



CENTRO DE INVESTIGACIONES BIOLÓGICAS
DEL NOROESTE, S.C.

Programa de Estudios de Posgrado

IDENTIFICACIÓN DE LA BATERÍA BÁSICA DE PROTEÍNAS
INVOLUCRADAS EN LA FORMACIÓN DE LA CONCHA DE
MOLUSCOS BIVALVOS Y GASTERÓPODOS.

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

Ivis Laura Méndez Bernal

La Paz, Baja California Sur, julio de 2022.

ACTA DE LIBERACIÓN DE TESIS

En la Ciudad de La Paz, B. C. S., siendo las 12:00 horas del día 23 del Mes de junio 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:

"Identificación de la batería básica de proteínas involucradas en la formación de la concha de moluscos gasterópodos y bivalvos"

Presentada por la alumna:

Ivis Laura Méndez Bernal

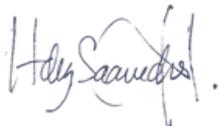
Aspirante al Grado de MAESTRO EN CIENCIAS EN EL USO, MANEJO Y PRESERVACIÓN DE LOS RECURSOS NATURALES CON ORIENTACIÓN EN **Biotecnología**

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.

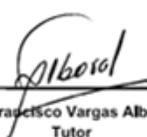
LA COMISIÓN REVISORA



Dra. Crisalejandra Rivera Pérez
Co-Directora de Tesis



Dra. Norma Yolanda Hernández Saavedra
Co-Directora de Tesis



Dr. Francisco Vargas Albores
Tutor



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

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

Los miembros del comité de tesis del (la) estudiante Ivis Laura Méndez Bernal 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:
No se aplicaron filtros
- Porcentajes de similitud:
18%
Se muestra captura de pantalla

Identificación de la batería básica de proteínas involucradas en la formación la concha de moluscos bivalvos y gasterópodos

INFORME DE ORIGINALIDAD

18%
ÍNDICE DE SIMILITUD

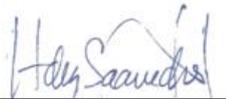
FUENTES PRIMARIAS

1	cibnor.repositorioinstitucional.mx Internet	1106 palabras — 5%
2	link.springer.com Internet	194 palabras — 1%
3	peerj.com Internet	122 palabras — 1%
4	www.cicimar.ipn.mx Internet	117 palabras — < 1%
5	hdl.handle.net Internet	112 palabras — < 1%
6	www.biorxiv.org Internet	95 palabras — < 1%
7	tel.archives-ouvertes.fr Internet	73 palabras — < 1%
8	Ignacio G. Barroso, Carla S. Santos, Mauro Bertotti, Clelia Ferreira, Walter R. Terra. "Molecular mechanisms associated with acidification and alkalization along the larval midgut of <i>Musca domestica</i> ",	66 palabras — < 1%

Firmas del comité



Dra. Crisalejandra Rivera



Dra. Norma Hernández Saavedra



Dr. Francisco Vargas

Conformación de Comités

Comité Tutorial

Dra. Crisalejandra Rivera Pérez
CONACyT - Centro de Investigaciones Biológicas del Noreste, S.C.
Co-Directora de Tesis

Dra. Norma Yolanda Hernández Saavedra
Centro de Investigaciones Biológicas del Noreste, S.C.
Co-Directora de Tesis

Dr. Francisco Vargas Albores
Centro de Investigación en Alimentación y Desarrollo, A.C.
Co-Tutor de Tesis

Comisión Revisora de Tesis

Dra. Crisalejandra Rivera Pérez
Dra. Norma Yolanda Hernández Saavedra
Dr. Francisco Vargas Albores

Jurado de Examen

Dra. Crisalejandra Rivera Pérez
Dra. Norma Yolanda Hernández Saavedra
Dr. Francisco Vargas Albores

Suplente

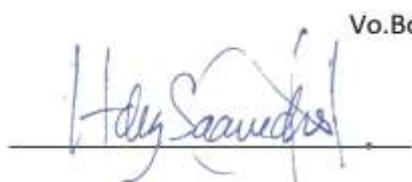
Dra. Liliana Carolina Rojo Arreola

Resumen

Los moluscos poseen una concha calcárea sólida con excepcionales propiedades ópticas, mecánicas, magnéticas, osteogénicas y osteoinductivas, muy atractivas para ciencias como la biomedicina. La biomineralización de estos organismos es un proceso de deposición de materia inorgánica controlado por una compleja matriz orgánica (proteínas, quitina y polisacáridos). Las proteínas de la matriz son los principales reguladores de este proceso (nucleación, crecimiento, polimorfismos) y hasta la fecha, no se ha descrito una batería básica de proteínas encargada de la biomineralización de los moluscos. Sin embargo, la presencia de diferentes microestructuras en las capas nacarada y prismática en las clases de moluscos más estudiadas (por ejemplo: nácar columnar en gasterópodos y en forma de placas en bivalvos), sugiere la presencia de una batería básica taxón-específica. El objetivo de este trabajo es identificar la batería básica de proteínas que intervienen en la formación de concha de los moluscos, bivalvos y gasterópodos. Para ello, se creó una base de datos con proteínas de la matriz orgánica de diferentes especies de bivalvos y gasterópodos, las cuales fueron caracterizadas mediante herramientas bioinformáticas y agrupadas utilizando diagramas de Venn y diagramas de cuerdas. Así mismo, se hicieron cladogramas con las proteínas con mayor representatividad de las bases de datos. Las bases de datos fueron conformadas por cerca de 1,500 proteínas, de 32 especies de gasterópodos y 37 de bivalvos, de las cuales numerosas proteínas presentaron modificaciones postraduccionales; sugiriendo que las proteínas son extracelulares y migran al sitio de formación de la concha. No se logró esclarecer una posible batería básica de proteínas de la concha para moluscos, sin embargo, se encontró que existen dominios comunes contenidos en las secuencias de las proteínas de la concha de bivalvos y gasterópodos, siendo los dominios más representativos C1q, *Carbonic anhydrase*, ChtBD2, EGF, KU y vWA, además del motivo *Coiled coil region* y Efh y las regiones de baja complejidad (LCR) y las regiones transmembranales (TR). No se determinó la relación entre las proteínas de la concha con la microestructura de su concha debido a la falta de información existente.

Palabras claves: bioinformática, bivalvos, concha, gasterópodos, matriz orgánica, moluscos, proteína.

ORCID: 0000-0001-8553-1799



Dra. Norma Yolanda Hernández Saavedra

Vo.Bo. Co-Director de Tesis



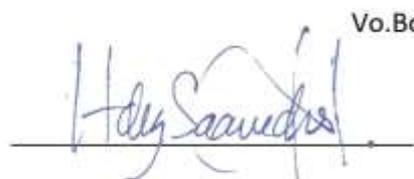
Dr. Crisalejandra Rivera Pérez

Summary

Mollusks have a solid calcareous shell with exceptional optical, mechanical, magnetic, osteogenic, and osteoinductive properties that are attractive for sciences such as biomedicine. The biomineralization of these organisms is a deposition process of inorganic matter controlled by a complex organic matrix (proteins, chitin, and polysaccharides). The matrix proteins are the main regulators of the process (nucleation, growth, polymorphisms) and to date, a basic toolkit of proteins responsible for the biomineralization of mollusks have not been described. However, the presence of different microstructures in the layers: nacreous and prismatic in the most studied classes of mollusks (for example columnar nacre in gastropods and plate-shaped in bivalves), suggests the presence of a basic taxon-specific toolkit. The objective of this work is to identify the basic toolkit of proteins that mediates the formation of the shell of mollusks, bivalves, and gastropods. To do this, a database with organic matrix proteins from different species of bivalves and gastropods was created, which were characterized using bioinformatics tools and grouped using Venn diagrams and string diagrams. Also, cladograms were performed with the most representative proteins from the database. The databases were made up of nearly 1,500 proteins, from 32 species of gastropods and 37 of bivalves, of which many presented posttranslational modifications; suggesting that proteins are extracellular and migrate to the site of shell formation. It was not possible to determine a possible protein basic toolkit of shell matrix proteins for mollusks, however, it was found common domains in the shell matrix proteins from bivalves and gastropods, being the most representative domains: C1q, Carbonic anhydrase, ChtBD2, EGF, KU, and vWA, also the motif Coiled-coil region and Efh, and the low complexity regions (LCR) and transmembranal regions (TR). No relationship was determined between the shell matrix proteins and the microstructure of their shell due to the lack of information.

Keywords: bioinformatics, bivalves, shell, gastropods, organic matrix, mollusk protein.

ORCID: 0000-0001-8553-1799



Dra. Norma Yolanda Hernández Saavedra

Vo.Bo. Co-Director de Tesis



Dr. Crisalejandra Rivera Pérez

Dedicatoria

A mi Madre y mi Padre por su guía. A mi Esposo por su paciencia.

Agradecimientos

- A mi Comité Tutorial, por sus acertados consejos y paciencia infinita.
- A CIBNOR por brindarme la oportunidad de estudiar en sus prestigiosas instalaciones.
- A CONACyT por la beca otorgada (número de becario: 1011770).
- A mis compañeros de clase: por su cordial acogida, especialmente a: Brianna, Daniela y mi cubanita Dailenys, por sus ocurrencias y su cariño desmedido.
- A Ana, Daylí, Leo, Enmailyn, Anabel y su familia y a todos los cubanos que conocí en La Paz, por siempre hacerme sentir como en casa. A Lupita y Juan, gracias a ustedes sobreviví a la pandemia y a mucho más.
- A México, que me robó el corazón.

Contenido

Resumen	i
Summary	ii
Dedicatoria	iii
Agradecimientos	iv
Contenido	v
Lista de figuras	vii
Lista de tablas	x
Abreviaturas	xi
1. INTRODUCCIÓN	1
2. ANTECEDENTES	6
2.1 Mecanismo de formación de la concha de los moluscos	6
2.2 Proteínas involucradas en la formación de la concha	11
2.3 SMPs de gasterópodos y bivalvos	13
3. JUSTIFICACIÓN	21
4. HIPÓTESIS	23
5. OBJETIVOS	24
5.1 Objetivo general	24
5.2 Objetivos particulares.....	24
6. MATERIAL Y MÉTODOS	25
6.1 Construcción de una base de datos de secuencias proteicas de gasterópodos y bivalvos	25
6.2 Determinación de la batería básica de proteínas.....	26
6.2.1 Caracterización de las secuencias de las bases de datos.....	26
6.2.2 Identificación de proteínas comunes entre bivalvos y gasterópodos.....	26
6.3 Caracterización de las proteínas y su relación con la microestructura de la concha.....	28
7. RESULTADOS	29
7.1 Construcción de una base de datos de secuencias proteicas de gasterópodos y bivalvos	29
7.1.1 Gasterópodos.....	30
7.1.2 Bivalvos	31
7.2 Determinación de la batería básica de proteínas.....	32
7.2.1 Caracterización de las secuencias de las bases de datos.....	32
7.2.2 Identificación de proteínas comunes entre bivalvos y gasterópodos.....	38
7.2.2.1 Diagramas de Venn para visualizar relaciones entre dominios/regiones y especies	38
7.2.2.2 Diagramas de cuerdas para análisis cualitativo y cuantitativo	41
7.2.2.3 Cladogramas.....	45
7.3 Caracterización de las proteínas y su relación con la microestructura de la concha.....	51
7.3.1 Propuesta de batería básica de bivalvos y descripción de los dominios constituyentes.....	51
7.3.2 Microestructura y batería básica	55
8. DISCUSIÓN	56
8.1 Construcción de una base de datos de secuencias proteicas de bivalvos y gasterópodos	56
8.2 Determinación de las proteínas de la batería básica de moluscos, gasterópodos y bivalvos y su relación con la microestructura de la concha	57

8.2.1 Caracterización de las secuencias de la base de datos.....	57
8.2.2 Identificación de proteínas comunes entre bivalvos y gasterópodos	61
8.3 Caracterización de la batería básica de proteínas que interviene en la formación de la concha de moluscos, gasterópodos y bivalvos y relación con su microestructura	62
9. CONCLUSIONES	66
10. LITERATURA CITADA	67
11. ANEXOS	76
Anexo A: Tabla VI. SMPs de gasterópodos y bivalvos, cuyas secuencias no se encuentran disponibles en el momento de la creación de las bases de datos y por tanto no se incluyeron en los análisis bioinformáticos.	76

Lista de figuras

- Figura 1.** Mecanismo hipotético de biomineralización de la capa nacarada de moluscos. A: Ensamblaje de la matriz orgánica; la matriz orgánica se ensambla en el sitio de mineralización (espacio extrapaleal) originando la matriz interlamelar (proteínas hidrofóbicas + quitina) y matriz intratabular (proteínas hidrofílicas). B: Formación primaria de los minerales; el carbonato de calcio amorfo es transportado hasta el sitio de mineralización donde se desestabiliza para su posterior cristalización. C: Nucleación de las tabletas de nácar; ocurre la nucleación de las tabletas de nácar a partir de un punto central o matriz intratabular (proteínas hidrofílicas). D: Crecimiento de las tabletas de nácar; las tabletas de nácar crecen guiadas por interacciones hidrofóbicas que inhiben la mineralización descontrolada. En el nácar maduro, las proteínas hidrofílicas o matriz intratabular quedan atrapadas dentro de las tabletas individuales (Addadi et al., 2006). Autora: Ivis Laura Méndez Bernal. 11
- Figura 2.** Caracterización bioinformática de las proteínas de las bases de datos de bivalvos y gasterópodos. Resumen de la cantidad de secuencias con posibles péptidos señales, O- y N-glicosilaciones y puentes disulfuro detectadas a través de SignalP versión 5.0 (Almagro Armenteros et al., 2019), NetOGlyc 4.0 Server (Steentoft et al., 2013), NetNGlyc 1.0 Server (Gupta and Brunak, 2002) y DiANNA versión 1.1 (Ferre y Clote, 2005), respectivamente. 33
- Figura 3.** Predicción de localizaciones celulares de las SMPs contenidas en la base de datos de gasterópodos y bivalvos. A: Localizaciones más comunes. B: Localizaciones menos comunes. Identificación mediante ProtComp 9.0. UM: unido a membrana (<http://www.softberry.com/berry.phtml?topic=protcompan&group=programs&subgroup=proloc>) y WoLF PSORT (Horton et al., 2007)..... 34
- Figura 4.** Distribución de los principales dominios y regiones identificadas en SMPs de las bases de datos de gasterópodos y bivalvos. LCR: Regiones de baja complejidad (Low complexity regions); TR: Regiones transmembranales (Transmembrane regions); Otros dominios identificados. Caracterización mediante herramienta SMART (<http://smart.embl-heildeberg.de>) y/o con la herramienta de predicción integrada al NCBI (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>). 35
- Figura 5.** Dominios y regiones identificadas en SMPs de gasterópodos y bivalvos. A: Dominios menos representativos. B: Regiones más representativos. ZP: Zona pellucida; WAP: Whey Acidic Protein; vWA: von Willebrand Factor A; TGFB: Transforming growth factor-beta; EGF: Epidermal growth factor; KU: Kunitz; CLECT: C-type lectin; LCR: Low complexity regions. TR: Transmembrane region. Identificación mediante herramienta SMART (<http://smart.embl-heildeberg.de>) y/o con la herramienta de predicción integrada al NCBI (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>). 36
- Figura 6.** Dominios identificados exclusivos de proteínas de gasterópodos y bivalvos, respectivamente. Identificación mediante herramienta SMART (<http://smart.embl-heildeberg.de>) y/o con la herramienta de predicción integrada al NCBI (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>). 37

- Figura 7.** Familias de proteínas identificadas en SMPs de gasterópodos y bivalvos. Identificación mediante Pfam 34.0: protein families (Sonnhammer, Eddy y Durbin, 1997)... 38
- Figura 8.** Diagrama de Venn representando dominios/regiones conservados de SMPs de la base de datos de gasterópodos. El diagrama muestra el número de especies en común donde fueran detectados dichos dominios/regiones. Los colores representan cada dominio/regiones o grupo de dominios/regiones, mientras que los números representan el número de especies en los que se detectaron dichos dominios/regiones; a su vez los números en las intersecciones representan la cantidad de especies en común en las que se presentan los dominios/regiones de la subsecuente intersección. Diagrama generado con: VennPainter versión 1.2.0 (Lin et al., 2016). 39
- Figura 9.** Diagrama de Venn anidado representando los dominios/regiones conservados de SMPs de la base de datos de bivalvos. El diagrama muestra el número de especies en común donde fueran detectados dichos dominios. A: Dominios y motivo identificados entre seis u ocho especies y su relación. B: Regiones identificadas en numerosas especies. Los colores representan cada dominio o grupo de dominios, mientras que los números representan el número de especies en los que se detectaron dichos dominios; a su vez los números en las intersecciones representan la cantidad de especies en común en las que se presentan los dominios de la subsecuente intersección. Diagrama generado con: VennPainter versión 1.2.0 (Lin et al., 2016). 40
- Figura 10.** Diagrama de cuerdas de SMPs de gasterópodos. Las líneas individuales del ideograma enlazan dominios/regiones presentes en las secuencias de SMPs de diferentes especies de gasterópodos. ZP: Zona pellucida; LCR: Low complexity regions. TR: Transmembrane region. WAP: Whey Acidic Protein. vWA: von Willebrand Factor A; KU: Kunitz; TGFB: Transforming growth factor-beta; CLECT: C-type lectin; EGF: Epidermal growth factor. La identificación de dominios/regiones se realizó mediante la herramienta SMART (<http://smart.embl-heidelberg.de>) y/o con la herramienta de predicción integrada al NCBI (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>). Diagrama generado con la librería circlize (Gu et al., 2014) en R Studio versión 4.0.1. 42
- Figura 11.** Diagrama de cuerdas de SMPs de bivalvos. Las líneas individuales del ideograma enlazan dominios/regiones/motivos estructurales presentes en diferentes SMPs de especies de bivalvos. ZP: Zona pellucida; LCR: Low complexity regions. TR: Transmembrane region. WAP: Whey Acidic Protein. vWA: von Willebrand Factor A; KU: Kunitz; TGFB: Transforming growth factor-beta; CLECT: C-type lectin; EGF: Epidermal growth factor. Identificación de dominios mediante la herramienta SMART (<http://smart.embl-heidelberg.de>) y/o con la herramienta de predicción integrada al NCBI (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>). Diagrama generado con la librería circlize (Gu et al., 2014) en R Studio versión 4.0.1. 44
- Figura 12.** Comparación estructural de SMPs que intervienen en la biomineralización de bivalvos, que poseen dominios del tipo ChtBD2 (dominio de unión a quitina). Dominios conservados identificados con SMART (<http://smart.embl-heidelberg.de>) y/o con la herramienta de predicción integrada al NCBI (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>); obsérvense los dominios ChtBD2

en las diferentes secuencias. SMP: Shell matrix protein; Collag a4: collagen alpha 4; CMP: Cartilage matrix protein; mp1: matrix protein 1; slp: sushi like protein; Cg: *Crassostrea gigas*; Pf: *Pinctada fucata*; Pmg: *P. margaritifera*; Mysyes: *Mizuhopecten yessoensis*; Myc: *Mytilus coruscus*. Figura editada con Inkscape versión 1.1.1. Autora: Ivis Laura Méndez Bernal. 47

Figura 13. Comparación de SMPs que intervienen en la biomineralización de bivalvos, que poseen dominios del tipo Carb_anhydrase (anhidrasa carbónica). A: Cladograma empleando el método de Neighbor-Joining (Saitou y Nei, 1987) con Mega versión 11 (Tamura et al., 2021), bootstrap (10,000 réplicas) (Felsenstein, 1985). Este análisis involucró 28 secuencias de aminoácidos. B: Comparación de los dominios conservados presentes en SMPs del tipo anhidrasa carbónica de bivalvos; obsérvense los dominios Carb_anhydrase en las diferentes secuencias. Cg: *Crassostrea gigas*; Cnp: *C. nippona*; Pf: *Pinctada fucata*; Pmg: *P. margaritifera*; Pmx: *P. maxima*; Pmz: *P. mazatlanica*; Pst: *Pteria sterna*; Tsq: *Triadacna squamosa*; Myc: *Mytilus coruscus*; Mysyes: *Mizuhopecten yessoensis*. Largo de secuencias no especificada para secuencias genómicas. Figura editada con Inkscape versión 1.1.1. 48

Figura 14. Comparación de SMPs con dominio del tipo Kunitz_BPTI que intervienen en la biomineralización de bivalvos. A: Cladograma empleando el método de Neighbor-Joining (Saitou y Nei, 1987) con Mega versión 11 (Tamura et al., 2021), bootstrap (10,000 réplicas) (Felsenstein, 1985). Este análisis involucró 20 secuencias de aminoácidos. B: Dominios conservados presentes en proteínas de bivalvos; obsérvense los dominios KU en las diferentes secuencias. Cg: *Crassostrea gigas*; Pf: *Pinctada fucata*; Pmg: *P. margaritifera*; Pmx: *P. maxima*; Pimb: *P. imbricata*; Myc: *Mytilus coruscus*; Mygall: *Mytilus galloprovincialis*. Solg: *Solen grandis*. Largo de secuencias no especificada para secuencias genómicas. Figura editada con Inkscape versión 1.1.1. 49

Figura 15. Comparación de SMPs con dominios del tipo vWA. A: Cladograma realizado por el método de Neighbor-Joining (Saitou y Nei, 1987) con Mega versión 11 (Tamura et al., 2021), bootstrap (10,000 réplicas) (Felsenstein, 1985). Este análisis involucró 20 secuencias de aminoácidos. Cg: *Crassostrea gigas*; Pf: *Pinctada fucata*; Pmg: *P. margaritifera*; Myc: *Mytilus coruscus*; Mygall: *Mytilus galloprovincialis*; Rudd: *Rudipates desucatus*. Figura editada con Inkscape versión 1.1.1. 50

Figura 16. Diagrama de Venn que representa las posibles baterías básicas de SMPs involucradas en la biomineralización de moluscos gasterópodos y bivalvos y la relación entre estas. Como se puede observar, no se determinaron SMPs taxa específicos para gasterópodos, sin embargo, las LCR y TR mostraron relevancia y probablemente integren la batería básica de esta clase taxonómica. La posible batería básica de SMPs responsable de la biomineralización de moluscos podría integrarse por LCR y TR. Carb_anhydrase: anhidrasa carbónica; ChtBD2: dominios de unión a quitina (del inglés: chitin binding domain 2); EFh: dominio EF-hand; EGF: factor de crecimiento epidérmico (del inglés: epidermic growth factor); KU: dominios Kunitz; LCR: Regiones de baja complejidad (del inglés: low complexity regions); TR: regiones transmembranales (del inglés: transmembrane regions); vWA: von Willebrand Factor A. 55

Lista de tablas

Tabla I. Representación de proteínas de la matriz orgánica de la concha de moluscos bivalvos y gasterópodos.....	13
Tabla II. Proteínas/dominios-estructurales comunes en la matriz orgánica de moluscos bivalvos y gasterópodos.....	19
Tabla III. Características de las bases de datos construidas de SMPs de gasterópodos y bivalvos.....	29
Tabla IV. Diversidad de especies y número de secuencias por especie en la base de datos de gasterópodos.....	30
Tabla V. Diversidad de especies y número de secuencias por especie en la base de datos de bivalvos.....	31
Tabla VI. SMPs de gasterópodos y bivalvos, cuyas secuencias no se encuentran disponibles en el momento de la creación de las bases de datos y por tanto no se incluyeron en los análisis bioinformáticos.....	76

Abreviaturas

BMPS: *bone morphogenic protein 2*

ChtBD2: dominios de unión a quitina (*chitin binding domain 2*)

Cys: Cisteína

Efh: EF-hand, motivo de unión a calcio

EGF: factor de crecimiento epidérmico (*epidermal growth factor*)

KU: dominio de la familia BPTI/Kunitz de inhibidores de serino proteasas

LCR: regiones de baja complejidad (*low complexity regions*)

SMPS: proteínas de la matriz orgánica de la concha (*shell matrix protein*)

TR: regiones transmembranales

vWA: factor de von *Willebrand* tipo A

1. INTRODUCCIÓN

La biomineralización es el proceso a través del cual los organismos producen minerales con propósitos funcionales como: soporte mecánico de tejidos, sensibilidad magnética o sensorial, y/o almacenamiento o protección (Tao, 2013). Durante este proceso biológico se combina materia orgánica e inorgánica de forma regulada en organismos vivos (Estroff, 2008), siendo una estrategia funcional la evolución biológica, que permite a los organismos adaptarse mejor al ambiente donde se desarrollan (Chen *et al.*, 2019). Los huesos y dientes de los humanos, la testa de los erizos de mar, el esqueleto de corales, la concha de los moluscos, el esqueleto de los equinodermos, etc., son ejemplos de biominerales.

Los biominerales se diferencian de los minerales geológicos o sintéticos, por sus excelentes propiedades mecánicas, ópticas, magnéticas y térmicas. Estas cualidades son el resultado de micro- y nanoestructuras multiniveles altamente ordenadas con morfologías específicas, consistentes con la macroescala (Chen *et al.*, 2019). La adquisición de esta configuración espacial depende del efecto sinérgico de las biomoléculas que intervienen en el mecanismo de síntesis del biomaterial (Marin *et al.*, 2012; Chen *et al.*, 2019).

La biomineralización ha inspirado el desarrollo de nanomateriales, que persiguen recrear y potenciar estas excepcionales propiedades. La generación de nanomateriales inspirados en este proceso se ha catapultado en los últimos años, llegando a múltiples ramas de la ciencia como: ingeniería mecánica, ingeniería eléctrica, ingeniería ambiental e ingeniería biomédica. La biomedicina lidera el desarrollo de bio-nanomateriales debido a su importancia en la salud y bienestar humanos. Los bio-nanomateriales son utilizados en: terapia celular y de drogas (donde brindan protección, almacenamiento y sistema de entrega); ingeniería de blanco cáncer/tumor (como plataformas de diagnóstico y tratamiento); ingeniería de enzimas (donde aportan estabilidad y sistema de entrega); ingeniería de impresión tridimensional (siendo biocompatibles con hueso artificial); e ingeniería regenerativa de tejido óseo (como implantes óseos de alta biocompatibilidad, alta fuerza mecánica, con propiedades osteoinductivas y osteogénicas). Estas aplicaciones permiten el tratamiento de enfermedades causadas por la mineralización anormal del hueso en humanos como: osteoporosis, osteomalacia, hipofosfatasa, cálculos renales y arteriosclerosis (Chen *et al.*,

2019). Debido a estas múltiples aplicaciones, el estudio del proceso de biomineralización de diversos organismos, representa la vanguardia en el desarrollo de nanomateriales.

