



CENTRO DE INVESTIGACIONES BIOLÓGICAS
DEL NOROESTE, S.C.

Programa de Estudios de Posgrado

METALOPEPTIDASAS DIGESTIVAS DE LA FAMILIA DE
LAS ASTACINAS EN *Penaeus vannamei*.

T E S I S

Que para obtener el grado de

Maestra en Ciencias

Uso, Manejo y Preservación de los Recursos Naturales
(Orientación en Biotecnología)

P r e s e n t a

Ana Patricia Olguín León

La Paz, Baja California Sur, noviembre de 2022.

ACTA DE LIBERACIÓN DE TESIS

En la Ciudad de La Paz, B. C. S., siendo las 9:00 horas del día 24 del Mes de octubre del 2022, se procedió por los abajo firmantes, miembros de la Comisión Revisora de Tesis avalada por la Dirección de Estudios de Posgrado y Formación de Recursos Humanos del Centro de Investigaciones Biológicas del Noroeste, S. C., a liberar la Tesis de Grado titulada:

"METALOPEPTIDASAS DIGESTIVAS DE LA FAMILIA DE LAS ASTACINAS EN *Penaeus vannamei*"

Presentada por el alumno:

Ana Patricia Olguín León

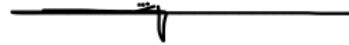
Aspirante al Grado de MAESTRO EN CIENCIAS EN EL USO, MANEJO Y PRESERVACIÓN DE LOS RECURSOS NATURALES CON ORIENTACIÓN EN **BIOTECNOLOGÍA**

Después de intercambiar opiniones los miembros de la Comisión manifestaron su **APROBACIÓN DE LA TESIS**, en virtud de que satisface los requisitos señalados por las disposiciones reglamentarias vigentes.

LA COMISIÓN REVISORA



Dra. Liliana Carolina Rojo Arreola
Directora Tesis



Dr. José Arturo Sánchez Paz
Co-Tutor



Dr. Reinhard Saborowski
Co-Tutor



Dra. Gracia Alicia Gómez Anduro,
Directora de Estudios de Posgrado y
Formación de Recursos Humanos


La Paz, Baja California Sur, a 28 de octubre de 2022.

Los miembros del comité de tesis del (1a) estudiante Ana Patricia Olguín León del Programa de Maestría en Uso, Manejo y Preservación de los Recursos Naturales, revisamos el contenido de la tesis y otorgamos el Vo.Bo., dado que la tesis no representa un plagio de otro documento como lo muestra el reporte de similitud realizado:

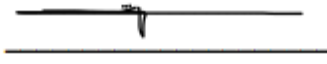
- Herramienta antiplagio:
iThenticate
- Filtros utilizados:
Citas excluidas, bibliografía excluida
- Porcentajes de similitud: 8 %
Se muestra captura de pantalla

Citas incluidas		8%
Bibliografía excluida		8%
Resumen de Coincidencias		
1	Internet 332 palabras Copiado el 15-Oct-2022 cibnorpostdoctoralinstitucional.mx	2%
2	Internet 85 palabras hdl.handle.net	1%
3	Internet 55 palabras Copiado el 12-Abr-2021 springerlink.com	<1%
4	Internet 49 palabras Copiado el 15-Oct-2022 ciad.repositorioinstitucional.mx	<1%
5	Internet 42 palabras Copiado el 22-Jun-2022 bibliotecadigital.exactas.uba.ar	<1%
6	Internet 41 palabras Copiado el 25-Nov-2020 koc.pub	<1%
7	Internet 39 palabras Copiado el 19-Oct-2022 www.frontiersin.org	<1%
8	Internet 29 palabras Copiado el 08-Feb-2019 kamilbodo.blogspot.com	<1%
9	Internet 28 palabras Copiado el 15-Oct-2022 ipcyt.repositorioinstitucional.mx	<1%
10	Internet 21 palabras Copiado el 01-Oct-2021 academic.oup.com	<1%
11	Internet 21 palabras Copiado el 24-Oct-2022 cioyrepositorioinstitucional.mx	<1%
12	Internet 21 palabras Copiado el 24-Oct-2022 revistas.sgoeru.sjg.pe	<1%
13	Internet 20 palabras Copiado el 26-Oct-2022 ri.usamex.mx	<1%
14	Internet 18 palabras Copiado el 21-Jul-2016 www.scribd.com	<1%


Firmas del comité



Dra. Liliana Carolina Rojo Arreola



Dr. José Arturo Sánchez Paz



Ph.D. Reinhard Saborowski

Conformación de Comités

Comité tutorial

Dra. Liliana Carolina Rojo Arreola
Centro de Investigaciones Biológicas del Noroeste S. C.
Directora de Tesis

Dr. José Arturo Sánchez Paz
Centro de Investigaciones Biológicas del Noroeste S. C.
Co-Tutor de Tesis

PhD. Reinhard Saborowski
Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research
Co-Tutor de Tesis

Comité Revisor de Tesis

Dra. Liliana Carolina Rojo Arreola
Dr. José Arturo Sánchez Paz
PhD. Reinhard Saborowski

Jurado de Examen

Dra. Liliana Carolina Rojo Arreola
Dr. José Arturo Sánchez Paz
PhD. Reinhard Saborowski

Suplente

Dra. Martha Patricia Hernández Cortés

Resumen

Las enzimas de la familia de las astacinas (M12) son metalopeptidasas extracelulares presentes en bacterias, invertebrados y vertebrados. Estas enzimas tienen una amplia diversidad estructural y funcional, en crustáceos se han descrito como enzimas participantes en la digestión del alimento, morfogénesis y formación de la matriz extracelular. Además, en *Penaeus vannamei*, se ha identificado una zinc metalopeptidasa cuya actividad está relacionada a los procesos de ontogenia. Esta enzima se expresa en tejidos no digestivos, sin embargo, la expresión génica es significativamente mayor en la glándula digestiva.

En el presente estudio, se identificaron secuencias de astacinas en varias especies de decápodos, las secuencias fueron localizadas en la base de datos de proteínas Non Redundant Protein Sequences del NCBI, usando la herramienta BLASTP y con base a la astacina de *Astacus astacus*, el prototipo de la familia M12 y la de mayor evidencia experimental. Se confirmó la presencia de astacinas varias especies de decápodos, al menos un representante en las familias Astacidae, Portunidae, Oregoniidae, Cambaridae, Nephropidae, Lithodidae y Penaeidae.

Además, se realizó un análisis más detallado de las secuencias de zinc metalopeptidasas de *P. vannamei*. Las secuencias fueron localizadas en la base de datos de baja redundancia UniProt, y fueron caracterizadas *in silico*, de esta manera se identificaron los patrones de variación de las secuencias y la heterogeneidad estructural de las proteínas, así como las variaciones de la estructura primaria y su efecto en la estructura secundaria y terciaria.

Se identificaron cuatro tipos de arquitecturas en las secuencias de astacinas de *P. vannamei*, que además de presentar los dominios catalíticos Astacina, se diferencian por la presencia de los dominios (módulos) no catalíticos CUB, ShK y EGF. El tipo I consta de 1 o 2 dominios CUB, el grupo II presenta 1 o 3 dominios ShK. En este trabajo se reporta por primera vez esta arquitectura de la astacina de un invertebrado no hematófago. El grupo III contiene a las astacinas monodominio, y finalmente el grupo IV incluye a secuencias con el dominio astacina trunco (no funcional) y aquellas que incluyen dominios EGF.

Finalmente, se realizó una búsqueda de astacinas monodominio de *P. vannamei* en las bases de datos de nucleótidos Reference RNA sequences y Transcriptome Shotgun Assembly de NCBI, para ello se usó la herramienta tBlastn con base a la secuencia de la metalopeptidasa Mpc1 de *P. vannamei*. Se encontraron varias secuencias que se agruparon en 16 variantes, a las cuales se les realizaron análisis funcionales *in silico*. Estas secuencias presentan características típicas de astacinas digestivas: El motivo (HEXXHXXGFXHEXXRXDR), la secuencia de giro de metionina (Met-turn, SXMHY), el dominio catalítico de astacina, el péptido señal y el propéptido. Las secuencias de aminoácidos tienen una variabilidad mayor al 5 % por lo que se propone que son isoformas y su función es digestiva.

Palabras clave: astacina, metalopeptidasas, zinc peptidasas, *Penaeus vannamei*

ORCID: 0000-0002-5583-8789



Dra. Liliana Carolina Rojo Arreola
Vo.Bo. Directora de Tesis

Summary

The enzymes of the astacin family (M12) are extracellular metalloproteinases present in bacteria, invertebrates and vertebrates. A wide structural and functional diversity have been described for the M12 family, in crustaceans they are involved in food digestion, morphogenesis and the formation of the extracellular matrix. Furthermore, in *Penaeus vannamei*, a zinc metalloproteinase has been identified with activity is related to the ontogeny development. This enzyme is expressed in non-digestive tissues; however, gene expression is significantly higher in the digestive gland.

In the present study, astacin sequences from several species of decapods were identified, the sequences were located in the NCBI Non-Redundant Protein Sequences database, using the BLASTP tool, the sequence of *Astacus astacus* astacin (the prototype of the M12 family and the one with the most experimental evidence) was used as query. The presence of astacins was confirmed in at least one species from the Astacidae, Portunidae, Oregoniidae, Cambaridae, Nephropidae, Lithodidae and Penaeidae families.

A more detailed analysis of *P. vannamei* zinc metalloproteinases sequences was performed. The sequences were located in UniProt, a protein database with the low redundancy, and characterized *in silico*. The patterns of sequence variation and the structural heterogeneity of the proteins were identified, as well as the variations of the primary structure and its effect on the secondary and tertiary structure.

We identified four main architectures in astacin sequences of *P. vannamei*, which in addition to the Astacin catalytic domains, the non-catalytic domains CUB, ShK and EGF are present. Type I consists of 1 or 2 CUB domains, group II has 1 or 3 ShK domains, in this work, this architecture is reported for the first time in a non-hematophagous invertebrate. Group III contains single-domain astacins, and finally, group IV includes sequences with an incomplete astacin domain (non-functional) and those bearing EGF domains.

Finally, a search for *P. vannamei* monodomain astacins was performed in the NCBI Reference RNA sequences and Transcriptome Shotgun Assembly nucleotide databases, for this, the tBlastn tool and *P. vannamei* metalloproteinase Mpc1 as query were used. Several sequences were found, they were grouped into 16 variants. *In silico* functional analyzes were performed on these. These sequences exhibit the characteristics of astacin: The motif (HEXXHXXGFXHEXXRXDR), the methionine turn sequence (Met-turn, SXMHY), the astacin catalytic domain, the signal peptide, and the propeptide. The amino acid sequences showed a variability greater than 5%, so they can be consider as isoforms with a putative digestive function.

Keyword: astacin, metalloproteinase, zinc metalloproteinase, *Penaeus vannamei*

ORCID: 0000-0002-5583-8789



Dra. Liliana Carolina Rojo Arreola
Vo.Bo. Directora de Tesis

Dedicatoria

“La mejor herencia que te puedo dejar es la educación”

Javier Olgún Ríos

A mi familia

A Leo

Agradecimientos

Al Centro de Investigaciones Biológicas del Noroeste, S.C., por ser la institución receptora de los estudios de Posgrado y el apoyo otorgado No. 1247 para la culminación de tesis y al Consejo Nacional de Ciencia y Tecnología por la beca (1009085) otorgada para la realización del Posgrado.

Un agradecimiento especial a todo el personal de Posgrado en CIBNOR.

Quiero expresar mi agradecimiento al Dr. Fernando Luis García Carreño por haberme dado la oportunidad de involucrarme en este tema y por su estimulante apoyo intelectual. Al equipo de trabajo del Laboratorio de Bioquímica, Ann Navarrete y el Dr. Julio Humberto Córdova.

Un agradecimiento especial a mi Comité Tutorial, por su apoyo ante la crisis y no dejarme atrás. Además, gracias por el gran apoyo para la culminación de este trabajo.

Dra. Liliana Carolina Rojo, quien, con su empatía y comprensión, me ayudó cuando el mundo se me venía encima. Gracias doctora, por entender el proceso en el cual estaba mi vida en estos años y ayudarme a seguir adelante.

PhD. Reinhard Saborowski, quien sin conocerme aceptó ser parte de mi comité y sus comentarios ayudaron a este trabajo.

Dr. José Arturo Sánchez, gracias de corazón. Usted ha sido mi coach todo este tiempo, desde que nos conocimos hace muchos años. Tuve la fortuna de tenerlo de Director de Tesis en la Licenciatura y ahora, de nuevo, coincidimos en la etapa de Maestría. Gracias infinitas por sus múltiples y minuciosas revisiones, que fueron alrededor de 10 para este escrito, siempre con el afán de enseñarme a crecer, ser mejor persona y ser una mejor profesional. Gracias por su amistad y confianza; siempre valoro las largas horas de escuchar música para concentrarnos, sus recomendaciones de libros, películas y series, nuestras pláticas y discusiones y, sobre todo, muchísimas gracias por nunca dejarme renunciar y acompañarme en este proceso.

A mis amigos y compañeros de Maestría *“Los Miraculous”*: Aldahir Núñez, Ayenía Rosales, Bárbara Cúmez, Daniela Montealegre, David Rosales, Giovanni Quintino, Jaqueline Ramírez y Miranda Villa. Su compañía me ayudó en los momentos más difíciles, gracias infinitas.

A Alejandro Palomares y Mónica Buendía, mis amiguitos del Lab. Gracias por enseñarme y guiarme a través del Laboratorio y darme su amistad. Me divertí mucho en su compañía, con nuestras pláticas y aventuras.

Quiero expresar un agradecimiento muy especial al equipo del Laboratorio de Virología del CIBNOR Campus Hermosillo, que siempre me abre las puertas y siempre están dispuestos a ayudar.

A Trinidad Encinas, por resolver mis dudas y generarme algunas más. Tus conversaciones y apoyo ayudaron a este proyecto. A Fernando Mendoza, Juan Bautista Vega y Doña Claudia por hacer las tareas difíciles más amenas y divertidas con la plática y compañía.

A mi Psicóloga Elsa Figueroa que, con su trabajo, me ayudó a conocerme mejor y abrirme los ojos para descubrir mis habilidades y capacidades que en el proceso había olvidado que tenía y, sobre todo, a darme cuenta que puedo cumplir lo que me proponga.

A todos los Investigadores que han contribuido a alimentar las bases de datos de nucleótidos y aminoácidos y todos los transcriptomas: gracias por su generosidad. Sin ustedes este trabajo de bioinformática no hubiera sido posible.

Gracias a mi Familia, quienes son el fundamento de mi persona y siempre están conmigo. A mis amigas Natalia Cuevas y Beatriz Ramonetti, por todo el apoyo y su amistad.

Quiero mencionar con mucho cariño a Leonardo Luken, Leo de mi corazón, tú siempre has creído en mí. Tus aportaciones durante estos años en el programa de Maestría son valiosísimas para este proyecto. Gracias por acompañarme y apoyarme cuando te dije que era hora de entrar a la Maestría, sin pensarlo me ayudaste a hacerlo posible. Te agradezco cada uno de los minutos, horas, días y meses que pasamos a distancia para que yo pudiera cumplir este sueño, no fue nada fácil estar lejos de ti, pero tu hiciste posible que nos sintiéramos más cerca. Gracias por tu infinito amor, paciencia, por ser mi persona favorita, por acompañarme en estos años donde mi vida ha cambiado de mil maneras, por darme confianza en todas las decisiones que he y hemos tomado juntos.

Finalmente, quiero agradecer a todas las personas que no mencioné. Una disculpa por mi falta de memoria. Muchas gracias.

Contenido

Resumen	i
Summary	ii
Dedicatoria.....	iii
Agradecimientos.....	iv
Contenido.....	vi
Lista de figuras	viii
Lista de tablas	x
1. INTRODUCCIÓN	1
2. ANTECEDENTES.....	3
2.1 El orden Decápoda.....	3
2.2 Sistema digestivo de decápodos	5
2.3 Enzimas digestivas (peptidasas en Decápoda).....	7
2.4 Metalopeptidasas	10
2.5 Familia Astacina.....	16
2.6 Proteasa Astacina	18
3. JUSTIFICACIÓN.....	21
4. HIPÓTESIS	22
5. OBJETIVO	23
5.1 Objetivo general	23
5.2 Objetivos particulares.....	23
6 MATERIAL Y MÉTODOS	24
6.1 Identificación de zinc metalopeptidasas con dominio astacina con bases de datos de genes y proteínas.....	24
6.1.1 <i>Identificación de zinc metalopeptidasas de la familia Astacina (M12) en Decápodos.....</i>	24
6.1.2 <i>Identificación de zinc metalopeptidasas de la familia Astacina (M12) y multidominio en Penaeus vannamei.....</i>	24
6.1.3 <i>Identificación de zinc metalopeptidasas monodominio de astacina en Penaeus vannamei.....</i>	24
6.2 Traducción de secuencias de nucleótidos a proteínas y selección de secuencias.....	25
6.3 Alineamiento de secuencias y cladogramas.....	26
6.3.1 <i>Alineamiento de secuencias de zinc metalopeptidasas de decápodos</i>	26
6.3.2 <i>Alineamiento de secuencias con dominio astacina de zinc metalopeptidasas y multidominio en Penaeus vannamei.....</i>	26
6.3.3 <i>Alineamiento de secuencias de zinc metalopeptidasas monodominio de astacina en Penaeus vannamei</i>	26
6.4 Análisis fisicoquímicos	26
6.5 Variabilidad de proteínas	28
6.6 Análisis filogenético.....	28
6.6.1 <i>Análisis filogenético para zinc metalopeptidasas con dominio astacina en decápodos</i>	28
6.6.2 <i>Análisis filogenético para zinc metalopeptidasas con dominio astacina y multidominio en Penaeus vannamei.....</i>	28

6.6.3	<i>Análisis filogenético para zinc metalopeptidasas monodominio de astacinasa en Penaeus vannamei</i>	29
6.7	Modelo estructural.....	29
7	RESULTADOS	30
7.1	<i>Secuencias de astacinas (M12) en Decápodos</i>	30
7.2	<i>Secuencias de astacinas (M12) mono y multidominios en Penaeus vannamei</i>	30
7.3	<i>Secuencias de astacinas (M12) monodominio en Penaeus vannamei</i>	31
7.4	<i>Alineamiento de astacinas monodominio en Penaeus vannamei</i>	34
7.5	Características fisicoquímicas identificadas <i>in silico</i>	38
7.6	Variabilidad de proteínas	42
7.7	<i>Filogenia de las astacinas de la familia M12 en decápodos</i>	43
7.8	<i>Modelo estructural de la astacina M12 de P. vannamei</i>	44
8.	DISCUSIÓN	46
8.1	Secuencias de astacinas monodominio de <i>P. vannamei</i>	48
9.	CONCLUSIÓN	55
10.	LITERATURA CITADA	56

Lista de figuras

- Figura 1.** Cuerpo segmentado de los decápodos. El cefalotórax está compuesto por 13 segmentos: caparazón, antenas, anténulas, maxilípedos, pereiópodos, quelas, ojos y rostro. La región del abdomen está compuesta por el pleuron, telson, urópodos, endópodos, exópodos, pleópodos y el petasma en el caso de los machos.4
- Figura 2.** Sistema digestivo de los decápodos a escala macroscópica. La región anterior se compone por el aparato bucal el esófago y el estómago; la región media se conforma por la mayor parte del tubo intestinal, la glándula digestiva o hepatopáncreas y dos o tres ciegos intestinales; la región posterior está formada por la última parte del intestino, el recto y el ano. Foto modificada de FAO 2022.....6
- Figura 3.** Sistema digestivo de decápodos, dividido en: intestino anterior compuesto por esófago (EF), estomago cardiaco (EC) y estómago pilórico (EP); el intestino medio lo compone el hepatopáncreas (HP), el tubo de intestino medio (TIM), así como el ciego dorsal anterior (CDA) y posterior (CDP); el último apartado lo compone el intestino posterior, el cual se encuentra revestido por una cutícula. Modificado de: Vogt, 2021.....7
- Figura 4.** Estructura y clasificación básica de peptidasas. Tomado de: Machado de Castro *et al.*, 2018.....9
- Figura 5.** Representación esquemática de la estructura de astacina (1AST, PDB). Detalle del sitio activo de la incisión de astacina. En azul se representan las histidinas que están unidas al zinc (esfera gris), en amarillo la metionina M147 del Met-turn (giro de metionina) además de la tirosina 149 señalado en rojo. El ácido glutámico (E93) que corresponde al sitio activo se encuentra indicado en color rojo. Creado en Mol* en PDB (Sehna et al., 2021).11
- Figura 6.** Clasificación de la subclase de metalopeptidasas mononucleares (que requieren un sólo ion metálico). Solo se representan las tribus $\alpha\beta\alpha$ -exopeptidasas, metalopeptidasas LAS y las zincinas. A su vez, las zincinas se dividen en dos clanes: metzincinas y gluzincinas. El clan metzincina contiene la familia de las astacinas.14
- Figura 7.** Composición modular de las astacinas. En la imagen están representados los distintos dominios que se han observado en los miembros de la familia de las astacinas. El péptido (negro) y propéptido (gris claro) señal N-terminal. Los dominios C-terminal se encuentran unidos al dominio de la proteasa catalítica (rojo) y estos son el dominio ShK (morado), CUB (verde), LC (rojo oscuro), EGF (azul), TSP (café), MAM (amarillo), TM (gris oscuro), C (naranja). Imagen tomada de Gomis-Rüth et al., 2012.17
- Figura 8.** Vista 3D de la cadena A de astacina (PDB 1AST).20
- Figura 9.** Diagrama esquemático de los componentes estructurales de varios miembros representativos de la familia astacina. La estructura secundaria se compone de alfa hélices (cilindros) y hojas beta(flechas), las tijeras indican el sitio de corte de activación final. La región de color naranja corresponde al péptido señal y al propéptido; la región azul es la proteína madura y dominio catalítico de astacina. En fondo negro se encuentran la secuencia del propéptido y el residuo conservado de aspártico (en amarillo), el sitio de activación, las cisteínas conservadas que hacen puentes disulfuro, la región del motivo, el giro de metionina (met-turn) y en rosa el subsitio S_1' . Tomado de Guevara *et al.*, 2010.....27

- Figura 10.** Cladograma y representación esquemática de secuencias homólogas y distribución de los dominios en metalopeptidasas de la familia de las astacinas en *P. vannamei*. La escala del diagrama está relacionada con la longitud de la secuencia. Pre y proproteínas, recuadros amarillos; dominio de astacina, rectángulo verde; dominios CUB, rectángulos rojos CUB; dominio ShKT, cuadros rojos; dominio EGF, cuadros morados; secuencias de baja complejidad, en azul; no especificado, en gris.31
- Figura 11.** Alineamiento de secuencias de astacinas monodominio de *P. vannamei*. En recuadros negros se enmarcan varios sitios de la estructura de astacina. Péptido señal, el propéptido y el sitio de activación, cisteínas que participan en la formación de enlaces disulfuro, el motivo donde se señala el sitio activo y los sitios de unión al zinc, el giro de metionina y el subsitio S1'.37
- Figura 12.** Cladograma de secuencias homólogas de zinc metalopeptidasas monodominio de la familia astacina monodominio de *P. vannamei*.37
- Figura 13.** Alineamiento de las secuencias de las variantes de astacinas monodominio de *P. vannamei* que contienen el propeptido y péptido señal. En los recuadros rojos se indica el sitio de corte del péptido señal y en los recuadros azules el pro péptido38
- Figura 14.** Gráfico de la variabilidad en la composición de las astacinas de *P. vannamei* con relación a los datos de la Tabla 5.41
- Figura 15.** Índice de variabilidad de proteínas Wu-Kabat para cada uno de los 205 residuos de aminoácidos de las zinc metalopeptidasas monodominio de astacina.42
- Figura 16.** Secuencia consenso de la variabilidad de las secuencias de zinc metalopeptidasas monodominio astacinas. En rojo: cisteínas, motivo y giro de metionina.42
- Figura 17.** Análisis filogenético de secuencias de astacinas de diferentes especies del orden Decápoda. Se incluyen los valores de Bootstrap. Cada una de las cuatro ramas principales en donde se agrupan las secuencias de astacinas de diversos decápodos está representada por colores diferentes. El grupo A se representa de color azul, el grupo B de color rojo, el grupo C de color verde y el grupo D de color morado.44
- Figura 18.** Modelo de homología la MPC1 de *P. vannamei* con una molécula de zinc como ligando. El modelo se construyó utilizando la estructura de rayos X la astacina de *Astacus astacus*. A) Las placas- β y las α -hélices se muestran en púrpura y cian, respectivamente. Se resaltan las regiones del motivo HEXXHXXGXXHEXXRXDR (rojo). Las cadenas laterales de histidinas que se unen al zinc (azul), la región del met-turn (verde), cisteínas unidas por enlaces disulfuro (bastones naranjas), el ion de zinc se muestra como esfera plateada. B) Acercamiento al ion de zinc catalítico en donde se distingue que está putativamente asociado a las 3 histidinas del motivo catalítico, los iones hidrógeno de las histidinas que estabilizan el sitio catalítico se representan como bastones blancos45
- Figura 19.** Regiones variables (en color rojo) de la secuencia del motivo de las secuencias de astacinas del grupo C de la Fig. 1753
- Figura 20.** Regiones variables (en color rojo) de la secuencia del motivo de las secuencias de astacinas del grupo D de la Fig. 17.53

Lista de tablas

Tabla 1. Jerarquía taxonómica del orden Decápoda de acuerdo con Bowman y Abele (1982) y McLaughlin et al., (2005).....	3
Tabla 2. Clasificación de Metalopeptidasas MEROPS.....	15
Tabla 3. Números de acceso (GenBank) de las secuencias de las variantes de zinc metalopeptidasas monodominio de <i>P. vannamei</i> seleccionadas para el análisis, se indica el porcentaje de similitud con respecto a astacina.	33
Tabla 4. Número de aminoácidos (# aa), peso molecular (MW) y punto isoeléctrico teórico de las secuencias de variantes de astacinas.....	39
Tabla 5. Composición de aminoácidos de las secuencias de zinc metalopeptidasas en <i>P. vannamei</i> representados en porcentajes.	40
Tabla 6. Fragmentos conservados con residuos no variables. ($V \leq 1$).....	43

1. INTRODUCCIÓN

A lo largo de la historia evolutiva, los organismos que han habitado nuestro planeta han enfrentado diversos desafíos, uno de ellos es la obtención de nutrientes que les permita generar la energía para realizar las funciones bioquímicas, inmunes y fisiológicas necesarias para sobrevivir y reproducirse. Es por ello que muchos organismos presentan diversas adaptaciones en sus sistemas digestivos que les permiten optimizar la cantidad de energía obtenida del alimento y canalizarla para sobrevivir.

