo de la beca otorgada por el Consejo Nacional de Ciencia y Tecnología (CONACyT) con número de CVU 540036.

A mi madre...

A mi abuela (hasta el cielo) ...

Agradecimientos

Al Dr. Ernesto Ortiz Suri por brindarme la oportunidad, por la confianza, consejos y por todo el apoyo brindado, por contribuir de gran manera a mi formación y sobre todo por su amistad.

Al Dr. Baltazar Becerril Lujan y al Dr. Lourival Domingos Possani Postay por todo el apoyo brindado para la realización de este trabajo y por su contribución en mi formación doctoral.

A la Dra. Marcela Ayala Aceves y al Dr. Jorge Luis Folch Mallol por formar parte de mi comité tutorial y por sus valiosos consejos durante todo el doctorado.

A los integrantes del jurado de examen: Dr. Lorenzo Patrick Segovia Forcella, Dra. Rosa María Gutiérrez Ríos, Dra. Blanca Itzelt Taboada Ramírez, Dra. Rosario Adelaida Muñoz Clares y Dr. Diego Claudio Cortez Quezada por el tiempo dedicado a la revisión y corrección de esta tesis.

A la Dra. Georgina Gurrola Briones, Dra. Rita Restano-Cassulini, M.C. Leopoldo Guereca Gurrola, Dra. Juana María Jiménez Vargas, Dr. Edson Norberto Carcamo Noriega, T.L Fredy Ingerborg Coronas Valderrama y al Dr. Fernando Zamudio Zuñiga por el apoyo técnico y asesoría en el laboratorio.

Al Lic. J. Antonio Bolaños Guillen y a Gloria Villa Herrera de la unidad de docencia y Formación de Recursos Humanos por su apoyo administrativo.

A María del Carmen Martínez Segura por todo su apoyo en el laboratorio y por siempre contagiarnos con una sonrisa.

A Leslie Mariella Montes por su apoyo y motivación durante la escritura de esta tesis.

A mis compañeros de laboratorio: Alberto, Ignacio, Teresa, Guillermo, Leonel, Hugo, Linda, Cipriano, Ale, Ilse, Jimena. Gracias por tantos y tan buenos momentos compartidos.

Al M.C. Jerome Jean Verleyen de la Unidad de Secuenciación Masiva y Bioinformática del Instituto de Biotecnología de la UNAM por brindar acceso al cluster "Teopanzolco".

Al "Programa de Apoyo a los Estudios del Posgrado" por el apoyo otorgado para la asistencia al XXXII Congreso Nacional de Bioquímica, Ixtapa, Zihuatanejo, Guerrero. 4-9 noviembre 2018.

Al Consejo Nacional de Ciencia y Tecnología (CONACyT) por la beca otorgada para la realización de este doctorado

Contenido

Resumen	10
Abstract	11
1. Introducción	12
1.1 Reacción de amidación.....	13
1.2 Enzimas amidantes dentro del dominio Eukarya	14
2. Antecedentes	16
2.1 Enzimas amidantes en organismos venenosos.	16
3. Hipótesis	17
4. Objetivos	18
5. Materiales y métodos	19
5.1 Especies analizadas.....	19
5.2 Filtrado de lecturas, ensamble de novo, limpieza y evaluación de los transcriptomas	20
5.3 Identificación y anotación de enzimas amidantes	20
5.4 Organización genómica de las enzimas PAM, PHM y PAL de alacrán	21
5.5 Modelado de las estructuras tridimensionales de las enzimas amidantes	21
5.6 Reconstrucción filogenética de las enzimas PAM, PHM y PAL	22
5.7 Amplificación y clonación de la secuencia PAM de <i>Centruroides noxius</i>	22
5.8 Niveles de expresión de las enzimas amidantes	23
5.9 Diversidad de péptidos amidados	23
6. Resultados	24
6.1 Filtrado de lecturas, ensamble <i>de novo</i> y evaluación de los transcriptomas	24
6.2 El sistema enzimático dual de α -amidación en el orden Scorpiones	26
6.3 Análisis transcriptómico de las glándulas de veneno	27
6.4 Estructuras tridimensionales de las enzimas amidantes	34
6.5 Reconstrucción filogenética de las enzimas amidantes en arácnidos	37
6.6 Niveles de expresión de las enzimas amidantes en la glándula de veneno	40
6.7 Diversidad de péptidos amidados	41
7. Discusión	45
8. Conclusiones	47
9. Perspectivas	48
Referencias	50
Apéndices	57
A1. Remoción de contigs contaminantes.....	57

A2. Ensamble y extensión manual de contigs.....	58
A3. Comparación de los niveles de expresión de la enzima PHM _m	59
A4. Niveles individuales de expresión de las enzimas amidantes	60
A5. Relación entre los niveles de expresión de las enzimas amidantes y la profundidad de secuenciación	61
A6. Clonación, expresión y caracterización bioquímica de la enzima peptidil-glicina α -amidante monooxigenasa (PAM) de alacrán.....	62
Selección de la enzima PAM de alacrán para su expresión y caracterización.....	62
Región de la enzima PAM a expresar (dominios catalíticos PHM-PAL).....	62
Selección del sistema heterólogo de expresión	63
Diseño de “primers” para la amplificación de los dominios PHM-PAL	64
Extracción y digestión de los plásmidos.....	65
Transformación de células electrocompetentes de <i>Pichia pastoris</i> X-33	66
PCR de colonia para identificar integrantes	66
Determinación del fenotipo “methanol utilization” (<i>mut</i>).....	66
Ensayo de expresión	67
Análisis por SDS-PAGE.....	68
Identificación de clonas que expresan los dominios PHM-PAL de <i>Centruroides noxius</i>	68
Identificación de clonas que expresan los dominios PHM-PAL de <i>Rattus norvegicus</i>	70
Ensayo de actividad enzimática	70
Síntesis del sustrato (Dansyl-YVG).....	73
Anexos.....	74
Cartel.....	74

Figuras

Figura 1. Procesamiento enzimático de la señal de amidación.....	14
Figura 2. Número de lecturas, tamaño e integridad de los transcriptomas ensamblados.....	25
Figura 3 Tamaño y número de exones en los genes de las enzimas amidantes.	26
Figura 4. Reconstrucción y extensión manual de secuencias PAM parciales posterior al ensamble <i>de novo</i>	29
Figura 5. Alineamiento múltiple de secuencias aminoacídicas de las enzimas PAM, PHM <i>m</i> y PAL <i>m</i>	30
Figura 6. Topología de los precursores del sistema enzimático de amidación.....	32
Figura 7. Alineamiento de los dominios catalíticos del sistema enzimático de amidación del alacrán <i>C. noxius</i>	33
Figura 8. Predicción de la estructura terciaria de la enzima bifuncional PAM en mamíferos y en alacranes.	35
Figura 9. Modelos de las estructuras terciarias las enzimas <i>phm</i> -PAM/PHM <i>m</i> y <i>pal</i> -PAM/PAL <i>m</i> del alacrán <i>C. noxius</i>	36
Figura 10. Árbol de máxima verosimilitud de los dominios <i>phm</i> -PAM y PHM <i>m</i> de alacrán.	38
Figura 11. Reconstrucción de la historia evolutiva de los dominios <i>pal</i> -PAM y PAL <i>m</i>	39
Figura 12. Niveles normalizados de expresión de las enzimas PAM, PHM <i>m</i> y PAL <i>m</i> en las glándulas de veneno.....	40
Figura 13. Diversidad y abundancia de componentes peptídicos de veneno.	41
Figura 14. Estructuras representativas de los sustratos peptídicos del sistema enzimático dual de amidación.	44

Tablas

Tabla 1. Péptidos amidados en el veneno de alacrán.	12
Tabla 2. Enzimas de la ruta de amidación presentes en las glándulas de veneno de alacrán	27
Tabla 3. Porcentaje de identidad entre los dominios catalíticos de las enzimas PAM, PHM <i>m</i> y PAL <i>m</i>	33

Abreviaturas

C. sc	<i>Centruroides sculpturatus</i>	M. ma	<i>Mesobuthus martensii</i>
C. he	<i>Centruroides hentzi</i>	H. az	<i>Hoffmannihadrurus aztecus</i>
C. no	<i>Centruroides noxius</i>	H. sp	<i>Hadrurus spadix</i>
C. li	<i>Centruroides limpidus</i>	H. co	<i>Hadrurus concolorus</i>
C. or	<i>Centruroides orizaba</i>	M. ge	<i>Megacormus gertschi</i>
C. oc	<i>Centruroides ochraceus</i>	S. ge	<i>Serradigitus gertschi</i>
C. hi	<i>Centruroides hirsutipalpus</i>	P. sc	<i>Paravaejovis schwenkmeyeri</i>
C. po	<i>Centruroides possani</i>	T. in	<i>Thorellius intrepidus</i>
C. vi	<i>Centruroides vittatus</i>	D. me	<i>Diplocentrus melici</i>
T. tr	<i>Tityus trivittatus</i>	U. ya	<i>Urodacus yaschenkoi</i>
L. ab	<i>Leiurus abdullahbayrami</i>		

Resumen

Muchos de los péptidos en el veneno de los alacranes están amidados en su extremo C-terminal. Estos péptidos se caracterizan por tener una señal de amidación en el interior o en el C-terminal de su precursor, esta señal está compuesta por un residuo de glicina que puede estar sucedido por uno o más residuos básicos (típicamente 2). La reacción de amidación ocurre en dos etapas e involucra la acción secuencial de las enzimas peptidilglicina α -hidroxilante monooxigenasa (PHM) y peptidil- α -hidroxiglicina α -amidante liasa (PAL). Esta modificación postraducciona confiere una mayor resistencia a la degradación por carboxipeptidasas y es esencial para la completa función biológica de muchas toxinas peptídicas que actúan sobre diversos canales iónicos, péptidos antimicrobianos, defensinas, y otros componentes presentes en el veneno. Las señales canónicas de amidación identificadas en los precursores peptídicos que fueron deducidos a partir de análisis transcriptómicos sugerían que un sistema enzimático de α -amidación debería ser responsable de esta modificación postraducciona. Información genómica del alacrán *Centruroides sculpturatus* reveló que en los alacranes existen 3 genes que codifican para un sistema enzimático dual de α -amidación, el primero de estos genes codifica para una enzima bifuncional, la peptidil-glicina α -amidante monooxigenasa (PAM), enzima que contiene ambos dominios catalíticos, PHM y PAL (indicados como *phm*-PAM y *pal*-PAM), un dominio de anclaje a la membrana celular (MSD) y un dominio citosólico (CD) que regula su tráfico a través de la vía exocítica. Los dos genes restantes son el resultado de la duplicación de un gen PAM ancestral y codifican para las enzimas monofuncionales PHM_m y PAL_m que carecen de dominios MSD y CD. El porcentaje de identidad aminoacídica entre los dominios catalíticos de las enzimas de los alacranes estudiados se encuentra entre el 27.7 - 32.1% para *phm*-PAM/PHM_m y 31.7 - 33.7% para *pal*-PAM/PAL_m. Los residuos implicados en la catálisis y la coordinación de metales están conservados en cada uno de los dominios catalíticos. Estas enzimas pudieron evolucionar para reconocer sustratos específicos o ser activas en determinadas condiciones fisiológicas. La expresión de los tres genes se confirmó mediante el análisis transcriptómico de las glándulas de veneno de alacranes de las dos ramas basales del orden Scorpiones (parvorden Buthida e Iurida), siendo los bítidos quienes tuvieron los niveles de expresión más altos de estas enzimas y la mayor diversidad y abundancia de péptidos amidados.

Abstract

Many of the peptides in scorpion venom are amidated at their C-terminal end. These peptides are characterized by having an amidation signal within or at the C-terminus of their precursor, this signal is composed of a glycine residue that may be succeeded by one or more basic residues (typically 2). The amidation reaction occurs in two steps and involves the sequential reaction of the enzymes peptidylglycine α -hydroxylating monooxygenase (PHM) and peptidyl- α -hydroxyglycine α -amidating lyase (PAL). This posttranslational modification confers enhanced resistance to degradation by carboxypeptidases and is essential for the full biological function of many peptide toxins that act on various ion channels, antimicrobial peptides, defensins, and other components present in venom. The canonical amidation signals identified in peptide precursors that were deduced from transcriptomic analyses suggested that an α -amidation enzyme system should be responsible for this posttranslational modification. Genomic information from the scorpion *Centruroides sculpturatus* revealed that in scorpions, 3 genes encode for a dual α -amidation enzyme system, the first of these encodes for a bifunctional enzyme, peptidyl-glycine α -amidating monooxygenase (PAM), an enzyme that contains both PHM and PAL catalytic domains (indicated as *phm*-PAM and *pal*-PAM), a membrane spanning domain (MSD) and a cytosolic domain (CD) that regulates PAM trafficking through the exocytic pathway. The remaining two genes are the result of duplication of an ancestral PAM gene and code for monofunctional PHM m and PAL m enzymes that lack MSD and CD domains. The percent amino acid identity between the catalytic domains of the scorpion enzymes is between 27.7 - 32.1% for *phm*-PAM/PHM m and 31.7 - 33.7% for *pal*-PAM/PAL m . Residues involved in catalysis and metal coordination are conserved in each of the catalytic domains. These enzymes could have evolved to recognize specific substrates or be active under certain physiological conditions. Expression of the three genes was confirmed by transcriptomic analysis of scorpion venom glands from the two basal branches of the order Scorpiones (parvorder Buthida and Iurida). Buthids had the highest expression levels of these enzymes and the highest diversity and abundance of amidated peptides.

1. Introducción

El orden Escorpiones representa a uno de los linajes más antiguos dentro del phylum Artrópoda (Sharma *et al.*, 2015; Waddington *et al.*, 2015). La clave del éxito ecológico de estos depredadores reside en gran medida en la producción de potentes venenos que utilizan para alimentarse, así como para defensa y disuasión de potenciales competidores (Oldrati *et al.*, 2016; Ortiz *et al.*, 2015). El veneno de alacrán (o escorpión) es una mezcla compleja de varios componentes, entre ellos péptidos bioactivos con potencial aplicabilidad terapéutica (Ortiz *et al.*, 2015), enzimas (Ortiz *et al.*, 2014), metabolitos, y un relevante arsenal de toxinas con actividad sobre canales iónicos (Na⁺, K⁺, Ca²⁺ y Cl⁻) (Fuller *et al.*, 2007; Gurevitz, 2012; Possani *et al.*, 1999; Rodríguez de la Vega y Possani, 2004, 2005; Xiao *et al.*, 2016). El veneno es producido y secretado por dos glándulas simétricas ubicadas en el telson, el último segmento del metasoma (Chippaux, 2012). En estas glándulas, los componentes peptídicos del veneno son sintetizados y madurados; dicha maduración es un proceso que implica una serie de modificaciones postraduccionales (MPT) que dan lugar a la compleja mezcla de moléculas biológicamente activas (Oldrati *et al.*, 2016). Entre las MPT más comunes que se encuentran en los péptidos del veneno de alacrán están la formación de puentes de disulfuro, escisiones proteolíticas y la amidación del C-terminal (α -amidación). Toxinas y péptidos sin puentes disulfuro (NDBP) son componentes amidados bien conocidos en el veneno de los alacranes (Tabla 1).

Tabla 1. Péptidos amidados en el veneno de alacrán. El péptido señal está subrayado. El péptido maduro se indica en negritas. El propéptido se indica en itálicas y la señal de amidación está indicada en azul.

Tipo	péptido	Uniprot	Secuencia del precursor	C-terminal
Toxinas de sodio	AaH2	<u>P01484</u>	<u>MNYLVMISLALLFVTGVESVKDGYIVDDVNCTYFCGRNAYCNEECTKLKLGESGYCQWASPYGNACYCYKLPDHVRTKGPGRCH</u> GR	H-NH ₂
	LqhIT2	<u>Q26292</u>	<u>MKLLLLLIVSASMLIESLVNADGYIKRRDGGCKVACLIGNEGCDKECKAYGGSYGYCWTWGLACWCEGLPDDKTWKSETNTCG</u> GKK	G-NH ₂
	BmKlTa	<u>Q9XY87</u>	<u>MKLFLLLLISASMLIDGLVNAADGYIRGNSGCKVSLWGNEGCNKECRAYGASGYCWTWGLACWCGQLPDDKTWKSESNCTCG</u> GKK	G-NH ₂
	Cn2	<u>P01495</u>	<u>LLIITACLALIGTVWAKEGYLVDKNTGCKYECLKLGDN DYCLRECKQYQYKGGAGGYCYAFACWCWTHLYEQAIWVPLPNKRCS</u> GK	S-NH ₂
	Ts1	<u>P15226</u>	<u>MKGMILFISCLLLIGIVVECKEGYLMDEHGCKLSCFIRPSGYCGRECGIKKSSGYCAWPACYCYGLPNVWVWDRATNKC</u> GKK	C-NH ₂
	CsEI	<u>P01491</u>	<u>MNSLLMITACLVLIGTVWAKDGYLVEKTGCKKTCYKLGENDFCNRECKWKHIGGSYGYCYGFCYCEGLPDSTQTWPLPNKTC</u> GKK	C-NH ₂
	CsEv3	<u>P01494</u>	<u>MNSLLMITACLFLIGTVWAKEGYLVNKSTGCKYKGLGENEGCDKECKAKNQGGSYGYCYAFACWCEGLPESTPTYPLPNKSC</u> GKK	C-NH ₂
Toxinas de Potasio	NTx	<u>P08815</u>	<u>MKAFYGLIILLFCSMFNLNESTIINVKCTSPKQCSKPKELYGSSAGAKCMNGKCKCYNNG</u>	N-NH ₂
	BmKTX	<u>Q9NII7</u>	<u>MKVFFAVLITLIFCSMIIGHGVGINVKCKHSGQLKPKCKDAGMRFGKCKINGKCDCTPK</u> G	K-NH ₂
	CoTx1	<u>O46028</u>	<u>MEGIAKITLILLFLFVTMHTFANWNTEAAVVCVYRTCDKDKRRRGYRSGKCNACKCYPY</u> GK	Y-NH ₂
	OcKTx5	<u>Q6XLL5</u>	<u>MNAKFILLLVLTMMLLPDTKGAEVIRCSGSKQCYGPKQQTGCTNSKCMNKVKCKCYGC</u> G	C-NH ₂
	OcKTx1	<u>Q6XLL9</u>	<u>MNAKFILLLVVATMMLLPDTQGAEVIKRTPKDCAGPCRKQTCGPHGKCMNRTCRCNRC</u> G	C-NH ₂
Péptidos sin puentes disulfuro	IsCT	<u>Q8MMJ7</u>	<u>MKTQFAILLVALVLFQMFQSDAILGKIWEGIKSLF</u> GKR GLSDLGLDELFDGEISKADRDFLRELMR	F-NH ₂
	BmKb1	<u>Q718F4</u>	<u>MEIKYLLTVFLVLLIVSDHCOAFLFSLIPSAISGLISAFK</u> GRRKR DLNGYIDHFNFRKRDAELEEELSKLPIY	K-NH ₂
	Hp1090	<u>PODJ02</u>	<u>MKTQFAIFLITLVLVLFQMFQSDAIFKAIWVGSIKSLF</u> GKR GLSDLDDLDESFDGEVQADIDFLKELMQ	F-NH ₂
	IsCT2	<u>Q8MTX2</u>	<u>MKTQFAILLVALVLFQMFQSEAFIFGAIWNGIKSLF</u> GRR ALNNDLDDLDELFDGEISQADVDFLKELMR	F-NH ₂
	VAMP-2	<u>E4VP07</u>	<u>MKSOTPFLLFLVFLLAITOSEAIFGAIAGLLKNIF</u> GKR SLRDMTMYLYDPSLSAADLTKLQKLMENY	F-NH ₂

La α -amidación se encuentra en muchos péptidos bioactivos de diversos taxones eucariotas (Jekely, 2013). Esta MPT confiere a los péptidos una mayor resiliencia ante la degradación por carboxipeptidasas, aumentando así su vida media y disminuyendo su tasa de recambio (Gutte, 1995). Se sabe que la amidación es esencial para el correcto funcionamiento de muchos neuropéptidos y hormonas peptídicas de mamíferos (Merkler, 1994). Las toxinas de alacrán también han mostrado requerir de la amidación de su C-terminal para su completa actividad biológica, sin la cual, dicha actividad puede verse reducida (Coelho *et al.*, 2014; Liu *et al.*, 2003). La amidación en el C-terminal resulta en un cambio con dos posibles implicaciones funcionales: el aminoácido terminal amidado podría estar directamente involucrado en eventos de reconocimiento molecular, o la amidación podría simplemente actuar reduciendo la carga negativa del extremo carboxilo y del péptido en su conjunto (Estrada *et al.*, 2011).

1.1 Reacción de amidación

En general, los péptidos amidados se traducen como precursores polipeptídicos de mayor tamaño que contienen una señal de amidación compuesta por un residuo de glicina sucedido típicamente por uno o dos residuos básicos consecutivos (R-X-Gly-Basic o R-X-Gly-Basic-Basic) y en algunos casos por una secuencia propéptido (Tabla 1 y Figura 1A). La señal de amidación es procesada inicialmente por endoproteasas de la familia protein-convertasa (PC1/PC2), que escinden después del o los residuos básicos. Estos residuos a su vez son sustratos de carboxipeptidasas tipo E (CPE) que remueven esos aminoácidos exponiendo al residuo de glicina en el C-terminal del precursor (Eipper *et al.*, 1992b; Fricker, 2005) (Figura 1A). La reacción de amidación ocurre en dos etapas e involucra la acción secuencial de dos actividades enzimáticas. En primer lugar, la enzima dependiente de cobre peptidilglicina α -hidroxilante monooxigenasa (PHM, EC 1.14.17.3, pfam01082 - pfam03712), cataliza la hidroxilación del residuo de glicina, utilizando ascorbato y oxígeno molecular como cosustratos (Figura 1B y 1C). A continuación, la enzima dependiente de zinc peptidil- α -hidroxiglicina α -amidante liasa (PAL, EC 4.3.2.5, pfam01436) cataliza la escisión del residuo hidroxiglicina liberando el péptido amidado y glioxilato (Eipper *et al.*, 1991; Perkins *et al.*, 1990) (Figura 1B y 1D).

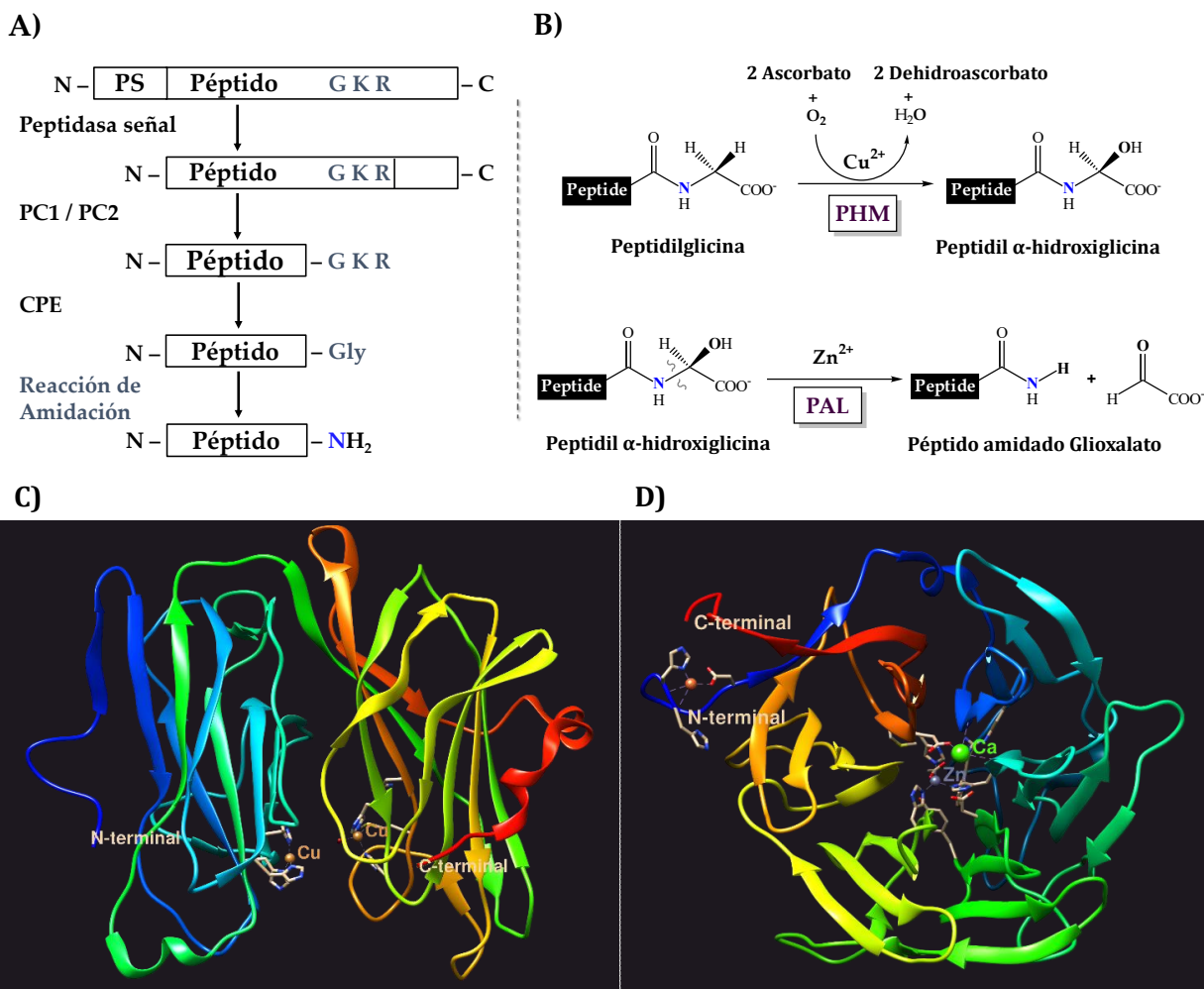


Figura 1. Procesamiento enzimático de la señal de amidación. **A)** Ruta completa que conduce a la amidación de péptidos. PS, Péptido Señal; PC1/PC2, Protein-convertasas tipo 1/2; CPE, carboxipeptidasa E. **B)** Reacción secuencial de amidación catalizada por las enzimas PHM y PAL. **C)** y **D)** Estructuras tridimensionales obtenidas por cristalografía de rayos X de los dominios *phm*-PAM y *pal*-PAM de *R. norvegicus* (PDB: 1SDW/3FVZ).

1.2 Enzimas amidantes dentro del dominio Eukarya

Al igual que los péptidos amidados, las enzimas amidantes (PHM y PAL) se encuentran ampliamente distribuidas en diversos taxones eucariotas (excepto en amebozoos, coanoflagelados, hongos y plantas) (Attenborough *et al.*, 2012). Sin embargo, la forma en que se expresan estas enzimas difiere entre taxones. En vertebrados, como *Bos taurus*, *Rattus norvegicus*, *Xenopus laevis*, *Homo sapiens* y otros, un único gen codifica para un polipéptido con ambas actividades en forma de una enzima bifuncional, la peptidil-glicina α-amidante monooxigenasa (PAM), enzima que además cuenta con un dominio integral de membrana y un dominio citosólico hacia el C-terminal (Eipper *et al.*, 1992a; Eipper *et al.*, 1987; Glauder *et al.*, 1990; Mizuno *et al.*, 1987). Se informó de la

misma estructura PAM de dos dominios catalíticos en el gasterópodo *Aplysia californica* (Fan *et al.*, 2000). Curiosamente, otro gasterópodo, *Lymnaea stagnalis* produce un zimógeno que consta de cuatro dominios PHM diferentes seguidos de un único dominio PAL, que endógenamente se convierte en una mezcla de isoenzimas monofuncionales (Spijker *et al.*, 1999). Se han descrito hasta 7 isoformas PAM en *R. norvegicus* generadas por “splicing” alternativo. Estas isoformas incluyen configuraciones con y sin sitios proteolíticos internos, éstos dan como resultado enzimas PHM y PAL independientes y solubles (*phm*-PAM y *pal*-PAM) (Eipper *et al.*, 1992a; Eipper *et al.*, 1993). Entre los artrópodos, las enzimas PHM y PAL, pero no la enzima bifuncional PAM se expresan en los insectos *Apis mellifera*, *Drosophila melanogaster* y otros (Han *et al.*, 2004; Kolhekar *et al.*, 1997; Zabriskie *et al.*, 1994), mientras que tanto las enzimas monofuncionales, indicadas en adelante con el sufijo “m” (*e.g.*, PHMm/PALm), como la enzima bifuncional PAM se expresan en crustáceos como la pulga de agua *Daphnia pulex* (Attenborough *et al.*, 2012) y *D. magna*, cuyo porcentaje de similitud aminoacídica entre los dominios *phm*-PAM/PHMm y *pal*-PAM/PALm es 28.8 y 36.1% respectivamente. Análisis filogenómicos sugieren que la enzima bifuncional PAM pudo estar en el último ancestro común eucariota (LECA) y que las enzimas PHMm/PALm en ciertos linajes son resultado de eventos de duplicación del gen PAM ancestral (Attenborough *et al.*, 2012). No se dispone de información sobre el sistema enzimático de amidación en otros subfilos *e.g.*, quelicerados y, en particular, en arácnidos.