La biomineralización de los moluscos es sin lugar a duda, una de las más estudiadas (Marin *et al.*, 2012). Los moluscos son los invertebrados acuáticos y terrestres más abundantes, después de los artrópodos. Se caracterizan por presentar un cuerpo blando y pie musculoso modificado para reptar, cavar o nadar; están recubiertos por una epidermis o manto que secreta una concha calcárea sólida (Brusca y Brusca, 2003). La concha se desarrolló en estos organismos como una estrategia compleja para preservar su masa visceral blanda, protegerse contra la depredación y evitar la desecación (Addadi *et al.*, 2006).

La concha de los moluscos está formada fundamentalmente por cristales de carbonato de calcio depuesto en forma de capas calcáreas, a menudo recubiertas externamente por una capa orgánica denominada periostraco (o hipostraco). Los cristales de carbonato de calcio tienen tres polimorfos típicos: aragonita, calcita y vaterita, los cuales son químicamente idénticos, sin embargo, bajo el microscopio su forma de cristalización es diferente (Brusca y Brusca, 2003). Generalmente la concha consta de dos capas calcáreas: 1) capa prismática externa (calcita y/o aragonita) y 2) capa nacarada, lamelar interna o madreperla (aragonita), esta última presente únicamente en bivalvos, gasterópodos, cefalópodos y una especie de monoplacóforos (Marin *et al.*, 2012); el nácar surgió independientemente en cada una de estas clases (Checa, 2018). A la matriz de carbonato de calcio se incorporan, además, compuestos inorgánicos como: fosfatos, sulfatos de calcio, carbonato de magnesio, sales de aluminio, hierro, cobre, estroncio, bario, silicio, manganeso, yodo y flúor; y otros compuestos raros (poco abundantes) como: pirroles o porfirinas que se asocian con la coloración (Brusca y Brusca, 2003).

La concha de los moluscos posee propiedades mecánicas y ópticas directamente asociadas a la estructura jerárquica altamente organizada de sus capas calcáreas. La capa prismática posee una resistencia moderada a la fractura, con cierta flexibilidad reforzada por acción del periostraco orgánico externo (Marin *et al.*, 2012); está compuesta por cristales de calcita estables (Suzuki *et al.*, 2009) y presumiblemente constituye una estrategia eficiente de rápida mineralización y engrosamiento de la concha (Marin *et al.*, 2012). Por otro lado, la

capa nacarada de la concha protege el cuerpo blando del molusco. Está compuesta por carbonato de calcio en forma de cristales de aragonita diminutos e indistinguibles individualmente, y termodinámicamente metaestables (Suzuki *et al.*, 2009), lo que la hace altamente resistente a la fractura. La aragonita de la capa nacarada, es mil veces más resistente a la aragonita de origen geológico (Marin *et al.*, 2012), y más resistente que el hueso, con una resistencia a la fractura de 1.47 GPa contra solo 150 MPa del hueso (Bahn *et al.*, 2015). Posee macromoléculas como el polisacárido quitina y proteínas que proveen alta fuerza mecánica y propiedades ópticas inusuales (refleja la luz y causa interferencia) que lo convierten en atractivo para la joyería e imagenología (Suzuki *et al.*, 2009; Marin *et al.*, 2012). Posee, además, propiedades osteogénicas y osteoinductivas con potencial en la regeneración ósea, probablemente como fuente de moléculas orgánicas bioactivas (Marin *et al.*, 2012). Las diferentes texturas de las capas impiden que una fractura pueda propagarse a través de toda la extensión de la concha (Marin *et al.*, 2012). Esta combinación de propiedades, hacen de la concha de los moluscos un material rígido y dúctil a la vez, sin embargo, se desconocen exactamente los desencadenantes de dichas propiedades, pero se atribuyen a la presencia de factores bioactivos difusos en la matriz de nácar (matriz orgánica).

La biomineralización de los moluscos está biológicamente controlada, lo que significa que es causada y mediada por macromoléculas, que constituyen la matriz orgánica. En aspectos de biomineralización, se define como matriz orgánica a cualquier matriz localizada en la superficie de los constituyentes, como proteínas, fosfolípidos, colágeno y/o carbohidratos, que actúa como mediadores en la mineralización de los sistemas biológicos. Numerosas matrices orgánicas compuestas fundamentalmente por proteínas involucradas en la formación de materiales inorgánicos durante la biomineralización; las proteínas actúan regulando la nucleación, crecimiento, cristalización y su direccionamiento (este proceso es conocido como reconocimiento molecular) (Chen *et al.*, 2019). La matriz orgánica de los moluscos media el proceso de formación y mineralización de la concha en general.

La matriz orgánica de la concha de los moluscos representa una fracción menor (1-5 %) de sus componentes y está constituida principalmente por quitina, y proteínas hidrofílicas e hidrofóbicas (Addadi *et al.*, 2006). Addadi *et al.* (2006) propusieron el mecanismo hipotético

de formación de la concha, definiendo el papel de cada uno de estos compuestos en el proceso de biomineralización e identificando el papel de las proteínas de la matriz tanto en la nucleación del carbonato de calcio, como en el control de la nucleación. Desde mediados de la década de los 90's, se inició la caracterización de las proteínas de la matriz de la concha de los moluscos (Berland *et al.*, 2011). Actualmente, se han descrito numerosas proteínas como: Pif97, Pif80, nacreína, perlina, tirosinasa, proteínas de unión a quitina, anhidrasas carbónicas, etc. (Suzuki *et al.*, 2009; Bahn *et al.*, 2015; Arivalagan *et al.*, 2016). Sin embargo, las proteínas de la matriz orgánica muestran numerosas diferencias intra- e interespecies, lo que fomenta el debate sobre el estado de conservación del mecanismo de biomineralización de los moluscos (Berland *et al.*, 2011; Marie *et al.*, 2013; Arivalagan *et al.*, 2016); y hasta la fecha no se ha descrito una batería básica de proteínas encargada de la biomineralización de estos organismos.

Por otro lado, las clases de moluscos presentan diferencias en la forma, disposición de sus placas (prismática, nácar lamelar, placas de nácar, lamelar cruzada, lamelar cruzada compleja, y foliada) y composición de la concha (no todos poseen nácar); el mayor número de especies con nácar se agrupan en las clases Bivalvia y Gastropoda, que son moluscos de amplia distribución y elevado número de especies. Las diferencias en la concha de estos taxa son apreciables a simple vista, pues los bivalvos ostentan una concha formada por dos valvas unidas dorsalmente a través de ligamentos, mientras que los gasterópodos presentan una concha asimétrica espiralizada (Brusca y Brusca, 2003). Estas diferencias, se extienden a la microestructura de la concha, donde las capas prismática y nacarada son polimórficas. Los bivalvos presentan casi exclusivamente un tipo de nácar conocido como “pared de ladrillos” u “hoja de nácar”, en el que las tabletas se disponen en placas escalonadas, como ladrillos en una pared. Sin embargo, los gasterópodos poseen comúnmente nácar de tipo columnar, donde las tabletas planas crecen por encima de las tabletas subyacentes, formando pilas o torres columnares. En una misma pila, las tabletas no se encuentran completamente alineadas, sino que presentan una pequeña desviación lateral que permite la interpenetración y asociación con las tabletas de las columnas vecinas. Por otro lado, la capa prismática posee una mayor variabilidad de microestructuras entre estas clases: lamelar, cruzada, homogénea (granulada), esferulítica, fibrosa, etc., con numerosos subtipos y combinaciones (Marin *et al.*, 2012).

Las proteínas de la matriz orgánica de la concha actúan regulando la nucleación de los cristales de carbonato de calcio, su crecimiento y la generación de los polimorfismos de su microestructura (Arivalagan *et al.*, 2016). Si se toma en cuenta que la presencia de una concha calcárea es un rasgo característico de los moluscos, y se analiza la diversidad de proteínas de la matriz orgánica de la concha, es probable la existencia de una batería básica de proteínas conservada encargada de la biomineralización. Bajo este mismo precepto, y observando la diversidad de microestructuras de las conchas, es posible que gasterópodos y bivalvos, presenten baterías básicas exclusivas encargadas de su biomineralización específica y relacionadas a las microestructuras características de sus conchas. De esta manera, las diferencias entre las baterías básicas de gasterópodos y bivalvos podrían explicar las diferentes microestructuras de sus capas calcáreas.

En el presente estudio se emplea un enfoque descriptivo-comparativo para identificar la batería básica de proteínas de la matriz orgánica de moluscos, gasterópodos y bivalvos, partiendo de la información disponible en bases de datos, repositorios públicos y literatura científica. Las proteínas fueron agrupadas atendiendo a sus taxa, capa de la concha encontrada y función basada en la presencia de dominios, y asignadas como “comunes”, taxa-específicas o como parte de la batería básica; o “exóticas” (especie-específicas, género-específicas, etc.).

La determinación de una batería básica de proteínas de moluscos permitirá avanzar en el conocimiento del fenómeno de la biomineralización a nivel molecular y cuáles son los componentes imprescindibles del proceso; y las proteínas diferentes entre gasterópodos y bivalvos que permitan proponer un modelo sobre la radiación evolutiva causante de la diversidad estructural y funcional de la concha de moluscos. Además, la descripción de las proteínas contribuirá al desarrollo de protocolos de biomineralización *in vitro* eficientes, capaces de recrear y potenciar las propiedades del biomineral en forma de nanomateriales, a utilizarse como implantes óseos, sistemas de diagnóstico y tratamiento de enfermedades óseas, etc., o en la joyería para la creación de perlas sintéticas reduciendo los efectos ambientales de la pesca y/o el cultivo de ostras perleras.

2. ANTECEDENTES

2.1 Mecanismo de formación de la concha de los moluscos

La matriz orgánica es una mezcla de macromoléculas extracelulares secretadas para ayudar y guiar la mineralización, o al menos, para ser usada como sustrato para la deposición del mineral (Marin *et al.*, 2012). La formación de la concha de los moluscos, en general, está regulada por la matriz orgánica constituida fundamentalmente por β -quitina, proteínas hidrofóbicas similares a la seda y proteínas ácidas hidrofílicas ricas en ácido aspártico (Addadi *et al.*, 2006).

La capa prismática es la capa calcárea más externa de la concha y está compuesta por cristales elongados perpendiculares a la superficie de la concha, que pueden ser de calcita o aragonita (Marin *et al.*, 2012); los cristales se disponen en capas organizadas y paralelas (Evans, 2019), formando en ocasiones dos capas o más capas, de ambos polimorfos (Marin *et al.*, 2012).

El mecanismo de biomineralización de los moluscos fue hipotetizado por (Addadi *et al.*, 2006) y explica la formación de la concha, a partir de su capa nacarada. La capa nacarada de la concha de los moluscos ostenta una microestructura laminar bien definida, compuesta de tabletas, de poligonales a redondeadas, dispuestas en láminas anchas, paralelas y regulares. Las tabletas son cristales de aragonita con un grosor que varía entre medio micrón y un micrón, con una extensión lateral de pocos micrones (Marin *et al.*, 2012). Los cristales interactúan con complejos orgánicos (matriz orgánica) que se disponen entre los complejos inorgánicos poligonales (o tabletas), a modo de ladrillos y cemento (Bahn *et al.*, 2015); la estructura está densamente empaquetada sin intersticios entre los bloques (Marin *et al.*, 2012).

El nácar maduro presenta dos formas de matriz orgánica, una que rodea a las tabletas de nácar (matriz interlamelar) y otra en el interior de estas (matriz intratabular). La matriz interlamelar o intercrystalina es una capa delgada (20-50 nm de ancho) no homogénea, continua, dispuesta en subcapas perpendiculares al eje *c* de las tabletas de nácar. Está compuesta de una porción central de quitina (material denso en electrones) contenida entre

dos capas de material proteínico (menos denso en electrones; compuesto por proteínas hidrofóbicas). Es la plantilla para la nucleación y el crecimiento de las tabletas de nácar. Por otro lado, la matriz intratabular o intracristalina, es una red continua, similar a espuma, que interrumpe el mineral en nanómetros (45 nm), o sea, cada tableta de nácar está a su vez compuesta por nano-tabletas o nano-bloques que exhiben una organización jerárquica. Esta matriz es rica en carboxilatos y sulfatos que teóricamente actúan como nucleadores del cristal (proteínas hidrofílicas) (Marin *et al.*, 2012).

La teoría del mesocristal intenta explicar cómo se forma el nácar (Addadi *et al.*, 2006). Un mesocristal se define como cristales coloidales construidos a partir de nanocristales individuales alineados en un registro cristalográfico común. El mesocristal corresponde con la estructura del nácar formada por dos capas orgánicas, discutida anteriormente (Figura 1) y fue descubierta por Cölfen y Antonietti, (2005). La teoría del mesocristal se conoce como cristalización no clásica, donde nanopartículas amorfas cristalinas transientes son estabilizados por polímeros orgánicos que adsorben sus superficies. Estas nanopartículas se ensamblan y co-orientan idénticamente en una superestructura denominada mesocristal, por acción de los polímeros, los cuales quedan atrapados en su interior (matriz intratabular) (Marin *et al.*, 2012). A continuación, se describe brevemente el mecanismo (Figura 1. Mecanismo hipotético de biomineralización de la capa nacarada de moluscos. A: Ensamblaje de la matriz orgánica; la matriz orgánica se ensambla en el sitio de mineralización (espacio extrapaleal) originando la matriz interlamelar (proteínas hidrofóbicas + quitina) y matriz intratabular (proteínas hidrofílicas). B: Formación primaria de los minerales; el carbonato de calcio amorfo es transportado hasta el sitio de mineralización donde se desestabiliza para su posterior cristalización. C: Nucleación de las tabletas de nácar; ocurre la nucleación de las tabletas de nácar a partir de un punto central o matriz intratabular (proteínas hidrofílicas). D: Crecimiento de las tabletas de nácar; las tabletas de nácar crecen guiadas por interacciones hidrofóbicas que inhiben la mineralización descontrolada. En el nácar maduro, las proteínas hidrofílicas o matriz intratabular quedan atrapadas dentro de las tabletas individuales (Addadi *et al.*, 2006).

- 1) Ensamblaje de la matriz orgánica: El primer paso en la síntesis de la capa nacarada es el ensamblaje de la matriz orgánica (Figura 1A); la matriz orgánica guía el proceso de mineralización.

Esta fase inicia con el autoensamblaje de la quitina en varias capas sucesivas perpendiculares al eje *c* de las futuras tabletas de nácar (Marin *et al.*, 2012). Simultáneamente al ensamblaje de las hojas de quitina, un gel hidratado pre-rellena el espacio donde se formarán las tabletas, manteniendo las placas de quitina separadas. El gel está compuesto por proteínas hidrofóbicas desordenadas y proteínas polianiónicas hidrofílicas (Marin *et al.*, 2012) (Figura 1A).

La quitina es un polisacárido resistente y flexible, que sirve como esqueleto de la matriz tridimensional (Marin *et al.*, 2012). Este polímero no modifica la nucleación o crecimiento debido a que carece de grupos funcionales con la capacidad de unión al Ca^{+2} (Bahn *et al.*, 2015), por lo que se requiere de las proteínas para la formación de las tabletas de aragonita (Marin *et al.*, 2012). La quitina se combina con las proteínas hidrofóbicas dando lugar a la matriz interlamelar que ejercerá una función de molde plantilla en la formación de las tabletas, mientras que las proteínas hidrofílicas actuarán como centro de nucleación para la síntesis de los cristales y constituyen la matriz intratabular (Marin *et al.*, 2012) (Figura 1A).

La formación de la concha estará vinculada al epitelio secretor calcificante o capa de células más externas del manto. Este epitelio es una monocapa de células con microvellosidades típicas intercaladas con células caliciformes (mucositos), que se encarga de la secreción de la matriz orgánica y de los materiales precursores del carbonato de calcio (Marin *et al.*, 2012).

El crecimiento de nácar puede ocurrir simultáneamente en varios niveles a la vez, por lo que las células del manto se encuentran secretando minerales y macromoléculas a través de las capas de quitina y lateralmente entre las capas (Addadi *et al.*, 2006).

- 2) Formación primaria de los minerales: Esta etapa se caracteriza por la formación de los primeros minerales, que son amorfos y se encuentran en la escala nanométrica (Figura 1B).

Los cristales se formarán en el espacio relleno por el gel proteico y contenido entre las capas de matriz interlamelar. La forma en la que el carbonato de calcio amorfo (ACC, del inglés *amorphous calcium carbonate*) es introducido en el gel aún se desconoce (difusión o vesículas), pero podría ocurrir a través de canales y bombas ATPasas en el epitelio calcificante; estas bombas permitirían el transporte activo de los precursores del carbonato de calcio, previamente adquirido por el molusco vía metabolismo (sistema digestivo) y/o filtración (branquias y epitelio interno del manto). Una vez en el peculiar ambiente del gel, la fase ACC transitoria puede ser desestabilizada para ser más susceptible a la cristalización. Las proteínas hidrofóbicas del gel impiden entonces la mineralización descontrolada y mantienen el contenido de agua necesario para la posterior cristalización (Marin *et al.*, 2012).

El sitio de mineralización se encuentra en la cavidad extrapaleal y se define como el espacio confinado entre el manto y el periostraco orgánico externo, y se encuentra lleno de un líquido de composición desconocida. Es en este sitio donde ocurren los procesos de ensamblaje de la matriz orgánica y la posterior formación del mineral, sirviendo el periostraco como molde primario para la deposición de ambos.

- 3) Nucleación de las tabletas de nácar: En este punto se produce la formación de las tabletas de aragonita (Figura 1C).

Se asume que cada tableta crece a partir de un punto central que contiene grupos reactivos específicos del tipo polianiones: carboxilatos y sulfatos, o sea, un punto central que contiene proteínas hidrofílicas (matriz intratabular), que controlan la nucleación y el crecimiento (sección central compuesta por carboxilatos, rodeada de un anillo de sulfatos que atraen al Ca^{+2}). Los nano-elementos se auto-ensamblan desde el centro, auto-orientados, formando el mesocristal de tabletas de nácar. Es

un proceso centrífugo y la unión de los nanoelementos ocluye una parte de la matriz (Marin *et al.*, 2012).

- 4) Crecimiento de las tabletas: Una vez formadas las tabletas de nácar, se produce el crecimiento de estas que culmina con el nácar maduro (Figura 1D).

Cada tableta crece en dirección vertical hasta alcanzar la capa de quitina superior, y luego se expande lateralmente (crecimiento centrífugo). El crecimiento lateral ocurre a expensas del gel que es desplazado. Todo el fenómeno es guiado por interacciones hidrofóbicas que inhiben la cristalización descontrolada de la calcita, favoreciendo de esta forma el crecimiento de la aragonita. Al mismo tiempo, las proteínas polianiónicas quedan atrapadas en el interior de las tabletas. Cuando las tabletas vecinas confluyen, el gel se sella, polimeriza y transforma en una matriz insoluble. Probablemente la matriz entre las tabletas se mantiene en condiciones amorfas debido a las interacciones con las impurezas expelidas durante el crecimiento lateral de las tabletas (Marin *et al.*, 2012).

Esta teoría, descrita por Addadi *et al.* (2006) es especulativa y aún no se ha podido comprobar. Además, actualmente la información disponible sobre las proteínas de la concha se ha ampliado considerablemente y se conoce que conjuntamente a la formación de los cristales mediada por matrices orgánicas, actúan otros factores físicos y biológicos como: competición de cristales (durante el crecimiento cristalino), procesos de auto-organización de las matrices orgánicas, actividad celular directa, etc. (Checa, 2018).

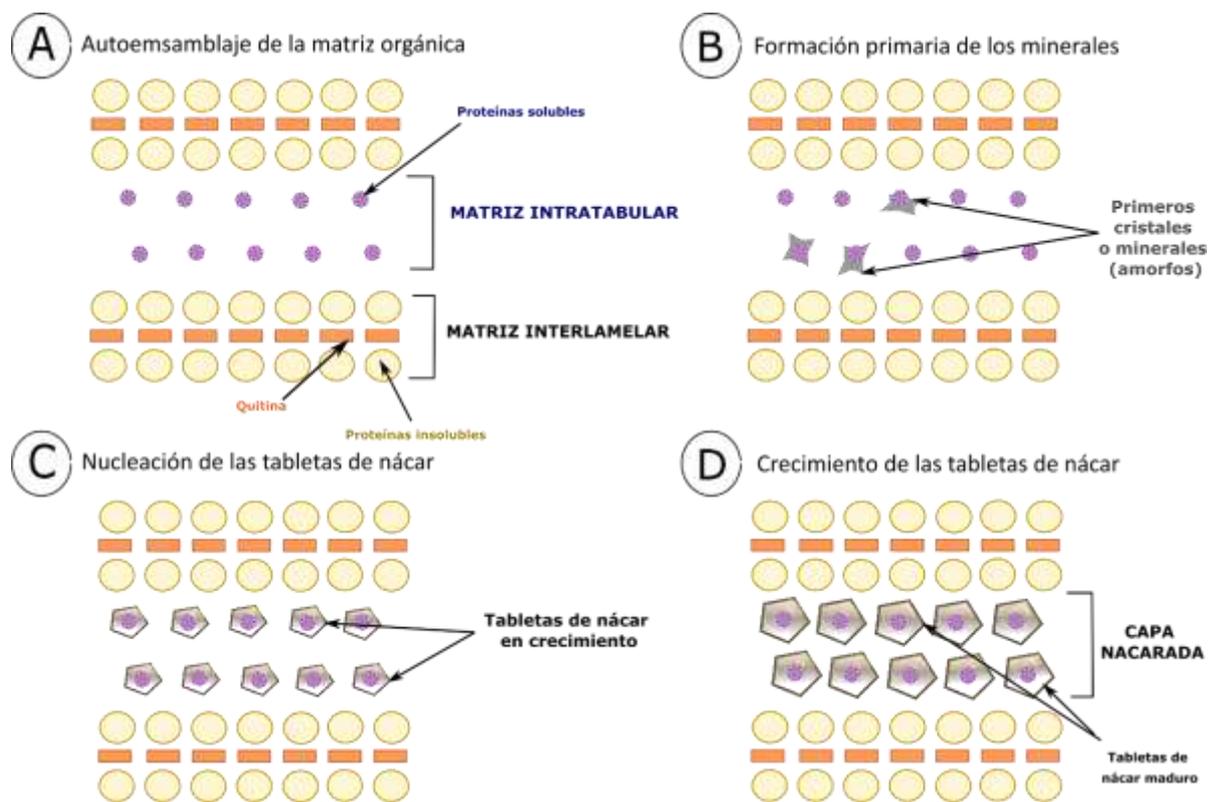


Figura 1. Mecanismo hipotético de biomineralización de la capa nacarada de moluscos. A: Ensamblaje de la matriz orgánica; la matriz orgánica se ensambla en el sitio de mineralización (espacio extrapaleal) originando la matriz interlamelar (proteínas hidrofóbicas + quitina) y matriz intratabular (proteínas hidrofílicas). B: Formación primaria de los minerales; el carbonato de calcio amorfo es transportado hasta el sitio de mineralización donde se desestabiliza para su posterior cristalización. C: Nucleación de las tabletas de nácar; ocurre la nucleación de las tabletas de nácar a partir de un punto central o matriz intratabular (proteínas hidrofílicas). D: Crecimiento de las tabletas de nácar; las tabletas de nácar crecen guiadas por interacciones hidrofóbicas que inhiben la mineralización descontrolada. En el nácar maduro, las proteínas hidrofílicas o matriz intratabular quedan atrapadas dentro de las tabletas individuales (Addadi *et al.*, 2006). Autora: Ivis Laura Méndez Bernal.

2.2 Proteínas involucradas en la formación de la concha

La matriz orgánica de la concha representa una pequeña fracción de los componentes de la concha (entre 1-5%) y está compuesta por: polisacáridos como quitina y polímeros solubles ácidos, lípidos, pigmentos, pequeños péptidos, aminoácidos libres y proteínas (Marie *et al.*, 2010; Marin *et al.*, 2012). De estos, los componentes principales son la quitina y las proteínas, y el papel en la mineralización del resto permanece poco estudiado.

Las proteínas que actúan en la formación de la concha se sintetizan en el epitelio calcificante que las excreta hacia el medio de mineralización, donde actúan regulando la deposición de

carbonato de calcio. Las proteínas de la matriz de la concha o SMPs (del inglés: *shell matrix proteins*) se clasifican en dos grupos principales, con base al tipo de solución usada para descalcificar la concha de moluscos: proteínas hidrofóbicas o insolubles y proteínas hidrofílicas o solubles. Las proteínas hidrofóbicas o proteínas semejantes a la seda se caracterizan por presentar abundantes residuos de Ala y/o Gly y dominios de tipo hoja- β y se posicionan en la matriz intralamelar estableciendo interacciones con la β -quitina y ejercen una función de molde/plantilla. Por otro lado, las proteínas hidrofílicas ácidas son polianiónicas (poseen grupos sulfatos y carbonatos, de unión a minerales) y exhiben abundantes residuos de Asp (Bahn *et al.*, 2015). Asumen conformaciones hoja- β en presencia de Ca^{+2} (Addadi *et al.*, 2006). También, presentan regiones altamente repetitivas, motivos repetidos o regiones repetidas de baja complejidad o LCRs (del inglés: *low complexity regions*) (Marie *et al.*, 2010), dominios desordenados intrínsecamente y secuencias propensas a la agregación, sin embargo, carecen de dominios comunes o similitudes en sus secuencias (Bahn *et al.*, 2015). Además de su solubilidad, pueden presentar modificaciones postraduccionales como, péptido señal, multidominios con multifunciones, o funcionan como proteínas mosaico.