Los organismos del orden Decápoda (langostas, cangrejos, langostinos, y camarones, entre otros) son un claro ejemplo de éxito evolutivo, pues se los encuentra en la mayoría de los ecosistemas marinos (desde el litoral hasta las profundidades del mar), hábitats de agua dulce e incluso en ambientes terrestres. Los crustáceos destacan por su compleja historia de divergencia, presentando un amplio margen de adaptabilidad al medio ambiente, así como a las fuentes de nutrientes. Así mismo, la mayoría de las especies exhiben hábitos alimenticios del tipo omnívoro y en menor medida, herbívora, carnívora y detritívora (De Grave et al., 2009; Mente, 2008; Deshmukh, 2013; Štrus et al., 2019).

Las especies del orden Decápoda tienen una creciente importancia ecológica y económica que, a la vez, es relevante social y culturalmente. La acuicultura es el sector de producción de alimentos con el crecimiento más acelerado. En 2018, el cultivo de crustáceos representó casi el 12 % de la producción mundial en toneladas de organismos cultivados en el mundo con 9.4 millones de toneladas cultivadas, mientras que en ese mismo año se capturaron 6 mil toneladas más de crustáceos. Durante 2018, México ocupó el séptimo lugar en la producción mundial de crustáceos, con 157.9 mil toneladas cultivadas (FAO, 2020). Cabe mencionar que el cultivo de la especie de este estudio, *P. vannamei*, aporta en gran medida (en algunos lugares hasta el 90 %) a esta producción.

La descripción del proceso digestivo en crustáceos permite a los productores la formulación de alimentos funcionales y mejora en la eficiencia de la cadena productiva. Una parte importante en la digestión son las proteasas las cuales, son enzimas que hidrolizan proteínas y contribuyen en la degradación de proteínas lo cual puede mejorar o perder la calidad del organismo para su venta.

Este trabajo se enfocó en el estudio de las zinc metalopeptidasas de la familia de las astacinas que se encuentran en *Penaeus vannamei*, además de una visión global de diferentes decápodos, con el objetivo de analizar, a través de herramientas bioinformáticas, los patrones de variación, la arquitectura de las enzimas, los efectos funcionales, y enfatizar en las metalopeptidasas digestivas del camarón blanco.

Una de las principales aportaciones de esta tesis es la identificación de 16 secuencias únicas de astacina monodominio en camarón blanco, *P. vannamei*, las cuales pueden ser consideradas isoformas, debido a la alta variabilidad entre ellas, además, se sugiere que su función es digestiva. Por otro lado, se localizaron también secuencias de astacina con múltiples dominios, que presentan diferentes tipos de arquitecturas y que se agrupan en 4 grupos. Se propone que la función de estas es diferente a la digestiva. El grupo I de 1 o 2 dominios CUB, el grupo II con 1 o 3 dominios ShK, el grupo III son secuencias de 1 dominio astacina (monodominio) y finalmente, el grupo IV se consideran no funcionales.

2. ANTECEDENTES

Durante la larga historia de la vida en el planeta, ocurrieron eventos críticos que provocaron tanto procesos de extinción, como circunstancias que propiciaron el surgimiento de nuevas especies. Actualmente podemos encontrar un gran número de agrupaciones taxonómicas dentro del reino animal, pero uno en especial se considera de gran éxito evolutivo, dadas las adaptaciones anatómicas, fisiológicas y bioquímicas que son resultado de la presión de selección ejercida por los diversos ecosistemas en que habitan: los crustáceos. Este grupo taxonómico se caracteriza por presentar un exoesqueleto duro, cuerpo segmentado y extremidades articuladas (Thiel y Poore, 2020)

2.1 El orden Decápoda

Los crustáceos del orden Decápoda constituyen el taxón más diverso del subfilo Crustácea (Tabla 1). Estos organismos, se han diversificado a través de millones de años, llegando a dominar una extraordinaria variedad de hábitats; se les puede encontrar en litorales y en el fondo marino, así como otros ecosistemas como arrecifes, manglares, estuarios, depósitos de agua dulce e incluso, en ambientes semiterrestres. Si bien dentro de este grupo de animales se encuentran especies tan reconocidas como los camarones, langostas y cangrejos, también se incluyen cerca de 15,000 especies actuales, además de otras 3,000 extintas, muchas de las cuales poseen peculiaridades anatómicas únicas (Bracken *et al.*, 2009; Wolfe *et al.*, 2019).

Tabla 1. Jerarquía taxonómica del orden Decápoda de acuerdo con Bowman y Abele (1982) y McLaughlin *et al.*, (2005).

Reino	Animalia
Subreino	Bilateria
Infrareino	Protostomia
Superfilo	Ecdysozoa
Filo	Arthropoda
Subfilo	Crustácea
Clase	Malacostraca
Subclase	Eumalacostraca
Superorden	Eucarida
Orden	Decápoda
Suborden	Dendrobranchiata
	Pleocyemata

La característica principal de este grupo, y por la cual se deriva la denominación de Decápoda, es que poseen 10 pares de pereiópodos, o patas locomotoras. Los decápodos tienen exoesqueletos compuestos de quitina, así como el cuerpo segmentado y apéndices articulados que están conectados a través de membranas. El cuerpo en estos organismos se divide en cefalotórax (que comprende los segmentos de la cabeza y el tórax, que están fusionados) y el abdomen.

El cefalotórax se compone de 13 segmentos fusionados que están compuestos por ocho pares de fragmentos torácicos, donde los primeros tres segmentos están modificados para la alimentación y se conocen como maxilípedos. Los últimos cinco segmentos torácicos son conocidos como pereiópodos y se emplean en la locomoción del organismo (Feldman, 2003).

Por su parte, el abdomen incluye seis segmentos en los que se encuentran los pleópodos (cinco pares de apéndices abdominales, es decir, los apéndices del pleon) que son utilizados en la natación, copulación o transporte de los huevos, y termina en un telson (Fig. 1) (Feldman, 2003).

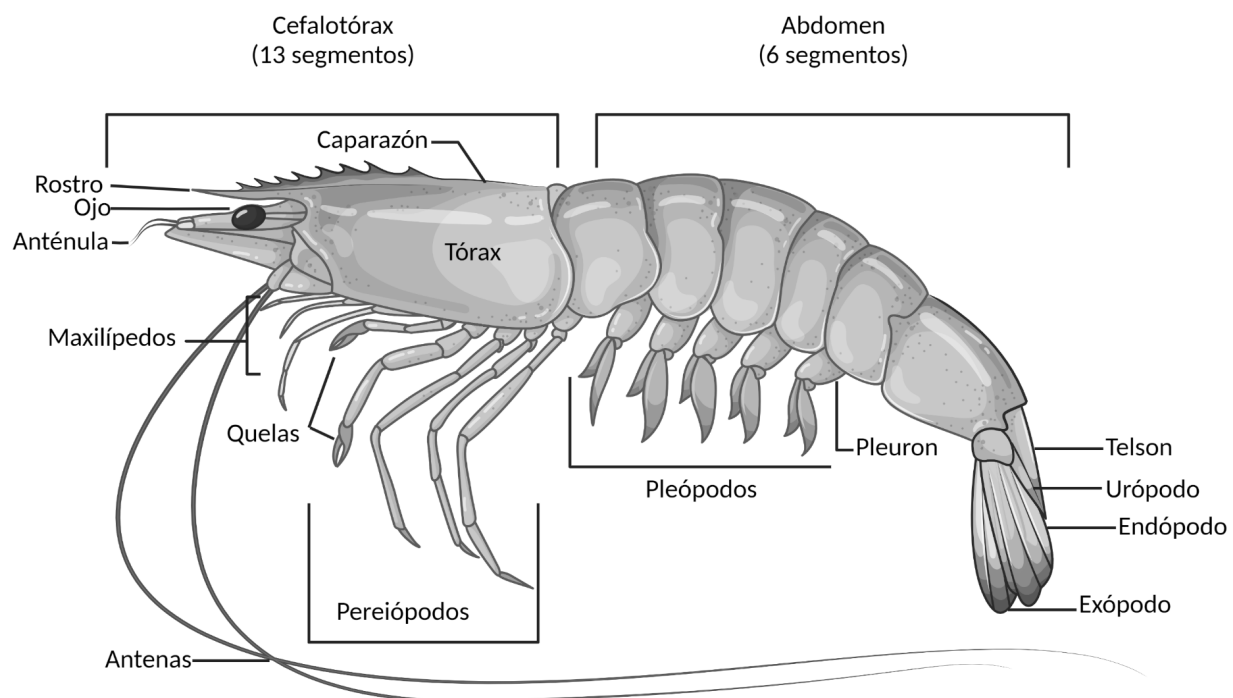


Figura 1. Cuerpo segmentado de los decápodos. El cefalotórax está compuesto por 13 segmentos: caparazón, antenas, anténulas, maxilípedos, pereiópodos, quelas, ojos y rostro. La región del abdomen está compuesta por el pleuron, telson, urópodos, endópodos, exópodos, pleópodos y el petasma en el caso de los machos.

En la gastronomía, los decápodos son considerados un manjar, por lo que las poblaciones son explotadas mediante pesquería y para muchas especies se ha tecnificado su cultivo, por lo que el estudio de cualquier aspecto biológico de los decápodos repercute a nivel económico. De acuerdo con la FAO (2020), durante 2018 se capturaron cerca de 6 millones de toneladas de decápodos, además, en ese mismo año se cultivaron más de 9 millones de toneladas de decápodos para consumo humano (FAO, 2020). Actualmente, la actividad económica del comercio de camarones, cangrejos y langostas tiene un valor aproximado de 24.4 mil millones de dólares, es decir 0.15 % del comercio global (OEC, 2020).

2.2 Sistema digestivo de decápodos

El sistema digestivo es responsable de las funciones relacionadas a la captación de nutrientes. Estas funciones incluyen la ingestión, digestión mecánica, digestión química, absorción celular y transformación de excretas (Ceccaldi, 1989; Fernández-Gimenez, 2013).

Las fuentes de nutrientes de los decápodos pueden provenir de otros animales (carnívoros), plantas (herbívoros), y detritos (detritívoros), o una combinación de estos (omnívoros). El alimento es obtenido a través de diversas estrategias alimenticias, de modo que se han descrito especies depredadoras, carroñeras, suspensívoras (obtienen su alimento de las partículas en suspensión en la columna de agua), y parásitas (Saborowski, 2015).

Los nutrientes presentes en el alimento modulan en gran manera el tipo de enzimas encontradas en el tracto digestivo. Dichas enzimas participan en la hidrólisis de moléculas de carbohidratos, proteínas y lípidos. La morfología y estructura general del sistema digestivo es similar entre las especies del orden Decápoda, el cual está dividido en tres compartimentos principales: anterior, medio y posterior (Fig. 2) (McGaw y Curtis, 2013; Štrus *et al.*, 2019; Vogt, 2021).

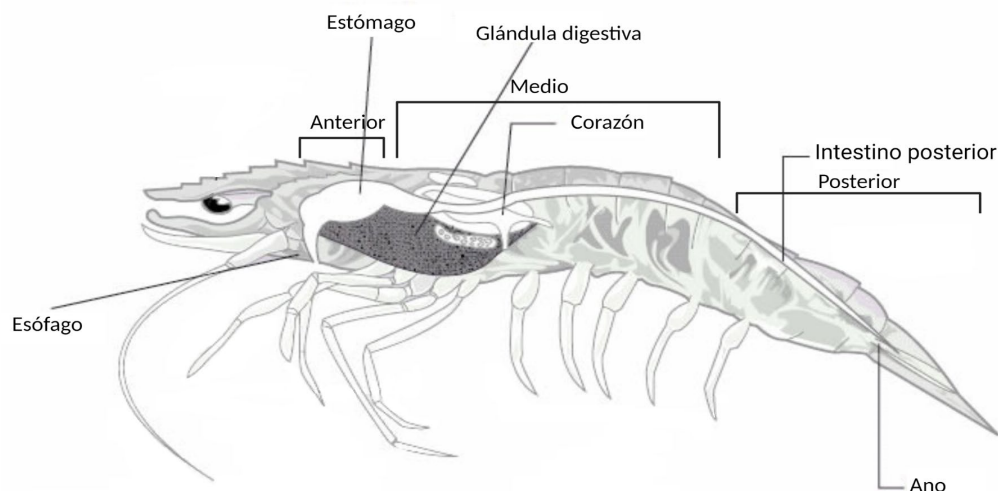


Figura 2. Sistema digestivo de los decápodos a escala macroscópica. La región anterior se compone por el aparato bucal el esófago y el estómago; la región media se conforma por la mayor parte del tubo intestinal, la glándula digestiva o hepatopáncreas y dos o tres ciegos intestinales; la región posterior está formada por la última parte del intestino, el recto y el ano. Foto modificada de FAO 2022.

En los decápodos, los maxilípedos son extremidades auxiliares usadas para atrapar, sostener, desgarrar y dirigir el alimento hacia la cavidad bucal. El alimento pasa entonces por el esófago, en el cual desembocan unas glándulas que secretan moco (llamadas glándulas tegumentarias) para facilitar la ingestión. El esófago conecta a la boca con la primera cámara gástrica, conocida como estómago cardíaco, se trata de una estructura parcialmente distensible recubierta de quitina, en donde se almacena y procesa parcialmente el alimento, en la que se encuentra una estructura rígida conocida como molino gástrico, que se encarga del procesamiento mecánico del alimento, el equivalente a la masticación. Es en el molino gástrico donde se mezcla el alimento con las enzimas digestivas provenientes del hepatopáncreas. El alimento parcialmente procesado es filtrado hacia la segunda cámara digestiva, conocida como estómago pilórico cuya función principal es continuar filtrando partículas de alimento digerido y dirigirlo hacia el hepatopáncreas. Cabe mencionar que solo la ingesta procesada y en forma líquida entra al hepatopáncreas en donde continúa la digestión química y en donde se absorben los nutrientes (Fig. 3) (MacGaw y Curtis, 2013).

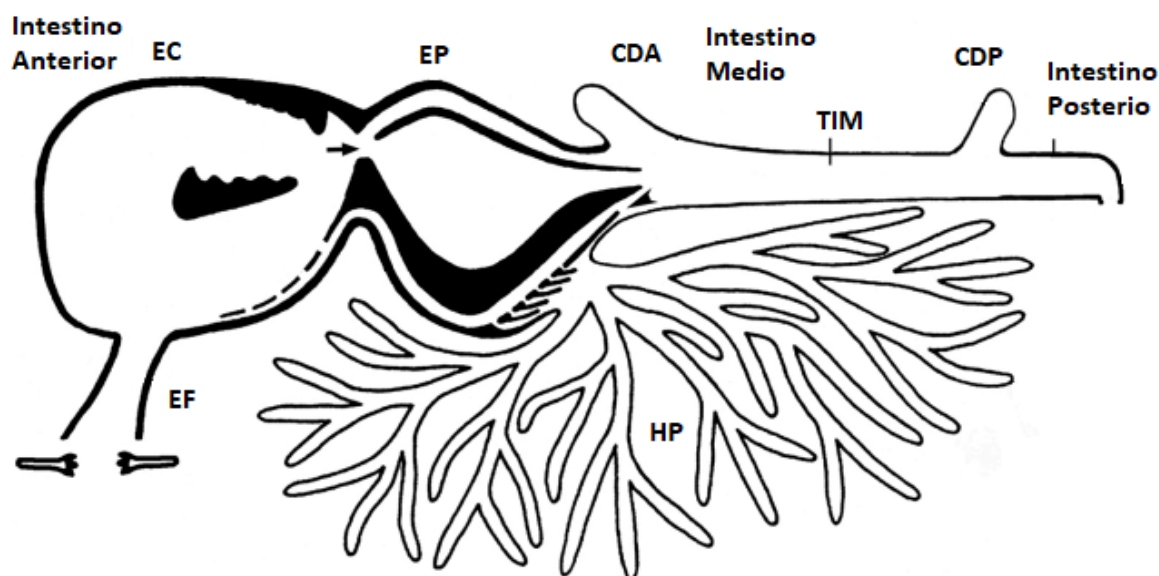


Figura 3. Sistema digestivo de decápodos, dividido en: intestino anterior compuesto por esófago (EF), estómago cardíaco (EC) y estómago pilórico (EP); el intestino medio lo compone el hepatopáncreas (HP), el tubo de intestino medio (TIM), así como el ciego dorsal anterior (CDA) y posterior (CDP); el último apartado lo compone el intestino posterior, el cual se encuentra revestido por una cutícula. Modificado de: Vogt, 2021.

El hepatopáncreas, también conocido como glándula digestiva, es un órgano multifuncional. En términos digestivos, es aquí donde se lleva a cabo la digestión enzimática, la absorción de nutrientes y la secreción de enzimas digestivas, cada una de estas funciones las realizan células especializadas (conocidas como células B, R, F y E) acomodadas en las paredes de cada uno de los cientos de túbulos ciegos que conforman este órgano (MacGaw y Curtis, 2013). Entre las funciones no digestivas del hepatopáncreas se distinguen la reserva de nutrientes como el glucógeno, la eliminación de toxinas, y la síntesis de proteínas y hormonas (Vogt, 2019), entre otras. Finalmente, el intestino, que conecta el estómago pilórico y corre hasta el ano, es un tubo cubierto por la membrana mucoperiotrópica que se expele junto con las heces. El intestino puede estar también involucrado en el transporte de iones (MacGaw y Curtis, 2013; Watling y Thiel, 2013).

2.3 Enzimas digestivas (peptidasas en Decápoda)

Como se mencionó anteriormente, los decápodos habitan una gran diversidad de nichos y, por lo tanto, son capaces de utilizar diversas fuentes de alimento (Saborowski, 2015). Esta presión de selección ha favorecido el desarrollo de sistemas digestivos complejos, por lo que en los

decápodos se han descrito una variedad de enzimas digestivas, como la amilasa, para la hidrólisis de carbohidratos (incluyendo celulosa, hemicelulosa, y quitina), proteasas para la hidrólisis de proteínas, y lipasas para la hidrólisis de lípidos (Ceccaldi, 1989; Rojo-Arreola *et al.*, 2019; Van Wormhoudt *et al.*, 1995; Vogt, 2021).

Las proteínas, son componentes esenciales de la dieta. En la glándula digestiva de varias especies de decápodos, las proteasas son las enzimas más abundantes en relación con otras (Sánchez-Paz *et al.*, 2007). En esta tesis se aborda el estudio de un grupo particular de enzimas involucradas en la degradación de las proteínas del alimento y su incorporación como nutrientes, las metalopeptidasas.

Los crustáceos, para su crecimiento, requieren el consumo diario de proteínas en alimentos naturales o artificiales. Por otro lado, las proteínas son uno de los insumos más costosos en granjas de cultivo de camarón. En *P. vannamei*, los requerimientos proteicos cambian dependiendo del estadio de vida en el que se encuentren, como las etapas de intermuda, y el ambiente en el que se encuentran (Štrus *et al.*, 2019). En la formulación de alimentos para camarones Peneidos de cultivo se recomiendan porcentajes de proteína entre 30 al 57 % para la optimización de su crecimiento (Sánchez-Paz *et al.*, 2006).

Las proteasas (también conocidas como proteinasas o peptidasas) son enzimas que hidrolizan los enlaces peptídicos que unen los aminoácidos en las proteínas (Kanost y Clarke, 2005). Sus funciones en los seres vivos van mucho más allá de la digestión de los alimentos, ya que participan en diversos procesos elementales entre los que se incluyen desde el ciclo celular hasta el sistema inmune (King *et al.*, 1996; Patel, 2017). Se han propuesto varias clasificaciones para las proteasas de acuerdo con sus diversos atributos, pero de manera general se pueden clasificar en *exopeptidasas* y *endopeptidasas*, de acuerdo con la ubicación del sitio de corte dentro de una proteína. Así, si el corte ocurre en los extremos de una proteína, la proteasa se clasifica como *exopeptidasa*, mientras que si el corte ocurre en la secuencia interior de la proteína se le considera como *endopeptidasa* (Comité de Nomenclatura de la Unión Internacional de Bioquímica y Biología Molecular) (Barret, 1992). Las *exopeptidasas* se dividen en dos grandes grupos, las *aminopeptidasas* y las *carboxilopeptidasas*, denotando al extremo en el que reconocen e hidrolizan a una proteína, ya sea en el extremo amino-terminal (N-terminal) o carboxilo-terminal (C-terminal), respectivamente. Por su parte, las *endopeptidasas* son

clasificadas de acuerdo con la naturaleza y las propiedades químicas del sitio catalítico, por lo que se les ha dividido en las siguientes seis familias: aspártico-, cisteín-, glutámico-, metalo-, serin- y treonin-peptidasas (Fig. 4). Cabe mencionar que existen peptidasas con mecanismos catalíticos mixtos y también algunos permanecen desconocidos, por lo que se han agrupado dentro de una categoría especial (The Merops Database: <https://www.ebi.ac.uk/merops/index.shtml>) (Sekton, 2010; Machado de Castro *et al.*, 2018; Rawlings *et al.*, 2018).

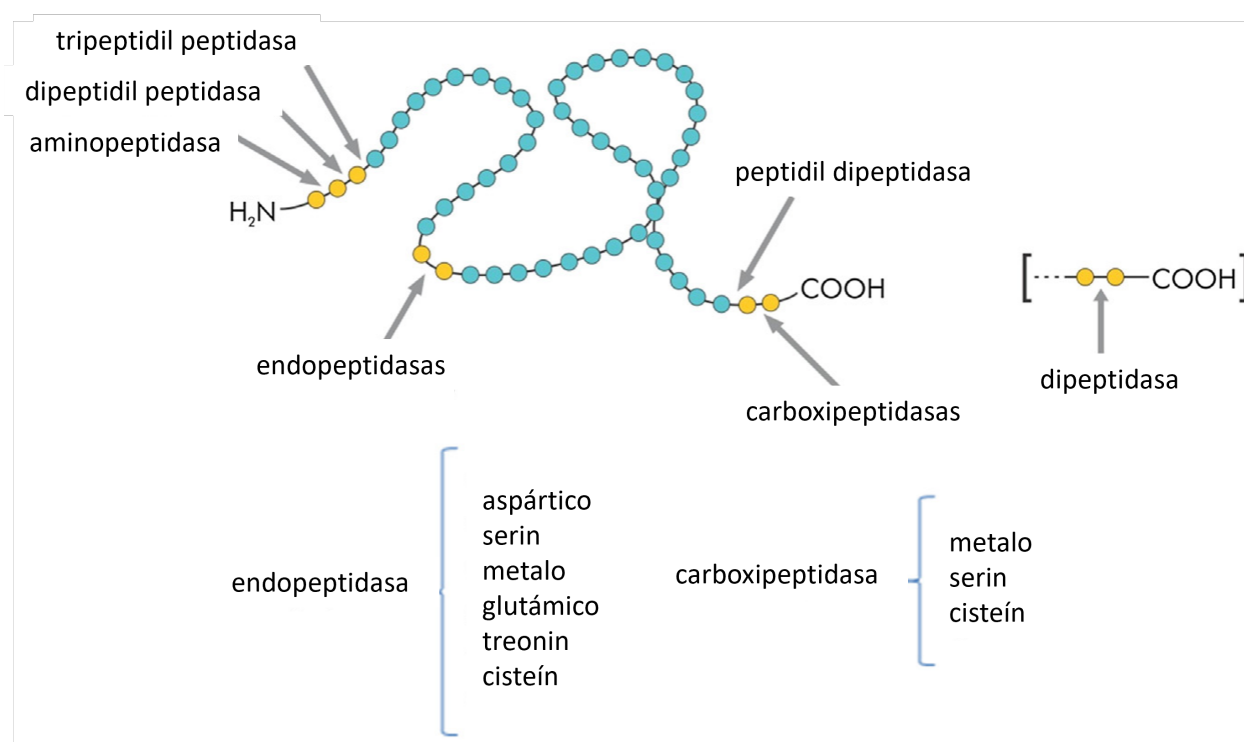


Figura 4. Estructura y clasificación básica de peptidasas. Tomado de: Machado de Castro *et al.*, 2018.