2. Antecedentes

2.1 Enzimas amidantes en organismos venenosos.

Los gasterópodos marinos del género *Conus* producen una gran variedad de toxinas peptídicas (conotoxinas), de las cuales una fracción significativa está amidada. Ul-Hasan *et al.*, en 2013, clonaron cDNAs codificantes para enzimas PAM a partir de los conductos del veneno de estos caracoles, cuya expresión heteróloga demostró actividad α -amidante. Ward *et al.*, en 2018, realizaron un análisis transcriptómico y proteómico de la glándula de veneno del escorpión *Centruroides hentzi*. En este trabajo se reportó una monooxigenasa en el veneno (anotada como GFWZ01000197.1 TSA: *Centruroides hentzi* Chent_MonoO). Un análisis de secuencia riguroso demostró que esta corresponde a una enzima PHMm de secuencia completa (ver materiales y métodos, sección 5.3). En 2018, de Oliveira *et al.*, reportaron una PAM de secuencia corta (347 aminoácidos) en las glándulas de veneno del escorpión *Tityus obscurus* (GenBank: JAT91064). Tras el análisis de secuencia se demostró que ésta no corresponde a una enzima PAM, sino a una PHMm. Del mismo modo, Kuhn-Nentwig *et al.*, en 2019, realizaron un análisis transcriptómico y proteómico de la glándula de veneno de la araña *Cupiennius salei*, reportando otra enzima PAM de secuencia corta (353 aminoácidos), anotada como PAM_CUPSA [MH766628]. El análisis de secuencia mostró que esta enzima también correspondía a una PHMm. Estos resultados demostraron la presencia de la enzima PHMm en los venenos de alacranes y arañas; sin embargo, no hay reportes de la enzima PALm o PAM o del subdominio *pal*-PAM que complete la reacción de α -amidación en las glándulas de veneno de los alacranes.

3. Hipótesis

A pesar de que no se ha descrito ningún sistema de amidación en arácnidos, la correlación comprobada entre la presencia de las señales canónicas de α -amidación en transcritos obtenidos por diferentes análisis transcriptómicos y la obtención de péptidos maduros amidados confirmados mediante análisis bioquímicos y proteómicos (Abdel-Rahman *et al.*, 2013; Becerril *et al.*, 1995; Bougis *et al.*, 1989; de Oliveira *et al.*, 2018; Romero-Gutierrez *et al.*, 2017), indican que los sistemas PHMm más PALm y/o PAM deben estar presentes en las glándulas de veneno de los escorpiones. Por lo tanto, la hipótesis de este proyecto propone que, utilizando técnicas de Secuenciación Masiva de Nueva Generación, Bioinformática y Biología Molecular, seremos capaces de identificar, amplificar y clonar, así como caracterizar estructural y filogenéticamente las secuencias codificantes para la o las enzimas que participan en el proceso de α -amidación en escorpiones.

4. Objetivos

General

Describir el sistema enzimático de amidación en el orden Scorpiones

Particulares

- Identificar la o las enzimas que se expresan en las glándulas de veneno de los alacranes, que son responsables de catalizar la reacción de α -amidación en los componentes peptídicos del veneno
- Explorar la historia evolutiva de las enzimas amidantes de los alacranes
- Evaluar y comparar los niveles de expresión de las enzimas amidantes en las glándulas de veneno de alacranes representativos de las dos ramas basales del orden Scorpiones: parvódenes Buthida e Iurida
- Evaluar la diversidad y abundancia de los componentes amidados en el veneno de bítidos e iúridos

5. Materiales y métodos

5.1 Especies analizadas

Partimos de las lecturas crudas generadas para el análisis transcriptómicos de las glándulas de veneno de 4 especies de alacranes previamente publicados del género *Centruroides*: *C. limpidus* (ERR3289195 / ERR3289194) (Cid-Urbe *et al.*, 2019), *C. hirsutipalpus* (Valdez-Velazquez *et al.*, 2020), *C. vittatus* (SRR5345824 / SRR5338072) (McElroy *et al.*, 2017), *C. hentzi* (SRR6041834 / SRR6041835) (Ward *et al.*, 2018), así como de lecturas de secuenciación para 4 especies inéditas de este género, *C. noxius*, *C. ochraceus*, *C. possani*, *C. orizaba*, y de los datos genómicos de *C. sculpturatus* (BioProject: PRJNA168116) disponibles en el NCBI. Estas especies están distribuidas principalmente en el territorio mexicano y el sur de Estados Unidos. Adicional a esta información, incluimos datos de especies estrechamente relacionadas, *Tityus trivittatus* (especie inédita) (centro, centro-sur del continente americano), *Mesobuthus martenssi* (SRR4188636) distribuido en el continente asiático y *Leiurus abduhbayrami* (especie inédita) distribuido en el sureste de Europa y Asia occidental. Un aspecto evolutivo relevante es que estas especies pertenecen a una de las dos ramas basales del orden Scorpiones, parvorden Buthida, y con la excepción de *C. hentzi*, las especies antes mencionadas son consideradas como peligrosas para el humano.

Datos de secuenciación de otros alacranes distribuidos en México, *Hoffmanniadrurus aztecus* (ERR3534794), *Hadrurus concolorus* (ERR3561754) (Santibanez-Lopez *et al.*, 2019b), *Hadrurus spadix* (SRR4069278/SRR4069277) (Rokyta y Ward, 2017), *Paravaejovis schwenkmeyeri* (ERR2653951) (Cid-Urbe *et al.*, 2018), *Thorellius cristimanus* (reportado como *T. atrox*) (ERR2367086) (Romero-Gutierrez *et al.*, 2017), *Serradigitus gertschi* (ERR2843384) (Romero-Gutierrez *et al.*, 2018), *Superstitionia donensis* (SRR4381683) (Santibanez-Lopez *et al.*, 2016), *Megacormus gertschi* (SRR3657526) (Santibanez-Lopez *et al.*, 2017), así como de *Urodacus yaschenkoi* (Australia) (SRR1557168) (Luna-Ramirez *et al.*, 2015), se incluyeron como representantes de la segunda rama basal del orden, parvorden Iurida. Este segundo grupo solo tiene una especie considerada de importancia médica (*Hemiscorpius lepturus*, país: Irán) (Torabi *et al.*, 2020) y no forma parte de este trabajo. En total, se contó con información de 24 especies de escorpiones del viejo y nuevo mundo pertenecientes a 7 de las 20 familias actualmente reconocidas (Santibanez-Lopez *et al.*, 2015).

5.2 Filtrado de lecturas, ensamble *de novo*, limpieza y evaluación de los transcriptomas

Todas las lecturas incluidas en este trabajo se generaron por tecnología Illumina y son de extremo pareado. Las lecturas cuentan con una longitud de 72 bases para la mayoría de las especies, con la excepción de *C. hentzi*, *C. vittatus*, *H. spadix* y *M. martenssi*, cuyas lecturas tienen una longitud de 150, 101, 150 y 89 bases respectivamente. La calidad de las lecturas se evaluó con FastQC (Andrews, 2010). La limpieza de las lecturas se realizó usando fastp 0.21.0 (Chen *et al.*, 2018). Entre los parámetros de limpieza se consideró la remoción de adaptadores y las colas poli G generadas por secuenciación Illumina (Chen *et al.*, 2018), así como la remoción de lecturas de baja calidad y complejidad (lecturas que contienen secuencias repetitivas largas. Ver Apéndice A2). Se retuvieron lecturas cuya longitud final fuera mayor o igual a 25 bases y con valores de calidad Phred mayores o iguales a Q20 y con niveles de complejidad superiores a 30.

Las lecturas se ensamblaron *de novo* utilizando rna-SPAdes 3.15.0 (Bushmanova *et al.*, 2019). El tamaño de los k-meros fue determinado por rna-SPAdes, utilizando k-meros de longitud 29 y 35 para las lecturas de 72 bases, 29 y 43 para las lecturas de 89 bases y de 49 y 73 para las lecturas de 150 bases.

Se removieron secuencias contaminantes alineando los contigs ensamblados contra una serie de genomas representativos de mamíferos, insectos, nematodos, protistas, bacterias, hongos y plantas (*Homo sapiens*, *Rattus norvegicus*, *Mus musculus*, *Drosophila melanogaster*, *Coccinella septempunctata*, *Caenorhabditis elegans*, *Acanthamoeba astronyxis*, *Escherichia coli*, *Bradyrhizobium pachyrhizi*, *Streptomyces albus*, *Bacillus subtilis*, *Klebsiella variicola*, *Aspergillus terreus*, *Trametes hirsuta*, *Saccharomyces cerevisiae*, *Arabidopsis thaliana*, *Jatropha curcas*) usando el programa deconseq (Schmieder y Edwards, 2011).

La estimación de la integridad y redundancia de los transcriptomas ensamblados se realizó mediante la búsqueda de las secuencias de 1,013 ortólogos conservados entre artrópodos usando BUSCO 5.0.0 (Benchmarking Universal Single Copy Orthologs) (Seppey *et al.*, 2019).

5.3 Identificación y anotación de enzimas amidantes

Homólogos de las enzimas PAM, PHM y PAL se identificaron mediante TBLASTN utilizando la secuencia PAM de *R. norvegicus* (Uniprot, P14925) y las secuencias PHM y PAL2 de *Drosophila melanogaster* (Uniprot, O01404 y Q9W1L5) como referencia (evalue=1e-11). La organización de los dominios se evaluó con el servidor NCBI-CDART (Geer *et al.*, 2002). Las enzimas PAM se caracterizan por tener la organización de dominios Cu2_monooxygen (pfam01082)-Cu2_monoox_C (pfam03712) seguidos del dominio NHL (pfam01436), mientras las PHMm

incluyen únicamente los dominios Cu₂_monooxygen (pfam01082)–Cu₂_monoox_C (pfam03712) y las PAL_m solo cuentan con el dominio NHL (pfam03712). Se identificaron secuencias adicionales distintivas como el péptido señal de secreción (SP) con SignalP 4.1 y Phobius (Kall *et al.*, 2004; Petersen *et al.*, 2011), la región del propéptido (Pp) con ArachnoServer v. 3.0 (Pineda *et al.*, 2018) y el dominio de anclaje a membrana de las enzimas PAM con TMHMM 2.0 (Moller *et al.*, 2001). La identificación y delimitación de los dominios catalíticos y los residuos implicados en la coordinación de metales y la formación de puentes disulfuro se realizó manualmente mediante el alineamiento de las secuencias recuperadas con la secuencia de referencia de *R. norvegicus*, la cual cuenta con la anotación más completa de residuos y ensayos enzimáticos. Los sitios potenciales de glicosilación se predijeron con NetNGlyc 1.0 (<http://www.cbs.dtu.dk/services/NetNGlyc/>).

En los casos en que las secuencias se encontraran incompletas o parciales, se procedió a alinear las lecturas de secuenciación contra una secuencia de una especie del mismo género, o familia, o consenso usando bowtie2 (Langmead y Salzberg, 2012). Las lecturas recuperadas en formato SAM (Sequence Alignment Map) se ordenaron usando SAMtools 1.11 (Danecek *et al.*, 2021). El archivo SAM se convirtió a formato FASTA usando el script de Python sam2fasta.py (<https://sourceforge.net/projects/sam2fasta/files/>). Las secuencias ordenadas y alineadas se usaron para ensamblar manualmente nuevos contigs (por consenso de secuencias) o extender la longitud de contigs previamente generados por rna-SPAdes, así como para fusionar dos o más contigs.

5.4 Organización genómica de las enzimas PAM, PHM y PAL de alacrán

Los genes de las enzimas PAM, PHM y PAL se recuperaron de los genomas de *C. esculentatus* (GeneID:111617460, 111638043 y 111633922), *R. norvegicus* (Gene ID: 25508) y *Drosophila melanogaster* (Gene ID:37823, 36033 y 37746). La identificación y organización de intrones y exones se realizó con Splign (Kapustin *et al.*, 2008).

5.5 Modelado de las estructuras tridimensionales de las enzimas amidantes

Las secuencias aminoacídicas de las cadenas maduras de las enzimas PAM PHM_m y PAL_m se modelaron con ayuda del “Colab notebook” de Alpha fold 2 (Jumper *et al.*, 2021). Los metales divalentes para cada modelo se transfirieron de los archivos pdb de las estructuras cristalográficas de las enzimas PHM (*phm*-PAM) y PAL (*pal*-PAM) de *R. norvegicus* (PDB: 1SDW/3FVZ) a los modelos generados.

5.6 Reconstrucción filogenética de las enzimas PAM, PHM y PAL

La reconstrucción filogenética de la enzima PAM y sus dos dominios catalíticos *phm*-PAM y *pal*-PAM (con el sufijo '-PAM' utilizado para diferenciarlos de aquéllos de las enzimas monofuncionales), y de las enzimas independientes PHM m y PAL m (con el sufijo ' m ', para 'monofuncional') se realizó utilizando el método de máxima verosimilitud (ML) con secuencias de nucleótidos. Se obtuvieron secuencias adicionales de organismos relacionados filogenéticamente (grupos externos) del NCBI o se ensamblaron a partir de lecturas de secuenciación de transcriptomas depositados en el SRA-NCBI. Todas las secuencias se alinearon con MAFFT 7.407 (Kato *et al.*, 2019). La búsqueda del mejor modelo de sustitución (GTR + F + I + G4) y el análisis ML se realizaron con IQ-TREE 1.6.9 (Kalyaanamoorthy *et al.*, 2017; Nguyen *et al.*, 2015) utilizando el método de arranque ultrarrápido (UFBoot2) (Hoang *et al.*, 2018) con 10,000 réplicas.

5.7 Amplificación y clonación de la secuencia PAM de *Centruroides noxius*

Se extrajo el RNA total del telson de una hembra del alacrán *C. noxius* con el kit SV Total RNA Isolation System (Promega Corporation, Madison, WI, EE. UU.). El cDNA se amplificó con el kit cDNA First Strand para RT-PCR (AMV) (Roche, Basilea, Suiza). Los primers utilizados, Cen-Fw3 (5'-GAT CTT GTA AAC GGC GTA TTT CCC TT-3 ') y Cen-Rv4 (5'-CCG ATA TCC TCC CAA CCA TCC TTT C-3 '), se diseñaron a partir del consenso de las secuencias PAM de dos escorpiones del género *Centruroides* (*C. limpidus* y *C. orizaba*). Las condiciones de amplificación fueron 3 min a 96 °C, seguido de 30 ciclos de 30 s a 96 °C, 1 min a 56 °C y 2 min a 68 °C, más un paso final de 5 min a 68 °C. Se utilizó una polimerasa Pfu recombinante producida en nuestro laboratorio. El producto de PCR se purificó con el kit de extracción en gel QIAQuick (QIAGEN GmbH, Hilden, Alemania), se ligó en el vector pBluescript II KS (+) digerido con EcoRV y se electroporó en células DH5 α de *Escherichia coli* electrocompetentes. Las clonas positivas se seleccionaron mediante el sistema azul/blanco sembrando las células en medio LB/ampicilina, complementado con X-Gal/IPTG. Los plásmidos se prepararon por lisis alcalina y se sometieron a secuenciación con los 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') y Cnox-In3 (5'-ATT GAT GCT GAT GAT GCC TA-3 ').

5.8 Niveles de expresión de las enzimas amidantes

La cuantificación de los niveles de expresión, referidos como transcritos por millón (TPM), se realizó usando Salmon 1.3.0 (Patro *et al.*, 2017) con las opciones `-seqBias` y `--gcBias`. Los niveles de expresión fueron normalizados usando como medida normalizadora los TPM de un gen “housekeeping”. Para ello, se evaluó la presencia, completitud y singularidad de copia de los siguientes genes candidatos: GAPDH (gliceraldehído 3-fosfato deshidrogenasa), SDHA (succinato deshidrogenasa), HPRT1 (hipoxantina fosforribosil transferasa 1) (Silver *et al.*, 2006), PGK1 (fosfoglicerato quinasa 1), CANX (Calnexina), ATP5B (ATP sintasa F1 subunidad beta), TBP (proteína de unión a TATA), YWHAZ (Proteína de activación Zeta de tirosin 3-monooxigenasa/triptófano 5-monooxigenasa), enunciados en el “qPCR Reference Gene Selection Protocol, Sigma-Aldrich Co”, y actina.

5.9 Diversidad de péptidos amidados

Se construyó una base de datos con secuencias aminoacídicas representativas para cada tipo de toxinas y otros componentes que se han reportado en los venenos de distintos alacranes (Cid-Uribe *et al.*, 2020). Se recuperaron las secuencias incluidas en el “Animal toxin annotation project” de UniProt (Jungo *et al.*, 2012) [actualizado a Nov/2020] y de “Kalium: Toxins Active on Potassium Channels” (Tabakmakher *et al.*, 2019) [actualizado a Nov/2020]. Las secuencias aminoacídicas redundantes entre ambas bases de datos se removieron con CD-HIT 4.8.1 (Fu *et al.*, 2012) con las opciones: `-c 1.0` y `-s 1.0` (opciones que permiten remover secuencias con el 100% de similitud y de igual longitud). La identificación de toxinas y otros componentes de veneno se realizó con la opción `blastp` de diamond 2.0.6.144 (Buchfink *et al.*, 2021) con un e-value de $1e-18$.

A partir de las secuencias identificadas y clasificadas como componentes peptídicos de veneno *e.g.*, toxinas (NaTx, KTx, CaTx, ClTx, TRPTx), péptidos de defensa del huésped (NDBP1-4, defensinas, péptidos aniónicos, péptidos tipo waprina) y otros, se buscaron señales de amidación dentro de las secuencias de los precursores o en el C-terminal de los mismos. Aunque se ha descrito que en mamíferos la señal de amidación está compuesta por un residuo de glicina seguido típicamente por uno o dos residuos básicos consecutivos, los precursores de algunos péptidos en el veneno de alacrán cuentan con hasta 4 residuos básicos que suceden al residuo de glicina (Tabla 1), por lo que se consideraron las combinaciones posibles para estos 4 residuos básicos en la señal de amidación (G, GK, GR, GKK, GKR, GRR, GRK, GK RK, GK RR, GRRR, GRKR, GRKK, GK KK, GK KR, GRRK, GK KKK, GK KKR, GK KRR, GK RRR, GRRRR, GK RRR, GK KKR, GK RKR, GRRRK, GK RRR).

6. Resultados

6.1 Filtrado de lecturas, ensamble *de novo* y evaluación de los transcriptomas

Se encontraron diferencias en el número de lecturas (profundidad de secuenciación) disponibles para cada uno de los transcriptomas analizados. En *C. hirsutipalpus* se tuvo la menor profundidad de secuenciación con 5 millones de lecturas, mientras que en *T. trivittatus* se contó con 112 millones de lecturas, siendo el transcriptoma con la mayor profundidad de secuenciación (Figura 2A). La relación entre las lecturas retenidas tras la remoción de adaptadores y secuencias de baja calidad se muestra en la Figura 2A. El transcriptoma del alacrán *T. trivittatus* tuvo el mayor número de contigs ensamblados con 89,350, mientras que para el alacrán *C. hirsutipalpus* se tuvo el menor número de contigs con 16,057 (Figura 2B). La tasa de mapeo de las lecturas en los ensambles fue superior al 90%, con la excepción de *H. spadix* (hembra) cuya tasa de mapeo fue de solo 84% (Figura 2C). Se encontró un bajo número de secuencias contaminantes en los ensambles (<2,000 contigs), la mayoría de estas secuencias son de origen bacteriano y de humano (Apéndice A1). En los transcriptomas con una profundidad de secuenciación menor a 10 millones de lecturas (*C. hirsutipalpus*, *S. donensis* y *C. noxius*) se identificaron menos del 50% de los ortólogos BUSCO completos (Figura 2D). La relación entre los ortólogos de secuencia única y duplicados, así como los que se encuentran fragmentados o que no fueron identificados se muestran a detalle en la Figura 2D. Se observó que la profundidad de secuenciación guarda una estrecha correlación con el número e integridad de los contigs ensamblados, donde un menor número de lecturas dificulta la identificación de ortólogos de secuencia completa (Figura 2E y 2F). En la mayoría de los transcriptomas con más de 45 millones de lecturas se lograron identificar entre el 85 y 98% de los ortólogos BUSCO completos, con la excepción de *H. aztecus* y *C. Orizaba* en los que otros factores metodológicos pudieron afectar el resultado del análisis de integridad de los transcriptomas (Figura 2F).

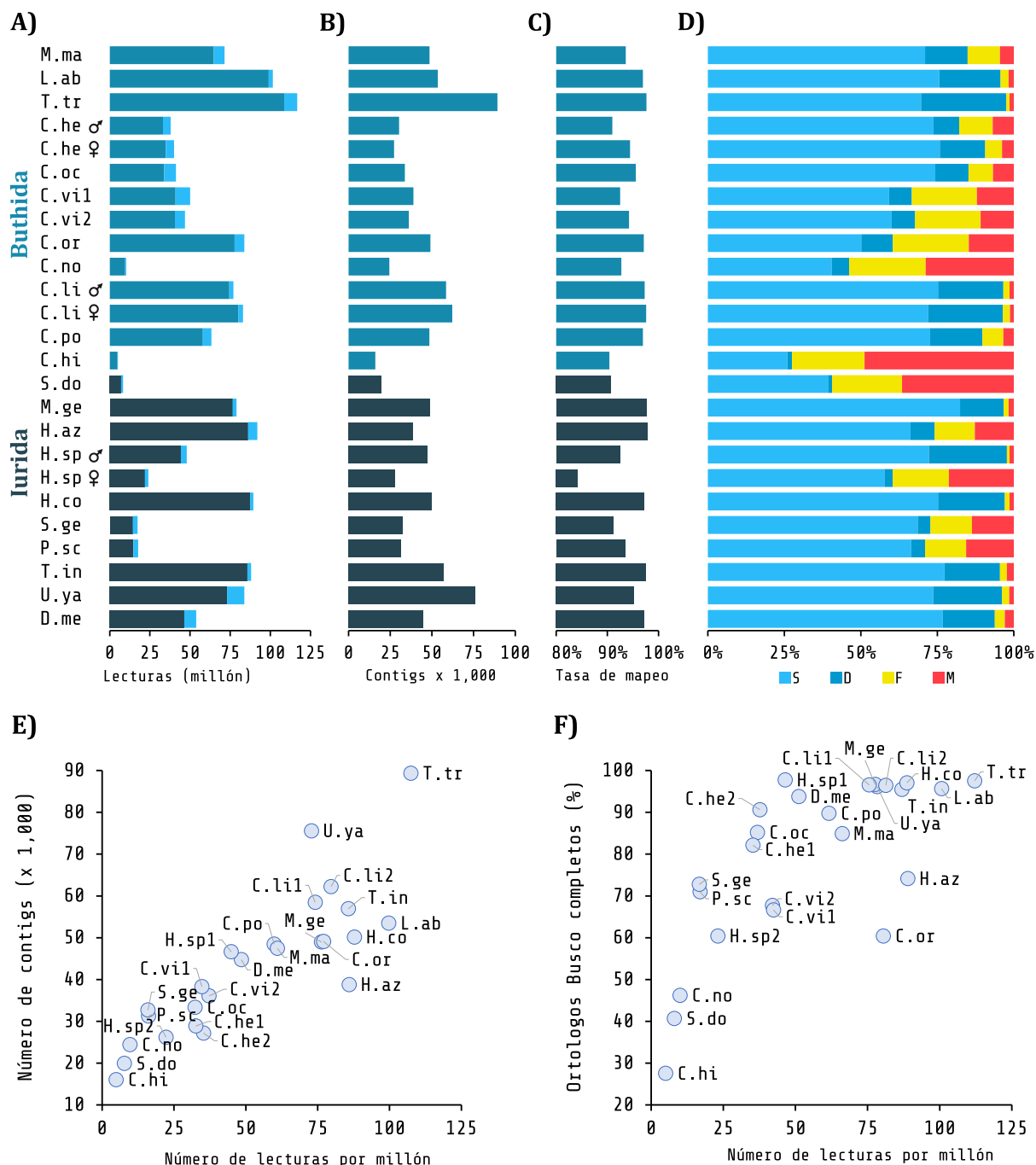


Figura 2. Número de lecturas, tamaño e integridad de los transcriptomas ensamblados.

A) Número de lecturas retenidas tras la remoción de adaptadores y secuencias baja calidad en **■**) Bútididos; **■**) Iúrididos; **■**) Número de lecturas que fueron removidas por no cumplir con los requisitos de calidad y longitud; **B)** Número de contigs ensamblados por rnaSPAdes; **C)** Tasa de mapeo (porcentaje de lecturas que fueron mapeadas de forma única en los ensamblados); **D)** Porcentaje acumulado de los 1,013 ortólogos conservados en artrópodos incluidos en la base de datos BUSCO; **■**) S) Ortólogos completos de copia única; **■**) D) Ortólogos completos que se encuentran duplicados; **■**) F) Ortólogos incompletos o fragmentados; **■**) M) Ortólogos no presentes en los ensamblados; **E)** Relación entre la profundidad de secuenciación y el número de contigs ensamblados por rnaSPAdes; **F)** Relación entre la profundidad de secuenciación y porcentaje de ortólogos completos identificados por BUSCO.

6.2 El sistema enzimático dual de α -amidación en el orden Scorpiones

Se identificaron tres genes que codifican para un doble sistema enzimático de α -amidación en el genoma de *C. sculpturatus*. El primero de estos genes codifica para una enzima bifuncional PAM y los dos genes restantes (genes parálogos) codifican para enzimas PHM y PAL monofuncionales (PHM m y PAL m). La Figura 3A muestra la estructura de los genes PAM, PHM m y PAL m en *C. sculpturatus*, incluyendo el tamaño, número y arreglo de exones. Los genes PAM de rata y los genes PHM, PAL1 y PAL2 de la mosca de la fruta se muestran para su comparación (Figura 3B y 3C). Aunque el gen PAM de alacrán es más pequeño que el de mamífero (rata), ambos tienen un número parecido de exones, con 22 y 25 respectivamente. Los genes de las enzimas PHM m de alacrán y de la mosca de la fruta cuentan con 11 y 8 exones respectivamente. En el caso de PAL m , se observó el mismo número de exones en el gen PAL de alacrán y el gen PAL2 de mosca (8 exones) y 6 exones para el gen PAL1 de mosca.

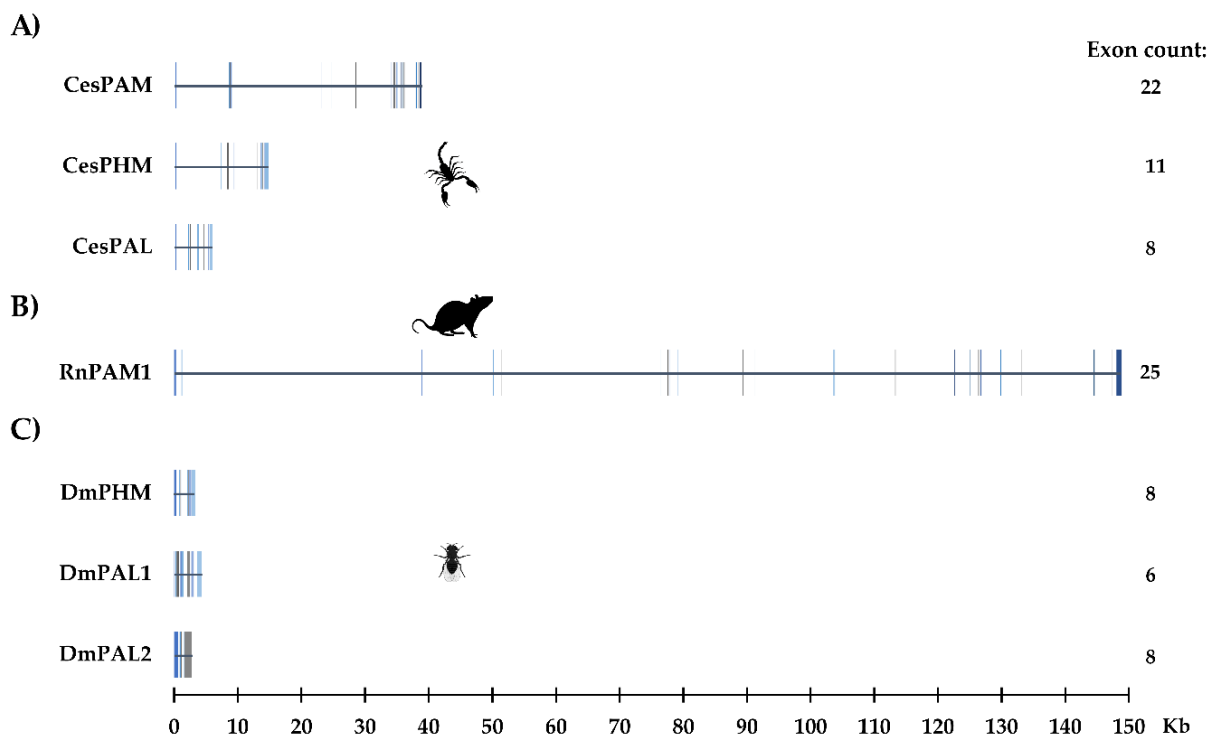


Figura 3 Tamaño y número de exones en los genes de las enzimas amidantes.