En 1960 Watabe y Wilbur (1960) encontraron que la matriz orgánica de la capa nacarada induce la formación de la concha. Posteriormente, se sugirió que los residuos de Asp de la matriz orgánica, en interacción con los iones Ca^{+2} del carbonato, regulan el polimorfismo cristalino (Suzuki *et al.*, 2009). A partir de entonces, se describieron numerosas proteínas de la matriz orgánica de la concha de los moluscos, sin embargo, la mayor parte de esos estudios (~90%) se realizaron en especies modelo como *Haliotis* spp. (gasterópodos) y *Pinctada* spp. (bivalvos), no abarcando la gran variedad de microestructuras del nácar (Marin *et al.*, 2012). Adicionalmente, se puede mencionar que el modelo del mesocristal, descrito por Addadi *et al.* (2006), no considera la gran variedad de microestructuras que pueden adoptar las tabletas de aragonita.

Actualmente, los estudios sobre las proteínas involucradas en la biomineralización en moluscos se han ampliado hacia otras especies: *Pinctada margaritifera* (Bédouet *et al.*, 2007; Joubert *et al.*, 2010; Berland *et al.*, 2011); *P. mazatlanica* (Rivera-Pérez *et al.*, 2019b) *P. fucata* (Miyamoto *et al.*, 2013; Liu *et al.*, 2015); *Crassostrea gigas* (Wang *et al.*, 2013;

Arivalagan *et al.*, 2016; Feng *et al.*, 2017); *Pteria sterna* (Arroyo-Loranca *et al.*, 2020a), *Haliotis asinina* (Marie *et al.*, 2010); *H. laevigata* (Mann *et al.*, 2018); *Lottia gigantea* (Mann y Edsinger, 2014; Marie *et al.*, 2013), etc.; incluso hacia gasterópodos pulmonados como *Cepaea nemoralis* (Mann y Jackson, 2014), sin embargo, aún constituyen una pobre representación del taxón y se mantiene en debate, si el mecanismo de biomineralización de los moluscos, se encuentra conservado entre todas las especies (Arivalagan *et al.*, 2016).

2.3 SMPs de gasterópodos y bivalvos

Con los avances ómicos en ácidos nucleicos y proteínas, muchas proteínas que intervienen en la formación de la concha de bivalvos y gasterópodos han sido caracterizadas (*Tabla I*), permitiendo una mejor comprensión del mecanismo de calcificación de los moluscos (Bédouet *et al.*, 2007; le Pabic *et al.*, 2017; Mann y Jackson, 2014). Las proteínas de la matriz de la concha o SMPs ostentan una gran variedad en cuanto a composición y propiedades (Bédouet *et al.*, 2007) como se observa en la *Tabla I*.

Tabla I. Representación de proteínas de la matriz orgánica de la concha de moluscos bivalvos y gasterópodos.

Proteína	Especie	Referencia
BIVALVIA		
N66	<i>P. sterna</i>	Rivera-Pérez <i>et al.</i> (2019a)
Ps19	<i>P. sterna</i>	Arroyo-Loranca <i>et al.</i> , (2020)
Pif97, Pif80	<i>P. fucata</i>	Bahn <i>et al.</i> (2015)
Chematrina	<i>P. fucata</i>	Yano <i>et al.</i> (2006)
Anhidrasa carbónica	<i>P. fucata</i>	Zhao <i>et al.</i> (2018)
Proteína de unión a calcio EF-hand	<i>P. fucata</i>	Huang <i>et al.</i> (2007)
Proteína con dominio de unión a quitina y vWA	<i>P. fucata</i>	Zhao <i>et al.</i> (2018)
Quitina sintetasa	<i>P. fucata</i>	Liu <i>et al.</i> (2015)
N66	<i>P. mazatlanica</i>	Rivera-Pérez <i>et al.</i> (2019b)
N66	<i>P. mazatlanica</i>	Rivera-Pérez <i>et al.</i> (2019a)
Chematrina, KRMP1	<i>P. maxima</i>	McDougall <i>et al.</i> (2013)
Pif97, Pif80	<i>P. margaritifera</i>	Marie <i>et al.</i> (2012)
Pif177-similar, nacreína (A1, B3, B4, C5), tirosinasa, chematrina, KRMP	<i>P. margaritifera</i>	Jeoubert <i>et al.</i> (2010)

Proteína	Especie	Referencia
Proteína de unión a quitina, MNRP	<i>P. imbricata</i>	Liu. No publicado (AIF72920, AIT18245 respectivamente)
Proteína de ácido retinoico insoluble	<i>C. gigas</i>	Zhang <i>et al.</i> (2014)
Nacreína	<i>C. gigas</i>	Song <i>et al.</i> (2014)
Tirosinasa	<i>C. gigas</i>	Huan <i>et al.</i> (2013)
Pif97	<i>C. gigas</i>	Wang <i>et al.</i> (2013)
Proteína poly-Ala Shelk2	<i>C. virginica</i>	Takahashi <i>et al.</i> No publicado (BAM66399)
Proteína similar a perlucina, inhibidor a proteasa	<i>M.gallorprovincialis</i>	Gao <i>et al.</i> (2015)
BMSP	<i>M. galloprovincialis</i>	Suzuki <i>et al.</i> (2011)
GASTROPODA		
HasCL10contig2; P0012N13_463; P0006O07_675; P0025F23_658; ML3A11; ML6A10; ML8B1; P0011O14_517; 6G3;	<i>H. asinina</i>	Marie <i>et al.</i> (2010)
Proteína con dominio tipo BPTI/Kunitz	<i>H. asinina</i>	Jackson <i>et al.</i> (2010)
Perlwin, lustrina	<i>H. asinina</i>	Jackson <i>et al.</i> (2006)
Lustrina	<i>H. tuberculata</i>	Gaume <i>et al.</i> (2014)
Anhidrasa carbónica	<i>H. tuberculata</i>	Le Roy <i>et al.</i> No publicado (AEL22201)
Perlucina	<i>H. laegivata</i>	Weiss <i>et al.</i> (2001)
Lustrina A	<i>H. rufescens</i>	Shen <i>et al.</i> (1997)
AP24	<i>H. rufescens</i>	Michenfelder <i>et al.</i> No publicado (AAK00634)
Anhidrasa carbónica	<i>L. gigantea</i>	Marie <i>et al.</i> (2013)
BMSP, nacreína, perlucina, perlustrina, Pif	<i>L. gigantea</i>	Mann <i>et al.</i> (2012)

A pesar de que los proteomas de moluscos son altamente variables, las SMPs pueden presentar similitudes, como son la presencia de ciertos dominios funcionales repetidos (Tabla II). Por ejemplo,

- Las anhidrasas carbónicas están conservadas en los metazoos (Arivalagan *et al.*, 2016) y pertenecen a una superfamilia de metalo-enzimas que catalizan la

hidratación inversa de CO₂; su función en la biomineralización ha sido asociada a la calcificación y como regulador negativo de este proceso. El dominio que contienen las anhidrasas carbónicas (dominio cl00012, PF00194) fue identificado por primera vez en nacreínas de *P. fucata* por (Miyamoto *et al.*, 1996), y posteriormente se detectaron en el gasterópodo *Patella vulgata* (Werner *et al.*, 2013); proteínas similares a nacreínas del bivalvo *C. gigas* (Song *et al.*, 2014); bivalvo *Mytilus coruscus* (Liao *et al.*, 2015), etc.

- Las SMPs también presentan dominios de unión a carbohidratos, con función probablemente asociada a la calcificación. Ejemplo de estos, son los dominios de unión a quitina (CBM_14, PF01607) o ChtBD2 (del inglés: *chitin binding 2*) aislados de diferentes especies de bivalvos como *C. gigas*, *Mytilus edulis*, *Mytilus truncata* y *Penion maximus* (Arivalagan *et al.*, 2016); *M. coruscus* (Liao *et al.*, 2015); *P. fucata* (Liu *et al.*, 2015) y gasterópodos como *H. asinina* (Marie *et al.*, 2010); como su nombre lo indica, estos dominios interactúan con la quitina, el carbohidrato mejor representado en la matriz orgánica

El dominio ChtBD2 se ha reportado en el gasterópodo terrestre *Cepaea nemoralis*, representados en la proteína isotig_821, homóloga a SMPs de moluscos marinos (Mann y Jackson, 2014). Estos dominios interactúan con la quitina a nivel de matriz interlamelar, creando las capas sucesivas dentro de las cuales crecerán los cristales de CaCO₃.

- Otro dominio común en las SMPs son los del tipo vWA (por sus siglas en inglés: *Von Willebrand factor type A*). Los dominios vWA (PF00092) se encuentran en glicoproteínas hidrofílicas y tienen función de adhesión asociada a la interacción proteína-proteína; estos suelen encontrarse formando complejos proteicos como redes de seda, que actúan de conjunto con la quitina y participan en el acomodamiento de los cristales de carbonato de calcio (Arivalagan *et al.*, 2016). Estos dominios usualmente se encuentran, junto a los dominios de unión a quitina, en las proteínas Pif (Arivalagan *et al.*, 2016).

- Pif es una proteína de la matriz insoluble, y es un factor regulador de la biomineralización en *P. fucata*, donde fuera identificada por primera vez. Pif ha sido localizada en especies de bivalvos como la mencionada *P. fucata* (Wang *et al.*, 2013; Liu *et al.*, 2015) y *C. gigas* (Wang *et al.*, 2013); *Perna viridis* (Liao *et al.*, 2019); *P. margaritifera*, *P. maxima*, *Pteria penguin* y en la proteína BMPS de *Mytilus galloprovincialis* (Suzuki *et al.*, 2013), lo que sugiere que sus homólogos probablemente tengan un rol importante en la biomineralización de los moluscos. Pif es un complejo proteico formado por Pif80 (C-terminal) y Pif97 (N-terminal). Pif97 posee dominios del tipo vWA y dominios de unión a quitina conservados en homólogos, mientras que la proteína Pif80 que interactúa con el CaCO₃ varía entre las especies (Bahn *et al.*, 2015; Arivalagan *et al.*, 2016). Además, Pif posee diferentes isoformas con diversos arreglos de sus dominios (Suzuki *et al.*, 2009).
- Las proteínas de la matriz pueden presentar dominios de inhibición a proteasas, como Kunitz (PF00014), WAP (PF00095) (del inglés: *Whey Acidic Protein*; nombre correcto 4-DSC [debido a su dominio central de cuatro puentes disulfuro]), Kazal (PF00050), etc. Estos dominios se encuentran ampliamente conservados en Metazoa, p.e. testa de erizos de mar, huevos de gallina, etc. (Marie *et al.*, 2010). Los inhibidores de proteasas poseen diferentes funciones y probablemente en el contexto de la matriz orgánica, estén asociadas a la formación y/o mantenimiento del biomineral, al inhibir el efecto de exopeptidasas. Funciones alternativas pueden ser el remodelaje de la concha o regulación (activación-inactivación) de componentes multidominios de la matriz (Arivalagan *et al.*, 2016). Bédouet, *et al.* (2007) reportaron la presencia de una población heterogénea de inhibidores de proteasas de alto y bajo peso molecular, por primera vez, en SMPs de *P. margaritifera*. Presumiblemente, los inhibidores de proteasas proceden de hemocitos de tejidos del manto, los cuales secretan el Ca²⁺ que formará la concha, arrastrando estas proteínas que finalmente quedarán atrapadas en el interior del nácar.
- Marie *et al.* (2010) detectaron dos proteínas en el gasterópodo *H. asinina* con dominios de inhibición a proteasas del tipo Kunitz y una de estas con dominios 4-DSC. (Mann y Jackson, 2014) encontraron dominios 4-DSC en el gasterópodo *C. nemoralis*,

donde probablemente su presencia se debiera a mecanismos de defensa frente a enzimas digestivas secretadas por organismos incrustantes y depredadores, que podrían disolver y penetrar a través de la concha como gasterópodos natácidos, anélidos poliquetos y esponjas (Bédouet *et al.*, 2007; Mann y Jackson, 2014). Los dominios WAP se han identificado en proteínas de *Perna viridis* (Liao *et al.*, 2019).

En cuatro especies de bivalvos *C. gigas*, *M. edulis*, *P. truncata* y *M. truncata* estudios de proteómica de Arivalagan *et al.* (2016) revelaron la presencia de dominios de inhibición a proteasas como TIMP (PF00965) [del inglés: *tissue inhibitor of metalloproteinase*]; inhibidores de inter-alfa-tripina (PF08487) o VIT (del inglés: *vault protein inter-alpha-trypsin domain*); dominios Serpin (PF00079), Kazal (PF00050) o Kunitz (PF00014); macroglobulina alfa-2 (PF00207) o A2M (del inglés: *alpha-2-macroglobulin family*; un tipo de inhibidor de endopeptidasa que actúa en la defensa contra infecciones). La presencia de estos dominios se asoció a un rol de protección de la estructura de la concha en los moluscos, como la defensa bioquímica o como proteínas inmunomoduladoras. En este estudio se identificaron, además, proteínas de PRR (del inglés *pattern recognition receptors*) que actúan en la señalización inicial ante infecciones, proteínas de unión a lípidos, lectinas, receptores tipo Toll (TLR), *nod-like receptors* o receptores en forma de cabeza (NLR) como parte de la inmunidad innata, el cual desencadena respuestas de melanización, fagocitosis, encapsulamiento, opsonización, etc., con función similar a los dominios anteriormente mencionados.

- Las proteínas con dominios de inversión (*epedymine-related proteins*; PF00811) pertenecen a una familia de glicoproteínas extracelulares. Estos dominios poseen aproximadamente 200 aminoácidos, con seis residuos de Cys, de los cuales cuatro son altamente conservados y probablemente forman puentes disulfuro; presentan, además, dos sitios de glicosilación putativos. Estas proteínas tienen dos funciones: 1) se unen a Ca^{+2} a través de residuos de ácido siálico-N unidos y 2) se polimerizan en fibras insolubles actuando probablemente en la organización de la matriz orgánica

(Marie *et al.*, 2010). Estos dominios fueron detectados por Marie *et al.* (2010) en el gasterópodo *H. asinina*.

- Las regiones de baja complejidad o LCR (del inglés: *low complexity regions*) están también presentes en SMPs. Las LCR confieren alta flexibilidad a las proteínas facilitando su transporte a través de las membranas (las proteínas con conformaciones más flexibles migran más fácil que las proteínas plegadas). Debido a que las proteínas de la matriz son sintetizadas en el epitelio calcificante, y posteriormente trasladadas hasta el sitio de mineralización, la presencia de estos dominios podría ayudar en el proceso. La plasticidad intrínseca conferida por estos dominios podría proporcionar la propiedad de reconocer varios blancos biológicos, sin sacrificar especificidad. Las LCRs, además, evolucionan a gran velocidad vía eventos de recombinación, y pueden constituir una respuesta de adaptación al ambiente (Arivalagan *et al.*, 2016). Estas regiones se han identificado en varias especies de bivalvos como *C. gigas*, *M. edulis*, *P. truncata*, *M. truncata*, *Perna viridis*, *Mytilus galloprovincialis*, *Mytilus spp*, *Crassostrea gigas* y *P. fucata* (Arivalagan *et al.* (2016); Liao *et al.*, 2019).

En un estudio realizado en el cefalópodo *Sepia officinalis*, se aislaron varias SMPs que mostraron cierta homología hacia secuencias pre-existentes en la base de NCBI; estas proteínas se caracterizaron por presentar un dominio de unión a quitina tipo 2, un vWA y 3 Kunitz, además de un dominio O-glicosil hidrolasa (PF00232) y un dominio transferrina (PF00405). Todos estos son dominios previamente descritos para especies de bivalvos y gasterópodos, lo que sugiere la presencia de un mecanismo ancestral, conservado, en la biomineralización de los moluscos (le Pabic *et al.*, 2017).

Tabla II. Proteínas/dominios-estructurales comunes en la matriz orgánica de moluscos bivalvos y gasterópodos.

Especie	Proteína/Dominio	Función	Referencia
<i>P. fucata</i>	Pif	Regulador de la biomineralización (posee vWA y dominios de unión a quitina)	Bahn <i>et al.</i> (2015)
<i>C. gigas</i>			Wang <i>et al.</i> (2013)
<i>P. margaritifera</i>	Inhibición a proteasas (tipo Kunitz o vWA)	Formación y/o mantenimiento del biomineral; remodelaje de la concha o regulación (activación-inactivación) de componentes multidominios de la matriz; defensa contra depredadores	Bedouet <i>et al.</i> (2007b)
<i>H. asinina</i>			Marie <i>et al.</i> (2010)
<i>C. nemoralis</i>			Mann y Jackson (2014)
<i>C. gigas, M. edulis, P. maximus y M. truncata</i>			Arivalagan <i>et al.</i> (2016)
<i>H. asinina</i>	Dominios de inversión	Unión a Ca ²⁺ , polimerización en fibras insolubles actuando probablemente en la organización de la matriz orgánica	Marie <i>et al.</i> (2010)
<i>C. gigas, M. edulis, P. maximus y M. truncata</i>	Anhidrasa carbónica	Regulador negativo de la calcificación	Arivalagan <i>et al.</i> (2016)
<i>C. gigas, M. edulis, P. maximus y M. truncata</i>	Dominio de unión a quitina (ChtBD2)	Unión al polisacárido quitina para formar el esqueleto tridimensional donde crecerán los cristales de carbonato de calcio	Arivalagan <i>et al.</i> (2016)
<i>C. nemoralis</i>			Mann y Jackson (2014)
<i>C. gigas, M. edulis, P. maximus y M. truncata</i>	Tirosinasa	Por identificar	Arivalagan <i>et al.</i> (2016)

Como se observa en la Tabla II existen dominios funcionales comunes entre las SMPs de gasterópodos y bivalvos, lo que sugiere la presencia de una batería básica involucrada en la biomineralización de los moluscos. En este sentido, un estudio realizado por (Arivalagan *et al.*, 2016) comparó por primera vez los proteomas de cuatro especies divergentes de bivalvos *Crassostrea gigas*, *Mytilus edulis*, *Mytilus truncata* y *Penion maximus* extraídas y analizadas utilizando el mismo procedimiento. En el mismo, se determinó la presencia de cuatro dominios funcionales comunes en las SMPs de las mencionadas especies: tirosinasa, anhidrasa carbónica, dominio de unión a quitina 2 y Von Willebrand factor-A, fortaleciendo

la hipótesis de la batería básica ancestral responsable de la mineralización de los moluscos, pues las especies comparadas son divergentes y muestran diferente mineralogía y microestructura en sus conchas.

Hasta la fecha, no se ha descrito la batería básica mínima responsable de la biomineralización de los moluscos, y se desconoce si pudiesen existir baterías asociadas a las diferentes clases taxonómicas de moluscos y/o incluso proteínas y/o dominios funcionales asociados a las diferentes microestructuras detectadas en las capas calcáreas de la concha. La presente investigación persigue identificar y comparar las baterías básicas de proteínas de la concha de moluscos bivalvos y gasterópodos y establecer su relación con las microestructuras de la concha. Para ello, se realizó una búsqueda de SMPs (proteínas y ácidos nucleicos) reportadas en bases de datos y artículos científicos, las cuales fueron comparadas para identificar las proteínas comunes que pudieran formar la batería básica para la formación de la concha de moluscos.

3. JUSTIFICACIÓN

La biomineralización de los moluscos es un proceso de calcificación controlado por moléculas orgánicas que se entrelazan y actúan mediando la formación de los cristales de CaCO_3 . Las proteínas de la matriz orgánica de la concha o SMPs son los principales reguladores del proceso de biomineralización y se han descrito en numerosas especies del filo. Muchas de estas proteínas, aunque diferentes, poseen dominios previamente descritos para SMPs de otros moluscos conchíferos, lo que sugiere la presencia de mecanismos antiguos y ancestrales en su proceso de biomineralización. El presente estudio intenta mediante un enfoque comparativo la identificación de las proteínas y/o dominios comunes en diferentes especies de moluscos que constituyan la batería básica imprescindible para la biomineralización en estos organismos.

La identificación de una batería básica permitirá: A) comprender mejor el proceso de biomineralización de los moluscos, organismos en los que las conchas son muy diferentes entre sus clases taxonómicas, géneros e incluso especies; B) la identificación de las proteínas imprescindibles para la biomineralización impulsará la formulación *in vitro* de biominerales y el consecuente desarrollo de bio-nanomateriales que son utilizados en diversas industrias. Por ejemplo, en la biomedicina para el desarrollo de tecnologías eficientes para el tratamiento de cáncer, enfermedades óseas (hipofosfatasa, osteoporosis, osteomalacia), desarrollo de vacunas, terapias enzimáticas, etc.; C) la creación en condiciones de laboratorio de nácar es de gran atractivo para la joyería, donde la identificación de la batería básica representaría un hito en la producción de perlas y la disminución de la explotación marítima de moluscos con estos fines. Adicionalmente, en cosmetología, el nácar es utilizado en la producción de cremas y jabones con propiedades de despigmentación y antienvjecimiento.

Teniendo en cuenta que existen polimorfos microestructurales en las capas calcáreas de la concha de moluscos, el estudio de las baterías proteicas asociadas a las principales clases (gasterópodos y bivalvos) permitirá hipotetizar las posibles proteínas asociadas a las diferentes microestructuras y, por ende, comprender mejor el proceso de biomineralización desde un ángulo evolutivo y entender el impacto de las microestructuras en la capacidad de los organismos para hacer frente a los cambios ambientales. Por otro lado, la recreación en

condiciones *in vitro* de diferentes microestructuras permitiría jugar con las propiedades de los bio-nanomateriales como resistencia, dureza, flexibilidad, etc. ampliando su gama de aplicaciones. Por último, muchas SMPs se caracterizan por presentar regiones de baja complejidad y alta repetitividad lo que hace su captura mediante oligonucleótidos, secuenciación proteómica y estrategias de visualización procedimientos complejos. Al realizar la descripción bioinformática o *in silico* de sus rasgos permitirá ampliar el catálogo de información disponible sobre las proteínas de la concha de moluscos, aspecto crítico en la dilucidación de su papel evolutivo en cuestiones de plasticidad adaptativa.

4. HIPÓTESIS

Si las proteínas de la matriz orgánica de la concha de los moluscos regulan la nucleación, cristalización y direccionamiento del carbonato de calcio que formará la concha, entonces existe una batería básica de proteínas responsables de la formación de la concha de moluscos, bivalvos y gasterópodos; y las diferencias en las baterías básicas de los mismos estarán relacionadas a las diferentes microestructuras presentes en las capas calcáreas de cada especie en particular.

5. OBJETIVOS

5.1 Objetivo general

Identificar la batería básica de proteínas que intervienen en la formación de la concha de moluscos gasterópodos y bivalvos.

5.2 Objetivos particulares

1. Construir una base de datos de secuencias proteicas de gasterópodos y bivalvos recopilando información disponible en repositorios públicos y literatura científica.
2. Determinar la batería básica de proteínas que interviene en la formación de la concha de moluscos, gasterópodos y bivalvos.
3. Hipotetizar las posibles proteínas y sus respectivos dominios estructurales responsables de las diferentes microestructuras de la concha de moluscos gasterópodos y bivalvos.

6. MATERIAL Y MÉTODOS

6.1 Construcción de una base de datos de secuencias proteicas de gasterópodos y bivalvos

Para identificar las proteínas encargadas de la biomineralización de los moluscos de las clases gasterópodos y bivalvos, se construyeron dos bases de datos (una de gasterópodos y otra de bivalvos), usando el paquete Access de Microsoft Office.

Se realizó una búsqueda en la base de datos de proteínas y nucleótidos del *National Center for Biotechnology Information* (NCBI) (<https://www.ncbi.nlm.nih.gov/>), empleando las palabras clave: *shell matrix*, *protein*, *bivalves* y *gastropods*. Además, se seleccionaron 10 proteínas de interés (proteínas seleccionadas atendiendo a su relevancia en la biomineralización de moluscos: *chitin synthase*, *carbonic anhydrase*, *EGF-like domain containing protein*, *BPTI-Kunitz domain-containing protein*, *nacreina*, *perlucin*, *Pif97*, *shematrina*, *tyrosinase* y *vWA and chitin binding domain-containing protein*) para búsqueda mediante Blastp en repositorios de genomas de especies de interés: *C. gigas* (bivalvo) y *L. gigantea* (gasterópodo) en Ensembl Metazoa (<http://metazoa.ensembl.org/index.html>).

Las secuencias de proteínas y/o nucleótidos fueron obtenidas en formato FASTA y se recopiló información significativa de las mismas procedente del NCBI (especie, número de acceso, tejido de procedencia, capa calcárea donde se localiza en la concha y dominios conservados). Se determinó el hábitat de las especies utilizando *Worms: World Register of Marine Species* (<http://www.marinespecies.org/>). Las secuencias proceden de datos proteómicos y transcriptómicos, aisladas de la concha y rastreadas hasta sus genomas ensamblados respectivamente (Mann y Jackson, 2014) y para su selección se consideró la calidad (secuencias parciales o completas), disponibilidad de los proteomas, y descripción de las proteínas de acuerdo con el proceso de biomineralización (le Pabic *et al.*, 2017).

Posteriormente, las bases de datos se depuraron, manualmente, eliminando parciales y/o incompletas de menor importancia, incluyendo un elevado porcentaje de secuencias correspondientes a NADH deshidrogenasas, ATP sintasas y citocromos b y c. Se eliminaron secuencias parciales bajo el precepto del desconocimiento exacto de la información contenida en las secuencias y la posibilidad de introducción de errores en las relaciones a

establecerse entre las secuencias. Para un estudio enfocado por primera vez en el análisis del basto grueso de SMPs descritas para estos organismos, se consideró apropiado partir de secuencias completas. Se excluyeron también secuencias repetidas. Se hicieron cladogramas de las proteínas más representativas para bivalvos y gasterópodos mediante mediante MEGA X (MEGA7: *Molecular Evolutionary Genetics Analysis version 7.0* (Kumar *et al.*, 2016) usando *Neighbor Joining* (parámetros por defecto).