Los perfiles proteolíticos¹, son muy variables entre las especies de decápodos, y dichas variaciones son atribuidas a los diversos hábitats en que estos organismos se distribuyen y, por lo tanto, a sus variados hábitos alimenticios. En el sistema digestivo de crustáceos decápodos se han descrito exopeptidasas (García-Carreño y Haard, 1994) y endopeptidasas responsables de la digestión del alimento. Entre las endopeptidasas, las más comúnmente descritas pertenecen a

¹ Perfiles proteolíticos se refiere los patrones de bandas de actividad enzimática de proteasas al ser revelados por un zimograma. Zimograma es un método electroforético para determinar actividad enzimática, un gel de poliacrilamida es impregnado con un sustrato proteico que es degradado por las enzimas de la muestra, observándose una banda.

la clase serino (Hernández-Cortés *et al.*, 1997), cisteíno (Hu *et al.*, 2007; Stephens *et al.*, 2012), aspártico (Rojo *et al.*, 2010) y metalo (Geier y Zwilling, 1998), siendo las tripsinas y quimotripsinas (clase serino) las más frecuentemente caracterizadas (Hernández-Cortés *et al.*, 1997; Sainz-Hernández y Córdova-Murrieta 2009).

Además de la variación entre especies, se han descrito otros factores que contribuyen a cambios en la actividad proteolítica. Por ejemplo, se sabe que en juveniles de camarones Peneidos, las tripsinas y quimotripsinas, son responsables de aproximadamente un 60 % en la hidrólisis de proteínas del alimento, pero este porcentaje varía de acuerdo con el estadio de desarrollo en que se encuentran los organismos, pues sus hábitos alimenticios están estrechamente relacionados con la ontogenia. En 2017, Hernández-Cortés *et al.*, describieron que los cambios en la dieta ocurren a la par que los cambios anatómicos que sufren los camarones Peneidos (por ejemplo, *Penaeus monodon*, *P. japonicus*, *P. paulensis*, *P. schmitti*) durante el desarrollo larvario y que están relacionados con la maduración del sistema digestivo. Para *P. vannamei*, durante el estadio *nauplio*, la actividad proteolítica es prácticamente nula, esto es atribuido a la presencia del saco vitelino que se utiliza como reserva energética durante este estadio; cuando los organismos pasan a la siguiente etapa, el estadio *protozoa*, se alimentan principalmente de fitoplancton, y es en esta etapa donde comienza a observarse actividad proteolítica, lo que se ha sido asociado a la diferenciación celular del hepatopáncreas en formación. En la etapa *misís*, se observa una disminución en la actividad proteolítica, que permanece baja debido a la reducción de los ciegos hasta que los organismos alcanzan la primera etapa de *postlarva*, cuando el sistema digestivo está completamente diferenciado, demostrando diferentes capacidades metabólicas de los estados larvales.

2.4 Metalopeptidasas

Las metalopeptidasas son un grupo diverso de enzimas, presentes en organismos procariotas y eucariotas, y se les ha implicado en procesos tan diversos como el cáncer, enfermedades cardiovasculares, digestión, e incluso pueden ser toxinas (como las del tétanos, botulismo y venenos de serpientes) (Hooper, 1996; Gutiérrez y Rucavado, 2000; Nagase y Murphy, 2006; Rawlings y Barrett, 2013) y su distribución taxonómica se expandió ampliamente durante la evolución.

Las metalopeptidasas son hidrolasas, cuyo mecanismo de acción implica un ataque nucleofílico sobre el enlace peptídico. Este ataque es mediado por una molécula de agua que se activa con un catión metálico que normalmente es divalente (usualmente zinc, pero puede ser de cobalto, cobre, níquel o manganeso). Dicho ion se mantiene en el sitio activo de la enzima unido usualmente por 3 ligandos de aminoácidos (Barret *et al.*, 2013). Además, estos iones pueden formar enlaces covalentes con residuos de histidina, glutamato, aspartato, lisina o arginina; facilitando la coordinación con los ligandos (Rawlings y Barrett, 2013; Ramos y Selistre-de-Araujo, 2001).

Generalmente, las metalopeptidasas se clasifican en dos grandes grupos con base a los iones metálicos que necesitan para catalizar: las bimetálicas que requieren de dos iones metálicos para que actúen conjuntamente (co-catalíticamente) o las monometálicas, que sólo requieren un ion metálico. La mayoría de las metalopeptidasas que actúan con dos iones metálicos son endopeptidasas, y las que actúan con uno solo pueden ser endo- o exopeptidasas (Barret *et al.*, 2013; Lian *et al.*, 2019).

El zinc en las metalopeptidasas monometálicas está ligado tetraédricamente por tres átomos de las cadenas laterales de la proteína (Fig. 5), usualmente tres histidinas y en la parte debajo de la hendidura del sitio activo se une por una molécula de disolvente catalítico que además está unido a un ácido/base, que por lo general es un glutamato. (Arolas *et al.*, 2018).

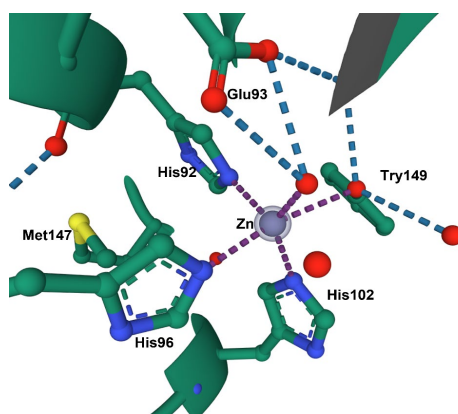


Figura 5. Representación esquemática de la estructura de astacina (1AST, PDB). Detalle del sitio activo de la incisión de astacina. En azul se representan las histidinas que están unidas al zinc (esfera gris), en amarillo la metionina M147 del Met-turn (giro de metionina) además de la tirosina 149 señalado en rojo. El ácido glutámico (E93) que corresponde al sitio activo se encuentra indicado en color rojo. Creado en Mol* en PDB (Sehnal *et al.*, 2021).

Los primeros intentos realizados para clasificar las zinc-metalopeptidasas ocurrieron en los años noventa. Esto se basó en la existencia de distintos motivos y afinidad de unión al motivo con el zinc y se realizaron comparaciones de secuencias y topologías (Jongeneel *et al.*, 1989; Spyroulias *et al.*, 2004). Como resultado, se identificó la secuencia primaria del motivo que estas enzimas poseen en común, HEXXH (X se refiere a cualquier residuo aminoacídico) (Jongeneel *et al.*, 1989).

Actualmente, una clasificación ampliamente aceptada de las enzimas proteolíticas y las metalopeptidasas es la que propone la base de datos MEROPS (<https://www.ebi.ac.uk/merops/>) (Rawlings *et al.*, 2018) que también se adopta en el Manual de Enzimas Proteolíticas (Barrett *et al.*, 2013) en la que se agrupan por conjuntos de homólogos de enzimas proteolíticas e inhibidores, en función de la similitud de la distancia evolutiva y las secuencias en orden jerárquico por clanes, familias y especies.

Dependiendo del número y naturaleza de los iones metálicos, las metalopeptidasas han sido clasificadas en 14 clanes (Tabla 2). Hasta hoy, el clan con mayor número de familias es el clan MA con 49 (Rawlings *et al.*, 2018; MEROS, 2022).

Además de esta clasificación, Cerdá-Costa y Gomis-Rüth (2013), proponen la clasificación de las metalopeptidasas con base a la arquitectura del sitio activo y la similitud general del pliegue en el orden jerárquico de subclase, tribu, clan, familia. Esta clasificación primero divide a las metalopeptidasas en mono y binucleares. Las metalopeptidasas mononucleares (que requieren un sólo ion metálico), incluyen a las tribus $\alpha\beta\alpha$ -exopeptidasas conteniendo el clan de las funelinas, las metalopeptidasas LAS, las relacionadas a factores de maduración de la hidrogenasa (HybD) que generalmente tienen iones de níquel, las inverzincinas y las zincinas. Las tribus se distinguen por el residuo del enlace al metal, que por lo general pueden ser histidinas, aspartato, y glutamato, y se incluyen en el motivo de enlace al zinc y son característicos de tribus, clanes y familias. Las zincinas se clasifican y se nombran así por su motivo de unión al zinc y se dividen en dos clanes: las gluzincinas y las metzincinas. En el primer clan se encuentran las toxinas de botulinum y de tétanos, la termolisina, la enzima convertidora de angiotensina y la oligopeptidasa de thimet (Semenova y Rudenskaya, 2009).

En la Fig. 6 se representan tres de los clanes mencionados: funelinas, metzincinas y gluzincinas. El clan de las funelinas (M14 - MEROPS), el clan gluzincinas, que abarcan las familias de termolisinas (M4), fungalicinas, entre otros, y finalmente el clan de las metzincinas, en el cual nos enfocaremos en este trabajo. Este clan incluye a las familias de las metaloproteinasas de matriz (MMP, M10), las desintegrinas y metalopeptidasas (ADAMs), la proteína desintegrina y metaloproteinasas con motivos de trombospondina (ADAMTSs) (Rivera 2010), la familia fragilicinas y la familia astacina (M12) (Arolas *et al*, 2018; Rawling *et al.*, 2018).

El nombre del clan de las metzincinas deriva del giro de metionina conservado (Met-turn) y del sitio de unión al zinc (Bode *et al.*, 1993). Bode y colaboradores en 1993, después de elucidar las estructuras cristalinas de la enzima astacina y adamalisina II, que cada una pertenece a una familia distinta, observaron que ambas peptidasas tienen una equivalencia topológica en común además que el ambiente entorno al zinc es casi idéntico, así como el giro de metionina. Además de estas dos peptidasas, en el mismo estudio se compararon las secuencias y estructura con cuatro familias de metalopeptidasas, como la astacinas, adamalisinas, matrixinas y serralisinas, que también comparten estas mismas características, por lo cual se sugirió agruparlas dentro de las metzincinas.

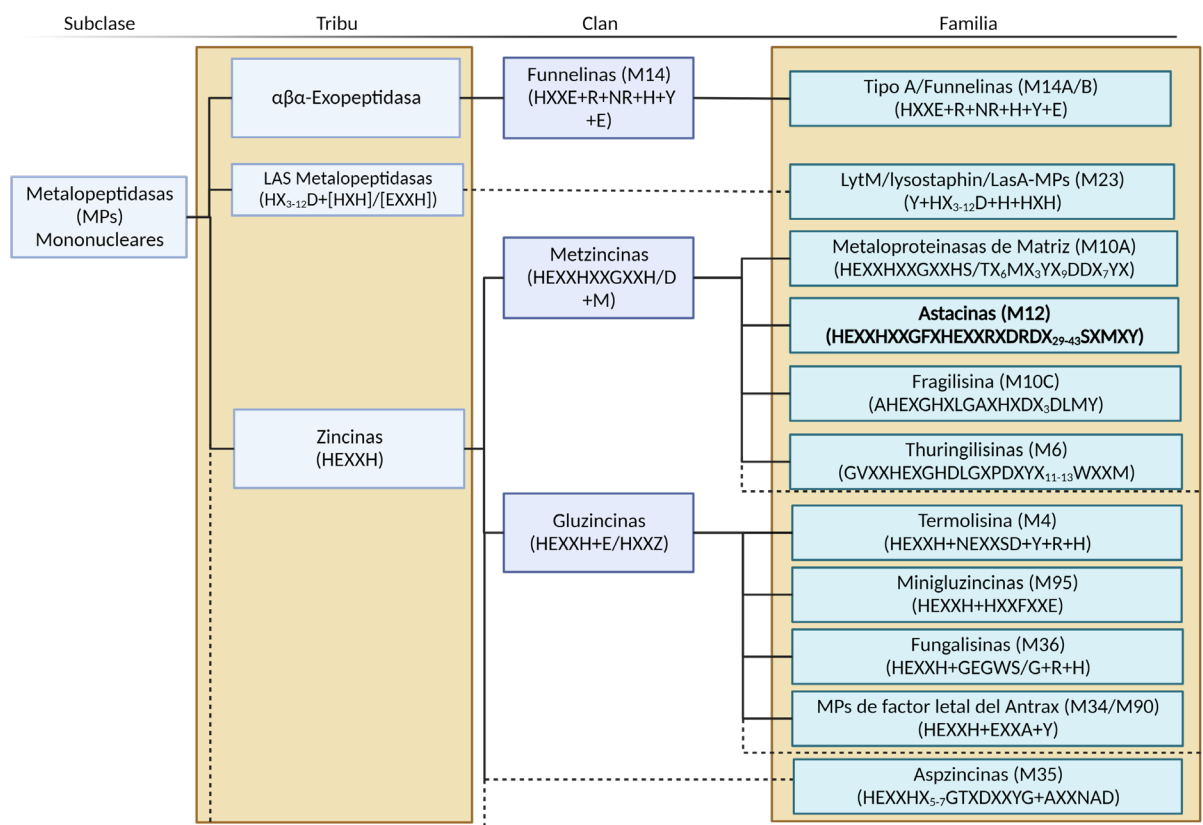


Figura 6. Clasificación de la subclase de metalopeptidasas mononucleares (que requieren un sólo ion metálico). Solo se representan las tribus $\alpha\beta$ -exopeptidasas, metalopeptidasas LAS y las zincinas. A su vez, las zincinas se dividen en dos clanes: metzincinas y gluzincinas. El clan metzincina contiene la familia de las astacinas.

Tabla 2. Clasificación de Metalopeptidasas MEROPS

Clan	Familia	Tipo de peptidasa	Clan	Familia	Tipo de peptidasa
MA	M1	aminopeptidase N (<i>Homo sapiens</i>)	M87		chloride channel accessory protein 1 (<i>Homo sapiens</i>)
	M2	angiotensin-converting enzyme peptidase unit 1 (<i>Homo sapiens</i>)	M88		IMPa peptidase (<i>Pseudomonas aeruginosa</i>)
	M3	thimet oligopeptidase (<i>Rattus norvegicus</i>)	M90		MtfA peptidase (<i>Escherichia coli</i>)
	M4	thermolysin (<i>Bacillus thermoproteolyticus</i>)	M91		NleD peptidase (<i>Escherichia coli</i>)
	M5	mycolysin (<i>Streptomyces cacaoi</i>)	M93		BACCAC_01431 g.p. and similar (<i>Bacteroides caccae</i>)
	M6	immune inhibitor A peptidase (<i>Bacillus thuringiensis</i>)	M95		abylysin (<i>Pyrococcus abyssi</i>)
	M7	snapalysin (<i>Streptomyces lividans</i>)	M97		EcxAB peptidase (<i>Escherichia coli</i>)
	M8	leishmanolysin (<i>Leishmania major</i>)	M98		YghJ g.p. (<i>Escherichia coli</i>)
	M9	collagenase V (<i>Vibrio alginolyticus</i>)	M100		spartan peptidase (<i>Homo sapiens</i>)
	M10	matrix metallopeptidase-1 (<i>Homo sapiens</i>)	M102		DA1 peptidase (<i>Arabidopsis thaliana</i>)
	M11	gametolysin (<i>Chlamydomonas reinhardtii</i>)	M105		MAP1 peptidase (<i>Myxococcus xanthus</i>)
	M12	astacin (<i>Astacus astacus</i>)	M106		FP0506 g.p. (<i>Flavobacterium psychrophilum</i>)
	M13	nepriylisin (<i>Homo sapiens</i>)	MC	M14	carboxypeptidase A1 (<i>Homo sapiens</i>)
	M26	IgA1-specific metallopeptidase (<i>Streptococcus sanguinis</i>)		M86	PghP gamma-polyglutamate hydrolase (Bacillus phage phiNIT1)
	M27	tentoxilysin (<i>Clostridium tetani</i>)		M99	Csd4 peptidase (<i>Helicobacter pylori</i>)
	M30	hycolysin (<i>Staphylococcus hyicus</i>)	MD	M15	zinc D-Ala-D-Ala carboxypeptidase (<i>Streptomyces albus</i>)
	M32	carboxypeptidase Taq (<i>Thermus aquaticus</i>)		M74	murein endopeptidase (<i>Escherichia coli</i>)
	M34	anthrax lethal factor (<i>Bacillus anthracis</i>)	ME	M16	pitrilysin (<i>Escherichia coli</i>)
	M35	deuterolysin (<i>Aspergillus flavus</i>)		M44	pox virus metallopeptidase (<i>Vaccinia virus</i>)
	M36	fungalyisin (<i>Aspergillus fumigatus</i>)	MF	M17	leucyl aminopeptidase 3 (<i>Bos taurus</i>)
	M41	FtsH peptidase (<i>Escherichia coli</i>)	MG	M24	methionyl aminopeptidase 1 (<i>Escherichia coli</i>)
	M43	cytophagalyisin (<i>Cytophaga sp.</i>)		M18	aminopeptidase I (<i>Saccharomyces cerevisiae</i>)
	M48	Ste24 peptidase (<i>Saccharomyces cerevisiae</i>)		M20	glutamate carboxypeptidase (<i>Pseudomonas sp.</i>)
	M49	dipeptidyl-peptidase III (<i>Rattus norvegicus</i>)		M28	aminopeptidase S (<i>Streptomyces griseus</i>)
	M54	archaelysin (<i>Methanocaldococcus jannaschii</i>)		M42	glutamyl aminopeptidase (<i>Lactococcus lactis</i>)
	M56	BlaR1 peptidase (<i>Staphylococcus aureus</i>)	MJ	M19	membrane dipeptidase (<i>Homo sapiens</i>)
	M57	prtB g.p. (<i>Myxococcus xanthus</i>)		M38	isoaspartyl dipeptidase (<i>Escherichia coli</i>)
	M60	enhancin (Lymantria dispar nucleopolyhedrovirus)	MM	M50	site 2 peptidase (<i>Homo sapiens</i>)
	M61	glycyl aminopeptidase (<i>Sphingomonas capsulata</i>)	MN	M55	D-aminopeptidase DppA (<i>Bacillus subtilis</i>)
	M64	IgA peptidase (<i>Clostridium ramosum</i>)	MO	M23	beta-lytic metallopeptidase (<i>Achromobacter lyticus</i>)
	M66	StcE peptidase (<i>Escherichia coli</i>)		M67	PSMD14 peptidase (<i>Saccharomyces cerevisiae</i>)
	M72	peptidyl-Asp metallopeptidase (<i>Pseudomonas aeruginosa</i>)	MP	M67	PSMD14 peptidase (<i>Saccharomyces cerevisiae</i>)
	M76	Atp23 peptidase (<i>Homo sapiens</i>)	MQ	M29	aminopeptidase T (<i>Thermus aquaticus</i>)
	M78	ImmA peptidase (<i>Bacillus subtilis</i>)	MS	M75	imelysin (<i>Pseudomonas aeruginosa</i>)
	M80	Wss1 peptidase (<i>Saccharomyces cerevisiae</i>)	MT	M81	microcystinase MlrC (<i>Sphingomonas sp.</i> ACM-3962)
	M84	MpriBi peptidase (<i>Bacillus intermedius</i>)			
	M85	NleC peptidase (<i>Escherichia coli</i>)	unassigned	M73	camelysin (<i>Bacillus cereus</i>)
				M82	PrsW peptidase (<i>Bacillus subtilis</i>)

2.5 Familia Astacina

El objeto de estudio de esta tesis son las peptidasas de *P. vannamei* pertenecientes a la familia Astacina (M12A, MEROPS, Tabla 1). Esta familia es la segunda más grande en cuanto al número de proteasas que la componen y son miembros del subclan MA(M) que tomó este nombre por una proteína extraída del jugo gástrico del langostino *Astacus astacus*, la astacina, al ser el miembro más simple y el primero en ser secuenciado. Este grupo de proteasas fue reconocido así cuando se analizaron las secuencias y estructuras de la astacina junto con otras 4 metaloendopeptidasas, y el alineamiento de las secuencias mostró que pertenecían a un nuevo grupo de metaloproteinasas estructuralmente diferentes (Dumermuth *et al.*, 1991). La astacina ha servido como prototipo para el estudio de los demás miembros de esta familia ya que fue la primera en ser secuenciada y en descifrar su estructura cristalográfica por rayos X (Bode *et al.*, 1992; Stöcker *et al.*, 1993).

Los miembros originales de esta familia fueron identificados por similitudes en las secuencias: la enzima digestiva astacina del langostino (*Astacus astacus*), la proteína morfogénica ósea 1 (BMP-1, por sus siglas en inglés) de humano, las meprinas de intestino humano y de riñón de ratón (Meprina A) y la proteína UVS.2 involucrada en el desarrollo de embriones de *Xenopus laevis* (Dumermuth *et al.*, 1991; Bond, 2019).

Los miembros de la familia Astacina son endopeptidasas dependientes de un ion de zinc que se encuentra en el sitio activo, el cual se une a tres residuos de histidina (H) y adquiere una configuración pentacoordinada (a excepción de BMP-1). El sitio activo de esta familia incluye un motivo cuya secuencia de 18 aminoácidos (HEXXHXXGFXHEXXRXDR) inicia con el motivo de unión a zinc HEXXH, que, como ya se mencionó, es específico para las metalopeptidasas. Adicionalmente, en la secuencia de la astacina se identifican tres estructuras tipo hélice α , una hoja β de cinco hebras y dos o tres enlaces disulfuro (Bond y Beynon, 1995; Gomis-Ruth *et al.*, 2012). Estas enzimas son sintetizadas como precursores inactivos y en algunas la inactividad se debe a la interacción entre un sitio conservado de cisteína en el propéptido y el sitio catalítico del zinc para prevenir la unión a la molécula de agua (Rawlings y Barrett, 2013).

Además del dominio principal, la mayoría de los integrantes de la familia de las astacinas M12, contienen una o más copias del dominio no catalítico, denominadas módulos, mismos que son

independientes al módulo de astacina, y funcionales. Las proteasas pueden combinar varios módulos en las regiones N-terminal y C-terminal, y el módulo que ocurre con mayor frecuencia en la familia M12 es el CUB y ShK (se denomina así por los subcomponentes del complemento C1r/C1s, la proteína de erizo de mar embrionario UEGF y BMP-1) (Bork y Beckmann, 1993). Además del dominio CUB, existen otros módulos como el de factor de crecimiento epidérmico (EGF), el dominio de la toxina ShK (detectado en celenterados), los dominios TRAF y MAM (que se encuentran en las meprinas). Los módulos pueden presentarse de manera individual en algunas astacinas, pueden repetirse en la secuencia o bien, combinarse varios módulos. Son importantes para las interacciones proteína-proteína o enzima-sustrato, además de la variación de tamaños y la unión al calcio (Fig. 7) (Bond, 2019; Gomis-Rüth y Stöcker, 2012; Gerdol *et al.*, 2019; Martín-Galiano y Sotillo, 2021).

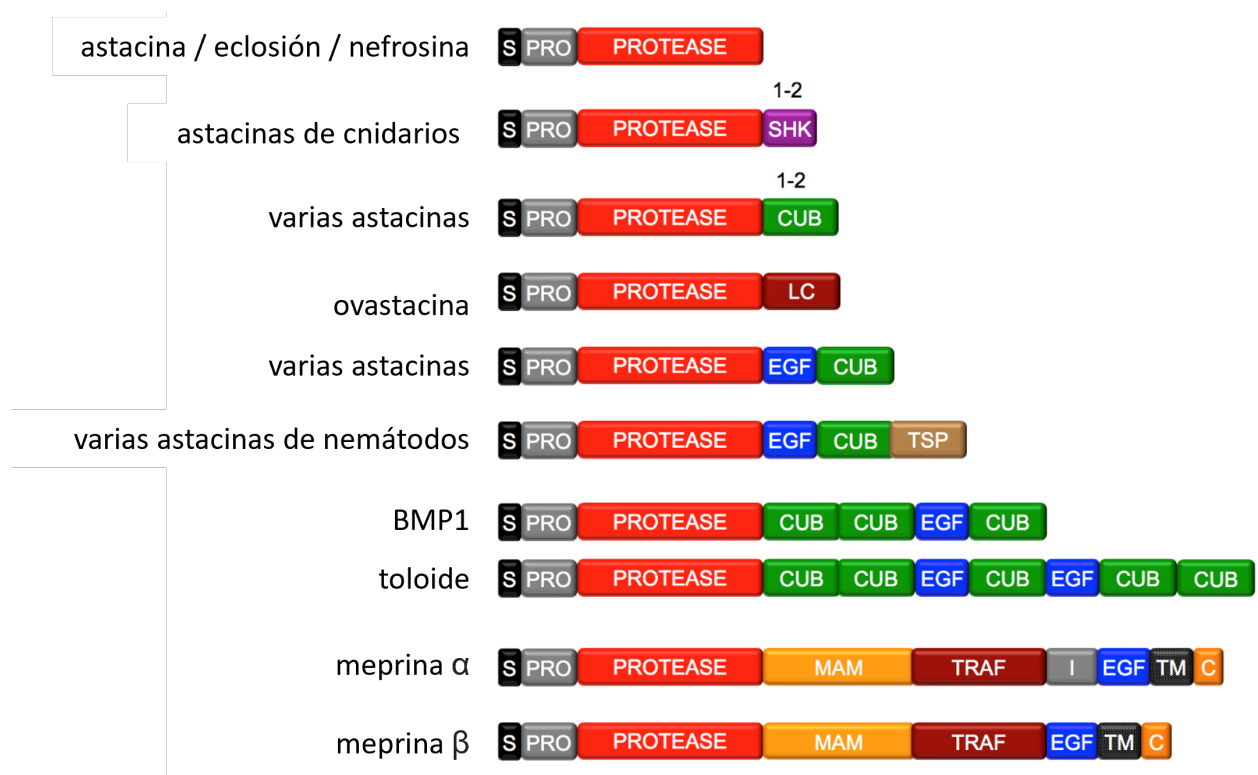


Figura 7. Composición modular de las astacinas. En la imagen están representados los distintos dominios que se han observado en los miembros de la familia de las astacinas. El péptido (negro) y propéptido (gris claro) señalan N-terminal. Los dominios C-terminal se encuentran unidos al dominio de la proteasa catalítica (rojo) y estos son el dominio ShK (morado), CUB (verde), LC (rojo oscuro), EGF (azul), TSP (café), MAM (amarillo), TM (gris oscuro), C (naranja). Imagen tomada de Gomis-Rüth *et al.*, 2012.