A) *C. sculpturatus* (CesPAM, CesPHM y CesPAL); B) *R. norvegicus* (PAM1); C) *D. melanogaster* (DmPHM, DmPAL1 y DmPAL2). Los exones se indican como bloques verticales.

6.3 Análisis transcriptómico de las glándulas de veneno

El análisis transcriptómico mostró que los tres genes se coexpresan en las glándulas de veneno de los alacranes. Se identificaron transcritos PAM de secuencia completa en 18 de las 24 especies analizadas, secuencias parciales se identificaron en las 6 especies restantes. Dos secuencias PAM fueron amplificadas por RT-PCR a partir del RNA total de la glándula de veneno de los alacranes *Centruroides noxius* y *Centruroides limpidus* (Tabla 2 y apéndice A6). Transcritos más cortos que codifican para enzimas PHM y PAL monofuncionales (PHM_m y PAL_m) se identificaron en la mayoría de los transcriptomas analizados (Tabla 2). De manera adicional, se identificaron secuencias codificantes para otros componentes de la vía de α -amidación en los transcriptomas analizados. Estos transcritos codifican para convertasas tipo 1 y 2 (PC1 y PC2) y para carboxipeptidasa E (CPE), enzimas que operan río arriba en la vía de la α -amidación (Figura 1A, Tabla 2).

Tabla 2. Enzimas de la ruta de amidación presentes en las glándulas de veneno de alacrán

Parvorder	Familia	Especie	PAM	PHM	PAL	PC1	PC2	CPE
Buthida	Buthidae	<i>Mesobuthus martensii</i> [*]						
		<i>Leiurus abduallahbayrami</i> [*]						
		<i>Tityus trivittatus</i>						
		<i>Centruroides hentzi</i>						
		<i>Centruroides ochraceus</i>						
		<i>Centruroides vittatus</i>						
		<i>Centruroides orizaba</i>						
		<i>Centruroides noxius</i> ^a						
		<i>Centruroides sculpturatus</i>						
		<i>Centruroides limpidus</i> ^b						
		<i>Centruroides possani</i>						
		<i>Centruroides hirsutipalpus</i>						
Iurida	Superstitionidae	<i>Superstitionia donensis</i>						
	Chactidae	<i>Megacormus gertschi</i>						
	Hadruridae	<i>Hoffmannihadrurus aztecus</i>						
		<i>Hadrurus spadix</i>						
		<i>Hadrurus concolorus</i>						
	Vaejovidae	<i>Thorellius cristimanus</i>						
		<i>Paravaejovis schwenkmeyeri</i>						
		<i>Serradigitus gertschi</i>						
	Diplocentridae	<i>Diplocentrus melici</i>						
Urodacidae	<i>Urodacus yaschenkoii</i> [*]							

(y): Secuencias completas de transcritos PAM, PHM_m y PAL_m; (): Secuencias PAM parciales con más del 60% de la longitud total estimada; (): Secuencias PC1, PC2 y CPE completas; (): Secuencias PC1, PC2 y CPE con más del 60% de la longitud estimada; (✓): Secuencias parciales menores o igual al 60% de la secuencia total estimada; ^aSecuencia PAM obtenida por RT-PCR; ^bSecuencia PAM verificada mediante secuenciación de DNA; * Escorpiones del Viejo Mundo.

La completitud final de las secuencias nucleotídicas PAM mostrada en la Tabla 2 se obtuvo mediante el alineamiento de las lecturas con las secuencias PAM completas de otros alacranes. La Figura 4A muestra como ejemplo el alineamiento de las secuencias parciales identificadas en el transcriptoma de *Paravaejovis schwenkmeyeri* con una secuencia PAM consenso de especies de la misma familia (Vaejovidae). En el alineamiento se observan regiones que no son cubiertas por los contigs ensamblados por rnaSPAdes. La baja cobertura observada durante el alineamiento manual de lecturas impidió la conexión y extensión de secuencias durante el ensamble *de novo*. El alineamiento de las lecturas con la secuencia PAM consenso permitió construir nuevos contigs que cubren regiones previamente desconocidas y fusionar contigs en otros de mayor longitud (Figura 4B). La identificación de secuencias PAM parciales ocurrió principalmente en los transcriptomas con una profundidad de secuenciación menor a los 25 millones de lecturas (Figura 4C). Este resultado guardó relación con la estimación de la integridad general de los transcriptomas evaluada en el punto 6.1, donde la completitud de secuencias se vio afectada por el bajo número de lecturas disponibles para realizar los ensambles *de novo*. El porcentaje de completitud de las secuencias nucleotídicas PAM recuperadas inicialmente de los ensambles se situó entre un 26 y 77 % de la longitud total estimada (considerando la longitud de las secuencias de otros alacranes relacionados) (Figura 4C). El porcentaje de completitud después de la reconstrucción y extensión manual de contigs se situó entre el 57 y 94 % (Figura 4D). El alineamiento de lecturas permitió incrementar entre un 23 y 36 % la completitud de secuencias PAM (con excepción de *D. melici* en el que no se recuperó información adicional) (Figura 4E). Se observó que en los alacranes *C. hirsutipalpus* y *C. noxius* (parvorden Buthida) a pesar de tener un número menor de lecturas en comparación con otros alacranes, como *S. donensis* y *S. gertschi* (parvorden Iurida), cuentan con una mayor completitud de las enzimas PAM sugiriendo que esta enzima podría tener niveles más altos de expresión en los alacranes bítidos (ver sección 6.6).

El ensamble y extensión manual de contigs para el resto de los alacranes (*C. hirsutipalpus*, *C. noxius*, *S. donensis*, *S. gertschi* y *D. melici*) puede consultarse en el Apéndice A2.

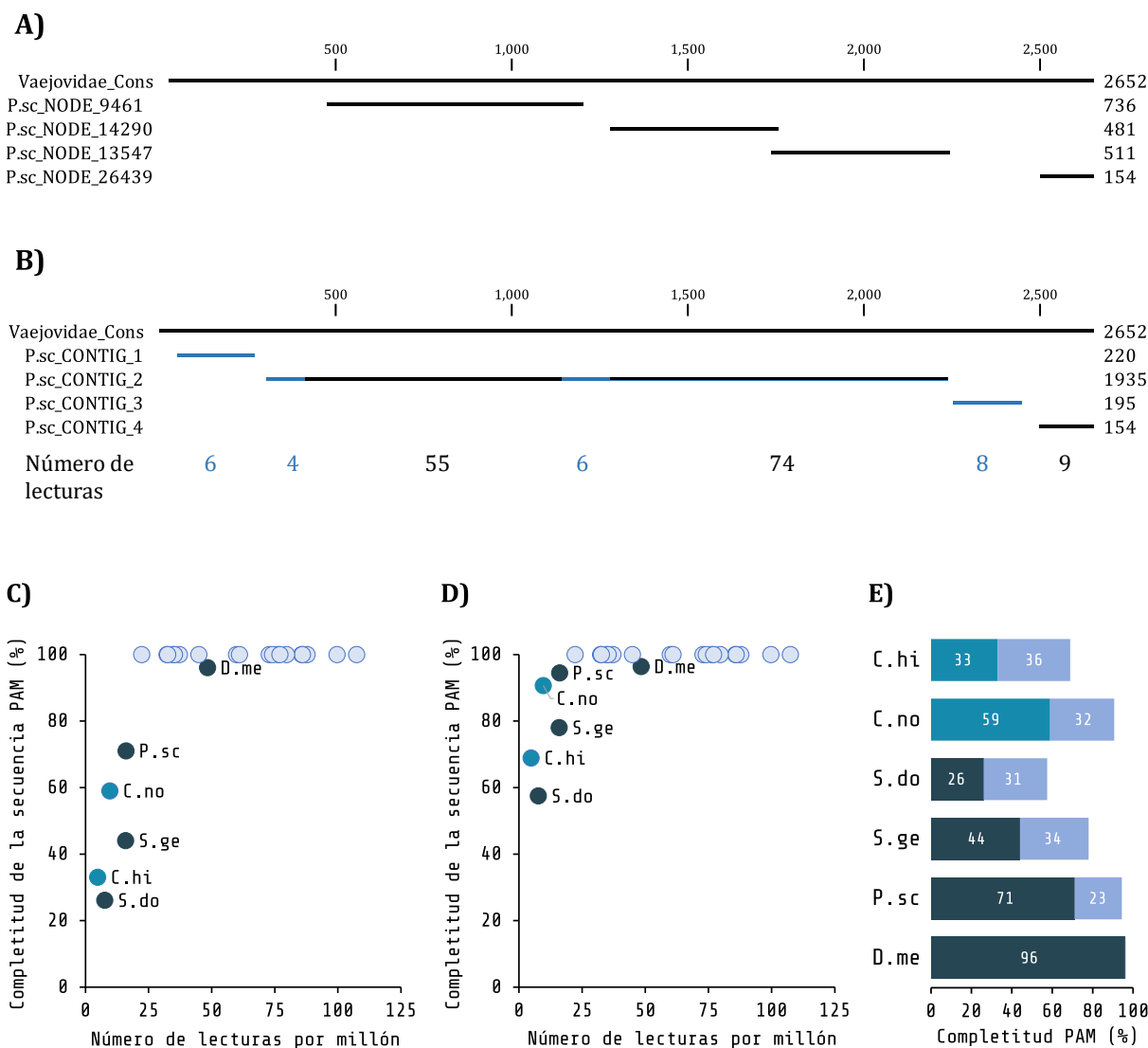


Figura 4. Reconstrucción y extensión manual de secuencias PAM parciales posterior al ensamble *de novo*. **A)** Alineamiento de las secuencias PAM parciales identificadas en el transcriptoma de *Paravaejovis schwenkmeyeri* con una secuencia PAM consenso de especies de la familia Vaejovidae; **B)** Reconstrucción y extensión manual de secuencias PAM mediante el alineamiento de las lecturas de secuenciación; —) Regiones previamente desconocidas; **C)** Relación entre la profundidad de secuenciación y la completitud de secuencias PAM obtenidas a partir de los contigs ensamblados por rnaSPAdes; **D)** Relación entre la profundidad de secuenciación y la completitud de secuencias PAM después del ensamble y extensión manual de contigs; **E)** Porcentaje de completitud inicial de secuencias PAM en Bútidos (■) e Íúridos (■), e incremento porcentual en la completitud debida al ensamble y extensión manual de contigs (■).

La información genómica confirmó que las enzimas monofuncionales (PHM*m* y PAL*m*) están codificadas por genes parálogos y no son resultado del “splicing” alternativo o procesamiento proteolítico; mecanismo reportado en la generación de isoformas individuales en *R. norvegicus* (Eipper *et al.*, 1992a; Spijker *et al.*, 1999). Los transcritos encontrados se traducen a proteínas de entre 861-887 aminoácidos para PAM, y a proteínas más cortas cuya longitud está entre 343-347 aminoácidos para PHM*m* y entre 356-366 aminoácidos para PAL*m* (Figura 5).

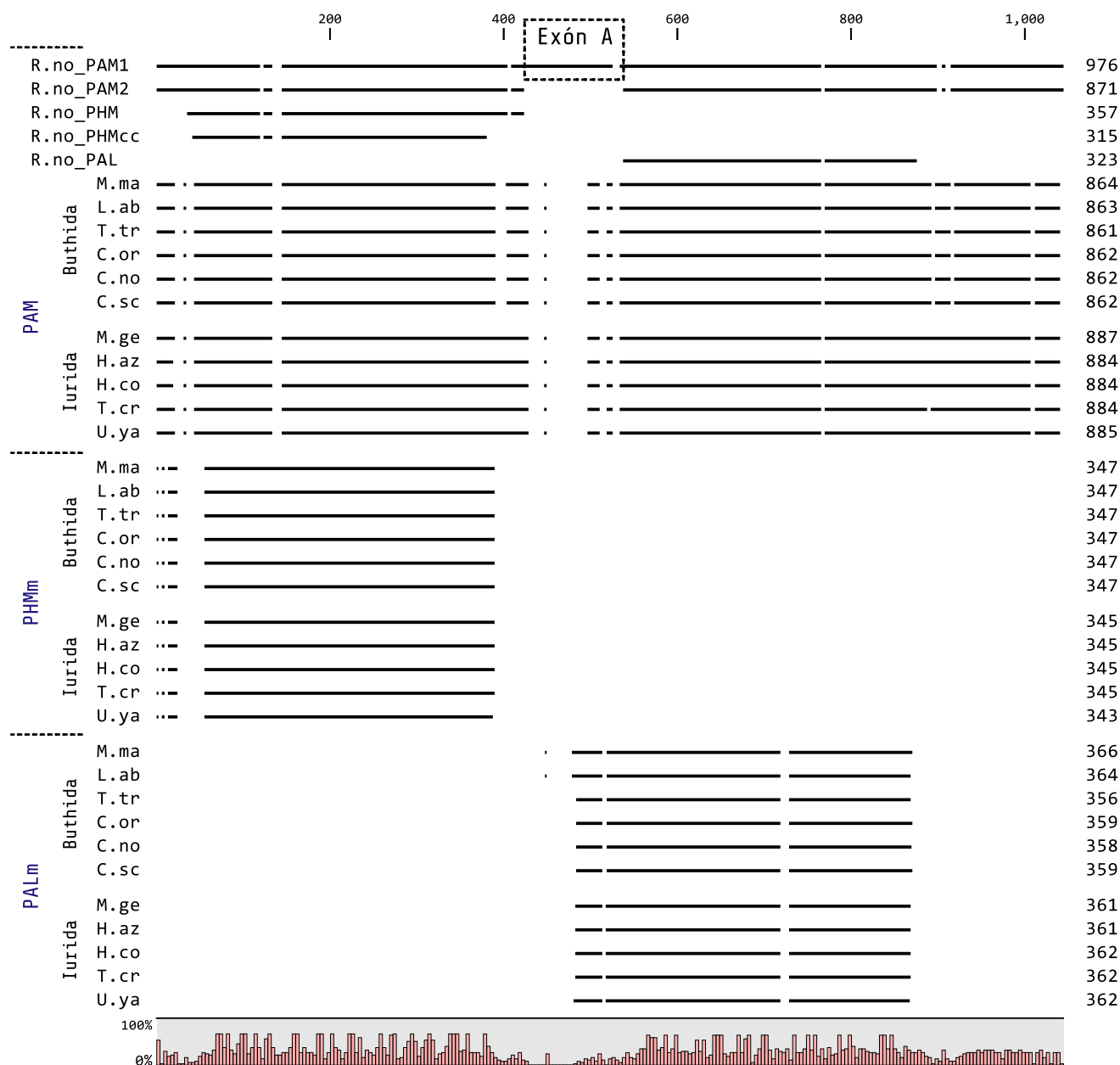
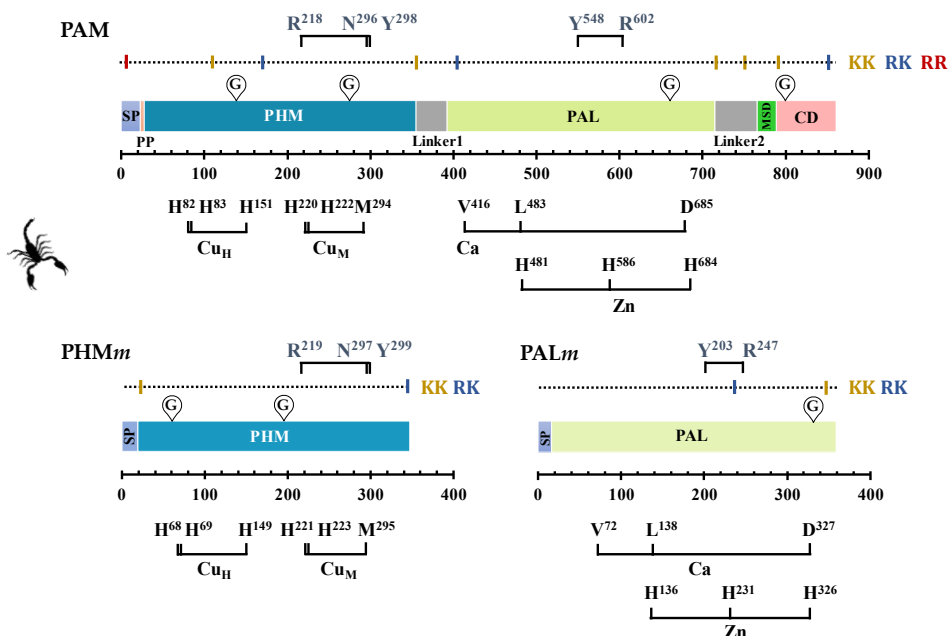


Figura 5. Alineamiento múltiple de secuencias aminoacídicas de las enzimas PAM, PHM*m* y PAL*m*. El alineamiento incluye como referencias a las isoformas PAM1, PAM2, los dominios *phm*-PAM, *pal*-PAM y PHMcc (núcleo catalítico del dominio PHM) de *R. norvegicus*. Se incluyeron secuencias completas de especies representativas de los parvódenes Buthida e Iurida.

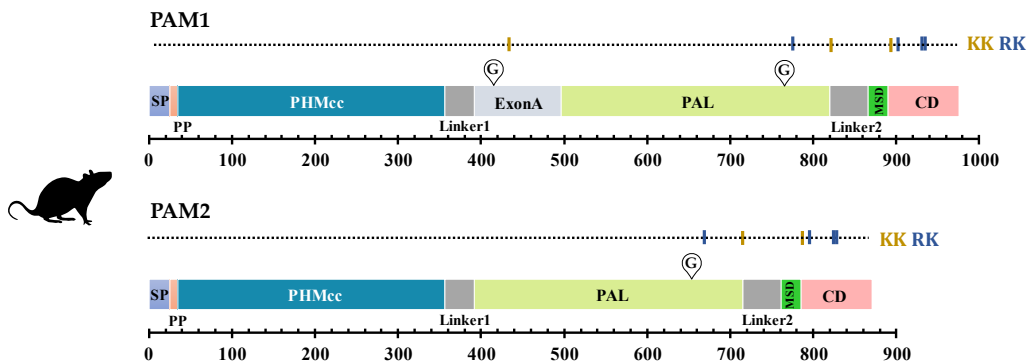
La topología deducida del precursor PAM en los escorpiones es idéntica a la de la isoforma PAM-2 descrita en *R. norvegicus* (Figura 5 y 6A-B). Un péptido señal (SP) seguido por una región propéptido corta (PP), el dominio PHM, una secuencia “linker” (Linker 1), el dominio PAL, una segunda secuencia “linker” (Linker 2), un dominio de anclaje a membrana (MSD) y un dominio citosólico (CD). La isoforma PAM-2 de rata difiere de la isoforma PAM-1 en la ausencia del Linker 1 codificado por el Exón-A (Figura 5). Esta región extra contiene un sitio de endoproteólisis que al ser procesado separa a los dominios PHM y PAL del polipéptido. El Exón-A solo ha sido descrito en vertebrados (Attenborough *et al.*, 2012), y no tiene una región homóloga en las PAM de escorpión (Figura 5). Es notable que, aunque la enzima PAM de escorpión carece de esta región (Exón-A), dos sitios potenciales de endoproteólisis están presentes en la secuencia PAM de escorpión (Figura 6A). El primero de estos sitios, definido por una diada de lisinas (KK), se encuentra entre los dominios PHM y PAL, y se propone que delimita el dominio PHM. El segundo sitio, situado entre la secuencia PAL y el MSD, también está definido por una diada KK, y al ser procesado liberaría a la enzima PAL de los dominios MSD y CD. Por lo tanto, la enzima PAM bifuncional de escorpión podría procesarse para generar enzimas PHM y PAL independientes y solubles, referidas como *phm*-PAM y *pal*-PAM.

La topología de las enzimas monofuncionales guarda similitud con las de las isoformas PHM y PAL-2 de *D. melanogaster* (Figura 6A-C). Las proteínas incluyen un péptido señal y el dominio catalítico. No se identificaron dominios MSD ni CD; por lo tanto, se prevé que las enzimas monofuncionales sean solubles. Los residuos clave implicados en la catálisis y la coordinación de metales se encuentran conservados en ambos sistemas de amidación (Figura 6A), lo que sugiere que estas enzimas son potencialmente funcionales. Dentro de los artrópodos, un sistema dual similar al de los escorpiones está presente en los crustáceos, pero no en los insectos (Zabriskie *et al.*, 1994), los cuales se presume perdieron el gen de la enzima bifuncional PAM y solo conservan los genes de las enzimas monofuncionales *PHMm* y *PALm* (Attenborough *et al.*, 2012).

A) Alacranes *Centruroides noxius*



B) Mamíferos *Rattus norvegicus*



C) Insectos *Drosophila melanogaster*

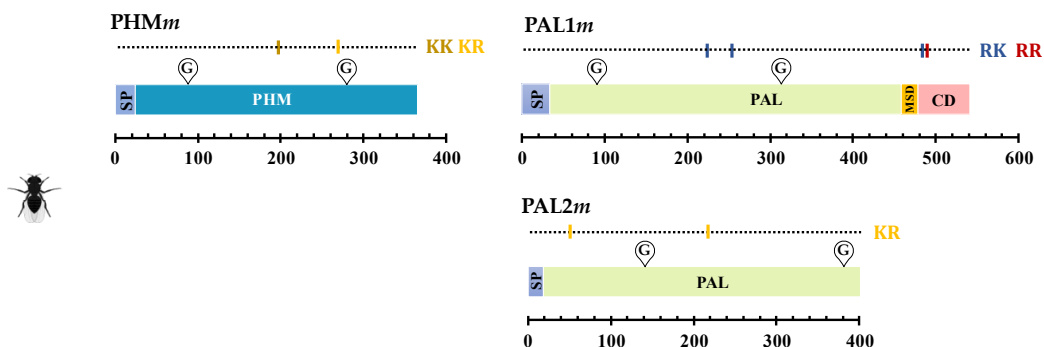


Figura 6. Topología de los precursores del sistema enzimático de amidación.

A) PAM, PHMm y PALm de escorpión (*C. noxius*); **B)** Isoformas PAM-1 y PAM-2 de mamíferos (*R. norvegicus*); **C)** PHMm, PAL1m y PAL2m en *D. melanogaster*. SP, Péptido señal; PP, Propéptido; MSD, Dominio de anclaje a membrana; CD, Dominio Citosólico; KK, RK, KR y RR, diadas de aminoácidos básicos, sitios potenciales de escisión por protein-convertasas; G, sitio potencial de glicosilación; PHMcc, núcleo catalítico del dominio PHM. Las unidades de la escala indican el número de aminoácidos.

El porcentaje de identidad aminoacídica entre los dominios de las enzimas bifuncionales y sus parálogos monofuncionales se muestra en la Tabla 3.

Tabla 3. Porcentaje de identidad entre los dominios catalíticos de las enzimas PAM, PHMm y PALm

Parvorden	Familia	Especie	% de identidad	
			<i>phm</i> -PAM / PHMm	<i>pal</i> -PAM / PALm
Buthida	Buthidae	<i>Mesobuthus martensii</i> *	30.9	31.8
		<i>Leiurus abduhbayrami</i> *	30.3	31.8
		<i>Tityus trivittatus</i>	30.1	32.0
		<i>Centruroides Orizaba</i>	31.0	32.3
		<i>Centruroides noxius</i>	32.1	32.5
		<i>Centruroides sculpturatus</i>	31.0	32.8
Lurida	Euscorpiidae	<i>Megacormus gertschi</i>	26.4	33.7
	Hadruridae	<i>Hoffmannihadrurus aztecus</i>	29.6	34.6
		<i>Hadrurus spadix</i>	30.1	34.5
		<i>Hadrurus concolorus</i>	30.1	34.2
	Vaejovoidea	<i>Thorellius cristimanus</i>	27.7	31.7
Urodacidae	<i>Urodacus yaschenkoi</i> *	29.2	32.1	

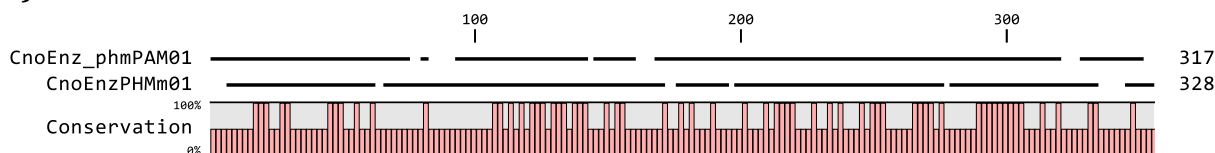
Se muestra información de alacranes representativos cuyos dominios están completos.

El porcentaje de identidad corresponde a las secuencias de aminoácidos.

*Alacranes del viejo mundo.

Por ejemplo, para el alacrán *C. noxius* el porcentaje de identidad entre los subdominios PAM (*phm*-PAM/*pal*-PAM) y los dominios PHMm y PALm son 32.1% y 32.5%, respectivamente (Figura 7).

A)



B)

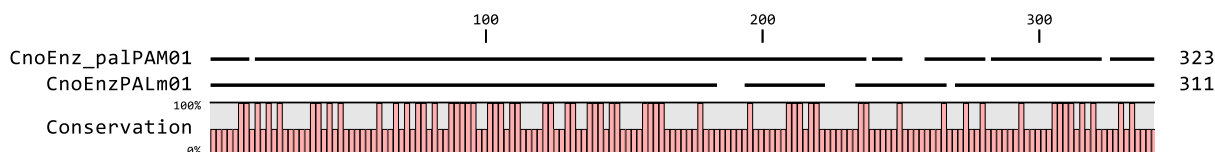


Figura 7. Alineamiento de los dominios catalíticos del sistema enzimático de amidación del alacrán *C. noxius*. **A)** Alineamiento de los dominios *phm*-PAM/PHMm; porcentaje de similitud: 32.1%; **B)** Alineamiento de los dominios *pal*-PAM/PALm; porcentaje de similitud: 32.5%.

6.4 Estructuras tridimensionales de las enzimas amidantes

Las estructuras tridimensionales de las enzimas PAM2 de *R. norvegicus* y PAM del alacrán *C. noxius* predichas por AlphaFold2 muestran la disposición de los dominios catalíticos *phm*-PAM y *pal*-PAM unidos por el linker1, y este complejo catalítico unido al dominio de anclaje a la membrana vesicular/celular (MSD) por el Linker 2 y, por último, el dominio citosólico (CD). Se remarca la localización de los sitios de escisión proteolítica (EPS) que tras ser procesados liberarían las unidades catalíticas (Figura 8A). La disposición de los dominios *phm*-PAM y *pal*-PAM uno enfrente del otro podría facilitar la reacción de amidación en los sustratos reconocidos por el complejo enzimático dada la proximidad de ambos sitios activos, y en el caso de la enzima PAM de alacrán, estos dominios podrían reconocer y modificar otros sustratos cuando ambas subunidades se encuentren separadas (Figura 8B).

Las estructuras de las enzimas monofuncionales, así como los dominios catalíticos PAM muestran un sitio activo conservado accesible mediante un poro, donde la disposición de los residuos que coordinan metales, o que están implicados en la catálisis, están conservados en ambos sistemas enzimáticos (Figuras 6 y 9). La Figura 9 muestra las estructuras de los dominios catalíticos de la enzima PAM y de las enzimas monofuncionales para su comparación. Dado el bajo porcentaje global de similitud entre las secuencias aminoacídicas *phm*-PAM/PHMm y *pal*-PAM/PALm (Tabla 3 y Figura 7), las enzimas monofuncionales podrían reconocer una población adicional o específica de sustratos, teniendo a este sistema enzimático dual de α -amidación actuando en conjunto para la modificación postraducciona de toda la diversidad de péptidos amidados (y posiblemente de otros sustratos de naturaleza no proteica) contenidos en el veneno.