6.2 Determinación de la batería básica de proteínas

6.2.1 Caracterización de las secuencias de las bases de datos

Las secuencias colectadas de bivalvos y gasterópodos fueron caracterizadas utilizando algunas herramientas de bioinformática. El peso molecular y punto isoeléctrico (pI) fueron estimados usando *ExPASy ProtParam* (Gasteiger *et al.*, 2005); la presencia de péptidos señal se determinó usando el programa *SignalP* versión 5.0 (Almagro-Armenteros *et al.*, 2019); y la localización subcelular se estableció usando los programas *ProtComp* 9.0 (<http://www.softberry.com/berry.phtml?topic=protcompan&group=programs&subgroup=proloc>) y *Wolf Psort* (Horton *et al.*, 2007). También, se determinó la presencia de puentes disulfuro con *Dianna* versión 1.1 (Ferre y Clote, 2005); glicosilaciones con *NetNGlyc 1.0 Server* (Gupta y Brunak, 2002) y *NetOGlyc 4.0 Server* (Steentoft *et al.*, 2013). Los dominios conservados se buscaron con SMART (<http://smart.embl-heidelberg.de>) y/o con la herramienta de predicción integrada al NCBI (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>); mientras que las familias de proteínas se determinaron usando *Pfam 34.0: protein families* (Sonnhammer *et al.*, 1997). Según la información disponible en la literatura científica, y en caso necesario, fueron designados nombres para los dominios estructurales no descritos.

6.2.2 Identificación de proteínas comunes entre bivalvos y gasterópodos

Se identificaron proteínas y/o dominios comunes entre las diferentes especies de bivalvos y gasterópodos, las cuales fueron consideradas como potencial batería básica de dichas clases taxonómicas. Se utilizaron tres estrategias: 1) generación de diagramas de Venn para visualizar las relaciones entre dominios y especies, posteriormente 2) diagramas de cuerdas

para un análisis más profundo de esta relación (cualitativo y cuantitativo) y, por último, 3) cladogramas de las proteínas más representativas de las bases de datos.

Los diagramas de Venn se generaron con el software VennPainter versión 1.2.0 (Lin *et al.*, 2016). Los grupos por representar en los diagramas de Venn se seleccionaron de acuerdo con los dominios conservados identificados en seis o más especies (16 dominios y dos regiones en gasterópodos y siete dominios, dos regiones y un motivo en bivalvos). Se utilizaron listados de datos basados en dominios y no en especies; el número de especies (~30 en cada base de datos) no podía ser disminuido sin restar información importante, mientras que el número de dominios según su relevancia por presencia en más de una secuencia.

Para evaluar la relación entre los dominios conservados identificados en proteínas de las diferentes especies de gasterópodos y bivalvos, se construyeron diagramas de cuerdas utilizando la función *circlize* (Gu *et al.*, 2014) en R Studio versión 4.0.1. El diagrama de cuerdas fue construido en base a una matriz de presencia/ausencia de los dominios conservados en las diferentes especies (incluyendo datos cuantitativos: número de observaciones para cada dominio) por lo que no representa similitudes desde el punto de vista de las secuencias proteicas (matriz no incuida).

Por último, las proteínas con mayor representatividad entre bivalvos y gasterópodos fueron analizadas mediante un cladograma. Los análisis fueron realizados entre proteínas similares, que comparten dominios funcionales (según los agrupamientos obtenidos anteriormente). Los alineamientos se realizaron con Clustal Omega versión 1.2.4 y los cladogramas en Jalview versión 2.11.1.4 (Waterhouse *et al.*, 2009). Se utilizó el método de Neighbor Joining con una matriz de sustitución de aminoácidos del tipo BLOSUM62.

Con la información recopilada en los análisis anteriores, se seleccionaron las proteínas/dominios funcionales integrantes de la batería básica encargadas de la biomineralización de gasterópodos y bivalvos.

6.3 Caracterización de las proteínas y su relación con la microestructura de la concha

Una vez determinadas las proteínas y/o dominios funcionales integrantes de la batería básica, se procedió a su caracterización bioinformática. Dicha caracterización se fundamentó en la información disponible en diferentes programas como SMART (<http://smart.embl-heidelberg.de>), la herramienta de predicción integrada al NCBI (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>), *Pfam 34.0: protein families* (Sonnhammer *et al.*, 1997) y la literatura científica; justificándose su presencia en la batería básica (Mann y Jackson, 2014; Arivalagan *et al.*, 2016). La relación de las proteínas con la microestructura fue obtenida también de literatura científica.

7. RESULTADOS

7.1 Construcción de una base de datos de secuencias proteicas de gasterópodos y bivalvos

Las palabras claves utilizadas para la generación de las bases de datos fueron: *shell protein + bivalves* o *gastropods*. Con estas palabras, la búsqueda en la base de datos de NCBI arrojó un total de 5,392 secuencias de proteínas (3,286 de gasterópodos y 2,106 de bivalvos). A este total se sumaron: 10 proteínas de *L. gigantea* (gasterópodo) y 10 de *C. gigas* (bivalvo) extraídas de *Ensembl Metazoa*; para cada especie fueron las siguientes proteínas: *chitin synthase*, *carbonic anhydrase*, *EGF-like domain containing protein*, *BPTI-Kunitz domain-containing protein*, *nacreina*, *perlucin*, *Pif97*, *shematrina*, *tyrosinase* y *vWA and chitin binding domain-containing protein*. Además, a cada base de datos se incorporaron 19 secuencias proteicas (gasterópodos) y 73 secuencias (bivalvos) procedentes de búsquedas específicas en repositorios públicos y literatura científica (

Tabla III). Se eliminaron un total de 4,081 secuencias parciales y seis secuencias de organismos no moluscos (secuencias parciales específicas fueron conservadas debido a su función vinculada a la biomineralización de moluscos); las secuencias parciales eliminadas pertenecían mayormente a proteínas relacionadas a la maquinaria energética.

Tabla III. Características de las bases de datos construidas de SMPs de gasterópodos y bivalvos.

Base de datos	Palabras clave	Secuencias por grupo		
		Bivalvos	Gastrópodos	Total
<i>National Center for Biotechnology Information</i>	<i>shell protein, bivalves, gastropods</i>	2,106	3,286	5,392
<i>Ensembl Metazoa</i>	<i>chitin synthase, carbonic anhydrase, EGF-like domain containing protein, BPTI-Kunitz domain-containing protein, nacreina, perlucin, Pif97, shematrina, tyrosinase y vWA and chitin binding domain-containing protein</i>	<i>C. gigas</i> (10)	<i>L. gigantea</i> (10)	20
Repositorios públicos y literatura científica	Mismas	73	19	92
				5,504
				Parciales y no moluscos
				- 4,081
Número de secuencias finales		1,121	301	1,423

7.1.1 Gasterópodos

La base de datos de gasterópodos, originalmente con un total de 3,314 secuencias fue depurada de 3,006 secuencias parciales (se mantuvieron tres secuencias parciales de interés) y siete secuencias repetidas, quedando integrada por 301 secuencias. La base de datos agrupó un total de 32 especies, de las cuales el 91% (29) son marinas, el 6% (2) dulceacuícolas y el 3% (1) terrestre. Las especies mejor representadas fueron: *Lottia gigantea* (50), *Amalda northlandica* (26), *Oxymeris dimidiata* (26), *Penion maximus* (26), *Penion sulcatus* (26), *Littorina saxatilis* (20), *Haliotis asinina* (17) y *Reishia clavigera* (17) (Tabla IV).

Tabla IV. Diversidad de especies y número de secuencias por especie en la base de datos de gasterópodos.

Especies	No. secuencias	Especies	No. secuencias
<i>Amalda northlandica</i>	26	<i>Littorina saxatilis</i>	20
<i>Aplysia californica</i>	11	<i>Lottia gigantea</i>	50
<i>Arianta arbustorum</i>	2	<i>Lymnaea stagnalis</i>	2
<i>Biomphalaria glabrata</i>	2	<i>Nipponacmea fuscoviridis</i>	1
<i>Bivetiella cancellata</i>	13	<i>Omphalius pfeifferi</i>	4
<i>Conus borgesii</i>	13	<i>Oxymeris dimidiata</i>	26
<i>Conus geographus</i>	1	<i>Patella vulgata</i>	2
<i>Conus gloriamaris</i>	1	<i>Penion maximus</i>	26
<i>Conus magus</i>	1	<i>Penion sulcatus</i>	26
<i>Conus radiatus</i>	1	<i>P. sulcatus mandarinus</i>	13
<i>Conus stercusmuscarum</i>	1	<i>Pomacea canaliculata</i>	2
<i>Haliotis asinina</i>	17	<i>Reishia clavigera</i>	17
<i>Haliotis diversicolor</i>	1	<i>Solen strictus</i>	1
<i>Haliotis laevigata</i>	7	<i>Steromphala varia</i>	4
<i>Haliotis rufescens</i>	5	<i>Strombus decorus persicus</i>	1
<i>Haliotis tuberculata</i>	3	<i>Turbo marmoratus</i>	1

La base de datos de gasterópodos contiene ~60 proteínas diferentes, con abundantes: NADH deshidrogenasas (diferentes subunidades), citocromo c (diferentes subunidades), ATP sintasa F0 (diferentes subunidades), citocromo b, las cuales agrupan aproximadamente 177 secuencias; las proteínas más representadas fueron: proteínas no caracterizadas (13).

Muchas de estas proteínas recibieron su nombre de acuerdo con sus características distintivas, pero esto no significa que no sean similares entre ellas o pertenezcan a un mismo grupo o familia; por tanto, se requieren análisis adicionales. Las proteínas del metabolismo energético se concentraron fundamentalmente en las especies *Amalda northlandica*, *Littorina saxatilis*, *Penion maximus*, *Oxymeris dimidiata*, *Penion sulcatus*, *P. maximus*, *Conus borgesii*, *Reishis clavigera*, *Aplysia californica* y *Bivetiella cancellata*.

7.1.2 Bivalvos

La base de datos de bivalvos contenía 2,189 de las cuales se eliminaron: 1,029 secuencias parciales; 33 secuencias repetidas y seis secuencias de organismos no moluscos (bacterias y simbiontes), quedando finalmente integrada por 1,121 secuencias. La base de datos agrupó un total de 37 especies (Tabla V), siendo cerca del 98% de hábitat marino y el 2% de una especie dulceacuícola: *Hyriopsis cumingii*. La especie *Pinctada fucata* fue la más abundante con 520 secuencias (~47%); seguida de *P. margaritifera* (99), *Crassostrea gigas* (69), *P. maxima* (46), *Mytilus galloprovincialis* (44) y *Solen grandis* (43).

Tabla V. Diversidad de especies y número de secuencias por especie en la base de datos de bivalvos.

Especie	No. secuencias	Especie	No. secuencias
<i>Adamussium sp.</i>	1	<i>Mytilus galloprovincialis</i>	44
<i>Atrina pectinata</i>	27	<i>Pecten maximus</i>	5
<i>Atrina rigida</i>	11	<i>Perna canaliculus</i>	4
<i>Azumapecten farreri</i>	2	<i>Pinctada fucata</i>	520
<i>Crassostrea gigas</i>	69	<i>Pinctada imbricata</i>	10
<i>Crassostrea nippona</i>	5	<i>Pinctada margaritifera</i>	99
<i>Crassostrea virginica</i>	11	<i>Pinctada maxima</i>	46
<i>Cyclina sinensis</i>	1	<i>Pinctada mazatlanica</i>	1
<i>Hyriopsis cumingii</i>	12	<i>Pinna nobilis</i>	4
<i>Isognomon perna</i>	1	<i>Pteria penguin</i>	28
<i>Meretrix meretrix</i>	1	<i>Pteria sterna</i>	1
<i>Meretrix petechialis</i>	2	<i>Ruditapes decussatus</i>	3
<i>Mizuhopecten yessoensis</i>	12	<i>Ruditapes philippinarum</i>	7
<i>Mya arenaria</i>	25	<i>Scapharca broughtonii</i>	30
<i>Mytilisepta virgata</i>	2	<i>Solen grandis</i>	43
<i>Mytilus californianus</i>	9	<i>Solen strictus</i>	30
<i>Mytilus coruscus</i>	44	<i>Tridacna squamosa</i>	2

<i>Mytilus edulis</i>	4	<i>Unio pictorium</i>	5
-----------------------	---	-----------------------	---

La base de datos de bivalvos mostró una diversidad de proteínas, siendo las mejores representadas las proteínas no caracterizadas del tipo: proteínas de nácar, proteínas de la matriz y proteínas de la matriz insoluble. Similar a la base de datos de gasterópodos, las proteínas energéticas fueron representativas (NADH deshidrogenasas, ATP sintasas y citocromos).

7.2 Determinación de la batería básica de proteínas

7.2.1 Caracterización de las secuencias de las bases de datos.

Se caracterizaron las proteínas de las bases de datos de bivalvos y gasterópodos de acuerdo con su punto isoeléctrico, presencia/ausencia de péptido señal y modificaciones postraduccionales, tales como O- y N-glicosilaciones y puentes disulfuro. Las proteínas de la base de datos de gasterópodos y bivalvos se caracterizaron por presentar pesos moleculares y puntos isoeléctricos variables; gasterópodos con pI entre 11.69 y 3.53, con ~53% de proteínas en el rango de pH básico y un ~47% en el rango ácido; mientras que bivalvos entre 1.79 y 11.79, con ~44% de proteínas en el rango básico y 56% en el rango ácido. Se predijeron un total de 55 secuencias con posibles péptidos señales (18%) en gasterópodos y 800 secuencias en bivalvos (71%) (Figura 2). El tamaño del péptido señal osciló entre 15 - 25 aminoácidos en gasterópodos y 15 - 30 en bivalvos. El 72% de las secuencias de gasterópodos posiblemente presenta puentes disulfuros (218), además de 69 secuencias (23%) con posibles O-glicosilaciones y 175 (58%) con N-glicosilaciones. En bivalvos, se predijeron 962 secuencias con posibles puentes disulfuro en bivalvos (86%), así como 381 secuencias con O-glicosilaciones (34%) y 348 con N-glicosilaciones (31%).

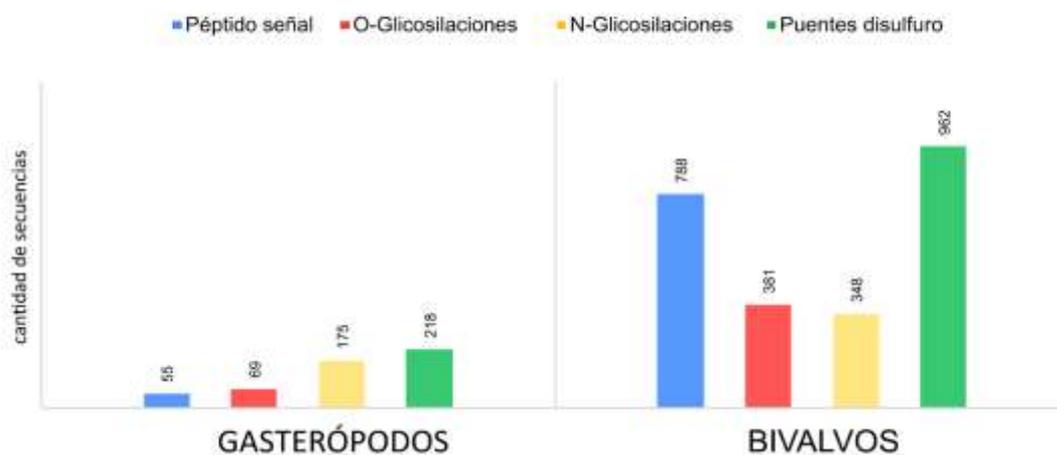


Figura 2. Caracterización bioinformática de las proteínas de las bases de datos de bivalvos y gasterópodos. Resumen de la cantidad de secuencias con posibles péptidos señales, O- y N-glicosilaciones y puentes disulfuro detectadas a través de SignalP versión 5.0 (Almagro Armenteros et al., 2019), NetOGlyc 4.0 Server (Steentoft *et al.*, 2013), NetNGlyc 1.0 Server (Gupta and Brunak, 2002) y DiANNA versión 1.1 (Ferre y Clote, 2005), respectivamente.

El péptido señal se asocia a la localización celular de las proteínas. Las principales localizaciones celulares asociadas a SMPs con péptido señal fueron extracelular secretada y extracelular secretada de unión a membrana, para ambas bases de datos (~76% de todas las secuencias con péptido señal para gasterópodos y ~95% de bivalvos fueron descritas para estas localizaciones). Se identificaron 25 SMPs de gasterópodos con localización celular del tipo extracelular secretada y 17 con extracelular secretada de unión a membrana, mientras que, en bivalvos, se identificaron 532 SMPs con localización extracelular secretada y 216 con localización celular extracelular secretada de unión a membrana (con base a su péptido señal).

En el caso de las glicosilaciones, la representatividad de dicha modificación fue variable entre ambas bases de datos, atendiendo a la composición de estas, así como las características intrínsecas de sus secuencias constituyentes. Por otro lado, los puentes disulfuro son una modificación postraduccional más común que las glicosilaciones para las proteínas de la concha de bivalvos y gasterópodos, un 70% de las secuencias contenidas en las bases de datos generadas en este estudio poseen posibles puentes disulfuro. Para gasterópodos, los puentes disulfuros se predijeron para un conjunto variado de proteínas incluidas proteínas ricas en Pro, Ser, Gly, precursores de aragonita, nacreína, perlihibina,

fosfolipasas, perlustrinas, perlawinas, anhidrasas carbónicas y deshidrogenasas, entre otras. En bivalvos, se predijo la presencia de puentes disulfuro en proteínas del tipo: no caracterizadas, proteínas de nácar, de la matriz insoluble, perlucinas, nacreínas, Pif, metaloproteinasas, KRMP1, NADH deshidrogenasas, quitina sintetasas con dominios EGF (PF00008), tirosinasas (PF00264), quitinasas (PF00182), BPTI/Kunitz domain (PF00014), etc. La presencia de este tipo de modificación postraducciona en algunas de estas proteínas se relaciona directamente a la presencia de ciertos dominios cuyos motivos centrales poseen Cys como los dominios WAP (4-DSC) (PF00095), EGF (PF00008), etc.

Retomando la localización celular, se predijo que una gran parte de las proteínas en ambas bases de datos son extracelulares secretadas [gasterópodos: 61 (20%) y bivalvos: 596 (53%)]; seguidas por proteínas extracelulares unidas a membranas (secretadas) (gasterópodos: 27 [10%] y bivalvos: 246 [22%]), y proteínas unidas a membrana plasmática (gasterópodos: 75 [25%] y bivalvos: 73 [7%]) (Fig. 3A). Por otro lado, las cito-localización menos común fue la de organelos como: núcleo, citoplasma, vacuola, peroxisoma, lisosoma, etc., y a membranas de organelos como: Golgi, lisosoma, peroxisoma, etc. (Figura 3B).

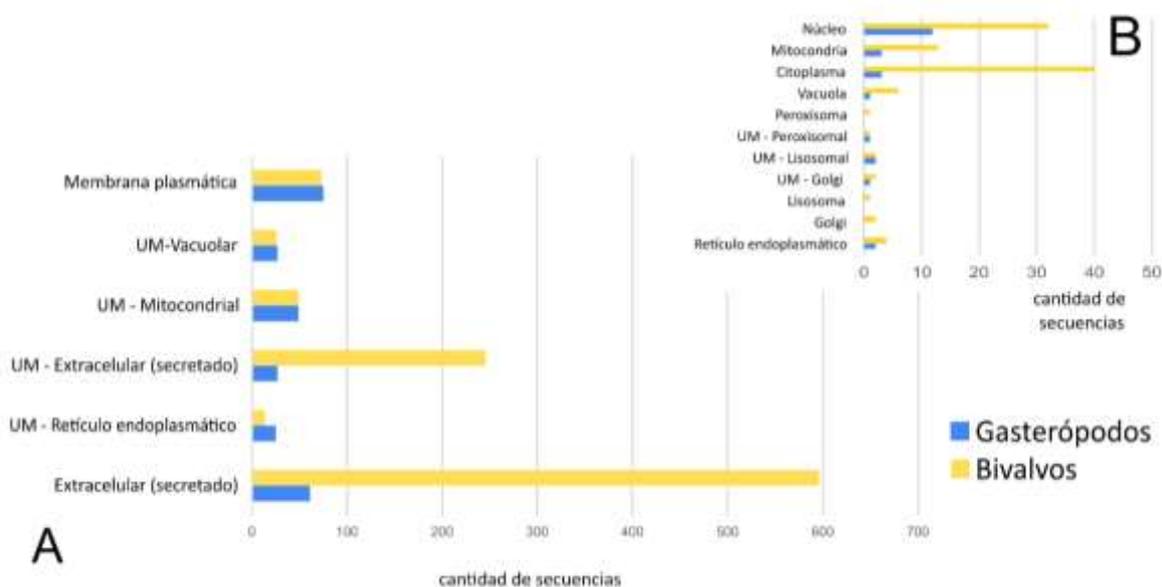


Figura 3. Predicción de localizaciones celulares de las SMPs contenidas en la base de datos de gasterópodos y bivalvos. A: Localizaciones más comunes. B: Localizaciones menos comunes. Identificación mediante ProtComp 9.0. UM: unido a membrana (<http://www.softberry.com/berry.phtml?topic=protcompan&group=programs&subgroup=proloc>) y WoLF PSORT (Horton *et al.*, 2007).

Se identificaron 56 dominios conservados en secuencias de SMPs de gasterópodos y 163 en bivalvos. Además, se identificaron regiones de regiones de baja complejidad (LCRs) y regiones transmembranales (TR) (Figura 4).

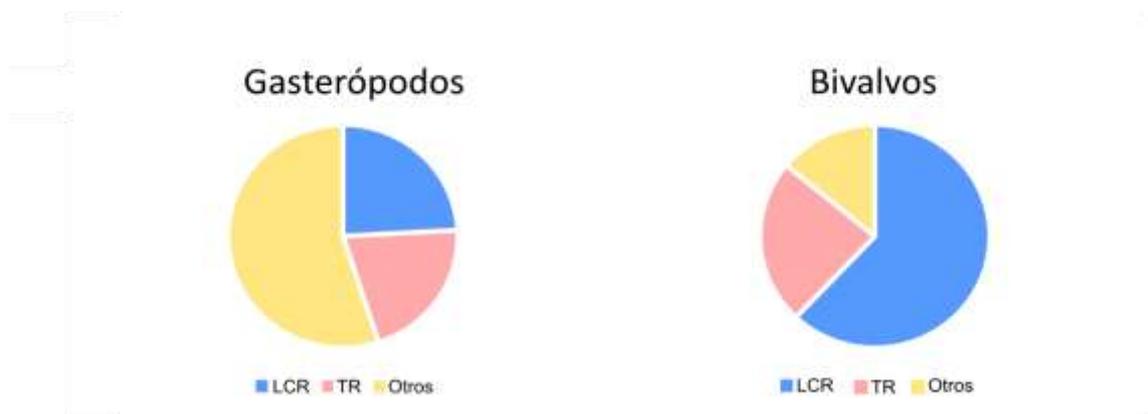


Figura 4. Distribución de los principales dominios y regiones identificadas en SMPs de las bases de datos de gasterópodos y bivalvos. LCR: Regiones de baja complejidad (*Low complexity regions*); TR: Regiones transmembranales (*Transmembrane regions*); Otros dominios identificados. Caracterización mediante herramienta SMART (<http://smart.embl-heidelberg.de>) y/o con la herramienta de predicción integrada al NCBI (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>).

Como se observa en la Figura 5, se identificaron dominios “comunes” para las SMPs de ambas bases de datos (32 dominios), mientras que un grupo de dominios fueron “exclusivos” de cada clase (gasterópodos: 26 y bivalvos: 131) (Figura 6).

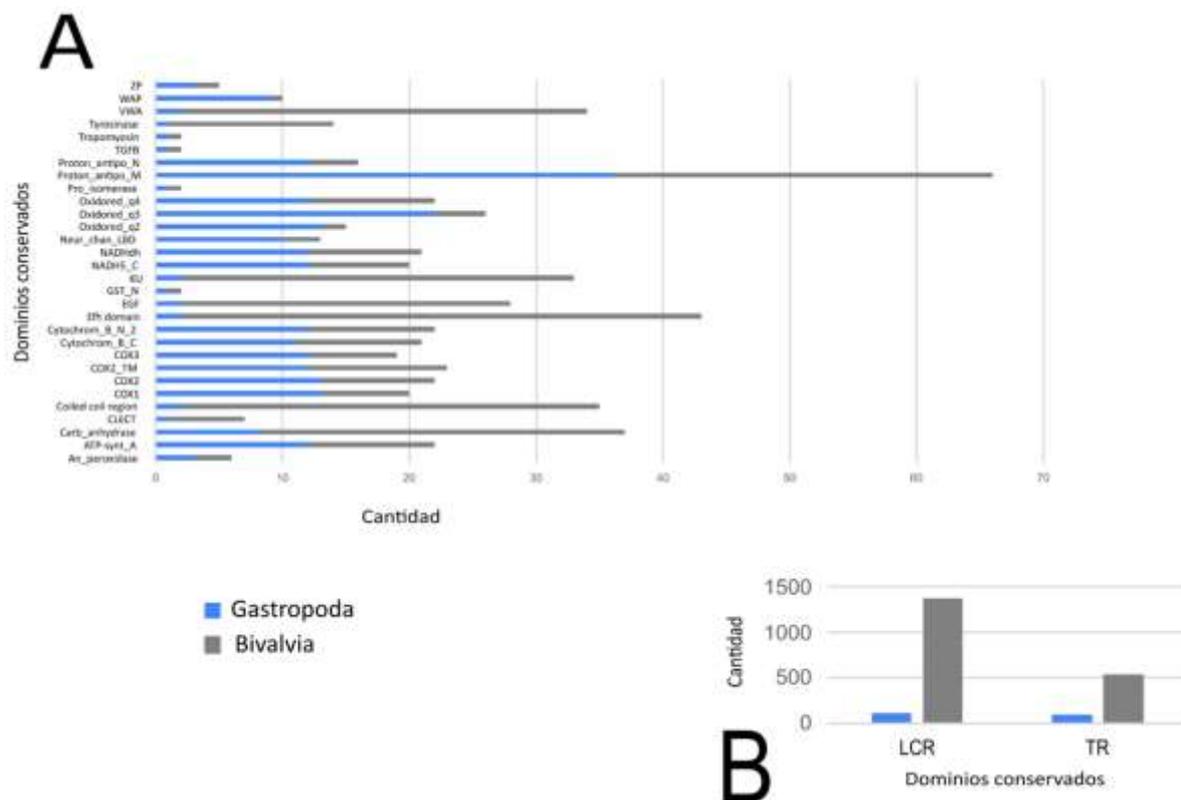


Figura 5. Dominios y regiones identificadas en SMPs de gasterópodos y bivalvos. A: Dominios menos representativos. B: Regiones más representativos. ZP: *Zona pellucida*; WAP: *Whey Acidic Protein*; vWA: *von Willebrand Factor A*; TGFβ: *Transforming growth factor-beta*; EGF: *Epidermal growth factor*; KU: *Kunitz*; CLECT: *C-type lectin*; LCR: *Low complexity regions*. TR: *Transmembrane region*. Identificación mediante herramienta SMART (<http://smart.embl-heidelberg.de>) y/o con la herramienta de predicción integrada al NCBI (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>).