2.6 Proteasa Astacina

Históricamente, el primer reporte de una metalopeptidasa en el tracto digestivo de langostino europeo *Astacus astacus*, fue presentado por Zwilling y colaboradores en 1967. En dicho reporte se describió una proteína con un peso molecular de 11 kDa por lo que se nombró *pequeña molécula de proteasa de langostino*. Una vez conocida la secuencia de esta proteína su nombre cambió a *proteasa Astacus*, y finalmente, cuando se demostró que era una enzima dependiente de zinc, se le denominó *Astacina* (Stöcker *et al.*, 1993).

La astacina, una zinc metaloendopeptidasa, se encuentra en alta concentración en el jugo gástrico del langostino *Astacus astacus* y su función es la digestión de las proteínas del alimento. Esta enzima degrada gelatina, la triple hélice del colágeno tipo I y otras proteínas nativas, en condiciones de pH neutro (Stöcker y Yiallourous, 2011). En un ensayo realizado por Vogt y colaboradores en 1989, se estimuló la secreción enzimática en langostinos mediante el vaciado manual del jugo gástrico o como resultado de la ingestión de alimento, y se propuso que su síntesis ocurre en los túbulos del hepatopáncreas y se almacenan extracelularmente de forma activa en el estómago cardiaco. Además, se sugirió que dicha astacina poseen una función digestiva (Zwilling y Neurath, 1981; Vogt *et al.*, 1989).

La astacina es estable a la digestión proteolítica y la autólisis, es soluble hasta concentraciones de 20 mg/mL en buffers diluidos. Mientras que por debajo de un pH 4, la enzima se precipita. La astacina tiene una absorbancia de $42,800 \text{ M}^{-1} \text{ cm}^{-1}$ con coeficiente de absorbancia $\epsilon_{280 \text{ nm}}$. El pH óptimo para la catálisis está cercano a valores neutros dependiendo del sustrato que sea empleado. Por ejemplo, con el sustrato STANA (succinil-Alanil-Alanil-Alanil-4-Nitroanilida), el valor de pH óptimo es aproximadamente 8.0 mientras que con Alanil-Alanil-Alanil-4-Nitroanilida es de aproximadamente 6.0. Para la proteína en su conformación sin desnaturalizar (nativa), el punto isoeléctrico es de pH 3.5. La masa molecular estimada en gel de poliacrilamida desnaturalizante de la enzima es de 23 kDa, mientras que el tamaño putativo con base a la secuencia de aminoácidos es de 22.6 kDa (Stöcker y Zwilling, 1995).

Un potente inhibidor sintético de la astacina son los pseudopéptidos análogos al estado de transición fosfínico como FMOC-Pro-Lys-Phe- ψ (PO₂-CH₂)-Ala-Pro-Leu-Val-OH (FMOC=9-fluorenilmetiloxycarbonilo) ($K_i=42\text{nM}$); y su comportamiento de unión es extremadamente lento

(Yiallouro *et al.*, 1998). El inhibidor natural más potente reportado hasta la fecha es α 2-macroglobulina. Otros inhibidores naturales son la proteína de plasma fetuina A y la cistatina C también han sido identificados como inhibidores de astacina. Además, se ha reportado inhibidores en las semillas de *Ecballium elaterium* y en los tubérculos de papa. La astacina y otros miembros de la familia de las astacinas no son inhibidos por los inhibidores de tejidos de metalopeptidasas (TIMPs) (Hedrich *et al.*, 2010; Stöcker y Yiallourous, 2011).

Además de la secuencia proteica del langostino *A. astacus*, existen reportes de secuencias similares a la astacina en otros crustáceos decápodos, como el cangrejo azul de las nieves (*Chionoecetes opilio*), el cangrejo rojo gigante o centollo de Alaska (*Paralithodes camtschaticus*), el camarón blanco del Pacífico (*P. vannamei*); así como en otros organismos como el nemátodo *Caenorhabditis elegans*, el cangrejo herradura del Atlántico (*Limulus polyphemus*), la araña *Argiope aurea*, y el mosquito *Anopheles gambiae* (Stöcker y Yiallourous, 2013).

En 1992, Bode y colaboradores reportaron la estructura de la astacina, que elucidaron mediante cristalografía de rayos X. Este trabajo demostró que la astacina es una molécula esférica que se subdivide por la incisión profunda del sitio activo donde el zinc se encuentra al fondo ligado por tres histidinas, una molécula de agua y una tirosina.

El mecanismo catalítico de la astacina (Figura 6) involucra el ion metálico del sitio activo, que se encuentra pentacoordinado por tres residuos de histidina (H93, H96 y H102), un residuo de tirosina (Y149) que actúa como un cuarto ligando en el precursor y, como base general durante la catálisis, un ácido glutámico (E93) que participa como un electrófilo para la estabilización del complejo en el estado de transición. Además, este mecanismo implica la participación de una molécula de agua (H₂O) en una geometría bipiramidal trigonal (toma esa estructura cuando está en un estado de apoenzima²) (Rawlings y Barrett, 2013).

En su forma madura, la astacina posee una secuencia de 200 aminoácidos (Bode *et al.*, 1992), que forman el dominio catalítico, el cual puede extenderse en el propéptido N-terminal, que le brinda la propiedad de latencia (es decir, la enzima se sintetiza con un prepropéptido de

² Una apoenzima es la parte proteica de la enzima desprovista de los cofactores o grupos prostéticos que puedan ser necesarios para que la enzima sea funcionalmente activa. La apoenzima es catalíticamente inactiva. (Devlin, 2019)

activación), siendo secretada como un zimógeno. El dominio catalítico está entrecruzado con dos enlaces disulfuro en C42 - C198 el cual une el segmento C-terminal con el N-terminal en el bucle de la hélice αA con la hoja $\beta 2$, mientras que el enlace C64 - C84 une L $\beta 3\beta 4$ con el comienzo de la hoja $\beta 4$ a L $\beta 5\alpha B$ y así ayudan al plegamiento para formar la hendidura del sitio activo y unir el sustrato (Gomis-Rüth y Stöcker, 2012).

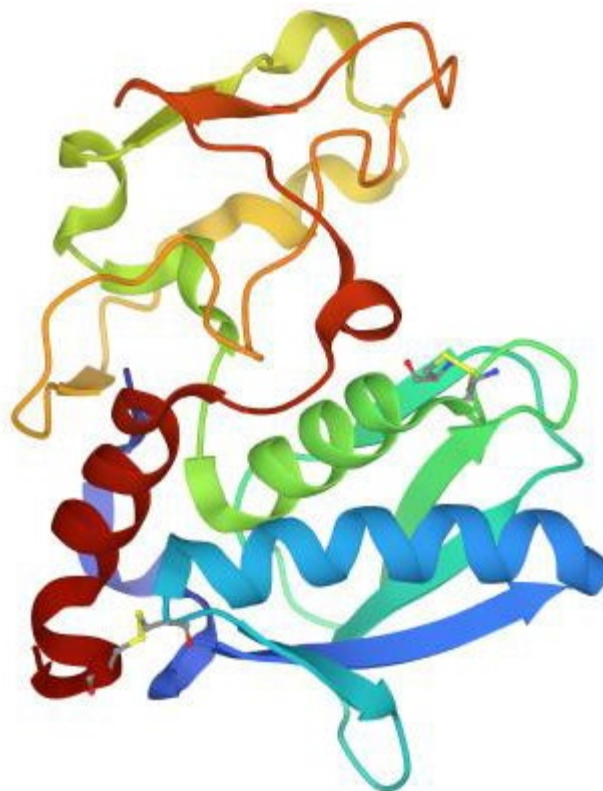


Figura 8. Vista 3D de la cadena A de astacina (PDB 1AST).

El gen que codifica la astacina del langostino contiene 2,616 pares de bases (pb). Este gen está conformado de cinco exones y cuatro intrones, y es esta conformación que corresponde a la estructura secundaria a nivel de proteína. El primer intrón interrumpe la secuencia codificante del péptido señal, los intrones 2 y 3 tienen posiciones iguales que en otras astacinas y el intrón 4 se caracteriza por ser diferente (Stöcker y Yiallourous, 2013).

3. JUSTIFICACIÓN

En el transcurso de las últimas dos décadas ha quedado en evidencia la falta de información sobre las características bioquímicas, fisiológicas y evolutivas de las astacinas de decápodos, y en particular de la del camarón blanco *Penaeus vannamei*. Esto ha obstaculizado la comprensión integral de la función de esta enzima. Dado que el conocimiento sobre la estructura de una proteína es crucial para comprender su función, el uso de algunas herramientas bioinformáticas puede servir como punto de partida para contribuir al conocimiento de las características moleculares y evolutivas relacionadas con la funcionalidad de esta enzima en el camarón *Penaeus vannamei*.

4. HIPÓTESIS

Si la estructura primaria de las zinc metalopeptidasas monodominio (astacinas) de *Penaeus vannamei* encontradas en las bases de datos tienen una alta variabilidad, entonces a partir del análisis de las secuencias variantes se podrá inferir la adaptabilidad fisiológica y digestiva de esta especie.

5. OBJETIVO

5.1 Objetivo general

Analizar, a través de herramientas bioinformáticas, los patrones de variación, heterogeneidad estructural y efectos funcionales de las variantes en la secuencia de aminoácidos de metalopeptidasas digestivas del camarón blanco *Penaeus vannamei*.

5.2 Objetivos particulares

1. Caracterizar *in silico* las secuencias de proteínas de las zinc metalopeptidasa de *Penaeus vannamei* depositadas con bases de datos.
2. Estimar la variabilidad de las secuencias de aminoácidos de las metalopeptidasas digestivas de *P. vannamei*.
3. Comparar los elementos estructurales de las variantes de las metalopeptidasas digestivas de *P. vannamei*.

6 MATERIAL Y MÉTODOS

6.1 Identificación de zinc metalopeptidasas con dominio astacina con bases de datos de genes y proteínas

Las secuencias de metalopeptidasas de *Penaeus vannamei* analizadas en este trabajo fueron recuperadas de las bases de datos del NCBI (National Center for Biotechnology Information, <https://www.ncbi.nlm.nih.gov/>) y UNIPROT (<https://www.uniprot.org/blast>). La búsqueda se realizó en orden de lo general a lo particular: zinc metalopeptidasas en decápodos y después se realizaron búsquedas específicas en *P. vannamei* que se describen a continuación.

6.1.1 Identificación de zinc metalopeptidasas de la familia Astacina (M12) en Decápodos

La presencia de astacinas en especies de decápodos fue identificada mediante una búsqueda en la base de datos de proteínas de NCBI. Se utilizó la herramienta BLASTp (Altschul *et al.*, 1990) usando como término de referencia a la astacina de *Astacus astacus* (UNIPROT: P07584), la identificación por alineamiento se realizó en la base de datos Non Redundant Protein Sequences, restringiendo al taxa Decápoda (ID: 6683). Se descargaron solo aquellas secuencias que tienen un porcentaje de identidad >40% y una cobertura >60%.

6.1.2 Identificación de zinc metalopeptidasas de la familia Astacina (M12) y multidominio en *Penaeus vannamei*

Se realizó una búsqueda en la base de datos curada y con muy baja redundancia UniProt (The UniProt Consortium, 2021) con el fin de observar las astacinas multidominio que se encuentran en *P. vannamei*. Se utilizó la herramienta BLAST de dicha base de datos, usando la secuencia de la astacina de *A. astacus* (UniProt: P07584) como base para el alineamiento. La búsqueda se restringió por taxonomía para *P. vannamei* (ID: 6689), bajo los parámetros predeterminados. Para los análisis posteriores se seleccionaron solamente aquellas secuencias que contenían el motivo HEXXHXXGXXHEXXRXDR y presentaban uno o más módulos diferentes al dominio astacina.

6.1.3 Identificación de zinc metalopeptidasas monodominio de astacina en *Penaeus vannamei*

Para una búsqueda más detallada de las variantes de transcritos de metalopeptidasas monodominio y putativamente digestivas de *P. vannamei*, se utilizó la base de datos del NCBI, con ayuda de la herramienta BLAST (Basic Local Alignment Search Tool), utilizando el programa

tBlastn. El término de búsqueda fue la zinc metalopeptidasa identificada en *P. vannamei* como Mpc1 (ABD65301). Los alineamientos se restringieron a las bases de datos Nucleotide collection (nr/nt), Reference RNA sequences (refseq_rna) (Pruitt *et al.*, 2007) y Transcriptome Shotgun Assembly (TSA), así como la restricción taxonómica de *P. vannamei* (ID:6689).

Las bases de datos del NCBI contienen secuencias altamente redundantes por lo que solamente se descargaron aquellas secuencias que tienen un porcentaje de identidad >50% y una cobertura >60% respecto a la Mpc1 (ABD65301), con estos parámetros se evitó analizar proteínas que contienen secuencias homólogas a las astacinas pero que son clasificadas en familias diferentes a la astacina (M12).

6.2 Traducción de secuencias de nucleótidos a proteínas y selección de secuencias

Las secuencias de nucleótidos obtenidas de las bases de datos en NCBI fueron traducidas a proteínas usando la herramienta EMBOSS Transeq (https://www.ebi.ac.uk/Tools/st/emboss_transeq/). Se seleccionó la opción de traducir en los 6 marcos de lectura. Visualmente se identificaron las regiones no traducidas en los extremos amino y carboxilo de las secuencias y fueron eliminadas.

Las secuencias que carecían de las características distintivas (el motivo HEXXHXXGXXHEXXRXDR y met-turn (SXMHY)), se consideraron homólogos no funcionales y se excluyeron de los análisis posteriores. A continuación, se eliminaron las secuencias con inserciones largas (>15 aminoácidos) que parecían no tener homólogos en ninguna de las secuencias analizadas.

Finalmente, dado que NCBI es una base de datos altamente redundante, esta se redujo utilizando la herramienta Expsy Decrease Redundancy (https://web.expasy.org/decrease_redundancy/) con una alineación múltiple en formato FASTA como entrada creada en la herramienta Tcoffe (<https://tcoffee.crg.eu/apps/tcoffee/do:regular>) (Di Tommaso *et al.*, 2011) en un 90 % de similitud máxima.

6.3 Alineamiento de secuencias y cladogramas

6.3.1 Alineamiento de secuencias de zinc metalopeptidasas de decápodos

Para las secuencias de astacinas de decápodos se realizó un alineamiento múltiple utilizando el programa T-coffee regular (<https://www.ebi.ac.uk/Tools/msa/tcoffee/>) bajo los parámetros predeterminados.

6.3.2 Alineamiento de secuencias con dominio astacina de zinc metalopeptidasas y multidominio en *Penaeus vannamei*

Con base a las secuencias de proteínas seleccionadas de la base de datos UNIPROT, es decir las metalopeptidasas de *P. vannamei*, se realizó un alineamiento múltiple utilizando el programa T-coffee regular y se seleccionó la matriz de sustitución de proteínas Blosum.

6.3.3 Alineamiento de secuencias de zinc metalopeptidasas monodominio de astacina en *Penaeus vannamei*

Por otro lado, las secuencias de astacinas de *P. vannamei* putativamente digestivas obtenidas de la base de datos del NCBI fueron analizadas mediante un alineamiento múltiple realizado en T-coffee que fue analizado mediante Jalview 2.11.2.3 (Waterhouse *et al.*, 2009).

6.4 Análisis fisicoquímicos

Para la identificación y anotación de dominios de las secuencias de proteínas y el análisis de arquitecturas de dichos dominios se utilizaron las herramientas EMBL-EBI Pfam (<http://pfam.xfam.org/ncbiseq/398365647>) (Mistry *et al.*, 2021) e InterPro (Blum *et al.*, 2021).

La predicción de características moleculares y fisicoquímicas fue realizada solo en las secuencias maduras que contienen únicamente el dominio de astacina y la región de péptido señal y propéptido. Para el análisis fisicoquímico se utilizó la herramienta ExPASy-Protparam (<https://web.expasy.org/protparam/>). Esta herramienta estima el peso molecular, punto isoeléctrico teórico, y la composición de aminoácidos de una proteína, entre otras particularidades. Las características que indican la presencia de un péptido señal fueron analizadas utilizando la herramienta SignalIP 5.0 DTU (<https://services.healthtech.dtu.dk/service.php?SignalP-5.0>). La estructura secundaria fue modelada con Jpred4 (<https://www.compbio.dundee.ac.uk/jpred/>). La predicción de presencia de regiones transmembranales fue realizada con la herramienta DeepTMHMM

(<https://dtu.biolib.com/DeepTMHMM>). Finalmente, la presencia de enlaces disulfuro y su conectividad fue predicha con el software Disulfind (<http://disulfind.disi.unitn.it/>).

La estructura secundaria fue comparada con la descripción de Guevara y col., (2010) y el diagrama de la Figura 9 que representa un alineamiento de varios representantes de la familia de las astacinas, además de la estructura secundaria.

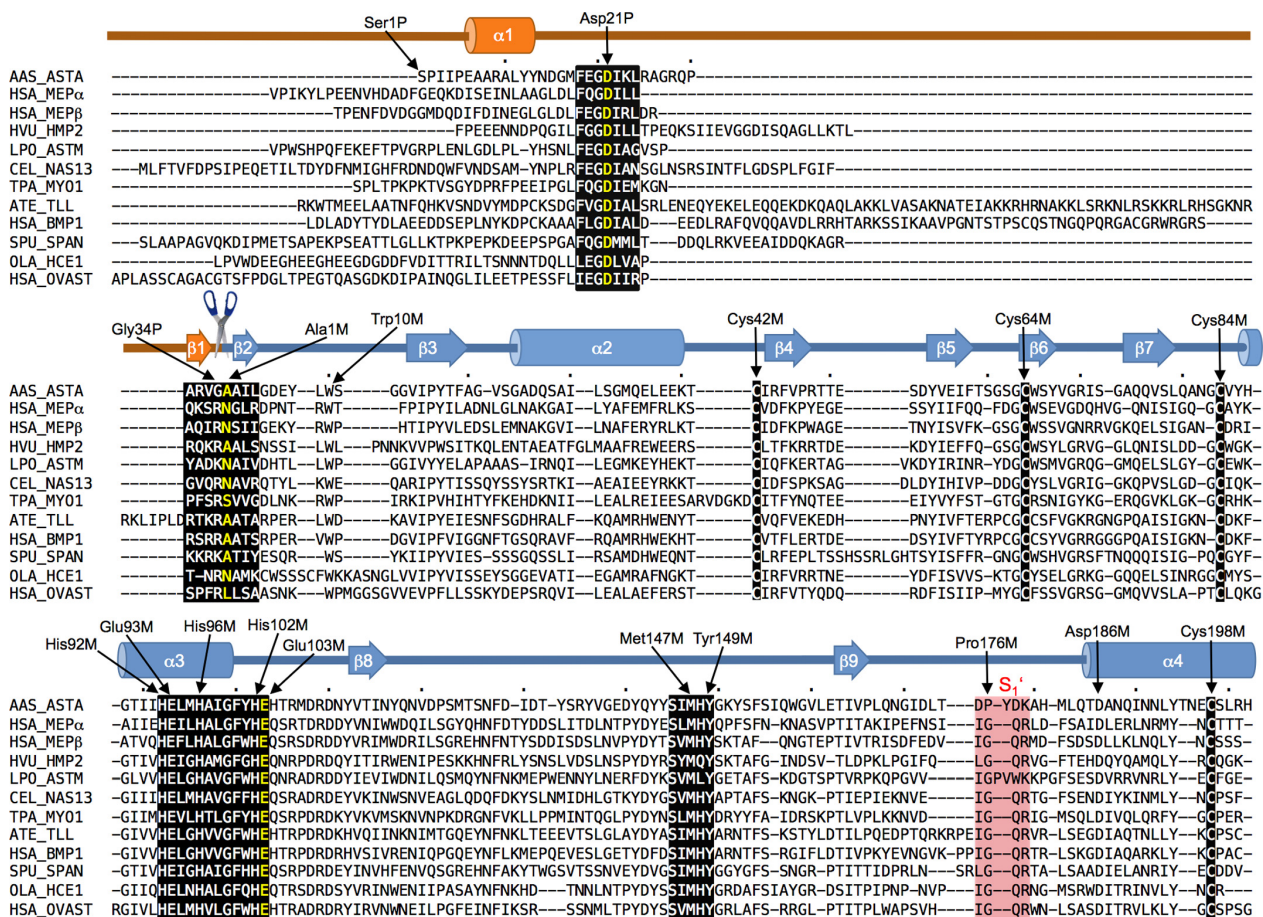


Figura 9. Diagrama esquemático de los componentes estructurales de varios miembros representativos de la familia astacina. La estructura secundaria se compone de alfa hélices (cilindros) y hojas beta (flechas), las tijeras indican el sitio de corte de activación final. La región de color naranja corresponde al péptido señal y al propéptido; la región azul es la proteína madura y dominio catalítico de astacina. En fondo negro se encuentran la secuencia del propéptido y el residuo conservado de aspártico (en amarillo), el sitio de activación, las cisteínas conservadas que hacen puentes disulfuro, la región del motivo, el giro de la metionina (met-turn) y en rosa el subsitio S₁'. Tomado de Guevara *et al.*, 2010.

6.5 Variabilidad de proteínas

Inicialmente, se realizó un alineamiento múltiple con las secuencias maduras seleccionadas en la herramienta T-coffee y posteriormente el resultado se utilizó en formato FASTA como entrada en Protein Variability Server (PVS, <http://imed.med.ucm.es/PVS/>). Se empleó el coeficiente de variabilidad Wu-Kabat (W) para calcular la variabilidad de la posición de aminoácidos, y se incluyó la secuencia de astacina de *A. astacus* como referencia. El coeficiente de variabilidad se calculó mediante la siguiente fórmula:

(1)

$$W = \frac{Nk}{n}$$

Donde la variable N indica el número total de secuencias en el alineamiento múltiple de secuencias, la variable k es el número de diferentes aminoácidos dada una posición y, por último, n se define como la frecuencia del aminoácido más común en una posición dada. El valor mínimo de W es 1 y el máximo incrementa con el número de secuencias en el alineamiento múltiple de secuencias (García-Boronat *et al.*, 2008).

El índice de Shannon fue utilizado para que el programa arrojara una secuencia con los aminoácidos no variables.

6.6 Análisis filogenético

6.6.1 Análisis filogenético para zinc metalopeptidasas con dominio astacina en decápodos

Con base al alineamiento de las secuencias de astacinas de decápodos, se realizó un análisis filogenético de máxima similitud, usando el programa PhyML-SMS (<http://www.atgc-montpellier.fr/phyml/>), bajo el algoritmo Smart Model Selection y un análisis de bootstrap de 1000 sustituciones. El árbol se visualizó usando la herramienta online iTOL (<https://itol.embl.de/>).

6.6.2 Análisis filogenético para zinc metalopeptidasas con dominio astacina y multidominio en *Penaeus vannamei*

Con los resultados generados por el alineamiento múltiple, se construyó un cladograma en UniProt usando el método Neighbor-Joining sin corrección de distancias, utilizando la secuencia de la astacina de *A. astacus* con fines comparativos.

6.6.3 Análisis filogenético para zinc metalopeptidasas monodominio de astacinasa en *Penaeus vannamei*

Se utilizó el modelo el valor de p como modelo de sustitución para la construcción del árbol. Se obtuvo un árbol filogenético, que representa la posible historia evolutiva de las metalopeptidasas de *P. vannamei* con MEGA-X, utilizando la secuencia P07584 de *A. astacus* como referencia y las secuencias encontradas y seleccionadas, con el modelo Neighbor-Joining y bootstrap de 1,000 repeticiones.

6.7 Modelo estructural

Para obtener más información sobre los elementos estructurales de las astacinas M12 de *P. vannamei*, se construyó un modelo de la estructura tridimensional de la secuencia usando el programa SWISS-MODEL (<http://swissmodel.expasy.org/>), empleando como templado a la astacina de *A. astacus* (PDB: 3IQ0.1.A). El modelo se interpretó con ayuda del software Pymol.