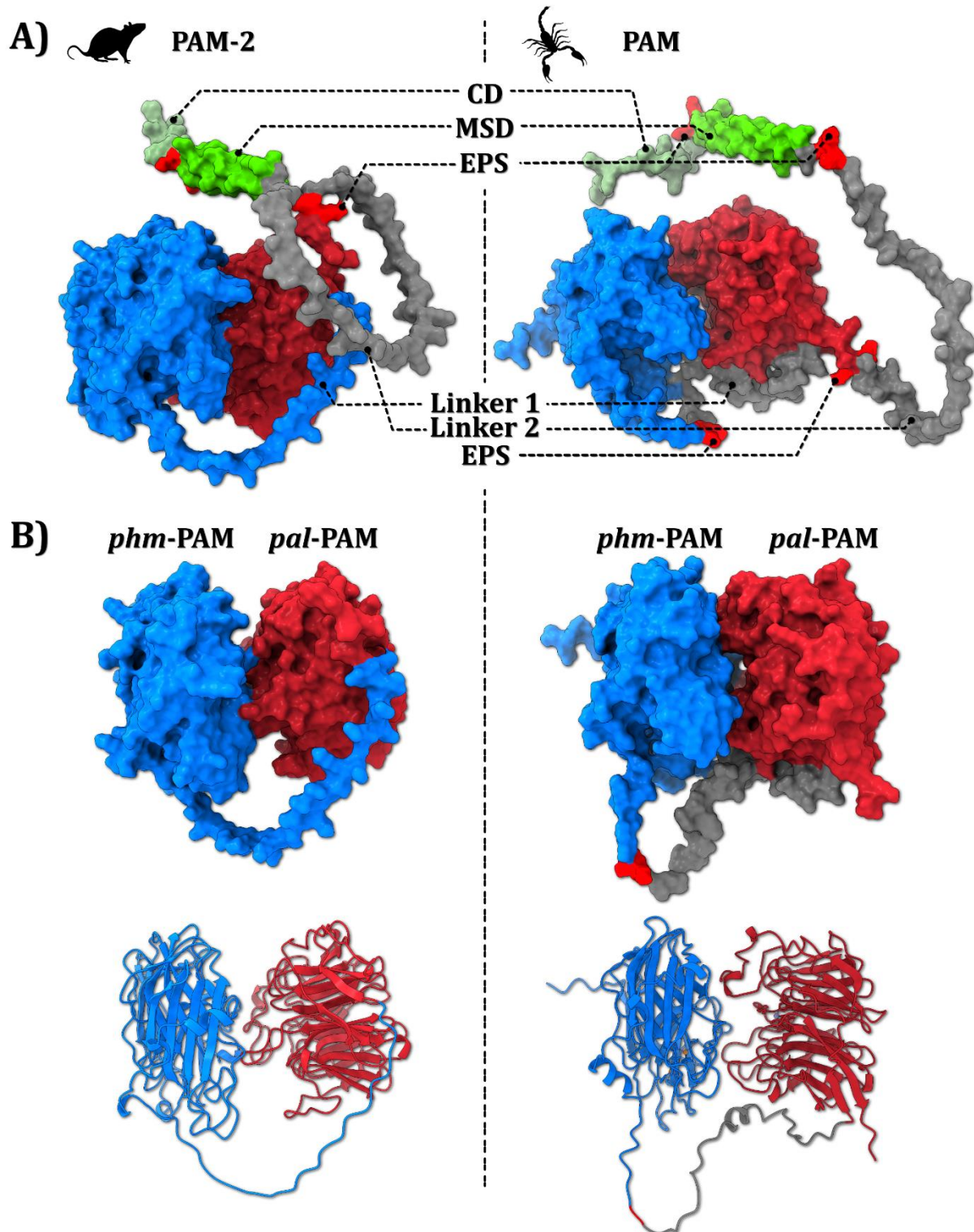


Figura 8. Predicción de la estructura terciaria de la enzima bifuncional PAM en mamíferos y en alacranes. **A)** Estructura tridimensional de superficie de los dominios catalíticos *phm*-PAM, *pal*-PAM, Linker 1 y 2, dominio de anclaje a membrana (MSD), dominio citosólico y de los sitios potenciales de endoproteólisis (EPS) presentes en la isoforma PAM2 de *R. norvegicus* y PAM de *Centruroides noxius*; **B)** Estructura tridimensional de superficie y de listón de los dominios catalíticos PAM solubles (en ausencia del Linker2, del dominio de anclaje a membrana y del dominio citosólico).

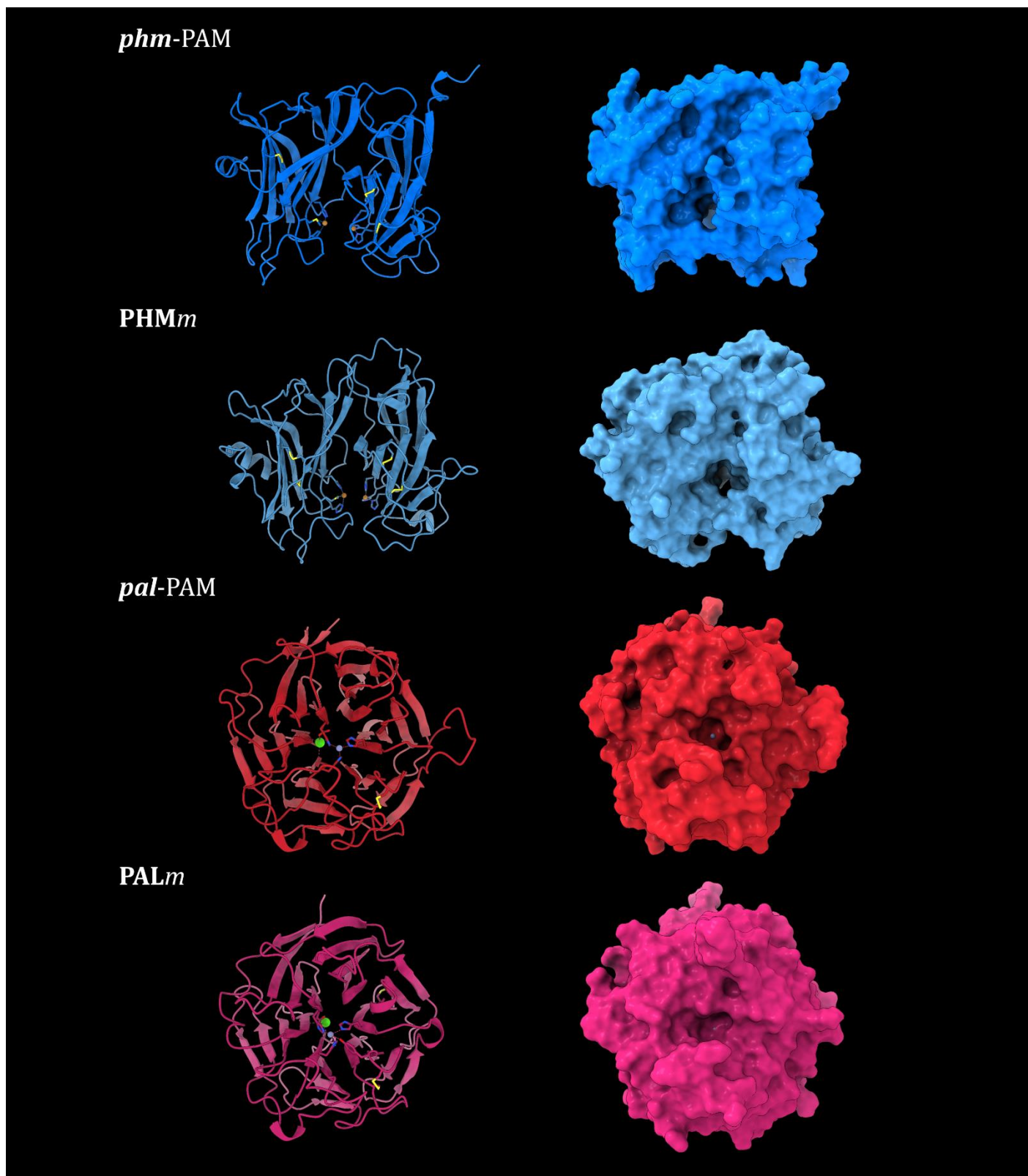


Figura 9. Modelos de las estructuras terciarias de las enzimas *phm*-PAM/**PHM***m* y *pal*-PAM/**PAL***m* del alacrán *C. noxius*. Las estructuras están orientadas de frente a los sitios activos que contienen los residuos que coordinan los metales (2 iones de Cu^{2+} en PHM y 1 ion de Cu^{2+} y 1 de Ca^{2+} en PAL)

6.5 Reconstrucción filogenética de las enzimas amidantes en arácnidos

El análisis de máxima verosimilitud de las secuencias nucleotídicas de los dominios PHM y PAL de la enzima PAM (*phm*-PAM y *pal*-PAM) y las secuencias de las enzimas monofuncionales PHM*m* y PAL*m* de escorpión muestran una correlación entre la filogenia de estas enzimas y la filogenia de las familias taxonómicas de las especies de las que proceden (Santibanez-Lopez *et al.*, 2019a). Las figuras 10 y 11 muestran los dos clados principales en los que las secuencias de *phm*-PAM y *pal*-PAM se separan de las secuencias de sus parálogos PHM*m* y PAL*m* respectivamente. Dentro de esos clados hay una clara divergencia entre las secuencias de especies pertenecientes a la familia Buthidae (parvorden Buthida) y las secuencias de especies pertenecientes a familias del parvorden Iurida (Vaejovidae, Hadruridae, Euscorpidae, Chactidae, Superstitionidae, Diplocentridae, Urodauroe, Scorpionidae). Dentro de la familia Butidae las secuencias de los escorpiones del Viejo Mundo *Leiurus abduallahbayrami* (Turquía) y *Mesobuthus martensii* (países de Asia Oriental) se colocan en un clado independiente que precede al clado con las especies del Nuevo Mundo. Las secuencias de *Tityus trivittatus* (distribuidas en la parte centro-sur de América, Argentina y Brasil) están en clados independientes bien soportados respecto a los del género *Centruroides* (distribuidos en Centroamérica, el Caribe y América del Norte).

Secuencias para un sistema dual de amidación similar al de los escorpiones también se encontraron en otros arácnidos, incluyendo miembros del orden Araneae (*Liphistius malayanus*, *Frontinella communis*, *Parasteatoda tepidariorum*, *Leucauge venusta*), Opiliones (*Siro boyerae*, *Trogulus martensi*), Ricinulei (*Ricinoides atewa*), y el Xiphosura, *Limulus polyphemus* (recientemente colocado dentro de la clase Arachnida) (Ballesteros y Sharma, 2019). Este resultado también se extiende a *Rhagoletis zephyria* (Orden: Díptera), único insecto identificado que aún conserva tanto la enzima PAM como sus parálogos monofuncionales PHM*m* y PAL*m*.

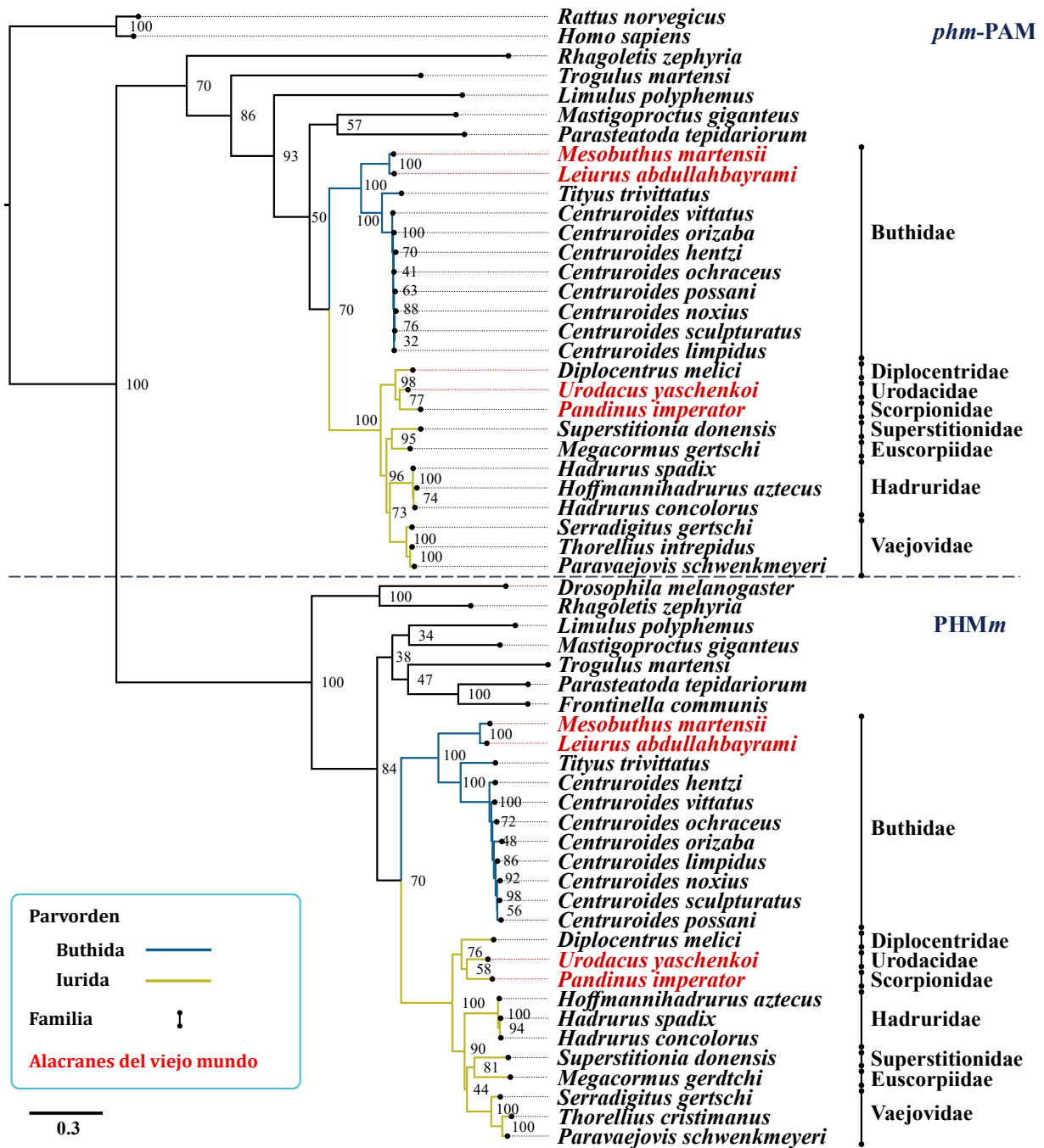


Figura 10. Árbol filogenético de máxima verosimilitud de los dominios *phm*-PAM y PHMm de alacrán. Análisis realizado con las secuencias nucleotídicas de los dominios *phm*-PAM y PHMm. Los números debajo de los nodos indican los valores de ultrafast bootstrap (UFBoot).

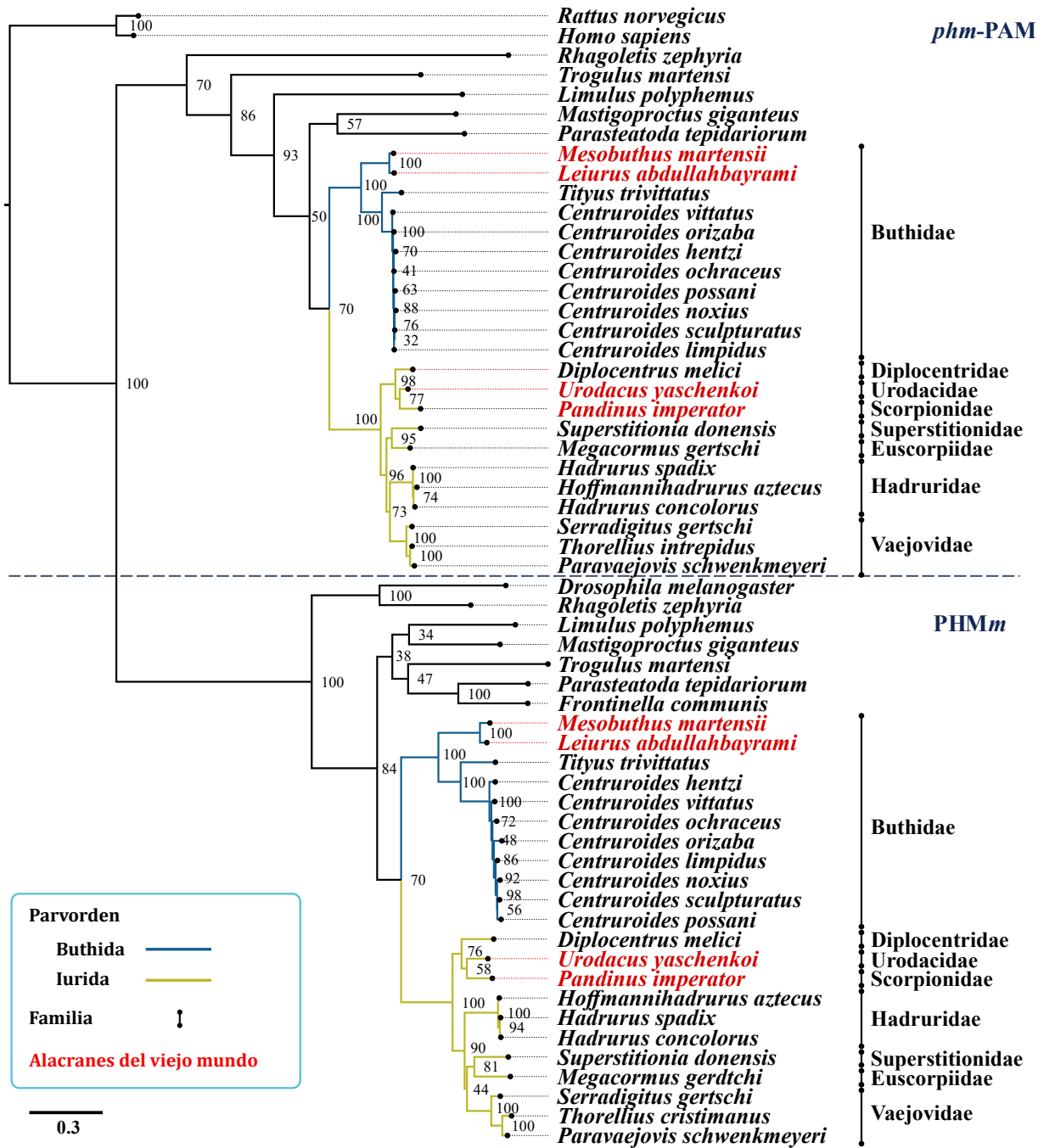


Figura 11. Árbol filogenético de máxima verosimilitud de los dominios *pal*-PAM y PALm. Análisis realizado con las secuencias nucleotídicas de los dominios *pal*-PAM y PALm. El número debajo de los nodos indica los valores de ultrafast bootstrap (UFBoot).

6.6 Niveles de expresión de las enzimas amidantes en la glándula de veneno

La cuantificación de todos los transcritos se normalizó usando las cuentas del gen PGK1 (fosfoglicerato quinasa 1). En el caso particular de las enzimas amidantes, se encontró que los niveles de expresión de la enzima PAM fueron consistentemente bajos (<1 TPM), especialmente entre iúridos, con la excepción de *Hoffmannihadrurus aztecus* (Haz) y *Hadrurus concolorus* (Hco), hadrúridos de gran tamaño en los que se observaron los niveles de expresión más altos para la enzima PAM (Figura 12). Los niveles de expresión de la enzima monofuncional PAL m tuvieron valores apenas mayores a 1 TPM entre iúridos, mientras que entre bütidos los niveles de expresión se encuentran entre 1 y 5 TPM. La enzima PHM m mostró los niveles de expresión más altos de las tres enzimas y sus niveles de expresión no normalizados son parecidos a los reportados previamente por (Ward *et al.*, 2018) (Apéndice A3). Los niveles de expresión de estas enzimas son consistentes entre los alacranes de cada parvorden (Figura 12 y Apéndice A4) y la normalización de sus valores redujo las diferencias en sus profundidades de secuenciación (Apéndice A5). El análisis estadístico mostró que existen diferencias significativas en los niveles de expresión de las enzimas PAM ($p = 0.0133$), PHM m ($p = 0.0246$) y PAL m ($p < 0.0001$) entre bütidos e iúridos (prueba U Mann-Whitney), siendo los bütidos los que tuvieron los mayores niveles de expresión para las tres enzimas (Figura 12 y Apéndice A5). Este resultado es consistente con el mayor número de péptidos amidados reportados en los venenos de los alacranes del parvorden Buthida (Ver sección 6.7).

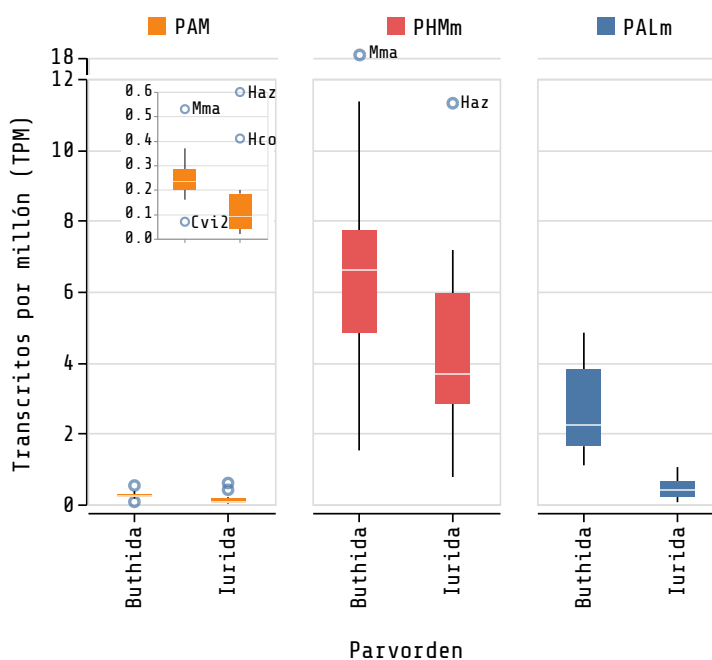


Figura 12. Niveles normalizados de expresión de las enzimas PAM, PHM m y PAL m en las glándulas de veneno. Se observan los niveles de expresión los alacranes agrupados por parvorden y por enzima

6.7 Diversidad de péptidos amidados

La identificación de secuencias no redundantes de los componentes peptídicos en el veneno mostró la mayor abundancia de toxinas de sodio (neurotoxinas) (NaTx) y de potasio (KTx) que caracteriza a los alacranes del parvorden Buthida. En el caso de los íuridos, estos contienen una mayor abundancia de otras toxinas menos representadas como calcinas, liotoxinas, omegaescorpinas y TRPs (Figura 13A). En cuanto a los péptidos de defensa del huésped (HDP), en íuridos se encontró una mayor abundancia de péptidos sin puentes disulfuro (NDBPs) de la familia 4 (péptidos antimicrobianos). Señales de amidación se identificaron en toxinas de sodio (α -NaTx y β -NaTx), potasio (familias α -KTx, δ -KTx, ε -KTx, κ -KTx, y λ -KTx), calcio (liotoxinas y omegaescorpinas) y en toxinas de cloro (α -ClTx). Dentro de los péptidos de defensa del huésped se identificaron señales de amidación en defensinas, en waprinas de íuridos y en todos los péptidos sin puentes disulfuro de la familia 4 (NDBP4) (Figura 13B).

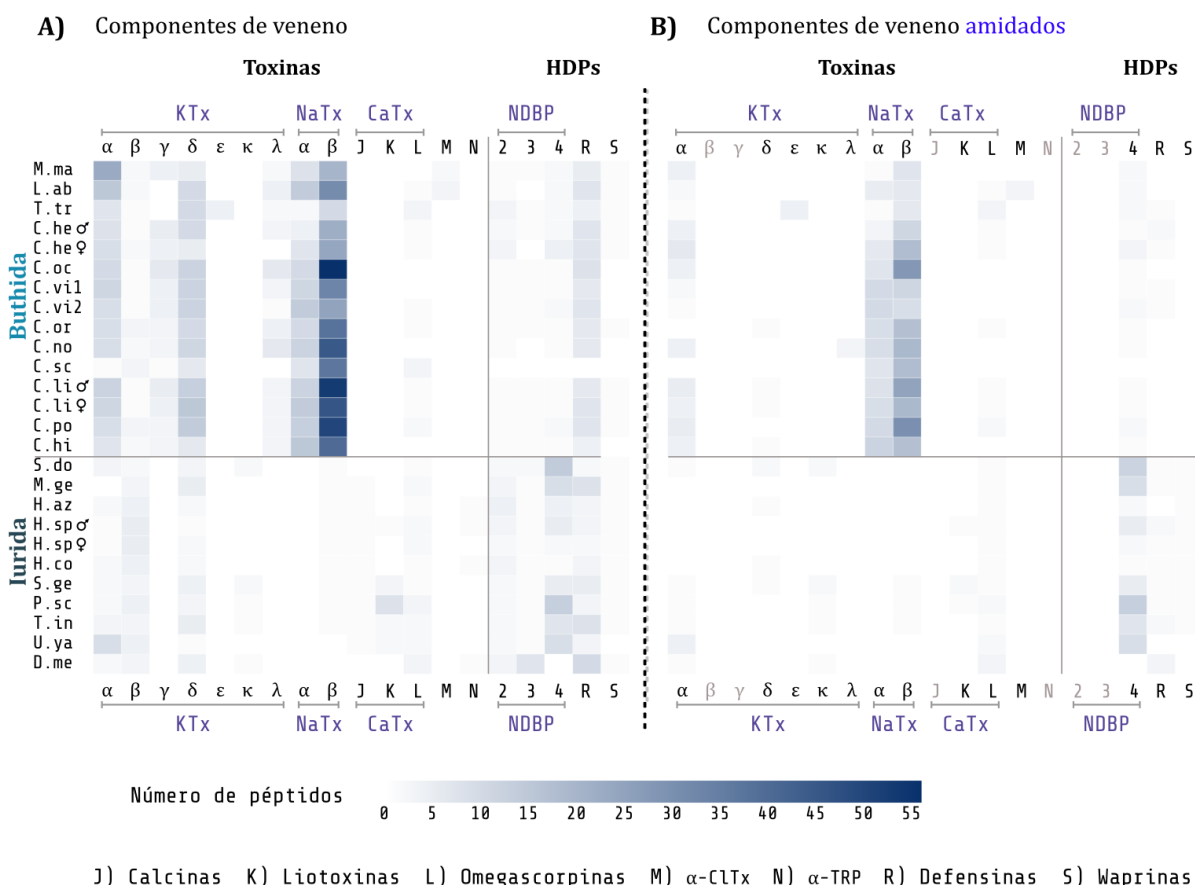


Figura 13. Diversidad y abundancia de componentes peptídicos de veneno.

A) Número de péptidos de cada una de las familias y subfamilias de toxinas y HDP; **B)** Número de péptidos en los que se identificaron señales de amidación

Los modelos de las estructuras terciarias de las cadenas maduras de péptidos amidados representativos (toxinas y HDP) se muestran en la Figura 14. En los modelos se observa la compleja diversidad de sustratos potenciales para el doble sistema enzimático de amidación en las glándulas de veneno. Se resalta el residuo de glicina en el C-terminal sobre el que ocurren las reacciones secuenciales catalizadas por las enzimas PHM y PAL que dan paso a la obtención del péptido amidado. Entre los sustratos identificados se encuentran péptidos cortos como los NDBP4 con una longitud de 18 aminoácidos y péptidos como las toxinas de sodio, cuya longitud esta entre 58-76 aminoácidos (Cid-Uribe *et al.*, 2020). En comparación, las hormonas peptídicas en los humanos varían considerablemente en longitud, desde 3 aminoácidos para la hormona liberadora de tirotrópina (TRH) hasta 52 aminoácidos para adrenomedulina (Merkler *et al.*, 2021).