Gasterópodos	Bivalvos				
ATP-synt_8	1-cyxPrx_C	CARD_2	EF-hand_8	Jacalin	Peptidase_C1
Conotoxin	A2M	Cation ATPase_N	EGF Lam	KAZAL	Peptidase_C14
Cytochrom_B	A2M_recep	Cation_ATPase	EGF like	LamB	Pahge_lysozyme
DERM	A2M_comp	Cation_ATPase_C	Fam20C	LamG	PNTB
EF-hand5	A2M_N	CCP	Ferritin	Laminin_EGF	PNTB_4TM
Ependymin	A2M_N_2	CH	Filament	Laminin_G_3	polyprenyl_synt
GST_3	AAA	CHB_HEX	FN3	LDLa	Ribonuc_L-PSP
GST_C	AhpC-TSA	Chitin_synth_2	Gal_Lectin	LER	Ribosomal_S19e
HOLI	AlaDh_PNT_C	ChtBD2	Gal-bind lectin	LITAF	RIG-I_C-RD
Homeobox	AlaDh_PNT_N	Cpn60_TCP1	GLECT	LMPO_10	S_TKc
HOX	Ami_2	Cu_amine_oxid	Glyco_18	LP_N	SAM
IB	Amino_oxidase	Cu_amine_oxidN2	Glyco_hidro_18	LTD	SEA
IDO	ANK	Cu_monoox_C	Glyco_hidro_20	Macin	SERPIN
IRPT	Antistasin	Cu_monooxygen	Glyco_hidro_20b	MreB_Mbl	Sod_Fe_C
Lustrin_Cysteine	ApeC	Cu-oxidase_2	Glyco_tranf_2_3	MYS_C	SPARC_Ca_bdg
Methyltransf_FA	ATP-gua_Ptrans	Cu-oxidase_3	GST_C_2	Na_Ca_ex	SRP40_C
Oxidored_q5_N	ATP-gua_PtransN	Death	HELICc	Neur_chan_memb	TCTP
Oxidored_q5_N	ATP-synt_ab	Death_2	HintN	NTR	TGFb_propeptide
SCP	ATP-synt_ab_N	Defensin_beta_2	HLH	Oxidored_q5	Thiol-ester_cl
TGBF	BPI1	Defensin_big	Hydrolase	p450	THY
TGFb	BPI2	DERM	IG	P53	TIMP
TGFb_propeptide	BRLZ	DEXDc	IG_FLMN	P53_TAD	TIR
Tropomyosin_1	C1q	DUF1081	IGc2	P53_tetramer	TIR_2
WR1	Calponin	DUF1943	IL17	PAC	Tryp_SPc
ZnF_C4	Calx beta	E1-E2_ATPase	Inhibitor_I29	PAS	TY
			Integrin_B_tail	PDZ	WD40
					ZM

Figura 6. Dominios identificados exclusivos de proteínas de gasterópodos y bivalvos, respectivamente. Identificación mediante herramienta SMART (<http://smart.embl-heidelberg.de>) y/o con la herramienta de predicción integrada al NCBI (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>).

Se identificaron 42 familias de proteínas en gasterópodos y 92 en bivalvos, con 22 familias comunes a ambas clases. Los nombres de muchas familias identificadas coinciden con los nombres de los dominios detectados en SMPs de gasterópodos y bivalvos. Por ejemplo: Proton_antipo_N (PF00662) y _M (PF00361), COX1 (PF00115), 2 (PF00116) y 3 (PF00510), vWA (PF00092), entre otras. Las familias comunes y exclusivas de cada clase se muestran en la Figura 7.

et al., 2016) que permitía utilizar hasta ocho listados de datos (los criterios de selección se especifican en la sección 6.2.2). En las Figura 8 y Figura 9 se representaron 16 dominios y 2 regiones (LCR y TR) de gasterópodos y 7 dominios, 2 regiones (LCR y TR) y 1 motivo de bivalvos.

La Figura 8 representa el diagrama de Venn generado entre 16 dominios y 2 regiones identificadas en gasterópodos. Como se puede observar, las regiones más observadas fueron: LCR con presencia en 15 especies y TR con presencia en 10 especies; seguidos de los dominios: COX1 (PF00115) y COX2 (PF00116) identificados en nueve especies; y el conjunto de dominios ATP-synt_A (PF00119), COX2_TM (PF02790), COX3 (PF00510), Cytochrom_B (PF00033), Cytochrom_B_C (PF00032), Cytochrom_B_N_2 (PF13631), NADH5_C (PF06455), NADHdh (PF00146), Oxidored_q2 (PF00420), Oxidored_q4 (PF00507), Oxidored_q5_N (PF01059), Proton_antipo_M (PF00361) y Proton_antipo_N (PF00662) en ocho especies. El resto de los dominios identificados en SMPs de la base de datos, se detectaron en menos de seis especies (algunos tuvieron alta representatividad en la misma secuencia o en diferente secuencia, pero en una misma especie, por lo que quedaron excluidos).

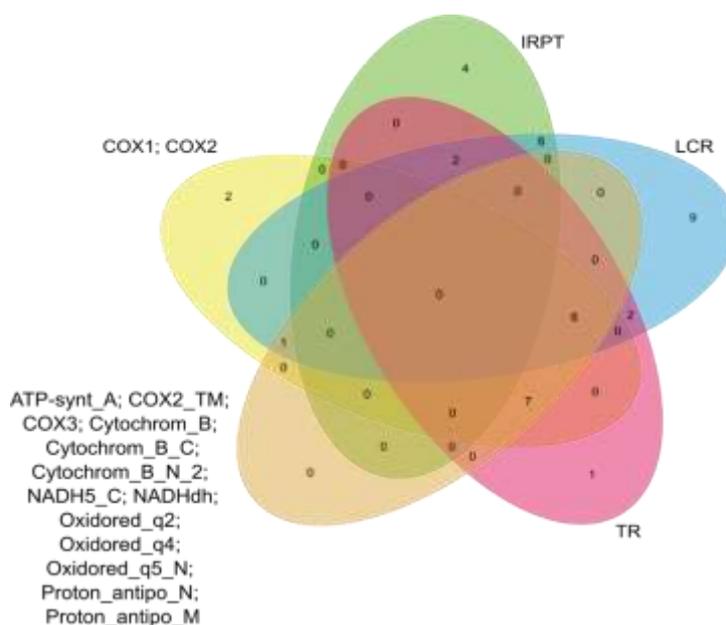


Figura 8. Diagrama de Venn representando dominios/regiones conservados de SMPs de la base de datos de gasterópodos. El diagrama muestra el número de especies en común donde fueran detectados dichos dominios/regiones. Los colores representan cada dominio/regiones o grupo de dominios/regiones, mientras que los números representan el número de especies en los que se detectaron dichos dominios/regiones; a su vez los números en las intersecciones representan la cantidad de especies en común en las que se presentan los dominios/regiones de la subsecuente intersección. Diagrama generado con: VennPainter versión 1.2.0 (Lin *et al.*, 2016).

La Figura 9 representa dos diagramas de Venn generados entre 7 dominios, 2 regiones y un motivo identificados en diferentes especies de las SMPs de la base de datos de bivalvos. Como se puede observar, las regiones más representadas entre SMPs de bivalvos fueron las regiones TR y LCR, detectadas en 26 y 29 especies respectivamente (detectadas en casi la totalidad de las especies de la base de datos de bivalvos) (Figura 9B). El motivo *Coiled coil region* fue la segunda más representada, encontrándose en 12 especies, mientras que los restantes oscilaron entre estar presentes en siete u ocho especies de 32 especies estudiadas (Figura 9A).

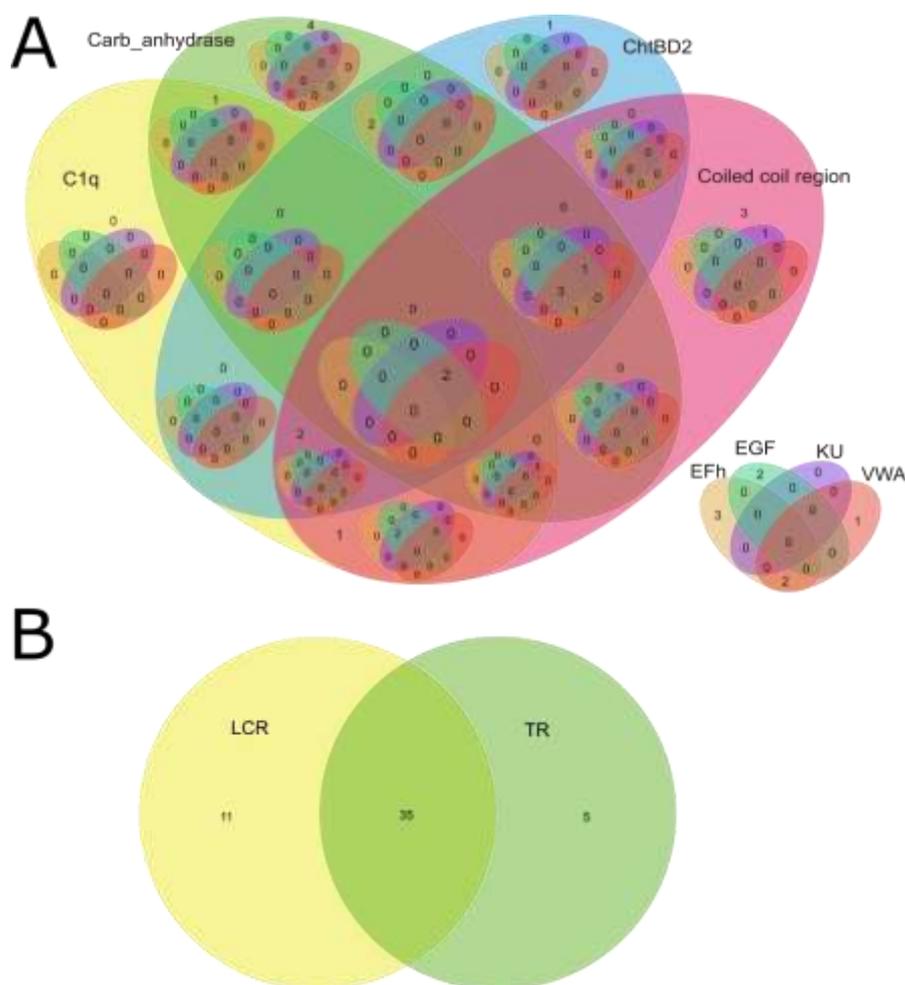


Figura 9. Diagrama de Venn anidado representando los dominios/regiones conservados de SMPs de la base de datos de bivalvos. El diagrama muestra el número de especies en común donde fueran detectados dichos dominios. A: Dominios y motivo identificados entre seis u ocho especies y su relación. B: Regiones identificadas en numerosas especies. Los colores representan cada dominio o grupo de dominios, mientras que los números representan el número de especies en los que se detectaron dichos dominios; a su vez los números en las intersecciones representan la cantidad de especies en común en las que se presentan los dominios de la subsecuente intersección. Diagrama generado con: VennPainter versión 1.2.0 (Lin *et al.*, 2016).

7.2.2.2 Diagramas de cuerdas para análisis cualitativo y cuantitativo

Los diagramas de cuerdas de las Figura 10 y Figura 11, representan dos conjuntos de datos, por un lado, las especies contenidas en las bases de datos de gasterópodos y bivalvos y por otro los dominios/regiones identificados en dichas especies, representando por tanto la relación entre ambos. A continuación, se analizan los diagramas atendiendo al comportamiento de los dominios/regiones y las especies de las bases de datos.

Como se observa en la Figura 10, en la base de datos SMPs de gasterópodos las LCR fueron las regiones más representadas en las secuencias de SMPs. Las LCR con 111 repeticiones para el 20% de todos los dominios detectados, se observaron en 15 especies (47% del total de especies) (entiéndase por repeticiones el número total de LCR detectados independientemente de que se detectara múltiples veces en una misma secuencia proteica o en secuencias de una misma especie). La segunda región más representada fue TR, con 96 repeticiones para un 17% de todas las regiones encontrándose múltiples veces en 10 especies de gasterópodos (31%). El tercer dominio mejor representado fue el dominio Proton_antipo_M, con 36 repeticiones en ocho especies (25%). Los siguientes dominios son representatividad fueron: ATP-synt_A (PF00119), COX1 (PF00115), COX2 (PF00116), COX2_TM PF02790), COX3 (PF00510), Cytochrom_B (PF00033), Cytochrom_B_C (PF00032), Cytochrom_B_N_2 (PF13631), NADH5_C (PF06455), NADHdh (PF00146), Oxidored_q2 (PF00420), Oxidored_q5_N PF01059) y Proton_antipo_N (PF00662) (atendiendo a número de repeticiones y distribución entre las especies). El dominio Oxidored_q3 (PF00499) aunque con un alto número de repeticiones, se concentró en solo dos especies (*Littorina saxatilis* y *Amalda northlandica*).

Si se observa, por otro lado, las SMPs de las especies contenidas en la base de datos, *Lottia gigantea* (50 secuencias) es la especie más común, con una amplia variedad de dominios (14 dominios con 112 repeticiones). Su amplia variedad de dominios está asociada probablemente a su dominancia numérica dentro de la base de datos, lo que aumenta la probabilidad de detección. Las siguientes especies mejor representadas en cuanto a dominios y su cantidad, fueron: *Amalda northlandica*, *Oxymeris dimidiata*, *Penion maximus*, *P. sulcatus* y *Reishia clavigera*, con más de 15 dominios detectados (algunos en múltiples ocasiones).

De las 32 especies de gasterópodos contenidas en la base de datos de SMPs, y atendiendo a la distribución del número de secuencias entre dichas especies, ningún dominio conservado parece mostrar prevalencia en todas las especies o se manifiesta una representación variada y extendida en todas las especies de los diferentes dominios. Los únicos dominios y regiones que parecieran constituir la batería básica son: LCR, TR y en menos proporción, Proton_antipo_M (PF00361).

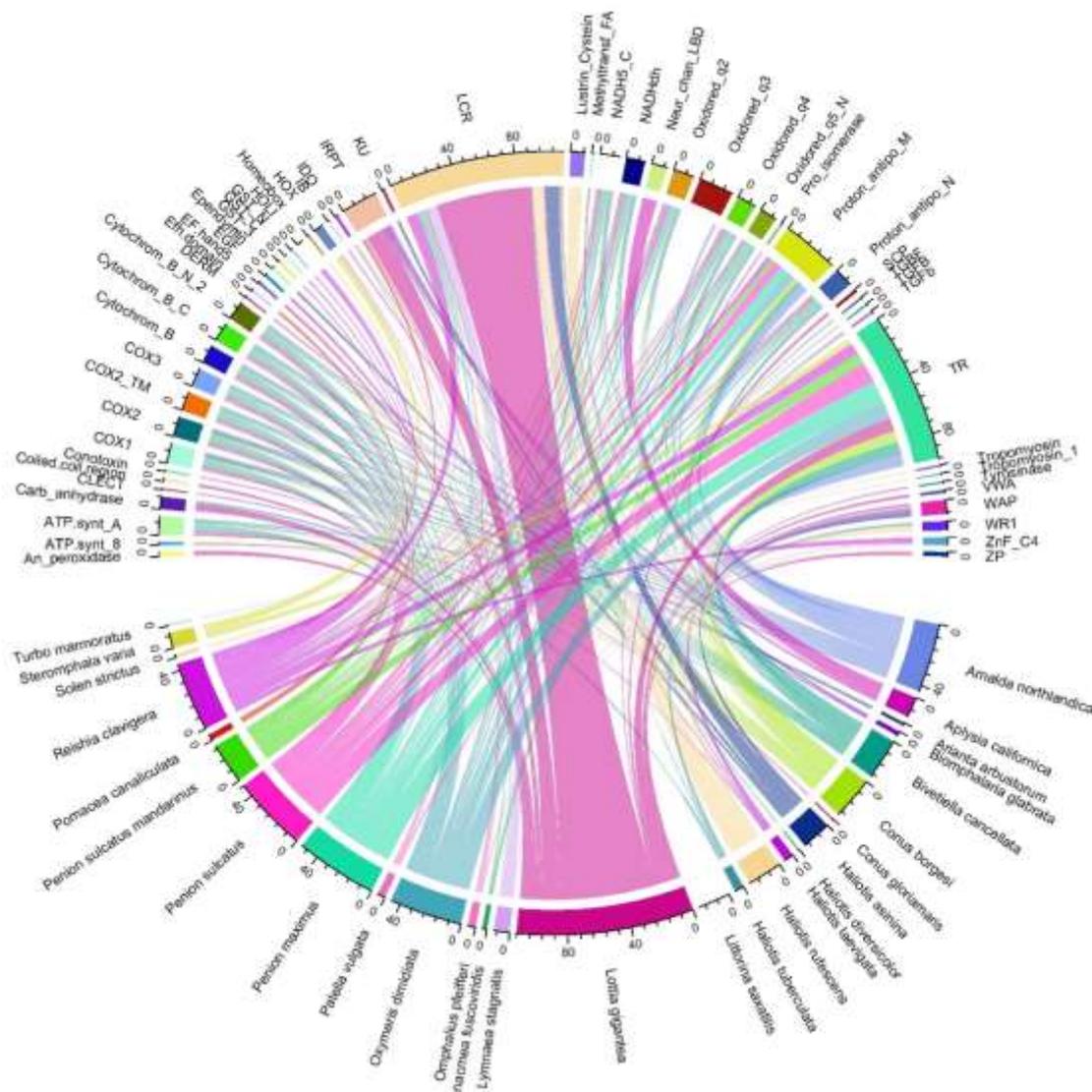


Figura 10. Diagrama de cuerdas de SMPs de gasterópodos. Las líneas individuales del ideograma enlazan dominios/regiones presentes en las secuencias de SMPs de diferentes especies de gasterópodos. ZP: Zona pellucida; LCR: Low complexity regions. TR: Transmembrane region. WAP: Whey Acidic Protein. vWA: von Willebrand Factor A; KU: Kunitz; TGFB: Transforming growth factor-beta; CLECT: C-type lectin; EGF: Epidermal growth factor. La identificación de dominios/regiones se realizó mediante la herramienta SMART (<http://smart.embl-heidelberg.de>) y/o con la herramienta de predicción integrada al NCBI (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>). Diagrama generado con la librería circlyze (Gu et al., 2014) en R Studio versión 4.0.1.

En la Figura 11, se observa, que el número de dominios y regiones conservadas en bivalvos es casi tres veces mayor que los identificados en la base de datos de gasterópodos (esta base de datos es casi cuatro veces mayor en número de secuencias). Sin embargo, aunque el diagrama de cuerdas (Figura 11), posee una distribución referente a los dominios y regiones, las LCR y TR destacan por presentarse en múltiples especies, al igual que en gasterópodos. Las LCR de bivalvos, se encuentran en el 80.5% de las especies de la base de datos (29 especies) y concentran el 50% de todos los dominios y regiones detectados (1,376 repeticiones). Por otro lado, las TR se encuentran en un 72% de las especies (26) y representan el 19% del total de los dominios/regiones descritos. Los siguientes dominios más abundantes son: C1q (PF00386), *Carbonic anhydrase* (PF00194), ChtBD2 (PF01607), EFh (PF00036), EGF (PF00008), Kunitz_BPTI (PF00014) y vWA (PF00092) y el motivo *Coiled coil region* representan del 1-2% de los dominios y regiones descritos (12-33% de las especies). En cuanto a las especies, *P. fucata* ocupa una gran porción de la base de datos con aproximadamente 500 secuencias y amplia variedad de dominios asociada a su dominancia numérica. *C. gigas* y *P. margaritifera* son las siguientes especies con más secuencias y también muestran variedad de dominios.

Como se analizó anteriormente, algunos dominios y regiones identificados parecen prevalecer en presencia y abundancia dentro de las especies de gasterópodos (regiones LCR y TR) y bivalvos (regiones LCR y TR y dominios C1q, *Carbonic anhydrase*, ChtBD2, EFh, EGF, KU, vWA y motivos *Coiled coil region*).

7.2.2.3 Cladogramas

En las Figuras 12-15 se muestran cladogramas para SMPs seleccionadas de acuerdo a la abundancia de información en las bases de datos. En la Figura 12 se muestra las estructuras de las SMPs de la base de datos de bivalvos con dominios de unión a quitina o ChtBD2 (PF01607), que incluyen a las proteínas tipo Pif97 y Pif177, BMSP, así como otras proteínas de la matriz orgánica de la concha. Las SMPs muestran además diferentes combinaciones de dominios; este comportamiento es común en SMPs como Pif donde dominios ChtBD2 se encuentran en combinación a dominios vWA (PF00092). Por ejemplo, Pif97 (una de las subunidades de Pif) posee dos dominios ChtBD2 en y un dominio vWA en *P. fucata*, mientras que en *C. gigas* posee uno de cada dominio (Wang *et al.*, 2013). En la Figura 12 se observa, como Pif97 y Pif177 de *P. margaritifera* poseen similares estructuras, composición, longitud y posicionamiento de dominios, al ser SMPs de la misma especie.

En la Figura 13 se muestra el cladograma y estructura de SMPs con el dominio anhidrasa carbónica (PF00194), como nacreínas, N66 y anhidrasas carbónicas. Las nacreínas, son una de las SMPs más estudiadas y de las primeras en determinarse su actividad enzimática. Poseen dominios anhidrasa carbónica y dominios repetidas del tipo GlyXAsn o GlyAsn, en dependencia de la SMPs y especie en particular (Marin *et al.*, 2007); siendo similares por tanto a las anhidrasas carbónicas. Se distingue en este cladograma (Figura 13) dos grupos fundamentales, por un lado, las nacreínas y proteínas similares a nacreínas, y por otro las anhidrasas carbónicas, precursores de nacreínas y proteínas similares a nacreínas, las cuales engloban a las primeras. En la Figura 13B, se distingue la estructura de dichas secuencias, las cuales suelen presentar péptido señal con un dominio anhidrasa carbónica que ocupa gran parte de esta, difiriendo en la longitud de las proteínas, pero a simple vista parecieran muy similares. Los alineamientos de estas secuencias arrojan homología entre los dominios anhidrasa carbónica, atendiendo a la extensión de estos, observando secuencias conservadas hacia el inicio y final de las proteínas, con una región intermedia rica en repeticiones AnsGly o GlyXAsn.

En la Figura 14 se muestra el cladograma de proteínas con dominios del tipo Kunitz (PF00014), que como se puede observar agrupa proteínas del tipo inhibidores de proteasas del tipo serina. Las secuencias oscilan entre 150 y 250 aa, y suelen presentar dos dominios

Kunitz. Como se observa en Figura 14A, grupo de proteasas tipo serina y BPTI/Kunitz SMPs de *P. margaritifera* principalmente, que se forma con un alto porcentaje de bootstrap, incluyendo a las secuencias más cortas y con un único dominio Kunitz. Los siguientes grupos de parentesco, se crean a partir de la presencia de dos dominios y sus agrupamientos tienen menor porcentaje de *bootstrap* debido a la presencia de regiones de baja complejidad.

En la Figura 15, por otro lado, se muestra el cladograma de proteínas con dominios del tipo vWA (PF00092), siendo proteínas del tipo Pif97, Pif177, BMSP y perlucin, entre otras. Se observa en este análisis la presencia de proteínas Pif, de las cuales se conoce la presencia de dominios vWA y ChtBD2 en este tipo de SMPs, como se describió anteriormente. Las proteínas Pif se agrupan juntas en dos ramas principales, una para las Pif97 de *C. gigas* y otra para cuyo tamaño de secuencias es menor, y no presenta otros dominios, mientras que las Pif de *P. margaritifera*, cuyas secuencias son más variables en cuanto a composición y poseen mayor longitud se observan más cercanas a BMSP y otras SMPs (proteínas similares al colágeno). Una vez más el nivel de homología entre las secuencias disminuye con la introducción de las regiones de baja complejidad.