7 RESULTADOS

7.1 Secuencias de astacinas (M12) en Decápodos

En total se recuperaron 81 secuencias de la base de datos de GenBank del NCBI y después de eliminar aquellas redundantes se seleccionaron 60. Se debe considerar que el número de estas secuencias presenta un sesgo por el nivel de anotación de cada especie, es decir, el número de Bioprojects reportados en NCBI para Penaeidae es mucho mayor que para cualquier otra familia de decápodos. Así, de los 34 genomas encontrados en dicha base de datos, 14 son de Penaeidos, y como consecuencia, las especies de esta familia están mucho mejor representadas en este análisis. Es posible concluir que las astacinas de la familia M12 están presentes en todos los subórdenes del orden Decápoda, ya que se identificó la presencia de secuencias homólogas en los subórdenes Pleocyemata y Dendrobranchiata, con al menos un representante en las familias Astacidae, Portunidae, Oregoniidae, Cambaridae, Nephropidae, Lithodidae y Penaeidae.

7.2 Secuencias de astacinas (M12) mono y multidominios en *Penaeus vannamei*

Mediante la búsqueda en la base de datos UNIPROT (con baja redundancia) se encontraron 42 secuencias de *P. vannamei* que contienen el dominio astacina, estas secuencias fueron filtradas por la presencia del motivo HEXXHXXGXXHEXXRXDR completo. De esta manera, se obtuvieron y analizaron 27 secuencias de metalopeptidasas que contienen el dominio astacina. En la Figura 10 se puede observar el agrupamiento de las secuencias de acuerdo con sus características estructurales, por lo que fueron denominadas *monodominio* aquellas secuencias que sólo cuentan con el dominio Astacina y *multidominio* a aquellas que cuentan con dominios adicionales, entre ellos dominios CUB, ShKT y EGF.

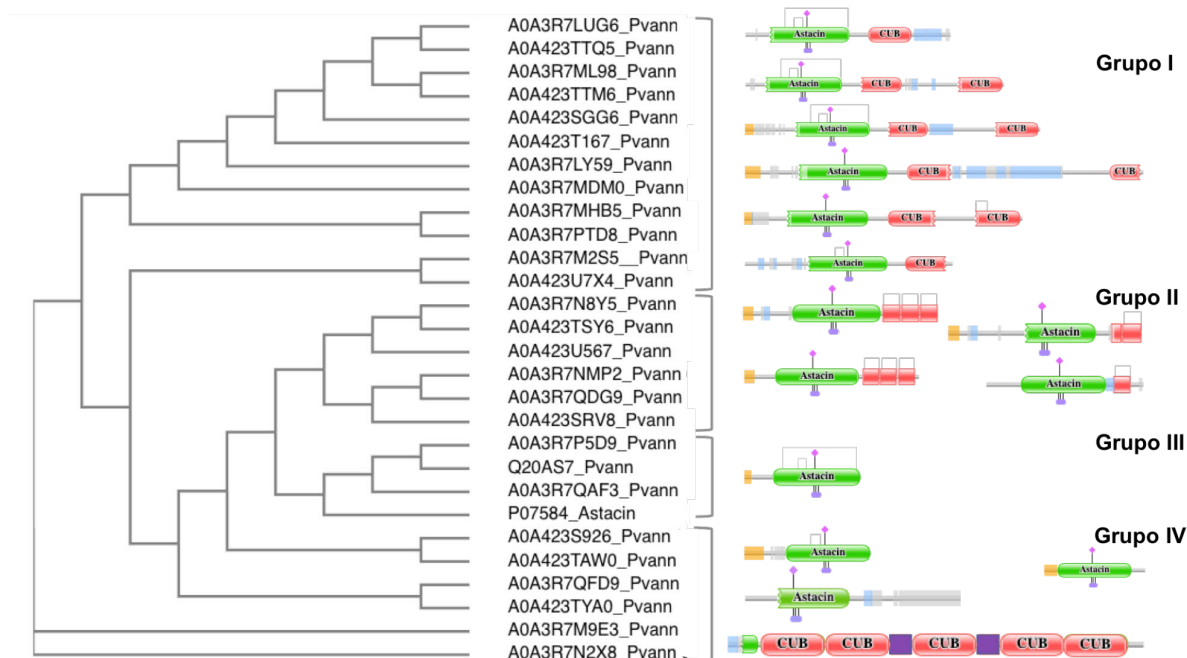


Figura 10. Cladograma y representación esquemática de secuencias homólogas y distribución de los dominios en metalopeptidasas de la familia de las astacinas en *P. vannamei*. La escala del diagrama está relacionada con la longitud de la secuencia. Pre y proproteínas, recuadros amarillos; dominio de astacina, rectángulo verde; dominios CUB, rectángulos rojos CUB; dominio ShKT, cuadros rojos; dominio EGF, cuadros morados; secuencias de baja complejidad, en azul; no especificado, en gris.

7.3 Secuencias de astacinas (M12) monodominio en *Penaeus vannamei*

Mediante la búsqueda más precisa de secuencias nucleotídicas homólogas realizadas en BLAST, se identificaron 153 secuencias con regiones que contienen el dominio de metalopeptidasas de la familia astacina de *P. vannamei*. Todas las secuencias se tradujeron a proteínas en los seis marcos de lectura, el marco correcto fue seleccionado para los siguientes análisis. Una tercera selección de secuencias se realizó tomando en cuenta la presencia del motivo HEXHXXGXXHEXXRXDR y la secuencia met-turn (SXMHY). Después de este filtrado se obtuvieron 69 secuencias proteicas. Posteriormente, se eliminaron las secuencias duplicadas o con un porcentaje de similitud >80 %, obteniendo 22 secuencias únicas. Finalmente, para este análisis no se tomaron en cuenta secuencias multidominio (de acuerdo con un análisis con Prosite) o que contienen insertos >15 aminoácidos no homólogos con otras secuencias de la lista, obteniendo 16 secuencias putativamente digestivas que se enlistan en la Tabla 3. Estas secuencias únicas tienen los motivos correspondientes y son monodominio, es decir, tienen solamente el dominio de astacina.

Las secuencias monodominio de astacina enlistadas en la Tabla 3, se agrupan en 16 variantes. Se eligió a un representante de cada variante para realizar un alineamiento múltiple (Fig. 11) y un cladograma (Fig. 12) y en ambos se incluyó la secuencia de astacina de *A. astacus* (P07584) para comparación.

Tabla 3. Números de acceso (GenBank) de las secuencias de las variantes de zinc metalopeptidasas monodominio de *P. vannamei* seleccionadas para el análisis, se indica el porcentaje de similitud con respecto a astacina.

No. Var.	Número de acceso en GenBank	% sim con AST	No. Var.	Número de acceso en GenBank	% sim con AST
1	GGKO01006419 ¹	67.62	11	GETZ01008581	57.66
	GGUK01034411			HAAW01015232	
	GGUK01030978			GGKO01006417	
	GGUK01031094			XM_027379199	
	GFRP01040203			GETD01026251	
	GGUK01031023				
	GGUK01002983				
	GETZ01007828				
	HAAW01015231				
	GGKO01006413				
2	GGUK01030804	66.19	12	JR494627 ⁹	57.66
				DQ398567	
3	GGKO01006418 ²	65.98		XM_027379201	
	GGUK01003068		GFRP01010103		
	GGUK01002924		GGUK01031011		
	GGUK01002988		GGUK01030798		
4	GETZ01046399 ³	65.98		GGUK01030798	
	HAAW01015233		GDTK01008983		
	GETD01026252		GGKO01006414		
			GGUK01033162		
			GDUV01005935		
			GETD01026254		
5	GGUK01030696 ⁴	64.19		GFRP01010104	
	GGUK01003104		GGUK01005017		
			GGUK01031041		
			GGUK01030650		
6	XM_027371679	63.87		HAAW01015228	
			GETD01026253		
7	GETZ01046361 ⁵	61.48		GFRP01010105	
	HAAW01015230				
8	GGUK01001973	60.87	13	GGUK01030711	56.76
9	GGUK01003085 ⁶	60.54	14	GGUK01030725	54.55
	GGUK01030979				
10	GGUK01000053 ⁷	58.8	15	GGUK01031049	47.96
	GGUK01003473		16	XM_027373234	41.5

Notas: Las secuencias ^{1,2,3,4,5,6,7,8,9} fueron utilizadas como representantes de una variante para realizar un cladograma.

7.4 Alineamiento de astacinas monodominio en *Penaeus vannamei*

A partir del alineamiento múltiple (Fig. 11) se distinguen algunas regiones conservadas características de la familia de las Astacinas. En la posición 156-173 se encuentra la región del motivo de astacina HEXXHXXGXXHEXXRXDR, que incluye un ácido glutámico (E) en la posición 157 que corresponde al sitio activo y solo se encuentra en la familia de Astacinas y tres histidinas (H) que se unen al zinc en las posiciones 156, 160 y 166; el subsitio S_1' que es responsable de la especificidad del corte, además de la región del propéptido donde está involucrada un ácido aspártico (D, posición 35-38); el sitio del propéptido (FXGD) la región del cambio de aspartato (aspartate-switch); la región que se describe como giro de metionina (Met-turn) se encuentra en la posición 212 al 216 y la metionina conservada en la posición 214. La tirosina (Y), dos lugares abajo (216), también participa en la catálisis y unión al zinc. Las cisteínas que forman los dos enlaces disulfuro identificados se encuentran en las posiciones 96, 128, 148 y 256.

	10	20	30	40	50			
XM_027373234/1-160	1	MYAAHLLS P		A		HPT	13	
GGUK01030711/1-188	1	MYQR				ACG P	8	
GGUK01030696*/1-234	1	R		DERGGDT	H TSGV RGNR	LRKEMNGQKPG	E 29	
GGK001006418*/1-246	1	MYG V L L L L A L	- V S L A G A A P F A P E A A R A L	- Y N P G L W E G D I	- K G V A G Q K P G	- E	46	
GGK001006419*/1-248	1	R F G I L L L L A V	- V A A A G A A P E V P A A A R A L	- Y N P H L F E G D I	- K G V A G Q K P G	- E	46	
GGUK01001973/1-208	1					PG	H 3	
GGUK01030725/1-221	1		A P	LAARTSRHT P	HDAATAM	YNPDLFQGD I	- K G I A G Q E P G	K 39
XM_027371679/1-246	1	MIWVRRVA	A G F G V L A G A A P F A P E A A R A L	- Y N P G L W E G D I	- K G V A G Q K P G	- E	48	
GGUK01030840*/1-249	1	M L L V T L L A V	- V A V A G A T P A V P R A A T A M	- Y N P D L F Q G D I	- K G I A G Q E P G	- K	46	
GGUK01030804/1-139								
GETZ01046399*/1-242	1	MI L V V L L A V	- A A V A S A T P T I P R A A Q A M	- Y N P K L F Q G D I	- M G I A G Q E P G	- H	46	
GGUK01031049/1-223	1	HQ						
GETZ01046361*/1-248	1	M F G I L L L L A V	- V A A A G A A P E V P A A A R A L	- Y N P H L F E G D I	- K G V A G Q K P G	- E	46	
GGUK01000053*/1-251	1	MI L V V L L A V	- A A V A S A T P T I P R A A Q A M	- Y N P K L F Q G D I	- M G I A G Q E P G	- H	46	
GGUK01003085*/1-186								
IR494627*/1-249	1	M L L V T L L A V	- V A V A G A T P A V P R A A T A M	- Y N P D L F Q G D I	- K G I A G Q E P G	- K	46	
P07584: Astacin/1-251	1	MQCAVLLV L	- L G V V A A S P I P E A A R A L Y N D G M F E G D I	- K L R A G R Q P A	- R		47	
Consensus		M+L V+L L A V G F V A V A G A + P + V P + A A R A + Y Y N P G L L				F+G D I R K G I A G Q + P G A E		

	60	70	80	90	C	100		
XM_027373234/1-160	14			S Q E R A V I A R A M A	E Y R Q S T	C I X L V P R	38	
GGUK01030711/1-188	9	R						
GGUK01030696*/1-234	30	E R S A I I G N E Y L W T G G V V P Y V I A G L		- Y E D Y I L A G M Q E I M D K T C I R F V E R			41	
GGK001006418*/1-246	47	E K S A I I G D Q F L W T D G V I P Y I L A G G		- S D A E I V A A M D E I M S K T C I K F I P R			93	
GGK001006419*/1-248	47	E R S A I I G N E Y L W T G G V V P Y V I A G L		- Y E D Y I L A G M Q E I M D K T C I R F V E R			93	
GGUK01001973/1-208	4	E R A A I L G D E Y L W S G G V V P Y I F G P S V D S Y Q K S V I L A G M S D F H D R T C I R F I E R					54	
GGUK01030725/1-221	40	E R A A I L G E Q Y L W P G G V V P Y V F G N S V N S Y E R S L I I S A M D F H A R T C I R F V E R					90	
XM_027371679/1-246	49	E K S A I I G D X F L W T D G V I P Y I L A G G		- S D A E I V A A M D E I M S K T C I K F I P R			95	
GGUK01030840*/1-249	47	E R A A I L G E Q Y L W P G G V V P Y V F G N S V N S Y E R S L I I S A M D F H A R T C I R F V E R					97	
GGUK01030804/1-139								
GETZ01046399*/1-242	47	E R A A I L G D E Y L W S G G V V P Y I F G P S V D S Y Q K S V I L A G M S D F H D R T C I R F I E R					97	
GGUK01031049/1-223	26	R R R S L C V R Q F C Q Q L R E V P H H F R H G C H		- Q S - H P G T S P C R A R A L R H R R R R			71	
GETZ01046361*/1-248	47	E R S A I I G N E Y L W T G G V V P Y V I A G L Y E		- D Y I L A G M Q E I M D K T C I R F V E R			93	
GGUK01000053*/1-251	47	E R A A I L G D E Y L W S G G V V P Y I F G P S V D S Y Q K S V I L A G M S D F H D R T C I R F I E R					97	
GGUK01003085*/1-186	1							
IR494627*/1-249	47	E R A A I L G E Q Y L W P G G V V P Y V F G N S V N S Y E R S L I I S A M D F H A R T C I R F V E R					97	
P07584: Astacin/1-251	48	V G A A I L G D E Y L W S G G V I P Y T F A G V - S G A D Q S A I L S G M Q E L E K T C I R F V P R					97	
Consensus		E R A A I L G D E Y L W T G G V V P Y V F G G S V N S Y E R S + I L A G M D + F H D R T C I R F V E R						

	110	120	C	130	140	C	150	
XM_027373234/1-160	39	S - S H A D Y I H I L R G Q G Q M L Y I R V P L G	V P C A V G R S G G V Q V V S L G	- Y G C V Q I G V				87
GGUK01030711/1-188	42	T - T E A N Y I Y I T T S G A		- S G G G C W S Y V G M M G G W Q E V S L D Q Y G C I Y H G T				85
GGUK01030696*/1-234	77	T - T E A N Y I Y I T T S G A		- S G G G C W S Y V G M M G G W Q E V S L D Q Y G C I Y H G T				120
GGK001006418*/1-246	94	S V G E H N Y I R I T T N G		- D G C W S Y V G M M D G G Q E I S L A R N G C I Y H G T				135
GGK001006419*/1-248	94	T - T E A N Y I Y I T T S G A		- S G G G C W S Y V G M M G G W Q E V S L D Q Y G C I Y H G T				137
GGUK01001973/1-208	55	T - T Q A N Y L E I V T N D		- S G C W S Y V G T I G G M Q R L S L D T N G C I Y K G T				95
GGUK01030725/1-221	91	T - S Q S N Y I E I V S N D		- S G C W S Y V G T I G G K Q R V S L D S N G C M Y K G T				131
XM_027371679/1-246	96	S V G E H N Y I R I T T N G		- D G C W S Y V G M M D G G Q E I S L A R N G C I Y H G T				137
GGUK01030840*/1-249	98	T - S Q S N Y I E I V S N D		- S G C W S Y V G T I G G K Q R V S L D S N G C M Y K G T				138
GGUK01030804/1-139	1							
GETZ01046399*/1-242	98	T - T Q A N Y L E I V T N D		- S G C W S Y V G T I G G M Q R L S L D T N G C I Y K G T				138
GGUK01031049/1-223	72	T - S Q S N Y I E I V S N D		- S G C W S Y V G T I G G K Q R V S L D S N G C M Y K G T				112
GETZ01046361*/1-248	94	T - T E S N Y I Y I T T S G A		- S G G G C W S Y V G M M G G W Q E V S L D Q Y G C I Y H G T				137
GGUK01000053*/1-251	98	T - T Q A N Y L E I V T N D		- S G C W S Y V G T I G G M Q R L S L D T N G C I Y K G T				138
GGUK01003085*/1-186	35	T - S Q S N Y I E I V S N D		- S G C W S Y V G T I G G K Q R V S L D S N G C M Y K G T				75
IR494627*/1-249	98	T - S Q S N Y I E I V S N D		- S G C W S Y V G T I G G K Q R V S L D S N G C M Y K G T				138
P07584: Astacin/1-251	98	T - T E S D Y V E I F T S G		- S G C W S Y V G R I S G A Q Q V S L Q A N G C V Y H G T				138
Consensus		T V T Q + N Y I E I V T N D A Q M L Y I R S G S G C W S Y V G T I G G K Q R V S L D S N G C I Y K G T						

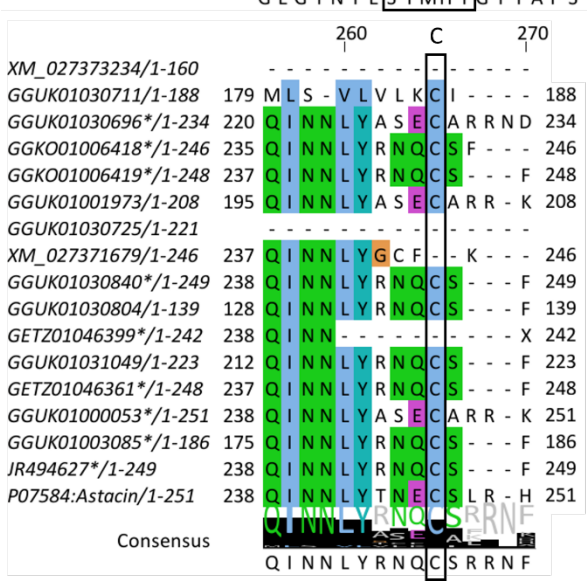
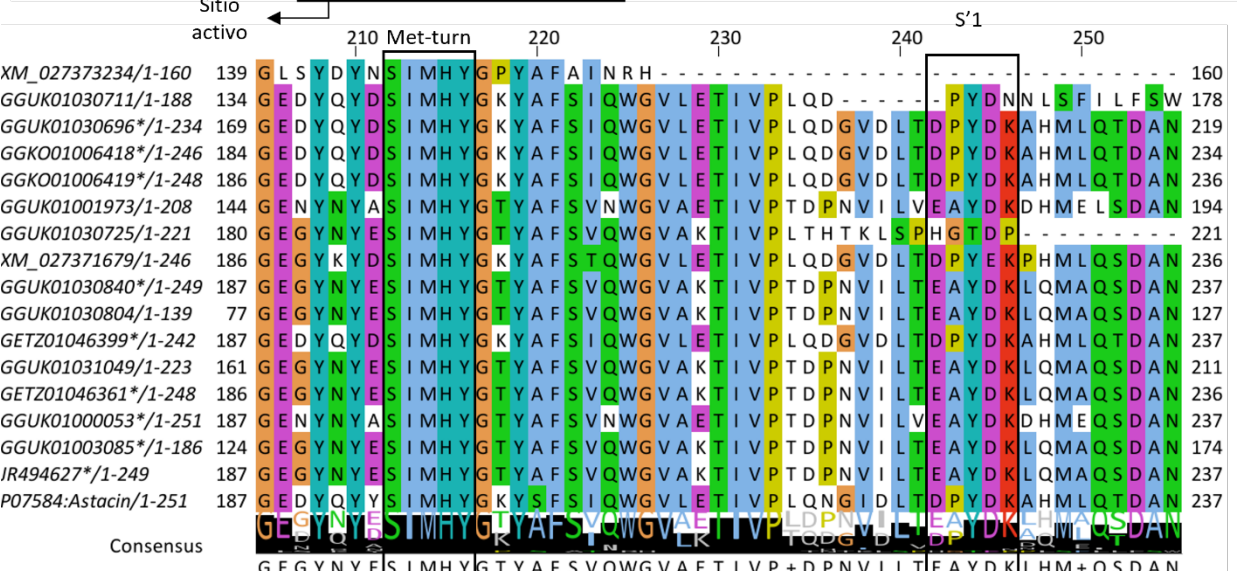
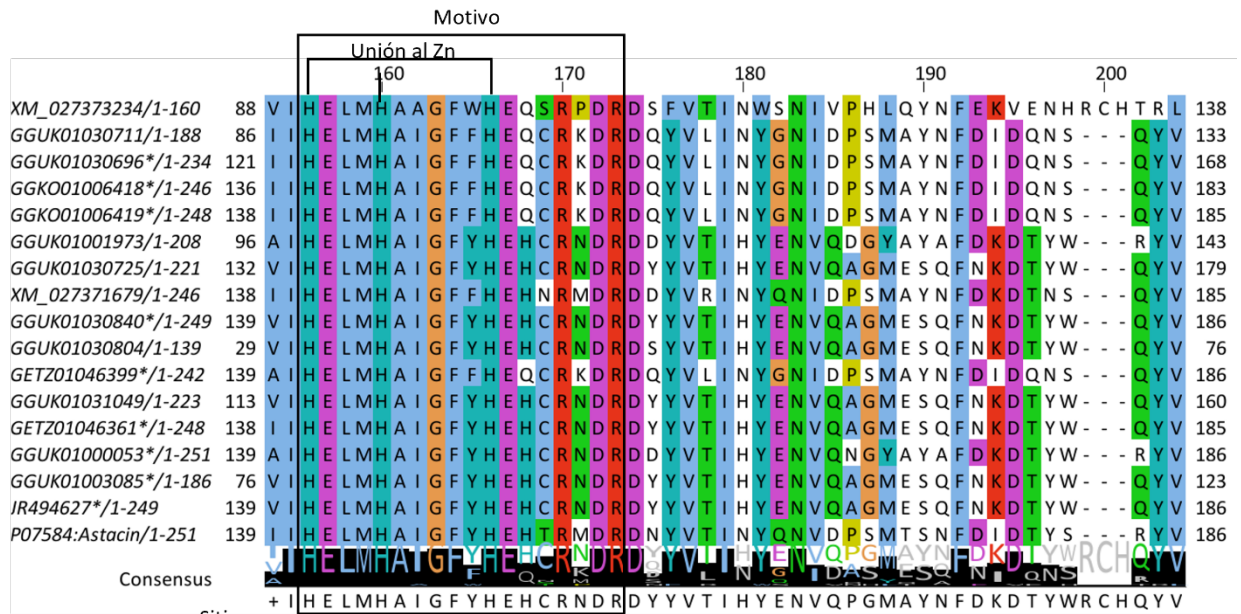


Figura 11. Alineamiento de secuencias de astacinas monodominio de *P. vannamei*. En recuadros negros se enmarcan varios sitios de la estructura de astacina. Péptido señal, el propéptido y el sitio de activación, cisteínas que participan en la formación de enlaces disulfuro, el motivo donde se señala el sitio activo y los sitios de unión al zinc, el giro de metionina y el subsitio S1'.

Para la construcción de un cladograma (Fig. 12) de las secuencias monodominio de zinc metalopeptidasas se empleó el método Neighbor-Joining con 1000 repeticiones bootstrap. La secuencia de *A. astacus* (P07584) fue incluida en los análisis con fines comparativos. Las distancias evolutivas se calcularon utilizando el método de distancia p y están en las unidades del número de diferencias de aminoácidos por sitio. La variación de la tasa entre sitios se modeló con la distribución gamma (parámetro de forma =1) y se eliminaron los gaps o sitios sin datos.