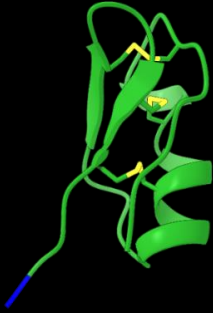
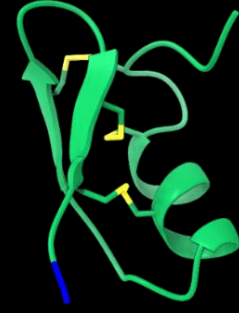
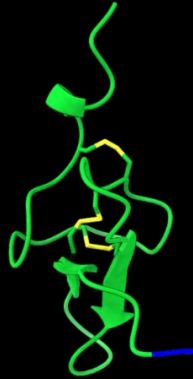
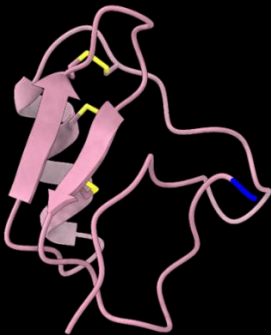
α -KTx (CnoKTxAlp01) **δ -KTx** (CliKTxDel01) **ϵ -KTx** (TtrKTxEps01) **κ -KTx** (PscKTxKap01) **λ -KTx** (CnoKTxLam01) **α -ClTx** (LabClTxAlp01) **α -NaTx** (CnoNaTxAlp01) **β -NaTx** (CnoNaTxBet02 (Cn2))**Defensina** (CnoHDPDef01)**Liotoxina** (HspCaTxLio01)**Omegaescorpina** (CliCaTxOme01)**NDBP4** (CnHDPNDBP401)

Figura 14. Estructuras representativas de los sustratos peptídicos del sistema enzimático dual de amidación. Se muestran los nombres de los componentes peptídicos, número de aminoácidos y peso molecular. CnoKTxAlp01 (39 aa; 4.30 kDa), CliKTxDel01 (63 aa; 7.07 kDa), TtrKTxEps01 (24 aa; 2.66 kDa), PscKTxKap01 (24 aa; 2.66 kDa), CnoKTxLam01 (42 aa; 4.69 kDa), LabClTxAlp01 (37 aa; 4.06 kDa), CnoNaTxAlp01 (55 aa; 6.29 kDa), CnoNaTxBet02 (67 aa; 7.66 kDa), CnoHDPDef01 (37 aa; 4.22 kDa), HspCaTxLio01 (35 aa; 3.93 kDa), CliCaTxOme01 (43 aa; 4.96 kDa); CnHDPNDBP401 (18 aa; 1.82 kDa); —) Residuo de glicina en el C-terminal de los péptidos

7. Discusión

Análisis transcriptómicos y proteómicos de las glándulas de veneno de alacrán han demostrado la enorme diversidad de compuestos presentes en este fluido (Cid-Uribe *et al.*, 2020). La información bioquímica disponible sobre componentes de veneno y las secuencias de los precursores peptídicos inferidos por análisis transcriptómicos señalan a la amidación como una de las MPT más frecuentes en los péptidos de veneno de estos emblemáticos arácnidos. La descripción de señales canónicas de amidación en los precursores sugería que un sistema conservado de α -amidación estaría presente en las glándulas de veneno de los alacranes. La identificación de genes PAM, PHM m y PAL m en el genoma del alacrán *Centruroides sculpturatus* mostró que las enzimas monofuncionales no son el resultado de “splicing” alternativo ni del procesamiento proteolítico de la enzima PAM, tal y como ocurre en mamíferos. La presencia de parálogos involucrados en diversos procesos del desarrollo y funciones celulares dentro de los órdenes Scorpiones y Araneae se sabe que ha sido consecuencia de la duplicación completa del genoma en el ancestro común de escorpiones y arañas (Leite *et al.*, 2018; Schwager *et al.*, 2017). Sin embargo, análisis filogenómicos sugieren que la duplicación del gen ancestral PAM que pudo dar paso a los genes parálogos que codifican para PHM m y PAL m encontrados en algunos órdenes podría remontarse hasta de la separación de Cnidarios y Bilaterios (Attenborough *et al.*, 2012). La identificación de transcritos para las tres enzimas sugiere que éstas podrían expresarse conjuntamente en la glándula de veneno. Sin embargo, el análisis de los niveles de expresión de PAM y PAL m fueron, en comparación con los PHM m , consistentemente bajos en todos los alacranes, especialmente entre miembros del parvorden Iurida, los cuales también tienen un número menor de componentes peptídicos amidados. Análisis proteómicos han confirmado la presencia de enzimas PHM m en los venenos de *Tityus obscurus* y *Centruroides hentzi* (de Oliveira *et al.*, 2018; Ward *et al.*, 2018). En concordancia con estos resultados, también se identificaron secuencias PHM m en análisis proteómicos de los venenos de *Centruroides limpidus*, *Centruroides hirsutipalpus* (bútidos) y *Superstitionia donensis* (iúrido) (Datos no mostrados), sin encontrar coincidencias con secuencias PAL. Los péptidos con hidroxiglicina en su C-terminal no se han detectado *in vivo* (Yin *et al.*, 2011), lo que indica que la escisión del enlace N-C de la hidroxiglicina debe llevarse a cabo (Figura 1). Guardando una correlación con sus bajos niveles de expresión, es posible que la enzima PAL m se encuentre en una muy baja proporción en el veneno dificultando su detección, esto debido probablemente a una mayor eficiencia catalítica; en ensayos enzimáticos se han demostrado que el dominio *pal*-PAM de *R. norvegicus* tiene una k_{cat} hasta 10 veces mayor que el dominio *phm*-PAM (sustrato: α -Nacetyl-Tyr-Val-Gly), indicando que el dominio PHM es quien limita la formación de péptidos amidados

(Husten *et al.*, 1993). En el caso de la enzima bifuncional PAM, los dominios transmembranal y citosólico desempeñan un papel esencial en el tráfico de esta enzima por la vía exocítica y endocítica. La degradación del dominio citosólico provoca la acumulación de la enzima PAM en la membrana plasmática (Milgram *et al.*, 1993) y en cilios primarios de células epiteliales (estructuras sensoriales y de señalización críticas, importantes para detectar y responder a los estímulos ambientales); mientras que la fosforilación del dominio citosólico controla la entrada de la enzima a endosomas que pueden devolver la enzima a la red trans-Golgi y a los gránulos secretorios o degradarse en lisosomas (Kumar *et al.*, 2016). El recirculamiento de la enzima PAM supone un método eficiente para la amidación de péptidos durante el tráfico vesicular, reduciendo quizá la síntesis *de novo* de esta enzima, lo cual estaría apoyado por los bajos niveles de expresión consistentes en todos los alacranes analizados. Las enzimas monofuncionales, al carecer de un dominio transmembranal y citosólico, tienen un tráfico unidireccional a través de la vía exocítica. Esto, en principio, favorece su acumulación en el veneno en función de sus niveles de expresión. No se conocen las propiedades catalíticas de ninguna de las enzimas amidantes de alacrán o de ningún otro arácnido (ver Apéndice 6). Es posible que estas enzimas hayan evolucionado para reconocer sustratos específicos o ser activas en determinadas condiciones fisiológicas. Aunque no está claro qué funciones adicionales podrían desempeñar en el veneno, se sabe que la enzima PAM bovina es capaz de catalizar tres reacciones *in vitro* alternativas a la α -amidación: sulfoxidación, N-dealquilación de aminas y O-dealquilación (Katopodis y May, 1990). Esto abre la posibilidad para encontrar nuevos sustratos naturales para este conjunto de enzimas y aprovechar sus capacidades catalíticas para la síntesis o modificación química de moléculas de interés médico y biotecnológico.

8. Conclusiones

- Se describió un doble sistema enzimático de α -amidación en las glándulas de veneno de los alacranes, conformado por la proteína bifuncional de membrana peptidil-glicina α -amidante monooxigenasa (PAM) y de las enzimas monofuncionales peptidilglicina α -hidroxilante monooxigenasa (PHMm) y peptidil- α -hidroxiglicin α -amidante liasa (PALm), codificadas por genes independientes.
- Se encontraron ortólogos de las enzimas PC1, PC2 y carboxipeptidasas E (CPE) que actúan río arriba de la reacción de amidación.
- No se encontró evidencia de isoformas generadas por “splicing” alternativo en las enzimas PAM, PHMm y PALm expresadas en las glándulas de veneno, ni en datos transcriptómicos, ni por RT-PCR.
- Los niveles de expresión de las enzimas PAM, PHMm y PALm en las glándulas de veneno son significativamente diferentes entre los alacranes de las dos ramas basales del orden Scorpiones (Buthida e Iurida), siendo los miembros del parvorden Buthida los alacranes con los niveles de expresión más altos y con el mayor número de componentes de veneno amidados.
- La enzima PHMm mostró los niveles de expresión más altos en comparación con las otras dos enzimas (PALm y PAM) y es la única que hasta el momento se ha confirmado por proteómica en el veneno de 4 bítidos (*C. hentzi*, *C. limpidus*, *C. hirsutipalpus*, *Tityus obscurus*) y en 1 iúrido (*Superstitionia donensis*)

9. Perspectivas

- Establecer un modelo de expresión para la obtención de enzimas amidantes
- Expresar y caracterizar al menos una enzima amidante de un alacrán representativo
- Comparar las características catalíticas del doble sistema de α -amidación presente en los alacranes
- Comparar las características catalíticas de las enzimas de alacrán con las de una enzima de mamífero

En el Apéndice A3 se muestran los avances obtenidos en el proceso de expresión de las enzimas amidantes que se propone como perspectiva.

Los resultados de este trabajo fueron publicados en la revista “Toxins” (Factor de impacto 3.531, cuartil Q1).

El artículo completo se incluye como anexo al final de este documento.





toxins



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



Referencias

- Abdel-Rahman, M. A., Quintero-Hernandez, V., & Possani, L. D. **2013**. Venom proteomic and venomous glands transcriptomic analysis of the Egyptian scorpion *Scorpio maurus palmatus* (Arachnida: Scorpionidae). *Toxicon*, *74*, 193-207. doi:10.1016/j.toxicon.2013.08.064
- Andrews, S. **2010**. FastQC: a quality control tool for high throughput sequence data.
- Attenborough, R. M., Hayward, D. C., Kitahara, M. V., Miller, D. J., & Ball, E. E. **2012**. A "neural" enzyme in nonbilaterian animals and algae: preneural origins for peptidylglycine alpha-amidating monooxygenase. *Molecular Biology and Evolution*, *29*(10), 3095-3109. doi:10.1093/molbev/mss114
- Ballesteros, J. A., & Sharma, P. P. **2019**. A Critical Appraisal of the Placement of Xiphosura (Chelicerata) with Account of Known Sources of Phylogenetic Error. *Systematic Biology*, *68*(6), 896-917. doi:10.1093/sysbio/syz011
- Becerril, B., Corona, M., García, C., Bolívar, F., & Possani, L. D. **1995**. Cloning of Genes Encoding Scorpion Toxins: An Interpretative Review. *Journal of Toxicology: Toxin Reviews*, *14*(3), 339-357. doi:10.3109/15569549509019468
- Bougis, P. E., Rochat, H., & Smith, L. A. **1989**. Precursors of *Androctonus australis* scorpion neurotoxins. Structures of precursors, processing outcomes, and expression of a functional recombinant toxin II. *Journal of Biological Chemistry*, *264*(32), 19259-19265.
- Buchfink, B., Reuter, K., & Drost, H. G. **2021**. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods*, *18*(4), 366-368. doi:10.1038/s41592-021-01101-x
- Bushmanova, E., Antipov, D., Lapidus, A., & Pribelski, A. D. **2019**. rnaSPAdes: a de novo transcriptome assembler and its application to RNA-Seq data. *Gigascience*, *8*(9). doi:10.1093/gigascience/giz100
- Chen, S., Zhou, Y., Chen, Y., & Gu, J. **2018**. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, *34*(17), i884-i890. doi:10.1093/bioinformatics/bty560
- Chippaux, J. P. **2012**. Emerging options for the management of scorpion stings. *Drug Design, Development and Therapy*, *6*, 165-173. doi:10.2147/DDDT.S24754
- Cid-Urbe, J. I., Meneses, E. P., Batista, C. V. F., Ortiz, E., & Possani, L. D. **2019**. Dissecting Toxicity: The Venom Gland Transcriptome and the Venom Proteome of the Highly Venomous Scorpion *Centruroides limpidus* (Karsch, 1879). *Toxins*, *11*(5). doi:10.3390/toxins11050247
- Cid-Urbe, J. I., Santibanez-Lopez, C. E., Meneses, E. P., Batista, C. V. F., Jimenez-Vargas, J. M., Ortiz, E., & Possani, L. D. **2018**. The diversity of venom components of the scorpion species *Paravaejovis schwenkmeyeri* (Scorpiones: Vaejovidae) revealed by transcriptome and proteome analyses. *Toxicon*, *151*, 47-62. doi:10.1016/j.toxicon.2018.06.085
- Cid-Urbe, J. I., Veytia-Bucheli, J. I., Romero-Gutierrez, T., Ortiz, E., & Possani, L. D. **2020**. Scorpion venomics: a 2019 overview. *Expert Rev Proteomics*, *17*(1), 67-83. doi:10.1080/14789450.2020.1705158
- Coelho, V. A., Cremonez, C. M., Anjolette, F. A., Aguiar, J. F., Varanda, W. A., & Arantes, E. C. **2014**. Functional and structural study comparing the C-terminal amidated beta-neurotoxin Ts1 with its isoform Ts1-G isolated from *Tityus serrulatus* venom. *Toxicon*, *83*, 15-21. doi:10.1016/j.toxicon.2014.02.010

- Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. **2021**. Twelve years of SAMtools and BCFtools. *Gigascience*, *10*(2). doi:10.1093/gigascience/giab008
- 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. **2018**. Proteomic endorsed transcriptomic profiles of venom glands from *Tityus obscurus* and *T. serrulatus* scorpions. *PLoS One*, *13*(3), e0193739. doi:10.1371/journal.pone.0193739
- Eipper, B. A., Green, C. B., Campbell, T. A., Stoffers, D. A., Keutmann, H. T., Mains, R. E., & Ouafik, L. **1992a**. Alternative splicing and endoproteolytic processing generate tissue-specific forms of pituitary peptidylglycine alpha-amidating monooxygenase (PAM). *Journal of Biological Chemistry*, *267*(6), 4008-4015.
- Eipper, B. A., Milgram, S. L., Husten, E. J., Yun, H. Y., & Mains, R. E. **1993**. Peptidylglycine alpha-amidating monooxygenase: a multifunctional protein with catalytic, processing, and routing domains. *Protein Science*, *2*(4), 489-497. doi:10.1002/pro.5560020401
- Eipper, B. A., Park, L. P., Dickerson, I. M., Keutmann, H. T., Thiele, E. A., Rodriguez, H., Schofield, P. R., & Mains, R. E. **1987**. Structure of the precursor to an enzyme mediating COOH-terminal amidation in peptide biosynthesis. *Molecular Endocrinology*, *1*(11), 777-790. doi:10.1210/mend-1-11-777
- Eipper, B. A., Perkins, S. N., Husten, E. J., Johnson, R. C., Keutmann, H. T., & Mains, R. E. **1991**. Peptidyl-alpha-hydroxyglycine alpha-amidating lyase. Purification, characterization, and expression. *Journal of Biological Chemistry*, *266*(12), 7827-7833.
- Eipper, B. A., Stoffers, D. A., & Mains, R. E. **1992b**. The biosynthesis of neuropeptides: peptide alpha-amidation. *Annual Review of Neuroscience*, *15*, 57-85. doi:10.1146/annurev.ne.15.030192.000421
- Estrada, G., Restano-Cassulini, R., Ortiz, E., Possani, L. D., & Corzo, G. **2011**. Addition of positive charges at the C-terminal peptide region of CstII, a mammalian scorpion peptide toxin, improves its affinity for sodium channels Nav1.6. *Peptides*, *32*(1), 75-79. doi:10.1016/j.peptides.2010.11.001
- Fan, X., Spijker, S., Akalal, D. B., & Nagle, G. T. **2000**. Neuropeptide amidation: cloning of a bifunctional alpha-amidating enzyme from *Aplysia*. *Brain Research: Molecular Brain Research*, *82*(1-2), 25-34. doi:10.1016/s0169-328x(00)00173-x
- Fricker, L. D. **2005**. Neuropeptide-processing enzymes: applications for drug discovery. *AAPS J*, *7*(2), E449-455. doi:10.1208/aapsj070244
- Fu, L., Niu, B., Zhu, Z., Wu, S., & Li, W. **2012**. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, *28*(23), 3150-3152. doi:10.1093/bioinformatics/bts565
- 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., Pohl, J., Kubanek, J., & McCarty, N. A. **2007**. State-dependent inhibition of cystic fibrosis transmembrane conductance regulator chloride channels by a novel peptide toxin. *Journal of Biological Chemistry*, *282*(52), 37545-37555. doi:10.1074/jbc.M708079200
- Geer, L. Y., Domrachev, M., Lipman, D. J., & Bryant, S. H. **2002**. CDART: protein homology by domain architecture. *Genome Research*, *12*(10), 1619-1623. doi:10.1101/gr.278202

- Glauder, J., Ragg, H., Rauch, J., & Engels, J. W. **1990**. Human peptidylglycine alpha-amidating monooxygenase: cDNA, cloning and functional expression of a truncated form in COS cells. *Biochemical and Biophysical Research Communications*, *169*(2), 551-558. doi:10.1016/0006-291x(90)90366-u
- Gurevitz, M. **2012**. Mapping of scorpion toxin receptor sites at voltage-gated sodium channels. *Toxicon*, *60*(4), 502-511. doi:10.1016/j.toxicon.2012.03.022
- Gutte, B. **1995**. *Peptides: Synthesis, Structures, and Applications* (1st edition ed.). San Diego, California: Academic Press.
- Han, M., Park, D., Vanderzalm, P. J., Mains, R. E., Eipper, B. A., & Taghert, P. H. **2004**. *Drosophila* uses two distinct neuropeptide amidating enzymes, dPAL1 and dPAL2. *Journal of Neurochemistry*, *90*(1), 129-141. doi:10.1111/j.1471-4159.2004.02464.x
- Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. **2018**. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution*, *35*(2), 518-522. doi:10.1093/molbev/msx281
- Husten, E. J., Tausk, F. A., Keutmann, H. T., & Eipper, B. A. **1993**. Use of endoproteases to identify catalytic domains, linker regions, and functional interactions in soluble peptidylglycine alpha-amidating monooxygenase. *Journal of Biological Chemistry*, *268*(13), 9709-9717.
- Jekely, G. **2013**. Global view of the evolution and diversity of metazoan neuropeptide signaling. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(21), 8702-8707. doi:10.1073/pnas.1221833110
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., & Hassabis, D. **2021**. Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583-589. doi:10.1038/s41586-021-03819-2
- Jungo, F., Bougueleret, L., Xenarios, I., & Poux, S. **2012**. The UniProtKB/Swiss-Prot Tox-Prot program: A central hub of integrated venom protein data. *Toxicon*, *60*(4), 551-557. doi:10.1016/j.toxicon.2012.03.010
- Kall, L., Krogh, A., & Sonnhammer, E. L. **2004**. A combined transmembrane topology and signal peptide prediction method. *Journal of Molecular Biology*, *338*(5), 1027-1036. doi:10.1016/j.jmb.2004.03.016
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermin, L. S. **2017**. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*, *14*(6), 587-589. doi:10.1038/nmeth.4285
- Kapustin, Y., Souvorov, A., Tatusova, T., & Lipman, D. **2008**. Splign: algorithms for computing spliced alignments with identification of paralogs. *Biology Direct*, *3*, 20. doi:10.1186/1745-6150-3-20
- Katoh, K., Rozewicki, J., & Yamada, K. D. **2019**. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Briefings in Bioinformatics*, *20*(4), 1160-1166. doi:10.1093/bib/bbx108
- Katopodis, A. G., & May, S. W. **1990**. Novel substrates and inhibitors of peptidylglycine alpha-amidating monooxygenase. *Biochemistry*, *29*(19), 4541-4548. doi:10.1021/bi00471a006

- Kolhekar, A. S., Roberts, M. S., Jiang, N., Johnson, R. C., Mains, R. E., Eipper, B. A., & Taghert, P. H. **1997**. Neuropeptide amidation in *Drosophila*: separate genes encode the two enzymes catalyzing amidation. *Journal of Neuroscience*, *17*(4), 1363-1376. doi:10.1523/JNEUROSCI.17-04-01363.1997
- Kumar, D., Mains, R. E., & Eipper, B. A. **2016**. 60 YEARS OF POMC: From POMC and alpha-MSH to PAM, molecular oxygen, copper, and vitamin C. *Journal of Molecular Endocrinology*, *56*(4), T63-76. doi:10.1530/JME-15-0266
- Langmead, B., & Salzberg, S. L. **2012**. Fast gapped-read alignment with Bowtie 2. *Nat Methods*, *9*(4), 357-359. doi:10.1038/nmeth.1923
- 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., & McGregor, A. P. **2018**. Homeobox Gene Duplication and Divergence in Arachnids. *Molecular Biology and Evolution*, *35*(9), 2240-2253. doi:10.1093/molbev/msy125
- Liu, Z., Yang, G., Li, B., Chi, C., & Wu, X. **2003**. Cloning, co-expression with an amidating enzyme, and activity of the scorpion toxin BmK ITa1 cDNA in insect cells. *Molecular Biotechnology*, *24*(1), 21-26. doi:10.1385/MB:24:1:21
- Luna-Ramirez, K., Quintero-Hernandez, V., Juarez-Gonzalez, V. R., & Possani, L. D. **2015**. Whole Transcriptome of the Venom Gland from *Urodacus yaschenkoi* Scorpion. *PLoS One*, *10*(5), e0127883. doi:10.1371/journal.pone.0127883
- McElroy, T., McReynolds, C. N., Gullledge, A., Knight, K. R., Smith, W. E., & Albrecht, E. A. **2017**. Differential toxicity and venom gland gene expression in *Centruroides vittatus*. *PLoS One*, *12*(10), e0184695. doi:10.1371/journal.pone.0184695
- Merkler, D., Hawley, A., Eipper, B., & Mains, R. **2021**. doi:10.22541/au.162749074.40639174/v1
- Merkler, D. J. **1994**. C-terminal amidated peptides: production by the in vitro enzymatic amidation of glycine-extended peptides and the importance of the amide to bioactivity. *Enzyme and Microbial Technology*, *16*(6), 450-456. doi:10.1016/0141-0229(94)90014-0
- Milgram, S. L., Mains, R. E., & Eipper, B. A. **1993**. COOH-terminal signals mediate the trafficking of a peptide processing enzyme in endocrine cells. *Journal of Cell Biology*, *121*(1), 23-36. doi:10.1083/jcb.121.1.23
- Mizuno, K., Ohsuye, K., Wada, Y., Fuchimura, K., Tanaka, S., & Matsuo, H. **1987**. Cloning and sequence of cDNA encoding a peptide C-terminal alpha-amidating enzyme from *Xenopus laevis*. *Biochemical and Biophysical Research Communications*, *148*(2), 546-552. doi:10.1016/0006-291x(87)90911-9
- Moller, S., Croning, M. D., & Apweiler, R. **2001**. Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics*, *17*(7), 646-653. doi:10.1093/bioinformatics/17.7.646
- Nguyen, L. T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. **2015**. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), 268-274. doi:10.1093/molbev/msu300
- Oldrati, V., Arrell, M., Violette, A., Perret, F., Sprungli, X., Wolfender, J. L., & Stocklin, R. **2016**. Advances in venomomics. *Molecular Biosystems*, *12*(12), 3530-3543. doi:10.1039/c6mb00516k
- Ortiz, E., Gurrola, G. B., Schwartz, E. F., & Possani, L. D. **2015**. Scorpion venom components as potential candidates for drug development. *Toxicon*, *93*, 125-135. doi:10.1016/j.toxicon.2014.11.233

- Ortiz, E., Rendon-Anaya, M., Rego, S. C., Schwartz, E. F., & Possani, L. D. **2014**. Antarease-like Zn-metalloproteases are ubiquitous in the venom of different scorpion genera. *Biochimica et Biophysica Acta*, *1840*(6), 1738-1746. doi:10.1016/j.bbagen.2013.12.012
- Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. **2017**. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*, *14*(4), 417-419. doi:10.1038/nmeth.4197
- Perkins, S. N., Husten, E. J., & Eipper, B. A. **1990**. The 108-kDA peptidylglycine alpha-amidating monooxygenase precursor contains two separable enzymatic activities involved in peptide amidation. *Biochemical and Biophysical Research Communications*, *171*(3), 926-932. doi:10.1016/0006-291x(90)90772-f
- Petersen, T. N., Brunak, S., von Heijne, G., & Nielsen, H. **2011**. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods*, *8*(10), 785-786. doi:10.1038/nmeth.1701
- 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., Anangi, R., Senff, S., Gorse, D., & King, G. F. **2018**. ArachnoServer 3.0: an online resource for automated discovery, analysis and annotation of spider toxins. *Bioinformatics*, *34*(6), 1074-1076. doi:10.1093/bioinformatics/btx661
- Possani, L. D., Becerril, B., Delepierre, M., & Tytgat, J. **1999**. Scorpion toxins specific for Na⁺-channels. *European Journal of Biochemistry*, *264*(2), 287-300. doi:10.1046/j.1432-1327.1999.00625.x
- Rodriguez de la Vega, R. C., & Possani, L. D. **2004**. Current views on scorpion toxins specific for K⁺-channels. *Toxicon*, *43*(8), 865-875. doi:10.1016/j.toxicon.2004.03.022
- Rodriguez de la Vega, R. C., & Possani, L. D. **2005**. Overview of scorpion toxins specific for Na⁺ channels and related peptides: biodiversity, structure-function relationships and evolution. *Toxicon*, *46*(8), 831-844. doi:10.1016/j.toxicon.2005.09.006
- Rokyta, D. R., & Ward, M. J. **2017**. 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*, *128*, 23-37. doi:10.1016/j.toxicon.2017.01.014
- Romero-Gutierrez, M. T., Santibanez-Lopez, C. E., Jimenez-Vargas, J. M., Batista, C. V. F., Ortiz, E., & Possani, L. D. **2018**. Transcriptomic and Proteomic Analyses Reveal the Diversity of Venom Components from the Vaejovid Scorpion *Serradigitus gertschi*. *Toxins*, *10*(9). doi:10.3390/toxins10090359
- Romero-Gutierrez, T., Peguero-Sanchez, E., Cevallos, M. A., Batista, C. V. F., Ortiz, E., & Possani, L. D. **2017**. A Deeper Examination of *Thorellius atrox* Scorpion Venom Components with Omic Technologies. *Toxins*, *9*(12). doi:10.3390/toxins9120399
- Santibanez-Lopez, C. E., Cid-Urbe, J. I., Batista, C. V., Ortiz, E., & Possani, L. D. **2016**. Venom Gland Transcriptomic and Proteomic Analyses of the Enigmatic Scorpion *Superstitionia donensis* (Scorpiones: Superstitioniidae), with Insights on the Evolution of Its Venom Components. *Toxins*, *8*(12), 367. doi:10.3390/toxins8120367
- Santibanez-Lopez, C. E., Cid-Urbe, J. I., Zamudio, F. Z., Batista, C. V. F., Ortiz, E., & Possani, L. D. **2017**. Venom gland transcriptomic and venom proteomic analyses of the scorpion *Megacormus gertschi* Diaz-Najera, 1966 (Scorpiones: Euscorpiidae: Megacorminae). *Toxicon*, *133*, 95-109. doi:10.1016/j.toxicon.2017.05.002

- Santibanez-Lopez, C. E., Francke, O. F., Ureta, C., & Possani, L. D. **2015**. Scorpions from Mexico: From Species Diversity to Venom Complexity. *Toxins*, *8*(1), 2. doi:10.3390/toxins8010002
- Santibanez-Lopez, C. E., Gonzalez-Santillan, E., Monod, L., & Sharma, P. P. **2019a**. Phylogenomics facilitates stable scorpion systematics: Reassessing the relationships of Vaejovidae and a new higher-level classification of Scorpiones (Arachnida). *Molecular Phylogenetics and Evolution*, *135*, 22-30. doi:10.1016/j.ympev.2019.02.021
- Santibanez-Lopez, C. E., Graham, M. R., Sharma, P. P., Ortiz, E., & Possani, L. D. **2019b**. Hadrurid Scorpion Toxins: Evolutionary Conservation and Selective Pressures. *Toxins*, *11*(11). doi:10.3390/toxins11110637
- Schmieder, R., & Edwards, R. **2011**. Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PLoS One*, *6*(3), e17288. doi:10.1371/journal.pone.0017288
- Schwager, E. E., Sharma, P. P., Clarke, T., Leite, D. J., Wierschin, T., Pechmann, M., Akiyama-Oda, Y., Esposito, L., Bechsgaard, J., Bilde, T., Buffry, A. D., Chao, H., Dinh, H., Doddapaneni, H., Dugan, S., Eibner, C., Extavour, C. G., Funch, P., Garb, J., Gonzalez, L. B., Gonzalez, V. L., Griffiths-Jones, S., Han, Y., Hayashi, C., Hilbrant, M., Hughes, D. S. T., Janssen, R., Lee, S. L., Maeso, I., Murali, S. C., Muzny, D. M., Nunes da Fonseca, R., Paese, C. L. B., Qu, J., Ronshaugen, M., Schomburg, C., Schonauer, A., Stollewerk, A., Torres-Oliva, M., Turetzek, N., Vanthournout, B., Werren, J. H., Wolff, C., Worley, K. C., Bucher, G., Gibbs, R. A., Coddington, J., Oda, H., Stanke, M., Ayoub, N. A., Prpic, N. M., Flot, J. F., Posnien, N., Richards, S., & McGregor, A. P. **2017**. The house spider genome reveals an ancient whole-genome duplication during arachnid evolution. *BMC Biology*, *15*(1), 62. doi:10.1186/s12915-017-0399-x
- Seppy, M., Manni, M., & Zdobnov, E. M. **2019**. BUSCO: Assessing Genome Assembly and Annotation Completeness. *Methods in Molecular Biology*, *1962*, 227-245. doi:10.1007/978-1-4939-9173-0_14
- Sharma, P. P., Fernandez, R., Esposito, L. A., Gonzalez-Santillan, E., & Monod, L. **2015**. Phylogenomic resolution of scorpions reveals multilevel discordance with morphological phylogenetic signal. *Proceedings: Biological Sciences*, *282*(1804), 20142953. doi:10.1098/rspb.2014.2953
- Silver, N., Best, S., Jiang, J., & Thein, S. L. **2006**. Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. *BMC Molecular Biology*, *7*, 33. doi:10.1186/1471-2199-7-33
- Spijker, S., Smit, A. B., Eipper, B. A., Malik, A., Mains, R. E., & Geraerts, W. P. **1999**. A molluscan peptide alpha-amidating enzyme precursor that generates five distinct enzymes. *FASEB Journal*, *13*(6), 735-748. doi:10.1096/fasebj.13.6.735
- Tabakmakher, V. M., Krylov, N. A., Kuzmenkov, A. I., Efremov, R. G., & Vassilevski, A. A. **2019**. Kalium 2.0, a comprehensive database of polypeptide ligands of potassium channels. *Sci Data*, *6*(1), 73. doi:10.1038/s41597-019-0074-x
- Torabi, E., Behdani, M., Khalaj, V., Pooshang Bagheri, K., & Shahbazzadeh, D. **2020**. Complete neutralization of the lethality of *Hemiscorpius lepturus* crude venom by a novel anti-recombinant phospholipase D1 IgGs. *Toxicon*, *183*, 36-43. doi:10.1016/j.toxicon.2020.05.010
- Valdez-Velazquez, L. L., Cid-Urbe, J., Romero-Gutierrez, M. T., Olamendi-Portugal, T., Jimenez-Vargas, J. M., & Possani, L. D. **2020**. Transcriptomic and proteomic analyses of the venom and venom glands of

Centruroides hirsutipalpus, a dangerous scorpion from Mexico. *Toxicon*, 179, 21-32. doi:10.1016/j.toxicon.2020.02.021

- Waddington, J., Rudkin, D. M., & Dunlop, J. A. **2015**. A new mid-Silurian aquatic scorpion-one step closer to land? *Biology Letters*, 11(1), 20140815. doi:10.1098/rsbl.2014.0815
- Ward, M. J., Ellsworth, S. A., & Rokyta, D. R. **2018**. 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*, 142, 14-29. doi:10.1016/j.toxicon.2017.12.042
- Xiao, L., Gurrola, G. B., Zhang, J., Valdivia, C. R., SanMartin, M., Zamudio, F. Z., Zhang, L., Possani, L. D., & Valdivia, H. H. **2016**. Structure-function relationships of peptides forming the calcin family of ryanodine receptor ligands. *Journal of General Physiology*, 147(5), 375-394. doi:10.1085/jgp.201511499
- Yin, P., Bousquet-Moore, D., Annangudi, S. P., Southey, B. R., Mains, R. E., Eipper, B. A., & Sweedler, J. V. **2011**. Probing the production of amidated peptides following genetic and dietary copper manipulations. *PloS One*, 6(12), e28679. doi:10.1371/journal.pone.0028679
- Zabriskie, T. M., Klinge, M., Szymanski, C. M., Cheng, H., & Vederas, J. C. **1994**. Peptide amidation in an invertebrate: purification, characterization, and inhibition of peptidylglycine alpha-hydroxylating monooxygenase from the heads of honeybees (*Apis mellifera*). *Archives of Insect Biochemistry and Physiology*, 26(1), 27-48. doi:10.1002/arch.940260104

Apéndices

A1. Remoción de contigs contaminantes

El alineamiento de los contigs ensamblados de cada uno de los transcriptomas a los genomas enlistados en el punto 5.2 permitió la identificación y posterior remoción de secuencias contaminantes. Los grupos más representados fueron el bacteriano y el correspondiente a los mamíferos (humano principalmente).