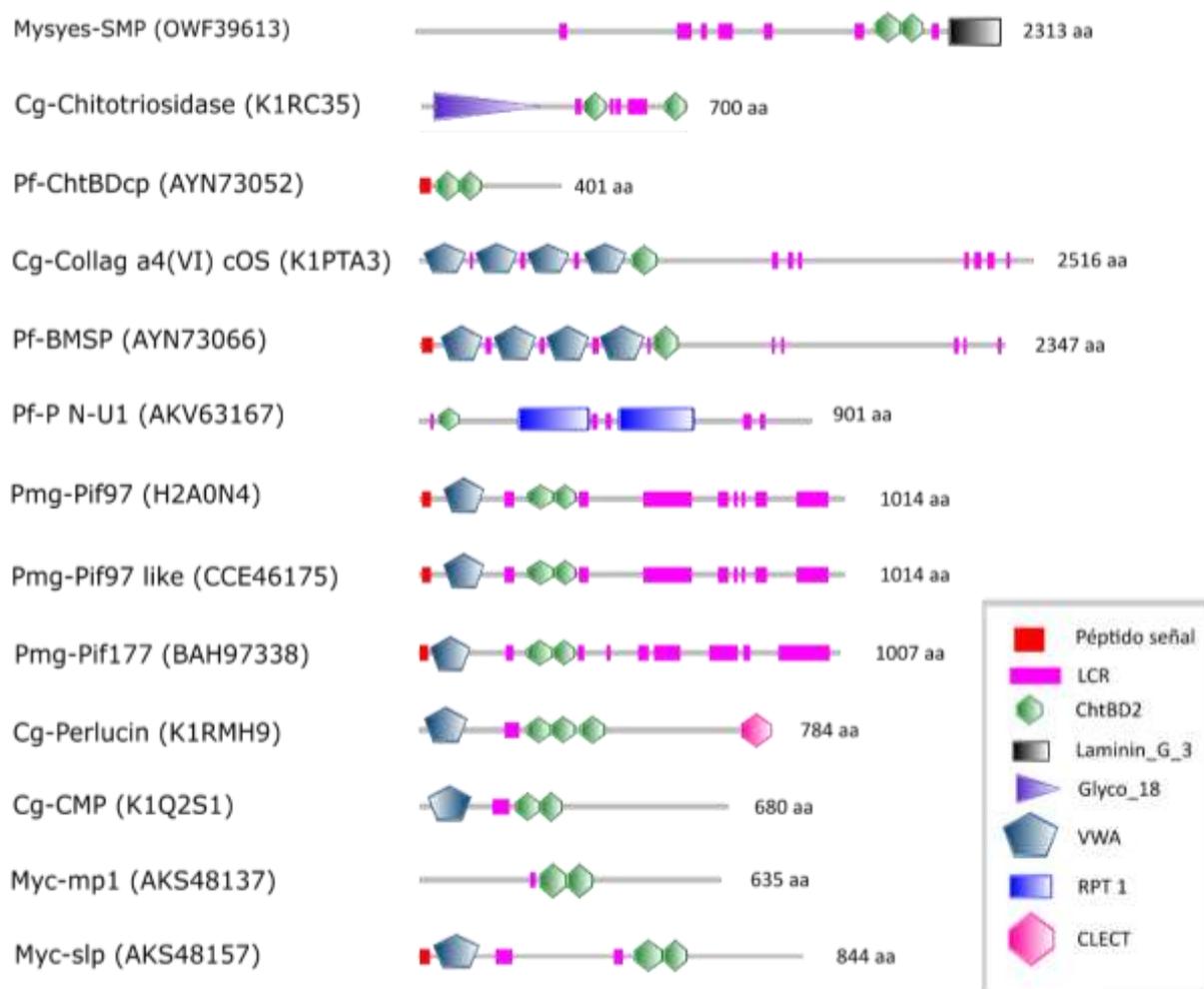
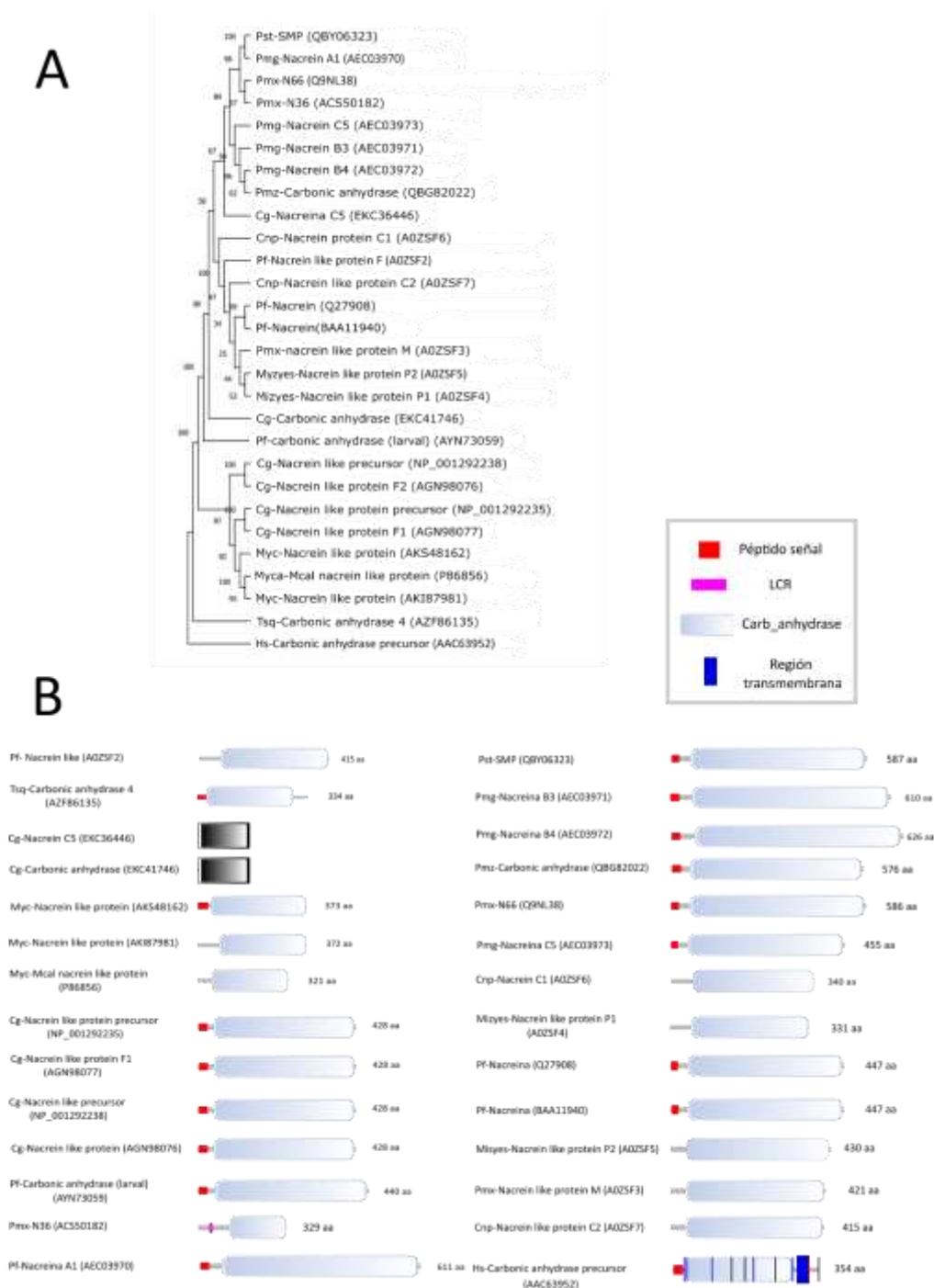


Figura 12. Comparación estructural de SMPs que intervienen en la biomineralización de bivalvos, que poseen dominios del tipo ChtBD2 (dominio de unión a quitina). Dominios conservados identificados con SMART (<http://smart.embl-heidelberg.de>) y/o con la herramienta de predicción integrada al NCBI (<https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>); obsérvense los dominios ChtBD2 en las diferentes secuencias. SMP: *Shell matrix protein*; Collag a4: *collagen alpha 4*; CMP: *Cartilage matrix protein*; mp1: *matrix protein 1*; slp: *sushi like protein*; Cg: *Crassostrea gigas*; Pf: *Pinctada fucata*; Pmg: *P. margaritifera*; Mysyes: *Mizuhopecten yessoensis*; Myc: *Mytilus coruscus*. Figura editada con Inkscape versión 1.1.1. Autora: Ivis Laura Méndez Bernal.



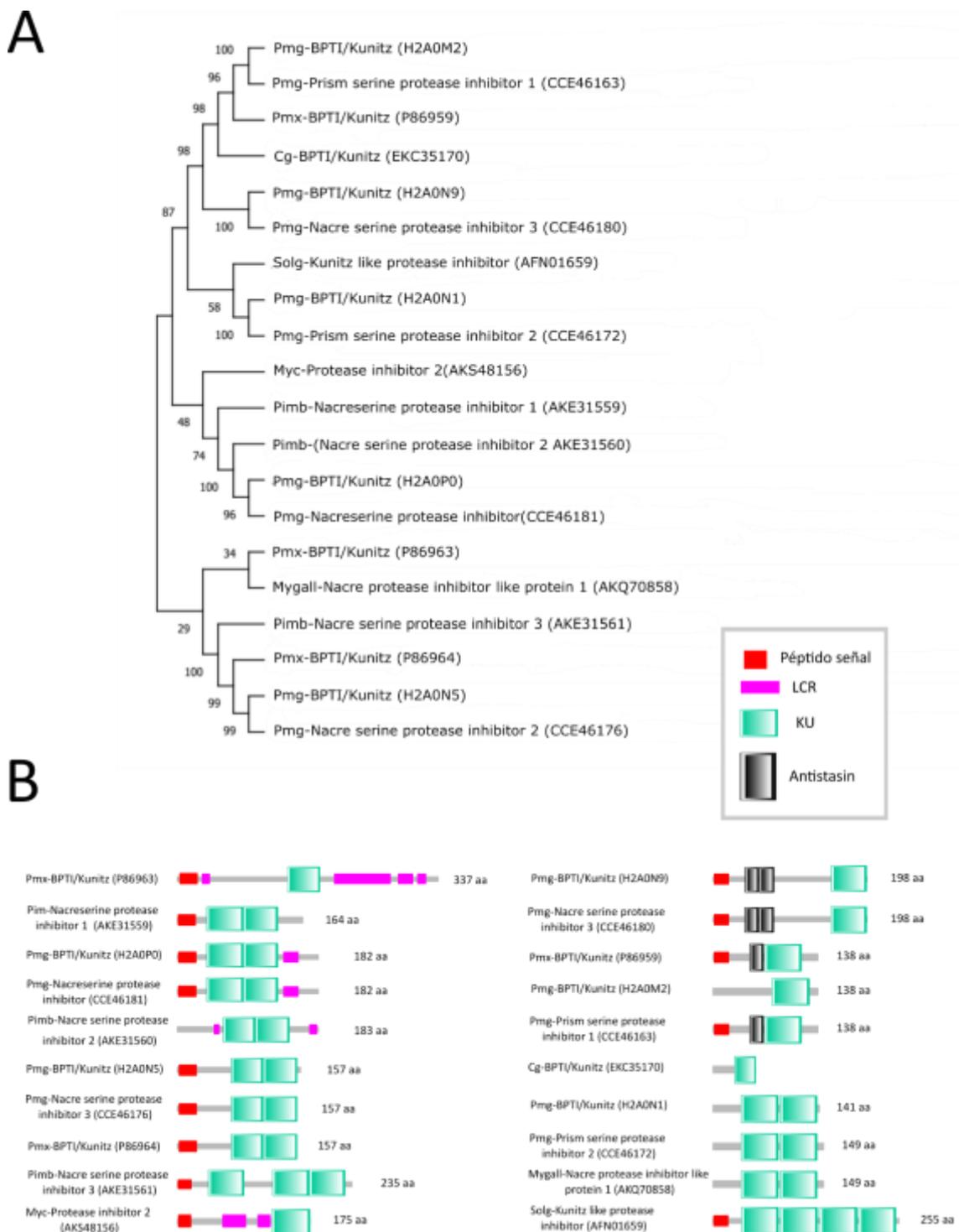
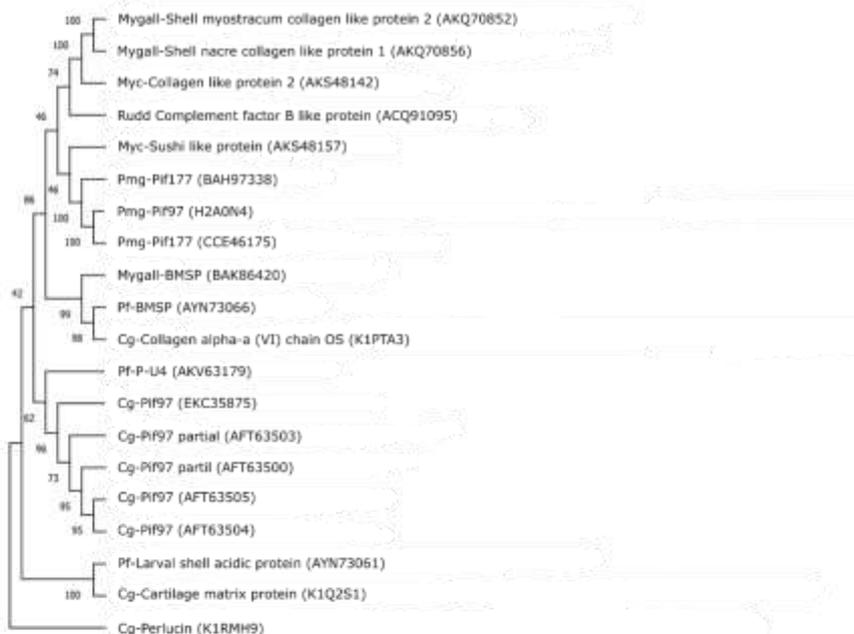


Figura 14. Comparación de SMPs con dominio del tipo Kunitz_BPTI que intervienen en la biomineralización de bivalvos. **A:** Cladograma empleando el método de Neighbor-Joining (Saitou y Nei, 1987) con Mega versión 11 (Tamura et al., 2021), *bootstrap* (10,000 réplicas) (Felsenstein, 1985). Este análisis involucró 20 secuencias de aminoácidos. **B:** Dominios conservados presentes en proteínas de bivalvos; obsérvense los dominios KU en las diferentes secuencias. Cg: *Crassostrea gigas*; Pf: *Pinctada fucata*; Pmg: *P. margaritifera*; Pmx: *P. maxima*; Pimb: *P. imbricata*; Myc: *Mytilus coruscus*; Mygall: *Mytilus galloprovincialis*. Solg: *Solen grandis*. Largo de secuencias no especificada para secuencias genómicas. Figura editada con Inkscape versión 1.1.1.

A



B

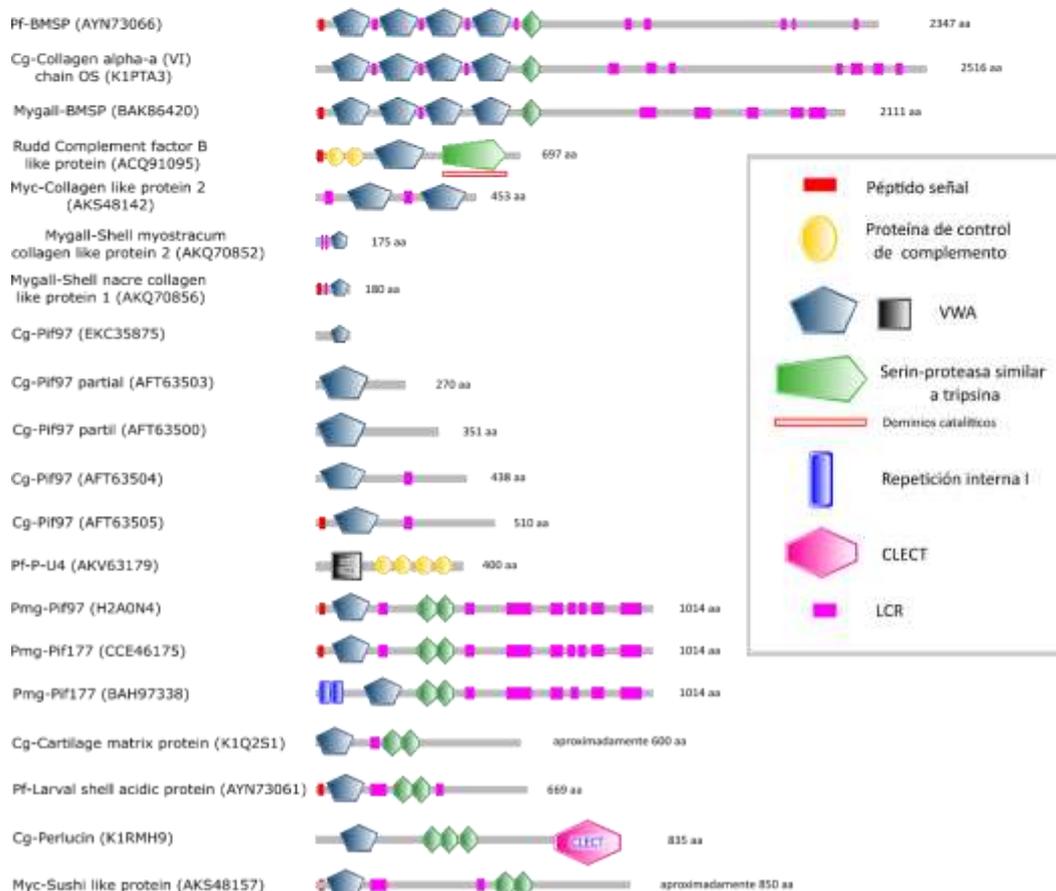


Figura 15. Comparación de SMPs con dominios del tipo vWA. **A:** Cladograma realizado por el método de Neighbor-Joining (Saitou y Nei, 1987) con Mega versión 11 (Tamura *et al.*, 2021), *bootstrap* (10,000 réplicas) (Felsenstein, 1985). Este análisis involucró 20 secuencias de aminoácidos. Cg: *Crassostrea gigas*; Pf: *Pinctada fucata*; Pmg: *P. margaritifera*; Myc: *Mytilus coruscus*; Mygall: *Mytilus galloprovincialis*; Rudd: *Rudipates desucatus*. Figura editada con Inkscape versión 1.1.1.

De acuerdo con los análisis anteriores, no se identifica claramente una batería básica de SMPs involucrada en la biomineralización para moluscos. Sin embargo, si fue posible identificar dominios o regiones comunes en las SMPs de bivalvos, de las cuales se identificaron las regiones LCR, TR; los dominios C1q, *Carbonic anhydrase*, ChtBD2, Efh, EGF, KU y vWA y el motivo *Coiled coil region*. Para el caso de gasterópodos, se identificaron las regiones LCR y TR, además del dominio vWA y el motivo *coiled coil region*. Por lo que, como parte de la batería básica de moluscos, bivalvos y gasterópodos, se tiene hasta ahora que puede incluir SMPs con dominios vWA, regiones LCR y TR y motivos *Coiled coil region*, los cuales se encontraron en ambas bases de datos (bivalvos y gasterópodos).

7.3 Caracterización de las proteínas y su relación con la microestructura de la concha

Después de analizar la distribución de dominios/regiones identificadas en las SMPs entre las especies de la base de datos de gasterópodos y bivalvos, se describen cuáles y por qué constituyen la batería básica responsable de la biomineralización en estos taxa.

7.3.1 Propuesta de batería básica de bivalvos y descripción de los dominios constituyentes

Después de analizar la base de datos de gasterópodos, atendiendo a que esta base de datos solo se encuentra formada por 301 secuencias de las cuales 177 constituyen proteínas bioenergéticas, distribuidas entre 32 especies, donde la mayor representación de proteínas la contiene la especie *Lo gigantea*, y después de observar que los dominios estructurales comunes a varias especies (dígase un mínimo de ocho especies: *L. gigantea*, *Amalda noirthlandica*, *Bivitiella cancellata*, *Conus borguesi*, *Penion maximus*, *Penion sulcatus*, *Reishia clavigera*, etc.), se consideró que la información no fue suficiente para determinar de forma eficiente una batería básica para esta clase taxonómica. A pesar de esto, se detectó que las LCR y TR se destacaron entre las proteínas no energéticas, lo que sugiere cierta relevancia que debería profundizarse en estudios posteriores. Se recomienda ampliar los criterios de búsqueda de proteínas de gasterópodos en bases de datos científicas para profundizar el análisis sobre las proteínas de la concha de moluscos gasterópodos.

Por otro lado, se determinó que la batería específica de SMPs involucrada en la biomineralización de bivalvos está probablemente integrada por proteínas con dominios:

C1q, anhidrasa carbónica, ChtBD2, Efh, EGF, Kunitz y vWA, las regiones LCR y TR y el motivo *Coiled coil region*. A continuación, se describe la función asociada a estos:

- 1) Los dominios C1q se encuentran en muchas proteínas tipo colágeno y en complementos C1q entre otros. Se sintetiza en células productoras de colágeno e interviene en funciones de plegamiento y alineamiento correcto de la doble hélice de ADN, así como en eventos de reconocimiento proteína-proteína y plegamiento de proteínas (<http://smart.embl-heidelberg.de>).
- 2) Los dominios anhidrasa carbónica intervienen en la catalización inversa de CO₂ según a ecuación $\text{CO}_2 + \text{H}_2\text{O} \leftrightarrow \text{HCO}_3^- + \text{H}^+$, lo que permite la precipitación en carbonato de calcio, a través de la reacción del bicarbonato con los iones Ca²⁺ (Marin *et al.*, 2012). Su función en la biomineralización ha sido asociada a la calcificación y como regulador negativo de este proceso. Además de intervenir en la regulación del pH en el microambiente inmediato al sitio de biomineralización (le Pabic *et al.*, 2017).
- 3) Los dominios ChtBD2 intervienen en la biomineralización, creando el molde plantilla hidrofóbico, dentro del cual crecerá el mineral. Este tipo de dominios interactúa con la quitina, el carbohidrato mejor representado en la matriz orgánica (Liao *et al.*, 2019). Además, interviene en la unión a los compuestos de mineral en precipitación y/o en mecanismos de reparación de la concha (le Pabic *et al.*, 2017).
- 4) Los dominios vWA, son dominios de unión a carbohidratos, usualmente vinculados a proteínas extracelulares (ancestrales). Su rango de funciones en diferentes proteínas incluye: adhesión celular, migración, formación de patrones, señalización, transducción, interacción con ligandos, etc. Se conoce su interacción con el carbohidrato quitina durante la biomineralización de los moluscos, interviniendo en el control de la cristalización de CaCO₃. Ambos dominios, vWA y ChtBD2 suelen encontrarse unidos en proteínas del tipo Pif, por ejemplo, (Wang *et al.*, 2013) compararon la proteína Pif97 de *C. gigas* y *P. fucata*, encontrando en ambas este tipo de dominios (Pif97: 1vWA + 2ChTBD2; cgPif97: 1vWA + 1ChtBD2). Además, en las Figuras 12 y 15, se observa que esta combinación de dominios no es exclusiva de las

proteínas Pif o de estas especies. En un estudio realizado por (Liu y Zhang, 2021a) se reportó que uno de los dominios más abundantes en SMPs de invertebrados es el vWA, el cual suele encontrarse combinado a dominios de unión a quitina, entre otros. Los dominios vWA a nivel de biomineralización, probablemente actúen en cuestiones de interacción proteína-proteína, formando un complejo de proteínas el enrejado de matriz orgánica) a partir del cual se generará la concha.

- 5) *Coiled coil* son motivos estructurales comunes, formado por aproximadamente el 3-5 % de todos los aminoácidos en las proteínas. Son motivos peptídicos repetitivos ubicuos presentes en todos los dominios de la vida (Truebestein y Leonard, 2016). Generalmente constan de dos a cinco hélices α envueltas entre sí en una hélice levógira para formar un superenrollamiento. Mientras que las hélices α regulares pasan por 3,6 residuos por cada vuelta completa de la hélice, la distorsión impuesta sobre cada hélice dentro de una bobina enrollada a la izquierda reduce este valor a alrededor de 3,5. Así, una repetición de heptada ocurre cada dos vueltas de la hélice; se encuentran entre las estructuras más importantes que participan en interacciones cruciales como el control transcripcional, crecimiento celular y proliferación (Mason y Arndt, 2004); su estructura y posicionamiento específico en la secuencias de aminoácidos ejecuta funciones como influenciar la estructura de organelos, anclar vesículas de transporte, propagar cambios conformacionales, funcionar como reglas moleculares en enzimas posicionando la actividad catalítica a distancias específicas (Truebestein y Leonard, 2016).
- 6) El motivo Efh es asociado a la unión a Ca^{2+} que se encuentran regularmente en pares. Estos motivos sufren cambios conformacionales una vez unidos al Ca^{2+} que conducen a la activación o inactivación de proteínas blanco ensemble.
- 7) Los dominios EGF están formados por un péptido corto (50 aminoácidos) con un motivo específico de seis Cys que forma tres puentes disulfuro. Se encuentra en múltiples proteínas de un amplio rango de funciones: suele encontrarse en proteínas de membrana, interviniendo en cascadas de señalización <http://smart.embl-heildeberg.de>; interviene en reconocimiento proteína-proteína, agregación de

proteínas, señalización molecular o capacidad de unión Ca^{2+} (Marin *et al.*, 2012); en proteínas extracelulares se relaciona a funciones inmunes (Liu *et al.*, 2015). Junto al dominio vWA, es de los dominios más comunes dentro de los biominerales de invertebrados (Liu y Zhang, 2021a). Un estudio realizado por (Liu y Zhang, 2021b) sobre las SMPs reveló que los dominios EGF y vWA en 12 de 17 biominerales estudiados, junto a otros dominios parecer ser de los más comunes dentro de las SMPs, así como los más abundantes dentro de los proteomas. Una vez más sugiriendo la relevancia de los dominios aquí mencionados.

- 8) Los dominios Kunitz pertenecen a la familia de inhibidores de serinoproteasas. Suelen ser proteínas cortas de aproximadamente 50 aa con pocas estructuras secundarias (<http://smart.embl-heidelberg.de>). En la biomineralización, este tipo de dominios interviene en la formación, mantenimiento y regulación del biomineral, así como en la defensa contra depredadores (Arivalagan *et al.*, 2016), y defensa contra proteinasas procedentes de microorganismos en el microambiente del sitio de mineralización que pueden degradar las SMPs (le Pabic *et al.*, 2017).
- 9) Las LCR son regiones estructuralmente desordenadas e inestables. Poseen baja afinidad hacia otras macromoléculas (como proteínas y polisacáridos) pero se unen débilmente a superficies minerales e iones en ambientes acuosos, como es el caso de la cavidad extrapaleal (Marin *et al.*, 2012). Estos dominios suelen presentar una alta representación de ciertos aminoácidos como Glu, Asp, Gly, Ala, Met, Arg, Thr y Ser (Liao *et al.*, 2019). Aportan flexibilidad, facilitan la migración transmembranal, especificidad de reconocimiento y rápida evolución (Arivalagan *et al.*, 2016).
- 10) Las regiones TR, son regiones que atraviesan la bicapa lipídica con diferentes funciones como transporte o funciones de reconocimiento y señalización.