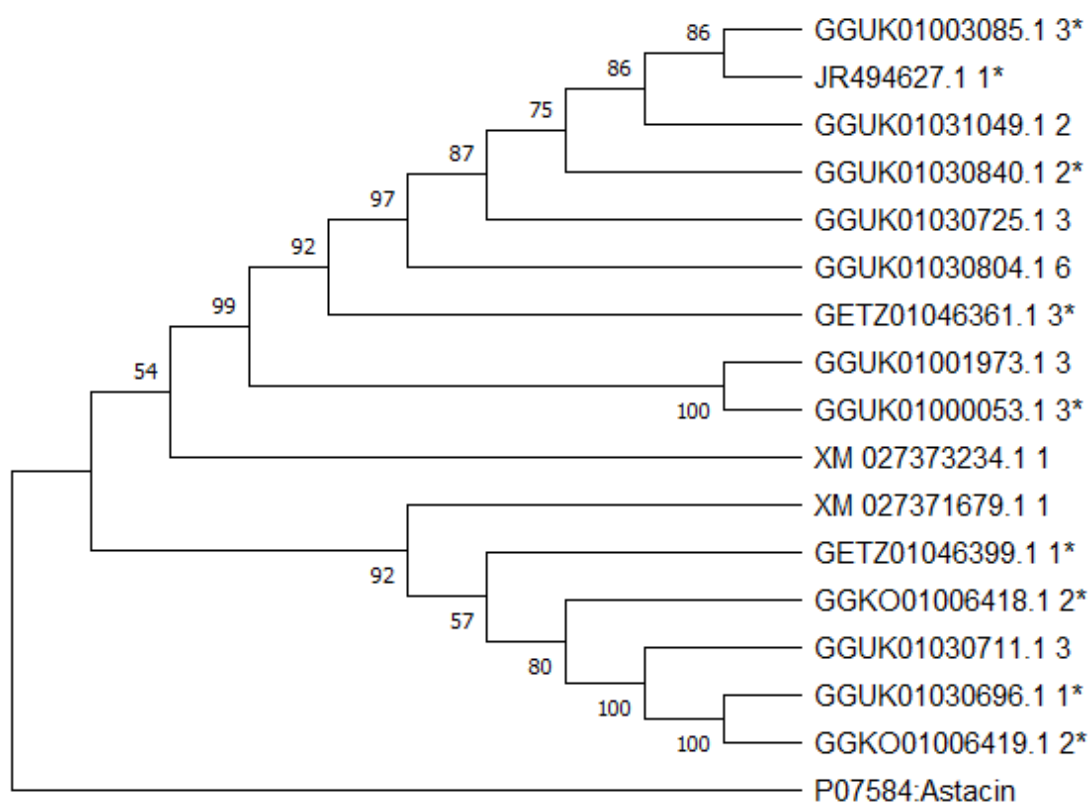


Figura 12. Cladograma de secuencias homólogas de zinc metalopeptidasas monodominio de la familia astacina monodominio de *P. vannamei*.

7.5 Características fisicoquímicas identificadas *in silico*

Los análisis fisicoquímicos que se reportan a continuación se realizaron en las secuencias monodominio de astacina de *P. vannamei*. El número de aminoácidos, el peso molecular y el punto isoelectrónico calculado para cada secuencia de las variantes se enlistan en la Tabla 4; en la Tabla 5 se indica la composición de aminoácidos. Los propéptidos y los péptidos señal detectados en cada secuencia se muestran en el alineamiento de la Fig. 13

Solamente ocho variantes contienen péptido señal, las cuales se muestran en la Figura 13, mientras que siete secuencias contienen el propéptido. En conjunto, seis secuencias contienen ambas regiones de péptido señal y pro-péptido.

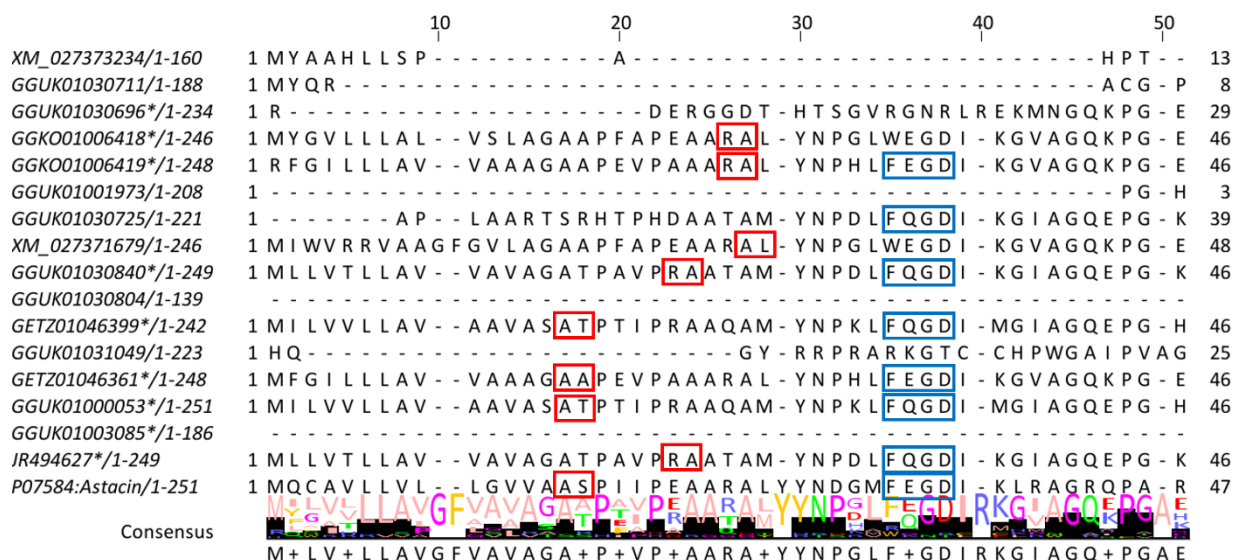


Figura 13. Alineamiento de las secuencias de las variantes de astacinas monodominio de *P. vannamei* que contienen el propeptido y péptido señal. En los recuadros rojos se indica el sitio de corte del péptido señal y en los recuadros azules el pro péptido

En promedio, las secuencias se conforman de 198 aminoácidos, con un promedio de 22.7 KDa de peso molecular (MW). Los valores de punto isoelectrónico (pI) teórico computados en la herramienta ProtParam de Expassy (Tabla 4) tienen un rango estrecho que van desde el 4.39 al 5.25, lo que indica características ácidas.

Tabla 4. Número de aminoácidos (# aa), peso molecular (MW) y punto isoeléctrico teórico de las secuencias de variantes de astacinas.

	Variantes	# aa	MW	pl
1	XM_027373234	-	-	-
2	GGUK01030711	-	-	-
3	GGUK01030696	-	-	-
4	GGKO01006418	-	-	-
5	GGKO01006419	199	22772.50	4.39
6	GGUK01001973	-	-	-
7	GGUK01030725	-	-	-
8	XM_027371679	-	-	-
9	GGUK01030840	200	22953.54	5.25
10	GGUK01030804	-	-	-
11	GETZ01046399	192	21943.67	4.49
12	GGUK01031049	-	-	-
13	GETZ01046361	199	22826.51	4.70
14	GGUK01000053	202	23118.63	4.92
15	GGUK01003085	-	-	-
16	JR494627	200	22953.54	5.25
	Promedio	198	22761.39	4.83

En la Tabla 5, se presenta la composición de aminoácidos de las secuencias de astacinas maduras de *P. vannamei*. Estas secuencias son pobres en triptófano (W) y ricas en tirosina (Y). Se observa una baja frecuencia de cisteína (C), fenilalanina (F), histidina (H), lisina (K), metionina (M), prolina (P), arginina (R) y triptófano (W). Con relación a la Figura 14, sobre la variabilidad en la composición de las astacinas de *P. vannamei* se observa que los aminoácidos con frecuencia variables son metionina con valores de 2.48 a 4.02, ácido aspártico con valores de 5 a 9.4, ácido glutámico con valores de 3.65 a 6.53, leucina con valores de 4 a 6.25, glutamina con valores de 2.97 a 6.53, arginina con valores de 2.51 a 4.46, serina con valores de 4.52 a 8, valina con valores de 5.53 a 8 y tirosina con valores de 7.81 a 10.55. En general son ricos en ácido aspártico (D), glicina (G), isoleucina (I) y tirosina (Y).

Tabla 5. Composición de aminoácidos de las secuencias de zinc metalopeptidasas en *P. vannamei* representados en porcentajes.

	Ala	Cys	Asp	Glu	Phe	Gly	His	Ile	Lys	Leu	Met	Asn	Pro	Gln	Arg	Ser	Thr	Val	Trp	Tyr	Total
5⁵	5.03	2.51	7.54	5.03	3.02	10.05	3.02	9.55	2.01	5.53	4.02	5.03	2.01	6.53	2.51	4.52	5.03	5.53	2.01	9.55	199
9⁹	5.00	2.5	5	5.5	3.5	8	3.5	7	2.5	4	3	7	2	5.5	4	8	5	8	2	9	200
11¹¹	4.69	2.08	9.38	3.65	3.65	8.85	3.13	8.85	2.6	6.25	3.13	5.21	2.6	6.25	3.13	6.25	5.21	5.73	1.56	7.81	192
13¹³	6.93	2.48	7.43	5.45	2.97	7.92	3.96	7.43	2.48	4.46	2.48	6.44	1.98	2.97	4.46	5.94	5.94	6.93	1.98	9.41	202
14¹⁴	5	2.5	5	5.5	3.5	8	3.5	7	2.5	4	3	7	2	5.5	4	8	5	8	2	9	200
16¹⁶	5.03	2.51	7.54	5.03	3.02	10.05	3.02	9.55	2.01	5.53	4.02	5.03	2.01	6.53	2.51	4.52	5.03	5.53	2.01	9.55	199
Avg	5.24	2.37	6.46	5.31	3.16	8.76	3.52	8.04	2.23	4.88	3.23	5.89	2.08	5.38	3.45	6.6	5.67	6.68	1.94	9.12	199

Nota: ⁵ GGK001006419*, ⁹ GGUK01030840*, ¹¹ GETZ01046399*, ¹³ GETZ01046361*, ¹⁴ GGUK01000053*, ¹⁶ JR494627*. Las secuencias marcadas con * indican con son los representantes de ese número de variable

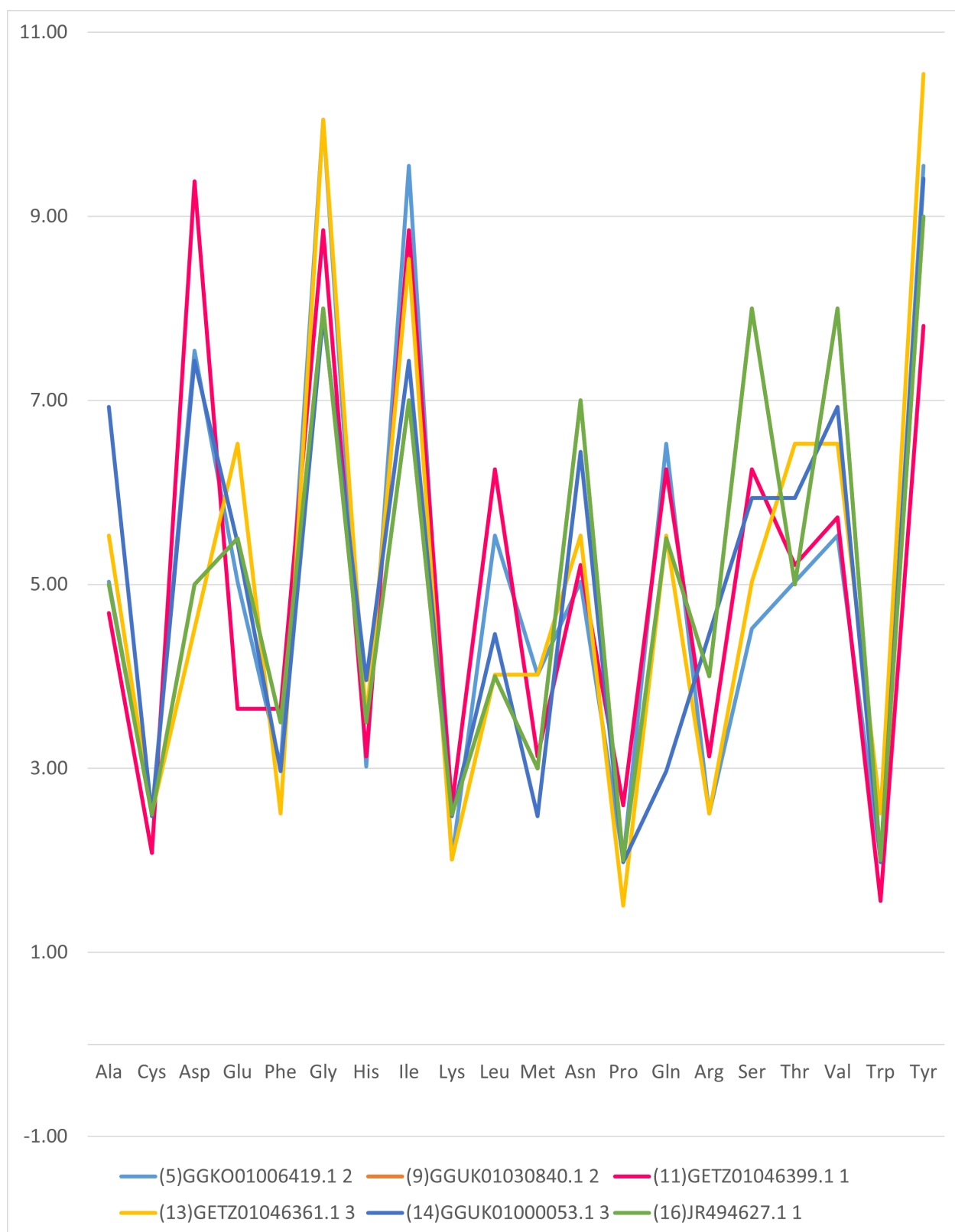


Figura 14. Gráfico de la variabilidad en la composición de las astacinas de *P. vannamei* con relación a los datos de la Tabla 5.

7.6 Variabilidad de proteínas

La variabilidad de las secuencias maduras de zinc metalopeptidasas monodominio de astacina en *P. vannamei* fueron calculadas con el coeficiente de variabilidad Wu-Kabat que se presentan en la Figura 15. La secuencia consenso de astacina de *P. vannamei* se derivó del alineamiento múltiple de secuencias utilizando el servidor PVS (Fig. 15).

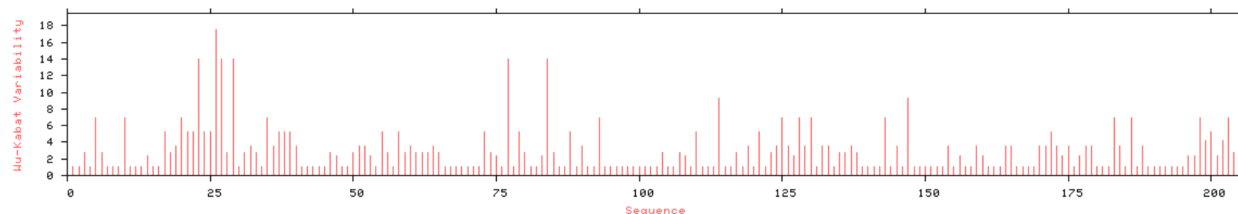


Figura 15. Índice de variabilidad de proteínas Wu-Kabat para cada uno de los 205 residuos de aminoácidos de las zinc metalopeptidasas monodominio de astacina.

En la secuencia (en formato FASTA), se muestran los aminoácidos donde la variabilidad es mayor o igual que el umbral de variabilidad (en este caso 1 índice de Shannon) y se representan con un punto (·) (Fig. 16).

```
>142237272022_query
AILG.EYLW.GGVVPY.FG.....S·.ILAGM.D...RT·CIRFVERTTQSNY·EI·TNSGC·WSYVG·
IGG·Q·VSLD·NGC·YKGT·I·HELMHAIGFYHEHCR·DRD·YVTIHY·NVQ·GM·S·FDKDTYWQYVGE·
YNY·SIMHY·GTYAFSVQWGVAETIVPTD·NVILTEAYDK·HM·QSDANQINNLY...C...
```

Figura 16. Secuencia consenso de la variabilidad de las secuencias de zinc metalopeptidasas monodominio astacinas. En rojo: cisteínas, motivo y giro de metionina.

En la Tabla 6 se identifican los fragmentos de longitud mínima de seis aminoácidos en la secuencia consenso que consta únicamente de residuos consecutivos con variabilidad por debajo del umbral de variabilidad establecido (<1).

Tabla 6. Fragmentos conservados con residuos no variables. ($V \leq 1$).

N	Start	End	Sequence
1	11	16	GGVIPY
2	40	54	KTCIRFVPRTTESDY
3	62	72	TSGSGCWSYVG
4	94	109	IHELMHAIGFYHEHTR
5	115	120	YVTINY
6	131	142	FDIDTYSRYVGE
7	148	171	SIMHYGKYSFSIQWGV LETIVPLQ
8	173	182	GIDLTPYDK
9	187	197	QTDANQINNLY

7.7 Filogenia de las astacinas de la familia M12 en decápodos

El análisis filogenético las astacinas del orden Decápoda indica que estas se agrupan en 4 ramas claramente diferenciadas (Fig. 17), las nombramos grupo A (color azul), B (color rojo), C (color verde) y D (color morado). El árbol corresponde a las características de las secuencias, las cuales son altamente variables. El grupo D contiene la mayoría de las secuencias con representantes de los infraórdenes Anomura, Astacidea, Brachyura, así como de la familia Penaeidae; mientras que el grupo A tiene solamente dos especies de la familia Penaeidae.

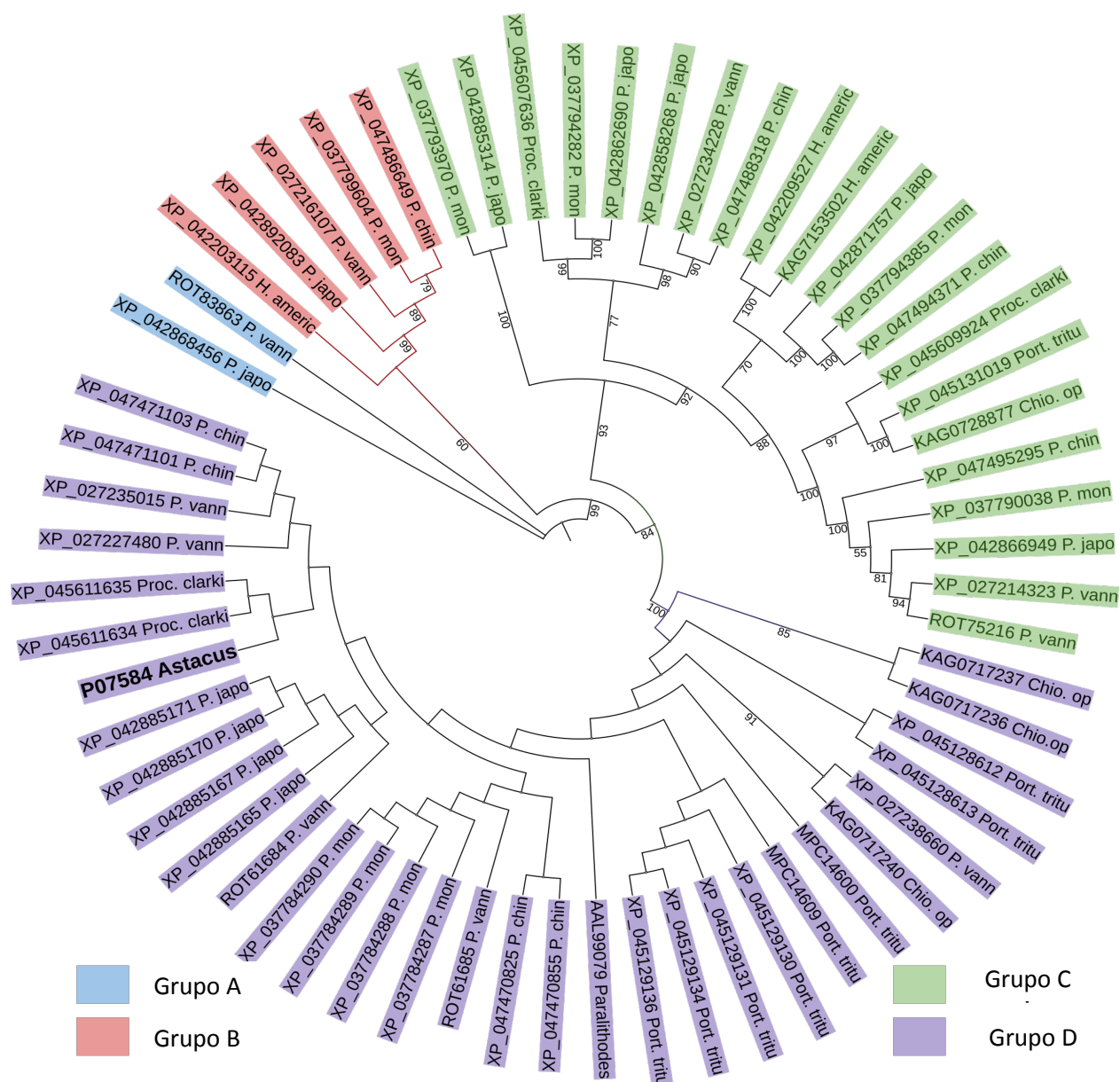


Figura 17. Análisis filogenético de secuencias de astacinas de diferentes especies del orden Decápoda. Se incluyen los valores de Bootstrap. Cada una de las cuatro ramas principales en donde se agrupan las secuencias de astacinas de diversos decápodos está representada por colores diferentes. El grupo A se representa de color azul, el grupo B de color rojo, el grupo C de color verde y el grupo D de color morado.

7.8 Modelo estructural de la astacina M12 de *P. vannamei*

Se realizó una predicción del modelo de la estructura espacial de un representante de las astacinas de *P. vannamei*. Dicho modelo se construyó con la finalidad de analizar la posible unión al zinc y otras características estructurales. La estructura molecular de la astacina de *Astacus astacus* en complejo con zinc obtenida por difracción de rayos X (código PDB: 3LQ0), se

utilizó como molde. Se modeló la secuencia proteica de la zinc proteasa Mpc1 de *P. vannamei* (número de acceso GenBank: ROT61684 Uniprot: A0A3R7QAF3), como referencia ya que tienen una identidad de 58.8 % con la secuencia de *Astacus astacus*.

Como se observa en la figura 18, el modelo estructural de la Mpc1 de *P. vannamei* muestra una arquitectura conservada comparable a la de las proteasas de la familia M12.

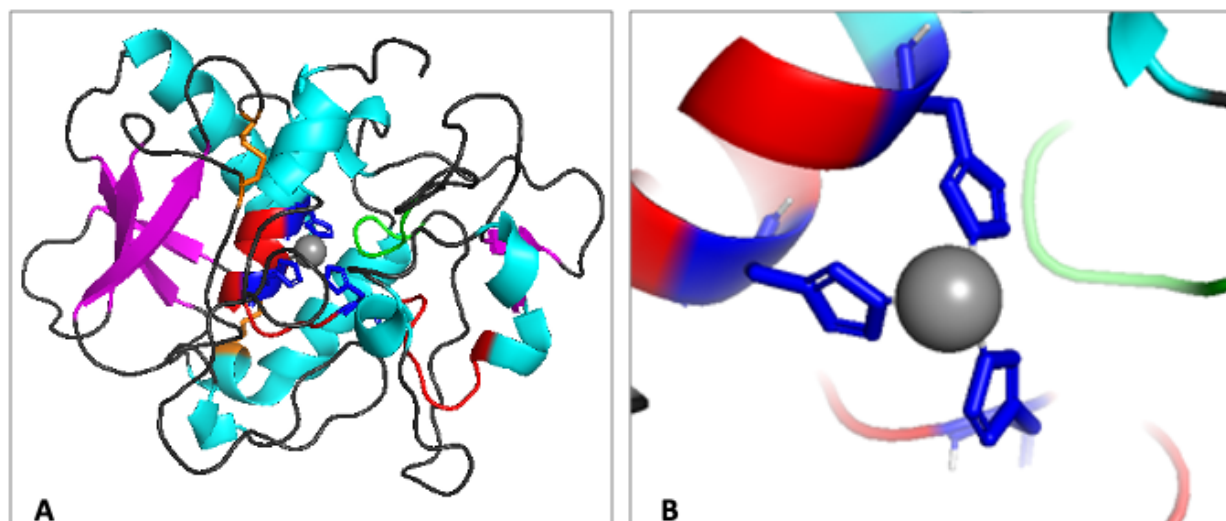


Figura 18. Modelo de homología la MPC1 de *P. vannamei* con una molécula de zinc como ligando. El modelo se construyó utilizando la estructura de rayos X la astacina de *Astacus astacus*. A) Las placas- β y las α -hélices se muestran en púrpura y cian, respectivamente. Se resaltan las regiones del motivo HEXXHXXGXXHEXXRXDR (rojo). Las cadenas laterales de histidinas que se unen al zinc (azul), la región del met-turn (verde), cisteínas unidas por enlaces disulfuro (bastones naranjas), el ion de zinc se muestra como esfera plateada. B) Acercamiento al ion de zinc catalítico en donde se distingue que está putativamente asociado a las 3 histidinas del motivo catalítico, los iones hidrógeno de las histidinas que estabilizan el sitio catalítico se representan como bastones blancos.

8. DISCUSIÓN

Las metalopeptidasas del tipo astacina son ubicuas en el reino animal, pero sus relaciones filogenéticas y funciones no se han definido claramente aún (Möhrlen *et al.*, 2006). Para el caso de las astacinas de decápodos, un grupo muy diverso de crustáceos, la información sobre su origen evolutivo y función fisiológica es menos clara aún. Para el caso particular de *P. vannamei*, la primer secuencia de nucleótidos de una astacina se reportó en 2007 por Zhao y colaboradores. Sin embargo, hasta la fecha se desconoce si esta enzima posee una función digestiva o participa en otros procesos fisiológicos. Por esa razón el estudio de las características estructurales, bioquímicas y evolutivas de estas enzimas es relevante y se aborda en este trabajo, analizamos a través de herramientas bioinformáticas, las variantes de las secuencias de astacinas reportadas para *Penaeus vannamei*.