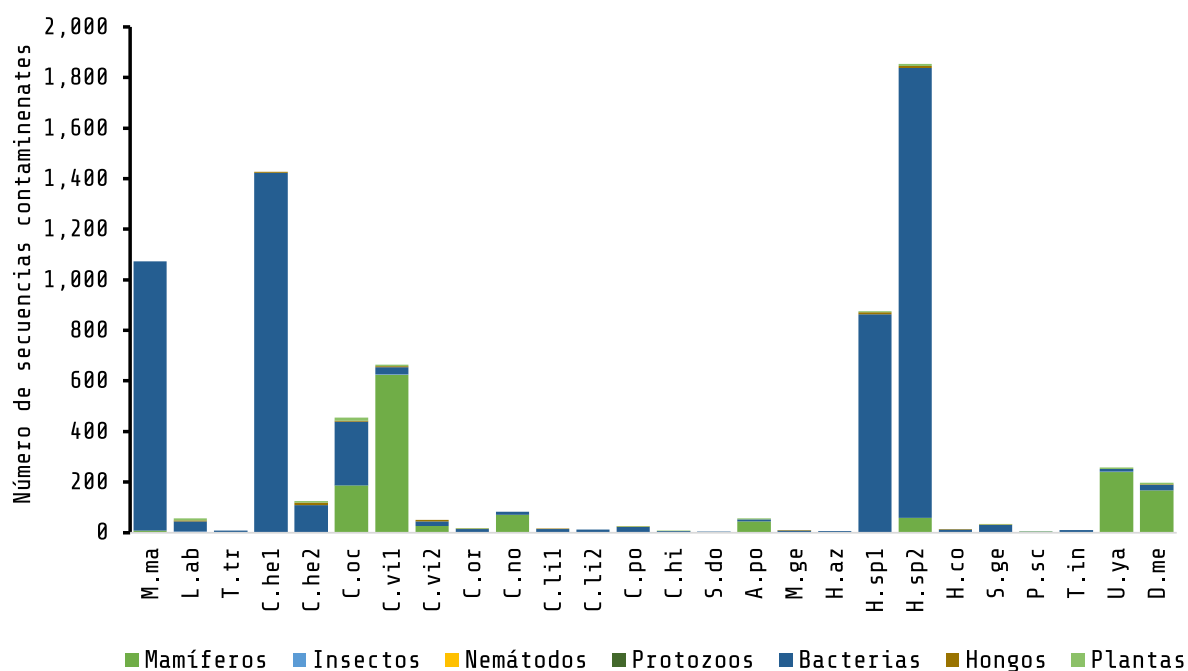


Figura A1.1. Identificación de secuencias contaminantes

A2. Ensamble y extensión manual de contigs

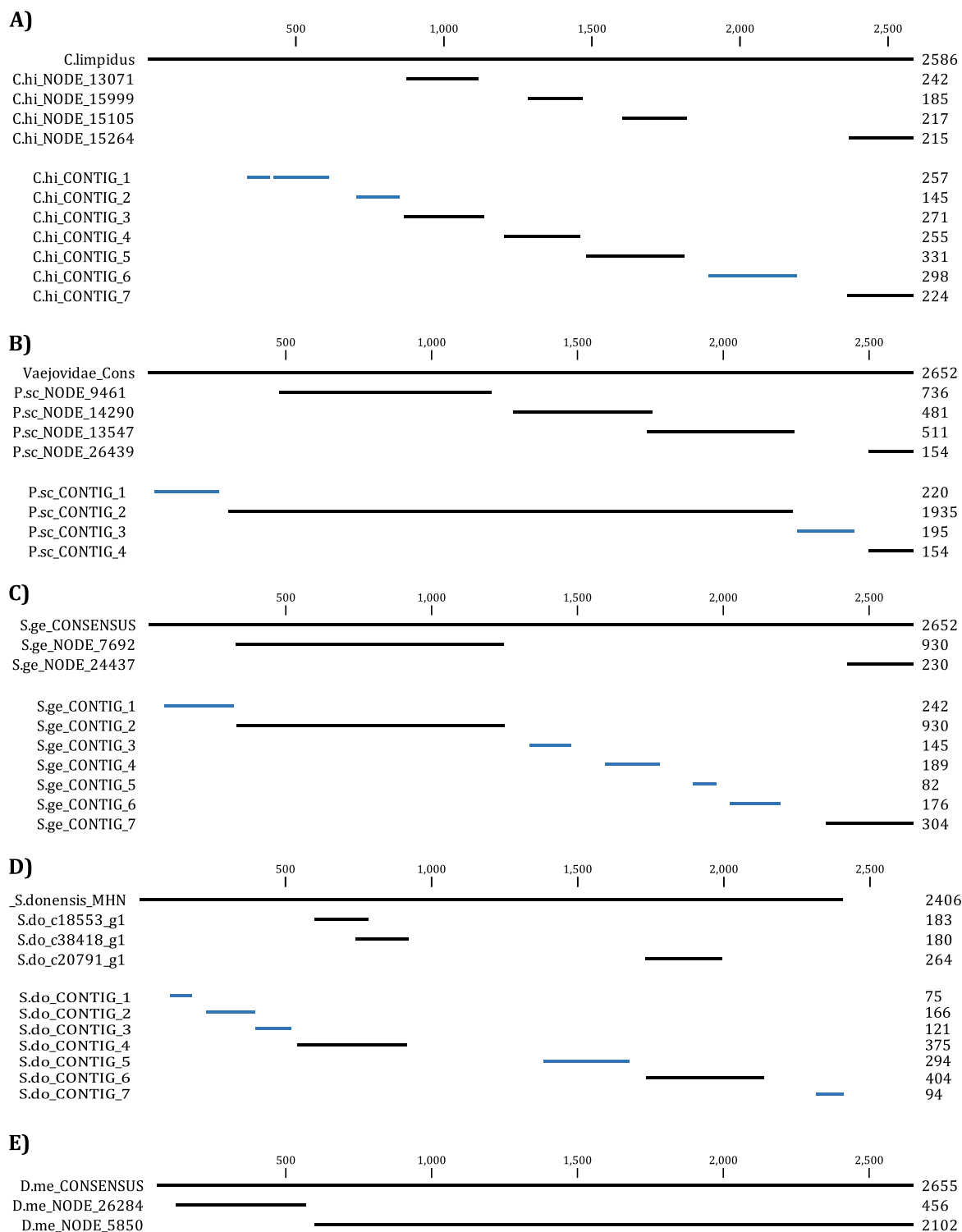


Figura A2.1. Reconstrucción y extensión manual de secuencias PAM parciales posterior al ensamble *de novo*. Alineamiento de las secuencias PAM de referencia con las secuencias parciales identificadas en el transcriptoma de **A)** *Centruroides hirsutipalpus*; **B)** *Paravaejovis schwenkmeyeri*; **C)** *Serradigitus gertschi*; **D)** *Superstitionia donensis*; **E)** *Diplocentrus melici*; —) Regiones previamente desconocidas

A3. Comparación de los niveles de expresión de la enzima PHMm

La enzima PHMm fue identificada previamente en los transcriptomas del alacrán *Centruroides hentzi* (hembra y macho) por Ward *et al.*, en 2018. Los niveles de expresión no normalizados son parecidos a los obtenidos en este trabajo, sin embargo, para el análisis comparativo estos valores fueron normalizados considerando los TPM de PGK1 (fosfoglicerato quinasa 1) obteniendo valores de 5.2 y 5.4 TPM para macho y hembra respectivamente.

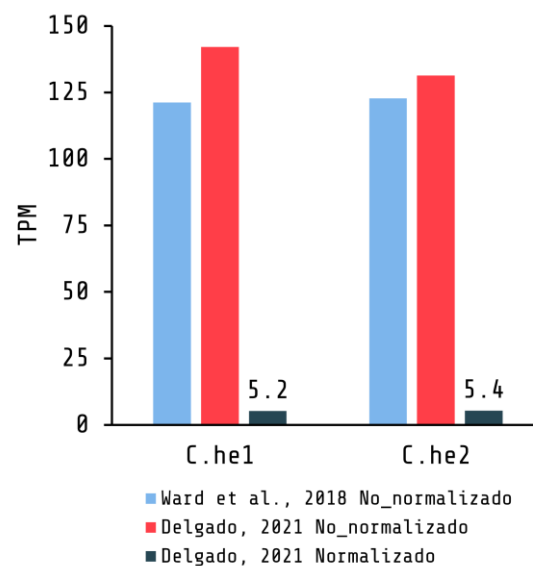


Figura A3.1. Niveles de expresión de la enzima PHMm en *C. hentzi* (Che1: macho; C.he2: hembra) Se muestran los niveles no normalizados reportados por Ward *et al.*, y los valores normalizados y no normalizados obtenidos en este trabajo.

A4. Niveles individuales de expresión de las enzimas amidantes

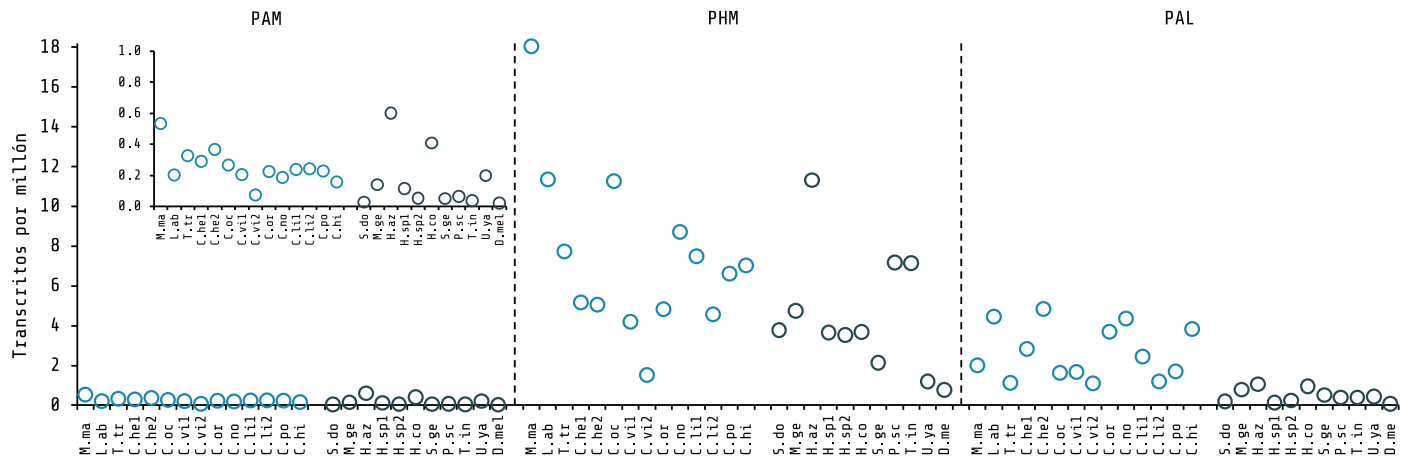


Figura A4.1. Niveles normalizados de expresión de los transcritos codificantes para las enzimas PAM, PHMm y PALm en cada alacrán; ○) alacranes del parvorden Bithida (bútidós); ○) alacranes del parvorden Iurida (iúridós)

A5. Relación entre los niveles de expresión de las enzimas amidantes y la profundidad de secuenciación

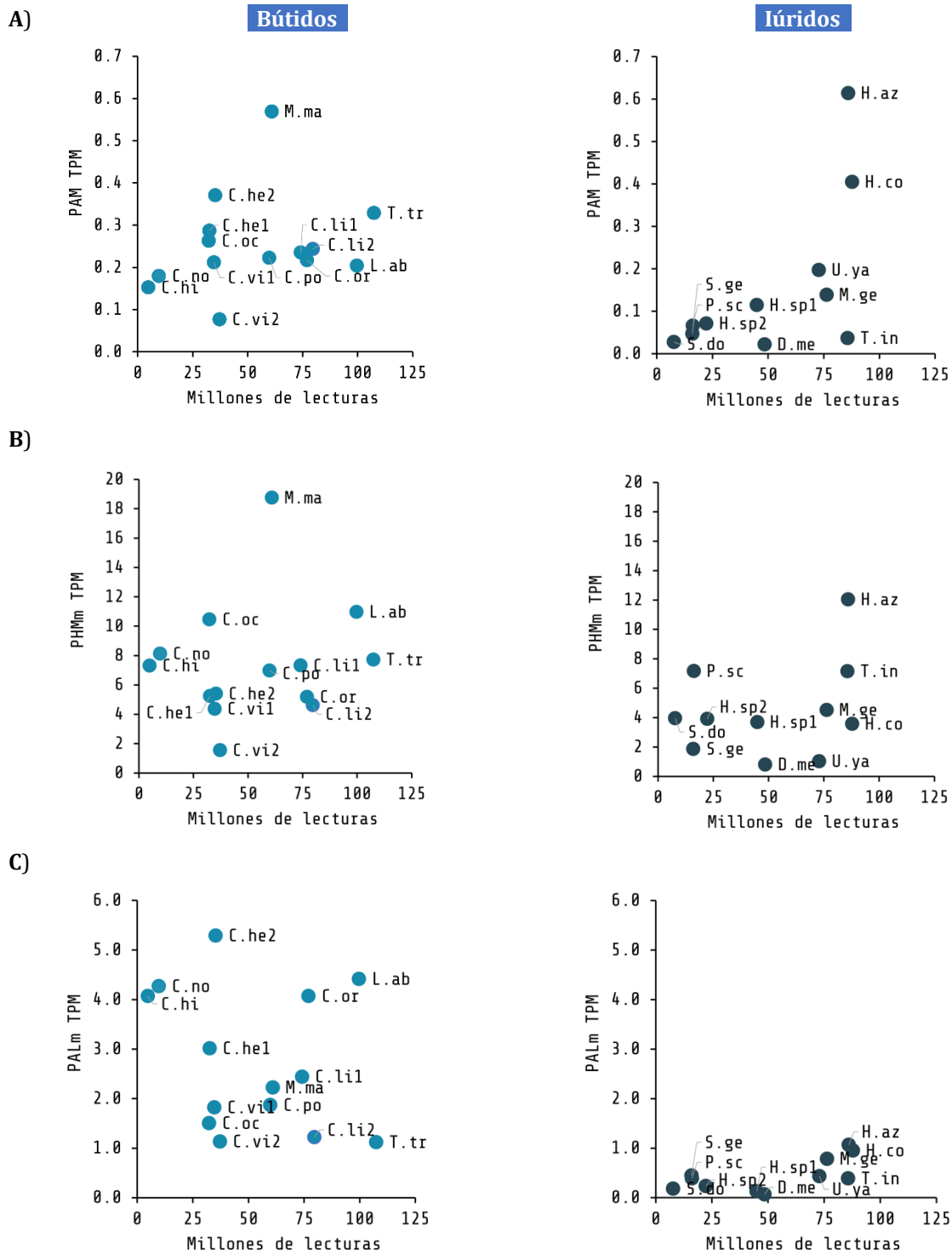


Figura A5.1. Efecto de la profundidad de secuenciación sobre los niveles de expresión de las enzimas A) PAM; B) PHMm; C) PALm entre bütidos e iúridos. No se observa correlación entre las variables.

A6. Clonación, expresión y caracterización bioquímica de la enzima peptidil-glicina α -amidante monooxigenasa (PAM) de alacrán

La caracterización y potencial uso de estas enzimas supone una alternativa para la obtención *in vitro* de péptidos amidados, ya que actualmente muchos de estos péptidos se obtienen mediante síntesis química.

Objetivo General

Expresar y caracterizar los dominios catalíticos PHM-PAL de un alacrán representativo y comparar sus parámetros cinéticos con los de las unidades catalíticas de la enzima PAM de *Rattus norvegicus*.

Selección de la enzima PAM de alacrán para su expresión y caracterización

La selección de la enzima PAM de un alacrán representativo para la expresión y posterior caracterización se basó en el conocimiento sobre la abundancia de péptidos amidados presentes en los venenos de alacranes (Figura 13). En particular, en el veneno del alacrán mexicano *Centruroides noxius* está presente una toxina amidada muy abundante. Esta toxina (Cn2, β -NaTx), representa el 6.5% del total del veneno (datos no publicados), y en ningún otro veneno actualmente estudiado, un componente amidado tiene una abundancia cercana a la observada en *C. noxius*. Tomando como base esta información, se propuso la enzima PAM de este alacrán para su expresión y caracterización bioquímica. La extracción del RNA total, amplificación y secuenciación de la secuencia PAM de *C. noxius* puede consultarse en Materiales y métodos, sección 5.7.

Región de la enzima PAM a expresar (dominios catalíticos PHM-PAL)

Las proteínas ancladas a membranas, como la enzima PAM, suelen presentar complicaciones y bajos rendimientos en su purificación. Satani *et al.*, (2003), describen la expresión de la enzima PAM humana en células CHO, eliminando el dominio transmembranal (MSD), esto al introducir un codón de paro en la posición del Glu⁸³⁵, residuo localizado en la región que conecta el dominio PAL y dominio de anclaje a membrana (MSD), logrando obtener una enzima soluble y activa (Figura A6.1), este trabajo es la base para la expresión de la enzima amidante de alacrán sin el dominio transmembranal. Los ensayos de expresión y caracterización bioquímica de la enzima PAM del alacrán *C. noxius* se realizaron a la par de la correspondiente enzima de *R. norvegicus*. La secuencia PAM-1 de *R. norvegicus* fue clonada previamente por Milgram *et al.*, en 1992, el plásmido fue depositado en Addgene con el nombre pCIS-krPAM1 y fue solicitado por la sección de compras del IBT-UNAM. Las regiones para expresar estarán delimitadas por el inicio de la cadena madura (inicio

del dominio *phm*-PAM) y terminarán una posición antes del glutámico correspondiente. A dicha región se le refiere como Cn-PHM-PAL-Glu⁷⁴¹ (Cn → *Centruroides noxius*) y Rn-PHM-PAL-Glu⁸³⁵ (Rn → *Rattus norvegicus*) (Figura A6.1).

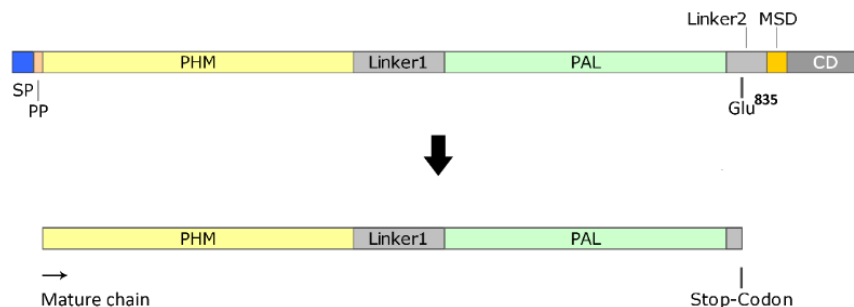


Figura A6.1. Región de la enzima PAM que se expresará: *Cn*-PHM-PAL-Glu⁷⁴¹ y *Rn*-PHM-PAL-Glu⁸³⁵.

Selección del sistema heterólogo de expresión

Se empleó como sistema de expresión a la levadura *Pichia pastoris* X-33. El plásmido pPICZ α A fue seleccionado para la subclonación de los dominios PHM-PAL. El plásmido permite la fusión de la proteína con el péptido señal del α -factor de *Saccharomyces cerevisiae*, el cual gestiona su secreción. La expresión del producto está regulada por el promotor de la enzima alcohol deshidrogenasa (AOX1), inducible con metanol (Figura A6.2).

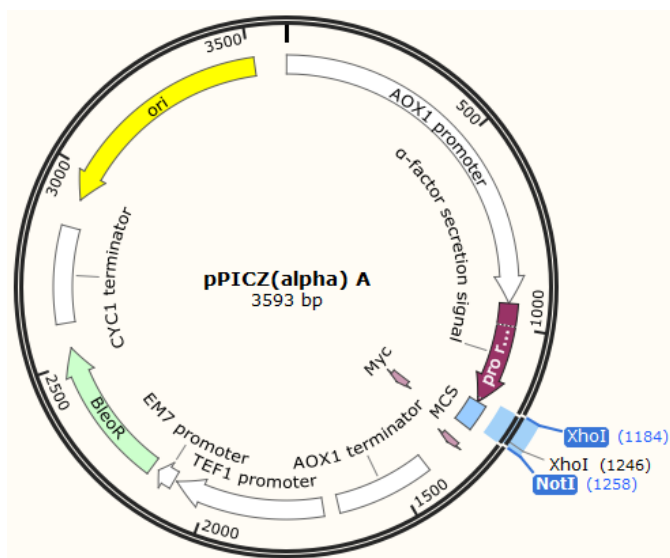


Figura A6.2. Plásmido pPICZ α A, se muestran los sitios NotI y XhoI dentro del sitio múltiple de clonación (MCS)

Diseño de “primers” para la amplificación de los dominios PHM-PAL

Se diseñaron los primers Rat-Fw7 (GAC TCG AGA AAA GAT TTA AAG AAA CTA CCA GAT CAT TTT CC) / Rat-Rv8 (TCG CGG CCG CCT ATT AAA CAA CTG CCT CGG CTT C) y Cen-Fw5 (GAC TCG AGA AAA GAG CAG AAC CAC CCA TGC AAA) / Cen-Rv6 (TCG CGG CCG CCT ATT AAG GTG TTA TTT TTT GCG ACG ATA ATG A) para amplificar las regiones de interés y agregar los sitios de corte XhoI / NotI que permitirán la clonación en el plásmido pPICZ α A. Los primers directos Rat-Fw7 / Cen-Fw5 incluyen un par de nucleótidos (GA) río arriba del sitio de corte XhoI para facilitar el posicionamiento de la enzima de restricción e incluyen la secuencia que regenera el sitio de corte del α -factor (-AAAAGA-) que codifica para los residuos lisina y arginina en la secuencia diana KR|EA). Los primers Rat-Rv8 y Cen-Rv6, por su parte, incluyen un par de bases (GA) río abajo del sitio de corte NotI y dos codones de paro, el primero de ellos fue introducido en la posición correspondiente al Glu⁸³⁵ en la secuencia de *R. norvegicus* y del Glu⁷⁴¹ en la secuencia de *C. noxius* seguido de un segundo codón de paro. Los amplificados fueron digeridos con las enzimas XhoI y NoI, posteriormente fueron ligados al plásmido pPICZ α A y transformados en *E. coli* DH5 α , las células fueron sembradas en placas con medio LB + Zeocina (25 μ g/mL). Después del periodo de incubación (18 h), se realizó una PCR de colonia con los primers Rat-Fw7 / Rat-Rv8 y Cen-Fw5 / Cen-Rv6 para identificar clonas transformantes, 5 μ L de las reacciones se cargaron en un gel de agarosa al 1% y teñida con Gel-Red al 3% v/v. Como control se cargaron reacciones de PCR realizadas con los primers 5'AOX1/3'AOX1 que generan amplicones de mayor tamaño (2,671 pb para la secuencia de alacrán y 2,941 pb para la secuencia de rata) (Figura A6.3).

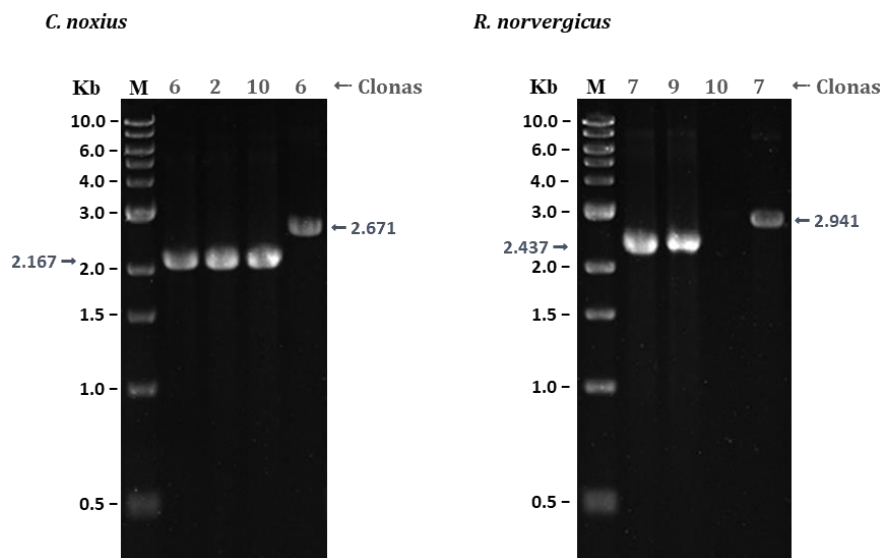


Figura A6.3. Resultado de la PCR de colonia. Productos esperados: 2,167 pb (Cn-PHM-PAL-Glu⁷⁴¹) y 2,437 pb (Rn-PHM-PAL-Glu⁸³⁵).

Los insertos clonados fueron secuenciados para asegurar que la construcción fuera correcta. Las secuencias clonadas en pPICZ α A pueden integrarse al genoma de *P. pastoris* por recombinación, la cual se ve favorecida al transformar con el plásmido lineal.

Extracción y digestión de los plásmidos.

Se inocularon tubos con 12.5 mL de medio LB bajo en sales + Zeocina (25 μ g/mL) con clonas cuya secuencia fue verificada por secuenciación. La extracción de los plásmidos se realizó con el High Pure Plasmid Isolation Kit (Roche). La digestión se realizó con 2 U de la enzima SacI en Buffer NEB 1.1, incubando a 37°C. Tras 2 y 4 horas de incubación se tomaron 0.5 μ L de cada digestión para ser cargadas en un gel de agarosa al 1% y evaluar la eficiencia de la digestión (Figura A6.4).