Como se puede observar en la Figura 16, predecir una batería básica proteica responsable de la biomineralización de los moluscos como filo fue imposible, debido a factores como el tamaño de la base de datos de gasterópodos y la representación de proteínas involucradas en la biomineralización. Sin embargo, se predijo la presencia de regiones LCR y TR en

proteínas de ambas clases taxonómicas, lo que sugiere relevancia de estos dentro del filo y debería estudiarse cuidadosamente en estudios futuros. Es importante destacar que las regiones LCR no son las mismas en todas las proteínas, y que, de hecho, su alta variabilidad, y la diversidad de proteínas que las presentan impide relacionarlas filogenéticamente y definir regiones específicas para ubicarlas en una batería básica.

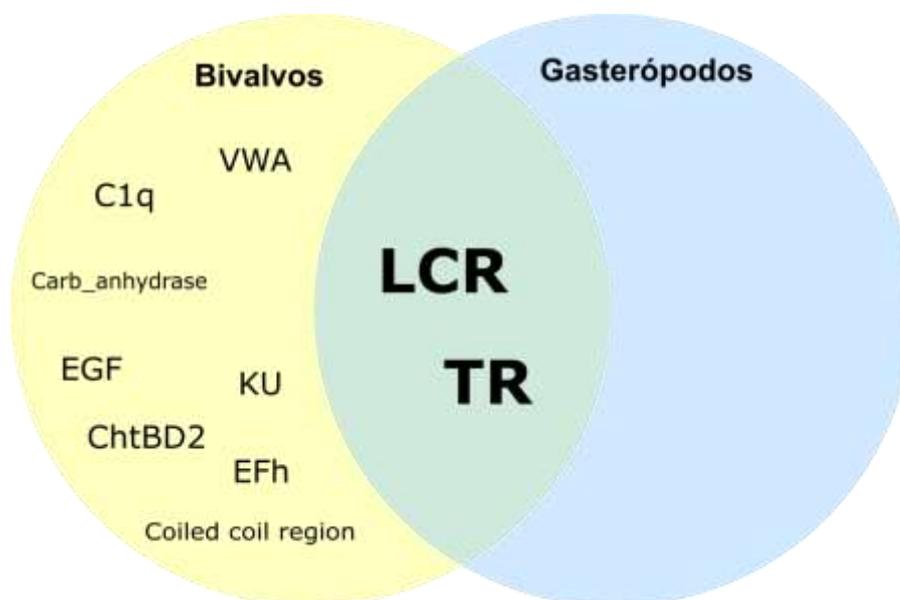


Figura 16. Diagrama de Venn que representa las posibles baterías básicas de SMPs involucradas en la biomineralización de moluscos gasterópodos y bivalvos y la relación entre estas. Como se puede observar, no se determinaron SMPs taxa específicos para gasterópodos, sin embargo, las LCR y TR mostraron relevancia y probablemente integren la batería básica de esta clase taxonómica. La posible batería básica de SMPs responsable de la biomineralización de moluscos podría integrarse por LCR y TR. Carb_anhydrase: anhidrasa carbónica; ChtBD2: dominios de unión a quitina (del inglés: *chitin binding domain 2*); EFh: dominio EF-hand; EGF: factor de crecimiento epidérmico (del inglés: *epidermic growth factor*); KU: dominios Kunitz; LCR: Regiones de baja complejidad (del inglés: *low complexity regions*); TR: regiones transmembranales (del inglés: *transmembrane regions*); vWA: von Willebrand Factor A.

7.3.2 Microestructura y batería básica

La imposibilidad para detectar una batería básica de SMPs en gasterópodos impidió relacionar las diferencias entre el conjunto de proteínas esenciales de ambas clases taxonómicas con las microestructuras de la concha. A pesar de las dificultades presentadas, se observan regiones comunes repetidos en ambas clases taxonómicas, dígame regiones LCR y TR (Figura 16).

8. DISCUSIÓN

En la presente investigación, se analizaron cerca de 1,500 secuencias proteicas de 68 especies de gasterópodos y bivalvos marinos y/o de otros hábitats con la intención de identificar las principales proteínas involucradas en la biomineralización de ambas clases taxonómicas y su relación con la microestructura de sus conchas.

8.1 Construcción de una base de datos de secuencias proteicas de bivalvos y gasterópodos

Las bases de datos de gasterópodos y bivalvos recopilaron las secuencias proteicas resultantes de la búsqueda en diferentes plataformas, disponibles en el momento de la realización de dicha búsqueda (en la actualidad este número puede ser diferente). Durante este proceso fueron eliminadas secuencias parciales para evitar la introducción de errores en los análisis comparativos. La disponibilidad de secuencias proteicas en bases de datos, gracias a estudios de proteómica, transcriptómica y genómica permitió el estudio profundo de SMPs de moluscos y parece prometedor en el mejor entendimiento del proceso de biomineralización de moluscos. Sin embargo, mientras aumenta la disponibilidad de información se plantean nuevas preguntas que sugieren que el conocimiento sobre este proceso es aún incipiente. No obstante, en este estudio se planteó por primera vez estudiar la extensión de proteínas descritas buscando identificar una batería básica de proteínas involucradas en la formación de la concha. Las bases de datos quedaron integradas por 301 y 2,189 proteínas de 32 especies de gasterópodos y 37 de bivalvos respectivamente, con variedad entre las mismas. Teniendo en cuenta las diferentes metodologías de extracción empleadas en la literatura científica, así como la disponibilidad de las secuencias, es posible que en este estudio se esté subestimado el número real de proteínas disponibles.

La generación de las bases de datos de gasterópodos y bivalvos utilizando palabras claves en la búsqueda en NCBI como fuente primaria de secuencias arrojó diferencias en la cantidad de información disponible, y las proteínas constituyentes finales de ambas bases fueron diferentes. La mayor disponibilidad de SMPs de bivalvos, confirma lo altamente estudiadas que son las SMPs de bivalvos, en comparación de gasterópodos. Por otro lado, se conoce que las SMPs son abundantes en regiones de baja complejidad, las cuales estadísticamente serán menos clivados por tripsina en las metodologías de extracción y caracterización

proteómicas, generando largos péptidos que son más difíciles de analizar bajo MS-MS (Karakostis *et al.*, 2016). Tomando en cuenta lo anterior, es probable que la proporción real de SMPs de moluscos sea superior, y este estudio este subestimando el número real de las mismas (hasta la fecha de creación de las bases de datos, enero de 2021).

8.2 Determinación de las proteínas de la batería básica de moluscos, gasterópodos y bivalvos y su relación con la microestructura de la concha

8.2.1 Caracterización de las secuencias de la base de datos

Las SMPs de ambas bases de datos fueron caracterizadas a través de herramientas bioinformáticas atendiendo a punto isoeléctrico y peso molecular, modificaciones postraduccionales del tipo glicosilaciones, presencia de puentes disulfuro, presencia de péptido señal, localización subcelular, dominios conservados y familias proteicas.

Las proteínas de ambas bases de datos se caracterizaron por presentar variados pI, en correspondencia con sus tamaños, peso molecular, composición de aminoácidos, estructura y función. Las SMPs se caracterizaron por presentar en su mayoría péptidos señales, lo que coincide con la otras SMPs ya descritas, así como el hallazgo de que la predicción de localización celular predominante sea extracelular secretada. Se observó, además, la elevada presencia de modificaciones postraduccionales de los tipos glicosilaciones y puentes disulfuro, y aunque no se conoce con exactitud la función que ejercen por sí solas, en esta caracterización de largo alcance se hace evidente su importancia.

La variada presencia de pI en las SMPs está relacionada con el amplio rango de acidez y basicidad de estas proteínas y con que este sistema representa una clasificación artificial asociada al proceso de extracción de las proteínas. Según Marin *et al.* (2007) y las SMPs poseen un pI según la cual se clasifican como ácidas o básicas, y su función específica en la biomineralización se vincula a la formación de una microestructura específica a nivel de las capas calcáreas de la concha. Por ejemplo, las proteínas cuyo pI varía entre muy ácidas y muy básicas ($4.5 < \text{pH} < 7$), inducen la formación del polimorfo de CaCO_3 calcita, mientras que las proteínas que varían entre ligeramente ácidas o básicas ($4.5 > \text{pH} < 7$), se asocian con el polimorfo aragonita (esta clasificación es artificial) (Marin *et al.*, 2007, 2012). Un estudio realizado por Monteaalgre *et al.* (2021; no publicado) demostró que las SMPs de

gasterópodos no cumplieran con este patrón descrito en la literatura científica, por tanto, esta clasificación es ambigua. Si se analiza al amplio espectro y variabilidad de las proteínas de la matriz orgánica de la concha de bivalvos, y las diferencias en el número de proteínas, composición, dominios, y otros factores aquí analizados, se puede esperar un resultado semejante en esta clase de moluscos, así como en el filo en general. Por otro lado, la variabilidad de pesos moleculares se asoció a los diferentes tipos de proteínas presentes en la matriz, así como a las modificaciones postraduccionales que sufren muchas de estas proteínas, que modifican su peso.

Hasta la fecha se han identificado numerosas SMPs de moluscos disponibles en bases de datos públicas, sin embargo, en comparación, la cantidad de proteínas caracterizadas *in vitro* para determinar sus propiedades bioquímicas (como modificaciones postraduccionales) y funcionalidad es sumamente inferior (Rivera-Pérez y Hernández-Saavedra, 2021); por lo que este estudio contribuyó a la caracterización de las SMPs de las bases de datos de bivalvos y gasterópodos disponibles en bases de datos públicas.

Las glicosilaciones han sido descritas en SMPs de diferentes especies y clases de moluscos, usualmente en proteínas ácidas (bajo pI, pH < 4.5) de la matriz soluble (ASM) y con alto contenido de aminoácidos del tipo Gly, Ser y Asp (aminoácidos ácidos). Por ejemplo, (Marie *et al.*, 2013) encontró dominancia de las glicoproteínas LUSP-23, LUSP-10 y LUSP-9 en la matriz orgánica del gasterópodo *L. gigantea*; estas proteínas poseen azúcares ácidas, son ricas en Asp y sus funciones se asocian con la formación de cristales de CaCO₃. Resultados similares se describieron para las proteínas del bivalvo *Union pictorum*: P50, P95 y P29; al realizar ensayos de desglicosilación química, generando cambios en los patrones de formación de cristales, demostrando su actividad de unión a Ca²⁺ (Marin *et al.*, 2007). En la especie de bivalvo *Patinopecten yessoensis*, la glicoproteína MSP-1 se ha descrito como uno de los principales componentes de la matriz soluble, esta proteína posee de tres a cinco sitios potenciales de N-glicosilación y su función está vinculada a nucleación/inhibición del crecimiento del biomineral (Sarashina y Endo, 2001). Por otro lado, en el cefalópodo *Nautilus macrocephalus*, se identificaron azúcares ácidos en la glicoproteína Nautilin-63 la cual no posee actividad de unión a Ca²⁺ ni similitudes con otras proteínas involucradas en biomineralización en otros moluscos (hasta julio de 2001).

La función exacta de las glicosilaciones como parte de las SMPs en la biomineralización aún no está completamente definida (Marin *et al.*, 2007; Rivera Pérez y Hernández-Saavedra, 2021) sin embargo, se han propuesto diferentes funciones para los azúcares en SMPs. Un experimento realizado por (Addadi *et al.*, 2006) demostró que los polisacáridos sulfatados y/o carboxilados de glicoproteínas de la ASM favorecen la concentración y nucleación de iones Ca^{2+} en cristales. Los azúcares ácidos de la ASM parecen concentrar los iones en la vecindad de las proteínas ácidas, induciendo la nucleación de los cristales, pero también los polisacáridos podrían ejercer funciones de reconocimiento superficial del mineral, selección de polimorfos, interacción manto-matriz y señalización celular (Marin *et al.*, 2007). Por otro lado, los puentes disulfuro se han descrito en varias SMPs de diferentes especies de moluscos. Por ejemplo, la Lustrina A del gasterópodo *Haliotis rufescens* clonada y caracterizada por (Shen *et al.*, 1997). Esta proteína posee una estructura modular globular de diez dominios ricos en Cys que la estabilizan a través de puentes disulfuro intra-dominios (posiblemente establezca puentes entre cadenas). Estos autores hipotetizaron la funcionalidad de esta proteína relacionada a interacción proteína-proteína, permitiendo la formación y autoensamblaje de la compleja matriz multiproteica que constituye la matriz orgánica. Una función similar, es ejecutada por la glicoproteína MPP1 de *Crassostrea nipponna* que posee en una estructura modular de dominios SG (ricos en Ser y Gly) repetidos e intercalados con Cys individuales; las Cys forman puentes disulfuro que crean polímeros con otras MPP1s, estableciendo un enrejado para la atracción y captura y de Ca^{2+} durante la mineralización (Samata *et al.*, 2008).

Lo anterior sugiere que los puentes disulfuro no solo participan en la estabilidad de las proteínas, sino también favorecen cambios conformacionales específicos para la funcionalidad de las proteínas, y su reversibilidad ayuda a controlar la estructura dinámica de las proteínas durante la nucleación mineral (Rivera Pérez y Hernández-Saavedra, 2021), sin embargo, se necesita profundizar sobre la función de estas, no tanto a nivel individual de las proteínas, sino general en la biomineralización. La presencia de glicosilaciones en este estudio, coincide con lo descrito por variados autores, resaltando la importancia de las mismas como parte de la estructura de las SMPs.

En cuanto a la caracterización de SMPs en cuanto a presencia de péptido señal y citolocalización, se determinó que la identificación de los péptidos en las proteínas estuvo relacionada a localizaciones celulares de tipo extracelular secretada y extracelular secretada de unión a membrana. Esto coincide con lo descrito por (Addadi *et al.*, 2006) y (Marin *et al.*, 2012) que plantean que las proteínas de la matriz orgánica de la concha son producidas y secretadas por las células más externas del manto (epitelio calcificante externo), que luego migran y se ensamblan originando el “enrejado” donde ocurrirá la nucleación y crecimiento de los cristales de CaCO_3 ; por lo que esta localización celular coincide con este comportamiento. Por otro lado, la dominancia de esta localización coincide con la presencia de péptidos señales (migración de proteínas) y la predicción de sitios de glicosilación son comunes en proteínas secretadas extracelularmente (según (Rivera-Pérez y Hernández-Saavedra, 2021).

Se describieron 56 dominios conservados en gasterópodos y 163 en bivalvos, para 42 y 92 familias de proteínas respectivamente. Los dominios conservados detectados se correspondieron con muchas de las familias proteicas predichas, y la diferencia en el número de dominios y familias se relacionó directamente con el tamaño de las bases de datos. Muchos de los dominios aquí identificados, fueron descritos previamente para SMPs como se muestra en la Tabla II y en la sección 2.3 SMPs de gasterópodos y bivalvos. Aunque estudios previos se han centrado en caracterizar proteomas de especies particulares (Bédouet, *et al.*, 2007; Marie *et al.*, 2010; Liao *et al.*, 2015, 2019; Liu *et al.*, 2015; le Pabic *et al.*, 2017; Mann *et al.*, 2018), describir proteínas específicas (Treccani *et al.*, 2006; Mann *et al.*, 2007; Suzuki *et al.*, 2009; Bahn *et al.*, 2015; Jin *et al.*, 2019; Rivera-Perez, *et al.*, 2019a; Rivera-Perez *et al.*, 2019b; Arroyo-Loranca *et al.*, 2020b), o estudiar relaciones filogenéticas entre la misma SMPs de diferentes especies de moluscos (Suzuki *et al.*, 2013), esta es la primera investigación que aborda SMPs de gasterópodos y bivalvos disponibles en repositorios públicos y busca identificar tanto por presencia/ausencia como por relaciones de parentesco entre las SMPs descritas hasta la fecha, una posible batería responsable de la biomineralización.

8.2.2 Identificación de proteínas comunes entre bivalvos y gasterópodos

Aunque existe un alto volumen de información disponible sobre SMPs y biomineralización en moluscos, y muchas proteínas y/o dominios reaparecen en proteomas de diferentes especies, no se ha identificado o estudiado, hasta la fecha una batería proteica básica responsable de la formación de la concha. Este estudio, por primera vez, busca arrojar luz en el asunto. A pesar del esfuerzo realizado, al depurar y analizar el grosor de información disponible, no se logró identificar una batería básica para gasterópodos y se recomienda ampliar los parámetros de búsqueda de SMPs para esta clase, en aras de identificar una batería básica en la misma. Por otro lado, se determinó que una posible batería básica encargada de la biomineralización de los bivalvos podría estar integrada por proteínas que contengan: LCR (*Low complexity regions*) y TR (*transmembrane regions*), además de dominios como: C1q, anhidrasa carbónica (*carbonic anhydrase*), dominio de unión a quitina (ChtBD2), Efh, EGF (*epidermal growth factor*), Kunitz y vWA, además del motivo *coiled coil region*. Aunque no se identificaron dominios representativos para las SMPs que formen parte de la batería básica para gasterópodos, se observó que las regiones LCR y TR de las SMPs mostraban cierta relevancia entre las proteínas de la base de datos, al igual que en bivalvos. Lo que resalta la importancia de estos, de relevancia a tomarse en cuenta en estudios posteriores.

En la literatura científica, solo un estudio realizado por (Arivalagan *et al.*, 2016) describe una posible batería básica para bivalvos. En este estudio fueron estudiadas cuatro especies de bivalvos (*C. gigas*, *M. edulis*, *M. truncata* y *Pecten maximus*) y se definió como posibles integrantes de una batería básica de SMPs que contienen los dominios: tirosinasa, anhidrasa carbónica, ChtBD2 y vWA. Si comparamos dicho resultado con el presente análisis, al menos tres de estos dominios se ratifican como posibles dominios contenidos en las SMPs que conforman la batería básica de bivalvos: vWA, ChtBD2 y anhidrasa carbónica, además, (Arivalagan *et al.*, 2016), también reportaron que muchas proteínas presentan LCRs; y dos de estos dominios se mostraron en gasterópodos, aunque no con la suficiente repetitividad como para integrar una posible batería básica.

Por otro lado, Liu *et al.* (2021) después de un análisis de proteomas de la concha de moluscos divide los dominios/regiones de SMPS en siete grupos principales, dentro de los

cuales se encuentran los dominios/regiones aquí identificados: 1) secuencias con LCR; 2) dominios extracelulares (vWA, TR, etc.); 3) dominios inmunológicos (EF-hand, C1q); 4) dominios de interacción a polisacáridos (ChtBD2); 5) enzimas (anhidrasa carbónica); 6) inhibidores de proteinasas (KU); 7) otros dominios. Esta es solo una de las clasificaciones existentes, que intentan describir y profundizar sobre el mecanismo de biomineralización de los moluscos, no obstante, engloba el rango de funciones necesarias para la producción del biomineral (ensamblaje, crecimiento, defensa y reparación de la concha). Como se observa, estos grupos de dominios y/o regiones están presentes en la batería básica de bivalvos propuesta.

Además, estos dominios han sido descritos en especies de otras clases, como es el caso de *Sepia officinales*. En este cefalópodo, se detectó la presencia de ChtBD2, vWA y Kunitz, lo que sugiere la existencia de un mecanismo ancestral conservado involucrado en la formación de la concha de moluscos (Liu y Zhang, 2021). Estos autores destacan, que la proporción de información disponible para esta clase taxonómica es minúscula comparada con las muy estudiadas clases Bivalvia y Gastropoda, y aun así se observaron similitudes entre las mismas.

Después de relacionar a través de relaciones de presencia/ausencia de los dominios y/o regiones detectadas en SMPs de gasterópodos y bivalvos, fue posible hipotetizar al menos una batería básica para bivalvos. Esto no coincide con lo planteado por Suzuki *et al.* (2013) que plantea que las proteínas con funciones fisiológicas particulares, como enzimas, hormonas, anticuerpos, etc., están altamente conservadas entre las especies de moluscos, sin embargo, las SMPs de al menos la capa nacarada de gasterópodos y bivalvos no poseen dominios comunes, y no hay prácticamente similitudes entre sus secuencias.

8.3 Caracterización de la batería básica de proteínas que interviene en la formación de la concha de moluscos, gasterópodos y bivalvos y relación con su microestructura

Por otro lado, los dominios identificados como parte de la batería básica de bivalvos no parecen tener un efecto en la microestructura de la concha, pues no se reporta en la literatura ningún efecto de este tipo atribuido a los mismos, a diferencia de las regiones de baja complejidad que sí parecen tener relevancia. De hecho, se ha sugerido que las LCR de

tipo repeticiones -GY y -GN de las nacreínas se unen débilmente al Ca^{2+} de la concha, mientras que los dominios de Pif se unen a minerales aragoníticos. Este tipo de dominios probablemente cumplan funciones mayores ya seas en la formación de la concha, y/o impartiendo las propiedades físicas como resistencia a la fractura (Marin *et al.*, 2012). Además, la glicoproteína MPP1 de *Crassostrea nipona* se conoce induce la formación de cristales tabulares (capa prismática [calcita]). Esta proteína forma un polímero multimérico a través de enlaces disulfuro, la repetitividad de la molécula en el polímero, unido a la estructura intrínseca de regiones LCR repetidas con residuos de Asp fosforilados que interactúan con el Ca^{2+} inducen la formación de microestructuras específicas.

Un estudio en *P. fucata* (Liu *et al.*, 2015) sugirió que la batería básica responsable de la biomineralización de las diferentes capas calcáreas en la concha de moluscos, dígase prisma y nácar, puede ser diferente. Aunque este no es el objetivo de este estudio, y no se realizaron distinciones entre las capas de procedencia de extracción de las proteínas en el momento de comparación de presencia/ausencia de sus dominios, sería relevante comparar este aspecto en un futuro.

Considerando la cantidad total de SMPs disponibles hasta la fecha, y la cantidad aisladas de proteomas individuales, es de esperarse que las proteínas sean exclusivas de ciertas capas calcáreas y que por tanto controlen las microestructuras específicas de las mismas (Liao *et al.*, 2015), a pesar de la creciente disponibilidad de información, ningún estudio ha identificado proteínas relacionadas a la microestructura.

Es importante destacar, que la matriz orgánica no es única y exclusivamente responsable de la microestructura de la concha, sino que existen otros fenómenos que influyen en el polimorfismo estructural. Se conoce que algunas microestructuras están favorecidas grandemente por algunos de estos factores físicos, mientras que otras están favorecidas por factores biológicos como las proteínas de la concha (Checa, 2018). Estos factores físico-biológicos pueden abarcar la competencia de cristales y la auto-organización de las matrices orgánicas (Checa, 2018), pH (Chen *et al.*, 2019) o procesos biológicos como la deposición de materia orgánica y mineral, la interacción de las proteínas sobre los cristales, o el reconocimiento de contacto subcelular y secreción diferencial, o una combinación de todos. (Checa, 2018). Los aportes de este trabajo solo se centran en las proteínas que componen la

matriz orgánica de la concha, pero comprender al 100% la biomineralización de los moluscos, implica analizar conjuntamente estos factores físico-biológicos.

Aunque aquí no se describe, ninguna relación aparente de nuestra propuesta de batería básica para bivalvos con la microestructura de la concha, atendiendo a la información disponible, la aparente simplicidad y repetitividad de las regiones de baja complejidad, así como su presencia en proteínas de la matriz soluble (responsable de la nucleación), pareciera involucrados en los patrones de biominerales a nivel microestructural.

En el futuro, sería prudente estudiar el efecto de proteínas con LCR asociadas a microestructuras de la concha. Partiendo de un conjunto de especies con variabilidad de microestructuras y cierta distancia filogenética, utilizando mecanismos de extracción similares, podrían realizarse procedimientos de extracción y caracterización de proteomas previos a ensayos de calcificación para estudiar los efectos de las SMPs en la microestructura.

Las regiones de baja complejidad se destacaron en este estudio, no únicamente por estar presente en ambas clases taxonómicas estudiadas, sino por presentarse en abundancia dentro de los bivalvos. Estas regiones se encuentran en SMPs que usualmente carecen de otros dominios conocidos y/o caracterizados (Liao *et al.*, 2019). Aunque, pueden encontrarse en proteínas combinadas con otros dominios como ChtBD2, vWA, etc.

Según Arivalagan *et al.* (2016) la presencia de LCR en SMPs confiere alta flexibilidad a las proteínas, facilitando su transporte a través de las membranas (las proteínas con conformaciones más flexibles pasan más fácilmente que las proteínas plegadas). Debido a que las proteínas de la matriz son sintetizadas en el epitelio calcificante y posteriormente trasladadas hasta el sitio de mineralización, la presencia de estos dominios podría ayudar en el proceso. La plasticidad intrínseca conferida por estos dominios podría conferir además a las SMPs la propiedad de reconocer varios blancos biológicos, sin sacrificar su especificidad. LCRs tienen además a evolucionar a gran velocidad vía eventos de recombinación, y pueden constituir una respuesta de adaptación al ambiente (Arivalagan *et al.*, 2016).

Según Liu *et al.* (2015), las LCR son una de las características de las proteínas con desorden intrínseco, las cuales son importantes para el ensamblaje de la matriz orgánica. Estas

aportan flexibilidad y diferentes funciones al estar repetidos con cierta regularidad a través de las proteínas, y facilitar la correcta interacción a nivel de posicionamiento estructural ya sea con otras proteínas o con el mismo CaCO_3 . Por ejemplo, la proteína MSP-1 de *Petinopecten yessoensis*, posee LCR repetidas como parte de una estructura modular altamente conservada (probablemente plegada) y repetida en cuatro tándems. Esta estructura parece permitir que estas moléculas se extiendan para cubrir e interactuar con el cristal de CaCO_3 ejecutando funciones de nucleación y/o inhibición en baja proporción y con alta efectividad (Sarashina y Endo, 2001). En comparación, la glicoproteína MPP1 de *Crassostrea nipona* también posee una estructura que alterna dominios internos repetidos y LCR, con dominios -SG y dominios DE (ricos en Asp y Glu interrumpidos por Cys) alternados. En esta proteína, el posicionamiento de los residuos Asp en dominios DE puede relacionarse con el posicionamiento del Ca^{2+} en el CaCO_3 de los cristales lacitos (Samata *et al.*, 2008).