De acuerdo con la información de la base de datos MEROPS (especializada en peptidasas), hasta julio de 2022, se han depositado un total de 1,142,788 secuencias, de las cuales 390,108 (34.1 %) son metalopeptidasas, lo que significa que, de las secuencias anotadas en esta base de datos, las metalopeptidasas son la segunda familia de peptidasas más abundante, después de las serino peptidasas (36.8 %). En esta misma base de datos están depositadas 74 secuencias de peptidasas de decápodos, de las cuáles solo el 17.5 % son metalopeptidasas, siendo las cisteíno peptidasas la familia con más representantes para este taxón (37.8 %), mientras que las secuencias serino peptidasas representan el 36.4 % del total de entradas para el orden Decápoda. Para el caso específico de *P. vannamei*, hay 32 secuencias depositadas, 3 de las cuales corresponden a metalopeptidasas, y dos más son secuencias a las que no les ha sido asignada una identidad. Estos datos resaltan la relevancia de las secuencias caracterizadas en este trabajo de manera general en el degradoma³ de *P. vannamei*, y de manera particular su función en la digestión del alimento.

La mayoría de las secuencias homólogas de astacina encontradas en *P. vannamei* no tienen ninguna anotación en las bases de datos, solamente se indica que es un transcrito, pero no se especifica que sea de astacina o que sea una zinc metalopeptidasa, así que las búsquedas por palabras son poco exitosas e insustanciales. Por esta razón, en primera instancia se eligió buscar

³ El degradoma se ha definido como el conjunto de genes de proteasas que están presentes en un organismo o el repertorio de proteasas expresadas por un tejido o célula en particular (López-Otín y Overall, 2002)

homólogos de la proteína de astacina de *A. astacus* (P07584), la cual está completamente caracterizada y anotada. La mayoría de las secuencias encontradas en las bases de datos son multidominio, por lo que en este trabajo se hizo una distinción entre las proteínas con un solo dominio catalítico astacina y las que tenían uno o más módulos no catalíticos en las secuencias. Encontramos que en los análisis de cladograma las secuencias se agrupan en 4 grupos (Fig. 10).

Como se mencionó en los antecedentes, las peptidasas de la familia de las astacinas suelen contener dominios adicionales en el extremo C-terminal. Entre los dominios más comúnmente identificados están los dominios EGF y una o varias copias de los dominios CUB. Dichos dominios podrían estar involucrados en interacciones de unión a calcio, proteína-proteína o enzima-sustrato (Gomis-Ruth *et al.*, 2012). Las astacinas exhiben una amplia variedad de longitudes y secuencias, lo que sugiere una alta diversidad de funciones dentro de una sola familia (Martín-Galiano y Sotillo, 2021). En helmintos, la mayoría de las proteínas que se secretan para facilitar la infección y la colonización en el hospedero pertenecen a la familia astacina. Cabe mencionar que de manera global, el dominio astacina ha sido encontrado en asociación con otros 110 dominios diferentes, constituyendo 145 arquitecturas de dominios únicas en donde los módulos CUB y Shk son los más comúnmente reportados, por lo que se ha sugerido que debido a la presencia de dominios coexistentes (en forma de módulos) tienen funciones adaptables promiscuas asociadas a varios roles (Martín-Galiano y Sotillo, 2021).

El análisis del cladograma de las secuencias de astacinas mono- y multidominio de *P. vannamei* (Fig. 10), las agrupa claramente en cuatro grupos, que comparten entre sí dominios adicionales al de astacina. En el grupo I se encuentran secuencias que contienen uno o dos dominios CUB, distribuidos en secuencias con tamaños variables y que pueden estar separados por regiones de baja complejidad, esta distribución es relativamente común en diferentes taxa de vertebrados e invertebrados (Geier y Zwillig, 1998). Se ha propuesto la hipótesis de que una de las funciones del dominio CUB es el reconocimiento de sustratos y otros ligandos, guiando a las astacinas hacia sustratos específicos y/o incrementando de manera local el número de sitios activos de astacina mediante oligomerización (Galeano *et al.*, 2021).

En el grupo II se encuentran secuencias asociadas a uno o tres módulos del dominio ShK, y la función de las astacinas con esta arquitectura es relativamente poco conocida. Se sabe que las astacinas asociadas al dominio ShK se encuentran solamente en invertebrados (Gerdol *et al.*,

2019) y específicamente, las astacinas asociadas al dominio ShK se han descrito solo en invertebrados hematófagos (Gerdol *et al.*, 2019; Galeano *et al.*, 2021). En esta tesis se reporta por vez primera una secuencia de astacina con dicha arquitectura en un invertebrado no-hematófago.

En el grupo III está comprendidas las astacinas monodominio, consideradas aquí como digestivas, pues comparten dicha arquitectura con la astacina de *A. astacus*, la astacina mejor estudiada a nivel funcional y estructural, cuya función exclusivamente digestiva ha sido descrita previamente (Zwilling y Neurath, 1981; Vogt *et al.*, 1989; Dumermuth *et al.*, 1991; Stöcker y Zwilling, 1995; Stöcker y Yiallourous, 2011). Los subsecuentes análisis y modelaciones bioinformáticas fueron realizadas en secuencias de este grupo.

En el grupo IV se encuentran las astacinas de un solo dominio que contienen elementos estructurales con arquitectura atípica considerada no funcional. Finalmente, en la base del cladograma se encuentra una secuencia que contiene el dominio astacina truncado, asociado con dominios CUB y EGF, una estructura típica de las astacinas del grupo BMP1/Tolloid, que tienen un arreglo de cinco dominios CUB y dos segmentos EGF (Hopkins y Greenspan, 2007), pero que en *P. vannamei* puede no ser funcional por no contener el dominio astacina completo. Una explicación plausible para esto es que, como se observa en la Figura 10, dicha secuencia carece de los últimos 2 aminoácidos del motivo astacina.

La arquitectura particular de los subgrupos I al IV, parecen proporcionar una función específica para cada gen de astacina de *P. vannamei* y los miembros de cada subgrupo pueden tener funciones similares o idénticas.

8.1 Secuencias de astacinas monodominio de *P. vannamei*

En el análisis de las secuencias monodominio de *P. vannamei*, se observa que existen cuatro grandes grupos con características particulares. La primera sección la forman el pre- y pro-segmento, la segunda es la cadena catalítica del dominio de astacina y finalmente, la extensión C-terminal, la cual contiene las regiones funcionales de la proteína como el subsitio S₁.

Ocho secuencias contienen el pre-segmento (péptido señal) (Fig. 13) (GGKO01006418, GGKO01006419, XM_027371679, GGUK01030840, GETZ01046399, GETZ01046361, GGUK01000053 y JR494627). El propéptido (pro-segmento) se encuentra en siete secuencias

(GGKO01006419, GGUK01030725, GGUK01030840, GETZ01046399, GETZ01046361, GGUK01000053 y JR494627) y ambos se encuentran solamente en seis secuencias (GGKO01006419, GGUK01030840, GETZ01046399, GETZ01046361, GGUK01000053 y JR494627). En este caso, probablemente la falta de péptido señal en algunas secuencias se deba a artefactos relacionados con el método de secuenciación y ensamblaje o en la traducción de nucleótido a aminoácidos.

En el dominio catalítico de astacinas de *P. vannamei* se encuentran los residuos esenciales para la actividad catalítica, mientras que en la región C-terminal se encuentra la región estructural correspondiente al giro de metionina y el subsitio S'_1 . Estas estructuras se han identificado experimentalmente por cristalografía en la astacina de *A. astacus* y se encuentran conservadas en el camarón, por lo tanto, se sugiere la función digestiva de las astacinas monodominio de *P. vannamei*.

La función digestiva de las astacinas ha sido estudiada en algunos invertebrados. Así, Fuzita y cols., (2015a) detectaron en el proteoma de la araña *Nephilingis cruentata* 25 astacinas en el fluido digestivo y sugieren estas variantes pudieran desempeñar un papel importante en la digestión extracorpórea en Araneae. Antes de este estudio estas astacinas solo habían sido asociadas a las glándulas de veneno de esta araña. Además, en el escorpión *Tityus serrulatus*, se reportan 16 secuencias de astacinas siendo las peptidasas más abundantes encontradas en la glándula digestiva (Fuzita *et al.*, 2015b).

En 2016, explorando más a fondo a las proteasas digestivas en arañas, Fuzita y colaboradores, publicaron el primer reporte de un gran número de astacinas que son empleadas en la digestión de un metazoario. En el caso del estudio, se analizaron astacinas de la araña *N. cruentata*, se encontraron 46 genes diferentes (parálogos) cuyas secuencias tienen al menos 5 % de diferencia entre ellas (a nivel de aminoácidos) y son consideradas isoformas. Siete secuencias que tuvieron menos del 5 % de diferencias fueron consideradas como variantes de un mismo locus. Dicho estudio demuestra la dependencia de las astacinas en el proceso de digestión extraoral ya que en el fluido digestivo constituyen hasta el 56 % de las enzimas digestivas, además de ser las hidrolasas más abundantes en los divertículos del intestino medio.

Considerando que las muestras de donde se obtuvieron las secuencias que analizamos en este trabajo, no provienen exclusivamente de muestras de hepatopáncreas, el cual tiene básicamente una función digestiva, nuestros resultados no permiten confirmar definitivamente que las secuencias analizadas tienen una función digestiva. Sin embargo, en otros trabajos se encuentran datos que apoyan la hipótesis de función digestiva, por ejemplo, tanto Hernández-Cortés y cols., (2017) y Rojo y cols., (2020) reportan un aumento en la expresión del gen *Mpc1* durante las etapas tardías del desarrollo ontogénico de *P. vannamei*, específicamente en etapas postlarva, en las que el tracto digestivo termina de madurar, los autores también reportan que el aumento de expresión genética está relacionado a un aumento de actividad proteolítica específica y detección de una banda de actividad encontrada por zimografía, cuya secuencia determinada por espectrometría de masas corresponde al transcrito de *Mpc1* (Hernández-Cortés *et al.*, 2017).

A pesar de identificar 16 variantes de secuencias únicas de astacinas que contienen las regiones características propias de esta enzima, como el motivo y el giro de metionina, solamente seis están completas, es decir contienen la región péptido señal y pro-péptido, el resto no contiene esta señalización, lo que se podría deber al método de secuenciación o a la traducción de nucleótido a proteína.

Seis secuencias maduras fueron seleccionadas para realizar los análisis subsecuentes (Fig. 13). Estas secuencias tienen un promedio de 198 aminoácidos en la región del dominio, lo que concuerda con el número de aminoácidos de la astacina de *A. astacus* que es de 200. La astacina es sintetizada como una pre-pro-enzima. Cuando la enzima se secreta a los túbulos del hepatopáncreas, se escinde el péptido señal (15 aminoácidos) del extremo N-terminal para dar lugar a la pro-enzima o zimógeno. El zimógeno se activa después de la remoción del pro-pépéptido, que consiste en 34 residuos en el N-terminal para dar lugar a la enzima madura (Dumermuth *et al.*, 1991; Stöcker *et al.*, 1993; Yiallourous *et al.*, 2002).

El peso molecular promedio de las enzimas maduras es 22 kDa, similar a la astacina de *A. astacus*, mientras que el punto isoeléctrico (pI) teórico promedio es de 4.8 para las secuencias de *P. vannamei* y de 4.70 para *A. astacus*, aunque las variantes de *P. vannamei* tienen un rango de 4.39 a 5.25. Sin embargo, hay que considerar la localización en la célula de las proteínas

analizadas y su ambiente. Las astacinas se localizan en la matriz extracelular, donde son secretadas o están ancladas a la membrana y contribuyen en la proteólisis extracelular.

Otro dato interesante por mencionar es el perfil de uso de aminoácidos de un gen y los perfiles de uso de codones. Di Giulio (2005) propone que el número de codones que codifican los aminoácidos en el código genético, son positivamente correlacionados (teniendo presente el peso molecular) con el punto isoeléctrico o dicho de otra manera, que en promedio, los aminoácidos que son codificado con un mayor número de codones son los que poseen características básicas además observa que el ambiente en el que la proteína es funcional es importante ya que los organismos que tengan un medio intracelular ácido no vivirán en ambiente alcalinos o neutros y es por esto que evolutivamente los organismos no desarrollan medios intracelulares ácidos si viven en ambientes alcalinos, al contrario, los organismos que se desarrollan en ambientes ácidos, el pH del ambiente extracelular debió haber sido ácido, por lo que el uso de perfil de aminoácidos de un gen sería útil para ver el reflejo de donde ha evolucionado para ver la forma en la que ocurrió la transferencia horizontal o vertical de genes a través de esos ambientes (Di Giulio, 2005; Goodazi *et al.*, 2008). Usando esta información, entonces se puede sugerir que las secuencias de astacina monodominio de *P. vannamei* se desarrollan en ambientes ácidos.

Como se observa en la Tabla 5 que indica la composición aminoacídica de las enzimas maduras de astacina de *P. vannamei*, las secuencias son ricas en ácido aspártico (D), glicina (G), isoleucina (I) y tirosina (Y) y pobres en cisteína (C) fenilalanina (F), histidina (H), lisina (K), metionina (M), prolina (P), arginina (R) y triptófano (W).

Los resultados de variabilidad de las secuencias monodominio maduras de astacina en *P. vannamei*, de acuerdo con el índice de Wu-Kabat con base al alineamiento múltiple se muestran en la Fig. 15. El valor mínimo de W es 1 y el máximo es 17. Las regiones con una mayor variabilidad concuerdan con las regiones no conservadas en el alineamiento (Fig. 16). Los fragmentos conservados se indican en la Tabla 6, el fragmento 4 contiene la región motivo que se interrumpe del motivo extendido solamente por un aminoácido variable. El giro de metionina también se mantiene conservado, corresponde al fragmento 7 (Tabla 6). Lo que muestran estos resultados es que hay un gran número de sitios conservados en la proteína madura, la mayor variabilidad se encuentra en la región de péptido señal y propéptido.

Fuzita y colaboradores en 2016, consideran que, para las astacinas de arácnidos, si la diferencia entre secuencias de aminoácidos es mayor al 5 %, pueden ser consideradas isoformas, mientras que por debajo de este valor pueden ser consideradas variantes de una misma isoforma. Considerando que las seis secuencias completas de astacinas de *P. vannamei* tienen una variabilidad mayor al 5 %, se puede sugerir que son isoformas, de acuerdo con el criterio de empleado por Fuzita *et al.*, (2016) en otras astacinas.

Como ya se mencionó al principio de esta discusión, las astacinas (familia M12) son el segundo grupo de proteasas más abundante en los seres vivos según la base de datos MEROPS. Este dato no es trivial porque sugiere que los organismos se adaptan a funciones y ambientes diferentes. Es evidente que en el árbol filogenético las astacinas del orden Decápoda se agrupan en 4 ramas claramente diferenciadas (Fig. 16), y dicha diferenciación no obedece necesariamente a la filogenia de Decápoda sino a las características de las secuencias.

En la Fig. 17, el grupo A (azul) tiene solo dos secuencias representantes que corresponden al género *Penaeus*. Al analizar estas secuencias, vemos que el motivo de la secuencia (HEFMHAAGFWHEQSRDYDR) es conservado, esto se puede deber al número de secuencias en comparación con los demás grupos que tienen más representantes.

El grupo B (rojo) tiene cinco secuencias representantes, de las cuales cuatro son del género *Penaeus* y uno es del género *Homarus*. La región motivo de estas secuencias es HELMHAAGFWHEQSRXDR y el aminoácido variable (X) puede ser F, S o P.

El grupo C (verde) tiene 21 secuencias de aminoácidos que corresponden a organismos de los géneros *Penaeus*, *Homarus*, *Procambarus*, *Portunus*, y *Chionoecetes*. Al tener un mayor grupo de secuencias (que el grupo A y B) de diferentes organismos y de distintos géneros, la región del motivo es aún más variable. La secuencia conservada del motivo es HEXMHXXGFWHEXSRXDR.

H	E	X	M	H	X	X	G	F	W	H	E	X	S	R	X	D	R
M					V	V						Q			F		
L					A										Y		
F					G										W		
															Q		
															A		
															S		

Figura 19. Regiones variables (en color rojo) de la secuencia del motivo de las secuencias de astacinas del grupo C de la Fig. 17.

El grupo D (morado), es el grupo con el mayor número de secuencias, distribuidas en los géneros *Penaeus*, *Procambarus*, *Portunus*, *Chionoecetes*, *Paralithodes* y *Astacus*. La secuencia conservada del motivo es HELMHXXGFXHEXXRXDR y a pesar de tener el mayor número de secuencias, se nota que esta región es menos variable que las secuencias del grupo C.

H	E	L	M	H	X	X	G	F	X	H	E	X	X	R	X	D	R
					A	V			F			Q	T		N		
					T	I			Y			H	C		K		
					L							N			M		

Figura 20. Regiones variables (en color rojo) de la secuencia del motivo de las secuencias de astacinas del grupo D de la Fig. 17.

Los miembros de Decápoda son los más diversos del taxón crustácea por lo que poseen una amplia variedad de hábitos alimenticios como carnívoros, herbívoros, detritívoros, omnívoros (Saborowski, 2015) e incluso canibalismo. La gran variedad de macronutrientes que los decápodos pueden digerir sugiere que las enzimas digestivas, encargadas de mediar y limitar la absorción de los nutrientes y la energía de los alimentos, puedan ser resultado de la adaptabilidad de los genes que los codifican en los crustáceos.

Las especies analizadas en la (Fig. 17) son organismos que ocupan diferentes hábitats por lo que su alimentación varía, además de las variaciones alimenticias relacionadas a la ontogenia. La langosta *Homarus americanus* se alimenta de moluscos, equinodermos, poliquetos, ofiuros y cnidarios (Holthuis, 1991), mientras que el langostino *Procambarus clarki* es omnívoro alimentándose de vegetales, animales y carroña (Correia, 2003) y el langostino *Astacus astacus* se alimenta de algas, perifiton, plantas acuáticas vasculares macrófitas, invertebrados, vertebrados, peces, huevos de peces, carroña y detritos (Zimmerman, 2012). Los juveniles y

adultos del género *Penaeus*, se alimentan de fito y zooplancton bentónicos, además de gusanos, bivalvos y crustáceos, e incluso pueden ser caníbales (Holthuis, 1980). El cangrejo de la nieve (*Chionoecetes opilio*), que habita en las profundidades (4 a 1400 m) se alimentan de pequeños invertebrados de los fondos marinos, otros crustáceos, bivalvos, anélidos, ofiúridos, poliquetos, fitobentos y foraminíferos (Kon, Adachi y Susuki, 2003). El cangrejo rojo gigante (*Paralithodes camtschaticus*), otra especie que habita en las profundidades (hasta los 461 m) es un oportunista y depredador generalista. Su alimentación consiste en otros crustáceos, bivalvos, poliquetos, moluscos, equinodermos, esponjas, ofiuras, rizópodos, briozoos y algas (Jørgensen, 2013). El cangrejo gazami (*Portunus trituberculatus*) es bentopelágico y habita en hábitats menos profundo de 0 a 50 m y su alimentación consiste en mayor medida en bivalvos, crustáceos bénticos, peces y cefalópodos (Ng, 1998). La variación de la fuente de nutrientes en los estadios de vida de estos organismos podría influir en los perfiles proteolíticos de las proteasas digestivas. En el caso de las astacinas, se refleja una diversidad de las secuencias en el cladograma de la Figura 17 y sería interesante hacer un estudio donde se analice a fondo la diversidad de astacinas en los demás subórdenes de Decápoda, ya que el mejor representado es el la familia Penaeidae.

Y finalmente, la estructura del modelo tridimensional (Fig. 18) construido con Swiss-Model confirmó una alta homología con la estructura tridimensional de la astacina de *A. astacus*, esto aporta aún más evidencias para proponer que las secuencias encontradas en este trabajo son putativamente digestivas.

9. CONCLUSIÓN

En las bases de datos de aminoácidos y nucleótidos, se encuentran depositadas un gran número de secuencias con dominios de astacina que aún no han sido anotadas. En este trabajo se obtuvieron 81 secuencias de aminoácidos de astacinas en diferentes decápodos, de las cuales 60 fueron seleccionadas para análisis. El árbol filogenético las agrupa en cuatro grupos que son diferentes en cuanto a la cantidad de representantes que los conforman y la diversidad del dominio catalítico. Es posible concluir que las astacinas de la familia M12 están presentes en todos los subórdenes del orden Decápoda, ya que se identificó la presencia de secuencias homólogas en los subórdenes Pleocyemata y Dendrobranchiata, con al menos un representante en las familias Astacidae, Portunidae, Oregoniidae, Cambaridae, Nephropidae, Lithodidae y Penaeidae.

Para *Penaeus vannamei*, se obtuvieron 42 secuencias de aminoácidos de astacinas las cuales se agrupan en cuatro grupos por su arquitectura y el primero (I) presenta 1 o 2 dominios no catalíticos CUB. El grupo II contiene secuencias con 1 o 3 dominios no catalíticos ShK y este trabajo es el primer reporte de esta arquitectura en un invertebrado no-hematófago. El grupo III se caracteriza por contener a las astacinas con el dominio astacina catalítico sin otros módulos, por lo cual las denominamos *monodominio*. Y finalmente, el grupo IV, contiene secuencias donde tiene un solo dominio astacina no funcional y 5 módulos CUB y 2 EGF.

En una búsqueda de secuencias de nucleótidos homólogas a astacinas de *P. vannamei*, encontramos 153 secuencias de las cuales 69 contienen la región del motivo y el giro de metionina (Met-turn) completo. Se analizaron a detalle solamente las que secuencias monodominio, putativamente digestivas. Se encontraron 16 variantes (secuencias únicas), de las cuales solamente 6 están completas, es decir contienen la región de péptido señal y propéptido. La variabilidad de las secuencias es mayor al 5 % por lo que se sugiere considerarlas como isoformas, aunque se recomienda realizar análisis confirmatorios posteriores. Las secuencias monodominio no son exclusivas de *Penaeus vannamei* y son prevalentes en otros decápodos. El análisis bioinformático de las características bioquímicas y estructurales del conjunto de secuencias recuperadas para la realización de este trabajo puede indicar su función, la cual se propone que es putativamente digestiva.

10. LITERATURA CITADA

- Altschul, S.F., Gish, W., Miller, W., Myers, E.W. y Lipman, D.J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, 215(3): 403-410. doi: 10.1016/S0022-2836(05)80360-2.
- Arolas, J.L., Goulas, T., Cuppari, A. y Gomis-Rüth, F.X. (2018). Multiple architectures and mechanisms of latency in metallopeptidase zymogens. *Chemical Reviews*, 118(11): 5581-5597. doi: 10.1021/acs.chemrev.8b00030.
- Barrett, A.J. (1995). Enzyme Nomenclature. Recommendations 1992. *European Journal of Biochemistry*, 232: 1-6, Disponible en: <http://dx.doi.org/10.1111/j.1432-1033.1995.tb20774.x>.
- Barrett, A.J., Rawlings, N.D., Salvesen, G. y Woessner, J.F. (2013). Introduction. En: Rawlings N.D., Salvesen, G. (Eds) *Handbook of proteolytic enzymes*, 3.^a ed. Elsevier Ltd., Amsterdam, pp li-liv.
- Blum, M., Chang, H.Y., Chuguransky, S., Grego, T., Kandasamy, S., Mitchell, A., Nuka, G., Paysan-Lafosse, T., Qureshi, M., Raj, S., Richardson, L., Salazar, G.A., Williams, L., Bork, P., Bridge, A., Gough, J., Haft, D.H., Letunic, I., Marchler-Bauer, A., Mi, H., Natale, D.A., Necci, M., Orengo, C.A., Pandurangan, A.P., Rivoire, C., Sigrist, C.J.A., Sillitoe, I., Thanki, N., Thomas, P.D., Tosatto, S.C.E., Wu, C.H., Bateman, A. y Finn, R.D. (2021). The InterPro protein families and domains database: 20 years on. *Nucleic Acids Research*, 49(D1): D344-D354. doi: 10.1093/nar/gkaa977.
- Bode, W., Gomis-Rüth, F.X., Huber, R., Zwilling, R. y Stöcker, W. (1992). Structure of astacin and implications for activation of astacins and zinc-ligation of collagenases. *Nature*, 359: 710-713. doi: <https://doi.org/10.1038/358164a0>.
- Bode, W., Gomis-Rüth, F.X. y Stöckler, W. (1993). Astacins, serralysins, snake venom and matrix metalloproteinases exhibit identical zinc-binding environments (HEXXHXXGXXH and Met-turn) and topologies and should be grouped into a common family, the «metzincins». *FEBS Letters*, 331(1-2): 134-140. doi: 10.1016/0014-5793(93)80312-I.
- Bond, J.S. (2019). Proteases: History, discovery, and roles in health and disease. *Journal of Biological Chemistry*, 294(5): 1643-1651. doi: 10.1074/jbc.TM118.004156.
- Bond, J.S. y Beynon, R.J. (1995). The astacin family of metalloproteinases. *Biomeditsinskaya Khimiya*, 54(5): 531-554.
- Bork, P. y Beckmann, G. (1993). The CUB domain: a widespread module in developmentally regulated proteins. *Journal of Molecular Biology*, 231(2): 539-545. doi: <https://doi.org/10.1006/jmbi.1993.1305>.