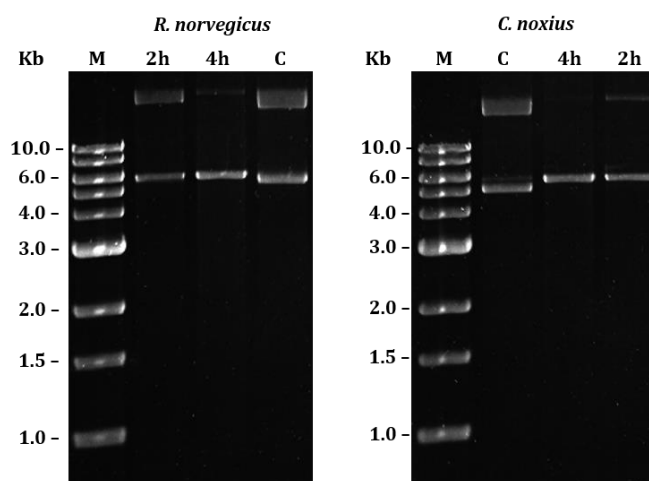


Figura A6.4. Digestión de los plásmidos con las secuencias de interés a las 2 y 4 horas de incubación; C, plásmido sin digerir.

Se realizó la limpieza del DNA digerido con el High Pure PCR Product Purification Kit (Roche), la cantidad de DNA fue determinada considerando la concentración determinada en Nanodrop y el volumen de elución (Tabla A1).

Tabla A1. Cantidad de DNA digerido para su transformación en *P. pastoris* X-33

Construcción	DNA antes de la digestión (μ g)	DNA después de la digestión (μ g)
pPICZ α A – CnPHM-PAL-Glu ⁷⁴¹	20.2	17.4
pPICZ α A – RnPHM-PAL-Glu ⁸³⁵	18.7	15.2

Transformación de células electrocompetentes de *Pichia pastoris* X-33

Los plásmidos digeridos se concentraron en una centrifuga de vacío, el volumen se redujo de ~200 μL a ~5 μL . Se prepararon células electrocompetentes según el EasySelect Pichia Expression Kit (Invitrogen), las células se cosecharon a una $\text{DO}_{600\text{nm}} = 1.6$. Se transfirieron 80 μL de la suspensión celular a los tubos con los plásmidos previamente concentrados, se dejaron incubando en hielo durante 5 min previo a la electroporación. Las células transformadas se sembraron en placas con YPDS + Zeocina (800 $\mu\text{g}/\text{mL}$) para favorecer la selección de multicopias.

PCR de colonia para identificar integrantes

La evaluación de integrantes se realizó con un primer que hibrida en el plásmido pPICZ α A (5'AOX1) y un primer que hibrida con el inserto correspondiente (Rat-Rv8/Cen-Rv6) (Figura A6.5).

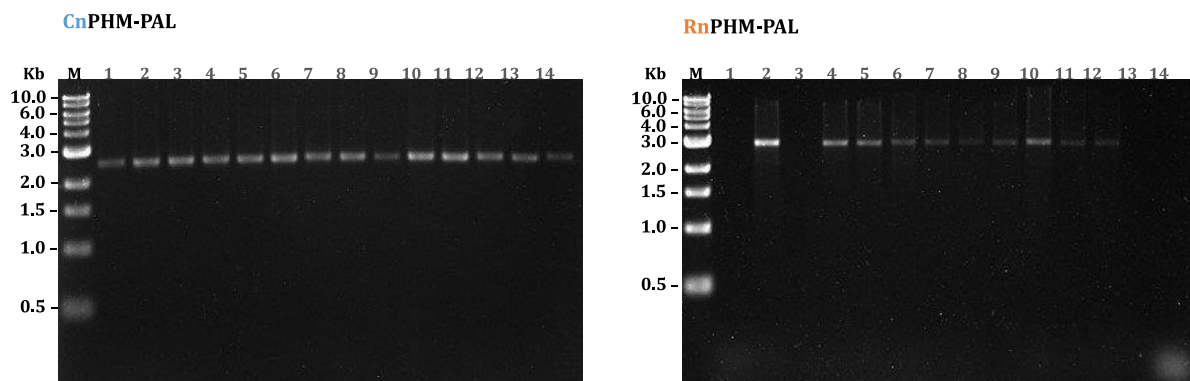


Figura A6.5. Resultado de la PCR de colonia. Productos esperados: 2,671 pb (Cn-PHM-PAL-Glu⁷⁴¹) y 2,941 pb (Rn-PHM-PAL-Glu⁸³⁵).

Determinación del fenotipo “methanol utilization” (*mut*)

La transformación de la cepa *P. pastoris* X-33 con construcciones linealizadas favorece la recombinación cruzada simple en el locus AOX1. La mayoría de las transformantes deberían tener el fenotipo denominado *mut*⁺ (methanol utilization plus), sin embargo, con la presencia de las secuencias de AOX1 en el plásmido pPICZ α A, existe la posibilidad de que la recombinación se produzca en la región 3'AOX1, alterando el gen AOX1 y creando transformantes *mut*^s (methanol utilization slow). La determinación del fenotipo mediante pruebas de crecimiento en placas con medio mínimo con metanol (MMH, YNB 1.34X, Biotina 0.00004%, Metanol 0.5%, Histidina 0.004%) indicadas por el fabricante no permitieron confirmar de manera reproducible el fenotipo *mut*. El crecimiento aparente de las clonas dependía en gran medida de la cantidad de inóculo sembrado en la placa. Se procedió a evaluar el fenotipo *mut* de las clonas transformantes en medio líquido usando microplacas de 96 pozos. Células clonales suficientes fueron suspendidas en medio MMH y

se ajustaron los inóculos a una $DO_{595nm} = 0.15 \pm 0.03$. Las células fueron incubadas a $30^{\circ}C$ en agitación constante (150 rpm) y se midió el crecimiento después de 24 h. A las lecturas obtenidas se les restó la lectura inicial. En el caso de las clonas recombinantes Cn-PHM-PAL-Glu⁷⁴¹, se identificaron 7 clonas (C-1, C-6, C-15, C-16 y C-20) con un crecimiento superior al de la clona control (X-33 WT), a éstas se les consideró como el fenotipo *mut*⁺. A su vez, en este experimento se observaron clonas con un menor crecimiento respecto a la clona control, siendo C-12 la que mostró el menor crecimiento de todas ellas, esta fue seleccionada junto con las clonas *mut*⁺ para el ensayo de expresión (Figura A6.6A). En el caso de las clonas recombinantes Rn-PHM-PAL-Glu⁸³⁵, ninguna de las clonas evaluadas mostró un crecimiento superior a la cepa control (X-33 WT), las clonas R-4, R-6 y R-7 fueron seleccionadas para el ensayo de expresión (Figura A6.6B).

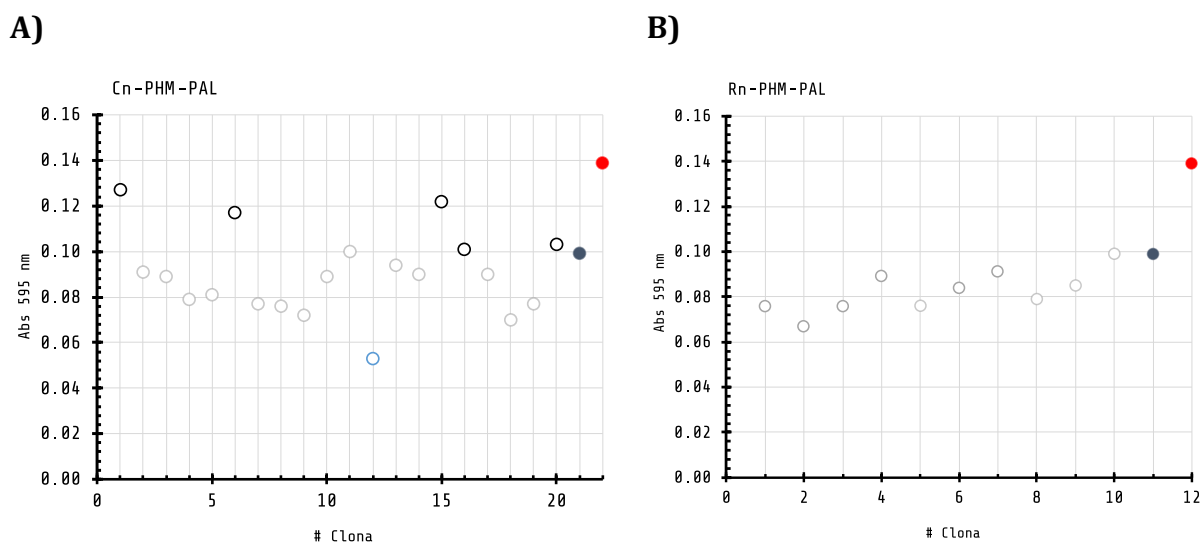


Figura A6.6. Determinación del fenotipo *mut*⁺/*mut*⁻.

Crecimiento de clonas en medio mínimo (líquido) con metanol (MMH). A) Clonas con la construcción Cn-PHM-PAL-Glu⁷⁴¹; B) Clonas con la construcción Rn-PHM-PAL-Glu⁸³⁵; ○) Clonas con el mayor crecimiento medido respecto a la clona WT, *mut*⁺; ◯) Clona con el menor crecimiento medido, *mut*⁻; ●) cepa control X-33, WT; ●) Clona X-33 que expresa una lacasa, donada por la Dra. Marcela Ayala (control positivo de expresión).

Ensayo de expresión

Se inocularon 12.5 mL de medio BMGY (extracto de levadura 1%, peptona 2%, buffer MES 0.1M pH 6, YNB 1.34%, biotina 0.00004%, glicerol 1%) en tubos cónicos de 50 mL con las clonas seccionadas, estas se incubaron a $29^{\circ}C$ con agitación (250 rpm) hasta que el cultivo alcanzó una $DO_{600nm} \sim 12$. Las células fueron cosechadas centrifugando a $2,000 \times g$ durante 5 minutos a temperatura ambiente. El sobrenadante se decantó y se realizó un primer lavado del sedimento celular con 20 mL de agua Milli-Q estéril, se centrifugó a $2,000 \times g$ durante 5 min, el sobrenadante

se decantó y se realizó un segundo lavado con 10 mL de agua Milli-Q siguiendo las mismas condiciones de centrifugación. El sedimento celular se suspendió en 20.0 mL de medio amortiguado BMMY (extracto de levadura 1%, peptona 2%, buffer MES 0.1M pH 6, YNB 1.34%, biotina 0.00004%, metanol 1.5%, CuSO₄ 0.7 mM, CaCl₂ 0.5 mM, ZnCl₂ 0.5 mM). Se ajustó la OD_{600nm} ~ 10, el volumen para el ensayo de expresión fue 20 mL en matraces de 500 mL.

En cada uno de los momentos indicados (0, 24, 48, 72, 96, 120, 144 horas), se transfirieron alícuotas de 1.0 mL del cultivo a 2 microtubos de 1.5 mL y se adicionó un volumen equivalente (2 mL) de una solución de metanol al 15% v/v a los cultivos. Los tubos con las muestras fueron centrifugados a velocidad máxima durante 2 minutos a temperatura ambiente, el sobrenadante clarificado se transfirió a un nuevo microtubo de 1.5 mL. Las muestras fueron almacenadas a -70°C. Estas muestras se utilizaron para analizar los niveles de expresión mediante SDS-PAGE y para la determinación de la actividad enzimática por HPLC (ver más adelante). Como control se incluyó a la cepa X33 WT.

Análisis por SDS-PAGE

Se tomaron 150 µL de los sobrenadantes muestreados a las 96 h de iniciado el ensayo de expresión, las proteínas se precipitaron siguiendo el protocolo de Metanol/Cloroformo (Fic, E., *et al.*, 2010). Los pellets fueron resuspendidos en 7 µL de Buffer de carga 2X (125 mM Tris-HCl, pH 6.8, 4% SDS, 20% v/v glicerol, 10mM β-mercaptoetanol, 0.05% azul de bromofenol), las muestras se calentaron a 80 °C durante 5 min. Se cargaron ~7 µL en cada pozo de un gel de SDS-PAGE al 10.5%.

Identificación de clonas que expresan los dominios PHM-PAL de *Centruroides noxius*

Las clonas C-1, C-6, C-15, C-16 y C-20 fueron previamente catalogadas con el fenotipo *mut*⁺, lo que hizo suponer que estas clonas expresarían en mayor cantidad a la proteína de interés. Por el contrario, la clona C-12 (*mut*^s) se esperaba que mostrara los niveles más bajos de proteína recombinante. Sin embargo, la clona con C-12, expresó una proteína abundante respecto al resto de clonas, cuyo tamaño es cercano al esperado para la enzima de alacrán (Cn-PHM-PAL-Glu⁷⁴¹) (~80 KD). El resto de clonas evaluadas tienen un patrón de bandeo menos evidente en el rango de tamaño esperado. La cepa X-33 WT no presenta proteínas del tamaño esperado (Figura A6.7).

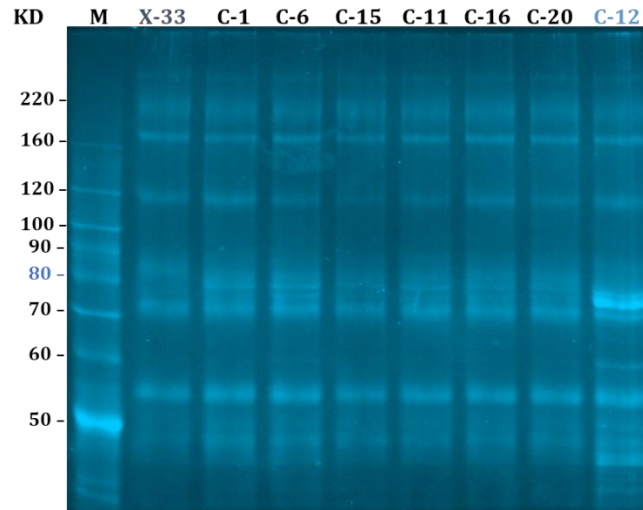


Figura A6.7. Perfil de expresión a las 96 h de inducción de las clonas C-1, C-6, C-15, C-16 y C-20 de *P. pastoris* con las secuencias PHM-PAL de *C. noxius*; X-33) Expresión de fondo de la cepa WT. 96 h de expresión.

Una vez identificada una clona que potencialmente sobreexpresa a la proteína de interés (clona C-12), se evaluó el perfil de proteínas secretadas a lo largo del tiempo (Figura A6.8). Este resultado mostró que a las 24 horas hay dos bandas del tamaño cercano al esperado, siendo la banda superior la que es consistente con el tamaño teórico de la proteína (80.4 KD) después de las 24 horas, la banda superior desaparece y se observa una banda de tamaño ligeramente inferior.

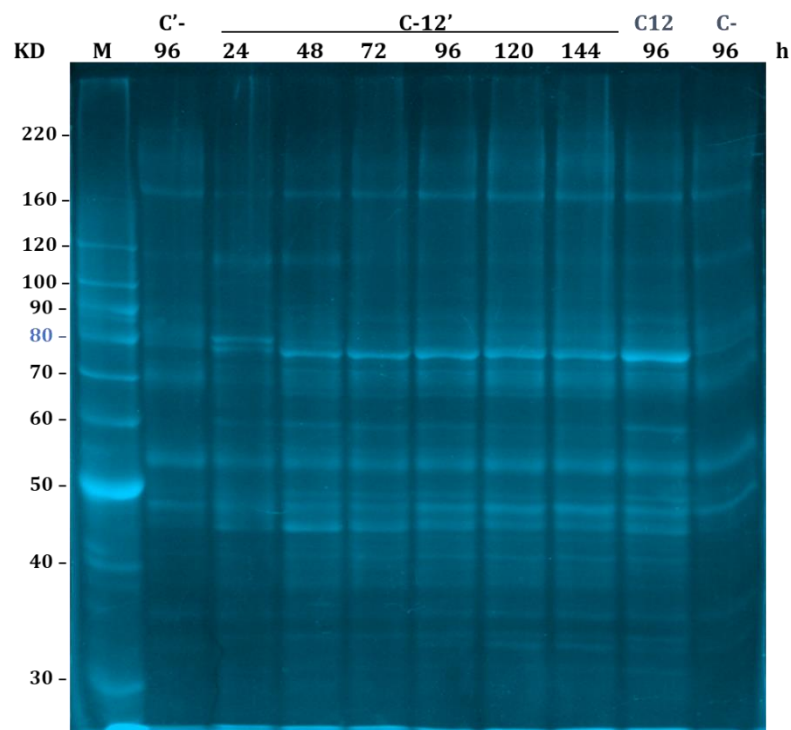


Figura A6.8. Perfil de expresión de la clona C-12 a lo largo del tiempo. C-) expresión de fondo de la cepa WT X-33.

Identificación de clonas que expresan los dominios PHM-PAL de *Rattus norvegicus*

La Figura A10 muestra el perfil de proteínas secretadas por las clonas R-4, R-6 y R-7. Este perfil de expresión mostró un claro incremento en las proteínas secretadas, el mayor número de bandas observadas quizá se deba a la acción proteolítica. Entre las bandas observadas se encuentra una del tamaño esperado (90 KD) la cual no se observa en el perfil de proteínas secretadas de la cepa WT. En el ensayo se incluyó a la clona Rn24, la cual fue seleccionada en ensayos previos como potencial productora de las enzimas PHM-PAL de rata (Rn-PHM-PAL-Glu⁸³⁵), además de clonas con la construcción de la enzima PHM monofuncional (PHM_m) de *C. noxius* (M1, M2, M8, 36 KD) y la clona que expresa una lacasa (Lac) como control positivo (Figura A6.9).

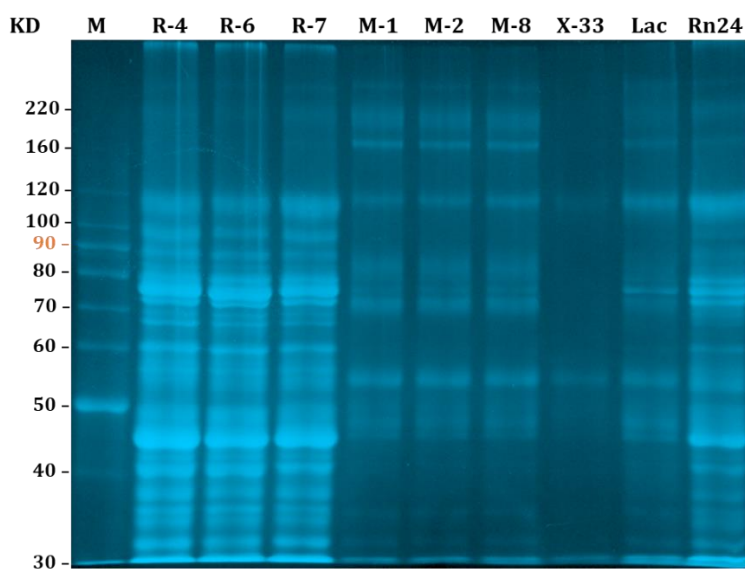


Figura A6.9. Perfil de expresión a las 96 h de inducción de las clonas *P. pastoris* X-33 con las secuencias PHM-PAL de *C. noxius* X-33; expresión de fondo de la cepa WT.

Ensayo de actividad enzimática

Aunque el análisis por SDS-PAGE del patrón de proteínas secretadas es una primera aproximación para seleccionar clonas que potencialmente expresen y secreten al medio la proteína de interés, la presencia de una enzima funcional se evaluó mediante el ensayo de actividad enzimática, este ensayo consiste en evaluar por HPLC de fase reversa, la conversión del sustrato Dansyl-YVG a Dansyl-YV-NH₂. Esta conversión tiene como consecuencia un cambio en la hidrofobicidad de la

molécula, por lo que el tiempo de retención dentro de la columna difiere entre el producto y el sustrato, siendo el producto el que presenta un menor tiempo de retención (Figura A6.10).

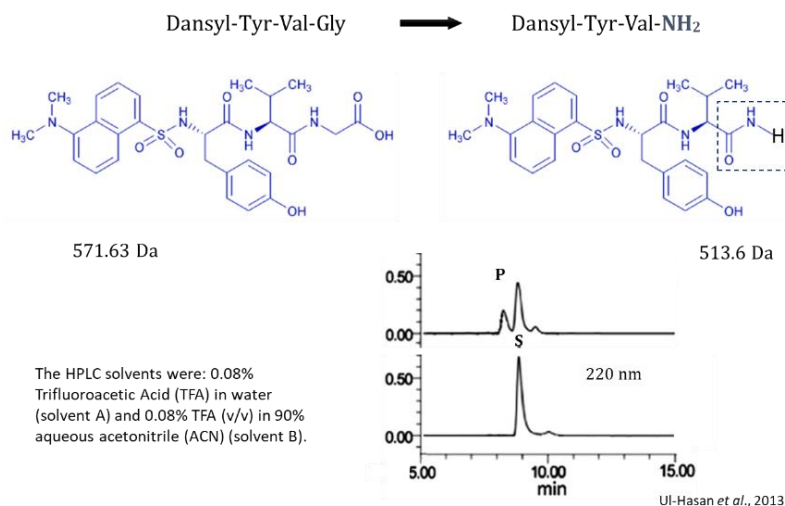


Figura A6.10. Conversión del sustrato Dansyl-YVG (S) a Dansyl-YV-NH₂(P).

Para el ensayo de actividad se preparó una mezcla de reacción 2x (100 mM NaTES pH 6.5, 0.02% Tween 20, 5 mM ácido ascórbico, 4 μ M CuSO₄, 200 mg/mL catalasa, 0.01 μ M Dansyl-YVG) y se combinó con un volumen de sobrenadante de 96 horas de cultivo. El volumen de reacción (300 μ L), se pasó por un filtro para centrífuga Amicon® 3K. Un volumen de 30 μ L fue cargado a la columna (C-18) y se tomaron registros durante 40 min. Las Figuras A6.11A y A6.11B muestran los cromatogramas de los componentes de *P. pastoris* X-33 (control negativo) eluidos en ausencia y presencia del sustrato respectivamente. Este resultado permitió saber el tiempo de retención del sustrato (aproximadamente 16.7 min) y que *P. pastoris* no tiene actividad amidante nativa. La Figura A6.11C muestra los componentes de la clona C-12 (PHM-PAL de *C. noxius*) eluidos en ausencia de sustrato. La Figura A6.11D corresponde al cromatograma de la clona C-12 en presencia del sustrato, se observa un perfil similar al visto previamente en el control negativo, el pico de 16.7 min fue colectado y su masa fue determinada, dando como resultado la masa del sustrato. Un pico con un menor tiempo de retención que corresponda al producto no fue observado, concluyendo en una ausencia de actividad en el sobrenadante de la clona C-12. Las Figuras A6.11E y A6.11F muestran los componentes de la clona Rn-24 (PHM-PAL de *R. norvegicus*), en ausencia y presencia de sustrato, el análisis de tiempos de retención y comprobación de la masa molecular del único pico diferencial, indicado con una flecha, comprueba que el sustrato no fue amidado, indicando la ausencia de actividad en el sobrenadante. Se evaluaron clonas adicionales (C-6, PHM-PAL de *C. noxius* y R-6, PHM-PAL de *R. norvegicus*) (datos no mostrados) sin encontrar actividad.

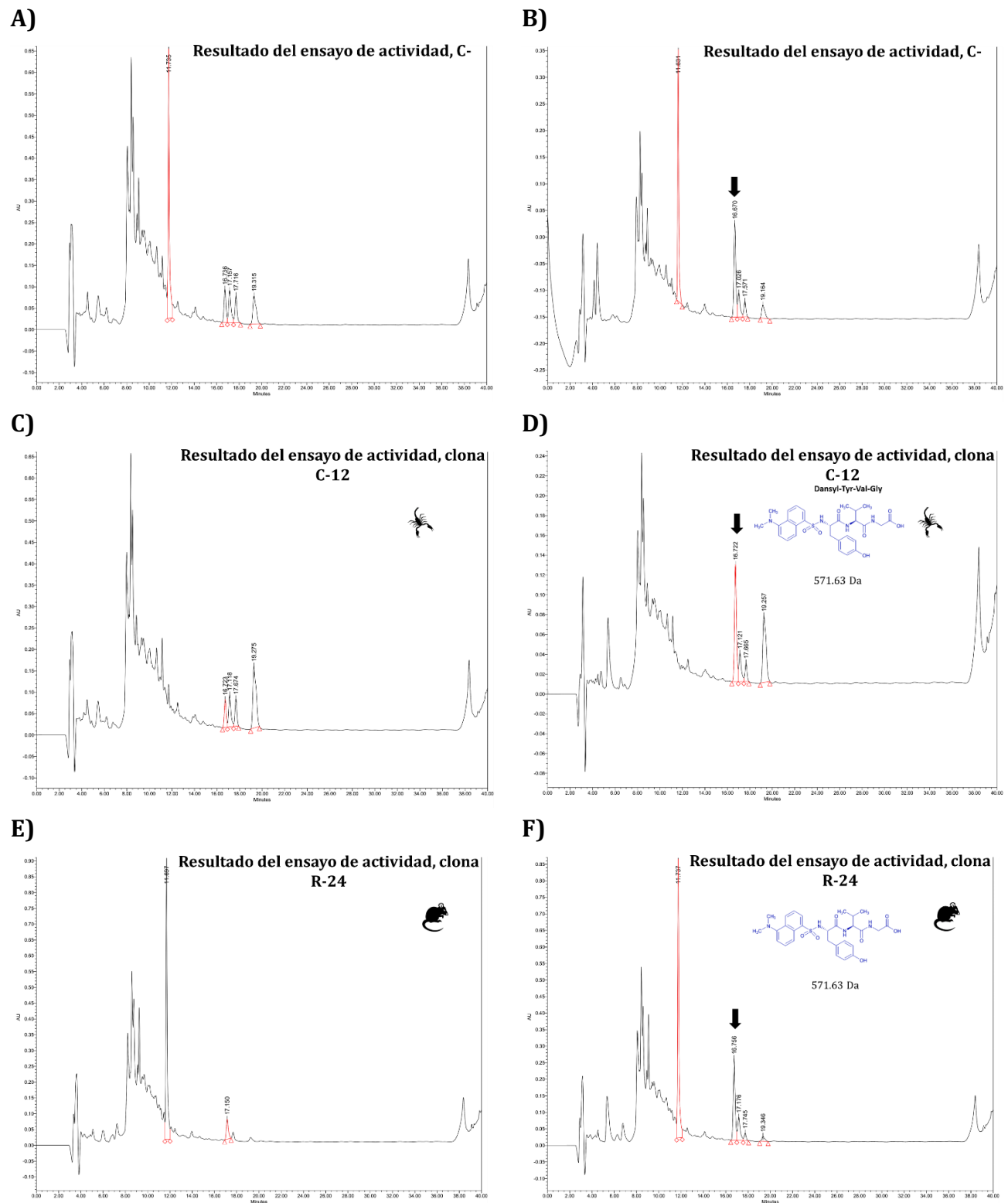


Figura A6.11. Ensayos de actividad enzimática en ausencia y presencia de sustrato A) y B) *P. pastoris* X-33 WT; C) y D) Clona C-12 (Cn-PHM-PAL-Glu⁷⁴¹); E) y F) Clona R-24 (Rn-PHM-PAL-

Glu⁸³⁵). La flecha indica el pico correspondiente al tiempo de retención del sustrato (Dansyl-YVG) de 16.7 min.

Una causa probable por la que no se detectó actividad es por la baja concentración de sustrato empleada para los ensayos de actividad (0.01 μM), esta pudo no ser suficiente para alcanzar la mitad de la velocidad máxima de las enzimas (K_m).

Síntesis del sustrato (Dansyl-YVG)

Los ensayos realizados agotaron el stock disponible de sustrato Dansyl-YVG. Se sintetizó el tripéptido aprovechando la experiencia que se tiene en nuestro grupo para la síntesis, purificación y plegamiento de péptidos. La síntesis del tripéptido se realizó con éxito (Merrifield B., 1986), sin embargo, la incorporación del grupo Dansilo en el N-terminal (Figura A6.12) está inconclusa al momento de escribir este documento.

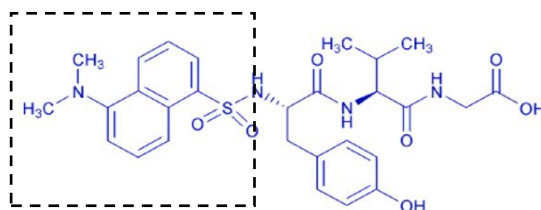


Figura A6.12. Fórmula desarrollada del sustrato “Dansyl-YVG.” Se muestra el grupo dansilo encerrado en un rectángulo punteado.