Las LCR no se encuentran únicamente en moluscos, sino que se han reportado en braquiópodos (Isowa *et al.*, 2015) y en un amplio rango de proteínas no biomineralizantes de metazoos: proteínas de la seda, *mussel byssus* o el insecto corion. Sin embargo, pareciera que su abundancia no es exclusiva de todos los moluscos, pues en el cefalópodo *Sepa officinalis*, parece carecer de dominancia de este tipo de regiones y/o la abundancia de aminoácidos Asp y Glu que suelen repetirse en los mismos (en su lugar las proteínas fueron ricas en Leu/Ice) (le Pabic *et al.*, 2017), sugiriendo comportamientos específicos a sus líneas evolutivas y funciones.

Las LCR son una característica predominante de los proteomas de moluscos estudiados hasta la fecha. Por ejemplo, en 13/39 proteínas del proteoma de *P. fucata* fueron detectados este tipo de regiones (Marie *et al.*, 2013). Las LCR detectados hasta la fecha parecen ser linaje-específicas, sugiriendo que han evolucionado independientemente en los diferentes modelos de moluscos. Por otro lado, se ha descrito, que los secretomas de cada especie de moluscos son único y resultante de genes evolucionados y nuevos productos de genes para crear una barrera específica a su ambiente (Evans, 2019). Este combinado a LCR específicos sugieren que la biomineralización, como proceso biológico en invertebrados posee un mecanismo ancestral con cambios linajes específicos y evolución de genes (Liu y Zhang, 2021a).

9. CONCLUSIONES

- Se construyeron dos bases de datos constituidas por 301 secuencias proteicas de 32 especies de gasterópodos y 1,118 secuencias de 37 especies de bivalvos. Las proteínas se caracterizaron por mostrar puntos isoeléctricos variables, péptidos señales, diferentes glicosilaciones, puentes disulfuro y por localizarse posiblemente como proteínas extracelulares secretadas. Se detectaron 56 dominios conservados en gasterópodos y 163 en bivalvos, pertenecientes a 42 y 99 familias respectivamente.
- No fue posible identificar una batería básica de proteínas involucradas en la formación de la concha de moluscos gasterópodos. Las regiones LCR y TR parecen relevante y deberían estudiarse a profundidad.
- Se hipotetizó regiones/dominios que pudieran ser parte de las SMPs de la batería básica para bivalvos integrada por los dominios: C1q, anhidrasa carbónica, unión a carbohidratos (ChtBD2), Efh, factor de crecimiento epidérmico (EGF,) Kunitz (KU) y Von Willebrand Factor A (vWA), el motivo *Coiled coil region* y las regiones de baja complejidad (LCR) y transmembranales (TR).
- Las TR y LCR fueron relevantes en las SMPs de gasterópodos y bivalvos.
- No se determinó relación entre los dominios/regiones de las SMPs que pudieran formar parte de la batería básica de bivalvos con microestructura de las capas calcáreas de la concha. La inexistencia de una batería de gasterópodos impidió evaluar esta relación.

10. LITERATURA CITADA

Addadi, L., Joester, D., Nudelman, F. y Weiner, S. (2006). Mollusk shell formation: A source of new concepts for understanding biomineralization processes. *Chemistry - A European Journal*. John Wiley & Sons, Ltd, 12(4), pp. 980–987. doi:10.1002/chem.200500980.

Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., von Heijne, G. y Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology*, 37(4). doi:10.1038/s41587-019-0036-z.

Arivalagan, J., Yarra, T., Marie, B., Sleight, V. A., Duvernois-Berthet, E., Clark, M. S., Marie, A. y Berland, S. (2016). Insights from the shell proteome: biomineralization to adaptation Downloaded from. *Molecular Biology and Evolution*, 34(1), pp. 66–77. Available at: <http://mbe.oxfordjournals.org/>.

Arroyo-Loranca, R. G., Hernandez-Saavedra, N. Y., Hernandez-Adame, L. y Rivera-Perez, C. (2020a). Ps19, a novel chitin binding protein from *Pteria sterna* capable to mineralize aragonite plates in vitro. *PLoS ONE*. Public Library of Science, 15(3). doi:10.1371/journal.pone.0230431.

Bahn, S. Y., Jo, B. H., Hwang, B. H., Choi, Y. S. y Cha, H. J. (2015). Role of Pif97 in Nacre Biomineralization: In Vitro Characterization of Recombinant Pif97 as a Framework Protein for the Association of Organic-Inorganic Layers in Nacre. *Crystal Growth and Design*. American Chemical Society, 15(8), pp. 3666–3673. doi:10.1021/acs.cgd.5b00275.

Bédouet, L., Duplat, D., Marie, A., Dubost, L., Berland, S., Rousseau, M., Milet, C. y Lopez, E. (2007a). Heterogeneity of proteinase inhibitors in the water-soluble organic matrix from the oyster nacre. *Marine Biotechnology*, 9(4), pp. 437–449. doi:10.1007/s10126-007-7120-y.

Bédouet, L., Marie, A., Dubost, L., Péduzzi, J., Duplat, D., Berland, S., Puisségur, M., Boulzaguet, H., Rousseau, M., Milet, C. y Lopez, E. (2007b). Proteomics analysis of the nacre soluble and insoluble proteins from the oyster *Pinctada margaritifera*. *Marine Biotechnology*, 9(5), pp. 638–649. doi:10.1007/s10126-007-9017-1.

Berland, S., Marie, A., Duplat, D., Milet, C., Sire, J. Y. y Bédouet, L. (2011). Coupling Proteomics and Transcriptomics for the Identification of Novel and Variant Forms of Mollusk Shell Proteins: A Study with *P. margaritifera*. *ChemBioChem*, 12(6), pp. 950–961. doi:10.1002/cbic.201000667.

Brusca, R. C. y Brusca, G. J. (2003). *Invertebrates*. 2nd ed. Sunderland, M.A.: Sinauer Associates.

Checa, A. G. (2018). Physical and biological determinants of the fabrication of Mollusk shell microstructures. *Frontiers in Marine Science*. Frontiers Media S.A. doi:10.3389/fmars.2018.00353.

Chen, Y., Feng, Y., Deveaux, J. G., Masoud, M. A., Chandra, F. S., Chen, H., Zhang, D. y Feng, L. (2019). Biomineralization forming process and bio-inspired nanomaterials for biomedical application: A review. *Minerals*. MDPI AG. doi:10.3390/min9020068.

Cölfen, H. y Antonietti, M. (2005). Mesocrystals: Inorganic Superstructures Made by Highly Parallel Crystallization and Controlled Alignment. *Angewandte Chemie International Edition*, 44(35), pp. 5576–5591. doi:10.1002/anie.200500496.

Estroff, L. A. (2008). Introduction: Biomineralization. *Chemical Reviews*. American Chemical Society, pp. 4329–4331. doi:10.1021/cr8004789.

Evans, J. S. (2019). The Biomineralization Proteome: Protein Complexity for a Complex Bioceramic Assembly Process. *Proteomics*. Wiley-VCH Verlag. doi:10.1002/pmic.201900036.

Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*. Wiley, 39(4), pp. 783–791. doi:10.1111/j.1558-5646.1985.tb00420.x.

Feng, D., Li, Q., Yu, H., Kong, L. y Du, S. (2017). Identification of conserved proteins from diverse shell matrix proteome in *Crassostrea gigas*: Characterization of genetic bases regulating shell formation. *Scientific Reports*. Nature Publishing Group, 7. doi:10.1038/srep45754.

Ferre, F. y Clote, P. (2005). DiANNA: a web server for disulfide connectivity prediction. *Nucleic Acids Research*, 33(Web Server). doi:10.1093/nar/gki412.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. y Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server, in Walker, J. M. (ed.) *The Proteomics Protocols Handbook*. 1st ed. Humana Press, pp. 571–607.

Gu, Z., Gu, L., Eils, R., Schlesner, M. y Brors, B. (2014). Circlize implements and enhances circular visualization in R. *Bioinformatics*. Oxford University Press, 30(19), pp. 2811–2812. doi:10.1093/bioinformatics/btu393.

Gupta, R. y Brunak, S. (2002). Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac Symp Biocomput.*, pp. 310–322.

Heberle, H., Meirelles, G. V., da Silva, F. R., Telles, G. P. y Minghim, R. (2015). InteractiVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics*, 16(1), p. 169. doi:10.1186/s12859-015-0611-3.

Horton, P., Park, K.-J., Obayashi, T., Fujita, N., Harada, H., Adams-Collier, C. J. y Nakai, K. (2007). WoLF PSORT: protein localization predictor. *Nucleic Acids Research*, 35(Web Server). doi:10.1093/nar/gkm259.

Hulsen, T., de Vlieg, J. y Alkema, W. (2008). BioVenn – a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genomics*, 9(1), p. 488. doi:10.1186/1471-2164-9-488.

Isowa, Y., Sarashina, I., Oshima, K., Kito, K., Hattori, M. y Endo, K. (2015). Proteome analysis of shell matrix proteins in the brachiopod *Laqueus rubellus*. *Proteome Science*. BioMed Central Ltd., 13(1). doi:10.1186/s12953-015-0077-2.

Jin, C., Zhao, J., Pu, J., Liu, X. y Li, J. (2019). Hichin, a chitin binding protein is essential for the self-assembly of organic frameworks and calcium carbonate during shell formation. *International Journal of Biological Macromolecules*. Elsevier B.V., 135, pp. 745–751. doi:10.1016/j.ijbiomac.2019.05.205.

Joubert, C., Piquemal, D., Marie, B., Manchon, L., Pierrat, F., Zanella-Cléon, I., Cochennec-Laureau, N., Gueguen, Y. y Montagnani, C. (2010). Transcriptome and proteome analysis of *Pinctada margaritifera* calcifying mantle and shell: Focus on biomineralization. *BMC Genomics*. BioMed Central Ltd., 11(1). doi:10.1186/1471-2164-11-613.

Karakostis, K., Zanella-Cléon, I., Immel, F., Guichard, N., Dru, P., Lepage, T., Plasseraud, L., Matranga, V. y Marin, F. (2016). A minimal molecular toolkit for mineral deposition? Biochemistry and proteomics of the test matrix of adult specimens of the sea urchin *Paracentrotus lividus*. *Journal of Proteomics*. Elsevier B.V., 136, pp. 133–144. doi:10.1016/j.jprot.2016.01.001.

Kumar, S., Stecher, G. y Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0. *Molecular Biology and Evolution* (submitted). *Mol. Biol. Evol.*, 33(7), pp. 1870–1874.

Liao, Z., Bao, L. Fei, Fan, M. Hua, Gao, P., Wang, X. Xing, Qin, C. Li y Li, X. min (2015). In-depth proteomic analysis of nacre, prism, and myostracum of *Mytilus* shell. *Journal of Proteomics*. Elsevier, 122, pp. 26–40. doi:10.1016/j.jprot.2015.03.027.

Liao, Z., Jiang, Y. Ting, Sun, Q., Fan, M. Hua, Wang, J. Xin y Liang, H. Ying (2019). Microstructure and in-depth proteomic analysis of *Perna viridis* shell. *PLoS ONE*. Public Library of Science, 14(7). doi:10.1371/journal.pone.0219699.

Lin, G., Chai, J., Yuan, S., Mai, C., Cai, L., Murphy, R. W., Zhou, W. y Luo, J. (2016). Vennpainter: A tool for the comparison and identification of candidate genes based on Venn diagrams. *PLoS ONE*. Public Library of Science, 11(4). doi:10.1371/journal.pone.0154315.

Liu, C., Li, S., Kong, J., Liu, Y., Wang, T., Xie, L. y Zhang, R. (2015). In-depth proteomic analysis of shell matrix proteins of *Pinctada fucata*. *Scientific Reports*. Nature Publishing Group, 5. doi:10.1038/srep17269.

Liu, C. y Zhang, R. (2021a). Biomineral proteomics: A tool for multiple disciplinary studies. *Journal of Proteomics*. Elsevier B.V. doi:10.1016/j.jprot.2021.104171.

Liu, C. y Zhang, R. (2021b). Biomineral proteomics: A tool for multiple disciplinary studies. *Journal of Proteomics*. Elsevier B.V. doi:10.1016/j.jprot.2021.104171.

Mann, K., Cerveau, N., Gummich, M., Fritz, M., Mann, M. y Jackson, D. J. (2018). In-depth proteomic analyses of *Haliotis laevis* (greenlip abalone) nacre and prismatic organic shell matrix. *Proteome Science*, 16(1). doi:10.1186/s12953-018-0139-3.

Mann, K. y Edsinger, E. (2014). The *Lottia gigantea* shell matrix proteome: re-analysis including MaxQuant iBAQ quantitation and phosphoproteome analysis. *Proteome Science*. Available at: <http://www.cbs.dtu.dk/services/SignalP/>.

Mann, K. y Jackson, D. J. (2014). Characterization of the pigmented shell-forming proteome of the common grove snail *Cepaea nemoralis*. *BMC Genomics*. BioMed Central Ltd., 15(1). doi:10.1186/1471-2164-15-249.

Mann, K., Siedler, F., Treccani, L., Heinemann, F. y Fritz, M. (2007). Perlinhibin, a cysteine-, histidine-, and arginine-rich miniprotein from abalone (*Haliotis laevis*) nacre, inhibits in vitro calcium carbonate crystallization. *Biophysical Journal*. Biophysical Society, 93(4), pp. 1246–1254. doi:10.1529/biophysj.106.100636.

Marie, B., Jackson, D. J., Ramos-Silva, P., Zanella-Cléon, I., Guichard, N. y Marin, F. (2013). The shell-forming proteome of *Lottia gigantea* reveals both deep conservations and lineage-specific novelties. *FEBS Journal*, 280(1), pp. 214–232. doi:10.1111/febs.12062.

Marie, B., Marie, A., Jackson, D. J., Dubost, L., Degnan, B. M., Milet, C. y Marin, F. (2010). Proteomic analysis of the organic matrix of the abalone *Haliotis asinina* calcified shell. *Proteome Science*, 8. doi:10.1186/1477-5956-8-54.

Marin, F., Luquet, G., Marie, B. y Medakovic, D. (2007). Molluscan Shell Proteins: Primary Structure, Origin, and Evolution. *Current Topics in Developmental Biology*, pp. 209–276. doi:10.1016/S0070-2153(07)80006-8.

Marin, F., le Roy, N. y Marie, B. (2012). The formation and mineralization of mollusk shell. *Frontiers in Bioscience*, 4, pp. 1099–1125. Available at: https://www.researchgate.net/publication/51925970_The_formation_and_mineralization_of_mollusk_shell (Accessed: March 11, 2020).

Mason, J. M. and Arndt, K. M. (2004). Coiled coil domains: Stability, specificity, and biological implications, in *ChemBioChem*, pp. 170–176. doi:10.1002/cbic.200300781.

Miyamoto, H., Endo, H., Hashimoto, N., Imura, K., Isowa, Y., Kinoshita, S., Kotaki, T., Masaoka, T., Miki, T., Nakayama, S., Nogawa, C., Notazawa, A., Ohmori, F., Sarashina, I., Suzuki, M., Takagi, R., Takahashi, J., Takeuchi, T., Yokoo, N., Satoh, N., Toyohara, H., Miyashita, T., Wada, H., Samata, T., Endo, K., Nagasawa, H., Asakawa, S. y Watabe, S. (2013). The Diversity of Shell Matrix Proteins: Genome-Wide Investigation of the Pearl Oyster, *Pinctada fucata*. *Zoological Science*. Zoological Society of Japan, 30(10), p. 801. doi:10.2108/zsj.30.801.

Miyamoto, H., Miyashita, T., Okushimat, M., Nakanoi, S., Morita, T. y Matsushiro, A. (1996). A carbonic anhydrase from the nacreous layer in oyster pearls. *Genetics Communicated by Igor Dawid*.

Oliveros, J. (2007) *Venny. An interactive tool for comparing lists with Venn Diagrams*, <https://bioinfoq.cnb.csic.es/tools/venny/index.html>. le Pabic, C., Marie, A., Marie, B., Percot, A., Bonnaud-Ponticelli, L., Lopez, P. J. y Luquet, G. (2017). First proteomic analyses of the dorsal and ventral parts of the *Sepia officinalis* cuttlebone. *Journal of Proteomics*. Elsevier B.V., 150, pp. 63–73. doi:10.1016/j.jprot.2016.08.015.

Pirooznia, M., Nagarajan, V. y Deng, Y. (2007). GeneVenn-A web application for comparing gene lists using Venn diagrams. *Bioinformatics*, 1(10), p. 420.

Rivera Pérez, C. y Hernandez-Saavedra, N. Y. H. (2021). Review: Post-translational modifications of marine shell matrix proteins. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, 256.

Rivera-Perez, C., Magallanes-Dominguez, C., Dominguez-Beltran, R. V., de Areyano, J. J. O. R. y Hernandez-Saavedra, N. Y. (2019a). Biochemical and molecular characterization of N66 from the shell of *Pinctada mazatlanica*. *PeerJ*. PeerJ Inc., 2019(6). doi:10.7717/peerj.7212.

Rivera-Perez, C., Ojeda-Ramirez de Areyano, J. J. y Hernandez-Saavedra, N. Y. (2019b). Purification and functional analysis of the shell matrix protein N66 from the shell of the pearl oyster *Pteria sterna*. *Comparative Biochemistry and Physiology Part - B: Biochemistry and Molecular Biology*. Elsevier Inc., 235, pp. 19–29. doi:10.1016/j.cbpb.2019.05.007.

Saitou, N. y Nei, M. (1987). The Neighbor-joining Method: A New Method for Reconstructing Phylogenetic Trees'. *Molecular Biology and Evolution*, 4, pp. 406–425. Available at: <https://academic.oup.com/mbe/article/4/4/406/1029664>.

Samata, T., Ikeda, D., Kajikawa, A., Hideyoshi, S., Nogawa, C., Yamada, D., Yamazaki, R. y Akiyama, T. (2008). A novel phosphorylated glycoprotein in the shell matrix of the oyster *Crassostrea nippona*. *The FEBS Journal*, 275, pp. 2977–2989.

Sarashina, I. y Endo, K. (2001). The complete primary structure of molluscan shell protein 1 (MSP-1), an acidic glycoprotein in the shell matrix of the scallop *Patinopecten yessoensis*. *Marine Biotechnology*, 3(4), pp. 362–369. doi:10.1007/s10126-001-0013-6.

Shen, X., Belcher, A. M., Hansma, P. K., Stucky, G. D. y Morse, D. E. (1997). Molecular cloning and characterization of Lustrin A, a matrix protein from shell and pearl nacre of *Haliotis rufescens*. *Journal of Biological Chemistry*, 272(51), pp. 32472–32481. doi:10.1074/jbc.272.51.32472.

Song, X., Wang, X., Li, L. y Zhang, G. (2014). Identification two novel nacrein-like proteins involved in the shell formation of the Pacific oyster *Crassostrea gigas*. *Molecular Biology Reports*. Kluwer Academic Publishers, 41(7), pp. 4273–4278. doi:10.1007/s11033-014-3298-z.

Sonnhammer, E. L. L., Eddy, S. R. y Durbin, R. (1997). Pfam: A comprehensive database of protein domain families based on seed alignments. *Proteins: Structure, Function, and Genetics*, 28(3). doi:10.1002/(SICI)1097-0134(199707)28:3< 405:AID-PROT10>3.0.CO;2-L.

Steentoft, C., Vakhrushev, S. Y., Joshi, H. J., Kong, Y., Vester-Christensen, M. B., Schjoldager, K. T.-B. G., Lavrsen, K., Dabelsteen, S., Pedersen, N. B., Marcos-Silva, L., Gupta, R., Paul Bennett, E., Mandel, U., Brunak, S., Wandall, H. H., Lavery, S. B. y Clausen, H. (2013). Precision mapping of the human O-GalNAc glycoproteome through SimpleCell technology. *The EMBO Journal*, 32(10). doi:10.1038/emboj.2013.79.

Suzuki, M., Iwashima, A., Kimura, M., Kogure, T. y Nagasawa, H. (2013). The Molecular Evolution of the Pif Family Proteins in Various Species of Mollusks. *Marine Biotechnology*, 15(2), pp. 145–158. doi:10.1007/s10126-012-9471-2.

Suzuki M, Iwashima A, Tsutsui N, Ohira T, Kogure T, Nagasawa H. (2011). Identification and characterisation of a calcium carbonate-binding protein, blue mussel shell protein (BMSP), from the nacreous layer. *Chembiochem: a European Journal of Chemical Biology*. 12(16): 2478-2487. DOI: 10.1002/cbic.201100317. PMID: 21932217.

Suzuki, M., Saruwatari, K., Kogure, T., Yamamoto, Y., Nishimura, T., Kato, T. y Nagasawa, H. (2009). An acidic matrix protein, Pif, is a key macromolecule for nacre formation. *Science*, 325(5946), pp. 1388–1390. doi:10.1126/science.1173793.

Tao, J. (2013). FTIR and Raman studies of structure and bonding in mineral and organic-mineral composites, in *Methods in Enzymology*. Academic Press Inc., pp. 533–556. doi:10.1016/B978-0-12-416617-2.00022-9.

Treccani, L., Mann, K., Heinemann, F. y Fritz, M. (2006). Perlwapin, an abalone nacre protein with three four-disulfide core (whey acidic protein) domains, inhibits the growth of calcium carbonate crystals. *Biophysical Journal*. Biophysical Society, 91(7), pp. 2601–2608. doi:10.1529/biophysj.106.086108.

Truebestein, L. y Leonard, T. A. (2016). Coiled-coils: The long and short of it. *BioEssays*. John Wiley and Sons Inc., 38(9), pp. 903–916. doi:10.1002/bies.201600062.

Wang, X., Song, X., WANG, T., Zhu, Q., Miao, G., Chen, Y., Fang, X., Que, H., Li, L. y Zhang, G. (2013). Evolution and functional analysis of the Pif97 gene of the pacific oyster *Crassostrea gigas*. *Current Zoology*. Chinese Academy of Sciences, 59(1), pp. 109–115. doi:10.1093/czoolo/59.1.109.

Watabe, N. y Wilbur, K. M. (1960). Influence of the Organic Matrix on Crystal Type in Molluscs. *Science*, 133, p. 1960.

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), pp. 1189–1191.

Werner, G. D. A., Gemmell, P., Grosser, S., Hamer, R. y Shimeld, S. M. (2013). Analysis of a deep transcriptome from the mantle tissue of *Patella vulgata* Linnaeus (Mollusca: Gastropoda: Patellidae) reveals candidate biomineralising genes. *Marine Biotechnology*. Springer Science and Business Media, LLC, 15(2), pp. 230–243. doi:10.1007/s10126-012-9481-0.

Zuckerlandl, E. y Pauling, L. (1965) . Evolutionary Divergence and Convergence, in Proteins, in Bryson V.; and Vogel, H. J. (eds.). New York: Academic Press, pp. 97–166.

11. ANEXOS

Anexo A: Tabla VI. SMPs de gasterópodos y bivalvos, cuyas secuencias no se encuentran disponibles en el momento de la creación de las bases de datos y por tanto no se incluyeron en los análisis bioinformáticos.

Espece	Proteína	Tejido	Observaciones	Referencia
GASTROPODA				
<i>H. rufescens</i>	AP8	Concha	Rica en NG. Procedente del nácar.	Fu <i>et al.</i> (2005)
<i>H. asinina</i>	Chitin-binding protein	Concha		Marie <i>et al.</i> (2010)
<i>H. rufescens</i>	Conchiolin	Concha	Rica en NG. Procedente de nácar y prisma.	Cariolou y Morse, (1988)
<i>H. levigata</i>	Perlinhibin	Concha	Rica en CHR. Procedente del nacar.	Mann <i>et al.</i> (2007)
<i>H. laevigata</i>	Perlwapin	Concha	PGGYPR. Procedente del nácar.	Treccani <i>et al.</i> (2006)
<i>Strombus decorus persicus</i>	ACLS40	Concha	Toda la concha.	Pokroy <i>et al.</i> (2006)
<i>L. gigantea</i>	BMSP	Concha		Mann <i>et al.</i> (2012)
<i>L. gigantea</i>	dermatopotin	Concha	a-CA	Mann <i>et al.</i> (2012)
<i>L. gigantea</i>	nacrein	Concha		Mann <i>et al.</i> (2012)
<i>L. gigantea</i>	perlustrin	Concha		Mann <i>et al.</i> (2012)
<i>L. gigantea</i>	perlucin	Concha		Mann <i>et al.</i> (2012)
<i>L. gigantea</i>	Pif	Concha		Mann <i>et al.</i> (2012)
BIVALVIA				
<i>P. margaritifera</i>	Calcine			Cochennec-Laureau <i>et al.</i> (2005)
<i>P. fucata</i>	Carbonic Anhydrase precursor parcial	Manto		Li <i>et al.</i> (2016)
<i>P. margaritifera</i>	MSI60	Manto	Procedente del nácar.	Sudo <i>et al.</i> (1997); Zhang <i>et al.</i> (2006); Marie <i>et al.</i> (2012); Joubert <i>et al.</i> (2014); Li <i>et al.</i> (2017)
<i>P. margaritifera</i> / <i>P. maxima</i>	MRNP34	Manto	Procedente del nácar.	Marie <i>et al.</i> (2012)
<i>P. fucata</i>	N44	ND	Procedente del nácar.	Yan <i>et al.</i> (2007)
<i>P. margaritifera</i>	N44	Manto		McGinty <i>et al.</i> (2012)

Especie	Proteína	Tejido	Observaciones	Referencia
<i>ra/ P. maxima</i>				
<i>P. fucata</i>	P60 partial	Concha	Procedente del nácar.	Lao <i>et al.</i> (2007)
<i>Pinna nobilis</i>	Calprismín partial	Concha	Procedente del prisma.	Marin <i>et al.</i> (2005)
<i>P. nobilis</i>	Caspartin partial	Concha	Procedente del prisma.	Marin <i>et al.</i> , (2005)
<i>M. edulis</i>	ND	ND		Yin <i>et al.</i> (2005)
<i>M. coruscus</i>	fp-2	Manto		Inoue <i>et al.</i> (2000)
<i>M. galloprovin cialis</i>	Mgal-perlwapin	Manto		Marie <i>et al.</i> (2011)
<i>Adamussiu m sp.</i>	RP-1			Halloran y Donachy, (1995)