- Bowman, T.E. y Abele, L.G. (1982). Classification of the recent crustacea. En: Abele L.G. (Ed). The Biology of Crustacea: Volume 1: Systematics, The Fossil Record, And Biogeography. Academic Press, Inc., United States of America, pp 1-27.
- Bracken, H.D., Toon, A., Felder, D.L., Martin, J.W., Finley, M., Rasmussen, J., Palero, F. y Crandall, K.A. (2009). The decapod tree of life: compiling the data and moving toward a consensus of decapod evolution. *Arthropod Systematics and Phylogeny*, 67(1): 99-116.
- Ceccaldi, H.J. (1989). Anatomy and physiology of digestive tract of Crustaceans Decapods reared in aquaculture. En: *Actes de colloques Ifremer*. Tahiti, French Polynesia, pp 243-259.
- Cerdà-Costa, N. y Gomis-Rüth, F.X. (2014). Architecture and function of metallopeptidase catalytic domains. *Protein Science*, 23(2): 123-144. doi: 10.1002/pro.2400.
- Correia, A.M. (2003). Food choice by the introduced crayfish *Procambarus clarkii*. *Annales Zoologici Fennici*, 40(6): 517-528.
- De Grave, S., Pentcheff, N.D., Ah Yong, S.T., Chan, T.Y., Crandall, K.A., Dworschak, P.C., Felder, D.L., Feldmann, R.M., Franssen, C.H.J.M., Goulding, L.Y.D., Lemaitre, R., Low, M.E.Y., Martin, J.W., Ng, P.K.L., Schweitzer, C.E., Tan, S.H., Tshudy, D. y Wetzer, R. (2009). A classification of living and fossil genera of decapod crustaceans. *Raffles Bulletin of Zoology*, (21): 1-109. doi: 10.2495/ICACC130631.
- Devlin, T.M. (2019). Bioquímica: libro de texto con aplicaciones clínicas., 4.^a ed. Reverté, Barcelona. ISBN: 978-84-291-7213-3.
- Di Giulio, M. (2005). Structuring of the genetic code took place at acidic pH. *Journal of Theoretical Biology*, 237(2): 219-226. doi: 10.1016/j.jtbi.2005.04.009.
- Di Tommaso, P., Moretti, S., Xenarios, I., Orobítg, M., Montanyola, A., Chang, J.M., Taly, J.F. y Notredame, C. (2011). T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension. *Nucleic Acids Research*, 39 (2): 13-17. doi: 10.1093/nar/gkr245.
- Dumermuth, E., Sterchi, E.E., Jiang, W., Wolz, R.L., Bond, J.S., Flannery, A.V. y Beynon, R.J. (1991). The astacin family of metalloendopeptidases. *Journal of Biological Chemistry*, 266(32): 21381-21385. doi: 10.1016/s0021-9258(18)54648-2.
- FAO. (2016). Indian white prawn - Natural food and feeding habits. Disponible en: <https://www.fao.org/fishery/affris/perfiles-de-las-especies/indian-white-prawn/natural-food-and-feeding-habits/es/>. [consulta: 30 junio 2022].
- FAO. (2020). El estado mundial de la pesca y la acuicultura 2020. La sostenibilidad en acción. FAO, Roma. ISBN: 9789251327562. doi: 10.4060/ca9229es.

- Feldmann, R.M. (2003). The decapoda: new initiatives and novel approaches. *Journal of Paleontology*, 77(6): 1021-1039. doi: 10.1666/0022-3360(2003)077<1021:tdnian>2.0.co;2.
- Fernández Gimenez, A.V. (2013). Digestive physiology of three species of decapod crustaceans of Argentina. *Journal of Shellfish Research*, 32(3): 767-777. doi: 10.2983/035.032.0320.
- Fuzita, F.J., Pinkse, M.W.H., Verhaert, P.D.E.M. y Lopes, A.R. (2015)a. Cysteine cathepsins as digestive enzymes in the spider *Nephilengys cruentata*. *Insect Biochemistry and Molecular Biology*, 60: 47-58. doi: 10.1016/j.ibmb.2015.03.005.
- Fuzita, F.J., Pinkse, M.W.H., Patane, J.S.L., Juliano, M.A., Verhaert, P.D.E.M. y Lopes, A.R. (2015)b. Biochemical, transcriptomic and proteomic analyses of digestion in the scorpion *Tityus serrulatus*: insights into function and evolution of digestion in an ancient arthropod. *PLoS ONE*, 10(4): 1-31. doi: 10.1371/journal.pone.0123841.
- Fuzita, F.J., Pinkse, M.W.H., Patane, J.S.L., Verhaert, P.D.E.M. y Lopes, A.R. (2016). High throughput techniques to reveal the molecular physiology and evolution of digestion in spiders. *BMC Genomics*, 17(1): 1-19. doi: 10.1186/s12864-016-3048-9.
- Garcia-Boronat, M., Diez-Rivero, C.M., Reinherz, E.L. y Reche, P.A. (2008). PVS: a web server for protein sequence variability analysis tuned to facilitate conserved epitope discovery. *Nucleic acids research*, 36: 35-41. doi: 10.1093/nar/gkn211.
- Garcia-Carreno, F.L. y Haard, N.F. (1994). Preparation of an exopeptidase-enriched fraction from the hepatopancreas of decapods. *Process Biochemistry*, 29(8): 663-670. doi: 10.1016/0032-9592(94)80044-8.
- Geier, G., Jacob, E., Stöcker, W. y Zwilling, R. (1997). Genomic organization of the zinc-endopeptidase astacin. *Archives of Biochemistry and Biophysics*, 337(2): 300-307.
- Geier, G. y Zwilling, R. (1998). Cloning and characterization of a cDNA coding for *Astacus* embryonic astacin, a member of the astacin family of metalloproteases from the crayfish *Astacus astacus*. *European Journal of Biochemistry*, 253(3): 796-803. doi: 10.1046/j.1432-1327.1998.2530796.x.
- Gerdol, M., Cervelli, M., Mariottini, P., Oliverio, M., Dutertre, S. y Modica, M.V. (2019). A recurrent motif: diversity and evolution of ShKT domain containing proteins in the vampire snail *Cumia reticulata*. *Toxins*, 11(2): 1-18. doi: 10.3390/toxins11020106.
- Gomis-Rüth, F.X., Trillo-Muyo, S. y Stöcker, W. (2012). Functional and structural insights into astacin metallopeptidases. *Biological Chemistry*, 393(10): 1027-1041. doi: 10.1515/hsz-2012-0149.
- Goodarzi, H., Torabi, N., Najafabadi, H.S. y Archetti, M. (2008). Amino acid and codon usage profiles: adaptive changes in the frequency of amino acids and codons. *Gene*, 407(1-2): 30-41. doi: 10.1016/j.gene.2007.09.020.

- Guevara, T., Yiallourous, I., Kappelhoff, R., Bissdorf, S., Stöcker, W. y Gomis-Rüth, F.X. (2010). Proenzyme structure and activation of astacin metallopeptidase. *Journal of Biological Chemistry*, 285(18): 13958-13965. doi: 10.1074/jbc.M109.097436.
- Gutiérrez, J.M. y Rucavado, A. (2000). Snake venom metalloproteinases: their role in the pathogenesis of local tissue damage. *Biochimie*, 82(9-10): 841-850. doi: 10.1016/S0300-9084(00)01163-9.
- Hedrich, J., Lottaz, D., Meyer, K., Yiallourous, I., Jahnen-Dechent, W., Stöcker, W. y Becker-Pauly, C. (2010). Fetuin-a and cystatin C are endogenous inhibitors of human meprin metalloproteases. *Biochemistry*, 49(39): 8599-8607. doi: 10.1021/bi1004238.
- Hernandez-Cortes, P., Rivera-Pérez, C., García-Carreño, F. y Martínez-Alarcón, D. (2017). Proteinases during early development of the Pacific whiteleg shrimp *Penaeus vannamei*. *Biological Bulletin*, 232(1): 2-11. doi: 10.1086/691381.
- Hernandez-cortes, P., Whitaker, J.R. y Garcia-Carreño, F.L. (1997). Purification and characterization of chymotrypsin from *Penaeus vannamei* (Crustacea: Decapoda). *Journal of Food Biochemistry*, 21(1 12): 497-514.
- Holthuis, L.B. (1991). FAO Species Catalogue. Marine lobsters of the world. An annotated and illustrated catalogue of species of interest to fisheries known to date. *FAO Fisheries Synopsis*, Rome, (125) Vol. 13: 292.
- Holthuis, L.B. (1980). FAO Species Catalogue. Shrimps and prawns of the world. An annotated catalogue of species of interest to fisheries. *FAO Fisheries Synopsis*, Rome, 1(125): 271.
- Hooper, N.M. (1996). Zinc metalloproteases in health and disease. En: Hooper N.M. (Ed). *The biological roles of zinc and families of zinc metalloproteases*, 1.^a ed. CRC Press, Boca Raton, pp 21-42
- Hopkins, D.R., Keles, S. y Greenspan, D.S. (2007). The bone morphogenetic protein 1/tolloid-like metalloproteinases. *Bone*, 26(7): 508-523. doi:10.1016/j.matbio.2007.05.004.
- Hu, K.J. y Leung, P.C. (2007). Food digestion by cathepsin L and digestion-related rapid cell differentiation in shrimp hepatopancreas. *Comparative Biochemistry and Physiology part B: Biochemistry and Molecular Biology*, 146(1): 69-80. doi: 10.1016/j.cbpb.2006.09.010.
- Jongeneel, C.V., Bouvier, J. y Bairoch, A. (1989). A unique signature identifies a family of zinc-dependent metallopeptidases. *FEBS Letters*, 242(2): 211-214. doi: 10.1016/0014-5793(89)80471-5.
- Jørgensen, L.L. (2013). NOBANIS - invasive alien species fact sheet *Paralithodes camtschaticus*. En: Online Database European Network. Invasive Alien Species - NOBANIS. Disponible en: www.nobanis.org. [consulta: 18 octubre 2022].

- Josileen, J. y Pillai, S.L. (2013). Principles of Crustacean Taxonomy. En: Josileen J., Pillai S.L. (Eds). *Taxonomy and identification of commercially important crustaceans of India*. Central Marine Fisheries Research Institute, Kerala. pp. 28-39.
- Kanost, M.R. y Clarke, T.E. (2005). Proteases. En: Gilbert L.I.(Ed). *Comprehensive Molecular Insect Science*. Elsevier, Amsterdam, pp 247-265.
- King, R.W., Deshaies, R.J., Peters, J. y Kirschner, M.W. (1996). How proteolysis drives the cell cycle. *Science*, 274(5293): 1652-1659.
- Kon, T., Adachi, T. y Susuki, Y. (2003). Fecundity of snow crab *Chionoecetes opilio* off Wakasa Bay in the Sea of Japan. *Nippon Suisan Gakkaishi (Japanese Edition)*, 72(4): 673-680. doi: 10.2331/suisan.72.673.
- Kozlowski, L.P. (2022). Proteome-pl 2.0: proteome isoelectric point database update. *Nucleic Acids Research*, 50(D1): D1535-D1540. doi: 10.1093/nar/gkab944.
- Lian, Y.Y., He, H.H., Zhang, C.Z., Li, X.C. y Chen, Y.H. (2019). Functional characterization of a matrix metalloproteinase 2 gene in *Litopenaeus vannamei*. *Fish and Shellfish Immunology*, 84: 404-413. doi: 10.1016/j.fsi.2018.10.021.
- López-Otín, C. y Overall, C.M. (2002). Protease degradomics: a new challenge for proteomics. *Nature Reviews Molecular Cell Biology*, 3(7): 509-519. doi: 10.1038/nrm858.
- Machado de Castro, A., Fragoso dos Santos, A., Kachrimanidou, V., Koutinas, A.A. y Freire, D.M.G. (2018). Solid-state fermentation for the production of proteases and amylases and their application in nutrient medium production. En: Pandey, A., Larroche, C., Soccol, C.R. (Eds). *Current developments in biotechnology and bioengineering*. Elsevier B.V., Amsterdam, pp. 185-210.
- Martín-Galiano, A.J. y Sotillo, J. (2021). Insights into the functional expansion of the astacin peptidase family in parasitic helminths. *International Journal for Parasitology*, 52(4): 243-251. doi: 10.1016/j.ijpara.2021.09.001.
- McGaw, I.J. y Curtis, D.L. (2013). A review of gastric processing in decapod crustaceans. *Journal of Comparative Physiology B: Biochemical, Systemic, and Environmental Physiology*, 183(4): 443-465. doi: 10.1007/s00360-012-0730-3.
- McLaughlin, P.A., Camp, D.K., Angel, M.V., Bousfield, E.L., Brunel, P., Brusca, R.C., Cadien, D., Cohen, A.C., Conlan, K., Eldredge, L.G., Felder, D.L., Goy, J.W., Haney, T., Hann, B., Heard, R.W., Hendrycks, E.A., Hobbs III, H.H., Holsinger, J.R., Kensley, B., Laubitz, D.R., LeCroy, S.E., Lamaitre, R., Maddocks, R.F., Martin, J.W., Mikkelsen, P., Nelson, E., Newman, W.A., Overstreet, R.M., Poly, W.J., Price, W.W., Reid, J.W., Robertson, A., Rogers, D.C., Ross, A., Schotte, M., Schram, F.R., Shih, C.T., Watling, L., Wilson, G.D.F. y Turgeon, D.D. (2005). Common and scientific names of aquatic invertebrates from the United States and Canada: mollusks, 2nd edition. American Fisheries Society, Bethesda, Maryland.

- Mente, E. (2008). Reproductive biology of crustaceans: case studies of decapod crustaceans. Science Publishers, Enfield, N.H., USA. ISBN: 0367452774 9780367452773.
- MEROPS. MEROPS - The Peptidase Database. Disponible en: https://www.ebi.ac.uk/merops/cgi-bin/family_index?type=P#M. [consulta: 1 julio 2022].
- Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., Finn, R.D. y Bateman, A. (2021). Pfam: the protein families database in 2021. *Nucleic Acids Research*, 49(D1): D412-D419. doi: 10.1093/nar/gkaa913.
- Möhrlen, F., Maniura, M., Plickert, G., Frohme, M. y Frank, U. (2006). Evolution of astacin-like metalloproteases in animals and their function in development. *Evolution and Development*, 8(2): 223-231. doi: 10.1111/j.1525-142X.2006.00092.x.
- Nagase, H., Visse, R. y Murphy, G. (2006). Structure and function of matrix metalloproteinases and TIMPs. *Cardiovascular Research*, 69(3): 562-573. doi: 10.1016/j.cardiores.2005.12.002.
- Ng, P.K.L. (1998). Crabs. En: Carpenter K.E., Niem V.H. (Eds). *FAO species identification guide for fishery purposes. The living marine resources of the Western Central Pacific. Volume 2. Cephalopods, crustaceans, holothurians and sharks*. FAO, Rome. pp. 1045-1155.
- OEC. (2020). *Crustaceans*. OEC - The Observatory of Economic Complexity. Disponible en: <https://oec.world/en/profile/hs/crustaceans>. [consulta: 29 junio 2022].
- Patel, S. (2017). A critical review on serine protease: key immune manipulator and pathology mediator. *Allergologia et Immunopathologia*, 45(6): 579-591. doi: 10.1016/j.aller.2016.10.011.
- Pruitt, K.D., Tatusova, T. y Maglott, D.R. (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research*, 35(1): 61-65. doi: 10.1093/nar/gkl842.
- Ramos, O.H.P. y Selistre-de-Araujo, H.S. (2001). Identification of metalloprotease gene families in sugarcane. *Genetics and Molecular Biology*, 24(1-4): 285-290. doi: 10.1590/s1415-47572001000100037.
- Rawlings, N.D. y Barrett, A.J. (2013). Introduction: metalloproteases and their clans. En: Rawlings, N.D., Salvesen, G. (Eds). *Handbook of Proteolytic Enzymes*, 3.^a ed. Elsevier Ltd, Amsterdam, pp 325-370.
- Rawlings, N.D., Barrett, A.J. y Bateman, A. (2012). MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research*, 40(D1): 343-350. doi: 10.1093/nar/gkr987.

- Rawlings, N.D., Barrett, A.J., Thomas, P.D., Huang, X., Bateman, A. y Finn, R.D. (2018). The MEROPS database of proteolytic enzymes, their substrates and inhibitors in 2017 and a comparison with peptidases in the PANTHER database. *Nucleic Acids Research*, 46(D1): D624-D632. doi: 10.1093/nar/gkx1134.
- Rivera, S., Khrestchatsky, M., Kaczmarek, L., Rosenberg, G.A. y Jaworski, D.M. (2010). Metzincin proteases and their inhibitors: foes or friends in nervous system physiology? *Journal of Neuroscience*, 30(46): 15337-15357. doi: 10.1523/JNEUROSCI.3467-10.2010.
- Rojo, L., Muhlia-Almazan, A., Saborowski, R. y García-Carreño, F. (2010). Aspartic cathepsin D endopeptidase contributes to extracellular digestion in clawed lobsters *Homarus americanus* and *Homarus gammarus*. *Marine Biotechnology*, 12(6): 696-707. doi: 10.1007/s10126-010-9257-3.
- Rojo-Arreola, L., Navarrete del Toro, M.A., Cordova-Murueta, J. y García-Carreño, F. (2019). Techniques for protein digestion research in decapoda: a review. *Trends in Food Science and Technology*, 89: 65-75. doi: 10.1016/j.tifs.2019.05.004.
- Saborowski, R. (2015). Nutrition and digestion. En: Chang E.S., Thiel, M. (Eds). *The Natural History of the Crustacea, Volumen 4: Physiology*. Oxford University Press, New York, USA, p 528.
- Sainz Hernández, J.C. y Cordova Murueta, J.H. (2009). Activity of trypsin from *Litopenaeus vannamei*. *Aquaculture*, 290(3-4): 190-195. doi: 10.1016/j.aquaculture.2009.02.034.
- Sánchez-Paz, A., García-Carreño, F., Hernández-López, J., Muhlia-Almazán, A. y Yepiz-Plascencia, G. (2007). Effect of short-term starvation on hepatopancreas and plasma energy reserves of the Pacific white shrimp (*Litopenaeus vannamei*). *Journal of Experimental Marine Biology and Ecology*, 340(2): 184-193. doi: 10.1016/j.jembe.2006.09.006
- Sánchez-Paz, A., García-Carreño, F., Muhlia-Almazán, A., Peregrino-Uriarte, A.B., Hernández-López, J. y Yepiz-Plascencia, G. (2006). Usage of energy reserves in crustaceans during starvation: status and future directions. *Insect Biochemistry and Molecular Biology*, 36 (4): 241-249. doi: 10.1016/j.ibmb.2006.01.002.
- Sehnal, D., Bittrich, S., Deshpande, M., Svobodová, R., Berka, K., Bazgier, V., Velankar, S., Burley, S.K., Koča, J. y Rose, A.S. (2021). Mol*Viewer: modern web app for 3D visualization and analysis of large biomolecular structures. *Nucleic Acids Research*, 49(W1): W431-W437. doi: 10.1093/nar/gkab314.
- Sekton, B. (2010). Matrix metalloproteinases - an overview. *Research and Reports in Biology*, 1: 1-20 . doi: 10.2147/rrb.s12043.
- Spyroulias, G., Galanis, A., Pairas, G., Manessi-Zoupa, E. y Cordopatis, P. (2004). Structural features of angiotensin-I converting enzyme catalytic sites: conformational studies in

- solution, homology models and comparison with other zinc metallopeptidases. *Current Topics in Medicinal Chemistry*, 4(4): 403-429. doi: 10.2174/1568026043451294.
- Stephens, A., Rojo, L., Araujo-Bernal, S., Garcia-Carreño, F. y Muhlia-Almazan, A. (2012). Cathepsin B from the white shrimp *Litopenaeus vannamei*: cDNA sequence analysis, tissues-specific expression and biological activity. *Comparative Biochemistry and Physiology part B: Biochemistry and Molecular Biology*, 161(1): 32-40. doi: 10.1016/j.cbpb.2011.09.004.
- Stöcker, W., Gomis-Rüth, F.X., Bode, W. y Zwillig, R. (1993). Implications of the three-dimensional structure of astacin for the structure and function of the astacin family of zinc-endopeptidases. *European Journal of Biochemistry*, 214(1): 215-231. doi: 10.1111/j.1432-1033.1993.tb17915.x.
- Stöcker, W. y Yiallourous, I. (2011). Astacin. *Encyclopedia of Inorganic and Bioinorganic Chemistry*, 1-15. doi: DOI: 10.1002/9781119951438.eibc0494.
- Stöcker, W. y Yiallourous, I. (2013). Astacin. En: Rawlings N.D. (Ed). *Handbook of Proteolytic Enzymes*, 3.^a ed. Elsevier Ltd, Amsterdam, pp 895-900.
- Stöcker, W. y Zwillig, R. (1995). Astacin. *Methods in Enzymology*, 248: 305-325. doi: 10.1016/0076-6879(95)48021-8.
- Štrus, J., Žnidaršič, N., Mrak, P., Bogataj, U. y Vogt, G. (2019). Structure, function and development of the digestive system in malacostracan crustaceans and adaptation to different lifestyles. *Cell and Tissue Research*, 377(3): 415-443. doi: 10.1007/s00441-019-03056-0.
- The UniProt Consortium. (2021). UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research*, 49(1): D480-D489. doi: 10.1093/nar/gkaa1100.
- Thiel, M. y Poore, G. (eds) (2020). *Evolution and Biogeography: Volume 8*. Oxford University Press, New York. ISBN: 9780190637842. doi: 10.1093/oso/9780190637842.001.0001.
- Titani, K., Torff, H-J., Hormel, S., Kumar, S., Walsh, KA., Rödl, J., Neurath, H. y Zwillings, R. (1987). Amino acid sequence of a unique protease from the crayfish *Astacus fluviatilis*. *Biochemistry*, 26 (1): 222-226.
- Van Wormhoudt, A., Bourreau, G. y Le Moullac, G. (1995). Amylase polymorphism in crustacea decapoda: electrophoretic and immunological studies. *Biochemical Systematics and Ecology*, 23(2): 139-149. doi: 10.1016/0305-1978(94)00090-4.
- Vogt, G., Stöcker, W., Storch, V. y Zwillig, R. (1989). Biosynthesis of *Astacus* protease, a digestive enzyme from crayfish. *Histochemistry*, 91(5): 373-381. doi: 10.1007/BF00493824.

- Vogt, G. (2019). Functional cytology of the hepatopancreas of decapod crustaceans. *Journal of Morphology*, 280(9): 1405-1444. doi: 10.1002/jmor.21040.
- Vogt, G. (2021). Synthesis of digestive enzymes, food processing, and nutrient absorption in decapod crustaceans: a comparison to the mammalian model of digestion. *Zoology (Jena, Germany)*, 147: 125945. doi: 10.1016/j.zool.2021.125945.
- Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M. y Barton, G.J. (2009). Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics*, 25(9): 1189-1191. doi: 10.1093/bioinformatics/btp033.
- Watling, L. y Thiel, M. (2013). Functional morphology and diversity. Oxford University Press, New York, USA. ISBN: 9780199875450.
- Wolfe, J.M., Breinholt, J.W., Crandall, K.A., Lemmon, A.R., Lemmon, E.M., Timm, L.E., Siddall, M.E. y Bracken-Grissom, H.D. (2019). A phylogenomic framework, evolutionary timeline and genomic resources for comparative studies of decapod crustaceans. *Proceedings of the Royal Society B: Biological Sciences*, 286(1901): 1-10. doi: 10.1098/rspb.2019.0079.
- Yiallourous, I., Kappelhoff, R., Schilling, O., Wegmann, F., Helms, M.W., Auge, A., Brachtendorf, G., Berkhoff, E.G., Beermann, B., Hinz, H.J., König, S., Peter-Katalinic, J. y Stöcker, W. (2002). Activation mechanism of pro-astacin: role of the pro-peptide, tryptic and autoprolytic cleavage and importance of precise amino-terminal processing. *Journal of Molecular Biology*, 324(2): 237-246. doi: 10.1016/S0022-2836(02)01102-6.
- Yiallourous, I., Vassiliou, S., Yiotakis, A., Zwilling, R., Stöcker, W. y Dive, V. (1998). Phosphinic peptides, the first potent inhibitors of astacin, behave as extremely slow-binding inhibitors. *Biochemical Journal*, 331 (2): 375-379.
- Zhao, Z.Y., Yin, Z.X., Weng, S.P., Guan, H.J., Li, S.D., Xing, K., Chan, S.M. y He, J.G. (2007). Profiling of differentially expressed genes in hepatopancreas of white spot syndrome virus-resistant shrimp (*Litopenaeus vannamei*) by suppression subtractive hybridization. *Fish and Shellfish Immunology*, 22(5): 520-534. doi: 10.1016/j.fsi.2006.07.003.
- Zimmerman, JKM. (2012). *Noble Crayfish (Astacus astacus) in a changing world - implications for management. Tesis Doctoral*. Mid Sweden University, Department of Natural Sciences, Engineering and Mathematics, Sweden.
- Zwilling, R., Dörsam, H., Torff, H.J. y Rödl, J. (1981). Low molecular mass protease: evidence for a new family of proteolytic enzymes. *FEBS Letters*, 127(1): 75-78.
- Zwilling, R. y Neurath, H. (1981). Invertebrate proteases. *Methods in Enzymology*, 80(C): 633-664. doi: 10.1016/S0076-6879(81)80050-X