Anexos

Cartel

Presentado en el XXXII Congreso Nacional de Bioquímica
Ixtapa, Zihuatanejo, Guerrero. 4-9 noviembre 2018

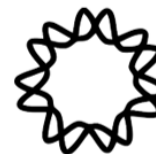
Artículo científico

Publicado en la revista "Toxins"
Factor de impacto 3.531, cuartil Q1



C-terminal amidation in scorpions – Identification of the amidating enzyme (PAM)

Gustavo Delgado-Prudencio¹, Lourival Domingos Possani Postay¹, Baltazar Becerril Lujan¹, Ernesto Ortiz Suri^{1,2,3}
¹Departamento de Medicina Molecular y Bioprocesos, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Avenida Universidad, 2001, Colonia Chamilpa, Cuernavaca, Morelos, 62210, México, ^{2,3}erne@ibt.unam.mx



INTRODUCTION

α -amidation is an important posttranslational modification present in different bioactive peptides of many eukaryotic taxa. Many neuropeptides, hormones and toxins require to be amidated for their biological function. The amidation signals (-X-Gly-Basic- or -X-Gly-Basic) are processed for the production of the amidated peptides (Fig 1 A). The amidation reaction occurs on a glycine residue at the C-terminus of the precursor and involves the sequential action of two enzymes: peptidylglycine α -hydroxylating monooxygenase, PHM (EC 1.14.17.3) and peptidyl- α -hydroxyglycine α -amidating lyase, PAL (EC 4.3.2.5) (Fig 1 B). In mammals, such as *Rattus norvegicus* and *Homo sapiens*, a single gene encodes the enzyme peptidyl-glycine α -amidating monooxygenase, PAM, a bifunctional enzyme that includes the catalytic domains PHM, PAL and a membrane spanning domain in the same polypeptide. Insects, such as *Apis mellifera* and *Drosophila melanogaster*, having lost the bifunctional PAM, have genes separately coding for PHM and PAL instead. Although amidating enzymes are widely distributed in eukaryotic taxa, PAM, PHM and PAL are not present in plants and fungi. No information is available on amidation pathways for arachnids, including scorpions. The objective of this work is to explore the enzymatic amidation system present in the venom glands of scorpions, which contain a wide variety of amidated peptides (e. g. toxins and antimicrobial peptides).

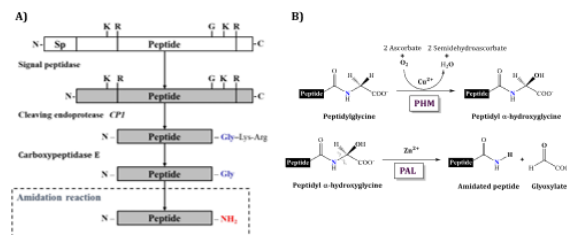
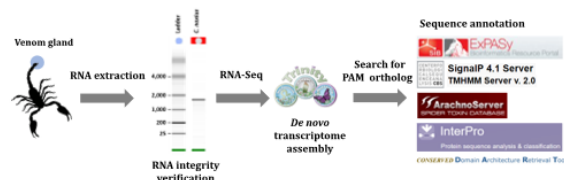


Figure 1 | A) Processing of the amidation signal from a larger inactive precursor; B) Sequential reactions catalyzed by the PHM and PAL enzymes.

METHODS



RESULTS

Sequences coding for bifunctional PAM and monofunctional PHM and PAL were identified in the transcriptomes of 20 scorpion species from the New and Old Worlds (Table 1).

Table 1 | Scorpions Amidating Enzymes.

Family	Species	PAM	PHM	PAL
Buthidae	<i>Centruroides noxius</i>	■	■	■
	<i>Centruroides sculpturatus</i>	■	■	■
	<i>Centruroides limpidus</i>	■	■	■
	<i>Centruroides orizaba</i>	■	■	■
	<i>Centruroides ochraceus</i>	■	■	■
	<i>Centruroides hirsutipalpus</i>	■	✓	■
	<i>Tityus trivittatus</i>	■	■	■
Vaejovidae	<i>Leiurus abduallahbayrami</i> *	■	■	■
	<i>Thorellius cristimanus</i>	■	■	■
	<i>Paravaeovis schwenkmeyeri</i>	■	✓	■
	<i>Chihuahuanus coahuilae</i>	■	■	■
Caraboctonidae	<i>Serradigitus gertschi</i>	■	✓	■
	<i>Hoffmannhadrurus aztecus</i>	■	■	■
Euscorpidae	<i>Hadrurus concolorus</i>	■	■	■
	<i>Megacormus gertschi</i>	■	■	■
Chactidae	<i>Anuroctonus pococki bajae</i>	■	✓	■
	<i>Superstitionia donensis</i>	■	■	■
Diplocentridae	<i>Diplocentrus melici</i>	■	■	■
	<i>Urodacrus yaschenko</i> *	■	■	■
Scorpionidae	<i>Pandinus imperator</i> *	■	■	■

(■) and (■): Indicate complete PAM, PHM and PAL sequences; (■) indicates PAM sequences with > 90% of the estimated total length; (✓) indicates partial sequences with < 50% of the estimated total length; * Old World scorpion.

PAM, PHM and PAL are encoded by independent genes. The PHM and PAL genes are the products of a duplication of the ancestral PAM gene.

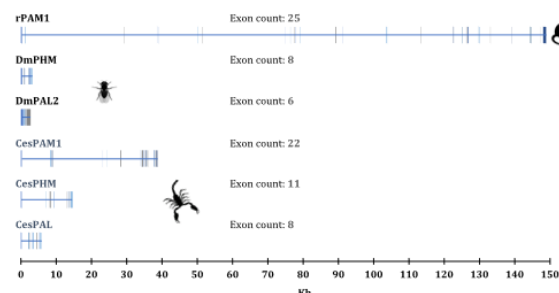


Figure 2 | Genes coding for the amidating enzymes of *R. norvegicus* (rPAM1), *D. melanogaster* (DmPHM and DmPAL2) and *C. sculpturatus* (CesPAM, CesPHM and CesPAL). Exons are indicated as vertical blocks

The topology of the scorpion PAM precursors is identical to that of the PAM-2 isoform described in mammals. Monofunctional PHM and PAL was similar to the precursors described for the insect *D. melanogaster* (Figure 3A and 3B). The sequence analysis showed the conservation of all residues essential for catalysis (Figure 3C).

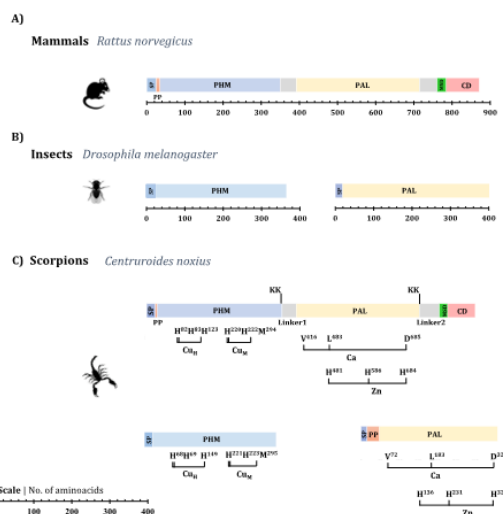


Figure 3 | Representative structure of A) mammalian PAM-2 isoform; B) Mono-functional PHM and PAL2 in *D. melanogaster*; C) precursors of scorpion bifunctional PAM, Monofunctional PHM and PAL; SP, Signal peptide; PP, Propeptide; MSD, Membrane Spanning Domain; CD, Cytosolic Domain.

CONCLUSIONS

We have shown that the amidation pathway in scorpions is conserved. The dual enzymatic amidation system present in the venom glands of scorpions is reminiscent of the mammalian system in terms of the use of both, the bifunctional and the independent enzymes. While the topology of the bifunctional PAM relates to that of the mammalian PAM, the monofunctional enzymes are more similar to the ones in insects.

References

- Eipper, B. A., Milgram, S. L., Jean Hueton, E., Yun, H.-Y., & Mains, R. E. (1993). Peptidylglycine α -amidating monooxygenase: A multifunctional protein with catalytic, processing, and routing domains. *Protein Science*, 2(4), 489–497.
- Attenborough, R. M. F., Hayward, D. C., Kitahara, M. V., Millic, D. J., & Ball, E. E. (2012). A "neuronal" enzyme in nonbilaterian animals and algae: preneuronal origins for peptidylglycine α -amidating monooxygenase. *Molecular Biology and Evolution*, 29(10), 3095–3109.
- Ortiz, E., Guerrero, G. B., Schwartz, E. F., & Possani, L. D. (2015). Scorpion venom components as potential candidates for drug development. *Toxicol*, 93, 125–135.

Funding

GDP is the recipient of CONACYT scholarship 540036

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 G R C H GR	H-NH ₂
	LqhIT2	Q26292	<u>MKLLLLLIVSASMLIESLVNADGYIKRRDGGCKVACLIGNEGCDKECKAYGGSYGYCWTWGLACWCE</u> GLPDDKTWKSETNTCG GKK	G-NH ₂
	BmKITa	Q9XY87	<u>MKLFLLLLISASMLIDGLVNADGYIRGSNGCKVSLWGNEGCNKECRAYGASYGYCWTWGLACWC</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>MNSLLMITACLALVGTVWAKEGYLVNSYTGCKFECFKLGDNDYCLRECRQQYGGKSGGYCYAFG</u> CWCTHLYEQAVVW P L P N K T C N GK	N-NH ₂
	CsEI	P01491	<u>MNSLLMITACLVLIGTVWAKDGYLVEKTGCKKTCYKLGENDFCNRECKWKHIGGSYGYCYGFGC</u> YCEGLPDSTQTW P L P N K T C GKK	C-NH ₂
	CsEv3	P01494	<u>MNSLLMITACLFLIGTVWAKDGYLVNKSTGCKYGCCLKLGENEGCDKECKAKNQGGSYGYCYAFA</u> CWCEGLPESTPTY P L P N K S GKK	C-NH ₂
	Ts1	P15226	<u>MKGMILFISCLLLIGIVVECKEGYLMDEHGCKLSCFIRPSGYCGRECGIKKSSSGYCAWPACYCY</u> GLPNWVKVWDRATN KCGK	C-NH ₂
	Ts3	P01496	<u>LVVVCLLTAGTEGKKDGYPVEYDNCAYICWNYDNAYCDKLCCKDKKADSGYCYVWHILCYCY</u> GLPDSEPTKTNGK CSGKK	S-NH ₂
Potassium toxins	NTx	P08815	<u>MKAFYGILILLFCSMFNLNESTIINVKCTSPKQCSKPCKELYGSSAGAKCMNGKCKCYNNG</u>	N-NH ₂
	BmKTX	Q9NII7	<u>MKVFFAVLITLFCMSIIGHGVGINVKCKHSGQCLKPKDAGMRFKGCINGKCDCTPKG</u>	K-NH ₂
	CoTx1	O46028	<u>MEGIAKITLILLFLVMTMHTFANWNTEAAVCVYRTCDKDKCKRRGYRSGKINNACKCYPYGK</u>	Y-NH ₂
	OcKTx5	Q6XLL5	<u>MNAKFILLLVLTMMMLLPDTKGAEVIRCSGSKQCYGPKQQTGCTNSKCMNKVCKCYCG</u>	C-NH ₂
	OcKTx1	Q6XLL9	<u>MNAKFILLLVVATTMLLPDTQGAEVIKCRTPKDCAGPCRKQTGCPHGKCMNRTCRNCG</u>	C-NH ₂
Non disulfide bridged peptides	IsCT	Q8MMJ7	<u>MKTQFAILLVALVLFQMFAQSDAILGKIWEGIKSLFGKRGLSDLDGLDELFDGEISKADRDFLRELMR</u>	F-NH ₂
	BmKb1	Q718F4	<u>MEIKYLLTVFLVLLIVSDHCQAFSLIPSAISGLISAFKGRRKRDLNGYIDHFKNFRKRDAELEEELLSKLPY</u>	K-NH ₂
	Hp1090	P0DJ02	<u>MKTQFAIFLITLVLFQMFSQSDAIFKAIWSGIKSLFGKRGLSDLDLDESFDGEVSQADIDFLKELMQ</u>	F-NH ₂
	IsCT2	Q8MTX2	<u>MKTQFAILLVALVLFQMFAQSEAFGAIWNGIKSLFGRRALNNDLDDGLDELFDGEISQADVDFLKELMR</u>	F-NH ₂
	VAMP-2	E4VP07	<u>MKSQTFLLFLVVFLLAITQSEAFGAIAGLLKNIFGKRSLRDMDTMKYLYDPSLSAADLKTLQKLMENY</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).

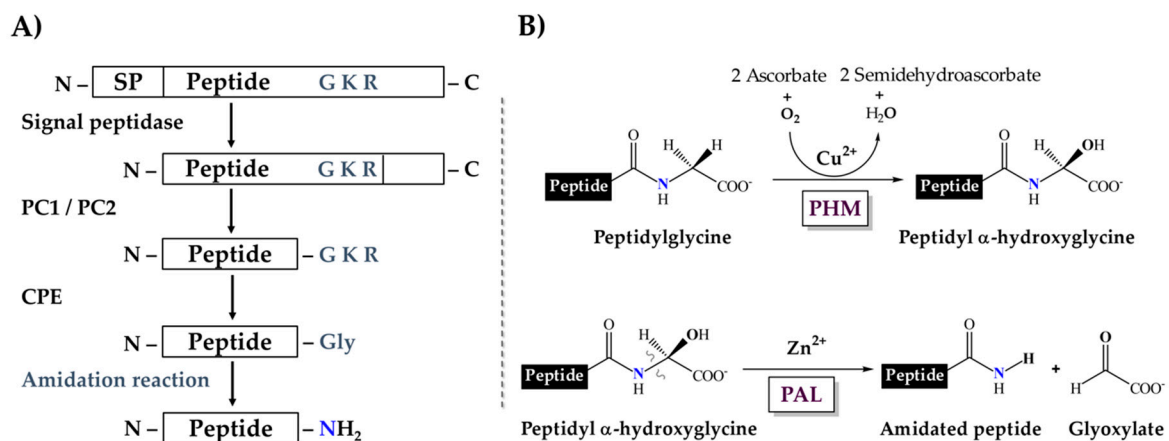


Figure 1. Enzymatic processing of the amidation signal. (A) The complete pathway leading to amidation. SP, Signal Peptide; PC1/PC2, proprotein convertases 1/2; CPE, carboxypeptidase E. (B) Sequential amidation reaction catalyzed by the PHM and PAL domains.

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.

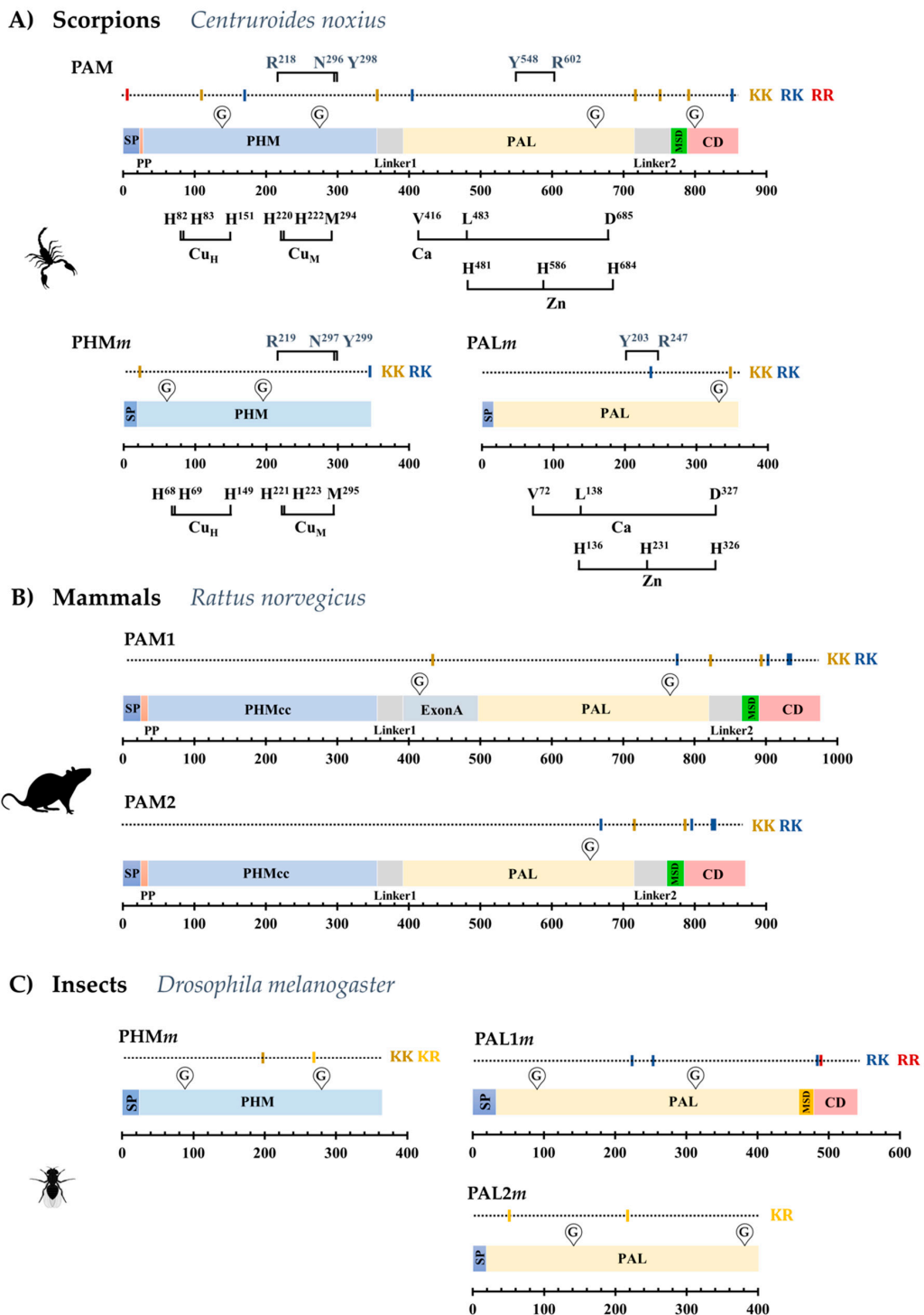


Figure 2. Representative structures of the precursors of (A) scorpion (*C. noxius*) bifunctional PAM and mono-functional PHM m and PAL m ; (B) mammalian (*R. norvegicus*) PAM-1 and PAM-2 isoforms; (C) monofunctional PHM m , PAL1 m and PAL2 m in *D. melanogaster*. Structural and sequence features are 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 PHMm and PALm 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-, PHMm- and PALm-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, PHMm and PALm 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).

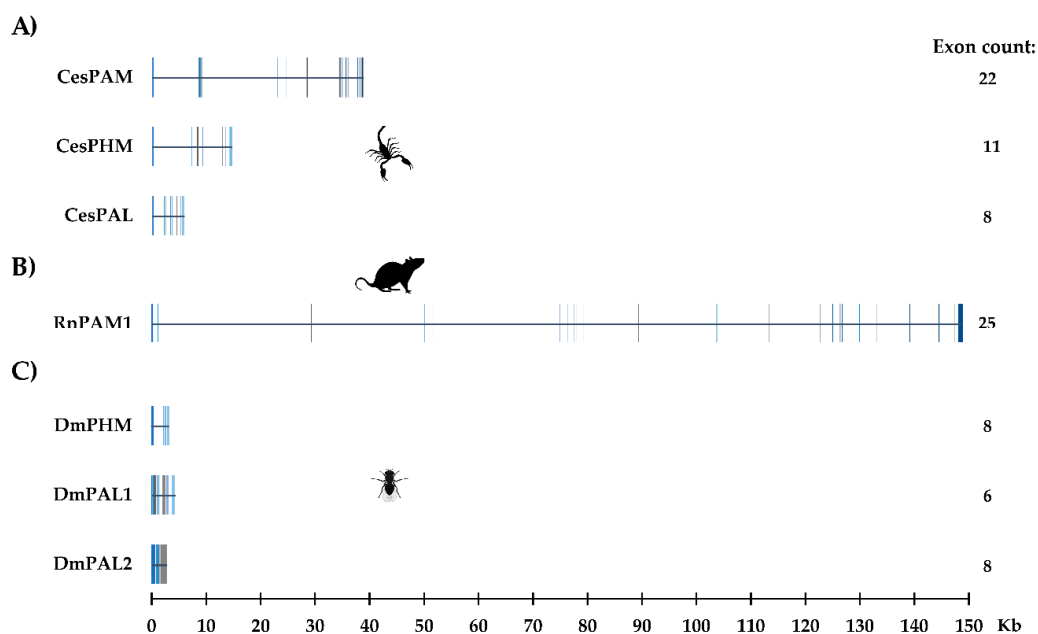


Figure 3. The structure of genes encoding the amidating enzymes of (A) *C. sculpturatus* (CesPAM, CesPHM and CesPAL); (B) *R. norvegicus* (PAM1); (C) *D. melanogaster* (DmPHM, DmPAL1 and DmPAL2). Exons are indicated as vertical blocks.

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 PHMm and PALm 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 PHMm and PALm, 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.

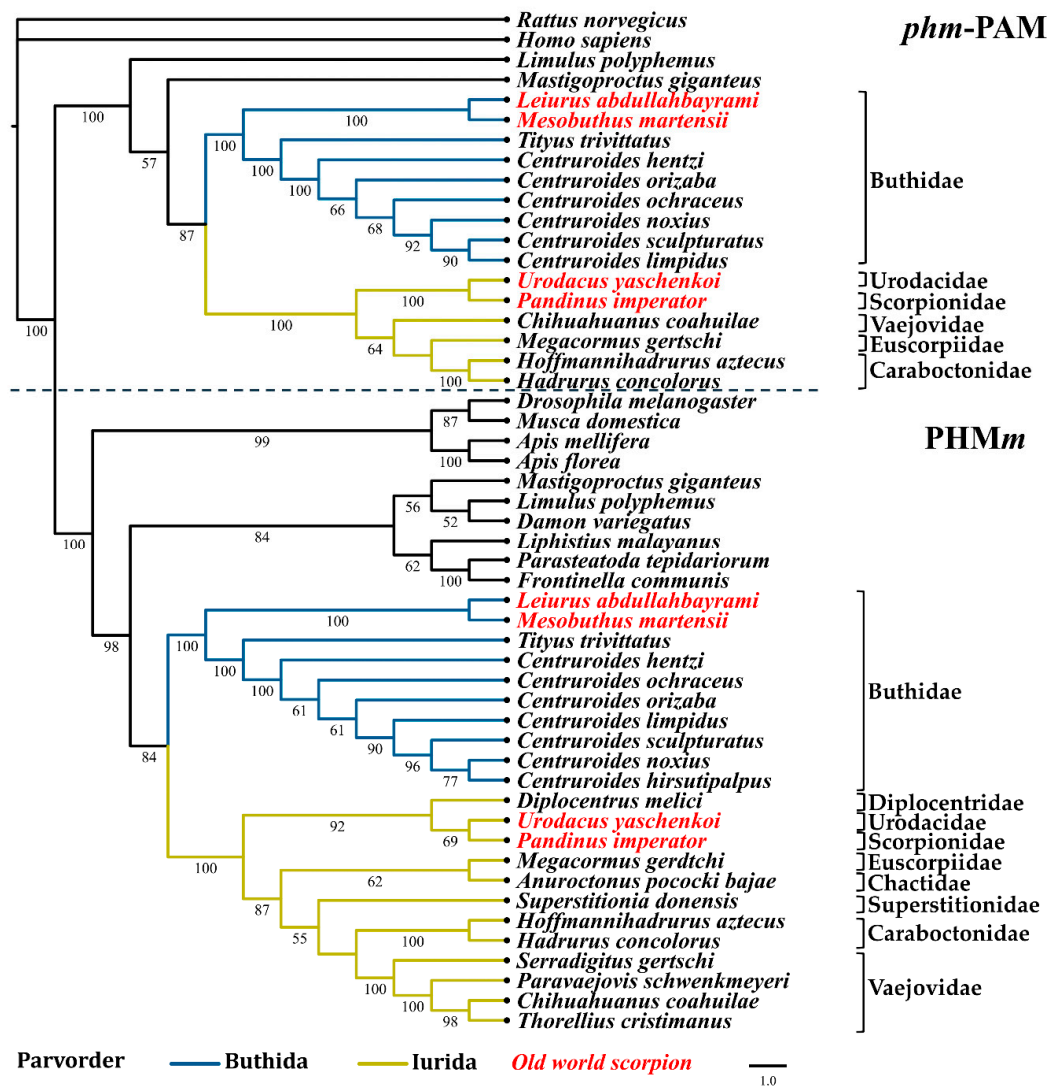


Figure 4. Reconstructed evolutionary history of the phm-PAM and PHMm domains. Maximum likelihood analyses were performed with nucleotide sequences corresponding to the respective domains. Numbers under the nodes indicate the values of ultrafast bootstrap (UFBoot) (only branches with values higher than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

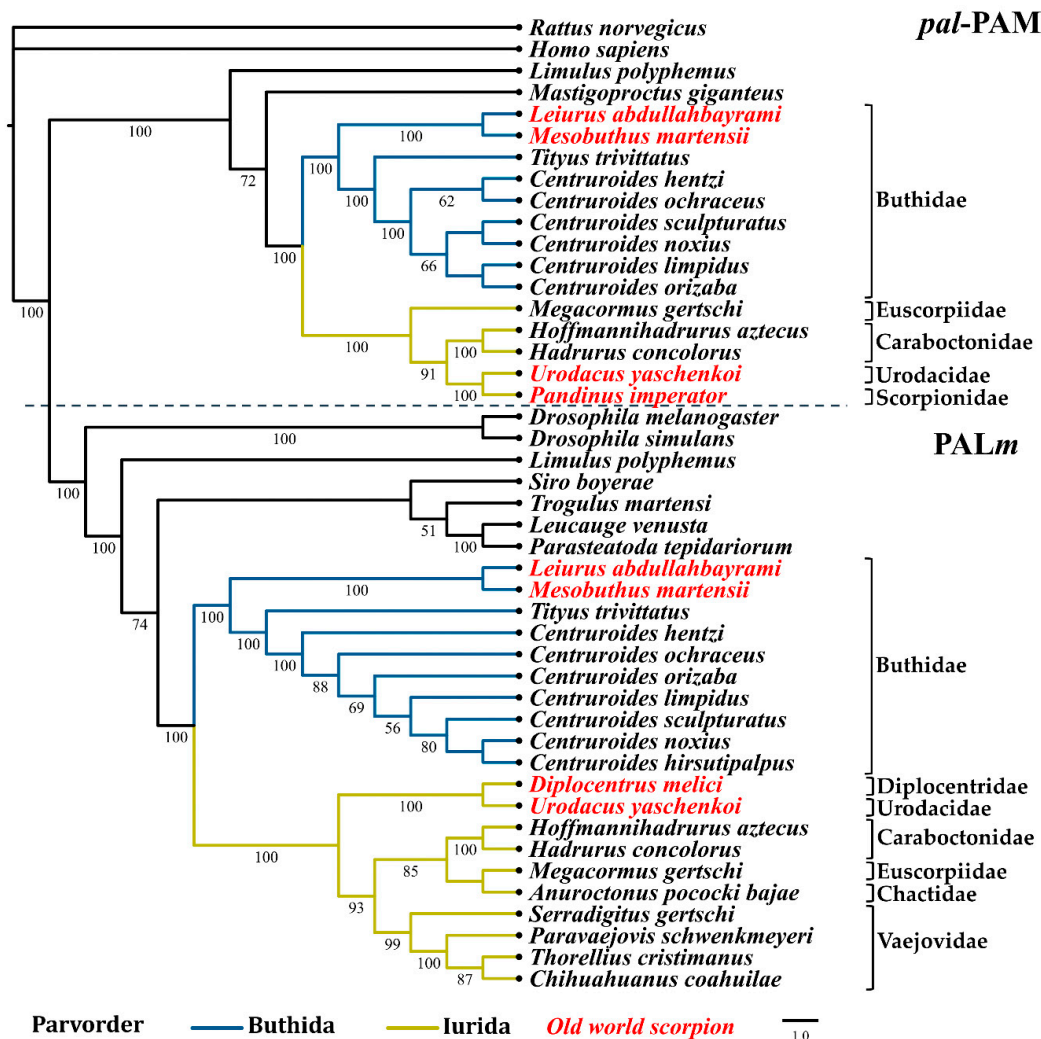


Figure 5. Reconstructed evolutionary history of the pal-PAM and PALm domains. Maximum likelihood analyses were performed with the nucleotide sequences corresponding to the respective domains. Number under the nodes indicate the values of ultrafast bootstrap (UFBoot) (only branches with values higher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

- 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](#)]
- 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](#)]
- 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](#)]
- 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](#)]
- 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](#)]
- 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](#)]
- Gurevitz, M. Mapping of scorpion toxin receptor sites at voltage-gated sodium channels. *Toxicon* **2012**, *60*, 502–511. [[CrossRef](#)] [[PubMed](#)]
- Possani, L.D.; Becerril, B.; Delepierre, M.; Tytgat, J. Scorpion toxins specific for Na⁺-channels. *Eur. J. Biochem.* **1999**, *264*, 287–300. [[CrossRef](#)]
- 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](#)]
- Rodríguez de la Vega, R.C.; Possani, L.D. Current views on scorpion toxins specific for K⁺-channels. *Toxicon* **2004**, *43*, 865–875. [[CrossRef](#)]
- Chippaux, J.P. Emerging options for the management of scorpion stings. *Drug Des. Dev. Ther.* **2012**, *6*, 165–173. [[CrossRef](#)] [[PubMed](#)]
- 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 *Urodacus 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](#)]

